Научная статья
УДК 621.396
https://doi.org/10.31854/1813-324X-2024-10-5-14-23

CC BY 4.0

Обнаружение аномалий трафика на основе обработки их фреймовых вейвлет-преобразований

Инна Михайловна Жданова [1] ✉, inna_zhdan@icloud.com
Сергей Сергеевич Дворников [1, 2], dvornik.92@mail.ru
Сергей Викторович Дворников [1, 2], practicdsv@yandex.ru

[1]Военная академия связи им. С.М. Буденного,
Санкт-Петербург, 194064, Российская Федерация
[2]Санкт-Петербургский государственный университет аэрокосмического приборостроения,
Санкт-Петербург, 190000, Российская Федерация

Аннотация

Актуальность. *Активный переход к массовой цифровой инфраструктуре, основанной на технологии интернета вещей (IoT), вывел телекоммуникационные сети на уровень доминирующих информационных ресурсов. Одновременное увеличение количества существующих интернет-сервисов неразрывно связано с ростом разнообразия сетевых аномалий на телекоммуникационное оборудование. В свою очередь, существующие методы обнаружения сетевых угроз не позволяют своевременно оценить сетевой трафик, который характеризуется большим количеством параметров, а выявляемые аномалии от внешнего вторжения не имеют явно выраженных закономерностей.* ***Целью (исследования)*** *является повышение эффективности обнаружения аномалий трафика по результатам обработки его фреймового вейвлет-преобразования. Научная задача состоит в разработке научно-методических подходов, позволяющих эффективно проводить анализ и своевременное обнаружение аномалий в сетевом трафике. В интересах исследования был проведен сравнительный обзор* ***методов*** *поиска обнаружения аномалий сетевого трафика, а также применены алгоритмы обнаружения неконтролируемых аномалий, методы анализа трафика на основе локального коэффициента выброса, бинарных деревьев, оптической эмиссионной спектроскопиии.*

Решение. *Рассматриваются результаты исследования возможности обнаружения аномалий в трафике битового потока по результатам его кратномастабного преобразования в базисе вейвлета Хаара. Обоснован выбор для дальнейшей обработки коэффициентов матрицы декомпозиции трафика вдоль переменной временно́го сдвига. Доказано, что кратномасштабные преобразования не только повышают структурные различия трафиков, но и открывают возможность локализации аномалий, вызвавшие указанные различия.* ***Научная новизна*** *работы определяется авторским подходом к обнаружению аномалий сетевого трафика при переходе от непосредственного представления сигнала в виде его дискретных отсчетов к коэффициентам, сформированным из матриц его вейвлет-преобразований, и, как результат, повышения его контрастности по отношению к другим сигналам с близкой структурой.*

Теоретическая значимость. *Доказана необходимость и достаточность использования вейвлет-коэффициентов вместо временны́х отсчетов сигналов в базисе материнского вейвлета из матрицы формируемого фрейма. Установлена взаимосвязь между показателями Херста и коэффициентами функций взаимной корреляции.*

Практическая значимость. *Полученные в работе результаты в перспективе могут быть использованы при построении моделей оценки сетевого трафика в условиях преднамеренных воздействий, а также методик поиска и синтеза эффективных методов защиты от них.*

Ключевые слова: *аномалии трафика, обнаружение аномалий битового потока, статистические параметры трафика, локализация положения аномалий в структуре трафика*

Ссылка для цитирования: Жданова И.М., Дворников С.С., Дворников С.В. Обнаружение аномалий трафика на основе обработки их фреймовых вейвлет-преобразований // Труды учебных заведений связи. 2024. Т. 10. № 5. С. 14–23. DOI:10.31854/1813-324X-2024-10-5-14-23. EDN:BJFZSE

© Жданова И.М., Дворников С.С., Дворников С.В., 2024

Original research
https://doi.org/10.31854/1813-324X-2024-10-5-14-23

Detection of Traffic Anomalies Based on Their Frame Wavelet Transformations Processing

Inna M. Zhdanova [1] ✉, inna_zhdan@icloud.com
Sergey S. Dvornikov [1, 2], dvornik.92@mail.ru
Sergey V. Dvornikov [1, 2], practicdsv@yandex.ru

[1]Military Academy of Communications,
St. Petersburg, 194064, Russian Federation
[2]Saint Petersburg State University of Aerospace Instrumentation,
St. Petersburg, 190000, Russian Federation

Annotation

Relevance. *The active transition to a massive digital infrastructure based on Internet of Things (IoT) technology has brought telecommunications networks to the level of dominant information resources. The one-time increase in the number of existing Internet services is inextricably linked to the growing variety of network anomalies on telecommunications equipment. In turn, existing methods of detecting network threats do not allow timely assessment of network traffic, which is characterized by a large number of parameters, and the detected anomalies from external interference do not have pronounced patterns.*

The purpose *of the study is to increase the efficiency of detecting traffic anomalies based on the results of processing its frame wavelet transform. The scientific task is to develop scientific and methodological approaches that allow effective analysis and timely detection of anomalies in network traffic. A comparative review of search* ***methods*** *for detecting network traffic anomalies, algorithms for detecting uncontrolled anomalies, traffic analysis methods based on local emission factor, binary trees, optical emission spectroscopy.*

Decision. *The results of the study of the possibility of detecting anomalies in the bitstream traffic based on the results of its multiple-variable transformation in the Haar wavelet basis are considered. The choice for further processing of the coefficients of the traffic decomposition matrix along the time shift variable is justified. It is proved that multiple-scale transformations not only increase the structural differences in traffic, but also open up the possibility of localization of anomalies that caused these differences.*

The scientific novelty *of the work is determined by the author's approach to detecting network traffic anomalies during the transition from the direct representation of a signal in the form of its discrete samples to coefficients formed from the matrices of its wavelet transformations, and, as a result, increasing its contrast with other signals with a similar structure.*

Theoretical significance. *The necessity and sufficiency of using wavelet coefficients instead of time samples of signals in the basis of the parent wavelet from the matrix of the generated frame is proved. The relationship between the Hurst indicators and the coefficients of the cross-correlation functions has been established.*

Practical significance*. The results obtained in the work, in the future, can be used in the construction of models for evaluating network traffic in conditions of deliberate, as well as methods for searching and synthesizing effective methods of protection against them.*

Keywords: *traffic anomalies, detection of bit stream anomalies, statistical traffic parameters, localization of the position of anomalies in the traffic structure*

For citation: Zhdanova I.M., Dvornikov S.S., Dvornikov S.V. Detection of Traffic Anomalies Based on Their Frame Wavelet Transformations Processing. *Proceedings of Telecommunication Universities.* 2024;10(5):14–23. (in Russ.) DOI:10.31854/1813-324X-2024-10-5-14-23. EDN:BJFZSE

Введение

Развитие цифровых технологий привело к существенному возрастанию объема передаваемой информации в сетях передачи данных [1]. А с учетом активного перехода к массовой инфраструктуре, основанной на технологии интернета вещей (IoT, *аббр. от англ.* Internet of Things), телекоммуникационные сети становятся доминирующим информационным ресурсом, оказывающим значительное влияние на общество и экономику любого современного государства [2, 3].

В свою очередь, развитие интернет-сервисов привело к росту разнообразия сетевых аномалий, обусловленных не только низким качеством сетевых услуг, но различными видами вредоносных атак [4]. Все это не только негативно сказывается на работе веб-сервисов, но и вызывает экономические потери, а также приводит к социальным потрясениям. Таким образом, аномалии выступают в качестве своеобразного индикатора, характеризующего состояние не только информационного трафика, но и телекоммуникационной сети в целом. Согласно [5], под аномалией понимается любое отклонение статистических показателей трафика от номинально установленных значений.

В качестве основных причин возникновения аномалий, как правило, выступают или технологический сбой работы оборудования, или злонамеренное вторжение в виде DDoS-атаки [5–7]. При этом следует понимать, что сам по себе факт наличия аномалии нельзя рассматривать в качестве гарантированного признака несанкционированного воздействия. Однозначное заключение можно сделать только по результатам анализа аномалий.

На основании указанного следует заключить, что поиск научно-методических подходов, обеспечивающих эффективное проведение анализа трафика, является актуальной задачей в области информационной безопасности, обеспечивающей своевременное обнаружение причин, приводящих к возникновению аномалий трафика.

Анализ подходов к выявлению аномалий

В настоящее время проблема обнаружения аномалий, как тема исследования, рассматривалась в различных научных областях знаний, о чем свидетельствуют многочисленные публикации [1–8]. В частности, в [4] представлен анализ подходов обнаружения аномалий в сетевом трафике. По результатам указанного анализа авторами сделан обоснованный вывод об объективных сложностях получения однозначных закономерностей аномалий с несанкционированным вторжением.

Прежде всего, это связано с тем, что сетевой трафик характеризуется большим количеством параметров, а выявляемые аномалии, обусловленные внешним вторжением, не имеют явно выраженных закономерностей в их структуре [4, 5, 9]. Однако, аномалия является причиной угроз служебному трафику.

Вместе с тем не все подходы, используемые при анализе трафика, в равной мере применимы для выявления аномалий [4, 8]. А поскольку разработка новых подходов требует проведения их сравнительного анализа, то целесообразно из арсенала существующих способов отобрать те из них, которые позволяют непосредственно обрабатывать битовые потоки.

Простейший способ обнаружения аномалий основан на расчете статистических параметров трафика [10, 11], к которым относят используемые в описательной статистике для характеристики данных следующие статистические меры (показатели): местоположение; разброс местоположения; форма.

В свою очередь, мера местоположения определяется средним значением, математическим ожиданием, значением медианы и величиной моды.

Среднее значение трафика $f(x)$ на интервале от $a = x_1$ до $b = x_2$ определяется как:

$$[\overline{f(x)}] = \frac{1}{b-a}\sum_{i\,=\,a}^{b} x_i. \qquad (1)$$

Для расчета математического ожидания трафика как функции $f(x)$ используется значение вероятности p_i проявления каждого из значений x_i функции трафика:

$$\mathbf{E}[f(x)] = \sum_{i\,=\,a}^{b} x_i p_i. \qquad (2)$$

При равномерном распределении величина математического ожидания вырождается в среднее значение.

Медиана представляет собой среднее значение при упорядочении элементов функции $f^*(x)$ от наименьшего значения к наибольшему:

$$\mathrm{med}\left[f\left(x_{\frac{N+1}{2}}\right)\right], \qquad (3)$$

где N – общее количество данных трафика $f(x)$ на интервале от $a = x_1$ до $b = x_2$.

Мода определяет одно или несколько значений функции $f(x)$, которые наиболее часто проявляются в обрабатываемом трафике.

Учитывая битовую последовательность трафика, тем не менее, целесообразность использования показателей медианы и моды оправдана в том случае, когда анализ производится на уровне пакетов [12, 13].

Гораздо больший интерес представляют показатели меры разброса местоположения, в качестве которых выступают:

– дисперсия, как средняя величина квадрата среднего значения:

$$\mathbf{D}[f(x)] = \sum_{i=a}^{b} p_i(x_i - \mathbf{E}[f(x)])^2 = \\ = \frac{1}{b-a-1}\sum_{i=a}^{b}(x_i - \mathbf{E}[f(x)])^2; \quad (4)$$

– среднеквадратическое отклонение, как квадратный корень из дисперсии:

$$S[f(x)] = \sqrt{\sum_{i=a}^{b} p_i(x_i - \mathbf{E}[f(x)])^2} = \\ = \sqrt{\frac{1}{b-a-1}\sum_{i=a}^{b}(x_i - \mathbf{E}[f(x)])^2}; \quad (5)$$

– минимальные и максимальные размахи распределений функции $f(x)$, значения дисперсии рассчитаны в соответствии с формулой (4):

$$\min(\mathbf{D}[f(x)]) \div \max(\mathbf{D}[f(x)]); \quad (6)$$

– интерквартильный размах, представляющий разность между третьим $f(x_{0,75})$ и первым $f((x_{0,25});)$ квартилями, т. е.:

$$f(x_{0,75}) - f(x_{0,25}); \quad (7)$$

– размах вариаций, т. е. различия значений какого-либо показателя:

$$R = \max[f(x)] - \min[f(x)]; \quad (8)$$

– для практики интерес представляет так называемый коэффициент осцилляции (относительный размах вариации):

$$\rho = \frac{R}{\mathbf{E}[f(x)]} \quad (9)$$

и коэффициент вариации:

$$V = \frac{S[f(x)]}{\mathbf{E}[f(x)]}, \quad (10)$$

а также доверительный интервал, определяющий границы, в пределах которых обеспечивается значение математического ожидания с заданной вероятностью.

К мерам формы относят коэффициенты асимметрии и перекоса:

$$A_S = \frac{\mu_3}{S^3[f(x)]}, \quad (11)$$

где μ_3 – третий центральный момент:

$$\mu_3 = \mathbf{E}[(f(x) - \mathbf{E}[f(x)])^3].$$

Коэффициент перекоса (эксцéсса) характеризует меру остроты пика распределения, и рассчитывается по формуле [14]:

$$A_\text{A} = (\mu_4/S^4[f(x)]) - 3, \quad (12)$$

где μ_4 – четвертый центральный момент:

$$\mu_4 = \mathbf{E}[(f(x) - \mathbf{E}[f(x)])^4].$$

Из представленных характеристик, согласно [10, 11], наиболее информативными для рассматриваемой проблематики являются среднее значение трафика, величина его дисперсии, среднеквадратическое отклонение, а также коэффициенты вариаций, асимметрии и перекоса.

На практике в качестве критерия обнаружения аномалий используют пороговые значения, рассчитанные по результатам статистической обработки потока данных [11, 15, 16]. Вместе с тем, существуют и более прогрессивные алгоритмы обнаружения неконтролируемых аномалий, основанных на анализе локального коэффициента выбросов (LOF, *аббр. от англ.* Local Outlier Factor) [17, 18], и оценки выбросов по результатам обработки гистограмм (HBO, *аббр. от англ.* Histogramm-Based Outlier) [4, 19, 20].

Широкое применение получил алгоритм обнаружения аномалий, основанный на использовании так называемых бинарных деревьев Isolation forest (iForest). Алгоритм разработан Фей Тони Лю, Кай Мин Тоном и Чжи-Хуа Чжоу в 2008 г. [21]; имеет линейную временну́ю сложность и низкие требования к памяти, что позволяет работать с большими объемами данных. Его работа основана на предположении, что характеристики аномалий существенно отличаются друг от друга, что позволяет их идентифицировать.

Аналитически алгоритм iForest может быть записан в следующем виде [21–23]:

$$c(m) = \begin{cases} 2H(m-1) - \dfrac{2(m-1)}{n} & \text{for } m > 2; \\ 1 & \text{for } m = 2; \\ 0 & \text{otherwise,} \end{cases} \quad (13)$$

где n – объем тестируемых данных; m – размер набора выборок; H – число оцениваемых гармоник согласно выражению:

$$H(i) = \ln(i) + \gamma,$$

где $\gamma = 0{,}5772156649$ – постоянная Эйлера – Маскерони [24].

Особенность алгоритма iForest в том, что его реализация не предполагает оценку плотности распределения трафика. Метод отсекает изолированные точки от нормальных кластеризованных. Поэтому алгоритм iForest работает быстро, поскольку

просто разделяет трафик данных, выбирая для этого случайным образом точку деления.

Алгоритм iForest неплохо зарекомендовал себя в области производства полупроводников. В частности, размерные и массивные характеристики данных оптической эмиссионной спектроскопии (OES, *аббр. от англ.* Optical Emission Spectroscopy) ограничивают достижимую производительность систем обнаружения аномалий. Для решения данной проблемы Л. Паггини и С. Маклун в [25] предложили осуществлять выбор переменных по результатам применения алгоритма iForest, позволившего не только обнаруживать аномалии, но и уменьшить размерность данных.

Также хорошо зарекомендовал себя алгоритм iForest при обработке больших объемов данных, обеспечивая приемлемую точность обнаружения аномалий. Однако в [26] выявлено ограничение: iForest позволяет установить, является ли выборка данных выбросом, вычисляя оценку аномалий.

Т. е. эффективность работы алгоритма iForest в значительной степени зависит от настройки оценки показателя аномалий, которая рассчитывается как:

$$s(x, m) = 2^{\frac{-\mathbf{E}[h(x)]}{c(m)}},$$

где $\mathbf{E}[h(x)]$ – среднее значение длины пути при рекурсивном секционировании данных, которое представляется древовидной структурой, именуемой деревом изоляции [21]; величина $c(m)$ рассчитывается в соответствии с выражением (13).

Поскольку в реальной промышленной сетевой среде можно полагаться только на опыт, полученный в ходе подбора соотношения $s(x, m)$, то, разумеется, нет гарантий того, что алгоритм iForest сможет обеспечить наилучшее быстродействие при расчетах.

Неточные настройки параметров также могут приводить к уменьшению точности и увеличению количества ложных обнаружений в iForest. Кроме того, широкое применение в методах исследования сетевого трафика находят алгоритмы анализа плотности распределений временны́х рядов, в ходе которых оценивается стационарность, различия, определяется тренд, используется «снятие сезонности» и методы скользящего среднего, экспоненциального среднего, обнаружение и устранение выбросов, выравнивание по времени, преобразование и масштабирование данных, нормализация. Следует отметить, что для обнаружения аномалий методы исследования сетевого трафика практически не применяются [27].

Информативной мерой, характеризующей фрактальную размерность обрабатываемой выборки, является показатель Херста $X_{\text{п}}$, позволяющий выявлять долгосрочные зависимости временны́х рядов [8]. Чем ближе значение $X_{\text{п}} \to 1$, тем устойчивее зависимость. С указанных позиций показатель Херста в какой-то мере можно рассматривать как некоторую оценку устойчивости таких зависимостей, определяемых статистическими закономерностями.

Уникальность показателя Херста в том, что он тесно связан с такими широко известными параметрами временны́х рядов, как спектральный и фрактальный показатели $S_{\text{п}}$ и $F_{\text{п}}$, а также корреляционный параметр $K_{\text{п}}$ (таблица 1).

Связь показателя Херста с известными параметрами временны́х рядов можно представить в аналитическом виде:

$$K_{\text{п}} = 2(1 - X_{\text{п}});$$
$$S_{\text{п}} = 2X_{\text{п}} + 1;$$
$$F_{\text{п}} = 3 - 2X_{\text{п}}.$$

Аналитически расчет показателя Херста осуществляется по формуле [28]:

$$\mathbf{E}\left[\frac{R_{[a;\,b]}}{S[f(x)_{[a;\,b]}]}\right] = C[a;b]^{X_{\text{п}}}, \qquad (14)$$

где $[a;\ b]$ – интервал временно́го анализа; C – константа.

Согласно выражению (14), показатель Херста лишь опосредованно может характеризовать аномалии трафика.

И, наконец, корреляционный подход, предполагающий расчет как автокорреляционных функций, так и функций взаимной корреляции.

Автокорреляционная функция для трафика, согласно [24], рассчитывается в вид:

$$r(k) = \frac{\sum_{i=}^{N-\tau}{}_{1}(x_i - \mathbf{E}[f(x)])(x_{i+k} - \mathbf{E}[f(x)])}{(N - \tau)\mathbf{D}[f(x)])}. \qquad (15)$$

В качестве числового показателя, характеризующего близость двух трафиков, например, текущего $f(x)$ и контрольного (эталонного) $g(x)$, используют коэффициент корреляции Пирсона [24, 29]:

$$r_{fg} = \frac{\sum_{i=}^{N}{}_{1}(x_i - \mathbf{E}[f(x)])(x_i - \mathbf{E}[g(x)])}{\sqrt{\sum_{i=}^{N}{}_{1}(x_i - \mathbf{E}[f(x)])^2 \sum_{i=}^{N}{}_{1}(x_i - \mathbf{E}[g(x)])^2}}. \qquad (16)$$

Рассмотренный арсенал широко известных инструментов анализа трафика позволяет сделать вывод, что для исследования аномалий трафика, представленного в виде битового потока к уже отмеченным подходам, с использованием среднего значения трафика, величина его дисперсии, а также коэффициенты вариаций, асимметрии и перекоса, следует применять алгоритм iForest и коэффициент корреляции Пирсона.

Обоснование подхода к обнаружению аномалий трафика по результатам обработки его фреймового вейвлет-преобразования

Практическое применение в обработке сигналов вейвлет-преобразование получило благодаря работам С. Маллата [9, 30], который обосновал, что в ряде приложений целесообразно использовать для обработки вместо временны́х отсчетов сигналов их вейвлет-коэффициенты.

В продолжение развития этого направления в [31] предложено вместо вейвлет-коэффициентов выбирать коэффициенты преобразования сигнала в базисе материнского вейвлета из матрицы формируемого фрейма.

Если принимать непрерывное вейвлет-преобразование как свертку обрабатываемого процесса $z(t)$ с кратномасштабными копиями формирующего материнского вейвлета [32]:

$$W(k,n) = \frac{1}{\sqrt{k}} \int_{-\infty}^{\infty} z(t)\psi_{k,n}(t)\,dt, \qquad (17)$$

где

$$\psi_{k,n}(t) = \frac{1}{\sqrt{k}}\psi\left(\frac{t-n}{k}\right), \qquad (18)$$

то с математических позиций параметры n и k являются непрерывными.

Таким образом в [31], применительно к дискретному набору данных предложено работать не со всем обрабатываемым процессом, а с его частью, выборкой или фреймом, что позволило выражение (17) преобразовать к виду:

$$W_k(n) = \frac{1}{\sqrt{k}} \sum_{i=a}^{b} z(x_i)\psi\left(\frac{x_i - n}{k}\right). \qquad (19)$$

Формально выражение (19) сохраняет бóльшую часть положительных свойств непрерывного вейвлет-преобразования, что позволяет его считать кратномасштабным преобразованием процесса $z(t)$ на длительности фрейма в интервале от a до b. В качестве примера на рисунке 1а показана матрица фреймового преобразования битового потока в виде тестового сигнала $z1$ в интервале 32 отсчетов, содержащего последовательность повторяющихся нулей и единиц. В качестве формирующего определен материнский вейвлет Хаара.

Переход от непосредственного представления сигнала в виде его дискретных отчетов к коэффициентам, сформированным из матриц его вейвлет-преобразований, обусловлен тем, что в результате происходит повышение его контрастности по отношению к другим временны́м рядам с близкой структурой, определяемой их элементами. Данный переход обоснован в [33].

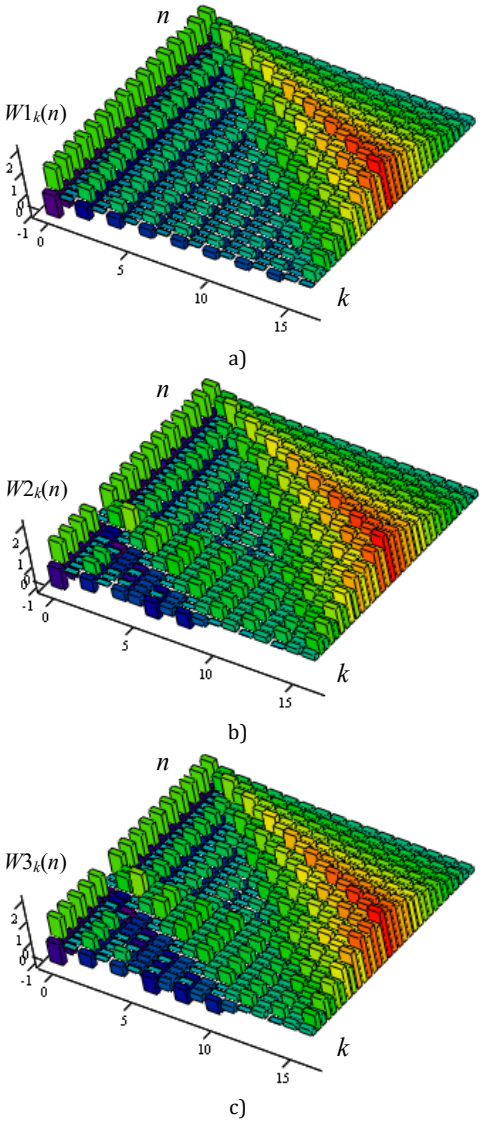


Рис. 1. Матрица фреймового вейвлет-преобразования тестового сигнала: a) *z*1; b) *z*2; c) *z*3

Fig. 1. Frame Wavelet Transform Matrix of the Test Signal: a) z1; b) z2; c) z3

Следовательно, можно выдвинуть гипотезу, что если рассматривать два битовых потока с близким, но не совпадающем набором 0 и 1, то после вейвлет-преобразований этих потоков контрастность их различий усилится.

Рассмотрим три битовых потока:

$z1$ = {01010101010101010101010101010101};

$z2$ = {01010101011101010101010101010101};

$z3$ = {01010101010111010101010101010101}.

Будем полагать, что поток $z1$ является эталонным, а потоки $z2$ и $z3$ содержат аномалии в виде трех последовательных 1. Для потоков $z2$ и $z3$, в соответствии с выражением (19), были реализованы их фреймовые вейвлет-преобразования (для потока $z1$ фреймовое вейвлет-преобразование аналогично изображенному на рисунке 1а).

Полученные матрицы явились исходными данными для формирования на их основе преобразованных потоков $s1$, $s2$ и $s3$, посредством выбора соответствующих коэффициентов вдоль оси сдвигов n. В результате предоставляется возможность отбора коэффициентов матриц, обеспечивающих наибольшую информативность об аномалии исходного потока. Критерием такого отбора были определены коэффициенты корреляции Пирсона (16). Результаты представлены в таблице 1.

Результаты таблицы 1 ограничены n = 13, поскольку при дальнейшем увеличении n коэффициент Пирсона сохраняет свое значение, равное 1. Важным моментом полученных результатов является то, что, несмотря на одинаковую аномалию в абсолютном значении исходных потоков, расчетные величины коэффициентов Пирсона их кратномасштабных коэффициентов различны. В то же время значение коэффициентов Пирсона для исходных потоков одинаковы: $rz1z3 = rz1z2 = 0{,}939$.

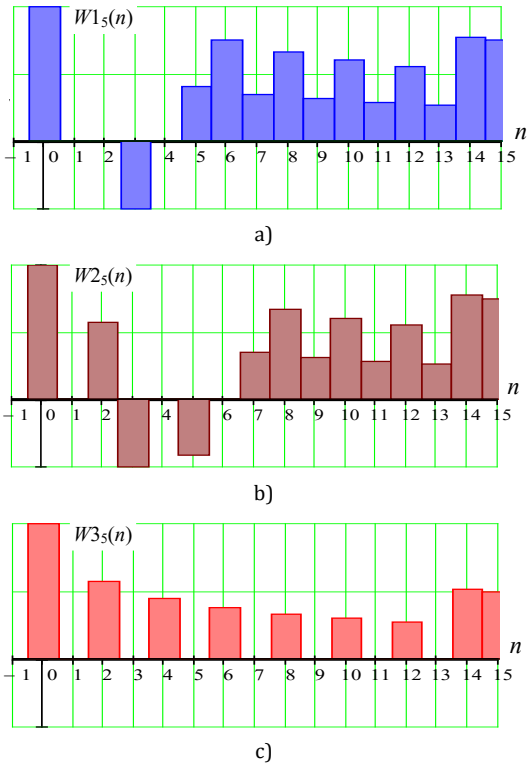
Данный факт открывает перспективу не только повысить вероятность обнаружения аномалии, поскольку, согласно таблице 1, различия между коэффициентами матриц кратномасштабного преобразования потоков гораздо сильнее, чем между самими потоками.

ТАБЛИЦА 1. Значение коэффициента Пирсона между коэффициентами матриц кратномасштабного преобразования

TABLE 1. Pearson Coefficient Value between the Coefficients of the Multiple-Scale Transformation Matrices

Потоки	n	0	1	2	3	4	5	6	7	8	9	10	11	12	13
$z1$ и $z2$	$rs1s2$	0,762	0,722	0,76	0,635	0,815	0,667	0,835	0,634	0,877	0,715	0,888	0,65	0,964	1
$z1$ и $z3$	$rs1s2$	0,723	0,626	0,773	0,612	0,782	0,548	0,843	0,624	0,848	0,513	0,945	1	1	1

В качестве примера на рисунке 2 показаны потоки коэффициентов кратномасштабного преобразования, соответствующие значению k = 5, при котором обеспечиваются наибольшие различия.



a)

b)

c)

Рис. 4. Коэффициенты матрицы фреймового вейвлет-преобразования тестовых потоков: a) $z1$; b) $z2$; c) $z3$

Fig. 4. Coefficients of the Frame Wavelet Transform Matrix of the Test Stream: a) z1; b) z2; c) z3

Следует отметить, что выбор коэффициентов матрицы возможен и вдоль оси k, но получаемые в этом случае коэффициенты менее контрастны, т. е. информативны. Различия в абсолютных значениях коэффициентов Пирсона потоков, составленных из коэффициентов матриц их кратномасштабных преобразований, открывают возможность не только повышать вероятность их обнаружения ввиду большей контрастности, но и позиционировать местоположение аномалий в пределах потока данных.

Необходимо добавить, что статистические параметры потока, рассмотренные в предыдущем разделе, еще менее информативны. В таблице 2 представлены результаты расчетов статистических параметров потоков $z1$, $z2$ и $z3$, согласно которым следует, что все параметры потоков $z2$ и $z3$, традиционно рассчитываемые при статистической обработке, одинаковы, несмотря на различия расположения аномалий в их структуре. Поэтому использование результатов статистического анализа, включая расчетные значения коэффициента корреляции Пирсона, не позволят идентифицировать положение локации аномалии.

В то же время значения коэффициента корреляции Пирсона, рассчитанные на основе элементов матриц фреймового вейвлет-преобразования указанных потоков по отношению к эталонному, существенно отличаются. При этом указанные различия настолько существенны, что могут выступать индикаторами, не только характеризующими положение возникшей аномалии в пределах обрабатываемого потока, но и идентифицировать структуру самой аномалии.

ТАБЛИЦА 2. Значение статистических параметров, характеризующих поток *z*1, *z*2 и *z*3

TABLE 2. The Value of Statistical Parameters Characterizing the Flow z1, z2 and z3

Параметр	**E**[*]	**D**[*]	*V*	A_S	A_A	*rz*1*z*2	*rz*1*z*3	$X_п$	*c*(*m*)
*z*1	0,5	0,25	1	3	7	0,939	0,939	–	0,984
*z*2	0,563	0,246	0,882	3,15	7,635	0,939	–	0,53	0,986
*z*3	0,563	0,246	0,882	3,15	7,635	–	0,939	0,53	0,985

Заключение

Анализ результатов таблицы 2 позволяет заключить, что среди стандартных широко используемых статистических характеристик потоков наиболее информативным является алгоритм iForest. Однако предложенный подход позволяет получить более ярко выраженные различия.

Представленные исследования открывают новый взгляд на решение вопроса не только обнаружения аномалии, но и ее локализации в структуре трафика битового потока. Задача локализации положения потока может быть обеспечена на основе приложения элементов распознавания образов, реализованных в [20]. Данное направление, по мнению авторов, является перспективным для дальнейшего исследования. Кроме того, повышение различий трафиков авторы связывают с использованием методов частотно-временно́го анализа [31].

Список источников

1. Колчина О.А., Лесничая М.А. Оценка дифференциации развития муниципальных образований по уровню и качеству жизни населения в условиях цифровой трансформации // Экономика и предпринимательство. 2020. № 11(124). С. 578–584. DOI:10.34925/EIP.2020.124.11.108. EDN:WCNVSU

2. Askaruly B., Abitova G.A. Hybrid information systems modeling technology for business process analysis based on the internet of things // Bulletin of Shakarim University. Technical Sciences. 2023. Iss. 3(11). PP. 19–28. DOI:10.53360/2788-7995-2023-3(11)-2. EDN:QGULYH

3. Larsson E., Bratt E., Palmqvist J., Söderberg A., Hall A. Internet of things as a complement to increase safety // Journal of the Belarusian State University. International Relations. 2020. Iss. 1. PP. 88–93. EDN:NYVGQK

4. Лизнев Д.С. Обзор методов прогнозирования сетевых аномалий // Вестник СибГУТИ. 2023. Т. 17. № 2. С. 44–50. DOI:10.55648/1998-6920-2023-17-2-44-50. EDN:RPMMTF

5. Орехов А.В., Орехов А.А. Автоматическое обнаружение аномалий сетевого трафика при DDoS-атаках // Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления. 2023. Т. 19. № 2. С. 251–263. DOI:10.21638/11701/spbu10.2023.210. EDN:XYNCXN

6. Гайфулина Д.А., Котенко И.В. Анализ моделей глубокого обучения для задач обнаружения сетевых аномалий интернета вещей // Информационно-управляющие системы. 2021. № 1(110). С. 28–37. DOI:10.31799/1684-8853-2021-1-28-37. EDN:DTPPJY

7. Ажмухамедов И.М., Марьенков А.Н. Поиск и оценка аномалий сетевого трафика на основе циклического анализа // Инженерный вестник Дона. 2012. № 2(20). С. 17–26. EDN:PCRPQT

8. Kotenko I.V., Saenko I.B., Kushnerevich A.S. Parallel big data processing system for security monitoring in Internet of Things networks // Journal of Wireless Mobile Networks, Ubiquitous Computing and Reliable Applications (JoWUA). 2017. Vol. 8. Iss. 4. PP. 60–74. DOI:10.22667/JOWUA.2017.12.31.060

9. Mallat S.G. A Theory of multiresolution signal decomposition: the wavelet representation // IEEE Transactions on Pattern Analysis and Machine Intelligence. 1989. Vol. 11. Iss. 7. PP. 674–693. DOI:10.1109/34.192463

10. Поздняк И.С., Плаван А.И. Выявление DOS-атак с помощью анализа статистических характеристик трафика // Инфокоммуникационные технологии. 2021. Т. 19. № 1. С. 73–80. DOI:10.18469/ikt.2021.19.1.10. EDN:CTTSUN

11. Шелухин О.И., Судариков Р.А. Анализ информативных признаков в задачах обнаружения аномалий трафика статистическими методами // T-Comm: Телекоммуникации и транспорт. 2014. Т. 8. № 3. С. 14–18. EDN:SGIHFZ

12. Дворников С.В., Погорелов А.А., Вознюк М.А., Иванов Р.В. Оценка имитостойкости каналов управления с частотной модуляцией // Информация и космос. 2016. № 1. С. 32–35. EDN:VPQCFF

13. Симаков Д.В., Кучин А.А. Анализ статистических характеристик Интернет-трафика в магистральном канале // T-Comm: Телекоммуникации и транспорт. 2014. Т. 8. № 11. С. 95–98. EDN:TESPPD

14. Калистратова А.В., Никитин А.А. Исследование уравнения Дикмана с интегральными ядрами, имеющими переменное значение коэффициентов эксцесса // Доклады Академии наук. 2016. Т. 470. № 6. С. 628–631. DOI:10.7868/S086956521630006X. EDN:WOSDQV

15. Воробьев В.И., Грибунин В.Г. Теория и практика вейвлет-преобразования. СПб.: ВУС, 1999. 204 с.

16. Попов И.Ю. Метод настройки параметров алгоритма локальных коэффициентов выбросов для поиска сетевых аномалий // Наука и бизнес: пути развития. 2019. № 8(98). С. 88–91. EDN:DWUTUV

17. Нестерова Е.С. Алгоритм локальных коэффицентов выбросов // Фундаментальные и прикладные исследования в современном мире. 2019. № 26-2. С. 41–43. EDN:FWQJQH

18. Денисова А.Ю., Мясников В.В. Обнаружение аномалий на гиперспектральных изображениях // Компьютерная оптика. 2014. Т. 38. № 2. С. 287–296. DOI:10.18287/0134-2452-2014-38-2-287-296. EDN:SFAZCT

19. Васильева Д.В., Дворников С.С., Толстуха Ю.Е., Обрезков П.С., Дворников С.В. Формирование векторов признаков для систем видеонаблюдения // Вопросы радиоэлектроники. Серия: Техника телевидения. 2023. № 4. С. 62–68. EDN:NZSBEJ

20. Liu F.T., Ting K.M., Zhou Z.H. Isolation Forest // Proceedings of the 8th International Conference on Data Mining (Pisa, Italy, 15–19 December 2008). IEEE, 2008. PP. 413–422. DOI:10.1109/ICDM.2008.17

21. Краснов Ф.В., Смазневич И.С., Баскакова Е.Н. Оптимизационный подход к выбору методов обнаружения аномалий в однородных текстовых коллекциях // Информатика и автоматизация. 2021. Т. 20. № 4. С. 869–904. DOI:10.15622/ia.20.4.5. EDN:XWIJOS

22. Шелухин О.И., Полковников М.В. Применение алгоритма «Изолирующий Лес» для решения задач обнаружения аномалий // Решение. 2019. Т. 1. С. 186–188. EDN:SSIRSY

23. Bol G. Deskriptive Statistik. Oldenbourg: Oldenburg Verlag, 2004.

24. Puggini L., McLoone S. An enhanced variable selection and Isolation Forest based methodology for anomaly detection with OES data // Engineering Applications of Artificial Intelligence. 2018. Vol. 67. PP. 126–135. DOI:10.1016/j.engappai.2017.09.021

25. Spiekermann D., Keller J. Unsupervised packet-based anomaly detection in virtual networks // Computer Networks. 2021. Vol. 192. P. 108017. DOI:10.1016/j.comnet.2021.108017

26. Ahmed M., Mahmood A., Hu J. A Survey of Network Anomaly Detection Techniques // Journal of Network and Computer Applications. 2016. Vol. 60. Iss. 1. PP. 19–31. DOI:10.1016/j.jnca.2015.11.016

27. Калуш Ю.А., Логинов В.М. Показатель Херста и его скрытые свойства // Сибирский журнал индустриальной математики. 2002. Т. 5. № 4(12). С. 29–37. EDN:HZOKUN

28. Dvornikov S.-Jr., Dvornikov S. Detection Range Estimation of Small UAVs at a Given Probability of Their Identification // Proceedings of Telecommunication Universities. 2023. Vol. 9. Iss. 4. PP. 6–13. DOI:10.31854/1813-324X-2023-9-4-6-13. EDN:YLBWOS

29. Умбиталиев А.А., Дворников С.В., Оков И.Н., Устинов А.А. Способ сжатия графических файлов методами вейвлет-преобразований // Вопросы радиоэлектроники. Серия: Техника телевидения. 2015. № 3. С. 100–106. EDN:UMOIVJ

30. Mallat S.G., Zhang Z. Matching pursuits with time-frequency dictionaries // IEEE Transactions on Signal Processing. 1993. Vol. 41. Iss. 12. PP. 3397–3415. DOI:10.1109/78.258082

31. Alzhanov A., Nugumanova A., Sutula M. Research on crop classification methods based on machine learning using wavelet transformations // Eurasian Journal of Applied Biotechnology. 2023. Iss. 2. PP. 52–60. DOI:10.11134/btp.2.2023.7. EDN:ZKFOAC

32. Новиков И.Я., Протасов В.Ю., Скопина М.А. Теория всплесков. М.: Физматлит, 2005. 616 с.

References

1. Kolchina O.A., Lesnichaya M.A. Assessment of the Differentiation of Municipalities Development by People's Living Standards and Their Quality of Life in the Digital Transformation Conditions. *Economics and Entrepreneurship*. 2020;11(124):578–584. (in Russ.) DOI:10.34925/EIP.2020.124.11.108. EDN:WCNVSU

2. Askaruly B., Abitova G.A. Hybrid information systems modeling technology for business process analysis based on the internet of things. *Bulletin of Shakarim University. Technical Sciences*. 2023;3(11):19–28. DOI:10.53360/2788-7995-2023-3(11)-2. EDN:QGULYH

3. Larsson E., Bratt E., Palmqvist J., Söderberg A., Hall A. Internet of things as a complement to increase safety. *Journal of the Belarusian State University. International Relations*. 2020;1:88–93. EDN:NYVGQK

4. Lisnev D.S. Review of methods for forecasting network anomalies. *The Herald of the Siberian State University of Telecommunications and Information Science*. 2023;17(2):44–50. (in Russ.) DOI:10.55648/1998-6920-2023-17-2-44-50. EDN:RPMMTF

5. Orekhov A.V., Orekhov A.A. Automatic detection of network traffic anomalies in DDoS attacks. *Vestnik of Saint Petersburg University. Applied Mathematics. Computer Science. Control Processes*. 2023;19(2):251–263. (in Russ.) DOI:10.21638/11701/spbu10.2023.210. EDN:XYNCXN

6. Gaifulina D.A., Kotenko I.V. Analysis of deep learning models for the tasks of detecting network anomalies of the Internet of Things. *Information and Control Systems*. 2021;1(110):28–37. (in Russ.) DOI:10.31799/1684-8853-2021-1-28-37. EDN:DTPPJY.

7. Azhmukhamedov I.M., Marienkov A.N. Search and evaluation of network traffic anomalies based on cyclic analysis. *Inzhenernyi vestnik Dona*. 2012;2(20):17–26. (in Russ.) EDN:PCRPQT

8. Kotenko I.V., Saenko I.B., Kushnerevich A.S. Parallel big data processing system for security monitoring in Internet of Things networks. *Journal of Wireless Mobile Networks, Ubiquitous Computing and Reliable Applications (JoWUA)*. 2017;8(4):60–74. DOI:10.22667/JOWUA.2017.12.31.060

9. Mallat S.G. A Theory of multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1989;11(7):674–693. DOI:10.1109/34.192463

10. Pozdnyak I.S., Plavan A.I. Identification of DOS attacks by analyzing statistical characteristics of traffic. *Infocommunication Technologies*. 2021;19(1):73–80. (in Russ.) DOI:10.18469/ikt.2021.19.1.10. EDN:CTTSUN

11. Shelukhin O.I., Sudarikov R.A. Analysis of informative signs in the tasks of detecting traffic anomalies by statistical methods. *T-Comm*. 2014;8(3):14–18. (in Russ.) EDN:SGIHFZ

12. Dvornikov S.V., Pogorelov A.A., Voznyuk M.A., Ivanov R.V. Assessment of the immunity of control channels with frequency modulation. *Information and Space*. 2016;1:32–35. (in Russ.) EDN:VPQCFF

13. Simakov D.V., Kuchin A.A. Analysis of statistical characteristics of Internet traffic in the trunk channel. *T-Comm.* 2014;8(11):95–98. (in Russ.) EDN:TESPPD

14. Kalistratova A.V., Nikitin A.A. Investigation of the Dikman equation with integral kernels having variable values of kurtosis coefficients. *Doklady Akademii nauk*. 2016;470(6):628–631. (in Russ.) DOI:10.7868/S086956521630006X. EDN:WOSDQV

15. Vorobyov V.I., Gribunin V.G. *Theory and Practice of the Wavelet Transform*. St. Petersburg: VUS Publ.; 1999. 204 p. (in Russ.)
16. Popov I.Y. Method of setting parameters of the algorithm of local emission coefficients for searching for network anomalies. *Science and Business: Ways of Development*. 2019;8(98):88–91. (in Russ.) EDN:DWUTUV
17. Nesterova E.S. Algorithm of local emission coefficients. *Fundamental and Applied Research in the Modern World*. 2019; 26-2:41–43. (in Russ.) EDN:FWQJQH
18. Denisova A.Yu., Myasnikov V.V. Detection of anomalies in hyperspectral images. Computer Optics. 2014;38:2:287–296. DOI:10.18287/0134-2452-2014-38-2-287-296. (in Russ.) EDN:SFAZCT
19. Vasilyeva D.V., Dvornikov S.S., Tolstukha Yu.E., Obrezkov P.S., Dvornikov S.V. Formation of feature vectors for video surveillance systems. *Voprosy radioelektroniki Seriia Tekhnika televideniia*. 2023;4:62–68. (in Russ.) EDN:NZSBEJ
20. Liu F.T., Ting K.M., Zhou Z.H. Isolation Forest. *Proceedings of the 8th International Conference on Data Mining, 15–19 December 2008, Pisa, Italy*. IEEE; 2008. p.413–422. DOI:10.1109/ICDM.2008.17
21. Krasnov F.V., Smaznevich I.S., Baskakova E.N. Optimization approach to the choice of methods for detecting anomalies in homogeneous text collections. *Informatics and Automation*. 2021;20(4):869–904. (in Russ.) DOI:10.15622/ia.20.4.5. EDN:XWIJOS
22. Shelukhin O.I., Polkovnikov M.V. Application of the "Isolating Forest" algorithm for solving problems of anomaly detection. *Reshenie*. 2019;1:186–188. (in Russ.) EDN:SSIRSY
23. Bol G. *Deskriptive Statistik*. Oldenbourg: Oldenburg Verlag; 2004.
24. Puggini L., McLoone S. An enhanced variable selection and Isolation Forest based methodology for anomaly detection with OES data. *Engineering Applications of Artificial Intelligence*. 2018;67:126–135. DOI:10.1016/j.engappai.2017.09.021
25. Shpikerman D., Keller J. Anomaly detection based on uncontrolled packets in virtual networks. *Computer Networks*. 2021;192:108017. DOI:10.1016/j.comnet.2021.108017
26. Ahmed M., Mahmoud A, Hu J. A Survey of Network Anomaly Detection Techniques. *Journal of Network and Computer Applications*. 2016;60:19–31. DOI:10.1016/j.jnca.2015.11.016
27. Kalush Yu.A., Loginov V.M. The Hurst index and its hidden properties. *Journal of Applied and Industrial Mathematics*. 2002;5(4(12)):29–37. (in Russ.) EDN:HZOKUN
28. Dvornikov S.-Jr., Dvornikov S. Detection Range Estimation of Small UAVs at a Given Probability of Their Identification. *Proceedings of Telecommunication Universities*. 2023;9(4):6–13. DOI:10.31854/1813-324X-2023-9-4-6-13. EDN:YLBWOS
29. Umbitaliev A.A., Dvornkov S.V., Okovs I.N., Ustinov A.A. Compression method graphic files using wavelet transform. *Voprosy radioelektroniki Seriia Tekhnika televideniia*. 2015;3:100–106. (in Russ.) EDN:UMOIVJ
30. Mallat S.G., Zhang Z. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*. 1993;41(12):3397–3415. DOI:10.1109/78.258082
31. Dvornikov S.V., Stepynin D.V., Dvornikov A.S., Bukareva A.P. Formation of signal feature vectors from the wavelet coefficients of their frame transformations. *Information Technologies*. 2013;5:46–49. EDN:QAQWUF
32. Novikov I.Ya., Protasov V.Yu., Skopina M.A. *Theory of Bursts*. Moscow: Fizmatlit Publ.; 2005. 616 p. (in Russ.)

Статья поступила в редакцию 21.07.2024; одобрена после рецензирования 06.09.2024; принята к публикации 30.09.2024.
The article was submitted 21.07.2024; approved after reviewing 06.09.2024; accepted for publication 30.09.2024.

Информация об авторах:

ЖДАНОВА Инна Михайловна
адъюнкт Военной академии связи имени Маршала Советского Союза С.М. Буденного
https://orcid.org/0009-0007-3997-2332

ДВОРНИКОВ Сергей Сергеевич
кандидат технических наук, доцент института радиотехники, электроники и связи (институт 2) Санкт-Петербургского государственного университета аэрокосмического приборостроения, научный сотрудник научно-исследовательского отдела Военной академии связи имени Маршала Советского Союза С.М. Буденного
https://orcid.org/0000-0001-7426-6475

ДВОРНИКОВ Сергей Викторович
доктор технических наук, профессор, профессор института радиотехники, электроники и связи (институт 2) Санкт-Петербургского государственного университета аэрокосмического приборостроения, профессор кафедры Военной академии связи имени Маршала Советского Союза С.М. Буденного
https://orcid.org/0000-0002-4889-0001

Дворников С.В. является членом редакционного совета журнала «Труды учебных заведений связи» с 2016 г., но не имеет никакого отношения к решению опубликовать эту статью. Статья прошла принятую в журнале процедуру рецензирования. Об иных конфликтах интересов авторы не заявляли.
Dvornikov S.V. has been a member of the journal "Proceedings of Telecommunication Universities" Editorial Council since 2016, but has nothing to do with the decision to publish this article. The article has passed the review procedure accepted in the journal. The authors have not declared any other conflicts of interest.