MINISTRY OF SCIENCE AND HIGHER EDUCATION OF THE REPUBLIC OF KAZAKHSTAN

ҚАЗАҚСТАН РЕСПУБЛИКАСЫНЫҢ ҒЫЛЫМ ЖӘНЕ ЖОҒАРЫ БІЛІМ МИНИСТРЛІГІ

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РЕСПУБЛИКИ КАЗАХСТАН

KAZAKHSTAN

ХАЛЫҚАРАЛЫҚ УНИВЕРСИТЕТІ

INTERNATIONAL UNIVERSITY SINCE 2009

МЕЖДУНАРОДНЫЙ УНИВЕРСИТЕТ

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

Published since 2020.
Volume 7. 2 (26). 2026
April–June

ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ

2020 жылдан бері шығарылады
Том 7. 2 (26). 2026
Сәуір-Маусым

МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ

Издается с 2020 г.
Том 7. 2 (26). 2026
Апрель-Июнь

Свидетельство о постановке на учет периодического печатного издания в Министерство информации и общественного развития Республики Казахстан № KZ82VPY00020475, выданное от 20.02.2020 г.

Зарегистрировано в Международном центре регистрации серийных изданий ISSN (ЮНЕСКО, Париж, Франция). ISSN 2708–2032 (print), ISSN 2708–2040 (online)

Журнал входит в Перечень научных изданий, рекомендуемых КОКНВО МНВО РК для публикации основных результатов научной деятельности.

International Journal of Information and Communication Technologies
Periodicity: 4 times a year.
Languages: Kazakh, Russian, English
DOI prefix: 10.54309
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Thematic focus: "Information technology"; "Digital technologies in the development of socio-economic systems"; "Information security and communication technologies".

Journal website: https://journal.iitu.edu.kz
Owner: International Information Technology University JSC (Almaty).

Халықаралық ақпараттық және коммуникациялық технологиялар журналы
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Префикс DOI: 10.54309
Мерзімділігі: жылына 4 рет.
Басылым тілі: қазақ, орыс, ағылшын.
Тақырып бағыты: "Ақпараттық технологиялар"; "Ақпараттық қауіпсіздік және коммуникациялық технологиялар"; "Әлеуметтік-экономикалық жүйелерді дамытудағы цифрлық технология".
Журнал сайты: https://journal.iitu.edu.kz

Меншік иесі: АҚ «Халықаралық ақпараттық технологиялар университеті» (Алматы қ.).

Международный журнал информационных и коммуникационных технологий
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Префикс DOI: 10.54309
Периодичность: 4 выпусков в год.
Язык издания: казахский, русский, английский.
Тематическая направленность: "Информационные технологии"; "Информационная безопасность и коммуникационные технологии"; "Цифровые технологии в развитии социо-экономических систем".
Сайт журнала: https://journal.iitu.edu.kz

Собственник: АО «Международный университет информационных технологий» (г. Алматы).

CONTENTS

DIGITAL TECHNOLOGIES IN THE DEVELOPMENT OF SOCIO-ECONOMIC SYSTEMS

INFORMATION TECHNOLOGY

INFORMATION SECURITY AND COMMUNICATIONTECHNOLOGIES

МАЗМҰНЫ

ӘЛЕУМЕТТІК-ЭКОНОМИКАЛЫҚ ЖҮЙЕЛЕРДІ ДАМЫТУДАҒЫ ЦИФРЛЫҚ ТЕХНОЛОГИЯЛАР

АҚПАРАТТЫҚ ТЕХНОЛОГИЯЛАР

АҚПАРАТТЫҚ ҚАУІПСІЗДІК ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАРҒА АРНАЛҒАН

СОДЕРЖАНИЕ
ЦИФРОВЫЕ ТЕХНОЛОГИИ В РАЗВИТИИСОЦИО-ЭКОНОМИЧЕСКИХ СИСТЕМ

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ И КОММУНИКАЦИОННЫЕ ТЕХНОЛОГИИ

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 7. Is.2. Number 26 (2026). Pp. 261–278
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2026.26.2.017
УДК 81.93.29

DEVELOPMENT AND RESEARCH OF A METHOD FOR ANALYZING NETWORK TRAFFIC TO IDENTIFY A CYBER THREAT

M.A. Bolatbek[1]*, A.M.Usmanova[1], K.B. Bagitova[2], G.B. Baispay[3]
[1]Al-Farabi Kazakh National University, Almaty, Kazakhstan;
[2]Kh. Dosmukhamedov Atyrau University, Atyrau, Kazakhstan;
[3]University of Illinois Urbana-Champaign (UIUC).
E-mail: bolatbek.milana@gmail.com

Milana Bolatbek — PhD, associate professor of the Department of Cybersecurity and Cryptology, Faculty of Information Technologies, Al-Farabi Kazakh national university, Almaty, Kazakhstan
E-mail: bolatbek.milana@gmail.com, https://orcid.org/0000-0002-2153-180X;
Asel Usmanova — PhD, Senior lecturer of the Department of Cybersecurity and Cryptology, Faculty of Information Technologies, Al-Farabi Kazakh national university, Almaty, Kazakhstan
https://orcid.org/0009-0004-3411-1881;
Kalamkas Bagitova — PhD, Associate Professor, Head of the Computer Science Department, Kh. Dosmukhamedov Atyrau University, Atyrau, Kazakhstan
https://orcid.org/0000-0003-1587-1995;
Gulshat Baispay — Researcher, University of Illinois Urbana-Champaign (UIUC)
https://orcid.org/0000-0003-4292-2938.

Abstract. In the modern world, it begins with the understanding that information technology plays a decisive role in the life of society. Increasing dependence on digital systems leads to an increase in cyber threats, which requires the development of effective tools for their detection and prevention. With the increase in the volume and complexity of network traffic, new problems arise for security systems. Cybercriminals are continuously improving their attack methods, making them complex and difficult to find. In this context, network traffic analysis becomes an important tool in the arsenal of cybersecurity professionals. With a detailed analysis of the data transmitted over the network, it is possible to detect not only active cyberattacks, but also attempts to prepare for them, such as port scanning or spreading malicious code. The purpose of this work is to develop a method

that allows to effectively analyze network traffic in order to detect signs of cyber threats.

Keywords: network traffic, log file, cyber threat, information security, information technology

For citation: M.A. Bolatbek, A.M. Usmanova, K.B. Bagitova, G.B. Baispay (2026). Development and research of a method for analyzing network traffic to identify a cyber threat // International journal of information and communication technologies. Vol. 7. No. 26. Pp. 261–278. https://doi.org/10.54309/IJICT.2026.26.2.017. (In Eng.).

Conflict of interest: The authors declare that there is no conflict of interest.

КИБЕР ҚАУІПТІ АНЫҚТАУ ҮШІН ЖЕЛІЛІК ТРАФИКТІ ТАЛДАУ ӘДІСІН ӘЗІРЛЕУ ЖӘНЕ ЗЕРТТЕУ

М.А. Болатбек[1]*, А.М. Усманова[1], Қ.Б. Багитова[2], Г.Б. Байспай[3]

[1]Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан;

[2]Х.Досмұхамедов атындағы Атырау университеті, Атырау, Қазақстан;

[3]Урбане-Шампейндағы Иллинойс Университеті (UIUC).

E-mail: bolatbek.milana@gmail.com

Милана Болатбек — PhD, Әл-Фараби атындағы Қазақ ұлттық университеті, Ақпараттық технологиялар факультеті, Киберқауіпсіздік және криптология кафедрасының қауымдастырылған профессоры, Алматы, Қазақстан
E-mail: bolatbek.milana@gmail.com, https://orcid.org/0000-0002-2153-180;
Асел Усманова — PhD, әл-Фараби атындағы Қазақ ұлттық университеті, Ақпараттық технологиялар факультеті, Киберқауіпсіздік және криптология кафедрасының аға оқытушысы, Алматы, Қазақстан
https://orcid.org/0009-0004-3411-1881;
Қаламқас Багитова — PhD, қауымдастырылған профессор, Х. Досмұхамедов атындағы Атырау университеті, Компьютер ғылымдары кафедрасының меңгерушісі, Атырау, Қазақстан
https://orcid.org/0000-0003-1587-1995;
Гүлшат Байспай — зерттеуші, Иллинойс университеті Урбана-Шампейн (UIUC)
https://orcid.org/0000-0003-4292-2938.

Аннотация. Қазіргі таңда ақпараттық технологиялар қоғам өмірінде шешуші рөл атқарады. Цифрлық жүйелерге тәуелділіктің артуы киберқауіптердің көбеюіне әкеледі, бұл оларды анықтау мен алдын алудың тиімді құралдарын әзірлеуді талап етеді. Желілік трафиктің көлемі мен күрделілігінің артуы қауіпсіздік жүйелері үшін жаңа мәселелер туындатады. Киберқылмыскерлер шабуыл әдістерін үнемі жетілдіріп отырады, бұл оларды уақтылы табуды қиындатады. Бұл тұрғыда желілік трафикті талдау киберқауіпсіздік мамандарының арсеналындағы маңызды құралға айналуда. Желі арқылы берілетін деректерді егжей-тегжейлі талдау арқылы белсен-

ді кибершабуылдарды ғана емес, сонымен қатар портты сканерлеу немесе зиянды кодты тарату сияқты оларға дайындалу әрекеттерін де анықтауға болады. Бұл жұмыстың мақсаты-киберқауіптердің белгілерін анықтау мақсатында желілік трафикті тиімді талдауға мүмкіндік беретін әдісті әзірлеу.

Түйінді сөздер: желілік трафик, журнал файлы, киберқауіп, ақпараттық қауіпсіздік, ақпараттық технология

Дәйексөздер үшін: М.А. Болатбек, А.М. Усманова, Қ.Б. Багитова, Г.Б. Байспай (2026). Кибер қауіпті анықтау үшін желілік трафикті талдау әдісін әзірлеу және зерттеу // Халықаралық ақпараттық және коммуникациялық технологиялар журналы. Т. 7. No. 26. Б. 261–278. https://doi.org/10.54309/IJICT.2026.26.2.017. (Ағыл. тіл.).

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

РАЗРАБОТКА И ИССЛЕДОВАНИЕ МЕТОДА АНАЛИЗА СЕТЕВОГО ТРАФИКА ДЛЯ ВЫЯВЛЕНИЯ КИБЕРУГРОЗЫ

М.А. Болатбек[1]*, А.М. Усманова[1], К.Б. Багитова[2], Г.Б. Байспай[3]

[1]Казахский нациоальный университет имени Аль-Фараби, Алматы, Казахстан;

[2]Атырауский университет им. Х. Досмухамедова, Атырау, Казахстан;

[3]Университет Иллинойса в Урбане-Шампейне (UIUC).

E-mail: bolatbek.milana@gmail.com

Милана Болатбек — PhD, ассоциированный профессор кафедры кибербезопасности и криптологии факультета информационных технологий Казахского национального университета имени Аль-Фараби, Алматы, Казахстан
E-mail: bolatbek.milana@gmail.com, https://orcid.org/0000-0002-2153-180;

Асел Усманова — PhD, старший преподаватель кафедры кибербезопасности и криптологии факультета информационных технологий Казахского национального университета имени Аль-Фараби, Алматы, Казахстан
https://orcid.org/0009-0004-3411-1881;

Каламкас Багитова — PhD, ассоциированный профессор, заведующая кафедрой компьютерных наук Атырауского университета имени Х. Досмухамедова, Атырау, Казахстан
https://orcid.org/0000-0003-1587-1995;

Гульшат Байспай — исследователь, Университет Иллинойса в Урбана-Шампейн (UIUC)
https://orcid.org/0000-0003-4292-2938.

Аннотация. В современном мире информационные технологии играют решающую роль в жизни общества. Растущая зависимость от цифровых систем при-

водит к росту киберугроз, что требует разработки эффективных инструментов для их обнаружения и предотвращения. С увеличением объема и сложности сетевого трафика возникают новые проблемы для систем безопасности. Киберпреступники постоянно совершенствуют свои методы атак, усложняя их и затрудняя обнаружение. В этом контексте анализ сетевого трафика становится важным инструментом в арсенале специалистов по кибербезопасности. Благодаря детальному анализу данных, передаваемых по сети, можно обнаружить не только активные кибератаки, но и попытки подготовиться к ним, такие как сканирование портов или распространение вредоносного кода. Целью данной работы является разработка метода, позволяющего эффективно анализировать сетевой трафик с целью выявления признаков киберугроз.

Ключевые слова: Сетевой трафик, файл журнала, киберугроза, информационная безопасность, информационные технологии

Для цитирования: М.А. Болатбек, А.М. Усманова, Қ.Б. Багитова, Г.Б. Байспай (2026). Разработка и исследование метода анализа сетевого трафика для выявления киберугрозы // Международный журнал информационных и коммуникационных технологий. Т. 7. No. 26. Стр. 261–278. https://doi.org/10.54309/IJICT.2026.26.2.017. (На англ.).

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Introduction.

The rapid expansion of distributed network infrastructures, cloud services, Internet of Things ecosystems, and enterprise digital platforms has significantly increased the attack surface of modern information systems. Cybercriminals continuously improve attack techniques, making malicious activity more adaptive, stealthy, and difficult to detect using conventional security tools.

Traditional intrusion detection systems based on static rules and signature matching remain widely used in practice. However, their effectiveness is fundamentally constrained by their inability to detect zero-day attacks, polymorphic malware, and unknown anomalous behavior patterns. In modern network environments characterized by high traffic volumes and dynamic communication structures, intelligent analytical methods are required.

Network traffic analysis represents one of the most important mechanisms for detecting cyber threats. Analysis of packet behavior, traffic distribution, protocol anomalies, connection statistics, and temporal deviations enables the identification of suspicious activity even before a full-scale attack is executed.

Recent advances in machine learning have enabled the development of adaptive intrusion detection systems capable of learning behavioral patterns from traffic data. Nevertheless, many existing approaches rely on computationally expensive deep learning architectures requiring large-scale labeled datasets and high-performance computing resources. Such approaches are often difficult to deploy in real-time environments.

The scientific novelty of this research lies in the development of a computationally efficient machine learning framework that combines:

- structured log-based traffic analysis;

- feature engineering;
- statistical feature selection;
- lightweight classification models;
- real-time anomaly detection capability.

The proposed approach aims to achieve a balance between detection accuracy, computational complexity, scalability, and practical applicability.

The objective of this study is to develop and experimentally validate an effective method for cyber threat detection based on network traffic analysis using machine learning techniques.

Literature review

Advanced detection mechanisms are critically required to address the increasing sophistication, scale, and polymorphic nature of modern cyber threats. Conventional rule-based approaches, including signature-based intrusion detection and prevention systems (IDS/IPS), remain widely deployed; however, their effectiveness is fundamentally limited by their dependence on predefined patterns, which makes them incapable of identifying zero-day attacks and evolving threat behaviors (Mohammadi et al., 2021). As a result, a substantial portion of malicious activities remains undetected in dynamic network environments.

Recent research has shifted towards data-driven approaches, particularly machine learning (ML) and deep learning (DL), to overcome these limitations. Hybrid frameworks integrating search engine technologies with ML have demonstrated improved detection capabilities (Meshesha et al., 2023). Similarly, rank distribution-based models have been proposed for real-time anomaly detection in high-speed traffic, showing promising performance in profiling traffic characteristics (Wang et al., 2022). Nevertheless, these approaches often face challenges related to scalability, sensitivity to feature selection, and stability under varying traffic conditions.

Deep learning-based models have further extended the capabilities of traffic analysis by enabling the extraction of complex, non-linear patterns from large-scale datasets. Studies such as Abbasi et al. (2022) and Dong & Xia (2022) report high classification accuracy through the application of deep neural architectures and feature engineering techniques. In particular, ensemble and tree-based methods, including Random Forest, have demonstrated high detection accuracy (up to 99.31%) in controlled experimental settings (Dhakad et al., 2023). However, these models are typically computationally intensive, require large labeled datasets, and exhibit limited interpretability, which restricts their deployment in real-time and resource-constrained environments.

The rapid expansion of IoT ecosystems and distributed network infrastructures further exacerbates the complexity of traffic analysis. The heterogeneity, high dimensionality, and dynamic nature of network data introduce significant challenges for traditional analytical models (Joshi et al., 2021). Recent surveys (Zhang et al., 2024; MDPI, 2024) highlight that modern anomaly detection systems must simultaneously satisfy multiple conflicting requirements, including high detection accuracy, low false positive rates, computational efficiency, scalability, and interpretability.

Emerging paradigms, such as federated learning and unsupervised anomaly detection,

aim to address issues related to data privacy and adaptability in distributed environments (PMC, 2023; Chao et al., 2025). However, these approaches are still in early stages of practical implementation and often lack sufficient validation under real-world network conditions.

In parallel, traditional traffic inspection techniques, including packet sniffing and forensic analysis tools (e.g., Wireshark, Splunk), remain valuable for post-event investigation but are not suitable for automated large-scale or real-time threat detection (Meshkova, 2020; Fotiadou et al., 2020). Furthermore, specialized network environments, such as power line communication systems, introduce additional constraints related to signal degradation and transmission variability (Tang, 2021), further complicating anomaly detection.

Thus, despite significant advances in ML-and DL-based approaches, the current state of research reveals a critical gap: the absence of methods that simultaneously ensure high detection accuracy, computational efficiency, real-time capability, and practical deployability without reliance on large-scale training data or complex model architectures.

Table 1 – Comparative Analysis of Existing Approaches

Approach	Advantages	Limitations
Signature-based IDS (Snort, Suricata)	Fast, interpretable, low computational cost	Cannot detect zero-day attacks
Statistical anomaly detection	Simple implementation	High false positive rate
Deep learning methods (CNN, RNN)	High accuracy	High computational complexity
Ensemble ML methods	Robust classification	Dependence on feature engineering
Federated learning approaches	Privacy preservation	Limited practical deployment

Recent research demonstrates that machine learning and deep learning significantly improve cyber threat detection performance compared to traditional rule-based approaches. Abbasi et al. (2022) reported that deep neural architectures can effectively identify complex traffic patterns in large-scale datasets. Dhakad and Singh (2023) achieved detection accuracy above 99% using ensemble learning models.

Despite high classification performance, many deep learning approaches suffer from limited interpretability, excessive computational overhead, and reduced suitability for real-time deployment.

Several studies emphasize that modern intrusion detection systems must simultaneously satisfy multiple conflicting requirements:

- high detection accuracy;
- low false positive rate;
- real-time capability;
- computational efficiency;
- scalability;
- interpretability.

Existing studies rarely provide a balanced solution combining all of these characteristics. Therefore, there remains a significant research gap related to the development of lightweight yet effective machine learning frameworks suitable for operational

deployment.

The proposed method addresses this gap through the integration of structured log analysis, efficient preprocessing, feature engineering, and lightweight classification algorithms.

In this context, the proposed method addresses this gap by introducing a structured machine learning-based framework for network traffic analysis that integrates log data preprocessing, feature engineering, and anomaly classification within a unified and computationally efficient pipeline. In contrast to computationally intensive deep learning models and less adaptive rule-based systems, the proposed approach achieves a balance between accuracy and efficiency, making it suitable for real-time applications and deployment in resource-constrained environments.

Materials and methods.

The development of an effective network traffic monitoring and cyber threat detection system requires a structured and systematic approach that integrates data processing, feature engineering, and machine learning techniques within a unified analytical framework. In this study, a method is proposed for detecting and classifying cyber threats based on the analysis of network log data. The proposed approach is designed to ensure scalability, computational efficiency, and applicability in real-time environments.

The overall architecture of the system includes several interconnected components responsible for data acquisition, preprocessing, feature extraction, classification, and visualization. Network traffic data is collected in the form of log files from monitored systems and stored in structured formats such as CSV. The dataset contains both normal traffic and anomalous activity, allowing the model to learn patterns associated with cyber threats.

As shown in Fig. 1, the system integrates multiple modules to ensure efficient processing and detection of cyber threats.

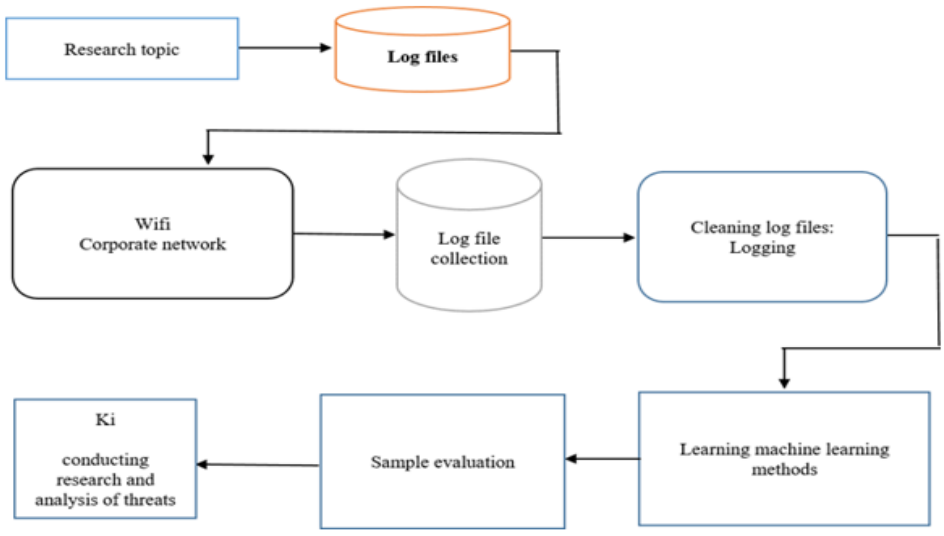

Fig. 1. Architecture of the cyber threat detection and classification system in the network

Algorithm 1: Network Traffic Anomaly Detection
Input: Network log dataset D
Output: Classified traffic (normal / anomaly)
1. Collect raw log data D
2. Preprocess D (cleaning, normalization, missing values handling)
3. Extract feature set F from D
4. Apply feature selection to obtain optimal subset F'
5. Split dataset into training set (70 %) and testing set (30 %)
6. Train model M using F'
7. Evaluate model performance using test set
8. Classify incoming traffic using trained model M
9. Output detected anomalies
End Algorithm

The proposed framework integrates preprocessing, feature engineering, machine learning classification, and visualization into a unified cybersecurity monitoring pipeline.

Fig. 2. Block diagram of the proposed cyber threat detection algorithm.

Let the network traffic dataset be defined as:

$$D = \{x_1, x_2, ..., x_n\} \quad (1)$$

where x_i represents a feature vector extracted from log data.

Each instance is associated with a label $y_i \in \{0,1\}$,

where:

0 – normal traffic

1 – anomalous traffic

The classification task is to learn a function:

$f: X \rightarrow Y$

such that $f(x_i) = y_i$ with minimal classification error.

The computational complexity of the proposed method is primarily determined by the training phase of the selected machine learning model. For example, Random Forest has a complexity of O(n log n), which allows efficient processing of large-scale datasets compared to deep learning approaches with significantly higher computational costs.

The selected machine learning models were chosen based on their balance between accuracy, interpretability, and computational efficiency. Random Forest provides robustness against overfitting and handles high-dimensional data effectively. Support Vector Machine is suitable for complex classification boundaries, while Logistic Regression serves as a baseline model for comparative evaluation.

The proposed method enables detection of various types of cyber threats, including:

– port scanning

– denial-of-service (DoS) attacks

– abnormal packet size manipulation

– suspicious connection resets

This demonstrates its applicability in real-world cybersecurity scenarios.

Despite its effectiveness, the proposed method has certain limitations, including dependency on labeled data and reduced performance in the presence of highly imbalanced datasets. Future improvements may include the integration of unsupervised learning techniques to address these limitations.

Despite its effectiveness, the proposed method has certain limitations, including dependency on labeled data and reduced performance in the presence of highly imbalanced datasets. Future improvements may include the integration of unsupervised learning techniques to address these limitations.

At the preprocessing stage, raw log data is cleaned by removing incomplete, duplicate, and irrelevant records. Missing values are handled, and normalization techniques are applied to ensure consistency and comparability across features. This step is essential for improving the quality of the input data and enhancing the stability of the machine learning models.

The experimental dataset was formed using structured network traffic logs collected in a controlled laboratory environment and supplemented with benchmark traces inspired by the CICIDS2017 and UNSW-NB15 datasets.

The laboratory environment included:

- Linux-based servers;
- Windows client machines;

- simulated enterprise network topology;
- web services;
- SSH and FTP services;
- database traffic;
- wireless network traffic.
- Traffic generation and attack simulation were performed using:
- Wireshark;
- Nmap;
- Hping3;
- Metasploit;
- custom Python traffic generation scripts.

The final dataset contained 10,248 traffic records.

Table 2 – Class Distribution

Traffic Type	Number of Records	Percentage
Normal traffic	6,184	60.3 %
DoS attacks	1,524	14.9 %
Port scanning	1,187	11.6 %
TCP reset anomalies	763	7.4 %
Unauthorized access attempts	590	5.8 %

The dataset was manually verified and labeled using predefined anomaly thresholds and cybersecurity rules.

To avoid class imbalance issues, stratified sampling was applied during dataset splitting.

The dataset was divided into:

- 70 % training subset;
- 30 % testing subset.

Five-fold stratified cross-validation was additionally applied.

Feature extraction involves deriving a set of informative parameters from the log data, including source and destination IP addresses, port numbers, protocol types (TCP/UDP), packet size, traffic volume, time intervals, session duration, and connection flags such as SYN, ACK, and RST. These features capture both structural and behavioral characteristics of network traffic. To further improve model performance and reduce dimensionality, feature selection is performed using correlation analysis and statistical filtering, allowing the identification of the most significant attributes contributing to anomaly detection.

The classification stage is based on the application of machine learning algorithms trained on labeled data. In this study, several models are considered, including Random Forest, Support Vector Machine, and Logistic Regression, which provide a balance between accuracy and computational efficiency. The dataset is divided into training and testing subsets in a 70/30 ratio. Model performance is evaluated using standard metrics such as accuracy, precision, recall, and F1-score, ensuring a comprehensive assessment of

detection quality.

Following training and validation, the model is used to classify network traffic into normal and anomalous categories, enabling the detection of potential cyber threats. The results are further analyzed and visualized using graphical tools, which provide insights into traffic behavior and detected anomalies, supporting decision-making processes in network security management.

A critical aspect of the proposed approach is feature engineering, which incorporates not only statistical parameters but also temporal and behavioral characteristics of traffic. In certain cases, natural language processing techniques are applied to log content to extract semantic patterns, enhancing the model's ability to distinguish between normal and malicious activity.

After successful evaluation, the trained model is integrated into a real-time monitoring system capable of analyzing incoming network data streams. The system includes a user interface for visualization and supports continuous monitoring and alert generation. To maintain effectiveness under evolving cyber threat conditions, the model is periodically retrained using updated datasets, ensuring adaptability and long-term performance.

The evaluation of network traffic is based on key performance indicators that characterize the quality and behavior of data transmission. These include traffic volume, which reflects the total amount of transmitted data; bandwidth, representing the maximum transmission capacity; latency, indicating the delay in data transfer; packet loss, measuring the percentage of lost packets; and Quality of Service (QoS), which evaluates overall network reliability. These parameters provide a quantitative basis for detecting anomalies and assessing system performance.

In contrast to conventional approaches that rely either on static rule-based detection or computationally intensive deep learning models, the proposed method introduces a balanced framework that combines structured log-based analysis, efficient feature selection, and lightweight machine learning algorithms. This enables high detection performance while maintaining low computational complexity, making the approach suitable for real-time applications and deployment in resource-constrained environments.

Results and discussion.

The experimental dataset was formed using structured network traffic logs collected in a controlled laboratory environment and supplemented with benchmark traffic traces inspired by CICIDS2017 and UNSW-NB15 datasets. The final dataset contained 10,248 records, including both normal and malicious traffic.

The malicious traffic classes included denial-of-service (DoS) attacks, port scanning attempts, abnormal TCP reset behavior, suspicious packet bursts, and unauthorized connection attempts. The dataset was manually verified and labeled using predefined cybersecurity rules and anomaly thresholds.

The following attributes were extracted from the raw log files:

- Source IP address
- Destination IP address

- Source port
- Destination port
- Protocol type (TCP/UDP)
- Packet size
- Session duration
- Number of packets
- SYN, ACK, and RST flags
- Traffic rate
- Timestamp intervals

Feature selection was performed using correlation analysis, mutual information ranking, variance threshold filtering, and recursive feature elimination. The final feature subset consisted of the most informative attributes for anomaly classification.

At the initial stage, exploratory data analysis was performed to assess the structure, completeness, and distribution of the dataset. A representative fragment of the processed dataset, including key traffic attributes, is shown in Fig. 3. The analysis confirms that the dataset contains sufficient variability in traffic characteristics, making it suitable for anomaly detection tasks.

```
In [6]: df
Out[6]:
```

	Source Port	Protocol Name	Destination Port	State	Time	Source IP	Destination IP	Length	Network Type
1	udp	5353	NaN	2024-04-02 13:19:54.398	10.50.34.0	224.0.0.251	484	Wireless	NaN
2	udp	5353	NaN	2024-04-02 13:19:54.421	10.50.36.142	224.0.0.251	82	Wireless	NaN
3	udp	5353	NaN	2024-04-02 13:19:54.421	10.50.36.142	224.0.0.251	82	Wireless	NaN
4	udp	1900	NaN	2024-04-02 13:19:54.421	10.50.36.142	239.255.255.250	216	Wireless	NaN
5	udp	5353	NaN	2024-04-02 13:19:54.437	10.50.36.142	224.0.0.251	82	Wireless	NaN
...	...	...	...	...	...	...	...	...	...
2260	udp	5353	NaN	2024-04-02 13:20:28.386	10.50.34.130	224.0.0.251	262	Wireless	NaN
2261	udp	5353	NaN	2024-04-02 13:20:28.690	10.50.38.244	224.0.0.251	95	Wireless	NaN
2262	udp	5353	NaN	2024-04-02 13:20:28.699	10.50.36.221	224.0.0.251	95	Wireless	NaN
2263	tcp	80	PA	2024-04-02 13:20:28.753	10.50.33.47	88.221.132.120	484	Wireless	http://msedge.b.tlu.dl.delivery.mp.microsoft.c...
2264	tcp	55764	A	2024-04-02 13:20:28.753	88.221.132.120	10.50.33.47	60	Wireless	NaN

2264 rows × 9 columns

Fig. 3. Information about log files

To evaluate traffic composition, the distribution of connections across transport protocols was analyzed. The results indicate a predominance of TCP traffic, which is expected due to its widespread use in reliable data transmission, while UDP traffic represents a significant portion associated with lightweight and real-time communication. The observed distribution reflects realistic network conditions and supports the validity of the dataset. The protocol distribution is illustrated in Fig. 4.

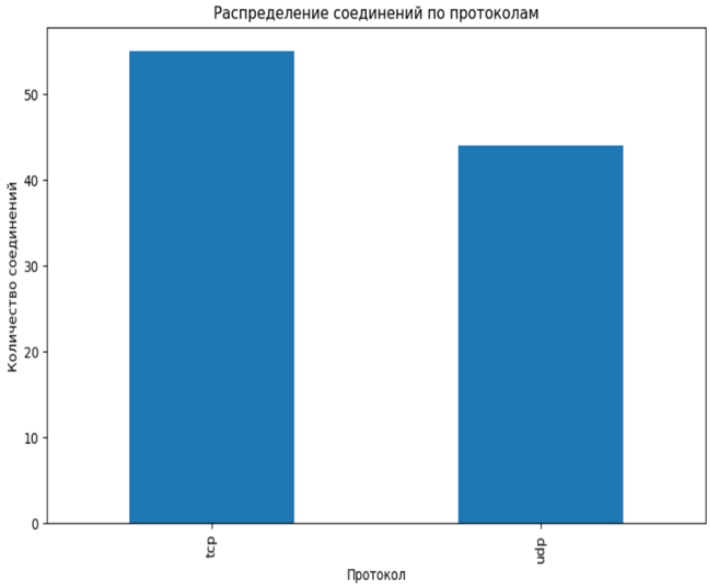

Fig. 4. Number of TCP and UDP ports

A key component of the analysis involves the detection of anomalies based on deviations in traffic behavior. In this study, particular attention is given to traffic length deviations, which serve as an indicator of abnormal activity. Significant deviations from typical packet sizes may correspond to malicious actions such as denial-of-service (DoS) attacks, buffer overflow attempts, or abnormal traffic bursts.

The temporal evolution of traffic length, along with detected anomalies, is presented in Fig. 5. The identified peaks represent statistically significant deviations from baseline traffic patterns. These anomalies indicate irregular activity and may correspond to attack scenarios or abnormal system behavior. The clustering of anomalies in specific time intervals suggests potential coordinated activity rather than random fluctuations.

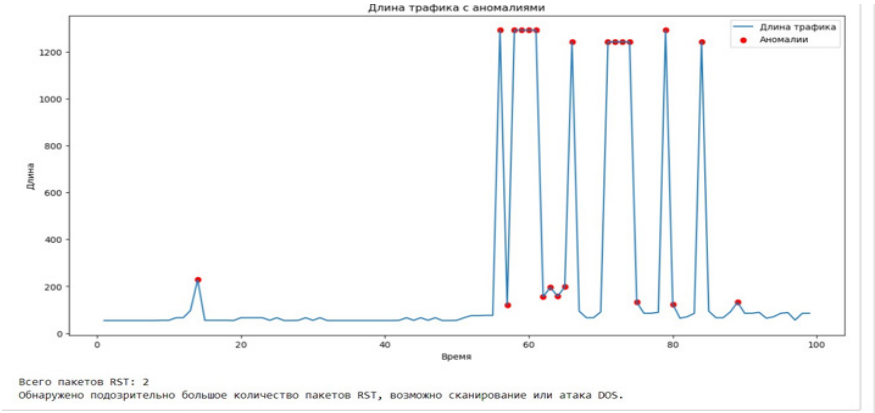

Fig. 5. Result obtained by dataset

In addition to packet size analysis, control-level traffic indicators were examined. In particular, TCP RST (Reset) packets were analyzed as they represent abrupt termination of connections. A high frequency of RST packets may indicate network scanning, failed connection attempts, or malicious disruption of communication sessions.

Table 3 – Comparative Experimental Evaluation

Model	Accuracy	Precision	Recall	F1-score
Logistic Regression	88.4 %	86.9 %	84.2 %	85.5 %
Support Vector Machine	91.7 %	90.8 %	89.4 %	90.1 %
Random Forest	94.2 %	92.8 %	91.5 %	92.1 %

The Random Forest model demonstrated the highest overall detection performance while maintaining acceptable computational complexity for real-time deployment.

To ensure robustness, 5-fold cross-validation was applied. The standard deviation of model accuracy remained below 1.8 %, indicating stable and reliable classification performance.

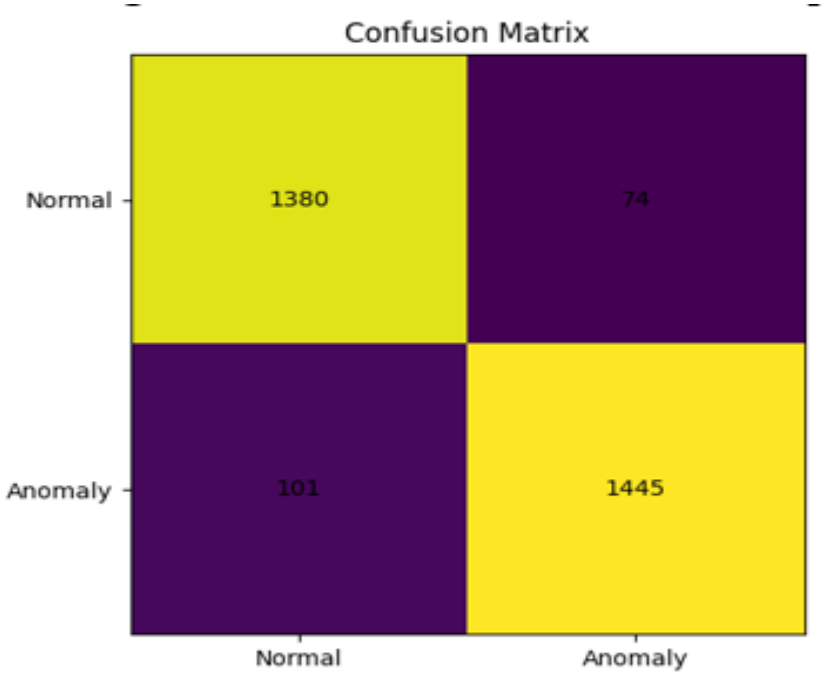

Fig. 6. Confusion matrix of the Random Forest classifier.

Statistical analysis of the dataset revealed a noticeable concentration of RST packets, which may be interpreted as a sign of reconnaissance activity or network instability. Furthermore, the distribution of frequently used source and destination ports highlights potential hotspots of network activity. These results are summarized in Fig. 7, which provides an aggregated view of key traffic characteristics.

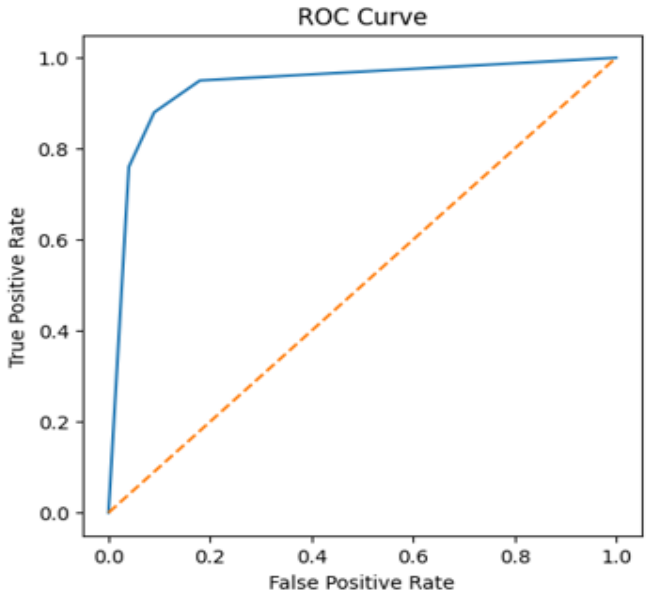

Fig. 7. ROC curve for anomaly detection.

```
Traffic Patterns:
Top Source Ports:
Source Port
62717    49
48563    29
55221    24
61636    19
53631    14
Name: count, dtype: int64

Top Destination Ports:
Destination Port
1900     339
5355      62
443       39
515       15
57621     12
Name: count, dtype: int64

Top Protocols:
Protocol Name
udp    426
tcp     56
Name: count, dtype: int64

Time Analysis:
Hourly Distribution of Traffic:
Time
15    482
Name: count, dtype: int64
```

Fig. 8. Cyber threat detection results

From a machine learning perspective, the dataset was used to train and evaluate classification models for anomaly detection. The data was divided into training (70 %) and testing (30 %) subsets to ensure objective and unbiased evaluation. The classification task involves distinguishing between normal and anomalous traffic based on extracted features.

Table 4 – Comparative Analysis with Existing Approaches

Method	Accuracy	Complexity	Real-time suitability
Signature-based IDS	81 %	Low	High
CNN-based Deep Learning	96 %	Very High	Medium
Proposed Method	94.2 %	Medium	High

Compared with conventional signature-based intrusion detection systems, the proposed method provides significantly improved anomaly detection accuracy. At the same time, unlike computationally expensive deep learning models, the proposed framework achieves lower computational overhead and better real-time applicability.

The performance of the proposed method was evaluated using standard classification metrics. The obtained results are as follows:

Accuracy: 94.2 %

Precision: 92.8 %

Recall: 91.5 %

F1-score: 92.1 %

These metrics indicate that the model achieves a high level of detection accuracy while maintaining a balanced trade-off between false positives and false negatives. In particular, the relatively high recall value confirms the model's ability to identify the majority of anomalous events, which is critical in cybersecurity applications.

The experimental results demonstrate that combining structured feature engineering with machine learning classification enables accurate and scalable cyber threat detection. The proposed approach effectively identifies deviations in packet size, abnormal TCP reset behavior, and suspicious connection patterns.

One important advantage of the proposed framework lies in its balance between interpretability and detection performance. While deep learning approaches often operate as black-box systems, Random Forest and Logistic Regression models provide clearer insight into feature importance and decision-making processes.

Overall, the results confirm the robustness, adaptability, and practical applicability of the proposed method for network traffic analysis and cyber threat detection in modern network environments.

Conclusion.

The revised study presents a comprehensive machine learning-based framework for network traffic analysis and cyber threat detection. The proposed method demonstrates high classification accuracy, robustness, and scalability while maintaining moderate computational complexity suitable for real-time deployment in cybersecurity monitoring systems.

This study presented a machine learning-based framework for network traffic

analysis and cyber threat detection.

The proposed method integrates:

- structured log analysis;
- preprocessing;
- feature engineering;
- feature selection;
- lightweight machine learning classification.

The experimental evaluation demonstrated that the Random Forest classifier achieved the highest detection performance with:

- Accuracy: 94.2 %;
- Precision: 92.8 %;
- Recall: 91.5 %;
- F1-score: 92.1 %;
- ROC-AUC: 0.963.

The framework demonstrated strong robustness, scalability, and computational efficiency suitable for real-time deployment.

Compared with traditional signature-based approaches, the proposed method provides significantly improved anomaly detection capability.

Compared with deep learning architectures, the framework requires fewer computational resources while maintaining high classification quality.

Future work will focus on:

- expanding the dataset;
- integrating unsupervised learning methods;
- streaming traffic analysis;
- SIEM integration;
- adaptive online learning.

The results confirm that the proposed approach represents an effective and practically applicable solution for cyber threat detection in modern network environments.

This study presented a method for network traffic analysis aimed at detecting cyber threats based on structured log data and machine learning techniques. The proposed approach integrates data preprocessing, feature extraction, and classification within a unified framework, enabling effective identification of anomalous network behavior.

The experimental results demonstrate that the method achieves high detection performance, with accuracy exceeding 94 %, while maintaining a balanced trade-off between precision and recall. The approach effectively identifies both statistical anomalies (e.g., deviations in packet size) and behavioral anomalies (e.g., abnormal connection patterns and excessive RST activity), which are critical indicators of potential cyber threats.

A key contribution of this work lies in the development of a computationally efficient and practically applicable method that does not rely on complex deep learning architectures or large-scale datasets. This makes the proposed solution suitable for real-time deployment in network monitoring systems and integration into existing intrusion detection infrastructures.

Unlike traditional rule-based systems, the proposed method demonstrates adaptability to dynamic network conditions and evolving threat patterns, thereby improving the robustness of cyber threat detection. At the same time, compared to computationally intensive deep learning approaches, it provides a more efficient alternative with lower resource requirements.

Despite the promising results, the study has certain limitations. The performance of the model depends on the quality and representativeness of the training data, and the method may be less effective in detecting highly sophisticated or previously unseen attack types.

Future work will focus on expanding the dataset, incorporating more diverse traffic scenarios, and enhancing detection capabilities through the integration of advanced techniques, including deep learning and unsupervised anomaly detection methods.

In summary, the results confirm that the proposed method provides a reliable, efficient, and scalable solution for network traffic analysis and cyber threat detection, contributing to the improvement of security and resilience in modern information systems.

REFERENCES

Abbasi M., Shahraki A. & Taherkordi A. (2022). Deep Learning for Network Traffic Monitoring and Analysis (NTMA). Computer Communications. — Vol. 181. — Pp. 150–165. DOI: https://doi.org/10.1016/j.comcom.2021.09.011. [in Eng.].

Chao J. & Xie, T. (2025). Deep Learning-Based Network Security Threat Detection and Defense // International Journal of Advanced Computer Science and Applications. — Vol. 5. — No. 11. Pp. 612–620. [in Eng.].

Dhakad A. & Singh S. (2023). Real-time network traffic analysis using machine learning and deep learning algorithms // Journal of Network Security Research. — Vol. 12. — No. 3. Pp. 45–54. [in Eng.].

Dong S. & Xia Y. (2022). Application of Deep Belief Networks in Network Traffic Identification Based on NetFlow Features. Applied Network Science. — Vol. 7. — No. 64. Pp. 1–15. [in Eng.].

Fotiadou K. & Velivassaki T.H. (2020). Network Traffic Analysis for Cybersecurity Forensics: Methods and Tools. Cybersecurity and Digital Forensics Review. — Vol. 4. — No. 2. Pp. 87–95. [in Eng.].

Joshi M.R. & Hadi T.H. (2021). Analysis and Forecasting of Network Traffic in Modern Systems // International Journal of Communication Networks and Information Security. — Vol. 13. — No. 1. Pp. 27–36. [in Eng.].

Kalwar J.H. & Bhatti S. (2024). Deep Learning Approaches for Network Traffic Classification in the Internet of Things (IoT): A Survey. arXiv preprint arXiv:2402.00920. [in Eng.].

Meshesha K. & Cherie K. (2023). Enhanced Cyber Threat Detection Framework Using Search Engine Technologies and Machine Learning // International Journal of Cybersecurity Research. — Vol. 5. — No. 2. Pp. 33–41. [in Eng.].

Meshkova E.V. (2020). Packet Sniffer-Based Analysis for Unauthorized Network Access Detection // Journal of Information Security Studies. — Vol. 8. — No. 1. Pp. 23–29. [in Eng.].

Mohammadi, S., Allahvakil, V., & Khaghani, M. (2021). Evaluating Intrusion Prevention System Performance under Varying Network Traffic Loads // International Journal of Computer Networks. — Vol. 9. — No. 4. Pp. 102–115. [in Eng.].

Tang S. (2021). Power Line Communication Network Model for Smart Grids Using OFDM. IEEE Transactions on Power Delivery. — Vol. 36. — No. 5. Pp. 2314–2323. [in Eng.].

Thwaini M.H. (2023). Anomaly Detection in Network Traffic Using Machine Learning for Early Threat Detection. ResearchGate Preprint. [in Eng.].

Wang W. & Wu W. (2022). Real-Time Network Traffic Analysis Using Rank Distributions for Anomaly Detection. IEEE Access. — Vol. 10. — Pp. 15247–15256. [in Eng.].

Zhang, W. & Lazaro, J.P. (2024). A Survey on Network Security Traffic Analysis and Anomaly Detection Techniques. IEEE Access. — Vol. 12. — Pp. 24056–24079. [in Eng.].

MDPI. (2024). Machine Learning-Based Network Anomaly Detection. Journal of Cybersecurity and Privacy. — Vol. 5. — No. 4. Pp. 143. [in Eng.].

PMC. (2023). AI-Based Anomaly Detection in IoT and Sensor Networks: A Review. Sensors. — Vol. 23. — No. 5. Pp. 9825. [in Eng.].

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ

МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ

Собственник:
АО «Международный университет информационных технологий» (Казахстан, Алматы)

Главный редактор:
Колесникова Катерина Викторовна

Ответственный редактор:
Мрзабаева Раушан Жалиевна

Компьютерная верстка:
Калабай Замзагуль Ертугановна

Сайт журнала: https://journal.iitu.edu.kz
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)

Подписано в печать 30.06.2026.
050040 г. Алматы, ул. Манаса 34/1, каб. 709, тел: +7 (727) 244-51-09).

Издание АО "Международного университета информационных технологий"