СИСТЕМНЫЙ АНАЛИЗ, УПРАВЛЕНИЕ И ОБРАБОТКА ИНФОРМАЦИИ / SYSTEM ANALYSIS, MANAGEMENT AND PROCESSING OF INFORMATION

DOI: https://doi.org/10.23670/IRJ.2023.132.18

СРАВНИТЕЛЬНЫЙ АНАЛИЗ СИСТЕМ ОБНАРУЖЕНИЙ АНОМАЛИЙ С ИСПОЛЬЗОВАНИЕМ ПОТОКА СЕТЕВОГО ТРАФИКА И ПРОТОКОЛА NETFLOW

Научная статья

Голованов А.А.[1, *], Мельникова О.И.[2]
[1] ORCID : 0000-0002-8810-4909;
[1, 2] Государственный университет Дубна, Дубна, Российская Федерация

* Корреспондирующий автор (golovanov-asus-p5b-mx2011[at]yandex.ru)

Аннотация

Обнаружение аномалий в потоке сетевого трафика является важнейшей задачей в современной сетевой безопасности, где основной целью является выявление любого ненормального поведения трафика и подача сигналов тревоги для предотвращения потенциальных угроз безопасности. Существуют различные методы, доступные для обнаружения аномалий в потоке сетевого трафика, включая обнаружение на основе сигнатур, статистическое обнаружение и обнаружение на основе машинного обучения. В последние годы системы обнаружения аномалий на основе протокола NETFLOW приобрели значительную популярность благодаря своей способности предоставлять подробную информацию о поведении сетевого трафика.

В данной статье рассматриваются и сравниваются две системы определения аномалий в сетевом трафике с использованием искусственных нейронных сетей. Мы использовали общедоступные наборы данных для обучения систем. Затем мы сравнили результаты и проанализировали преимущества и недостатки каждой системы. Каждая система использует различный тип нейронных сетей: многослойная нейронная сеть и рекуррентная нейронная сеть. Основными критериями оценки качества системы были выбраны ROC метрика и площадь под кривой ROC, которые позволили определить эффективность используемых методов в определении аномалий.

Ключевые слова: аномалии, сетевой трафик, искусственные нейронные сети, системы обнаружения аномалий.

A COMPARATIVE ANALYSIS OF ANOMALY DETECTION SYSTEMS USING NETWORK TRAFFIC FLOW AND NETFLOW PROTOCOL

Research article

Golovanov A.A.[1, *], Melnikova O.I.[2]
[1] ORCID : 0000-0002-8810-4909;
[1, 2] Dubna State University, Dubna, Russian Federation

* Corresponding author (golovanov-asus-p5b-mx2011[at]yandex.ru)

Abstract

Anomaly detection in network traffic flow is the most important task in today's network security, where the main goal is to detect any abnormal traffic behaviour and raise alerts to prevent potential security threats. There are various methods available to detect anomalies in the network traffic flow, including signature-based detection, statistical detection, and machine learning-based detection. In recent years, NETFLOW-based anomaly detection systems have gained significant popularity due to their ability to provide detailed information about the behaviour of network traffic.

This article examines and compares two systems for detecting anomalies in network traffic using artificial neural networks. We used publicly available datasets to train the systems. We then compared the results and analysed the pros and cons of each system. Each system uses a different type of neural network: multilayer neural network and recurrent neural network. The ROC metric and the area under the ROC curve were chosen as the main criteria for evaluating the quality of the system, which allowed to determine the effectiveness of the methods used in detecting anomalies.

Keywords: anomalies, network traffic, artificial neural networks, anomaly detection systems.

Введение

Объемы интернет-трафика и количество пользователей сети интернет продолжает неуклонно расти во всем мире. По данным сервиса Telegeography, отслеживающего изменения во всемирной сети Интернет, сделано заключение, что средний международный интернет-трафик увеличился примерно со 120 Тбит/с до 170 Тбит/с с 2019 по 2020 год, также за 2021 год увеличилась пропускная способность мировой сети на 29%, тем самым достигнув отметки в 786 Тбит/с [4]. Приведенные данные говорят о все увеличивающемся объеме данных, проходящих через сетевую инфраструктуру, а значит и об увеличении сетевых аномалий в проходящем трафике. Огромной задачей для всей сетевой инфраструктуры является определение сетевых аномалий, оповещение и ее ликвидация. Поставленную задачу решают системы обнаружения аномалий (СОА), чаще построенные на сигнатурном методе определения, однако, в связи с развитием вычислительных мощностей, все чаще используется нейросетевой метод [3], [4], [5]. Инструменты, использующиеся в анализе больших данных, могут помочь в создании системы обнаружения аномалий, но в данном случае время обработки данных будет увеличиваться пропорционально объему сетевого трафика, проходящего за единицу времени. В работе проведено исследование различных подходов к построению системы обнаружения аномалий на основе нейронных сетей. В первом случае система на вход использует «сырые данные», то есть

необработанный сетевой трафик. Во втором случае используется протокол Netflow, позволяющий получить потоки данных из трафика уже в обработанном виде и передать их на вход в систему.

Цель исследования заключается в сравнении двух СОА на идентичных данных сетевого трафика и выявлении лучшей системы в способе определении аномалии таких типов атак: DDoS, Bruteforce, PortScan.

Мы предполагаем, что каждая из систем имеет свои преимущества и недостатки, и задачей является анализ этих систем и проведение эксперимента по выявлению преимуществ и недостатков в каждой из них. Входные данные будут пропущены через разные модули предобработки данных и переданы двум основным компонентам СОА: модулям обнаружения аномалий и модулям классификации аномалий Нами определяются критерии оценки систем и анализируется результат работы каждой системы.

Обзор информационной системы определения сетевых атак на основе рекуррентных нейронных сетей

Первая система использует архитектуру рекуррентной сети LSTM (Long short-term memory). Преимуществом нейронных сетей LSTM является преодоление проблемы долговременной зависимости, при которой сеть теряет способность связывать информацию [8]. С такой проблемой сталкиваются рекуррентые нейронные сети (RNN) [5], [7]. LSTM сети используют фильтры, которые позволяют пропускать информацию на основе некоторых условий и изменять внутреннее состояние блоков называемыми ячейками памяти. Фильтры состоят из слоя сигмоидальной нейронной сети и операции поточечного умножения. Выделяют 3 основных фильтра: фильтры ввода (input gates), фильтры вывода (output gates) и фильтры забывания (forget gates) (Рисунок 1).

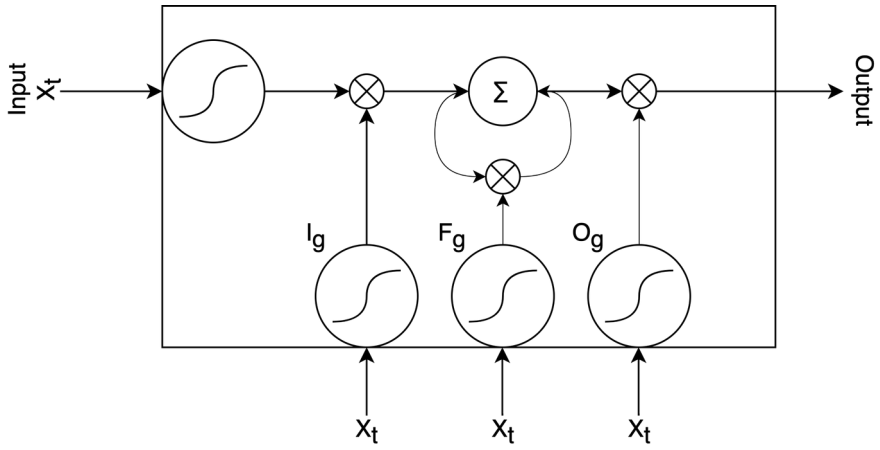

Рисунок 1 - Структура LSTM сети
DOI: https://doi.org/10.23670/IRJ.2023.132.18.1

Система обнаружения аномалий на основе рекуррентных нейронных сетей состоит из 5 основных модулей:

– модуль коллектора потока трафика – собирается сетевой трафик и сохраняет дампы трафика;

– модуль подготовки данных – обрабатывает сырые дампы трафика и подготавливает данные для модуля обнаружения аномалий;

– модуль обнаружения аномалий – функция, разделяющая данные на нормальный поток трафика и аномальный поток (сетевая атака);

– модуль классификации аномалий – дополнительный модуль, состоящий из 3 функций классифицирующий аномальный поток;

– модуль оповещения – оповещает администратора системы о наличии аномалии в потоке данных.

Модель обучалась на наборе данных CICIDS2017 и CICIDS2018 (CIC) [5], [7], который подвергался предварительной обработке: масштабирование и нормализация набора данных, разделение набора на обучающий и тестируемый, преобразования типов данных.

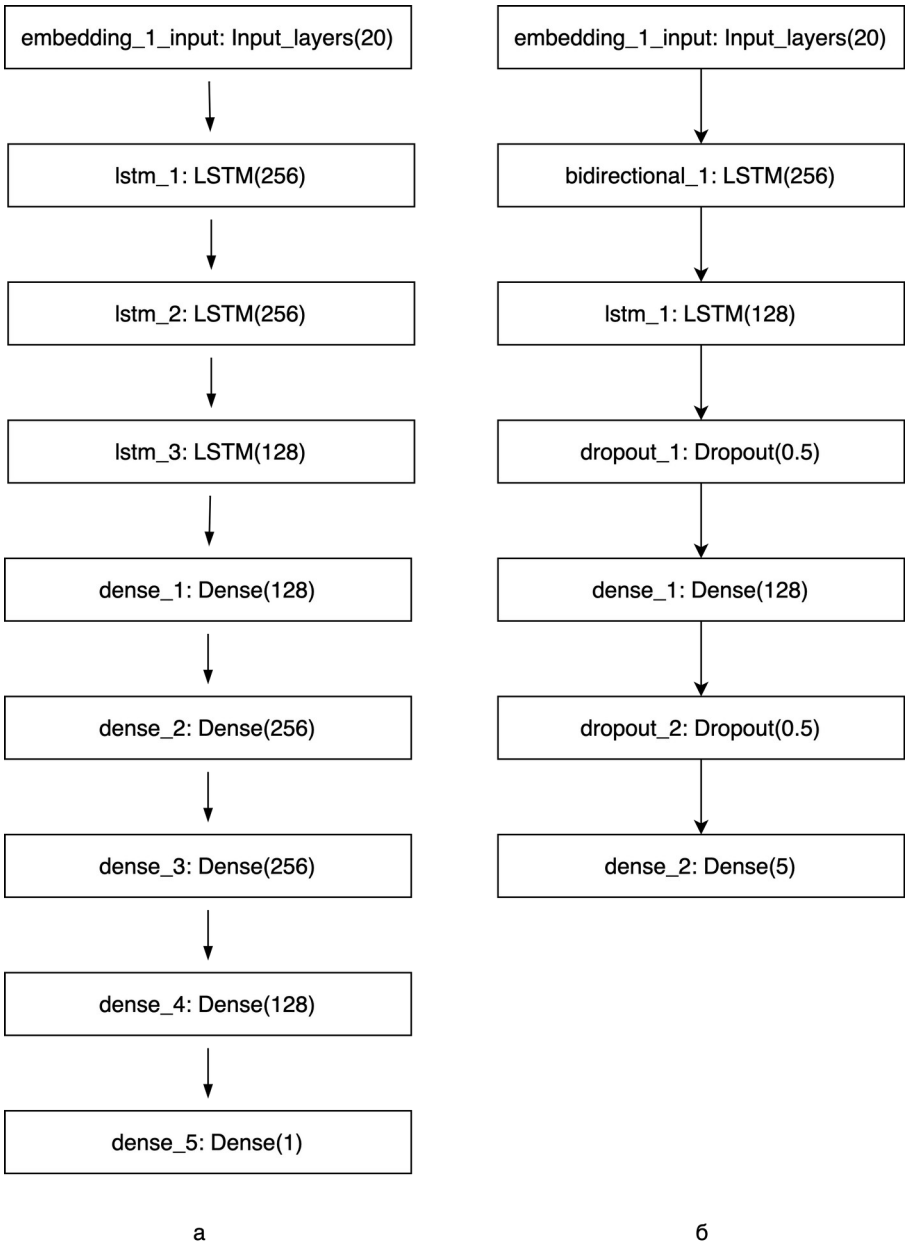

Рисунок 2 - Архитектура модуля обнаружения (*а*) и модуля классификации аномалии (*б*)
DOI: https://doi.org/10.23670/IRJ.2023.132.18.2

В обучении сети использовалось 2 подхода:

– Обучение с 32 гиперпараметрами, имеющими ненулевое влияние на целевую переменную;

– Обучение с 20 гиперпараметрами, имеющими 93% влияния на целевую переменную.

Модуль классификатора аномалий обучился с точностью 82%, а модуль обнаружения аномалий обучился с точностью определения 72%.

Обзор информационной системы обнаружения аномалий IP-трафика по протоколу Net-Flow v9 с использованием глубокого обучения

Вторая система использует классическую многослойную искусственную сеть. Архитектура системы обнаружения аномалий с использованием протокола Netflow похожа на архитектуру вышеописанной системы, в ее состав входит:

– роутер BRAS с сенсером Netflow – обеспечивает детерминирования абонентской сессии и передачи потоков Netflow;

– коллектор Netflow – виртуальная машина с установленным коллектором nfdump для сбора Netflow потоков;

– конвертор – скрипт на языке bash преобразующий данные из формата nfcapd в csv формат;

– анализатор – модуль нейронной сети, занимающийся обработкой и классификацией данных;

– модуль оповещения – производит оповещение администратора.

Для обучения нейронной сети использовался набор данных UGR16, он также разделен на обучающий набор – 7 млн. строк, тестовый набор – 1.5 млн. строк и контрольный набор – 1.5 млн. строк. В качестве архитектуры использовалась многослойная сеть размерностью (64, 32, 32, 2) и размерностью (64, 32, 32 3).

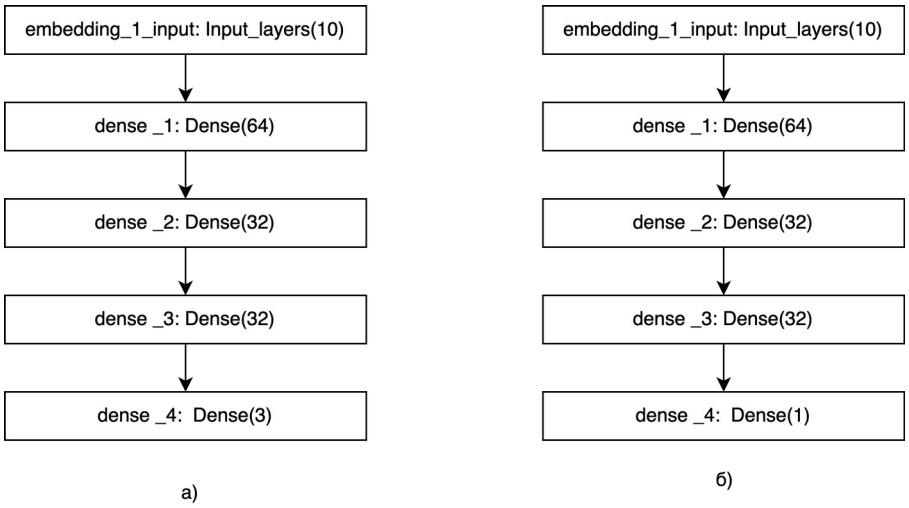

Рисунок 3 - Архитектура модуля обнаружения (*б*) и модуля классификации аномалии (*а*)
DOI: https://doi.org/10.23670/IRJ.2023.132.18.3

В первом случае сеть используется для определения аномалии, во втором случае для классификации аномалии (Рисунок 3). На контрольных данных нейронная сеть классифицировала DoS с точностью 92%, Portscan 76%, Bruteforce 74%, с точностью 82,7% сеть определила аномалию и с точностью 83,32% определила нормальный трафик. Результаты проведенных исследований можно объединить для проверки работы СОА на идентичных наборах реальных данных.

Подготовка эксперимента

Описанные СОА имеют между собой много общего: количество модулей, одинаковые выполняемые функции каждого модуля, главным отличием между ними является модуль классификации, в которой расположена искусственная нейронная сеть.

Для проверки поставленной цели, разработана новая архитектура системы обнаружения аномалий, включающая в себя две архитектуры из вышеуказанных исследований (Рисунок 4). В обновленной архитектуре системы разбиты на 2 контура, в которые будут поступать данные.

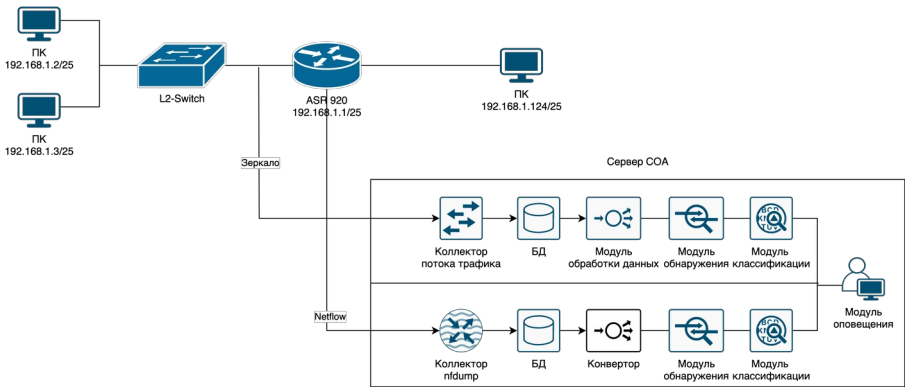

Рисунок 4 - Обновленная архитектура СОА
DOI: https://doi.org/10.23670/IRJ.2023.132.18.4

Роутер ASR920 соединен двумя линками Ten Gigabit Ethernet с сервером СОА. По одному линку происходит зеркалирование входящего трафика с роутера BRAS на модель коллектора потока трафика сервер СОА, по другому происходит передача данных Netflow на коллектор nfdump сервера СОА. При такой коммутации мы обеспечиваем идентичное поступление данных на сервер. После получения входящих данных на коллекторы происходит их отправка на модули обнаружения и классификации аномалий, и после обнаружения аномалий отправляется оповещение администратору системы.

В системе будут определяться аномальный и нормальный трафик, а аномальный трафик классифицироваться на следующие группы: DoS, Bruteforce, Port Scan.

В качестве компонентов тестового стенда использовались:

– 3 ПК генерирующие нормальный и аномальный трафик;

– роутер Cisco ASR920 в качестве BRAS и сенсора Netflow;

– сервер СОА с характеристиками 64 ГБ оперативной памяти, – Intel Core i7-11700KF, GPU – GeForce RTX 3070 Ti 8ГБ.

Проведение эксперимента

В качестве критериев оценки результата эксперимента будут использоваться:

– истинно положительные результаты (TP);

– истинно отрицательны результаты (TN);

– ложноположительные результаты (FP);

– ложноотрицательные результаты (FN);

– True Positive Rate – метрика, показывающая процент среди всех истинно положительных результатов верно предсказанных моделью;

– False Positive Rate – метрика, показывающая процент среди всех ложноположительных результатов неверно предсказанных моделью

– Receiver operating characteristic (ROC) – метрика, показывающая соотношение TPR и FPR;

– Area Under Curve (AUC) – площадь под кривой ROC, показывающая, что случайно выбранный экземпляр негативного класса будет иметь меньшую вероятность быть распознанным как позитивный класс, чем случайно выбранный позитивный класс. Значение AUC ограничено от 0 до 1, чем выше значение AUC, тем модель более предсказательна;

– суммарное время обработки данных и получение оповещения от систем.

Целевым критерием оценки нейросетевой модели мы берем метрику AUC, остальные метрики являются вспомогательными для ее вычисления.

В ходе проведения эксперимента с хостов 192.168.1.2/25, 192.168.1.3/25, генерировались атаки типа DoS, Bruteforce, Port scan в сторону хоста 192.168.1.124/25 каждый 10 минут в течение 24 часов, в остальное время генерировался нормальный трафик (Рисунок 4).

Результатом эксперимента стали следующие значения по заявленным критериям, представленные в Таблице 1. Дополнительно ROC кривые показаны на рисунках 5-8.

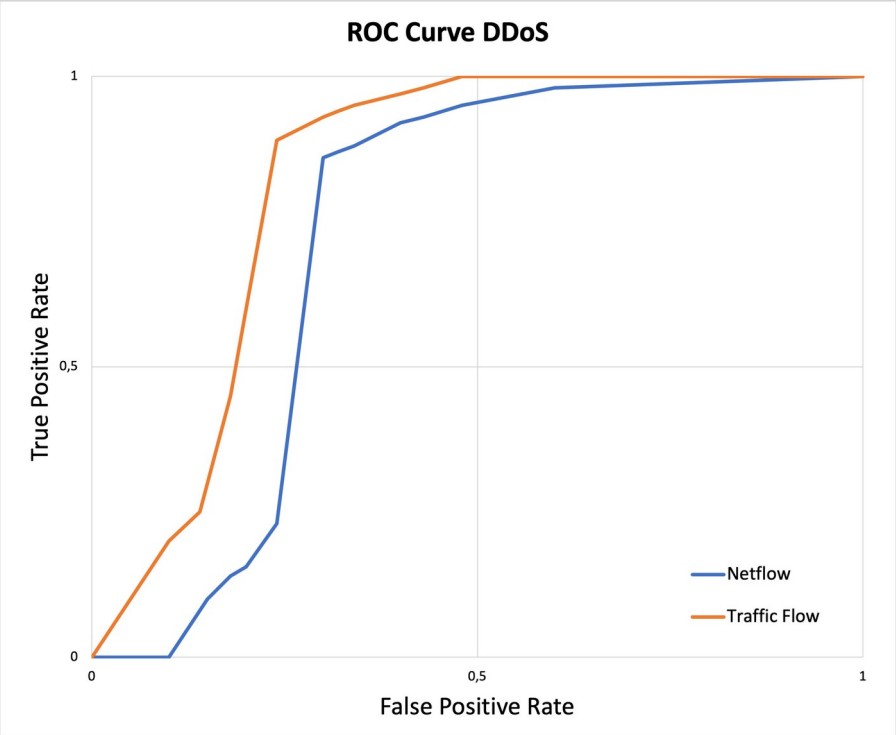

Рисунок 5 - Кривые ROC алгоритма DDoS

DOI: https://doi.org/10.23670/IRJ.2023.132.18.5

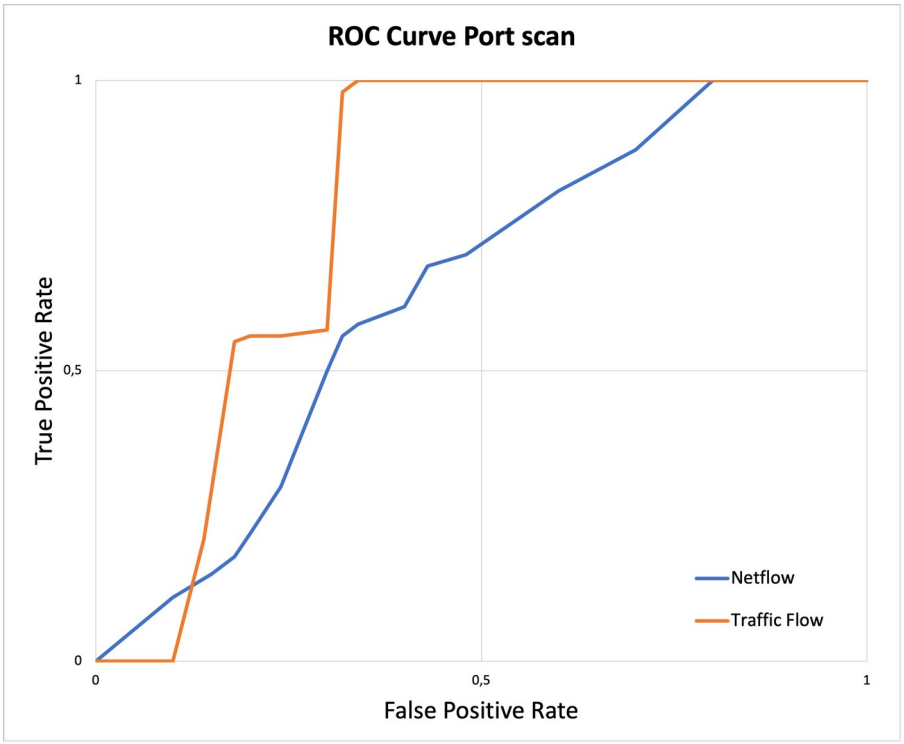

Рисунок 6 - Кривые ROC алгоритма Port scan
DOI: https://doi.org/10.23670/IRJ.2023.132.18.6

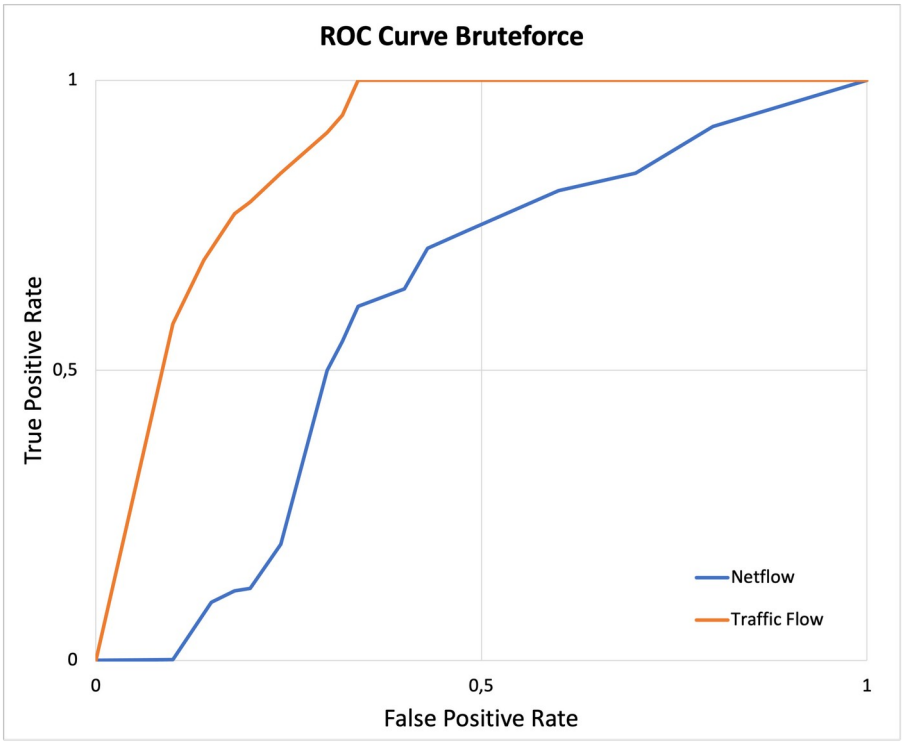

Рисунок 7 - Кривые ROC алгоритма Bruteforce
DOI: https://doi.org/10.23670/IRJ.2023.132.18.7

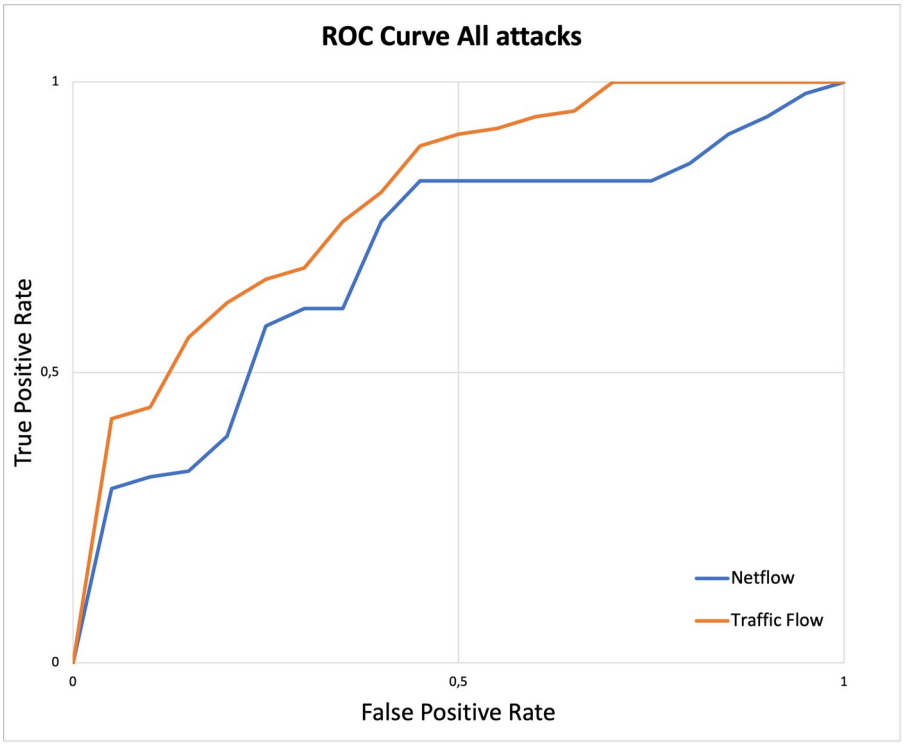

Рисунок 8 - Кривые ROC алгоритма All attacks
DOI: https://doi.org/10.23670/IRJ.2023.132.18.8

Таблица 1 - Критерии AUC

DOI: https://doi.org/10.23670/IRJ.2023.132.18.9

COA		Все атаки, %	DDoS, %	Bruteforce, %	Port scan, %
AUC	С использованием потока трафика	82,80	84,68	91,92	79,78
	С использованием протокола Netflow	72,00	88,66	64,22	67,03

По результатам проведенного исследования имеются следующие выводы:

- при определении атаки DDoS метрика AUC в методе с использованием протокола Netflow выше на 5%, чем при использовании метода с потоком трафика;

- при определении атак Bruteforce и Port Scan значение AUC больше у метода с использованием потока трафика на 30% и 16% соответственно, чем у метода с использованием протокола Netflow. Также при определении всех типов атак метод с использованием потока трафика показал лучшие значения.

Использование СОА с потоком трафика позволяет получить лучшие результаты в атаках Bruteforce и port scan, это вероятно связано с тем, что сеть может иметь больше входных данных, поступающих с сетевого трафика и различных уровней модели OSI. Однако для лучшего определения DDoS оказалась СОА с использованием протокола Netlfow.

Заключение

Использование СОА на основе нейронных сетей показывают хорошие результаты в определении различных аномалий в сетевом трафике. В ходе исследования проведено сравнение системы обнаружения аномалий с использованием рекуррентной LSTM сети и системы обнаружения аномалий с использованием глубокого обучения и протокола Netflow, которое позволило проверить работу систем на реальных данных, а также выявить достоинства и недостатки каждой из систем в поиске аномалий в потоке сетевого трафика:

– для атак, более сложных в определении и требующих глубоко анализа пакетов, подходит СОА с рекуррентной нейронной сетью и использованием потока трафика;

– для атак типа DDoS, имеющие огромную скорость и объем трафика, подходит СОА с глубоким обучением и использованием протокола Netflow.

Конфликт интересов

Не указан.

Рецензия

Все статьи проходят рецензирование. Но рецензент или автор статьи предпочли не публиковать рецензию к этой статье в открытом доступе. Рецензия может быть предоставлена компетентным органам по запросу.

Conflict of Interest

None declared.

Review

All articles are peer-reviewed. But the reviewer or the author of the article chose not to publish a review of this article in the public domain. The review can be provided to the competent authorities upon request.

Список литературы / References

1. Гафаров Ф.М. Искусственные нейронные сети и приложения / Ф.М. Гафаров, А.Ф. Галимянов. — Казань: Изд-во Казан. ун-та, 2018. — 121 с.

2. Тимочкина Т.В. Применение нейронных сетей для обнаружения сетевых атак. / Т.В. Тимочкина, Т.М. Татарникова, Е.Д. Пойманова // Изв. вузов. Прибостроение. — 2021. — 64(5). — с. 357-363.

3. Andropov S. Network anomaly detection using artificial neural networks [Electronic source] / S. Andropov, A. Guirik, M. Budko et al. // 20th Conference of Open Innovations Association (FRUCT). — 2017. — URL: https://ieeexplore.ieee.org/document/8071288. (accessed: 10.09.22) doi: 10.23919/FRUCT.2017.8071288

4. 2021 Global Internet Map Tracks Global Capacity, Traffic, and Cloud Infrastructure [Electronic source] // TeleGeography. — 2021. — URL: https://blog.telegeography.com/2021-global-internet-map-tracks-global-capacity-traffic-and-cloud-infrastructure. (accessed: 01.09.22)

5. CSE-CIC-IDS2018 on AWS [Electronic source] // University of New Brunswick. — 2018. — URL: https://www.unb.ca/cic/datasets/ids-2018.html. (accessed: 01.09.22)

6. Kwon D. An Empirical Study on Network Anomaly Detection Using Convolutional Neural Networks [Electronic source] / D. Kwon, K. Natarajan, S.C. Suh et al. // IEEE 38th International Conference on Distributed Computing Systems (ICDCS). — 2018. — URL: https://ieeexplore.ieee.org/abstract/document/8416441. (accessed: 13.06.23) doi: 10.1109/ICDCS.2018.00178

7. Intrusion Detection Evaluation Dataset (CIC-IDS2017). // University of New Brunswick. — 2017. — URL: https://www.unb.ca/cic/datasets/ids-2017.html (accessed: 01.09.2022).

8. Hochreiter S. Long Short-Term Memory / S. Hochreiter, J. Schmidhuber // Neural Computation. — 1997. — 9(8). — p. 1735-1780.

9. Pearlmutter B.A. Gradient Calculations for Dynamic Recurrent Neural Networks / B.A. Pearlmutter // A survey. IEEE Transactions on Neural Networks. — 1995. — 6(5). — p. 1212-1228.

10. Schmidhuber J. A local Learning Algorithm for Dynamic Feedforward and Recurrent Networks / J. Schmidhuber // Connection Science. — 1995. — 1(4). — p. 403-412.

Список литературы на английском языке / References in English

1. Gafarov F.M. Iskusstvennie neironnie seti i prilozheniya [Artificial Neural Networks and Applications] / F.M. Gafarov, A.F. Galimyanov. — Kazan: Publishing House of Kazan University, 2018. — 121 p. [in Russian]

2. Timochkina T.V. Primenenie nejronny'x setej dlya obnaruzheniya setevy'x atak [Using Neural Networks to Detect Network Attacks]. / T.V. Timochkina, T.M. Tatarnikova, E.D. Pojmanova // Izv. vuzov. Pribostroenie [Proceedings of Universities. Instrumentation]. — 2021. — 64(5). — p. 357-363. [in Russian]

3. Andropov S. Network anomaly detection using artificial neural networks [Electronic source] / S. Andropov, A. Guirik, M. Budko et al. // 20th Conference of Open Innovations Association (FRUCT). — 2017. — URL: https://ieeexplore.ieee.org/document/8071288. (accessed: 10.09.22) doi: 10.23919/FRUCT.2017.8071288

4. 2021 Global Internet Map Tracks Global Capacity, Traffic, and Cloud Infrastructure [Electronic source] // TeleGeography. — 2021. — URL: https://blog.telegeography.com/2021-global-internet-map-tracks-global-capacity-traffic-and-cloud-infrastructure. (accessed: 01.09.22)

5. CSE-CIC-IDS2018 on AWS [Electronic source] // University of New Brunswick. — 2018. — URL: https://www.unb.ca/cic/datasets/ids-2018.html. (accessed: 01.09.22)

6. Kwon D. An Empirical Study on Network Anomaly Detection Using Convolutional Neural Networks [Electronic source] / D. Kwon, K. Natarajan, S.C. Suh et al. // IEEE 38th International Conference on Distributed Computing Systems (ICDCS). — 2018. — URL: https://ieeexplore.ieee.org/abstract/document/8416441. (accessed: 13.06.23) doi: 10.1109/ICDCS.2018.00178

7. Intrusion Detection Evaluation Dataset (CIC-IDS2017). // University of New Brunswick. — 2017. — URL: https://www.unb.ca/cic/datasets/ids-2017.html (accessed: 01.09.2022).

8. Hochreiter S. Long Short-Term Memory / S. Hochreiter, J. Schmidhuber // Neural Computation. — 1997. — 9(8). — p. 1735-1780.

9. Pearlmutter B.A. Gradient Calculations for Dynamic Recurrent Neural Networks / B.A. Pearlmutter // A survey. IEEE Transactions on Neural Networks. — 1995. — 6(5). — p. 1212-1228.

10. Schmidhuber J. A local Learning Algorithm for Dynamic Feedforward and Recurrent Networks / J. Schmidhuber // Connection Science. — 1995. — 1(4). — p. 403-412.