Державний університет “Київський авіаційний інститут”
Міністерство освіти і науки України

Кваліфікаційна наукова
праця на правах рукопису

ПЕТЛЯК НАТАЛІЯ СЕРГІЇВНА

УДК 004.056.5

ДИСЕРТАЦІЯ

МОДЕЛІ ТА МЕТОДИ ВИЯВЛЕННЯ АНОМАЛЬНОГО ТРАФІКУ В ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ СИСТЕМАХ

125 – Кібербезпека
12 – Інформаційні технології

Подається на здобуття наукового ступеня доктора філософії

Дисертація містить результати власних проваджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело.

______________ Н.С. Петляк
підпис

Наукові керівники:
Лазаренко Сергій Володимирович, доктор технічних наук, професор;
Кльоц Юрій Павлович, кандидат технічних наук, доцент

Київ – 2025

АНОТАЦІЯ

Петляк Н.С. Моделі та методи виявлення аномального трафіку в інформаційно-комунікаційних системах. – Кваліфікаційна наукова праця на правах рукопису. Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 125 – Кібербезпека. – Державний університет “Київський авіаційний інститут”. Міністерство освіти і науки України, м. Київ, 2025.

У сучасному цифровому середовищі інформаційна безпека набуває важливого значення для більшості організацій незалежно від їхнього розміру або сфери діяльності. Серед головних напрямів забезпечення безпеки домінує концепція виявлення аномалій у мережевому трафіку, яка дозволяє ефективно ідентифікувати загрози, що не мають ознак, властивих відомим шаблонам атак. Системи виявлення вторгнень, які активно використовуються в межах архітектур інформаційного захисту, здатні здійснювати аналіз мережевого трафіку в режимі реального часу. Проте, незважаючи на їх розповсюдження, більшість з них все ще спирається на сигнатурний аналіз, ефективність якого обмежується наявністю попередньо зареєстрованих шаблонів атак. Такий підхід втрачає ефективність у випадках нових або модифікованих загроз, а також не здатен виявляти атаки, які розподілені у часі або мають прихований характер. Це підкреслює актуальність розробки інтелектуальних систем, орієнтованих саме на виявлення аномальної активності, що дозволить значно покращити рівень кіберзахисту. Особливу увагу в дослідженні приділено аналізу вихідного трафіку, який часто залишається поза зоною основного моніторингу, хоча може бути вагомим у контексті безпеки. Статистика останніх років свідчить про стрімке зростання частоти та складності таких атак, що завдають компаніям значних фінансових втрат, знижують довіру споживачів і створюють загрози безперервності бізнес-процесів. У зв’язку з цим актуальним є формування ефективної системи виявлення аномалій у вихідному трафіку як окремого напряму захисту мережевої інфраструктури.

У *вступі* обґрунтовано актуальність дисертаційної роботи, сформульовано мету та основні завдання дослідження, наведено відомості щодо зв'язку роботи з науковим темами. Подано анотацію отриманих в дисертації результатів, висвітлено їх наукову новизну і практичну значимість, відзначено особистий внесок здобувача у спільних публікаціях, наведено кількісні показники стосовно опублікованих наукових праць, структури та обсягу дисертації.

У першому розділі дисертаційного дослідження представлено комплексний огляд сучасного стану інформаційної безпеки в контексті зростаючих загроз для інформаційно-комунікаційних систем. Особливу увагу приділено аналізу викликів, пов'язаних із виявленням аномалій у мережевому трафіку, що є важливим чинником у попередженні складних і нових типів кібератак. Окреслено недоліки традиційного моніторингу, який орієнтований переважно на загальну оцінку продуктивності системи, на відміну від методів, що передбачають виявлення саме аномальної поведінки. У рамках розділу проведено аналіз існуючих моделей виявлення аномалій: статистичних, кластеризаційних, поведінкових, моделей на основі машинного навчання тощо. Розглянуто їхні обмеження, зокрема в умовах високої динамічності мережевого трафіку та появи нових, ще невідомих атак. Зроблено акцент на особливості виявлення аномалій саме у вихідному трафіку, який часто не охоплюється традиційними системами захисту, але може містити важливу інформацію про зловмисну активність або використання локальних систем як платформи для атак на інші ресурси. Таким чином, у розділі обґрунтовано наукову та практичну значимість вдосконалення моделей і методів виявлення аномалій з орієнтацією на вихідний трафік, що стало основою для постановки цілей дослідження. Визначено напрямки вдосконалення існуючих підходів та необхідність розроблення нових, адаптивних рішень на основі гібридних методик і моделей нечіткого аналізу.

У другому розділі представлено моделі процесів, що описують поведінку користувачів і потенційних зловмисників у мережі, а також моделей класифікації мережевого трафіку з метою виявлення аномалій. Розроблені моделі охоплюють різні сценарії функціонування мережі та типи взаємодії в ній, що дозволяє сформувати базу для подальшої побудови ефективних методів виявлення аномального трафіку. Запропоновано вдосконалену модель поведінки типового користувача у відкритому сегменті мережі, яка враховує реальні кроки користувача при з'єднанні з мережею, а також його активність під час сесії. Також розроблено модель поведінки порушника, яка описує характерні дії порушника в контексті доступу до мережевих ресурсів, що дозволяє чітко відокремити законну і потенційно загрозливу активність. Окрему увагу приділено формуванню моделі сигнатури пакету, яка виступає вагомим елементом у процесі класифікації трафіку. Важливим етапом є вдосконалення моделей класифікації трафіку за ознаками та на основі самоподібності. Зокрема, впроваджено концепцію нечіткої класифікації, що дозволяє враховувати невизначеність і неповноту вхідних даних. Це особливо актуально в умовах, коли класичні методи демонструють суперечливі результати або є недостатньо чутливими до слабких, але потенційно небезпечних відхилень. Розділ завершується формалізацією всіх моделей, які у подальшому використовуються як складові частини при розробці методів виявлення аномалій.

У третьому розділі дисертації здійснено безпосередню розробку та вдосконалення методів виявлення аномального трафіку, що базуються на моделях, описаних у попередньому розділі. Покращено метод класифікації трафіку за ознаками, що дозволяє ефективно сортувати вхідний потік даних на дозволений, заборонений і невизначений. Метод враховує множини сигнатур, а також реалізує сценарії перевірки залежно від рівня довіри до джерела та характеру трафіку. Вдосконалено метод класифікації на основі самоподібності, який аналізує мережевий трафік у часових вікнах, дозволяючи виявити відхилення від звичних моделей поведінки у мережі. Запропоновано нечіткий метод виявлення аномалій, який оперує правилами нечіткої логіки для оцінки рівня загрози. Цей метод продемонстрував високу точність в умовах невизначених або суперечливих даних. На основі аналізу переваг усіх підходів створено гібридний метод, що об’єднує метод класифікації трафіку за ознаками, на основі самоподібності та нечіткий метод, в єдину систему. Така інтеграція дозволяє мінімізувати хибні спрацювання, підвищити адаптивність системи та її стійкість до нових загроз. Ефективність кожного з методів була перевірена шляхом експериментальних розрахунків і моделювань, що створило надійне підґрунтя для реалізації системи у наступному розділі.

У четвертому розділі дисертації реалізовано структурну модель системи виявлення аномального трафіку на основі розробленого гібридного методу, яка включає основні модулі обробки мережевого трафіку, механізми класифікації, систему логування результатів і керування множинами сигнатур. Описано особливості реалізації кожного з етапів функціонування системи — від захоплення та попередньої обробки трафіку до прийняття рішення щодо безпечності з'єднання. Окрему увагу приділено реалізації модулів на практиці, зокрема у тестовому середовищі. Проведено низку експериментів, що включали різні сценарії роботи мережі, типи атак і варіації навантаження. Результати оцінки достовірності свідчать про підвищення точності виявлення аномалій при використанні гібридного підходу порівняно з окремими методами. Продемонстровано здатність системи до адаптації в умовах динамічних змін середовища, зменшення кількості хибнопозитивних спрацювань та стабільність при обробці великого обсягу трафіку. Таким чином, у межах розділу забезпечено повноцінну верифікацію розробленої системи та підтверджено її ефективність для впровадження в інформаційно-комунікаційні системи з метою підвищення рівня кіберзахисту.

У *висновках* узагальнено результати виконаного дослідження, обґрунтовано наукову новизну розроблених моделей та методів, а також доведено їхню практичну ефективність. Запропонований гібридний метод виявлення аномального трафіку, що базується на поєднанні класифікації трафіку за ознаками, на основі самоподібності та нечіткого аналізу, дозволяє суттєво підвищити рівень захисту інформаційно-комунікаційних систем. Особливу увагу приділено виявленню аномалій у вихідному мережевому трафіку, що забезпечує своєчасне виявлення внутрішніх загроз та мінімізує ризики несанкціонованого використання ресурсів мережі для зовнішніх атак. Практичне впровадження запропонованих рішень підтвердило їхню ефективність в умовах реального мережевого середовища.

Наукова новизна отриманих результатів

На підставі теоретичних і практичних досліджень, виконаних у дисертаційній роботі, отримано такі нові результати:

– *Вдосконалено* модель сигнатури пакету для пошуку аномального трафіку, що за рахунок виключення з параметрів сигнатури розміру заголовка, контрольної суми заголовку, корисного розміру пакета, мітки потоку, пріоритету пакету, контрольної суми пакету за принципом Парето, забезпечило зменшення часу аналізу трафіку.

– *Вдосконалено* модель процесу нечіткого виявлення аномального трафіку, в якій за рахунок використання експертного підходу сформована множина правил та набір відповідних лінгвістичних змінних, що дозволило розробити нові підходи до виявлення аномального трафіку в інформаційно-комунікаційних системах.

– *Вперше* розроблено гібридний метод виявлення аномального трафіку в інформаційно-комунікаційних системах в якому за рахунок інтеграції методу класифікації трафіку за ознаками, методу самоподібності та розробленої моделі процесу нечіткого виявлення дозволило динамічно формувати множини сигнатур при класифікації трафіку за ознаками та підвищити показники виявлення аномального трафіку.

– *Удосконалено* структурну модель системи виявлення аномального трафіку в інформаційно-комунікаційних системах, що за рахунок інтегрування розробленого гібридного методу виявлення аномального трафіку в інформаційно-комунікаційних системах та динамічного варіювання множиною дозволених та заборонених з'єднань у режимі реального часу дозволило зменшити навантаження на процесор.

Практичне значення отриманих результатів

Одержані в дисертаційній роботі результати стали підґрунтям для забезпечення зменшення обсягів аномального трафіку в інформаційно-комунікаційні системі.

Практично вагомими вважаються такі результати:

- розроблено алгоритмічне забезпечення системи виявлення аномального трафіку в інформаційно-комунікаційних системах, що реалізує гібридний метод, його складові метод класифікації трафіку за ознаками, метод самоподібності, нечіткий метод та моделі, що в них використовуються;

- на основі алгоритму розроблено програмний застосунок, що дозволяє проводити аналіз трафіку в інформаційно-комунікаційних системах;

- можливість практичного використання розробленого програмного застосунку сумісно з маршрутизаторами в інформаційно-комунікаційній системі.

Результати дисертаційної роботи апробовано і використано у таких організаціях та установах:

- у ТОВ «X-CITY» (м. Хмельницький) для аналізу трафіку абонентів за допомогою гібридного методу в інформаційно-комунікаційні системі, що дозволило без зміни архітектури інформаційно-комунікаційної системи ТОВ «X-CITY» та апаратного забезпечення зменшити аномальний трафік, звернення провайдерів вищого рівня про атаки, що надходять з інформаційно-комунікаційної системи ТОВ «X-CITY». Акт впровадження від 10.04.2025р.

- у навчальному процесі кафедри кібербезпеки Хмельницького національного університету під час викладання дисциплін «Технології виявлення вразливостей та вторгнень», «Безпека безпроводових мереж та інтернет речей» та «Моніторинг та менеджмент інформаційної безпеки». Акт впровадження від 25.03.2025р.

Ключові слова: кібербезпека, інформаційна безпека, аномалії, виявлення аномалій, вразливість, кібератаки, кіберзагрози, виявлення вторгнень, нейронні мережі, нечітка логіка, нечіткі множини, лінгвістична змінна, мережа, інформаційно-комунікаційна система, аналіз трафіку, модель, самоподібність, сигнатурний аналіз, метод, гібридний метод, система, Snort, Suricata.

ANNOTATION

Petliak N.S. Models and methods for detecting abnormal traffic in information and communication systems. Dissertation for the degree of Doctor of Philosophy in speciality 125 - Cybersecurity - State University “Kyiv Aviation Institute”. Ministry of Education and Science of Ukraine, Kyiv, 2025.

In today's digital environment, information security is becoming increasingly important for most organisations, regardless of their size or field of activity. The concept of detecting anomalies in network traffic dominates among the main areas of security, allowing to effectively identify threats that do not have the features inherent in known attack patterns. Intrusion detection systems, which are widely used within information security architectures, are capable of analysing network traffic in real time. However, despite their proliferation, most of them still rely on signature analysis, which is limited by the availability of previously registered attack patterns. This approach loses its effectiveness in the presence of new or modified threats, and is also unable to detect attacks that are distributed over time or have a covert nature. This underscores the urgency of developing intelligent systems focused on detecting anomalous activity, which will significantly improve the level of cyber defence in the security context. Recent statistics show a rapid increase in the frequency and complexity of such attacks, which cause significant financial losses to companies, reduce consumer confidence and pose threats to the continuity of business processes. In this regard, it is important to develop an effective system for detecting anomalies in outgoing traffic as a separate area of network infrastructure protection.

The *introduction* substantiates the relevance of the dissertation, formulates the purpose and main objectives of the study, and provides information on the relationship of the work with scientific topics. It provides a summary of the results obtained in the dissertation, highlights their scientific novelty and practical significance, notes the personal contribution of the applicant in joint publications, and provides quantitative indicators of published scientific papers, structure and scope of the dissertation.

The first section of the dissertation research provides a comprehensive overview

of the current state of information security in the context of growing threats to information and communication systems. Particular attention is paid to the analysis of challenges related to the detection of anomalies in network traffic, which is an important factor in preventing complex and new types of cyberattacks. The article outlines the disadvantages of traditional monitoring, which is focused mainly on the overall assessment of system performance, as opposed to methods that involve detecting anomalous behaviour. This section analyses existing anomaly detection models: statistical, clustering, behavioural, machine learning, etc. Their limitations are considered, in particular, in the context of high dynamism of network traffic and the emergence of new, yet unknown attacks. Emphasis is placed on the peculiarities of detecting anomalies in outbound traffic, which is often not covered by traditional security systems, but may contain important information about malicious activity or the use of local systems as a platform for attacks on other resources. Thus, the section substantiates the scientific and practical significance of improving models and methods for detecting anomalies with a focus on outbound traffic, which became the basis for setting the research objectives. It identifies the directions for improving existing approaches and the need to develop new, adaptive solutions based on hybrid techniques and fuzzy analysis models.

The second section presents models of processes that describe the behaviour of users and potential attackers in the network, as well as models for classifying network traffic to detect anomalies. The developed models cover various scenarios of network functioning and types of interaction in it, which allows forming a basis for further development of effective methods for detecting anomalous traffic. An improved model of the behaviour of a typical user in an open network segment is proposed, which takes into account the user's real steps when connecting to the network, as well as his activity during the session. A model of intruder behaviour has also been developed and describes the typical actions of an intruder in the context of access to network resources, which allows for a clear separation of legitimate and potentially threatening activity. Special attention is paid to the formation of a packet signature model, which is an important element in the traffic classification process. An important step is to improve traffic classification models based on features and self-similarity. In particular, the concept of unclear classification was introduced, which allows taking into account the uncertainty and incompleteness of the input data. This is especially relevant when classical methods demonstrate contradictory results or are not sufficiently sensitive to weak but potentially dangerous deviations. The section concludes with the formalisation of all the models that are subsequently used as components in the development of anomaly detection methods.

The third section of the thesis deals with the direct development and improvement of anomalous traffic detection methods based on the models described in the previous chapter. The method of traffic classification by signs has been improved, which allows to effectively sort the incoming data stream into allowed, prohibited and undefined. The method takes into account sets of signatures and implements verification scenarios depending on the level of trust in the source and the nature of the traffic. The self-similarity-based classification method is improved, which analyses network traffic in time windows, allowing to detect deviations from the usual patterns of network behaviour. A vague anomaly detection method is proposed that operates on vague logic rules to assess the level of threat. This method has demonstrated high accuracy in the face of uncertain or contradictory data. Based on the analysis of the advantages of all approaches, a hybrid method was created that combines the feature-based traffic classification method, self-similarity, and vague method into a single system. This integration minimises false positives and increases the system's adaptability and resilience to new threats. The effectiveness of each method was tested through experimental calculations and simulations, which provided a solid basis for the implementation of the system in the next section.

Section four of the thesis implements a structural model of an anomalous traffic detection system based on the developed hybrid method, which includes the main modules for processing network traffic, classification mechanisms, a system for logging results and managing signature sets. The features of the implementation of each stage of the system's functioning are described - from traffic capture and pre-processing to making a decision on the security of the connection. Special attention is paid to the implementation of the modules in practice, in particular in a test environment. A number of experiments were conducted, including various network scenarios, types of attacks and load variations. The results of the reliability assessment indicate an increase in the accuracy of anomaly detection when using a hybrid approach compared to individual methods. The system's ability to adapt to dynamic changes in the environment, reduce the number of false positives, and stability when processing a large amount of traffic is demonstrated. Thus, this section provides a full verification of the developed system and confirms its effectiveness for implementation in information and communication systems to improve the level of cyber defence.

The conclusions summarise the results of the study, substantiate the scientific novelty of the developed models and methods, and prove their practical effectiveness.

Scientific novelty of the results obtained

Based on the theoretical and practical research carried out in this thesis, the following new results were obtained:

– The packet signature model for detecting anomalous traffic has been improved, which, by excluding the header size, header checksum, payload size, flow label, packet priority, and Pareto checksum from the signature parameters, has reduced the time for traffic analysis.

– The model of the process of fuzzy detection of anomalous traffic was improved, in which a set of rules was formed by expert determination of the number of terms and their values for each linguistic variable, which allowed to develop a hybrid method for detecting anomalous traffic in information and communication systems.

– For the first time, a hybrid method for detecting anomalous traffic in information and communication systems has been developed, which, by integrating the method of traffic classification by features, the method of self-similarity and the fuzzy method, allowed to dynamically form sets of signatures for the method of traffic classification by features, which reduced the time of traffic analysis.

– The structural model of the system for detecting anomalous traffic in information and communication systems has been improved, which, by integrating the developed hybrid method for detecting anomalous traffic in information and communication systems, has increased the reliability of detecting anomalous traffic and reduced the load on the processor by changing the set of allowed and prohibited connections in real time.

Practical significance of the results.

The results obtained in this thesis have become the basis for reducing the volume of abnormal traffic in information and communication systems.

The following results are considered practically significant:

- practical implementation of the proposed solution in the form of a software application;
- the possibility of practical use in conjunction with routers in information and communication systems.

The results of the dissertation have been tested and used in the following organisations and institutions

- in X-CITY LLC (Khmelnytskyi) to analyse subscriber traffic using a hybrid method in the information and communication system, which allowed to reduce abnormal traffic, appeals from higher-level providers about attacks coming from the information and communication system of X-CITY LLC and hardware without changing the architecture of the information and communication system of X-CITY LLC. Act of implementation dated 10.04.2025.
- in the educational process of the Department of Cyber Security of Khmelnytsky National University when teaching the disciplines ‘Vulnerability and Intrusion Detection Technologies’, ‘Wireless Network and Internet of Things Security’ and ‘Information Security Monitoring and Management’. Act of implementation dated 25.03.2025.

Keywords: cybersecurity, information security, abnormalities, abnormality detection, vulnerability, cyberattacks, cyber threats, intrusion detection, neural networks, fuzzy logic, fuzzy sets, linguistic variable, network, information and communication system, traffic analysis, model, self-similarity, signature analysis, method, hybrid method, system, Snort, Suricata.

СПИСОК ПУБЛІКАЦІЙ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

Статті у наукових фахових виданнях України

1. Тітова В.Ю., Кльоц Ю.П., Петляк Н.С., Капустян М.В.. Fuzzy inference subsystem for classifying threats to computer information. Вимірювальна та обчислювальна техніка в технологічних процесах. 2022. № 1. С. 57-61

2. Кльоц Ю.П., Петляк Н.С.. Виявлення аномального трафіку у загальнодоступних комп’ютерних мережах. Вимірювальна та обчислювальна техніка в технологічних процесах. 2022. № 3. С. 79-86. DOI: 10.31891/2219-9365-2022-71-3-9

3. Мостовий С.В., Петляк Н.С., Голота І.О.. Дослідження ефективності інструментів виявлення і запобігання вторгнень на вузли в корпоративних мережах. Вимірювальна та обчислювальна техніка в технологічних процесах. 2023. № 2. С. 5-8. DOI: 10.31891/2219-9365-2023-74-1

4. Nataliia Petliak, Yuliia Khokhlachova. METHOD OF ANALYSIS OF OUTGOING TRAFFIC PACKAGE SIGNATURES. Захист інформації. 2024. № 1. С. 179-187. DOI: 10.18372/2410-7840.26.18841

5. Петляк Н.С. Аналіз моделей виявлення аномалій трафіку в сучасних інформаційно-комунікаційних системах та мережах. Вимірювальна та обчислювальна техніка в технологічних процесах. 2025. № 1. С. 180-186. DOI: 10.31891/2219-9365-2025-81-21

6. Петляк Н.С.. Гібридний метод та система виявлення аномального трафіку в інформаційно-комунікаційних системах. Вісник Хмельницького національного університету. Серія: Технічні науки. 2025. №2. С. 561-569. DOI: 10.31891/2307-5732-2025-349-82.

Наукові праці, які додатково відображають наукові результати дисертації

7. Klots Y., Titova V., Petliak N., Cheshun V., Salem A.-B.M.. Research of the

Neural Network Module for Detecting Anomalies in Network Traffic. 3rd International Workshop on Intelligent Information Technologies and Systems of Information Security (IntelITSIS 2022), Khmelnytskyi, 23-25 March 2022. Vol. 3156, P. 378 - 389. ISSN 16130073 (SCOPUS)

8. Petliak N., Klots Y., Titova V., Cheshun V., Boyarchuk A.. Signature-based Approach to Detecting Malicious Outgoing Traffic. 4th International Workshop on Intelligent Information Technologies and Systems of Information Security (IntellTSIS 2023), Khmelnytskyi, 22-23 March 2023. Vol. 3373. P. 486 - 506. ISSN 16130073 (SCOPUS)

9. Klots Y., Petliak N., Titova V.. Evaluation of the efficiency of the system for detecting malicious outgoing traffic in public networks. 13th International Conference on Dependable Systems, Services and Technologies (DESSERT 2023), Athens, 13-15 October 2023. DOI: 10.1109/DESSERT61349.2023.10416502 (IEEE)

10. Klots Y., Petliak N., Martsenko S., Tymoshchuk V., Bondarenko I.. Machine Learning system for detecting malicious traffic generated by IoT devices. nd International Workshop on Computer Information Technologies in Industry 4.0 (CITI 2024), Ternopil, 12 June 2024. Vol. 3742. P. 97 - 110. ISSN 16130073 (SCOPUS)

11. Titova V., Klots Y., Petliak N., Cheshun V., Salem A.-B.M.. Detection of network attacks in cyber-physical systems using a rule-based logical neural network. 1st International Workshop on Intelligent and CyberPhysical Systems (ICyberPhyS 2024), Khmelnytskyi, 28 June 2024. Vol. 3736. P. 255 - 268. ISSN 16130073 (SCOPUS)

12. Petliak N., Klots Y., Titova V., Salem A.-B.M.. Attack detection system based on network traffic analysis by means of fuzzy inference. 1st International Workshop on Advanced Applied Information Technologies (AdvAIT 2024), Khmelnytskyi, 5 December 2024. Vol. 3899. P. 201-213.ISSN 16130073 (SCOPUS)

13. Петляк Н.С., Кльоц Ю.П., Хохлачова Ю.Є.. Виявлення аномального трафіку у загальнодоступних комп’ютерних мережах. Інформаційна, функційна і кібербезпека: тези доп. II студ. наук.-техн. к. (м. Харків, 30 листопада – 1 грудня 2022 р.) Х., 2022. С. 45-46

14. Петляк Н.С., Кльоц Ю.П.. Підхід до аналізу вихідного трафіку на основі сигнатур. Інформаційна безпека та комп'ютерні технології: тези доп. VI міжнар. наук.-пр. конф. (м. Кропивницький, 20-21 квітня 2023 р.) Кр., 2023. С. 3-4

15. Петляк Н.С., Кльоц Ю.П., Хохлачова Ю.Є.. Підхід до аналізу вихідного трафіку. Безпека інформаційних технологій: тези доп. XII міжнар. наук.-техн. конф. (м. Ужгород, 2-4 травня 2023 р.). К., 2023. С. 97-99

16. Петляк Н.С., Кльоц Ю.П., Тітова В.Ю., Чешун В.М. Виявлення зловмисного вихідного трафіку мережі на основі нечіткого логічного висновку. Захист інформації і безпека інформаційних систем: тези доп. IX міжнар. наук.-техн. конф. (м. Львів, 25-26 травня 2023 р.) Л., 2023. С. 33-35

17. Петляк Н.С., Кльоц Ю.П.. Метод ідентифікації вторгнень на основі алгоритму визначення самоподібності трафіку та алгоритмів нечіткої логіки. Безпека інформаційних технологій: тези доп. XIII міжнар. наук.-техн. конф. (м. Львів, 9-11 травня 2024 р.) Л., 2024. С. 117-119.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

ІКС - інформаційно-комунікаційна система

ЛЗ – лінгвістична змінна

ФН – функція належності

ПЗ - програмне забезпечення

НСД - несанкціонований доступ

IoT - Інтернет речей

НЛ - нечітка логіка

МН - машинне навчання

ВСТУП

Обґрунтування теми дослідження. У сучасному цифровому середовищі забезпечення інформаційної безпеки є пріоритетним завданням для більшості організацій, незалежно від їх масштабу та галузі діяльності. Одним із найважливіших напрямів у цьому контексті є виявлення аномалій у мережевому трафіку, яке відіграє вагому роль у попередженні загроз, що не мають явних ознак, притаманних відомим шаблонам атак.

На відміну від традиційного мережевого моніторингу, який переважно орієнтований на оцінку стану систем та продуктивності мережевих ресурсів, виявлення аномалій передбачає глибокий аналіз поведінки об'єктів у мережі з метою ідентифікації відхилень від типових характеристик. Такий підхід дозволяє виявляти приховані загрози, що не піддаються виявленню за допомогою класичних сигнатурних методів, та своєчасно реагувати на потенційні інциденти безпеки.

На сьогоднішній день одним з основних інструментів мережевого моніторингу є системи виявлення вторгнень, які дозволяють захоплювати мережевий трафік і представляти його у вигляді потоку даних, і застосовуючи до нього певні правила виявляти аномалії у потоці. Вони зазвичай застосовуються в межах концепції інформаційного захисту, здатні здійснювати захоплення й аналіз мережевого трафіку в реальному часі. Проте сигнатурному методу притаманні наступні недоліки: неможливість виявляти нові несанкціоновані впливи, що не зустрічалися раніше; нестійкість до модифікацій вже відомих атак; нездатність визначати розподілені у часі атаки та аномалії. Такий підхід виявляється малоефективним у випадках нових або модифікованих атак, що не мають точного відповідника в сигнатурному профілі.

Отже, реалізація системи виявлення аномального трафіку, як самостійного або комбінованого з сигнатурним аналізом продукту, є актуальною науковою задачею.

Вирішенню даної проблеми присвячені роботи провідних науковців, в

Україні та за кордоном, а саме: Бурячка В.Л., Смірнова О. А., Конопліцької-Слободенюк О. К., Єфименка А.А., Гнатюка С.О., Корченко А.О., Noemí Marta Fuentes-García, José Camacho, Gabriel Maciá-Fernández та інших.

Проте більшість існуючих досліджень присвячені моніторингу вхідного трафіку, у той час, як вихідний мережевий трафік теж може нести в собі загрози цілісності та конфіденційності мережі, а також слугувати джерелом для зловмисної діяльності на інші мережі, зокрема використовуючись для організації хакерських атак або атак на відмову в обслуговуванні.

DDoS-атаки (розподілена відмова в обслуговуванні) є одним з найпоширеніших і небезпечних видів кіберзагроз, які продовжують зростати в частоті та потужності. Наприклад, у першій половині 2024 року зафіксовано 445000 інцидентів, пов’язаних з DDoS-атаки, що на 46% більше, ніж у першій половині 2023 року. Одна така атака коштує бізнесу близько 6130 доларів за хвилину простою, враховуючи втрати в доходах та витрати на відновлення. Також, подібні атаки можуть завдати шкоди репутації компанії, знижуючи довіру клієнтів.

Отже, можна зробити висновки, що такі атаки продовжують бути серйозним викликом для бізнесу в усіх сферах. Підвищення їх частоти та потужності вимагає розробки та впровадження надійних стратегій захисту. А тому, розробка системи виявлення аномалій у вихідному мережевому трафіку є актуальною на сьогоднішній день науковою задачею, вирішенню якої і присвячене дане дослідження.

Мета і завдання дослідження. Метою дисертаційної роботи є підвищення достовірності виявлення аномального трафіку в інформаційно-комунікаційних системах шляхом розроблення гібридного методу виявлення аномального трафіку в інформаційно-комунікаційних системах, який враховує поведінкові характеристики користувачів і порушників. Для досягнення поставленої мети необхідно вирішити такі завдання:

– провести аналіз сучасних методів та моделей виявлення аномального трафіку в інформаційно-комунікаційних системах і моделей типового користувача та потенційного порушника;

– вдосконалити модель сигнатури пакету з урахуванням інформативності полів заголовку та пакету щодо визначення їх аномальності;

– вдосконалити модель процесу визначення аномального трафіку на основі самоподібності з урахуванням потенційних загроз та особливостей трафіку в інформаційно-комунікаційних системах;

– вдосконалити модель процесу нечіткого визначення аномального трафіку, що враховує поведінку типового користувача та потенційного порушника;

– розробити гібридний метод визначення аномального трафіку на основі нечіткого аналізу, визначеного як не самоподібний та імплементувати його в архітектуру системи захисту мережі;

– розробити структурну модель системи виявлення аномального трафіку в інформаційно-комунікаційних системах;

– розробити алгоритмічне та програмне забезпечення системи виявлення аномального трафіку та провести експериментальні дослідження.

Об’єкт дослідження: процес виявлення аномального трафіку в інформаційно-комунікаційних системах.

Предмет дослідження: моделі та методи виявлення аномального трафіку в інформаційно-комунікаційних системах.

Методи дослідження. Для розв’язання поставлених задач використовуються основні положення методів аналізу даних, нечіткої логіки, теорії графів, теоретико-множинний підхід, статистичного аналізу даних, експертні оцінки, теорії множин та теорії комп’ютерних мереж.

Наукова новизна одержаних результатів:

1. *Вдосконалено* модель сигнатури пакету для пошуку аномального трафіку, що за рахунок виключення з параметрів сигнатури розміру заголовка, контрольної суми заголовку, корисного розміру пакета, мітки потоку, пріоритету пакету, контрольної суми пакету за принципом Парето, забезпечило зменшення часу аналізу трафіку.

2. *Вдосконалено* модель процесу нечіткого виявлення аномального трафіку, в якій за рахунок використання експертного підходу сформована множина правил та набір відповідних лінгвістичних змінних, що дозволило розробити нові підходи до виявлення аномального трафіку в інформаційно-комунікаційних системах.

3. *Вперше* розроблено гібридний метод виявлення аномального трафіку в інформаційно-комунікаційних системах в якому за рахунок інтеграції методу класифікації трафіку за ознаками, методу самоподібності та розробленої моделі процесу нечіткого виявлення дозволило динамічно формувати множини сигнатур при класифікації трафіку за ознаками та підвищити показники виявлення аномального трафіку.

4. *Удосконалено* структурну модель системи виявлення аномального трафіку в інформаційно-комунікаційних системах, що за рахунок інтегрування розробленого гібридного методу виявлення аномального трафіку в інформаційно-комунікаційних системах та динамічного варіювання множиною дозволених та заборонених з'єднань у режимі реального часу дозволило зменшити навантаження на процесор.

Особистий внесок автора. Всі результати, які становлять основний зміст дисертації, автор отримав особисто. У роботах, написаних у співавторстві, автору дисертації належить: [1] – описано підсистему логічного виведення результатів для реалізації набору правил; [2] – проведено аналіз статистики кіберінцидентів та запропоновано структуру ІКС з використанням системи виявлення вихідного аномального трафіку; [3] – налаштування тестового середовища та порівняння результатів роботи систем Snort та Suricata, що дозволило обрати оптимальний варіант для інтегрування розробленого гібридного методу; [4] – описано вдосконалені модель сигнатури пакету для пошуку аномального трафіку за принципом Парето та модель процесу нечіткого виявлення аномального трафіку; [5] – проведено аналіз наявних моделей виявлення аномалій мережевого трафіку в ІКС; [6] – розроблено гібридний метод виявлення аномального трафіку в ІКС за рахунок інтеграції методу класифікації трафіку за ознаками, методу самоподібності та нечіткого методу та запропоновано удосконалену структурну модель системи виявлення аномального трафіку в ІКС; [7] – запропоновано модифікацію системи виявлення вторгнень Snort для можливості аналізу вихідного трафіку; [8,16] - вдосконалено модель сигнатури пакету для пошуку аномального трафіку з оптимізацією параметрів за принципом Парето; [9, 12] – виконано моделювання процесу нечіткого виявлення аномального трафіку; [10] – запропоновано рішення аналізу вихідного трафіку для пошуку ботів та проведено аналіз наявних наборів даних; [11] – запропоновано метод виявлення мережевих атак в локальній мережі та проведено розрахунок достовірності пропонованого рішення; [13] – проаналзовано наявні рішення систем аналізу мережевого трафіку та сформовано проблемні задачі при їх використанні, [14] – описано алгоритм роботи нечіткого методу виявлення аномального трафіку, [15] – запропоновано та описано комбінацію роботи методу класифікації трафіку за ознаками та нечіткого методу виявлення аномального трафіку; [17] - запропоновано та описано комбінацію роботи методу класифікації трафіку на основі самоподібності та нечіткого методу виявлення аномального трафіку.

Апробація результатів дисертації. Результати дисертації доповідались та обговорювались на конференціях, серед яких: ІІ студентська науково-технічна конференція «Інформаційна, функційна та кібербезпека» (СКІФіК-2022) (Харків, 2022), The International Workshop on Intelligent Information Technologies & Systems of Information Security (Хмельницький, 2022), VI Міжнародна науково-практична конференція «Інформаційна безпека та комп’ютерні технології» (Кропивницький, 2023), XII Міжнародна науково-технічна конференція ITSec: Безпека інформаційних технологій (Ужгород, 2023), IX Міжнародна науково-практична конференція Захист інформації і безпека інформаційних систем (Львів, 2023), 13th International Conference on Dependable Systems, Services and Technologies (Athens, 2023), XIII Міжнародна науково-технічна конференція ITSec: Безпека інформаційних технологій (Львів, 2024), 4the International Workshop on Intelligent Information Technologies & Systems of Information Security (Хмельницький, 2024), 1st International Workshop on Advanced Applied

Information Technologies (Khmelnytskyi, 2024), 1st International Workshop on Intelligent and CyberPhysical Systems (Khmelnytskyi, 2024), 2nd International Workshop on Computer Information Technologies in Industry 4.0 (Ternopil, 2024).

Публікації. Основні наукові результати дисертаційної роботи опубліковані у 17 наукових працях, із них: 4 наукових статей, надрукованих у вітчизняних фахових виданнях, 3 статей включених до наукометричної бази Scopus, 5 тез доповідей на міжнародних і науково-практичних конференціях.

Зв'язок роботи з науковими програмами, планами, темами. Дослідження проведено в межах науково-дослідної роботи, яка велася в Національному авіаційному університеті за темою: «Моделі кіберзахисту інформаційних систем» (Державний реєстраційний номер: 0122U201817 від 24.12.2022).

Практичне значення одержаних результатів. Запропоновано удосконалення методу виявлення аномального трафіку на основі самоподібності та нечіткий метод визначення аномального трафіку, що дозволяє покращити достовірність виявлення аномального трафіку. Розроблено гібридний метод виявлення аномального трафіку в інформаційно-комунікаційних системах, який забезпечив зменшення аномального трафіку без додаткових витрат на обладнання. Впроваджено в навчальний процес для проведення лабораторних занять матеріали щодо моделей та методів виявлення аномалій в сучасних інформаційно-комунікаційних системах.

Практична цінність роботи підтверджена актами впровадження від Хмельницького національного університету (акт від 25.03.2025 р.) та від ТОВ «X-CITY» (акт від 10.04.2025 р.).

Структура і обсяг дисертації. Дисертація складається з анотації, переліку умовних позначень, вступу, 4 розділів, висновків, списка використаних джерел, 3 додатки. Загальний обсяг дисертації становить 154 сторінок, в тому числі 46 рисунків, 14 таблиць та 104 джерела використаної літератури. Основний текст роботи викладено на 124 сторінках..

РОЗДІЛ 1
АНАЛІЗ ВИКОРИСТАННЯ І ОБҐРУНТУВАННЯ НАПРЯМІВ ВДОСКОНАЛЕННЯ МОДЕЛЕЙ ТА МЕТОДІВ ВИЯВЛЕННЯ АНОМАЛІЙ ТРАФІКУ В ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ СИСТЕМАХ

1.1 Аналіз загроз безпеки інформації в сучасних інформаційно-комунікаційних системах та мережах

Загрози інформаційній безпеці в сучасних системах і мережах є одними з найбільших викликів для організацій, урядів та окремих користувачів [1-3]. У сучасну епоху стрімкої цифрової трансформації, будь-яка сфера — від збереження приватної інформації до функціонування критичної інфраструктури — може стати об'єктом кібератаки [4-6]. Атаки на інформаційні системи можуть мати серйозні наслідки, включаючи фінансові збитки, витоки конфіденційної інформації, порушення роботи систем і навіть загрозу національній безпеці [7-9]. Одним із видів загроз є кібератаки, зокрема хакерські атаки, які використовують вразливості в програмному забезпеченні (ПЗ) або недостатню безпеку мережі [10-11]. Зловмисники можуть отримати несанкціонований доступ (НСД) до даних, змінювати або видаляти їх, впроваджувати шкідливе ПЗ, такі як віруси, трояни чи програми-вимагачі, які блокують доступ до системи або даних і вимагають викупу. Збільшення кількості таких атак, зокрема на організації, що обробляють великі обсяги чутливої інформації, як-от банки, медичні установи та урядові структури, створює значні ризики для безпеки та конфіденційності. Суттєву небезпеку становить і соціальна інженерія – сукупність психологічних прийомів, за допомогою яких зловмисники змушують людей добровільно надати конфіденційні відомості або відкрити доступ до систем [12-14]. Фішинг, одна з найбільш популярних форм соціальної інженерії, використовує фальшиві електронні листи чи вебсайти, щоб обманути користувачів і змусити їх передати особисті дані, паролі або фінансову інформацію [15-17]. Також важливою є проблема безпеки мобільних і розподілених систем [18]. З розширенням використання мобільних гаджетів і переходом до дистанційного формату роботи, питання захисту даних у таких середовищах набуло ще більшої складності. [19]. Відсутність належного контролю за мобільними пристроями та хмарними сервісами може призвести до витоків важливої інформації та кібератак на всю організацію [20-21].

Кіберстатистика є важливим інструментом для аналізу кіберзагроз, оцінки ризиків і розробки ефективних стратегій кібербезпеки. Вона охоплює збір, аналіз і інтерпретацію даних про інциденти, загрози, вразливості та тренди, дозволяючи фахівцям і організаціям приймати обґрунтовані рішення [22]. Відповідно до звітів Оперативного центру реагування на кіберінциденти Державного центру кіберзахисту Державної служби спеціального зв’язку та захисту інформації України за 2021-2023 роки можна зробити наступні висновки. У 2021 році було оброблено 650 мільярдів подій, у 2022 році цей показник знизився до 58 мільярдів, а у 2023 році — до 18 мільярдів, що свідчить про вдосконалення методів фільтрації та зосередження на більш значущих подіях [23]. Водночас кількість підозрілих подій у 2022 році різко зросла до 181 мільйона (у 4,4 рази більше, ніж у 2021-му), але в 2023 році знизилася до 133 мільйонів, що можна пояснити покращенням автоматичного аналізу. Кількість критичних подій залишалася відносно стабільною між 2021 і 2022 роками (160 тисяч і 179 тисяч відповідно), однак у 2023 році вона скоротилася до 148 тисяч завдяки превентивним заходам. Кількість зареєстрованих кіберінцидентів стрімко зростала: зі 147 у 2021 році до 415 у 2022-му і до 1105 у 2023-му, що свідчить про посилення атак на критичну інфраструктуру [24-26].

Аналіз бази даних загальновідомих вразливостей інформаційної безпеки демонструє, що станом на серпень 2024 року у світі було виявлено понад 52000 нових поширених вразливостей у безпеці (CVE), що підкреслює стрімке зростання кібератак. Найбільша кількість вразливостей була зафіксована у 2023 році — понад 29000 випадків. Особливу загрозу становлять програми-вимагачі, які стали найпоширенішим типом кібератак, зокрема у 2023 році [27].

У період з 2005 по 2023 рік у США спостерігався різкий ріст кількості компрометацій даних, зокрема випадків зламу, витоку та викриття даних. У 2023 році постраждало понад 353 мільйони осіб, чия конфіденційна інформація опинилася у НСД. Ця динаміка вказує на зростаючу вразливість до кіберзагроз, особливо в галузях, що зберігають великий обсяг чутливої інформації [28].

Аналізуючи статистику кібербезпеки за останні роки, можна помітити постійне зростання масштабів кібератак та їхнього впливу на організації та приватних осіб. У 2023 році щорічна кількість кібератак сягнула 800000, що свідчить про значне зростання порівняно з попередніми роками. Однією з основних кіберзагроз залишається електронна пошта, через яку доставляється понад 92% шкідливого ПЗ, а кількість нових шкідливих програм щодня сягає 300000. У 2022 році було зафіксовано понад 255 мільйонів фішингових атак лише за пів року, а загальна кількість атак зловмисного ПЗ перевищила 2,8 мільярда. Програми-вимагачі та атаки на об'єкти Інтернету речей (IoT) також демонструють стрімке зростання. Наприклад, у 2020 році понад 20% організацій зазнали атак на IoT-пристрої, тоді як у 2022 році кількість порушень IoT вже перевищила 1,5 мільярда за пів року. Загалом, ці дані демонструють, що кіберзагрози постійно розвиваються, стаючи більш масштабними та складними. Статистика кібербезпеки малого та середнього бізнесу свідчить про критичний рівень вразливості цієї категорії підприємств перед кіберзагрозами. Значна частина компаній, попри використання антивірусного ПЗ, не здатна захиститися від зловмисних атак, що демонструє обмежену ефективність базових захисних заходів [29].

Аналіз статистики показує, що середня вартість кібератак у 2024 році зросла на 10% порівняно з 2023 роком і становить $4,88 мільйона, досягнувши найбільшого стрибка після пандемії. Головними чинниками стали зростаючі витрати на бізнес-перерви та заходи з реагування на атаку, зокрема обслуговування клієнтів після інциденту та штрафи за регуляторні порушення [30].

Рекордна кількість вразливостей нульового дня, зафіксована у 2021 році, свідчить про еволюцію кіберзагроз і зміну динаміки їх використання. Зростання до 80 таких вразливостей, що більш ніж удвічі перевищує попередній рекорд, вказує на зростаючу складність і обсяги підключених до Інтернету систем, зокрема хмарних технологій, мобільних пристроїв і IoT. Близько третини атак у 2021 році мали фінансове підґрунтя, що вказує на комерціалізацію експлойтів через брокерів і приватні компанії. Сукупність цих факторів підкреслює необхідність адаптації кібербезпеки до нових загроз, які стають все більш поширеними та різноманітними [31].

Звіт NETSCOUT за перше півріччя 2024 року демонструє значне ускладнення ландшафту загроз DDoS, що проявляється як у збільшенні частоти атак, так і в їхній складності. Особливо помітним є зростання атак на прикладному рівні на 43% порівняно з 2023 роком, що підкреслює їхню адаптацію до сучасних мережевих умов. Стрімке включення нових мереж у механізми DDoS за короткі терміни, поряд із суттєвим зростанням інфікованих пристроїв, свідчить про швидку еволюцію інфраструктури ботнетів [32]. Усе це разом вимагає від організацій посиленого підходу до захисту, зокрема інвестицій у передові системи виявлення загроз, які можуть впоратися з новими викликами в кібербезпеці [33].

Найпоширенішим наслідком кібератак є НСД до конфіденційної інформації, що призводить до витоків даних. Цей тип інциденту зростає, як показує статистика, і може мати серйозні репутаційні та фінансові наслідки для компаній. Витік даних, пов’язаний з порушенням облікових записів, може призвести до ще більш серйозних наслідків, таких як крадіжка коштів або збої в операційних системах. Загалом, результативність кіберзахисту в фінансовому секторі має критичне значення для зменшення наслідків атак, оскільки будь-яка атака може не тільки призвести до фінансових втрат, а й поставити під загрозу репутацію компанії та її взаємини з клієнтами [34].

Check Point Research оприлюднила дані щодо глобальних тенденцій кібератак у 2022 році, які показали зростання кількості атак на 38% порівняно з 2021 роком. У середньому організації зазнавали 1168 атак щотижня, що є історичним максимумом.

Найбільше атак зазнали галузі освіти та досліджень, урядові структури та охорона здоров'я, причому сектор освіти збільшив кількість атак на 43%, з 2314 атак на організацію щотижня. Однією з тенденцій стало зростання атак на інструменти для бізнес-співпраці, такі як Slack, Teams, OneDrive і Google Drive, а також на навчальні заклади, які не були готові до масового переходу на онлайн-навчання через пандемію [35].

У [36] представлено аналіз наявних рішень, технологій і методів, які можуть сприяти організаціям та приватним користувачам у протидії соціотехнічним загрозам. Результати дослідження свідчать, що основну роль у зменшенні ризику успішних атак відіграють освітні та організаційні заходи. Крім того, важливим залишається постійне вдосконалення існуючих засобів безпеки та впровадження інноваційних технологій для підвищення захищеності інформаційних систем.

Аналіз сучасного стану кібербезпеки вказує на низку проблем, які залишаються актуальними для організацій у всьому світі. Різноманітність і складність кібератак постійно зростають, включаючи атаки на локальні мережі підприємств, використання соціальної інженерії та експлуатацію вразливостей ПЗ. Активне впровадження IoT, хмарних технологій і мобільних пристроїв створює нові вектори загроз, які часто залишаються недооціненими через слабкі заходи захисту [37].

1.2 Аналіз моделей виявлення аномалій трафіку в сучасних інформаційно-комунікаційних системах та мережах

Сучасні інформаційно-комунікаційні системи (ІКС) є невід'ємною складовою життєдіяльності суспільства, забезпечуючи комунікацію, обробку та передачу даних у реальному часі. Водночас із ростом складності та масштабів мережевих структур зростає і кількість загроз їхній безпеці [38-39]. Аномалії у трафіку, які можуть бути індикаторами кібератак, технічних несправностей або збоїв у роботі систем, становлять серйозну проблему для забезпечення стабільної та безперебійної роботи мереж [40-42]. А виявлення аномалій у мережевому трафіку є важливим аспектом забезпечення кібербезпеки та ефективного функціонування ІКС. Основна складність цього завдання полягає у різноманітності типів аномалій, великих обсягах даних, що генеруються мережею, та необхідності оперативного виявлення можливих загроз [43].

Аналіз існуючих моделей виявлення аномалій трафіку є важливим завданням у контексті забезпечення інформаційної безпеки ІКС та адаптації до нових викликів. Вивчення існуючих рішень дозволяє уникнути дублювання зусиль, оскільки розробники отримують змогу зосередитися на вирішенні нових проблем, враховуючи вже наявні напрацювання. Крім того, аналіз існуючих моделей сприяє виявленню їхніх потенційних слабких місць, таких як вразливість до певних типів атак чи обмеження в масштабованості. Одним із аспектів аналізу є можливість інтеграції найкращих практик, зокрема інноваційних рішень та підходів, які продемонстрували ефективність у реальних умовах. Застосування таких ідей, наприклад, гібридних алгоритмів, що забезпечують обробку даних у реальному часі, може значно покращити результати.

Робота [44] стосується аналізу мережевого трафіку для виявлення атак SQL-ін'єкцій у мережевих потоках, використовуючи протокол NetFlow V5. Дослідники зосередилися на побудові моделей машинного навчання (МН) для автоматичного виявлення атак на основі даних про потоки. У дослідженні було створено два набори даних: для тренування та тестування моделей. Вони збирали інформацію про трафік, створений під час атак SQL-ін'єкцій та звичайних запитів. Потоки NetFlow аналізувалися з використанням алгоритмів МН, таких як лінійна регресія, методи ансамблю та алгоритми підтримки векторів. Дослідження підкреслює важливість адаптації моделей до різних умов, таких як використання різних баз даних та атак різних типів. .

У дослідженні [45] пропонується модель для виявлення шкідливого ПЗ в операційній системі Android. Використовується підхід на основі нечіткої логіки

(НЛ) та методів МН, які включають шість класифікаторів: логістичну регресію, нейронну мережу, підтримуючі векторні машини, деревоподібні моделі, імовірнісний байєсівський метод та метод ансамблів дерев рішень. Головна мета роботи — інтеграція результатів роботи зазначених моделей через нечітку систему виведення, щоб ефективно і точно виявляти шкідливі програми. Це досягається завдяки розподілу обчислень між класифікаторами, обранню найкращих результатів голосування та подальшому об'єднанню їх через НЛ. Використання кількох моделей та їх комбінація збільшують обчислювальне навантаження, однак експериментальні результати демонструють високу ефективність.

У роботі [46] обговорюється створення та тестування гібридної моделі WK-FNN для виявлення аномалій у трафіку комп'ютерних мереж. Модель побудована на основі двох класифікаторів — зваженого k-найближчого сусіда і прямого нейронного шару. У моделі WK-FNN два класифікатори паралельно оцінюють трафік, а потім їх результати обробляються XOR-блоком. Таким чином, модель WK-FNN спрямована на виявлення аномалій у режимі реального часу та формування сповіщень про потенційні загрози.

Робота [47] присвячена розробці системи виявлення атак Multi-Channel Man-in-the-Middle (MC-MitM), яка загрожує безпеці Wi-Fi мереж, особливо в контексті Інтернету речей. Система працює на основі пасивного аналізу мережевого трафіку та використовує сигнатурний підхід для виявлення специфічних ознак атак MC-MitM. Пропоноване рішення є сумісним із різними операційними системами та підтримує інтеграцію з іншими системами безпеки.

Запропонована у [48-49] модель базується на рекурентних нейронних мережах і спеціалізується на виявленні аномалій у мережах IoT. У її основі лежать три підходи: Long Short-Term Memory (LSTM), Bidirectional Long Short-Term Memory (BiLSTM) та Gated Recurrent Unit (GRU). Пропонована модель використовує оптимізовану архітектуру для багатокласової класифікації та бінарного розпізнавання.

Корнієнко та інші [50] аналізують моделі самоподібного мережевого трафіку, які використовуються для виявлення аномалій у системах виявлення та запобігання атак. Вони використовують комбінацію методів штучного інтелекту, таких як нейронні мережі, а також методи глобальної та локальної оптимізації, включаючи генетичні алгоритми, метод імітації відпалу та інші. Для створення моделей застосовуються критерії мінімуму зсуву і регулярності, які забезпечують точність і надійність виявлення аномалій у реальному часі.

Гайдур та інші [51] пропонують модель виявлення шкідливої активності в інформаційних системах організацій, яка базується на гібридній класифікації. Основою цієї моделі є метод ансамблевого навчання, зокрема стекінг. Модель спрямована на підвищення точності виявлення ботнетів шляхом аналізу мережевого трафіку, включаючи моніторинг DNS-запитів, використання honeypots та виявлення аномалій у системних ресурсах.

Запропонована у [52] модель нечіткої нейронної мережі призначена для виявлення аномалій у даних через гібридний підхід, що поєднує нейронні мережі та НЛ. Особливістю моделі є здатність витягувати знання з великих даних і перетворювати їх на набір інтерпретованих лінгвістичних правил. Це робить її придатною для управління знаннями в системах кіберзагроз і розробки програмованих пристроїв для інформаційного захисту.

Запропонована модель виявлення аномалій на основі МН [53] використовує гібридний підхід із застосуванням кластеризації k-середніх та алгоритму SMO для ідентифікації аномалій у мережевих даних. Завдяки цьому підходу модель зменшує обсяг навчальних даних та оптимізує процес навчання. Наступним етапом є класифікація, яка виконується алгоритмом SMO для поділу трафіку на нормальний і аномальний. Така комбінація дозволяє ефективно працювати навіть із великими обсягами даних, характерними для IoT і WSN.

Автори [54] пропонують модель для виявлення мережевих аномалій базується на п’ятишаровому автокодері, який використовує механізм реконструкції даних для розпізнавання аномальних шаблонів мережевого трафіку. Архітектура моделі складається з двох основних етапів — кодування та декодування, що дозволяє стискати вхідні дані до латентного простору з меншим розміром і відновлювати їх із мінімальною втратою інформації. Особливістю моделі є її 5-рівнева структура [122-32-5-32-122], яка дозволяє ефективно зменшувати розмірність даних і відображати їх у компактному латентному просторі. Модель оснащена методологією попередньої обробки даних, яка включає кодування категоріальних змінних, видалення викидів і нормалізацію даних, що зменшує дисбаланс у наборі функцій та покращує продуктивність.

У [55] запропоновано модель виявлення аномалій у мережевому трафіку яка базується на ансамблі механізмів навчання та прогнозування. Її метою є підвищення точності виявлення мережевих вторгнень через поєднання підходів автоматизованого МН та алгоритмів фільтра Калмана. У рамках моделі реалізовано автоматизований пошук архітектури нейронної мережі, що дозволяє оптимізувати гіперпараметри та створювати оптимальні моделі глибокого навчання.

Модель виявлення аномалій у кіберфізичних системах [56] спрямована на покращення детекції атак на цілісність даних. Методологія базується на принципі зворотного аналізу в часі, що дозволяє виявляти аномалії, аналізуючи стан системи на момент до початку атаки. Для цього використовуються віртуальні значення стану, які оцінюються за допомогою $H\infty$-фільтра згладжування. Головним елементом моделі є генерація залишкових величин і порогів, які визначають, чи є виявлені зміни результатом атаки. Підхід із використанням $H\infty$-фільтра згладжування є обчислювально затратним.

Представлена у [57] модель розширює можливості XAI на неконтрольоване навчання. Вона дозволяє отримати зрозумілі правила для OneClass SVM (OCSVM), що відкриває можливість застосування в реальних умовах. Запропонована функція для агрегування метрик є базовою, і її можна вдосконалити для досягнення кращого балансу між різними аспектами якості.

Sajad Einy у [58] пропонує модель, що інтегрує системи виявлення вторгнень, які базуються на аномаліях та сигнатурах, з гібридним підходом до аналізу, щоб покращити захист мережі. Це рішення поєднує методи МН, зокрема нейронні мережі та НЛ, для ідентифікації та запобігання різним видам кіберзагроз. Використовуючи Suricata IDS/IPS разом із нейронною мережею, система генерує правила для виявлення відомих атак і обробляє нерозпізнаний трафік за допомогою НЛ для виявлення аномалій. Однак, система залежить від навчальних даних і потребує регулярного оновлення, щоб виявляти нові загрози.

STEP-GAN є моделлю для виявлення аномалій, яка базується на генеративних змагальних мережах. Особливістю STEP-GAN є її здатність використовувати кілька генераторів, які кожен моделює різні вузли в розподілі даних. Ефективність STEP-GAN залежить від припущення, що нормальні дані в майбутньому будуть розподілені подібно до навчального набору. А в разі змін у характеристиках нормальних даних потрібна додаткова адаптація моделі [59].

Shaojun Sheng [60] пропонує модель виявлення аномалій мережевого трафіку із використанням хаотичної нейронної мережі, яка інтегрує механізми попередньої обробки даних, адаптивної вибірки та вибору ключових ознак для покращення якості класифікації. Основна ідея моделі полягає у використанні хаосу для моделювання складних і нелінійних взаємозв’язків у даних мережевого трафіку, що дозволяє краще фіксувати приховану інформацію між ключовими характеристиками даних. Недоліком моделі є її складність і потреба в значних обчислювальних ресурсах для навчання.

Модель згорткової нейронної мережі, яку представлено у [61], призначена для виявлення атак типу Bruteforce-SSH і Bruteforce-FTP у мережевому трафіку. Особливістю моделі є її структура, яка включає п’ять типів шарів: Conv1D для обчислення скалярного добутку між вагами і вхідними сигналами, BatchNormalization для нормалізації виходів попередніх шарів, MaxPooling1D для зменшення розмірності даних, Flatten для перетворення багатовимірного тензора в одновимірний вектор, і Dense для лінійного перетворення вхідних даних. Функції активації (ReLU, SiLU, Sigmoid) були досліджені разом із трьома оптимізаторами (Adagrad, Adam, Nadam), що дозволило створити дев’ять варіантів моделі. Однак суттєвим недоліком є значна залежність точності від якості попередньої обробки даних, що використовуються для навчання.

У [62] пропонується модель виявлення аномалій у даних, яка базується на виявленні відхилень у поведінці, що не відповідають легітимним процесам, шляхом порівняння моделі з заданими властивостями. На першому етапі для кожного користувача створюється автомат, який відображає його операції з базою даних на основі журналів, включаючи часові мітки, типи операцій і дані, з якими вони взаємодіяли. На другому етапі система перевіряється на відповідність заданим властивостям. Якщо відхилення виявлені, система вважається аномальною. Перевагою є адаптивність моделі, оскільки вона може бути застосована до різних типів даних, включаючи великі дані та гетерогенні джерела. Генерація автоматів і перевірка властивостей є складними процесами, які можуть бути трудомісткими для масштабних баз даних.

В роботі [63] запропоновано модель виявлення низькошвидкісних DDoS-атак у програмно-конфігурованих мережах за допомогою онлайн-методів машинного навчання. Основу моделі становить алгоритм класифікації Passive-Aggressive, що дозволяє адаптуватися до змін у мережевому трафіку в режимі реального часу. У моделі реалізовано етапи попередньої обробки даних, які включають видалення нерелевантних ознак, нормалізацію та підготовку даних до навчання. Проте вона потребує додаткових досліджень для забезпечення універсальності та ефективності у ширшому спектрі реальних сценаріїв.

Anzhelika Mezina та інші [64] пропонують дві моделі для виявлення аномалій у мережевому трафіку: U-Net та Temporal Convolutional Network (TCN) у поєднанні з Long Short-Term Memory. Модель U-Net адаптована для класифікації мережевого трафіку. Модель Temporal Convolutional Network у поєднанні з LSTM використовує дві окремі гілки для аналізу часових рядів. TCN виділяє довготривалі патерни за допомогою ієрархії темпоральних фільтрів, тоді як LSTM дозволяє враховувати контекст попередніх станів. До недоліків можна віднести складність реалізації та навчання моделі TCN+LSTM через використання двох гілок.

Таблиця 1.1 представляє порівняльний аналіз моделей для виявлення аномалій і атак у мережевому трафіку за визначеними критеріями [65].

Таблиця 1.1 – Порівняння моделей відповідно до заданих критеріїв

Критерій / Джерело	Критерії				
№	Кількість типів атак	Обчислювальна складність («-» складний, «+-» середній, «+» легкий)	Простота налаштування та впровадження	Відсутність впливу на продуктивність мережі («-» впливає, «+» не впливає)	Аналіз вихідного трафіку
[44]	1	+-	-	-	-
[45]	1	-	-	-	-
[46]	3	-	+	-	-
[47]	1	+	+	-	-
[48]	1	+-	-	-	-
[50]	1	-	+	-	-
[51]	1	-	-	+	-
[52]	2	+-	+	-	-
[53]	1	+	-	+	-
[54]	1	+-	-	-	-
[55]	1	+-	-	-	-
[56]	1	-	-	-	-
[57]	1	+-	-	+	-
[58]	3	+-	+	-	-
[59]	1	-	-	+	-
[60]	1	+-	-	-	-
[61]	2	-	-	-	-
[62]	1	+-	-	-	-
[63]	1	+-	+	-	-
[64]	1	-	-	+	-

Наявні моделі мають низку переваг та недоліків, які визначають їхню ефективність у різних контекстах. Більшість моделей здатні аналізувати вхідний трафік. Водночас ці моделі мають і певні недоліки. Наприклад, складність їхнього налаштування часто потребує залучення висококваліфікованих фахівців, що може бути викликом для організацій із обмеженими ресурсами. Деякі підходи зосереджені лише на окремих видах атак, що обмежує їхнє застосування в умовах більш комплексних загроз. Наприклад, автоенкодери, хоча і є сучасним методом, демонструють нижчу точність у порівнянні з іншими моделями, що може бути критичним у системах із підвищеними вимогами до безпеки.

Подальші дослідження та розробки моделей для виявлення аномалій у мережевому трафіку доцільно зосередити на кількох напрямах, які відповідають викликам кібербезпеки та потребам ІКС. Одним із завдань є вдосконалення адаптивності моделей до нових, раніше невідомих атак. Це передбачає активне впровадження методів самонавчання та безперервного оновлення знань моделей у реальному часі. Наприклад, розробка гібридних моделей, які об'єднують переваги МН, глибокого навчання та НЛ, дозволить створювати рішення, що одночасно є точними та гнучкими. Варто також досліджувати способи зниження навантаження на мережеву інфраструктуру, що може бути досягнуто за рахунок оптимізації алгоритмів обробки даних, впровадження розподілених обчислювальних систем або використання хмарних технологій. Такі підходи дозволять забезпечити обробку великих обсягів трафіку без зниження продуктивності системи. Таким чином, подальші дослідження мають зосереджуватися на розробці адаптивних, масштабованих і високоточних моделей, здатних працювати в умовах реального часу та змінюваних умов сучасних мережевих середовищ.

1.3 Аналіз методів виявлення аномалій трафіку в сучасних інформаційно-комунікаційних системах та мережах

Вибір методів для виявлення аномалій є суттєвим аспектом у забезпеченні ефективності та точності аналізу. Аномалії можуть бути ознакою подій інформаційної безпеки, таких як вихід з ладу обладнання, НСД чи кібератаки, і їх виявлення потребує врахування особливостей досліджуваних даних і системи [66]. Вибір методу залежить від природи аномалій, характеру даних і вимог до обчислювальних ресурсів. Методи виявлення аномалій класифікуються за кількома ознаками, включаючи тип навчання (контрольоване, напівконтрольоване чи неконтрольоване), підхід до аналізу даних (статистичний, машинне навчання, графічний) та спосіб обробки (в режимі реального часу чи ні). Статистичні методи базуються на теоретико-ймовірнісних моделях і аналізі відхилень від нормального розподілу даних. Вони ефективні для дослідження передбачуваних моделей, однак можуть бути недостатніми для складних або високодинамічних середовищ. Однією з характеристик є параметри, які залучаються для виявлення аномалій. Це можуть бути частота та інтенсивність трафіку, розподіл даних у просторі ознак, кореляції між параметрами, а також рівень шуму. Висока якість даних та належна попередня обробка, зокрема очищення та нормалізація, сприяють підвищенню точності детекції. Під час вибору параметрів важливо також враховувати специфіку задачі, наприклад, чи орієнтовані методи на виявлення одноразових викидів або нових, ще не зафіксованих аномалій. Розробка і впровадження методів виявлення аномалій повинні враховувати ризики та потенційні наслідки помилок. Виявлення хибнопозитивних або хибнонегативних результатів може мати наслідки, зокрема в кібербезпеці чи управлінні складними системами, тому баланс між чутливістю і специфічністю є пріоритетним завданням. Автоматизація процесів і здатність систем до самонавчання на основі змін у поведінці даних також сприяють підвищенню ефективності виявлення аномалій.

Aklil Kiflay та інші у дослідженні [67] пропонують метод виявлення мережевих вторгнень за допомогою мультимодальної системи на основі машинного навчання. Головними елементами системи є два моделі з алгоритмом Random Forest, які працюють окремо з метаданими потоку і першими 32 байтами протоколу, об'єднуючи результати через soft voting. Зміна правил або методів виявлення можлива лише шляхом перенавчання. Проте масштабованість мережі обмежена через обробку великих обсягів даних.

Авторами [68] представлено метод виявлення мережевих вторгнень, який базується на автономній НЛ із використанням "м'яких прототипів". Цей підхід, відомий як Soft Prototype-based Autonomous Fuzzy Inference System (SPAFIS), дозволяє в режимі реального часу адаптуватися до нових даних та змінювати свою структуру. Особливістю є здатність до автономного навчання: SPAFIS додає нові прототипи, видаляє застарілі або дублюючі елементи і таким чином оптимізує обчислювальні ресурси. Метод підтримує аналіз вхідного трафіку, адаптуючись до змінюваних патернів даних. Для налаштування потрібно мати фахівців, а базові параметри легко адаптуються. SPAFIS здатний працювати з даними у реальному часі, але здебільшого орієнтований на аналіз потоків у режимі онлайн. SPAFIS ефективно масштабується для великих мереж, забезпечуючи мінімальний вплив на продуктивність.

У статті [69] досліджується гібридна система виявлення мережевих вторгнень, яка використовує метод МН з функцією самовідновлення. Основна ідея полягає у поєднанні методів виявлення аномалій та сигнатурного аналізу для підвищення ефективності виявлення атак. Особливістю є механізм самовідновлення, що дозволяє автоматично додавати нові сигнатури до бази даних, зокрема для невідомих атак, таким чином підвищуючи точність та ефективність виявлення. Підтримка масштабованості забезпечується через використання оптимізованих алгоритмів, що зменшують навантаження на ресурси. Система підтримує гнучке налаштування конфігурації відповідно до потреб користувача.

Робота [70] стосується виявлення та запобігання атак нульового дня на мережі за допомогою методів МН. Аналіз мережевого трафіку використовується для підтримки систем виявлення мережевих вторгнень. Використання закону Бенфорда дозволяє ідентифікувати мережеві параметри, що свідчать про аномальну активність. Описані методи підтримують налаштування, масштабованість, сумісність із різними операційними системами та іншими системами безпеки. Точність роботи системи залежить від правильного вибору параметрів аналізу. Зокрема, пропонуються напівавтоматичні методи МН, які здатні навчатися на обмежених обсягах маркованих даних, що полегшує інтеграцію нових правил виявлення атак.

Mahmoud Said El Sayed та інші [71] акцентують увагу на розробці методу виявлення аномалій для боротьби з DDoS-атаками в мережах, побудованих за принципом програмно-визначених мереж. Основна увага приділяється використанню технологій глибокого навчання для створення ефективної системи виявлення вторгнень, що аналізує поведінку мережевого трафіку для розпізнавання загроз. Особливістю є застосування методів вибору характеристик, таких як Information Gain і Random Forest, що дозволяє оптимізувати набір вхідних даних, зменшуючи кількість оброблюваних параметрів без втрати точності.

Автори [72] пропонують покращення продуктивності системи виявлення вторгнень завдяки об'єднанню сигнатурного та на основі аномалій підходів до виявлення. Система аналізує вхідний трафік, що дозволяє класифікувати мережеві пакети на нормальні та такі, що містять атаки. Адміністратори можуть отримувати детальні звіти та журнали роботи, а також інформацію про виявлені загрози в реальному часі, що сприяє швидкому реагуванню на зловмисні дії.

Sivasankari Nitiynandan [73] зосереджує увагу на виявленні та запобіганні атак типу "людина посередині" у мережах IoT за допомогою методів регресійного моделювання. Дослідження пропонує використовувати три техніки МН — лінійну регресію, багатовимірну лінійну регресію та регресію Гаусівського процесу. Робота включає моделювання трафіку IoT з використанням інструменту NS2, що дозволяє генерувати дані як для нормального, так і для шкідливого трафіку.

Дослідження [74] спрямоване на створення гібридної системи виявлення вторгнень, яка поєднує в собі методи НЛ, штучних нейронних мереж та генетичного алгоритму. Система аналізує вхідний трафік, дозволяючи класифікувати дані на нормальні або зловмисні. Виявлення атак може здійснюватися в реальному часі, особливо у випадках, коли це стосується певних аномалій у поведінці трафіку. Модель підтримує різні операційні системи, а також може інтегруватися з іншими системами безпеки, підвищуючи їхню загальну ефективність.

Робота Радівілової [75] присвячена аналізу статистичних методів для виявлення аномалій у телекомунікаційному трафіку. Основними методами, що розглядалися, є аналіз часових рядів, класифікація за допомогою дерев рішень, кластерний аналіз та метод на основі ентропії. Для оцінки методів враховувалися такі показники, як імовірність виявлення аномалій, кількість хибнопозитивних спрацьовувань та час роботи.

Автори [76] пропонують систему для виявлення та пом'якшення аномалій у програмно-визначених мережах, що базується на поєднанні методів довгої короткочасної пам'яті для прогнозування нормальної поведінки мережевого трафіку та НЛ для виявлення аномалій. Основна ідея системи полягає у створенні цифрового підпису мережевих сегментів, що базується на атрибутах мережевих потоків. Система використовує вікно трафіку, уникаючи потреби в історичній базі даних, і визначає порогові значення нормальності за допомогою нерівності Бієнаме-Чебишева. НЛ дозволяє адаптивно оцінювати відхилення від нормального профілю, що допомагає класифікувати аномалії, такі як DDoS-атаки чи сканування портів. Для пом'якшення наслідків атак використовується політика динамічного управління потоками, яка блокує шкідливі пакети або перенаправляє їх до спеціальних вузлів, гарантуючи стабільність мережі.

Запропонований метод [77] класифікації мережевого трафіку, заснований на МН, має низку важливих переваг і певні обмеження, які впливають на його ефективність і сферу застосування. Випадкові ліси, які виявилися найбільш ефективними, демонструють збалансовану точність понад 87%.

У [78] автори пропонують метод FuzHD++, який поєднує відновлення відсутніх даних і виявлення аномальних вузлів у середовищі IoT. Для відновлення відсутніх даних застосовується профіль матриці, який виявляє повторювані шаблони серед даних різних вузлів. Потім використовується алгоритм k-найближчих сусідів для заповнення відсутніх значень на основі інформації від сусідніх вузлів. Водночас виявлення аномальних вузлів базується на нечітких правилах із використанням експертних знань і прихованих закономірностей у даних. Недоліком методу є його обмежена здатність виявляти лише певні типи атак, наприклад, атаки на цілісність даних (FDIA).

Liangchen Chen [79] представляє метод кластеризації для виявлення аномалій у мережевому трафіку DPC-GS-MND. Він базується на алгоритмі кластеризації за піковою щільністю, який було покращено за допомогою врахування взаємного ступеня сусідства. Взаємний ступінь сусідства забезпечує визначення центрів кластерів, аналізуючи локальні зв'язки між даними. Крім того, метод використовує автоматизований вибір центрів кластерів, що зменшує ризик помилок, спричинених людським фактором. Для обробки дисбалансованих даних можуть виникнути труднощі, а додатковий час на обчислення взаємного ступеня сусідства збільшує загальні витрати часу.

Запропонований у [80] метод виявлення аномалій у кіберфізичних системах базується на інтеграції алгоритму найменших квадратів та ентропійного аналізу часових вікон. Метод найменших квадратів оптимізований шляхом додавання функцій, таких як функції Гауса та дробові функції, які сприяють досягненню розрідженості параметрів стану та мінімізують перенавчання. Коли система виявляє аномальну поведінку, алгоритм аналізує дані у зворотному порядку, використовуючи ентропійний аналіз для визначення джерела аномалії.

У [81] запропоновано метод, який використовує поведінковий аналіз для визначення нормальної поведінки мережі та виявлення відхилень, що можуть свідчити про аномалії. Методологія базується на об'єднанні декількох алгоритмів прогнозування, таких як просте експоненціальне згладжування, подвійне експоненціальне згладжування та метод Холта-Вінтерса. Вони інтегровані в адаптивний алгоритм, що враховує змінність сезонних даних та може налаштовуватись відповідно до специфіки трафіку.

Шпінарева та інші [82] пропонують метод створення каскаду глибинних нейронних мереж для виявлення та класифікації мережевих атак. Він складається з двох етапів: спочатку використовується гібридна мережа, що об'єднує згорткову нейронну мережу та мережу з довгою короткостроковою пам'яттю, для виявлення атак. Тестування методу проводилося з використанням набору даних UNSW-NB15, а також на реальних мережевих даних, що дозволило перевірити його ефективність. Водночас існують певні обмеження. Зокрема, складно забезпечити аналіз великих обсягів даних у реальному часі. Крім того, навчання та налаштування таких глибоких моделей вимагають значних обчислювальних ресурсів.

У [83] автори пропонують метод створення гібридної системи для аналізу та діагностики загроз типу Zero-Day. Він базується на інтеграції кількох компонентів, які забезпечують автоматичне прийняття рішень, статичний аналіз з використанням мультисканера та аналіз у ізольованому середовищі Sandbox. Система використовує модульний підхід, що включає аналіз трафіку, дампи пам’яті та створення унікального набору правил для виявлення аномалій. Інтеграція з багатьма антивірусними рішеннями забезпечує паралельний аналіз файлів, а ізольоване середовище Sandbox дозволяє глибоко досліджувати поведінку шкідливих об'єктів. Слід зазначити, що метод потребує значних обчислювальних ресурсів для функціонування ізольованого середовища та аналізу великих обсягів даних. Використання багатьох компонентів може ускладнити налаштування і підтримку системи. Крім того, обмеження щодо точності правил Yara для деяких складних поліморфних модифікацій можуть впливати на загальну ефективність.

У роботі [84] автори представляють метод централізованої розподіленої системи виявлення мережевих атак у корпоративних мережах на основі мультифрактального аналізу. Для цього застосовується метод максимумів модулів вейвлет-перетворення, який дає змогу виявити особливості сигналу та оцінити ступінь його самоподібності за допомогою коефіцієнта Херста. Недоліком підходу є залежність від кількості доступних даних. За недостатньої кількості значень в автокореляційній послідовності неможливо точно сформувати мультифрактальний спектр, що може знижувати ефективність для рідкісних або слабких атак. Також виявлено необхідність повторного налаштування параметрів самоподібності для кожного нового масштабу аналізу, що може ускладнити впровадження системи в умовах динамічного трафіку.

Таблиця 1.2 містить порівняльний аналіз методів виявлення аномалій у мережевому трафіку відповідно до визначених критеріїв, що дозволяє оцінити їхню функціональність, ефективність і сумісність із різними умовами застосування. Таблиця дозволяє забезпечити зручне порівняння досліджуваних підходів залежно від вимог системи та характеристик середовища, що сприяє прийняттю обґрунтованих рішень щодо подальших досліджень та розробки нових методів виявлення аномалій у сучасних інформаційно-комунікаційних мережах.

Таблиця 1.2 – Порівняння методів відповідно до заданих критеріїв

Критерії / Джерело	Критерії										
№	Аналіз вихідного трафіку	Простота налаштування та впровадження	інформування адміністратора	в режимі реального часу	Здатність запобігати атакам	Наявність звіту	Можливість модифікації під час роботи	Підтримка масштабованості мережі	Вплив на продуктивність	Кросплатформеність	Сумісність із іншими системами безпеки
[67]	-	-	+	+	+	+	+	+	+	-	+
[68]	-	-	+	+	+	+	+	+	+	-	+
[69]	-	-	-	-	-	-	+	+	-	-	+
[70]	-	+	-	+	-	-	-	+	-	+	+
[71]	-	-	-	+	-	-	-	-	-	-	-
[72]	-	-	-	+	-	-	+	+	-	-	-
[73]	-	-	-	+	-	+	-	+	-	-	-
[74]	-	-	-	+	-	+	-	+	+	-	+
[75]	-	-	-	-	-	-	-	-	+	-	-
[76]	-	-	+	+	+	-	-	+	-	+	+
[77]	-	+	-	+	+	-	-	-	+	-	+
[78]	-	+	-	+	-	-	+	-	+	-	+
[79]	-	+	-	-	-	-	-	-	-	-	-
[80]	-	+	-	+	-	-	+	+	+	-	-
[81]	-	+	-	+	-	-	+	+	-	-	+
[82]	-	+	+	+	-	-	-	+	+	-	-
[83]	-	+	-	-	-	-	+	+	+	-	+
[84]	-	+	+	-	-	-	+	+	-	-	+

Пропоновані методи виявлення мережевих аномалій демонструють значний прогрес у забезпеченні мережевої безпеки завдяки використанню сучасних технологій МН, НЛ та глибокого навчання. Багато систем пропонують високий рівень точності (до 98-99%) при виявленні атак, таких як сканування портів, DoS, віруси чи експлойти, що забезпечується завдяки комбінованому використанню аналізу мережевого потоку, корисного навантаження та сигнатурного аналізу. Деякі підходи, як-от SPAFIS, демонструють адаптивність до нових загроз у реальному часі, завдяки механізмам автономного навчання та динамічного оновлення правил. Інші методи, зокрема засновані на гібридних системах, пропонують гнучку конфігурацію, масштабованість та мінімальний вплив на продуктивність мережі. Інноваційні підходи, такі як використання автокодерів, LSTM або аналізу закономірностей за допомогою закону Бенфорда, дозволяють ефективно обробляти великі обсяги трафіку з високою точністю та мінімальною кількістю хибнопозитивних спрацьовувань. Попри високий рівень ефективності, деякі методи мають суттєві обмеження. Використання складних алгоритмів, як-от LSTM чи ResNet50, може вимагати значних обчислювальних ресурсів, що ускладнює застосування у великих або обмежених у ресурсах мережах. Методи, що базуються на сигнатурному аналізі, обмежені у здатності виявляти нові атаки, що вимагає регулярного оновлення бази сигнатур [85]. Інші підходи, наприклад, SPAFIS, потребують залучення кваліфікованих фахівців для налаштування і моніторингу, що ускладнює впровадження для некваліфікованих користувачів. Деякі моделі, такі як гібридні системи на основі дерева рішень і LSTM, демонструють вищий рівень хибнопозитивних спрацьовувань (до 17%), що може створювати додаткове навантаження на адміністраторів. Крім того, масштабованість у методах із великим обсягом обробки даних інколи залишається обмеженою, як у випадку з системами, заснованими на аналізі корисного навантаження пакетів.

Робота [86] присвячена аналізу методів виявлення аномалій у середовищі Інтернету речей. Її мета – огляд підходів до детекції аномалій, визначення основних напрямів застосування та виявлення прогалин, які потребують подальшого дослідження. Результати порівняння методів свідчать, що найбільш ефективними є підходи, які базуються на глибокому навчанні та алгоритмах МН, особливо у випадках, коли дані містять високу кількість шуму або змінні закономірності. Поєднання хмарних обчислень із локальною обробкою на рівні сенсорів забезпечує оптимальне співвідношення між затримкою і точністю. Робота також вказує на ключові виклики, такі як потреба в адаптації до змін у патернах даних і недостатність розмічених даних для тренування моделей. Усі методи мають свої сильні та слабкі сторони, і вибір залежить від конкретного контексту застосування. Поєднання різних підходів, таких як гібридні системи із самовідновленням або адаптивні моделі на основі НЛ, є перспективним шляхом для покращення мережевої безпеки. Подальші дослідження можуть зосереджуватися на підвищенні ефективності обробки даних, зниженні рівня хибнопозитивних спрацьовувань і спрощенні інтеграції систем для широкого кола користувачів.

Проаналізовані методи виявлення аномалій у мережевому трафіку в [87-88] демонструють як сильні сторони, так і недоліки виявлення DDoS-атак. До основних переваг відносяться високий рівень точності завдяки використанню алгоритмів МН, таких як опорні векторні машини, нейронні мережі та методи на основі LSTM. Однак, поряд із цим, існують певні недоліки. По-перше, деякі методи складними у налаштуванні. По-друге, для багатьох методів є проблемою висока обчислювальна складність, яка зростає пропорційно до розміру мережі та обсягу трафіку. Крім того, у випадках обробки незбалансованих даних, ефективність деяких підходів може суттєво знижуватися, що ускладнює їх використання у реальних умовах. Методи виявлення аномального трафіку в IoT [89-92] забезпечують як конструктивні рішення для поліпшення функціонування систем, так і ефективність у захисті від шкідливих дій. Проте їх адаптація до різноманітних сценаріїв потребує значних зусиль. Водночас, недоліком багатьох моделей та методів є їхня залежність від обчислювальних ресурсів. Ще одним викликом є використання застарілих або синтетичних наборів даних, які не завжди відображають реальні умови застосування. Оскільки значна частина експериментів все ще базується на старих або синтетичних наборах, таких як KDD Cup 1999, що обмежує актуальність отриманих результатів у сучасних умовах.

РОЗДІЛ 2
РОЗРОБКА МОДЕЛЕЙ ПРОЦЕСІВ КЛАСИФІКАЦІЇ ТРАФІКУ

2.1 Вдосконалення моделі поведінки типового користувача

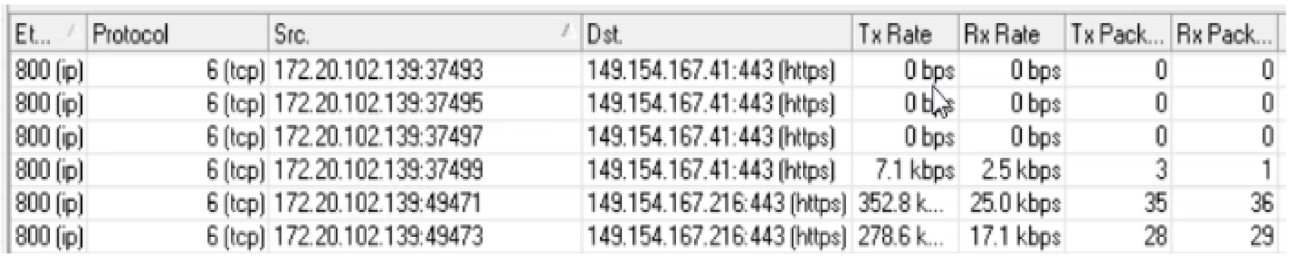
Формування моделі поведінки типового користувача у відкритому сегменті мережі включає декілька послідовних етапів. Першочергово користувач має переглянути перелік доступних Wi-Fi точок на своєму пристрої. На цьому етапі пристрій ще не є пристроєм у мережі. Наступним кроком відбувається приєднання до мережі. У разі відсутності паролюз’єднання здійснюється шляхом вибору SSID і підтвердження підключення. Якщо точка доступу має ключ безпеки, то пристрій приєднується до мережі після авторизації. Пароль чи QR-код, зазвичай, є у відкритому доступі чи без труднощів можна взяти у працівника організації, до мережі якої приєднано точку. Після успішної авторизації пристрою автоматично присвоюється IP-адреса, і він буде вважатися користувачем мережі. За замовчуванням усіх авторизованих користувачів слід вважати типовими. Завершенням роботи користувача в мережі є від’єднання (вимкнення Wi-Fi на пристрої, перехід в режим «Літака» тощо) чи вихід із зони дії точки доступу.

Проаналізувавши поведінку користувачів у відкритих сегментах за допомогою програми Wireshark, можна визначити найбільш використовувані застосунки.

Використання застосунку Telegram. Відповідно до технічної документації даний продукт використовує IP-адреси двох підмереж 91.108.4.0/22 та 149.154.167.0/20, використовує порти 80, 88 та 8443. У таблиці 2.1 представлено фрагмент аналізу трафіку при роботі з Telegram. Під час моніторингу було виявлено, що при роботі з додатком IP-адреса джерела залишається незмінною, хоча задіяних портів з діапазону портів від 49152 до 65535 може бути щонайменше 3 (табл. 2.1, рис. 2.1). IP-адреса одержувача залишається незмінною для додатка та для роботи із версією для браузер. Порт одержувача у цьому випадку 443 та використовується протокол TCP. Аналіз отриманих даних дозволив зробити висновок, що поєднання IP-адреси однієї із підмереж та одного із вказаних портів є характерною ознакою роботи з Telegram та можна вважати за типовий трафік.

Таблиця 2.1 - Фрагмент аналізу мережі при роботі з застосунком Telegram

IP адреса відправника	Порт відправника	IP отримувача (сервер застосунку)	Порт отримувача (сервер застосунку)
172.20.110.72	61184	149.154.167.41	443
172.20.110.72	61065	149.154.167.41	443
172.20.110.72	61064	149.154.167.99	443
192.168.16.107	53562	149.154.167.41	443
192.168.16.107	54096	149.154.167.41	443
192.168.16.107	54149	149.154.167.99	443



Et...	Protocol	Src.	Dst.	Tx Rate	Rx Rate	Tx Pack...	Rx Pack...
800 (ip)	6 (tcp)	172.20.102.139:37493	149.154.167.41:443 (https)	0 bps	0 bps	0	0
800 (ip)	6 (tcp)	172.20.102.139:37495	149.154.167.41:443 (https)	0 bps	0 bps	0	0
800 (ip)	6 (tcp)	172.20.102.139:37497	149.154.167.41:443 (https)	0 bps	0 bps	0	0
800 (ip)	6 (tcp)	172.20.102.139:37499	149.154.167.41:443 (https)	7.1 kbps	2.5 kbps	3	1
800 (ip)	6 (tcp)	172.20.102.139:49471	149.154.167.216:443 (https)	352.8 k...	25.0 kbps	35	36
800 (ip)	6 (tcp)	172.20.102.139:49473	149.154.167.216:443 (https)	278.6 k...	17.1 kbps	28	29

Рисунок 2.1 − Фрагмент аналізу мережі при роботі з версією для мобільного застосунку Telegram

Використання застосунку Viber. Аналіз даних, що було отримано у результаті моніторингу мережі (табл. 2.2), дозволяє зробити висновок, що одноразово встановлюється з’єднання між пристроєм та сервером застосунку для обміну даними. Далі використовується одна і та ж комбінація IP-адреси та порта одержувача й використовуються поки користувач знаходиться в поточній мережі.

Використання застосунку Zoom. У технічній документації на сайті розробника вказано, що IP-адреси та порти є чітко визначені в залежності від того, яким чином використовуються ресурси (мобільний додаток, програмне забезпечення, браузер). Для налаштування брандмауерів використовують порти

TCP 80, 443, 8801, 8802 та UDP 3478, 3479, 8801-8810. При роботі з додатком Zoom із мобільного пристрою задіяні порти TCP 5091, 390 та UDP 20000-64000. Відповідно до правил брандмауера для Zoom Contact Center задіяні порти UDP 20000-64000 та TCP 443, 5091. Брандмауер при використанні через сайт застосунок Zoom використовує порти TCP 80, 443 і вже не визначені IP-адреси, а домен *.zoom.us. Правила брандмауера для додатків Zoom застосовують порти TCP 443 та домен *.zoomapp.cloud. Дослідивши поведінку мережевого трафіку при роботі з додатком Zoom (табл. 2.3) можна відслідкувати з’єднання з підмережею в діапазоні IP-адрес та використання порту, що підтверджує дані з документації.

Таблиця 2.2 - Фрагмент аналізу мережі при роботі з застосунком Viber

IP адреса відправника	Порт відправника	IP отримувача (сервер застосунку)	Порт отримувача (сервер застосунку)
172.20.110.72	61099	52.0.252.41	443
52.0.252.41	443	172.20.110.72	61099
192.168.16.107	53597	44.192.202.0	443
44.192.202.0	443	192.168.16.107	53597

Таблиця 2.3 - Фрагмент аналізу мережі при роботі з Zoom

IP адреса відправника	Порт відправника	IP отримувача (сервер застосунку)	Порт отримувача (сервер застосунку)
172.20.110.72	51800	170.114.52.4	443
172.20.110.72	51800	170.114.52.4	443
172.20.110.72	61128	170.114.15.95	443
192.168.16.107	63924	170.114.2.8	443
192.168.16.107	63954	170.114.14.58	443

Під час роботи з пошуковими системами переважно використовуються порти 80, 443 та IP-адресу, що відповідає тому чи іншому сайту (табл. 2.4).

Таблиця 2.4 - Фрагмент аналізу мережі при роботі з пошуковими системами

IP адреса відправника	Порт відправника	IP отримувача (сервер застосунку)	Порт отримувача (сервер застосунку)
172.20.110.72	61166	35.245.210.119	443
172.20.110.72	61473	216.58.209.16	443
172.20.110.72	61505	18.66.121.64	443
172.20.110.72	61151	173.194.73.188	80
172.20.110.72	61504	142.250.203.197	443
172.20.110.72	61504	142.250.203.197	443

Et...	Protocol	Src.	Dst.	VLAN Id	DSCP	Tx Rate	Rx Rate	Tx Pack...	Rx Pack...
800 (ip)	6 (tcp)	172.20.102.139:40258	159.203.145.121:443 (https)			0 bps	0 bps	0	0
800 (ip)	6 (tcp)	172.20.102.139:40260	159.203.145.121:443 (https)			0 bps	0 bps	0	0
800 (ip)	6 (tcp)	172.20.102.139:40268	159.203.145.121:443 (https)			0 bps	0 bps	0	0
800 (ip)	17 (udp)	172.20.102.139:40276	157.240.224.7:443 (https)			0 bps	0 bps	0	0
800 (ip)	17 (udp)	172.20.102.139:40357	157.240.224.7:443 (https)			240.7 k...	9.2 kbps	25	14
800 (ip)	17 (udp)	172.20.102.139:40428	142.250.186.193:443 (https)			0 bps	0 bps	0	0
800 (ip)	17 (udp)	172.20.102.139:40879	104.26.7.133:443 (https)			4.1 Mbps	174.0 k...	460	81
800 (ip)	17 (udp)	172.20.102.139:40969	34.117.98.198:443 (https)			0 bps	0 bps	0	0
800 (ip)	17 (udp)	172.20.102.139:41043	104.21.52.152:443 (https)			6.7 Mbps	100.3 k...	689	119
800 (ip)	6 (tcp)	172.20.102.139:41465	157.240.224.12:443 (https)			776 bps	0 bps	1	0
800 (ip)	6 (tcp)	172.20.102.139:41467	157.240.224.12:443 (https)			776 bps	0 bps	1	0
800 (ip)	6 (tcp)	172.20.102.139:41671	157.240.224.12:443 (https)			0 bps	0 bps	0	0
800 (ip)	17 (udp)	172.20.102.139:41730	216.58.215.100:443 (https)			0 bps	0 bps	0	0
800 (ip)	6 (tcp)	172.20.102.139:41878	35.186.231.97:443 (https)			44.7 kbps	9.4 kbps	8	9
800 (ip)	17 (udp)	172.20.102.139:41889	172.20.0.2:53 (dns)			0 bps	0 bps	0	0
800 (ip)	17 (udp)	172.20.102.139:42150	172.20.0.2:53 (dns)			0 bps	0 bps	0	0
800 (ip)	6 (tcp)	172.20.102.139:42335	18.66.233.2:443 (https)			91.3 kbps	16.7 kbps	17	14
800 (ip)	6 (tcp)	172.20.102.139:42436	78.152.183.46:443 (https)			15.8 Mbps	371.9 k...	1317	688
800 (ip)	17 (udp)	172.20.102.139:42450	142.250.186.200:443 (https)			0 bps	0 bps	0	0
800 (ip)	17 (udp)	172.20.102.139:42868	18.66.233.57:443 (https)			47.2 kbps	19.2 kbps	8	8
800 (ip)	6 (tcp)	172.20.102.139:42953	172.67.75.202:443 (https)			0 bps	0 bps	0	0
800 (ip)	17 (udp)	172.20.102.139:43103	142.250.203.206:443 (https)			0 bps	0 bps	0	0
800 (ip)	6 (tcp)	172.20.102.139:43196	142.250.203.206:443 (https)			0 bps	0 bps	0	0
800 (ip)	17 (udp)	172.20.102.139:43285	172.20.0.2:53 (dns)			0 bps	0 bps	0	0

Рисунок 2.2 − Фрагмент аналізу мережі при роботі з пошуковими системами використовуючи мобільний телефон

При використанні додатків Facebook і Messenger (рис.2.3) застосовуються протоколи TCP та UDP у залежності від виконуваних дій, одноразово встановлюється з'єднання за певною IP-адресою, задіяно 443 порт.

Et...	Protocol	Src.	Dst.	Tx Rate	Rx Rate	Tx Pack...	Rx Pack...
800 (ip)	6 (tcp)	172.20.102.139:37605	157.240.224.7:443 (https)	0 bps	0 bps	0	0
800 (ip)	6 (tcp)	172.20.102.139:37627	157.240.224.7:443 (https)	47.8 kbps	12.1 kbps	14	11
800 (ip)	6 (tcp)	172.20.102.139:37629	157.240.224.7:443 (https)	31.6 kbps	6.6 kbps	7	8
800 (ip)	6 (tcp)	172.20.102.139:37854	157.240.224.16:443 (https)	0 bps	0 bps	0	0
800 (ip)	17 (udp)	172.20.102.139:40404	157.240.224.63:443 (https)	631.2 k...	8.5 kbps	64	8
800 (ip)	17 (udp)	172.20.102.139:40869	157.240.224.63:443 (https)	126.9 k...	74.6 kbps	28	19
800 (ip)	6 (tcp)	172.20.102.139:40902	157.240.224.63:443 (https)	0 bps	0 bps	0	0
800 (ip)	17 (udp)	172.20.102.139:47288	157.240.224.63:443 (https)	0 bps	0 bps	0	0

Рисунок 2.3 − Фрагмент аналізу мережі при роботі з додатками Facebook

При використанні додатку поштового сервісу (рис.2.4) чи сайту одноразово встановлюється з'єднання з IP-адресою та виключно її використання у поєднанні з 443 портом, порти мобільного пристрою можуть змінюватись. Протокол, що застосовується TCP.

Et...	Protocol	Src.	Dst.	Tx Rate	Rx Rate	Tx Pack...	Rx Pack...
800 (ip)	6 (tcp)	172.20.102.139:37449	89.184.85.91:443 (https)	5.4 kbps	8.3 kbps	1	2
800 (ip)	6 (tcp)	172.20.102.139:37457	89.184.85.91:443 (https)	5.9 kbps	7.6 kbps	2	2
800 (ip)	6 (tcp)	172.20.102.139:37459	89.184.85.91:443 (https)	0 bps	0 bps	0	0
800 (ip)	6 (tcp)	172.20.102.139:37465	89.184.85.91:443 (https)	4.4 Mbps	10.5 kbps	369	20

Рисунок 2.4 − Фрагмент аналізу мережі при роботі з додатком поштового сервісу

Під час роботи із застосунком банку (рис.2.5) використовуються аналогічно до попередніх пунктів TCP-протокол та 443 порт одержувача. Але у залежності від операцій, що виконує користувач, IP-адреси змінюються. Оскільки, до прикладу, архів транзакцій зберігається за однією адресою, а при переказі коштів чи перегляді стану рахунку використовуються інші адреси.

Et...	Protocol	Src.	Dst.	Tx Rate	Rx Rate	Tx Pack...	Rx Pack...
800 (ip)	6 (tcp)	172.20.102.139:40827	3.72.185.176:443 (https)	6.3 Mbps	134.2 k...	527	245
800 (ip)	6 (tcp)	172.20.102.139:40831	3.72.185.176:443 (https)	3.3 Mbps	71.3 kbps	280	126
800 (ip)	6 (tcp)	172.20.102.139:46869	3.65.157.177:443 (https)	592 bps	592 bps	1	1
800 (ip)	6 (tcp)	172.20.102.139:46873	3.65.157.177:443 (https)	1120 bps	2.1 kbps	2	4
800 (ip)	6 (tcp)	172.20.102.139:48005	52.31.131.184:443 (https)	4.7 kbps	2.4 kbps	2	3
800 (ip)	6 (tcp)	172.20.102.139:48213	52.223.11.138:443 (https)	111.8 k...	11.4 kbps	19	12
800 (ip)	6 (tcp)	172.20.102.139:48215	52.223.11.138:443 (https)	42.2 kbps	6.7 kbps	8	9
800 (ip)	6 (tcp)	172.20.102.139:48217	52.223.11.138:443 (https)	42.2 kbps	6.7 kbps	8	9
800 (ip)	6 (tcp)	172.20.102.139:48219	52.223.11.138:443 (https)	42.2 kbps	6.7 kbps	8	9

Рисунок 2.5 − Фрагмент аналізу мережі при роботі з додатком банку

Отож, на основі наявного аналізу можна зробити висновок, що при поєднанні певних (типових) IP-адрес та відповідних портів з'єднання можна вважати безпечним та ці дані можна приймати за типовий дозволений.

2.2 Вдосконалення моделі поведінки порушника у відкритому сегменті мережі

Після того, як у параграфі 2.1. описано модель поведінки типового користувача у відкритому сегменті мережі та поведінку типового трафіку слід охарактеризувати модель поведінки порушника у відкритому сегменті мережі відповідно до найпоширеніших атак [93].

Все частіше здійснюються атаки доступу, зокрема атака на паролі. Оскільки кількість облікових записів та цінність інформації, яку вони містять, з переходом до цифровізації суттєво зросла. При реалізації такої атаки у більшості випадків атака йде на стандартні сервіси та стандартні порти (табл. 2.5). Типовими методами злому паролів є метод грубої сили, підбір за словником та перевірка за шаблоном.

Таблиця 2.5 – Типове використання протоколів та портів при атаці на паролі

Використання пари протоколів	Атака на порти
SSH і TCP	22, 80
PPTP і TCP	1723
L2TP і UDP	500, 1701, 4500
RDP і TCP	3389
VNC і TCP	5900

Приклад трафіку зі сторони зловмисника при спробі підбору логіна та пароля до сайту на CMS Wordpress зібраний за допомогою програми Wireshark зображено на рис. 2.6.

Під час атаки етапи виконання порушником буде сформовано наступним чином:

Крок 1.Підготовка до здійснення атаки. На цьому етапі порушник обирає об’єкт по відношенню до якого буде здійснюватися атака, керуючись власними мотивами здійснення даної атаки, та збирає необхідну інформацію, в тому числі за допомогою соціальної інженерії.

Крок 2. Розробка плану атаки. Порушник визначається із мережею, до якої буде приєднуватись для здійснення атаки, та способом здійснення атаки, підготовкою словників чи налаштуванням програмного забезпечення.

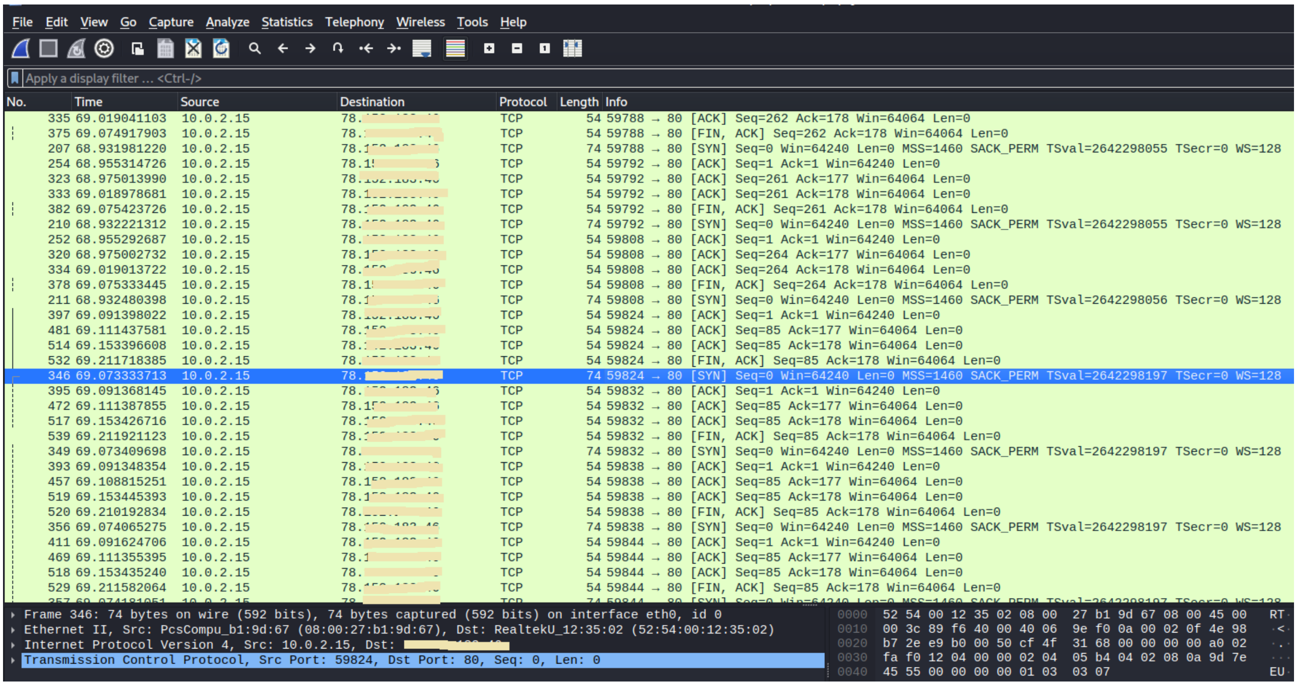


Рисунок 2.6 – Трафік зі сторони зловмисника при спробі підбору логіна та пароля

Крок 3. Реалізація. Порушник приєднується до відкритого сегменту мережі та розпочинає безпосередню атаку. Під час якої можна за аналізом трафіка виявити факт порушення. Для цього слід при скануванні вихідного мережевого трафіку аналізувати вхідні та вихідні порти та IP, частоту звернень, протоколи та інші параметри. Час, який потрібно використати під час підбору пароля, безпосередньо залежить від ентропії паролю та від пристрою з якого запускається атака, швидкості мережі з якої йде атака.

Крок 4. Після успішного завершення атаки користувач може використовувати отримані дані залежно до раніше визначених мотивів.

Крок 5. Порушник може намагатись приховати інформацію щодо своїх дій зі сторони об’єкта атаки.

Крок 6. Порушник покидає вже скомпрометовану мережу.

Слід зазначити, що пункти 1, 2, 4, 5 та 6 шляхом аналізу трафіка не ідентифікуються, бо є подібними до дій типового користувача.

На основі попередньо викладеного матеріалу сформуємо модель порушника під час атаки на пароль, що зображено на рис. 2.7.

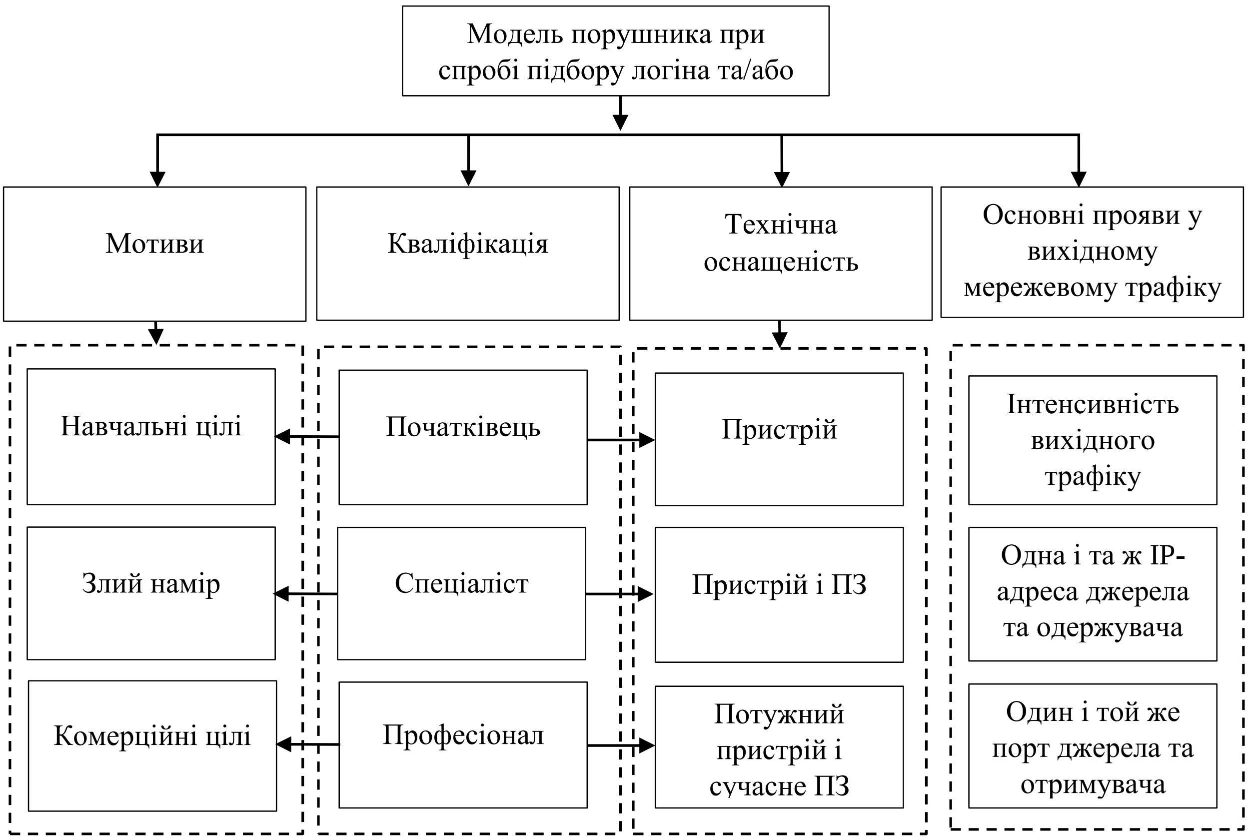


Рисунок 2.7 - Модель порушника при атаці на пароль

Незважаючи на простоту реалізації та наявність базових правил для захисту від атаки, популярність даної атаки не зменшується. Адже вчиняти дії можна не лише до кореневих серверів чи IP, а й до певних чистин, які виконують життєвоважливі завдання чи до місць, де попередньо було знайдено «дірки» безпеки.

При аналізі ICMP-флуду у мережевому трафіку можна спостерігати однотипні запити ICMP ECHO, що матимуть одну і ту ж адресу одержувача. Варто зазначити, що запити будуть надходити з однаковим інтервалом. Але інтервал надсилання запитів налаштовується порушником перед початком атаки. Слід звернути увагу також на розмірність пакету. Перевищення розмірності пакету може вивести з ладу сервер, якщо на ньому не налаштовано фрагментацію пакетів.

SYN-флуд ініціалізує велику кількість одночасних TCP-з’єднань через посилку SYN-пакету з неіснуючою зворотною адресою. Для більш ефективної атаки порушник спочатку може проводити розвідувальну атаку з метою виявлення найбільш вразливих вузлів мережі. Після того, як атака успішно реалізована, об’єкт атаки перестає функціонувати через велику кількість ненадісланих відповідей та розриває усі нові з’єднання.

UDP-флуд при аналізі мережевого трафіку відслідковується за рахунок великої кількості UDP-пакетів на різні порти з однією IP-адресою.

HTTP-повідомлення GET з великою інтенсивністю на 80 порт характеризують атаку типу HTTP-флуд. Її основною ціллю є завантаження ключових елементів системи до стану, коли вони будуть не в змозі обробляти будь-які інші запити.

Під час атаки Ping of Death зловмисник намагається зупинити роботу сервера чи вивести його з ладу, надіславши звичайний запит ping, який або фрагментований, або негабаритний. Коли надсилається пінг більший за максимальний розмір, цільовий сервер буде фрагментувати файл. Пізніше, коли сервер сформулює відповідь, повторна збірка цього більшого файлу може спричинити перевантаження буфера та збій. При аналізі трафіку цей тип атаки можна відстежити за розміром пакета та його складових.

DoS атаки зі сторони зловмисника є дещо схожими до сканування портів якщо аналізувати мережевий трафік, зокрема час та протоколи, що використовуються. Проте можна встановити інтенсивність атаки, вказати кількість запитів і відмовитись від отримання відповіді [94]. Саме ці параметри є головними для виявлення при аналізі мережевого трафіка.

Отож, модель порушника при атаці на відмову в обслуговуванні буде виглядати відповідно до рис. 2.8.

Під час розвідувальних атак порушник намагається виявити активні вузли мережі, тому надсилає ping-запити на весь діапазон IP-адрес потрібної йому мережі. Зазвичай, використовують інструменти розгортання ping-запитів, що послідовно перебирають усі IP-адреси мережі послідовно один за одним. Під час сканування IP-адрес йде певна кількість запитів на кожну адресу.

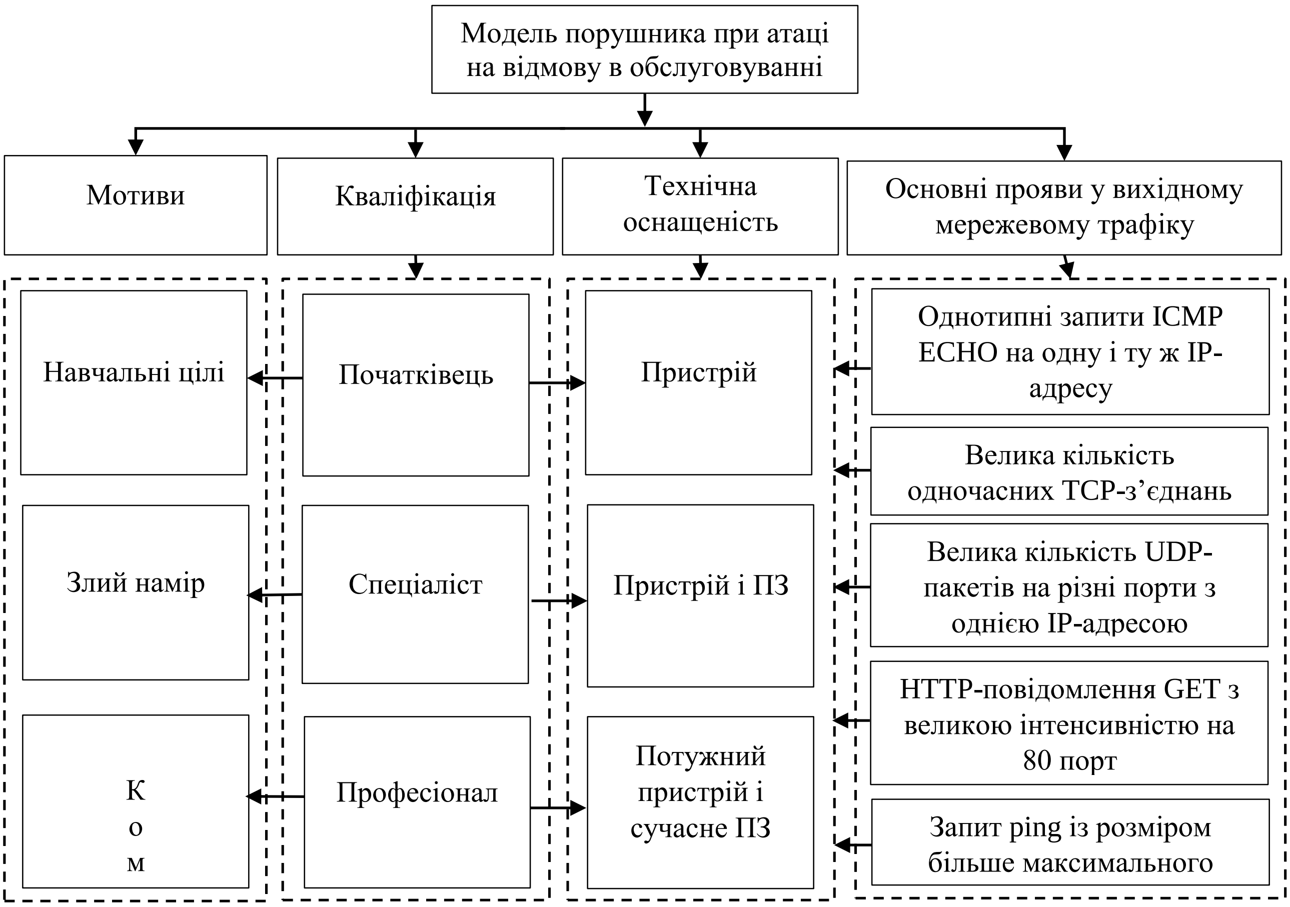
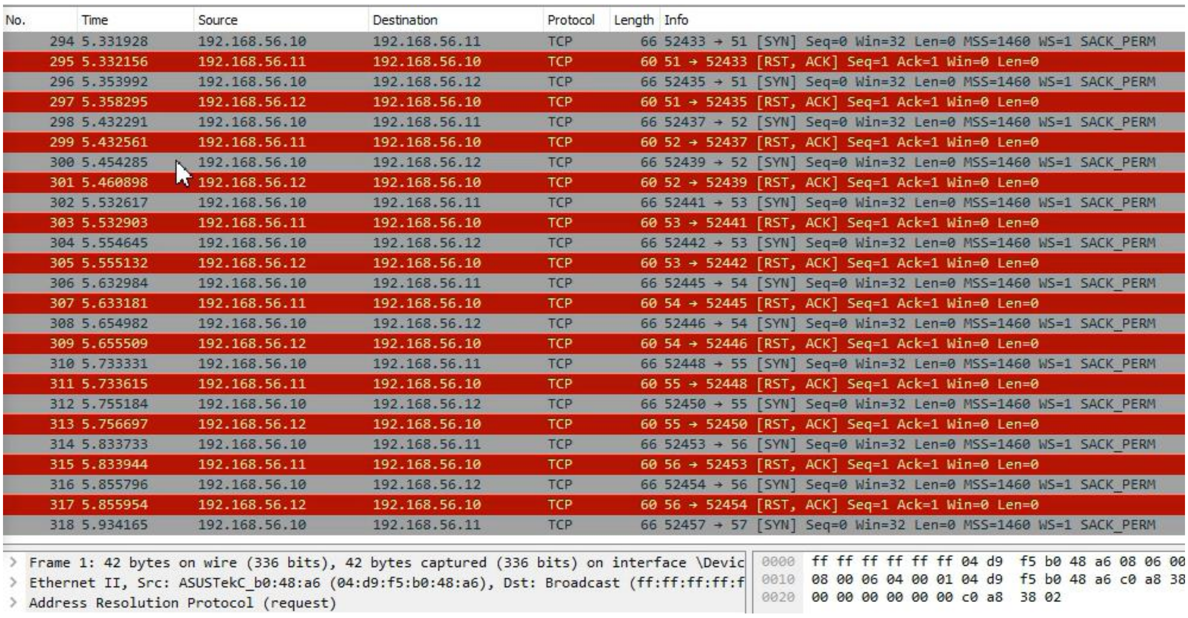


Рисунок 2.8 - Модель порушника при атаці на відмову в обслуговуванні

Оскільки відсутність однієї відповіді ще не свідчить про його вимкнений стан чи відсутність в мережі, оскільки міг бути зайнятий іншим запитом. Часовий інтервал з яким здійснювати запити на одну і ту ж адресу та кількість запитів для кожного сканування може відрізнятися, оскільки порушник може їх легко змінити від потреб та можливостей. До прикладу, велика кількість одночасних звертань, що йдуть через один мережевий пристрій, може вивести з ладу мережеве обладнання, але бажаного результату порушник не отримає. Тому сканування за IP-адресами «розтягнуте» в часі для отримання потрібного результату. Під час сканування використовується протокол ARP для запиту та ICMP протокол для відповіді, якщо адреса доступна (працює, включена). Коли IP-адресу ідентифіковано як робочу, то може розпочатися процес сканування портів. Під час нього використовуються одна IP-адреса, змінюються лише номери портів (зазвичай при скануванні використовують діапазон портів 0-1023). Кількість потоків сканувань адрес та портів може налаштовувати порушник залежно від пропускної спроможності мережі. Приклад трафіку зі сторони порушника при скануванні портів наведено на рис.2.9.



No.	Time	Source	Destination	Protocol	Length	Info
294	5.331928	192.168.56.10	192.168.56.11	TCP	66	52433 → 51 [SYN] Seq=0 Win=32 Len=0 MSS=1460 WS=1 SACK_PERM
295	5.332156	192.168.56.11	192.168.56.10	TCP	60	51 → 52433 [RST, ACK] Seq=1 Ack=1 Win=0 Len=0
296	5.353992	192.168.56.10	192.168.56.12	TCP	66	52435 → 51 [SYN] Seq=0 Win=32 Len=0 MSS=1460 WS=1 SACK_PERM
297	5.358295	192.168.56.12	192.168.56.10	TCP	60	51 → 52435 [RST, ACK] Seq=1 Ack=1 Win=0 Len=0
298	5.432291	192.168.56.10	192.168.56.11	TCP	66	52437 → 52 [SYN] Seq=0 Win=32 Len=0 MSS=1460 WS=1 SACK_PERM
299	5.432561	192.168.56.11	192.168.56.10	TCP	60	52 → 52437 [RST, ACK] Seq=1 Ack=1 Win=0 Len=0
300	5.454285	192.168.56.10	192.168.56.12	TCP	66	52439 → 52 [SYN] Seq=0 Win=32 Len=0 MSS=1460 WS=1 SACK_PERM
301	5.460898	192.168.56.12	192.168.56.10	TCP	60	52 → 52439 [RST, ACK] Seq=1 Ack=1 Win=0 Len=0
302	5.532617	192.168.56.10	192.168.56.11	TCP	66	52441 → 53 [SYN] Seq=0 Win=32 Len=0 MSS=1460 WS=1 SACK_PERM
303	5.532903	192.168.56.11	192.168.56.10	TCP	60	53 → 52441 [RST, ACK] Seq=1 Ack=1 Win=0 Len=0
304	5.554645	192.168.56.10	192.168.56.12	TCP	66	52442 → 53 [SYN] Seq=0 Win=32 Len=0 MSS=1460 WS=1 SACK_PERM
305	5.555132	192.168.56.12	192.168.56.10	TCP	60	53 → 52442 [RST, ACK] Seq=1 Ack=1 Win=0 Len=0
306	5.632984	192.168.56.10	192.168.56.11	TCP	66	52445 → 54 [SYN] Seq=0 Win=32 Len=0 MSS=1460 WS=1 SACK_PERM
307	5.633181	192.168.56.11	192.168.56.10	TCP	60	54 → 52445 [RST, ACK] Seq=1 Ack=1 Win=0 Len=0
308	5.654982	192.168.56.10	192.168.56.12	TCP	66	52446 → 54 [SYN] Seq=0 Win=32 Len=0 MSS=1460 WS=1 SACK_PERM
309	5.655509	192.168.56.12	192.168.56.10	TCP	60	54 → 52446 [RST, ACK] Seq=1 Ack=1 Win=0 Len=0
310	5.733331	192.168.56.10	192.168.56.11	TCP	66	52448 → 55 [SYN] Seq=0 Win=32 Len=0 MSS=1460 WS=1 SACK_PERM
311	5.733615	192.168.56.11	192.168.56.10	TCP	60	55 → 52448 [RST, ACK] Seq=1 Ack=1 Win=0 Len=0
312	5.755184	192.168.56.10	192.168.56.12	TCP	66	52450 → 55 [SYN] Seq=0 Win=32 Len=0 MSS=1460 WS=1 SACK_PERM
313	5.756697	192.168.56.12	192.168.56.10	TCP	60	55 → 52450 [RST, ACK] Seq=1 Ack=1 Win=0 Len=0
314	5.833733	192.168.56.10	192.168.56.11	TCP	66	52453 → 56 [SYN] Seq=0 Win=32 Len=0 MSS=1460 WS=1 SACK_PERM
315	5.833944	192.168.56.11	192.168.56.10	TCP	60	56 → 52453 [RST, ACK] Seq=1 Ack=1 Win=0 Len=0
316	5.855796	192.168.56.10	192.168.56.12	TCP	66	52454 → 56 [SYN] Seq=0 Win=32 Len=0 MSS=1460 WS=1 SACK_PERM
317	5.855954	192.168.56.12	192.168.56.10	TCP	60	56 → 52454 [RST, ACK] Seq=1 Ack=1 Win=0 Len=0
318	5.934165	192.168.56.10	192.168.56.11	TCP	66	52457 → 57 [SYN] Seq=0 Win=32 Len=0 MSS=1460 WS=1 SACK_PERM

> Frame 1: 42 bytes on wire (336 bits), 42 bytes captured (336 bits) on interface \Devic
> Ethernet II, Src: ASUSTekC_b0:48:a6 (04:d9:f5:b0:48:a6), Dst: Broadcast (ff:ff:ff:ff:f
> Address Resolution Protocol (request)

0000 ff ff ff ff ff ff 04 d9 f5 b0 48 a6 08 06 00
0010 08 00 06 04 00 01 04 d9 f5 b0 48 a6 c0 a8 38
0020 00 00 00 00 00 00 c0 a8 38 02

Рисунок 2.9 – Мережевий трафік зі сторони зловмисника при опитування портів

При зборі інформації про хост-жертву методом активного сканування порушник надсилає запити на адресу хоста з використанням усіх можливих протоколів та отримує відповіді. Головною метою є пошук відкритих баз даних чи портів, аналіз програмного забезпечення, що використовується та структури мережі з метою виявлення вразливостей для здійснення подальших зловмисних дій. Етапи виконання атаки будуть аналогічні до моделі порушника при атаці на пароль за винятком третього пункту. Для поточної атаки він буде наступним:

- Порушник приєднується до відкритого сегменту мережі та запускає програму сканування IP-адрес.

- Якщо IP-адреса дає відповідь, то за цим адресом запускається сканування портів.

Під час аналізу трафіка дані дії проявлятимуться у послідовних IP-адресах одержувача та великій кількості запитів на одну IP-адресу з використанням послідовно певного діапазону портів. Тому модель порушника можна зобразити к на рис. 2.10.

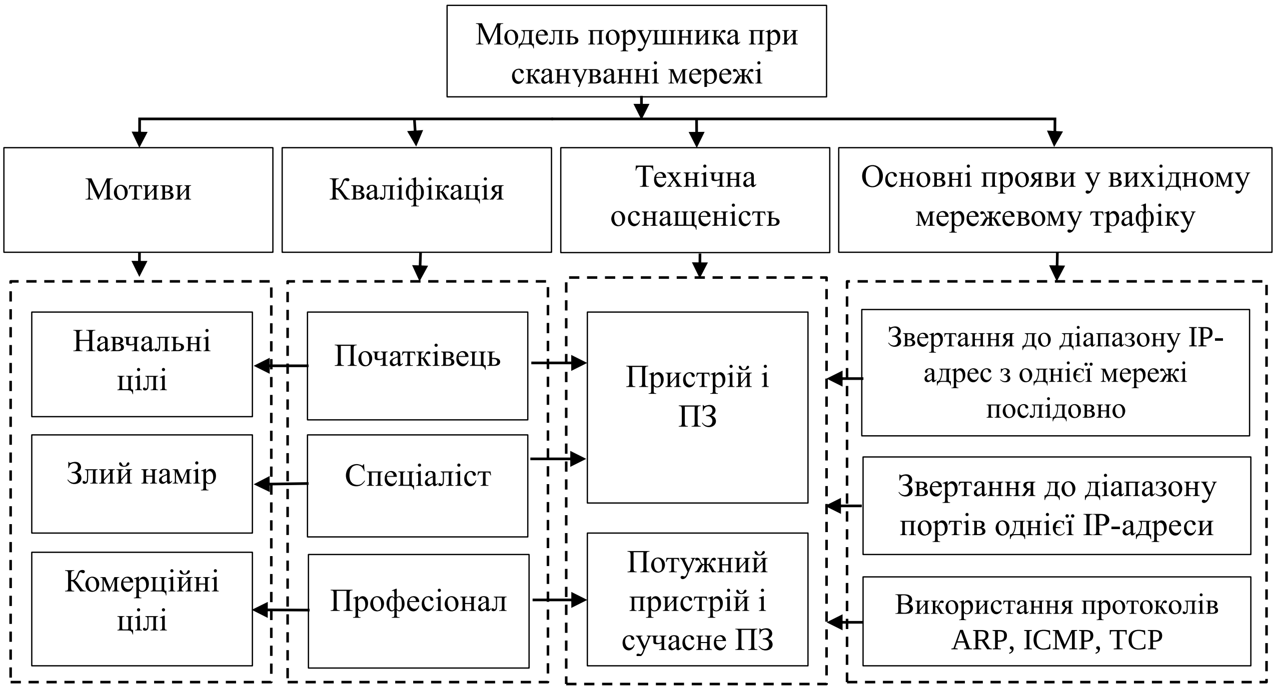


Рисунок 2.10 - Модель порушника при скануванні мережі

Багато атак здійснюється за допомогою використання шкідливого ПЗ чи зміною привілеїв облікового запису системи на роботу якої намагаються вплинути порушники. Проте атаки такого типу досить важко відслідкувати, аналізуючи вихідний трафік. Оскільки пошук відомих сигнатур шкідливого програмного забезпечення при аналізі даних вмісту пакета є довготривалим процесом у порівнянні з просто передачею пакета. Тому це буде додаткове навантаження на мережеве обладнання, сповільнення роботи всієї мережі. Але характерною ознакою початку чи наміру зловмисних дій є налаштування віддаленого доступу до комп’ютера чи будь-якого іншого пристрою, що знаходиться поза мережею аналіз якої здійснюється. Саме тому слід відслідковувати протоколи, які використано при з’єднанні. Під час вивчення даного питання та проведення дослідження (рис. 2.11) можна сказати, що при роботі з віддаленим пристроєм переважно функціонують протоколи SSL, TPKT, TCP, RDP-UDP.

Етапи виконання атак з використанням в тому числі віддаленого робочого столу будуть аналогічні до етапів виконання атаки на пароль. На основі отриманих даних будуємо модель порушника, що зображена на рис. 2.12.

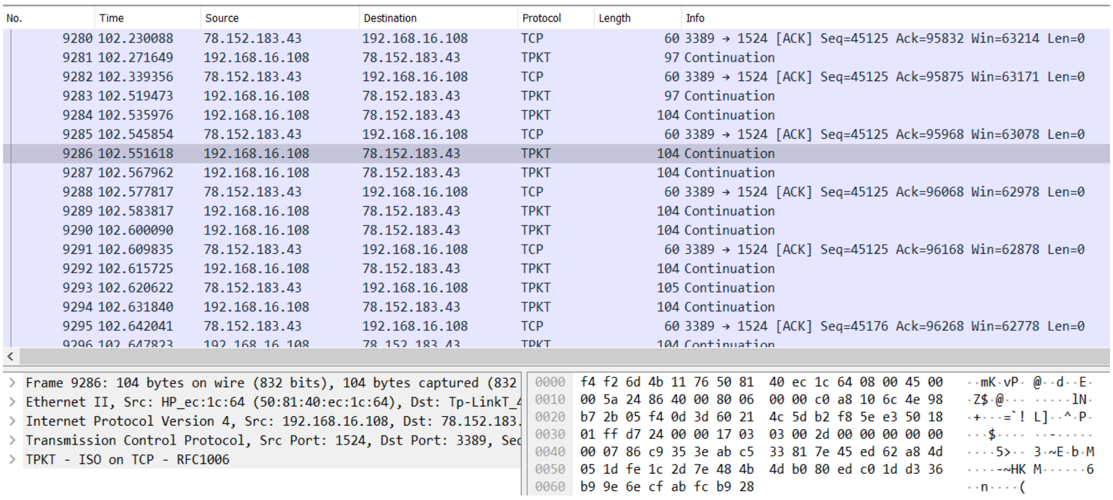


Рисунок 2.11 – Трафік зі сторони зловмисника при віддаленій роботі

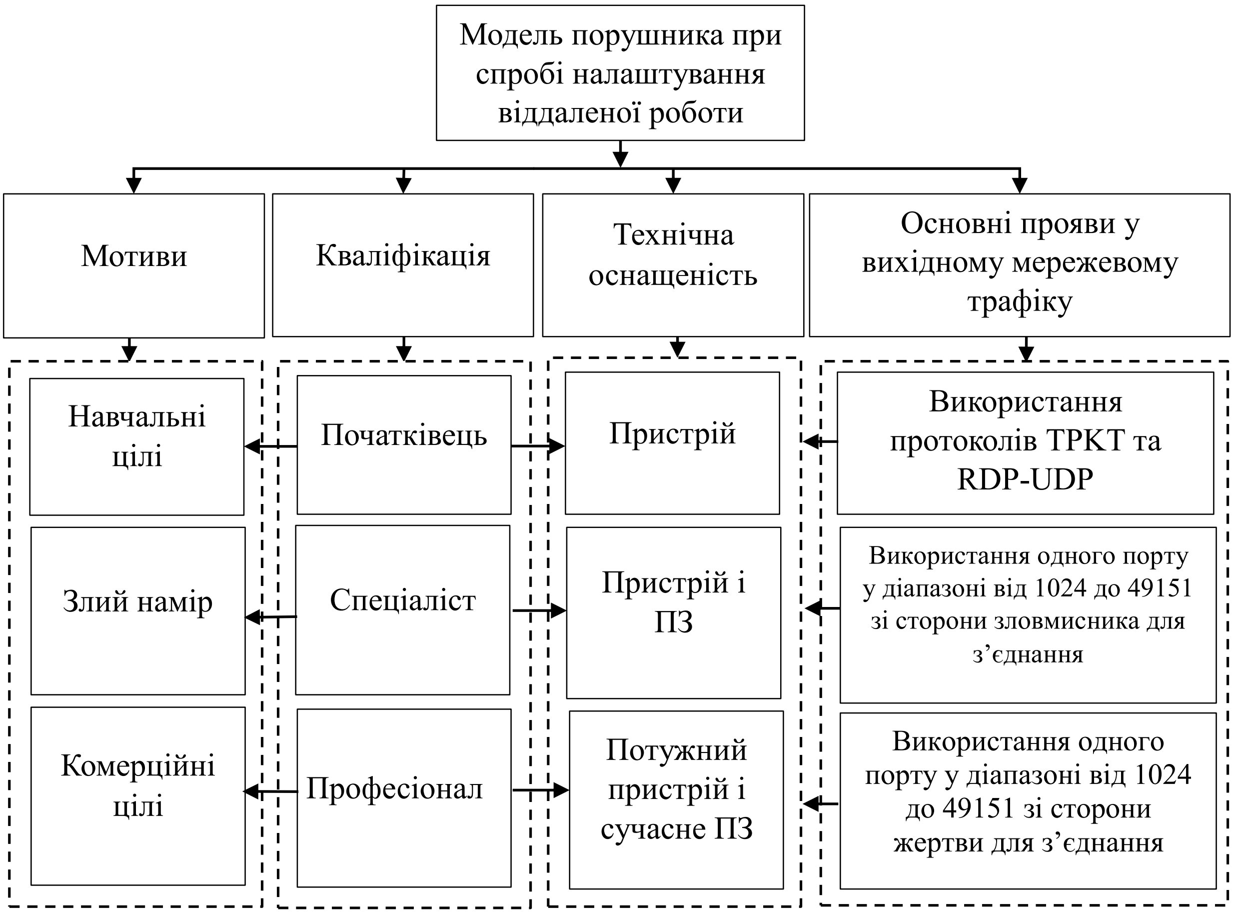


Рисунок 2.12 - Модель порушника під час налаштування віддаленої роботи

Електронну пошту для розсилки спаму використовують не лише у рекламних цілях різні організації, де активно використовуються різні лінгвістичні засоби та власник поштової скриньки міг здійснювати реєстрацію на певному сайті чи робив замовлення. Також під час розсилки листів можуть бути надіслані посилання для здійснення фішингових атак або ж містити вкладення, які являють собою шкідливе програмне забезпечення. Переважно теми листів у таких випадках є популярні чи соціально важливі. Згідно CERT-UA у листах з шкідливим вмістом за 2022 рік переважали наступні теми: грошові виплати, допомога від Червоного Хреста та від імені Держспецзв'язку, СБУ чи інших органів влади.

При аналізі трафіку розсилку спаму через пошту можна відслідкувати за допомогою великої кількості запитів на один чи кілька поштових серверів, оскільки листи не надходять напряму до одержувача. Також використання SMTP-протоколу з 25 порту на 25 порт, оскільки це службовий порт що відповідає за передачу e-mail між поштовими серверами.

Під час такої атаки кроки виконання буде сформовано наступним чином:

Крок 1. Підготовка до здійснення атаки. На цьому етапі порушник буде займатись збором адрес на які буде здійснено розсилку. Збір адрес можна проводити різними методами: від купівлі баз клієнтів до парсингу сайтів.

Крок 2. Розробка плану атаки. Порушник визначається із мережею, до якої буде приєднуватись для здійснення атаки, та способом розсилки повідомлень. Адже від поштового сервісу залежить максимально можливий ліміт відправлених листів за добу. Якщо це безкоштовна розсилка з сайту, то ліміт сягатиме 500 листів на добу. Якщо звичайна поштова скринька, то 5000 листів на добу. Ці значення вказано за замовчуванням та їх можна коригувати. Для платних профілів кількість листів за добу залежить від тарифу. Але при цьому слід враховувати розкриття інформації про порушника згідно якої він проводитиме реєстрацію та оплату. Також до початку атаки потрібно підготувати текст. Для того, щоб теми листів автоматично змінювалися та було важче ідентифікувати лист як спам порушники можуть використовувати шкідливий програмний код.

Крок 3. Реалізація. Порушник приєднується до відкритого сегменту мережі та розпочинає безпосередню атаку. Під час якої можна за аналізом трафіка виявити факт порушення. Для цього слід при скануванні вихідного мережевого трафіку аналізувати вхідні та вихідні порти та IP, частоту звернень, порти.

Швидкість здійснення розсилки безпосередньо залежить від кількості листів, що буде розіслано та можливостей мережі.

Крок 4. Після успішного завершення атаки порушник намагається приховати ознаки своєї участі в розсилці листів та очікує дій зі сторони отримувачів. Це може бути перехід за посиланням чи запуск шкідливого коду.

На основі попередньо викладеного матеріалу сформуємо модель порушника, що зображена на рис. 2.13.

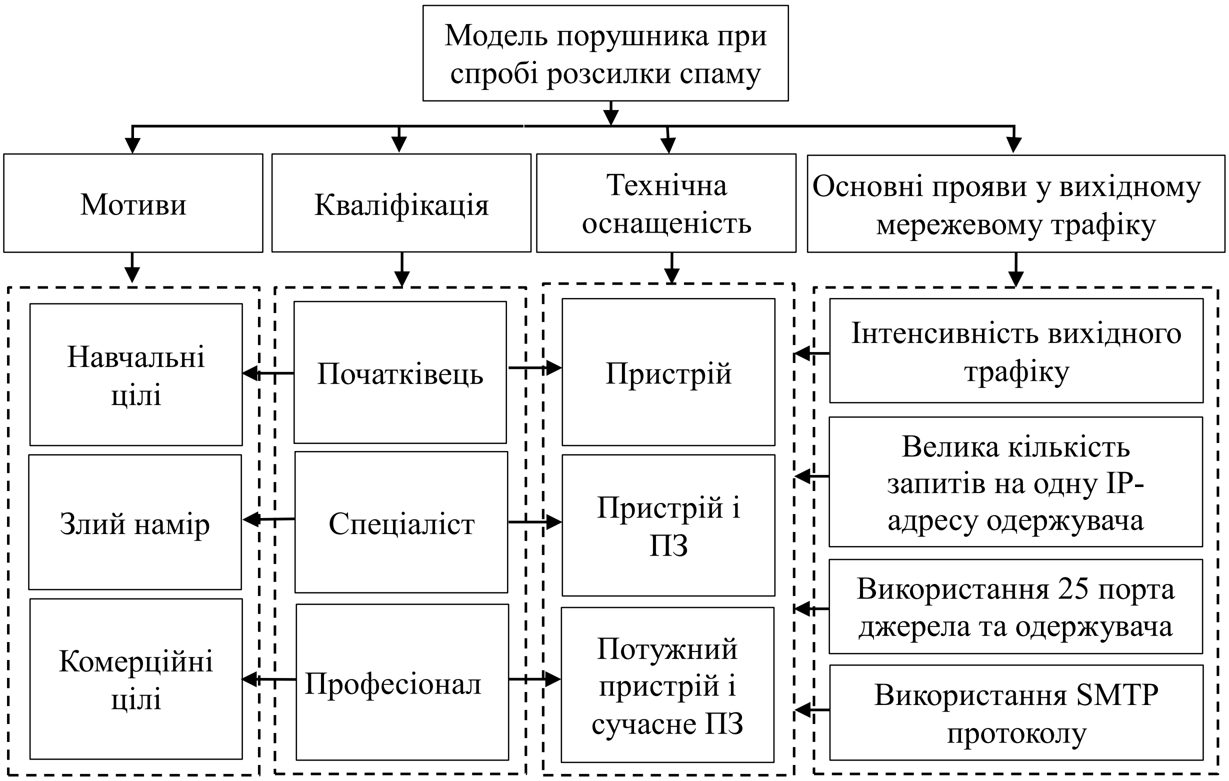


Рисунок 2.13 - Модель порушника при спробі розсилки спаму

Тому можна сформувати узагальнену модель порушника (рис.2.14) при здійсненні атак із відкритого сегменту мережі.

Оскільки більшість вмісту пакета займають дані, то для аналізу трафіку використано тільки заголовки. У табл. 2.6 представлено елементи заголовків, які потрібні при відслідковуванні, та типи атак. Із таблиці видно, що частина даних не використовується під час аналізу трафіку. Це пов’язано із тим, що вони містять виключно службову інформацію чи їх аналіз буде завдавати додаткове навантаження на систему.

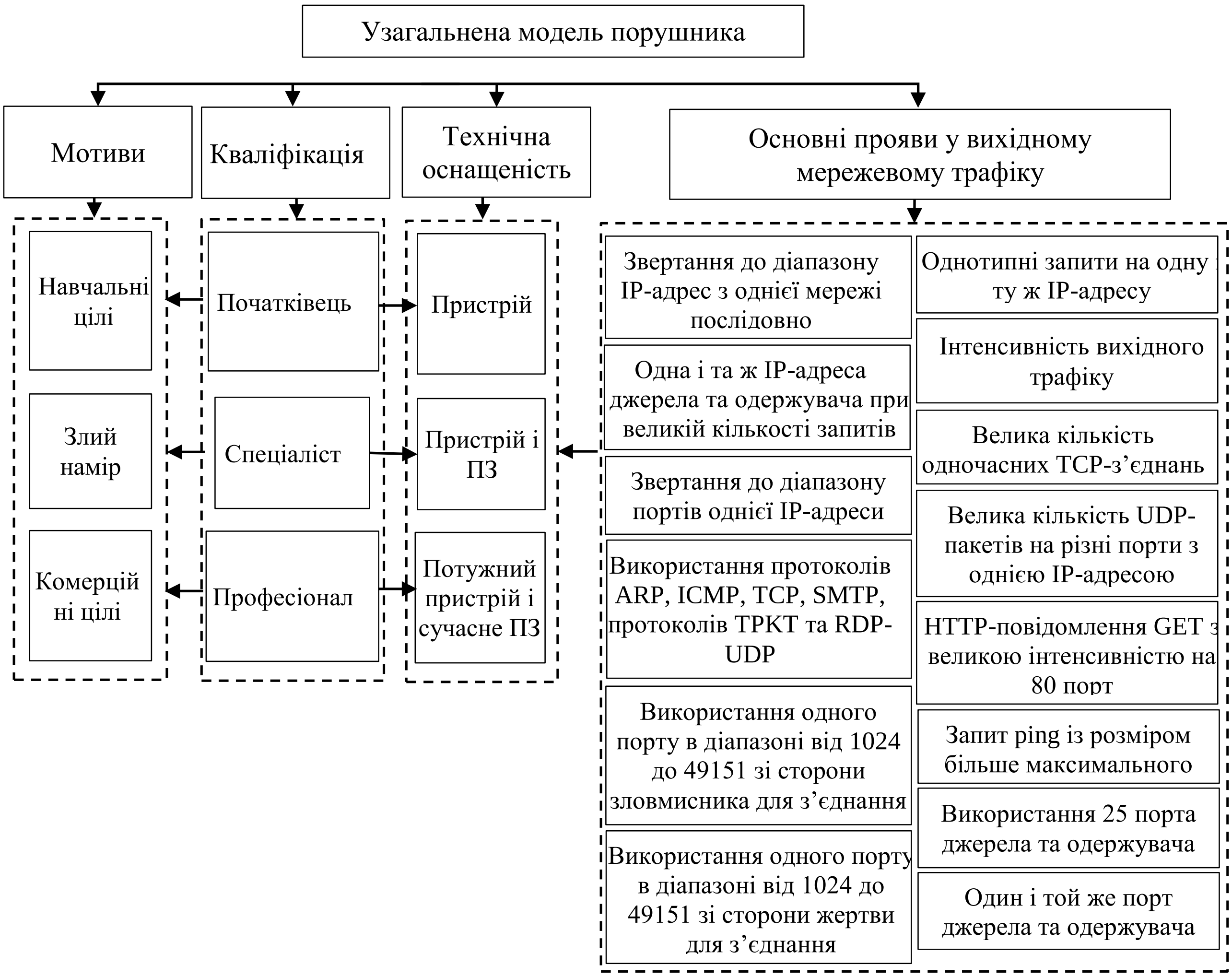


Рисунок 2.14 – Узагальнена модель порушника

Таблиця 2.6 – Співвідношення елементів заголовків до типів атак

№ пп	Назва атаки / Елемент заголовка	Атака на пароль	Відмова в обслуговуванні	Розвідувальні атаки	Віддалена робота	Розсилка спаму
1	2	3	4	5	6	7
1	Acknowledgment Number	ні	ні	ні	ні	ні
2	Checksum	ні	так	ні	ні	ні
3	Data	ні	ні	ні	ні	ні
4	Data offset	ні	так	ні	ні	ні
5	Destination IP address	так	так	так	так	так
6	Destination Port	так	так	так	так	так

Кінець таблиці 2.6

1	2	3	4	5	6	7
7	DSCP	ні	ні	ні	ні	ні
8	ECN	ні	ні	ні	ні	ні
9	Flags	ні	ні	ні	ні	ні
10	Flow Label	ні	ні	ні	ні	ні
11	Fragment Offset	ні	ні	ні	ні	ні
12	Header Checksum	ні	так	ні	ні	ні
13	Header Length	ні	так	ні	ні	ні
14	Hop Limit	ні	ні	так	ні	ні
15	Identification	ні	ні	ні	ні	ні
16	Next Header	ні	ні	ні	ні	ні
17	Options	ні	ні	ні	ні	ні
18	Payload Length	ні	так	ні	ні	ні
19	Protocol		так	так	так	так
20	Reserved	ні	ні	ні	ні	ні
21	Sequence Number	ні	ні	ні	ні	ні
22	Source IP address	так	так	так	так	так
23	Source Port	так	так	так	так	так
24	Time to live	ні	ні	ні	ні	ні
25	Total Length	ні	так	ні	ні	ні
26	Traffic Class	ні	ні	ні	ні	ні
27	Urgent Point	так	ні	ні	так	ні
28	Version	ні	ні	ні	ні	ні
29	Window size	ні	ні	ні	ні	ні

Отож, залежно від типу атаки може змінюватися модель порушника та поведінка шкідливого трафіку. У більшості випадків атаки зазнають відомі порти. Проте є атаки, де не здійснюється атака на порти, а відбувається велика кількість запитів на одну і ту ж IP. Також важливим параметром для усіх типів атак є інтенсивність трафіку, інформацію про який не можна отримати із заголовків пакетів. Саме тому слід розробити сигнатуру пакету за якою можна буде аналізувати мережевий трафік.

2.3 Розробка моделі сигнатури пакету для пошуку аномального трафіку

На початковому етапі буде побудовано граф залежності, який ілюструє взаємозв'язок між елементами заголовків мережевих пакетів та типами атак (рис. 2.15). Цей граф демонструє значення кожного параметра у процесі ідентифікації різних типів атак [95].

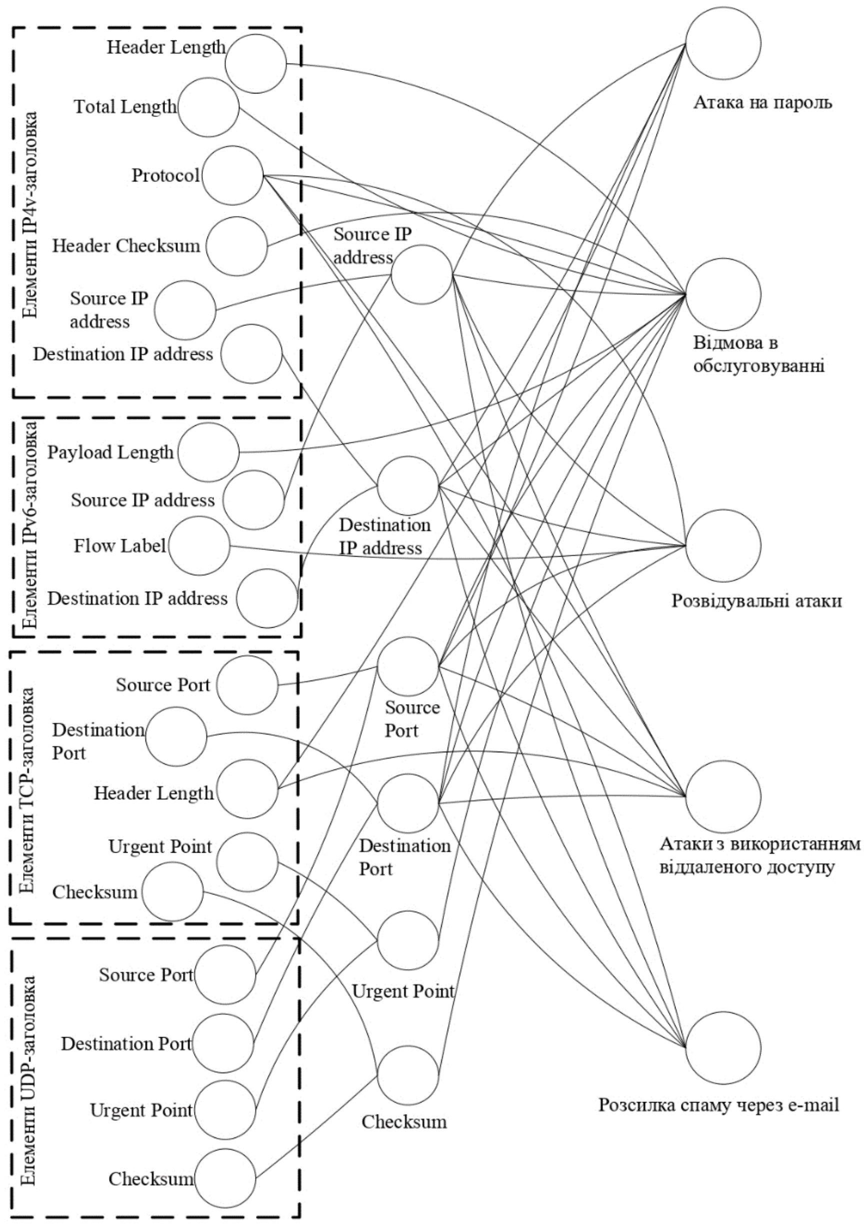


Рисунок 2.15 - Граф залежності елементів заголовків до типів атак

Велика кількість параметрів сповільнить аналіз трафіку та збільшить навантаження на обладнання. Тому доцільно обрати параметри з найбільшою ефективністю з метою збереження стабільної роботи мережі (рис.2.16). Використаємо принцип Парето, який говорить що 20% параметрів забезпечують 80% ефективності. Одиниці дорівнює один тип атаки. Максимальному значенню п’ять буде рівний параметр при умові, що він потрібен для пошуку порушника для всіх типів атак. Слід зазначити, що при цьому аналізі з’явиться показник інтенсивності трафіку [96].

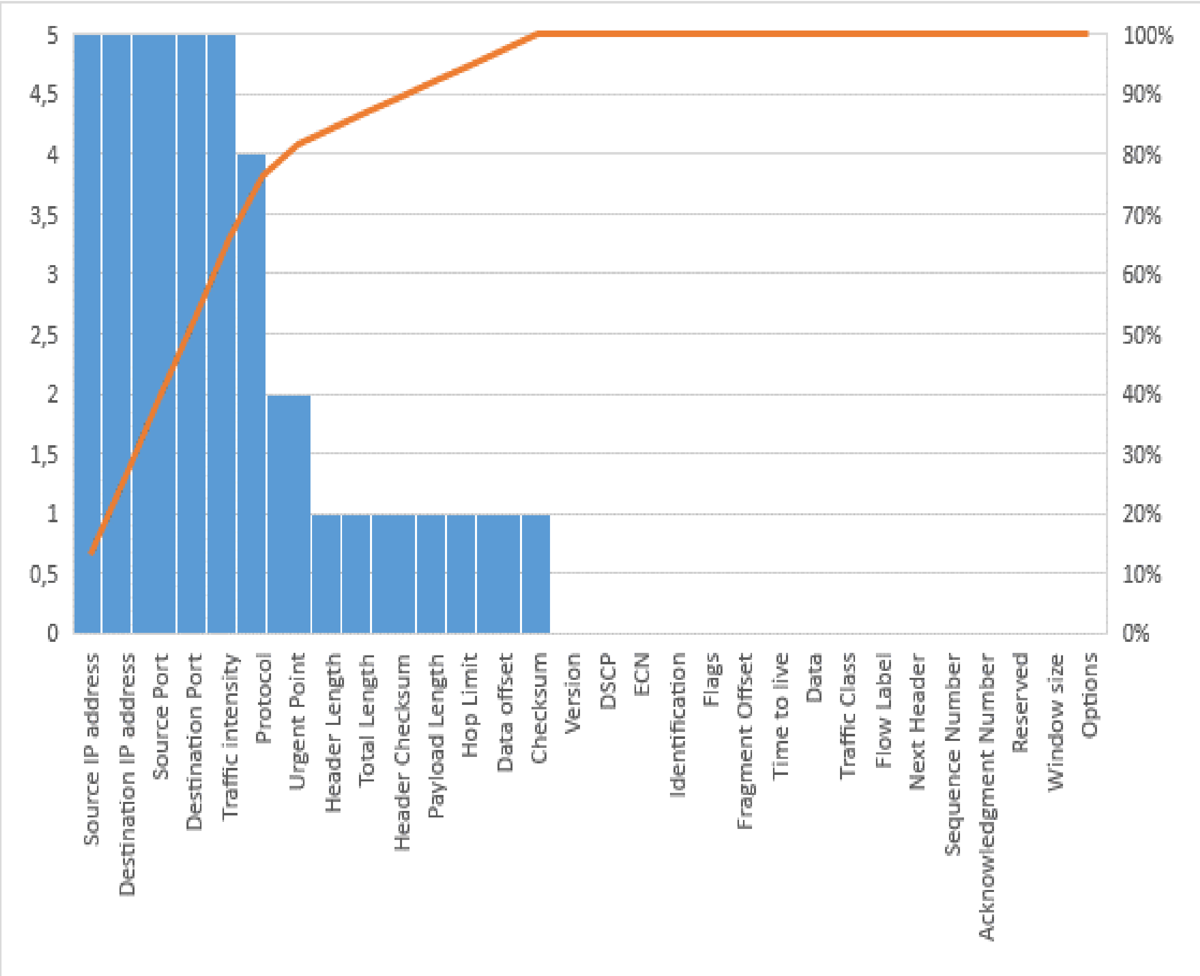


Рисунок 2.16 – Діаграма з використанням принципу Парето

З діаграми видно, що оптимально використовувати наступні параметри (рис. 2.17): порти джерела та одержувача, IP-адреси джерела та одержувача, протокол, інтенсивність трафіку.

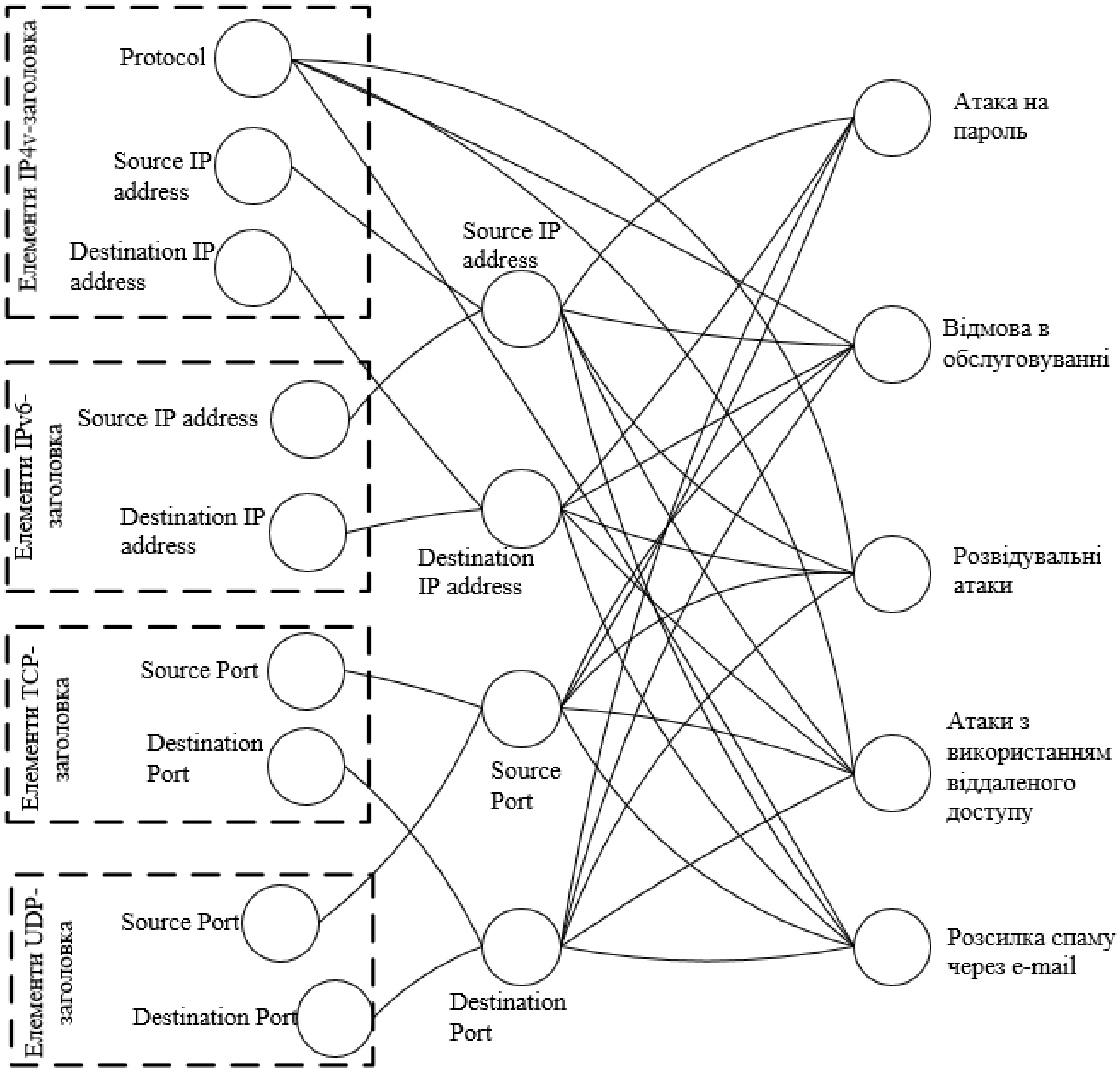


Рисунок 2.17 – Оптимізований граф залежності елементів заголовків до типів атак після використання принципу Парето

Сформуємо сигнатуру пакету, яка потрібна для подальшого аналізу. Потрібно додати часову відмітку, яка не впливає на аналіз, але буде використана для тривалості зберігання параметрів у базі.

Сигнатуру пакету буде представлено у вигляді багатокомпонентного кортежу [97]:

$$d = < d^1, d^2, \ldots, d^k >, \tag{2.1}$$

де k – кількість компонентів в кортежі.

Для розглянутого випадку приймемо $\mathrm{k} = 9$ та визначимо ці компоненти:

d^1 - IP-адреса джерела, який надсилає запит на з'єднання та обмін інформацією,

d^2 - IP-адреса призначення, тобто до якої IP-адреси надсилаються запити,

d^3 - порт джерела який використано для встановлення з'єднання,

d^4 - порт призначення,

d^5 – протокол, який використано для передачі даних,

d^6 – інтенсивність трафіку, яка визначається в біт/с,

d^7 - час надходження пакету на перевірку у 24х-годинному форматі гг-хх-сс,

d^8 – MAC-адреса пристрою, який надсилає дані із мережі,

d^9 – розмір пакету.

Значення параметрів d^1, d^3, d^8 використовуються для блокування пристрою, що генерує аномальний трафік. Значення із множини d^7 призначено для зберігання IP-адрес у білих та чорних списках протягом певного часу після надходження аномальних даних від пристрою, по завершенню вказаного часового проміжку дані про пристрій буде видалено із списку заборонених пристроїв тим самим час на перевірку заборонених з'єднань не буде зростати з часом. Оскільки використовується лінійний пошук перевірки елемента з усіма елементами бази даних, то час змінюється пропорційно до кількості елементів у базі:

$$t = k * w, \tag{2.2}$$

де t – загальний час перевірки,

k - константа, що залежить від швидкості виконання однієї операції перевірки,

w - кількість елементів у базі даних.

Сигнатура пакету дозволить однозначно ідентифікувати джерело трафіку, що його ініціює, визначати розмір пакету та час, коли він потрапив на аналіз [98].

Вхідний трафік представимо у вигляді множини сигнатур:

$$D = \bigcup_{i=0}^{N_d} \{d_i\}\,, \tag{2.3}$$

де d_i – елемент множини, вхідна сформована сигнатура,

i – початкове значення,

N_d – кількість елементів множини вхідних сигнатур.

Кожна сформована сигнатура буде належати до множини D:

$$\forall d_i \in D \tag{2.4}$$

2.4 Вдосконалення моделі процесу класифікації трафіку за ознаками

З метою класифікації трафіку (2.3) на дозволений, заборонений, та такий, що потребує подальшого аналізу використаємо множини дозволених, заборонених та невизначених сигнатур.

Модель класифікації трафіку за ознаками представимо у вигляді:

$$MO = < D, D^G, D^B, D^U >$$

де D – множина вхідного трафіку,

D^G – множина дозволених сигнатур,

D^B – множина заборонених сигнатур,

D^U – множина невизначених сигнатур.

Представимо множину дозволених сигнатур (D^G) наступним чином:

$$D^G = \bigcup_{i=0}^{N_g} \{d_i\}\,, \tag{2.5}$$

де N_g – кількість дозволених сигнатур,

d_i – сигнатура, віднесена до множини дозволених сигнатур (D^G).

Представимо множину заборонених сигнатур (D^B) наступним чином:

$$D^B = \bigcup_{i=0}^{N_b} \{d_i\}, \tag{2.6}$$

де N_b – кількість заборонених сигнатур,

d_i – сигнатура, віднесена до множини заборонених сигнатур (D^B).

Представимо множину невизначених сигнатур (D^U) наступним чином:

$$D^U = \bigcup_{i=0}^{N_u} \{d_i\}, \tag{2.7}$$

де N_u – кількість невизначених сигнатур,

d_i – сигнатура, віднесена до множини невизначених сигнатур (D^U).

Процес класифікації трафіку на основі аналізу сигнатур представимо у вигляді розбиття множини D:

Згідно цього розбиття якщо сигнатура аналізованого пакету (d_j) містить IP адреси джерела (d^1) і отримувача (d^2) такі, що є в множині D, то така сигнатура переноситься до множини дозволених (D^G):

$$D^G = \left\{d_j \in D \middle| \text{Якщо } \exists d_i \in D^G \text{ така, що } d_i^1 = d_j^1 \wedge d_i^2 = d_j^2 \right\} \tag{2.8}$$

де d_j – сигнатура вхідного трафіку, щодо якої проводиться аналіз.

Якщо сигнатура аналізованого пакету (d_j) містить IP адреси джерела (d^1) або отримувача (d^2) такі, що є в множині заборонених сигнатур (D^B), то така сигнатура переноситься в множину заборонених (D^B):

$$D^B = \left\{d_j \in D \middle| \text{Якщо } \exists d_i \in D^B \text{ така, що } d_i^1 = d_j^1 \vee d_i^2 = d_j^2 \right\} \tag{2.9}$$

В усіх інших випадках сигнатура (d_j) додається до множини невизначених сигнатур (D^U):

$$D^U = \{d_j \in D \mid \text{Якщо } d_j \notin D^G \wedge d_j \notin D^B\} \quad (2.10)$$

Оскільки множини D^G, D^B та D^U є розбиттям множини D то справедливими будуть такі твердження:

$$D = D^G \cup D^B \cup D^U \quad (2.11)$$

$$D^G \cap D^B = \emptyset,\ D^G \cap D^U = \emptyset,\ D^B \cap D^U = \emptyset \quad (2.12)$$

2.5 Вдосконалення моделі процесу класифікації трафіку на основі самоподібності

Самоподібність є одним із критеріїв аналізу мережевого трафіку, що дозволяє оцінити подібність нових сигнатур до раніше відомих, що знаходяться у множині дозволених сигнатур яку описано в параграфі 2.4. Використання самоподібності дає змогу класифікувати нові сигнатури, зменшуючи необхідність перевірки сигнатури відповідно до нечіткого методу виявлення аномального трафіку, який описано у параграфі 3.3. Основна перевага самоподібності полягає в її здатності виявляти стійкі закономірності в трафіку, навіть якщо вона змінюється в часі. Аналізуючи часові вікна потоку даних, можна використовувати раніше отримані дані для перевірки нових, що дозволяє адаптивно оновлювати множину дозволених сигнатур без потреби в повному перегляді всього набору даних.

Показник Херста використовується для кількісного визначення ступеня самоподібності потоку даних. Він дає можливість обчислювати рівень подібності між сигнатурами, що дозволяє з високою точністю оцінювати, чи слід відносити нові пакети до вже відомої множини дозволених даних або ж вони потребують додаткового аналізу. Це важливо для виявлення кіберзагроз, оскільки швидкість і точність класифікації трафіку виконує вагому роль у кіберзахисті. Завдяки самоподібності можна значно скоротити обсяг невизначеного трафіку, який потребує подальшого аналізу, що спрощує процес моніторингу мережевого трафіку та покращує загальну продуктивність системи виявлення аномалій.

З метою врахування самоподібності трафіку визначимо такі часові параметри (рис. 2.18).

$t_1 - t_3$ – часовий інтервал аналізу сигнатур на самоподібність, що складається з двох підінтервалів,

$t_1 - t_2$ – часовий інтервал раніше аналізованих на самоподібність сигнатур,

$t_2 - t_3$ – часовий інтервал нових сигнатур.

Рисунок 2.18 – Часові інтервали для класифікації трафіку

Модель процесу класифікації трафіку на основі самоподібності:

$$MS = < D, D^G, D^U, f(T), t_1, t_2, t_3, M, M' >,$$

де $f(T)$ – функція подібності,

t_1 – початок часового інтервалу аналізу,

t_2 – момент, який розділяє вже проаналізовані раніше сигнатури та нові, які щойно з'явилися,

t_3 – кінець інтервалу аналізу,

M – множина раніше аналізованих сигнатур на самоподібність,

M'– множина нових сигнатур, що раніше не аналізувались на самоподібність.

Визначимо множину M сигнатур, раніше аналізованих на самоподібність:

$$M = \left\{ d_i \in D^G \cup D^U \,\middle|\, t_1 < d_i^7 \leq t_2 \right\} \qquad (2.13)$$

та множину M' нових сигнатур, що раніше не аналізувались на самоподібність

$$M' = \{d_i \in D^G \cup D^U \,|\, t_2 < d_i^7 \leq t_3\}, \tag{2.14}$$

де d_i^7 – час надхлодження пакету на перевірку.

Визначимо функцію подібності

$$y = f(T), \tag{2.15}$$

де T – множина сигнатур, T⊂D.

y – числова міра подібності, $y \in [0,1]$.

З метою визначення подібності сигратури пакету, множині раніше визначених сигнатур пакетів, задамо функцію $S(n)$ стандартного відхилення:

$$S^j(n) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(d_i^j - \overline{d^J})^2}, \tag{2.16}$$

де $n = |T|$ – кількість сигнатур в множині T,

j – номер компонента в кортежі,

d_i^j – значення j-го компонента i-го кортежу множини,

$\overline{d^J}$ – середня значення j-го компонента кортежів множини.

Та функцію $R(n)$ розмах накопиченої відхилення функції.

$$R^j(n) = \max_{i=1,n} \sum_{m=1}^{k} d_i^j - \overline{d^J} - \min_{i=1,n} \sum_{m=1}^{k} d_i^j - \overline{d^J},$$

$$k = 1, n \tag{2.17}$$

Визначимо показник Херста для кожного компонента кортежу:

$$H^j = \frac{\ln(R^j(n)/S^j(n))}{\ln(n)} \quad (2.18)$$

Визначимо показник Херста для сигнатури пакету як усереднене значення показників Херста для кожного компонента кортежу:

$$H = f(T) = \frac{1}{k}\sum_{j=1}^{k} H^j \quad (2.19)$$

де j – кількість компонентів в кортежі.

Процес класифікації невизначеного трафіку (2.15) на основі аналізу подібності сигнатур представимо у вигляді:

$$D^G = \{D^G \cup M' | \, f(M \cup M') > 0{,}5\}, \quad (2.20)$$

$$D^U = \{D^U \backslash M' | \, f(M \cup M') > 0{,}5\}, \quad (2.21)$$

По завершенню аналізу в множині D^U залишаться сигнатури, що невизначені або неподібні дозволеним сигнатурам і потребують подальшого аналізу.

2.6 Розробка моделі процесу нечіткого виявлення аномального трафіку

Структуру системи реалізації нечіткого логічного висновку для класифікації ознак елементів сигнатури за допомогою Matlab зображено на рисунку 1. Вона складається з наступних блоків: фазифікація, функції приналежності, алгоритм нечіткого логічного висновку та дефазифікація. Загалом, послідовність виконання кроків системи (рис. 2.19) наступна: вхідні дані X подаються на блок фазифікації; на етапі фазифікації перетворюються вхідні дані на нечіткі значення X' за допомогою функцій приналежності; алгоритм нечіткого логічного висновку виконує імплікацію, агрегацію, активацію і акумуляцію на основі нечіткої бази знань для отримання нечітких вихідних значень Y'; дефазифікація перетворює нечіткі вихідні значення Y' на чітке вихідне значення Y; вихідне значення Y є результатом роботи системи, який використовується для прийняття рішень [99].

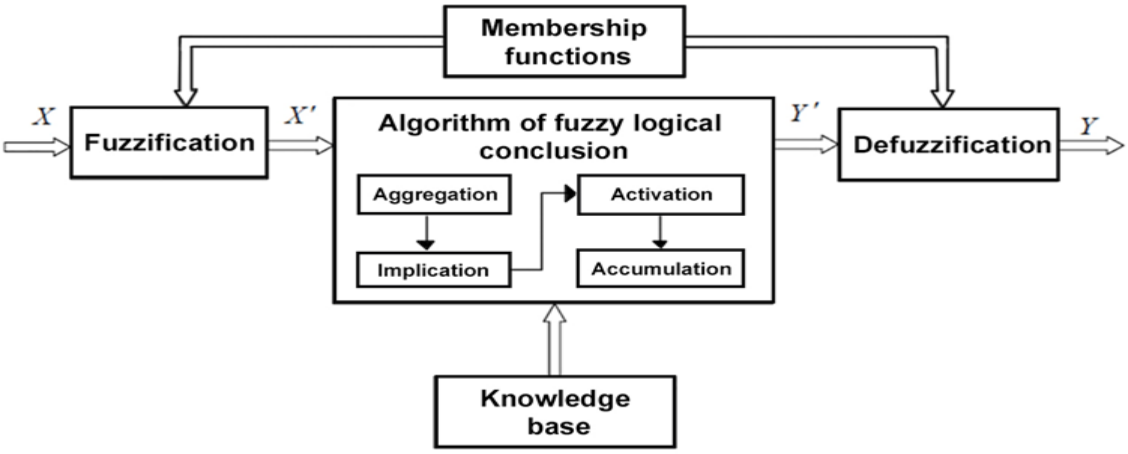


Рисунок 2.19 - Структура системи реалізації нечіткого логічного висновку

Під час фазифікації відбувається перетворення чітких вхідних даних X на нечіткі значення X' за допомогою функцій приналежності. Це дозволяє врахувати невизначеність і нечіткість у вхідних даних. До прикладу, d^6 дорівнює значенню 10000 біт/с, на етапі фазифікації за допомогою функцій приналежності вказана інтенсивність здійснює перетворення у нечітке значення «вище середнього».

Алгоритм нечіткого логічного висновку виконує процес отримання нечіткого логічного висновку на основі нечітких правил. Він складається з чотирьох етапів: імплікації, агрегації, активації та акумуляції. Імплікація перетворює вхідні нечіткі значення X' на вихідні нечіткі множини. Агрегація об'єднує результати імплікації від різних правил для кожного вихідного терму. Активація застосовує ступінь відповідності кожного правила до вихідної нечіткої множини із використанням методів мінімуму для визначення ступеня активації правила. Акумуляція об'єднує вихідні нечіткі множини в одну нечітку множину Y' для кожного вихідного терму.

Дефазифікація перетворює вихідні нечіткі значення Y' на чіткі вихідні значення Y. Це потрібно для отримання конкретного рішення на основі нечіткого висновку. Результатом роботи системи буде одне значення result, що приймає один із двох параметрів: «дозволено» [0, 50] чи «заборонено» [50, 100]. Отож, якщо сигнатура по завершенню перевірки системою отримала результат «дозволено», то відбувається дозвіл на передачу даних, а сигнатуру записано в базу дозволених з'єднань. Якщо ж сигнатуру визначено як «заборонену», то блокується з'єднання за IP та MAC адресою відправника, а сигнатуру добавлено до бази заборонених.

Висновки за розділом 2

1. Вдосконалено модель поведінки типового користувача та порушника у відкритому сегменті мережі, що дозволяє класифікувати мережевий трафік за поведінковими характеристиками.
2. Розроблено модель сигнатури пакета для пошуку аномального трафіку, яка оптимізує процес пошуку, що дозволяє ідентифікувати потенційні загрози на рівні окремих пакетів даних шляхом аналізу характерних ознак трафіку. Використання даної моделі забезпечує можливість швидкого та точного виявлення аномальних сигнатур без надмірного навантаження на обчислювальні ресурси системи моніторингу мережевого трафіку.
3. Вдосконалено модель процесу класифікації за ознаками, що дозволить класифікувати вже відомі сигнатури.
4. Вдосконалено модель процесу класифікації трафіку на основі самоподібності. Враховано статистичні властивості мережевого трафіку, що дозволяє виявляти довготривалі залежності та аналізувати відхилення від нормальної поведінки в мережі.
5. Розроблено модель процесу нечіткого виявлення аномального трафіку, яка використовує нечітку логіку для аналізу мережевого трафіку в умовах невизначеності. Модель дозволяє оцінювати рівень аномальності трафіку на основі експертних правил і нечітких множин, що сприяє покращенню точності детекції загроз навіть у випадках, коли традиційні методи дають суперечливі результати.

РОЗДІЛ 3
РОЗРОБКА МЕТОДІВ ВИЯВЛЕННЯ АНОМАЛЬНОГО ТРАФІКУ В ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ СИСТЕМАХ

3.1 Вдосконалення методу класифікації трафіку за ознаками

Вхідними даними є: пакети із вихідного потоку даних, множина вихідних сигнатур (D), множина дозволених сигнатур D^G, множина заборонених сигнатур D^B, множина невизначених сигнатур D^U.

Крок 1. Підключення множини дозволених сигнатур (D^G). На першому кроці завантажується файл, що являє собою список дозволених сигнатур які визначено як безпечні.

Крок 2. Підключення множини заборонених сигнатур (D^B). На другому кроці завантажується файл, що являє собою список заборонених сигнатур, які визначено як небезпечні.

Крок 3. Формування сигнатури пакету (d_j) із вихідного потоку даних (D), яку потрібно перевірити. Алгоритм формування сигнатури представлено у форматі кортежу, що описано у параграфі 2.3.

Крок 4. Перевірка IP-адреси джерела (d_1) та IP-адреси отримувача (d_2) на відповідність множині дозволених сигнатур (D^G). На даному кроці відбувається перевірка значення IP-адреси джерела (d_1) та IP-адреси отримувача (d_2) на належність до множини дозволених сигнатур (D^G), що містить всі попередньо верифіковані та безпечні IP-адреси, яким можна довіряти в мережі. Якщо IP-адреса джерела (d_1) та IP-адреса отримувача (d_2) міститься в множині дозволених сигнатур (D^G), то це свідчить про те, що трафік, який надходить від цього джерела, є безпечним і відбувається перехід до кроку 6. Якщо IP-адреса джерела (d_1) та IP-адреса отримувача (d_2) не знайдені у множині дозволених сигнатур (D^G), то відбувається перехід на крок 5 для перевірки цих адрес в множині заборонених сигнатур (D^B).

Крок 5. Перевірка IP-адреси джерела (d_1) та IP-адреси отримувача (d_2) на відповідність множині заборонених сигнатур (D^B). Множина заборонених сигнатур (D^B) складається з IP-адрес, які пов'язані з небезпечним або аномальним трафіком, такими як адреси ботнетів, відомі джерела спаму чи інші кіберзагрози. На даному етапі відбувається перевірка значення IP-адреси джерела (d_1) та IP-адреса отримувача (d_2) на належність до множини заборонених сигнатур (D^B). Якщо IP-адресу джерела (d_1) або IP-адресу отримувача (d_2) знайдено серед елементів множини D^B, то з'єднання для цієї сигнатури класифікується як заборонене і відбувається перехід на крок 7. В іншому випадку відбувається перехід на крок 8.

Крок 6. Надається дозвіл для з'єднання та запис сигнатури пакету d_j до множини дозволених сигнатур (D^G). Далі перехід на крок 9.

Крок 7. Відбувається блокування з'єднання та запис сигнатури пакету d_j до множини заборонених сигнатур (D^B) та видалення сигнатури пакету d_j із множини дозволених сигнатур (D^G). Далі перехід на крок 9.

Крок 8. Якщо жоден з компонентів (d_1 чи d_2 сигнатури d_j) не потрапляє до множини заборонених сигнатур, то з'єднання класифікується як невизначене та сигнатура записується до множини D^U. Далі перехід на крок 9.

Крок 9. Перевірка на необхідність виходу. Після того, як сигнатура була класифікована, метод аналізує наступну дію: завершити обробку трафіку (крок 10) або повернутися до кроку 3 для перевірки наступного пакету.

Крок 10. Якщо прийнято рішення, що потрібно завершити обробку, то метод завершить свою роботу, і наступні пакети не будуть перевірятися.

Результатом роботи є розподіл вихідного мережевого трафіку на три категорії: дозволений, заборонений та невизначений.

Алгоритм роботи методу класифікації трафіку за ознаками представлено на рис.3.1.

Розглянемо декілька прикладів роботи методу для різних сценаріїв.

Приклад 1. Пакет класифіковано як дозволений. Першим етапом є завантаження файлу, який містить множину дозволених сигнатур. Ця множина, включає IP-адреси, яким можна довіряти.

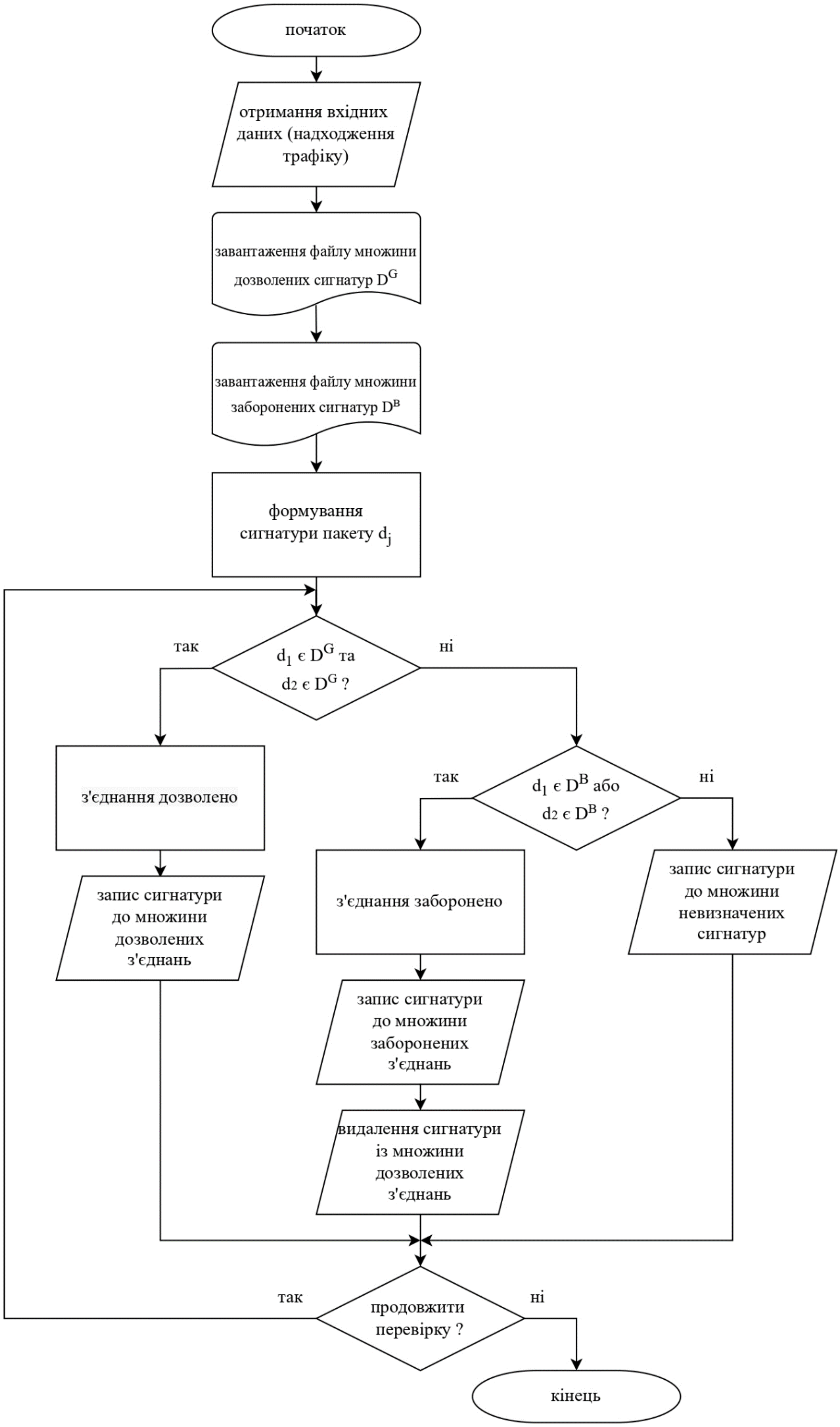


Рисунок 3.1– Алгоритм роботи методу класифікації трафіку за ознаками

У даному прикладі, до такого списку потрапляють наступні IP-адреси 103.168.1.10 та 192.168.1.15, що є адресами адміністратора мережі чи інших безпечних сервісів. На другому етапі завантажується файл із забороненими адресами. Прикладами таких адрес є 203.0.113.55 та 13.0.14.75, які можуть бути асоційовані з небажаними діями в мережі. Коли метод отримує пакет, то формує із нього сигнатуру. Формування сигнатури відбувається шляхом побудови кортежу, який містить основні компоненти пакету. У цьому випадку перший компонент сигнатури (d_1) містить IP-адресу джерела 103.168.1.10 та другий компонент сигнатури (d_2) IP-адресу отримувача 192.168.1.15. Наступним кроком метод перевіряє ці IP-адреси на наявність серед дозволених адрес. Вважатимемо, що IP-адреси 103.168.1.10 та 192.168.1.15 знаходяться в списку дозволених, отже метод класифікує з’єднання як дозволене. Тоді результатом ітерації виконання методу буде наступне: трафік від IP-адрес 103.168.1.10 та 192.168.1.15 вважатимуться безпечними, тому з’єднання дозволяється; пакет передається через мережу без блокувань; метод готовий перейти до наступної ітерації для перевірки наступного пакету і відбувається перехід на крок 3 якщо роботу методу продовжено.

Приклад 2. Пакет класифіковано як заборонений. Першим етапом є завантаження файлу, який містить множину дозволених сигнатур. У даному прикладі, до такого списку потрапляють IP-адреси, як-от 103.168.1.10 та 192.168.1.15. На другому етапі завантажується файл із забороненими адресами. Прикладами таких адрес є 203.0.113.55 та 13.0.14.75, які можуть бути асоційовані з небажаними діями в мережі. Коли метод отримує пакет, то формує із нього сигнатуру. У цьому випадку перший компонент сигнатури (d_1) містить IP-адресу джерела 203.0.113.55. Наступним кроком метод перевіряє цю IP-адресу на наявність серед дозволених адрес. Вважаючи, що IP-адреса 203.0.113.55 не входить у список дозволених, метод переходить до перевірки цієї адреси в списку заборонених сигнатур. Оскіцльки IP-адреса 203.0.113.55 виявляється в чорному списку, то метод класифікує з’єднання як заборонене. Тоді результатом ітерації виконання методу буде наступне: трафік від IP-адреси 203.0.113.55

вважається забороненим, тому з’єднання блокується; пакет не передається через мережу; метод готовий перейти до наступної ітерації для перевірки наступного пакету і відбувається перехід на крок 3 якщо роботу методу продовжено.

Приклад 3: Пакет класифіковано як невизначений. Першим етапом є завантаження файлу з дозволеними сигнатурами. У нашому випадку це адреси 103.168.1.10 та 192.168.1.15. Далі завантажується файл із забороненими адресами, такими як 203.0.113.55 та 13.0.14.75. Коли метод отримує мережевий пакет, він формує його сигнатуру. У цьому випадку перший компонент сигнатури (d_1) містить IP-адресу джерела, наприклад, 198.51.100.25, а другий компонент (d_2) — це адреса отримувача, наприклад, 203.0.113.75. Метод перевіряє ці адреси на наявність серед дозволених і заборонених сигнатур. Оскільки жодна з адрес не потрапляє до відповідних списків, сигнатура пакету класифікується як невизначена. Рішення: трафік з IP-адресами 198.51.100.25 та 203.0.113.75 не відноситься до категорій дозволених або заборонених. Тому метод записує цю сигнатуру до множини невизначених сигнатур D^U для подальшого аналізу. Після цього система повертається до наступного пакету для перевірки.

3.2 Вдосконалення методу класифікації трафіку на основі самоподібності

Метод, який аналізує блоки сигнатур, перевіряючи їх на самоподібність. Якщо блок відповідає критерію самоподібності, то з'єднання дозволяється, а сигнатуру блоку відносяться до множини дозволених сигнатур (D^G). У протилежному випадку пакет записується до множини невизначених сигнатур (D^U) [100]. Алгоритм роботи методу класифікації трафіку на основі самоподібності представлено на рис.3.2.

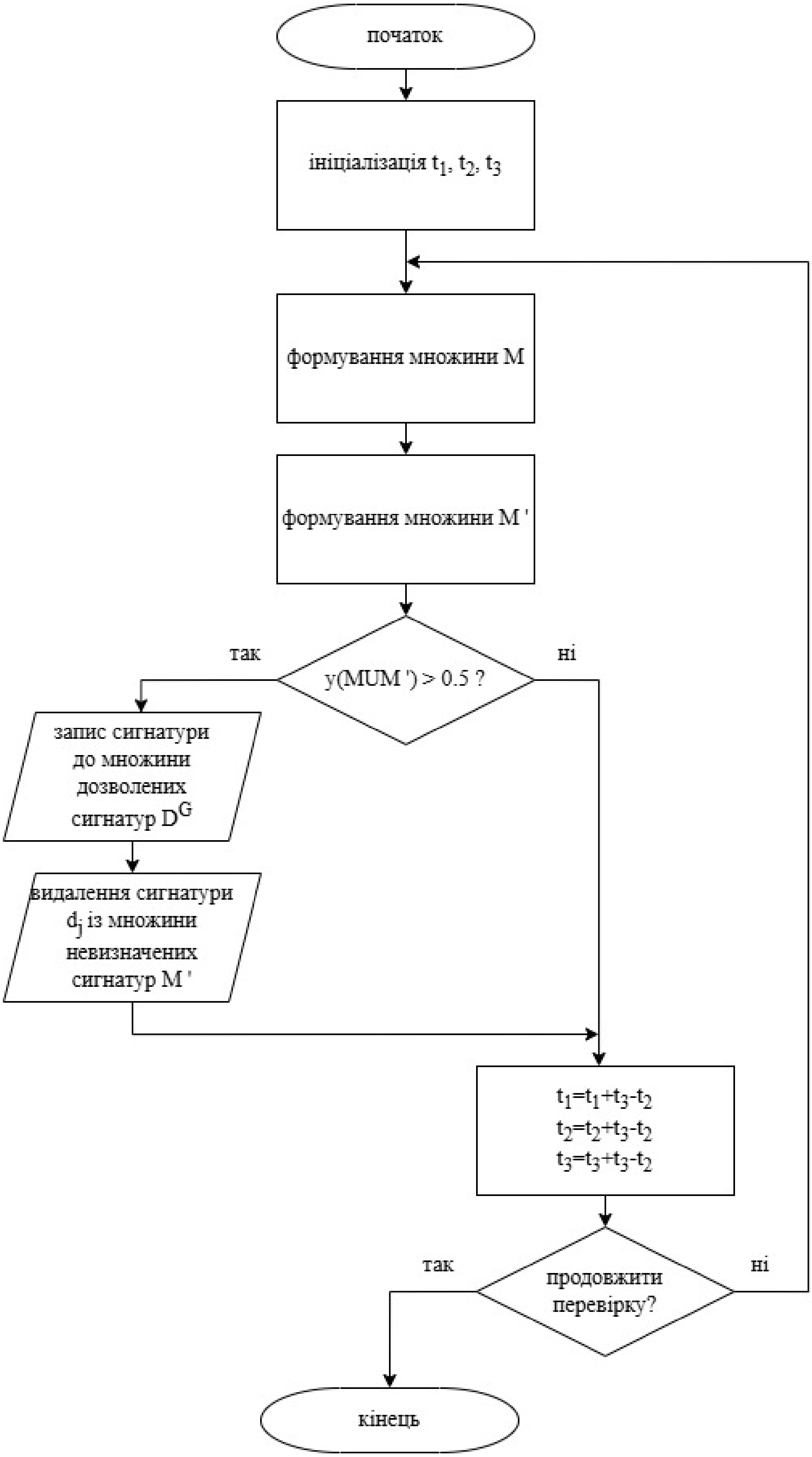


Рисунок 3.2 - Алгоритм роботи методу класифікації трафіку на основі самоподібності

Вхідні дані: множини дозволених (D^G) та невизначених сигнатур (D^U), часові інтервали опорної вибірки та вибірки для перевірки.

Крок 1. З потоку даних виділяються два суміжних послідовних часових інтервали. Перший інтервал містить опорні сигнатури в інтервалі t1-t2. Другий інтервал містить нові сигнатури на інтервалі t2-t3, які ще не піддавалися аналізу на самоподібність.

Крок 2. Формування об’єднаної множини сигнатур. Після визначення часових інтервалів створюється єдина множина, що об’єднує сигнатури з першого інтервалу та нові сигнатури з другого інтервалу. Ця спільна множина використовується для оцінки рівня подібності між сигнатурами.

Крок 3. Покомпонентний аналіз подібності сигнатур. Кожна сигнатура (d_j) представлена як набір характеристик. На цьому етапі здійснюється визначення показника Херста для кожного компонента сигнатури. Це дозволяє виявити, наскільки значення характеристик компонента сигнатури схожі між собою. Після оцінки окремих компонентів проводиться узагальнення отриманих результатів для всього набору. Результатом є загальна числова оцінка, що відображає ступінь самоподібності сигнатур в аналізованому інтервалі, причому більш високі значення свідчать про більшу подібність сигнатур. З отриманою оцінкою проводиться порівняння з встановленим пороговим значенням. Якщо загальний рівень самоподібності перевищує поріг, то це свідчить про те, що сигнатури нового блоку схожа на ті, що вже класифіковані як дозволені. У такому випадку нові сигнатури додаються до множини дозволених (D^G), а сигнатури з невизначеної множини (D^U), які демонструють подібну поведінку, вилучаються Якщо рівень подібності нижчий за поріг, сигнатури залишаються невизначеними і вимагають подальшого аналізу.

Крок 4. Адаптація та оновлення часових вікон. Після класифікації відбувається оновлення часових інтервалів: другий інтервал, що містив нові сигнатури, стає новою базою для наступного циклу аналізу, а попередній інтервал більше не використовується. Такий підхід дозволяє системі оперативно реагувати на зміни в мережевому трафіку, постійно оновлюючи множину дозволених сигнатур.

Крок 5. Перевірка завершення роботи. Якщо прийнято рішення, що потрібно завершити обробку, то метод завершить свою роботу, і наступні пакети не будуть перевірятися. В іншому випадку відбувається перехід до кроку 2.

Результатом роботи методу є визначення приналежності сигнатур із множини невизначених сигнатур (D^U) до множини дозволених сигнатур (D^G). Якщо сигнатуру із множини невизначених сигнатур (D^U) визначено як ту, що можна віднести до множини дозволених сигнатур (D^G), то сигнатура записується до множини дозволених сигнатур (D^G) й видаляється із множини невизначених сигнатур (D^U).

Таким чином, метод класифікації базується на послідовному аналізі нових сигнатур у контексті раніше отриманих даних, що дозволяє оперативно і точно визначати, чи належить новий трафік до дозволених, або він потребує додаткової перевірки.

Розглянемо приклад роботи методу для різних сценаріїв.

Приклад 1. Для роботи методу попередньо сформовано множину сигнатур D^G, що містить 785 сигнатур пакетів нормального трафіку. Визначимо $t_1 =$ 07.07.2024 3:30, $t_2 =$ 07.07.2024 3:40, $t_3 =$ 07.07.2024 3:41. З множини невизначеного трафіку D^U сформуємо множину M (згідно 2.13), що містить невизначені сигнатури в часовому діапазоні $t_1 - t_2$ та сигнатури нормального трафіку. Кількість сигнатур в часовому діапазоні $t_1 - t_2$ склала 856. Загальна потужність множини M склала 1355 сигнатур. Сформуємо множину M' (згідно 2.14), що містить раніше не аналізовані сигнатури з D^U в діапазоні $t_2 - t_3$. Потужність цієї множини склала 73 сигнатури. Загальна множина для аналізу $M \cup M'$ складає 1403 сигнатур.

Для кожного компонента множини $M \cup M'$ визначимо стандартне відхидення $S^j(n)$ (відповідно до 2.16)

Для першого компонента кортежу - IP-адреси джерела середнє значення числового представлення IP-адреси склало $\overline{d^1} = 3124844506{,}25$, при цьому стандартне відхидення $S^1(n) = 420698934{,}17$. Мінімальне накопичене відхилення склало $\underset{i=1,1403}{min} \sum_{m=1}^{k} d_i^1 - \overline{d^1} = -45161982751{,}91$, відповідно максимальне – $\underset{i=1,1403}{max} \sum_{m=1}^{k} d_i^1 - \overline{d^1} = 68793449107{,}67$. Розмах накопиченого відхилення функції згідно 2.17 склав $R^1(n) = 113955431859{,}58$.

Згідно 2.18 для даного прикладу показник Херста для перших компонентів кортежу склав $H^1 = 0{,}77$.

Аналогічно розраховуються показники Херста для інших компонентів кортежа аналізованої множини.

Для розглянутого прикладу отримано такі значення: $H^2 = 0{,}48$, $H^3 = 0{,}53$, $H^4 = 0{,}84$, $H^5 = 0{,}73$, $H^6 = 0{,}43$, $H^7 = 0{,}81$, $H^8 = 0{,}76$, $H^9 = 0{,}46$.

Відповідно до 2.19 визначимо усереднене значення показника Херста для аналізованої множини.

$$H = \frac{0{,}77 + 0{,}48 + 0{,}53 + 0{,}84 + 0{,}73 + 0{,}43 + 0{,}81 + 0{,}76 + 0{,}46}{9} = 0{,}65$$

В результаті аналізу на самоподібність множини $M \cup M'$ згідно 2.20 визначено, що вона є самоподібною, оскільки $0{,}65 > 0{,}5$ відповідно приймається рішення про віднімання множини M' від D^U та її додавання до D^G.

3.3. Розробка нечіткого методу виявлення аномального трафіку

Вхідними даними є: пакети із вихідного потоку даних, множина дозволених сигнатур D^G, множина заборонених сигнатур D^B, множина невизначених сигнатур D^U, перелік правил.

Функції приналежності визначають ступінь приналежності кожного вхідного значення до певних термів. Визначимо лінгвістичну змінну (ЛЗ) ld, що приймає одне із значень з терм-множини $ld \in \{ld^1, ld^2, \dots, ld^m\}$, де m – кількість термів.

Сформуємо ЛЗ, що використовуються для аналізу мережевого трафіку і відповідають компонентам кортежу (2.1): $ld^1, ld^2, \dots, ld^k$. Відповідно до 2.1 $k = 5$.

Слід зазначити, що для моделі процесу нечіткого виявлення аномального трафіку не використовуються перший наведений в кортежі компонент IP-адреса джерела (d^1), третій – порт джерела (d^3), сьомий - час надходження даних на перевірку (d^7) та восьмий - MAC-адреса пристрою (d^8). Оскільки IP-адреса джерела (d^1), яка визначає пристрій, що ініціює з'єднання та передає дані, не є надійним показником для аналізу, оскільки може змінюватися через використання динамічних IP, NAT або проксі-серверів, а також піддаватися спуфінгу. Порт джерела (d^3), який визначає вхідну точку зв'язку з боку відправника, зазвичай генерується операційною системою випадковим чином, що робить його непередбачуваним і малозначущим у контексті нечіткої класифікації аномалій. Час надходження пакету (d^7), виражений у 24-годинному форматі, залежить від багатьох зовнішніх факторів, таких як мережеві затримки, пікові навантаження та зміни у трафіку залежно від часу доби. Це створює значний рівень варіативності, що може ускладнити коректну нечітку оцінку аномальних шаблонів. MAC-адреса пристрою (d^8), який надсилає дані, також не використовується через її локальний характер, оскільки ця адреса не передається через маршрутизовані мережі і може бути легко підроблена або змінена на рівні пристрою. Виключення цих параметрів спрямоване на підвищення точності моделі шляхом усунення змінних, які можуть вносити зайвий шум. Таким чином, для нечіткого аналізу залишаються ті характеристики, що забезпечують стабільність та високу диференційованість між нормальним і потенційно загрозливим трафіком.

Другий наведений в кортежі компонент - IP-адреса призначення **(d^2)**, який можна відображати у вигляді символьної змінної, що приймає одне із значень множини ідентифікаторів $ld^2 \in \{ld_1^2, ld_2^2, ld_3^2\}$, а саме:

ld_1^2=«IP-адреса дозволеного трафіку»,

ld_2^2=«IP-адреса невизначеного трафіку»,

ld_3^2=«IP-адреса аномального трафіку».

Четвертий наведений в кортежі компонент – порт призначення (d^4)**,** який можна відображати у вигляді символьної змінної, що приймає одне із значень множини ідентифікаторів $ld^4 \in \{ld_1^4, ld_2^4, ld_3^4\}$, а саме:

ld_1^4=«порт нормального типового трафіку»,

ld_2^4=«порт невизначеного трафіку»,

ld_3^4=«порт аномального трафіку».

П'ятий наведений в кортежі компонент – протокол (d^5), який можна відображати у вигляді символьної змінної, що приймає одне із значень множини ідентифікаторів $ld^5 \in \{ld_1^5, ld_2^5, ld_3^5\}$, а саме:

ld_1^5=«безпечні протоколи»,

ld_2^5=«нейтральні протоколи»,

ld_3^5=«невизначені протоколи»,

ld_4^5=«підозрілі протоколи»,

ld_5^5=«небезпечні протоколи».

Інтенсивність трафіку в ІКС (d^6) може розглядатися з двох підходів: об'єктивного та суб'єктивного. Об'єктивна інтенсивність трафіку визначається як відношення обсягу переданої інформації через канал зв'язку за визначений період часу до загального доступного ресурсу мережі. Такий підхід базується на вимірюванні та аналізі реальних показників роботи системи, наприклад, кількості пакетів або байтів, що проходять через систему за одиницю часу. Суб'єктивна інтенсивність трафіку відображає оцінку експертів або користувачів системи щодо рівня завантаженості каналів зв'язку в певний момент або упродовж певного періоду. Вона ґрунтується на сприйнятті або оцінках ефективності роботи системи, а також на прогнозах щодо її продуктивності в залежності від поточного чи очікуваного обсягу переданої інформації. У ситуаціях, коли отримання точних статистичних даних ускладнене, експерти використовують логіко-лінгвістичний підхід. У цьому випадку інтенсивність трафіку подається через лінгвістичну змінну (ld^6) із базовою терм-множиною, що дозволяє формалізувати якісні оцінки та забезпечити їхню зручну інтерпретацію для подальшого аналізу:

$$ld^6 = \bigcup_{i=1}^{m} ld_i^6 \quad (2.22)$$

де m – кількість термів, для членів якого справедливе співвідношення порядку $ld_1^6 < ld_2^6 < \cdots < ld_m^6$.

Наприклад, при m=5 для зазначеної ЛЗ можна сформувати множини термів

$ld^6 = \bigcup_{i=1}^{5} ld_i^6 =${«низька (Н)», «нижче середньої (НС)», «середня (С)», «вище середньої (ВС)», «висока (В)»}, що відображається нечіткими числами (НЧ), для яких відповідні функції належності (ФН). Також можуть бути введені й інші значення термів, такі як «дуже низька (ДН)», «дуже висока (ДВ)» та ін.

Дев'ятий наведений в кортежі компонент – розмір пакету (d^9). Такий компонент можна відображати чисельно або через ЛЗ:

$$ld^9 = \bigcup_{i=1}^{y} ld_i^9 \quad (2.23)$$

де y – кількість термів, для членів якого справедливе співвідношення порядку $ld_1^9 < ld_2^9 < \cdots < ld_y^9$.

Наприклад, при y=5 для зазначеної ЛЗ можна сформувати множини термів $ld^9 = \bigcup_{i=1}^{3} ld_i^9 =${«малий», «нижче середнього», «середній», «вище середнього», «великий»}, що відображається НЧ, для яких відповідні функції належності (ФН).

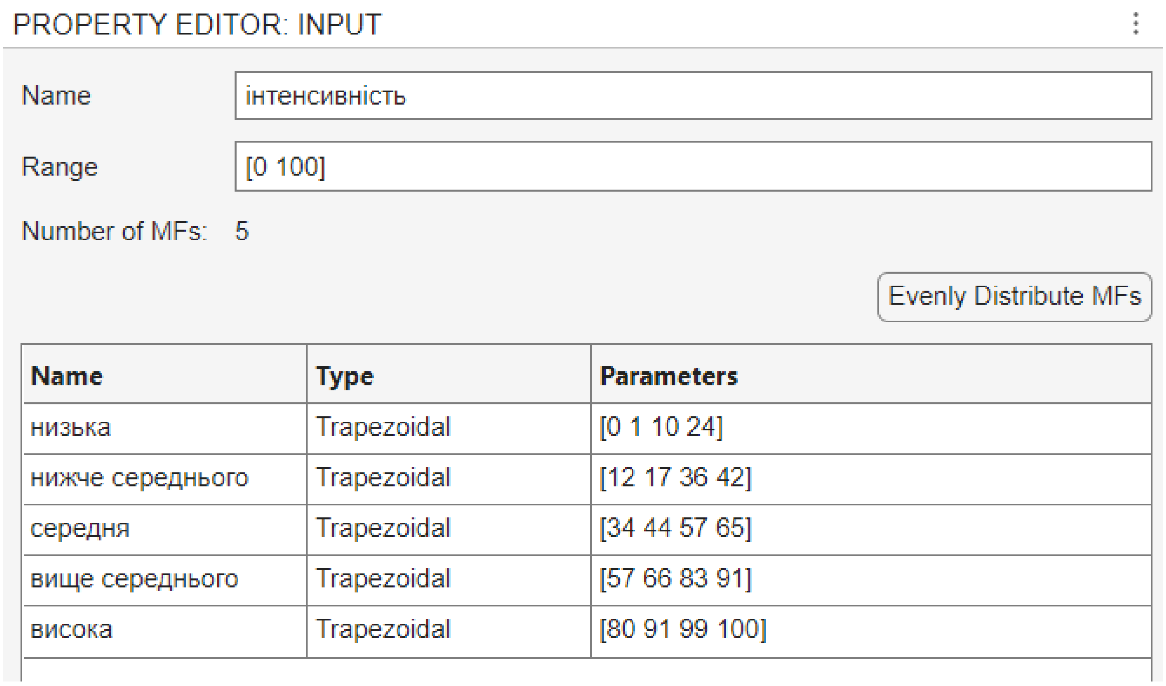
ФН обрана трапецієвидна, оскільки з її допомогою можна моделювати різні стани мережевого трафіку, що дозволяє більш детально аналізувати і класифікувати поведінку трафіку, що важливо для виявлення аномалій та забезпечення безпеки мережі. Параметри для кожної множини значень задано на основі експертних оцінок. Дані для всіх термів ЛЗ наведено у таблиці 3.1.

Оскільки моделювання реалізовано засобами Matlab, то представимо набір термів для однієї із необхідних терм-множин та їх налаштування (рис. 3.3).

Таблиця 3.1 - Дані термів ЛЗ

Терм-множина	Терм	Значення
d^2	Дозволені (ld_1^2)	[0 1 10 20]
	Невизначені (ld_2^2)	[10 23 72 84]
	Заборонені (ld_3^2)	[67 83 99 100]
d^4	Дозволені (ld_1^4)	[0 1 10 36]
	Невизначені (ld_2^4)	[9 24 78 89]
	Заборонені (ld_3^4)	[67 83 99 100]
d^5	Безпечні (ld_1^5)	[0 1 13 19]
	Нейтральні (ld_2^5)	[10 17 32 37]
	Невизначені (ld_3^5)	[31 38 55 61]
	Підозрілі (ld_4^5)	[52 63 81 89]
	Небезпечні (ld_5^5)	[81 89 99 100]
d^6	Низька (ld_1^6)	[0 1 10 24]
	Нижче середнього (ld_2^6)	[12 17 36 42]
	Середня (ld_3^6)	[34 44 57 65]
	Вище середнього (ld_4^6)	[57 66 83 91]
	Висока (ld_5^6)	[80 91 99 100]
d^9	Малий (ld_1^9)	[0 1 10 23]
	Нижче середнього (ld_2^9)	[10 19 31 42]
	Середній (ld_3^9)	[29 41 58 70]
	Вище середнього (ld_4^9)	[60 69 81 89]
	Великий (ld_5^9)	[81 89 99 100]

Наведемо приклад трапецієвидної функції належності для п’яти термів терм-множини d^6.


PROPERTY EDITOR: INPUT

Name: інтенсивність

Range: [0 100]

Number of MFs: 5

Evenly Distribute MFs

Name	Type	Parameters
низька	Trapezoidal	[0 1 10 24]
нижче середнього	Trapezoidal	[12 17 36 42]
середня	Trapezoidal	[34 44 57 65]
вище середнього	Trapezoidal	[57 66 83 91]
висока	Trapezoidal	[80 91 99 100]

Рисунок 3.3 - Набір термів для терм-множини d^6

Трапецієвидна функція належності для ЛЗ ld_1^6 із терм-множини d^6 матиме наступний вигляд:

$$\mu_{ld_1^6}(x) = \{0 \text{ при } x \leq 0; x \text{ при } 0 \leq x \leq 1;\ 1 \text{ при } 1 \leq x \leq 10;$$

$$^{(24-x)}/_{14} \text{ при } 10 \leq x \leq 24; 0 \text{ при } 24 \leq x\} \qquad (2.24)$$

Трапецієвидна функція належності для ЛЗ ld_2^6 із терм-множини d^6 матиме наступний вигляд:

$$\mu_{ld_2^6}(x) = \{0 \text{ при } x \leq 12; ^{(x-12)}/_{5} \text{ при } 12 \leq x \leq 17;$$

$$1 \text{ при } 17 \leq x \leq 36; ^{(42-x)}/_{6} \text{ при } 36 \leq x \leq 42; 0 \text{ при } 42 \leq x\} \qquad (2.25)$$

Трапецієвидна функція належності для ЛЗ ld_3^6 із терм-множини d^6 матиме наступний вигляд:

$$\mu_{ld_3^6}(x) = \{0 \text{ при } x \le 34; {(x-34)}/{10} \text{ при } 34 \le x \le 44;$$

$$1 \text{ при } 44 \le x \le 57; {(65-x)}/{8} \text{ при } 57 \le x \le 65; 0 \text{ при } 65 \le x\} \quad (2.26)$$

Трапецієвидна функція належності для ЛЗ ld_4^6 із терм-множини d^6 матиме наступний вигляд:

$$\mu_{ld_4^6}(x) = \{0 \text{ при } x \le 57; {(x-57)}/{9} \text{ при } 57 \le x \le 66;$$

$$1 \text{ при } 66 \le x \le 83; {(91-x)}/{8} \text{ при } 83 \le x \le 91; 0 \text{ при } 91 \le x\} \quad (2.27)$$

Трапецієвидна функція належності для ЛЗ ld_5^6 із терм-множини d^6 матиме наступний вигляд:

$$\mu_{ld_5^6}(x) = \{0 \text{ при } x \le 80; {(x-80)}/{11} \text{ при } 80 \le x \le 91;$$

$$1 \text{ при } 91 \le x \le 99;\ 100 - x \text{ при } 99 \le x \le 100; 0 \text{ при } 100 \le x\} \quad (2.28)$$

Графічно дані ФН представлено на рис. 3.4.

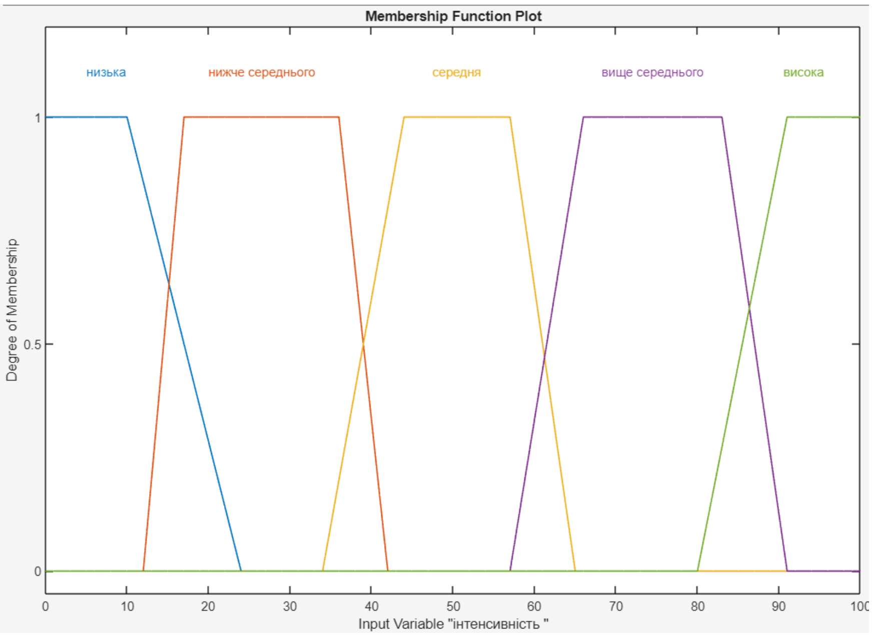


Рисунок 3.4 - ФН для терм-множини d^6

Для усіх нечітких множин, які використовуються, ФН визначається як перетин всіх множин:

$$\mu_{M_{d^2}} \cap \mu_{M_{d^4}} \cap \mu_{M_{d^5}} \cap \mu_{M_{d^6}} \cap \mu_{M_{d^9}} =$$
$$= \left(\mu_{M_{d^2}}(x), \mu_{M_{d^4}}(x), \mu_{M_{d^5}}(x), \mu_{M_{d^6}}(x), \mu_{M_{d^9}}(x)\right), \quad (2.29)$$

де $M_{d^2} = \{m_{d^2}{}_i\}_{i=0}^{4,3\cdot 10^9}$ – множина можливих значень IP-адреси призначення;

$M_{d^4} = \{m_{d^4}{}_i\}_{i=0}^{65535}$ - множина можливих значень портів призначення;

$M_{d^5} = \{m_{d^5}{}_i\}_{i=0}^{36}$ - множина можливих протоколів;

$M_{d^6} = \{m_{d^6}{}_i\}_{i=0}^{10\cdot 10^9}$ - множина можливих значень інтенсивності трафіку;

$M_{d^9} = \{m_{d^9}{}_i\}_{i=0}^{65535}$ - множина можливих значень розміру пакета.

Нечітка база знань містить набір нечітких правил, які визначають взаємозв'язки між вхідними та вихідними змінними. Кожне правило має вигляд "Якщо X, то Y".

Алгоритм роботи нечіткого методу виявлення аномального трафіку представлено на рис. 3.5.

Крок 1. Підключити файл із множиною дозволених сигнатур (D^G).

Крок 2. Підключити файл із множиною заборонених сигнатур (D^B).

Крок 3. Підключити файл із множиною невизначених сигнатур (D^U).

Крок 4. Підключити файл із переліком правил.

Крок 5. Отримання сигнатури пакету.

Крок 6. Фазифікація вхідних значень. Відбувається процес перетворення чітких значень параметрів (наприклад, інтенсивності трафіку, розміру пакета) в нечіткі значення за допомогою ФН. Це дозволяє системі обробляти дані з урахуванням невизначеності. Кожен параметр пакету отримує ступінь належності до певної категорії.

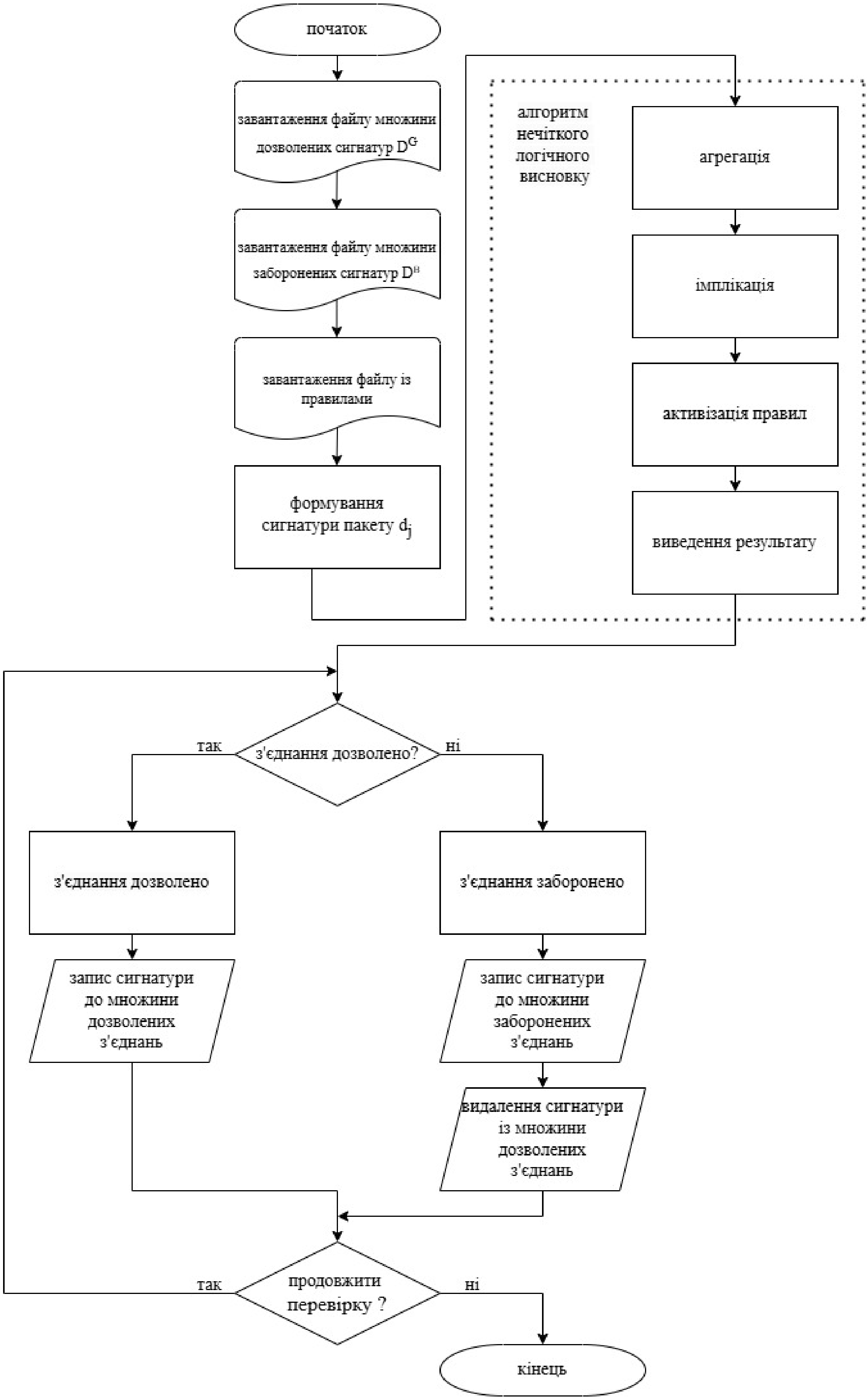


Рисунок 3.5 - Алгоритм роботи нечіткого методу виявлення аномального трафіку

Крок 7. Агрегація нечітких правил. Здійснюється об'єднання результатів застосування кількох нечітких правил для одного пакету. Кожне правило оцінює певний аспект трафіку (наприклад, IP-адресу, порт, інтенсивність трафіку). Після того, як правила оцінюють пакет, результати кожного з параметрів об'єднуються для визначення загального ступеня належності пакету до певної категорії (дозволений або заборонений).

Крок 8. Імплікація. Виконується перетворення нечіткої інформації в конкретні висновки на основі агрегації результатів. Для кожного пакету система приймає рішення про його класифікацію.

Крок 9. Активізація правил. Система оцінює ступінь належності кожного параметра пакету до відповідних термів у нечітких правилах. Якщо пакет відповідає всім вимогам правила, воно активується. Якщо хоча б одна умова не виконується, то правило не активується. Це дає змогу кожному правилу мати різну вагу в загальному рішенні.

Крок 10. Призначення рішення. Після застосування нечітких правил та активізації відповідних умов, система приймає рішення. Якщо більшість правил визначили пакет як дозволений (тобто всі параметри пакету відповідають дозволеному діапазону), то відбувається перехід на крок 11. Інакше відбувається перехід на крок 13.

Крок 11. Дозвіл з’єднання.

Крок 12. Запис сигнатури із множини невизначених до множини дозволених. Перехід до кроку 15.

Крок 13. Заборона з’єднання.

Крок 14. Запис сигнатури із множини невизначених до множини заборонених.

Крок 15. Якщо потрібно продовжити перевірку сигнатур, то відбувається перехід на крок 5. Інакше перехід на крок 16.

Крок 16. Завершення роботи.

Результатом роботи методу є: дозвіл з’єднань які відповідно до правил вважаються дозволеними; блокування з’єднань які відповідно до правил вважаються забороненими; додавання сигнатур до множини дозволених сигнатур D^G або множин заборонених сигнатур D^B [101].

Для формування висновку щодо наявності аномального трафіку, здійснюється класифікація елементів сформованої сигнатури засобами нечіткого логічного висновку, як показано на рис. 3.6.

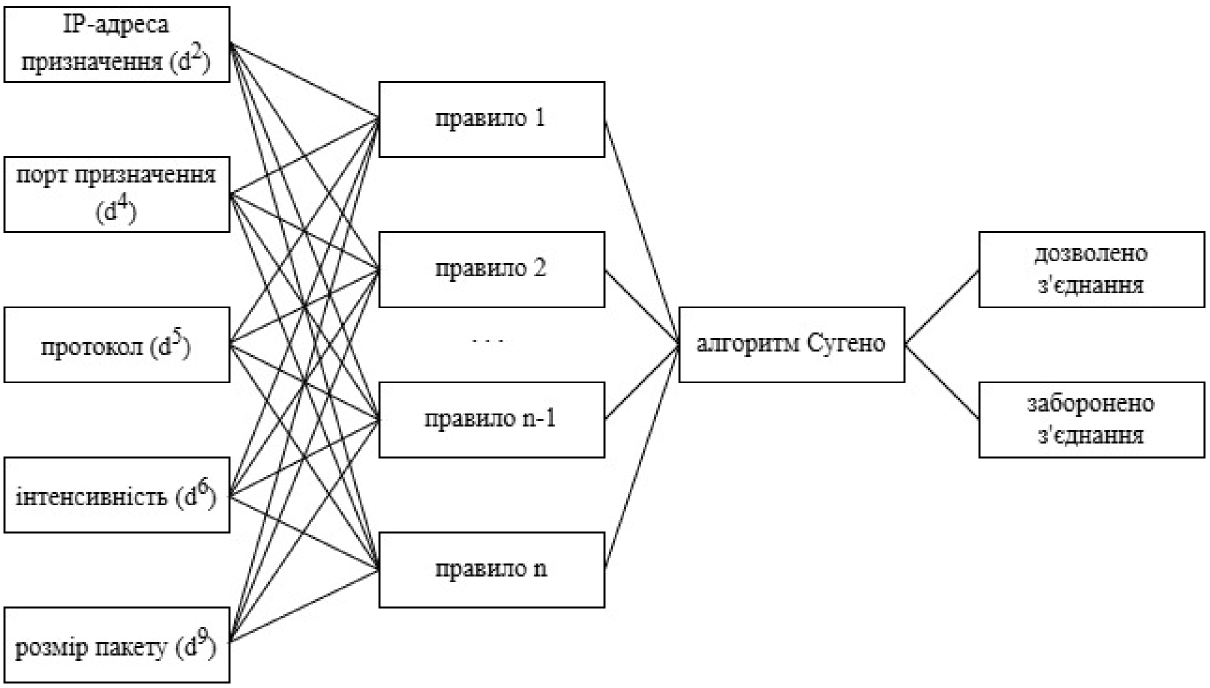


Рисунок 3.6 – Схема нечіткого логічного висновку для визначення приналежності сигнатури трафіку до аномального трафіку

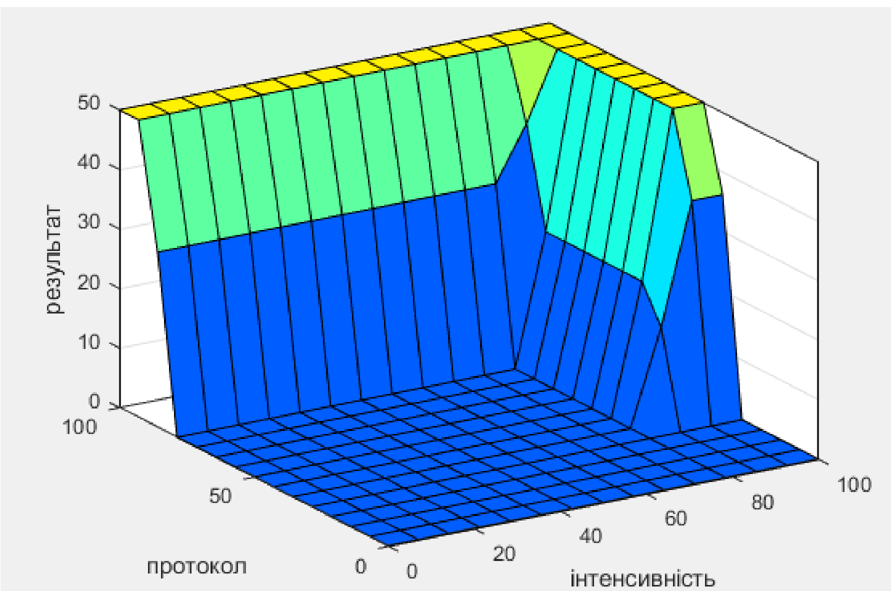


Рисунок 3.7 - Результати роботи у вигляді поверхні відгуку

Наведемо кілька прикладів.

Якщо IP-адреса призначення заборонена, порт дозволений, протокол нейтральний, інтенсивність трафіку низька, розмір пакету малий, то отримано результат заборонено з’єднання.

$$if\ \left((d^2\ is\ ld_3^2)\wedge(d^4\ is\ ld_1^4)\wedge\left(d^5\ is\ ld_2^5\right)\wedge(d^6\ is\ ld_1^6)\wedge(d^9\ is\ ld_1^9)\right)$$
$$\Rightarrow\ (result\ is\ \text{заборонено})$$

Якщо IP-адреса призначення є забороненою, порт заборонений, протокол небезпечний, інтенсивність трафіку нижче середнього, а розмір пакету малий, то отримано результат заборонено з’єднання.

$$if\ \left((d^2\ is\ ld_3^2)\wedge(d^4\ is\ ld_3^4)\wedge\left(d^5\ is\ ld_5^5\right)\wedge(d^6\ is\ ld_2^6)\wedge(d^9\ is\ ld_1^9)\right)$$
$$\Rightarrow\ (result\ is\ \text{заборонено})$$

Якщо IP-адреса призначення невизначена, порт не визначений, протокол безпечний, інтенсивність трафіку нижче середнього, а розмір пакету малий, то результат дозволено.

$$if\ \left((d^2\ is\ ld_2^2)\wedge(d^4\ is\ ld_2^4)\wedge\left(d^5\ is\ ld_1^5\right)\wedge(d^6\ is\ ld_2^6)\wedge(d^9\ is\ ld_1^9)\right)$$
$$\Rightarrow\ (result\ is\ \text{дозволено})$$

На рис. 3.8 представлено формування одного із правил засобами моделювання, а фрагмент сформованого набору правил показано на рис. 2.23.

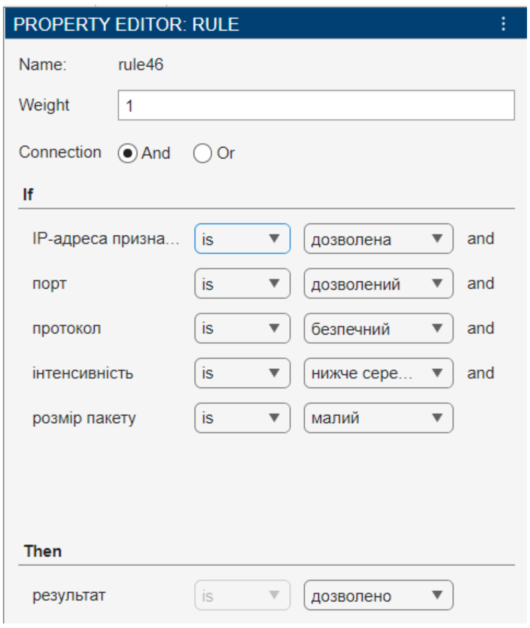


Рисунок 3.8 – Формування правила засобами моделювання

	Rule	Weight	Name
53	If IP-адреса призначення is невизначена and порт is заборонений and протокол is безпечн...	1	rule53
54	If IP-адреса призначення is заборонена and порт is заборонений and протокол is безпечни...	1	rule54
55	If IP-адреса призначення is дозволена and порт is дозволений and протокол is нейтральни...	1	rule55
56	If IP-адреса призначення is невизначена and порт is дозволений and протокол is нейтраль...	1	rule56
57	If IP-адреса призначення is заборонена and порт is дозволений and протокол is нейтральн...	1	rule57
58	If IP-адреса призначення is дозволена and порт is невизначений and протокол is нейтраль...	1	rule58
59	If IP-адреса призначення is невизначена and порт is невизначений and протокол is нейтра...	1	rule59
60	If IP-адреса призначення is заборонена and порт is невизначений and протокол is нейтрал...	1	rule60
61	If IP-адреса призначення is дозволена and порт is заборонений and протокол is нейтральн...	1	rule61
62	If IP-адреса призначення is невизначена and порт is заборонений and протокол is нейтрал...	1	rule62

Рисунок 3.9 – Фрагмент набору правил

Приклад 1. Пакет, який надходить на перевірку, має такі параметри. IP-адреса призначення — 10.0.0.5, це адреса одержувача пакету. Пакет має порт порт призначення 80, що вказує на HTTP-порт. Для цього пакету обраний протокол TCP, а інтенсивність трафіку становить 100 пакунків на секунду. Розмір пакету становить 512 байт. Для кожного з параметрів пакету застосовуються функції належності, що переводять чіткі значення в нечіткі категорії. IP-адреса призначення 10.0.0.5 має ступінь належності 5, оскільки належить до дозволених адрес. Для порту призначення 80 ступінь належності 3, оскільки це стандартний порт для HTTP. Протокол TCP має ступінь належності 2, оскільки він дозволений для HTTP/HTTPS з'єднань. Інтенсивність трафіку становить 100 пакетів на секунду, що класифікується як середня інтенсивність, з ступенем належності 45. Розмір пакету становить 512 байт, що класифікується як середній розмір пакету, зі ступенем належності 47. Далі відбувається агрегація ступені належності для кожного параметра пакету. Усі параметри мають високий ступінь належності до дозволених значень. Це означає, що пакет відповідає умовам дозволеного трафіку, оскільки більшість параметрів знаходяться на рівні ступеня належності близькому до 0. Наступним кроком здійснюється імплікація на основі агрегації результатів. Згідно з правилами, якщо всі параметри пакету належать до категорії дозволеного трафіку (дозволені IP-адреси, дозволені порти, нормальна інтенсивність трафіку, середній розмір пакету, робочий час), з'єднання класифікується як дозволене. Оскільки всі перевірки виконано успішно, з'єднання класифікується як дозволене. Далі правило, яке перевіряє відповідність пакету дозволеним умовам, активується, оскільки всі параметри пакету належать до дозволених категорій. Ступінь належності для кожного параметра вказує на те, що правило застосовне до цього пакету. Після того, як правило активовано, система приймає рішення. Оскільки пакет відповідає всім вимогам дозволеного трафіку, з'єднання дозволяється. Перевірка припиняється, оскільки з'єднання було класифіковано як дозволене. Подальша перевірка цього пакету не потрібна. У даному випадку інші правила не обробляються, оскільки з'єднання вже було класифіковано як дозволене. Пакет, який був класифіковано як дозволене з'єднання, додається до бази даних дозволених з'єднань для подальшого використання. Це дозволить прискорити перевірку подібних з'єднань у майбутньому. Перевірка для цього пакету завершена. Система готова перейти до перевірки наступного пакету.

Приклад 2. На перевірку надходить сигнатура із наступними параметрами. IP-адреса призначення – 192.168.50.200, що не входить до списку дозволених або невизначених адрес. Вона не використовується в локальній мережі та не належить до жодного відомого сервісу. Порт призначення, який використовується у сигнатурі, – 4444. Він не є стандартним для звичайних мережевих сервісів і часто застосовується у бекдорах, троянах або для створення зловмисних тунелів. Обраний протокол – UDP. Його використання для цього типу з'єднання є нетиповим, оскільки більшість легітимних сервісів, пов’язаних із веб-трафіком або критично важливими комунікаціями, працюють через TCP. Інтенсивність трафіку для сигнатури становить 10 000 пакетів на секунду, що суттєво перевищує середньостатистичні значення для стандартного мережевого обміну даними. Розмір пакета дорівнює 64 байти. Після надходження пакета вхідні параметри проходять фазифікацію, під час якої кожен із них отримує певний ступінь належності. IP-адреса отримує ступінь належності 67, оскільки вона може бути пов'язана з потенційною загрозою. Порт 4444 має значення належності 85, що вказує на його використання у зловмисних схемах. Протокол UDP отримує ступінь належності 71 через високий ризик його застосування у

DDoS-атаках. Інтенсивність трафіку оцінюється зі значенням 98, оскільки вона є надмірною для нормального мережевого функціонування. Розмір пакета отримує ступінь належності 2. На основі отриманих значень відбувається агрегація результатів. Більшість параметрів демонструє низьку ступінь належності до дозволених значень, що вказує на високий рівень ризику. Після цього здійснюється імплікація на основі агрегації даних. Система застосовує правила класифікації, які вказують, що за умови високої інтенсивності трафіку, використання забороненого порту та невідомої IP-адреси сигнатуру має бути віднесено до аномального трафіку. Оскільки всі зазначені умови виконуються, активується правило блокування з'єднання. Система ухвалює рішення про класифікацію пакета як аномального. Його передача блокується, а отримані параметри зберігаються у множині заборонених сигнатур. Таким чином, з'єднання перервано, а пакет не був допущений у мережу, оскільки його сигнатура класифікована як аномальна.

3.4. Розробка гібридного методу виявлення аномального трафіку в інформаційно-комунікаційних системах

Задля об'єднання методів, що описано в параграфах 3.1-3.3, розроблено гібридний метод виявлення аномального трафіку в ІКС.

Вхідними даними є: пакети із вихідного потоку даних, множина вихідних сигнатур (D), множина дозволених сигнатур D^G, множина заборонених сигнатур D^B, множина невизначених сигнатур D^U, часові інтервали опорної вибірки та вибірки для перевірки, перелік правил.

Етапи роботи методу наступні.

Крок 1. Підключити файл із множиною дозволених сигнатур (D^G).

Крок 2. Підключити файл із множиною заборонених сигнатур (D^B).

Крок 3. Підключити файл із множиною невизначених сигнатур (D^U).

Крок 4. Підключити файл із переліком правил.

Крок 5. Формування сигнатури пакету d_j із вихідного потоку даних (D), яку потрібно перевірити.

Крок 6. Виконання методу класифікації трафіку за ознаками. Якщо сигнатуру класифіковано, то відбувається перехід на крок 7. Інакше відбувається перехід на крок 9.

Крок 7. Якщо потрібно продовжити перевірку сигнатур, то відбувається перехід на крок 5. Інакше перехід на крок 8.

Крок 8. Завершення роботи.

Крок 9. Виконання методу класифікації трафіку на основі самоподібності. Якщо сигнатуру класифіковано, то відбувається перехід на крок 7. Інакше відбувається перехід на крок 10.

Крок 10. Виконання нечіткого методу виявлення аномального трафіку. Перехід на крок 7.

Результатом роботи є однозначна класифікація сигнатур як дозволених чи аномальних. Алгоритм роботи гібридного методу виявлення аномалій в ІКС представлено на рис.3.10. На рис.3.11 представлено функціональну схему гібридного методу виявлення аномального трафіку в ІКС.

Множина вихідних даних (D) є джерелом вхідної інформації для методу класифікації трафіку за ознаками (1). Також вхідними даними для цього методу є множина заборонених сигнатур $D^B(2)$ та множина дозволених сигнатур $D^G(3)$. Результатом роботи методу є множина невизначених сигнатур $D^U(4)$. Вхідними даними для методу класифікації трафіку на основі самоподібності є множина дозволених сигнатур $D^G(8)$ та множина невизначених сигнатур $D^U(9)$. В результаті роботи методу сигнатури записуються до множини дозволених сигнатур $D^G(7)$ та видаляються із множини невизначених сигнатур $D^U(10)$ або ж множини залишаються незмінними. Вхідними даними для нечіткого методу виявлення аномального трафіку є множина невизначених сигнатур $D^U(11)$ та набір правил (12). У результаті метод сигнатури із множини невизначених сигнатур D^U класифікує та перезаписує до множина заборонених сигнатур $D^B(5)$ або множини дозволених сигнатур $D^G(6)$, видаляючи їх із множини невизначених сигнатур $D^U(13)$.

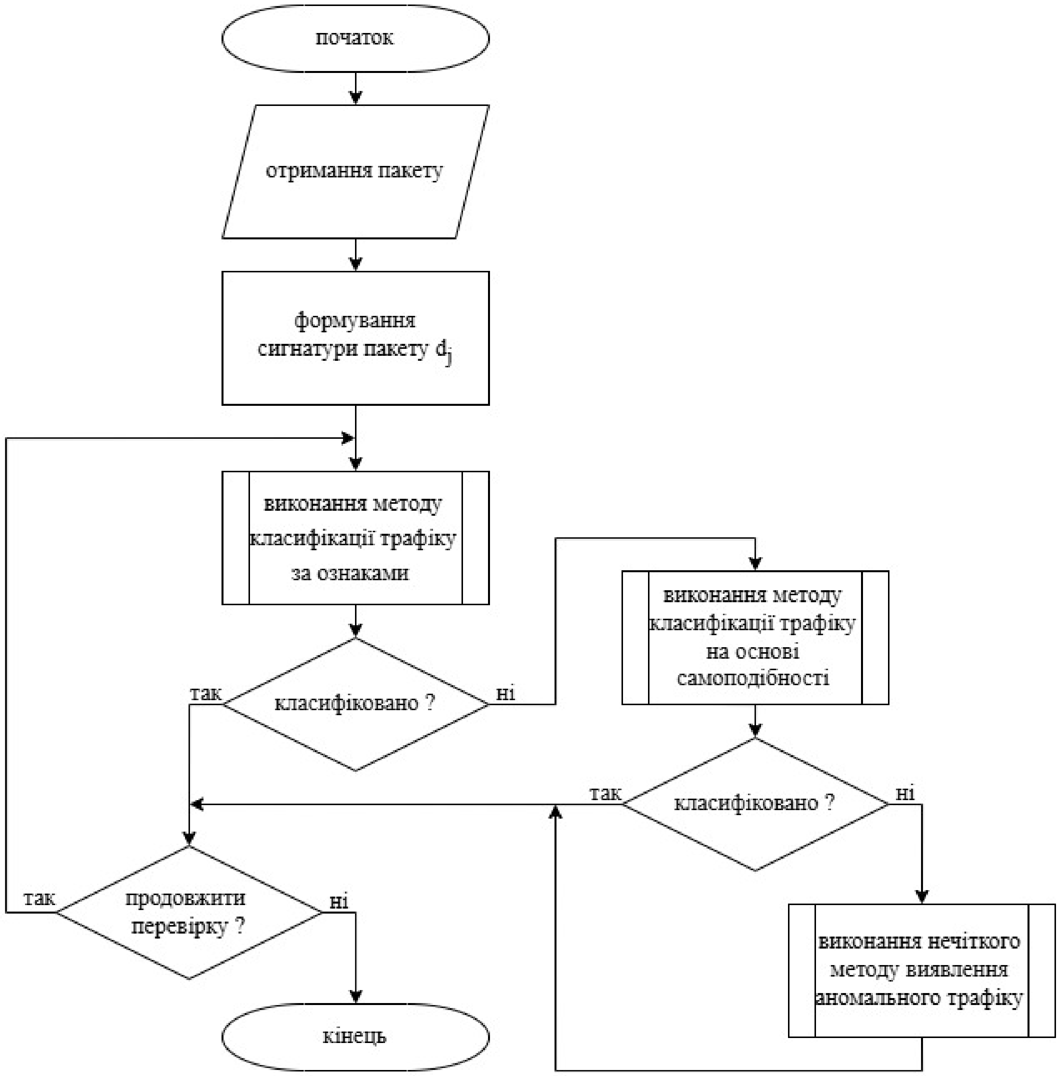


Рисунок 3.10 - Алгоритм роботи гібридного методу виявлення аномалій в ІКС

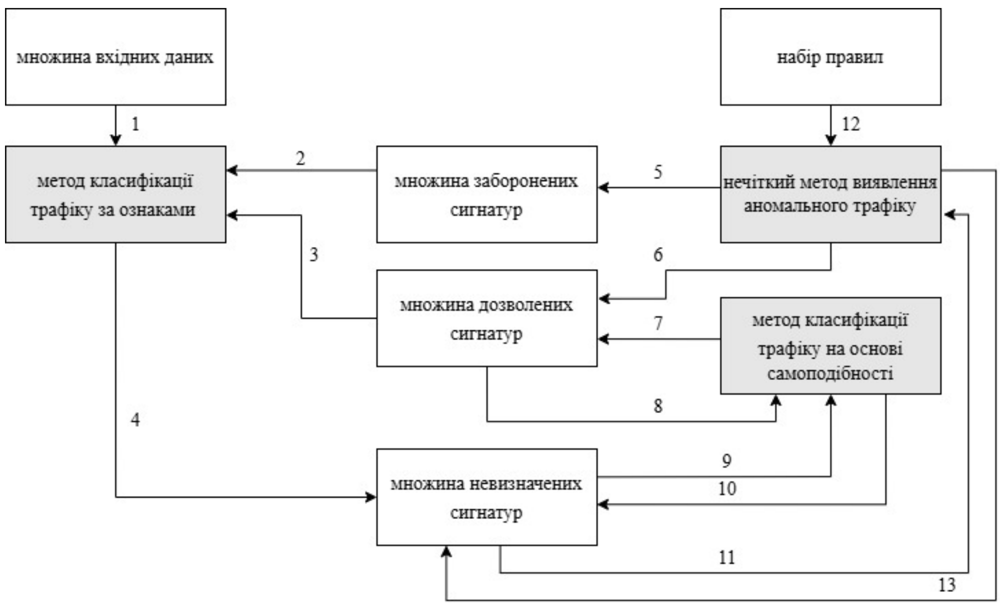


Рисунок 3.11 - Функціональна схема гібридного методу виявлення аномального трафіку в ІКС

Приклад 1. На перевірку надходить сигнатура із IP-адресою 156.44.1.78. Оскільки дана адреса є в множині заборонених, то на етапі роботи методу класифікації трафіку за ознаками з'єднання буде забороненим. Перехід на другий та третій етап перевірки не відбувається.

Приклад 2. Сигнатура <55.7.88.4, 157, TCP, 897, 385> надходить на перевірку. На етапі роботи методу класифікації трафіку за ознаками сигнатуру класифіковано як ту що не належить до множини дозволених чи заборонених, тому записано до множини невизначених сигнатур. На етапі роботи методу класифікації трафіку на основі самоподібності сигнатуру класифіковано як дозволену, тому з'єднання дозволено та сигнатуру записано до множини дозволених. Перехід на третій етап перевірки не відбувається.

Приклад 3. Сигнатура <167.57.8.14, 22, FTP, 248, 1764> надходить на перевірку. На етапі роботи методу класифікації трафіку за ознаками сигнатуру класифіковано як ту що не належить до множини дозволених чи заборонених, тому записано до множини невизначених сигнатур. На етапі роботи методу класифікації трафіку на основі самоподібності сигнатуру не класифіковано як дозволену, тому вона залишається у множині невизначених сигнатур. На етапі роботи нечіткого методу виявлення аномального трафіку відбувається перевірка сигнатури. Оскільки сигнатуру класифіковано як аномальну, то з'єднання блокується, а сигнатуру записано до множини заборонених та видалено із множини невизначених сигнатур.

Висновки за розділом 3

У третьому розділі дисертаційної роботи було здійснено безпосередню розробку, формалізацію та вдосконалення методів виявлення аномального трафіку в ІКС.Було удосконалено метод класифікації трафіку за ознаками, який є одним із базових підходів до аналізу мережевої активності в ІКС. Вдосконалено метод класифікації трафіку за ознаками, який використовує перевірку IP-адрес на відповідність дозволеним чи забороненим множинам. Вдосконалений алгоритм дозволяє реагувати на аномалії, блокуючи заборонений трафік, та забезпечує точне віднесення до категорії невизначеного трафіку для подальшого аналізу.

Вдосконалено метод класифікації трафіку на основі самоподібності, що дозволяє враховувати часову зміну мережевого потоку та адаптувати систему до нових даних. Це підтримує динамічне оновлення списку дозволених сигнатур, дозволяючи адаптувати систему до змін у мережі та забезпечуючи коректне визначення типових безпечних з’єднань навіть при зміні умов роботи мережі.

Розроблено нечіткий метод виявлення аномального трафіку в ІКС, що використовує систему правил нечіткого виведення з ЛЗ, що описують рівень загрози. Для кожної ЛЗ було обрано відповідну кількість термів на основі експертних оцінок, а також побудовано функції належності, які дозволяють враховувати невизначеність та нечіткість вхідної інформації. Перевагою методу є його здатність інтегрувати численні параметри в єдиний числовий показник ризику. Метод дозволяє ефективно працювати в умовах неповних або суперечливих даних, що робить його надзвичайно корисним для виявлення складних багатовекторних атак.

Розроблено гібридний метод виявлення аномального трафіку в ІКС який який інтегрує всі попередньо описані підходи: класифікацію за ознаками, класифікацію на основі самоподібності та нечіткий аналіз. Така інтеграція дозволила об’єднати сильні сторони кожного з методів і компенсувати їхні окремі недоліки. Гібридний метод передбачає початкове фільтрування трафіку на основі сигнатур, після чого дані, що не підпадають під встановлені шаблони, передаються для подальшого аналізу за допомогою методу самоподібності. У випадках, коли результат класифікації є невизначеним або близьким до порогових значень, до процесу долучається нечітка система, яка формує фінальну оцінку загрози. Це дозволяє істотно зменшити кількість хибнопозитивних спрацювань і підвищити стійкість системи до нових та складно детектованих атак.

РОЗДІЛ 4
РОЗРОБКА СТРУКТУРНОЇ МОДЕЛІ СИСТЕМИ ВИЯВЛЕННЯ АНОМАЛЬНОГО ТРАФІКУ В ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ СИСТЕМАХ ТА ОЦІНКА ДОСТОВІРНОСТІ

4.1 Архітектура системи виявлення аномального трафіку в ІКС на основі гібридного методу

Snort є системою виявлення та запобігання вторгнень, яка здійснює аналіз мережевого трафіку в реальному часі та здатна виявляти аномальні або зловмисні пакети. Його архітектура включає кілька компонентів, зокрема захоплення трафіку, декодування пакетів, попередню обробку даних, механізм виявлення атак, а також систему логування та виводу сповіщень. Підключення Snort до мережі може здійснюватися в кількох режимах, зокрема через мережевий міст, використовуючи функціональність порту SPAN на комутаторі або в режимі онлайн для запобігання атакам. У першому випадку Snort працює як пасивний монітор, який лише аналізує трафік, тоді як у режимі онлайн він активно втручається у процес передачі даних, блокуючи підозрілі пакети в режимі реального часу. Для налаштування Snort необхідно встановити відповідне програмне забезпечення та відредагувати конфігураційний файл, де вказується діапазон внутрішньої та зовнішньої мережі, а також підключаються файли з сигнатурами атак. Додатково можна інтегрувати сторонні бібліотеки для аналізу трафіку, що значно розширює можливості Snort. Процес роботи Snort починається із захоплення трафіку за допомогою бібліотеки libpcap, що дозволяє перехоплювати пакети на мережевому інтерфейсі. Далі здійснюється декодування заголовків мережевих протоколів для визначення характеристик пакетів та їх відповідності відомим стандартам. Після цього попередній аналіз дозволяє нормалізувати трафік та підготувати його для подальшої перевірки, що знижує ймовірність ухилення атакуючими від виявлення. Основний етап виявлення загроз відбувається за допомогою механізму правил, який перевіряє пакети на відповідність відомим сигнатурам атак, а також аналізує їхню поведінку для виявлення аномалій. Якщо виявлено загрозу, Snort формує сповіщення або записує відповідний запис у журнали, а у разі роботи в режимі IPS може заблокувати трафік, що не відповідає безпечним критеріям. Таким чином, Snort є багатофункціональною системою для аналізу мережевого трафіку, яка може бути використана як у класичних інфраструктурах, так і в хмарних середовищах. Завдяки можливості підключення сторонніх бібліотек та інтеграції з іншими системами безпеки Snort забезпечує високий рівень контролю за трафіком, що дозволяє своєчасно виявляти та запобігати атакам.

Архітектура Snort складається з чотирьох основних компонентів: сніфера пакетів, препроцесора, механізму виявлення загроз та модуля виводу. Потік трафіку між цими складовими представлений на рисунку 4.1.

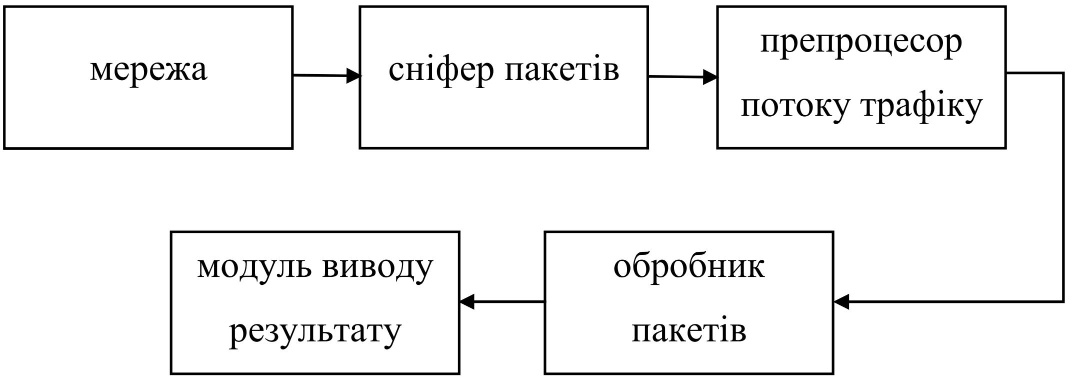


Рисунок 4.1 – Послідовність проходження трафіку в Snort

Snort 3 підтримує багатопотокову обробку трафіку в межах одного процесу, що значно підвищує ефективність роботи. Він аналізує трафік, що надходить на мережевий інтерфейс. Якщо мережевий інтерфейс працює в режимі перехоплення (promiscuous mode), Snort отримує доступ до всіх пакетів незалежно від їх адресата. Препроцесори відіграють важливу роль у попередній обробці трафіку перед тим, як він потрапить до механізму виявлення загроз. Послідовність роботи наведено на рис. 4.2. До прикладу, препроцесор frag3 здійснює дефрагментацію IP-пакетів, що допомагає ідентифікувати атаки, які були розділені на фрагменти для обходу систем захисту. Дефрагментація може бути як методом приховування атак, так і природним явищем у мережах.

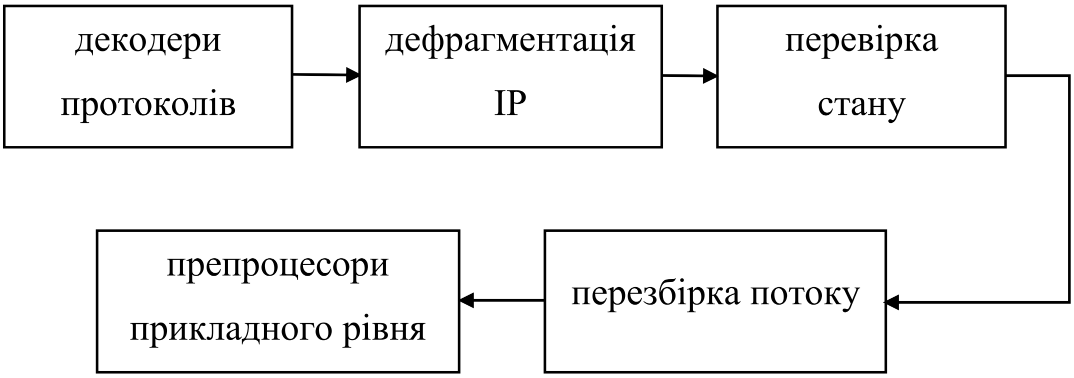


Рисунок 4.2 – Послідовність роботи препроцесора

Інтеграція власної бібліотеки до Snort 3 відбувається шляхом створення окремого модуля, адже Snort 3 використовує модульну архітектуру. Спочатку розроблено модуль, що використовує API Snort 3, для якого визначено функції ініціалізації, обробки вхідного мережевого трафіку та прийняття рішень на основі даних, що надходять із власної бібліотеки. Під час старту Snort модуль завантажується через конфігураційний файл, в якому задаються параметри ініціалізації, включаючи шлях до бібліотеки та необхідні налаштування логування й обробки помилок. У модулі відбувається виклик функції з власної бібліотеки безпосередньо під час обробки пакетів, що надходять у систему. Це дозволяє аналізувати трафік в режимі реального часу, приймати рішення щодо генерування сповіщень та застосування політик безпеки. Схему підключення гібридного методу виявлення аномального трафіку в ІКС до Snort наведено на рис. 4.3.

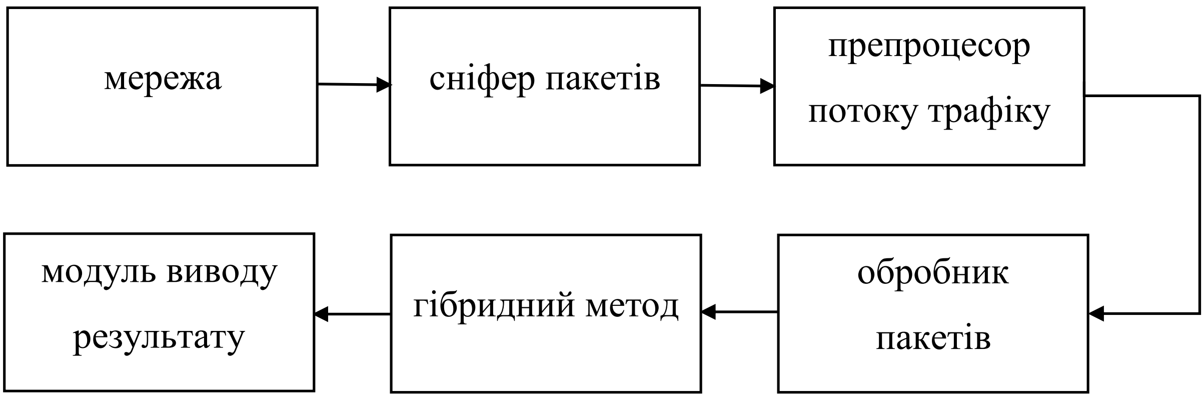


Рисунок 4.3 - Алгоритм роботи Snort із використанням гібридного методу виявлення аномального трафіку в ІКС

4.2 Структурна модель системи виявлення аномального трафіку в інформаційно-комунікаційних системах

Запропонована структурна модель системи виявлення аномального трафіку в ІКС (рис. 4.4) реалізує повний цикл обробки мережевого трафіку – від захоплення сирого трафіку до прийняття рішення щодо безпечності з’єднання. На першому етапі спеціалізований модуль захоплення трафіку інтегрується з мережевим середовищем і здійснює попередню обробку, що включає фільтрацію трафіку за ознаками, нормалізацію форматів даних і агрегацію пакетів у сесії, формуючи підґрунтя для подальшого аналізу.

Отримані дані передаються до модуля класифікації трафіку за ознаками, який виконує ідентифікацію сигнатур згідно з множинами дозволених і заборонених з'єднань. База сигнатур динамічно оновлюється в режимі реального часу, що дозволяє виявляти як відомі загрози, так і модифіковані або нові варіанти атак. Одночасно з цим задіюється метод класифікації трафіку на основі самоподібності, що порівнює поточні характеристики з моделями типової поведінки користувачів та зловмисників, виявляючи аномальні відхилення, зокрема атаки нульового дня.

Головним елементом моделі є нечіткий метод виявлення аномального трафіку, який використовує ЛЗ та ФН для формування інтегрованого ризикового показника. Система оперує базою правил НЛ, які дозволяють формувати узагальнену оцінку рівня загрози з урахуванням таких параметрів, як швидкість передавання, кількість спроб з’єднання, відповідність сигнатурам тощо.

Підсистема логічного виведення об’єднує результати модулів класифікації, самоподібності й нечіткої логіки, забезпечуючи комплексний механізм прийняття рішень. У разі виявлення невизначених або суперечливих ситуацій, система може скеровувати потік на додаткову перевірку, а при підвищеному рівні загрози – ініціювати взаємодію з іншими компонентами кіберзахисту, зокрема міжмережевими екранами та системами запобігання вторгненням.

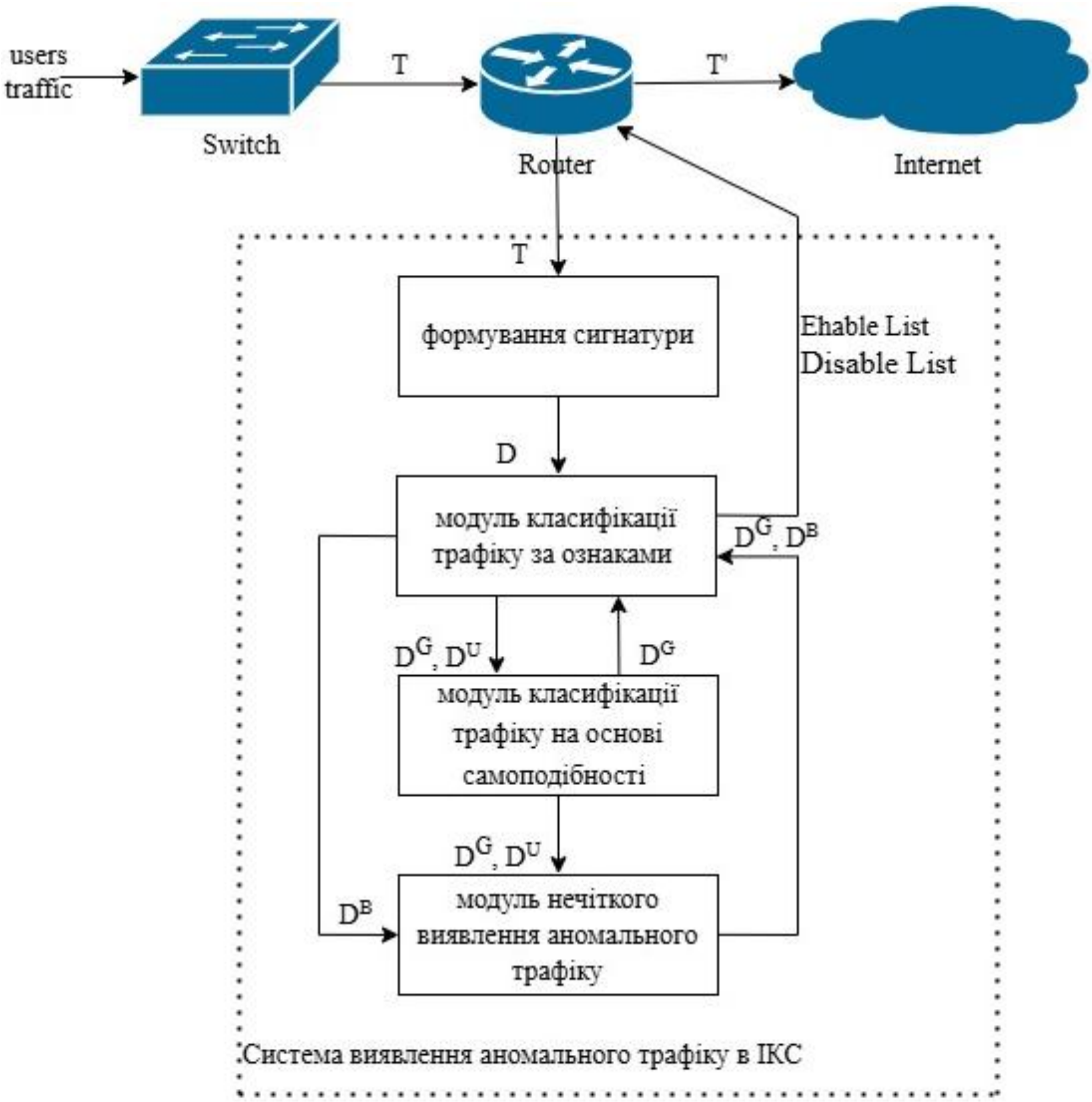


Рисунок 4.4 - Структурна модель системи виявлення аномального трафіку в ІКС

Всі події, що фіксуються у процесі обробки трафіку, записуються в централізовану систему логування, що включає IP-адреси, порти, часові мітки, типи протоколів. Ці дані використовуються не лише для аудиту та інцидент-аналізу, а й для зворотного зв’язку, що дозволяє уточнювати правила класифікації й оновлювати моделі самоподібності.

Запропонована модель, базуючись на гібридному методі виявлення аномалій (поєднання сигнатурного аналізу, аналізу самоподібності та НЛ), характеризується високим ступенем адаптивності до змін у трафіку ІКС, зниженням хибнопозитивних спрацювань і оптимізацією обчислювальних ресурсів. Особлива увага приділена аналізу вихідного трафіку, як окремому вектору загроз, з метою виявлення внутрішніх атак або витоку даних.

Тестова реалізація моделі підтвердила її ефективність у реальному мережевому середовищі, з можливістю подальшої інтеграції з іншими системами та зовнішніми індикаторами компрометації, що підвищує оперативність реагування на загрози.

Таким чином, структурна модель системи виявлення аномального трафіку в ІКС, реалізована на основі гібридного методу, забезпечує надійний рівень кіберзахисту, мінімізує ризики НСД і адаптується до умов високої динамічності мережевого середовища.

4.3 Опис тестового середовища для проведення дослідження

Для формування власного набору даних використовувалося локальне захищене середовище, на яке не впливали сторонні ресурси. Для формування типового дозволеного трафіку було запущено набір типових програм, що генерують дозволений трафік, який описано в параграфі 2.1. Для формування аномального трафіку було імітовано поведінку порушника, яку описано в параграфі 2.2, шляхом реалізації активної розвідки та здійснення імітації атак з одного пристрою на інший в мереж заданої мережі. Дані було зібрано за допомогою програмного продукту Snort [102]. Після завершення формування вибірки кожний елемент набору було позначено як дозволений чи аномальний. Загалом у створеному наборі даних було 5468657 записів, із яких 5160985 дозволених та 307672 класифікованих як заборонені. Фрагмент набору даних показано у таблиці 4.1, де тип трафіку 1 означає дозволений типовий трафік, а 2 – аномальний трафік.

Для тестування гібридного методу, який описано в 3.4, розроблено власну бібліотеку, яку інтегровано до Snort. Таким чином сніфер працює на одному пристрої у мережі та прослуховує весь вихідний трафік. Привівши вхідні дані у потрібний формат вони аналізуються з використанням підключеної бібліотеки.

Таблиця 4.1 – Фрагмент створеного набору даних

Дата і час	IP-адреса відправника	MAC відправника	Порт відправника	Протокол	IP-адреса отримувача	MAC отримувача	Порт отримувача	Розмір пакету, біт	Тип трафіку
1	2	3	4	5	6	7	8	9	10
Feb/14/2025 13:04:38	192.168.0.35	6A:BE:29:41:C3:F8	46836	443	104.18.4.62	F4:7A:21:8D:3E:66	443	572	1
Feb/14/2025 13:04:19	192.168.0.33	7F:92:41:DB:6C:30	64699	443	104.18.4.62	F4:7A:21:8D:3E:66	443	983	1
Feb/14/2025 13:03:28	192.168.0.203	B8:17:64:22:AF:3D	53900	443	157.240.224.174	3A:7F:20:BD:99:EE	443	1384	1
Feb/14/2025 13:03:24	192.168.0.202	98:2D:76:4B:9A:C3	58528	443	157.240.224.174	3A:7F:20:BD:99:EE	443	813	1
Feb/14/2025 13:03:19	192.168.0.180	3a:c8:38:d1:76:22	61064	443	149.154.167.99	9C:57:4D:AA:EE:62	443	599	1
Feb/14/2025 13:03:15	192.168.0.179	74:4d:28:46:bb:c0	54149	443	149.154.167.99	9C:57:4D:AA:EE:62	443	1567	1

Кінець таблиці 4.1

1	2	3	4	5	6	7	8	9	10
Feb/14/2025 13:45:16	192.168.2.11	00:15:22:02:AA:BB	443	80	232.23.116.133	00:15:22:02:CC:DD	443	1500	2
Feb/14/2025 13:45:17	192.168.1.13	00:14:22:04:AA:BB	21	21	218.12.6.133	00:14:22:04:CC:DD	21	1024	2
Feb/14/2025 13:45:17	192.168.0.53	42:40:88:67:89:66	8080	80	50.21.226.80	00:40:88:90:12:77	80	1100	2
Feb/14/2025 13:45:19	192.168.0.39	89:4F:30:77:AB:EC	22	22	104.21.73.81	A3:7B:15:CC:80:54	22	512	2
Feb/14/2025 13:45:19	192.168.0.37	89:4F:30:77:AB:EC	22	22	104.21.73.81	A3:7B:15:CC:80:54	22	512	2

Фрагмент коду, який показано на рис. 4.5, написаний на C++ і є частиною бібліотеки, що була розроблена для тестування гібридного методу, описаного у розділі 3.4 дослідження. Метод передбачає обчислення показника Херста – метрики, що застосовується для виявлення довготривалої залежності у часових рядах, зокрема мережевого трафіку. Бібліотека інтегрується до системи виявлення вторгнень Snort і дозволяє локальному сніферу аналізувати весь вихідний трафік пристрою у реальному часі. У наведеному коді реалізовано функцію realHerst, яка приймає вектор дійсних чисел і виконує статистичний аналіз, включаючи обчислення середнього значення, стандартного відхилення, мінімального та максимального накопиченого відхилення, а також остаточне визначення показника Херста.

```
double realHerst(const std::vector<double>& list) {
    if (list.empty()) return 0.0;

    double sum = std::accumulate(list.begin(), list.end(), 0.0);
    double mean = sum / list.size();

    double variance = 0.0;
    for (double v : list) {
        variance += std::pow(v - mean, 2);
    }
    variance /= list.size();
    double stdDeviation = std::sqrt(variance);

    auto maxIt = std::max_element(list.begin(), list.end(),
        [mean](double a, double b) { return (a - mean) < (b - mean); });
    auto minIt = std::min_element(list.begin(), list.end(),
        [mean](double a, double b) { return (a - mean) < (b - mean); });

    double maxv = *maxIt - mean;
    double minv = *minIt - mean;
    double r = maxv - minv;

    double herst = std::log(r / stdDeviation) / std::log(list.size());

    std::cout << "середнє - " << mean << std::endl;
    std::cout << "стандартне - " << variance << std::endl;
    std::cout << "max - " << maxv << std::endl;
    std::cout << "min - " << minv << std::endl;
    std::cout << "Розмах - " << r << std::endl;
    std::cout << "Херст - " << herst << std::endl;

    return herst;
}
```

Рисунок 4.5 – Фрагменто коду реалізації бібліотеки

Результати, які представлено на рис. 4.6, виводяться на екран для наочного представлення. Зокрема, згідно з прикладом виводу, значення показника Херста становить 0.77 — це може свідчити про наявність довготривалих залежностей у вхідному трафіку.

```
Середнє значення: 3124844506,25
Стандартне відхилення: 420698934,17
Мінімальне накопичене відхилення: -45161982751,91
Максимальне накопичене відхилення: 68793449107,67
Розмах накопиченого відхилення: 113955431859,58
Міра Херста: 0,77
```

Рисунок 4.6 – Приклад результату визначення показника Херста

Варто зазначити, що реалізація даного аналізу в основному проєкті не є доцільною. Це пояснюється тим, що основною метою є отримання результату та реагування, а не включення повноцінного графічного представлення в код снiфера.

4.4 Оцінка достовірності гібридного методу виявлення аномального трафіку в інформаційно-комунікаційних системах

Загальні вимоги до ефективності вирішення задачі, яку розв'язує система виявлення аномалій, можна виразити через такі відомі та часто використовувані метрики якості: TP (True Positive) – кількість подій, які правильно класифіковані як аномальні та дійсно є такими; FP (False Positive) – кількість подій, що були помилково визначені як аномальні, хоча насправді такими не є; TN (True Negative) – кількість подій, що не були класифіковані як аномальні та дійсно не є такими (тобто є нормальними); FN (False Negative) – кількість подій, які не були визначені як аномальні, хоча насправді вони є аномальними [103-104]. FP часто називають помилками першого роду, а FN – помилками другого роду.

Після налаштування тестового середовища запущено сформований набір даних як вхідна інформація для класифікації трафіку із використанням Snort та без використання даного ПЗ. Метрики якості показано в таблиці 4.2.

Таблиця 4.2 ілюструє результати тестування трьох систем — Snort із гібридною бібліотекою, оригінального Snort і Suricata — на одному й тому ж штучно сформованому наборі з 5468657 записів (з яких 307672 аномальні й 5160985 нормальні). Для Snort із підключеною гібридною бібліотекою кількість правильно виявлених аномалій (TP) склала 5141567, пропущених аномалій (FN) — 19418, хибних спрацювань на нормальний трафік (FP) — 28739 і коректно класифікованих нормальних подій (TN) — 278933. У випадку оригінального Snort ці показники становили 5027861 (TP), 133124 (FN), 46235 (FP) і 261437 (TN), а в Suricata — 5064128 (TP), 96857 (FN), 38138 (FP) і 269534 (TN). Порівняння демонструє, що гібридний метод значно зменшує кількість пропущених аномалій і хибних спрацювань у порівнянні з обома традиційними рішеннями.

Таблиця 4.2 - Тестування із використанням набору даних

набір даних	Кількість аномальних записів	Кількість дозволених записів	Всього записів	TP	TN	FP	FN
Snort із розробленою бібліотекою	307 672	5 160 985	5 468 657	5 141 567	278 933	28 739	19 418
Snort				5 027 861	261 437	46 235	133 124
Suricata				5 064 128	269 534	38 138	96 857

Для наступного етапу тестування розгорнуто дві ізольовані локальні мережі, які взаємодіють між собою (рис.4.7). Із мережі 1 (ІКС) буде відбуватися контрольований запуск типового трафіку та імітація атак із пристроїв мережі 1 на пристрої мережі 2 (Інтернет-ресурси). Результати тестування системи виявлення аномального трафіку в ІКС наведено в таблиці 4.3.

У таблиці 4.3 наведено метрики для реального мережевого середовища: контрольованої взаємодії двох ізольованих підмереж із загальною кількістю 1487888 записів (181139 аномальних і 1306749 нормальних). Snort із розробленою бібліотекою в цьому випадку досяг 1284963 правильно виявлених аномалій і пропустив 21786 випадків, тоді як хибні спрацювання становили лише 17514. Оригінальний Snort показав 1265976 TP, 40773 FN і 24215 FP, а Suricata — 1271295 TP, 35454 FN і 22767 FP. Ці дані підтверджують стійкість гібридного методу до шуму реального трафіку.

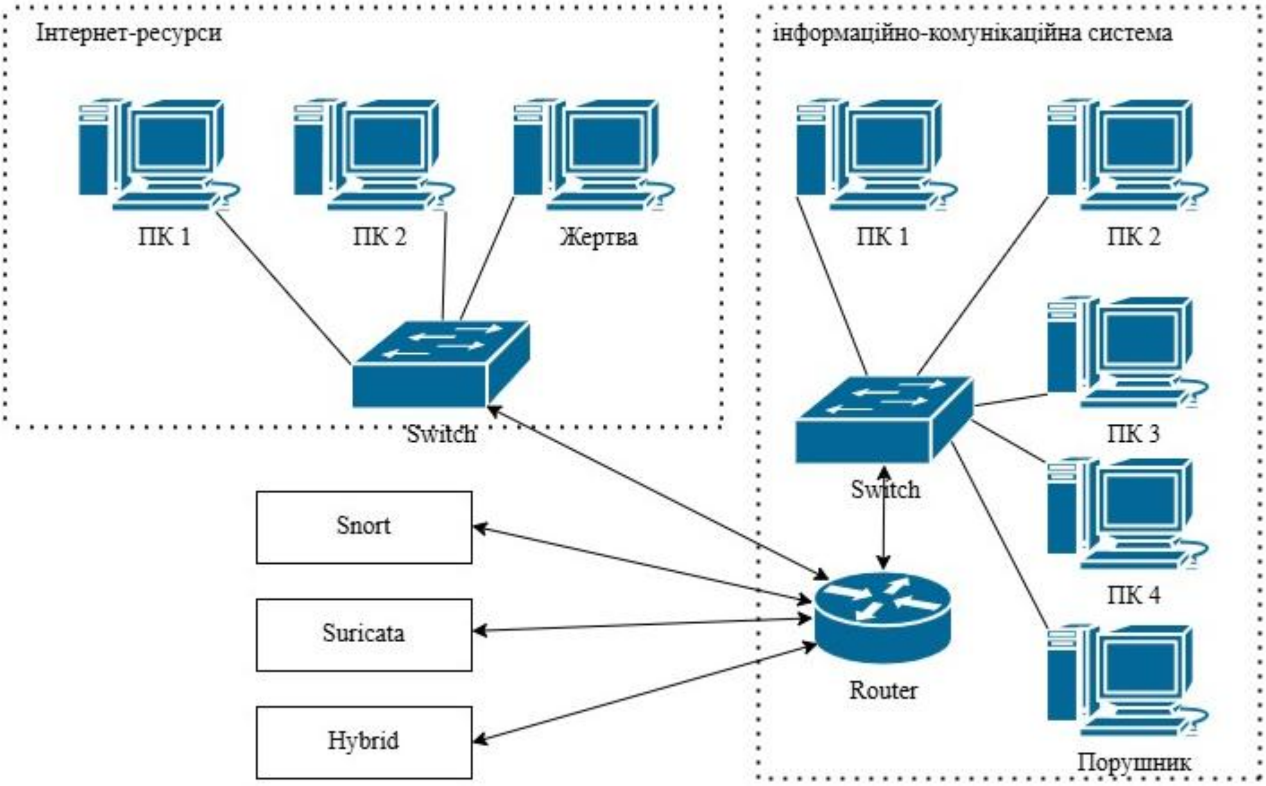


Рисунок 4.7– Схема мережі для тестування

Таблиця 4.3 – Метрики якості при використанні реального трафіку

реальний трафік	Кількість аномальних записів	Кількість дозволених записів	Всього записів	TP	TN	FP	FN
Snort із розробленою бібліотекою	181 139	1 306 749	1487 888	1 284 963	163 625	17 514	21 786
Snort				1 265 976	156 924	24 215	40 773
Suricata				1 271 295	158 372	22 767	35 454

На основі зазначених метрик можна обчислити наступні характеристики:

- повнота (recall) – співвідношення правильно класифікованих позитивних зразків до загальної кількості позитивних зразків:

$$Recall = \frac{TP}{TP+FN} \tag{4.1}$$

- точність (precision) – частка правильно визначених аномальних подій серед усіх подій, які система визначила як аномальні:

$$Precision = \frac{TP}{TP+FP} \tag{4.2}$$

- акуратність (accuracy) – частка правильно виявлених та правильно не виявлених подій серед усіх подій:

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{4.3}$$

- помилка (specificity) – вимірює здатність системи виявлення аномалій правильно ідентифікувати типові об'єкти або події як нормальні:

$$Specificity = \frac{FP+FN}{TP+FP+TN+FN} \tag{4.4}$$

- F1-міра (F1-score) – середнє значення між повнотою (Recall) та точністю (Precision):

$$F1\ score = \frac{Recall+Precision}{2} \tag{4.5}$$

Ці метрики дозволяють оцінити достовірність роботи системи виявлення та аналізувати її продуктивність у різних умовах. Достовірність при використанні набору даних представлено у таблицях 4.4 та 4.5

Таблиця 4.4 – Характеристики оцінки достовірності при використанні набору даних

Тип трафіку	метрика / система	Acccuracy	Precision	Specificity	recall	F1-score
набір даних	Snort із розробленою бібліотекою	0,9912	0,9944	0,9066	0,9962	0,9953
	Snort	0,9672	0,9909	0,8497	0,9742	0,9825
	Suricata	0,9753	0,9925	0,8760	0,9812	0,9868

У таблиці 4.4 представлено розраховані на основі даних із табл. 4.2 ключові показники якості роботи систем у лабораторних умовах. Загальна точність (Accuracy) гібридного модуля досягла 0,9912, точність виявлення аномалій (Precision) становила 0,9944, здатність правильно класифікувати нормальний трафік (Specificity) — 0,9066, повнота виявлення (Recall) — 0,9962, а F1-міра — 0,9953. У порівнянні з цим оригінальний Snort мав Accuracy 0,9672, Precision 0,9909, Specificity 0,8497, Recall 0,9742 і F1 0,9825, тоді як Suricata відповідно показала 0,9753, 0,9925, 0,8760, 0,9812 і 0,9868. Ці значення свідчать про перевагу гібридного методу в усіх п’яти метриках. Графічно представлено на рис .4.8.

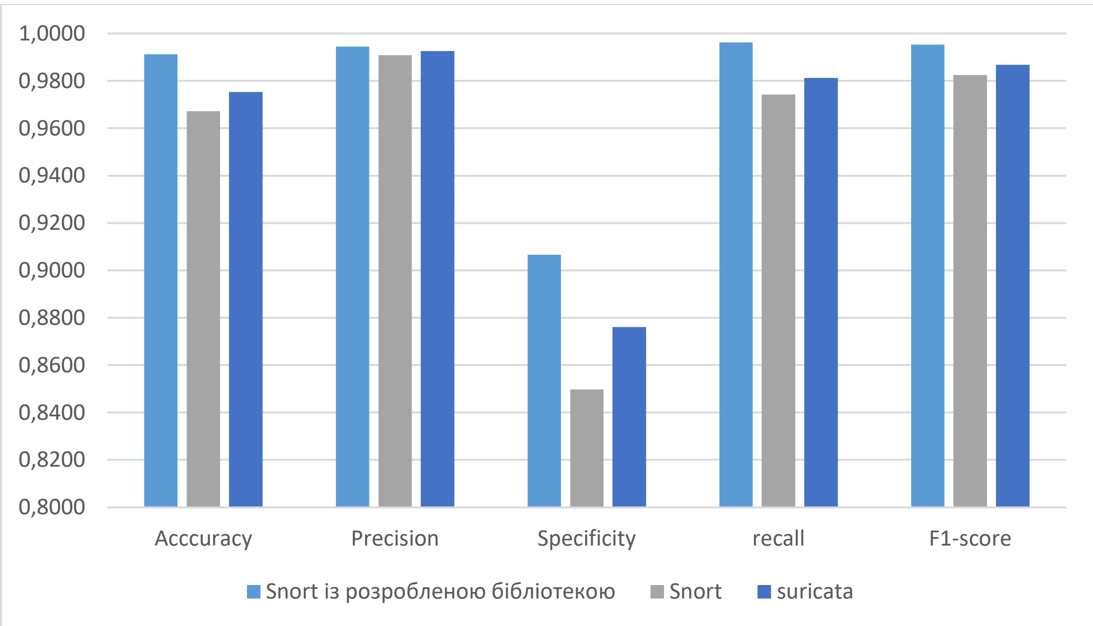


Рисунок 4.8 – Діаграма характеристик оцінки достовірності при використанні набору даних

Таблиця 4.5 – Характеристики оцінки достовірності в режимі реального часу

Тип трафіку	метрика / система	Acccuracy	Precision	Specificity	recall	F1-score
Реальна мережа	Snort із розробленою бібліотекою	0,9736	0,9866	0,9033	0,9833	0,9849
	Snort	0,9563	0,9812	0,8663	0,9688	0,9750
	Suricata	0,9609	0,9824	0,8743	0,9729	0,9776

Таблиця 4.5 містить розрахунки для умов реального часу на основі таблиці 4.3. Графічно представлено на рис .4.9.

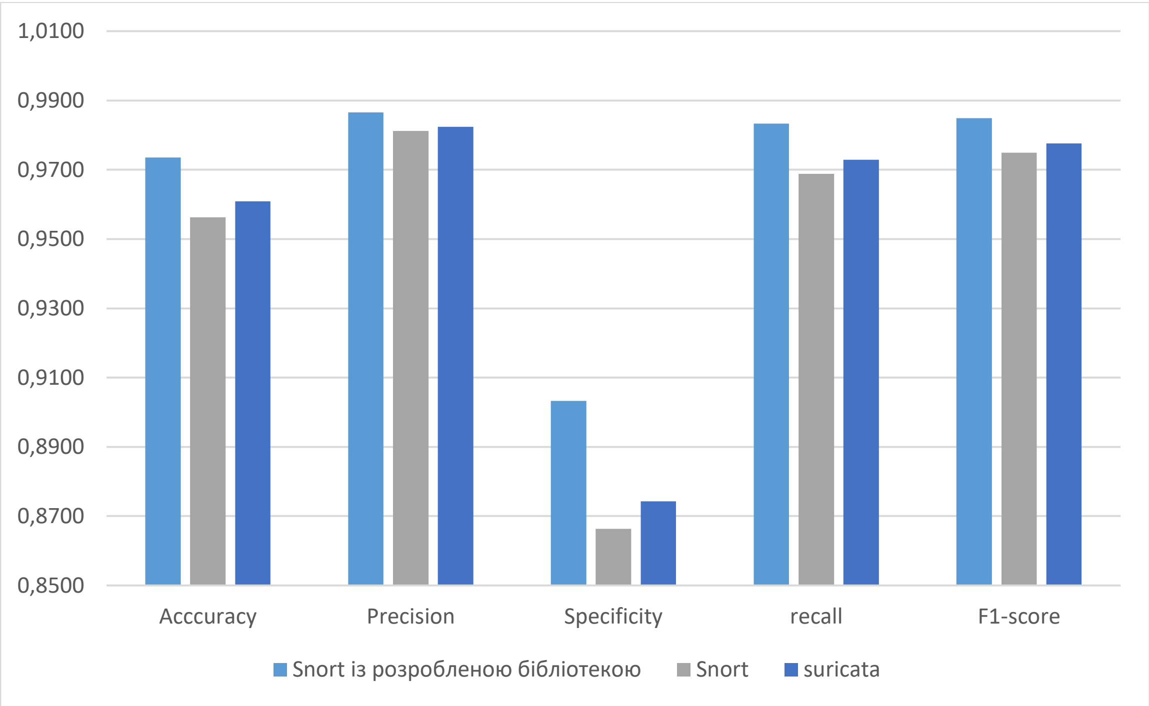


Рисунок 4.9 – Діаграма характеристик оцінки достовірності в режимі реального часу

Для Snort із розробленою бібліотекою Accuracy становила 0,9736, Precision 0,9866, Specificity 0,9033, Recall 0,9833 і F1-score 0,9849. Оригінальний Snort у цьому середовищі досяг Accuracy 0,9563, Precision 0,9812, Specificity 0,8663, Recall 0,9688 і F1 0,9750, а Suricata — 0,9609, 0,9824, 0,8743, 0,9729 і 0,9776 відповідно. Незначне зниження показників у порівнянні з лабораторними тестами пояснюється підвищеним рівнем фонових шумів та варіативністю трафіку, проте гібридний модуль усе одно залишається найефективнішим рішенням у реальних мережевих умовах.

4.5 Вплив гібридного методу виявлення аномального трафіку в інформаційно-комунікаційних системах на інформаційно-комунікаційну систему

Для оцінки впливу запропонованого методу було зібрано тестову систему в складі сегменту мережі, ПК, що використовуються в якості платформи для аналізу та жертв атак, маршрутизатора, де вихідний порт зеркалювався на сервери з запущеною системою аналізу трафіку.

Перевірка роботи систем відбувалась під навантаженням штатним трафіком (~200Мб/с) з послідовним виконанням експериментів.

Перший експеримент полягав в вивантаженні на зовнішні ресурси великих за обсягом файлів з швидкістю, що дорівнювала пропускній швидкості каналу (1Гб/с), тривалість експерименту склала 1,5с.

Під час другого експерименту проводилась атака на зовнішній ресурс типу «підбір паролю» трафік ід час атаки склав 0,5Гб/с, тривалість експерименту склала 1с.

Під час третього експерименту проводилась атака на зовнішній ресурс типу «сканування портів» трафік ід час атаки склав 0,7Гб/с, тривалість експерименту склала 1,3с.

Під час четвертого експерименту проводилась розсилка повідомлень в месенджери з низькою інтенсивністю, час атаки склав 0,3Гб/с, тривалість експерименту склала 2,6с.

Часова діаграма трафіку експерименту представлена на рис. 4.10.

При аналізі трафіку з використанням SIEM системи Snort були отримані результати, представлені на рис. 4.11, при чому завантаження процесора системи, на якій виконувався Snort представлена на рис 4.12.

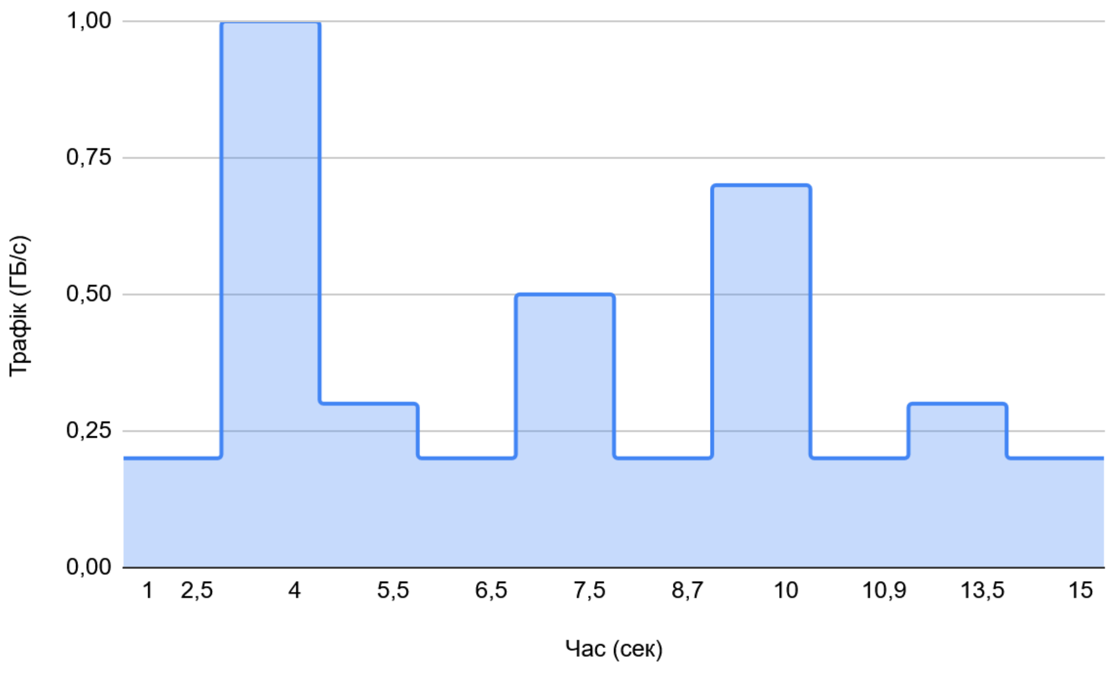


Рисунок 4.10 - Вихідний трафік в системі без аналізу

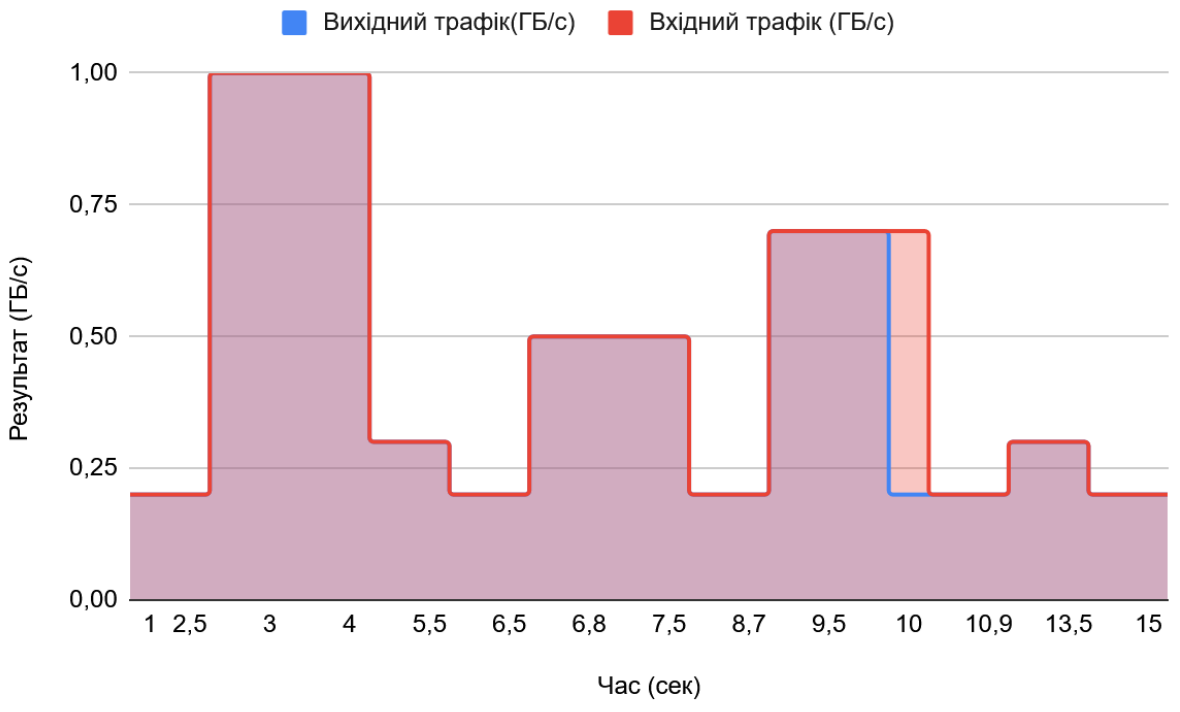


Рисунок 4.11 - Вихідний трафік з використанням Snort

Формула (4.6) дозволяє розрахувати середнє завантаження процесора під час роботи системи аналізу трафіку у рамках проведення експерименту. Вона враховує змінне навантаження впродовж різних часових інтервалів:

$$\overline{CPU} = \frac{\sum_{i=1}^{n-1} CPU_i * (t_{i+1} - t_i)}{\sum_{i=1}^{n-1} (t_{i+1} - t_i)}, \quad (4.6)$$

де CPU_i - завантаження процесора на часовому інтервалі;

$\overline{CPU}$ - середнє значення завантаження процесора за обраний проміжок часу. Воно не є простою арифметичною середньою, а зваженою середньою, яка враховує тривалість кожного інтервалу;

t_i та t_{i+1} - послідовні моменти часу, в які проводились вимірювання;

$t_{i+1} - t_i$ - це тривалість інтервалу, протягом якого фіксувалося певне навантаження.

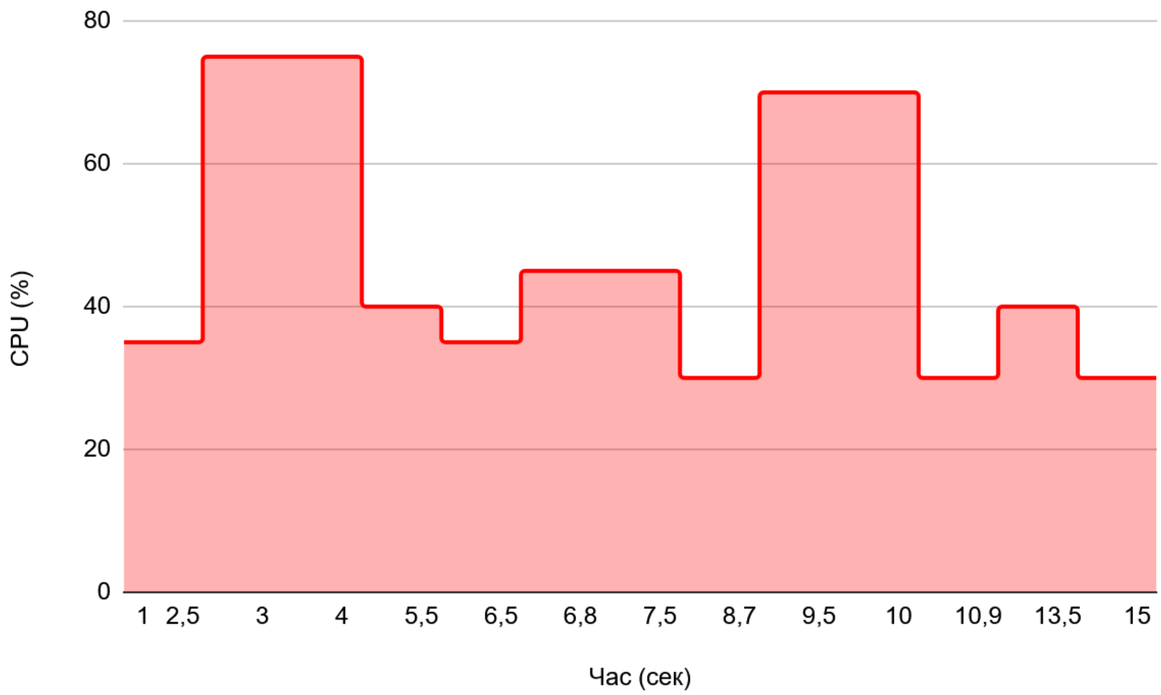


Рисунок 4.12 - Завантаження процесора з використанням Snort

Як видно з рис. 4.11 та 4.12 використання Snort не забезпечило виявлення двох з трьох атак, при чому навіть при виявленні атаки навантаження на процесор не зменшилось. Середнє завантаження процесора при використанні цієї системи склало 42,5%.

При аналізі трафіку з використанням системи Suricata були отримані результати, представлені на рис. 4.13, при чому завантаження процесора системи, на якій виконувалася Suricata представлена на рис 4.14.

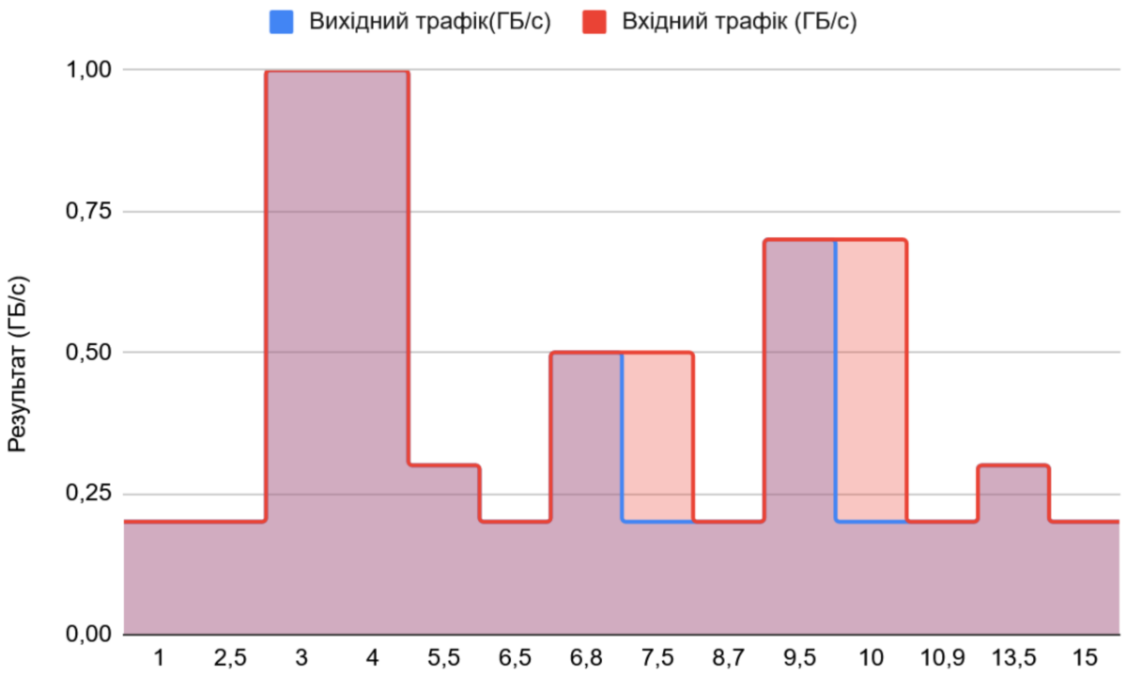


Рисунок 4.13 - Вихідний трафік з використанням Suricata

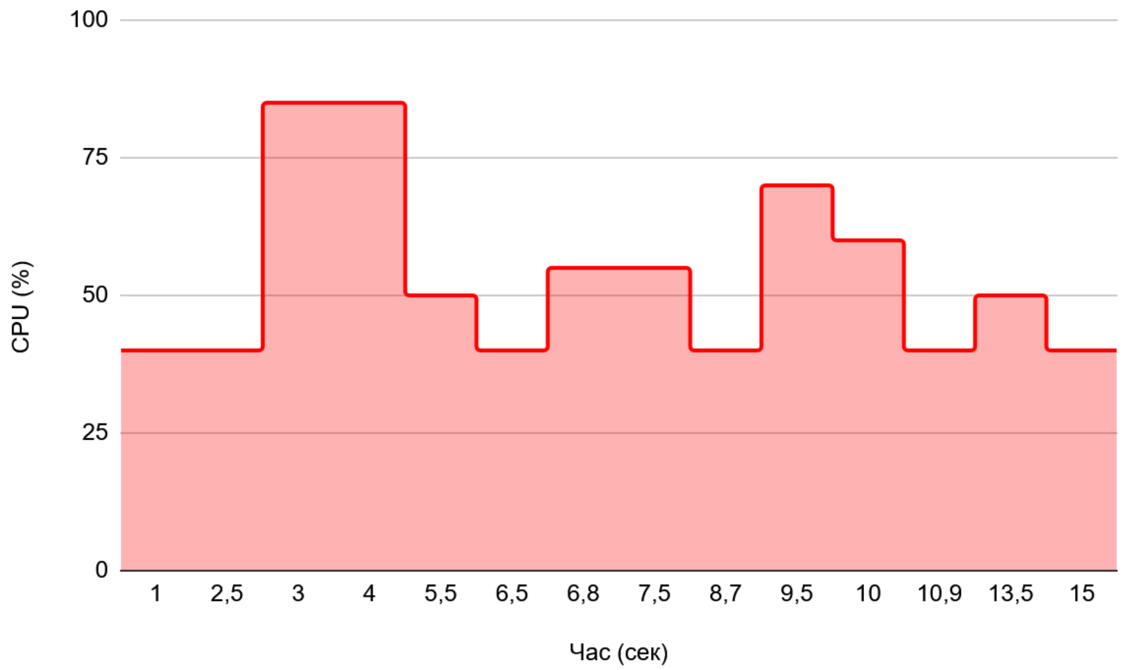


Рисунок 4.14 - Завантаження процесора з використанням Suricata

Як видно з рис. 4.13 та 4.14 використання Suricata дозволило виявити дві з трьох атак, при чому після виявлення другої атаки цастково знизилось навантаження на систему. Середнє завантаження процесора при використанні цієї системи склало 50,5%.

При аналізі трафіку з використанням розробленої системи були отримані результати, представлені на рис. 4.15, при чому завантаження процесора системи, на якій виконувалася власна розробка представлена на рис 4.16.

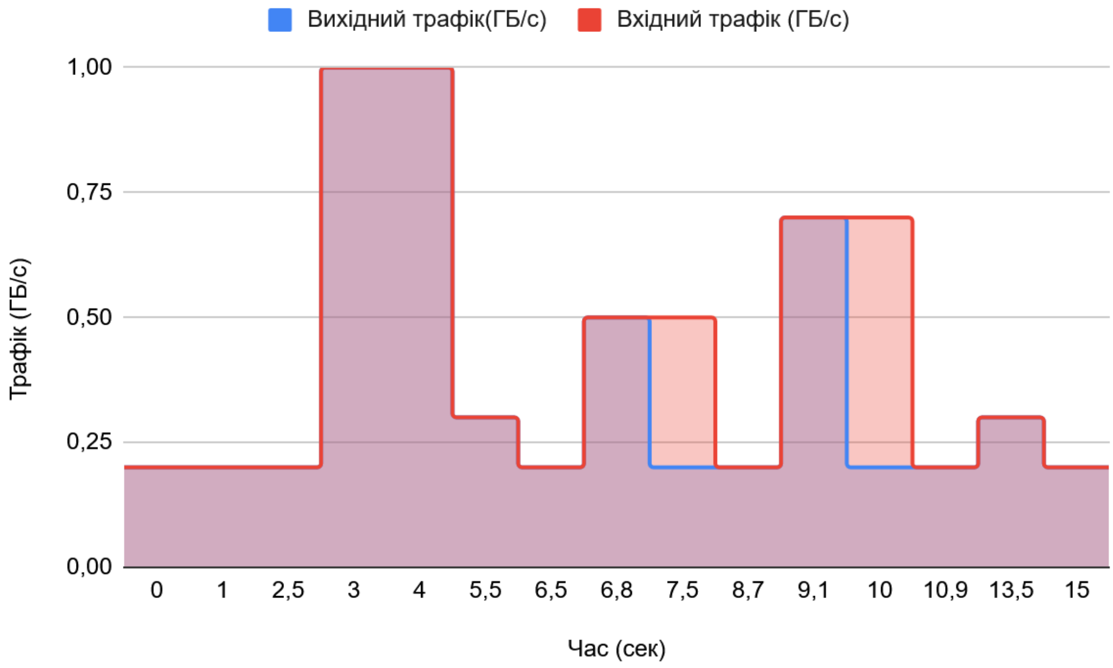


Рисунок 4.15 - Вихідний трафік з використанням власної системи

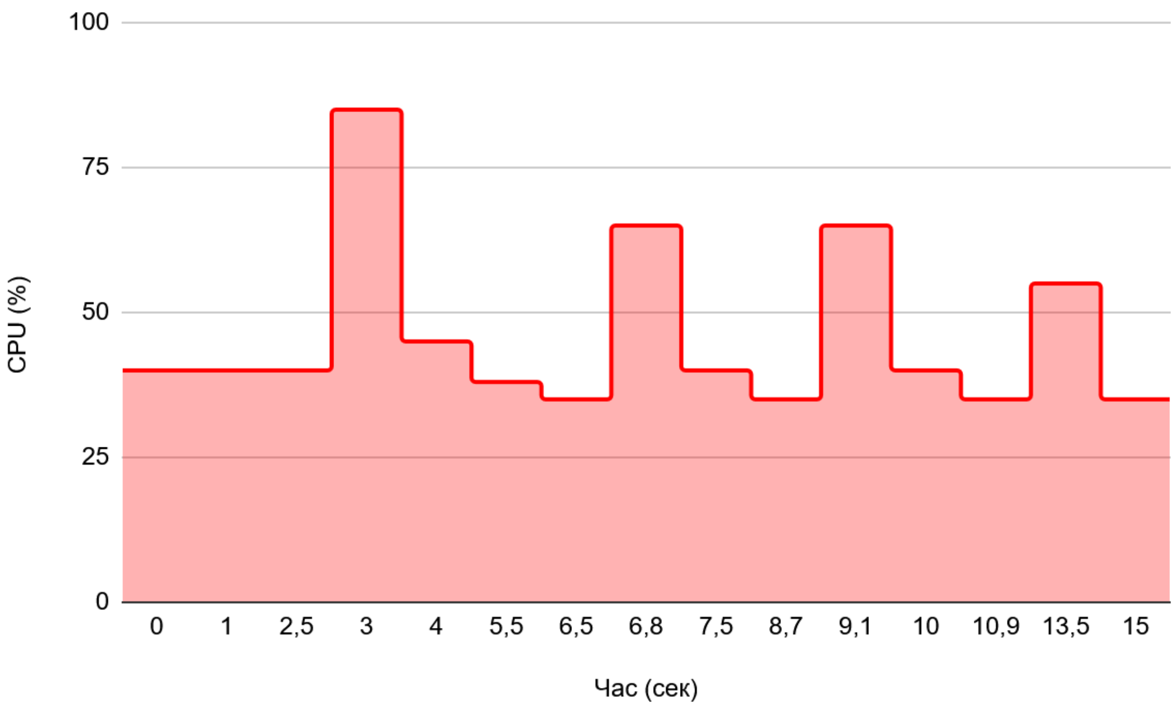


Рисунок 4.16 - Завантаження процесора з використанням власної системи

Як видно з рис. 4.14 та 4.15 використання власної системи дозволило виявити дві з трьох атак, при чому після виявлення аномалії суттєво зменшувалось навантаження на процесор, навіть про повному завантаженні каналу. Середнє завантаження процесора при використанні цієї системи склало 40,5%.

Проведений експеримент підтвердив ефективність розробленої системи аналізу мережевого трафіку з використанням гібридного методу. Порівняльний аналіз результатів засвідчив, що стандартна система Snort виявила лише одну з трьох атак, не зменшуючи навантаження на систему (середнє завантаження CPU — 42,5%). У той час як Suricata виявила дві атаки, але при цьому споживала більше ресурсів (CPU — 50,5%). Натомість, власна розробка забезпечила виявлення двох атак, аналогічно до Suricata, однак показала кращу оптимізацію: після виявлення аномалії навантаження на процесор знижувалось, навіть при високій інтенсивності трафіку. Середнє завантаження CPU при цьому склало 40,5%. Це свідчить про ефективніший розподіл ресурсів та потенціал подальшого розвитку підходу на базі гібридного аналізу трафіку.

Висновки за розділом 4

У даному розділі було детально розглянуто розробку структурної моделі системи виявлення аномального трафіку в ІКС з використанням гібридного методу, який поєднує традиційний сигнатурний аналіз, методи аналізу самоподібності та нечітку логіку. Архітектурна концепція базується на інтеграції системи виявлення вторгнень Snort із власноруч розробленою бібліотекою, яка забезпечує аналіз в режимі реального часу. Такий підхід дозволяє не лише виявляти відомі атаки за допомогою сигнатур, але й орієнтуватися на патерни поведінки, характерні для аномальних подій, що суттєво підвищує здатність системи реагувати на нові або модифіковані загрози.

Архітектура системи була побудована з урахуванням кількох компонентів, що охоплюють повний цикл обробки мережевого трафіку: від первинного захоплення пакетів і їх попередньої обробки (де відбувається фільтрація, нормалізація та агрегація даних) до етапів класифікації та логічного виведення рішень. Важливою особливістю запропонованої моделі є інтеграція власного модуля, який взаємодіє із стандартною платформою Snort через модульну архітектуру. Використання Snort дозволило адаптувати систему до умов реального часу, що забезпечило можливість оперативного аналізу та блокування підозрілих потоків трафіку.

Огляд роботи стандартного Snort, системи Suricata та модифікованої версії Snort із підключеною гібридною бібліотекою, проведено як у лабораторних умовах, так і на реальному мережевому середовищі. Результати експериментів показали, що інтеграція гібридного методу значно зменшує кількість пропущених аномалій та хибних спрацювань. Зокрема, для лабораторних тестів було продемонстровано, що модифікована версія Snort досягла точності (Accuracy) 0,9912 та F1-міри 0,9953, що є суттєвим покращенням у порівнянні з оригінальними рішеннями. Аналіз показників Precision, Recall і Specificity підтвердив, що застосування гібридного підходу дозволяє більш ефективно ідентифікувати аномальні зразки та знижувати вплив шуму нормального трафіку.

Особливу увагу було приділено аналізу ресурсних показників, зокрема навантаження на процесор під час роботи систем. Формула (4.6), що обчислює середнє завантаження CPU, враховує ваговий внесок кожного часовго інтервалу залежно від його тривалості, що дозволяє отримати більш точне значення середнього навантаження. Результати тестування показали, що стандартний Snort у лабораторних умовах демонстрував середнє завантаження процесора близько 42,5%, тоді як використання Suricata приводило до зростання цього показника до 50,5%. У випадку інтеграції гібридного методу, навіть при високому обсязі вихідного трафіку, система демонструвала значно кращу оптимізацію ресурсів, забезпечуючи коректне виявлення атак при середньому навантаженні CPU близько 40,5%.

Проведені експерименти як у лабораторних умовах, так і в режимі реального часу на ізольованих підмережах підтвердили, що використання гібридного методу дозволяє значно покращити показники якості виявлення аномалій. Метрики, розраховані за допомогою стандартних величин (TP, FP, TN, FN), свідчать про перевагу гібридного підходу: зменшення кількості пропущених атак та хибних спрацювань, що особливо важливо в умовах високої динамічності мережевого трафіку та зростання загроз кібербезпеки.

Таким чином, структурна модель системи виявлення аномального трафіку, базована на гібридному методі, демонструє високу адаптивність та ефективність у різних умовах експлуатації. Вона забезпечує не лише оперативне реагування на загрози, а й підвищує надійність системи завдяки комплексному підходу до аналізу трафіку. Результати дослідження свідчать про можливість подальшої інтеграції запропонованої моделі з іншими елементами комплексних систем кіберзахисту, що відкриває перспективи для розширення функціоналу та адаптації до нових типів атак.

ВИСНОВКИ

У межах виконаного дисертаційного дослідження було всебічно опрацьовано науково-практичну задачу виявлення аномального трафіку в ІКС. У процесі проведення дисертаційних досліджень отримані такі наукові та практичні результати.

1. Проведено аналіз сучасних методів та моделей виявлення аномального трафіку в інформаційно-комунікаційних системах та моделей типового користувача та потенційного порушника. На основі проведеного аналізу встановлено, що відомі методи виявлення аномального трафіку не дозволяють забезпечити високий рівень достовірності виявлення аномалій у вихідному трафіку. Обґрунтовано актуальність науково-практичного завдання дисертаційного дослідження. В результаті аналізу моделей поведінки типового користувача та порушника сформовано правила нечіткої логіки для аналізу трафіку.

2. Вдосконалено модель сигнатури пакету для пошуку аномального трафіку, що за рахунок виключення з параметрів сигнатури розміру заголовка, контрольної суми заголовку, корисного розміру пакета, мітки потоку, пріоритету пакету, контрольної суми пакету за принципом Парето. Це забезпечило зменшення часу аналізу трафіку.

3. Вдосконалено модель процесу визначення аномального трафіку на основі самоподібності з урахуванням потенційних загроз та особливостей трафіку в інформаційно-комунікаційних системах, за рахунок проведення аналізу компонентів сигнатур. Це дозволило формувати множину нормальних сигнатур.

4. Вдосконалено модель процесу нечіткого визначення аномального трафіку за рахунок використання експертного підходу сформована множина правил та набір відповідних лінгвістичних змінних, що враховують поведінку типового користувача та потенційного порушника. Це дозволило розробити нові підходи до виявлення аномального трафіку в інформаційно-комунікаційних

системах.

5. Розроблено гібридний метод виявлення аномального трафіку в інформаційно-комунікаційних системах в якому за рахунок інтеграції методу класифікації трафіку за ознаками, методу самоподібності та розробленої моделі процесу нечіткого виявлення. Це дозволило динамічно формувати множини сигнатур при класифікації трафіку за ознаками та забезпечити високі показники виявлення аномального трафіку, а саме точність – 0,9944, повнота – 0,9962.

6. Удосконалено структурну модель системи виявлення аномального трафіку в інформаційно-комунікаційних системах, що за рахунок інтегрування розробленого гібридного методу виявлення аномального трафіку в інформаційно-комунікаційних системах та динамічного варіювання множиною дозволених та заборонених з'єднань у режимі реального часу дозволило на 20-25% знизити середнє навантаження на процесор у порівнянні з Snort та Suricata.

7. Розроблене алгоритмічне та програмне забезпечення запропонованої системи дозволило провести експериментальні дослідження та впровадити розроблений гібридний метод.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Дячек О., Рябченко К., Доценко А. (2022). Безпека даних в інформаційно-комунікаційному середовищі та її складність для нових бізнес-моделей. Економіка та суспільство, (38). doi:10.32782/2524-0072/2022-38-16.

2. Barchenko N., Lubchak V., Lavryk T. (2022). Модель індикаторів оцінки національного рівня цифровізації та кібербезпеки держав світу. Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка», 2(18), 73–85. doi: 10.28925/2663-4023.2022.18.7385.

3. Гайдук О., Звєрєв В. (2024). Аналіз кіберзагроз в умовах стрімкого розвитку інформаційних технологій. Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка», 3(23), 225–236. doi: 10.28925/2663-4023.2024.23.225236.

4. Гарасимчук О., Партика А., Нємкова О., & Совин Я. (2023). Інтегрований підхід до забезпечення кібербезпеки та дослідження кібезлочинів критичної інфраструктури за допомогою системи моніторингу інцидентів вірусів-вимагачів. Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка», 1(21), 286–296. doi: 10.28925/2663-4023.2023.21.286296.

5. Кужда Т., Луциків І. (2022). Дослідження стану та перспектив цифрового розвитку в Україні. Галицький економічний вісник, 78–79(5–6), 146–155. doi: 10.33108/galicianvisnyk_tntu2022.05_06.146.

6. Andrukh A., Yurchenko Y. (2024). Cyber security of critical infrastructure: challenges of innovation and threats of digital technologies. Challenges and Issues of Modern Science, 3, 157–166. URL: https://cims.fti.dp.ua/j/article/view/210.

7. Ксьонжик І., Жовта Н., Павліна А. (2021). Страхування ризиків кібербезпеки діяльності суб’єктів господарювання в сучасному інформаційному просторі. Економіка та суспільство, (34). doi: 10.32782/2524-0072/2021-34-90.

8. Яровенко Г., Кобзенко В. (2022). Попередній аналіз і підготовка даних для прогнозування трендів кібератак. Економіка та суспільство, (45). doi: 10.32782/2524-0072/2022-45-42.

9. Середня вартість збитків від витоку даних від кібератак у світі в 2022 році зросла до $ 4,4 млн. Forinsurer. URL: https://forinsurer.com/news/23/02/20/42397 (дата звернення: 20.04.2024).

10. Aslan Ö., Aktuğ S.S., Ozkan-Okay M., Yilmaz A.A., Akin E.. A Comprehensive Review of Cyber Security Vulnerabilities, Threats, Attacks, and Solutions. Electronics 2023, 12, 1333. doi: 10.3390/electronics12061333.

11. Yaremchuk, K., Voskoboinykov, D., Melkozerova, O. (2022). Modern threats and ways to secure web applications. Computer Science and Cybersecurity, (2), 28-34. doi: 10.26565/2519-2310-2022-2-03.

12. Савченко В. А. (2024). Дослідження потенційного впливу соціальної інженерії на процеси цифрової трансформації. Зв’язок, (3), 12-17. doi: 10.31673/2412-9070.2024.031217.

13. Laptiev, S. (2022). Удосконалений метод захисту персональних даних від атак за допомогою алгоритмів соціальної інженерії. Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка», 4(16), 45–62. doi: 10.28925/2663-4023.2022.16.4562.

14. Лозовський Р., Мороз А., Дармограй Д. (2024). Кібербезпека: проблеми, виклики та можливі рішення. Measuring and computing devices in technological processes, (2), 378–383. doi: 10.31891/2219-9365-2024-78-43.

15. Кузик Я., Пшеничних С. Ефективність симуляцій RED TEAM для боротьби з фішинговими атаками: новітні підходи та технології. V Міжнародна студентська наукова конференція «Міждисциплінарні наукові дослідження та перспективи їх розвитку», 4 жовтня 2024 р., м. Черкаси, 2024. 103–107. URL: https://archive.liga.science/index.php/conference-proceedings/article/view/1229.

16. Івкова, В., Опірський, І. (2024). Дослідження проблематики забезпечення безпеки персональних даних та конфіденційної інформації в контексті протидії OSINT. Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка», 2(26), 189–199. doi: 10.28925/2663-4023.2024.26.682.

17. Алексеєнко О.А., Бурлін М.О., Кисельов О.В., Кравчук В.В., Мальгіна К.В. (2022). Методики фішингу в мобільних системах. Сучасний захист

інформації, 1, 44-49. doi: 10.31673/2409-7292.2022.014449.

18. Фішинг допомагає маніпулювати людьми заради даних. Як захиститися від кіберзагроз. URL: https://journal.gen.tech/post/sho-take-socialna-ingeneria (Дата звернення: 25.04.2024)

19. Klots Y., Petliak N., Martsenko S., Tymoshchuk V., Bondarenko I.. Machine Learning system for detecting malicious traffic generated by IoT devices. 2nd International Workshop on Computer Information Technologies in Industry 4.0, Ternopil, 12 June 2024, 3742, 97-110.

20. Костюк Ю. В., Бебешко Б. Т., Крючкова Л. П., Литвинов В.А., Оксанич І.М., Складанний П. М., Хорольська К.В. (2024). Захист інформації та безпека обміну даними в безпроводових мобільних мережах з автентифікацією і протоколами обміну ключами. Кібербезпека: освіта, наука, техніка, 1 (25). 229-252. ISSN 2663-4023.

21. A. Ali, N. A. Somroo, U. Farooq, M. Asif, I. Akour, W. Mansoor. (2022). Smartphone Security Hardening: Threats to Organizational Security and Risk Mitigation, 2022 International Conference on Cyber Resilience (ICCR), Dubai, United Arab Emirates, 1-12. doi: 10.1109/ICCR56254.2022.9995769.

22. Янко А., Прокудін А., Філь І., Крук О. (2024). Виявлення атак типу LDDOS за допомогою SDN мереж з елементами машинного навчання. Measuring and computing devices in technological processes, (4), 287–296. doi: 10.31891/2219-9365-2024-80-36.

23. Кльоц Ю., Петляк Н. (2022). Виявлення аномального трафіку у загальнодоступних комп’ютерних мережах. Measuring and computing devices in technological processes, (3), 79–86. doi: 10.31891/2219-9365-2022-71-3-9.

24. Звіт роботи системи виявлення вразливостей і реагування на кіберінциденти та кібератаки (2021). Державний центр кіберзахисту Державної служби спеціального зв’язку та захисту інформації України. URL: https://cert.gov.ua/files/pdf/SOC_Annual_Report_2022.pdf (дата звернення 01.02.2023).

25. Звіт про роботу системи виявлення вразливостей і реагування на

кіберінциденти та кібератаки (2022). Державний центр кіберзахисту Державної служби спеціального зв’язку та захисту інформації України. URL: https://scpc.gov.ua/api/docs/sseb6a10-b7aa-4396-8b04-e0e4b7fca1l1/sseb6a10-b7aa-4396-8b04-e0e4b7fca1l1.pdf (дата звернення 01.02.2023).

26. Звіт про роботу системи виявлення вразливостей і реагування на кіберінциденти та кібератаки (2023). Державний центр кіберзахисту Державної служби спеціального зв’язку та захисту інформації України. URL: https://scpc.gov.ua/api/files/9c21855d-74da-45d1-90f9-5d4f6795996a (дата звернення 21.02.2024).

27. Number of common IT security vulnerabilities and exposures (CVEs) worldwide from 2009 to 2024 YTD. Statista. URL: https://www.statista.com/statistics/500755/worldwide-common-vulnerabilities-and-exposures (дата звернення 17.03.2024).

28. Annual number of data compromises and individuals impacted in the United States from 2005 to 2024. Statista. URL: https://www.statista.com/statistics/273550/data-breaches-recorded-in-the-united-states-by-number-of-breaches-and-records-exposed/ (дата звернення 05.04.2025)

29. 239 Cybersecurity Statistics (2023). Packetlabs. URL: https://www.packetlabs.net/posts/239-cybersecurity-statistics-2023/ (дата звернення 31.04.2023).

30. Cost of a Data Breach Report 2024. IBM. URL: https://www.ibm.com/downloads/documents/us-en/107a02e94948f4ec (дата звернення 03.02.2025).

31. Zero Tolerance: More Zero-Days Exploited in 2021 Than Ever Before. Google Cloud Blog. URL: https://cloud.google.com/blog/topics/threat-intelligence/zero-days-exploited-2021 (дата звернення 11.05.2022).

32. B. Savenko, S. Lysenko, K. Bobrovnikova, O. Savenko, G. Markowsky. (2021). Detection DNS Tunneling Botnets. In 2021 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS). IEEE Press, 64–69. doi:

10.1109/IDAACS53288.2021.9661022.

33. NETSCOUT DDoS Threat Intelligence Report. Latest Cyber Threat Intelligence Report. URL: https://www.netscout.com/threatreport/key-findings/ (дата звернення 23.01.2024).

34. Cyberthreats in the financial industry: H2 2023 – H1 2024. Positive Technologies. URL: https://global.ptsecurity.com/analytics/financial-industry-cyberthreats-h2-2023-h1-2024 (дата звернення 08.08.2024).

35. Check Point Research Reports a 38% Increase in 2022 Global Cyberattacks. Check Point Blog. URL: https://blog.checkpoint.com/2023/01/05/38-increase-in-2022-global-cyberattacks/ (дата звернення 14.03.2023).

36. Корченко А., Мацюк С., Давиденко К. (2024). Огляд сучасних методів та засобів виявлення соціотехнічних атак. Information Technology: Computer Science, Software Engineering and Cyber Security, (3), 88-96. doi: 10.32782/IT/2024-3-9.

37. Петляк Н.С., Кльоц Ю.П., Хохлачова Ю.Є.. Виявлення аномального трафіку у загальнодоступних комп’ютерних мережах. Інформаційна, функційна і кібербезпека: тези доп. ІІ студ. наук.-техн. к., м. Харків, 30 листопада – 1 грудня 2022 р., Х., 2022. С. 45-46.

38. W. Duo, M. Zhou, A. Abusorrah. (2022). A Survey of Cyber Attacks on Cyber Physical Systems: Recent Advances and Challenges, IEEE/CAA Journal of Automatica Sinica, 9, 5, 784-800. doi: 10.1109/JAS.2022.105548.

39. Demertzi V, Demertzis S, Demertzis K. An Overview of Cyber Threats, Attacks and Countermeasures on the Primary Domains of Smart Cities. Applied Sciences. 2023, 13(2), 790. doi: 10.3390/app13020790.

40. Syfert M., Ordys A., Kościelny J.M., Wnuk P., Możaryn J., Kukiełka K.. (2022) Integrated Approach to Diagnostics of Failures and Cyber-Attacks in Industrial Control Systems. Energies 2022, 15, 6212. doi: 10.3390/en15176212.

41. R. Alguliyev, L. Sukhostat, A. Mammadov. (2022). Anomaly Detection in Cyber-Physical Systems based on BiGRU-VAE, 2022 IEEE 16th International Conference on Application of Information and Communication Technologies (AICT),

Washington DC, DC, USA, 1-5. doi: 10.1109/AICT55583.2022.10013581.

42. M. Ravinder, V. Kulkarni. (2023). A Review on Cyber Security and Anomaly Detection Perspectives of Smart Grid, 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2023, pp. 692-697. doi: 10.1109/ICSSIT55814.2023.10060871.

43. Pavlo R, Oleg S, Sachenko A, Andriy D, Petro V.. (2024). A trust model that ensures the correctness of computing in grid computing system. CEUR Workshop Proceedings, 3675, 388-401. URL: https://ceur-ws.org/Vol-3675/paper28.pdf.

44. Ignacio Samuel Crespo-Martínez, Adrián Campazas-Vega, Ángel Manuel Guerrero-Higueras, Virginia Riego-DelCastillo, Claudia Álvarez-Aparicio, Camino Fernández-Llamas. (2023). SQL injection attack detection in network flow data. Computers & Security, 127, 103093. doi: 10.1016/j.cose.2023.103093.

45. Atacak İ. (2023). An Ensemble Approach Based on Fuzzy Logic Using Machine Learning Classifiers for Android Malware Detection. Applied Sciences, 13(3), 1484. doi: 10.3390/app13031484.

46. Protic D., Stankovic M., Antic, V. (2022). WK-FNN design for detection of anomalies in the computer network traffic. Facta Universitatis, Series: Electronics and Energetics, 35(2), 269-282. doi: 10.2298/FUEE2202269P.

47. M. Thankappan, H. Rifà-Pous, C. Garrigues. (2024). A Signature-Based Wireless Intrusion Detection System Framework for Multi-Channel Man-in-the-Middle Attacks Against Protected Wi-Fi Networks. IEEE Access, 12, 23096-23121. doi: 10.1109/ACCESS.2024.3362803.

48. I. Ullah and Q. H. Mahmoud. (2022). Design and Development of RNN Anomaly Detection Model for IoT Networks. IEEE Access, 10, 62722-62750. doi: 10.1109/ACCESS.2022.3176317.

49. RNN & LSTM & BI-LSTM & GRU. URL: https://medium.com/%40antoniosmalak14/rnn-lstm-bi-lstm-gru-eb5869660691 (Дата звернення: 27.04.2023)

50. Корнієнко В., Герасіна О., Тимофєєв Д., Сафаров О., Ковальова Ю. (2022). Ідентифікація та прогнозування самоподібного трафіку інформаційно-

комунікаційних мереж для систем виявлення атак. Information Technology: Computer Science, Software Engineering and Cyber Security, 1, 20–29. doi: 10.32782/IT/2022-1-4

51. Гайдур Г.І., Гахов С.О., Гамза Д.Є. (2024). Модель виявлення шкідливої активності в інформаційній системі організації на основі гібридної класифікації. Сучасний захист інформації, 4(60), 30–38. doi: 10.31673/2409-7292.2024.040003.

52. de Campos Souza, Paulo Vitor, Augusto Junio Guimarães, Thiago Silva Rezende, Vinicius Jonathan Silva Araujo, Vanessa Souza Araujo. (2020). Detection of Anomalies in Large-Scale Cyberattacks Using Fuzzy Neural Networks. AI, 1(1), 92-116. doi: 10.3390/ai1010005.

53. Gadal S, Mokhtar R, Abdelhaq M, Alsaqour R, Ali ES, Saeed R. (2022). Machine Learning-Based Anomaly Detection Using K-Mean Array and Sequential Minimal Optimization. Electronics, 11(14), 2158. doi: 10.3390/electronics11142158.

54. W. Xu, J. Jang-Jaccard, A. Singh, Y. Wei and F. Sabrina. (2021). Improving Performance of Autoencoder-Based Network Anomaly Detection on NSL-KDD Dataset. IEEE Access, 9, 140136-140146, doi: 10.1109/ACCESS.2021.3116612.

55. Imran, Jamil, F., Kim, D. (2021). An Ensemble of Prediction and Learning Mechanism for Improving Accuracy of Anomaly Detection in Network Intrusion Environments. Sustainability, 13(18), 10057. doi: 10.3390/su131810057.

56. Kangkang Zhang, Marios M. Polycarpou, Thomas Parisini. (2020). Enhanced Anomaly Detector for Nonlinear Cyber-Physical Systems against Stealthy Integrity Attacks. IFAC-PapersOnLine, 53(2), 13682-13687, doi: 10.1016/j.ifacol.2020.12.870.

57. Alberto Barbado, Óscar Corcho, Richard Benjamins. (2022). Rule extraction in unsupervised anomaly detection for model explainability: Application to OneClass SVM. Expert Systems with Applications, 189, 116100, doi: 10.1016/j.eswa.2021.116100.

58. Einy S., Oz C., Navaei, Y. D. (2021). The anomaly-and signature-based IDS for network security using hybrid inference systems. Mathematical Problems in Engineering, 2021(1), 6639714. doi: 10.1155/2021/6639714.

59. Mohammad Adiban, Sabato Marco Siniscalchi, Giampiero Salvi. (2023). A step-by-step training method for multi generator GANs with application to anomaly detection and cybersecurity. Neurocomputing, 537, 296-308. doi: 10.1016/j.neucom.2023.03.056.

60. Shaojun Sheng, Xin Wang. (2023). Network traffic anomaly detection method based on chaotic neural network. Alexandria Engineering Journal, 77, 567-579, doi: 10.1016/j.aej.2023.07.019.

61. Волокита А., Меленчуков М. (2024). Дослідження моделей виявлення атак на розподілені системи за допомогою згорткових нейронних мереж. Measuring and computing devices in technological processes, (3), 224–229. doi: 10.31891/2219-9365-2024-79-29.

62. Madalina G. Ciobanu, Fausto Fasano, Fabio Martinelli, Francesco Mercaldo, Antonella Santone. (2019). Model Checking for Data Anomaly Detection. Procedia Computer Science, 159, 1277-1286, doi: 10.1016/j.procs.2019.09.297.

63. Davydiuk A.. Zubok V., Khokhlachova Y., Khudyntsev M., Komarov M. (2023). The cyber statistics in Ukraine. Current state. Ukrainian Scientific Journal of Information Security, 29(2). doi: 10.18372/2225-5036.29.17868.

64. A. Mezina, R. Burget, C. M. Travieso-González. (2021). Network Anomaly Detection With Temporal Convolutional Network and U-Net Model. IEEE Access, 9, 143608-143622. doi: 10.1109/ACCESS.2021.3121998.

65. Петляк Н.С. (2025). Аналіз моделей виявлення аномалій трафіку в сучасних інформаційно-комунікаційних системах та мережах. Measuring and computing devices in technological processes, (1), 180-186. doi: 10.31891/2219-9365-2025-81-21.

66. Klots, Y., Titova, V., Petliak, N., Cheshun, V., Salem, A.-B.M. (2022). Research of the Neural Network Module for Detecting Anomalies in Network Traffic. CEUR Workshop Proceedings, 3156, 378-389. URL: https://ceur-ws.org/Vol-3156/paper28.pdf.

67. Kiflay A., Tsokanos A., Fazlali M., Kirner R. (2024). Network intrusion detection leveraging multimodal features. Array, 22, 100349. doi:

10.1016/j.array.2024.100349.

68. Gu X., Howells G., Yuan H. (2024). A soft prototype-based autonomous fuzzy inference system for network intrusion detection. Information Sciences, 677, 120964. doi: 10.1016/j.ins.2024.120964.

69. S. Kushal, B. Shanmugam, J. Sundaram et al. (2024). Self-healing hybrid intrusion detection system: an ensemble machine learning approach. Discover Artificial Intelligence 4, 28. doi: 10.1007/s44163-024-00120-9.

70. I. Mbona, J. H. P. Eloff. (2022). Detecting Zero-Day Intrusion Attacks Using Semi-Supervised Machine Learning Approaches. IEEE Access, 10, 69822-69838. doi: 10.1109/ACCESS.2022.3187116.

71. M. S. E. Sayed, N. -A. Le-Khac, M. A. Azer and A. D. Jurcut. (2022). A Flow-Based Anomaly Detection Approach With Feature Selection Method Against DDoS Attacks in SDNs. IEEE Transactions on Cognitive Communications and Networking, 8, 4, 1862-1880. doi: 10.1109/TCCN.2022.3186331.

72. Shaikh A., Gupta P. (2023). Advanced Signature-Based Intrusion Detection System. In: Rajakumar, G., Du, KL., Vuppalapati, C., Beligiannis, G.N. (eds) Intelligent Communication Technologies and Virtual Mobile Networks. Lecture Notes on Data Engineering and Communications Technologies, 131. Springer, Singapore. doi: 10.1007/978-981-19-1844-5_24.

73. N. Sivasankari and S. Kamalakkannan. (2022). Detection and prevention of man-in-the-middle attack in iot network using regression modeling, Advances in Engineering Software, 169, 103126. doi: 10.1016/j.advengsoft.2022.103126.

74. Mohammed Ishaque, Md Gapar Md Johar, Ali Khatibi, Muhammed Yamin. (2023). A novel hybrid technique using fuzzy logic, neural networks and genetic algorithm for intrusion detection system, Measurement: Sensors, 30, 100933. doi: 10.1016/j.measen.2023.100933.

75. Radivilova T., Kirichenko L., Tawalbeh M., Ilkov, A. (2021). Виявлення аномалій в телекомунікаційному трафіку статистичними методами. Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка», 3(11), 183–194. doi: 10.28925/2663-4023.2021.11.183194.

76. M. P. Novaes, L. F. Carvalho, J. Lloret and M. L. Proença. (2020). Long Short-Term Memory and Fuzzy Logic for Anomaly Detection and Mitigation in Software-Defined Network Environment. IEEE Access, 8, 83765-83781. doi: 10.1109/ACCESS.2020.2992044.

77. Daniele Canavese, Leonardo Regano, Cataldo Basile, Gabriele Ciravegna, Antonio Lioy. (2022). Encryption-agnostic classifiers of traffic originators and their application to anomaly detection. Computers & Electrical Engineering, 97, 107621, doi: 10.1016/j.compeleceng.2021.107621.

78. N. Berjab, H. H. Le and H. Yokota. (2022). Recovering Missing Data via Top-k Repeated Patterns for Fuzzy-Based Abnormal Node Detection in Sensor Networks. IEEE Access, 10, 61046-61064, doi: 10.1109/ACCESS.2022.3181742.

79. Chen L., Gao S., Liu, B. (2022). An improved density peaks clustering algorithm based on grid screening and mutual neighborhood degree for network anomaly detection. Scientific Reports 12, 1409. doi: 10.1038/s41598-021-02038-z.

80. Jing Zhang, Yige Yuan, Jiahong Zhang, Yang Yang, Wenjin Xie. (2023). Anomaly detection method based on penalty least squares algorithm and time window entropy for Cyber–Physical Systems. Journal of King Saud University - Computer and Information Sciences, 35(10), 101860, doi: 10.1016/j.jksuci.2023.101860.

81. Mariusz Pelc, Dawid Galus, Magda Zolubak, Stepan Ozana, Wojciech Chlewicki, Katarzyna Cichon, Michal Podpora, Aleksandra Kawala-Sterniuk. (2019). Behavioural Approach to Network Anomaly Detection for Resource-Constrained System – Presentation of the Novel Solution – Preliminary Study. IFAC-PapersOnLine, 52(27), 121-126. doi: 10.1016/j.ifacol.2019.12.743.

82. Шпінарева І. М., Якушина А. О., Волощук Л. А., Рудніченко М. Д. (2021). Використання методів поглиблених навчання для виявлення і класифікації мережевих атак. Вісник сучасних інформаційних технологій, 4(3), 244-254. doi: 10.15276/hait.03.2021.4.

83. Саприкін О.С. (2021). Моделі і методи діагностування Zero-Day загроз в кіберпросторі. Вісник сучасних інформаційних технологій, 4(2), 155-167. doi: 10.15276/hait.02.2021.5.

84. Шагін В., Нічепорук А., Каштальян А. (2021). Централізована розподілена система виявлення атак у корпоративних комп’ютерних мережах на основі мультифрактального аналізу. Measuring and computing devices in technological processes, (1), 50–55. doi: 10.31891/2219-9365-2021-67-1-7.

85. Лужецький В., Савицька Л. (2024). Аналіз підходів щодо виявлення аномалій в децентралізованих комп’ютерних мережах. Measuring and computing devices in technological processes, (2), 395–401. doi: 10.31891/2219-9365-2024-78-46.

86. Ayan Chatterjee, Bestoun S. Ahmed. (2022). IoT anomaly detection methods and applications: A survey. Internet of Things, 19, 100568, doi: 10.1016/j.iot.2022.100568.

87. Савченко В., Поночовний П., Аверічев, І. (2024). Виявлення DDOS-атаки на високошвидкісну мережу: опитування. Прикладні проблеми комп’ютерних наук, безпеки та математики, (3), 71–81. URL: https://www.sciencedirect.com/science/article/abs/pii/S0045790622006814.

88. Haiping Lin, Chengwen Wu, Mohammad Masdari. (2022). A comprehensive survey of network traffic anomalies and DDoS attacks detection schemes using fuzzy techniques. Computers and Electrical Engineering, 104, 108466, doi: 10.1016/j.compeleceng.2022.108466.

89. Marchenko R., Kovalenko A., Znaidiuk V. (2024). Аналіз методів виявлення аномального трафіку в мережах IOT. Системи управління, навігації та зв’язку. Збірник наукових праць. 1. 133-136. doi: 10.26906/SUNZ.2024.1.133.

90. Fahim M., Sillitti A. (2019). Anomaly detection, analysis and prediction techniques in iot environment: A systematic literature review. IEEE Access, 7, 81664-81681. doi: 10.1109/ACCESS.2019.2921912.

91. Marchenko R., Kovalenko A., Znaidiuk V. (2024). Аналіз методів виявлення аномального трафіку в мережах IOT. Системи управління, навігації та зв’язку. Збірник наукових праць. 1. 133-136. doi: 10.26906/SUNZ.2024.1.133.

92. A. B. Nassif, M. A. Talib, Q. Nasir, F. M. Dakalbab. (2021). Machine Learning for Anomaly Detection: A Systematic Review. IEEE Access, 9, 78658-

78700. doi: 10.1109/ACCESS.2021.3083060.

93. Тітова В., Кльоц Ю., Петляк Н., Капустян М. (2022). Підсистема нечіткого висновку для класифікації загроз комп'ютерній інформації. Measuring and computing devices in technological processes, (1), 57–61. doi: 10.31891/2219-9365-2022-69-1-8.

94. Нічепорук А., Нічепорук А., Савенко О., Казанцев А. (2021). An intelligent system for detecting anomalies and identifying smart home devices based on the collective communication. Electrotechnic and Computer Systems, 34(110), 50-61. doi: 10.15276/eltecs.34.110.2021.6.

95. Petliak N., Klots Y., Titova V., Cheshun V., Boyarchuk A.. (2023). Signature-based Approach to Detecting Malicious Outgoing Traffic. 4th International Workshop on Intelligent Information Technologies and Systems of Information Security, IntellTSIS 2023. CEUR Workshop Proceedings, 3373, 486–506.

96. Petliak N., Khokhlachova Y.. (2024). Method of analysis of outgoing traffic package signatures. Ukrainian Information Security Research Journal, 26(1), 206-214. doi: 10.18372/2410-7840.26.18841.

97. Петляк Н.С., Кльоц Ю.П., Хохлачова Ю.Є.. Підхід до аналізу вихідного трафіку. Безпека інформаційних технологій: тези доп. XII міжнар. наук.-техн. конф., м. Ужгород, 2-4 травня 2023 р., К., 2023. С. 97-99.

98. Петляк Н.С., Кльоц Ю.П.. Підхід до аналізу вихідного трафіку на основі сигнатур. Інформаційна безпека та комп'ютерні технології: тези доп. VI міжнар. наук.-пр. конф., м. Кропивницький, 20-21 квітня 2023 р., Кр., 2023. С. 3-4.

99. Petliak N., Klots Y., Titova V., Salem A.-B.M.. Attack detection system based on network traffic analysis by means of fuzzy inference. 1st International Workshop on Advanced Applied Information Technologies, Khmelnytskyi, 5 December 2024, 3899, 201-213.

100. Петляк Н.С., Кльоц Ю.П.. Метод ідентифікації вторгнень на основі алгоритму визначення самоподібності трафіку та алгоритмів нечіткої логіки. Безпека інформаційних технологій: тези доп. XIII міжнар. наук.-техн. конф., м. Львів, 9-11 травня 2024 р., Л., 2024. С. 117-119.

101. Петляк Н.С., Кльоц Ю.П., Тітова В.Ю., Чешун В.М. Виявлення зловмисного вихідного трафіку мережі на основі нечіткого логічного висновку. Захист інформації і безпека інформаційних систем: тези доп. IX міжнар. наук.-техн. конф., м. Львів, 25-26 травня 2023 р., Л., 2023. С. 33-35.

102. Мостовий С.В., Петляк Н.С., Голота І.О. (2023). Дослідження ефективності інструментів виявлення і запобігання вторгнень на вузли в корпоративних мережах. Measuring and computing devices in technological processes, (2), 5-8. doi: 10.31891/2219-9365-2023-74-1.

103. Titova V., Klots Y., Petliak N., Cheshun V., Salem A.-B.M.. Detection of network attacks in cyber-physical systems using a rule-based logical neural network. 1st International Workshop on Intelligent and CyberPhysical Systems (ICyberPhyS 2024), Khmelnytskyi, 28 June 2024. Vol. 3736. P. 255 - 268. ISSN 16130073.

104. Y. Klots, N. Petliak, V. Titova. (2023). Evaluation of the efficiency of the system for detecting malicious outgoing traffic in public networks. 13th International Conference on Dependable Systems, Services and Technologies (DESSERT), Athens, Greece, 1-5. doi: 10.1109/DESSERT61349.2023.10416502.

ДОДАТОК А

АКТИ ВПРОВАДЖЕННЯ

ЗАТВЕРДЖУЮ
Ректор Хмельницького національного університету
________ Сергій МАТЮХ
(підпис) (Ім'я, ПРІЗВИЩЕ)
"25" березня 2025 року
М.П.

АКТ

Комісія у складі:
голови – проректор з науково-педагогічної роботи, к.е.н., доцент, Лопатовський Віктор Григорович;
членів комісії: – декан факультету інформаційних технологій, д.т.н., професор, Говорущенко Тетяна Олександрівна;
– доцент кафедри кібербезпеки, гарант ОП «Кібербезпека та захист інформації» магістерського рівня, к.т.н., доцент, Тітова Віра Юріївна;

цим Актом засвідчує, що результати дисертаційного дослідження аспіранта кафедри кібербезпеки державного некомерційного підприємства «Державний університет «Київський авіаційний інститут», асистента кафедри кібербезпеки Хмельницького національного університету

Петляк Наталії Сергіївни
Прізвище, ім'я, по батькові

на тему: "Моделі та методи виявлення аномального трафіку в інформаційно-комунікаційних системах"
Тема дисертації

на здобуття ступеня доктора філософії за спеціальністю 125 Кібербезпека
шифр і назва спеціальності

впроваджені в діяльності Хмельницького національного університету.
назва установи, організації

Зокрема, у процесі підготовки пропозицій щодо підготовки та викладання курсу лекцій з дисципліни «Технології виявлення вразливостей та вторгнень» використано: аналітичні матеріали щодо аналізу загроз безпеки інформації (розділ 1 дисертаційної роботи), моделей та методів виявлення аномалій в сучасних інформаційно-комунікаційних системах (розділи 2-3 дисертаційної роботи).

Голова комісії	к.е.н., доцент (науковий ступінь, вчене звання)	(підпис)	Віктор ЛОПАТОВСЬКИЙ (Ім'я, ПРІЗВИЩЕ)
Члени комісії	д.т.н., професор (науковий ступінь, вчене звання)	(підпис)	Тетяна ГОВОРУЩЕНКО (Ім'я, ПРІЗВИЩЕ)
	к.т.н., доцент (науковий ступінь, вчене звання)	(підпис)	Віра ТІТОВА (Ім'я, ПРІЗВИЩЕ)

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
вул. Інститутська 11, Хмельницький-16, 29016, тел.: (0382) 67-02-76, факс: (0382) 67-42-65
e-mail: centr@khmnu.edu.ua, код ЄДРПОУ 02071234

02.04.2025 № 022/16 На №____________ від____________________

ДОВІДКА ПРО ВПРОВАДЖЕННЯ

Результати дисертаційного дослідження *Петляк Наталії Сергіївни* на тему: "*Моделі та методи виявлення аномального трафіку в інформаційно-комунікаційних системах*" на здобуття ступеня доктора філософії за спеціальністю 125 – *Кібербезпека* впроваджені в діяльності *кафедри кібербезпеки та центру цифрових технологій*.

Зокрема, у процесі підготовки пропозицій щодо підготовки та викладання курсу лекцій з дисципліни «Технології виявлення вразливостей та вторгнень» використано: аналітичні матеріали щодо аналізу загроз безпеки інформації (розділ 1 дисертаційної роботи), моделей та методів виявлення аномалій в сучасних інформаційно-комунікаційних системах (розділи 2-3 дисертаційної роботи).

Ректор
к.е.н., доцент ____________ Сергій МАТЮХ
(підпис) (Ім'я, ПРІЗВИЩЕ)

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ • ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ • Ідентифікаційний код 02071234

ЗАТВЕРДЖУЮ
Директор
Ігор ЛИТВІНЧУК

"10" квітня 2025 року
М.П.

А К Т

Комісія у складі:

голови – директора ТОВ «Х-СІТУ» Ігоря ЛИТВІНЧУКА
членів комісії: – Заступник директора з технічних питань Івана ШТОГРИНА
– Заступник директора з комерційних питань Руслана КОНИКА

цим Актом засвідчує, що результати дисертаційного дослідження аспіранта кафедри кібербезпеки державного некомерційного підприємства «Державний університет «Київський авіаційний інститут»

Петляк Наталії Сергіївни
Прізвище, ім'я, по батькові

на тему: "Моделі та методи виявлення аномального трафіку в інформаційно-комунікаційних системах"
Тема дисертації

на здобуття ступеня доктора філософії за спеціальністю 125 Кібербезпека
шифр і назва спеціальності

впроваджені в діяльності Товариства з обмеженою відповідальністю «Х-СІТУ».
назва установи, організації

Зокрема, використано метод визначення аномального трафіку на основі самоподібності та нечіткий метод визначення аномального трафіку, що на 4-5% покращило виявлення аномального трафіку в порівнянні з раніше використовуваною системою.

Гібридний метод виявлення аномального трафіку в інформаційно-комунікаційних системах (3.4, 4.1) забезпечив зменшення аномального трафіку без додаткових витрат на обладнання.

Зазначені результати дають можливість підвищити рівень інформаційної безпеки підприємства.

Даний акт не є підставою для фінансових взаєморозрахунків.

Голова комісії Директор Ігор ЛИТВІНЧУК

Члени комісії Заступник директора з технічних питань Іван ШТОГРИН

Заступник директора з комерційних питань Руслан КОНИК

Товариство з обмеженою відповідальністю «Х-СІТУ»
ТОВ «Х-СІТУ» ЄДРПОУ 33761335
тел. (0382) 724-777, тел./факс (0382) 71-71-80
29019, м. Хмельницький, вул. Миколи Мазура, 8/1 ОФІС
ceo@x-city.ua, fin@x-city.ua

07.04.2025р. № 0704-1

ДОВІДКА
ПРО ВПРОВАДЖЕННЯ

Результати дисертаційного дослідження *Петляк Наталії Сергіївни* на тему: "*Моделі та методи виявлення аномального трафіку в інформаційно-комунікаційних системах*" на здобуття ступеня доктора філософії за спеціальністю 125 – *Кібербезпека* впроваджені в діяльності Товариства з обмеженою відповідальністю «Х-СІТУ».

Зокрема, використано метод визначення аномального трафіку на основі самоподібності та нечіткий метод визначення аномального трафіку, що на 4-5% покращило виявлення аномального трафіку в порівнянні з раніше використовуваною системою.

Гібридний метод виявлення аномального трафіку в інформаційно-комунікаційних системах (3.4, 4.1) забезпечив зменшення аномального трафіку без додаткових витрат на обладнання.

Зазначені результати дають можливість підвищити рівень інформаційної безпеки підприємства.

Директор Товариства з
обмеженою відповідальністю «Х-СІТУ» Ігор ЛИТВІНЧУК

ТОВАРИСТВО З ОБМЕЖЕНОЮ ВІДПОВІДАЛЬНІСТЮ * Х-СІТУ * №1 * 33761335 * Україна м.Хмельницький

ДОДАТОК Б

ПРОГРАМНИЙ КОД

```
// Signature.hpp
#pragma once
#include <string>
#include <sstream>
#include <ctime>
#include <queue>
#include <vector>

struct Signature {
    long ipSourse;
    long ipDst;
    long portSourse;
    long portDst;
    long protocol;
    long intens;
    long time;
    long mac;
    long size;

    Signature() = default;
    Signature(const std::string& csvLine);

    std::string toString() const;
    bool operator==(const Signature& other) const;
};

long ipToLong(const std::string& ip);
```

```
std::string longToIp(long ip);
long parseTime(const std::string& dateTime);

// Signature.cpp
#include "Signature.hpp"
#include <iomanip>
#include <iostream>
#include <sstream>
#include <vector>
#include <stdexcept>

Signature::Signature(const std::string& csvLine) {
    std::stringstream ss(csvLine);
    std::string part;
    std::vector<std::string> parts;

    while (std::getline(ss, part, ';')) {
        parts.push_back(part);
    }

    if (parts.size() < 9) throw std::runtime_error("Invalid CSV format");

    ipSourse = ipToLong(parts[0]);
    ipDst = ipToLong(parts[1]);
    portSourse = std::stol(parts[2]);
    portDst = std::stol(parts[3]);
    protocol = std::stol(parts[4]);
    intens = std::lround(std::stod(parts[5]));
    time = parseTime(parts[6]);
    mac = parts[7].empty() ? 0 : std::stol(parts[7], nullptr, 16);
```

```
    size = std::stol(parts[8]);
}

std::string Signature::toString() const {
    std::ostringstream oss;
    std::tm tm{};
    tm = *std::localtime(reinterpret_cast<const time_t*>(&time));
    char buf[20];
    std::strftime(buf, sizeof(buf), "%d.%m.%Y %H:%M", &tm);
    oss << longToIp(ipSourse) << ";"
        << longToIp(ipDst) << ";"
        << portSourse << ";"
        << portDst << ";"
        << protocol << ";"
        << intens << ";"
        << buf << ";"
        << (mac == 0 ? "" : std::hex << mac) << ";"
        << size;
    return oss.str();
}

bool Signature::operator==(const Signature& other) const {
    return ipSourse == other.ipSourse &&
           ipDst == other.ipDst &&
           portSourse == other.portSourse &&
           portDst == other.portDst &&
           protocol == other.protocol &&
           intens == other.intens &&
           time == other.time &&
           mac == other.mac &&
```

```
        size == other.size;
}

long ipToLong(const std::string& ip) {
    int b1, b2, b3, b4;
    sscanf(ip.c_str(), "%d.%d.%d.%d", &b1, &b2, &b3, &b4);
    return (b1 << 24) | (b2 << 16) | (b3 << 8) | b4;
}

std::string longToIp(long ip) {
    return std::to_string((ip >> 24) & 0xFF) + "." +
        std::to_string((ip >> 16) & 0xFF) + "." +
        std::to_string((ip >> 8) & 0xFF) + "." +
        std::to_string(ip & 0xFF);
}

long parseTime(const std::string& dateTime) {
    std::tm tm{};
    std::istringstream ss(dateTime);
    ss >> std::get_time(&tm, "%d.%m.%Y %H:%M");
    return std::mktime(&tm);
}

// PriorityQueueSignature.hpp
#pragma once
#include "Signature.hpp"
#include <queue>
#include <vector>

struct CompareSignatures {
```

```
    bool operator()(const Signature& a, const Signature& b) const {
        return a.intens < b.intens;
    }
};

using PriorityQueueSignature = std::priority_queue<Signature,
std::vector<Signature>, CompareSignatures>;

// main.cpp
#include "Signature.hpp"
#include "PriorityQueueSignature.hpp"
#include <fstream>
#include <iostream>
#include <string>

int main() {
    std::ifstream file("input.csv");
    if (!file.is_open()) {
        std::cerr << "Unable to open file" << std::endl;
        return 1;
    }

    PriorityQueueSignature pq;
    std::string line;
    while (std::getline(file, line)) {
        try {
            Signature sig(line);
            pq.push(sig);
        } catch (const std::exception& e) {
            std::cerr << "Error parsing line: " << e.what() << std::endl;
```

```
        }
    }

    file.close();

    while (!pq.empty()) {
        std::cout << pq.top().toString() << std::endl;
        pq.pop();
    }

    return 0;
}
```

ДОДАТОК В
СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА

Наукові праці, в яких опубліковані основні наукові результати дисертації:

1. Тітова В.Ю., Кльоц Ю.П., Петляк Н.С., Капустян М.В.. Fuzzy inference subsystem for classifying threats to computer information. Вимірювальна та обчислювальна техніка в технологічних процесах. 2022. № 1. С. 57-61
2. Кльоц Ю.П., Петляк Н.С.. Виявлення аномального трафіку у загальнодоступних комп'ютерних мережах. Вимірювальна та обчислювальна техніка в технологічних процесах. 2022. № 3. С. 79-86. DOI: 10.31891/2219-9365-2022-71-3-9
3. Мостовий С.В., Петляк Н.С., Голота І.О.. Дослідження ефективності інструментів виявлення і запобігання вторгнень на вузли в корпоративних мережах. Вимірювальна та обчислювальна техніка в технологічних процесах. 2023. № 2. С. 5-8. DOI: 10.31891/2219-9365-2023-74-1
4. Nataliia Petliak, Yuliia Khokhlachova. METHOD OF ANALYSIS OF OUTGOING TRAFFIC PACKAGE SIGNATURES. Захист інформації. 2024. № 1. С. 179-187. DOI: 10.18372/2410-7840.26.18841
5. Петляк Н.С. Аналіз моделей виявлення аномалій трафіку в сучасних інформаційно-комунікаційних системах та мережах. Вимірювальна та обчислювальна техніка в технологічних процесах. 2025. № 1. С. 180-186. DOI: 10.31891/2219-9365-2025-81-21
6. Петляк Н.С.. Гібридний метод та система виявлення аномального трафіку в інформаційно-комунікаційних системах. Вісник Хмельницького національного університету. Серія: Технічні науки. 2025. №2. С. 561-569. DOI: 10.31891/2307-5732-2025-349-82.

Наукові праці, які засвідчують апробацію матеріалів дисертації:

7. Klots Y., Titova V., Petliak N., Cheshun V., Salem A.-B.M.. Research of the Neural Network Module for Detecting Anomalies in Network Traffic. 3rd International Workshop on Intelligent Information Technologies and Systems of Information Security (IntelITSIS 2022), Khmelnytskyi, 23-25 March 2022. Vol. 3156, P. 378 - 389. ISSN 16130073 (SCOPUS)

8. Petliak N., Klots Y., Titova V., Cheshun V., Boyarchuk A.. Signature-based Approach to Detecting Malicious Outgoing Traffic. 4th International Workshop on Intelligent Information Technologies and Systems of Information Security (IntellTSIS 2023), Khmelnytskyi, 22-23 March 2023. Vol. 3373. P. 486 - 506. ISSN 16130073 (SCOPUS)

9. Klots Y., Petliak N., Titova V.. Evaluation of the efficiency of the system for detecting malicious outgoing traffic in public networks. 13th International Conference on Dependable Systems, Services and Technologies (DESSERT 2023), Athens, 13-15 October 2023. DOI: 10.1109/DESSERT61349.2023.10416502 (IEEE)

10. Klots Y., Petliak N., Martsenko S., Tymoshchuk V., Bondarenko I.. Machine Learning system for detecting malicious traffic generated by IoT devices. nd International Workshop on Computer Information Technologies in Industry 4.0 (CITI 2024), Ternopil, 12 June 2024. Vol. 3742. P. 97 - 110. ISSN 16130073 (SCOPUS)

11. Titova V., Klots Y., Petliak N., Cheshun V., Salem A.-B.M.. Detection of network attacks in cyber-physical systems using a rule-based logical neural network. 1st International Workshop on Intelligent and CyberPhysical Systems (ICyberPhyS 2024), Khmelnytskyi, 28 June 2024. Vol. 3736. P. 255 - 268. ISSN 16130073 (SCOPUS)

12. Petliak N., Klots Y., Titova V., Salem A.-B.M.. Attack detection system based on network traffic analysis by means of fuzzy inference. 1st International Workshop on Advanced Applied Information Technologies (AdvAIT 2024), Khmelnytskyi, 5 December 2024. Vol. 3899. P. 201-213.ISSN 16130073 (SCOPUS)

13. Петляк Н.С., Кльоц Ю.П., Хохлачова Ю.Є.. Виявлення аномального трафіку у загальнодоступних комп’ютерних мережах. Інформаційна, функційна і кібербезпека: тези доп. II студ. наук.-техн. к. (м. Харків, 30 листопада – 1 грудня

2022 р.) Х., 2022. С. 45-46

14. Петляк Н.С., Кльоц Ю.П.. Підхід до аналізу вихідного трафіку на основі сигнатур. Інформаційна безпека та комп’ютерні технології: тези доп. VI міжнар. наук.-пр. конф. (м. Кропивницький, 20-21 квітня 2023 р.) Кр., 2023. С. 3-4

15. Петляк Н.С., Кльоц Ю.П., Хохлачова Ю.Є.. Підхід до аналізу вихідного трафіку. Безпека інформаційних технологій: тези доп. XII міжнар. наук.-техн. конф. (м. Ужгород, 2-4 травня 2023 р.). К., 2023. С. 97-99

16. Петляк Н.С., Кльоц Ю.П., Тітова В.Ю., Чешун В.М. Виявлення зловмисного вихідного трафіку мережі на основі нечіткого логічного висновку. Захист інформації і безпека інформаційних систем: тези доп. IX міжнар. наук.-техн. конф. (м. Львів, 25-26 травня 2023 р.) Л., 2023. С. 33-35

17. Петляк Н.С., Кльоц Ю.П.. Метод ідентифікації вторгнень на основі алгоритму визначення самоподібності трафіку та алгоритмів нечіткої логіки. Безпека інформаційних технологій: тези доп. XIII міжнар. наук.-техн. конф. (м. Львів, 9-11 травня 2024 р.) Л., 2024. С. 117-119.

Відомості про апробацію результатів дисертації

1. Інформаційна, функційна і кібербезпека: тези доп. II студ. наук.-техн. к. (м. Харків, 30 листопада – 1 грудня 2022 р.) Х., 2022. С. 45-46

2. 4th International Workshop on Intelligent Information Technologies and Systems of Information Security (Khmelnytskyi, 22 March 2023). Khm., 2023. P. 486-506

3. 13th International Conference on Dependable Systems, Services and Technologies (Athens, 13 October 2023). Ath., 2023.

4. 2nd International Workshop on Computer Information Technologies in Industry 4.0 (Ternopil, 12 June 2024). T., 2024. P. 97-110

5. 1st International Workshop on Intelligent and CyberPhysical Systems (Khmelnytskyi, 28 June 2024). Khm., 2024. P. 255-268

6. 1st International Workshop on Advanced Applied Information Technologies

(Khmelnytskyi, 5 December 2024). Khm., 2024. P. 201-213

7. 3rd International Workshop on Intelligent Information Technologies and Systems of Information Security (Khmelnytskyi, 23 March 2022). Khm., 2022. P. 378-389

8. Інформаційна безпека та комп'ютерні технології: тези доп. VI міжнар. наук.-пр. конф. (м. Кропивницький, 20-21 квітня 2023 р.) Кр., 2023 . С. 3-4

9. Безпека інформаційних технологій: тези доп. XII міжнар. наук.-техн. конф. (м. Ужгород, 2-4 травня 2023 р.). К., 2023. С. 97-99

10. Захист інформації і безпека інформаційних систем: тези доп. IX міжнар. наук.-техн. конф. (м. Львів, 25-26 травня 2023 р.) Л., 2023. С. 33-35

11. Безпека інформаційних технологій: тези доп. XIII міжнар. наук.-техн. конф. (м. Львів, 9-11 травня 2024 р.) Л., 2024. С. 117-119