УДК 004.021

«Защита информации автоматизированных систем в ситуационном центре»

Университет Информационных технологий, механики и оптики.
Карманова Наталия Андреевна, 4 курс, факультет Безопасности информационных технологий, Университет ИТМО
Научный руководитель Бондаренко Игорь Борисович, к.т.н., доцент, факультет Безопасности информационных технологий, Университет ИТМО

В настоящее время наблюдается рост количества распределенных атак на глобальные компьютерные сети, значительная часть из которых приходится на атаки типа "Отказ в обслуживании" (DDoS). По данным российской компании Qrator Labs, уже сейчас проводятся реальные DDoS-атаки до 1 Тбит/с, значительно увеличивается количество сложных атак на протоколы прикладного уровня. DDoS-атаки влияют на один из основополагающих принципов информационной безопасности - доступность информационных ресурсов. В каждом конкретном случае DDoS может либо непосредственно причинить вред, либо создать угрозу и потенциальный риск нанесения убытков. Атакам такого рода могут быть подвержены любые сетевые ресурсы, вне зависимости от их владельцев направления деятельности. Таким образом, одним из наиболее актуальных на сегодняшний день направлений в области информационной безопасности является противодействие DDoS-атакам, в том числе разработка средств защиты, способных справиться со сложными высокоскоростными атаками. Целью работы является рассмотрение существующих алгоритмов классификации данных на основе машинного обучения, проведение оценки их эффективности и разработка метода детектирования сетевых распределенных атак типа "Отказ в обслуживании" с помощью алгоритмов машинного обучения. В работе рассмотрены существующие алгоритмы классификации данных на основе машинного обучения, проведена оценка их эффективности и разработан метод детектирования сетевых распределенных атак типа "Отказ в обслуживании" с помощью алгоритмов машинного обучения.

Использование механизмов машинного обучения помогает решить проблему обнаружения DDoS-атак как с помощью алгоритмов обучения с учителем, так и без учителя. Первый метод обладает более высокой точностью обнаружения, когда имеются точные и корректные атрибуты атак, но не может обнаруживать неизвестные атаки. Второй, напротив, может помочь в обнаружении неизвестных атак, однако он значительно теряет эффективность при наличии шумов и большого количества ложных данных. Также методы обучения без учителя, как правило, не могут работать в реальном времени, так как требуют более серьезной и продолжительной обработки данных. В данной работе были использованы классификаторы на основе алгоритмов дерева решений, k-ближайших соседей и случайного леса из библиотеки машинного обучения *scikit-learn*. Полный перечень программных инструментов, использованных в работе, приведен в табл. 1

Таблица 1

Программные инструменты, используемые для построения модели классификации

Модуль	Применение
pandas	Первичная обработка данных, работа с размеченными данными, формирование датафреймов. Работает в связке с NumPy, и помимо математических вычислений обеспечивает их агрегацию и визуализацию
sklearn	Отбор признаков, реализация алгоритмов классификации
numpy	Выполнение основных операций над n-массивами и матрицами. Использование механизмов векторизации NumPy повышает производительность и ускоряет выполнение операций над данными
matplotlib	Визуализация данных, формирование графиков и диаграмм

Разработанный метод состоит из следующих этапов:

1. формирование выборки трафика защищаемого приложения;
2. формирование выборки легитимного трафика;
3. разделение выборок на обучающую и тестовую;
4. выделение наиболее информативных признаков (атрибутов сетевого трафика);
5. фаза обучения;
6. оптимизация алгоритма;
7. оценка результатов классификации.

Таблица 2

Сравнительная таблица классификаторов сетевого трафика

Алгоритм	До оптимизации	После оптимизации
Дерево решений	86%	91%
k-ближайших соседей	91%	94%
Случайный лес	93%	95%

Во всех трех случаях примененные методы позволили определить наиболее оптимальные параметры моделей и повысить точность классификации сетевого трафика дополнительно на 2-5%.

В ходе работы были рассмотрены существующие алгоритмы классификации данных на основе машинного обучения, проведена оценка возможности их применения для детектирования сетевых распределенных атак типа «отказ в обслуживании». Наилучшие результаты на тестовых данных показал алгоритм случайного леса (95%). Результаты исследования применяются в работе по созданию автоматизированной программно аппаратной платформы защиты от 40 Гбит/с DDoS атак на основе сетевых плат FPGA PCI-E.

Литература

1. Muller, A. C. Introduction to Machine Learning with Python: Guide for Data Scientists / Andreas C. Muller, Sarah Guido. -
San Francisco: O'Reilly Media, 2016. - P. 268-328

2. Bhattacharyya, D. K. DDoS Attacks: Evolution, Detection, Prevention, Reaction, and Tolerance / Dhruba Kumar Bhattacharyya, Jugal Kumar Kalita. - London: Chapman and Hall/CRC, 2016. - P. 23-70
3. Garreta, R., Moncecchi G. Learning scikit-learn. Machine Learning in Python / R. Garreta. - Birmingham: Packt, 2013. P. 25-60
4. DDoS-атаки и электронная коммерция: современные подходы к защите [Электронный ресурс] / 1-С Битрикс / /Сайт Хабрахабр. - Режим доступа: https://habrahabr.ru/company/bitrix/blog/267947, свободный. - Загл. с экрана.
5. Машинное обучение: метод ближайших соседей [Электронный ресурс] / Сайт MachineLearning.ru. - Режим доступа:
http://www.machinelearning.ru/wiki/index.php?title=%D0%9C%D0%B5%D1%82%D0%BE%D0%B4_%D0%B1%D0%BB%D0%B8%D0%B6%D0%B0%D0%B9%D1%88%D0%B5%D0%B3%D0%BE_%D1%81%D0%BE%D1%81%D0%B5%D0%B4%D0%B0,
свободный. - Загл. с экрана.
6. Обучение на размеченных данных. Случайные леса [Электронный ресурс] / Сайт coursera.orgю - Режим доступа:
https://ru.coursera.org/learn/supervised-learning/lecture/bejGu/sluchainyie-liesa, свободный. - Загл. с экрана
7. Отчет за 2017 год компании Qrator Labs [Электронныйресурс] / Компания Qrator Labs. - Режим доступа: https://
qrator.net/presentations/QratorAnnualRepRus.pdf, свободный. -Загл. с экрана.
8. Технологии анализа данных. Алгоритм CART [Электронный ресурс] / Компания BaseGroup Labs. – Режим доступа: http://docplayer.ru/32842822-Algoritm-cart-m-154.html, свободный. - Загл. с экрана.
9. Отчет по безопасности за 2016 год [Электронный ресурс] /Cisco Systems. - Режим доступа: https://www.cisco.com/c/
dam/m/ru_ru/cisco_2016_asr_011116_ru.pdf, свободный. -Загл. с экрана.
10. Воробьева, А. А. История развития программно-аппаратных средств защиты информации / А.А. Воробьева, И.С. Пантюхин.- СПб: Университет ИТМО, 2017. - 62 с.
11. Ерохин, С. Д. Влияние фонового трафика на эффективность классификации приложений методами машинного
обучения / С.Д. Ерохин, А.В. Ванюшина / / T-Comm: Телекоммуникации и транспорт. - 2017. - Т. 11, № 12. - С. 31-36.
12. Кафтанников, И. Л. Особенности применения деревьев решений в задачах классификации / И.Л. Кафтанников,
А.В. Парасич / / Вестник ЮУрГУ. Серия «Компьютерные технологии, управление, радиоэлектроника». - 2015. - Т. 15,
№ 3. - С. 26-32.
13. Зубриенко, Г. А. Методы оптимизации выборки данных для определения аномального трафика / Г.А. Зубриенко,
О.Р. Лапонина / / International Journal of Open Information Technologies. - 2016. - Vol. 6, No. 3. - P. 1-8.
14. Котов, В. Д. Современное состояние проблемы обнаружения сетевых вторжений / В.Д. Котов, В.И. Васильев / /
Вестник УГАТУ. - 2012. - Т. 16, № 3. - С. 198-204.
15. Чистяков, С. П. Случайные леса: обзор / С.П. Чистяков / /Труды КарНЦ РАН. - 2013. - № 1. - С. 117-136.
16. Мухамедиев, Р. И. Таксономия методов машинного обучения и оценка качества классификации и обучаемости [Электронный ресурс] / Р.И. Мухамедиев, Е.Л. Мухамедиева, Я.И. Кучин / / Электронный журнал Cloud of science. - 2015. - Т. 2, №3. – Режим доступа: https://cyberleninka.ru/article/v/taksonomiya-metodovmashinnogoobucheniya-i-otsenka-kachestva-klassifikatsii-iobuchaemosti, свободный. - Загл. с экрана.