УДК 004.056.5
DOI: 10.25559/SITITO.14.201803.626-632

ВЫЯВЛЕНИЕ АТАК В КОРПОРАТИВНЫХ СЕТЯХ С ПОМОЩЬЮ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Н.Ф. Бахарева[1], В.Н. Тарасов[1], А.Е. Шухман[2], П.Н. Полежаев[2], Ю.А. Ушаков[2], А.А. Матвеев[2]
[1] Поволжский государственный университет телекоммуникаций и информатики, г. Самара, Россия
[2] Оренбургский государственный университет, г. Оренбург, Россия

ATTACK DETECTION IN ENTERPRISE NETWORKS BY MACHINE LEARNING METHODS

Nadezhda F. Bakhareva[1], Veniamin N. Tarasov[1], Aleksandr E. Shukhman[2], Petr N. Polezhaev[2], Yuri A. Ushakov[2], Artem A. Matveev[2]
[1] Povolzhskiy State University of Telecommunications & Informatics, Samara, Russia
[2] Orenburg State University, Orenburg, Russia

Ключевые слова

Защита корпоративных сетей; анализ трафика; классификация; машинное обучение; выявление атак.

Аннотация

Обнаружение сетевых атак является в данный момент одной из наиболее острых проблем безопасного применения корпоративных сетей. Сетевые системы обнаружения вторжений на основе сигнатурных правил не способны обнаруживать новые типы атак. Таким образом, актуальной является задача быстрой классификации сетевого трафика для обнаружения сетевых атак. В статье разрабатываются алгоритмы выявления атак в корпоративных сетях на основе анализа данных, которые могут быть в них собраны. Использован набор данных UNSW-NB15 для сравнения методов машинного обучения для классификации по принципу атака-обычный траффик, а также для выявления девяти наиболее популярных классов типовых атак, таких как Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode и Worms. В качестве основной метрики оценки точности классификации используется balanced_accuracy_score – сбалансированная точность. Основное преимущество данной метрики в адекватной оценке точности алгоритмов классификации с учетом сильного дисбаланса в количестве размеченных записей по каждому классу набора данных. В результате эксперимента было выявлено, что лучшим алгоритмом для идентификации наличия атаки является RandomForest, для уточнения ее типа – AdaBoost.

Keywords

Protection of enterprise networks; traffic analysis; classification; machine learning; detection of attacks.

Abstract

Detection of network attacks is currently one of the most important problems of secure use of enterprise networks. Network signature-based intrusion detection systems cannot detect new types of attacks. Thus, the urgent task is to quickly classify network traffic to detect network attacks. The article describes algorithms for detecting attacks in enterprise networks based on data analysis that can be collected in them. The UNSW-NB15 data set was used to compare machine learning methods for classifying attack or-normal traffic, as well as to identify nine more popular classes of typical attacks, such as Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode and Worms. Balanced accuracy is used as the main metric for assessing the accuracy of the classification. The main advantage of this metric is an adequate assessment of the accuracy of classification algorithms given the strong imbalance in the number of marked records for each class of data set. As a result of the experiment, it was found that the best algorithm for identifying the presence of an attack is RandomForest, to clarify its type – AdaBoost.

Об авторах:

Бахарева Надежда Федоровна, доктор технических наук, профессор, кафедра информатики и вычислительной техники, Поволжский государственный университет телекоммуникаций и информатики (443010, Россия, г. Самара, ул. Л.Толстого, д. 23), ORCID: http://orcid.org/0000-0002-9850-7752, Bahareva-nf@psuti.ru

Тарасов Вениамин Николаевич, доктор технических наук, профессор, кафедра управления и информатики в технических системах, Поволжский государственный университет телекоммуникаций и информатики (443010, Россия, г. Самара, ул. Л.Толстого, д. 23), ORCID: http://orcid.org/0000-0002-9318-0797, veniamin_tarasov@mail.ru

Шухман Александр Евгеньевич, кандидат педагогических наук, доцент, кафедра геометрии и компьютерных наук, Оренбургский государственный университет (460000, Россия, г. Оренбург, пр. Победы, д. 13), ORCID: http://orcid.org/0000-0002-4303-2550, shukhman@gmail.com

Полежаев Петр Николаевич, старший преподаватель, кафедра компьютерной безопасности и математического обеспечения информационных систем, Оренбургский государственный университет (460000, Россия, г. Оренбург, пр. Победы, д. 13), ORCID: http://orcid.org/0000-0001-7747-646X, newblackpit@mail.ru

Введение

Обнаружение сетевых атак является в данный момент одной из наиболее острых проблем безопасного применения корпоративных сетей. Масштабные эпидемии сетевых червей, DDoS атаки с бот-сетей, автоматизированные средства поиска уязвимостей в сетях – все это делает обеспечение безопасности локальных сетей весьма трудоемким делом. Сейчас трудно найти сеть, в которой отсутствуют такие активные средства предупреждения атак как антивирус, брандмауэр, системы предупреждения вторжений. К сожалению, одних активных средств отражения атак недостаточно. Поэтому, в дополнение к ним применяют пассивные средства борьбы с атаками – сетевые системы обнаружения вторжений.

Сетевые системы обнаружения вторжений (ССОВ) просматривают весь сетевой трафик (или трафик определенного участка сети) и при обнаружении каких-либо отклонений в нем сигнализируют об этом. Формальные ССОВ используют сигнатурные правила – пакеты, попадающие на сенсоры, сравниваются с БД сигнатур и, в случае обнаружения совпадения, объявляется тревога. К сожалению, даже формальных ССОВ становится недостаточно для надежной защиты сети. По данным CERT, количество известных новых методов вторжения только за 2010 год превысило 25000. Это значит, что в среднем, каждый день появляется порядка 70 новых атак. Физически невозможно обновлять БД сигнатур формальных ССОВ за такие промежутки времени. Кроме того, увеличение объема сигнатур отрицательно сказывается на производительности систем.

Таким образом, актуальной является задача быстрой классификации сетевого трафика для обнаружения сетевых атак. Задача может быть поставлена либо как задача бинарной классификации (нормальный или аномальный трафик), либо как более сложная задача многоклассовой классификации, когда аномальный трафик в свою очередь классифицируется по заранее выделенным типам атак.

Применение методов машинного обучения для выявления сетевых атак

В последнее время все чаще для классификации трафика и выявления сетевых атак используются современные методы машинного обучения.

Так, ученые из ЮФУ в статьях [1,2] предлагают метод обнаружения низкоинтенсивных (low-rate) атак типа «отказ в обслуживании» (DDoS). Особенностью метода является предварительная кластеризация пакетов с помощью самоорганизующихся карт Кохонена. Выходной вектор самоорганизующейся карты является входным вектором многослойного персептрона, который осуществляет бинарную классификацию – определяет, является ли набор сетевых пакетов нормальным или атакующим. В результате достигнута ошибка распознавания 0,84%. В статье [3] в качестве эффективного инструмента выявления DDoS-атак предложено использовать нейронную сеть. Для обучения и тестирования нейронной сети использовался набор данных «NSL-KDD». Точность классификации составила 97,87%. Отметим, что предлагаемые подходы рассчитаны на выявление только одного класса атак типа «отказ в обслуживании».

Значительное число публикаций посвящено возможностям использования методов глубокого обучения для обнаружения и классификации сетевых атак. В статье [4] приводится обзор современных публикаций по этой теме. В статье [5] рассмотрены две задачи классификации атак – бинарная классификация и классификация на 4 класса атак. Авторы используют рекуррентные нейронные сети для классификации большого объема данных. В результате для бинарной классификации достигнута точность менее 0,1% ошибок, для классификации по типу атак – 0,5%.

В статье [6] для обнаружения DDoS атак произведено сравнение рекуррентных нейронных сетей, в том числе LSTM сетей с традиционным методом случайного леса. LSTM сети показали самую высокую точность ¬ 98,4% правильного обнаружения атак. В статье [7] используются сети с автокодировщиком со стохастическим алгоритмом определения порога срабатывания. Этот метод позволил увеличить точность обнаружения атак на наборе NSL-KDD до 88,65%. В статье [8] предлагается система для обнаружения и классификации как известных, так и неизвестных аномалий по 4 классам. Экспериментально определена оптимальная архитектура нейронной сети. В статье [9] рассматривается возможность автоматической кластеризации пакетов для системы обнаружения аномалий в корпоративных сетях. Аномальными считаются большие кластеры с высокой плотностью, а также малые или разреженные кластеры. Далее на этих данных обучаются алгоритмы бинарной классификации. На наборе NSL-KDD удается получить точность классификации 88%.

В статье [10] рассматривается классификация атак в беспроводных сетях IEEE 802.11. Атаки классифицируются на 3 класса с помощью многослойного автокодировщика. В работе [11] для классификации вредоносного трафика используются сверточные нейронные сети. Идея состоит в том, что сырые данные трафика преобразуются в изображения, которые распознаются сверточными сетями. При этом точность детектирования атак достигает 99.41% В работе [12] для детектирования трудно обнаруживаемого типа атак – сканирования портов и поиска уязвимостей используются глубокие сети доверия, комбинирующие подходы обучения с учителем и без учителя. В статье [13] представлен метод бинарной классификации атак на основе метода нечеткой кластеризации С-средних. Для повышения точности алгоритма используется частичная ручная разметка небольшой части обучающих данных. Подробное сравнение различных алгоритмов машинного обучения, применяемых в системах информационной безопасности приведено в статье [14]. Авторы рассматривают три задачи – обнаружение вторжений, анализ вредоносных программ и обнаружение спама. Выводы – для каждой задачи лучше применять свои методы, которые требуют непрерывного обучения и тщательной настройки параметров. В статье [15] рассматриваются методы кластеризации для обнаружения вторжений на основе метода к-средних.

Отметим, что методы глубокого обучения не обладают высокой производительностью особенно на этапе обучения, также не исследованы методы, которые позволяют проводить классификацию более чем по 4 классам атак.

В нашем исследовании проведено сравнение традиционных методов машинного обучения как для бинарной, так и многоклассовой классификации сетевого трафика по 9 типам атак.

Бинарная классификация

Обучению подвергались следующие классификаторы из библиотеки scikit-learn:

- DecisionTree – алгоритм решающих деревьев;
- RandomForest – алгоритм случайного леса;
- AdaBoost – алгоритм AdaBoost для бустинга деревьев;

- LogisticRegression – логистическая регрессия;
- KNeighbors – алгоритм k-ближайших соседей;
- SVC – алгоритм машины опорных векторов SVM для классификации;
- VotingClassifier – ансамблевый метод на основе голосования других классификаторов.

Обучение происходило на размеченном наборе данных UNSW-NB15 [16]. Предварительно все данные были нормализованы с использованием алгоритма StandardScaler, который для количественных столбцов выполняет вычитание среднего значение и делит получившееся на стандартное отклонение.

Для двоичной классификации было выбрано два класса:

Attack – наличие атаки (любой из 9 классов атак, размеченных в UNSW-NB15), взято 30000 случайных записей из набора данных.

Normal – отсутствие атаки, взято также 30000 случайных записей.

75% случайных записей преобразованного для двоичной классификации набора данных использовалось для обучения, оставшиеся 25% - для оценки точности, которая проводилась с помощью метода k-fold кроссвалидации (4 блока). Для каждого алгоритма классификации, кроме VotingClassifier, перебором (с помощью алгоритма GridSearchCV) выполнялась оптимизация гиперпараметров. Рассмотренные параметры и значения сведены в таблицу 1.

Таблица 1. Гиперпараметры алгоритмов классификации и их возможные значения
Table 1. Hyperparameters of classification algorithms and their possible values

Алгоритм	**Гиперпараметр**	**Значения**
DecisionTree	criterion	«entropy», «gini»
	min_samples_split	2, 20, 100, 250, 300, 500
	max_depth	None, 2, 5, 10, 25, 50, 100, 200, 500, 750, 1000
	min_samples_leaf	1, 2, 5, 10, 25, 50, 100, 200
	max_leaf_nodes	None, 25, 50, 100, 250, 500, 750
RandomForest	max_depth	None, 50, 100
	max_features	1, 3, 10
	min_samples_split	2, 3, 10
	min_samples_leaf	1, 3, 10
	n_estimators	50, 100, 300
	criterion	«gini», «entropy»
AdaBoost	algorithm	«SAMME»,»SAMME.R»
	n_estimators	1, 10, 50, 100, 250, 500
	learning_rate	0.1, 0.2, 0.5, 1, 2, 3
LogisticRegression	penalty	«l1», «l2»
	C	0.001, 0.01, 0.1, 1, 10, 100, 1000
	solver	“liblinear”, “saga” – для penalty=”l1”; “newton-cg”, “lbfgs”, “sag” – для penalty=»l2»
KNeighbors	n_neighbors	2, 3, 4, 5, 10, 15, 25, 50, 75, 100
	weights	“uniform”, “distance»
	algorithm	“ball_tree”, “kd_tree»
SVM	C	0.01, 0.1, 1, 10
	Gamma	0.01, 0.1, 1, 10

Таблица 2. Лучшие гиперпараметры для алгоритмов и оценки их точности (бинарная классификация)
Table 2. The best hyperparameters for algorithms and their estimated accuracies (binary classification)

Алгоритм с лучшими гиперпараметрами	Сбалансированная точность на обучающем наборе	Сбалансированная точность на тестовом наборе
DecisionTree criterion: “gini”, max_depth: 5, max_leaf_nodes: 25, min_samples_leaf: 1, min_samples_split: 100	0.9870	0.9867
RandomForest **criterion: “entropy”, max_depth: None, max_features: 10, min_samples_leaf: 1, min_samples_split: 10, n_estimators: 100**	**0.9871**	**0.9868**
AdaBoost algorithm: “SAMME”, learning_rate: 0.1, n_estimators: 500	0.9870	0.9866
LogisticRegression C: 100, penalty: “l2”, solver: “newton-cg»	0.9869	0.9867
KNeighbors algorithm: “ball_tree”, n_neighbors: 10, weights: “distance»	0.9858	0.9859
SVM C: 10, gamma: 0.01	0.9867	0.9864
VotingClassifier	**-**	**0.9868**

В алгоритме AdaBoost в качестве базового использовалось наилучшее решающее дерево, полученное для DecisionTree.

Алгоритм VotingClassifier реализует голосование лучших вариантов всех остальных алгоритмов, которые выдают вероятности принадлежности объекта классам. При голосовании используется принцип “soft”, когда выбирается класс с наибольшей суммой сумой вероятности принадлежности ему.

В качестве основной метрики оценки точности классификации используется balanced_accuracy_score – сбалансированная точность.

Основное преимущество данной метрики в адекватной оценке точности алгоритмов классификации с учетом сильного дисбаланса в количестве размеченных записей по каждому классу набора данных.

Значения данной метрики для лучших вариантов алгоритмов (с учетом гиперпараметрической оптимизации) сведены в таблицу 2. В отдельных колонках представлены значения сбалансированной точности для обучающего и тестового наборов.

Анализ таблицы 2 показывает, что наибольшую точность классификации, как на обучающем, так и на тестовом наборах обеспечивают алгоритмы RandomForest и VotingClassifier. Остальные алгоритмы показывают также весьма близкую точность, большую 0.98. С точки зрения практики использование RandomForest более оправдано, чем VotingClassifier, т.к. не требуется обучение и обсчет алгоритмов всех остальных классификаторов, достаточно только одного.

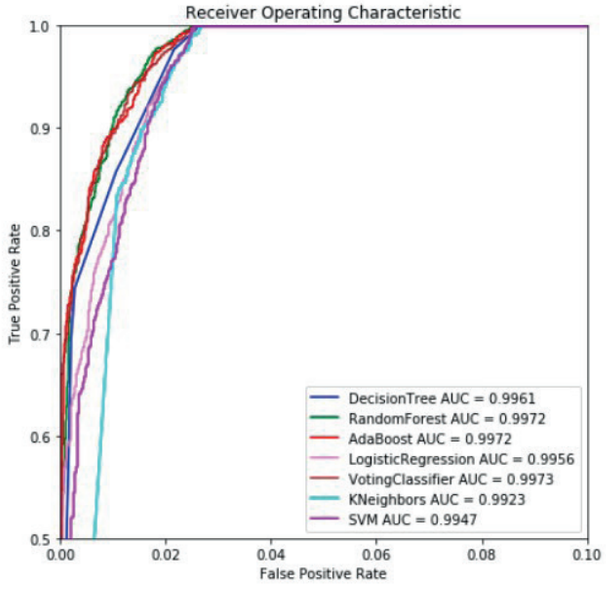

Рис. 1. ROC-кривые для алгоритмов двоичной классификации
Fig 1. ROC-curves for binary classification algorithms

На рисунке 2 показана матрица ошибок классификации алгоритма RandomForest, построенная для тестового набора данных.

Видно, что алгоритм RandomForest обеспечивает 0.01% ложно-отрицательных 2.6% ложно-положительных срабатываний, что является вполне приемлемым. Для исключения части ложно-положительных срабатываний может быть проведен дополнительный селективный анализ с использованием дальнейшей классификации трафика по основным категориям атак с последующим их анализом с использованием дополнительных инструментов (например, сигнатурных правил корреляций).

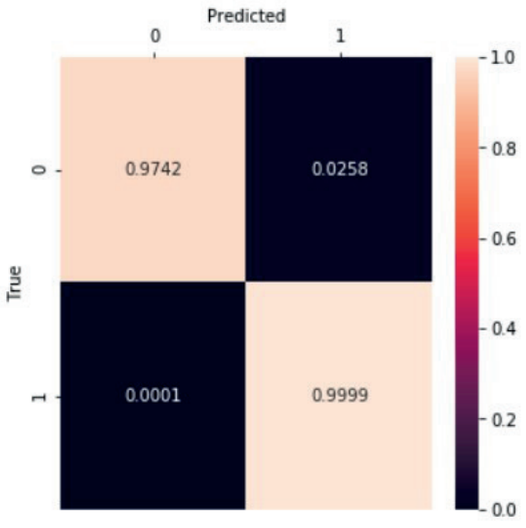

Рис. 2. Матрица ошибок двоичной классификации для RandomForest
Fig 2. Binary classification error matrix for RandomForest

Далее рассмотрим классификацию атак по категориям.

Многоклассовая классификация атак

В размеченном наборе данных UNSW-NB15 имеются 9 классов атак:

- Fuzzers – генерация случайных данных, чтобы вызвать отказ программы или сети.
- Analysis – содержит различные атаки, связанные со сканированием портов, спамом и внедрением в HTML-файлы.
- Bakdoors – обход механизмов защиты с целью скрытого доступа к данным или программам.
- DoS – отказ в обслуживании сервера или сетевого ресурса.
- Exploits – эксплуатация известных атакующему уязвимостей в операционной системе или программе.
- Generic – техника обнаружения трафика, шифрованного блочным шифром.
- Reconnaissance – разведывательные атаки.
- Shellcode – передача небольших частей кода, используемых для эксплуатации уязвимостей программ.
- Worms – атаки, связанные с самореплицируемыми вирусами.

В таблице 3 приведено количество размеченных записей об атаках каждого типа, взятых из набора данных. Из набора данных были удалены записи об обычном трафике, т.к. целью классификации на втором шаге является уточнение типа атаки.

Для многоклассовой классификации были использованы те же самые алгоритмы из библиотеки scikit-learn, которые были ранее применены для бинарной классификации. Аналогично двоичной классификации набор данных был поделен на обучающую и тестовую части в процентном соотношении 75% и 25%.

Также проводилась оптимизация гиперпараметров алгоритмов с использованием аналогичной техники кроссвалидации. В качестве метрики использовалась сбалансированная точность.

Таблица 3. Количество записей для атак каждого класса
Table 3. The number of entries for each class of attacks

Класс	Количество записей
Fuzzers	20960
Analysis	2032
Bakdoors	1880
DoS	5500
Exploits	27434
Generic	7603
Reconnaissance	9991
Shellcode	1456
Worms	171

В таблице 4 представлены результаты исследования алгоритмов. Анализ данной таблицы показывает, что наилучшим алгоритмом является AdaBoost. Он немного улучшает результаты DecisionTree, на базе которого построен. Оба алгоритма обеспечивают точность на тестовых данных, большую 0.691. Остальные алгоритмы показывают худшие результаты.

Таблица 4. Лучшие гиперпараметры для алгоритмов и оценки их точности (многоклассовая классификация)
Table 4. The best hyperparameters for algorithms and their estimated accuracies (binary classification)

Алгоритм с лучшими гиперпараметрами	Сбалансированная точность на обучающем наборе	Сбалансированная точность на тестовом наборе
DecisionTree criterion: «entropy», max_depth: None, max_leaf_nodes: 500, min_samples_leaf: 2, min_samples_split: 2»	0.6022	0.6916
RandomForest criterion: «gini», max_depth: 100, max_features: 10, min_samples_leaf: 1, min_samples_split: 2, n_estimators: 100	0.4913	0.5602
AdaBoost **algorithm: «SAMME», learning_rate: 0.1, n_estimators: 1**	**0.6024**	**0.6918**
LogisticRegression C: 1000, penalty: «l1», solver: «liblinear»	0.3659	0.5057
KNeighbors algorithm: “ball_tree”, n_neighbors: 4, weights: “uniform»	0.4042	0.6529
SVM C: 1, gamma: 0.1	0.3914	0.5409
VotingClassifier	-	**0.6394**

На рисунке 3 приведена матрица ошибок классификации для алгоритма AdaBoost.

Анализ рисунка позволяет сделать следующие выводы:

а) алгоритм AdaBoost достаточно точно выявляет атаки Fuzzers (90%), Exploits (88%), Generic(88%), чуть хуже Reconnaissance (73%), Shellcode (70%), гораздо хуже Worms (57%);

б) он довольно часто считает, что атака относится к классу Exploits вместо реального класса DoS (55%), Analysis (48%), Backdoor (47%), Worms (26%), Reconnaissance (20%), Shellcode (15%);

в) AdaBoost также ошибочно считает, что атака относится к классу Fuzzers, вместо реального класса Backdoor (24%), Analysis (18%).

Ошибки б) и в) возможно связаны с тем, что атаки эксплуатации уязвимостей и фазинга по признакам и последствиям похожи на атаки других классов.

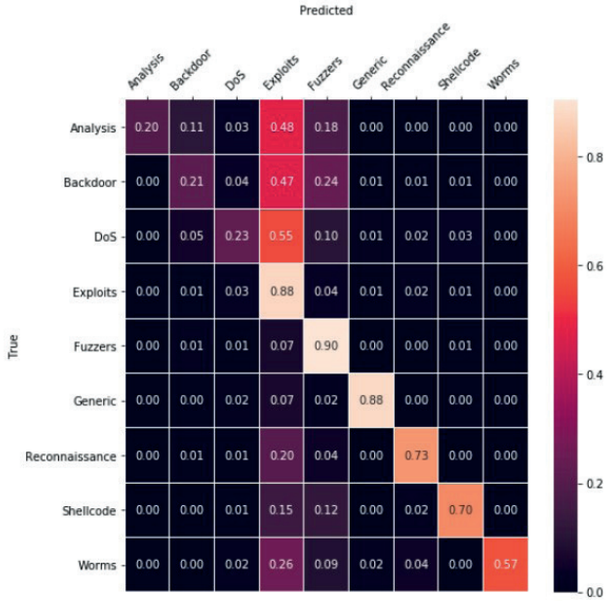

Рис. 3. Матрица ошибок многоклассовой классификации для AdaBoost
Fig. 3. Multi-class classification error matrix for AdaBoost

Заключение

В целом для анализа трафика в корпоративных сетях и выявления атак можно использовать алгоритм RandomForest, а для уточнения классов атак – AdaBoost. Основное достоинство данных алгоритмов в том, что они могут быть использованы для выявления новых типов или разновидностей атак в пределах описанных классов, что не могут обычные сигнатурные методы. Однако для известных атак сигнатурные методы могут быть полезны для их точной идентификации. В целом для корпоративных сетей рекомендуется использовать сочетание сигнатурных методов анализа и методов машинного обучения.

Благодарности

Исследование выполнено при финансовой поддержке РФФИ в рамках проектов 16-29-09639, 18-07-01446, 18-47-560017.

Список использованных источников

[1] *Абрамов Е.С., Тарасов Я.В.* Применение комбинированного нейросетевого метода для обнаружения низкоинтенсивных DDoS-атак на web-сервисы // Инженерный вестник Дона. 2017. Т. 46, № 3(46). С. 59. URL: https://elibrary.ru/item.asp?id=30753050 (дата обращения: 24.06.2018).

[2] *Тарасов Я.В.* Исследование применения нейронных сетей для обнаружения низкоинтенсивных DDoS-атак прикладного уровня // Вопросы кибербезопасности. 2017. № 5(24). С. 23-29. DOI: 10.21681/2311-3456-2017-5-23-29

[3] *Воробьёва Ю.Н.* и др. Нейросетевая модель выявления DDOS-атак // Вестник технологического университета. 2018. Т. 21, №. 2. С. 94-98. URL: https://elibrary.ru/item.asp?id=32683897 (дата обращения: 24.06.2018).

[4] *Bodström T., Hämäläinen T.* State of the Art Literature Review on Network Anomaly Detection with Deep Learning / O. Galinina, S. Andreev, S. Balandin, Y. Koucheryavy (Eds.) // Internet of Things, Smart Spaces, and Next Generation Networks and Systems. NEW2AN 2018, ruSMART 2018. Lecture Notes in Computer Science. Vol. 11118. Springer, Cham, 2018. Pp. 64-76. DOI: 10.1007/978-3-030-01168-0_7

[5] *Yin C., Zhu Y., Fei J., He X.* A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks // IEEE Access. 2017. Vol. 5. Pp. 21954-21961. DOI: 10.1109/ACCESS.2017.2762418

[6] *Yuan X., Li C., Li X.* DeepDefense: Identifying DDoS Attack via Deep Learning // 2017 IEEE International Conference on Smart Computing (SMARTCOMP). Hong Kong, 2017. Pp. 1-8. DOI: 10.1109/SMARTCOMP.2017.7946998

[7] *Aygun R.C., Yavuz A.G.* Network Anomaly Detection with Stochastically Improved Autoencoder Based Models // 2017 IEEE 4th International Conference on Cyber Security and Cloud Computing (CSCloud). New York, NY, 2017. Pp. 193-198. DOI: 10.1109/CSCloud.2017.39

[8] *Van N., Thinh T., Sach L.* An anomaly-based network intrusion detection system using Deep learning // 2017 International Conference on System Science and Engineering (ICSSE). Ho Chi Minh City, 2017. Pp. 210-214. DOI: 10.1109/ICSSE.2017.8030867

[9] *Baek S., Kwon D., Kim J., Suh S.C., Kim H., Kim I.* Unsupervised Labeling for Supervised Anomaly Detection in Enterprise and Cloud Networks // 2017 IEEE 4th International Conference on Cyber Security and Cloud Computing (CSCloud). New York, NY, 2017. Pp. 205-210. DOI: 10.1109/CSCloud.2017.26

[10] *Thing V.L.L.* IEEE 802.11 Network Anomaly Detection and Attack Classification: A Deep Learning Approach // 2017 IEEE Wireless Communications and Networking Conference (WCNC). San Francisco, CA, 2017. Pp. 1-6. DOI: 10.1109/WCNC.2017.7925567

[11] *Wang W., Zhu M., Zeng X., Ye X., Sheng Y.* Malware traffic classification using convolutional neural network for representation learning // 2017 International Conference on Information Networking (ICOIN). Da Nang, 2017. Pp. 712-717. DOI: 10.1109/ICOIN.2017.7899588

[12] *Viet H.N., Van Q.N., Trang L.L.T., Nathan S.* Using Deep Learning Model for Network Scanning Detection // Proceedings of the 4th International Conference on Frontiers of Educational Technologies (ICFET '18). ACM, New York, NY, USA, 2018. Pp. 117-121. DOI: 10.1145/3233347.3233379

[13] *Teoh T.T., Nguwi Y.Y., Elovici Y., Ng W.L., Thiang S.Y.* Analyst intuition inspired neural network based cyber security anomaly detection // International journal of innovative computing information and control. 2018. Vol. 14, no. 1. Pp. 379-386. DOI: 10.24507/ijicic.14.01.379

[14] *Apruzzese G., Colajanni M., Ferretti L., Guido A., Marchetti M.* On the effectiveness of machine and deep learning for cyber security // 2018 10th International Conference on Cyber Conflict (CyCon). Tallinn, 2018. Pp. 371-390. DOI: 10.23919/CYCON.2018.8405026

[15] *Makkar G., Jayaraman M., Sharma S.* Network Intrusion Detection in an Enterprise: Unsupervised Analytical Methodology / V. Balas, N. Sharma, A. Chakrabarti (Eds.) // Data Management, Analytics and Innovation. Advances in Intelligent Systems and Computing. Vol. 808. Springer, Singapore, 2019. Pp. 451-463. DOI: 10.1007/978-981-13-1402-5_34

[16] *Moustafa N., Jill S.* UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set) // 2015 Military Communications and Information Systems Conference (MilCIS). Canberra, ACT, 2015. Pp. 1-6. DOI: 10.1109/MilCIS.2015.7348942

Поступила 24.06.2018; принята в печать 20.08.2018;
опубликована онлайн 30.09.2018.

References

[1] Abramov E.S., Tarasov Y.V. Application of the combined neural network method for web-oriented low-rate DDoS-attack detection. *Engineering journal of Don*. 2017; 46(3):59. Available at: https://elibrary.ru/item.asp?id=30753050 (accessed 24.06.2018). (In Russian)

[2] Tarasov Ya.V. Investigation of the Use of Neural Networks for Detecting Low-Intensive DDoS-Atak of Applied Level. *Voprosy kiberbezopasnosti*. 2017; 5(24):23-29. (In Russian) DOI: 10.21681/2311-3456-2017-5-23-29

[3] Vorobeva Y.N., Kataseva D.V., Katasev A.S., Kirpichnikov A.P. Neural network model of detecting DDoS-Attacks. *Vestnik tekhnologicheskogo universiteta*. 2018; 21(2):94-98. Available at: https://elibrary.ru/item.asp?id=32683897 (accessed 24.06.2018). (In Russian)

[4] Bodström T., Hämäläinen T. State of the Art Literature Review on Network Anomaly Detection with Deep Learning. O. Galinina, S. Andreev, S. Balandin, Y. Koucheryavy (Eds.) *Internet of Things, Smart Spaces, and Next Generation Networks and Systems*. NEW2AN 2018, ruSMART 2018. Lecture Notes in Computer Science. Vol. 11118. Springer, Cham, pp. 64-76, 2018. DOI: 10.1007/978-3-030-01168-0_7

[5] Yin C., Zhu Y., Fei J., He X. A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks. *IEEE Access*. 2017; 5:21954-21961. DOI: 10.1109/ACCESS.2017.2762418

[6] Yuan X., Li C., Li X. DeepDefense: Identifying DDoS Attack via Deep Learning. *2017 IEEE International Conference on Smart Computing (SMARTCOMP)*. Hong Kong, pp. 1-8, 2017. DOI: 10.1109/SMARTCOMP.2017.7946998

[7] Aygun R.C., Yavuz A.G. Network Anomaly Detection with Stochastically Improved Autoencoder Based Models. *2017 IEEE*

4th International Conference on Cyber Security and Cloud Computing (CSCloud). New York, NY, pp. 193-198, 2017. DOI: 10.1109/CSCloud.2017.39

[8] Van N., Thinh T., Sach L. An anomaly-based network intrusion detection system using Deep learning. *2017 International Conference on System Science and Engineering (ICSSE)*. Ho Chi Minh City, pp. 210-214, 2017. DOI: 10.1109/ICSSE.2017.8030867

[9] Baek S., Kwon D., Kim J., Suh S.C., Kim H., Kim I. Unsupervised Labeling for Supervised Anomaly Detection in Enterprise and Cloud Networks. *2017 IEEE 4th International Conference on Cyber Security and Cloud Computing (CSCloud)*. New York, NY, pp. 205-210, 2017. DOI: 10.1109/CSCloud.2017.26

[10] Thing V.L.L. IEEE 802.11 Network Anomaly Detection and Attack Classification: A Deep Learning Approach. *2017 IEEE Wireless Communications and Networking Conference (WCNC)*. San Francisco, CA, pp. 1-6, 2017. DOI: 10.1109/WCNC.2017.7925567

[11] Wang W., Zhu M., Zeng X., Ye X., Sheng Y. Malware traffic classification using convolutional neural network for representation learning. *2017 International Conference on Information Networking (ICOIN)*. Da Nang, pp. 712-717, 2017. DOI: 10.1109/ICOIN.2017.7899588

[12] Viet H.N., Van Q.N., Trang L.L.T., Nathan S. Using Deep Learning Model for Network Scanning Detection. *Proceedings of the 4th International Conference on Frontiers of Educational Technologies (ICFET '18)*. ACM, New York, NY, USA, pp. 117-121, 2018. DOI: 10.1145/3233347.3233379

[13] Teoh T.T., Nguwi Y.Y., Elovici Y., Ng W.L., Thiang S.Y. Analyst intuition inspired neural network based cyber security anomaly detection. *International journal of innovative computing information and control.* 2018; 14(1):379-386. DOI: 10.24507/ijicic.14.01.379

[14] Apruzzese G., Colajanni M., Ferretti L., Guido A., Marchetti M. On the effectiveness of machine and deep learning for cyber security. *2018 10th International Conference on Cyber Conflict (CyCon)*. Tallinn, pp. 371-390, 2018. DOI: 10.23919/CYCON.2018.8405026

[15] Makkar G., Jayaraman M., Sharma S. Network Intrusion Detection in an Enterprise: Unsupervised Analytical Methodology. V. Balas, N. Sharma, A. Chakrabarti (Eds.) *Data Management, Analytics and Innovation. Advances in Intelligent Systems and Computing.* Vol. 808. Springer, Singapore, pp. 451-463, 2019. DOI: 10.1007/978-981-13-1402-5_34

[16] Moustafa N., Jill S. UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set) *2015 Military Communications and Information Systems Conference (MilCIS)*. Canberra, ACT, pp. 1-6, 2015. DOI: 10.1109/MilCIS.2015.7348942

Submitted 24.06.2018; revised 20.08.2018;
published online 30.09.2018.

About the authors:

Nadezhda F. Bakhareva, Doctor of Technical Sciences, Professor, Povolzhskiy State University of Telecommunications & Informatics (23 Tolstoy Str., Samara 443010, Russia), ORCID: http://orcid.org/0000-0002-9850-7752, Bahareva-nf@psuti.ru

Veniamin N. Tarasov, Doctor of Technical Sciences, Professor, Povolzhskiy State University of Telecommunications & Informatics (23 Tolstoy Str., Samara 443010, Russia), ORCID: http://orcid.org/0000-0002-9318-0797, veniamin_tarasov@mail.ru

Aleksandr E. Shukhman, Candidate of Pedagogic Sciences, Associate Professor, Head of the Department of Geometry and Computer Science, Orenburg State University (13 Pobeda Av., Orenburg 460000, Russia), ORCID: http://orcid.org/0000-0002-4303-2550, shukhman@gmail.com

Petr N. Polezhaev, Lecturer at the Department of Computer Security and Mathematical Maintenance of Information Systems, Orenburg State University (13 Pobeda Av., Orenburg 460000, Russia), ORCID: http://orcid.org/0000-0001-7747-646X, newblackpit@mail.ru

Yuri A. Ushakov, Candidate of Engineering Sciences, Associate Professor at the Department of Geometry and Computer Science, Orenburg State University (13 Pobeda Av., Orenburg 460000, Russia), ORCID: http://orcid.org/0000-0002-0474-8919, unpk@mail.ru

Artem A. Matveev, Student, Department of Computer Security and Mathematical Maintenance of Information Systems, Orenburg State University (13 Pobeda Av., Orenburg 460000, Russia), ORCID: http://orcid.org/0000-0002-5362-4373, artemi645@gmail.com