УДК 004.855.5

СРАВНЕНИЕ ЭФФЕКТИВНОСТИ АЛГОРИТМОВ КЛАССИФИКАЦИИ ДЛЯ ОБНАРУЖЕНИЯ ПРИЗНАКОВ DDOS-АТАК IOT-БОТНЕТОВ

С.Н. ПЕТРОВ[1], С.А. ШАВЛОВСКИЙ[2], А.О. РОДУЛЕВИЧ[2]

1 – Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь
2 – Национальный детский технопарк, Республика Беларусь

Поступила в редакцию 31 марта 2024

Аннотация. Проведен сравнительный анализ эффективности алгоритмов классификации для обнаружения признаков DDoS-атак IoT-ботнетов. Показан значительный разброс результатов классификации в зависимости от используемого датасета, что говорит о важности корректного подбора данных для обучения и тестирования моделей машинного обучения.

Ключевые слова: сетевые атаки, DDoS, IoT-ботнет, машинное обучение, классификация сетевого трафика.

Введение

Ботнет – это совокупность устройств, подключенных через Интернет, на каждом из которых работает один бот или серия ботов. Одним из преимуществ ботнета для нарушителя является использование вычислительной мощности сотен или тысяч вычислительных устройств. Распределенные атаки типа «отказ в обслуживании» (DDoS) являются наиболее распространенным применением ботнетов. Устройства интернета вещей (IoT) – это любые устройства с возможностью подключаться к интернету и обмениваться данными с другими устройствами и платформами. Согласно некоторым исследованиям и прогнозам, число устройств IoT может достигнуть от 20 до 30 миллиардов к 2025 году. Многие из этих устройств являются не безопасными. Анализ популярных моделей устройств выявил порядка 250 уязвимостей, включая открытые порты telnet, устаревшие прошивки и передачу конфиденциальных данных в открытом виде. Распространение небезопасных IoT-устройств привело к росту числа ботнет-атак [1]. Таким образом, изучение способов выявления вредоносного ботнет-трафика является актуальной темой. При этом, трафик IoT устройств отличается от трафика прочих устройств, например ноутбуков и смартфонов. Предлагается использовать машинное обучение для обнаружения характерного DDoS IoT сетевого поведения. Для достижения этого необходимо: провести анализ отличительных признаков сетевого трафика, характерных для IoT-ботнетов; провести анализ публичных датасетов для обнаружений аномалий сетевого трафика; обучить классификаторы для нахождения в трафике признаков DDoS-атаки IoT-ботнетов и выбрать оптимальный классификатор.

Методика проведения исследования

Обнаружение сетевых атак типа отказ в обслуживании можно свести к решению задачи классификации. Бинарной классификации (нормальный трафик, DDoS-трафик) или мультиклассовой классификации (нормальный трафик, DDoS-трафик различных типов). Были выбрали для изучения такие методы машинного обучения как метод опорных векторов (Support Vector Machines), метод ближайших соседей (Nearest Neighbors), деревья принятия решений (Decision Trees), многослойный персептрон (Multi-Layer Perceptron).

В качестве основного инструмента для обучения и тестирования моделей машинного обучения была выбрана библиотека машинного обучения TensorFlow. TensorFlow – это открытая библиотека машинного обучения, которая обладает богатым функционалом, поддерживает различные архитектуры моделей и предоставляет инструменты для эффективного обучения и развертывания моделей машинного обучения. Он является одним из наиболее популярных инструментов в данной области

Обучение моделей проводилось с использованием сервиса Google Colab, на основе Jupyter Notebook, который предоставляет бесплатный доступ к вычислительным ресурсам, включая графические процессоры (GPU) и тензорные процессоры (TPU, разработаны специально для задач машинного обучения). Сервис облегчает процесс работы с машинным обучением, предоставляя удобное и гибкое окружение для разработки и исполнения кода.

Использовались датасеты CIC-DDoS 2019 [2], Ton-IoT [3], EdgeIIoT [4], являющиеся популярными в области кибербезопасности и интернета вещей и используются для анализа различных видов угроз и атак. Из датасета Ton-IoT выбран набор данных Windows, собранный с помощью средств мониторинга ОС Windows 10, и набор данных, собранный с IoT-сенсоров Fridge (умные холодильники).

Результаты исследования классификаторов

Результаты исследования алгоритмов классификации представлены в табл. 1–4.

Табл. 1. **Значение Accuracy, полученное на различных датасетах**

Датасет	Классификаторы						
	Тип ядра SVM				KNN	Decision Trees	MLP
	Linear	RBF	Poly	Sigmoid			
CIC-DDoS 2019	85	53	54,5	66,5	98,9	93,5	99,2
Ton_IoT (IoT Fridge)	98,8	97,78	97,82	97,77	97,56	98,42	98
Ton_IoT (Windows 10)	99,94	99,9	99,47	99,93	98,78	99,88	99,9
EdgeIIoT	83,12	89,56	92,45	87,4	85,89	78,23	97,9

Табл. 2. **Значение Recall, полученное на различных датасетах**

Датасет	Классификаторы						
	Тип ядра SVM				KNN	Decision Trees	MLP
	Linear	RBF	Poly	Sigmoid			
CIC-DDoS 2019	99,4	99,1	99,2	99,1	58,8	49,7	57,5
Ton_IoT (IoT Fridge)	75,34	56	56,78	55,64	57,25	71,16	77
Ton_IoT (Windows 10)	87,2	86,7	82,87	87,02	85,22	98,95	99,22
EdgeIIoT	83	86	82	84,21	90	92	95,31

Табл. 3. **Значение F1-Score, полученное на различных датасетах**

Датасет	Классификаторы						
	SVM				KNN	Decision Trees	MLP
	Linear	RBF	Poly	Sigmoid			
CIC-DDoS 2019	87,3	55,5	58,1	66,4	55	51	54,5
Ton_IoT (IoT Fridge)	77,5	53,41	55	52,27	57,22	72,3	79
Ton_IoT (Windows 10)	87,2	86,7	84,6	87,02	84,12	98,93	99,1
EdgeIIoT	89	87,4	84,3	88,1	89	80,2	87,5

Табл. 4. **Значение Precision, полученное на различных датасетах**

Датасет	Классификаторы						
	SVM				KNN	Decision Trees	MLP
	Linear	RBF	Poly	Sigmoid			
CIC-DDoS 2019	66,3	99,5	99,6	66,3	99,5	53,4	74,6
Ton_IoT (IoT Fridge)	94,7	56,04	71,14	55	58,64	90,5	92,8
Ton_IoT (Windows 10)	99,68	99,35	99,2	99,54	95,6	98,92	99,02
EdgeIIoT	96	95,4	95,13	95,42	88	73	89

Была проведена валидация экспериментальных данных путем сравнения их данными из открытых источников. Рассмотрено использование классификатора SVM с различными ядрами для обнаружения признаков DDoS-атак. Результаты сравнения представлены в табл. 5.

Табл. 5. **Валидация экспериментальных данных на основе сравнения с открытыми источниками**

Источник данных	Тип ядра SVM	Метрика			
		Precision	Recall	F1-Score	Accuracy
Экспериментальные	Линейное	0,9033	0,8511	0,8739	0,9943
Статья [5]	Линейное	0,89	0,83	0,86	0,99
Экспериментальные	Сигмоидальное	0,7440	0,6806	0,7068	0,9838
Статья [5]	Сигмоидальное	0,95	0,88	0,91	0,998
Экспериментальные	RBF	0,9849	0,7737	0,7900	0,9915
Статья [5]	RBF	0,71	0,69	0,9	0,99

Как видно из таблицы, данные имеют значительное сходство.

Заключение

Классификаторы показали значительно отличающиеся результаты на различных датасетах. Например, многослойный персептрон (MLP) показал наилучший результат среди всех использованных в работе классификаторов на выборке из датасета Ton_IoT (Windows 10). При это показал очень низкий результат на датасете CIC-DDoS 2019. Это еще раз показывает важность корректного подбора данных для обучения и тестирования моделей машинного обучения. Некоторые задачи классификации могут оказаться более сложными из-за неоднородности данных или мультиклассовости, что может отразиться на производительности модели. Также разные датасеты могут содержать различные характеристики, аномалии и дисбаланс классов, что влияет на способность модели обобщить данные и точность классификации. При этом в открытых источниках нет информации о рекомендованном для исследовательских целей датасете для обучения моделей обнаружению DDoS-атак IoT-ботнетов.

COMPARISON OF THE EFFECTIVENESS OF CLASSIFICATION ALGORITHMS FOR DETECTING SIGNS OF DDOS ATTACKS BY IOT BOTNETS

S.N PETROV, S.A. SHAVLOVSKY, A.O. RODULEVICH

Abstract. A comparative analysis of the effectiveness of classification algorithms for detecting signs of DDoS-attacks by IoT-botnets has been carried out. A significant variation of classification results is shown depending on the dataset used, which indicates the importance of correct data selection for training and testing machine learning models.

Keywords: network attacks, DDoS, IoT-botnet, machine learning, network traffic classification.

Список литературы

1. Тенденции и аналитика DDoS-атак: обзор 2023 [Электронный ресурс]. – Режим доступа: https://ddos-guard.net/ru/blog/tendentsii-ddos-atak-2023 Дата доступа: 05.04.2024.
2. CIC-DDoS2019 Dataset [Электронный ресурс] – Режим доступа: https://data.mendeley.com/datasets/ssnc74xm6r/1 Дата доступа: 05.04.2024
3. The TON_IoT Datasets [Электронный ресурс] – Режим доступа: https://research.unsw.edu.au/projects/toniot-datasets Дата доступа: 05.04.2024
4. EDGE-IIOTSET: a new comprehensive realistic cyber security dataset of IoT and IIoT applications: centralized and federated learning [Электронный ресурс] – Режим доступа: https://ieee-dataport.org/documents/edge-iiotset-new-comprehensive-realistic-cyber-security-dataset-iot-and-iiot-applications#files Дата доступа: 05.04.2024
5. Dasari K.B., Devarakonda N. Detection of TCP-Based DDoS Attacks with SVM Classification with Different Kernel Functions Using Common Uncorrelated Feature Subsets [Электронный ресурс] – Режим доступа: DOI: https://doi.org/10.18280/ijsse.120213 Дата доступа: 05.04.2024