DOI https://doi.org/10.36059/978-966-397-357-9-126

ДОСЛІДЖЕННЯ АНОМАЛІЙ МЕРЕЖНОГО ТРАФІКА DDOS-АТАК

Соловська І. М.
кандидат технічних наук,
доцент кафедри комп'ютерних наук
Міжнародний гуманітарний університет
м. Одеса, Україна

Севрюков О. В.
здобувач вищої освіти другого (магістерського) рівня
за спеціальністю 125 – Кібербезпека
Міжнародний гуманітарний університет
м. Одеса, Україна

Сучасний технологічний розвиток інфокомунікаційних систем та мереж супроводжується постійною необхідністю забезпечення безпеки мережевої інфраструктури від різного виду загроз, зокрема атак, що призводять до відмови в обслуговуванні DDoS (Distributed Denial of Service Attack), таких як SYN-Flood, ICMP-Flood, UDP-Flood, HTTP-Flood [1–2].

Детектування DDoS-атак проводиться за допомогою аналізу аномалій мережного трафіку – пошуку відхилень від визначених характеристик трафіку. В той же час, DDoS-атаки рівнів L2-L4 (UDP-Flood, HTTP-Flood, SYN-Flood, ICMP-Flood) переважно генерують значний обсяг трафіку, а DDoS-атаки прикладного рівня L7 зовсім не вимагають генерації значного обсягу трафіку та складно ідентифікуються звичайними системами. Важливим питанням є дослідження аномалій мережного трафіку та вибір певного методу, який би дозволив забезпечити необхідну точність визначення атак [1–2].

Метою даної роботи є аналіз аномалій мережного трафіку DDoS-атак з метою підвищення точності їхнього детектування.

Для дослідження аномалій мережного трафіку DDoS-атак обґрунтовано обрано алгоритми класифікації даних машинного навчання, які дозволяють значно покращити результати детектування загроз [3–4], такі як: алгоритм логістичної регресії, алгоритм k-найближчих сусідів та алгоритм Random forest. Вихідними даними для дослідження є вибірки мережевого трафіку, зібрані за допомогою бібліотек, що забезпечують взаємодію з драйверами мережного

інтерфейсу та сформовані у Datasets мережних DDoS-атак, https://www.kaggle.com/datasets/jacobvs/ddos-attack-network-logs/data [5], який містить UDP-Flood, Smurf, SIDDOS, HTTP-Flood та легітимний (звичайний) трафік. Для дослідження використано вільне програмне забезпечення для аналізу даних та машинного навчання Weka 3.7.1 [6]. Візуалізація обраного Datasets в Weka 3.7.1 показано на рис. 1. Datasets має дані про характеристик трафіку: протоколи та інтерфейси, мережеві інтерфейси, види DDoS-атак, швидкості передавання даних, час затримки пакетів, час очікування та час надсилання.

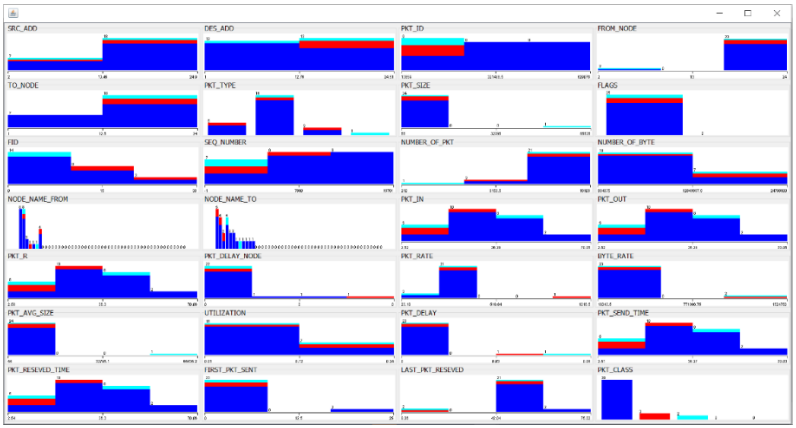

Рис. 1. Візуалізація Datasets/Ddos-attack-network

Проведена класифікація аномалій трафіку Datasets за допомогою алгоритму k-найближчих сусідів (рис. 2) з використанням пакету Weka 3.7.1, для загальної кількості випадків 25, серед яких, дані трьох DDoS-атак виду UDP-flood, двох Smurf та відсутні SIDDOS та HTTP-Flood, 20 даних нормального трафіку без атак. Отримані результати зведені в табл. 1.

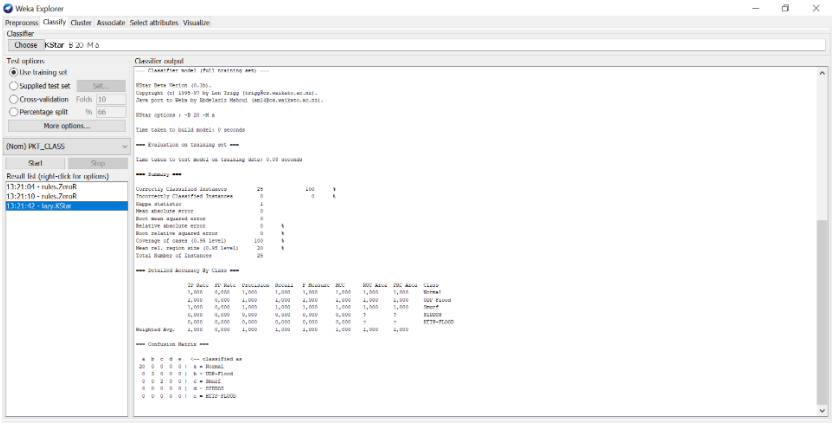

Рис. 2. Результати класифікації алгоритмом k-найближчих сусідів [6]

Таблиця 1

Результати класифікації алгоритмом k-найближчих сусідів

	Normal	UDP-flood	Smurf	SI DDoS	HTTP-flood
Normal	19	1	0	0	0
UDP-flood	2	0	1	0	0
Smurf	1	1	0	0	0
SI DDoS	0	0	0	0	0
HTTP-flood	0	0	0	0	0

Класифікація на основі алгоритму k-найближчих сусідів полягає у запам'ятовуванні навчальної вибірки даних. Для аналізу даних трафіку алгоритм виконує пошук найближчої до неї точки навчальної вибірки, тим самим знаходячи k-найближчих [6]. Згідно результатів класифікації наведених в табл. 1 можна стверджувати, що дана класифікація трафіку DDoS-атак має похибки. Аналогічно розглянуті алгоритми класифікації логістичної регресії та алгоритм Random forest. Результати порівняння точності визначення DDoS-атак зведені в табл. 2.

На закінчення можна зробити наступні висновки:

1. Дослідження аномалій трафіку DDoS-атак виконано за допомогою алгоритмів класифікації машинного навчання: алгоритм логістичної регресії, алгоритм k-найближчих сусідів та алгоритм Random forest.

Таблиця 2

Порівняння точності визначення DDoS-атак за допомогою різних методів класифікації

Значення середньої абсолютної похибки MAPE,%	Normal	UDP-flood	Smurf	SI DDoS	HTTP-flood
Класифікатор логістичної регресії	59,7%	64,3%	56,7%	-	-
Класифікатор k-найближчих сусідів	42,2%	52,3%	45,7%	-	-
Класифікатора Random forest	11,3%	9,8%	8,7%	-	-

2. Встановлено, що доцільним з точки зору підвищення точності детектування аномалій трафіку є використання алгоритм Random forest, який дозволяє досягти точності 8,7–11,3%.

Література:

1. ITU-T. Recommendation X.1208. X-series: Data Networks, Open Systems Interconnection and Security, 2019.

2. ITU-T. Recommendation X.1042. X-series: Data Networks, Open Systems Interconnection and Security, Information and Network Security – Network Security, 2019.

3. T. Radivilova, L. Kirichenko, O. Lemeshko and all, «Analysis of anomaly detection and identification methods in 5G traffic,» 2021 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), 2021.

4. Lemeshko O., Yevdokymenko M., Yeremenko M., Kuzminykh I. Cyber Resilience and Fault Tolerance of Artificial Intelligence Systems: EU Standards, Guidelines, and Reports. Proceedings of the Selected Papers on Cybersecurity Providing in Information and Telecommunication Systems (CPITS 2020). Kyiv, Ukraine. CEUR, 2020. P. 99–108.

5. Datasets. DDoS Attack Network. https://www.kaggle.com

6. Weka 3: Програмне забезпечення машинного навчання на Java. https://www.cs.waikato.ac.nz/ml/weka/