УДК 621.391

Анализ трафика демодулированного потока

Analysis of demodulated stream traffic

Жданова / Zhdanova I.
Инна Михайловна
(zhdanova.im@mail.ru)
ФГКВОУ ВО «Военная академия связи имени Маршала Советского Союза С. М. Буденного» МО РФ (ВАС им. С. М. Буденного), начальник отделения организации подготовки научно-педагогических кадров.
г. Санкт-Петербург

Дворников / Dvornikov S.
Сергей Викторович
(practicdsv@yandex.ru)
доктор технических наук, профессор.
ФГАОУ ВО «Санкт-Петербургский государственный университет аэрокосмического приборостроения» (ГУАП), профессор кафедры конструирования и технологий электронных и лазерных средств.
г. Санкт-Петербург

Симонова / Simonova K.
Карина Олеговна
(desire_8912@bk.ru)
ВАС им. С. М. Буденного, адъюнкт кафедры сетей связи и системы коммутации.
г. Санкт-Петербург

Васильева / Vasilyeva D.
Дина Владимировна
(profinst2guap@mail.ru.)
ГУАП, старший преподаватель кафедры радиотехнических систем.
г. Санкт-Петербург

Ключевые слова: статистический анализ данных – statistical data analysis; демодулированный трафик – demodulated traffic; коэффициенты ковариации и корреляции – covariance and correlation coefficients.

В статье представлены результаты обработки битовых последовательностей методами статистического анализа. Рассмотрены возможности статистических показателей для анализа структурных различий битовых потоков. Обосновано применение функции различия для выявления позиций, в которых элементы сравниваемых бинарных последовательностей имеют противоположные значения. Приведены результаты сравнительной оценки, характеризующей чувствительность статистических показателей, используемых при обработке бинарных данных.

The article presents the results of processing bit sequences using statistical analysis methods. The possibilities of statistical indicators for analyzing structural differences in bit streams are considered. The use of difference function for identifying positions in which the elements of the compared binary sequences have opposite values is substantiated. The results of a comparative assessment characterizing the sensitivity of statistical indicators used in processing binary data are presented.

Введение

Переход к IP-протоколам способствовал активному использованию интернет-контента в мобильных устройствах, в том числе в сетях передачи служебной информации [1, 2]. Вместе с тем наличие открытого сегмента в сетях служебного трафика предоставляет возможность несанкционированного доступа с целью нарушения работы сети [3].

Одним из признаков несанкционированного доступа является изменение структуры трафика, которая проявляется или в виде повторяющихся комбинаций битов, или длинных монотонных серий информационных символов. Для решения проблемы своевременного выявления аномалий трафика на практике применяют различные способы [4–6], в том числе основанные на результатах статистического анализа. Несмотря на то что данный подход достаточно хорошо изучен, доступный арсенал средств математической статистики открывает новые возможности по повышению его эффективности.

В настоящей работе представлены результаты исследования возможностей таких статистических показателей, как асимметрия и эксцесс.

Статистический анализ как инструмент оценки показателей трафика

Основной задачей любого анализа трафика является получение оценок характеризующих его показателей [7, 8]. Поскольку именно по результатам анализа полученных значений показателей предоставляется возможность проведения анализа состояния трафика.

В общем случае статистический анализ базируется на реализации процедур расчета как индивидуальных параметров, так и показателей результата сравнения [9, 10].

Статистический анализ является широко апробированным методом исследования данных, позволяющим получить показатели, характеризующие их местоположение, разброс местоположения и форму [4, 11].

К простейшим статистическим показателям относятся показатели местоположения, в частности, среднее значение и значение медианы, математическое ожидание и величина моды [4, 12, 13].

Как правило, средняя величина произвольного трафика s, упорядоченного на интервале значений i=0,…,k,…,l,…N, может быть рассчитана в граничных интервалах от k до l в соответствии со следующей формулой [13]:

$$s = \frac{1}{l-k}\sum_{i=k}^{l} s_i \text{, или } \bar{s} = \mathbf{E}[s]. \tag{1}$$

Медиана представляет собой среднее значение трафика (интервального ряда) при упорядочении элементов трафика s от наименьшего к наибольшему значению [13]:

$$M_e = s_{\text{Н}} + \Delta M_e \frac{\frac{\Sigma_f}{2} - S_{M_{e-1}}}{f_{M_e}}, \tag{2}$$

где $s_{\text{Н}}$ – нижняя граница медианного интервала; ΔM_e – величина медианного интервала; Σ_f – общее число единиц совокупности (сумма частот ряда); S_{Me-1} – сумма накопленных частот в интервалах до медианного интервала; S_{Me} – частота медианного интервала.

В свою очередь математическое ожидание представляет собой начальный момент первого порядка всех значений трафика $\mathbf{E}[s]$, и при рассмотрении каждого элемента из составляющих трафик как случайной величины характеризует его значения с позиций частости их проявления на заданном интервале [13, 14].

$$\mathbf{E}[\,s\,] = \sum_{i=k}^{l} s_i p_i\ , \tag{3}$$

где p_i – частость (вероятность) проявления s_i значения трафика.

При равномерном распределении значений вариационного ряда величина математического ожидания вырождается в среднее значение ряда, определяемое формулой (1).

Мода, как показатель статистической обработки, определяет одно или несколько наиболее часто проявляемых значений трафика s. Расчет моды как статистического показателя вариационного ряда производится по следующей формуле [13]:

$$M_0 = s_0 + \Delta h \frac{f_{M_0} - f_{M_{0-1}}}{(f_{M_0} - f_{M_{0-1}}) + (f_{M_0} + f_{M_{0-1}})}, \tag{4}$$

где s_0 – начало модального интервала; Δh – длина модального интервала; f_{M_0} – частота модального интервала; $f_{M_{0-1}}$ – частота интервала, предшествующего модальному интервалу; $f_{M_{0+1}}$ – частота интервала, последующего за модальным интервалом.

Учитывая, что в представленном исследовании рассматривается только битовая последовательность трафика, состоящая из бинарных значений, целесообразность использования показателей медианы и моды оправдана в том случае, когда анализ производится на уровне битовых комбинаций [15].

Для рассматриваемой ситуации интерес представляют такие показатели статистического анализа, как меры разброса местоположения [13].

Таковыми, в частности, являются:

– показатель дисперсии, характеризующий величину квадрата среднего значения:

$$\mathbf{D}[s] = \sum_{i=a}^{b} p_i\left(s_i - \mathbf{E}[s]\right)^2 = \frac{1}{b-a-1}\sum_{i=a}^{b}\left(s_i - \mathbf{E}[s]\right)^2; \tag{5}$$

– показатель среднего квадратического отклонения:

$$\mathbf{S}[s] = \sqrt{\mathbf{D}[s]} = \sqrt{\sum_{i=a}^{b} p_i\left(s_i - \mathbf{E}[s]\right)^2} = \sqrt{\frac{1}{b-a-1}\sum_{i=a}^{b}\left(s_i - \mathbf{E}[s]\right)^2}; \tag{6}$$

– показатели размаха распределения трафика, представляющие собой максимальное и минимальное значения дисперсии, рассчитанные в соответствии с формулой (5):

$$\max\left(\mathbf{D}[s]\right) \text{ и } \min\left(\mathbf{D}[s]\right); \tag{7}$$

– показатель разности третьего $s_{0,75}$ и первого $s_{0,25}$ квартилей, определяемый как интерквартильный размах:

$$\Delta s_{(0,75-0,25)} = s_{0,75} - s_{0,25}; \tag{8}$$

– показатель размаха вариаций, характеризующий различия между максимальным max[s] и минимальным min[s] значениями функции s:

$$R_{\text{var}} = \max[s] - \min[s]. \quad (9)$$

К особой категории следует отнести такие показатели разброса, как коэффициенты осцилляции и вариации.

Коэффициент осцилляции характеризует размах вариации значений функции, который описывается следующим выражением:

$$\rho_0 = \frac{R_{\text{var}}}{\mathbf{E}[s]} = \frac{\max[s] - \min[s]}{\frac{1}{N}\sum_{i=1}^{N} s_i}, \quad (10)$$

и коэффициент вариации:

$$V_{\text{var}} = \frac{s}{\mathbf{E}[s]}. \quad (11)$$

Кроме того, при рассмотрении указанных характеристик достаточно часто используют так называемый доверительный интервал, определяющий границы изменения показателей, в пределах которых значение математического ожидания обеспечивается с заданной вероятностью.

К мерам формы относят такие показатели, как коэффициенты асимметрии A_S и эксцесса (перекоса) A_A.

Коэффициент асимметрии A_S рассчитывается по следующей формуле [16]:

$$A_S = \frac{m_3}{s^3}, \quad (12)$$

где m_3 – третий центральный момент: $m_3 = \mathbf{E}[s - \mathbf{E}[s]]^3$.

$$A_S = \frac{N}{(N-1)(N-2)} \sum_{i=0}^{N-1} \left(\frac{s - \mathbf{E}[s]}{\mathbf{S}[s]} \right)^3. \quad (13)$$

Коэффициент эксцесса (перекоса) характеризует меру остроты пика распределения значений трафика. Расчет коэффициента эксцесса осуществляют по формуле [13, 16]:

$$A_A = \frac{m_4}{s^4 - 3}, \quad (14)$$

где m_4 – четвертый центральный момент $m_4 = \mathbf{E}[s - \mathbf{E}[s]]^4$.

На практике используют следующее выражение

$$A_A = \left[\frac{N(N+1)}{(N-1)(N-2)(N-3)} \sum_{i=0}^{N-1} \left(\frac{s - \mathbf{E}[s]}{\mathbf{S}[s]} \right)^4 \right] - \frac{3(N-1)^2}{(N-2)(N-3)}. \quad (15)$$

Рассмотренные выражения являются инструментом статистического анализа, используемого при обработке в том числе и бинарных потоков.

Инструментом сравнительного анализа выступают коэффициенты корреляции и ковариации.

В общем случае ковариация как корреляционный момент является мерой, характеризующей совместную зависимость двух случайных величин [17], которая рассчитывается как

$$\text{cov}(s1, s2) = \mathbf{E}[(s1 - \mathbf{E}[s1])(s2 - \mathbf{E}[s2])] = \\ = \frac{1}{N} \sum_{i=0}^{N-1} \left[(s1_i - \mathbf{E}[s1])(s2_i - \mathbf{E}[s2]) \right], \quad (16)$$

где $s1$ и $s2$ – сравниваемые случайные величины (анализируемые битовые потоки).

По своей сути ковариация характеризует направление изменения случайных величин. Применительно к анализу битовых потоков ковариация позволит сделать вывод о том, совместны ли происходящие в них изменения. Имеют ли они одну природу возникновения. Поэтому на практике важно не абсолютное

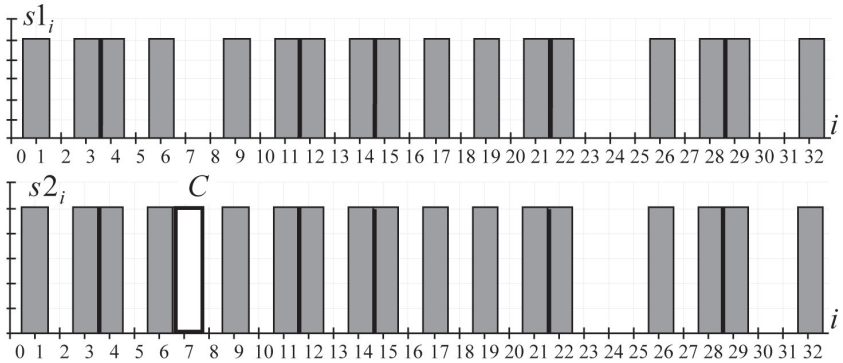

Рис. 1. Битовые последовательности с различным количеством «0» и «1»

значение ковариации, а знак. Если значение ковариации положительно, то это указывает на то, что изменения происходят в одном направлении, а если значение отрицательно – то в противоположных.

Коэффициент корреляции – это параметр, определяющий то, насколько «сильно» связаны случайные величины друг с другом [18]. В рамках рассматриваемой проблематики, как правило, используется коэффициент корреляции Пирсона [19], представляющий собой отношение ковариации величин к их среднеквадратическим значениям:

$$\mathrm{corr}(s1, s2) = \frac{\mathrm{cov}(s1, s2)}{\mathbf{S}[s1]\mathbf{S}[s2]}. \qquad (17)$$

Таким образом, значения ковариации и корреляции можно рассматривать в качестве дополнительных показателей статистического анализа, используемых для сравнительной оценки случайных величин [20].

Результаты моделирования

Для оценки возможностей представленного арсенала средств статистического анализа был проведен эксперимент по обработке битовых потоков $s1$ и $s2$ с близкой структурой, у которых на длительности 32 элементов различия состоят лишь в седьмой позиции: $s1_7=0$, а $s2_7=0$.

Структура битовых потоков представлена на рис. 1.

Битовые последовательности, представленные на рис. 1, отличаются лишь в пределах позиции № 7 (на рис. 1 отмечена буквой *C*).

В табл. 1 представлены статистические параметры, характеризующие последовательности $s1$ и $s2$, рассчитанные в соответствии с формулами (1), (2), (4), (5), (6), (12) и (14).

Анализ данных, представленных в табл. 1, позволяет заключить, что наиболее существенны различия у коэффициента асимметрии и эксцесса. Так, если разница в значении дисперсии составляет всего 1,2 %, то различия в значении математического ожидания у последовательностей $s1$ и $s2$ достигают 5,7 %. При этом различия коэффициента эксцесса всего 2,7 %, зато коэффициенты асимметрии различаются более, чем на 50 %.

Расчет коэффициента ковариации и корреляции Пирсона для последовательностей $s1$ и $s2$ позволяет сделать вывод, что направление изменений в анализируемых потоках совпадает. Их близость подтверждена и значением коэффициента корреляции Пирсона, равного $\mathrm{corr}(s1,s2)=0{,}939$.

Следует отметить, что статистические параметры характеризуют лишь общее соотношение «0» и «1» в обрабатываемой последовательности. Поэтому для потоков с одинаковым наполнением их «0» и «1» многие статистические параметры будут аналогичными. В качестве примера на рис. 2 показаны две последовательности $s3$ и $s2$, у которых одинаковое соотношение «0» и «1», но разное их положение (позиции *C* и *B* на рис. 2).

В табл. 2 представлены статистические параметры, характеризующие последовательности $s3$ и $s2$.

Данные табл. 2 позволяют заключить, что по своим статистическим характеристикам битовые последовательности $s3$ и $s2$ идентичны, что не соответствует действительности. Их различия проявляются только в расчетных значениях коэффициентов ковариации и корреляции $\mathrm{cov}(s3,s2)=0{,}215$, $\mathrm{corr}(s3,s2)=0{,}873$.

Важным моментом проведенного исследования является то, что результаты статистической обработки не содержат информации о структуре битовых потоков, что не позволяет рассматривать его в качестве оптимального инструмента анализа.

Альтернативным решением видится применение вида

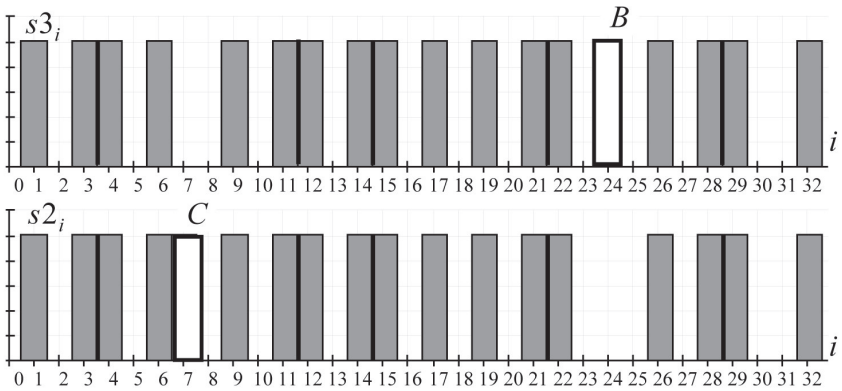

Рис. 2. Битовые последовательности с одинаковым количеством «0» и «1»

Таблица 1

Статистические характеристики анализируемых потоков s_1 и s_2

	E[*]	M_e	M_0	**D**[*]	**S**[*]	A_S	A_A
$s1$	0,531	1	1	0,257	0,507	-0,131	-2,119
$s2$	0,563	1	1	0,254	0,504	-0,265	-2,063

Таблица 2

Статистические характеристики анализируемых потоков $s3$ и $s2$

	E[*]	M_e	M_0	**D**[*]	**S**[*]	A_S	A_A
$s1$	0,563	1	1	0,254	0,504	-0,265	-2,063
$s2$	0,563	1	1	0,254	0,504	-0,265	-2,063

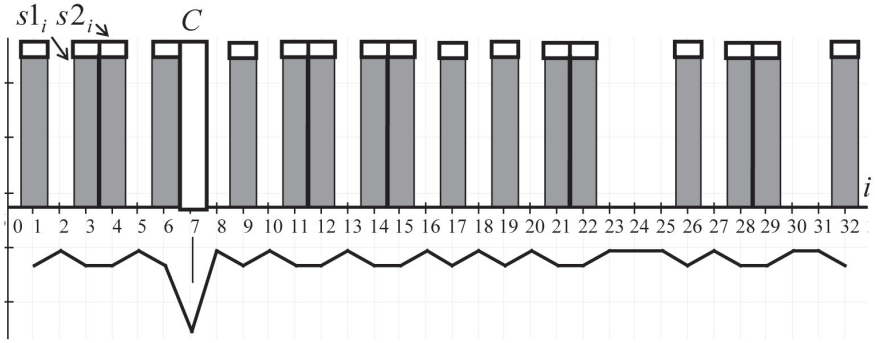

Рис. 3. Битовые последовательности s1, s2 и функция различий

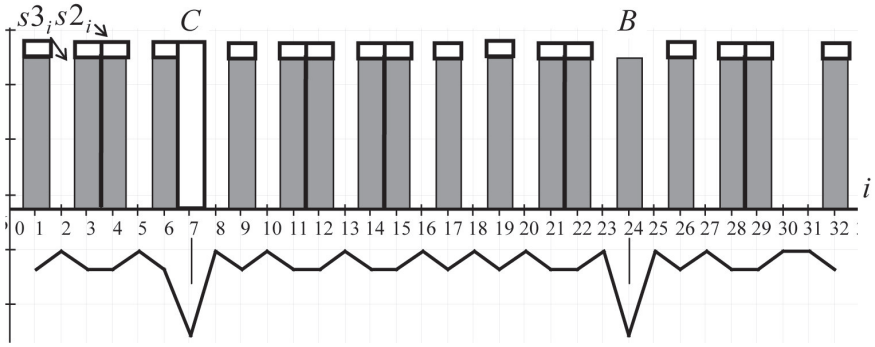

Рис. 4. Битовые последовательности s3, s2 и функция различий

$$K_i(s3, s2) = (s3_i - \mathbf{E}[s3])(s2_i - \mathbf{E}[s2]). \quad (18)$$

Функция $K(s3,s2)$ реализует поэлементную обработку последовательностей, что открывает возможность оценить уровень различий на каждой позиции.

На рис. 3 и 4 показаны битовые последовательности $s1$, $s2$ и $s3$, $s2$ с нанесенной функцией $K(*)$.

Полученный эффект объясняется следующим: поскольку элементы последовательностей имеют бинарную структуру, то, следовательно, при их различии (несовпадении) один из сомножителей будет иметь отрицательное значение. А значит, и само произведение будет отрицательно.

В том случае, если элементы совпадают, то и знаки сомножителей будут совпадать. При значениях «0» они будут отрицательны, а при значении «1» – положительны. Но в любом случае само произведение будет положительно.

Таким образом, функцию различий (18) можно рассматривать в качестве инструмента анализа структуры обрабатываемых битовых последовательностей.

Заключение

Полученные результаты аналитических возможностей статических способов обработки битовых последовательностей показали, что они способны выявлять лишь количественные различия в соотношении содержащихся в них информационных «1» и «0». При этом наибольшая чувствительность к структурным изменениям отмечена у коэффициента асимметрии.

Однако ни один из статистических показателей не способен выявить местоположение позиций [4], в которых наблюдаются различия. Коэффициенты ковариации и корреляции реагируют на структурные различия сравниваемых последовательностей [21–23], но также без указания мест изменений.

В такой ситуации предпочтительным видится использование разработанной функции различий, определяемой выражением (18), которая четко фиксирует позиции, в которых значения битов отличаются друг от друга.

Дальнейшее исследование авторы связывают со сравнением функции различий (18) и коэффициента Херста [4, 24, 25].

Литература

1. Шурыгин, С. А. Применение IP-протокола в сетях оперативно технологической связи / С.А. Шурыгин, Ю.В. Ширина // Автоматика, связь, информатика. – 2022. – № 5. – С. 30–32.

2. Корреляционный анализ параметров речевого трафика в IP-сети / К.А. Батенков, В.Ю. Головачев, О.Н. Катков [и др.] // Телекоммуникации. – 2020. – № 12. – С. 39–45.

3. Риск-модель атакуемого канала связи беспроводных сетей с применением технологий VPN / Н.М. Радько, Ю.С. Хирьянова, А.Н. Мокроусов, Е.А. Москалева // Информация и безопасность. – 2023. – Т. 26, № 2. – С. 191–202.

4. Жданова, И. М. Обнаружение аномалий трафика на основе обработки их фреймовых вейвлет-преобразований / И.М. Жданова, С.С. Дворников, С.В. Дворников // Труды учебных заведений связи. – 2024. – Т. 10, № 5. – С. 14–23.

5. Разработка модели обнаружения сетевых аномалий трафика в беспроводных распределенных самоорганизующихся сетях / Л.В. Легашев, Л.С. Гришина, Д.И. Парфенов, А.Ю. Жигалов // Научно-технический вестник информационных технологий, механики и оптики. – 2022. – Т. 22, № 4. – С. 699–707.

6. Жданова, И. М. Модель и условия возникновения аномалий в демодулированном трафике абонентских терминалов VSAT / И.М. Жданова, С.С. Дворников, С.В. Дворников // Системы управления, связи и безопасности. – 2025. – № 1. – С. 105–130.

7. Жданова, И. М. Модель возникновения аномалий в демодулированном трафике абонентских терминалов VSAT / И.М. Жданова, С.С. Дворников, С.В. Дворников // Морской вестник. – 2024. – № 4 (92). – С. 101–103.

8. Меньших, В. В. Обнаружение сетевых аномалий в трафике протокола удаленных рабочих столов в частотной области / В.В. Меньших, А.Ю. Телков // Вестник Воронежского института МВД России. – 2020. – № 1. – С. 48–56.

9. Поздняк, И. С. Выявление DOS-атак с помощью анализа статистических характеристик трафика / И.С. Поздняк, А.И. Плаван // Инфокоммуникационные технологии. – 2021. – Т. 19, № 1. – С. 73–80.

10. Черниговский, А. В. Статистический анализ сетевого трафика / А.В. Черниговский, М.В. Кривов // Математические методы в технике и технологиях – ММТТ. – 2019. – Т. 1. – С. 64–67.

11. Симаков, Д. В. Анализ статистических характеристик интернет трафика в магистральном канале / Д.В. Симаков, А.А. Кучин // T-Comm: Телекоммуникации и транспорт. – 2015. – Т. 9, № 5. – С. 31–35.

12. Веселова, В. А. Подход к обнаружению аномалий в самоподобном сетевом трафике / В.А. Веселова, В.С. Коломойцев // Надежность. – 2023. – Т. 23, № 2. – С. 57–63.

13. Андерсон, Т. Введение в многомерный статистический анализ / Т. Андерсон. – Москва : Физматгиз, 1963. – 500 с.

14. Дворников, С. В. Аппарат анализа частотного ресурса для режима псевдослучайной перестройки рабочей частоты / С.В. Дворников, С.С. Дворников, А.В. Пшеничников // Информационно-управляющие системы. – 2019. – № 4 (101). – С. 62–68.

15. Формирование векторов признаков для систем видеонаблюдения / Д.В. Васильева, С.С. Дворников, Ю.Е. Толстуха [и др.] // Вопросы радиоэлектроники. Серия: Техника телевидения. – 2023. – № 4. – С. 62–68.

16. Сикан, А. В. Оценка стандартных ошибок выборочных коэффициентов вариации и асимметрии при анализе гидрологических рядов / А.В. Сикан, Д.А. Щеглов // Гидрометеорология и экология. – 2024. – № 77. – С. 645–660.

17. Дворников, С. В. Модифицированные импульсные последовательности на основе кодов Баркера / С.В. Двор-

ников, С.С. Дворников, Е.В. Марков // Труды учебных заведений связи. – 2022. – Т. 8, № 1. – С. 8–14.

18. Dvornikov, S. S. SSB signals with controlled pilot level / S.S. Dvornikov, K.D. Zheglov, S.V. Dvornikov // T-Comm. – 2023. – Vol. 17, No. 3. – P. 41–47.

19. Svetunkov, S. G. On the problem of nominal data correlation / S.G. Svetunkov // Technoeconomics. – 2023. – Vol. 2, No. 2 (5). – P. 15–35.

20. Попукайло, В. С. Исследование линейной корреляционной связи в парных выборках малого объема / В.С. Попукайло // Технология и конструирование в электронной аппаратуре. – 2016. – № 1. – С. 27–32.

21. Новиков, В. А. Определение статистической связи между зависимыми нормально распределенными случайными величинами / В.А. Новиков, А.А. Козлов // Вестник воздушно-космической обороны. – 2023. – № 2 (38). – С. 43–45.

22. Дегтярев, А. М. Учет корреляции реперных данных при определении погрешности расчета с применением метода максимального правдоподобия / А.М. Дегтярев, О.А. Серянина // Атомная энергия. – 2023. – Т. 135, № 3–4. – С. 165–169.

23. Жердев, Г. М. Анализ ковариационных данных для урана-235 / Г.М. Жердев, Т.С. Кислицына, М.Н. Николаев // Вопросы атомной науки и техники. Серия: Физика ядерных реакторов. – 2020. – № 1. – С. 8–19.

24. Тянь, Ч. Х. Показатель Херста (Коэффициент Херста) / Ч.Х. Тянь // Современные аспекты экономики. – 2019. – № 5 (261). – С. 120–137.

25. Одоевский, С. М. Обработка и учет статистических характеристик мультимедийного трафика / С.М. Одоевский, М.И. Рафальская, И.В. Степанец // Известия Тульского государственного университета. Технические науки. – 2022. – № 2. – С. 385–390.