1 ВСТУП

1.1 Огляд продукту

Система, описана у кваліфікаційній роботі Трипілки А. В., є комплексним рішенням для онлайн‑моніторингу та виявлення розподілених атак відмови в обслуговуванні (DDoS) у зашифрованому TLS‑трафіку. На відміну від традиційних сигнатурних IDS, що аналізують payload, запропонований підхід базується виключно на метаданих мережевих потоків. Система виконує багатокроковий аналіз: спочатку AutoEncoder відсіює аномальні потоки, а далі рекурентна мережа LSTM класифікує тип атаки. Усі операції відбуваються у режимі реального часу у хмарному середовищі AWS, що доведено експериментально.

Ключові переваги рішення: робота без розшифрування пакету, масштабованість завдяки контейнерній орієнтації та можливість інтеграції з SOC через стандартизований REST API.

1.2 Мета

Метою роботи є дослідження методів аналізу зашифрованого мережевого трафіку та створення прототипу, що забезпечує точне й швидке виявлення DDoS‑атак без необхідності доступу до вмісту TLS‑з’єднань.

1.3 Межі

Система працює лише з метаданими NetFlow/IPFIX або згенерованих CICFlowMeter потоків; розшифрування трафіку або DPI‑аналіз не передбачені. Модель протестовано на 17 000 потоках із публічних датасетів CIC‑DDoS2019 та CSE‑CIC‑IDS2018, що підтверджує її застосовність до різних протоколів і сценаріїв атак.

Продуктивність обмежена 1 000 потоками за ≤5 с на одній інстанції; горизонтальне масштабування у Kubernetes дозволяє лінійно збільшувати пропускну здатність.

1.4 Посилання

В проєкті були використані такі посилання:

1. A Survey on Internet Traffic Identification / A. Callado та ін. IEEE Communications Surveys & Tutorials. 2009. Т. 11, № 3. С. 37–52. URL: https://doi.org/10.1109/surv.2009.090304 (дата звернення: 05.02.2025).

2. Bhatia S., Behal S., Ahmed I. Distributed Denial of Service Attacks and Defense Mechanisms: Current Landscape and Future Directions. SpringerLink. URL: https://link.springer.com/chapter/10.1007/978-3-319-97643-3\_3 (дата звернення: 06.02.2025).

3. Deep packet: a novel approach for encrypted traffic classification using deep learning / M. Lotfollahi та ін. Soft Computing. 2019. Т. 24, № 3. С. 1999–2012. URL: https://doi.org/10.1007/s00500-019-04030-2 (дата звернення: 08.02.2025).

4. Sharma A., Lashkari A. H. A survey on encrypted network traffic: A comprehensive survey of identification/classification techniques, challenges, and future directions. Computer Networks. 2025. Т. 257. С. 110984. URL: https://doi.org/10.1016/j.comnet.2024.110984 (дата звернення: 10.02.2025).

5. Anderson B., McGrew D. Machine Learning for Encrypted Malware Traffic Classification: Accounting for Noisy Labels and Non-Stationarity. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2017.

6. A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks / C. Yin та ін. IEEE Access. 2017. Т. 5. С. 21954–21961. URL: https://doi.org/10.1109/access.2017.2762418 (дата звернення: 15.02.2025)..

7. Adversarial Autoencoders / A. Makhzani та ін. arXiv.org. URL: https://arxiv.org/abs/1511.05644 (дата звернення: 05.04.2025).

8. Deep learning approach for Network Intrusion Detection in Software Defined Networking / T. A. Tang та ін. 2016 International Conference on Wireless Networks and Mobile Communications. 2016..

9. Lee M.-C., Lin J.-C., Katsikas S. Impact of Recurrent Neural Networks and Deep Learning Frameworks on Real-time Lightweight Time Series Anomaly Detection. 2024.

10. Лазар М. Технології big data у аналізі ризиків страхової компанії / М. Лазар, В. Кобзєв // Інформаційні системи та технології : матеріали статей 7-ї Міжнародної науково-технічної конференції, Коблеве-Харків, 10-15 вересня 2018 р. – Харків : ХНУРЕ, 2018. – С. 364–367.

11. Chala O., Bodyanskiy Y. Matrix Neo-Fuzzy-System and its Online Learning in Image Recognition Task. Information Technology and Management Science. 2021. Т. 24. С. 39–44. URL: https://doi.org/10.7250/itms-2021-0006 (дата звернення: 06.04.2025)..

12. An Approach to the Selection of Behavior Patterns Autonomous Intelligent Mobile Systems / O. Zolotukhin та ін. 2021 IEEE 8th International Conference on Problems of Infocommunications, Science and Technology (PIC S&T). 2021.

1.5 Означення та абревіатури

|  |  |
| --- | --- |
| TLS – Transport Layer Security | Протокол шифрування транспортного рівня. |
| Flow | Агрегований запис характеристик мережевого сеансу (5‑tuple). |
| AutoEncoder | Нейронна мережа без вчителя для відтворення нормального трафіку та пошуку аномалій. |
| LSTM – Long Short‑Term Memory | Тип RNN, що аналізує часові залежності. |
| SOC | Security Operations Center. |
| SIEM/SOAR | Платформи збору та кореляції подій безпеки та автоматизованого реагування. |
| TPR | Повнота виявлення атак. |
| FPR | Частка хибнопозитивних спрацювань. |

2 ЗАГАЛЬНИЙ ОПИС

2.1 Перспективи продукту

Подальший розвиток прототипу системи виявлення DDoS‑атак у зашифрованому трафіку передбачає перехід до хмарної багатозонної інфраструктури зі автоматичним масштабуванням: використання керованих сервісів Amazon RDS Multi‑AZ для PostgreSQL та Amazon Timestream для time‑series аналітики з архівацією історичних даних у S3 підвищить відмовостійкість і спростить довгострокове зберігання. Для обробки десятків мільйонів мережевих потоків буде реалізовано горизонтальне масштабування баз даних і потокової обробки шляхом шардінгу або використання TimescaleDB, а всі компоненти будуть контейнеризовані у Docker/Kubernetes.

Передбачено інтеграцію з потоковими платформами Apache Kafka або Pulsar, що забезпечить стійкий прийом NetFlow/IPFIX‑логів у реальному часі та розвантажить основні мікросервіси. Глибша інтеграція із SOC‑ланцюгом включатиме розширення webhook‑механізмів для автоматичного реагування в SIEM/SOAR‑платформах і передачу даних у форматах STIX/JSON поверх уже наявного REST‑API.

Моделі AutoEncoder і RNN регулярно перенавчатимуться без простою сервісу; планується тонке налаштування гіперпараметрів і експерименти з VAE, attention‑механізмами та GNN для підвищення точності понад 0,95 F1‑міри.

Єдиний стек спостережності на основі Prometheus і Grafana буде доповнено продуманими SLA‑алармами та бізнес‑дашбордами, а інтерфейс аналітика розшириться топологічними картами джерел атак, тепловими картами пікових навантажень і можливістю тонкого налаштування порогів спрацювання через GUI.

2.2 Функції продукту

* Збір PCAP / NetFlow/IPFIX у реальному часі.
* Автоматичне формування ознак CICFlowMeter 4.0.
* Попередня обробка й нормалізація в Apache Beam.
  + Виявлення аномалій AutoEncoder.
  + Класифікація LSTM RNN.
* Збереження результатів у PostgreSQL та InfluxDB.

2.3 Характеристики користувачів

Типові користувачі:

* Аналітик SOC‑центру, що відстежує поточні загрози.
* Інженер DevSecOps, який супроводжує інфраструктуру Kubernetes.

2.4 Загальні обмеження

Система не виконує активного блоку трафіку; інтеграція з фаєрволами можлива через SOAR. Виявлення конкретних IoT‑ботнетів, що змінюють шаблони у реальному часі, потребує перенавчання моделі.

2.5 Припущення та залежності

* Python 3.10 + PyTorch 2.1.0
* Docker 26.0; Docker Compose для локальних стендів
* Kubernetes 1.30 + Helm для хмари
* Prometheus 2.52, Grafana 11
* PostgreSQL 15, InfluxDB 2.7

3 КОНКРЕТНІ ВИМОГИ

3.1 Вимоги до зовнішніх інтерфейсів

3.1.1 Інтерфейс користувача

Головна сторінка дашборда показує метричну панель: TPR, FPR, графік потоків за останні 24 год, карту джерел атак та таблицю TOP‑IP. Деталізація потоку відкриває хронологію пакетів, значення ознак та результат нейромереж.

3.1.2 Програмний інтерфейс

REST API відповідає OpenAPI 3.0. Авторизація – JWT; токен видається окремим Auth‑сервісом. Максимальний розмір відповіді – 1 МБ, пагінація – 500 записів.

3.1.3 Продуктивність

Затримка від появи потоку до ALARM ≤ 5 с при 1 000 потоках/сек. Модель AE виконує інференс на CPU (i7‑11700) за 32 мс, LSTM – 44 мс на GPU T4.

3.2 Властивості програмного продукту

Надійність – робочі служби відокремлені від керуючих, а Kubernetes сам відновлює будь‑які збої.

Доступність – сервіс працює у двох дата‑центрах AWS і гарантує 99,9 % часу безвідмовної роботи.

Безпека – доступ до дашборда регулюється ролями, а всі мікросервіси спілкуються через TLS 1.3.

Супроводжуваність – понад 60 % коду покрито тестами, а GitHub Actions автоматично збирає та розгортає оновлення.

Переносимість – контейнери однаково запускаються на процесорах x86‑64 і ARM64.

Масштабованість – Kubernetes автоматично додає поди, а Kafka розподіляє черги повідомлень.

3.3 Вимоги до бази даних

У Postgres лежать дві основні таблиці:

* Flows – один рядок = один мережевий сеанс (час початку й кінця, адреси джерела й призначення, скільки байтів пішло в кожен бік тощо).
* Alerts – запис про спрацювання системи (який сеанс, коли “загорілося”, який тип атаки та наскільки впевнена модель).

У InfluxDB зберігаємо показники моделі щосекунди: наскільки великий збій у AutoEncoder (AE\_error) і наскільки впевнена RNN (RNN\_prob).

3.4 Алгоритми

AutoEncoder – 5 шарів, активація ReLU, оптимізатор Adam lr=0.001, поріг Reconstruction Error = 0.045 (квантіль 0.95). RNN – 2 шару, 64 units, bidirectional, softmax на виході (5 класів). Навчання: 50 епох, batch = 128, early stop = patience 7.

3.5 Експериментальні результати

Проведено 8 серій навчання/валідації на змішаному датасеті (17 032 потоки). Середня точність 87.4 %, TPR 89.6 %, FPR 9.2 %. Обробка всього датасету зайняла 43 с на хмарному екземплярі EC2 m6i.large.