ВІСНИК ПРИАЗОВСЬКОГО ДЕРЖАВНОГО ТЕХНІЧНОГО УНІВЕРСИТЕТУ
2025р. Серія: Технічні науки Вип. 50
p-ISSN: 2225-6733; e-ISSN: 2519-271X

УДК 004.738.5 DOI: 10.31498/2225-6733.50.2025.336234

ІДЕНТИФІКАЦІЯ ТРАФІКУ МЕРЕЖ ПЕРЕДАЧІ ДАНИХ У РЕАЛЬНОМУ ЧАСІ

Клюшник В.В. *канд. техн. наук, доцент, Український державний університет науки і технологій, м. Дніпро, ORCID: https://orcid.org/0000-0002-4197-7171, e-mail: baksik_evgen@ukr.net;*

Чернецький Є.В. *аспірант, Український державний університет науки і технологій, м. Дніпро, e-mail: aico@i.ua;*

Онищенко О.В. *викладач, Дніпропетровський фаховий коледж будівельно-монтажних технологій та архітектури, м. Дніпро, ORCID: https://orcid.org/0000-0001-9228-8459, e-mail: alekseyonischenko@gmail.com*

У статті розглянуто сучасні підходи до ідентифікації та класифікації мережевого трафіку в реальному часі. Дослідження підкреслює актуальність проблеми, пов'язаної з високою динамікою мережевого середовища, та необхідність точних методів для управління, захисту даних і виявлення аномалій. Представлено огляд існуючих моделей трафіку, включаючи моделі на основі розподілів, часових рядів, фрактальних моделей та ланцюгів Маркова. Проведено аналіз методів класифікації, таких як використання номерів портів, DPI (Deep Packet Inspection), машинне навчання та статистичний аналіз пакетів. Критичний аналіз виявив переваги та обмеження кожної моделі. Простота реалізації та ефективність моделей розподілу обмежуються врахуванням залежності між даними. Моделі часових рядів придатні для прогнозування, але потребують значних обчислювальних ресурсів. Фрактальні моделі забезпечують аналіз довготривалих залежностей, проте їх застосування у реальному часі є складним. Ланцюги Маркова демонструють високу точність, але вимагають точної ініціалізації параметрів. Особливу увагу приділено інтеграції методів машинного навчання та оптимізації обчислювальних процесів для роботи із зашифрованим трафіком і в умовах обмежених ресурсів. Автори наголошують на необхідності розробки гібридних рішень, що поєднують переваги різних підходів, та визначають перспективи для подальших досліджень, спрямованих на створення більш точних і продуктивних систем аналізу трафіку. Ключові напрями подальших досліджень включають автоматизацію параметризації моделей, врахування зашифрованого трафіку та розробку алгоритмів машинного навчання, оптимізованих для роботи в реальному часі. Стаття є вагомим внеском у розвиток методів ідентифікації мережевого трафіку, пропонуючи шляхи підвищення безпеки та якості обслуговування мереж.

Ключові слова: *трафік, модель, інформаційні системи, мережа передачі даних.*

Постановка проблеми

Мережевий трафік є основою функціонування сучасних інформаційних систем і мереж передачі даних. Його класифікація та ідентифікація мають важливе значення для вирішення завдань управління, захисту даних та виявлення аномалій. Висока динаміка і складність мережевого середовища ставлять перед дослідниками завдання адаптації методів ідентифікації для отримання точних і надійних результатів. Виконаний аналіз дозволить не лише виявити слабкі сторони існуючих рішень, але й сформулювати напрями для подальших досліджень, включаючи вдосконалення існуючих методів і розробку нових підходів до ідентифікації мережевого трафіку. Це відкриває перспективи для створення більш точних і продуктивних рішень, здатних забезпечувати безперервне функціонування інформаційних систем. Ця стаття спрямована на аналіз існуючих моделей і методів ідентифікації мережевого трафіку. У ній проводиться порівняння підходів на основі моделей розподілу, часових рядів, фрактальних моделей і ланцюгів Маркова. Особлива увага приділена аналізу існуючих методів класифікації трафіку, зокрема класифікації за номером порту, детекції пакетів за корисним навантаженням (DPI), використанню методів машинного навчання та статистичних властивостей пакетів.

Аналіз останніх досліджень та публікацій

Ідентифікація мережевого трафіку в реальному часі – це процес визначення типу даних, що передаються через мережу, у той самий момент, коли вони передаються [1]. Це ключовий елемент для багатьох мережевих технологій, таких як:

- системи виявлення вторгнень (IDS) – виявляють підозрілу або злочинну активність в мережі;
- системи запобігання вторгненням (IPS) – блокують шкідливий трафік;
- мережеві аналізатори – допомагають усунути проблеми з мережею та оптимізувати її роботу;
- системи управління мережею – автоматизують управління мережевими ресурсами.

Існує багато різних методів ідентифікації мережевого трафіку, але найпоширенішими є:

- Аналіз заголовків пакетів;
- Глибока інспекція пакетів (DPI);
- Машинне навчання.

Аналіз заголовків пакетів – це один з найдавніших і найпоширеніших методів ідентифікації мережевого трафіку. Він полягає у вивченні інформації, що міститься в заголовках даних пакетів, які передаються по мережі. Цей метод дозволяє визначити тип протоколу, джерело і призначення пакету, а також інші важливі характеристики [2].

p-ISSN: 2225-6733; e-ISSN: 2519-271X

Принцип роботи аналізу заголовків:

- Захоплення пакетів. Спеціальні програми (сніфери) захоплюють мережевий трафік і зберігають його у файли для подальшого аналізу;
- Розбір заголовків. Програма аналізу розбирає заголовки кожного захопленого пакету, витягуючи з них необхідну інформацію;
- Ідентифікація протоколів. На основі значень полів у заголовку визначається протокол, який використовується для передачі даних;
- Фільтрація. Захоплені пакети фільтруються за різними критеріями, наприклад, за типом протоколу, адресою джерела або призначення;
- Виявлення аномалій. Порівнюючи характеристики пакетів з відомими шаблонами, можна виявити аномалії, які можуть свідчити про наявність загроз.

Переваги методу: простота (аналіз заголовків є відносно простим методом, який не вимагає великих обчислювальних ресурсів); швидкість (аналіз заголовків може виконуватися в реальному часі, що дозволяє швидко виявляти зміни в мережі); ефективність (метод дозволяє ідентифікувати широкий спектр протоколів і служб).

З урахуванням переваг цей метод має свої обмеження. Зашифровані пакети ускладнюють аналіз, оскільки вміст даних недоступний для прямого перегляду. З'являються нові протоколи, які можуть використовувати нестандартні заголовки, що ускладнює їх ідентифікацію. Зловмисники можуть використовувати різні методи для обходу аналізу заголовків [3].

Аналіз заголовків пакетів широко використовується в системах виявлення вторгнень (IDS) і системах запобігання вторгненням (IPS). Ці системи можуть виявляти такі загрози, як сканування портів, DDoS-атаки, вторгнення.

Глибока інспекція пакетів (DPI) – це потужна технологія, яка дозволяє аналізувати вміст мережевих пакетів на більш глибокому рівні, ніж традиційний аналіз заголовків. Вона використовується для виявлення та класифікації різноманітних типів трафіку, включаючи веб-трафік, електронну пошту, VoIP, відео та ін. DPI відіграє ключову роль у забезпеченні мережевої безпеки, управлінні мережею та оптимізації продуктивності [4].

DPI передбачає декодування та аналіз всього вмісту пакету даних, включаючи заголовок і дані. Це дозволяє виявити не тільки тип протоколу, але й конкретні програми, які генерують цей трафік, типи файлів, що передаються, і навіть вміст повідомлень.

Основні етапи DPI:

1. Захоплення пакетів. Спеціальне обладнання або програмне забезпечення захоплює мережевий трафік і передає його на аналіз;
2. Декодування. Пакет розбирається на окремі компоненти, такі як заголовок, дані та контрольні суми;
3. Аналіз вмісту. Вміст пакету аналізується за допомогою різних методів, включаючи пошук патернів, розпізнавання сигнатур, аналіз протоколів і машинне навчання;
4. Класифікація. На основі аналізу пакет класифікується за типом, додатком або користувачем;
5. Прийняття рішення. На основі результатів класифікації приймається рішення про подальшу обробку пакету (наприклад, блокування, дозволу, зміну пріоритету).

DPI має широкий спектр застосування – безпека (виявлення і блокування шкідливого трафіку (віруси, шпигунське програмне забезпечення, ботнети)); фільтрація веб-контенту (захист від DDoS-атак); управління мережею (контроль доступу до мережевих ресурсів); оптимізація використання пропускної здатності); аналіз використання мережі; квалітет обслуговування (QoS) (приоритезація певних типів трафіку (наприклад, VoIP, відеоконференції), гарантування якості передачі даних для критично важливих додатків) [5].

До переваг можна DPI віднести:

- висока точність класифікації трафіку;
- можливість виявлення нових загроз;
- гнучкість налаштування;
- детальна статистика про використання мережі.

Недоліки DPI випливають із його переваг:

- висока обчислювальна складність;
- ризик порушення конфіденційності;
- можливість помилкових спрацювань.

Машинне навчання – це галузь штучного інтелекту, яка дозволяє комп'ютерам навчатися на даних і робити прогнози або приймати рішення без явного програмування. У контексті мережевої безпеки машинне навчання використовується для ідентифікації мережевого трафіку з високою точністю і швидкістю [6].

Алгоритм роботи машинного навчання для ідентифікації трафіку.

На першому етапі іде збір даних. Збирається велика кількість даних про мережевий трафік. Ці дані включають в себе заголовки пакетів, вміст пакетів (якщо доступний), статистику про трафік і іншу релевантну інформацію. Чим більше інформації, тим вища якість машинного навчання. Далі іде обробка даних. Дані очищаються, нормалізуються і перетворюються в формат, придатний для навчання моделей машинного навчання. Після цього вже іде безпосередньо навчання моделі. Вибирається відповідний алгоритм машинного навчання (наприклад, нейронні мережі, випадкові ліси, SVM) і навчається на підготовлених даних. Модель навчається виявляти закономірності і відмінності між різними типами трафіку [7].

Після навчання, обов’язково проводиться тестування моделі та деплоймент. Навчена модель тестується на нових даних, щоб оцінити її точність, а вже потім модель розгортається в реальній системі для

p-ISSN: 2225-6733; e-ISSN: 2519-271X

ідентифікації мережевого трафіку в режимі реального часу.

Машинне навчання для ідентифікації трафіку використовуються для:

- виявлення аномалій. Виявлення незвичайної поведінки в мережі, що може свідчити про атаку або збій;
- класифікація трафіку. Визначення типу трафіку (веб, пошта, P2P, VoIP тощо);
- виявлення ботнетів. Виявлення мережі заражених комп'ютерів, які використовуються для здійснення DDoS-атак або розповсюдження спаму;
- виявлення інфікованих хостів. Виявлення комп'ютерів, заражених вірусами або іншим шкідливим програмним забезпеченням.

Найбільш розповсюдженими типами алгоритмів машинного навчання є: нейронні мережі, випадкові ліси, SVM (Support Vector Machines).

Серед існуючих інструментів для ідентифікації трафіку можна визначити наступні:

- Wireshark: Популярний інструмент з відкритим кодом для аналізу мережевого трафіку.
- Tcpdump: Командна утиліта для захоплення та аналізу пакетів.
- Snort: Система виявлення вторгнень з відкритим кодом.
- Bro: Інструмент для аналізу мережевого трафіку, що використовує мову програмування для написання правил аналізу.
- Комерційні рішення: Існують різноманітні комерційні рішення від таких компаній, як Cisco, Fortinet, Palo Alto Networks.

Ідентифікація мережевого трафіку в реальному часі є важливою задачею для забезпечення безпеки мереж і оптимізації їх роботи. З розвитком технологій і зростанням кількості загроз, ця область продовжує активно розвиватися [8].

Мета статті

Метою статті є узагальнення та критичний аналіз сучасних підходів до ідентифікації та класифікації мережевого трафіку з метою розробки ефективних гібридних рішень, здатних працювати з зашифрованими даними в режимі реального часу та підвищувати безпеку й якість обслуговування мереж.

Виклад основного матеріалу

Існуючі моделі мережевого трафіку можна класифікувати наступним чином:

Моделі на основі законів розподілу:

- Приклад: Пуассонівська модель (простота реалізації, але відсутність пам'яті).
- Переваги: простота використання.
- Обмеження: не враховують залежність між пакетами.

На рисунку 1 представлено порівняння рівня складності різних моделей мережевого трафіку, включаючи моделі на основі законів розподілу, часових рядів, фрактальні моделі і моделі на основі ланцюгів Маркова. Це допомагає оцінити їх придатність для задач реального часу.

Моделі часових рядів:

- Приклад: AR, ARMA, ARIMA (прогнозування трафіку з урахуванням сезонних характеристик).
- Переваги: врахування короткострокових залежностей.
- Обмеження: висока обчислювальна складність.

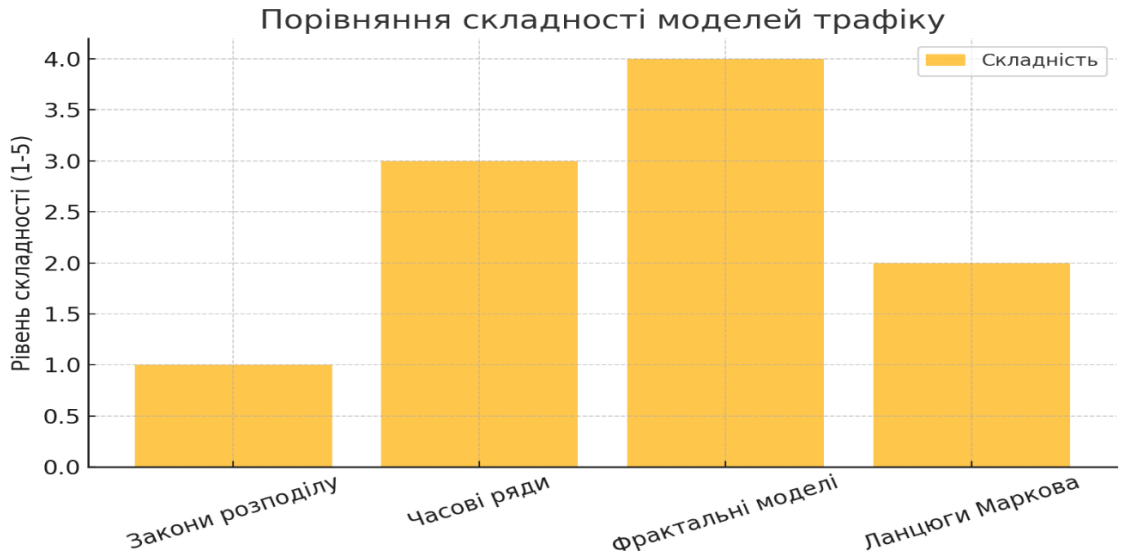

Рис.1 – Порівняння рівня складності різних моделей мережевого трафіку

p-ISSN: 2225-6733; e-ISSN: 2519-271X

На рисунку 2 показано порівняльну точність розглянутих моделей. Видно, що моделі на основі ланцюгів Маркова забезпечують найвищий рівень точності порівняно з іншими підходами.

Фрактальні моделі:

- Приклад: фрактальний броунівський рух (довготривала залежність).
- Переваги: опис явища пульсації трафіку.
- Обмеження: складність застосування у реальному часі.

Моделі на основі ланцюгів Маркова:

- Приклад: приховані Марковські моделі (врахування станів системи).
- Переваги: середня обчислювальна складність.
- Обмеження: необхідність точної ініціалізації параметрів.

Методи класифікації мережевого трафіку включають:

Класифікація за номером порту:

- Переваги: висока швидкість.
- Обмеження: непридатність для сучасних зашифрованих застосунків.

На рисунку 3 представлено аналіз складності різних методів класифікації мережевого трафіку. DPI і машинне навчання потребують більше обчислювальних ресурсів, тоді як класифікація за номером порту залишається найменш складною.

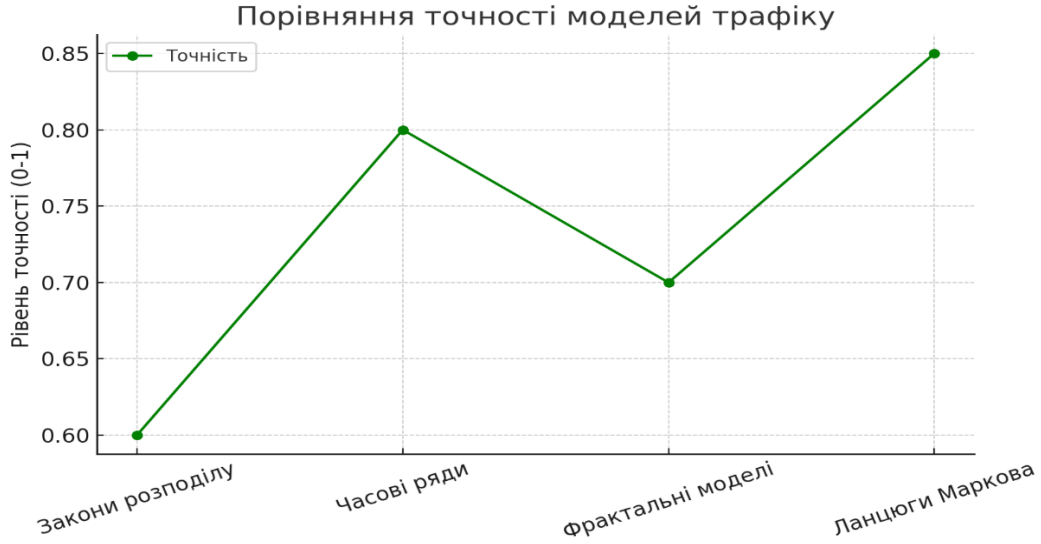

Рис. 2 – Порівняльна точність розглянутих моделей

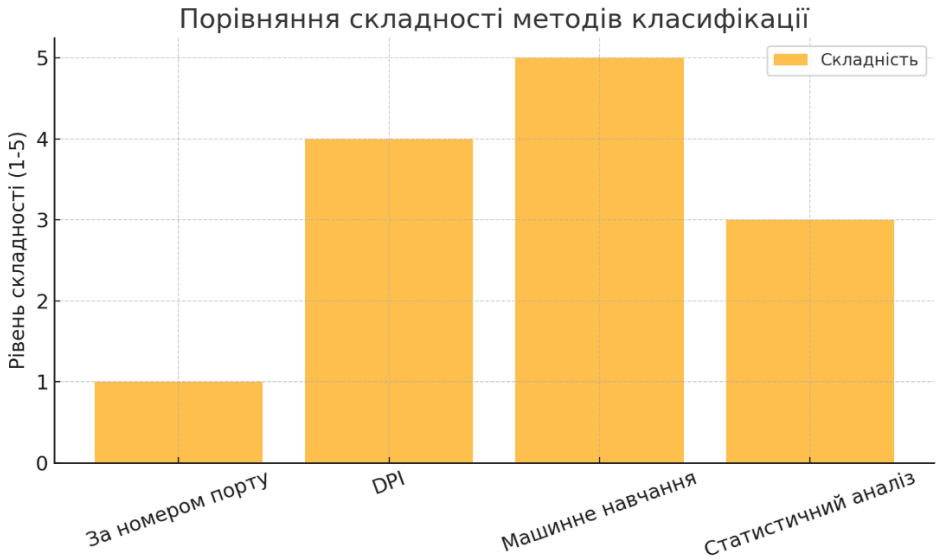

Рис. 3 – Аналіз складності різних методів класифікації мережевого трафіку

p-ISSN: 2225-6733; e-ISSN: 2519-271X

Глибокий аналіз вмісту пакетів (DPI):

- Переваги: висока точність для відомих застосунків.
- Обмеження: неможливість роботи із зашифрованим трафіком.

На рисунку 4 показано точність методів класифікації. DPI забезпечує максимальну точність, проте має обмеження в умовах роботи із зашифрованим трафіком.

Методи машинного навчання:

- Приклади: SVM, дерева рішень.
- Переваги: адаптація до різних типів трафіку.
- Обмеження: висока обчислювальна складність.

Статистичний аналіз пакетів:

- Переваги: незалежність від корисного навантаження.
- Обмеження: значні обчислювальні ресурси.

Проведений аналіз існуючих моделей і методів класифікації мережевого трафіку дозволив виявити їх ключові переваги та недоліки:

- Моделі на основі законів розподілу мають простоту реалізації, проте не враховують історичну залежність між пакетами. Це обмежує їх застосування для задач реального часу, де важливо враховувати контекст попередніх даних.
- Моделі часових рядів ефективно використовують короткострокові залежності, що робить їх придатними для прогнозування і аналізу певних типів трафіку. Проте їх висока обчислювальна складність обмежує використання в умовах високої пропускної здатності.
- Фрактальні моделі забезпечують довготривалі залежності, дозволяючи оцінювати мережеві ресурси і виявляти аномалії. Проте вони не підходять для ідентифікації трафіку в реальному часі через необхідність обробки великої кількості даних.
- Ланцюги Маркова, особливо приховані Марковські моделі, успішно застосовуються для класифікації різних протоколів і типів трафіку. Проте їх ефективність залежить від правильної ініціалізації параметрів, що вимагає додаткових обчислювальних зусиль.
- Методи машинного навчання мають високу точність і адаптивність, проте часто вимагають великого обсягу даних для навчання і значних ресурсів для обробки, що робить їх менш придатними в умовах реального часу.
- Статистичний аналіз пакетів демонструє гарні результати для класифікації трафіку, проте його використання обмежене складністю розрахунків, особливо при роботі із зашифрованими даними.

Висновки, зроблені на основі критичного аналізу, підкреслюють необхідність розробки нових методів, які зможуть поєднати переваги існуючих підходів, мінімізувавши їх недоліки. Це включає створення гібридних рішень, таких як комбінація статистичного аналізу і прихованих Марковських моделей, що дозволить досягти більш високої точності при зниженні обчислювальної складності.

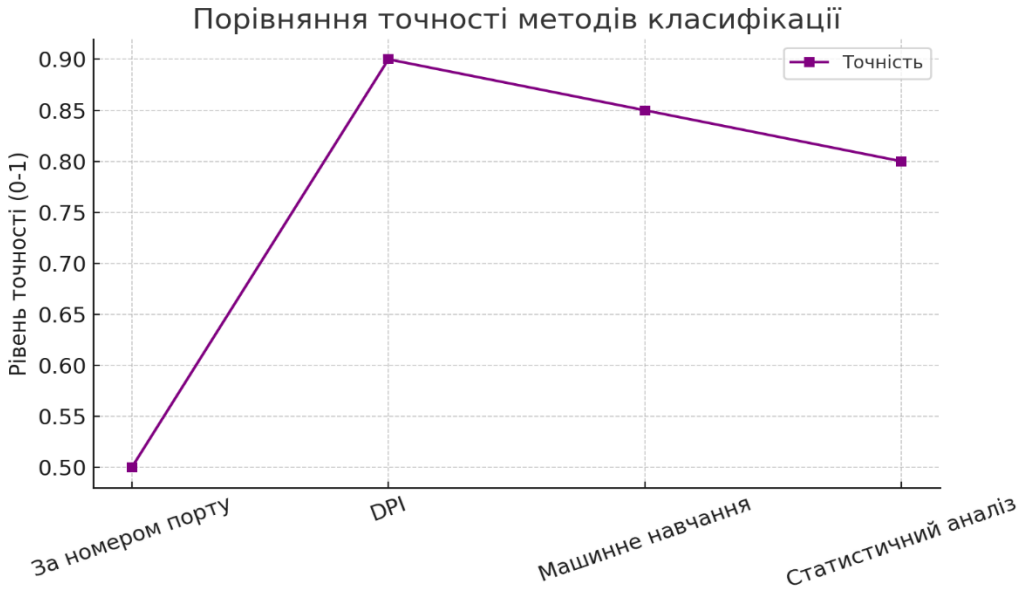

Рис. 4 – Точність методів класифікації

Висновки

Проведений аналіз існуючих моделей і методів класифікації мережевого трафіку показав, що кожна з них має свої переваги і обмеження. Простота використання класичних моделей, висока точність машинного навчання і адаптивність статистичних підходів створюють основу для розробки нових гібридних рішень. Необхідність у створенні методів, які враховують шифрування даних і забезпечують високу продуктивність у реальному часі, стає дедалі актуальнішою. Виділені напрями для подальших досліджень можуть закласти основу для нових рішень, здатних справлятися з сучасними викликами у сфері аналізу мережевого трафіку. Таким чином, створення гібридних підходів і оптимізація існуючих методів є ключовими кроками на шляху до більш ефективної ідентифікації мережевого трафіку, що, своєю чергою, сприяє поліпшенню якості обслуговування і безпеки мережевих систем.

Перелік використаних джерел

[1] Park J.W., Kwon M.W., Hong T. Queue congestion prediction for large-scale high performance computing systems using a hidden Markov model. *The Journal of Supercomputing*. 2022. Vol. 78. Pp. 12202-12223. DOI: https://doi.org/10.1007/s11227-022-04356-z.

[2] Network Traffic Classification Using Correlation Information. Parallel and Distributed Systems / J. Zhang et al. *IEEE Transactions on Parallel and Distributed Systems*. 2013. Vol. 24, iss. 1. Pp. 104-117. DOI: https://doi.org/10.1109/TPDS.2012.98.

[3] Технології забезпечення безпеки мережевої інфраструктури: Підручник / В.Л. Бурячок та ін. К. : КУБГ, 2019. 218 с.

[4] Abbasi M., Shahraki A., Taherkordi A. Deep Learning for Network Traffic Monitoring and Analysis (NTMA): A Survey. *Computer Communications*. 2021. Vol. 170. Pp. 19-41. DOI: https://doi.org/10.1016/j.comcom.2021.01.021.

[5] So-In C. A survey of network traffic monitoring and analysis tools. Washington University in St. Louis, 2009. 24 p.

[6] Big-DAMA: Big Data Analytics for Network Traffic Monitoring and Analysis / Casas P., D'Alconzo A., Zseby T., Mellia M. Proceedings of the 2016 Workshop on Fostering Latin-American Research in Data Communication Networks (LANCOMM '16), Florianopolis, Brazil, 22-26 August 2016. Pp. 1-3. DOI: https://doi.org/10.1145/2940116.2940117.

[7] Passive Flow Monitoring of Hybrid Network Connections regarding Quality of Service Parameters for the Industrial Automation / M. Ehrlich et al. *Kommunikation in der Automation*, Magdeburg, Germany, 2017. Pp. 1-11.

[8] The Data Domain Construction of Digital Twin Network / D. Chen et al. 2023 IEEE 3rd International Conference on Digital Twins and Parallel Intelligence (DTPI), Orlando, FL, USA, 07-09 November 2023. Pp. 1-5. DOI: https://doi.org/10.1109/DTPI59677.2023.10365474.

IDENTIFY DATA NETWORK TRAFFIC IN REAL TIME

Klyushnyk V.V. *PhD (Engineering), associate professor, Ukrainian State University of Science and Technologies, Dnipro, ORCID: https://orcid.org/0000-0002-4197-7171, e-mail: baksik_evgen@ukr.net;*

Chernetskyi E.V. *postgraduate student, Ukrainian State University of Science and Technologies, Dnipro, e-mail: aico@i.ua;*

Onishchenko O.V. *lecturer, Dnipro Applied College of Building and Mounting Technologies and Architecture, Dnipro, ORCID: https://orcid.org/0000-0001-9228-8459, e-mail: alekseyonischenko@gmail.com*

The article discusses modern approaches to real-time identification and classification of network traffic. The research emphasizes the urgency of the problem associated with the high dynamics of the network environment and the need for accurate methods for managing, protecting data, and detecting anomalies. An overview of existing traffic models is presented, including models based on distributions, time series, fractal models, and Markov chains. Classification methods such as port numbers, DPI (Deep Packet Inspection), machine learning, and statistical packet analysis are analyzed. The critical analysis revealed the advantages and limitations of each model. The ease of implementation and effectiveness of distribution models are limited by the consideration of dependencies between data. Time series models are suitable for forecasting but require significant computing resources. Fractal models provide an analysis of long-term dependencies, but their application in real time is difficult. Markov chains demonstrate high accuracy but require accurate initialization of parameters. Particular attention is paid to the integration of machine learning and computing process optimization methods for working with encrypted traffic and in conditions of limited resources. The authors emphasize the need to develop hybrid solutions that combine the advantages of different approaches and identify prospects for further research

aimed at creating more accurate and efficient traffic analysis systems. Key areas for further research include automating model parameterization, taking into account encrypted traffic, and developing machine learning algorithms optimized for real-time operation. This article is a significant contribution to the development of network traffic identification methods, suggesting ways to improve the security and quality of service of networks.
Keywords: *traffic, model, information systems, data transmission network.*

References

[1] J.W. Park, M.W. Kwon, and T. Hong, «Queue congestion prediction for large-scale high performance computing systems using a hidden Markov model», *The Journal of Supercomputing*, vol. 78, pp. 12202-12223, 2022. **doi: 10.1007/s11227-022-04356-z**.

[2] J. Zhang, Y. Xiang, Y. Wang, W. Zhou, Y. Xiang, and Y. Guan, «Network Traffic Classification Using Correlation Information. Parallel and Distributed Systems», *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, iss. 1, pp. 104-117, 2013. **doi: 10.1109/TPDS.2012.98**.

[3] V.L. Buryachok, *Tekhnolohii zabezpechennia bezpeky merezhevoi infrastruktury: Pidruchnyk* [Network Infrastructure Security Technologies: Textbook]. Kyiv, Ukraine: KUBG Publ., 2019. (Ukr.)

[4] M. Abbasi, A. Shahraki, and A. Taherkordi, «Deep Learning for Network Traffic Monitoring and Analysis (NTMA): A Survey», *Computer Communications*, vol. 170, pp. 19-41, 2021. **doi: 10.1016/j.comcom.2021.01.021**.

[5] C. So-In, *A survey of network traffic monitoring and analysis tools*. Washington University in St. Louis, 2009.

[6] P. Casas, A. D'Alconzo, T. Zseby, and M. Mellia, «Big-DAMA: Big Data Analytics for Network Traffic Monitoring and Analysis», in Proc. of the 2016 Workshop on Fostering Latin-American Research in Data Communication Networks (LANCOMM '16), Florianopolis, Brazil, 2016, pp. 1-3. **doi: 10.1145/2940116.2940117**.

[7] M. Ehrlich, A. Biendarra, H. Trsek, E. Wojtkowiak, and J. Jasperneite, «Passive Flow Monitoring of Hybrid Network Connections regarding Quality of Service Parameters for the Industrial Automation», in Proc. of the Kommunikation in der Automation, Magdeburg, Germany, 2017, pp. 1-11.

[8] D. Chen, C. Zhou, H. Yang, M. Li, and L. Lu, «The Data Domain Construction of Digital Twin Network», in Proc. of the 2023 IEEE 3rd International Conference on Digital Twins and Parallel Intelligence (DTPI), Orlando, FL, USA, 2023, pp. 1-5. **doi: 10.1109/DTPI59677.2023.10365474**.

Стаття надійшла 07.02.2025
Стаття прийнята 21.02.2025
Стаття опублікована 30.06.2025

Цитуйте цю статтю як: Клюшник В.В., Чернецький Є.В., Онищенко О.В. Ідентифікація трафіку мереж передачі даних у реальному часі. *Вісник Приазовського державного технічного університету. Серія: Технічні науки*. 2025. Вип. 50. С. 18-24. DOI: https://doi.org/10.31498/2225-6733.50.2025.336234.