Anzina Antonina Viktorovna – North Caucasus Federal University; e-mail: anttoniinav@gmail.com; 2, Kulakova street, Stavropol, 355041, Russia; student.

Medvedeva Anastasia Dmitrievna – e-mail: medvedeva.nastya26@mail.ru; student.

Emelyanov Evgeny Alekseevich – e-mail: eemelianov@ncfu.ru; phone: +79886261908; lecturer.

УДК 004.056.5 DOI 10.18522/2311-3103-2020-6-117-128

В.А. Буковшин, П.А. Чуб, Д.А. Короченцев, Л.В. Черкесова, Н.В. Болдырихин, О.А. Сафарьян

АНАЛИЗ ЗАШИФРОВАННОГО СЕТЕВОГО ТРАФИКА НА ОСНОВЕ ВЫЧИСЛЕНИЯ ЭНТРОПИИ И ПРИМЕНЕНИЯ НЕЙРОСЕТЕВЫХ КЛАССИФИКАТОРОВ

Анализ сетевого трафика позволяет решить множество задач, таких как: определение закономерности передачи данных по сети, сбор статистики об использовании веб–приложений, мониторинг и дальнейшее исследование сетевой нагрузки, определение потенциальных вредоносных программных средств и сетевых атак и т.д. На данный момент до 40% Интернет–траффика принадлежит неизвестным приложениям. Это говорит о том, что для области анализа сетевого трафика задача классификации приложений приобрела особую важность. Совершенствование программного обеспечения в области сетевых технологий способствовало обнаружению серьёзных уязвимостей в реализации некоторых сетевых протоколов, а именно: TCP и HTTP. С помощью анализаторов сетевого трафика злоумышленник получал доступ к содержимому пакетов данных, передающихся по сети. Однако с повышением квалификации информационного сообщества в области компьютерной безопасности, а также с развитием стандартов сетевых технологий, анализ сетевого трафика заметно усложнился. Возросшее применение математических методов защиты информации, таких как симметричные и ассиметричные криптографические протоколы, привела к тому, что большинство подходов к анализу сетевого трафика потеряли значение и перестали применяться. Поэтому актуален поиск новых решений задачи классификации сетевого трафика с учетом возможности его шифрования. Статья посвящена описанию нового смешанного подхода к анализу сетевого трафика, основанного на совокупном использовании теории информации и алгоритмов машинного обучения. Также приводится сравнительный анализ предложенного метода с уже существующими подходами, основанными как на теории информации, так и на машинном обучении. Целью исследований является разработка алгоритма, основанного на интеллектуальном подходе к анализу сетевого трафика. Предлагаемый алгоритм базируется на вычислении энтропии и применении нейросетевых классификаторов. Задачи исследований включают: проведение теоретического обоснования предложенного подхода в области теории информации, а также алгоритмов машинного обучения; проведение структурного описания реализованных алгоритмов вычисления энтропии и классификации приложений, генерирующих зашифрованный траффик; сравнительный анализ предложенного алгоритма с уже существующими подходами к анализу зашифрованного сетевого трафика. Результатом исследований является новый алгоритм, позволяющий с высокой степенью достоверности классифицировать различные виды зашифрованного трафика.

Сетевой трафик; машинное обучение; персептрон; нейронная сеть; нейрон; автокодировщик; функция активации; N-усечённая энтропия; функция максимального правдоподобия.

V.A. Bukovshin, P.A. Chub, D.A. Korochentsev, L.V. Cherkesova, N.V. Boldyrikhin, O.A. Safaryan

ANALYSIS OF ENCRYPTED NETWORK TRAFFIC BASED ON ENTROPY CALCULATION AND APPLICATION OF NEURAL NETWORK CLASSIFIERS

Network traffic analysis allows you to solve many problems, such as: determining the pattern of data transmission over the network, collecting statistics on the use of web applications, monitoring and further researching network load, identifying potential malicious software and network attacks, etc. 40% of Internet traffic belongs to unknown applications. This suggests that for the area of network traffic analysis, the task of classifying applications has acquired particular importance. Improvements in software in the field of network technologies have contributed to the discovery of serious vulnerabilities in the implementation of some network protocols, namely TCP and HTTP. By using network traffic analyzers, an attacker gained access to the contents of data packets transmitted over the network. However, with the increasing qualifications of the information community in the field of computer security, as well as with the development of network technology standards, the analysis of network traffic has become noticeably more complicated. The increased use of mathematical methods for protecting information, such as symmetric and asymmetric cryptographic protocols, has led to the fact that most approaches to the analysis of network traffic have lost their meaning and are no longer used. Therefore, the search for new solutions to the problem of classifying network traffic, taking into account the possibility of its encryption, is relevant. The article is devoted to the description of a new mixed approach to the analysis of network traffic, based on the combined use of information theory and machine learning algorithms. It also provides a comparative analysis of the proposed method with existing approaches based on both information theory and machine learning. The aim of the research is to develop an algorithm based on an intelligent approach to the analysis of network traffic. The proposed algorithm is based on calculating entropy and using neural network classifiers. Research objectives include: theoretical substantiation of the proposed approach in the field of information theory, as well as machine learning algorithms; carrying out a structural description of the implemented algorithms for calculating entropy and classifying applications that generate encrypted traffic; comparative analysis of the proposed algorithm with existing approaches to the analysis of encrypted network traffic. The result of the research is a new algorithm that allows classifying various types of encrypted traffic with a high degree of reliability.

Network traffic; intellectual algorithm; machine learning; perceptron; neural network; neuron; autoencoder; activation function; N-truncated entropy; maximum likelihood function.

Введение. Исследования в области анализа сетевого трафика вызывали интерес ещё со времён рождения Всемирной паутины. Рост вычислительной мощности и развитие телекоммуникационных технологий предоставили возможность объединять компьютеры в сети различного масштаба и передавать информацию по всему миру. Востребованность Интернет–технологий предопределила «информационный взрыв»: производство информации за последние пять лет превысило её объёмы за всю предшествующую историю человечества [1–21]. В сложившейся ситуации анализ сетевого трафика стал одной из самых часто встречающихся и важных задач, как сетевого администрирования, так и информационной безопасности [3, 6–12, 18]. Следует отметить, что решение задачи анализа и классификации сетевого трафика в последнее время заметно усложнилась в связи с тем, что многие приложения стали шифровать трафик. Поэтому представляется актуальным поиск новых подходов к решению этой задачи, позволяющих на основании косвенных признаков проводить анализ и классификацию трафика.

Основная часть. Клод Шеннон в своей работе [22] определил энтропию как показатель неопределённости. В общем случае, чем выше энтропия, тем выше зашумлённость канала передачи данных и тем меньше количество информации. Если имеется *m* возможных событий $A_1, A_2, ..., A_m$ с соответствующими вероятностями их появления $p_1, p_2, ..., p_m$ энтропия определяется следующей формулой:

$$H = -\sum_{i=1}^{m} p_i * log\left(p_i\right). \quad (1)$$

Информационный поток с равномерным распределением битов будет иметь максимальное значение энтропии, известное как идеальная энтропия. Непредсказуемость поведения информационного потока обуславливает неопределённость, тем самым усиливая энтропию. Однако для сетевых пакетов небольшого размера необходима более специфическая оценочная функция, которая известна под названием N-усечённой энтропии [17].

N-усечённая энтропия для случайного сетевого пакета будет определятся следующей формулой [17]:

$$H_N\left(u\right) = \frac{1}{m_N} * \sum_{N}\left[\frac{N!}{n_1!*...*n_m!} * \sum_{i=1}^{m}\left(\frac{-n_i}{N}\right) * log\frac{n_i}{N}\right], \quad (2)$$

где N – объём захватываемых данных, m – общее число возможных символов, m = 256, n_i –частота появления символа.

Если рассматривать содержимое пакета как случайную величину, то у каждого сетевого пакета будет своя соответствующая плотность распределения вероятностей символов. Какие-то сетевые пакеты могут описываться семейством нормальных распределений, другие же будут иметь распределение Пуассона и др. На момент захвата очередного сетевого пакета его распределение неизвестно, однако известна его длина, а также энтропия сетевого пакета такой же длины со случайным содержимым.

Функцией правдоподобия в математической статистике называют совместное распределение выборки из параметрического распределения, рассматриваемое как функция параметра. Выборкой в данном случае является множество сетевых пакетов, распределение которых описывается конкретными параметрами [11]. В случае с нормальным распределением параметрами будут математическое ожидание и отклонение, в случае с распределением Пуассона – математическое ожидание. В случае с зашифрованным трафиком имеем некоторое распределение сетевого пакета, которое зависит от неизвестного параметра Q, тогда функция, которая зависит от параметра Q, при фиксированном событии X (длина пакета, время его появления и другие характеристики сетевого траффика, которые будут описаны ниже), и является функцией правдоподобия, которая показывает, насколько правдоподобна гипотеза о параметре Q.

Опишем алгоритм вычисления функции правдоподобия поэтапно:

1. Генерация псевдослучайной последовательности чисел от 0 до 255 длиной 64 байта в цикле с 10000 итерациями.

2. Вычисление N-усечённой энтропии, определяемой формулой (2), при N = 64.

3. Сбор экспериментального набора данных при помощи программного обеспечения для анализа сетевых пакетов с помощью программы Wireshark.

4. Случайный выбор сетевого пакета из числа захваченных экспериментальных данных, извлечение 64 байтов содержимого сетевого пакета и их запись в матрицу размером 100*64, которая представляет матрицу входных данных для функции правдоподобия.

5. Вычисление энтропии $H_i\left(p\right), i = 1,\ldots,k,\ k = 100$ для каждого изъятого пакета по формуле (1).

6. Вычисление среднеквадратичного отклонения по формуле:

$$\partial = \sum_{i=1}^{k} \left(H_i(p) - H_u(p) \right)^2 / (k-1). \quad (3)$$

7. Проверка принадлежности рассчитанного значения энтропии промежутку $\left[H_N(u) - \partial; H_N(u) + \partial \right]$. Если принадлежит, то захваченный сетевой трафик зашифрован.

Для дальнейшей классификации захваченного зашифрованного траффика используются нейросетевые классификаторы.

Для решения задачи классификации зашифрованного трафика применялась глубокая нейронная сеть с обратным распространением ошибки.

Рассмотрим структуру реализованной нейронной сети и применяющихся алгоритмов. Для данной задачи было выделено восемь классов выходных объектов, а именно: потоковый аудио- и видео-контент под VPN, VPN-торрент, VPN-VoIP, VPN-чат, FTP, SMTP. В качестве архитектуры нейронной сети использовался многослойный автокодировщик (далее МАК).

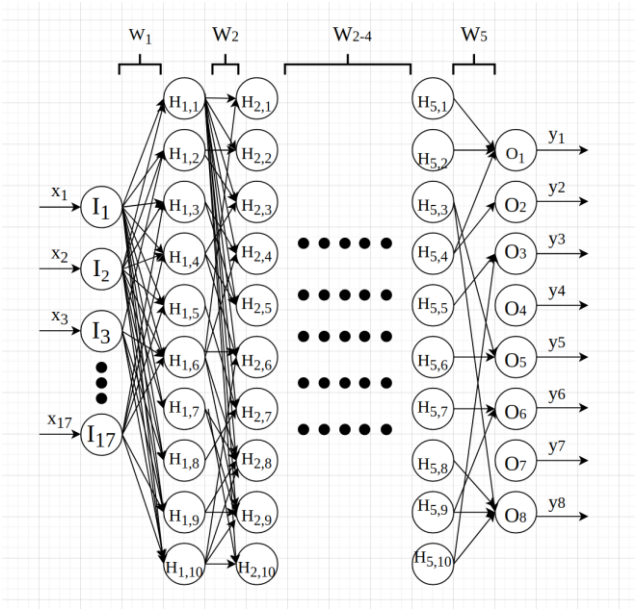

Рис. 1. Предлагаемая архитектура автокодировщика

Автокодировщик представляет собой нейронную сеть обратного распространения ошибки, которая обучается без учителя. Данная архитектура подходит для решения поставленной задачи, потому что отсутствует возможность корректно составить тренирующий набор данных, так как перехватываемый трафик зашифрован. Для улучшения точности и производительности нейронной сети автокодировщики обычно складывают в слои [15].

Предлагаемая нейронная сеть (рис. 1) состоит из 17 входных нейронов, которые представляют собой 17 характеристик сетевого трафика (выбранные характеристики будут описаны ниже); скрытый слой содержит 50 нейронов, распределённых на 5 скрытых слоёв, и выходной слой, состоящий из 8 нейронов, каждый из которых соответствует конкретному приложению, генерирующему зашифрованный трафик.

На рис. 2 можно увидеть, как три автокодировщика, структура которых представлена на рис. 1, и обозначенных как $АК_i$, $i=1,2,3$, последовательно объединены в единую глубокую нейронную сеть, при этом, выход каждого автокодировщика является входом последующего за ним.

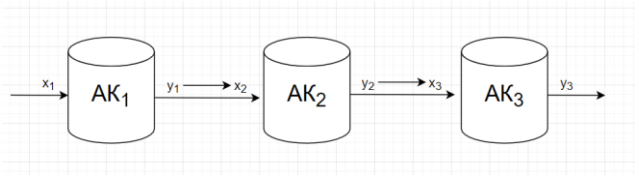

Рис. 2. Предлагаемая архитектура совокупного автокодировщика

На рис. 3 приведена блок-схема обучения описанной структуры. Она состоит в обучении каждой вложенной нейронной сети по отдельности в отдельных потоках, в то время, как её выходные значения (входные значения последующей нейронной сети) блокируются до момента окончания обучения. Подробное описание реализации данного процесса приведено ниже.

Для обучения каждого отдельного автокодировщика применяется логистическая функция активации, которую можно записать в виде следующей формулы:

$$f(s)=\frac{1}{1+e^{-as}},0<s\le 1. \tag{4}$$

Для расчёта погрешности вывода автокодировщика используется следующая функция потерь:

$$L(x)=\left|x-f\left(g(x)\right)\right|, \tag{5}$$

где $g(x)$ – функция декодирования результатов работы МАК, x – вектор выходных значений автокодировщика.

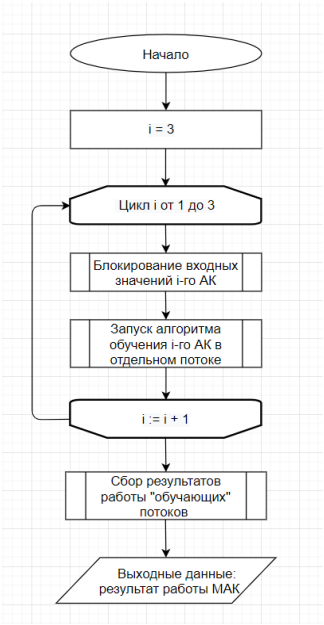

Рис. 3. Блок-схема обучения МАК

Рассмотрим алгоритм обучения многослойного автокодировщика поэтапно:

На вход МАК подаётся вектор из 17 выбранных характеристик сетевого трафика (формат входных данных будет подробно описан ниже).

Входные данные проходят процесс бинаризации, находится вектор $H^n = \left(h_0, h_1, ..., h_{(n-1)}\right)$, где $h_i = f\left(w_i * x_i + b\right), i = 0, ..., n-1$, а b – член смещения.

Запускаем работу МАК, получаем вектор выходных значений, декодируем его с помощью функции $g(x)$, описанной выше.

1. Вычисляем функцию потерь $L(x)$ по формуле (5).
2. В цикле для каждого автокодировщика МАК:
   2.1. Блокируем выходные значения;
   2.2. Запускаем алгоритм обратного распространения ошибки и корректируем веса нейронов;
   2.3. Разблокируем выходные значения и переходим к следующему автокодировщику.
3. Повторять процесс обучения 100 эпох.

Алгоритм обратного распространения ошибки представляет собой изменение весовых коэффициентов нейронов в соответствии с посчитанным значением функции потерь $L(x)$, это можно выразить следующей формулой:

$$w_{\left(i^{new}\right)} = w_i \pm L\left(y_i\right). \tag{6}$$

Опишем процесс обработки входных данных для реализованного программного средства. В качестве экспериментальных наборов данных решено было выбрать датасеты Канадского университета кибербезопасности, известные под названиями «ISCX VPN-nonVPN» и «ISCX-Tor/Non-Tor». Представленные наборы данных содержат такие типы сетевого траффика, как: траффик систем обнаружения вторжений, траффик сетей Tor-неTor, траффик виртуальных сетей VPN-неVPN и многие другие. Из данных наборов случайным образом было выбрано 5000 пакетов для обучения и тестирования нейронной сети. Соотношение обучающего набора данных к тестирующему – 9:1.

Важным моментом является определение характеристик сетевого траффика, которые определяют обучение нейронной сети. Исчерпывающее исследование важнейших сетевых характеристик было проведено в работе [15]. Все 17 характеристик, выбранных для данной работы, представлены в табл. 1. Выбор характеристик сетевого потока был сделан на основании исследования, проведённого в работе, а также на основании проведённых экспериментов по обучению вышеописанной нейронной сети. Характеристики, которые не повлияли на приближение к наилучшему результату классификации или повлияли мало, были отброшены.

Таблица 1

Выбранные характеристики сетевого траффика

1	DUR	Продолжительность передачи потока сетевых пакетов
2	flowPacketsPerSecond	Количество пакетов в потоке в секунду, pps
3	BIAT_total	Общее время между прибытием пакетов для обратного соединения
4	STD_flowiat	Стандартное отклонение времени между прибытием пакетов

5	MIN_active	Минимальное значение времени активной работы потока
6	flowBytesPerSecond	Количество байтов в потоке в секунду, pps
7	MEAN_active	Среднее значение времени активной работы потока
8	STD_active	Стандартное отклонение времени активной работы потока
9	MIN_idle	Минимальное значение времени бездействия потока
10	BIAT_max	Максимальное время между прибытием пакетов для обратного соединения
11	MIN_flowiat	Минимальное значение времени между прибытием пакетов
12	STD_idle	Стандартное отклонение времени бездействия потока
13	MAX_idle	Максимальное значение времени бездействия потока
14	BIAT_min	Минимальное значение времени между прибытием пакетов для обратного соединения
15	FIAT_total	Общее время между прибытием пакетов для прямого соединения
16	MAX_flowiat	Максимально значение времени между прибытием пакетов
17	MIN_idle	Минимальное значение времени бездействия потока

Обучение вышеописанной глубокой нейронной сети было реализовано с помощью асинхронного программирования. Особенностью асинхронного программирования является то, что можно запускать конкретные функции в разных потоках; таким образом, функции выполняются параллельно.

Для того, чтобы реализовать корректную асинхронность, необходимо учитывать характерные проблемы, связанные с параллельным выполнением инструкций, а именно, состояние гонки и нарушение целостности данных. В целях разрешить эти проблемы на этапе обучения автокодировщиков их выходы блокируются и освобождаются только по завершению функций обучения. Таким образом, можно обеспечить параллельное обучение автокодировщиков без нарушения целостности их входных значений.

Блокирование входных нейронов реализовано с помощью структуры, специально созданной для защиты целостности данных нейронов. Данная структура использует технологию так называемых мьютексов (с англ. to mute – заглушить, выключить). Если инструкции, выполняемые в одном потоке, модифицируют данные в конкретный момент времени, защищаемые мьютексом, то инструкции, выполняемые в другом потоке, в этот момент времени модифицировать их не могут.

Было проведено исследование асимптотической сложности вышеописанного программного средства. В результате получилось, что сложность обучения для одного автокодировщика зависит от числа входных, скрытых и выходных нейронов, числа тренировочных наборов данных и числа эпох обучения. Эта зависимость отражается следующей формулой:

$$O\left(m*n*j*\left(i+k\right)\right), \tag{7}$$

где i – число входных нейронов, j – число скрытых нейронов, k – число выходных нейронов, n – число тренировочных наборов данных и m – число эпох. Число автокодировщиков (в данном случае их три) не повлияет на общую сложность программы, а значит, её можно записать с точностью до константы:

$$O\left(100*5000*8*\left(17+50\right)\right)=268*10^{6}. \tag{8}$$

Как можно увидеть из формулы (8), количество элементарных операций для выполнения реализованного программного средства, учитывая, что количество обучающих эпох равно 100, а количество тренировочных наборов данных равно 5000, едва ли достигает порядка миллиарда. Зная, что большинство современных компьютерных процессоров выполняет приблизительно миллиард операций в секунду, указанная асимптотическая сложность является незначительной помехой для запуска описанного программного средства на большинстве персональных компьютеров.

Для оценки качества предложенной глубокой нейронной сети были использованы такие показатели, как отклик (Rc), точность (Pr) и F-мера (F_1), которые описываются следующими формулами:

$$Rc=\frac{TP}{TP+FN}, \tag{9}$$

$$Pr=\frac{TP}{TP+FP}, \tag{10}$$

$$F_1=\frac{2*Rc*Pr}{Rc+Pr}. \tag{11}$$

В приведённых выше формулах TP – число правильно классифицированных приложений (истинно-положительные результаты, когда экспертная оценка совпадает с результатом работы нейронной сети), FP – число неправильно классифицированных приложений (ложно-положительные результаты, когда экспертная оценка негативная, а результат работы нейронной сети показал положительное значение) и FN – число ложно-отрицательных решений (когда экспертная оценка положительная, однако нейронная сеть выдаёт ложное значение). Исследования оценок точности нейронных сетей и их сравнительная характеристика представлены в работах [19–21]. Обобщённые результаты оценки качества предлагаемой нейронной сети приведены в табл. 2.

Таблица 2

Результаты оценки качества классификации для VPN-траффика

Класс сетевого траффика	Rc	Pr	F_1
VPN-аудио/видеопоток	0.99	0.99	0.99
VPN-торрент	0.97	0.99	0.98
VPN-VoIP	1.00	0.99	0.99
VPN-чат	0.94	0.95	0.94
FTP	0.99	0.98	0.99
SMTP	0.93	0.97	0.95
Среднее значение	0.95	0.95	0.95

В табл. 3 можно увидеть результаты классификации приложений.

По содержимому табл. 2 и 3 можно сделать вывод, что качество распознавания трафика глубокой нейронной сетью достаточно высокое. Кроме того, реализация предложенного подхода имеет достаточно лояльную асимптотическую сложность, позволяющую снижать требования к аппаратным ресурсам.

Таблица 3

Результаты классификации приложений

Название приложения	Rc	Pr	F_1
Youtube	0.98	0.99	0.97
BitTorrent	0.97	0.94	0.98
Skype	1.00	0.92	0.99
Telegram	0.94	0.95	0.94
FTPS	0.99	0.93	0.99
GMail	0.93	0.97	0.95
Среднее значение	0.92	0.92	0.92

Однако предложенное программное средство имеет и свои недостатки. В результате проведённых экспериментов можно с уверенностью сказать, что разработанная нейронная сеть имеет затруднения с распознаванием траффика сетей Tor. Результаты замеров производительности для Tor-траффика можно увидеть в табл. 4.

Таблица 4

Результаты оценки качества классификации для Tor-траффика

Класс сетевого траффика	Rc	Pr	F_1
Tor:Google	0.74	0.46	0.64
Tor:Telegram	0.32	0.14	0.45
Tor:Youtube	0.58	0.17	0.82
Tor:Twitter	0.03	0.09	0.13
Tor:Vimeo	0.14	0.27	0.55
Среднее значение	0.35	0.18	0.53

Заключение. В данной работе был предложен новый смешанный подход для анализа сетевого траффика, основанный на совокупном применении теории информации, математической статистики и теории машинного обучения.

На основании результатов проведённых экспериментов по тестированию реализованной нейронной сети и оценки качества её работы можно утверждать, что данная сеть эффективно распознает такие распространенные виды трафика, как VPN-аудио/видеопоток, VPN-торрент, VPN-VoIP, VPN-чат, FTP, SMTP. Для них среднее значение обобщенного показателя качества распознавания (F-мера) составило 0,95.

Так же, использование данного подхода при распознавании таких популярных приложений, как Youtube, BitTorrent, Skype,Telegram, FTPS, GMail, показало высокие результаты. Среднее значение F-меры для них составило 0,92.

Вместе с тем, следует отметить, что для распознавания Tor-траффика этот подход оказался малоэффективным. Среднее значение F-меры для него составило 0,53.

Таким образам, в целом предложенный подход к анализу зашифрованного сетевого трафика, представляется весьма перспективным.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Информационные системы и технологии / под ред. Тельнова Ю.Ф. – М.: Юнити, 2017. – 544 с.
2. Модель угроз и нарушителя безопасности персональных данных, обрабатываемых в специальных информационных системах персональных данных отрасли. Министерство связи и массовых коммуникаций Российской Федерации. – М., 2019. – Режим доступа: http://minsvyaz.ru/common/upload/publication/1410084of.pdf (дата обращения: 4.12.2020).
3. *Буковшин В.А., Болдырихин Н.В.* Сравнительное исследование технологий анализа интенсивности сетевого трафика // Тр. Северо-Кавказского филиала Московского технического университета связи и информатики. – Ростов-на-Дону, 2019. – С. 104-107.
4. *Татарникова Т.М.* Статистические методы исследования сетевого трафика // Информационно-управляющие системы. – 2018. – № 5 (96). – С. 35-43.
5. *Алтунин Ф.А., Кносаль В.М., Давыдов Р.В., Болдырихин Н.В.* Анализ методов классификации трафика // Тр. Северо-Кавказского филиала Московского технического университета связи и информатики. – Ростов-на-Дону, 2017. – С. 23-27.
6. *Callado A., Kamienski C., Szabo G., Gero B., Kelner J., Fernandes S., Sadok D.A.* Survey on Internet Traffic Identification // Communications Surveys & Tutorials, IEEE. – 3rd Quarter 2009. – Vol. 11, Issue 3. – P. 37-52.
7. *Круглов В.В., Борисов В.В.* Искусственные нейронные сети: теория и практика. – М.: НИЦ ИНФРА-М, 2017. – 283 с.
8. *Круглов В.В., Борисов В.В.* Нечёткая логика и искусственные нейронные сети. – М.: НИЦ ИНФРА-М, 2016. – 233 с.
9. *Рутковская Д.А., Пилиньский М.В., Рутковский Л.А.* Нейронные сети, генетические алгоритмы и нечёткие системы. – М.: ДМК Пресс, 2018. – 512 с.
10. *Стивен Норткат, Джуди Новак.* Обнаружение нарушений безопасности в сетях. – 3-е изд.: пер. с англ. – М.: Издательский дом «Вильямс», 2017. – 448 с.
11. *Медведовский И.Д., Семьянов П.В., Леонов Д.Г.* Атака на Internet. – М.: ДМК Пресс, 2017. – 332 с.
12. *Скудис Э.* Противостояние хакерам. – М.: ДМК Пресс, 2003. –506 с.
13. *Shiguo L.* One-way hash function based on neural network. Department of Automation, Nanjing University of Science & Technology. – 2017. – Режим доступа: https://arxiv.org/abs/0707.4032 (дата обращения: 4.12.2020).
14. *Moore A., Zuev D., and Crogan M.* Discriminators for use in flow-based classification. Department of Computer Science Research Reports. RR-05-13, 2019. – Режим доступа: https://www.cl.cam.ac.uk/~awm22/publication/moore2005discriminators.pdf (дата обращения: 4.12.2020).
15. *Alshammari R., Zincir-Heywood A.N.* Can encrypted traffic be identified without port numbers, ip addresses and payload inspection? // Computer networks. – 2011. – Vol. 55 (6). – P. 1326-1350.
16. *Hinton G. et al.* Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups // IEEE Signal Processing Magazine. – 2012. – Vol. 29 (6). – P. 82-97.
17. *Sun Q. et al.* Statistical identification of encrypted web browsing traffic, in Proc // Conference: Security and Privacy, 2002.
18. *Paninski L.* Estimation of entropy and mutual information // Neural Computation. – 2003. – Vol. 15. – P. 1191-1253.
19. *Gil G.D. et al.* Characterization of encrypted and VPN traffic using time-related features // Conference: The International Conference on Information Systems Security and Privacy. At: Italy. – Vol. 2016.
20. *Sasaki Y.* The truth of the F-measure // School of Computer Science, University of Manchester MIB, 2007. – Режим доступа: https://www.toyota-ti.ac.jp/Lab/Denshi/COIN/people/yutaka.sasaki/F-measure-YS-26Oct07.pdf (дата обращения: 4.12.2020).

21. *Tharwat A.* Classification assessment methods // Faculty of Computer science and engineering, Frankfurt university of applied science, 2018. – Режим доступа: https://www.emerald.com/insight/content/doi/10.1016/j.aci.2018.08.003/full/pdf?title=classification-assessment-methods (дата обращения: 4.12.2020).
22. *Derczynski L.* Complementarity, F-score and NLP evaluation // University of Sheffield, Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), 2016. – Режим доступа: https://www.aclweb.org/anthology/L16-1040.pdf (дата обращения: 4.12.2020).
23. *Шеннон К.* Работы по теории информации и кибернетике. – М.: Изд-во иностр. лит., 1963. – 829 с.

REFERENCES

1. Informatsionnye sistemy i tekhnologii [Information systems and Technologies], ed. by Tel'nova Yu.F. Moscow: Yuniti, 2017, 544 p.
2. Model' ugroz i narushitelya bezopasnosti personal'nykh dannykh, obrabatyvaemykh v spetsial'nykh informatsionnykh sistemakh personal'nykh dannykh otrasli. Ministerstvo svyazi i massovykh kommunikatsiy Rossiyskoy Federatsii [The model of threats and violators of the security of personal data processed in special information systems of personal data of the industry. Ministry of Communications and Mass Media of the Russian Federation]. Moscow, 2019. Available at: http://minsvyaz.ru/common/upload/publication/1410084of.pdf (accessed 4 December 2020).
3. *Bukovshin V.A., Boldyrikhin N.V.* Sravnitel'noe issledovanie tekhnologiy analiza intensivnosti setevogo trafika [Comparative study of network traffic intensity analysis technologies], *Tr. Severo-Kavkazskogo filiala Moskovskogo tekhnicheskogo universiteta svyazi i informatiki* [Proceedings of the North Caucasus Branch of the Moscow Technical University of Communications and Informatics]. Rostov-on-Don, 2019, pp. 104-107.
4. *Tatarnikova T.M.* Statisticheskie metody issledovaniya setevogo trafika [Statistical methods of network traffic research], *Informatsionno-upravlyayushchie sistemy* [Information and control systems], 2018, No. 5 (96), pp. 35-43.
5. *Altunin F.A., Knosal' V.M., Davydov R.V., Boldyrikhin N.V.* Analiz metodov klassifikatsii trafika [Analysis of traffic classification methods] *Tr. Severo-Kavkazskogo filiala Moskovskogo tekhnicheskogo universiteta svyazi i informatiki* [Proceedings of the North Caucasus Branch of the Moscow Technical University of Communications and Informatics]. Rostov-on-Don, 2017, pp. 23-27.
6. *Callado A., Kamienski C., Szabo G., Gero B., Kelner J., Fernandes S., Sadok D.A.* Survey on Internet Traffic Identificationб, *Communications Surveys & Tutorials, IEEE,* 3rd Quarter 2009, Vol. 11, Issue 3, pp. 37-52.
7. *Kruglov V.V., Borisov V.V.* Iskusstvennye neyronnye seti: teoriya i praktika [Artificial neural networks: theory and practice]. Moscow: NITS INFRA-M, 2017, 283 p.
8. *Kruglov V.V., Borisov V.V.* Nechyotkaya logika i iskusstvennye neyronnye seti [Fuzzy logic and artificial neural networks]. Moscow: NITS INFRA-M, 2016, 233 p.
9. *Rutkovskaya D.A., Pilin'skiy M.V., Rutkovskiy L.A.* Neyronnye seti, geneticheskie algoritmy i nechyotkie sistemy [Neural networks, genetic algorithms, and fuzzy systems]. Moscow: DMK Press, 2018, 512 p.
10. *Stiven Nortkat, Dzhudi Novak.* Obnaruzhenie narusheniy bezopasnosti v setyakh [Detection of security violations in networks]. 3rd ed.: transl. from engl. Moscow: Izdatel'skiy dom «Vil'yams», 2017, 448 p.
11. *Medvedovskiy I.D., Sem'yanov P.V., Leonov D.G.* Ataka na Internet [Attack on the Internet]. Moscow: DMK Press, 2017, 332 p.
12. *Skudis E.* Protivostoyanie khakeram [Opposition to hackers]. Moscow: DMK Press, 2003, 506 p.
13. *Shiguo L.* One-way hash function based on neural network. Department of Automation, Nanjing University of Science & Technology, 2017. Available at: https://arxiv.org/abs/0707.4032 (accessed 4 December 2020).
14. *Moore A., Zuev D., and Crogan M.* Discriminators for use in flow-based classification. Department of Computer Science Research Reports. RR-05-13, 2019. Available at: https://www.cl.cam.ac.uk/~awm22/publication/moore2005discriminators.pdf (accessed 4 December 2020).
15. *Alshammari R., Zincir-Heywood A.N.* Can encrypted traffic be identified without port numbers, ip addresses and payload inspection?, *Computer networks,* 2011, Vol. 55 (6), pp. 1326-1350.

16. *Hinton G. et al.* Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups, *IEEE Signal Processing Magazine*, 2012, Vol. 29 (6), pp. 82-97.
17. *Sun Q. et al.* Statistical identification of encrypted web browsing traffic, in Proc, *Conference: Security and Privacy, 2002*.
18. *Paninski L.* Estimation of entropy and mutual information, *Neural Computation*, 2003, Vol. 15, pp. 1191-1253.
19. *Gil G.D. et al.* Characterization of encrypted and VPN traffic using time-related features, *Conference: The International Conference on Information Systems Security and Privacy. At: Italy*, Vol. 2016.
20. *Sasaki Y.* The truth of the F-measure, *School of Computer Science, University of Manchester MIB, 2007.* Available at: https://www.toyota-ti.ac.jp/Lab/Denshi/COIN/people/yutaka.sasaki/F-measure-YS-26Oct07.pdf (accessed 4 December 2020).
21. *Tharwat A.* Classification assessment methods, *Faculty of Computer science and engineering, Frankfurt university of applied science, 2018*. Available at: https://www.emerald.com/ insight/content/doi/10.1016/j.aci.2018.08.003/full/pdf?title=classification-assessment-methods (accessed 4 December 2020).
22. *Derczynski L.* Complementarity, F-score and NLP evaluation, *University of Sheffield, Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), 2016*. Available at: https://www.aclweb.org/anthology/L16-1040.pdf (accessed 4 December 2020).
23. *Shennon K.* Raboty po teorii informatsii i kibernetike [Works on information theory and cybernetics]. Moscow: Izd-vo inostr. lit., 1963, 829 p.

Статью рекомендовал к опубликованию д.т.н. А.В. Елисеев.

Буковшин Вадим Александрович – Донской государственный технический университет; e-mail: vadimbukovshin199813@gmail.com; 344000, г. Ростов-на-Дону, пл. Гагарина, 1; тел.: +78632381518, +78632381516; кафедра кибербезопасности информационных систем; студент.

Чуб Павел Андреевич – e-mail: pavel.chub.1997@mail.ru; кафедра кибербезопасности информационных систем; студент.

Короченцев Денис Александрович – e-mail: mytelefon@mail.ru; кафедра кибербезопасности информационных систем; зав. кафедрой; к.т.н.

Черкесова Лариса Владимировна – e-mail: chia2002@inbox.ru; кафедра кибербезопасности информационных систем; д.ф.-м.н.; профессор.

Болдырихин Николай Вячеславович – e-mail: boldyrikhin@mail.ru; кафедра кибербезопасности информационных систем; к.т.н.; доцент.

Сафарьян Ольга Александровна – e-mail: safari_2006@mail.ru; кафедра кибербезопасности информационных систем; к.т.н.; доцент.

Bukovshin Vadim Aleksandrovich – Don State Technical University; e-mail: vadimbukovshin199813@gmail.com; 1, pl. Gagarin, Rostov-on-Don, 344000, Russia; phones: +78632381518, +78632381516; the department of cybersecurity of information systems; student.

Chub Pavel Andreevich – e-mail: pavel.chub.1997@mail.ru; the department of cybersecurity of information systems; student.

Korochentsev Denis Aleksandrovich – e-mail: mytelefon@mail.ru; the department of cybersecurity of information systems; head of the department; cand. of eng. sc.

Cherkesova Larisa Vladimirovna – e-mail: chia2002@inbox.ru; the department of cybersecurity of information systems; dr. of phys. and math.; professor.

Boldyrikhin Nikolay Vyacheslavovich – e-mail: boldyrikhin@mail.ru; the department of cybersecurity of information systems; cand. of eng. sc. associate professor.

Safaryan Olga Aleksandrovna – e-mail: safari_2006@mail.ru; the department of cybersecurity of information systems; cand. of eng. sc. associate professor.