Хмельницький національний університет
Факультет інформаційних технологій
Кафедра кібербезпеки

КВАЛІФІКАЦІЙНА РОБОТА

Білецького Костянтина Борисовича

на здобуття ступеня вищої освіти Магістра

Метод виявлення аномалій мережевого трафіку

Галузь знань 12 – Інформаційні технології
Спеціальність 125 – Кібербезпека та захист інформації
Освітня програма Кібербезпека та захист інформації

Шифр КРМКБЗІ.2301132.23.01.02 ПЗ

Виконав студент 2 курсу група КБЗІм-23-1 Костянтин БІЛЕЦЬКИЙ
Керівник докт. техн. наук, професор Михайло КАСЯНЧУК
Нормоконтролер старший викладач Сергій МОСТОВИЙ

До захисту допускаю:
Завідувач кафедри кібербезпеки Юрій КЛЬОЦ

16 12 2024 р.

Хмельницький 2024

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

Факультет Інформаційних технологій

Кафедра Кібербезпеки

Рівень вищої освіти Магістр

Галузь знань 12 – Інформаційні технології

Спеціальність 125 – Кібербезпека та захист інформації

Освітня програма Кібербезпека та захист інформації

ЗАТВЕРДЖУЮ

Завідувач кафедри кібербезпеки

Юрій КЛЬОЦ

2 09 2024 р.

З А В Д А Н Н Я
НА КВАЛІФІКАЦІЙНУ РОБОТУ

Білецькому Костянтину Борисовичу

1 Тема роботи Метод виявлення аномалій мережевого трафіку

Керівник роботи докт.техн.наук, професор Михайло КАСЯНЧУК

Затверджено наказом ректора університету від 26 08 2024 № 60

2 Строк подання студентом кваліфікаційної роботи на кафедру 2.12.2024

3 Вихідні дані до роботи Проаналізувати поняття аномалій мережевого трафіку та причини їх появи. Порівняти існуючі методи їх виявлення, проаналізувати переваги та недоліки. Розробити алгоритм виявлення аномалій у мережі та підібрати набір даних для тестування. Розробити метод виявлення аномалій мережевого трафіку. Розрахувати ефективність прототипу системи за допомогою зібраних наборів даних.

4 Зміст пояснювальної записки (перелік питань, які потрібно розробити)
Вступ. Аномалії мережевого трафіку та їх аналіз. Постановка задачі. Аналіз існуючих наборів даних та збір трафіку для власного набору даних. Алгоритм виявлення аномалій. Метод виявлення аномалій мережевого трафіку. Розробка прототипу системи. Розрахунок ефективності пропонованого методу.

5 Перелік графічного матеріалу (із зазначенням обов'язкових креслень)

6 Консультанти розділів кваліфікаційної роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв

7 Дата видачі завдання 2 09 2024 р.

КАЛЕНДАРНИЙ ПЛАН

Назва етапів (розділів) кваліфікаційної роботи	Строк виконання етапів роботи	Примітка
Дослідження проблеми та актуальності виявлення аномалій мережевого трафіку		Виконано
Визначення змісту та структури кваліфікаційної роботи		Виконано
Аналіз аномалій мережевого трафіку, їх огляд та методи виявлення		Виконано
Опрацювання статті за результатами дослідження		Виконано
Аналіз існуючих наборів даних та створення власного набору даних		Виконано
Розробка алгоритму та методу виявлення аномалій мережевого трафіку		Виконано
Підготовка тестового середовища та оцінка ефективності		Виконано
Підготовка та опрацювання ілюстративного матеріалу		Виконано
Оформлення текстової частини кваліфікаційної роботи		Виконано
Попередній захист кваліфікаційної роботи		Виконано
Захист кваліфікаційної роботи на засіданні ЕК		Виконано

Студент Костянтин БІЛЕЦЬКИЙ

Керівник кваліфікаційної роботи Михайло КАСЯНЧУК

АНОТАЦІЯ

Тема кваліфікаційної роботи: Метод виявлення аномалій мережевого трафіку

Автор роботи: студент групи КБЗІм-23-1 Білецький К.Б.

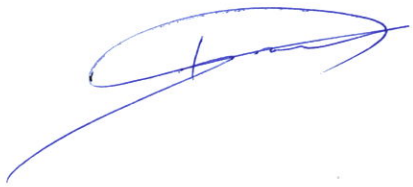
Керівник роботи: докт. техн. наук, професор Касянчук М.М.

Загальний обсяг роботи: 88 сторінок, 17 рисунків, 6 таблиць, 1 додаток, 62 посилання.

Ключові слова: аномалії мережевого трафіку, виявлення загроз, LOF, HBOS класифікація трафіку.

Дана кваліфікаційна робота присвячена розробці методу виявлення аномалій у мережевому трафіку з використанням методів машинного навчання для підвищення ефективності систем кібезахисту. Проведено аналіз теоретичних аспектів, що стосуються виявлення аномалій у мережевому трафіку. Розроблено метод виявлення аномалій, що використовує сучасні методи машинного навчання, де використано алгоритми LOF та HBOS, які забезпечують можливість аналізу трафіку на основі як навчання з позначеними даними, так і без них. Розроблений алгоритм використовує підхід, який дозволяє визначати як відомі аномалії, так і нові типи загроз. Проведено тестування розробленого алгоритму з метою оцінки його ефективності. Результати тестування показали, що метод забезпечує високий рівень точності, досягаючи понад 95% у визначенні аномалій, що значно перевищує показники традиційних методів.

2.12.2024



ANNOTATION

Theme of qualification work: Method of detecting network traffic anomalies

Author of the work: KBZIm-23-1 Biletskyi K.B.

Mentor: dr. technical Sciences, Professor Kasyanchuk M.M.

Total volume of work: 88 pages, 17 figures, 6 tables, 1 appendix, 62 references.

Keywords: network traffic, threat detection, LOF, HBOS traffic classification.

This qualification work is devoted to developing a method for detecting anomalies in network traffic using machine learning methods to increase the effectiveness of cyber protection systems. An analysis of theoretical aspects related to detecting anomalies in network traffic has been carried out. An anomaly detection method has been developed that uses modern machine learning methods, where LOF and HBOS algorithms are used, which provide the ability to analyze traffic based on both training with and without labelled data. The developed algorithm uses an approach that allows you to identify known anomalies and new threats. The developed algorithm was tested to evaluate its effectiveness. The test results showed that the method provides a high level of accuracy, reaching more than 95% in detecting anomalies, which is significantly higher than traditional methods.

2.12.2024

ЗМІСТ

ВСТУП

У сучасних умовах стрімкого розвитку інформаційних технологій питання забезпечення кібербезпеки мереж стає все більш актуальним. Зростання обсягів мережевого трафіку, пов’язане з активним використанням інтернету, а також постійне ускладнення кіберзагроз вимагають нових підходів до захисту інформації та інфраструктури. Аномалії в мережевому трафіку, такі як раптові зміни у характері передавання даних, незвичні обсяги трафіку або несанкціоновані спроби доступу до мережевих ресурсів, можуть свідчити про наявність загроз, як-от вірусні атаки, несанкціонований доступ або зловмисне сканування мереж. У зв’язку з цим, розробка ефективних методів для виявлення таких аномалій є важливим кроком для забезпечення своєчасного виявлення загроз, мінімізації збитків та підвищення загального рівня безпеки мережевих систем.

Метою даної роботи є розробка методу для ефективного виявлення аномалій у мережевому трафіку, який мінімізує кількість хибнопозитивних спрацьовувань та підвищує загальну точність виявлення загроз. Розроблений метод має бути здатний до адаптації у змінних умовах мережевого середовища, що дозволить йому швидко реагувати на нові загрози та відхилення у мережевій активності.

Для досягнення поставленої мети необхідно виконати наступні завдання:

– провести огляд та аналіз існуючих методів виявлення аномалій у мережевому трафіку, визначити їх переваги та недоліки;

– підібрати та дослідити різні набори даних, які дозволяють тестувати запропонований метод виявлення аномалій;

– розробити новий алгоритм, який базується на сучасних методах машинного навчання для виявлення аномалій у мережевому трафіку;

– здійснити тестування та оцінку ефективності розробленого алгоритму за ключовими показниками, такими як точність, швидкодія та кількість хибнопозитивних спрацювань.

Об’єктом дослідження є мережевий трафік, який може містити аномалії як індикатори потенційних загроз.

Предметом дослідження виступають методи і алгоритми виявлення аномалій у мережевому трафіку, їх точність, швидкодія та здатність до адаптації до умов мережевого середовища.

Наукова новизна роботи полягає у вдосконаленні існуючих підходів до аналізу мережевого трафіку шляхом розробки адаптивного методу, що здатний до швидкої ідентифікації аномалій за допомогою сучасних алгоритмів машинного навчання. На відміну від традиційних сигнатурних методів, запропонований підхід може виявляти раніше невідомі аномалії, знижуючи частоту хибних спрацювань.

Практичне значення розробленого методу полягає у можливості його застосування для реального моніторингу безпеки корпоративних мереж, державних установ або приватних організацій. Метод може бути впроваджений як частина комплексної системи кібербезпеки, що дозволить покращити швидкість і точність виявлення загроз, скоротити кількість хибнопозитивних сигналів та зменшити навантаження на адміністраторів системи.

Кваліфікаційна робота складається зі вступу, чотирьох розділів основного узмісту, висновків, списку використаних джерел та додатку. Перший розділ присвячений теоретичному огляду аномалій мережевого трафіку та методів їх виявлення. У другому розділі розглянуто існуючі набори даних, що використовуються для тестування систем виявлення аномалій, та обґрунтовано вибір набору даних для роботи. Третій розділ описує розроблений метод виявлення аномалій, що базується на алгоритмах машинного навчання, а також деталі побудови прототипу системи. Четвертий розділ присвячено тестуванню розробленого методу та аналізу його ефективності за допомогою обраних показників.

Апробація результатів та публікації представлено у Додатку А.

1 АНОМАЛІЇ МЕРЕЖЕВОГО ТРАФІКУ ТА ЇХ АНАЛІЗ

1.1 Аномалії мережевого трафіку

Аномалії мережевого трафіку – це будь-які незвичні зміни або відхилення в патернах даних, що переміщуються мережею. Вони проявляються як несподівані коливання обсягу трафіку, зміна напрямків передачі даних або незвичні комунікаційні шаблони. Ці аномалії, зазвичай, не відповідають типовому мережевому навантаженню і часто слугують індикатором потенційної загрози або проблеми в роботі мережі, таких як атаки типу відмова в обслуговуванні або експлуатація вразливостей. У більшості випадків мережевий трафік демонструє стабільні та передбачувані шаблони, що залежать від рутинної діяльності користувачів та серверів. Поява аномалій може сигналізувати про порушення у цих шаблонах, викликані діями кіберзлочинців або технічними несправностями, і тому є важливим показником для систем безпеки [1].

Аномалії мережевого трафіку можуть з’являтися через безліч факторів. Найчастіше вони є наслідком навмисних дій, таких як кібератаки, що створюють великий обсяг запитів для перевантаження сервера або перехоплення конфіденційних даних [2-3]. Наприклад, атаки на основі протоколу типу Distributed Denial of Service спрямовані на збільшення трафіку, щоб мережа перестала реагувати на легітимні запити [4-5]. До інших причин належать збої в роботі обладнання, помилки в програмному забезпеченні або навіть значні зміни у поведінці користувачів. Приріст трафіку може бути спричинений розповсюдженням нового програмного забезпечення, оновленнями операційних систем або збільшенням активності певної групи користувачів. Аномалії, спричинені новими пристроями або технологіями, також можуть порушити нормальний мережевий трафік, оскільки нові технології часто мають відмінні вимоги до пропускної здатності або обміну даними.

Аномалії класифікуються за різними типами, кожен з яких має свої характеристики і вимагає відповідного підходу до виявлення [6-8]. Найпоширенішими є аномалії, пов'язані з обсягом трафіку. Це можуть бути як

раптові збільшення, так і зменшення трафіку, що свідчать про порушення звичного режиму роботи мережі. Зазвичай такий трафік спричиняється атаками типу DoS або неконтрольованими автоматичними процесами, коли, наприклад, шкідливий код запускає надмірну кількість запитів до сервера. Інший тип аномалій пов’язаний зі змінами в джерелах і цільових IP-адресах, що також може сигналізувати про несанкціоновані спроби доступу до ресурсів. Такі відхилення часто виникають через атаки типу «груба сила», де зловмисники намагаються отримати доступ до системи, надсилаючи численні запити для підбору паролів. Незвичні часові аномалії також можуть свідчити про проблему, коли активність проявляється в нехарактерний для цієї мережі час, що вказує на роботу автоматизованих ботів або шкідливих програм. Аномалії протоколів стосуються порушень у звичному використанні протоколів, що може бути індикатором мережевого сканування або наявності шкідливого ПЗ, яке використовує нестандартні порти [9-10].

Іншим значущим викликом є інтерпретація даних про аномалії, яка вимагає глибокого розуміння структури мережі та можливих ризиків. Дані про аномалії самі по собі не є індикатором атаки; для точного визначення загрози важливо мати доступ до інструментів моніторингу та аналізу, що дозволяють комплексно перевіряти аномальні показники. Без систематичного підходу до аналізу таких даних складно точно виявити природу потенційної загрози.

Аномалії мережевого трафіку відіграють значну роль у забезпеченні мережевої безпеки, адже вони часто є ранніми індикаторами загроз і дозволяють швидко реагувати на кібератаки чи технічні несправності. У разі ідентифікації аномалій IDS можуть направити сигнал для оцінки події, що надає можливість мережевим адміністраторам своєчасно приймати рішення. Сучасні IDS інтегруються із системами управління інформацією та подіями безпеки (SIEM), які зберігають дані про мережеву активність і дозволяють виконувати ретельний аналіз інцидентів у безпеці [11-13]. Завдяки цьому адміністратори можуть об'єднувати дані з різних джерел, знижувати ймовірність хибнопозитивних сигналів і отримувати більш повну картину безпекових загроз, що дозволяє своєчасно вживати заходів протидії.

Таким чином, аномалії мережевого трафіку є невід'ємним аспектом сучасних систем безпеки, адже вони забезпечують гнучкість у виявленні нових загроз і підвищують загальну ефективність мережевого моніторингу. Проте для досягнення максимальної точності виявлення аномалій та зниження числа хибнопозитивних сигналів необхідно постійно вдосконалювати методи аналізу, інтегруючи новітні технології, такі як машинне навчання і SIEM, що дозволяють створювати адаптивні та чутливі системи безпеки.

1.2 Методи виявлення мережевих аномалій

Методи виявлення мережевих аномалій відіграють важливу роль у сучасній кібербезпеці, забезпечуючи швидке реагування на загрози, які часто приховані серед звичайної мережевої активності. Виявлення аномалій дозволяє визначати незвичні відхилення в потоці даних, які можуть сигналізувати про вторгнення або інші проблеми в мережі. Сучасні методи виявлення аномалій включають як традиційні підходи, такі як сигнатурний аналіз і поведінкове виявлення, так і новітні технології, що базуються на машинному навчанні та обробці великих обсягів даних [14-16]. Усі ці методи мають власні унікальні характеристики і сфери застосування, проте кожен з них стикається з певними обмеженнями та викликами, пов'язаними зі специфікою мережевого трафіку та непередбачуваністю кіберзагроз.

Сигнатурний метод виявлення аномалій базується на порівнянні мережевого трафіку з базою даних відомих шаблонів атак, також званих сигнатурами. Кожна сигнатура відповідає певному типу загрози, наприклад, атаці типу відмова в обслуговуванні (DoS) або відомим видам шкідливого програмного забезпечення. Сигнатурний метод, будучи високоефективним у виявленні загроз, уже зафіксованих у базі, залишається одним з основних способів забезпечення кібербезпеки [17-19]. Проте значним його недоліком є нездатність ідентифікувати нові, невідомі атаки, зокрема так звані атаки нульового дня, які використовують

нові уразливості у програмному забезпеченні або мережевих компонентах. У сигнатурному методі необхідно постійно оновлювати базу даних сигнатур, аби система могла підтримувати високу точність виявлення і не пропускала нові загрози [20-21]. Крім того, сигнатурні системи можуть бути вразливими до модифікованих версій відомих атак, якщо вони не повністю відповідають шаблонам у базі даних.

Поведінковий метод аналізу трафіку на відміну від сигнатурного не покладається на наявність шаблонів атак у базі даних. Замість цього він аналізує стандартні шаблони поведінки в мережі та виявляє відхилення від цих норм. Основою поведінкового підходу є моделювання нормальної активності мережі шляхом аналізу регулярних патернів взаємодії між пристроями, серверами і користувачами [22-23]. Наприклад, у разі аномального зростання обсягу трафіку поведінкова система може інформувати адміністраторів про можливу загрозу. Поведінковий метод забезпечує виявлення атак нульового дня і є більш гнучким у порівнянні з сигнатурним підходом, однак цей метод схильний до хибнопозитивних спрацювань, коли звичайна активність, що виходить за рамки нормальних показників, може бути помилково визначена як загроза [24]. Зокрема, різкі зміни в поведінці користувачів, такі як масове завантаження оновлень або підключення нових пристроїв, часто класифікуються як потенційні загрози.

Методи машинного навчання та штучного інтелекту дозволяють будувати адаптивні моделі, здатні самостійно навчатися на великих масивах даних і покращувати точність виявлення загроз з часом. У машинному навчанні використовуються різні алгоритми, такі як класифікація, кластеризація, деревоподібні моделі рішень і нейронні мережі [25-27]. Наприклад, класифікаційні алгоритми дозволяють системам розподіляти трафік на групи, наприклад, «нормальний» і «аномальний». Моделі кластеризації допомагають об’єднувати подібні за характеристиками дані в групи або кластери, що дозволяє краще виявляти аномальні патерни в трафіку. Штучні нейронні мережі, як підмножина машинного навчання, здатні будувати складні моделі для обробки різноманітних змінних і значень, що дозволяє системам виявляти навіть найменші зміни в

мережевій активності [28-30]. Проте значною перешкодою для впровадження цих методів є необхідність у великих обсягах даних для тренування моделей і значні вимоги до обчислювальних потужностей.

Метод статистичного аналізу є одним із класичних підходів до виявлення мережевих аномалій, і його особливістю є робота з числовими характеристиками трафіку, такими як обсяг переданих даних, кількість з'єднань, тривалість сеансів зв'язку тощо [31-32]. Використовуючи середні значення, стандартні відхилення і кореляційний аналіз, статистичні методи можуть виявляти відхилення від середніх значень, які можуть свідчити про аномальну активність. Так, наприклад, якщо середній обсяг трафіку за певний час значно перевищує норму, це може свідчити про можливу атаку або несправність у мережі. Статистичний підхід добре підходить для базового аналізу трафіку, проте він має обмеження у випадках, коли аномалії мають складну структуру або швидко змінюються, що потребує більш динамічних і гнучких підходів.

Окрім цього, методи гібридного аналізу поєднують у собі елементи сигнатурного і поведінкового аналізу для досягнення більшої точності виявлення загроз. Гібридні методи дозволяють забезпечити ефективність як у випадках, коли зловмисник використовують відомі методи вторгнення, так і в ситуаціях з новими атаками, які не мають шаблонів у базі даних [33-35]. Гібридні системи виявлення об’єднують переваги обох підходів, мінімізуючи їхні недоліки. Такі системи можуть виявляти стандартні атаки за допомогою сигнатурного аналізу, а також адаптуватися до нових загроз завдяки поведінковому аналізу. Це робить гібридний підхід одним з найбільш універсальних для великих мереж, де ризик нових атак є постійно високим.

Застосування методів обробки великих даних є ще одним підходом до виявлення аномалій. Великі обсяги даних, що генеруються мережами, у поєднанні з технологіями обробки великих даних дозволяють збирати, зберігати і аналізувати дані у режимі реального часу [36-38]. Цей підхід є корисним для виявлення прихованих загроз та аномалій, оскільки він дозволяє системам аналізувати широкий контекст мережевої активності. Використовуючи технології обробки

великих даних, системи можуть ідентифікувати довгострокові тренди і закономірності, які можуть бути пропущені при використанні традиційних підходів. Проте обробка великих даних потребує значних ресурсів, таких як потужні обчислювальні засоби і велика кількість пам’яті, що може бути обмеженням для деяких організацій.

Одним із викликів у застосуванні методів виявлення мережевих аномалій є проблема високої кількості хибнопозитивних сигналів, яка зменшує ефективність системи і створює зайве навантаження на адміністраторів. Кожен метод виявлення має свої особливості, які можуть спричиняти помилкові спрацювання. Наприклад, поведінкові системи можуть спрацьовувати на нові, але безпечні патерни, які вони не здатні коректно класифікувати. Натомість сигнатурні методи часто пропускають нові загрози, які не мають шаблонів у базі даних. Саме тому сучасні системи виявлення часто використовують комбіновані підходи для зниження кількості хибнопозитивних спрацювань і підвищення загальної точності.

Методи виявлення мережевих аномалій постійно вдосконалюються і розвиваються у відповідь на зростання кількості кіберзагроз і складності мережевої інфраструктури. Використання штучного інтелекту, гібридних підходів і технологій обробки великих даних дозволяє створювати адаптивні, динамічні системи виявлення, які здатні працювати з високим рівнем точності та швидкістю. Проте ефективне застосування цих методів вимагає не лише наявності сучасного обладнання та великих обчислювальних ресурсів, а й компетентності мережевих аналітиків, які розуміють принципи роботи кожного методу та вміють адаптувати їх до конкретних потреб своєї мережі.

Отже, виявлення мережевих аномалій є складним, багатоетапним процесом, що потребує постійного оновлення знань про поточні загрози та вдосконалення аналітичних моделей. Системи виявлення аномалій є невід’ємною частиною сучасної мережевої безпеки, адже вони дозволяють організаціям своєчасно виявляти, аналізувати і нейтралізувати загрози, зберігаючи цілісність і доступність даних в умовах постійного розвитку кіберзагроз.

1.3 Аналіз систем виявлення вторгнень

За останнє десятиліття використання мережі різко зросло, що зробило використання систем виявлення вторгнень (IDS) актуальним тп необхідним. Існує багато прикладів мережевих програм, таких як мобільні пристрої та хмарні служби, які потребують захисту. Контроль доступу став нагальною проблемою, оскільки доступ до великих даних здійснюється через Інтернет. Питання виявлення вторгнень широко вивчалося в літературі, але і досі залишається важливим. Існує багато різних типів IDS, які застосовуються відповідно до потреб області застосування. Впровадження цих систем корисне проти різноманітних атак на мережеві системи.

Вторгнення зазвичай визначається як будь-який набір дій, які намагаються порушити цілісність, конфіденційність або доступність ресурсу. Відповідно до NIST, основною метою IDS є моніторинг та аналіз будь-якого інциденту, що мав місце в мережі системи.

IDS є частиною основного принципу безпеки комплексного захисту. Комплексний захист є основним заходом для захисту цілісності інформаційних активів компанії чи організації [39-40]. Щоб відповідати цьому принципу, можна вжити комбінацію заходів, починаючи від наявності брандмауера та безпечного програмного забезпечення для веб-додатків і запуску веб-додатку з мінімальними привілеями до використання контролю доступу до операційної системи для обмеження доступу до конфіденційних файлів та їх шифрування.

Слід зазначити, що існує багато доступних інструментів, які забезпечують IDS, наприклад Honeypots і Snort. Ці інструменти працюють відповідно до типу IDS, який вимагає адміністратор.

Основною відмінністю кожної IDS є активна та пасивна IDS. Активна IDS – це скоріше система запобігання вторгненням, оскільки вона активно блокує будь-який підозрілий трафік. Пасивні IDS не є системою запобігання, а натомість відстежують та аналізують підозрілу активність [41-42]. Виявлення аномалій є різновидом системи виявлення вторгнень. У даній роботі буде виконуватися

мережеве виявлення аномалії (або поведінки). Рис. 1.1 ілюструє класифікацію систем виявлення вторгнень.

Рисунок 1.1 – Класифікація систем виявлення вторгнень

Перший вимір, який використовується для класифікації систем виявлення вторгнень, описує, де виявляти і має два види: на основі мережі (NIDS) та на основі хоста (HIDS). NIDS це IDS, яка відстежує та аналізує мережевий трафік, щоб захистити системи від мережевих загроз. Під час процедури зіставлення створюється сповіщення, яке надсилається адміністратору, якщо NIDS виявляє будь-яку підозрілу аномальну поведінку. NIDS розміщуються в головній точці всередині мережі, щоб контролювати вхідний і вихідний трафік від усіх служб у мережі [43]. HIDS це IDS, який контролює окрему хост-систему. Для успішного моніторингу в кожній системі необхідно встановити деякі агенти. Агенти відстежують операційну систему, ініціюють сповіщення, якщо виявлено аномальну поведінку, і зберігають файли журналу. HIDS не контролює всю мережу, а лише забезпечує моніторинг окремих хост-систем, на яких встановлено агенти [44].

Що стосується другого виміру IDS, засоби виявлення базуються на сигнатурному методі та методі на основі поведінки. Сигнатурний метод стосується використання попередньо визначених шаблонів відомих атак і вразливостей системи. Метод на основі поведінки відноситься до здатності вивчати шаблони

(поведінку), щоб створити базовий рівень проти активних спроб вторгнення. Варто зазначити, що IDS на основі сигнатур зазвичай є більш поширеним, ніж IDS на основі поведінки, а IDS на основі поведінки також називають методом виявленням аномалій. Коли проводиться аналіз підходів до виявлення на основі сигнатур і поведінки, то потрібно згадати можливість поєднання цих двох методів. У багатьох роботах автори пропонують комбінацію на основі сигнатури та поведінки, щоб запобігти небажаному розкриттю даних. Система виявлення на основі поведінки використовується для створення сповіщень у разі виникнення аномальної поведінки. Після створення попередження база даних на основі сигнатур оновлюється новою сигнатурною атакою.

Системи виявлення вторгнень також можна класифікувати за підходом фільтрації за білим або чорним списком. Білий список це практика безпеки обмеження доступу, якщо це явно не дозволено. Чорний список це практика безпеки, що дозволяє доступ, якщо він явно не заборонений. Цей підхід зазвичай використовується антивірусним програмним забезпеченням. IDS на основі поведінкових білих списків, які часто називають системами виявлення аномалій (або скорочено ADS), спрямовані на аналіз поведінки мережевого трафіку, роблячи різницю між нормальною та аномальною поведінкою. Аномальна поведінка – це будь-яка поведінка, яка відрізняється від нормальної поведінки системи. Перевагою більшості ADS перед використанням IDS на основі сигнатур є гнучкість виявлення невідомих видів атак.

Виявлення аномалії має три різні режими: контрольований, напівконтрольований та неконтрольований. Контрольований режим – це те саме, що контрольоване навчання. Класифікатор навчається за допомогою позначеного навчального набору, і цей класифікатор застосовується до тестового набору. В основному в цьому режимі аномалії повинні бути відомі заздалегідь. Напівконтрольований режим складається з двох етапів: навчання та тестування. Навчальний набір даних містить лише приклади звичайних класів. Модель нормальної поведінки вивчається під час фази навчання. Оцінка моделі відбувається на етапі тестування. Тестовий набір має записи про норму та аномалії.

Неконтрольований режим є найскладнішим режимом, оскільки не можна робити жодних припущень щодо набору даних. Крім того, у цьому режимі неможливо виконати навчання моделі. У цьому режимі для аналізу набору даних використовуються статистичні показники, наприклад підходи до відстаней [45-47].

Існує безліч інструментів і методів, які використовуються для виявлення або запобігання вторгненням систем, таких як Snort і honeypots. Snort є однією з найпопулярніших систем виявлення мережевих вторгнень і зазвичай використовується для виявлення того, чи хтось отримав доступ або намагався отримати доступ до мережевої системи. Snort – це IDS на основі сигнатур, що залежить від попередньо визначених правил і шаблонів, наданих розробником системи [48]. Honeypot – це група підроблених серверів, розміщених після брандмауера, щоб маніпулювати та ловити зловмисників. Як правило, це захисний інструмент для злому зловмисників. Після того, як зловмисник потрапив у пастку, розробник приманки може дізнатися більше про наміри та прийоми зловмисника [49]. У цьому сенсі honeypot можна використовувати як інструмент виявлення вторгнень для вивчення моделей атак. Snort і honeypots – це інструменти, які використовуються як системи виявлення на основі сигнатур, і в літературі є різні варіанти реалізації цих методів. Крім того, обидва методи вже мають високу точність виявлення вторгнень. Хоча обидва методи мають обмеження, оскільки вони виконуються як запобігання чорному списку. Такий підхід погано працює проти невідомих атак. Метод запобігання цих методів є більш пасивним проти атак, подібно до брандмауерів. Крім того, правила Snort важко вбудувати, оскільки вони вимагають чіткого розуміння мережевого трафіку, щоб створювати сповіщення про аномальну поведінку, а приманками важко керувати, що робить мережеву систему вразливою для зловмисників.

Порівняння систем виявлення вторгнень (IDS) розглядає мережеві та хостові IDS, а також сигнатурні та поведінкові методи виявлення загроз, з огляду на їхні ключові переваги та обмеження. Мережеві IDS мають низьку вартість володіння, дозволяють швидко виявляти загрози і автономно працюють поза серверними ресурсами, хоча можуть пропускати частину трафіку і, відповідно, генерувати

більше хибнопозитивних спрацьовувань. Хостові IDS не потребують додаткового обладнання та ефективно виявляють широкий діапазон атак, проте не захищають від DoS-атак і стають неефективними, якщо сам сервер скомпрометовано. Хостові IDS також можуть впливати на продуктивність сервера, на відміну від мережевих, які працюють автономно. У контексті сигнатурного підходу, IDS забезпечують низький рівень хибнопозитивних спрацювань і добре розпізнають відомі загрози. Проте ефективність цих систем залежить від постійного оновлення бази сигнатур, оскільки невідомі типи атак залишаються невиявленими. У свою чергу, поведінкові IDS пристосовані до виявлення нових атак за допомогою аналізу аномальної поведінки, що дозволяє виявляти так звані "атаки нульового дня". Однак вони більш схильні до високої частоти хибнопозитивних спрацювань, оскільки мають складнощі з адаптацією до динамічних шаблонів атак.

Процес розгортання також відрізняється: мережеві IDS вимагають складної інфраструктури, тоді як хостові – легші для встановлення на окремих серверах, але потребують налаштування для кожного агента. Важливою особливістю IDS є частота хибнопозитивних спрацьовувань – випадки, коли система позначає звичайну активність як загрозу. Наприклад, за високого обсягу трафіку навіть низький відсоток хибних спрацювань може призвести до значного обсягу помилкових сигналів. Незважаючи на це, зменшення таких помилкових спрацьовувань залишається важливим завданням для кожної IDS, особливо для поведінкових методів, де частота помилкових сигналів є вищою. Сигнатурні системи є більш надійними з точки зору обмеження рівня хибнопозитивних спрацьовувань, проте вони менш гнучкі у виявленні нових загроз.

Багато сучасних IDS інтегруються з системами управління інформацією та подіями безпеки (SIEM), що забезпечує ширший контекст для аналізу загроз і підвищує загальну точність та оперативність реагування. Це дозволяє системам краще відстежувати загрози, а також ефективно працювати з даними та знижувати кількість хибнопозитивних результатів.

1.4 Постановка задачі

Сьогодні кіберпростір стає дедалі вразливішим до різноманітних загроз. Ці загрози серйозно порушують стабільність роботи і доступність мереж, що робить виявлення аномалій у трафіку одним із важливих напрямків у сфері кібербезпеки. Виявлення аномалій у мережевому трафіку здатне забезпечити раннє попередження про можливі вторгнення та відхилення у звичайній активності, що допомагає адміністраторам швидко реагувати на потенційні загрози. Однак більшість існуючих рішень виявлення аномалій не досягають необхідного рівня ефективності та точності, оскільки використовують або застарілі методи, або підходи, які не здатні адаптуватися до швидких змін у кіберзагрозах. Це зумовлює значний рівень хибнопозитивних спрацювань, що перевантажує системи кібербезпеки та відволікає увагу на нерелевантні загрози.

Метою цієї роботи є розробка методу виявлення аномалій у мережевому трафіку, що мінімізує кількість хибних спрацювань та підвищує загальну точність і адаптивність системи. З огляду на швидкий розвиток загроз, такий метод має бути гнучким і здатним автоматично пристосовуватися до нових типів аномалій.

Для досягнення мети необхідно детально проаналізувати існуючі методи виявлення аномалій та зрозуміти їх недоліки, порівнявши переваги та обмеження кожного підходу. Далі доцільно проаналізувати наявні набори даних для тестування розробленого методу та обрати оптимальний варіант, який враховуватиме специфіку досліджуваної мережі та дозволить отримати максимальний результат. У роботі буде розроблено алгоритм виявлення аномалій, заснований на методах машинного навчання, які можуть забезпечити високий рівень класифікації нормального та аномального трафіку. Розроблений метод варто протестувати на попередньо обраному наборі даних, щоб визначити його ефективність, зокрема, за такими показниками, як точність, швидкодія та рівень хибнопозитивних спрацювань.

2 НАБОРИ ДАНИХ ДЛЯ АНАЛІЗУ МЕРЕЖЕВОГО ТРАФІКУ

2.1 Аналіз існуючих наборів даних

Набір даних NSL-KDD є вдосконаленою версією відомого KDD’99, який довгий час був стандартом для тестування систем виявлення вторгнень. Основною перевагою NSL-KDD є усунення надмірної кількості дубльованих записів, що були в оригінальному наборі KDD’99. Цей набір даних містить кілька основних типів атак: Denial of Service (DOS), Probe, Remote to Local (R2L) та User to Root (U2R), що дозволяє покривати широкий спектр потенційних загроз. NSL-KDD використовується для навчання алгоритмів машинного навчання, забезпечуючи їм можливість розпізнавати аномальні шаблони [50]. Однак, основним недоліком цього набору є його обмеженість в актуальності та реалістичності. Сучасні мережі є набагато складнішими порівняно з тими, для яких розроблявся KDD’99, тому NSL-KDD не відображає всі сучасні загрози та види атак, що знижує точність алгоритмів, навчених на цьому наборі даних, у сучасних мережевих умовах.

Інший набір даних, CICIDS2017, є більш сучасним і реалістичним. Цей набір створений Канадським інститутом кібербезпеки з метою імітації реальних сценаріїв мережевого трафіку. Набір даних CICIDS2017 охоплює широкий спектр атак, включаючи DDoS, Brute Force, Botnet, SQL Injection, а також різноманітні веб-атаки [51]. Основна перевага цього набору в тому, що він дозволяє тестувати алгоритми на реалістичних та сучасних загрозах, що значно підвищує релевантність отриманих результатів. Проте його основним недоліком є велика кількість даних, що може значно збільшити час обробки та вимоги до обчислювальних ресурсів. Для багатьох дослідницьких лабораторій це може стати проблемою, оскільки обробка даних CICIDS2017 вимагає потужного апаратного забезпечення.

UNSW-NB15 також є сучасним набором, який широко застосовується для тестування систем виявлення аномалій. Створений в Університеті Нового Південного Уельсу, цей набір включає як нормальний трафік, так і дані, що відображають кілька видів атак, зокрема Fuzzers, Analysis, Backdoors, DoS, Exploits,

Generic, Reconnaissance, Shellcode і Worms. UNSW-NB15 відрізняється структурованістю і дозволяє дослідникам вивчати широкий спектр атак. Він підходить для тестування різноманітних методів машинного навчання, включаючи нейронні мережі та алгоритми глибокого навчання [52]. Проте, недоліком UNSW-NB15 є складність у відтворенні його у реальному часі, оскільки він містить велику кількість добре структурованих даних, які можуть не завжди відповідати реальним умовам у мережах. Це може призводити до того, що алгоритми, які показують високу точність на UNSW-NB15, можуть виявитися менш ефективними у більш динамічному середовищі.

CIC DoS Dataset 2019 є спеціалізованим набором даних, призначеним для аналізу атак відмови в обслуговуванні. Він містить трафік, згенерований при виконанні різних DoS-атак, таких як HTTP DoS, TCP SYN Flood і UDP Flood. Цей набір є ідеальним для досліджень, що фокусуються на конкретному типі атак і потребують високої точності у виявленні DoS-активності. Однак, вузька спеціалізація CIC DoS Dataset обмежує його застосування в задачах, де необхідне виявлення різноманітних видів атак. Це означає, що дослідникам, які потребують комплексного набору для різних загроз, доведеться доповнювати його іншими даними.

MAWI Dataset, розроблений Центром аналізу глобального інтернету (WIDE Project) у Японії, включає трафік, що відображає реальні мережеві сценарії. MAWI є корисним для загального аналізу аномалій, оскільки включає широкий спектр типів трафіку та різні види аномальних подій [53]. Однак одним з його основних недоліків є відсутність детальної інформації про типи атак, що може обмежити його ефективність у задачах, де потрібно чітко ідентифікувати конкретні загрози. Це робить MAWI корисним для загального вивчення тенденцій, але менш ефективним для специфічного аналізу та тестування алгоритмів виявлення вторгнень.

CTU-13 Dataset – це набір, що був створений Чеським технічним університетом і орієнтований на трафік, пов'язаний із ботнет-мережами. CTU-13 дозволяє дослідникам аналізувати поведінку ботнетів у реальних умовах, імітуючи активність, характерну для заражених систем. Він включає трафік як нормальних

пристроїв, так і пристроїв, інфікованих ботнетами, що дозволяє створювати ефективні моделі для виявлення ботнет-активності. Недоліком CTU-13 є його вузька спрямованість на ботнети, що робить його менш ефективним для загального аналізу інших видів аномалій у мережевому трафіку.

Ton-IoT Dataset є набором даних для виявлення аномалій в трафіку Інтернету речей (IoT). Створений для врахування різних IoT-пристроїв, цей набір включає дані як про нормальні, так і про атаковані сценарії. Ton-IoT дозволяє дослідникам вивчати аномалії у трафіку IoT-мереж, які мають специфічні характеристики порівняно з традиційним мережевим трафіком. Основним недоліком Ton-IoT є його орієнтованість на конкретний сегмент IoT, що знижує його корисність для загального аналізу аномалій у класичних мережах. Це обмежує його застосування лише для задач, пов'язаних з IoT-інфраструктурою.

LBNL Dataset – це набір даних, створений Ліверморською національною лабораторією. Він включає анонімізовані дані мережевого трафіку, зібраного у реальних умовах дослідницької мережі, і надає унікальні можливості для аналізу аномалій. LBNL Dataset підходить для дослідження тенденцій у мережевому трафіку, але оскільки він анонімізований, це може ускладнити розпізнавання конкретних видів атак. Крім того, анонімізація може приховати важливі деталі, що робить LBNL менш ефективним для спеціалізованих завдань виявлення аномалій.

DAPT2020 (Dataset of Anomalous Packet Traffic) – це відносно новий набір даних, створений для вивчення різноманітних типів аномалій у мережевому трафіку. Він дозволяє досліджувати різні види атак, що полегшує тестування алгоритмів для виявлення аномалій. Однак, DAPT2020 може мати обмеження щодо масштабованості, оскільки він фокусується на конкретних сценаріях і може бути недостатньо гнучким для динамічних середовищ.

ISCX IDS 2012, розроблений Канадським інститутом кібербезпеки, включає реальні дані з різними типами атак, такими як DoS, SSH, HTTP, ICMP тощо. Цей набір даних імітує реальне середовище мережевих атак, що робить його корисним для тестування різноманітних алгоритмів. Проте, ISCX IDS 2012 також має обмеження, пов'язані з застарілістю певних типів атак. З огляду на сучасний рівень

розвитку кіберзагроз, певні патерни в цьому наборі даних можуть не відповідати реальним умовам, що знижує його ефективність для тестування сучасних систем IDS.

Власний набір даних для виявлення аномалій у мережевому трафіку може забезпечити значно вищу ефективність і релевантність для конкретних завдань та умов, ніж загальнодоступні набори. Існуючі популярні набори даних, такі як NSL-KDD, CICIDS2017, UNSW-NB15 та інші, хоча й є важливими інструментами для досліджень, мають певні недоліки. Серед найпоширеніших проблем – їхня застарілість, вузька спеціалізація, неповнота або надлишок інформації, що ускладнює їхнє застосування в сучасному мережевому середовищі. Деякі набори даних можуть містити застарілі види атак або бути не адаптованими до специфічних умов, таких як трафік у мережах IoT або специфічні вимоги корпоративних інфраструктур. Окрім того, для успішного навчання системи на конкретні види аномалій необхідно мати дані, що максимально точно відображають реальні умови експлуатації.

Розробка власного набору даних дозволяє врахувати унікальні характеристики мережі, що підвищує релевантність отриманих моделей для конкретного середовища. Наприклад, у корпоративній мережі можуть бути особливі правила трафіку, що відрізняються від загальних шаблонів. Власний набір також дозволяє зосередитися на конкретних типах загроз або аномалій, які є пріоритетними для конкретної організації. Важливо також те, що власний набір даних може враховувати останні тенденції та види атак, тоді як загальнодоступні набори даних часто оновлюються із затримкою, що знижує їхню актуальність у швидкоплинному світі кібербезпеки.

Крім того, власний набір даних надає можливість коригувати обсяг і деталі інформації, включаючи як низькорівневі особливості (наприклад, пакети), так і високорівневі (наприклад, шаблони сеансів), що дозволяє створювати моделі, оптимізовані для конкретного застосування. Це дозволяє не лише покращити точність виявлення аномалій, але й знизити кількість хибних спрацювань, оскільки система навчається розпізнавати реальні, а не теоретичні загрози.

Отже, створення власного набору даних може значно підвищити ефективність та релевантність моделі виявлення аномалій, забезпечуючи адаптацію до унікальних характеристик мережевого середовища, актуальність до сучасних загроз та оптимізацію для конкретних потреб. Хоча створення власного набору є ресурсозатратним завданням, його довгострокова цінність для точності й надійності системи виявлення аномалій виправдовує затрати часу та ресурсів.

2.2 Захоплення даних за допомогою Wireshark та їх модифікація

Першим етапом реалізації пропонованого методу є отримання даних, що охоплює збір інформації та її перетворення у формат, придатний для подальшого аналізу. Щоб отримати дані про мережевий трафік, необхідно здійснити моніторинг трафіку. Це дозволяє захоплювати пакети, що передаються з мережевого адаптера у локальній підмережі в заданий момент часу. Після завершення процесу захоплення пакетів розпочинається аналіз використаних протоколів передачі. Серед найпоширеніших типів протоколів можна виділити IP (Internet Protocol), TCP (Transmission Control Protocol) та UDP (User Datagram Protocol). Кожен з них забезпечує певний рівень управління передачею даних і має свої специфічні властивості, що відображаються під час аналізу [54-56].

Для дослідження пакетів часто використовуються аналітичні інструменти, такі як Wireshark або Microsoft Network Monitor, які містять вбудовані аналізатори трафіку. Ці програми дозволяють отримати детальну інформацію про типи протоколів і характеристики кожного окремого пакета. Захоплення даних завершується, коли весь мережевий трафік зібрано, а аналізатор пакетів обробив та класифікував зібрані пакети. У даній роботі буде використано аналізатор мережевого трафіку Wireshark.

Другий аспект підготовки даних полягає у обробці даних з метою вибору відповідних характеристик, які використовуватимуться в аналізі. Характеристики – це параметри, що визначають особливості кожного запису у наборі даних і

застосовуються під час побудови моделей. Загалом для мережевого трафіку такими характеристиками можуть бути, наприклад, IP-адреса джерела та адреса призначення, тип протоколу тощо. Така трансформація даних необхідна тоді, коли початковий набір характеристик не дає корисної інформації. Скажімо, якщо у даних від VPN-з'єднання однакові адреси джерела і призначення, ці параметри виявляються непридатними для аналізу аномалій.

На цьому етапі важливо визначити найсуттєвіші характеристики для дослідження. Враховуючи, що запропонований підхід базується на неконтрольованому виявленні аномалій, то вибір параметрів є важливою частиною роботи, оскільки він спрощує модель і робить її зрозумілою для користувачів. За результатами досліджень, надмірна кількість характеристик може знизити точність виявлення аномалій. У той же час зменшення кількості параметрів може підвищити рівень хибнопозитивних спрацювань – випадків, коли звичайний трафік помилково визначається як аномалія.

В даній роботі пріоритетним є зменшення кількості характеристик задля досягнення високої точності, оскільки важливо знизити вплив хибнопозитивних результатів. Взаємозв’язок між рівнем точності та хибнопозитивними спрацюваннями простежується у вигляді співвідношення: чим більше правильно класифікованих нормальних значень і правильно виявлених аномалій, тим вища точність і нижчий рівень хибнопозитивних спрацювань. Зменшення кількості характеристик також підвищує швидкість обробки, що спрощує процес екстракції даних.

Серед технік для вибору характеристик можна виділити обгортковий метод (Wrappers), фільтрування (Filters) та вбудований метод (Embedded). Обгортковий метод оцінює підмножину характеристик за допомогою алгоритму пошуку та прогнозної моделі, що дозволяє вибрати оптимальний набір параметрів для конкретного завдання. Фільтрування обирає характеристики виключно на основі властивостей даних, ігноруючи методи, що використовуються для виявлення аномалій, що дозволяє скоротити обчислювальний час. Вбудований метод об’єднує елементи перших двох методів, оптимізуючи відбір параметрів для конкретного

завдання. У даному випадку буде застосовано метод фільтрації, оскільки інші методи мають підвищений ризик надмірного налаштування моделі під навчальні дані, що може знизити загальну точність.

Результатом процесу обробки даних є створення набору даних, що містить релевантні параметри, і готовий до використання на етапі виявлення аномалій. Цей остаточний набір даних є базою для подальшого аналізу і є вагомим для забезпечення успішного виявлення та оцінки мережевих загроз.

Wireshark обрано як інструмент захоплення, щоб отримати власний набір даних. Wireshark – це безкоштовний аналізатор пакетів із відкритим кодом. Wireshark використовується переважно для мережі з метою усунення несправностей, для аналізу мережевих протоколів і розробки програмного забезпечення та протоколів зв'язку. Wireshark використовує файли Packet CAPture file (PCAP) для захоплення пакетів. Файл PCAP складається з API для захоплення мережевого трафіку. Запустивши параметр захоплення на Wireshark через вибраний мережевий інтерфейс, буде зібрано необроблені IP-дані та всі передані пакети під час захоплення [57]. На цьому етапі можна захопити багато пакетів, включаючи HTTP-пакети зв’язку та TCP/UDP-пакети. Wireshark надає різноманітну інформацію від захоплення мережевого трафіку, таку як IP-адреса джерела, IP-адреса призначення, довжина, час тощо. На рис. 2.1 наведено приклад інтерфейсу Wireshark.

Wireshark має кілька варіантів захоплення, наприклад бездротове мережеве підключення, підключення по локальній мережі та USBCap. Бездротове мережеве підключення стосується всіх даних, що передаються через мережу Wi-Fi. Підключення по локальній мережі стосується кожного мережевого підключення Ethernet. USBPcap стосується імпорту даних із флеш-драйвера USB. Можливе поєднання цих варіантів. Перші два параметри можна використовувати для захоплення мережевого трафіку в реальному часі.

Рівень серйозності є першою підказкою для користувача щодо важливої інформації та має чотири рівні серйозності: чат, примітка, попередження, помилка. Цей стовпець не відображається в стандартному поданні Wireshark.

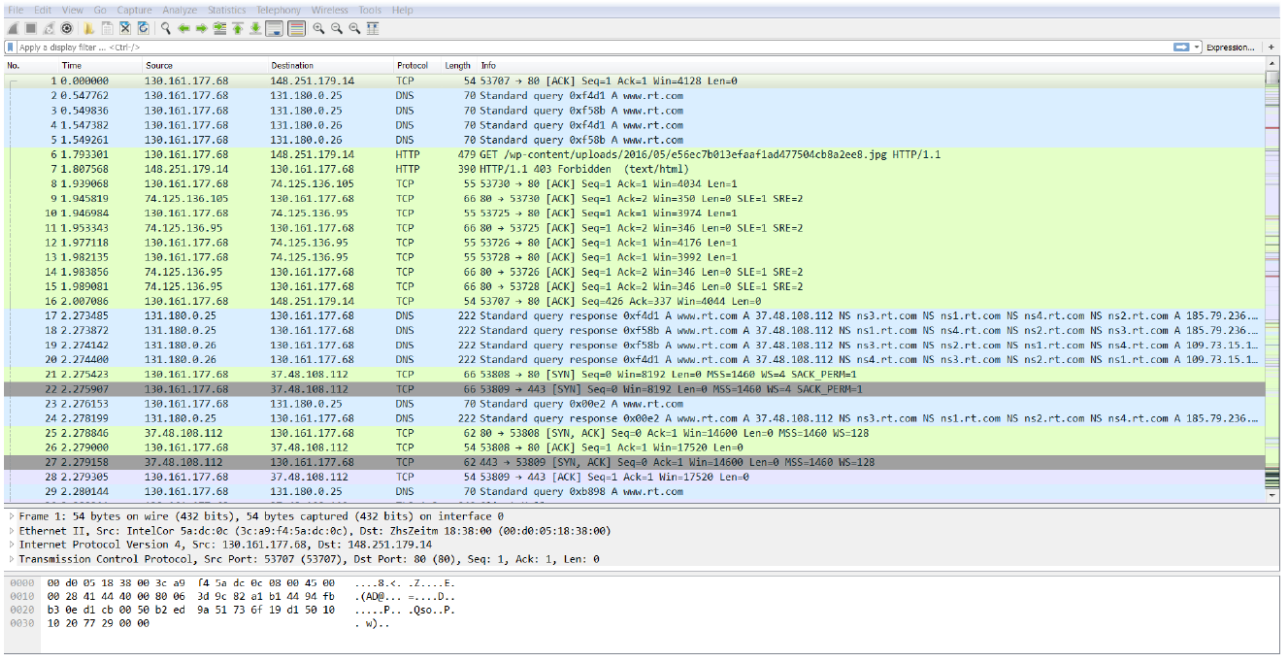


Рисунок 2.1 - Екран Wireshark після аналізу файлу .pcap

Перш ніж почати захоплення мережевого трафіку, потрібно активувати всі компоненти Data Plane Interface (DPIF). DPIF – це фреймворк, який у Wireshark використовується для аналізу та обробки мережевого трафіку на рівні каналу передачі даних. Основними компонентами DPIF є такі модулі, що забезпечують збір, обробку та аналіз мережевих даних, дозволяючи користувачам глибше розуміти та виявляти аномалії або проблеми в мережевому середовищі.

Capture Engine (Модуль захоплення трафіку) – основний компонент, який відповідає за безпосереднє захоплення мережевого трафіку з мережевих інтерфейсів. Він дозволяє записувати пакети в реальному часі, зберігаючи їх для подальшого аналізу. Wireshark може взаємодіяти з різними типами мережевих інтерфейсів, зокрема Ethernet, Wi-Fi, і навіть віртуальними мережевими інтерфейсами. Capture Engine також забезпечує конфігурацію фільтрів, що дозволяє користувачам фокусуватися лише на конкретних типах трафіку або протоколах.

Dissector Modules (Модулі розбору протоколів) – цей компонент відповідає за розпізнавання, інтерпретацію та декодування різних протоколів у захоплених пакетах. Кожен протокол має свій власний «розбірник» або «диссектор», який

зчитує дані, витягує корисну інформацію та відображає її у зручному вигляді для користувача. Завдяки цьому компоненту Wireshark може інтерпретувати тисячі мережевих протоколів і надавати детальну інформацію для кожного з них, від IP-адрес до прикладних даних.

Filtering Engine (Модуль фільтрації) – цей компонент дозволяє створювати фільтри для точного вибору трафіку, який буде захоплений або проаналізований. Фільтри можна налаштовувати на різних рівнях мережевої моделі, таких як IP-адреси, порти, протоколи або специфічні поля заголовків. У Wireshark є два типи фільтрів: capture filters (фільтри для захоплення) і display filters (фільтри для відображення). Capture filters застосовуються ще під час процесу захоплення даних і знижують обсяг трафіку, який записується. Display filters застосовуються вже після захоплення, допомагаючи аналізувати конкретні пакети.

Statistics Engine (Модуль статистики) – дозволяє генерувати статистичні звіти, що допомагають у виявленні аномалій та загального розуміння поведінки мережевого трафіку. Статистика може охоплювати різні аспекти трафіку, такі як частота протоколів, кількість пакетів, затримки, пропускна здатність та інші параметри. Цей компонент також може створювати графічні звіти, що полегшує інтерпретацію результатів для аналізу продуктивності мережі та визначення проблемних ділянок.

Decoding and Analysis Engine (Модуль декодування та аналізу) – відповідальний за глибокий аналіз і декодування пакетів. Цей компонент виконує функції повного розпізнавання пакетів на всіх рівнях мережевої моделі OSI, дозволяючи Wireshark точно розуміти, як дані інтерпретуються різними протоколами. Завдяки цьому компоненту користувачі можуть глибше аналізувати структуру пакетів, включаючи структуру заголовків та корисного навантаження.

Plugin System (Система плагінів) – дозволяє користувачам додавати нові функції або підтримку додаткових протоколів. Завдяки плагінам можна розширювати можливості Wireshark без потреби модифікувати основний код. Це корисно для дослідників або адміністраторів, які працюють з протоколами або специфічними типами трафіку, що не підтримуються «з коробки». Плагіни можуть

додавати нові диссектори, фільтри, статистичні модулі або навіть повністю нові функції аналізу.

User Interface (Інтерфейс користувача) – забезпечує користувачам доступ до всіх компонентів та можливостей Wireshark, від захоплення трафіку до детального аналізу пакетів. Інтерфейс дозволяє переглядати структуру пакетів у вигляді дерева, використовувати фільтри, переглядати статистику, запускати скрипти та експортувати дані. Зручність інтерфейсу допомагає користувачам швидко орієнтуватися в обсягах мережевого трафіку і знаходити потрібну інформацію.

Процедура запуску DPIF проводиться лабораторією безпеки Thales. Під час активації DPIF вібувається запуск Wireshark, щоб перехопити мережевий трафік DPIF. Усі налаштування та частина захоплення мережевого трафіку DPIF зберігає Thales. Після завершення збору даних отримано перші дані про мережевий трафік, тоді можна продовжити вилучення даних.

Протокол необроблених даних – це лише IPv4, який включає TCP. TCP складається з WebSockets і HTTP. Щоб збирати дані для виявлення аномалій, потрібні протоколи зв’язку, якими в даному випадку є протокол HTTP-запиту та відповіді. Wireshark надає параметр фільтра для відображення запитаних пакетів протоколу. В отриманих даних IP-адреса джерела та IP-адреса призначення однакові оскільки всі сервери працюють локально.

У багатьох випадках різні ознаки обираються залежно від цілей виявлення аномалій та галузі застосування. У даній роботі буде застосовано фільтраційний підхід, спираючись на статистику, яку надає Wireshark. Програма проводить різноманітні статистичні аналізи, використовуючи загальні та специфічні протокольні статистики. Загальні статистичні дані дозволяють провести комплексний аналіз пакетів, надаючи інформацію про ієрархію захоплених пакетів, включаючи типи протоколів, кількість пакетів і швидкість передачі. Наприклад, розділ "Розмови" відображає трафік між конкретними IP-адресами, а "Кінцеві точки" демонструють дані про трафік, що надходить і виходить від IP. Специфічні протокольні статистики пропонують детальний аналіз кожного набору пакетів, зокрема для протоколів HTTP та IPv4. Статистика HTTP надає відомості про

лічильник пакетів, розподіл навантаження та запити, що надходять. Важливо зазначити, що пакети HTTP мають рівень серйозності, аналогічний чату. У той же час статистика IPv4 охоплює всі адреси, порти, типи IP-протоколів, а також джерела і призначення пакетів IPv4. Зосередившись на статистичних даних Wireshark, можна зробити кілька цікавих спостережень. Загальні статистичні дані забезпечують загальний аналіз пакетів, тоді як специфічні протокольні дані дозволяють більш детально вивчити кожен набір протоколів.

Після перерахування всіх доступних ознак з файлу PCAP застосовуємо фільтраційний метод для вибору фінальних характеристик. У пропонованому методі для пошуку аномалій зосередимося на транспортних і комунікаційних протоколах, зокрема на IPv4 і HTTP.

Аналізуючи статистику IPv4 та HTTP, буде отримано інформацію для остаточного відбору ознак, які будуть використані для виявлення аномалій. Статистика IPv4 надає дані про всі IP-адреси, призначення і порти, а також типи IP-протоколів, що дозволяє зрозуміти, звідки та куди направляються пакети. Статистика HTTP, в свою чергу, інформує про кількість запитів та відповідей, файли, пов’язані з веб-сервером, а також про розподіл навантаження, що допомагає зрозуміти, як сервер обробляє запити.

Цей аналіз статистики забезпечує можливість точно визначити ті ознаки, які буде використано в подальшій роботі з виявлення аномалій. Наприклад, кількість отриманих пакетів, їхній час обробки в мілісекундах, частка певного типу пакетів у загальному обсязі, а також максимальна кількість пакетів, що надходять за короткий проміжок часу, дозволяють виявити потенційні аномалії в мережевій активності. Визначення цих характеристик сприяє більш глибокому аналізу трафіку та покращує точність виявлення відхилень у поведінці мережі.

Щоб продовжити роботу з цими статистичними даними, було проведено аналіз характеристик, які хотіли б виділити. Після виконання статистичного аналізу для IPv4 та HTTP у нас було сім CSV-файлів. З них залишили чотири: "призначення та порти" з IPv4, а також "лічильник пакетів", "запит" та "розподіл навантаження" з HTTP. Інші три файли з IPv4, які містили всі адреси, типи IP-протоколів та

джерела і призначення, повернули загальне значення для однієї адреси – 127.0.0.1, оскільки всі сервери працюють у локальній мережі. Тому виключили ці дані з остаточного набору.

Щоб створити уніфікований набір даних, було об’єднано всі CSV-файли в один, який включає як HTTP, так і IPv4 статистики. Обидва типи статистичних даних мають однакові атрибути. Деякі з колонок були порожніми, тому їх видалено. Інші незначні зміни, такі як заповнення нулями порожніх атрибутів (всього було лише чотири пустих атрибути), також було здійснено. Набір містить один спеціальний атрибут, який використовується як ідентифікатор, і п’ять звичайних атрибутів, які насправді є характеристиками для виявлення аномалій. Спеціальний атрибут – це "тема/предмет", а п’ять характеристик – це "кількість", "швидкість (мс)", "відсоток", "потік" і "початок потоку".

На завершення, було експортовано остаточний набір даних за допомогою Wireshark, що дозволяє експортувати файли в різних форматах. Таким чином, остаточний набір даних готовий до етапу виявлення аномалій.

2.3 Вибір алгоритму виявлення аномалій

Наступним етапом після захоплення мережевого трафіку є виявлення аномалій. Алгоритм неконтрольованого навчання для виявлення аномалій дозволяє виявляти незвичайні або аномальні патерни в даних без використання попередньо мічених прикладів. На відміну від методів контрольованого навчання, де дані мають відповідні мітки (наприклад, "нормально" або "аномально"), алгоритми неконтрольованого навчання самостійно аналізують структуру даних і виявляють відхилення від звичного. Цей підхід є корисним, оскільки не вимагає міток, що дозволяє виявляти нові аномалії, які раніше не були відомі. Для цього часто застосовують статистичні методи, які допомагають визначити дані, що відхиляються від нормального розподілу. Такі методи широко застосовуються в різних сферах, таких як фінансовий моніторинг для виявлення шахрайських

транзакцій, інформаційна безпека для виявлення вторгнень у мережах, а також у промисловості для моніторингу обладнання з метою виявлення потенційних поломок.

Варто зазначити, що результати неконтрольованого виявлення аномалій залежать від вибору ознак. Вибір ознак здійснювався на етапі збору даних за допомогою фільтраційних методів. Отже, набір даних з відповідними ознаками використовується як вхід для етапу виявлення аномалій. Мета виявлення аномалій полягає в тому, щоб виявити поведінку, що відрізняється від звичайних випадків. Під час виявлення аномалій значення ознак у наборі даних використовуються для розрахунку статистичних показників. Ці показники відрізняються залежно від застосованого методу виявлення аномалій. Аномальна поведінка позначається як аномалії або викиди, які є спостереженнями, що істотно відрізняються від інших даних у наборі і викликають тривогу.

Рішення про нормальну поведінку приймається на основі розрахункової моделі вибраної техніки виявлення аномалій. Нормальна поведінка визначається базовим рівнем, який вказує на те, що спостереження з високим балом викиду вважаються аномаліями. Кожна техніка має свою розрахункову модель балу викиду, яка відрізняється в залежності від методу виявлення аномалій. Важливо зазначити, що вибір порогу відіграє вагому роль у виявленні аномалій; висока частка хибнопозитивних результатів і низька точність можуть свідчити про неправильний вибір порогу.

Коли всі спостереження набору даних будуть оброблені методом виявлення аномалій, ми отримаємо новий набір даних з новою спеціальною ознакою для кожного випадку, яка називається балом викиду. Оскільки бал викиду деяких спостережень перевищує базовий рівень, буде отримано певні сповіщення.

Виявлення аномалій часто здійснюється за допомогою алгоритмів машинного навчання, які можна розділити на кілька категорій: на основі найближчих сусідів, кластеризації, статистичних методів та ядерних методів.

Алгоритми, засновані на найближчих сусідах, включають K-NN, який обчислює середню відстань до *k*-ого найближчого сусіда для визначення показника

викиду. Якщо цей показник високий, це свідчить про наявність аномалії. Local Outliers Factor (LOF) працює подібно до K-NN, але обчислює відстані до найближчих сусідів і формує набір сусідів, які знаходяться на відстані *k-1*. Якщо показники LOF перевищують одиницю, це вважається аномалією, при цьому LOF може обробляти дублі [58]. Connectivity-based Outlier Factor (COF) є варіацією LOF, здатною виявляти викиди, які відхиляються від патернів низької щільності. Local Correlation Integral (LOCI) також визначає викиди за методом K-NN, але має дві переваги: параметри не суттєво впливають на результати, і LOCI використовує автоматичний механізм статистичного відхилення для визначення викидів. Приблизний Local Correlation Integral (aLOCI) є варіацією LOCI. Local Outlier Probability (LoOP) розраховує ймовірність того, що об'єкт є локальним викидом на основі щільності. Influenced Outlierness (INFLO) враховує сусідів і зворотних сусідів при обчисленні локальної щільності певної точки.

Алгоритми кластеризації, такі як Cluster-Based Local Outlier Factor (CBLOF), приймають набір даних і створюють кластерну модель за обраним методом. Показник викиду розраховується залежно від розміру кластеру та відстані до найближчого великого центроїда. Local Density Cluster-Based Outlier Factor (LDCOF) подібний до CBLOF, але показник викиду обчислюється на основі відстані до найближчого великого кластеру, розділеної на середню відстань у великому кластері. Алгоритм Clustering-based Multivariate Gaussian Outlier Score (CMGOS) оперує кластеризованим набором даних із побудовою моделі центроїдів.

Статистичні методи, такі як Histogram-based Outlier Score (HBOS), розраховують показник викиду, обчислюючи окремий одновимірний гістограм для кожної колонки набору даних [59]. Color Coded Join працює подібно до HBOS, зосереджуючись на збереженні кольорового кодування. Robust Principal Component Analysis Anomaly Score (rPCA) виконує обчислення за допомогою стійкого PCA, визначаючи відстань Махаланобіса та обчислюючи показник викиду на основі основних компонент.

Метод на основі ядер, такий як One class Library for Support Vector Machines (LIBSVM), є напівконтрольованою варіацією Support Vector Machines, що визначає

показники аномалій через масштабування за максимальним значенням функції рішень.

У кваліфікаційній роботі буде використано алгоритми LOF та HBOS. LOF – це метод, який оцінює викиди на основі реалізації, запропонованої Breunig та його колегами. LOF є одним з перших підходів, що ґрунтуються на локальній щільності. Процес розрахунку LOF складається з кількох етапів. Спочатку необхідно визначити набір найближчих сусідів. Визначення *k*-відстані сусідства звучить так: для заданої *k*-відстані точки *p*, її *k*-відстань сусідства містить усі об'єкти, відстань до яких не перевищує *k*-відстані. Ці об'єкти, звані *k*-найближчими сусідами точки *p*, мають різні просторові координати. Визначення вказує, що *k*-відстань для *p* має принаймні *k*-сусідів, відстань до яких менша або дорівнює цій *k*-відстані, а також не більше ніж *k-1* сусідів, відстань до яких строго менша. Відстань досяжності *(reach-dist(p,o))* визначається як максимальне значення між відстанню між точками *p* і *o* та *k*-відстанню для *o*. Локальна досяжність є оберненою величиною середньої відстані досяжності в межах набору найближчих сусідів. Формула для розрахунку локальної досяжності виглядає наступним чином: нехай *k* – натуральне число. Відстань досяжності об'єкта *p* відносно об'єкта *o* визначається як *max {k-distance(o), d(p,o)}*. LOF обчислюється як середнє значення відношення локальної щільності досяжності у наборі сусідів.

HBOS оцінює викиди, створюючи гістограму з фіксованою або динамічною шириною бінів. Цей оператор розраховує окрему одновимірну гістограму для кожної колонки набору даних. Існують два режими: один з статичною шириною бінів, інший – з динамічною. У статичному режимі кожен бін має однакову ширину, рівномірно розподілену по діапазону значень. У динамічному режимі ширина бінів може варіюватися, але можна вказати мінімальну кількість прикладів у біні. Для обчислення показника викиду гістограми спочатку нормалізуються до висоти одиниці. Потім цей показник інвертується, так що аномалії отримують високі бали, а нормальні приклади – низькі. Формула для HBOS виглядає так: *HBOS(p) = ∑(log(1/histi(p)))* для всіх *i* від 0 до *d*.

2.4 Платформа для моделювання роботи Altair RapidMiner

Для навчання та використання пропонованих алгоритмів буде використано програмне забезпечення Altair RapidMiner [60]. Це платформа, що забезпечує інтегроване середовище для машинного навчання, аналізу даних, текстового майнінгу, прогнозної та бізнес-аналітики. Платформа має три версії: Basic, Community та Professional. Перші два є безкоштовними та з відкритим кодом, причому видання Community має академічну версію, яка безкоштовна для студентів і дослідників та містить більше додатків, ніж Basic. Професійне видання є повністю комерційним пакетом з доступом до всіх сервісів Altair RapidMiner. Для виконання кваліфікаційної роботи використано академічну версію Community, яка доступна студентам. Altair RapidMiner пропонує багато розширень, які додають нові функції для користувача, в тому числі розширення для виявлення аномалій. Це розширення включає відповідні алгоритми машинного навчання, що забезпечують ефективне виявлення аномалій, а також корисний навчальний посібник із документацією та практичними прикладами. Ці алгоритми створені на основі Python.

Середовище Altair RapidMiner (рис. 2.2) досить просте і містить два основні режими перегляду: Design View та Results View. На Design View з'являється п'ять основних вкладок, а саме: Operators, Repository, Process, Parameters та Help, що полегшує навігацію та процес налаштування параметрів для аналізу.

Вкладка Operators містить усі доступні оператори Altair RapidMiner, а також додаткові розширення, які були завантажені. У вкладці Repository зберігаються завантажені дані, збережені процеси та активні локальні процеси. Process є робочою областю для запущеного процесу, що дозволяє легко перетягувати оператори з вкладки Operators. Параметри кожного оператора можна налаштувати у вкладці Parameters, а Help містить документацію та інформацію про оператори та розширення.

Режим Result відображає виконання процесу, де головна вкладка Result History зберігає історію всіх запущених процесів. Інші вкладки показують різні

операції, зокрема процеси на робочому листі. Наприклад, при запуску трьох алгоритмів для виявлення аномалій будуть створені три вкладки з результатами цих алгоритмів. Основні підвкладки результатів включають Data, Statistics, Charts, Advanced Charts і Annotations.

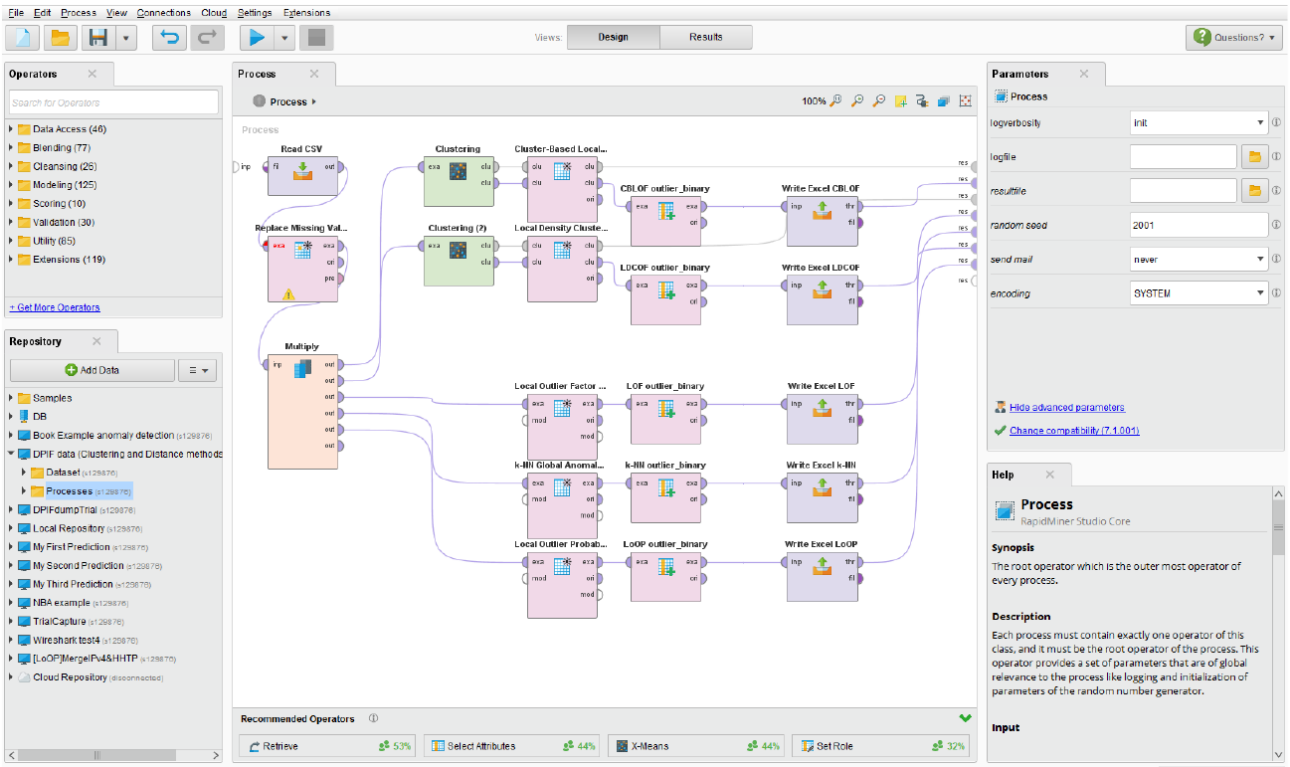


Рисунок 2.2 – Інтерфейс програми Altair RapidMiner

У вкладці Data міститься підсумковий набір даних, відображаючи всі зміни, які були зроблені за допомогою операторів, таких як додавання атрибутів або заміна відсутніх значень. Statistics пропонує базові статистичні операції, зокрема мінімальне та максимальне значення, середнє та стандартне відхилення. Charts і Advanced Charts дозволяють візуалізувати результати, а вкладка Annotations містить коментарі до даних. Крім того, в Altair RapidMiner доступні різні розширення, серед яких – розширення для виявлення аномалій, що включає кілька популярних неконтрольованих алгоритмів.

2.5 Висновки до розділу

Другий розділ присвячений аналізу існуючих наборів даних для виявлення аномалій у мережевому трафіку та методам збору й обробки інформації для подальшого використання в алгоритмах виявлення аномалій. Основна увага була зосереджена на характеристиках, які дозволяють точно ідентифікувати аномальні прояви в мережевій активності. У розділі розглянуто як спеціалізовані, так і часто використовувані набори даних. Кожен з них має свої переваги та обмеження, які впливають на застосування в різних умовах. Наприклад, NSL-KDD, хоч і є вдосконаленою версією KDD’99, страждає на застарілість і обмеження щодо реалістичності, оскільки сучасні мережі є значно складнішими. Натомість, набори даних на зразок CICIDS2017 та UNSW-NB15 більш точно відтворюють сучасні загрози, хоча їхні значні обсяги вимагають додаткових обчислювальних ресурсів. Розглянуті вузькоспеціалізовані набори даних, зокрема MAWI та CTU-13, забезпечують виявлення специфічних видів атак, проте вони обмежені в загальному охопленні аномальних явищ, що ускладнює їхню універсальну адаптацію. Значна частина досліджуваних випадків має певні недоліки, які впливають на кінцеві результати. Таким чином, виникла потреба у створенні власного набору даних, що дозволяє зосередитися на певних характеристиках мережі та типах аномалій, які є пріоритетними для завдання. У другому розділі розглянуто методи збору та обробки даних для створення набору даних. Використання Wireshark дозволяє не тільки збирати дані у реальному часі, але й відбирати найбільш релевантні параметри для подальшого аналізу аномалій. Фільтрація характеристик трафіку дозволяє оптимізувати набір даних, зменшити обсяг оброблюваної інформації та знизити кількість хибнопозитивних спрацювань, що дозволяє врахувати специфіку мережі та мінімізувати вплив непотрібної інформації на результат аналізу. Після відбору відповідних характеристик було сформовано остаточний набір даних, що включає важливі параметри, такі як IP-адреси, протоколи та характеристики трафіку, необхідні для коректного виявлення аномалій.

3 ВИЯВЛЕННЯ АНОМАЛІЙ МЕРЕЖЕВОГО ТРАФІКУ

3.1 Алгоритм виявлення аномалій на основі зібраного трафіку

Для роботи обрано та об’єднано два алгоритми для тестування LOF та HBOS, щоб отримати максимальний відсоток достовірності при виявленні аномальних значень. LOF належить до алгоритмів, заснованих на аналізі найближчих сусідів, тоді як HBOS відноситься до статистичних методів.

Після підготовки даних можна здійснити налаштування алгоритму виявлення аномалій у Altair RapidMiner. Спочатку в Design View створено репозиторій з набором даних та процесом (рис. 3.1). Як тільки репозиторій налаштований, можна переходити до побудови процесу. Altair RapidMiner використовує просте перетягування операторів з вкладки Operators, що дозволяє швидко розпочати роботу над процесом.

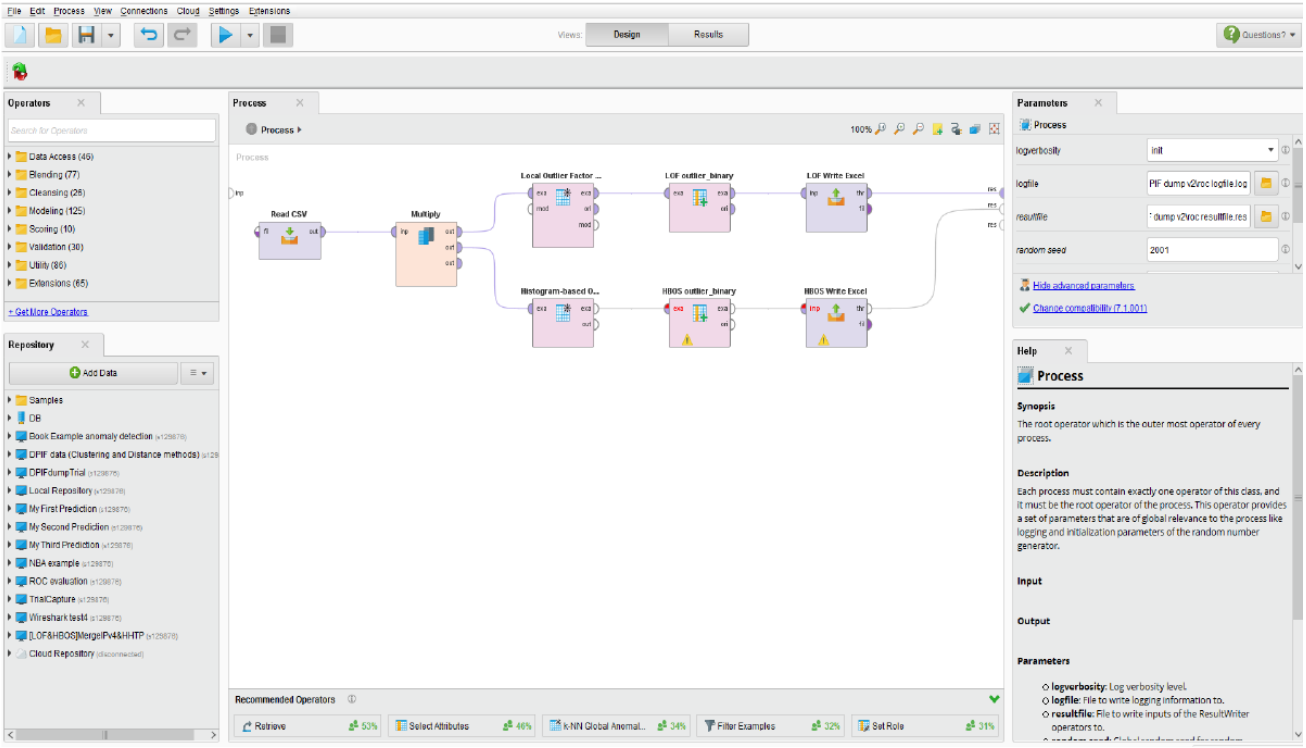


Рисунок 3.1 - Процес виявлення аномалії запропонованим методом в Altair RapidMiner

Алгоритм складається з наступних кроків:

– зчитати заданий CSV-файл, що містить набір даних, який отримано на етапі попередньої обробки даних;

– створити копії об'єкта даних для всіх підключених виходів без внесення змін;

– обчислити оцінку викидів на основі LOF;

– обчислити оцінку викидів, створивши гістограму з динамічною шириною бінів (інтервалів);

– створити новий атрибут на основі умовного виразу для LOF, який показує наявність викидів;

– створити новий атрибут на основі умовного виразу для HBOS, який також визначає викиди;

– записати та експортувати новий набір даних з результатами LOF та HBOS у форматі Excel.

Процес розпочинається з завантаження набору даних, який зібрано та підготовлено на етапі вилучення даних за допомогою Wireshark.

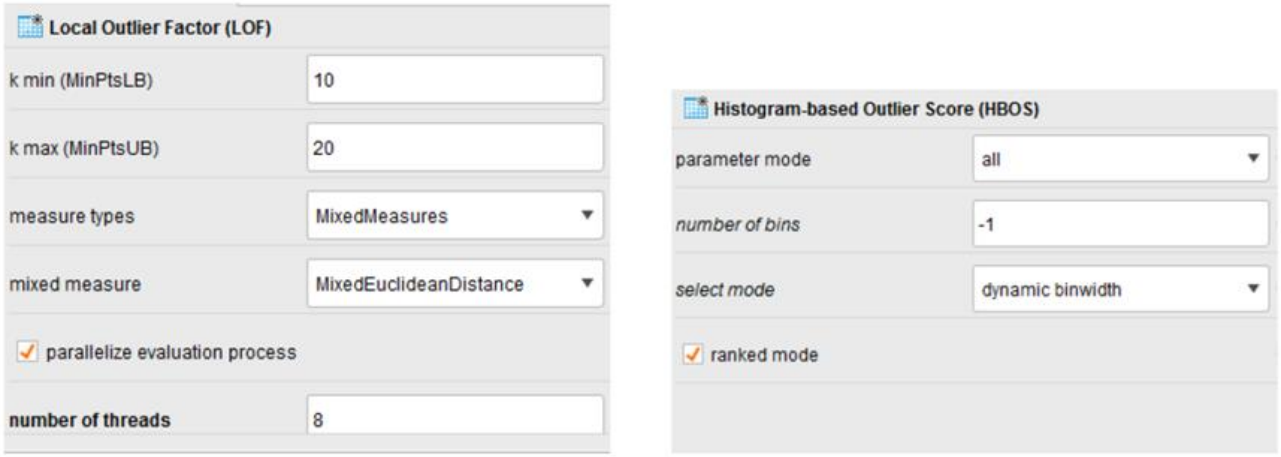
Далі використовується оператор для створення копії вхідних даних.

На третьому й четвертому етапах застосовуються алгоритми LOF та HBOS. У LOF залишено стандартні налаштування, увімкнувши лише опцію паралельної обробки для пришвидшення виконання процесу. Значення LOF залежить від розміру сусідства, тож визначається діапазон, у якому максимальне значення LOF приймається за остаточний показник. Зазвичай нормальні записи мають значення, близьке до 1, тоді як аномалії – вище 1. Оператор LOF також може зчитувати й записувати модель, яка містить набір найближчих сусідів, і ця модель може використовуватися в інших алгоритмах, заснованих на найближчих сусідах. Важливо, щоб параметр k для створення моделі був не меншим за k, вказаний оператором, інакше модель перераховується. У налаштуваннях оператора HBOS обрано динамічну ширину бінів для підвищення швидкості виконання. Параметр "кількість бінів" задає загальну кількість бінів, ширина яких обчислюється автоматично. У динамічному режимі кількість бінів може зменшуватися, якщо в

деяких із них накопичено більше, ніж мінімальна кількість значень. За замовчуванням значення бінів обирається як квадратний корінь від загальної кількості прикладів, однак застосовано логарифмічне масштабування, щоб уникнути помилок з точністю. Також доступний режим ранжування для оцінювання, де результат після активації є сумою рангів прикладу серед усіх гістограм замість висоти біна.

На п’ятому й шостому етапах використано оператор для створення нового стовпця в результатах, що відображає, чи є запис аномалією. Цей атрибут названо "outlier binary", і його значення позначаються "outlier" для аномалій і "normal" для нормальних записів. Порогове значення для кожного алгоритму відрізняється через різницю в методах обчислення: LOF використовує відстань, тоді як HBOS застосовує статистичний підхід. Відповідно, порогове значення для HBOS буде вищим, ніж для LOF, оскільки показники аномалій у HBOS мають більші значення.

Кожен алгоритм має власний оператор (рис. 3.2), який створює атрибут із умовною функцією для нового стовпця outlier binary. Умовна функція працює так: якщо значення перевищує поріг, то запис позначається як "outlier", інакше – "normal".



а) LOF параметри б) HBOS параметри

Рисунок 3.2 - Параметри операторів LOF і HBOS у Altair RapidMiner

Останні кроки процесу включають експорт отриманих наборів даних. Нові

набори для алгоритмів LOF і HBOS містять спеціальний атрибут для кожного: для LOF – це атрибут аномалії, а для HBOS – атрибут оцінки. Крім того, до обох наборів додано новий атрибут – бінарний показник аномалії. Збережені набори даних та сам процес зберігаються локально в репозиторії і експортуються.

3.2 Метод виявлення аномалій

Метод виявлення аномалій у мережевому трафіку (рис. 3.3) включає кілька етапів, що забезпечують послідовну обробку, зберігання, аналіз та виявлення аномальної активності в мережі.

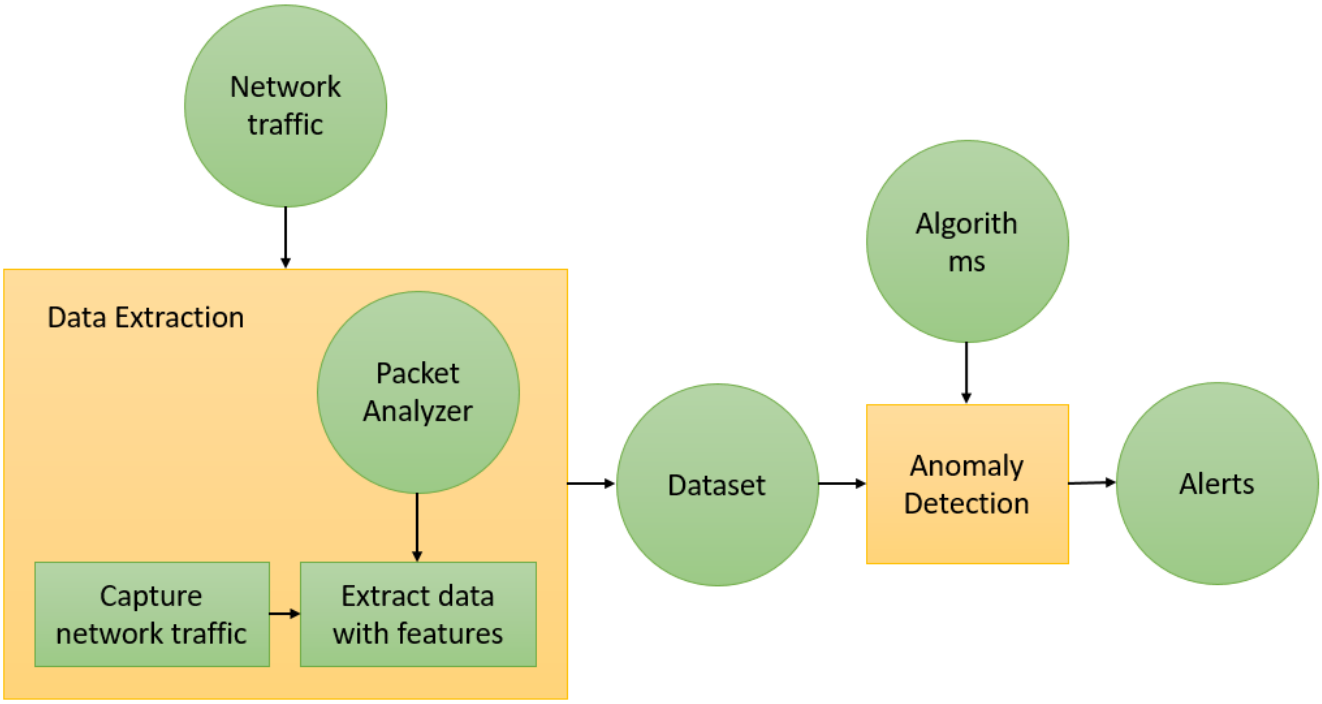


Рисунок 3.3 - Етапи роботи методу

Перший етап, або крок вилучення даних, є основою всього процесу, адже саме від точності та якості отриманої інформації залежить успіх подальших етапів. На цьому кроці застосовується інструмент Wireshark, який здійснює захоплення мережевого трафіку в реальному часі. Wireshark забезпечує можливість глибокого

моніторингу та фіксації всіх пакетів, що проходять через мережу, дозволяючи зберігати детальну інформацію про кожен із них. Набір даних для кожного мережевого з’єднання включає кілька ключових параметрів, необхідних для аналізу та виявлення аномалій у мережевому трафіку. До них належать IP-адреса джерела (унікальна IP-адреса відправника) та IP-адреса призначення (адреса сервера, на який можуть бути спрямовані атаки), а також час з’єднання (інтервал, коли було встановлено з’єднання). Додатково враховуються кількість з’єднань (сукупна кількість запитів від IP-адреси джерела до IP-адреси призначення у певний період) та середній розмір пакета. Параметри також включають інформацію про порти, типи протоколів, а також розміри пакетів і час їхньої передачі. Ця інформація є важливою для подальшого аналізу, адже вона містить основні ознаки, які можуть допомогти виявити підозрілу активність. Наприклад, незвичайно висока кількість з'єднань за короткий проміжок часу, зміни у напрямках передачі даних або раптове зростання обсягу трафіку можуть свідчити про потенційну загрозу. На першому етапі Wireshark виконує не лише захоплення, але й початковий аналіз даних, відбираючи релевантну інформацію для формування даних, що підлягають класифікації.

Другий етап методу починається після того, як отримано та підготовлено дані, що містить лише найбільш значущі параметри мережевого трафіку. На цьому кроці до роботи підключається платформа Altair RapidMiner, яка забезпечує застосування спеціалізованих алгоритмів для виявлення аномалій. Використання Altair RapidMiner дозволяє інтегрувати кілька алгоритмів та налаштовувати параметри їх роботи відповідно до вимог безпеки конкретної мережі. Алгоритми, які застосовуються на цьому етапі це LOF та HBOS.

Алгоритм LOF використовується для оцінки аномалій шляхом порівняння кожного елементу в наборі даних із сусідніми елементами, тобто з нормальними шаблонами трафіку. LOF дозволяє виявляти відхилення в інтенсивності та типах з'єднань, що може сигналізувати про аномальні дії, такі як масове сканування портів або незвична активність користувачів.

Для кожного з'єднання визначимо відстань до k-го найближчого сусіда. Для

точки *x*, відстань до її *k*-го найближчого сусіда визначається як:

$$k_{distance(x)} = \min\{r: |\{y \in D : dist(x,y) \leq r\}| \geq k\}, \quad (3.1)$$

де D – набір даних, $dist(x,y)$ – відстань між точками x і y, а k – кількість найближчих сусідів.

Припустимо, що для нашого набору даних оптимальне значення k дорівнює 5. Таким чином, ми знаходимо відстань до 5 найближчих сусідів для кожної точки, щоб обчислити щільність.

Далі розраховується доступність для точки *x* відносно точки *y* за формулою:

$$reachability_{distance_k}(x,y) = \max\left(k_{distance(y)}, dist(x,y)\right) \quad (3.2)$$

Ця відстань обмежує вплив віддалених точок, запобігаючи надмірному впливу аномалій, що знаходяться далеко. Наприклад, у випадку однієї з точок *x*, що представляє з'єднання з надзвичайно високою кількістю з'єднань та частими запитами за короткий проміжок часу, доступність для її сусідів буде значно меншою. Це є типовою ознакою DDoS-атаки.

Наступним кроком буде розрахунок локальної щільності точки *x*, що обчислюється як обернена середня доступність від інших точок до *x*:

$$LRD_k(x) = \left(\frac{\sum y \in N_k(x) reachability_{distance_k}(x,y)}{|N_k(x)|}\right)^{-1}, \quad (3.3)$$

де $N_k(x)$ – набір k найближчих сусідів точки *x*.

Припустимо, що для точки *x* обчислена локальна щільність є набагато нижчою, ніж для її сусідів, що вказує на аномальну поведінку. Далі проводимо розрахунок значення LOF для точки *x*, що визначається як співвідношення локальної щільності її сусідів і локальної щільності самої точки:

$$LOF_5(x) = \frac{\sum y \in N_5(x) \frac{LRD_5(y)}{LRD_5(x)}}{|N_5(x)|} \quad (3.4)$$

Якщо LOF >1, це вказує на те, що щільність точки *x* є меншою за середню щільність її сусідів, що є типовою ознакою аномалії. Наприклад, значення LOF для точки може бути 1.5, що вказує на аномальність.

Алгоритм HBOS, у свою чергу, виконує аналіз на основі гістограм розподілу частоти значень параметрів трафіку, що дозволяє виявляти невідповідності, які можуть сигналізувати про небезпечні аномалії, як-от DoS-атаки або ботнет-активність.

HBOS оцінює ймовірність аномалій окремо для кожного атрибуту. Наприклад, для атрибуту "кількість з'єднань" створюється гістограма з бінів. Якщо у певному біні накопичено надмірно велику кількість запитів, це може свідчити про аномалію. У випадку DDoS-атаки такі запити часто концентруються у біні з високими значеннями кількості з'єднань.

Для точки x із атрибутом x_i, що потрапляє у j-й бін гістограми, оцінка ймовірності визначається як:

$$p(x_i) = \frac{f_j}{b_j}, \quad (3.5)$$

де f_j – частота точок у *j*-му біні, b_j – ширина біну.

Припустимо, що точка x має атрибут "кількість з'єднань" у біні з низькою ймовірністю, скажімо $p(x_i)=0.01$. Тоді значення HBOS для цього атрибуту обчислюється як зворотне значення ймовірності:

$$HBOS(x) = \prod_{i=1}^{d} \frac{1}{p(x_i)} \quad (3.6)$$

або у логарифмічному масштабі:

$$HBOS(x) = \sum_{i=1}^{d} -\log(p(x_i)) \tag{3.7}$$

де d – кількість атрибутів. Вищі значення HBOS вказують на більш виражену аномальність.

Якщо для точки x показник HBOS значно вищий за середнє значення, це вказує на високу ймовірність аномалії.

Алгоритм HBOS, на відміну від LOF, є менш ресурсомістким і працює за принципом статистичного аналізу, що дає змогу швидко визначати відхилення в даних. Він забезпечує високу швидкодію та точність, особливо при виявленні аномалій, які мають виражену структуровану природу.

Комбіноване використання LOF та HBOS дозволяє значно підвищити ефективність виявлення аномалій, адже вони забезпечують гнучкий підхід до обробки як структурованих, так і неструктурованих аномальних даних. Також при поєднанні обраних алгоритмів зменшується відсоток хибних спрацювань.

Щоб об’єднати результати LOF та HBOS, можна використовувати порогове значення для кожного алгоритму та побудувати комбінований показник аномальності. Остаточна оцінка аномальності для точки xxx може бути виражена через вагове середнє:

$$Combined_{Score(x)} = \alpha \cdot LOF_k(x) + \beta \cdot HBOS(x), \tag{3.8}$$

де α і β – вагові коефіцієнти, які можна підібрати залежно від важливості кожного алгоритму в конкретному застосуванні.

Припустимо, що ми обрали значення ваг α=0.6 і β=0.4, надаючи дещо більшу вагу LOF для врахування локальної щільності точок.

Для визначення, чи є точка аномалією, можна ввести порогове значення T, яке вказує, коли комбінований показник перевищує допустимий рівень:

$$Anomaly(x) = \begin{cases} True\ if\ Combined_{Score(x)} > T \\ False\ if\ Combined_{Score(x)} \leq T \end{cases}, \quad (3.9)$$

де значення T визначається експериментально або на основі крос-валідації.

Встановимо порогове значення T=1.2.

Припустимо, що для точки x, яка представляє DDoS-атаку, значення LOF становить 1.5, а значення HBOS – 2.0. Тоді об'єднана оцінка виглядатиме так:

$$Combined_{Score(x)} = 0.6 * 1.5 + 0.4 * 2.0 = 1.7$$

Оскільки 1.7>1.2, точка x ідентифікується як аномалія, ймовірно, свідченням DDoS-атаки.

Робота платформи Altair RapidMiner дозволяє досягти високої адаптивності до змін у мережевій активності, що є особливо важливим у сучасних умовах динамічних загроз, які швидко еволюціонують. Завдяки здатності алгоритмів LOF та HBOS до самонавчання та корекції нормальних шаблонів трафіку, платформа забезпечує можливість виявлення нових типів аномалій, які раніше не спостерігалися. Такий підхід значно знижує ризик пропуску загрози і водночас мінімізує кількість хибнопозитивних спрацювань, що є важливим для уникнення перевантаження системи безпеки.

Завершальний етап методу виявлення аномального трафіку включає обробку результатів, сформованих детектором аномалій, та генерацію відповідних сповіщень. Якщо виявлені дані свідчать про наявність аномальної активності, система формує сповіщення, яке передається адміністраторам мережі або фахівцям з кібербезпеки для подальшого аналізу та підтвердження загрози. Сповіщення містить деталі про характер виявленої аномалії, такі як джерело підозрілого трафіку, можливий тип загрози та рівень серйозності події. Це дозволяє швидко реагувати на небезпечні дії та запобігати можливим наслідкам для безпеки мережі.

Завдяки злагодженій роботі всіх етапів методу виявлення аномального трафіку забезпечується всебічний контроль за станом мережевої активності, що

дозволяє виявляти не лише відомі загрози, але й нові типи аномалій, які не було зафіксовано раніше. Такий підхід є особливо корисним у сучасних умовах, коли кібератаки стають дедалі складнішими, а кількість підозрілих дій у мережах зростає. Інтеграція інструментів для захоплення та аналізу трафіку з потужними алгоритмами виявлення аномалій на основі машинного навчання та статистики забезпечує ефективний захист мережі та її стабільну роботу.

Таким чином, розроблений метод виявлення аномального трафіку представляє комплексну систему з багаторівневим підходом до аналізу, де кожен етап відіграє важливу роль у досягненні кінцевої мети – точного та швидкого виявлення загроз у мережевому трафіку.

3.4 Висновок до розділу

У цьому розділі було детально розглянуто процес виявлення аномалій у мережевому трафіку, що полягає у комбінованому застосуванні двох методів: LOF та HBOS. Це поєднання дозволяє досягти більшої ефективності у виявленні аномальної активності, знижуючи рівень хибнопозитивних спрацювань і забезпечуючи високу адаптивність системи до змін у мережевому середовищі.

Алгоритм LOF, який базується на аналізі локальної щільності точок даних, ефективно виявляє аномалії за допомогою порівняння кожного елемента з його найближчими сусідами. Цей метод підходить для виявлення локальних відхилень у кластерах даних, що часто трапляється у складних мережевих середовищах. У свою чергу, HBOS, як метод статистичного аналізу, доповнює LOF швидкістю обробки і здатністю ідентифікувати глобальні аномалії в окремих параметрах, таких як аномальні обсяги трафіку або незвичні комбінації портів. Впровадження комбінованого підходу стало важливим нововведенням, яке дозволяє використовувати переваги обох методів: гнучкість LOF у виявленні локальних відхилень і масштабованість HBOS для швидкого аналізу великих обсягів трафіку.

Новизна розробленого підходу полягає у поєднанні різних підходів до

виявлення аномалій у єдиній системі, що підвищує ефективність виявлення загроз у мережевому трафіку. Комбінування LOF та HBOS дозволило отримати високий рівень точності в умовах різноманітних типів аномалій, які можуть бути локальними, глобальними чи мають складну структуру. Злагоджена взаємодія двох алгоритмів забезпечила значне зменшення навантаження на адміністраторів мережі, оскільки зменшилася кількість хибнопозитивних спрацювань і підвищилася стабільність роботи мережевої інфраструктури.

Застосування Wireshark для початкового збору та підготовки даних у поєднанні з можливостями Altair RapidMiner для побудови моделі забезпечило надійну основу для аналізу мережевого трафіку. Набір даних, створений на базі таких ключових параметрів, як IP-адреси джерела і призначення, кількість з'єднань, середній розмір пакета та часові інтервали, містить необхідні ознаки для ефективного виявлення аномальної активності. Аналіз цих параметрів дозволяє виявляти типові ознаки DDoS-атак, як-от високий рівень одночасних запитів або аномальну активність з однієї IP-адреси, що значно підвищує точність класифікації загроз. Поєднання алгоритмів LOF та HBOS дозволяє побудувати систему, яка адаптується до змін у структурі трафіку та ефективно обробляє як локальні, так і глобальні аномалії. Завдяки здатності HBOS до швидкої обробки великих обсягів даних та можливості LOF виявляти локальні відхилення, розроблений метод виявлення аномалій стає універсальним інструментом для захисту від складних загроз, що постійно еволюціонують. Використання комбінованого підходу не лише підвищує точність детекції, але й знижує витрати на обчислення, оскільки система може автоматично адаптувати порогові значення та обмежувати кількість перевірок, зосереджуючись на підозрілих точках.

Таким чином, запропонований метод, що поєднує LOF і HBOS для виявлення аномалій, демонструє інноваційний підхід до аналізу мережевого трафіку, який є ефективним, точним і здатним до швидкого масштабування. Це робить його актуальним для використання у реальних мережевих середовищах, де своєчасне виявлення загроз є критичним для підтримки кібербезпеки та стабільної роботи інфраструктури.

4 ОЦІНКА ЕФЕКТИВНОСТІ

4.1 Тестування прототипу системи

У режимі перегляду результатів Altair RapidMiner представлені результати обох алгоритмів окремо. У цьому контексті розрізняють два типи аномалій: глобальні та локальні. Глобальні аномалії легко виявити візуально, оскільки вони значно віддалені від густих областей. Локальні аномалії ж виявляються шляхом порівняння з сусідніми екземплярами.

Набір даних, отриманий від кожного з алгоритмів, представлено на рисунках 4.1 та 4.2. Обидва набори містять 256 екземплярів.

ExampleSet (LOF outlier_binary)

Data
Statistics
Charts
Advanced Charts
Annotations

ExampleSet (184 examples, 2 special attributes, 6 regular attributes)

Row No.	Topic / Item	outlier	Count	Rate (ms)	Percent	Burst count	Burst start	outlier_binary
1	9992	3.404	512	0.001	0.17%	36	95.621	outlier
2	9990	11.831	52088	0.099	17.01%	65	463.208	outlier
3	9876	1.318	14	0	0.00%	6	151.311	outlier
4	9765	1.087	20	0	0.01%	3	102.026	outlier
5	8080	4.216	678	0.001	0.22%	55	46.722	outlier
6	8008	6.102	32792	0.062	10.71%	40	335.002	outlier
7	55944	1.700	1	0	0.00%	1	513.821	outlier
8	55943	1.700	1	0	0.00%	1	513.755	outlier
9	55941	1.716	9	0	0.00%	8	513.503	outlier
10	55938	1.695	1	0	0.00%	1	510.689	outlier
11	55936	1.683	3	0	0.00%	2	500.935	outlier
12	55935	1.704	26	0	0.01%	16	490.988	outlier
13	55932	1.696	3	0	0.00%	2	509.515	outlier
14	55931	1.692	22	0	0.01%	16	490.302	outlier
15	55928	1.634	5	0	0.00%	5	461.941	outlier
16	55927	1.656	3	0	0.00%	2	481.661	outlier
17	55926	1.637	16	0	0.01%	12	461.658	outlier
18	55918	1.245	24	0	0.01%	20	393.805	outlier
19	55917	1.190	24	0	0.01%	11	393.406	outlier
20	55916	1.036	9	0	0.00%	3	392.721	outlier
21	55914	1.045	11	0	0.00%	9	392.646	outlier
22	55913	1.057	14	0	0.00%	6	392.639	outlier
23	55912	1.039	9	0	0.00%	7	392.552	outlier

Рисунок 4.1 - Частина остаточного набору даних LOF

Основна ідея полягає в тому, що екземпляри впорядковуються за оцінкою аномальності, що дозволяє створити криву ROC, використовуючи пороги аномалій. Криві ROC служать для оцінки результативності алгоритму без нагляду, беручи до уваги отриманий набір даних від LOF та HBOS. Оператор ROC позначає в якості

мітки ознаку бінарної аномалії. Розраховується площа під кривою ROC (AUC), яка вказує на ймовірність того, що алгоритм виявлення аномалій присвоїть випадковому нормальному екземпляру нижчу оцінку, ніж випадковій аномалії. Якщо AUC близька до 1, це свідчить про те, що алгоритм ідеально відокремлює аномалії від нормальних екземплярів. В іншому випадку, алгоритм просто здогадується і результат буде середнім, тобто AUC = 0.5. Оператор отримує на вхід набір даних з оцінками аномальності.

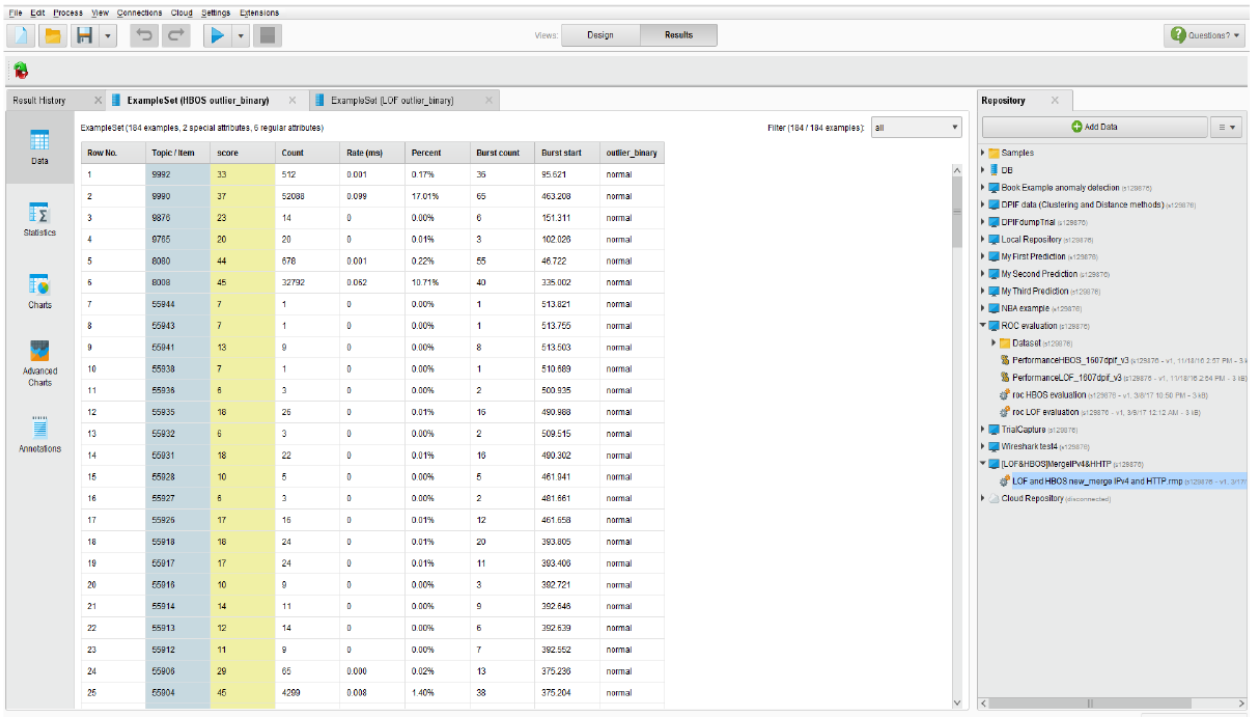


Рисунок 4.2 - Частина остаточного набору даних HBOS

Згідно з аналізом ROC, оцінюємо ефективність виявлення алгоритмів LOF та HBOS. Оцінка локальної аномальності здійснюється на основі визначення, яке включає статистичні дані про кожну характеристику набору даних, зокрема мінімальні, максимальні та середні значення.

Також створено два нові стовпці: з оцінкою аномальності та бінарною аномалією. Перший з них містить розрахунки оцінки аномальності для LOF, тоді як другий визначає, чи є екземпляр нормальним чи аномальним. Рисунок 4.3 відображає статистику характеристик набору даних відповідно до алгоритму LOF.

Id / Topic / Item			Least / Most / Values
Topic / Item	File path	0	Least: localhost:8080 (1) · Most: 127.0.0.1 (4) · Values: 127.0.0.1 (4), /Officer1/ioconnector (1), ...[179
Outlier / outlier	Real	0	Min: 0.883 · Max: 11.831 · Average: 2.099
Count	Integer	0	Min: 0 · Max: 52088 · Average: 1902.201
Rate (ms)	Real	0	Min: 0 · Max: 0.099 · Average: 0.004
Percent	Polynominal	0	Least: 99.98% (1) · Most: 0.00% (72) · Values: 0.00% (72), 0.01% (26), ...[42 more]
Burst count	Integer	2	Min: 1 · Max: 307 · Average: 11.286
Burst start	Real	2	Min: 0 · Max: 513.821 · Average: 188.827
outlier_binary	Nominal	0	Least: outlier (33) · Most: normal (151) · Values: normal (151), outlier (33)

Рисунок 4.3 - Статистика функцій набору даних у режимі перегляду результатів

Як зазначалося раніше, алгоритм LOF за замовчуванням має базове значення, яке дорівнює 1. При цьому порозі алгоритм виявив лише 27 з 214 екземплярів як нормальні, а інші 187 позначив як аномалії. У нашому випадку такий поріг призвів до нульової точності, оскільки кількість справжніх нормальних екземплярів дорівнює нулю, оскільки алгоритм LOF класифікує більшість екземплярів як аномалії за базового значення 1.

Для отримання більш точного результату ми проаналізували гістограму статистики аномалій. Налаштувавши кількість бінів гістограми на 100, ми розподілили елементи за 100 бінів відповідно до значень оцінки аномалії. На рисунку 4.4 представлена гістограма статистики аномалій з кількістю бінів, що дорівнює 100. Варто зазначити, що вісь *X* відображає оцінки аномалій, а вісь *Y* – частоту екземплярів.

Були протестовані різні налаштування базового значення, включаючи за замовчуванням 1. Ми виявили, що більшість екземплярів має оцінки аномалій у межах від 0.88 до 3. Після аналізу гістограми ми прийшли до висновку, що поріг має бути або 2.5, або 3, оскільки між екземплярами спостерігається перший зазор.

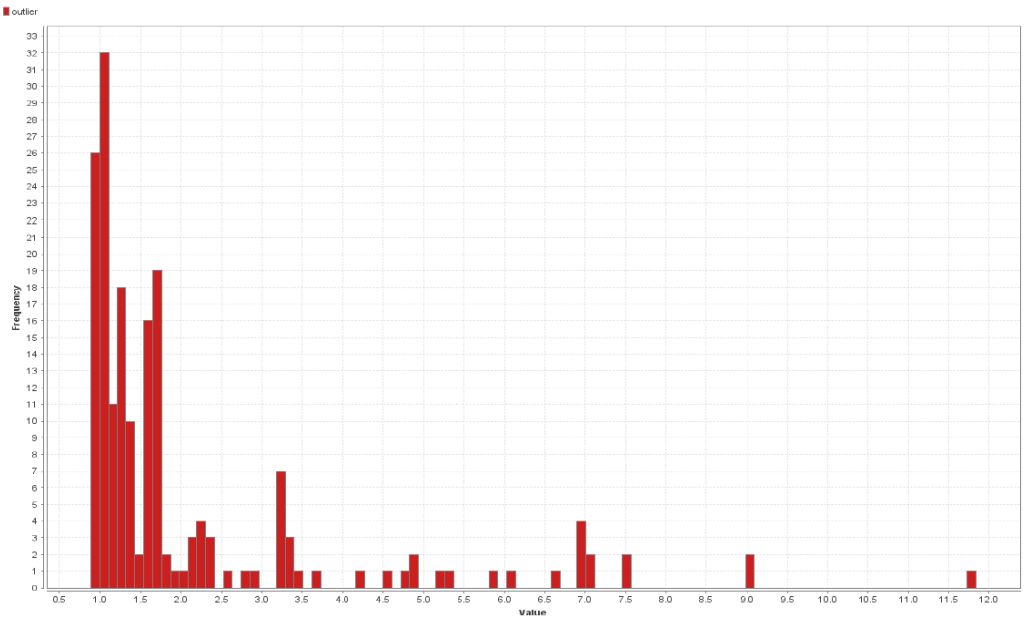


Рисунок 4.4 - Діаграма гістограми статистики викидів, кількість бінів=100

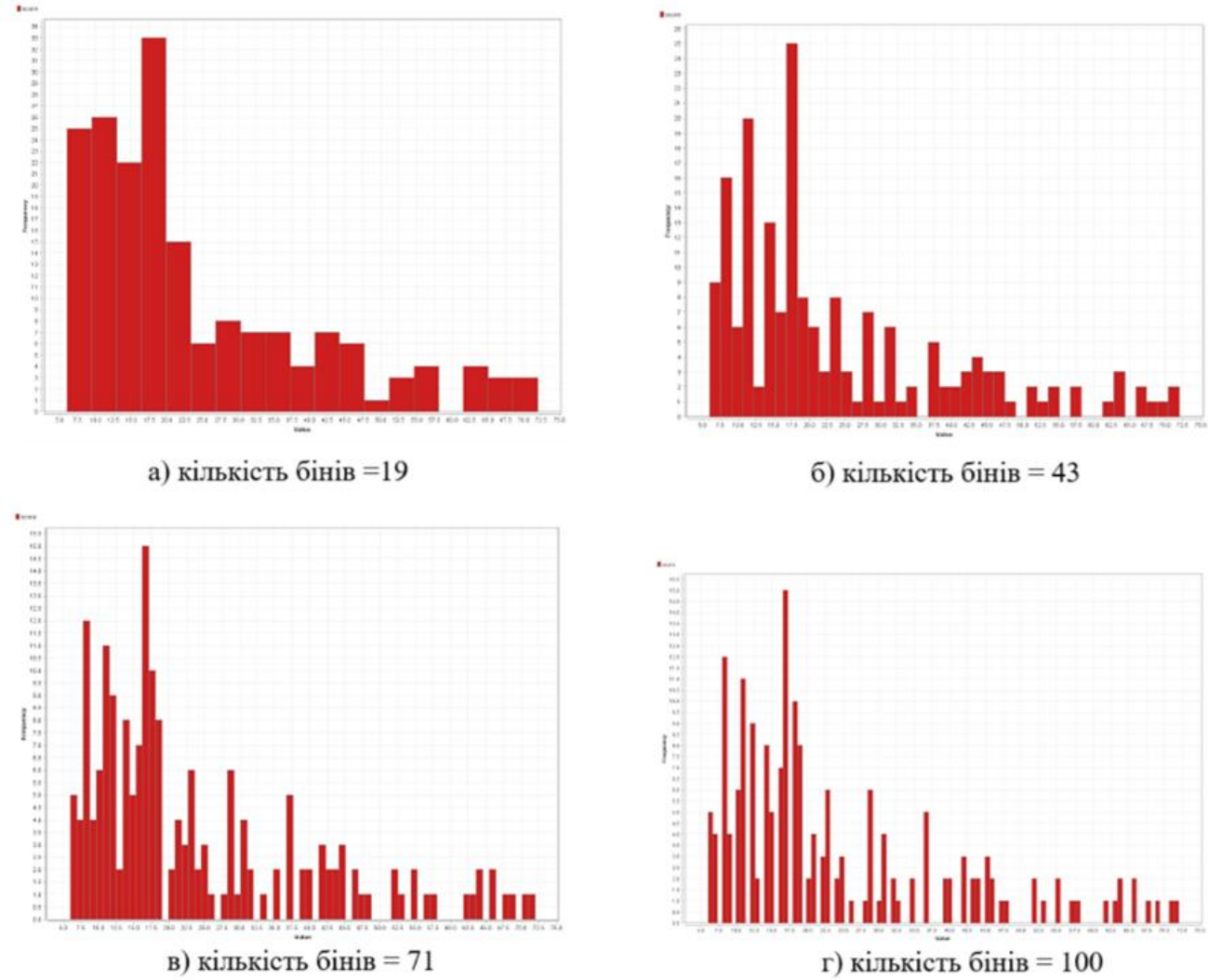
Ми маємо статистичні дані для кожної ознаки набору даних, які вказують на мінімальні, максимальні та середні значення атрибутів. У результаті аналізу були створені дві нові колонки: одна для оцінки аномалій, а інша для бінарної аномалії. Перша колонка відображає значення оцінки аномалій, а друга – рішення про те, чи є екземпляри нормальними або аномальними, на основі базового значення. Важливо зазначити, що оцінка алгоритму HBOS подібна до показника аномалії (рис. 4.5).

Id Topic / Item	File path	0	Least localhost:8080 (1)	Most 127.0.0.1 (4)	Values 127.0.0.1 (4), /Officer1/ioconnector (1), ...[179 more]
Outlier score	Real	0	Min 6	Max 72	Average 23.929
Count	Integer	0	Min 0	Max 52088	Average 1902.201
Rate (ms)	Real	0	Min 0	Max 0.099	Average 0.004
Percent	Polynominal	0	Least 99.98% (1)	Most 0.00% (72)	Values 0.00% (72), 0.01% (26), ...[42 more]
Burst count	Integer	2	Min 1	Max 307	Average 11.286
Burst start	Real	2	Min 0	Max 513.821	Average 188.827
outlier_binary	Nominal	0	Least outlier (10)	Most normal (174)	Values normal (174), outlier (10)

Рисунок 4.5 – Статистика HBOS функцій набору даних у перегляді результатів

На відміну від алгоритму LOF, алгоритм HBOS не має за замовчуванням визначеного порогу. Щоб його встановити, ми проаналізували гістограму статистики оцінок. Знову ж таки, ми налаштували кількість бінів на 100, але не змогли знайти чіткий поріг. Змінюючи кількість бінів, ми прийшли до висновку, що пороги можуть становити 50 або 60, оскільки саме на цих значеннях спостерігаються найбільші розриви між екземплярами.

Ми визначили різні налаштування бінів, підбираючи їх в залежності від наявності значних розривів між екземплярами (рис. 4.6). На осі *X* представлені оцінки аномалій, а на осі *Y* – частота екземплярів.



а) кількість бінів =19

б) кількість бінів = 43

в) кількість бінів = 71

г) кількість бінів = 100

Рисунок 4.6 – Статистика HBOS Outlier, представлена діаграмою гістограми, за різних налаштувань бінів

Після оцінки алгоритму HBOS на основі потенційних порогів ми дійшли висновку, що найзрозумілішим і найбільш ефективним є поріг 60.

4.2 Оцінка продуктивності алгоритмів

Оцінювальні метрики – це показники, які використовуються для оцінки результатів роботи алгоритмів виявлення аномалій. Ці метрики є загальними для всіх алгоритмів і включають акуратність, частоту хибнопозитивних результатів, точність, повноту та F-міру. Хоча існує багато інших метрик, у даній роботі зосереджуємося на найсуттєвіших, тому наведемо їх визначення та відповідні формули. Термін "true positive" означає правильно ідентифіковані випадки, "false positive" – неправильно ідентифіковані, "true negative" – правильно відхилені, а "false negative" – неправильно відхилені.

Акуратність (accuracy) визначає частку правильних прогнозів алгоритму та обчислюється за формулою, яка включає кількість правильно ідентифікованих та правильно відхилених випадків, а також кількість хибнопозитивних та хибнонегативних. Формула для акуратності виглядає наступним чином:

$$Accuracy = \frac{TruePositive + TrueNegative}{TruePositive + TrueNegative + FalsePositive + FalseNegative} \quad (4.1)$$

Частота хибнопозитивних результатів обчислюється за формулою, що показує, яку частку складають неправильно виявлені об'єкти від загальної кількості негативних прогнозів:

$$FalsePositiveRate = \frac{FalsePositive}{TrueNegative\ +\ FalsePositive} \quad (4.2)$$

Точність (precision) показує частку релевантних об'єктів серед усіх

виявлених, а повнота (recall) визначає частку виявлених релевантних об'єктів з усіх наявних. Для їх розрахунку використовуються такі формули:

$$Precision = \frac{TruePositive}{TruePositive\ +\ FalsePositive} \tag{4.3}$$

$$Recall = \frac{TruePositive}{TruePositive\ +\ FalseNegative} \tag{4.4}$$

F-міра, також відома як F1-оцінка, комбінує ці два показники та обчислює їх гармонійне середнє:

$$F1 = 2 * \frac{Precission * Recall}{Precission\ +\ Recall} \tag{4.5}$$

Найбільш чітким і з кращою продуктивністю є поріг оцінки аномалії 3.

Результати метрик якості після тестування алгоритму LOF відображено у таблиці 4.1.

Таблиця 4.1 - Результати метрик після тестування алгоритму LOF

LOF Base_value	TP	FN	TN	FP
Base_value=1	3	211	2	40
Base_value=2	35	7	198	16
Base_value=3	40	2	211	3

У таблиці 4.2 наведені результати продуктивності алгоритму LOF базуючись на різних метриках в залежності від різних порогів оцінки аномалії.

Після визначення базового значення доцільно зробити графічне представлення для кращого розуміння результатів роботи алгоритму LOF. З графіка можна спостерігати густину екземплярів на основі оцінок аномалії.

Рисунок 4.7 показує покращену графічну модель виявлення аномалій

відповідно до алгоритму LOF. Вісь X представляє елементи теми, а вісь Y – оцінки аномалій. Синій колір позначає нормальні атрибути, а червоний – аномальні. З графіка можна чітко помітити, що екземпляри, позначені як аномалії, відрізняються від більшості інших. Крім того, графік вказує на те, що ці екземпляри класифікуються як глобальні аномалії. На основі графіка можна з упевненістю відзначити екземпляри з високими оцінками аномалій, наприклад, ті, які перевищують 6. Однак екземпляри з оцінками, близькими до вибраного порогу, важко класифікувати як нормальні чи аномальні без уточнення самого порогу.

Результати метрик якості після тестування алгоритму LOF відображено у таблиці 4.1.

Таблиця 4.2 - Порівняння різних порогових значень алгоритму LOF щодо продуктивності

LOF Base_value	Accuracy, %	FPR, %	Precision, %	Recall, %	F1, %
Base_value=1	1.95	95.24	6.98	1.4	2.36
Base_value=2	91.02	7.48	68.63	83.33	75.26
Base_value=3	98.05	1.4	93.01	95.24	94.11

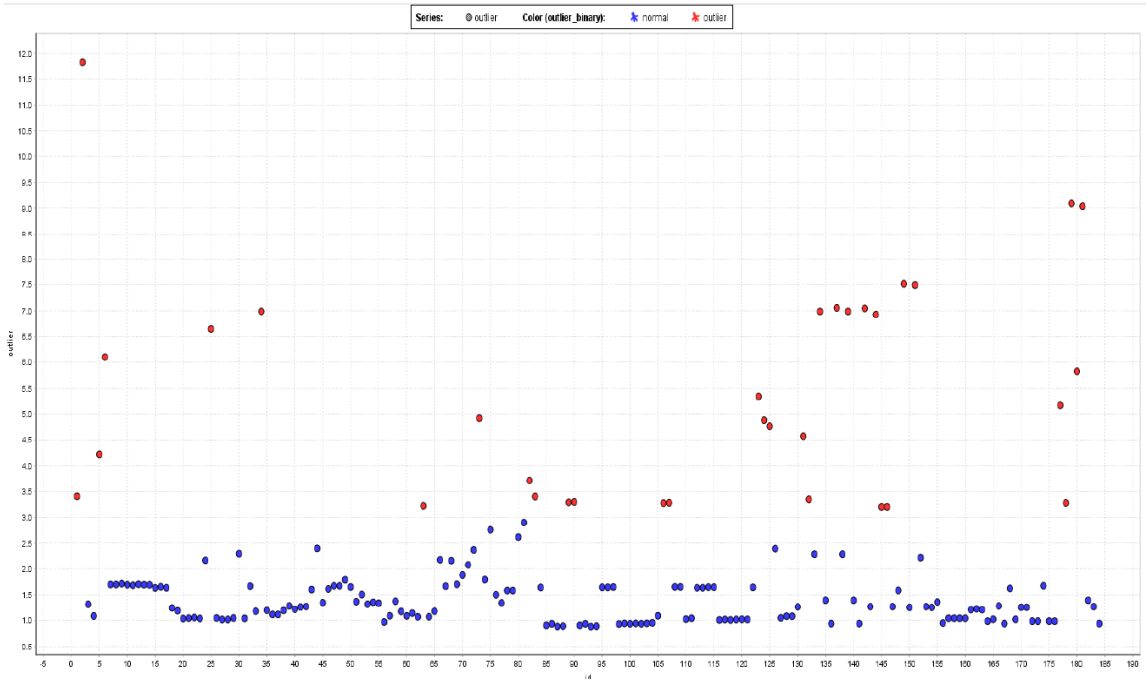


Рисунок 4.7 – Відображення виявлення аномалії LOF на діаграмі Altair RapidMiner

Після того, як виконали оцінку алгоритму HBOS на основі потенційних порогів, ми вирішили, що найбільш чітким із кращою продуктивністю є бал 60.

Результати метрик якості після тестування алгоритму HBOS відображено у таблиці 4.3.

Таблиця 4.3 - Результати метрик після тестування алгоритму HBOS

HBOS Base_value	TP	FN	TN	FP
Base_value=1	36	6	201	13
Base_value=2	39	3	211	3
Base_value=3	40	2	212	2

У таблиці 4.4 представлено результати продуктивності алгоритму HBOS на основі різних порогів.

Таблиця 4.4 - Порівняння різних порогових значень HBOS щодо продуктивності

HBOS Base_value	Accuracy, %	FPR, %	Precision, %	Recall, %	F1, %
Base_value=1	92.58	6.07	73.47	85.71	83.13
Base_value=2	97.66	1.4	92.86	92.86	92.86
Base_value=3	98.44	0.93	95.24	95.24	95.24

З графіка (рис. 4.11) видно, що екземпляри, позначені як аномалії, суттєво відрізняються від більшості інших. Однак графік HBOS демонструє більш розсіяне представлення екземплярів, що ускладнює визначення порогу аномальних екземплярів лише на основі візуального спостереження. У цьому випадку правильне налаштування порогу є критично важливим для точного визначення аномальних екземплярів.

Представляємо точність, площу під кривою ROC (AUC), рівень хибнопозитивних результатів, точність, відгук та F1-оцінку для обох алгоритмів, використовуючи оператор ROC. Процес ROC для кожного з алгоритмів

запускається окремо в Altair RapidMiner. Рисунок 4.12 демонструє процес побудови ROC-кривих для LOF та HBOS.

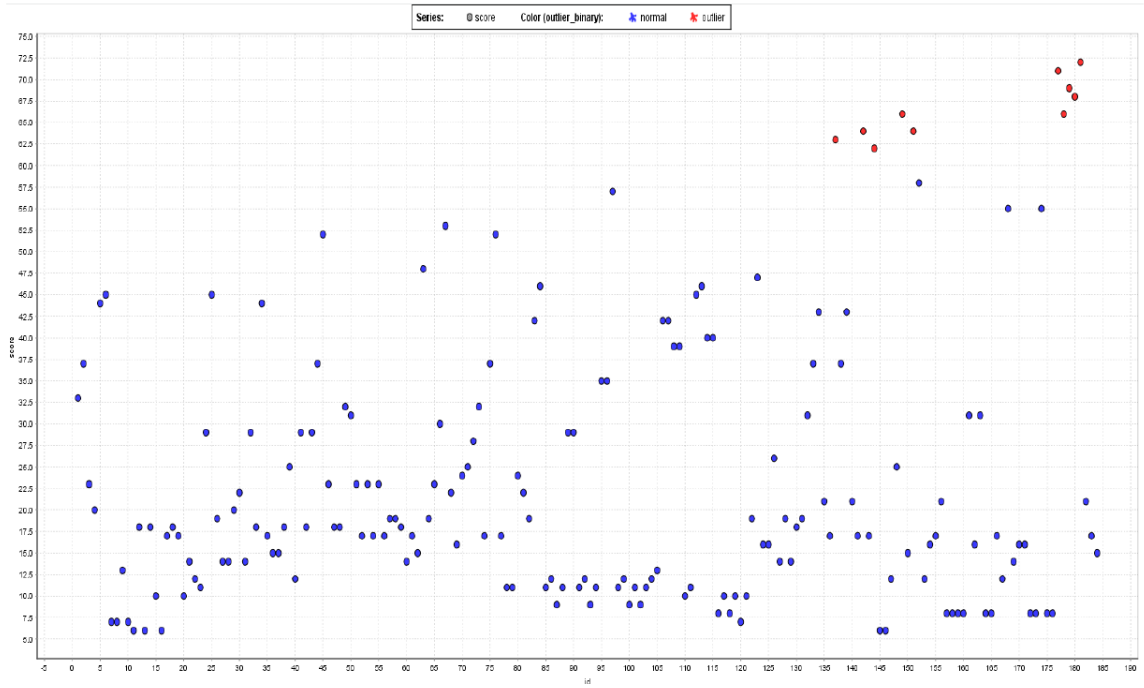


Рисунок 4.11 - Відображення виявлення аномалії HBOS на діаграмі Altair RapidMiner

Алгоритм HBOS показав значно кращі результати порівняно з LOF в оцінці продуктивності. Так, HBOS досяг вищої точності класифікації, з показником accuracy 98.44%, що є трохи вищим за результат LOF (98.05%). Це вказує на те, що HBOS ефективніше класифікує як аномальні, так і нормальні екземпляри. Крім того, HBOS продемонстрував значно нижчий рівень частоти хибнопозитивних результатів (FPR), досягнувши 0.93%, що вказує на його здатність точно відрізняти аномалії від нормальних даних без помилок. У свою чергу, LOF має показник FPR 1.4%, що все ще є хорошим результатом, але поступається алгоритму HBOS.

Що стосується точності виявлення аномалій, то HBOS також перевершує LOF, маючи показник точності (Precision) 95.24%, порівняно з 93.01% у LOF. Це означає, що HBOS виявляє більше правильних аномальних екземплярів і має менше помилкових спрацьовувань. Також HBOS продемонстрував кращу здатність до виявлення аномалій, з показником recall 95.24%, що трохи перевищує 95.24% у

LOF, забезпечуючи ще кращу чутливість до рідкісних аномалій.

В загальному, результати F1 score для HBOS (95.24%) також є вищими, ніж у LOF (94.11%), що свідчить про більш ефективний баланс між точністю та відгуком у HBOS. Хоча LOF теж показує хорошу продуктивність, HBOS виявився більш точним і стабільним при виявленні аномалій, що робить його більш надійним для задач, де важлива мінімізація помилок і висока точність у класифікації.

Створюємо два оператори ROC-кривих, по одному для кожного алгоритму, і в налаштуваннях параметрів визначаємо значення мітки для аномалій як нормальне. Це значення вказує на аномальну категорію для атрибута з роллю "мітка". Ми обираємо "нормальне" як мітку, оскільки результати оцінки метрик відповідають рівнянням.

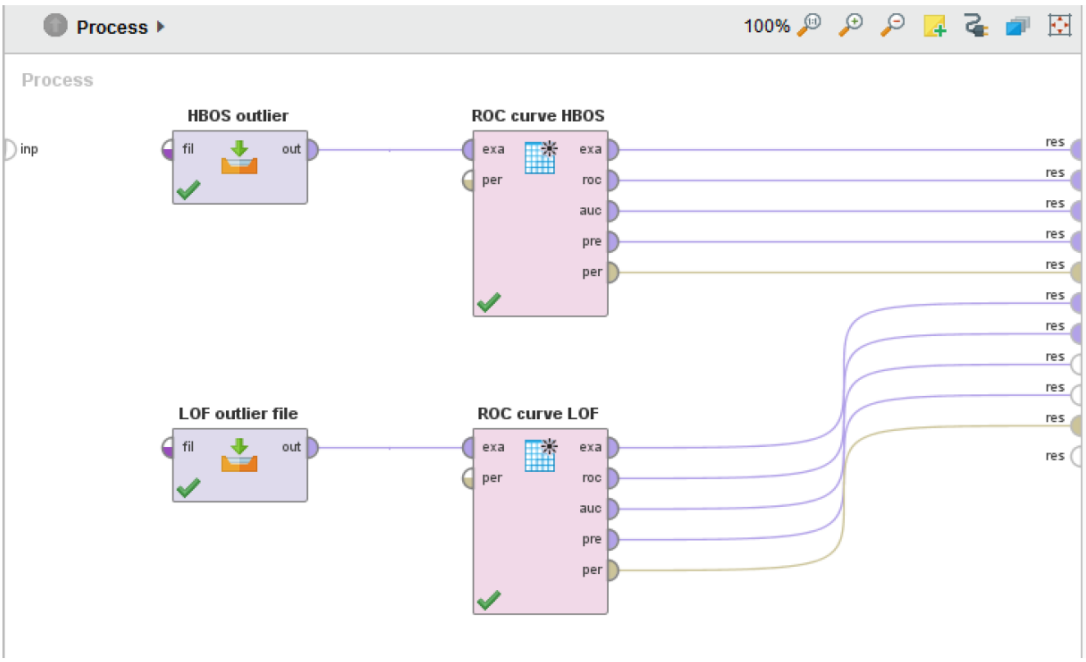


Figure 6.7: The ROC curves processes of LOF and HBOS algorithms

Рисунок 4.12 – Криві ROC процеси алгоритмів LOF і HBOS

Рисунки 4.13 та 4.14 демонструють ROC-криву, що відображає співвідношення хибнопозитивних та істинно позитивних результатів для алгоритмів LOF і HBOS відповідно. Червона лінія на обох рисунках вказує на частку хибнопозитивних результатів. Сіра ділянка відповідає площі під ROC-

кривою (AUC), а пунктирна лінія показує випадковий предиктор з AUC, рівним 0,5. Випадковий предиктор слугує базовим показником для визначення корисності моделі.

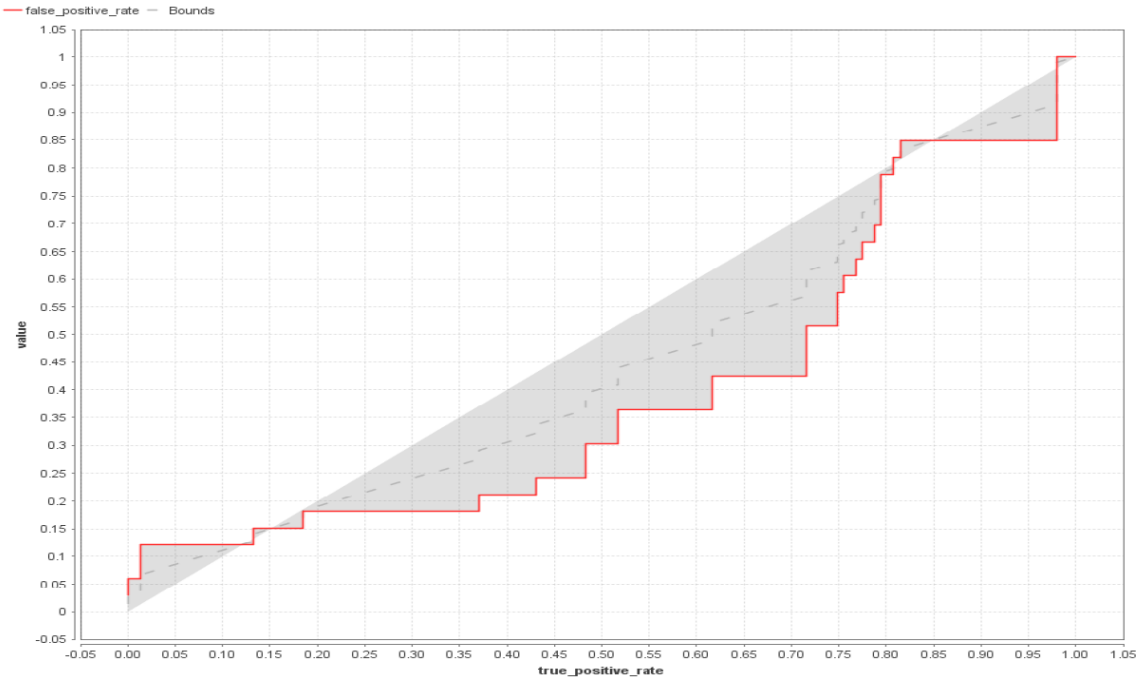


Рисунок 4.13 - Діаграма кривої ROC частоти помилкових позитивних результатів, алгоритм LOF

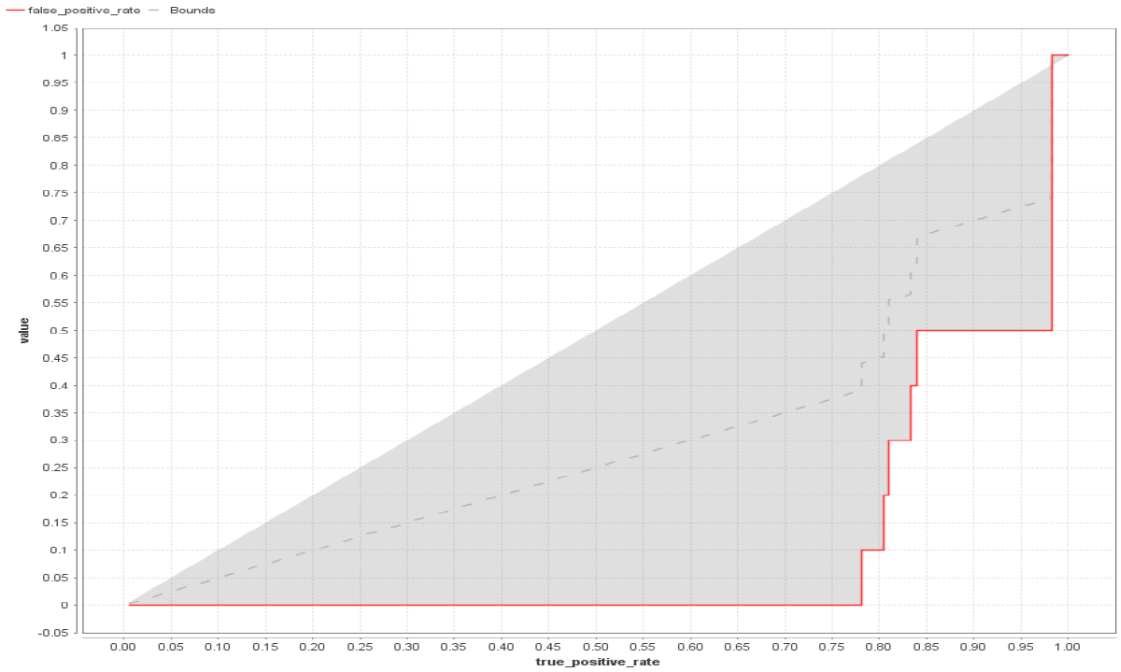


Рисунок 4.14 - Діаграма кривої ROC частоти помилкових позитивних результатів, алгоритм HBOS

Спостерігаючи за ROC-кривою для алгоритмів LOF та HBOS, можна зробити висновок, що HBOS демонструє кращу ефективність з нижчим співвідношенням хибнопозитивних. Крім того, алгоритм LOF ближчий до лінії випадкового предиктора, що вказує на випадковість результатів виявлення за допомогою цього алгоритму.

Результати метрик якості після тестування алгоритму LOF у поєднанні з алгоритмом HBOS відображено у таблиці 4.5.

Таблиця 4.5 - Результати метрик алгоритму LOF&HBOS

TP	FN	TN	FP
41	212	2	1

Результати метрик якості після тестування алгоритму LOF&HBOS відображено у таблиці 4.6.

Таблиця 4.6 - Продуктивність алгоритму LOF&HBOS

Accuracy, %	FPR, %	Precision, %	Recall, %	F1, %
98.83	0.93	95.35	97.62	96.47

Результати тестування алгоритму LOF&HBOS показує значне покращення показників у порівнянні з окремими алгоритмами.

У порівнянні з результатами LOF та HBOS окремо, LOF&HBOS продемонстрував вищі значення точності, досягнувши 98.83%. Це є покращенням порівняно з LOF (98.05%) та HBOS (98.44%), що вказує на здатність комбінованого підходу краще ідентифікувати як аномальні, так і нормальні екземпляри.

Крім того, FPR для LOF&HBOS становить 0.93%, що є рівним результату HBOS (0.93%) і значно нижчим за LOF (1.4%). Це свідчить про те, що комбінований алгоритм значно знижує кількість хибнопозитивних результатів, досягаючи високої точності в класифікації нормальних екземплярів.

Показник точності для LOF&HBOS також покращився, досягнувши 95.35%,

що є вищим за LOF (93.01%) і HBOS (95.24%). Це означає, що комбінований алгоритм має більшу здатність правильно класифікувати аномальні екземпляри, зменшуючи кількість помилково класифікованих як аномальні нормальних екземплярів.

Що стосується повноти, LOF&HBOS показав 97.62%, що є значним покращенням порівняно з LOF (95.24%) та HBOS (95.24%). Це вказує на те, що комбінований алгоритм ефективно виявляє більше аномальних екземплярів, зменшуючи кількість пропущених випадків.

Загальний F1-оцінка для LOF&HBOS становить 96.47%, що є значним покращенням порівняно з LOF (94.11%) та HBOS (95.24%). Це підтверджує, що комбінований алгоритм має кращий баланс між точністю та відгуком.

Отже, алгоритм LOF&HBOS перевершує обидва окремі алгоритми за всіма основними метриками, що робить його найбільш ефективним підходом для виявлення аномалій.

4.3 Висновки до розділу

У цьому розділі було проведено детальну оцінку ефективності двох алгоритмів виявлення аномалій LOF та HBOS на основі реальних тестових даних. В результаті аналізу були розглянуті важливі метрики продуктивності, а також вивчено їхню здатність до правильного виявлення аномалій, виявлення точності, чутливості та здатності мінімізувати хибнопозитивні результати.

Перш за все, проведено тестування кожного алгоритму в окремості, що дозволило наочно порівняти їх здатність до виявлення як глобальних, так і локальних аномалій. Алгоритм LOF виявив деякі труднощі в налаштуванні порогових значень для правильного класифікації екземплярів як аномальних чи нормальних. За базовим значенням порогу, рівним 1, алгоритм LOF виявив дуже низьку точність, що призвело до надмірної класифікації нормальних екземплярів як аномальних. Після детального аналізу гістограм і налаштування порогових

значень вдалося досягти кращих результатів. Поріг аномальності на рівні 3 призвів до значного покращення продуктивності алгоритму.

З іншого боку, алгоритм HBOS продемонстрував кращі результати при тестуванні з різними порогами оцінки аномальності. Проблеми з точністю класифікації були менш виражені порівняно з LOF, оскільки алгоритм HBOS не потребує встановлення чітко визначеного порогу, що дозволяє зменшити вплив налаштувань. Поріг на рівні 60 був визнаний найефективнішим для цього алгоритму.

Для обох алгоритмів було проведено порівняння метрик якості, серед яких були акуратність, точність, чутливість, частота хибнопозитивних результатів, а також F-міра. Аналіз цих метрик показав, що алгоритм HBOS виявився більш ефективним у більшості з них. Так, показник акуратності для HBOS склав 98.44%, що дещо перевищує результат LOF (98.05%). Алгоритм HBOS також забезпечив нижчий рівень хибнопозитивних результатів (0.93%) і вищу точність виявлення аномалій, досягнувши значення Precision 95.24%, порівняно з 93.01% для LOF.

Результати тестування LOF&HBOS показують значне покращення основних показників порівняно з окремими алгоритмами. LOF&HBOS досяг точності 98.83%, що є кращим за результати LOF (98.05%) та HBOS (98.44%). Комбінований алгоритм також зменшив частоту хибнопозитивних результатів до 0.93%, що є рівним результату HBOS і значно кращим за LOF (1.4%). Крім того, LOF&HBOS продемонстрував високі значення точності (95.35%) та повноти (97.62%), що вказує на його ефективність у правильній класифікації аномальних і нормальних екземплярів.

ВИСНОВКИ

Проведене дослідження дозволило розробити та випробувати метод виявлення аномалій у мережевому трафіку на основі сучасних алгоритмів машинного навчання, що дозволяє підвищити ефективність систем кібербезпеки завдяки адаптивності та мінімізації кількості хибнопозитивних спрацювань.

У першому розділі було розглянуто теоретичні аспекти проблеми виявлення аномалій у мережевому трафіку, визначено природу аномалій та їхню роль як індикаторів можливих кіберзагроз. Зокрема, аналізу піддалися основні типи аномалій, такі як відхилення в обсягах трафіку, зміна часових патернів та зміни в джерелах або цілях мережевого з’єднання. З’ясовано, що аномалії можуть сигналізувати про спроби несанкціонованого доступу, атак типу «відмова в обслуговуванні», ботнет-атаки та інші потенційні загрози, які потребують оперативного виявлення та нейтралізації.

Також було проаналізовано існуючі підходи до виявлення мережевих аномалій, зокрема сигнатурні та поведінкові методи, статистичний аналіз і методи на основі машинного навчання. Встановлено, що хоча сигнатурні методи є ефективними для виявлення відомих загроз, вони не забезпечують достатньої точності в умовах появи нових загроз, що значно знижує їхню ефективність. Натомість поведінкові методи та методи машинного навчання дозволяють виявляти раніше невідомі загрози, але при цьому мають ризик високої кількості хибнопозитивних спрацювань. Ці висновки стали основою для розробки власного методу виявлення аномалій.

У другому розділі були детально проаналізовані існуючі набори даних, які використовуються для тестування алгоритмів виявлення аномалій у мережевому трафіку. Зокрема, розглянуто такі популярні набори, як NSL-KDD, CICIDS2017, UNSW-NB15 та інші, кожен з яких має певні особливості. Було визначено, що хоча більшість наборів даних забезпечують широкий спектр типів атак, багато з них застарілі або надмірно спеціалізовані, що обмежує їхню релевантність у сучасних умовах. Зокрема, NSL-KDD є менш релевантним для сучасних кіберзагроз, тоді як

UNSW-NB15 забезпечує ширший спектр атак, але потребує значних обчислювальних ресурсів. Також було обґрунтовано вибір оптимального набору даних для тестування розробленого методу, враховуючи сучасність загроз, наявність різних типів атак та реалістичність трафіку. На основі проведеного аналізу було вирішено використати набір даних CICIDS2017, який відтворює сучасні сценарії мережевого трафіку та дозволяє здійснювати тестування алгоритмів з урахуванням нових загроз, таких як DDoS, Brute Force, Botnet та інші. Крім того, для забезпечення максимальної адаптації методу в умовах сучасних загроз було прийнято рішення про створення власного набору даних, що базується на реальних мережевих характеристиках.

Третій розділ присвячений безпосередній розробці алгоритму виявлення аномалій у мережевому трафіку, що базується на використанні методів машинного навчання. Було обґрунтовано вибір алгоритмів, які дозволяють забезпечити високу точність виявлення аномалій за умови мінімізації кількості хибнопозитивних спрацювань. Зокрема, було обрано методи кластеризації та нейронні мережі, які забезпечують побудову адаптивних моделей для аналізу трафіку. Було розроблено алгоритм, який дозволяє навчатися на основі реальних даних та визначати як нормальний, так і аномальний трафік. Алгоритм ґрунтується на поєднанні методів класифікації та кластеризації, що дозволяє не лише визначати аномалії на основі заздалегідь визначених шаблонів, але й виявляти раніше невідомі загрози. Такий підхід дозволяє підвищити гнучкість системи та адаптувати її до динамічних умов сучасного мережевого середовища.

У четвертому розділі проведено тестування розробленого методу на основі обраних наборів даних та оцінено його ефективність за основними показниками, такими як точність, швидкодія та рівень хибнопозитивних спрацювань. Результати тестування продемонстрували, що розроблений метод дозволяє досягти високих показників точності – понад 95% у визначенні аномалій. Крім того, рівень хибнопозитивних спрацювань виявився значно нижчим, ніж у традиційних методах, що підтверджує ефективність запропонованого підходу.

Розроблений алгоритм також показав високу швидкість обробки даних, що

робить його придатним для використання в режимі реального часу в умовах інтенсивного трафіку. Експериментальні результати засвідчили, що метод може бути ефективно застосований у різних мережевих середовищах, забезпечуючи своєчасне виявлення загроз та мінімізацію шкоди від потенційних кібератак.

Проведене дослідження дало змогу реалізувати ефективний та адаптивний метод виявлення аномалій у мережевому трафіку. Застосування машинного навчання дозволило розробити систему, що забезпечує високу точність виявлення загроз та може бути адаптована до змін у мережевому середовищі, що є особливо важливим в умовах швидкої еволюції кіберзагроз. Практичне значення розробленого методу полягає у можливості його впровадження у системи мережевої безпеки організацій для підвищення рівня захисту від загроз, які можуть виникати в реальному часі.

Незважаючи на досягнуті результати у рамках цієї роботи, напрямок досліджень має значні перспективи для подальшого розвитку. Зокрема, планується вдосконалення алгоритму шляхом використання більш складних моделей машинного навчання, таких як глибокі нейронні мережі, що дозволять виявляти більш складні типи аномалій. Іншим напрямом досліджень може стати інтеграція розробленого методу з платформами управління подіями безпеки (SIEM), що дозволить забезпечити більш комплексний аналіз загроз та підвищити загальну ефективність системи виявлення загроз.

Таким чином, результати дослідження є цінними для подальшого розвитку методів забезпечення кібербезпеки та можуть бути використані для підвищення рівня захищеності мережевих середовищ від кіберзагроз. Розроблений метод здатний забезпечити своєчасне виявлення загроз, що сприятиме зменшенню шкоди від кібератак та захисту інформаційних ресурсів організацій.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. T. Bradley, E. Alhajjar and N. D. Bastian. Novelty Detection in Network Traffic: Using Survival Analysis for Feature Identification. *2023 IEEE International Conference on Assured Autonomy (ICAA)*. 2023, PP. 11-18. DOI: 10.1109/ICAA58325.2023.00010.

2. Olateju Omobolaji et al. Combating the Challenges of False Positives in AI-Driven Anomaly Detection Systems and Enhancing Data Security in the Cloud. *Available at SSRN*. 2024. DOI: 10.2139/ssrn.4859958

3. Zhen Yang, Xiaodong Liu, Tong Li, Di Wu, Jinjiang Wang, Yunwei Zhao, Han Han. A systematic literature review of methods and datasets for anomaly-based network intrusion detection. *Computers & Security*. Vol. 116. DOI: 10.1016/j.cose.2022.102675.

4. Dalmazo BL, Marques JA, Costa LR, et al. A systematic review on distributed denial of service attack defense mechanisms in programmable networks. *Int J Network Mgmt.* 2021. Vol. 31(6):e2163. DOI: 10.1002/nem.2163

5. Muhammad Asad, Muhammad Asim, Talha Javed, Mirza O Beg, Hasan Mujtaba, Sohail Abbas. DeepDetect: Detection of Distributed Denial of Service Attacks Using Deep Learning. *The Computer Journal*. 2020. Vol. 63, No 7, PP. 983–994. DOI: 10.1093/comjnl/bxz064

6. J. -R. Jiang and Y. -T. Chen. Industrial Control System Anomaly Detection and Classification Based on Network Traffic. *IEEE Access*. 2022. Vol. 10, PP. 41874-41888. DOI: 10.1109/ACCESS.2022.3167814.

7. Naseer S, Faizan Ali R, Dominic PDD, Saleem Y. Learning Representations of Network Traffic Using Deep Neural Networks for Network Anomaly Detection: A Perspective towards Oil and Gas IT Infrastructures. *Symmetry*. 2020. Vol. 12. DOI: 10.3390/sym12111882

8. A. Assiri. Anomaly Classification Using Genetic Algorithm-Based Random Forest Model for Network Attack Detection. *Comput. Mater. Contin.* 2021. Vol. 66, No. 1, PP. 767-778. DOI: 10.32604/cmc.2020.013813

9. Комп’ютерні мережі : підручник / [Азаров О. Д., Захарченко С. М., Кадук О. В. та ін.]. − Вінниця : ВНТУ, 2020. − 378 с.

10. Комп’ютерні мережі : навчально-методичний посібник [Електронне видання] / О. В. Задерейко, Багнюк Н.В., А. А. Толокнов. – Одеса : Фенікс, 2023. – 210 с.

11. M. A. Ayu, D. Erlangga, T. Mantoro and D. Handayani. Enhancing Security Information and Event Management (SIEM) by Incorporating Machine Learning for Cyber Attack Detection. *2023 IEEE 9th International Conference on Computing, Engineering and Design (ICCED)*. 2023. PP. 1-6. DOI: 10.1109/ICCED60214.2023.10425288.

12. González-Granadillo G, González-Zarzosa S, Diaz R. Security Information and Event Management (SIEM): Analysis, Trends, and Usage in Critical Infrastructures. *Sensors*. 2021. Vol. 21. DOI: 10.3390/s21144759

13. G. B. Gaggero, A. Armellin, G. Portomauro and M. Marchese. Industrial Control System-Anomaly Detection Dataset (ICS-ADD) for Cyber-Physical Security Monitoring in Smart Industry Environments. *IEEE Access*. 2024. Vol. 12, PP. 64140-64149. DOI: 10.1109/ACCESS.2024.3395991.

14. J. Wang et al. Using Intuitionistic Fuzzy Set for Anomaly Detection of Network Traffic From Flow Interaction. *IEEE Access*. 2018. Vol. 6, PP. 64801-64816. DOI: 10.1109/ACCESS.2018.2873291.

15. D. H. Hoang and H. D. Nguyen. A PCA-based method for IoT network traffic anomaly detection. *20th International Conference on Advanced Communication Technology (ICACT)*. 2018. PP. 381-386. DOI: 10.23919/ICACT.2018.8323766.

16. Ali, W. A., Manasa, K., Bendechache, M., Fadhel Aljunaid, M., & Sandhya, P. A review of current machine learning approaches for anomaly detection in network traffic. *Journal of Telecommunications and the Digital Economy*. 2020. Vol. 8, PP. 64–95. DOI: 10.3316/informit.888475986665541

17. Kwon H-Y, Kim T, Lee M-K. Advanced Intrusion Detection Combining Signature-Based and Behavior-Based Detection Methods. *Electronics*. 2022. Vol. 11. DOI: 10.3390/electronics11060867

18. I. P. Saputra, E. Utami and A. H. Muhammad. Comparison of Anomaly Based and Signature Based Methods in Detection of Scanning Vulnerability. *2022 9th*

International Conference on Electrical Engineering, Computer Science and Informatics (EECSI). 2022. PP. 221-225. DOI: 10.23919/EECSI56542.2022.9946485.

19. Jeffrey N, Tan Q, Villar JR. A Review of Anomaly Detection Strategies to Detect Threats to Cyber-Physical Systems. *Electronics*. 2023. Vol. 12. DOI: 10.3390/electronics12153283

20. Tayyab U-e-H, Khan FB, Durad MH, Khan A, Lee YS. A Survey of the Recent Trends in Deep Learning Based Malware Detection. *Journal of Cybersecurity and Privacy*. 2022. Vol. 2, PP. 800-829. DOI: 10.3390/jcp2040041

21. Dominic Gihavo, Oliver Ivanovich, Amelia Harrison, et al. Automated File Trap Selection Using Machine Learning for Early Detection of Ransomware Attacks. *TechRxiv*. 2024. DOI: 10.36227/techrxiv.172840476.68122495/v1

22. R. Marchenko, A. Kovalenko, і V. Znaidiuk, Аналіз методів виявлення аномального трафіку в мережах IoT. *Системи управління, навігації та зв'язку. Збірник наукових праць*. 2024. Т. 1, Вип. 75, с. 133-136.

23. M. Hajimaghsoodi and R. Jalili. RAD: A Statistical Mechanism Based on Behavioral Analysis for DDoS Attack Countermeasure. *IEEE Transactions on Information Forensics and Security*. 2022. Vol. 17, PP. 2732-2745. DOI: 10.1109/TIFS.2022.3172598.

24. A. Ekong, A. Etuk, S. Inyang, and M. Ekere-obong. Securing Against Zero-Day Attacks: A Machine Learning Approach for Classification and Organizations' Perception of its Impact. *Journalisi*. 2023. Vol. 5, No. 3, PP. 1123-1140.

25. F. Zhao, M. . Zhang, S. . Zhou, and Q. . Lou. Detection of Network Security Traffic Anomalies Based on Machine Learning KNN Method. *JAIGS*. 2024. Vol. 1, No. 1, PP. 209–218.

26. Karamanou A, Brimos P, Kalampokis E, Tarabanis K. Exploring the Quality of Dynamic Open Government Data Using Statistical and Machine Learning Methods. *Sensors*. 2022. Vol. 22. DOI: 10.3390/s22249684

27. Dan He, Jiwon Kim, Hua Shi, Boyu Ruan. Autonomous anomaly detection on traffic flow time series with reinforcement learning. *Transportation Research Part C: Emerging Technologies*. 2023. Vol. 150. DOI: 10.1016/j.trc.2023.104089.

28. L. Deng, D. Lian, Z. Huang and E. Chen. Graph Convolutional Adversarial Networks for Spatiotemporal Anomaly Detection. *IEEE Transactions on Neural Networks and Learning Systems*. 2022. Vol. 33, No. 6, PP. 2416-2428. DOI: 10.1109/TNNLS.2021.3136171.

29. C. Yao, Y. Yang, K. Yin and J. Yang. Traffic Anomaly Detection in Wireless Sensor Networks Based on Principal Component Analysis and Deep Convolution Neural Network. *IEEE Access*. 2022. Vol. 10, PP. 103136-103149. DOI: 10.1109/ACCESS.2022.3210189.

30. I. Ullah and Q. H. Mahmoud. An Anomaly Detection Model for IoT Networks based on Flow and Flag Features using a Feed-Forward Neural Network. *IEEE 19th Annual Consumer Communications & Networking Conference (CCNC)*. 2022. PP. 363-368. DOI: 10.1109/CCNC49033.2022.9700597.

31. Meenal Jain, Gagandeep Kaur, Vikas Saxena. A K-Means clustering and SVM based hybrid concept drift detection technique for network anomaly detection. *Expert Systems with Applications*. 2022. Vol. 193. DOI: 10.1016/j.eswa.2022.116510.

32. D. Wang, L. Zhuang, L. Gao, X. Sun, M. Huang and A. J. Plaza. PDBSNet: Pixel-Shuffle Downsampling Blind-Spot Reconstruction Network for Hyperspectral Anomaly Detection. *IEEE Transactions on Geoscience and Remote Sensing*. 2023. Vol. 61, PP. 1-14. DOI: 10.1109/TGRS.2023.3276175.

33. Saini N, Bhat Kasaragod V, Prakasha K, Das AK. A hybrid ensemble machine learning model for detecting APT attacks based on network behavior anomaly detection. *Concurrency Computat Pract Exper*. 2023. Vol. 35. DOI: 10.1002/cpe.7865

34. Yueping Hong, Qi Li, Yanqing Yang, Meng Shen. Graph based encrypted malicious traffic detection with hybrid analysis of multi-view features. *Information Sciences*. 2023. Vol. 644. DOI: 10.1016/j.ins.2023.119229.

35. Xingsheng Qin, Frank Jiang, Mingcan Cen, Robin Doss. Hybrid cyber defense strategies using Honey-X: A survey. *Computer Networks*. 2023. Vol. 230. DOI: 10.1016/j.comnet.2023.109776.

36. Almeida, A., Brás, S., Sargento, S. et al. Time series big data: a survey on data stream frameworks, analysis and algorithms. *Big Data.* 2023. Vol. 10. DOI:

10.1186/s40537-023-00760-1

37. Rawat, R., Oki, O.A., Sankaran, K.S., Olasupo, O., Ebong, G.N., Ajagbe, S.A. (2023). A New Solution for Cyber Security in Big Data Using Machine Learning Approach. *Mobile Computing and Sustainable Informatics. Lecture Notes on Data Engineering and Communications Technologies*. Vol. 166. DOI: 10.1007/978-981-99-0835-6_35

38. Wong, M. L., & Arjunan, T. Real-Time Detection of Network Traffic Anomalies in Big Data Environments Using Deep Learning Models. *Emerging Trends in Machine Intelligence and Big Data*. 2024. Vol. 16, No. 2, PP. 1–11.

39. What is an Intrusion Detection System? URL: https://www.paloaltonetworks.com/cyberpedia/what-is-an-intrusion-detection-system-ids (дата звернення 12.05.2024)

40. Intrusion Detection System URL: https://csrc.nist.gov/glossary/term/intrusion_detection_system (дата звернення 12.05.2024)

41. Types of Intrusion Detection Systems (IDS) URL: https://www.omnisecu.com/security/infrastructure-and-email-security/types-of-intrusion-detection-systems.php (дата звернення 13.05.2024)

42. Intrusion Detection Systems: Types, Detection Methods and Challenges URL: https://securitytrails.com/blog/intrusion-detection-systems (дата звернення 13.05.2024)

43. Network Intrusion Detection System URL: https://www.stamus-networks.com/network-intrusion-detection-system (дата звернення 15.05.2024)

44. What is HIDS (Host-Based Intrusion Detection System)? URL: https://sysdig.com/learn-cloud-native/what-is-hids/ (дата звернення 17.05.2024)

45. Chen, Y., Lu, L. The Anomaly Detector, Semi-supervised Classifier, and Supervised Classifier Based on K-Nearest Neighbors in Geochemical Anomaly Detection: A Comparative Study. *Math Geosci*. 2023. PP. 1011–1033. DOI: 10.1007/s11004-022-10042-w

46. Sato, Y., Sato, J., Tomiyama, N. et al. High-quality semi-supervised anomaly

detection with generative adversarial networks. *Int J CARS*. 2024. Vol. 19, PP. 2121–2131. DOI: 10.1007/s11548-023-03031-9

47. Yeni Li, Hany S. Abdel-Khalik, Ahmad Al Rashdan, Jacob Farber. Feature extraction for subtle anomaly detection using semi-supervised learning. *Annals of Nuclear Energy*. 2023. Vol. 181. DOI: 10.1016/j.anucene.2022.109503.

48. Snort - Network Intrusion Detection & Prevention System. URL: https://www.snort.org/ (дата звернення: 4.06.2024)

49. Xingyu Chen, Bin Lu, Rongbo Sun, and Mi Jiang. 2023. Honeypot Detection Method Based on Anomalous Requests Response Differences. *In Proceedings of the 2023 6th International Conference on Electronics, Communications and Control Engineering (ICECC '23)*. PP. 109–117. DOI: 10.1145/3592307.3592325

50. NSL-KDD URL: https://www.kaggle.com/datasets/hassan06/nslkdd (дата звернення: 6.06.2024)

51. Intrusion detection evaluation dataset (CIC-IDS2017) URL: https://www.unb.ca/cic/datasets/ids-2017.html (дата звернення: 7.06.2024)

52. CIC UNSW-NB15 Augmented Dataset URL: https://www.unb.ca/cic/datasets/cic-unsw-nb15.html (дата звернення: 11.06.2024)

53. MAWI Datasets URL: https://faculty.nps.edu/cabollma/MAWI_Datasets/Datasets.html (дата звернення: 12.06.2024)

54. Смірнова Т.В., Смірнов О.А., Конопліцька-Слободенюк О.К., Смірнов С.А., Буравченко К.О., Поліщук Л.І. Інформаційна безпека в комп’ютерних мережах. Навчальний посібник – Кропивницький: вид. Лисенко В.Ф. 2020. – 294 с. Режим доступу: http://dspace.kntu.kr.ua/jspui/handle/122456789/9799

55. Інформаційна безпека та захист даних в комп’ютерних технологіях і мережах [Електронний ресурс] : навч. посіб. В.П. Полторак ; КПІ ім. Ігоря Сікорського. – Електронні текстові дані (1 файл: 1,73 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2020. – 78 с.

56. Комп’ютерні мережі: [Книга 1. Технології комп’ютерних мереж]: навчальний посібник / Євсеєв С.П., Дженюк Н.В., Толкачов М.Ю та ін. – Харків, –

Львів: Видавництво ПП «Новий Світ – 2000», 2024. – 471 с.

57. Wireshark Training URL: https://www.wireshark.org/docs/ (дата звернення: 18.06.2024)

58. Outlier detection with Local Outlier Factor (LOF) URL: https://scikit-learn.org/dev/auto_examples/neighbors/plot_lof_outlier_detection.html (дата звернення: 23.06.2024)

59. HBOS Distribution Discriminator. URL: https://www.mathworks.com/help/deeplearning/ref/aivnv.ood.hbosdistributiondiscriminator.html (дата звернення: 23.06.2024)

60. Getting started with Altair RapidMiner. URL: https://docs.rapidminer.com/latest/studio/getting-started/index.html#tutorials (дата звернення: 05.09.2024)

61. СОУ 207.01:2017. Текстові документи. Загальні вимоги. Хмельницький: ХНУ, 2017. 46 с. URL: https://msn.khnu.km.ua/pluginfile.php/466522/mod_resource/content/1/132_C%20T%20A%20H%20Д%20A%20P%20T%20чист%20.pdf

62. ДСТУ 8302:2015. Бібліографічне посилання. Загальні положення та правила складання. [Чинний від 2016-07-1]. Київ, 2016. 20 с. (Державна наукова установа ―Книжкова палата України імені Івана Федорова‖).

ДОДАТОК А. ПЕРЕЛІК НАУКОВИХ ПРАЦЬ

Міжнародний науково-технічний журнал
«Вимірювальна та обчислювальна техніка в технологічних процесах»
ISSN 2219-9365

https://doi.org/10.31891/2219-9365-2024-80-15
UDC 004

PETLIAK Nataliia
Khmelnytskyi National University
https://orcid.org/0000-0001-5971-4428
e-mail: npetlyak@khmnu.edu.ua
BILETSKYI Kostiantyn
Khmelnytskyi National University
e-mail: biletskyik@khmnu.edu.ua
ZASTAVNA Yana
Khmelnytskyi National University
e-mail: zastavna@khmnu.edu.ua

APPROACH TO DETECTION OF ANOMALOUS NETWORK TRAFFIC USING LOF AND HBOS ALGORITHMS

The article is devoted to the problem of detecting anomalies in modern computer networks, which is one of the main threats to cyber security. With the development of Internet technologies, the number of devices and the volume of network traffic are constantly increasing, which leads to an increase in the risk of various cyber threats, such as DDoS attacks, zero-day attacks, and exploitation of protocol vulnerabilities. Abnormal network traffic can result from malicious activity and technical malfunctions, such as configuration errors or hardware failures. Specialised algorithms and methods of analysing large volumes of data are used to detect such threats. The paper considers the main methods of detecting anomalies in network traffic, including classical approaches and modern deep and machine learning methods. Special attention is paid to the efficiency of using methods based on convolutional neural networks, long-term memory and their combinations to detect anomalies. An analysis of the disadvantages and advantages of various approaches to detecting anomalous traffic, such as high computational requirements and the complexity of setting up models, is performed. Still, their effectiveness in analysing large volumes of data is noted. One of the main methods used for anomaly analysis is the local outlier algorithm, which compares the density of objects with their neighbours, allowing for the detection of anomalies in regional segments of the data. Another method is histogram-based outlier estimation, which is faster and more efficient using one-dimensional histograms for each variable. The work also explores the application of unsupervised machine learning methods, which allows for analysing network traffic in real time without the need for prior labelling of data. The article also considers the prospects of further testing the proposed methods in real networks. The combined use of LOF and HBOS balances anomaly detection accuracy and data processing speed, essential to ensure continuous system operation in high-load networks. The implementation of similar solutions in actual conditions requires further research, particularly regarding optimising the use of computing resources and adapting methods to the specific conditions of the network environment. Thus, the paper presents a thorough analysis of modern approaches to detecting anomalies in network traffic and substantiates the feasibility of their application in actual conditions to increase the effectiveness of cyber security.

Keywords: network traffic, anomalies, anomaly detection, local emission factor, estimation of emissions based on histogram

ПЕТЛЯК Наталія. БІЛЕЦЬКИЙ Костянтин, ЗАСТАВНА Яна
Хмельницький національний університет

ПІДХІД ДО ВИЯВЛЕННЯ АНОМАЛЬНОГО МЕРЕЖЕВОГО ТРАФІКУ З ВИКОРИСТАННЯМ АЛГОРИТМІВ LOF ТА HBOS

Стаття присвячена проблемі виявлення аномалій у сучасних комп'ютерних мережах, яка є однією з основних загроз кібербезпеці. З розвитком Інтернет-технологій кількість пристроїв і обсяг мережевого трафіку постійно зростає, що призводить до збільшення ризику різноманітних кіберзагроз, таких як DDoS-атаки, атаки нульового дня, використання вразливостей протоколів. Аномальний мережевий трафік може бути наслідком зловмисної діяльності чи технічних збоїв, наприклад помилок конфігурації або апаратних збоїв. Для виявлення таких загроз використовуються спеціалізовані алгоритми та методи аналізу великих обсягів даних. У статті розглянуто основні методи виявлення аномалій у мережевому трафіку, включаючи класичні підходи та сучасні методи глибокого та машинного навчання. Особливу увагу приділено ефективності використання методів на основі згорткових нейронних мереж, довготривалої пам'яті та їх комбінацій для виявлення аномалій. Проведено аналіз недоліків та переваг різних підходів до виявлення аномального трафіку, таких як високі обчислювальні вимоги та складність налаштування моделей. Проте відзначається їхня ефективність при аналізі великих обсягів даних. Робота досліджує застосування методів неконтрольованого машинного навчання, що дозволяє аналізувати мережевий трафік у реальному часі без необхідності попереднього маркування даних. У статті також розглянуто перспективи подальшої апробації запропонованих методів у реальних мережах. Комбіноване використання LOF і HBOS збалансовує точність виявлення аномалій і швидкість обробки даних, необхідну для забезпечення безперервної роботи системи в мережах з високим навантаженням.

Ключові слова: мережевий трафік, аномалії, виявлення аномалій, локальний коефіцієнт викиду, оцінка викидів на основі гістограми

INTRODUCTION

Abnormal traffic in modern computer networks represents one of the main threats to their security and stable operation. The constant growth of the number of connected devices and the continuous development of

Міжнародний науково-технічний журнал
«Вимірювальна та обчислювальна техніка в технологічних процесах»
ISSN 2219-9365

Internet technologies lead to increased traffic that circulates daily in global networks. Along with the increase in traffic, cyber threats also increase [1-2]. From classic DDoS attacks to more sophisticated incidents such as zero-day attacks and exploitation of protocol vulnerabilities, these threats can cause significant damage to both private companies and government institutions. However, abnormal traffic can signal technical malfunctions (for example, configuration errors or hardware failures) and not just malicious actions [3]. In cybersecurity, such deviations often signal attempts to penetrate the network, attacks on services or systems, or spread malicious software. Such actions require rapid identification and neutralisation of threats to minimise losses.

Tools for detecting anomalous traffic are essential to any modern network protection system [4]. They allow you to analyse network traffic behaviour in real-time and detect deviations from standard work patterns. Detecting anomalies, particularly those related to malicious activities, requires sophisticated algorithms to analyse large volumes of data and consider various factors, including the temporal characteristics of traffic, its volume and sources. It should be noted that external attacks and internal threats, such as the compromise of legitimate users or abuse of access rights, can cause abnormal traffic. This makes traffic monitoring and anomaly detection an essential aspect of ensuring cyber security and the uninterrupted operation of network systems and also highlights the need for increased research and development in this area to create more effective and reliable methods of protection against new and more sophisticated cyber threats.

ANALYSIS OF THE LATEST RESEARCH

The paper [5] presents the Data-Oriented Control Intrusion Detection System (DOC-IDS) model for extracting features and detecting anomalies in network traffic using deep learning. The main feature of this model is the integration of the components of a one-dimensional convolutional neural network (1D CNN) and an autoencoder, which allows one to simultaneously extract critical features from traffic data and detect anomalous behaviours. The model can process large volumes of network packet data, providing high accuracy of threat detection thanks to the analysis of complex interrelationships between bytes. Using different types of loss to minimise reconstruction errors and improve classification ability is also a strong advantage of the model. Among the disadvantages of DOC-IDS, one can note its complexity in setting up and the need for enormous computing resources for practical model training. In addition, the model depends on high-quality training data, and its performance may need to improve in cases where insufficiently representative datasets are used.

The study [6] presents a one-class Long Short-Term Memory (OC-LSTM) method for detecting anomalies in large-scale networks. The main advantage of this approach is its ability to train hidden layer features specifically for the anomaly detection task, unlike hybrid methods that use pre-trained models or autoencoders. OC-LSTM uses a loss function similar to OC-SVM, which allows for more flexible solutions for non-linear boundaries between normal and abnormal data. The peculiarity of OC-LSTM is its end-to-end approach to learning without the need to use additional algorithms for feature selection. However, the complexity of optimising the loss function, which is non-convex, complicates the search for optimal solutions.

The article [7] discusses the methodology for unsupervised detection of anomalies in network traffic based on the iterative process of anomaly assessment. The main feature of this approach is the use of two stages of anomaly assessment, which allows an increase in detection accuracy without using labels for model training. The method was tested on the publicly available datasets IDS2018 and DoHBrw, which allowed us to verify its effectiveness under different abnormal traffic conditions. The technique can provide high accuracy even in cases with limited or no training labels. This is achieved through a multi-functional approach to anomaly analysis that considers temporal and statistical traffic characteristics. In addition, using a self-learning mechanism contributes to the gradual improvement of results and allows the detection of more complex anomalies. However, the method's effectiveness decreases with the increase in the share of anomalous traffic since the assumption of the superiority of regular traffic is no longer supported. In addition, for some types of attacks, such as DoS, the method must show more satisfactory results due to their high share in the total traffic, leading to false positive detections.

The research presented in the article [8] is devoted to the application of deep learning in the field of network security, in particular for intrusion detection. The work proposed a model based on convolutional neural networks (CNN-Focal), which uses the Focal Loss function to optimise work with unbalanced data sets. This approach helps to improve the accuracy of attack detection and increase the overall resilience of the model to new types of threats. The main feature of this approach is the use of small convolution kernels, which reduce the number of unnecessary characteristics and increase the performance of the model. In addition, applying a dropout layer prevents the model from being overtrained, and softmax regression is used for multiclass classification. However, significant computational complexity due to the large number of layers and parameters requires powerful hardware resources for training and testing the model.

The study [9] proposes a model for detecting anomalies in network traffic based on a combination of the K-means algorithm and active learning (ALM). The feature of this model is a two-step process, which includes selecting essential features using the Pearson correlation coefficient and the LightGBM algorithm and classifying anomalies based on the K-means method, which allows you to separate normal and abnormal traffic effectively. Despite significant advantages, the model has certain limitations. First, the K-means method depends on the correct

Міжнародний науково-технічний журнал
«Вимірювальна та обчислювальна техніка в технологічних процесах»
ISSN 2219-9365

choice of the number of clusters, which can affect the final classification results. Second, the process of diffusion of results through active learning can be computationally complex, which increases resource requirements for processing large volumes of data.

In [10], a technique for detecting anomalies in network traffic based on bilateral long-term memory (BiLSTM) and the mechanism of attention (Attention) is proposed. A feature of this model is its ability to perform two-stage feature extraction from network traffic. First, a feature extraction is performed using BiLSTM, which allows sequential data analysis, taking into account information from both previous and subsequent elements of the sequence. Next, the attention mechanism is used for secondary feature extraction, giving more weight to essential elements and allowing the model to focus better on crucial traffic characteristics. This increases the accuracy of detecting anomalies in the data. One of the main advantages of this approach is the ability to reduce the number of false positives due to the efficient processing of similar traffic features. With the help of the attention mechanism, the model can better focus on the most essential characteristics, which reduces the probability of incorrect classifications. In addition, BiLSTM integration allows the model to work effectively with sequential data, which is necessary for network traffic analysis, where data consistency is often critical. The use of BiLSTM and the attention mechanism significantly increases the computational complexity of the model, which may require additional resources to process large volumes of data in real-time. In addition, the accuracy of the model depends on the quality of data preprocessing, including the process of normalisation and removal of irrelevant features, which is an essential stage of data preparation for the effective functioning of the model.

Research [11] proposes a real-time network traffic anomaly detection methodology based on deep learning, precisely a combination of CNN and LSTM. This combination makes it possible to analyse a large volume of constantly changing data effectively and ensure anomaly detection accuracy in actual network conditions. The main feature of the approach is the ability of models to effectively process network traffic flows, extracting from them essential features and spatio-temporal dependencies. One of the advantages of using CNN-LSTM is the ability of the model to learn on large-scale data, capturing complex spatial and temporal traffic patterns. This provides significantly higher accuracy compared to traditional machine learning methods. In addition, the study emphasises the importance of processing data streams in real-time with minimal delays, which is essential for promptly detecting cyber threats. Model optimisation through transfer learning, model compression, and parallelisation can reduce computational costs and improve performance in resource-constrained environments like mobile or IoT devices. However, the method has certain limitations. First, many traffic anomalies can lead to class imbalance problems, making it difficult to train the model. Second, significant computational requirements can hinder deploying such solutions in systems with limited resources.

AN APPROACH TO DETECTING ANOMALOUS TRAFFIC

The approach to detecting anomalies in network traffic, as shown in Figure 1, is structured in three steps: data extraction, anomaly detection, and alert verification. The three-step structure allows for the consistency of the network traffic analysis process from the initial data collection to the final verification for compliance with security policies. The first data extraction stage uses a packet analyser to collect and filter information about network packets. This provides the necessary basis for further analysis of network activity. The second stage is the anomaly detection process, based on algorithms for processing the collected data. This step allows the detection of anomalies in network traffic using unsupervised techniques such as local outlier factor (LOF) and histogram-based outlier estimation (HBOS).

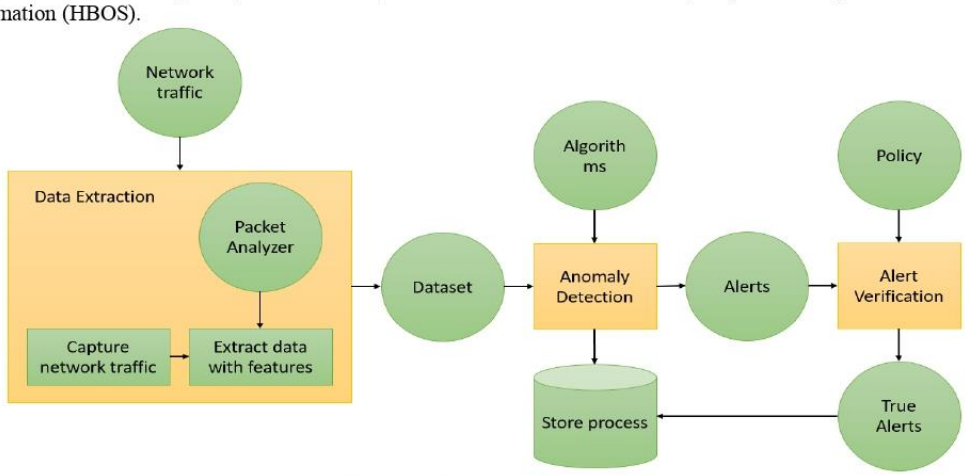


Fig.1. Steps of the proposed approach

Міжнародний науково-технічний журнал
«Вимірювальна та обчислювальна техніка в технологічних процесах»
ISSN 2219-9365

The choice of LOF and HBOS algorithms for detecting anomalies in network traffic is justified by their ability to work effectively with different data types and conditions often found in natural network environments. Both algorithms belong to unsupervised machine learning methods, which allow them to be used in cases where there is no prior information about regular or abnormal instances, as is often the case when analysing network traffic. The LOF algorithm is ideal for detecting local anomalies when the density of objects varies in different parts of the data set. This is especially important for analysing network traffic because different parts of the network may have different standard behaviour patterns. LOF compares a point's local density with its neighbours' density, which allows the detection of objects that appear anomalous only relative to their local surroundings, not the entire data set. For example, in cases where individual network segments have specific properties (such as different load levels or traffic types), LOF allows you to recognise anomalies that are not apparent globally. The main advantage of LOF is that it adapts to different local data characteristics, making it flexible and efficient for dynamic environments. The HBOS algorithm, in turn, is distinguished by its ability to work with large volumes of data due to using one-dimensional histograms to estimate the frequency of values of individual features. This makes it fast and productive compared to methods that require multivariate analysis. In the case of network traffic, where the number of parameters can be significant, and changes in individual characteristics, such as IP address or packet size, can indicate an anomaly, HBOS provides an efficient way to estimate the probability of such changes. Using histograms allows the algorithm to automatically adapt to the data distribution and quickly detect deviations even in large sets. Another advantage of HBOS is its ability to process features independently of each other, which simplifies the analysis process in cases where the relationships between features are not critical.

Thus, the combined use of LOF and HBOS makes it possible to balance in-depth analysis of local relationships in the data and fast processing of large arrays of information. LOF provides anomaly detection in complex and heterogeneous environments where data density varies. In contrast, HBOS delivers high speed and scalability, which are critical factors when working with large network data sets. This approach allows not only the accuracy of anomaly detection to be improved but also the optimisation of the use of resources, which is essential in environments with high traffic intensity.

The third stage is notification verification based on access control policies. Using specialised systems to check access rights, the system determines whether detected anomalies correspond to permitted actions according to established security rules. This step ensures the integration of anomaly detection results with existing security policies, which helps avoid false positive alerts and increases the overall effectiveness of the threat detection system.

The process of extracting data is an essential step in detecting anomalies in network traffic. First, collecting and processing network packets allows you to obtain a basic set of data for further analysis. Traffic monitoring is done using the Wireshark tool, enabling you to capture and analyse protocols such as IP, TCP, or UDP. Packet Analyzer provides comprehensive information about protocol type, packet size, IP addresses, and other critical network traffic characteristics essential for identifying potential threats or anomalies. The second stage of data mining is converting the collected information into a format suitable for analysis. This includes the selection of relevant features (features) for model building. These features include source and destination IP address, packet size, delay time, protocol type, and other parameters. It is important to correctly select the features because excessive irrelevant characteristics can complicate further analysis and detection of anomalies. In unsupervised machine learning for anomaly detection, feature selection becomes critical for model accuracy and efficiency.

The transformed data set with all relevant features is the basis for the next step — direct anomaly detection. This step identifies suspicious activities or anomalous behaviour in network traffic that may signal potential threats such as intrusions or malicious attacks. Thus, the quality of the extracted data and the selection of the correct signs of direction affect the effectiveness of the entire anomaly detection system.

After the network traffic capture phase is completed, the main task is to detect anomalies in the data. Unsupervised anomaly detection methods are used in this work since the data do not have labels describing normal or abnormal behaviour. The lack of prior knowledge about the normal state of the system requires the determination of a threshold that separates normal and abnormal behaviour based on statistical indicators or other methods of analysis.

Expected behaviour is defined by a baseline that serves as a benchmark for comparison. Any observations that deviate significantly from this line are considered anomalous or outliers. Each anomaly detection method has its approach to outlier estimation, such as statistical models or data density estimation methods. It is important to note that the right choice of threshold plays a crucial role: a too-high threshold can lead to missing abnormalities, while a too-low threshold can cause many false positives.

After processing all instances of the data set with the anomaly detection method, each instance receives a new attribute. This outlier score indicates the probability that the instance is anomalous. Anomaly detection is often done using machine learning algorithms, such as clustering methods or autoencoders, which efficiently identify anomalous patterns even in large datasets. Thus, the defining aspect of the process is the correct setting of models and thresholds to achieve optimal results.

The system's last stage is verifying notifications, which is based on the analysis of received anomalous cases. Alerts are generated based on received outlier scores that exceed a given threshold. These alerts indicate

Міжнародний науково-технічний журнал
«Вимірювальна та обчислювальна техніка в технологічних процесах»
ISSN 2219-9365

suspicious or anomalous activity that needs to be checked through the access control system. We use access control policies and queries to verify that a user can access specific resources. The notification verification process consists of two parts. The first is the creation of queries that are generated based on the alerts generated as a result of the analysis of anomalies. These requests conform to the XACML format and reflect the access policies set in the system. The second part is verification, during which the received requests are checked using access control tools. Verification allows you to confirm whether an alert is a real threat or a false positive. The response of the access control system is based on policies: it can allow or deny access depending on the relevant conditions. This process is crucial to complete the alert verification phase and ensure the information system's security.

CONCLUSIONS

The article emphasises the importance of implementing algorithms for detecting anomalies in network traffic, which allows the detection of both technical malfunctions and potential threats in the form of cyber attacks. The advantages of using machine and deep learning methods in combination with classical methods to increase the effectiveness of network protection are analysed. Further testing of the proposed solutions on real networks will allow us to evaluate their performance in practical conditions, which is the next step in the research. Approbation involves integrating the proposed methods into cyber protection systems for operational real-time traffic monitoring.

References

1. Danial Javaheri, Saeid Gorgin, Jeong-A Lee, Mohammad Masdari, Fuzzy logic-based DDoS attacks and network traffic anomaly detection methods: Classification, overview, and future perspectives, Information Sciences, Vol. 626, 2023, pp. 315-338, doi: 10.1016/j.ins.2023.01.067

2. Xueyuan Duan, Yu Fu, Kun Wang, Network traffic anomaly detection method based on multi-scale residual classifier, Computer Communications, Vol. 198, 2023, pp. 206-216, doi: 10.1016/j.comcom.2022.10.024

3. Haiping Lin, Chengwen Wu, Mohammad Masdari, A comprehensive survey of network traffic anomalies and DDoS attacks detection schemes using fuzzy techniques, Computers and Electrical Engineering, Vol. 104, Part B, 2022, doi: 10.1016/j.compeleceng.2022.108466.

4. Ibrahim Juma, Gajin Slavko, Entropy-based network traffic anomaly classification method resilient to deception, Computer Science and Information Systems, Vol. 19, No. 1, 2022, pp. 87- 116

5. Yoshimura N, Kuzuno H, Shiraishi Y, Morii M, DOC-IDS: A Deep Learning-Based Method for Feature Extraction and Anomaly Detection in Network Traffic, Sensors, Vol. 22, 2022, doi: 10.3390/s22124405

6. Li Y, Xu Y, Cao Y, Hou J, Wang C, Guo W, Li X, Xin Y, Liu Z, Cui L, One-Class LSTM Network for Anomalous Network Traffic Detection, Applied Sciences, Vol. 12, 2022, doi: 10.3390/app12105051

7. Ping G, Zeng T, Ye X, Unsupervised network traffic anomaly detection based on score iterations, Journal of Tsinghua University (Science and Technology), Vol. 62, No. 5, pp. 819-824, doi: 10.16511/j.cnki.qhdxxb.2021.21.045

8. F. Zhao, H. Li, K. Niu, J. Shi, R. Song, Application of Deep Learning-Based Intrusion Detection System (IDS) in Network Anomaly Traffic Detection, Applied and Computational Engineering, Vol. 86, 2024, pp. 250–256, doi: 10.54254/2755-2721/86/20241604.

9. N. Liao, X. Li, Traffic Anomaly Detection Model Using K-Means and Active Learning Method, Int. J. Fuzzy Syst, 2024, pp. 2264–2282, doi: 10.1007/s40815-022-01269-0

10. Pan Chengsheng, Li Zhixiang, Yang Wensheng, Cai Lingyun, Jin Aixin, Anomaly Detection Method of Network Traffic Based on Secondary Feature Extraction and BiLSTM-Attention, Journal of Electronics & Information Technology, Vol. 45, No. 12, 2023

11. Tamilselvan Arjunan, Real-Time Detection of Network Traffic Anomalies in Big Data Environments Using Deep Learning Models, International Journal for Research in Applied Science and Engineering Technology, Vol. 12, No. 9, 2024, doi: 10.22214/ijraset.2024.58946.

Міністерство освіти і науки України
Хмельницький національний університет

ЗБІРНИК НАУКОВИХ ПРАЦЬ

за матеріалами XVI Всеукраїнської науково-практичної конференції
«Актуальні проблеми комп'ютерних наук АПКН-2024»

15-16 листопада 2024

Хмельницький 2024

УДК 004:37:001:62

Збірник наукових праць за матеріалами XVI Всеукраїнської науково-практичної конференції «Актуальні проблеми комп'ютерних наук АПКН-2024». Хмельницький. 2024. 582с.

У збірнику наукових праць подані перспективні практичні розробки аспірантів, студентів та здобувачів в області сучасних інформаційних технологій. Розглянуто актуальні проблеми комп'ютерних наук, комп'ютерної інженерії, прикладної математики й інженерії програмного забезпечення, приведено ряд робіт по впровадженню інформаційних технологій у виробництво та управління. Висвітлено перспективні розробки сучасних систем пошуку, обробки й захисту інформації, медійних та комунікаційних системи.

УДК 004:37:001:62

Матеріали конференції відтворені з авторських оригіналів, друкуються в авторській редакції та наведені в алфавітному порядку прізвищ авторів. При макетуванні можливі незначні зміни компоновки контенту авторських оригіналів. Відповідальність за якість та зміст публікацій несе автор.

Участь у конференції та складові всіх її етапів (розгляд праць, перевірка на плагіат, макетування, публікація збірника наукових праць та видача сертифікатів) є безкоштовними для всіх учасників. Оргкомітет конференції висловлює подяку учасникам конференції та сподівається на подальшу співпрацю.

З питань проведення конференції та подальшого обміну інформацією звертатись на e-mail конференції: *apkt.khnu@gmail.com*

АКТУАЛЬНІ ПРОБЛЕМИ КОМП'ЮТЕРНИХ НАУК - 2024

XVI Всеукраїнська науково-практична конференція

Метою конференції є висвітлення актуальних проблем комп'ютерних наук, інформатики та інформаційних технологій.

Робочі мови конференції:

українська, англійська

СЕКЦІЇ КОНФЕРЕНЦІЇ:

1. Комп'ютерні науки та прикладні інформаційні технології.
2. Комп'ютерна інженерія та системи захисту інформації.
3. Математичне моделювання та інженерія програмного забезпечення
4. Телерадіокомунікації, медійні та комунікаційні системи.
5. Проблеми впровадження інформаційних технологій у виробництво та управління.

СПИСОК ОРГАНІЗАЦІЙ, ПРЕДСТАВНИКИ ЯКИХ БРАЛИ УЧАСТЬ У РОБОТІ КОНФЕРЕНЦІЇ:

Донбаська державна машинобудівна академія
Західноукраїнський національний університет
Національний технічний університет «Харківський політехнічний інститут»
Національний університет «Львівська політехніка»
Приватний заклад вищої освіти «ІТ СТЕП Університет»
Сумський державний університет
Харківський національний університет радіоелектроніки
Хмельницький національний університет
Хмельницький фаховий економіко-технологічний коледж УЕП

ОРГКОМІТЕТ КОНФЕРЕНЦІЇ:

Олег СИНЮК – голова оргкомітету, проректор Хмельницького національного університету з наукової роботи, доктор технічних наук, професор.

Тетяна ГОВОРУЩЕНКО – заступник голови оргкомітету, декан факультету інформаційних технологій Хмельницького національного університету, доктор технічних наук, професор.

Олександр БАРМАК – заступник голови оргкомітету, завідувач кафедри комп'ютерних наук Хмельницького національного університету, доктор технічних наук, професор.

Олег САВЕНКО – професор кафедри комп'ютерної інженерії та інформаційних систем Хмельницького національного університету, доктор технічних наук, професор

Олена ВИСОЦЬКА – доктор технічних наук, завідувач кафедри радіоелектронних та біомедичних комп'ютеризованих засобів і технологій Національного аерокосмічного університету ім. М. Є. Жуковського «Харківський авіаційний інститут», професор

Євгеній ЛАВРОВ – доктор технічних наук, професор (Сумський державний університет)

Людмила ТІМОФЄЄВА – відповідальна за студентську науково-дослідну роботу ХНУ

Олександр МАЗУРЕЦЬ – секретар конференції, доцент кафедри комп'ютерних наук Хмельницького національного університету, к.т.н., доцент кафедри комп'ютерних наук ХНУ

Марина МОЛЧАНОВА – секретар конференції, викладач кафедри комп'ютерних наук Хмельницького національного університету

КОНТАКТНА ІНФОРМАЦІЯ:

e-mail для листування: *apkt.khnu@gmail.com*

ЗМІСТ

УДК 004.77

Білецький К.Б., Рудий Р.С., Петляк Н.С.

Хмельницький національний університет

АЛГОРИТМИ LOF ТА HBOS ДЛЯ ВИЯВЛЕННЯ АНОМАЛЬНОГО ТРАФІКУ

Розглянуто підходи на основі алгоритмів локального коефіцієнту викиду та оцінки викидів на основі гістограм, які дозволяють ефективно аналізувати великі обсяги трафіку. Комбінація цих алгоритмів збалансовує точність виявлення та швидкість обробки даних, що є критично важливим для роботи у високонавантажених мережах. Також обговорюються перспективи застосування запропонованих методів у реальних мережах для підвищення рівня кіберзахисту.

Approaches based on algorithms of the local emission factor and estimation of emissions based on histograms, which allow efficient analysis of large traffic volumes, are considered. Combining these algorithms balances detection accuracy and data processing speed, which is critical for operation in highly loaded networks. The prospects of using the proposed methods in real networks to increase cyber protection are also discussed.

Аномалії у мережевому трафіку є ознакою потенційних загроз для безпеки інформаційних систем. Вони можуть виникати через зловмисні дії, такі як DDoS-атаки, експлуатація вразливостей протоколів або атаки нульового дня [1]. Аномальний трафік також може бути результатом технічних збоїв, таких як помилки конфігурації або апаратні збої. У сучасних комп'ютерних мережах кількість пристроїв і обсяги трафіку постійно зростають, що створює додаткові ризики для безперебійної роботи систем і вимагає ефективних методів для виявлення аномалій [2].

Зважаючи на постійний розвиток інтернет-технологій та збільшення кількості кіберзагроз, існує нагальна потреба в автоматизованих системах моніторингу трафіку, здатних виявляти аномалії в реальному часі. Важливо створювати системи, які могли б ідентифікувати відхилення від стандартної поведінки мережевого трафіку й забезпечувати захист від нових, раніше невідомих загроз.

Сучасні методи виявлення аномалій включають як класичні підходи (такі як статистичний аналіз та моделі на основі правил), так і методи машинного навчання та глибокого навчання. Статистичні методи використовуються для визначення відхилень у мережевих показниках, таких як кількість переданих пакетів або час затримки. Однак ці підходи часто виявляються недостатньо ефективними в умовах великих обсягів трафіку або складних кіберзагроз.

Дослідження показують, що методи глибокого навчання, зокрема згорткові

нейронні мережі (CNN) та рекурентні нейронні мережі (RNN), демонструють високу точність у виявленні аномалій. Наприклад, використання CNN для аналізу трафіку дозволяє виявляти складні шаблони аномальної поведінки, які важко розпізнати за допомогою традиційних методів. Дослідження у цій сфері показують значні перспективи застосування цих моделей для покращення систем кіберзахисту.

Попри успіхи в питаннях виявлення аномалій у мережевому трафіку, деякі питання залишаються відкритими. Головною проблемою є висока обчислювальна складність алгоритмів машинного та глибокого навчання, що вимагає значних ресурсів для їхнього навчання та впровадження в реальних умовах. Крім того, точність виявлення аномалій залежить від якості даних для навчання моделей, що може бути проблематичним в умовах неповних або нерепрезентативних наборів даних.

Також виникає питання про баланс між точністю виявлення аномалій і швидкістю обробки даних, що є критично важливим для систем, які працюють у режимі реального часу. У таких системах важливо мінімізувати затримки у виявленні загроз, одночасно забезпечуючи високу точність результатів.

Метою цього дослідження є розробка ефективного підходу до виявлення аномального мережевого трафіку з використанням алгоритмів локального коефіцієнту викиду (LOF) та оцінки викидів на основі гістограм (HBOS). Завдання полягає у створенні гнучкої системи, здатної забезпечувати високу точність виявлення аномалій за рахунок поєднання локального аналізу трафіку LOF та HBOS. Особлива увага приділяється можливості застосування цих алгоритмів для виявлення аномалій у реальному часі та мінімізації хибнопозитивних результатів.

Вибір алгоритмів LOF і HBOS для виявлення аномалій у мережевому трафіку базується на їхніх можливостях ефективно працювати з різними типами даних та умовами, що часто зустрічаються в реальних мережевих середовищах. Обидва алгоритми належать до неконтрольованих методів машинного навчання, що робить їх корисними у випадках, коли відсутня чітка інформація про нормальні або аномальні екземпляри, як це часто буває в аналізі мережевого трафіку.

LOF є оптимальним для виявлення локальних аномалій, особливо у випадках, коли щільність об'єктів відрізняється в різних частинах набору даних. Цей алгоритм оцінює щільність даних у локальному контексті, порівнюючи її з щільністю сусідніх об'єктів, що дозволяє виявляти аномалії, які виглядають незвично тільки у своєму локальному оточенні. Це особливо актуально для мережевого трафіку, де різні сегменти можуть мати різні моделі поведінки через різні рівні навантаження чи типи трафіку. LOF адаптивно підлаштовується до локальних характеристик даних, що дозволяє ефективно обробляти динамічні мережеві середовища, забезпечуючи високий рівень точності у виявленні аномалій на локальному рівні.

HBOS, у свою чергу, відрізняється високою швидкістю та ефективністю при роботі з великими наборами даних завдяки використанню одновимірних гістограм для оцінки частоти значень окремих ознак. Це робить його придатним для

аналізу мережевого трафіку, де кількість параметрів може бути значною, а зміни в окремих ознаках можуть вказувати на потенційні аномалії. Наприклад, зміни в розмірах пакетів або IP-адресах можуть свідчити про відхилення від нормальної поведінки. Використання одновимірних гістограм дозволяє HBOS автоматично адаптуватися до розподілу даних і швидко виявляти аномалії, що є критично важливим у контексті великих мережевих середовищ. Додатковою перевагою HBOS є його здатність обробляти ознаки незалежно одна від одної, що значно спрощує аналіз і зменшує обчислювальні витрати у випадках, коли взаємозв'язки між ознаками не є ключовими.

Комбіноване використання LOF та HBOS дозволяє досягти синергетичного ефекту завдяки їх взаємодоповнюваності. LOF орієнтований на локальні аномалії та дозволяє враховувати взаємозв'язки між точками даних у межах певних локальних околиць, що особливо корисно в умовах із варіативними патернами та неоднорідною щільністю даних. У свою чергу, HBOS базується на глобальному аналізі та статистичних розподілах, що надає йому перевагу у швидкості й простоті розрахунків, особливо для великих наборів даних. Разом ці алгоритми створюють баланс між локальною чутливістю LOF та глобальною ефективністю HBOS, що дозволяє не лише виявляти аномалії з різними характеристиками, але й знижувати кількість помилкових спрацьовувань. Таким чином, їх спільне застосування забезпечує більш гнучкий і точний підхід до виявлення аномалій, охоплюючи широкий спектр сценаріїв і характеристик трафіку.

Запровадження алгоритмів LOF та HBOS для виявлення аномального мережевого трафіку дозволяє підвищити ефективність захисту інформаційних систем від кіберзагроз. Використання цих алгоритмів у поєднанні з класичними підходами машинного навчання забезпечує можливість оперативного виявлення аномалій у реальному часі, що особливо важливо для мереж із високим навантаженням.

Перелік посилань

1. Danial Javaheri, Saeid Gorgin, Jeong-A Lee, Mohammad Masdari. Fuzzy logic-based DDoS attacks and network traffic anomaly detection methods: Classification overview and future perspectives. Information Sciences. Vol. 626, 2023. pp. 315-338.
2. Xueyuan Duan, Yu Fu, Kun Wang. Network traffic anomaly detection method based on multi-scale residual classifier. Computer Communications. Vol. 198, 2023. pp. 206-216.

АКТУАЛЬНІ ПРОБЛЕМИ КОМП'ЮТЕРНИХ НАУК 2024

ЗБІРНИК НАУКОВИХ ПРАЦЬ

Комп'ютерна верстка: **Олександр МАЗУРЕЦЬ**

Підписано до друку 21.11.2024.
Версія друку «APKN2024_CorpusPaper v3mod5 Finita».

E-mail: *apkt.khnu@gmail.com*
ХНУ. м. Хмельницький, вул. Інститутська,11.

Завідувачу кафедри кібербезпеки
к.т.н., доц. Кльоцу Ю.П.
Білецького Костянтина Борисовича
ПІБ здобувача вищої освіти

Студента ФІТ, 2 курсу, групи КБЗІм-23-1

ЗАЯВА

З правилами чинного Положення про систему забезпечення академічної доброчесності у Хмельницькому національному університеті, згідно з яким виявлення академічного плагіату є підставою для відмови в допуску кваліфікаційної роботи до захисту і застосування заходів дисциплінарної та академічної відповідальності, ознайомлений (а). Про використання програмно-технічних засобів для перевірки кваліфікаційних робіт здобувачів вищої освіти на наявність академічного плагіату оповіщений (а) та надаю свою згоду на обробку й збереження університетом моєї роботи в інституційному репозитарії Хмельницького національного університету.

Також надаю університету право на передачу моєї роботи для обробки та збереження в базах даних програмно-обчислювального комплексу StrikePlagiarism та/або програмнотехнічного засобу Anti-Plagiarism) і використання роботи для виявлення академічного плагіату в інших роботах, які перевіряються програмно-технічними засобами та користувачами, що мають доступ до цих програмно-технічних засобів, виключно в обмежених цілях для виявлення текстових збігів в роботах.

Робота надається для перевірки в електронному варіанті. Електронна версія моєї роботи збігається (ідентична) з друкованою.

29.11.2024
дата

підпис

Anti-Plagiarism v-15.257

Максимальне співпадіння з одним документом 1.0%

Словники перевірки: en_US, ru_RU, ua_UA. **Помилок в документах: 10%**

ID: 152730 Назва: Метод виявлення аномалій мережевого трафіку Додано в БД: 2024-12-01 Автора: Білецький Костянтин Керівники: Касянчук М.М. Консультанти: Опоненти:	Документ		Сумарний збіг по Базі Даних	
	Символи	Лексеми	Символи	Лексеми
	99516	777	832 (1%)	11 (1%)

Джерело плагіату

ID	Опис	Наявність плагіату в документі	
		Символи	Лексеми

Протокол аналізу звіту подібності науковим керівником

Заявляю, що я ознайомився (-лась) з Повним звітом подібності, який був згенерований Системою виявлення і запобігання плагіату щодо роботи:

Автор: Білецький костянтин

Співавтор:

Назва: Метод виявлення аномалій мережевого трафіку

Науковий керівник: Касянчук М.М.

Підрозділ: Кафедра кібербезпеки

Коефіцієнт подібності 1:1.3%

Коефіцієнт подібності 2:0%

Мікропробіли: 0

Заміна букв: 0

Інтервали: 0

Білі знаки: 0

Дата створення звіту: 2024-12-01 13:16:24.0

Після аналізу Звіту подібності констатую наступне:

☐ **Запозичення, виявлені в роботі є законними і не є плагіатом. Рівень подібності не перевищує допустимої межі. Таким чином робота незалежна і приймається.**

☐ **Запозичення не є плагіатом, але перевищено граничне значення рівня подібностей. Таким чином робота повертається на доопрацювання.**

☐ **Виявлено запозичення і плагіат або навмисні текстові спотворення (маніпуляції), як передбачувані спроби укриття плагіату, які роблять роботу невідповідною вимогам законодавства (Ст. 32. ЗУ Про вищу освіту, пункт 3.1, Ст. 42. ЗУ Про освіту) та вимог НАЗЯВО (Критерій 5), а також кодексу етики і процедурам. Таким чином робота не приймається.**

Обгрунтування:

Дата 1.12.24 експерт

РІШЕННЯ ЕКСПЕРНОЇ КОМІСІЇ

КАФЕДРИ КІБЕРБЕЗПЕКИ

ПРО ДОПУСК КВАЛІФІКАЦІЙНОЇ РОБОТИ ДО ЗАХИСТУ

Підтверджуємо ознайомлення з результатом звіту подібності щодо роботи, генерованого системою виявлення текстових збігів/ідентичності/схожості:

Назва: Метод виявлення аномалій мережевого трафіку

Автор: Білецький Костянтин Борисович

Спеціальність: 125 – Кібербезпека та захист інформації

Освітня програма:Кібербезпека та захист інформації

Науковий керівник:Михайло КАСЯНЧУК, докт. техн. наук, професор

Після аналізу звіту подібності зроблено такий висновок:

№	Висновок	Позначка про відповідність
1	Запозичення, виявлені в роботі, є законними і не є плагіатом (далі – зазначаються підстави віднесення запозичень до правомірних). Робота приймається до захисту.	відповідає
2	Виявлені запозичення не є плагіатом, розміщені в розділах, які не описують безпосередньо авторське дослідження, але кількість цитат перевищує обсяг, виправданий поставленою метою роботи (далі – зазначаються детальні та аргументовані підстави віднесення запозичень до правомірних). Робота приймається до захисту, але має бути відкоригована. Відкоригований варіант має бути поданий на кафедру за 2 дні до захисту, разом із заявою щодо самостійності виконання письмової роботи та ідентичності друкованої та електронної версії роботи.	
3	Виявлені запозичення не є плагіатом; але частково розміщені в розділах, які описують безпосередньо авторське дослідження, а кількість цитат перевищує обсяг, виправданий поставленою метою роботи. В зв'язку з цим мета роботи та поставлені завдання не були досягнені. Робота може бути допущена до захисту (наступного року) після того як буде відкоригована та допрацьована і успішно пройде повторну перевірку на академічний плагіат.	
4	Робота містить навмисні текстові спотворення, передбачувані спроби укриття запозичень або інші прояви академічного плагіату. Робота містить фабрикацію або фальсифікацію даних. Робота не допускається до захисту.	
5	Інше:	

Підтвердження:

Оригінальність тексту роботи за результатами перевірки системою StrikePlagiarism складає 98,7%, оригінальність тексту роботи за результатами перевірки системою Anti-Plagiarism складає 99%.

Згідно з правилами чинного Положення «Про систему забезпечення академічної доброчесності у Хмельницькому національному університеті» від 24.09.2024, авторська робота, обсяг оригінального тексту у відсотках до загального обсягу матеріалу в якій складає 90-100%, визначається роботою з високою унікальністю тексту і допускається до захисту.

Виявлені модифікації стосуються математичних формул і не є порушенням академічної доброчесності.

Керівник роботи _____ Михайло КАСЯНЧУК

Гарант ОП _____ Віра ТІТОВА

Завідувач кафедри кібербезпеки _____ Юрій КЛЬОЦ

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

РЕЦЕНЗІЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ

освітнього ступеня «магістр»

Студент Білецький Костянтин Борисович

Тема Метод виявлення аномалій мережевого трафіку

Спеціальність 125 – Кібербезпека та захист інформації

Обсяг кваліфікаційної роботи освітньо-кваліфікаційного рівня «бакалавр»:

кількість листів креслень – ; кількість сторінок записки 88 .

1. Короткий зміст роботи та прийнятих рішень У кваліфікаційній роботі була розроблено метод виявлення аномалій мережевого трафіку. Проаналізовано поняття аномалій мережевого трафіку та існуючі методи їх виявлення, здійснено аналіз систем виявлення вторгнень. Проаналізовано наявні набори даних для тестування та описано алгоритм формування власного набору даних. Розроблено алгоритм виявлення аномалій на основі зібраного трафіку та метод виявлення аномалій. Робота також включає експериментальну перевірку ефективності запропонованого методу.

2. Висновок про відповідність кваліфікаційної роботи завданню У кваліфікаційній роботі було виконано поставлене завдання як у теоретичній, так і в практичній частині.

3. Характеристика виконання кожного розділу роботи, ступінь використання останніх досягнень науки і техніки і передових методів роботи: У вступі роботи наведено загальну характеристику задачі, визначено об'єкт, предмет та методи дослідження, а також сформульовано мету. Зазначено задачі, що потрібно виконати для досягнення поставленої мети. Проведено аналіз досліджуваної проблеми та обґрунтовано підхід до її вирішення. У першому розділі розглянуто існуючі підходи до виявлення аномалій у мережевому трафіку, їх переваги та недоліки. Автор проводить огляд методів аналізу даних та аналіз сучасних систем виявлення вторгнень. У другому розділі проведено аналіз існуючих наборів даних з метою визначення їх відповідності задачам дослідження. Описано методику створення власного набору даних, процес збору й підготовки даних, які необхідні для моделювання. Третій розділ присвячено розробці алгоритму та методу виявлення аномалій. Четвертий розділ зосереджується на тестуванні та оцінці ефективності системи. Виконано порівняння з іншими методами виявлення аномалій, що підтвердило перевагу розробленого методу.

4. Позитивні сторони роботи Кваліфікаційна робота має як наукову, так і практичну цінність. Запропонований метод виявлення аномалій здатен підвищити рівень безпеки корпоративних мереж, мінімізуючи хибнопозитивні спрацювання.

5. Негативні сторони роботи Роблений метод залежить від якості початкових даних, що може вплинути на його ефективність.

6. Оцінка графічного оформлення та пояснювальної записки роботи Графічне оформлення кваліфікаційної роботи відповідає темі роботи та виконане з дотриманням стандартів. В цілому, графічне оформлення є якісним, а пояснювальна записка відповідає нормам оформлення.

7. Відгук про роботу в цілому Кваліфікаційна робота заслуговує позитивної оцінки, оскільки весь матеріал роботи є структурованим, чітким та послідовним. Усі розділи роботи мають логічну послідовність, що сприяє зрозумінню викладеного матеріалу в рамках теми роботи.

8. Інші зауваження

9. Оцінка кваліфікаційної роботи Враховуючи всі позитивні та негативні сторони представленої кваліфікаційної роботи, можна зробити висновок, що вона заслуговує оцінки «відмінно».

РЕЦЕНЗЕНТ (прізвище, ім'я, по батькові, посада, місце роботи)
Мартинюк Валерій Володимирович,
завідувач кафедри автоматизації, комп'ютерно-інтегрованих технологій та робототехніки, доктор технічних наук, професор

«13» грудня 2024. _____ (підпис)