Выбор типа материнского вейвлета при фрактальном анализе в задаче обнаружения компьютерных атак

С.Ю. Рыбаков

Московский технический университет связи и информатики, г. Москва

Аннотация: Исследование статистических характеристик сетевого трафика позволяет обнаружить его фрактальные особенности и оценить, как фрактальная размерность изменяется в условиях компьютерных атак (КА). Эти исследования освещают взаимосвязь между атаками и динамическими изменениями фрактальной размерности, что позволяет более глубоко понять, как атаки воздействуют на структуру и поведение сетевого трафика. Такое понимание критично для разработки эффективных методов мониторинга и защиты сети от потенциальных угроз. Эти наблюдения обосновывают применение методов фрактального анализа, включая дискретный вейвлет-анализ, для выявления КА. В частности, возможен мониторинг фрактальной размерности телекоммуникационного трафика в реальном времени с отслеживанием её изменений. Тем не менее, выбор наиболее подходящего материнского вейвлета для кратномасштабного анализа остаётся недостаточно исследованным аспектом. В статье оценивается влияние выбора типа материнских вейвлетов на оценку показателя Херста и достоверность обнаружения КА. Рассматриваются следующие типы материнских вейвлетов: Хаар, Добеши, Симлет, Мейер и Коифлет. В рамках исследования проводилась экспериментальная оценка показателя Херста на наборе данных, который включает в себя атаку типа SYN-flood и нормальный сетевой трафик. Показано, что минимальный разброс оценки показателя Херста для трафика с атаками типа SYN-flood достигается при использовании в качестве материнского вейвлета Мейера при окне анализа более 10000 выборок и вейвлетами Хаара при окне анализа менее 10000 выборок.
Ключевые слова: материнский вейвлет, компьютерная атака, сетевой трафик, показатель Херста, вейвлет-анализ, фрактальная размерность.

Постановка задачи

Множество работ, посвященных анализу статистических характеристик сетевого трафика, компьютерных атак (КА) и аномалий в сети свидетельствуют о наличии фрактальных свойств (самоподобия) у данных процессов, а также об изменчивости показателей фрактальной размерности (ФР) во время переходных процессов [1-3]. Оценка степени самоподобия осуществляется через такие параметры, как ФР множества D (по Хаусдорфу) и показатель Херста H. Эти два показателя связаны между собой следующим образом: $D = 2 - H$. В большинстве исследований в области телекоммуникаций [2-4] для анализа ФР предпочтение отдается показателю Херста H, поскольку он отличается от множества D на фиксированную

константу. Далее в качестве оценки ФР нормального трафика и КА будет использоваться показатель Херста.

Рекуррентный алгоритм оценки ФР

Если процесс $\{x_i, i = \overline{1, N_0}\}$ характеризующий поведение сетевого трафика является долговременно зависимым процессом с показателем Херста H_m, то график зависимости $\log_2(\mu_{j,m})$ от j, называемый логарифмической диаграммой (LD), имеет линейный наклон $2\widehat{H_m}$–1. Масштабный показатель $\hat{\alpha}_m = (2\widehat{H_m} - 1)$ может быть получен путем оценки наклона графика функции $\log_2(\mu_{j,m})$ от j при каждом m-м положении окна анализа.

Взвешенную оценку $\hat{a}_m$ для α на интервале $[j_1; j_2]$ при m-ом положении окна анализа можно найти воспользовавшись методикой [2-4]:

$$\hat{\alpha}_m = \sum_j w_j\, y_{j,m}, \tag{1}$$

$$\hat{c}_m = \sum_j v_j\, y_{j,m}, \tag{2}$$

$$w_j = \frac{Sj - S_1}{(SS_2 - S_1^2)\sigma_j^2}, \tag{3}$$

$$y_{j,m} = \log_2(\mu_{j,m}) - g(j), \tag{4}$$

$$g(j) = \psi(n_j/2)ln2 - \log_2\left(\frac{n_j}{2}\right) = \Gamma'(n_j/2)/\left(\Gamma(n_j/2)ln2\right) - \log_2\left(\frac{n_j}{2}\right) \sim -\frac{1}{n_j ln2} \tag{5}$$

$$\sigma_j^2 = \frac{\xi\left(2, \frac{n_j}{2}\right)}{\ln^2 2} \sim \frac{2}{n_j \ln^2 2}, \tag{6}$$

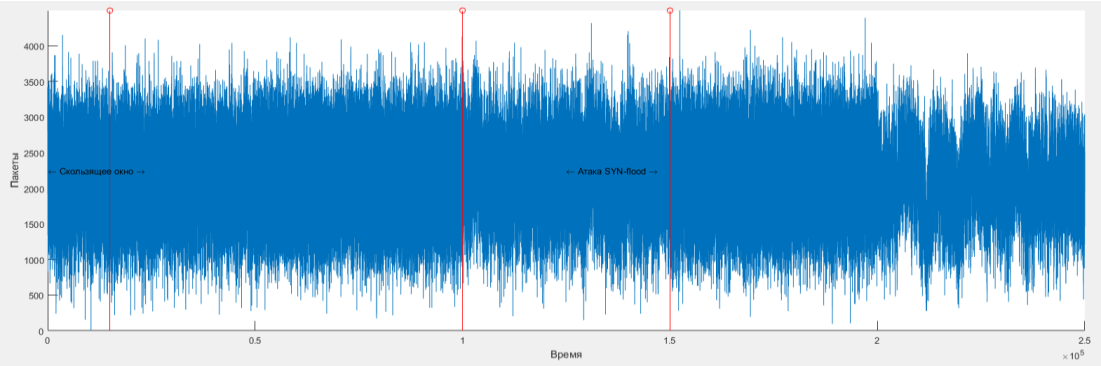
$$v_j = \frac{S_2 - jS_1}{(SS_2 - S_1^2)\sigma_j^2}, \tag{7}$$

$$S = \sum_{j=j_1}^{j_2} 1/\sigma_j^2,\ S_1 = \sum_{j=j_1}^{j_2} j/\sigma_j^2,\ S_2 = \sum_{j=j_1}^{j_2} j^2/\sigma_j^2, \tag{8}$$

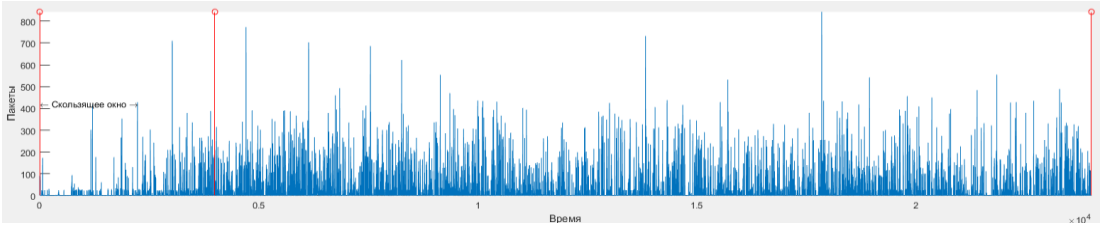
где $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt, \Gamma$ - гамма функция, Γ' - её производная, а $\xi(2,z) = \sum_0^\infty 1/(z+n)^2$ - обобщённая Зета функция Римана; $\psi(x) = \Gamma'(x)/\Gamma(x)$ - Psi функция (также имеющая название дигамма-функция); n_j – число коэффициентов - деталей на соответствующем уровне разложения (j).

Для анализа использовались следующие исходные данные, включающие сетевой трафик с атакой Syn-Flood (Рис. 1а) и обычный

(нормальный) сетевой трафик (Рис. 1б). Оценка проводилась с применением кратномасштабного анализа в пределах скользящего окна [5-7]. В процессе анализа использовались наиболее распространённые материнские вейвлеты, такие как Хаар, Симлет 4, Добеши 6, Мейера и Коифлет 4.



а



б

Рис. 1. - Исходные данные – реализация трафика: а) Syn-Flood, б) Нормальный трафик

Характеристики сетевого трафика представлены в Таблице 1.

Согласно описанному алгоритму (1) …(8), сетевой трафик внутри скользящего окна размером Δ проходит процесс вейвлет-декомпозиции, в ходе которого он разбивается на функции детализации d_i, охватывающие различные масштабы в пределах диапазона $[j_1, j_{max}]$. В данном контексте, j_{max} определяет предельное количество уровней разложения, а j_1 задаёт начальный уровень. В рамках данного исследования стартовое значение было установлено $j_1 = 3$, а максимально допустимый уровень разложения для анализируемого трафика ограничен значением $j_{max} = 10$.

Таблица 1.

Параметры анализируемого сетевого трафика

Трафик	Δ, сек х1000	Аномалия, сек х1000		Длительность реализации трафика, сек х1000
		Начало	Конец	
SYN-flood	2 - 9	100	150	200
Нормальный трафик	5 - 6	Аномалия отсутствует		24

На рисунке 2 представлена оценка показателя Херста после обработки данных, с применением разных типов материнских вейвлетов. Полученные оценки варьируются в диапазоне [0.5; 1], что еще раз доказывает наличие свойства самоподобия.

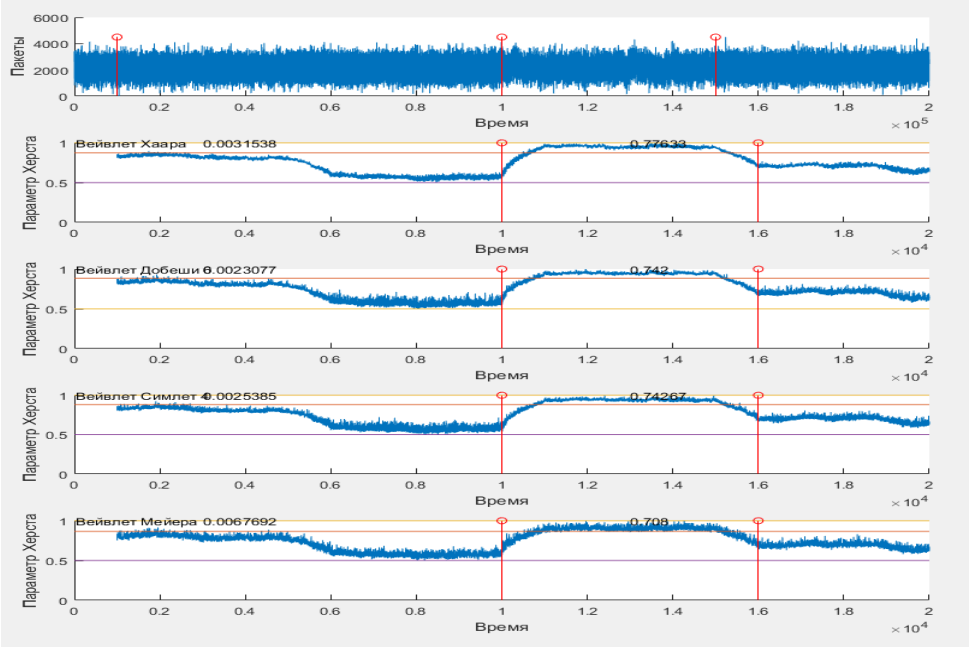


Рис. 2. - Оценка показателя Херста в скользящем окне при использовании 4-х типов вейвлетов для КА типа Syn-Flood

В исследованиях [4,5] установлено, что для оценки показателя Херста H ключевым свойством вейвлета является количество моментов, стремящихся к нулю. Является ли вейвлет симметричным или нет, ортогональным, полу- или би-ортогональным базисом, ни теоретической, ни значительной практической разницы не имеет. Однако оценка в задачах

обнаружения имеет свои особенности, связанные с постановкой задачи [8-10].Это связано с объемом анализируемой выборки.

Точность оценки показателя Херста

В рамках оценки показателя Херста была проведена серия вычислений стандартных отклонений (СКО) с использованием различных материнских вейвлетов. Эти расчёты охватывали все представленные реализации данных, что позволяло всесторонне исследовать их характеристики. Согласно данным, приведённым в Таблице 1, для каждой реализации сетевого трафика был установлен ряд параметров, включающий длительность "скользящего" окна и определены границы для атаки.

Для проведения анализа было подготовлено три реализации сетевого трафика: полная реализация сетевого трафика, трафик с КА SYN-flood, и нормальный трафик. Полученные результаты расчётов СКО для рассматриваемых материнских вейвлетов обобщены и представлены в Таблицах 2 и 3.

Таблица 2.

СКО для трафика с КА - SYN-flood

Тип материнского вейвлета	СКО $\sigma_{\hat{H}}$											
	Весь трафик, сек x1000				Трафик с КА, сек x1000				Нормальная часть трафика, сек x1000			
Размер окна	50	20	10	2	50	20	10	2	20	20	10	2
Хаар	0.12	0.14	0.13	0.10	0.06	**0.01**	**0.01**	**0.02**	0.12	0.09	0.10	
Симлет 4	0.12	0.13	0.13	0.09	0.06	0.02	0.03	0.04	0.12	0.09	0.09	0.07
Добеши 6	0.11	0.14	0.13	0.09	**0.06**	0.02	0.03	0.05	0.11	0.09	0.09	0.07
Мейер	**0.11**	**0.12**	**0.12**	**0.08**	**0.06**	0.02	0.02	0.03	**0.10**	**0.08**	**0.08**	**0.06**

Таблица 3.

СКО для нормального трафика

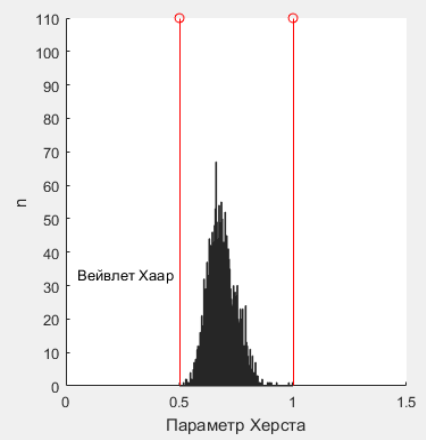
Тип материнского	СКО $\sigma_{\hat{H}}$	
	Весь трафик,	Нормальная часть трафика,

вейвлета	сек x1000				сек x1000			
Размер окна	6	4	1	0.5	6	4	1	0.5
Хаар	0.07	0.07	0.06	**0.03**	0.07	0.06	0.06	**0.02**
Симлет 4	0.05	0.04	0.06	0.06	0.04	0.04	0.05	0.06
Добеши 6	**0.04**	**0.03**	**0.04**	0.05	**0.03**	**0.03**	**0.04**	0.05
Мейер	0.05	0.06	0.06	0.05	0.05	0.06	0.05	0.05

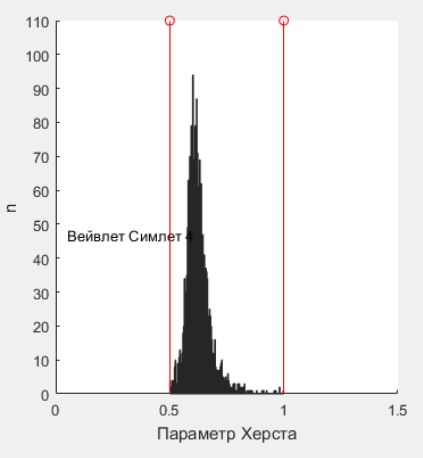
Анализ полученных результатов, изложенных в таблицах 2 и 3, показывают, что при размере окна свыше 1000 выборок вейвлеты Мейера и Добеши 6 демонстрируют наилучшие результаты для всех реализаций трафика. Однако при размере «скользящего» окна менее 1000 выборок лучшие результаты достигаются при использовании вейвлета Хаара.

Для оценки вероятности правильного обнаружения компьютерной атаки $P_{ПО}$ был рассчитан порог с использованием нормальной части графика и определялось количество точек, расположенных выше рассчитанного порога, который определялся показатель Херста.

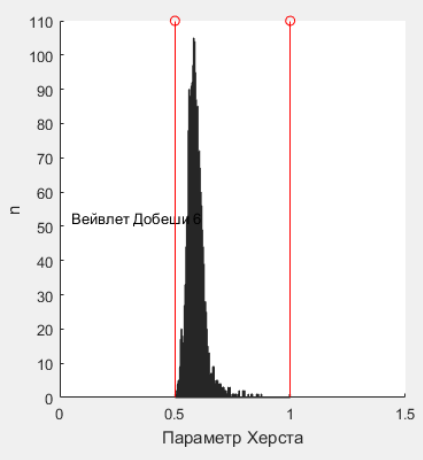
Методы оценки достоверности обнаружения КА иллюстрируют построенные гистограммы показателя Херста представленные на рисунках 3 и 4. Для определения достоверности полученных результатов были определены мат.ожидание оценки показателя Херста для аномальных и нормальных частей трафика, а также построены гистограммы анализируемых данных на рис.2. Средние значения параметра Херста для нормального трафика представлены в Таблице 4.



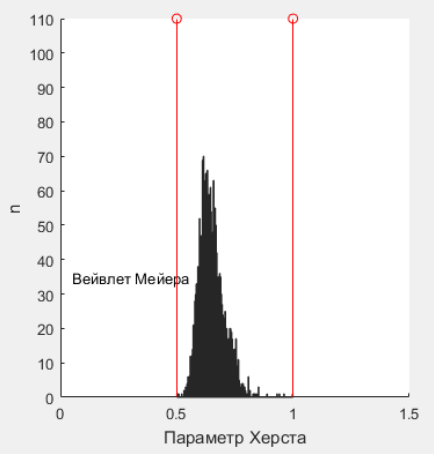
а



б

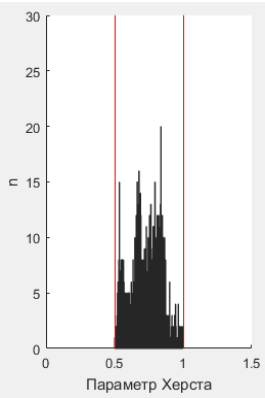


в

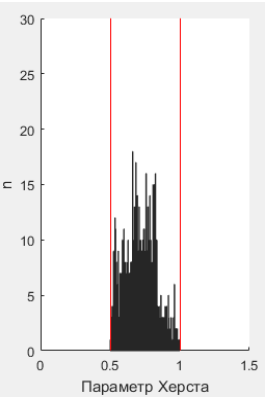


г

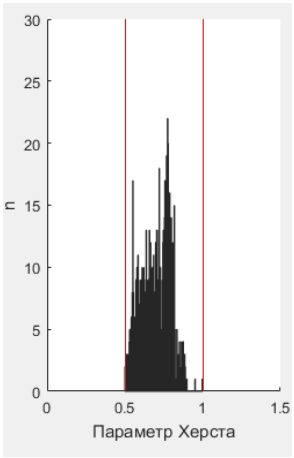
Рис. 3. – Распределение оценки показателя Херста нормального трафика для вейвлета: а) Хаар; б) Симлет 4; в) Добеши 6; г) Мейера.



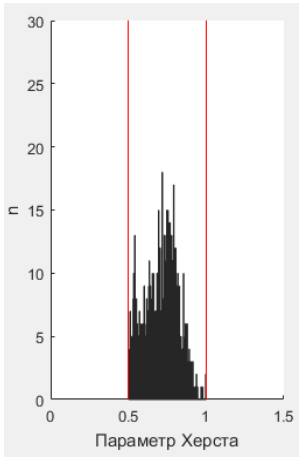
а



б



в



г

Рис. 4. - Распределение оценки показателя Херста трафика SYN-flood для вейвлета:а) Хаар; б) Симлет 4; в) Добеши 6; г) Мейера

На основе полученных гистограмм показателя Херста определены статистические характеристики оценки $\widehat{H}$ в «скользящем» окне $\Delta = 10000$ для различных типов вейвлетов. Результаты представлены в Таблице 4.

Таблица 4.

Статистические характеристики оценки $\widehat{H}$

Тип вейвлета	Нормальный трафик		Аномалия	
	$m_{аном}$	$\sigma_{норм}$	$m_{аном}$	$\sigma_{норм}$
Хаар	0.56	0.10	0.88	0.02
Добеши	0.57	0.10	0.87	0.01
Симлет	0.57	0.10	0.86	0.01
Мейер	0.57	0.09	0.84	0.02

Сравнив данные из Таблиц 2-4, можно сделать вывод, что минимальное значение стандартного отклонения при оценке ФР методом вейвлет разложения достигается при использовании вейвлетов Хаара и Мейера. Поэтому рекомендуется их использовать для обнаружения КА методом оценки ФР нормального трафика и КА.

Полученные результаты демонстрируют, что достоверность обнаружения компьютерной атаки достигается при использовании материнского вейвлета Мейера при размере «скользящего» окна, превышающем 10000. В случаях, когда размер «скользящего» окна составляет менее 10000, наилучшие результаты обеспечиваются с использованием вейвлета Хаара. Это свидетельствует о зависимости эффективности обнаружения КА от размеров применяемого окна и типа используемого вейвлета. В этом случае сочетание вероятности правильного обнаружения $P_{ПО}$ и вероятности ошибки первого рода $P_{ОШ1}$ оптимально.

Заключение

На основе проведенного анализа влияния различных типов материнских вейвлетов на оценку ФР сетевого трафика в условиях КА были сделаны следующие выводы.

Минимальный разброс оценки показателя Херста для трафика с КА SYN-flood достигается при использовании в качестве материнского вейвлета Мейера при окне анализа более 10000 выборок. При размере окна анализа менее 10000 выборок, целесообразно использовать вейвлеты Хаара

Литература

1. Шелухин О.И. Сетевые аномалии. Обнаружение, локализация, прогнозирование. – М.: Горячая линия – Телеком, 2019. – 448 с.

2. Abry P., Veitch D. Wavelet analysis of long-range dependent traffic, IEEE Trans. on Info. Theory, 1998. Vol. 44, № 1, pp. 2-15.

3. Abry P., Taqqu MS, Flandrin P., Veitch D. Wavelets for the analysis, estimation, and synthesis of scaling data, in Park K., Willinger W. (Eds.), Self - similar Network Traffic and Performance Evaluation, John Wiley & Sons. - 2000. P. 39-88.

4. Atayero A.A., Sheluhin O.I. Integrated Model for Information Communication Systems and Metworks. Desigtn and Development. IGI Global. USA, 2013. P. 462.

5. Котенко И.В., Саенко И.Б., Лаута О.С., Крибель А.М. Метод раннего обнаружения кибератак на основе интеграции фрактального анализа и статистических методов // Первая миля. 2021. № 6 (98). С. 64-71.

6. Карачанская Е. В., Соседова Н. И. Метод выявления аномалий сетевого трафика, основанный на его самоподобной структуре . Безопасность информационных технологий, IT Security, Том 26, № 1 (2019).], стр. 98-110

7. Перов Р.А., Лаута О.С., Крибель А.М., Федулов Ю.М. Комплексная методика обнаружения кибератак на основе интеграции фрактального анализа и статистических методов // Наукоемкие технологии в космических исследованиях Земли. 2022. Т. 14. № 2. С. 44-51.

8. Зегжда П.Д, Лаврова Д.С., Штыркина А.А.. Мультифрактальный анализ трафика магистральных сетей интернет для обнаружения атак отказа в обслуживании // Проблемы информационной безопасности. Компьютерные системы. – СПб., 2018. – № 2. – С. 48–58.

9. Семыкина Н.А., Садовникова Н.М. Особенности применения методов искусственного интеллекта при решении задачи мониторинга сетевого трафика с целью обнаружения атак // Инженерный вестник Дона. 2023. №4. URL: ivdon.ru/ru/magazine/archive/n4y2023/8367.

10. Табункова, М. П., Оганесян Л.Л. Технико-экономическое обоснование выбора оптимальных средств обнаружения атак (вторжений) для нужд центров мониторинга Российской Федерации // Инженерный вестник Дона. – 2023. – № 11. URL: ivdon.ru/ru/magazine/archive/n11y2023/8809.

References

1. Sheluhin O.I. Setevye anomalii. Obnaruzhenie, lokalizatsiya, prognozirovanie [Network anomalies. Detection, localization, prediction]. – M.: Goryachaya liniya – Telekom, 2019. – 448 p. ISBN 978-5-9912-0756-0.

2. Abry P., Veitch D. Wavelet analysis of long-range dependent traffic, IEEE Trans. on Info. Theory, 1998. Vol. 44, no. 1, P. 2-15.

3. Abry P., Taqqu MS, Flandrin P., Veitch D. Wavelets for the analysis, estimation, and synthesis of scaling data, in Park K., Willinger W. (Eds.), Self - similar Network Traffic and Performance Evaluation, John Wiley & Sons. 2000. P. 39-88.

4. Atayero A.A., Sheluhin O.I. Integrated Model for Information Communication Systems and Metworks. Desigtn and Development. IGI Global. USA, 2013. P. 462.

5. Kotenko I.V., Saenko I.B., Lauta O.S., Kribel' A.M. Pervaya milya. 2021. № 6 (98). pp. 64-71.

6. Karachanskaya E. V., Sosedova N. I. Bezopasnost' informacionnyh tekhnologij, IT Security, 2019. Vol. 26, № 1, pp. 98-110.

7. Perov R.A., Lauta O.S., Kribel A.M., Fedulov Yu.V. Naukoemkie texnologii v kosmicheskix issledovaniyax Zemli. 2022. Vol. 14. № 2. pp. 44-51.

8. Zegzhda P.D, Lavrova D.S., Shtyrkina A.A. Problemy informacionnoj bezopasnosti. Komp'yuternye sistemy. SPb. 2018. № 2. Pp. 48–58.

9. Semykina N.A., Sadovnikova N.M. Inzhenernyj vestnik Dona. 2023. №4. URL: ivdon.ru/ru/magazine/archive/n4y2023/8367.

10. Tabunkova M. P., Oganesyan L. L. Inzhenernyj vestnik Dona. 2023. № 11. URL: ivdon.ru/ru/magazine/archive/n11y2023/8809.

Дата поступления: 21.07.2024

Дата публикации: 12.10.2024