УДК 681.3

DOI: 10.18372/2073-4751.82.20365

Кулаков О.Ю., к.т.н.,
orcid.org/0009-0001-1647-0336,
e-mail: Kulakov.Oleksii@lll.kpi.ua,
Чередник В.Ю.,
e-mail: vitalijcerednik@gmail.com

АНАЛІЗ МЕТОДІВ КЛАСИФІКАЦІЇ ТРАФІКУ В ПРОГРАМНО-КОНФІГУРОВАНИХ МЕРЕЖАХ

**Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»**

Вступ

В умовах постійного зростання обсягів мережевого трафіку та різноманітності мережевих додатків, ефективна класифікація трафіку стає критично важливим завданням для забезпечення якості обслуговування (*QoS*) та безпеки мережі.

У сучасному світі, де обсяги мережевого трафіку постійно зростають, а вимоги до якості обслуговування стають все більш жорсткими, точна та ефективна класифікація трафіку стає критично важливою задачею. Особливо це актуально для програмно-конфігурованих мереж (*SDN*), де правильна класифікація трафіку безпосередньо впливає на якість надання послуг, безпеку мережі та ефективність використання ресурсів.

Мета дослідження

Метою даного дослідження є аналіз та систематизація сучасних методів класифікації мережевого трафіку в середовищі *SDN*. Дослідження спрямоване на визначення ефективності різних підходів до класифікації, їх переваг та недоліків, а також можливостей практичного застосування в умовах реальних мереж. Особлива увага приділяється методам, що використовують переваги архітектури *SDN* для підвищення точності та швидкодії класифікації.

Основна частина

Задача класифікації трафіку в програмно-конфігурованих мережах має певні особливості, що зумовлені специфічною архітектурою *SDN*, яка надає унікальні можливості для підвищення ефективності класифікації трафіку. По-перше, це наявність *централізованого управління***.** Контролер *SDN* може агрегувати статистику з усіх комутаторів мережі та приймати рішення на основі глобального представлення про стан мережі. По-друге, *програмованість* та можливість динамічно змінювати правила обробки трафіку дозволяє реалізовувати складні алгоритми класифікації та адаптувати їх до змінних умов. По-третє, *SDN* мережі дозволяють *розподіляти навантаження.* Тобто класифікація може виконуватися як на рівні контролера, так і на рівні комутаторів, що дозволяє оптимізувати використання ресурсів мережі.

Традиційні методи класифікації трафіку передусім використовують портову глибоку перевірку пакетів. Зі зростанням кількості застосунків і динамічністю портів традиційний метод класифікації на основі портів стає неефективним. Більшість методів глибокої перевірки пакетів використовують зіставлення регулярних виразів для ідентифікації даних пакетів, але існують два методи впровадження: недетерміновані скінченні автомати (*NFA*) і детерміновані скінченні автомати (*DFA*) [1]. Обидва мають певні обмеження; наприклад, *NFA* мають менший обсяг пам'яті, але потребують значного часу для зіставлення. Навпаки, *DFA* показують зворотну ситуацію, маючи вищий ризик займати занадто великий обсяг пам'яті. На сьогоднішній день основні методи класифікації трафіку головним чином фокусуються на використанні різноманітних методів штучного інтелекту. Хоча методи штучного

інтелекту можуть ефективно класифікувати трафік, зі зростанням обсягу даних застосунків та змінами у їхній формі висуваються вимоги до часової та просторової складності тренування алгоритмів [2].

Традиційні методи класифікації трафіку

Для класифікації трафіку в *SDN* мережах застосовуються методи, що відносяться до традиційних методів класифікації трафіку. Ці методи поділяються на кілька основних категорій.

Метод *класифікації на основі портів* базується на аналізі номерів портів *TCP*/*UDP*, які використовуються додатками. Хоча цей підхід є простим та швидким, він має обмежену ефективність через використання динамічних портів та тунелювання трафіку.

До традиційних методів належить метод *класифікації трафіку на основі корисного навантаження Deep Packet Inspection (DPI).* Характерною особливістю даного методу є аналіз вмісту пакетів для ідентифікації протоколів та додатків. Цей метод забезпечує високу точність, але вимагає значних обчислювальних ресурсів та має обмеження при роботі з шифрованим трафіком. Технологія *DPI* вимагає значних обчислювальних ресурсів, оскільки часто базується на алгоритмах машинного навчання. Наявні на дану тематику дослідження та практичні випробування [3] вказують на наявність певних проблем алгоритмів на основі технології *DPI*:

- Можуть ідентифікувати лише обмежену кількість типів трафіку.
- У деяких тестах демонструється недостатньо висока точність розпізнавання.
- Вартість комерційних бібліотек є значною, як і витрати на їхню підтримку.

Методи машинного навчання

Методи машинного навчання стали одним з найбільш перспективних напрямків у вирішенні задачі класифікації трафіку в *SDN* мережах [2]. Вони дозволяють виявляти складні закономірності в даних та адаптуватися до змін у характеристиках трафіку. Основна перевага машинного навчання полягає в можливості автоматизувати процес класифікації, мінімізуючи необхідність ручного налаштування правил та евристик. Виділяють дві основні групи методів машинного навчання, що застосовуються для класифікації трафіку в *SDN* мережах.

Навчання з учителем

Методи навчання з учителем використовують розмічені набори даних для побудови моделей, які використовуються для класифікації нового невідомого трафіку. Ці методи забезпечують високу ефективність при класифікації відомих типів трафіку.

Машина опорних векторів (Support Vector Machines, SVM) [4] – алгоритм для вирішення задач класифікації, який намагається знайти оптимальну гіперплощину, що розділяє дані різних класів у багатовимірному просторі ознак. У контексті класифікації мережевого трафіку *SVM* ефективно працює з великою кількістю ознак, витягнутих з мережевих пакетів та потоків. У *SDN* мережах *SVM* часто застосовується для класифікації трафіку за такими параметрами, як розмір пакетів, міжпакетні інтервали, тривалість сесій та інші статистичні характеристики потоків.

Перевагами застосування *SVM* для класифікації трафіку є висока точність класифікації, стійкість до перенавчання та ефективна робота у просторах ознак високої розмірності. До недоліків алгоритму відносять високу обчислювальну складність при навчанні на великих наборах даних та чутливість до шуму в даних [5].

У статті [6] наведені результати дослідження методу *Random Forest* для класифікації трафіку. Це ансамблевий метод, заснований на побудові множини дерев рішень. Кожне дерево навчається на випадковій підмножині ознак та прикладів, а фінальне рішення приймається шляхом голосування всіх дерев. У *SDN* мережах *Random Forest* успішно застосовується для класифікації потоків трафіку різних додатків та протоколів, а також для виявлення аномалій.

У контексті класифікації трафіку *Random Forest* має високу точність класифікації завдяки ансамблевому підходу, стійкість до перенавчання, вбудовану оцінку важливості ознак, що дозволяє визначити найбільш інформативні параметри трафіку. У якості недоліків виділяють: значні обчислювальні ресурси при великій кількості дерев, складність інтерпретації моделі в цілому, менша ефективність при наявності сильних нелінійних залежностей.

Нейронні мережі, особливо *глибокі (Deep Neural Networks, DNN)*, широко використовуються для задач класифікації трафіку завдяки їх здатності автоматично виділяти складні ієрархічні ознаки з даних. Результати дослідження ефективності застосування нейронних мереж для класифікації трафіку продемонстровані в роботах [7, 8]. Як показує практика, у *SDN* мережах нейронні мережі можуть бути реалізовані як на рівні контролера для глибокого аналізу агрегованих потоків, так і на рівні граничних пристроїв для попередньої класифікації в реальному часі.

Існують три основні типи нейронних мереж, що застосовуються для класифікації трафіку.

Згорткові нейронні мережі (CNN) дозволяють виявляти локальні шаблони в трафіку, які можуть бути характерними для певних додатків або протоколів. *CNN* показують високу точність при класифікації шифрованого трафіку, аналізуючи послідовності розмірів пакетів та міжпакетних інтервалів.

Рекурентні нейронні мережі (RNN), зокрема *LSTM (Long Short-Term Memory)* та *GRU (Gated Recurrent Units),* здатні ефективно моделювати часові залежності в послідовностях пакетів. Дана особливість є важливою для класифікації трафіку протоколів з складною поведінкою, таких як потокові мультимедійні сервіси або онлайн-ігри.

Глибокі нейронні мережі прямого поширення (DNN) використовуються для класифікації на основі агрегованих статистичних характеристик потоків трафіку. Вони дозволяють моделювати складні нелінійні залежності між параметрами трафіку та типом додатка.

Перевагами застосування нейронних мереж для класифікації трафіку є:

- Висока адаптивність до різних типів даних та задач.
- Можливість автоматичного вилучення ознак без необхідності ручного проектування.
- Висока точність класифікації при наявності достатньої кількості навчальних даних.
- Можливість навчання моделі в режимі онлайн, що важливо для адаптації до змін у трафіку.

Серед недоліків нейронних мереж виділяють:

- Необхідність великої кількості розмічених даних для навчання.
- Висока обчислювальна складність, особливо на етапі навчання.
- Складність інтерпретації результатів ("чорний ящик").
- Можливість перенавчання при неправильному виборі архітектури або гіперпараметрів.

Навчання без учителя

Методи навчання без учителя не потребують розмічених даних і здатні виявляти приховані структури та закономірності в трафіку. Завдяки даній особливості методи навчання без учителя є ефективними для виявлення нових типів додатків або аномалій.

K-means – один з алгоритмів кластеризації, суть якого полягає в розбитті даних на *K* кластерів, мінімізуючи внутрішньокластерну дисперсію. Алгоритми кластеризації групують потоки трафіку за подібністю їх характеристик, формуючи кластери, які можуть відповідати різним типам додатків або мережевої активності. У контексті програмно-конфігурованих мереж *K-means* може використовуватися для первинної кластеризації трафіку з подальшим аналізом отриманих кластерів та призначенням їм відповідних міток. Кластеризація трафіку дозволяє спростити задачу класифікації трафіку та зменшити час її виконання.

Результати досліджень [9, 10] доводять високу ефективність даного методу для класифікації трафіку. Перевагами *K-means* є простота реалізації, висока швидкість роботи та можливість масштабування для роботи з великими обсягами даних. До недоліків належать чутливість до початкового розподілу центроїдів та невисока ефективність при роботі з кластерами нестандартної форми.

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) – алгоритм кластеризації, заснований на щільності, який може виявляти кластери довільної форми та автоматично визначати кількість кластерів. Цей алгоритм має високу ефективність виявлення аномалій у трафіку, оскільки здатний ідентифікувати викиди як шум.

У статті [11], що присвячена дослідженню методів неконтрольованого навчання, було виділено такі переваги *DBSCAN*: не вимагає попереднього знання кількості кластерів, стійкість до шуму в даних, здатність виявляти кластери складної форми. У якості мінусів цього алгоритму зазначено чутливість до параметрів щільності та вищу обчислювальну складність порівняно з *K-means*.

Автоенкодери [12] – це тип нейронних мереж, які навчаються відтворювати вхідні дані через проміжний шар меншої розмірності. У контексті класифікації трафіку автоенкодери можуть використовуватися для зниження розмірності ознак трафіку, виявлення аномалій (трафік, який погано відтворюється, вважається аномальним), генерації компактних представлень (*embeddings*) для подальшої класифікації.

GenClass – метод класифікації трафіку, реалізований на мові програмування *C++*. Автори статті [13] у своєму дослідженні порівнюють його з іншими методами контрольованого навчання. Результати дослідження показали його ефективність для задачі класифікації трафіку.

Метод *GenClass* демонструє вищу точність порівняно з іншими методами, що підтверджується як високою точністю, так і повнотою класифікації. Головною перевагою *GenClass* є здатність автоматично створювати правила класифікації та виявляти приховані зв'язки між характеристиками проблеми та відповідними класами. Проте недоліком методів, заснованих на генетичному програмуванні, є потреба в значно більшій кількості часу на виконання у порівнянні з іншими підходами машинного навчання.

Порівняльний аналіз методів класифікації трафіку

Результати досліджень на тему класифікації трафіку в *SDN* мережах продемонстровані в табл. 1. Ефективність розглянутих методів класифікації трафіку оцінюється точністю класифікації для тих типів трафіку, що використовувалися в дослідженнях.

Таблиця 1. Порівняльний аналіз методів класифікації

Публікація	Метод класифікації	Тип трафіку	Точність
1	2	3	4
Data Traffic Classification in Software Defined Networks (*SDN*) *using supervised-learning*, 2020 [5]	*SVM*	*HTTP*, пошта, аудіо та відео	92,2%
SDN-Enabled FiWi-IoT Smart Environment Network Traffic Classification Using Supervised ML Models, 2021 [6]	*Random Forest*	*HTTP*, голосові повідомлення, відео, *IoT*	99,6%

Продовження таблиці 1

1	2	3	4
Elephant Flows Detection Using Deep Neural Network, *Convolutional Neural Network*, *Long Short-Term Memory*, *and Autoencoder*, 2023 [7]	*DNN*	Додатки в режимі реального часу	99,12%
	CNN		98,49%
	LSTM		98,78%
	Автоенкодер		97,95%
Traffic Classification in Software Defined Networks based on Machine Learning Algorithms, 2024 [14]	*SVM*	*DNS*, *WWW*, *VOIP*, *FTP*, *ICMP*, *P2P*	98,7%
	Random Forest		99,6%
	Decision Tree		99,8%
	KNN		99,1%
DBSCAN SMOTE LSTM: *Effective Strategies for Distributed Denial of Service Detection in Imbalanced Network Environments*, 2024 [15]	*DBSCAN*	*DDoS* атаки	98,3%
Traffic Classification in Software-Defined Networking Using Genetic Programming Tools, 2024 [13]	*GenClass*	*DNS*, *WWW*, *VOIP*, *FTP*, *ICMP*, *P2P*	99,42%

Висновки

У даній роботі було проведено аналіз сучасних методів класифікації мережевого трафіку в середовищі програмно-конфігурованих мереж. Визначено, що традиційні методи, такі як класифікація на основі портів та глибокий аналіз пакетів (*DPI*), мають низку обмежень, зокрема низьку ефективність у динамічних мережах та значні обчислювальні витрати.

Натомість, методи машинного навчання демонструють високу ефективність у класифікації трафіку в *SDN*. Метод *SVM* забезпечує високу точність, але потребує значних обчислювальних ресурсів. *Random Forest* показує надійність і точність при класифікації різних типів трафіку. Глибокі нейронні мережі (*DNN*, *CNN*, *LSTM*) довели свою ефективність при обробці великого обсягу трафіку, зокрема для аналізу шифрованих даних. Методи кластеризації, такі як *K-means* і *DBSCAN*, виявляють аномалії та нові типи трафіку, що є цінним для кібербезпеки.

Особливу увагу приділено методам, які використовують можливості *SDN* для покращення класифікації. Централізоване управління мережею дозволяє ефективно збирати та аналізувати великі обсяги даних у реальному часі, що значно підвищує точність класифікації та адаптивність до змін у мережевому середовищі.

Таким чином, використання штучного інтелекту, зокрема глибоких нейронних мереж та алгоритмів кластеризації, у поєднанні з особливостями *SDN* є перспективним напрямом для класифікації трафіку. Подальші дослідження мають бути зосереджені на оптимізації обчислювальних витрат та адаптації до змінюваних умов мережевого середовища.

Література

1. Luckner M. Conversion of Decision Tree into Deterministic Finite Automaton for High Accuracy Online SYN Flood Detection. *Computational Intelligence, IEEE Symposium Series on Cape Town*. 2015. P. 75–82.

2. Ospina Cifuentes B. J. et al. Analysis of the Use of Artificial Intelligence in Software-Defined Intelligent Networks: A Survey. *Technologies.* 2024. Vol. 12. P. 99.

3. Kulakov Y. O., Oboznyi D. M. DPI traffic classification technologies in SDN networks: a survey. *Problems of informatization and control*. 2023. Vol. 2, no. 74. P. 49–54.

4. Pradhan A. Network Traffic Classification using Support Vector Machine and Artificial Neural Network. *International Journal of Computer Applications*. 2011. Vol. 8. P. 8–12.

5. Raikar M. M. et al. Data traffic classification in software defined networks (SDN) using supervised-learning. *Procedia Computer Science.* 2020. Vol. 171. P. 2750–2759.

6. Ganesan E. et al. SDN-enabled FiWi-IoT smart environment network traffic classification using supervised ML models. *Photonics*. 2021. Vol. 8. P. 201.

7. Wassie Geremew G., Ding J. Elephant Flows Detection Using Deep Neural Network, Convolutional Neural Network, Long Short-Term Memory, and Autoencoder. *Journal of Computer Networks and Communications*. 2023. Vol. 1495642. P. 18.

8. Nuñez-Agurto D. et al. A Novel Traffic Classification Approach by Employing Deep Learning on Software-Defined Networking. *Future Internet.* 2024. Vol. 16. P. 153.

9. Wang Y. et al. A novel semi-supervised approach for network traffic clustering. *2011 5th International Conference on Network and System Security* : proceedings, Milan, Italy, 06–08 September 2011 / IEEE. 2011. P. 169–175.

10. Du Y., Zhang R. Design of a method for encrypted p2p traffic identification using k-means algorithm. *Telecommunication Systems.* 2013. Vol. 53. P. 163–168.

11. Canever H., Wang X. Network traffic classification using Unsupervised Learning: a comparative analysis of clustering algorithms. 2023. *hal-04149117.* P. 26.

12. D'Angelo G., Palmieri F. Network traffic classification using deep convolutional recurrent autoencoder neural networks for spatial–temporal features extraction. *Journal of Network and Computer Applications.* 2021. Vol. 173. P. 16.

13. Margariti S. V. et al. Traffic Classification in Software-Defined Networking Using Genetic Programming Tools. *Future Internet.* 2024. Vol. 16. P. 338.

14. Ashour M. M., Yakout M. A., AbdElhalim E. Traffic Classification in Software Defined Networks based on Machine Learning Algorithms. *International Journal of Telecommunications.* 2024. Vol. 4. P. 1–19.

15. Efendi R., Wahyono T., Widiasari I. R. DBSCAN SMOTE LSTM: Effective Strategies for Distributed Denial of Service Detection in Imbalanced Network Environments. *Big Data and Cognitive Computing.* 2024. Vol. 8. P. 118.

Кулаков О.Ю., Чередник В.Ю.

АНАЛІЗ МЕТОДІВ КЛАСИФІКАЦІЇ ТРАФІКУ В ПРОГРАМНО-КОНФІГУРОВАНИХ МЕРЕЖАХ

В даній роботі розглянуто методи класифікації трафіку в програмно-конфігурованих мережах (SDN). Метою роботи є аналіз та систематизація сучасних підходів до класифікації мережевого трафіку в середовищі SDN.

Проаналізовано традиційні методи класифікації, зокрема на основі портів та глибокого аналізу пакетів (DPI). Досліджено методи машинного навчання, включаючи SVM, Random Forest, нейронні мережі та алгоритми кластеризації, їх переваги та недоліки. Порівняно точність та ефективність різних методів класифікації на основі опублікованих результатів досліджень.

Авторами запропоновано подальше дослідження способів використання алгоритмів штучного інтелекту в програмно-конфігурованих мережах з метою підвищення

точності класифікації трафіку та адаптивності до змін у мережевому середовищі. Наведено результати порівняння різних методів класифікації трафіку з акцентом на їх ефективність у мережах SDN.

Ключові слова: *програмно-конфігуровані мережі; класифікація трафіку; машинне навчання; нейронні мережі; якість обслуговування; штучний інтелект.*

Kulakov O.Y., Cherednyk V.Y.

ANALYSIS OF TRAFFIC CLASSIFICATION METHODS IN SOFTWARE-DEFINED NETWORKS

This paper examines traffic classification methods in Software-Defined Networks (SDN). The purpose of this work is to analyze and systematize modern approaches to network traffic classification in the SDN environment.

Traditional classification methods, particularly port-based and Deep Packet Inspection (DPI), are analyzed. Machine learning methods, including SVM, Random Forest, neural networks, and clustering algorithms, along with their advantages and disadvantages, are investigated. The accuracy and efficiency of different classification methods are compared based on published research results.

The authors propose further research on the use of artificial intelligence algorithms in software-defined networks to improve traffic classification accuracy and adaptability to changes in the network environment. The results of comparing different traffic classification methods with a focus on their effectiveness in SDN networks are presented.

Keywords: *SDN; traffic classification; machine learning; neural networks; QoS; artificial intelligence.*