КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 3 (31), 2025
ISSN 2663 - 4023
DOI 10.28925/2663-4023.2025.31.1062
УДК 004.056

Бойко Анна Олександрівна
старший викладач кафедри систем та технологій кібербезпеки
Державний університет інформаційно-комунікаційних технологій, Київ, Україна
D: 0009-0001-3709-6283
a.boiko@duikt.edu.ua

МЕТОД ВИЯВЛЕННЯ АТАК НА КОРПОРАТИВНІ ВЕБ-ДОДАТКИ НА ОСНОВІ ГРАДІЄНТНОГО БУСТИНГУ

Анотація. У роботі розглянуто задачу виявлення веб-атак у мережевому трафіку корпоративних веб-додатків в умовах переваги кількості шифрованих з'єднань, коли аналіз вмісту запитів є обмеженим, а ключову роль відіграють потокові та поведінкові характеристики. Запропоновано підхід, що базується на ансамблевих моделях градієнтного бустингу дерев рішень і орієнтований на класифікацію мережевих потоків на нормальні та атакувальні. Експериментальну перевірку виконано на даних CSE-CIC-IDS2018; застосовано очищення та нормалізацію даних, а також балансування класів через контрольоване зменшення кількості нормальних потоків для зниження впливу дисбалансу. Для виявлення атак використано моделі LightGBM і XGBoost із типовими налаштуваннями для бінарної класифікації табличних ознак; оцінювання здійснено на відкладеній тестовій множині із застосуванням стандартних метрик якості та аналізом матриці помилок. Отримані результати демонструють дуже високу відокремлюваність класів і малу кількість помилок класифікації, причому XGBoost показує дещо кращий баланс між повнотою виявлення атак і кількістю помилкових спрацювань порівняно з LightGBM. Проаналізовано внесок ознак у рішення моделей; найбільш інформативними виявляються статистики довжин пакетів, часових інтервалів, співвідношення напрямків трафіку та атрибути транспортного рівня, що відображають структуру й динаміку мережевої взаємодії. Разом з тим відзначено, що близькі до граничних значення метрик у контрольованих умовах можуть частково залежати від специфіки формування датасету та наявності ознак, пов'язаних із сценаріями генерації трафіку, тому інтерпретацію результатів подано з урахуванням загроз валідності. Практична цінність роботи полягає у відтворюваному підході до побудови детектора веб-атак на потокових ознаках, порівнянні двох реалізацій градієнтного бустингу та формуванні рекомендацій щодо подальшої перевірки узагальнюваності на альтернативних схемах валідації та додаткових вибірках.

Ключові слова: веб-атаки; виявлення атак; мережевий трафік; HTTP; машинне навчання; градієнтний бустинг; LightGBM; XGBoost.

ВСТУП

Сучасні засоби захисту веб-додатків, зокрема WAF на основі правил і сигнатур, широко застосовуються в корпоративних середовищах, однак потребують постійного ручного супроводу. Для зниження кількості хибних спрацьовувань зазвичай виконуються тонке налаштування правил, коригування порогів, додавання виключень та адаптація під конкретні шаблони легітимного трафіку. Такий підхід є трудомістким, залежить від досвіду фахівців і не гарантує стійкості до модифікованих або нових варіантів атак, оскільки правила часто відстають від змін у поведінці зловмисників. У результаті виникає потреба у методах, здатних автоматизовано враховувати складні закономірності трафіку та зменшувати навантаження на процес експлуатації захисних механізмів.

Постановка проблеми. Корпоративні веб-додатки є критично важливими елементами інформаційної інфраструктури організацій, оскільки забезпечують доступ до сервісів, даних і бізнес-логіки. Зростання кількості веб-загроз, зокрема атак типу Brute Force, SQL Injection, Cross-Site Scripting (XSS) та їх модифікацій, підвищує вимоги до точності й адаптивності засобів виявлення та блокування атак на прикладному рівні. Додатковою складністю для мережевого моніторингу є домінування шифрованого трафіку, що обмежує застосовність підходів, орієнтованих на аналіз вмісту пакетів, і зумовлює перехід до аналізу ознак потоку та поведінкових характеристик з'єднань [3]. У практичних умовах це формує суперечність між потребою у високій якості детектування та обмеженнями щодо доступності інформативних ознак, продуктивності, а також прийнятного рівня помилкових спрацювань.

Аналіз останніх досліджень і публікацій. У сучасних роботах з веб-захисту підкреслюється, що класичні веб-аплікаційні брандмауери, орієнтовані на правила, залишаються ефективними проти відомих шаблонів атак, однак можуть продукувати хибні спрацювання та бути вразливими до обхідних прийомів на кшталт кодування й обфускації, що стимулює розвиток підходів на основі машинного навчання та глибинних моделей для аналізу веб-запитів і аномалій у трафіку [1], [2].

Для задач мережевого виявлення атак активно застосовуються підходи машинного навчання, що працюють із потоковими ознаками, зокрема на наборі даних CSE-CIC-IDS2018, сформованому в рамках співпраці Communications Security Establishment і Canadian Institute for Cybersecurity, який містить кілька сценаріїв атак, включно з веб-атаками, та супроводжується набором статистичних ознак мережевих потоків, отриманих засобами CICFlowMeter [4], [5], [6].

Водночас у літературі наголошується на ризику завищених оцінок якості, коли моделі демонструють близькі до ідеальних метрики на контрольованих датасетах, але такі результати можуть погано узагальнюватися при зміні середовища, сценаріїв атак або наборів ознак; зокрема показано, що вибір стандартного набору ознак суттєво впливає на переносимість детекторів між різними датасетами та умовами збору трафіку [7].

Окремим чинником, що ускладнює виявлення веб-атак у корпоративних мережах, є домінування шифрованих з'єднань, через що детектування часто спирається на побічні та агреговані характеристики потоків, а не на вміст корисного навантаження, що посилює вимоги до обґрунтованого відбору ознак і коректної методики оцінювання [3].

Мета статті. Метою статті є розроблення та експериментальна перевірка методу виявлення атак на корпоративні веб-додатки на основі градієнтного бустингу з використанням датасету CSE-CIC-IDS2018, із застосуванням моделей LightGBM та XGBoost як ефективних реалізацій бустингу дерев рішень для табличних ознак [8], [9].

ТЕОРЕТИЧНІ ОСНОВИ ДОСЛІДЖЕННЯ

Веб-атаки у корпоративних мережах реалізуються через різнорідні техніки, спрямовані на порушення конфіденційності, цілісності та доступності інформаційних ресурсів. До типових векторів належать атаки підбору облікових даних (brute force) під час автентифікації, ін'єкції SQL-запитів для несанкціонованого читання або модифікації даних у СУБД, міжсайтовий скриптинг (XSS) як ін'єкція шкідливого клієнтського коду, а також міжсайтове підроблення запитів (CSRF) і прикладні атаки на відмову в обслуговуванні на рівні HTTP-запитів (HTTP flood), що імітують або надмірно масштабують легітимну активність користувачів [11].

МЕТОДИКА ДОСЛІДЖЕННЯ

Для моделювання та аналізу вказаних загроз застосовано датасет CSE-CIC-IDS2018, який містить зразки веб-атак у межах сценаріїв датасету та представлений у вигляді мережевих потоків із набором ознак, що характеризують адресацію та порти джерела/призначення, часові й статистичні параметри потоків, а також похідні характеристики транспортного рівня, зокрема пов'язані з TCP. З метою підвищення повноти аналізу додатково включено HTTPS-трафік, який у практичних корпоративних середовищах переважно відповідає порту 443, що забезпечує застосовність моделі в умовах домінування шифрованих з'єднань і обмеженої доступності вмістових ознак. Попередній аналіз сформованої підвибірки засвідчив наявність вираженого дисбалансу класів: BENIGN – близько 1 159 309 потоків, тоді як сукупність атак становить 438 потоків.

Така нерівномірність потребує спеціальних технік балансування та коректного вибору метрик.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Для формування вибірки використано два файли датасету CICIDS2018: Thursday-22-02-2018.csv та Friday-23-02-2018.csv, що містять зразки трафіку з веб-атаками в межах сценаріїв датасету. Після попереднього очищення даних і балансування класів сформовано підвибірку обсягом 20 438 потоків**,** з яких 438 належать до класу *Attack*, а 20 000 – до класу *Benign* (контрольований *undersampling*). Подальший поділ даних на тренувальну та тестову множини виконано у співвідношенні 80/20**.**

Для розв'язання задачі бінарної класифікації (*Attack* / *Benign*) застосовано модель LightGBM з цільовою функцією objective="binary". Навчання виконано з параметрами: learning_rate=0.1, num_leaves=31, n_estimators=200, random_state=42, що забезпечує відтворюваність результатів та компроміс між швидкістю навчання і якістю узагальнення.

За результатами оцінювання на тестовій множині отримано такі значення метрик:

Accuracy: 0.9995

Precision: 0.9886

Recall: 0.9886

F1-score: 0.9886

AUC-ROC: 0.999997

Високі значення Accuracy і AUC-ROC свідчать про значну відокремлюваність класів у межах сформованої вибірки та обраної схеми розбиття даних.

Аналіз матриці помилок показав, що модель коректно класифікує 3999 із 4000 потоків класу Benign та правильно виявляє 87 із 88 потоків класу Attack. Таким чином, спостерігається лише 1 хибне спрацювання для нормального трафіку та 1 пропущена атака, що узгоджується зі значеннями Precision/Recall, близькими до 0.99.

Інтерпретація важливості ознак демонструє, що найбільший внесок у рішення моделі роблять статистичні характеристики трафіку, зокрема Bwd Packet Length Std та Fwd Packet Length Std, а також мережеві атрибути Src Port і Dst Port. Додатково значущими виявилися Down/Up Ratio та Flow IAT Min, які відображають співвідношення напрямків трафіку та мінімальні інтервали між пакетами. Сукупно ці ознаки

характеризують інтенсивність, структуру та часову динаміку потоків, що може бути корисним для відокремлення аномального трафіку від нормального.

Для порівняння використано модель XGBoost з логістичною функцією втрат для бінарної класифікації (objective="binary:logistic") та метрикою оптимізації eval_metric="auc". Параметри навчання встановлено як learning_rate=0.1, max_depth=6, n_estimators=200. Така конфігурація є типовою для задач із табличними ознаками та дозволяє моделі формувати нелінійні правила на основі ансамблю дерев рішень.

Отримані результати на тестовій множині:

Accuracy: 0.99975

Precision: 0.9887

Recall: 1.0

F1-score: 0.9943

AUC-ROC: 0.999997

Повнота Recall = 1.0 означає, що в межах тестової підвибірки не зафіксовано хибних негативів, тобто атаки не були пропущені. Водночас за матрицею помилок спостерігається лише 1 хибний позитив, тобто одне помилкове спрацювання на нормальний трафік.

За оцінкою важливості ознак найбільший внесок у рішення XGBoost мають Fwd PSH Flags, Attempted Category, Fwd Packet Length Std та Dst Port. Наявність серед лідерів таких атрибутів, як Dst Port і транспортні прапорці, вказує на високу інформативність мережевих патернів у даних. Окремо слід зазначити ознаку Attempted Category: якщо вона походить із сценарної/категорійної розмітки датасету, її використання може наближати модель до непрямого відтворення мітки та потенційно завищувати оцінку якості. Тому інтерпретація майже граничних значень AUC-ROC та Accuracy повинна виконуватися з урахуванням можливого впливу специфіки формування датасету.

Додатково слід враховувати, що майже граничні значення AUC-ROC та Recall можуть частково пояснюватися властивостями формування датасету та наявністю ознак, які відображають специфіку середовища або сценаріїв генерації трафіку. Зокрема, серед найбільш впливових ознак присутні параметри портів і транспортного рівня (наприклад, Src Port, Dst Port, TCP-прапорці), а також агреговані статистики таймінгу/довжин пакетів. Такі характеристики можуть виступати маркерами конкретних інструментів або умов збору даних і не завжди є універсальними ознаками суті атаки. Окремої уваги потребують службові/категорійні атрибути на кшталт Attempted Category, які потенційно можуть містити інформацію, близьку до цільової мітки, що підвищує ризик завищення оцінки якості. У межах поточної роботи зазначені фактори розглядаються як загрози валідності, а перевірка стійкості моделей шляхом виключення потенційно “маркерних” ознак та альтернативних схем розбиття даних визначається як напрям подальших досліджень.

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

У роботі реалізовано підхід до виявлення веб-атак у мережевому трафіку на основі градієнтного бустингу дерев рішень із використанням потокових ознак. Навчені моделі LightGBM та XGBoost показали високу якість класифікації на сформованій вибірці, а аналіз важливості ознак підтвердив значущість статистик довжин пакетів, часових характеристик і транспортних атрибутів.

Подальші дослідження доцільно спрямувати на перевірку узагальнюваності результатів: виключення потенційно маркерних ознак, застосування альтернативних

схем валідації (розбиття за часом або сценаріями), а також тестування на додаткових датасетах і вибірках, наближених до умов корпоративної експлуатації.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

pplebaum, S., Gaber, T., & Ahmed, A. (2021). Signature-based and machine-learning-based web application firewalls: A short survey. Procedia Computer Science, 189, 358–367.

elan, P., Čermák, M., Čeleda, P., & Drašar, M. (2015). A survey of methods for encrypted traffic classification and analysis. International Journal of Network Management, 25(5), 355–374.

characterization. Proceedings of the 4th International Conference on Information Systems Security and Privacy (ICISSP), 108–116. https://doi.org/10.5220/0006639801080116

Conference on Knowledge Discovery and Data Mining, 785–794.

riedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5), 1189–1232.

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 3 (31), 2025
ISSN 2663 - 4023

Anna Boiko
Senior Lecturer, Department of Cybersecurity Systems and Technologies
State University of Information and Communication Technologies, Kyiv, Ukraine
ORCID: 0009-0001-3709-6283
a.boiko@duikt.edu.ua

METHOD FOR DETECTING ATTACKS ON CORPORATE WEB APPLICATIONS BASED ON GRADIENT BOOSTING

Abstract. The paper addresses the problem of detecting web attacks in the network traffic of corporate web applications under the predominance of encrypted connections, where inspection of request contents is limited and flow-level and behavioral characteristics play a key role. An approach is proposed based on gradient boosting decision tree ensembles and aimed at classifying network flows as benign or malicious. The experimental evaluation is conducted on the CSE-CIC-IDS2018 dataset; the pipeline includes data cleaning and normalization, as well as class balancing through controlled reduction of benign flows to mitigate the impact of class imbalance. LightGBM and XGBoost models with standard configurations for binary classification on tabular features are used for attack detection; evaluation is performed on a held-out test set using common performance metrics and confusion-matrix analysis. The obtained results indicate very strong class separability and a low number of classification errors, while XGBoost provides a slightly better trade-off between attack detection completeness and the false alarm rate compared to LightGBM. Feature contributions to model decisions are analyzed; the most informative predictors are packet-length statistics, inter-arrival time measures, traffic directionality ratios, and transport-layer attributes reflecting the structure and dynamics of network interactions. At the same time, the near-ceiling metric values observed under controlled conditions may be partially influenced by dataset construction specifics and by features associated with traffic-generation scenarios; therefore, the findings are interpreted with explicit validity considerations. The practical value of the work lies in a reproducible flow-based web attack detection pipeline, a comparison of two gradient boosting implementations, and recommendations for further generalization checks using alternative validation schemes and additional datasets.

Keywords: web attacks; attack detection; network traffic; HTTP; machine learning; gradient boosting; LightGBM; XGBoost.

REFERENCES

1. Applebaum, S., Gaber, T., & Ahmed, A. (2021). Signature-based and machine-learning-based web application firewalls: A short survey. Procedia Computer Science, 189, 358–367. https://doi.org/10.1016/j.procs.2021.05.105
2. OWASP Core Rule Set Project. (n.d.). False positives and tuning. CRS Documentation. https://coreruleset.org/docs/
3. Velan, P., Čermák, M., Čeleda, P., & Drašar, M. (2015). A survey of methods for encrypted traffic classification and analysis. International Journal of Network Management, 25(5), 355–374. https://doi.org/10.1002/nem.1901
4. Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. Proceedings of the 4th International Conference on Information Systems Security and Privacy (ICISSP), 108–116. https://doi.org/10.5220/0006639801080116
5. Canadian Institute for Cybersecurity. (2018). IDS 2018 dataset. University of New Brunswick. https://www.unb.ca/cic/datasets/ids-2018.html
6. Registry of Open Data on AWS. (n.d.). A realistic cyber defense dataset (CSE-CIC-IDS2018). https://registry.opendata.aws/cse-cic-ids2018/
7. Sarhan, M., Layeghy, S., & Portmann, M. (2022). Evaluating standard feature sets towards increased generalisability and explainability of ML-based network intrusion detection. Big Data Research, 30, 100345. https://doi.org/10.1016/j.bdr.2022.100345
8. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 30.

9. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785
10. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5), 1189–1232.
11. OWASP. (2021). OWASP Top 10: 2021. https://owasp.org/Top10/