УДК 004.023
DOI: 10.26102/2310-6018/2022.36.1.019

Учебный стенд для анализа методов обнаружения аномалий на основе теории машинного обучения

М.М. Греков

Тульский государственный университет,
Тула, Российская Федерация

Резюме. На сегодняшний день актуальной задачей является своевременное выявление новых вредоносных воздействий на компьютерные сети. В связи с этим необходимо развитие методов обнаружения аномалий, которые позволяют выявлять неизвестные атаки. В работе представлена модель учебного стенда для анализа методов обнаружения аномалий на основе теории машинного обучения. Разработана модель генерации наборов данных с характеристиками реального сетевого трафика с помощью генеративно-состязательной нейронной сети. Генерируемый набор данных может применяться при обучении и тестировании моделей обнаружения, при этом выборка повторяет особенности реальной сети, что повышает эффективность детектирования аномалий. В учебном стенде также могут применяться общедоступные наборы данных: NSL-KDD, CICIDS2017. В качестве методов обучения используются машина опорных векторов, k-ближайших соседей, наивный Байес, логистическая регрессия, деревья решений, случайный лес, k-средних, а также реализована многослойная нейронная сеть на основе библиотеки PyTorch. Учебный стенд позволяет упростить процесс анализа методов машинного обучения, применяемых для получения моделей обнаружения аномалий. Разработанный программный продукт позволяет не только осуществлять обучение и тестирование на основе общедоступных наборов данных, но и реализует возможность сбора сетевого трафика и дополнения его сгенерированными данными с характеристиками реального трафика.

Ключевые слова: системы обнаружения аномалий, наборы данных, генеративно-состязательные нейронные сети, машинное обучение, безопасность компьютерных сетей.

Для цитирования: Греков М.М. Учебный стенд для анализа методов обнаружения аномалий на основе теории машинного обучения. *Моделирование, оптимизация и информационные технологии.* 2022;10(1). Доступно по: https://moitvivt.ru/ru/journal/pdf?id=1122 DOI: 10.26102/2310-6018/2022.36.1.019

A training device for the analysis of anomaly detection methods based on machine learning theory

M.M. Grekov

Tula State University, Tula, Russian Federation

Abstract: Nowadays, the timely detection of new malicious attacks on computer networks appears to be a relevant issue. In this regard, it is necessary to develop anomaly detection methods that enable the identification of unknown attacks. The paper presents a model of a training device for analyzing anomaly detection methods in reliance on machine learning theory. A model has been developed for generating datasets with characteristics of real network traffic by means of a generative adversarial neural network. The generated dataset can be employed to train and test detection models while the sample emulates the features of a real network, which increases the efficiency of anomaly detection. The training device can also use publicly available datasets: NSL-KDD, CICIDS2017. Support vector machine, k-nearest neighbors, naive Bayes, logistic regression, decision trees, random forest, k-means are utilized as training methods, and a multilayer neural network, based on the PyTorch library, is

implemented. The training device simplifies the process of analyzing machine learning methods, applied to obtain anomaly detection models. The developed software product facilitates not only training and testing with the aid of publicly available datasets, but also provides the ability to collect network traffic and supplements it with generated data with the characteristics of real traffic.

Keywords: anomaly detection systems, datasets, generative adversarial neural networks, machine learning, computer network security.

For citation: Grekov M.M. A training device for the analysis of anomaly detection methods based on machine learning theory. *Modeling, Optimization and Information Technology*. 2022;10(1). Available from: https://moitvivt.ru/ru/journal/pdf?id=1122 DOI: 10.26102/2310-6018/2022.36.1.019 (In Russ).

Введение

В области информационной безопасности немаловажной задачей является обеспечение защищенности компьютерных систем и сетей. Надежная защита сетевых ресурсов должна сопровождаться своевременным обнаружением состояний, которые могут приводить к нарушению конфиденциальности, доступности или целостности информации. В настоящее время актуальной задачей является развитие методов обнаружения аномалий, так как обнаружение злоупотреблений не позволяет выявлять неизвестные вредоносные воздействия, что в условиях постоянного развития и появления новых атак может потенциально привести к огромным убыткам для организаций.

Обнаружение аномалий позволяет выявлять неизвестные атаки, определяя их как отклонение от профиля нормальной активности. Аномалии в потоке сетевых данных могут быть вызваны случайными или преднамеренными действиями со стороны легитимных пользователей, неверной работой ПО, действиями злоумышленников и т. д. Применение систем обнаружения вторжений на основе аномалий является перспективным направлением, но существует необходимость анализа и развития методов обнаружения аномалий, так как в результатах детектирования достаточно высок процент ошибок второго рода (ложных срабатываний).

В настоящий момент актуальным подходом является применение аппарата машинного обучения [1]. При этом, профиль нормальной активности строится на этапе обучения на основе сетевого трафика, который может быть получен путем захвата реального трафика сети с помощью датчиков (сенсоров). Также для обучения могут быть использованы общедоступные наборы данных и искусственно сгенерированный трафик.

Таким образом, актуальной задачей является развитие методов обнаружения аномалий на основе теории машинного обучения, однако процесс сбора данных, предварительной обработки, обучения и тестирования является достаточно трудоемким. В связи с этим в данной работе предлагается модель учебного стенда, использование которого позволит упростить процесс сбора данных и анализа методов обнаружения аномалий на основе аппарата машинного обучения.

Материалы и методы

Учебный стенд состоит из четырех основных подсистем, структура представлена на Рисунке 1.

Подсистема управления является центральным компонентом системы, который обеспечивает доступ к остальным подсистемам через графический интерфейс программы. Модуль настройки позволяет вносить необходимые изменения в основные параметры разработанной системы и конфигурационные файлы методов обнаружения.

Подсистема обнаружения аномалий состоит из ядер обнаружения аномалий и модуля обучения и тестирования, который реализует возможность подготовки и проверки моделей детектирования с помощью множества алгоритмов. Ядра обнаружения аномалий представляют собой модели, полученные с помощью методов машинного обучения.

Могут быть использованы следующие методы обучения из библиотеки Scikit-learn [2,3]: машина опорных векторов, k-ближайших соседей, наивный Байес, логистическая регрессия, деревья решений, случайный лес, k-средних. Также на основе библиотеки PyTorch [4] реализована многослойная нейронная сеть с двумя скрытыми слоями.

Подсистема анализа методов состоит из модуля представления результатов, реализующего графическое отображение результатов внутри основной программы, а также модуля выбора, с помощью которого осуществляется настройка обучения и тестирования.

Подсистема агрегирования и обработки данных предназначена для объединения множества данных о сетевом трафике, получаемых из различных источников: наборы данных, захваченные сетевые пакеты, модуль генерации данных. Модуль предварительной обработки необходим для подготовки данных к дальнейшим операциям путем преобразования значений.

Подсистема сбора данных необходима для захвата сетевых пакетов, а также для записи полученных данных.

Рисунок 1 – Структура разработанного учебного стенда
Figure 1 – The structure of the developed training device

Подсистема генерации наборов данных основана на генеративно-состязательной нейронный сети, которая позволяет создавать образцы сетевого трафика для дополнения несбалансированных наборов данных или исследования устойчивости обученной модели к генерируемым образцам.

Данная подсистема анализирует нормальный трафик реальной сети и с помощью генеративно-состязательной нейронной сети (Generative Adversarial Network, GAN) [5, 6] генерирует атакующий трафик. На выходе подсистемы создается совмещенный набор данных со свойствами реального трафика. В дальнейшем сгенерированный набор данных может быть использован для обучения и тестирования моделей обнаружения аномалий [7].

На Рисунке 2 представлена структура модуля генерации наборов данных.

Основным модулем является конфигуратор, который инициализирует входные параметры модели, передает входной набор данных и основные параметры в модуль предварительной обработки. Характеристики реального трафика передаются в модуль объединения.

Входной PCAP-файл предварительно создается с помощью модуля сбора трафика на выбранном узле сети. Предполагается, что сбор данных осуществляется при нормальном функционировании узла. Иначе говоря, в момент захвата трафика не осуществляется сетевых атак.

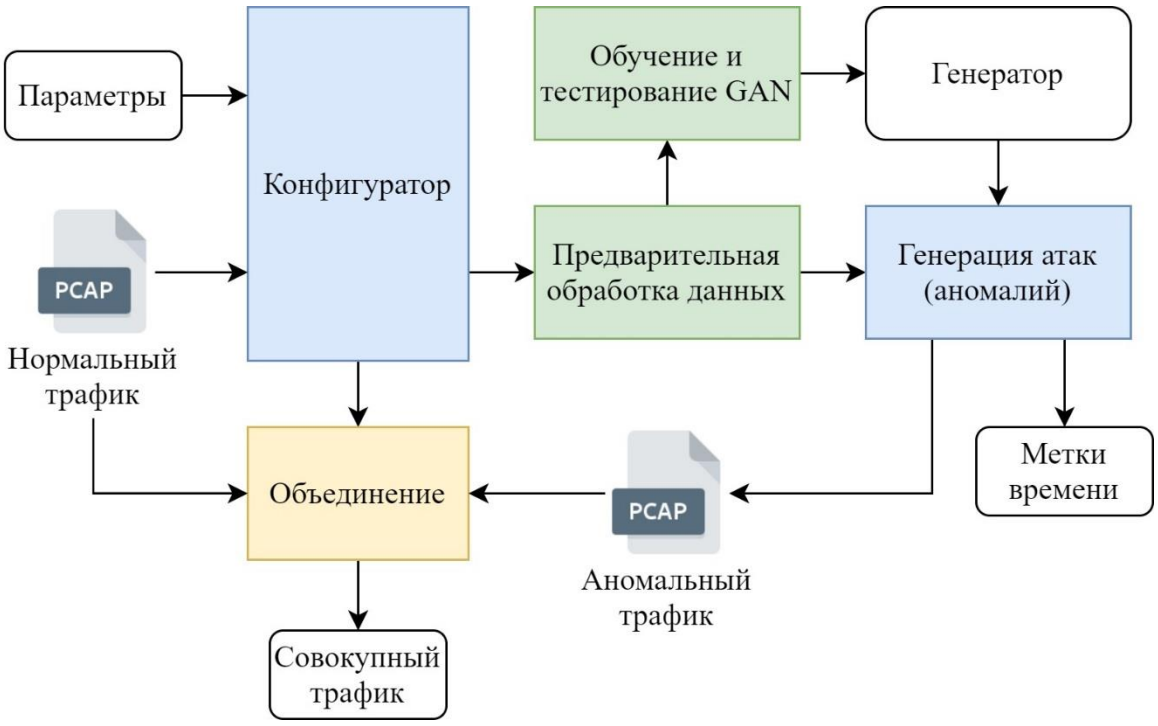

Рисунок 2 – Структура подсистемы генерации наборов данных на основе генеративно-состязательной нейронной сети

Figure 2 – The structure of the subsystem for generating datasets based on a generative adversarial network

Преобразование наборов данных выполняется модулем предварительной обработки в соответствии с характеристиками реального трафика. Предобработанный набор данных, содержащий сведения об атаках, необходим для первоначального обучения модели GAN.

Если модель GAN ранее не была создана, необходимо подготовить ее с помощью модуля обучения и тестирования GAN, который позволяет по заданным параметрам создать генератор выбранного типа атак. Если было выбрано несколько типов атак, выполняется последовательный вызов соответствующих генераторов. Выборка сгенерированных образцов атак поступает на вход модуля генерации атак (аномалий).

В модуле генерации аномалий с помощью обученного генератора выполняется запуск процесса создания образцов трафика, осуществляется проверка полученных наборов аномального трафика, производится объединение различных типов атак. Выход модуля генерации атак – метки времени и аномальный трафик. С помощью меток времени осуществляется маркировка (атака или нет) строк в наборе данных, который в дальнейшем применяется в системах обнаружения аномалий на основе машинного обучения. С помощью модуля объединения осуществляется слияние нормального и аномального наборов данных.

Модуль хранения результатов и параметров состоит из совокупности данных о сетевом трафике, моделей обнаружения аномалий, конфигурационных файлов и результатов работы анализируемых методов. В качестве основных выборок для обучения и тестирования используется два общедоступных набора данных: NSL-KDD [8] и CICIDS2017 [9].

Набор данных NSL-KDD был создан для решения проблем, связанных с избыточностью и дублированием данных в наборе KDD-99, который был создан Лабораторией Линкольна MIT и является одним из самых используемых наборов для исследования методов обнаружения сетевых вторжений.

NSL-KDD содержит записи сетевого соединения, где поток данных от источника к получателю зафиксирован и размечен в соответствии с TCP, UDP, ICMP протоколами. Для создания проверочной выборки из обучающего множества выделено подмножество из 21416 записей, что составляет примерно 17 % от изначального размера. Кроме записей нормальной активности, каждая из выборок содержит вредоносный трафик, разделенный на 4 категории атак: отказ в обслуживании (DoS, Denial of Service), сканирование для поиска уязвимостей (Probe), получение зарегистрированным пользователем повышенных привилегий (U2R, User to Root), получение удаленного доступа незарегистрированным пользователем к локальному компьютеру (R2L, Remote to Local).

Набор данных CICIDS2017 содержит нормальный трафик и самые распространенные на текущий момент атаки, похожие на реальные данные. CICIDS2017 подготовлен Канадским институтом кибербезопасности и Университетом Нью-Брансуика с использованием моделирования поведения 25 пользователей на основе протоколов HTTP, HTTPS, FTP, SSH и электронной почты. Реализованные атаки включают Brute Force FTP, Brute Force SSH, DoS, Heartbleed, Web Attack, Infiltration, Botnet и DDoS. Сбор данных осуществлялся в течение 5 дней.

На Рисунке 3 представлен графический интерфейс программной реализации на основе модели учебного стенда и продемонстрированы результаты обучения и тестирования с оценкой по основным метрикам.

Оценка эффективности методов обнаружения аномалий осуществляется с помощью ошибок первого и второго рода [10].

Для определения аномальности каждого объекта выборки из набора данных существует статистический критерий $a(x)$: $R\hat{}n \rightarrow \{H\_0, H\_1\}$ и стоит задача проверки двух гипотез: H_0 и H_1, где H_0 – нулевая гипотеза, соответствующая нормальной сетевой активности, а H_1 – альтернативная гипотеза, соответствующая аномальному поведению.

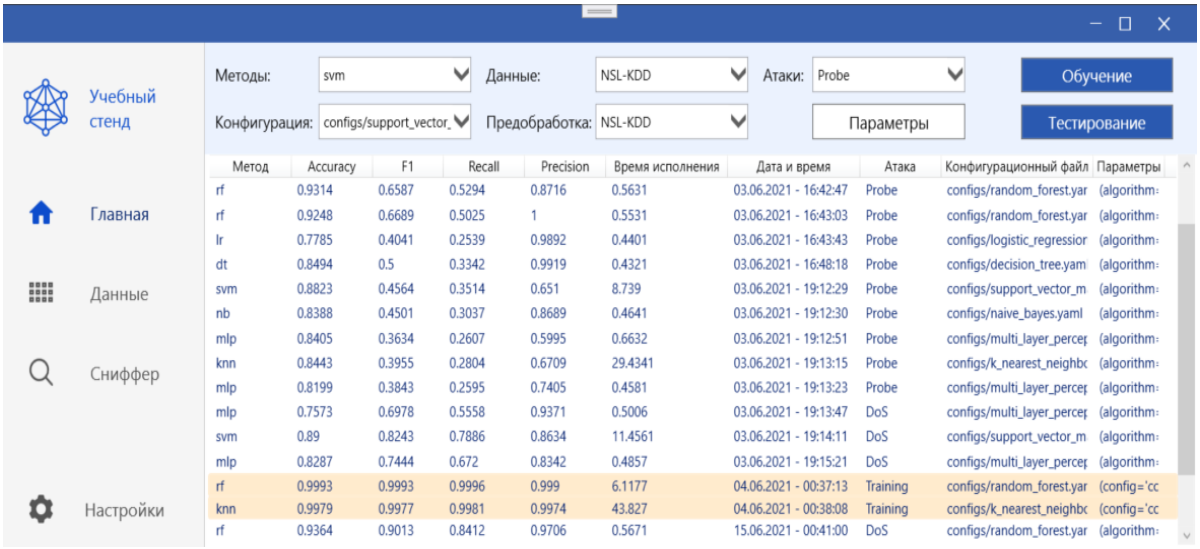

Рисунок 3 – Графический интерфейс программной реализации
Figure 3 – Graphical interface of software implementation

Четыре возможных исхода работы алгоритма:

1) объект x_i соответствует гипотезе H_0, верно определена критерием,
$$a(x) = H\_0$$
2) объект x_i соответствует гипотезе H_0, неверно отвергнута критерием,
$$a(x) = H\_1$$
3) объект x_i соответствует гипотезе H_1, верно принята критерием,
$$a(x) = H\_1$$
4) объект x_i соответствует гипотезе H_1, неверно отвергнута критерием,
$$a(x) = H\_0.$$

Второй случай называется ошибкой первого рода (ложная тревога), а четвертый – ошибкой второго рода (пропуск аномалии).

Оценка эффективности методов машинного обучения осуществлялась с помощью следующих метрик.

Достоверность (accuracy) является простой метрикой, отражающей долю правильных ответов, однако почти не используется без других метрик, так как бесполезна в случае несбалансированных классов.

$$accuracy = (TP + TN)/(TP + TN + FP + FN), \quad (1)$$

где TP – число верно определенных нормальных объектов; TN – число выявленных аномалий; FP – число пропущенных аномалий; FN – число ложных срабатываний.

Точность (precision) отражает долю истинно положительных решений среди всех положительно помеченных объектов:

$$precision = TP/(TP + FP) \quad (2)$$

Полнота (recall) отражает долю истинно положительных объектов среди всех объектов положительного класса:

$$recall = TP/(TP + FN). \quad (3)$$

Точность и полнота объединяются с помощью F-меры – в традиционном виде это гармоническое среднее значение двух метрик:

$$F\_1 = 2 \cdot (recall \cdot precision)/(recall + precision) \quad (4)$$

или в общем виде F_β, используя параметр β, определяющий вес полноты в метрике:

$$F\_\beta = ((1+\beta)\text{^}2 \cdot recall \cdot precision)/(\beta\text{^}2 \cdot recall + precision) \quad (5)$$

F-мера позволяет глубже оценить правильности классификации, по сравнению с полнотой и точностью.

ROC-кривая (Receiver Operating Characteristic, ROC Curve) является линией из точки (0,0) в точку (1,1) в координатах истинно положительных решений и ложно положительных решений.

Доля истинно положительных решений (TPR – true positive rate):

$$TPR = TP/(TP + FN) \quad (6)$$

Доля ложно положительных решений (FPR – false positive rate):

$$FPR = FP/(FP + TN) \quad (7)$$

Для оценки качества классификации вводится показатель AUC (Area Under Curve), соответствующий площади под ROC-кривой. Чем выше значение AUC, тем лучше выполняется классификация.

Заключение

Разработанный учебный стенд реализует широкий спектр функциональных возможностей и интуитивно понятный графический интерфейс, что позволяет упростить процесс анализа методов обнаружения аномалий в сетевом трафике с помощью методов машинного обучения. Модель генерации наборов данных позволяет создавать выборки с характеристиками реального сетевого трафика с помощью генеративно-состязательной нейронной сети. Генерируемые наборы данных повышают эффективность обнаружения аномалий за счет учета особенностей реальной сети. Программный продукт позволяет проводить обучение и тестирование моделей, захват, генерацию и предварительную обработку данных, анализ методов обнаружения аномалий в компьютерной сети на основе теории машинного обучения.

СПИСОК ИСТОЧНИКОВ

1. Buczak Anna L. and Erhan Guven. A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications surveys & tutorials*. 2015;18.2:1153-1176.
2. Рашка С. and Мирджалили В. *Python и машинное обучение. Машинное и глубокое обучение с использованием Python, scikit-learn и TensorFlow.* СПб.: ООО «Диалектика»; 2019.
3. Scikit-learn. Режим доступа: https://scikit-learn.org/stable/index.html (дата обращения: 11.12.2021).
4. PyTorch. Режим доступа: https://pytorch.org (дата обращения: 11.12.2021).
5. Goodfellow Ian, et al. Generative adversarial nets. *Advances in neural information processing systems.* 2014;27.
6. Arjovsky Martin, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. *International conference on machine learning*. PMLR; 2017.
7. Brauckhoff Daniela, Arno Wagner and Martin May. FLAME: A Flow-Level Anomaly Modeling Engine. *CSET*. 2008.
8. NSL-KDD. Режим доступа: https://www.unb.ca/cic/datasets/nsl.html (дата обращения: 11.12.2021).

9. CICIDS2017. Режим доступа: https://www.unb.ca/cic/datasets/ids-2017.html (дата обращения: 11.12.2021).
10. Шелухин О.И. *Сетевые аномалии. Обнаружение, локализация, прогнозирование*. М.: Горячая линия-Телеком; 2019.

REFERENCES

1. Buczak Anna L. and Erhan Guven. A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications surveys & tutorials*. 2015;18.2:1153-1176.
2. Rashka S., and Mirjalili V. *Python and machine learning. Machine and deep learning with Python, scikit-learn and TensorFlow.* St. Petersburg: Dialectics LLC; 2019. (In Russ.)
3. Scikit-learn. URL: https://scikit-learn.org/stable/index.html (accessed on 11.12.2021).
4. PyTorch. URL: https://pytorch.org (accessed on 11.12.2021).
5. Goodfellow Ian, et al. Generative adversarial nets. *Advances in neural information processing systems*. 2014;27.
6. Arjovsky Martin, Soumith Chintala and Léon Bottou. Wasserstein generative adversarial networks. *International conference on machine learning*. PMLR; 2017.
7. Brauckhoff Daniela, Arno Wagner and Martin May. FLAME: A Flow-Level Anomaly Modeling Engine. *CSET*. 2008.
8. NSL-KDD. URL: https://www.unb.ca/cic/datasets/nsl.html (accessed on 11.12.2021).
9. CICIDS2017. URL: https://www.unb.ca/cic/datasets/ids-2017.html (accessed on 11.12.2021).
10. Shelukhin O.I. *Network anomalies. Detection, localization, forecasting*. Moscow: Hotline-Telecom; 2019. (In Russ.)

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Греков Михаил Михайлович, аспирант, кафедра информационной безопасности, Тульский государственный университет Институт прикладной математики и компьютерных наук, Тула, Российская Федерация.
e-mail: grekov.web@yandex.ru
ORCID: 0000-0002-7301-1988

Mikhail M. Grekov, postgraduate, Information Security Department, Tula State University», Tula, Russian Federation.

Статья поступила в редакцию 23.12.2021; одобрена после рецензирования 15.02.2022; принята к публикации 04.03.2022.

The article was submitted 23.12.2021; approved after reviewing 15.02.2022; accepted for publication 04.03.2022.