Елена В. Карачанская[1], Надежда И. Соседова[2]
[1]*Дальневосточный государственный университет путей сообщения,*
ул. Серышева, 47, г. Хабаровск, 680027, Россия
e-mail: elena_chal@mail.ru, http://orcid.org/0000-0003-0815-3688
[2]*ООО «МАСКОМ-Техлайн»,*
ул. Яшина, 40, г. Хабаровск, 680027, Россия
e-mail: sosedowa.nadezhda@yandex.ru, http://orcid.org/0000-0002-4028-7304

МЕТОД ВЫЯВЛЕНИЯ АНОМАЛИЙ СЕТЕВОГО ТРАФИКА, ОСНОВАННЫЙ НА ЕГО САМОПОДОБНОЙ СТРУКТУРЕ

DOI: http://dx.doi.org/10.26583/bit.2019.1.10

Аннотация. Целью данной статьи является представление метода выявления аномалий сетевого трафика, основанного на утверждении о том, что трафик является фракталом. Предположено, что сетевой трафик является самоподобной структурой и моделируется фрактальным броуновским движением. В качестве инструментов при разработке данного метода были применены фрактальный анализ и математическая статистика. Проведен анализ существующих методов выявления сетевых аномалий на предмет их недостатков. Результатом работы является модифицированный метод выявления аномалий сетевого трафика. Данный метод относится к полуконтролируемой методике обнаружения аномалий, что позволяет процессу быть практически автономным от человеческого вмешательства. Кроме того, метод можно отнести к группе статистических методов, что делает его достаточно простым в реализации. В отличие от существующих представленный метод использует выборки оптимальных объемов (т.е. выборки за время, которое является одновременно достаточно малым, чтобы среагировать на аномалию вовремя, но при этом время, которое позволит получить такие выборки, которые будут достаточны для расчета параметра и снизят число ложных срабатываний), полученные за минимальное, но при этом достаточное время. Данный алгоритм выявления аномалий состоит из двух частей: расчета образцов (эталонных значений) и сравнения получаемого трафика с эталоном (анализ аномалий сетевого трафика). Расчет эталонов опирается на вычисление значений параметра самоподобия (параметра Хёрста) для некоторых показателей из заголовков пакетов. Алгоритм поиска аномалий, лежащий в основе метода, может применяться как для поиска входящих аномалий (сетевых атак), так и для поиска аномалий в исходящем трафике (DLP-системы).

Ключевые слова: сетевой трафик, фрактал, аномалии сетевого трафика, параметр Хёрста, самоподобие.

Для цитирования: КАРАЧАНСКАЯ, Елена В.; СОСЕДОВА, Надежда И. МЕТОД ВЫЯВЛЕНИЯ АНОМАЛИЙ СЕТЕВОГО ТРАФИКА, ОСНОВАННЫЙ НА ЕГО САМОПОДОБНОЙ СТРУКТУРЕ. Безопасность информационных технологий, [S.l.], p. 98-110, 2019. ISSN 2074-7136. Доступно на: <https://bit.mephi.ru/index.php/bit/article/view/1185>. Дата доступа: 19 feb. 2019. doi:http://dx.doi.org/10.26583/bit.2019.1.10.

Elena V. Karachanskaya[1], Nadezhda Iv. Sosedova[2]
[1]*Far-Easten State Transport University,*
Serysheva st., 47, Khabarovsk, 680027, Russia
e-mail: elena_chal2@mail.ru, http://orcid.org/0000-0003-0815-3688
[2]*«MASCOM-Techline» LLC,*
Yashina st., 40, Khabarovsk, 680027, Russia
e-mail: sosedowa.nadezhda@yandex.ru, http://orcid.org/0000-0002-4028-7304

Method for detection of network traffic anomalies which is based on its self-similar traffic structure

DOI: http://dx.doi.org/10.26583/bit.2019.1.10

Abstract. The paper presents a method for detecting network traffic anomalies taking into account its self-similar structure. It is assumed that network traffic is a self-similar structure and is modeled by fractal Brownian motion. Existing methods of detecting network anomalies are studied. The result of scientific work is a new method for detecting network traffic anomalies. This method is based on a semi-controlled method of anomaly detection, which allows the process to be almost autonomous from human interven-

tion. In addition, the method can be classified as a group of statistical methods, which makes it quite easy to implement. In contrast to the existing methods, this method uses samples of optimal volumes obtained in the minimum but sufficient time. This anomaly detection algorithm consists of two parts: calculation of samples (reference values) and comparison of the received traffic with the standard (analysis of network traffic anomalies). The calculation of standards is based on the calculation of the values of the self-similarity parameter (Hurst parameter) for some indicators from the package headers. The algorithm of anomaly search underlying the method can be used both to search for incoming anomalies (network attacks) and to search for anomalies in outgoing traffic (DLP-system).
Keywords: network traffic, fractal, network traffic anomalies, Hurst parameter, self-similar.
For citation: KARACHANSKAYA, Elena V.; SOSEDOVA, Nadezhda Iv. Method for detection of network traffic anomalies which is based on its self-similar traffic structure. IT Security (Russia), [S.l.], p. 98-110, 2019. ISSN 2074-7136. Available at: <https://bit.mephi.ru/index.php/bit/article/view/1185>. Date accessed: 19 feb. 2019. doi:http://dx.doi.org/10.26583/bit.2019.1.10.

Введение

Обнаружение аномалий сетевого трафика является одной из самых актуальных проблем в области защиты информации. Актуальность объясняется тем, что, согласно данным Лаборатории Касперского, почти 19,59 % всех компьютеров пользователей интернета в мире хоть раз подвергались веб-атаке [1]. Это объем только зафиксированных атак, малая их часть, что настораживает и вызывает определенное беспокойство у специалистов в области защиты информации. Для выявления и оперативного устранения таких атак создаются огромные аналитические центры. Проблема выявления аномалий обусловлена трудностью правильного выбора применяемого математического алгоритма. Кроме того, большинство высокоточных алгоритмов являются достаточно трудными в программной реализации.

Целью данного исследования является разработка (модификация существующего) метода, позволяющего производить выявление аномалий, как входящего, так и исходящего сетевого трафика различного рода с высокой степенью точности и низким уровнем ложных срабатываний. Данный метод построен на применении теории случайных процессов и учитывает фрактальную структуру сетевого трафика.

1. Самоподобная структура сетевого трафика

Под сетевым трафиком будем понимать объём информации, переданный по сети за определённое время. Многочисленные исследования показывают, что свойства сетевого трафика значительно отличаются от свойств трафика голосовых систем (трафика телефонных сетей). Сетевой трафик обладает свойством самоподобия, то есть процессы, протекающие в сетях передачи данных, имеют фрактальные свойства, которые легли в основу данного исследования.

Фрактал – множество, обладающее свойством самоподобия (объект, в точности или приближённо совпадающий с частью себя самого, то есть целое имеет ту же форму, что и одна или более частей). Применительно к сетевому трафику, фрактальность означает практически неизменность распределения трафика в течении времени при масштабировании временной шкалы.

Впервые о самоподобном трафике заговорили еще в 1993 году Leland, Taqqu, Willinger и Wilson [2] проводили исследования Ethernet-трафика в сети корпорации Bellcore и пришли к выводу, что на больших интервалах он обладает свойством самоподобия, то есть выглядит качественно одинаково при любых масштабах временной оси. Если из графика, приведенного на рис. 1 [3], выделим часть графика за небольшой промежуток времени (например, с 30 по 60 секунду), то получим приблизительно такой же по внешнему виду график (рис. 2). Наличие данного свойства у сетевого трафика означает, что простые модели, использующие пуассоновское распределение, неточны, и сети, построенные без учета самоподобия, могут функционировать в непредсказуемых режимах.

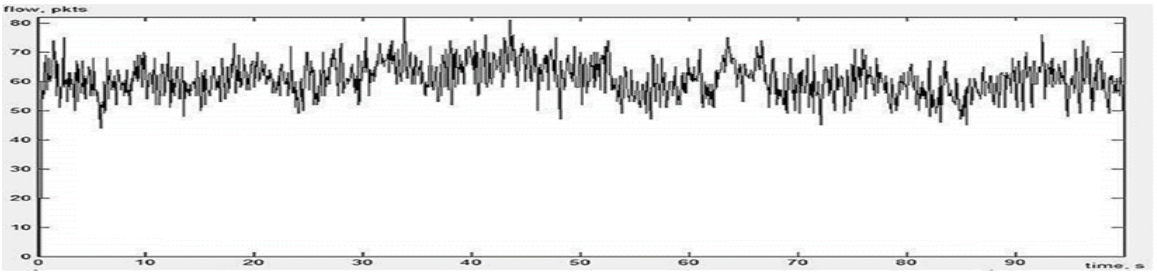

Рис. 1. Распределение сетевого трафика по временной шкале [3]
(Fig. 1. The distribution of network traffic on the time scale [3])

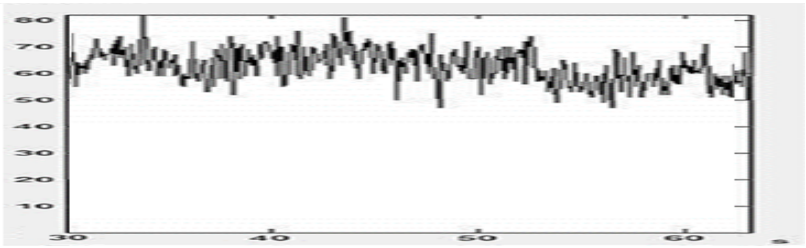

Рис. 2. Распределение сетевого трафика по временной шкале (30-60 секунды)
(Fig. 2. The distribution of network traffic on the time scale (30-60 seconds))

За длительное время наблюдения за поведением сетевого трафика, было создано большое число его математических моделей [4, 5], но, к сожалению, ни одна из них не является исчерпывающей.

Одна из наиболее корректных моделей – модель фрактального броуновского движения. Фрактальное броуновское движение было предложено как модель для «свободного трафика», агрегированного от множества независимых источников [6]. Модель привлекательна за счет простоты анализа и за счет того, что она характеризуется долгосрочной зависимостью.

Фрактальное броуновское движение (ФБД) – это случайный процесс, который обладает некой «памятью» в отличие от классического броуновского движения [7].

Случайный процесс *X(t)* называется фрактальным броуновским движением с параметром *H*, если:

1) $X\left(0\right)=0$ и его реализации $x\left(t\right)$почти всегда непрерывны;

2) $\Delta X=X\left(t_2\right)-X\left(t_1\right)$ имеет нормальное распределение с нулевым математическим ожиданием и дисперсией [6] –

$$\delta^2\cdot\left(t_2-t_1\right)^{2H}, \text{ где } 0\leq H\leq 1 \qquad (1)$$

Поскольку модель фрактального броуновского движения является одной из современных и лучше всего описывающих все свойства, которые присущи именно сетевому трафику, то она и была взята в качестве основной для разработки метода выявления аномалий.

2. Понятие аномальности сетевого трафика

Сетевой трафик можно рассматривать как случайный процесс, реализации которого обладают характеристиками, изменяющимися во времени, но остающимися детерминированными (например, математическое ожидание и дисперсия). Эти характеристики будем считать эталонными. Пусть $f(t)$ - фактическая характеристика реализации трафика в момент времени *t*. Аномальное состояние в сетевом трафике – состояние, при котором значение $f(t)$ в любой момент времени *t* отличается от

эталонного. Тогда аномалиями сетевого трафика будем называть любые отклонения показателей сети от заранее зафиксированных в качестве эталонных.

На рис. 3. приведена классификация аномалий сетевого трафика [8]. Кроме этого, в зависимости от количества аномалии можно разделить на три основных типа: точечные, групповые и аномалии в контексте. Предлагаемый метод выявления аномалий может выявлять все три типа аномалий.

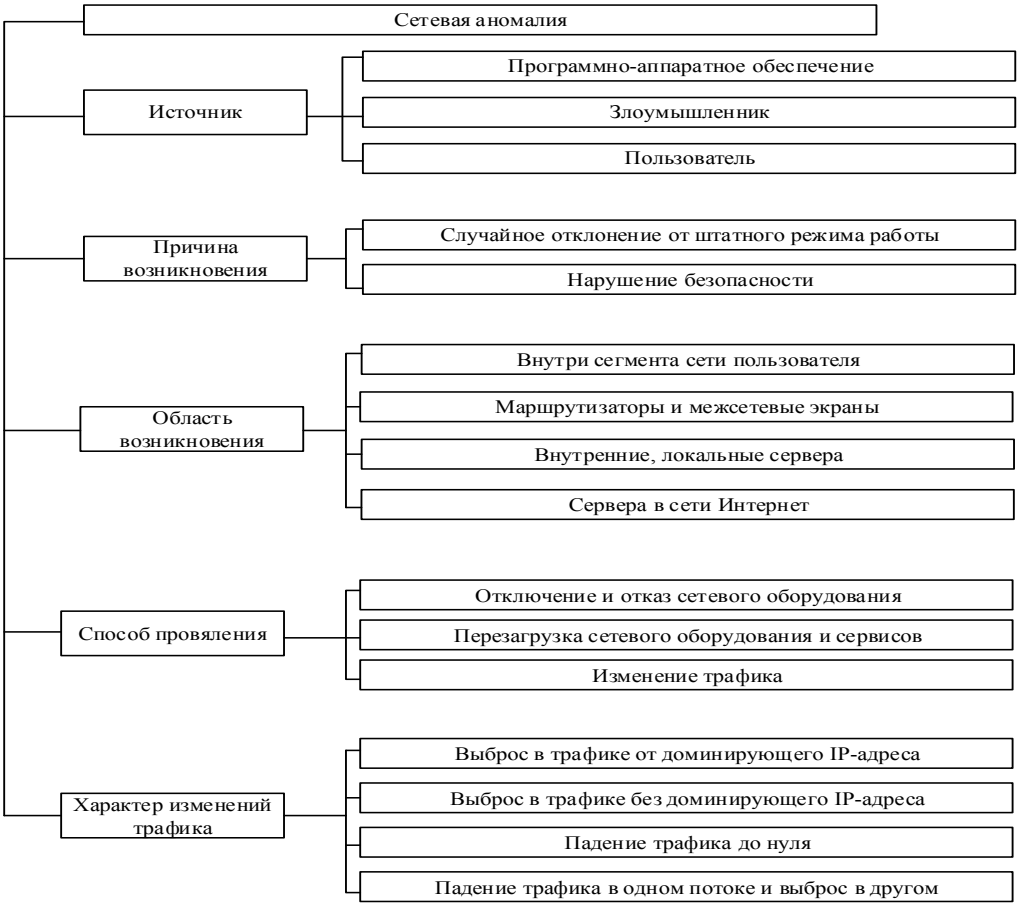

Рис. 3. Классификация аномалий [8].
(Fig. 3. Classification of anomalies [8])

Предлагаемый метод не позволяет определять категорию отслеживаемой аномалии сетевого трафика, поскольку для поставленных целей данная информация является избыточной и значительно сужает спектр областей его применения.

В табл. 1 проведено сравнение существующих современных методов выявления аномалий сетевого трафика. Принцип, по которому проведено сравнение методов, частично заимствовано из источника [9], поскольку данный сравнительный анализ проведен Унтеровым С. достаточно глубоко и совпадает с мнением авторов данной статьи. Представлены методы, относящиеся к статистическим, и методы, основанные на использовании фрактальных свойств сетевого трафика, проанализированы их характеристики для выявления слабых мест существующих алгоритмов с целью последующего их устранения. Приведенные методы являются достаточно близкими по своей структуре, и сравнение существующих методов выявления аномалий сетевого трафика является достаточно объемным, поэтому приведем положения по важному параметру, а именно – времени сбора информации. Как видно из табл. 1, параметр «требуемое время наблюдения» отличается от метода к методу. Данные, собранные в течение трех лет, нельзя считать корректными и оптимальными для различных систем, а данные, собираемые в течение 60 секунд, могут привести к огромному числу ложных срабатываний, поскольку выборки являются очень малыми для применения статистических методов.

Таблица 1. Сравнение методов выявления аномалий сетевого трафика

Характеристика	Naila Belhadj Aissaa, Mohamed Guerroumia [10]	Mazurek M., Dymora P. [11]	Унтеров Д. С. [9]
Простота расчетов	Нет	Да	Усложняется аппаратной реализацией метода
Скорость работы алгоритма	Низкая	Средняя	Средняя
Требуемое время наблюдения	Основывается на данных, собранных в течение 3 лет	Наблюдение в течение суток	60 секунд
Описан алгоритм действий при обнаружении аномалии	Нет	Нет	Да
Спектр анализируемого типа трафика	Широкий	Только http	Широкий
Метод расчета параметра самоподобия	–	–	RS-анализ

3. Вспомогательные положения для описания метода

В рамках данного исследования предлагается метод, в основе которого лежит алгоритм, построенный на свойствах самоподобия сетевого трафика. Метод можно отнести к группе статистических методов, поскольку он использует оценку ряда статистических параметров. Метод основан на полуконтролируемой технике выявления аномалий.

Данный метод может быть реализован в виде одного из модулей портативного программного обеспечения.

Работу разработанного метода выявления аномалий сетевого трафика можно разделить на несколько шагов:

1) запуск программы;
2) перехват трафика;
3) обработка трафика разработанным алгоритмом;
4) реакция на аномалию.

Для оценки аномальности предложено использовать не полезное содержимое пакетов, а взять за основу предположение Унтерова Д. С. [9] о том, что информации из заголовков пакетов (это определённым образом оформленный блок данных, передаваемый по сети на сетевом уровне) будет достаточно. В качестве такой информации будем использовать количественные значения передаваемых флагов (меток, указывающих на тип пакета).

4. Анализ трафика на аномальность

Рассмотрим более подробно алгоритм анализа трафика на аномальность. Данный алгоритм состоит из двух частей: расчет эталонных значений (обучение) и обнаружение аномалий. Для расчета будем опираться на данные транспортного (рис. 4) и сетевого уровней модели OSI, а именно на количество определенных флагов. Таким образом, единицей наблюдения будем считать количественное значение данных определенного типа (TCP-SYN, UDP, ICMP и так далее) за 1 секунду.

Данные, расчет которых производиться на этапе расчета эталонных значений, будут сравниваться с данными, которые рассчитываются в реальном времени на этапе обнаружения. Этап расчета эталонов является конечной процедурой, и его длительность составляет 4 минуты. Необходимое время для расчета эталонов вычислялось эмпирическим путем, причем за основу был взят опыт предшествующих методов. Временной интервал в 120 секунд позволяет уже достаточно точно определить наличие/отсутствие аномалий в сетевом трафике, данное значение было удвоено, взято с запасом точности. Важно понимать, что данные цифры актуальны для пользователей глобальной сети со средней и высокой активностью, в случае низкой активности данные параметры могут быть пересмотрены.

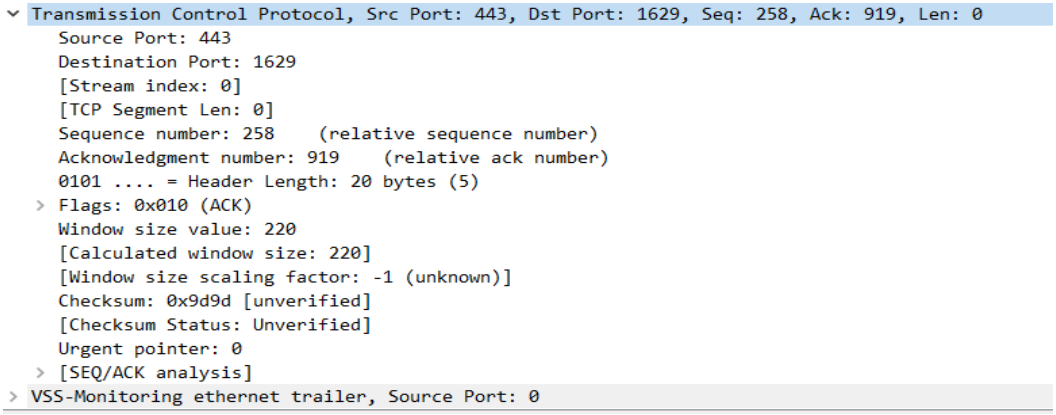

Рис. 4. Дамп кадра. Транспортный уровень
(Fig. 4. Frame dump. Transport level)

5. Оценка аномальности

Для оценки аномальности в методе используется параметр Хёрста [12], для расчета которого применяется RS-анализ (метод нормированного размаха). RS-анализ позволяет различить случайный и неслучайный временные ряды, а также делать выводы о наличии непериодических циклов и долговременной памяти [13].

Как известно, основной формулой RS-анализа является отношение:

$$R / S = \left(a \cdot N\right)^H, \tag{2}$$

где H – показатель Хёрста; S – среднеквадратичное отклонение ряда наблюдений; R – размах накопленного отклонения; N – дискретное время (объем выборки); a – заданная положительная константа.

Размах накопленного отклонения R является наиболее важным элементом формулы расчета показателя Хёрста. В общем виде R вычисляют следующим способом:

$$R = \max_{1 \le U \le N}(Z_u) - \min_{1 \le U \le N}(Z_u), \tag{3}$$

где Z_u – накопленное отклонение ряда от среднего X_{cp}.

Г.Э. Хёрст эмпирически рассчитал константу a, для сравнительно с краткосрочными временными рядами природных явлений $a = 0,5$. Таким образом, формула (2) примет вид:

$$R / S = \left(N / 2\right)^H. \tag{4}$$

Нахождение значения параметра H можно разбить на следующие этапы [13]:

1) исходный временной ряд разбиваем на блоки одинаковой длины;
2) для каждого блока вычисляем размах R;
3) для каждого блока вычисляем среднеквадратичное отклонение:

$$S = \sqrt{N^{-1}} \cdot \sum_{i=1}^{N}\left(X_i - X\right)^2; \tag{5}$$

4) принимаем за N объем выборки (дискретное время);
5) логарифмируем полученное выражение;
6) получаем искомый параметр:

$$H = \log\left(R/S\right) \Big/ \log\left(N/2\right). \tag{6}$$

Заведомо зная длительность этапа расчета эталонов – 4 минуты, можно для расчета использовать следующую схему разбиения на интервалы:

1) раз в 5 секунд;
2) раз в 15 секунд;

3) раз в 60 секунд;

4) раз в 120 секунд.

Схема «раз в *n*» секунд означает, что весь временной отрезок, равный четырем минутам (240 секундам), разбивается на интервалы по *n* секунд, и в каждом интервале производится расчет параметров Хёрста. Таким образом, для позиции «раз в 5 секунд» получаем сорок восемь интервалов, рассчитываем для каждого их них параметр Хёрста и среднее значение этих параметров. Аналогично для позиции «раз в 15 секунд» получаем шестнадцать интервалов, для позиции «раз в 60 секунд» - четыре интервала, для позиции «120 секунд» - два интервала.

Необходимость расчета параметров для четырех различных интервалов обусловлена тем, что в случае малого числа пакетов, которые могут быть получены на небольших временных отрезках, возможна ложная реакция, но, однако, не стоит игнорировать данные на этих числовых промежутках, так как при появлении аномалии скорость ее обнаружения является критичной.

После расчета эталонов, во время непосредственного поиска аномалий проводится сравнение с эталоном для трафика, полученного за 5 секунд, за 15 секунд, за 60 секунд и за 120 секунд.

Расчет эталонов основан на расчете параметра Хёрста для некоторых показателей из заголовков пакетов. Выводы о нормальности или аномальности трафика делаются исходя из удаленности получаемых фактических значений от эталонных. Несмотря на то, что данный способ дает лишь приближенное значение показателя Хёрста, решающим фактором стала простота расчетов. Можно предположить, что в дальнейшем развитии метода возможно уточнение способов расчета параметра Хёрста на более точный.

После получения значений записываются в файл, и для каждой группы по интервалам производится расчет среднего значения, а также стандартного отклонения S. Опираясь на эти данные, система построит интервал в соответствии с общеизвестным правилом 3σ (правило «трех сигм») [14]. В соответствии с этим правилом вероятность того, что нормально распределенная случайная величина отклонится от своего математического ожидания на большую величину, чем утроенное среднее квадратичное отклонение, которое составляет 0,9973 [14]:

$$P\left(\left|X-a\right|<3\sigma\right)=2\Phi\left(3\right)=2\cdot 0,49865=0,9973. \qquad (7)$$

Другими словами, вероятность того, что абсолютная величина отклонения превысит утроенное среднее квадратическое отклонение, очень мала. При большом числе наблюдений, когда среднее значение всей выборки неизвестно, правило «трех сигм» преобразуется в правило «трех *S*» при достаточно больших значениях выборки. И тогда получаем, что решение об аномальности принимается в зависимости от попадания, рассчитанного на этапе обнаружения значения в тот или иной диапазон в интервале, определяемом S.

Необходимо отметить, что метод будет работать только в случае, если на этапе расчета эталонов распределение трафика будет подчинено нормальному распределению.

Согласно правилу «трех сигм» известно, что почти 96 % случайных величин попадут в промежуток $(-2\sigma;\ 2\sigma)$, поэтому будем проверять попадание параметров Хёрста в интервал «два сигма». Поскольку параметр Хёрста является величиной положительной, то может быть рассмотрен только промежуток $(0;\ 2\sigma)$, что увеличит скорость расчета. В случае если значение параметра для анализируемого трафика выйдет за границы этого диапазона, система предупредит о возможной аномалии, и дополнительно будет проведена проверка по расширенному диапазону $(0;\ 3\sigma)$. Следует отметить, что правило 3σ имело место в работе Унтерова [9], однако, в данном случае предлагается использовать диапазон именно «два сигма», а не «один сигма», поскольку это позволяет выбрать диапазон с большей точностью. В случае, если и в этот промежуток параметр анализируемого трафика не попадет, система сообщит об аномалии.

6. Программная реализация метода

Для оценки работоспособности метода было разработано консольное приложение на языке Python под управлением ОС Kali Linux. Данный язык был выбран по причине его простоты и высокой функциональности для сбора и обработки необходимой информации. Данные для расчета эталонов и непосредственного анализа собирались при помощи Tshark –консольной утилиты, входящей в состав Wireshark, предназначенной для захвата и анализа сетевых пакетов [15]. С помощью этой утилиты можно выполнять захват пакетов в сети, сохранять пакеты в файле, просматривать ранее сохраненные файлы с информацией о сетевых пакетах.

Обучение написанной программы (расчет эталонов для написанной программы) выполняется в соответствии с разработанным алгоритмом (рис. 5).

На этапе обучения необходимыми к расчету данными является размах ряда R и среднее квадратическое отклонение S, а остальные данные рассчитываются в качестве вспомогательных.

Программа, производящая тестирование предложенного метода, работает на основе блок-схемы, представленной на рис. 6. Данный программный продукт является свободным к распространению и может быть использован для дальнейших исследований.

7. Тестирование метода

Целью тестирования является проверка работоспособности разрабатываемого метода, выявление его недостатков и достоинств.

Для тестирования метода запустим программу. Первый этап – обучение. Для генерации трафика произведем имитацию обычной для стандартного пользователя деятельности: запуск видео на видеохостинге, вход в почтовый ящик и выполнение некоторого количества поисковых запросов (рис. 7).

Значения полученных параметров Хёрста не округляются для того, чтобы при сравнении эталонных значений и текущего трафика можно было проследить разницу более отчетливо. Возможность настолько точных расчетов предоставляет язык программирования, на котором написана тест-программа.

После завершения расчета эталонов проводится тестирование обнаружения аномалий. Для этого продолжим ту же деятельность, которая проводилась на этапе обнаружения на протяжении четырех минут (продолжим генерировать нормальный трафик), не забывая, что на тестируемое устройство сторонний аномальный трафик не поступает. На протяжении данного времени сообщений об аномалии выдано не было, так как данный трафик система после обучения считает нормальной (рис. 8).

Для создания условий, близких к реальным аномалиям и сетевым атакам, была проведена симуляция ряда аномалий. Самый простой способ такой проверки – симуляция атаки SYN-flood. Данная атака была реализована при помощи генератора пакетов hping3 [16] (рис. 9).

По результатам работы тест-программы можно отметить, что метод является достаточно точным. Он легко распознает флуд-атаки на систему (рис. 10). Но на этапе в 5 секунд выдает ложное предупреждение о возможных аномалиях, это обусловлено тем, что 5 секунд – слишком малый промежуток времени, и решения о возможных аномалиях не могут быть приняты, основываясь на нем.

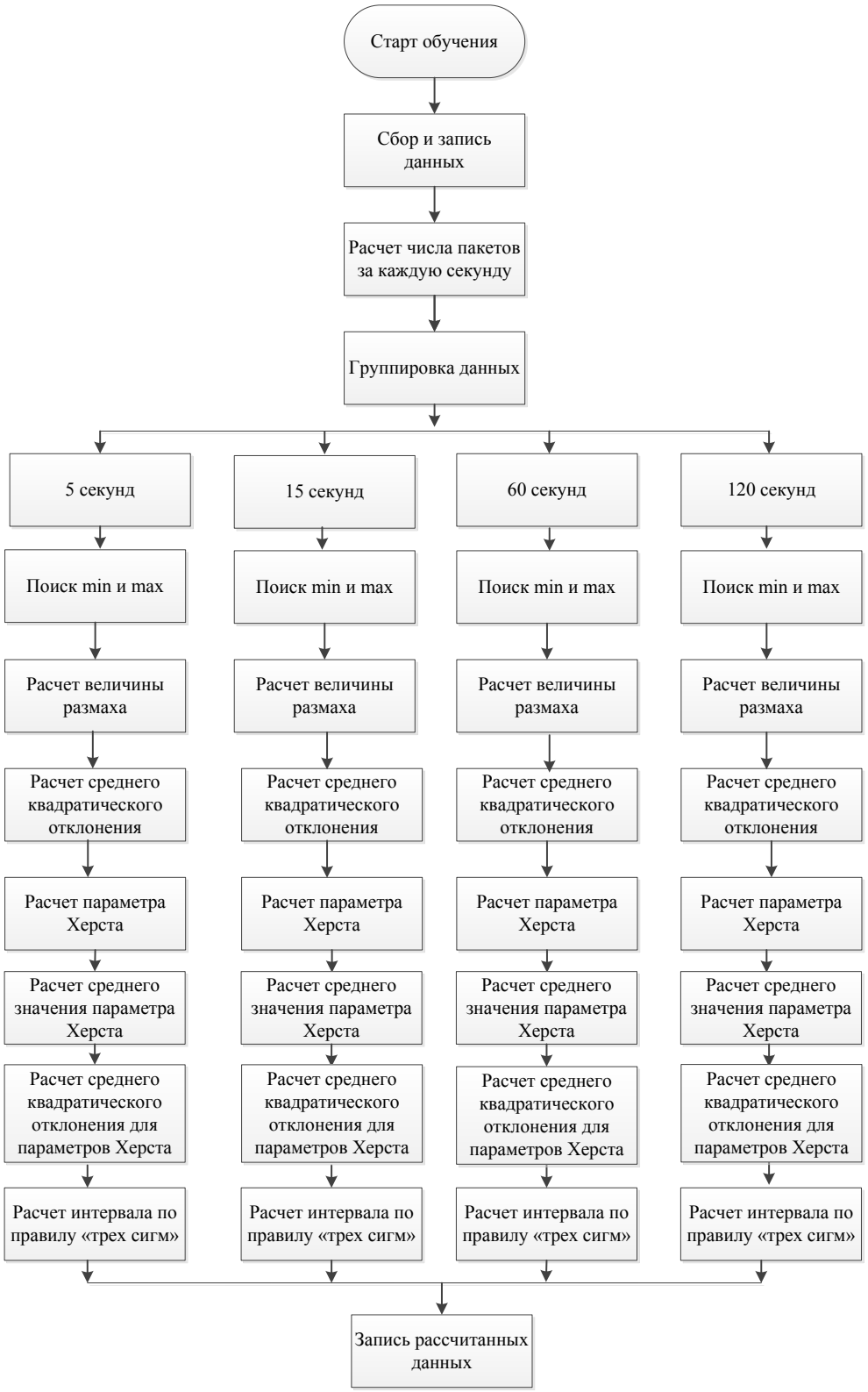

Рис. 5. Алгоритм метода расчета эталонов
(Fig. 5. The algorithm of the method of calculation of standards)

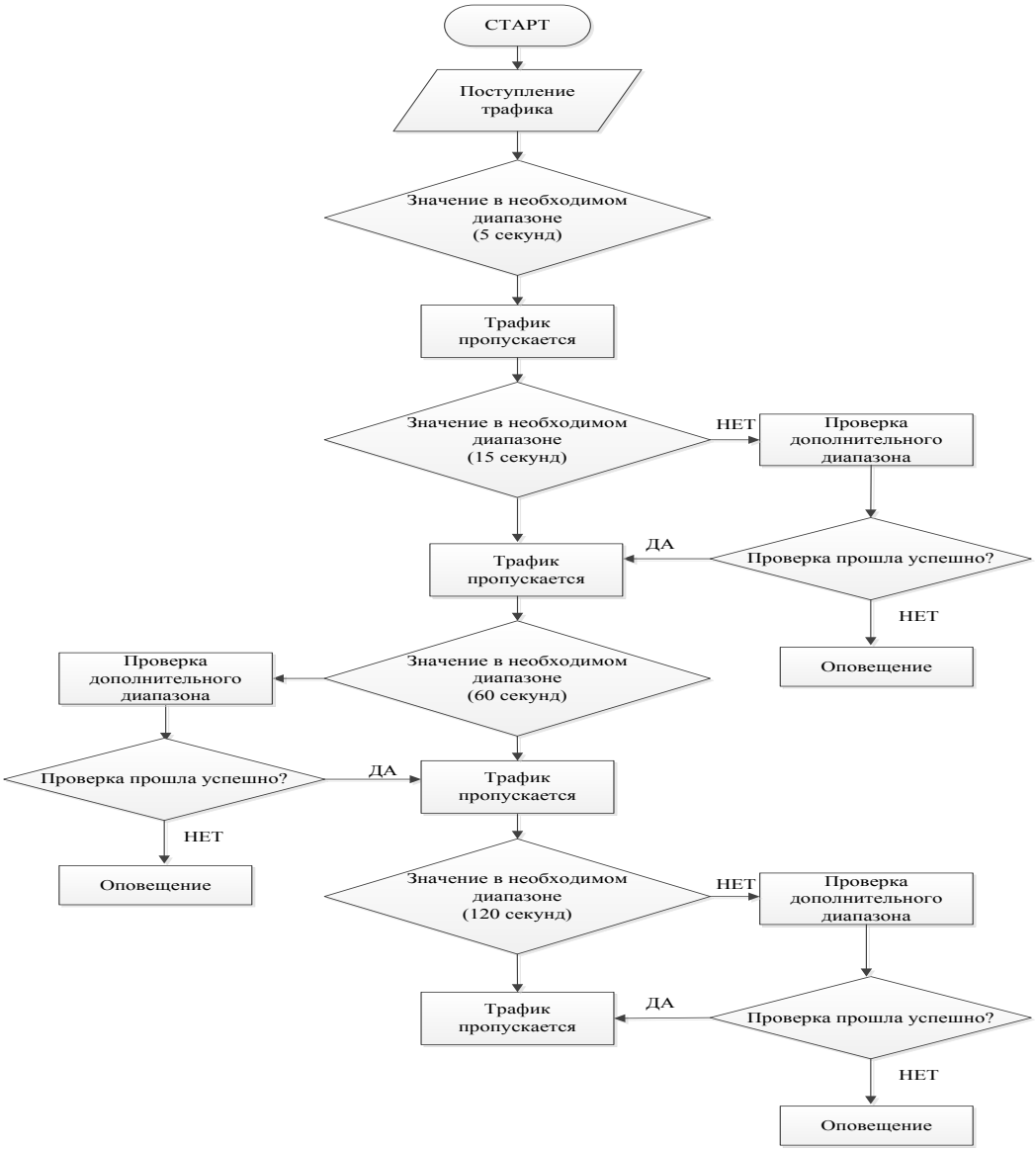

Рис. 6. Блок-схема «Обнаружение аномалий»
(Fig 6. A block diagram of the «Anomaly Detection»)

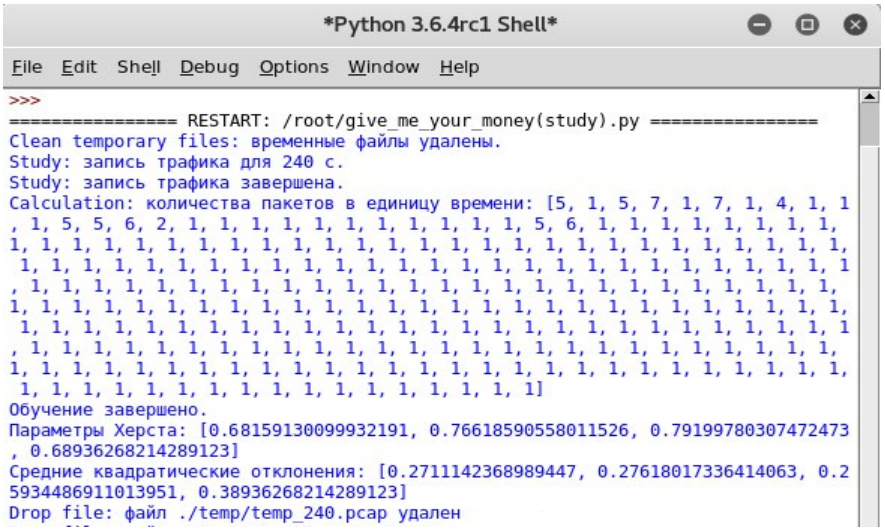

Рис. 7. Расчет данных на этапе обучения
(Fig. 7. Data calculation at the training stage)

```
Аномалий не обнаружено. Промежуток 15  с.
Drop file: файл ./temp/temp_15.pcap удален
Аномалий не обнаружено. Промежуток 60  с.
Drop file: файл ./temp/temp_60.pcap удален
Аномалий не обнаружено. Промежуток 120  с.
Drop file: файл ./temp/temp_120.pcap удален
```

Рис. 8. Реакция на нормальный трафик
(Fig. 8. Response to the normal traffic)

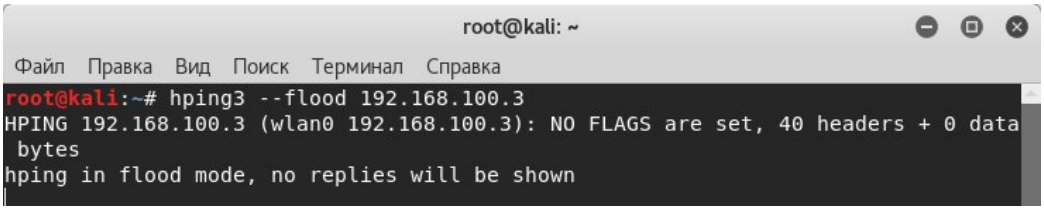

Рис. 9. Генерация аномального трафика
(Fig. 9. Simulation of the abnormal traffic)

```
ВОЗМОЖНА АНОМАЛИЯ.
 Промежуток 15  с. Параметр Херста:  1.3610616091280565
ОБНАРУЖЕНА АНОМАЛИЯ.
Drop file: файл ./temp/temp_15.pcap удален
ВОЗМОЖНА АНОМАЛИЯ.
 Промежуток 60  с. Параметр Херста:  0.9063063796773916
ОБНАРУЖЕНА АНОМАЛИЯ.
```

Рис. 10. Сообщение системы о выявленных аномалиях
(Fig. 10. The system message about the identified anomalies)

Заключение

По результатам проведенного тестирования можно сделать вывод, что метод работает исправно и может быть использован для обнаружения аномалий. Для повышения его эффективности рекомендуется проводить расчеты по нескольким флагам одновременно, а также использовать в комплексе с уже существующими методами.

Разработанный метод и программа для его тестирования являются бесплатными и находятся в открытом доступе, что может позволить убедиться в их работоспособности, а также взять за основу для новых исследований.

Данный метод может быть использован как модуль системы обнаружения вторжений, а также в DLP-системах, поскольку не ограничивает в выборе направления трафика, то есть может быть исследован как входящий, так и исходящий трафики. Полученные результаты могут быть использованы в процессе разработки программного или программно-аппаратного средства обнаружения вторжений.

Несмотря на то, что уже существуют методы, также использующие положения о самоподобности трафика, данное исследование обладает актуальностью и новизной, поскольку:

1. эмпирическим путем подобрано оптимальное время для обучения, т.е. время которое с одной стороны мало и позволяет оперативно обратить внимание на атаку, а с другой стороны снизит процент ложных срабатываний, которые могут возникать из-за слишком маленьких выборок;
2. метод учитывает несколько интервалов для сравнения – (0; 3σ) и (0; 2σ): в случае, если аномалий нет, то проверка промежутка (0; 2σ) ускоряет процесс анализа трафика, а в случае возникновения аномалий на этом промежутке проводится дополнительная проверка промежутка (0; 3σ), что помогает избежать ложных срабатываний;
3. метод учитывает степень активности пользователей, так как указано, что промежутки для малоактивных пользователей должны пересматриваться (является масштабируемым);
4. метод позволяет обнаруживать сетевые аномалии различного характера (аномалии каналов связи, сетевые атаки и повышение активности пользователей);
5. метод позволяет обнаруживать сетевые аномалии точечного типа.

СПИСОК ЛИТЕРАТУРЫ:

1. Чебышев Виктор, Синицын Федор, Паринов Денис, Лискин Александр, Купреев Олег. Развитие информационных угроз во втором квартале 2018 года. Статистика. URL: https://securelist.ru/it-threat-evolution-q2-2018-statistics/90919/ (дата обращения: 20.02.2019).
2. Leland W.E., Taqqu M.S., Willnger W., Wilson D.V. On the self-similar nature of Ethernet traffic. ACM Transaction on Networking. 1994. Vol. 2, no. 1. P. 1 – 15.
3. Моргайлов Д.Д., Ладыженский Ю.В., Юнис М. Моделирование самоподобного входного трафика сетевых процессоров в системе NS-2. Информатика и компьютерные технологии – 2012 (ИКТ – 2012) / Материалы VIII международной научно-технической конференции студентов, аспирантов и молодых ученых – 18-19 сентября 2012 – Донецк, ДонНТУ – 2012. С. 232 – 239. URL: http://uran.donntu.org/~masters/2013/fknt/morgajlov/library/selfsim.htm (дата обращения: 20.02.2019) .
4. Поршнев С.В. Математические модели информационных потоков в высокоскоростных магистральных интернет-каналах. М.: Горячая линия-Телеком, 2016. 232 с.
5. Сидорова О.И. Пуассоновская модель трафика с бесконечным числом неоднородных источников. Вестник ТвГУ. Серия: Прикладная математика. 2015. № 1. С. 47 – 66.
6. Осин А.В. Фрактальное движение Леви и его приложение к моделированию сетевого трафика. Электротехнические комплексы и системы. 2007. №1. Т. 3. С. 38 – 43.
7. Кроновер Р.М. Фракталы и хаос в динамических системах. Основы теории. М.: Постамаркет, 2000. 352 с.
8. Микова С.Ю, Оладько В.С. Нестеренко М.А. Подход к классификации аномалий сетевого трафика. Международный научный журнал «Инновационная наука». 2015. №11. С. 78 – 80.
9. Унтеров Д.С. Разработка метода обнаружения аномалий сетевого трафика на границе ЛВС предприятия: дис. магистерская. СПб.: 2016. 80 с.
10. Naila Belhadj Aissaa, Mohamed Guerroumia. Semi-Supervised Statistical Approach for Network Anomaly Detection – Procedia Computer Science 83. 2016. P. 1090 – 1095.
11. Mazurek M., Dymora P. Network anomaly detection based on the statistical self-similarity factor for HTTP protocol. Przeglad Elektrotechniczny. 2014. Vol. 90, no.1. P.127 – 130.
12. Федер Е. Фракталы. М.: Мир, 1991. 253 с.
13. Гончаренко А. RS-анализ (анализ фрактальной структуры временных рядов). URL: https://itnan.ru/post.php?c=1&p=256381 (дата обращения: 20.02.2019).
14. Справочник по теории вероятностей и математической статистике. В.С. Королев, Н.И. Портенко, А.В. Скороход, А.Ф. Турбин. -- М.: Наука, 1985. 640 с.
15. Котов Андрей. TShark - консольная версия программы Wireshark. URL: https://kb.zyxel.ru/hc/ru/articles/115002596254-TShark-Wireshark (дата обращения: 20.02.2019) .
16. Стресс-тест сети: DoS с использованием hping3 и спуфингом IP в Kali Linux. URL: https://codeby.net/blogs/stress-test-seti-dos-s-ispolzovaniem-hping3-i-spufingom-ip-v-kali-linux/ (дата обращения: 20.02.2019).

REFERENCES:

[1] Chebyshev Victor, Sinitsyn, Fedor, Parinov Denis, Liskin Alexander Kupreev Oleg. Development of information threats in the second quarter of 2018. Statistics. URL: https://securelist.ru/it-threat-evolution-q2-2018-statistics/90919/ (accessed: 20.02.2019). (in Russian)
[2] Leland W. E., Taqqu M.S., Willnger W., Wilson D.V. On the self-similar nature of Ethernet traffic. ACM Transaction on Networking. 1994. Vol. 2, no. 1. P. 1 – 15.
[3] Morgailov D.D., Ladyzenskiy Yu.V., Yunis M. Modelling self-similar input traffic to the network processor in the system of NS-2. Informatics and computer technologies-2012 (ICT – 2012). Proceedings of the VIII international scientific and technical conference of students, postgraduates and young scientists - September 18-19, 2012. Donetsk, DonNTU – 2012, P. 232 – 239.
URL: http://uran.donntu.org/~masters/2013/fknt/morgajlov/library/selfsim.htm (accessed: 20.02.2019). (in Russian)
[4] Porshnev S. V. Matematicheskie modeli informacionnyh potokov v vysokoskorostnyh magistral'nyh internet-kanalah [Mathematical models of information flows in high-speed main Internet channels]. M.:Gor. linija-Telekom, 2016. 232 p. (in Russian).
[5] Sidorova O. I. Puassonovskaja model' trafika s beskonechnym chislom neodnorodnyh istochnikov. [Poisson traffic model with an infinite number of inhomogeneous sources] Vestnik TvGU. Serija: Prikladnaja matematika. 2015. № 1. P. 47 – 66. (in Russian).
[6] Osin A. V. Fraktal'noe dvizhenie Levi i ego prilozhenie k modelirovaniju setevogo trafika. Jelektrotehnicheskie kompleksy i sistemy. [Levy's fractal motion and its application to network traffic modeling. Electrical systems and systems] 2007. №1, Vol. 3. P. 38 – 43. (in Russian).
[7] Crownover Richard M. Introduction to Fractals and Chaos. Jones and Bartlett Publisher, Inc. 1995. 299 p.
[8] Mikova S.Ju, Olad'ko V.S, Nesterenko M.A. Podhod k klassifikacii anomalij setevogo trafika. Mezhdunarodnyj nauchnyj zhurnal «Innovacionnaja nauka». [The approach to classification of anomalies in network traffic. International scientific journal "Innovative science".] 2015. №11. P. 78 – 80. (in Russian).

[9] Unterov D. S. Razrabotka metoda obnaruzhenija anomalij setevogo trafika na granice LVS predprija-tija: dis. magisterskaja. [Development of a method for detecting network traffic anomalies at the enterprise LAN border] SPb. 2016. 80 p. (in Russian).
[10] Naila Belhadj Aissaa, Mohamed Guerroumia. Semi-Supervised Statistical Approach for Network Anomaly Detection – Procedia Computer Science 83. 2016. P. 1090 – 1095.
[11] Mazurek M., Dymora P. Network anomaly detection based on the statistical self-similarity factor for HTTP protocol. Przeglad Elektrotechniczny. 2014. Vol. 90, no.1. P. 127 – 130.
[12] Feder J. Fractals. Springer Science + Business Media. Physics of Solids and Liquids. 1988. 305 p.
[13] Goncharenko A. RS-analysis (fractal structure analysis of time series). URL: https://itnan.ru/post.php?c=1&p=256381 (accessed: 20.02.2019). (in Russian)
[14] Jovanovic B. D., Levy P. S. A Look at the Rule of Three. The American Statistician, 1997. 51:2. P. 137 – 139.
[15] Kotov Andrey. TShark – is the console version of the program Wireshark. URL: https://kb.zyxel.ru/hc/ru/articles/115002596254-TShark-Wireshark (accessed: 20.02.2019). (in Russian)
[16] Network stress test: DoS using hping3 and IP spoofing in Kali Linux. URL: https://codeby.net/blogs/stress-test-seti-dos-s-ispolzovaniem-hping3-i-spufingom-ip-v-kali-linux/ (accessed: 20.02.2019). (in Russian)

Поступила в редакцию - 30 января 2019 г. Окончательный вариант – 28 февраля 2019 г.
Received – January 30, 2019. The final version – February 28, 2019.