УДК 004.056:519.24
doi:10.21685/2072-3059-2021-3-3

# Корреляционный тест на случайность кодовой последовательности, обладающий почти линейной вычислительной сложностью

**В. И. Волчихин[1], А. И. Иванов[2], А. П. Иванов[3], А. В. Строков[4]**

[1,3]Пензенский государственный университет, Пенза, Россия
[2]Пензенский научно-исследовательский электротехнический институт, Пенза, Россия
[4]ООО «Организационно-технологические решения 2000», Москва, Россия
[1]president@pnzgu.ru, [2]ivan@pniei.penza.ru, [3]ap_ivanov@pnzgu.ru, [4]strokov.alexey@otr.ru

**Аннотация.** *Актуальность и цели*. Целью данной статьи является создание и исследование нового теста качества случайной кодовой последовательности, имеющего низкую вычислительную сложность. *Материалы и методы*. Предложено, по аналогии с классическим вычислением автокорреляционной функции шума с континуальными отсчетами, использовать автокорреляционную функцию для дискретных шумов. Предложено почти «белый» шум получать от программного генератора псевдослучайных чисел. Образцы «окрашенного» шума предложено получать скользящей сверткой восьми рядом стоящих отсчетов «белого» шума без их взвешивания. *Результаты.* Сумма модулей семи первых отсчетов автокорреляционной функции анализируемого шума является мощным критерием проверки гипотезы независимости дискретных данных выборкой в 256 бит. Этот критерий имеет низкую, почти линейную вычислительную сложность и одновременно дает высокий уровень линейной разделимости зависимых и независимых данных. Мощность этого нового статистического критерия выше мощности аналогичных статистических критериев, построенных на вычислении расстояний Хэмминга. *Выводы.* Предложенный статистический критерий может быть использован при тестировании кодов биометрической аутентификации в малогабаритной доверенной вычислительной среде с низкой разрядностью, низким потреблением энергии, малым объемом долговременной и оперативной памяти.

**Ключевые слова**: оценка качества белого шума, статистические критерии проверки гипотезы независимости, низкая вычислительная сложность

**Для цитирования**: Волчихин В. И., Иванов А. И., Иванов А. П., Строков А. В. Корреляционный тест на случайность кодовой последовательности, обладающий почти линейной вычислительной сложностью // Известия высших учебных заведений. Поволжский регион. Технические науки. 2021. № 3. С. 25–33. doi:10.21685/2072-3059-2021-3-3

# Correlation test for code sequence randomness with almost linear computational complexity

**V.I. Volchikhin[1], A.I. Ivanov[2], A.P. Ivanov[3], A.V. Strokov[4]**

[1,3]Penza State University, Penza, Russia
[2]Penza Research Institute of Electrical Engineering, Penza, Russia
[4]Organizational & Technological Solutions 2000 LLC, Moscow, Russia
[1]president@pnzgu.ru, [2]ivan@pniei.penza.ru, [3]ap_ivanov@pnzgu.ru, [4]strokov.alexey@otr.ru

**Abstract.** *Background*. The purpose of this article is to create and study a new test of random code sequence quality with low computational complexity. *Materials and methods*. It is proposed, by analogy with the classical calculation of the autocorrelation function of noise with continual samples, to use the autocorrelation function for discrete noise. It is proposed to receive almost "white" noise from a software pseudo-random number generator. Samples of "colored" noise are proposed to be obtained by a sliding convolution of eight adjacent readings of "white" noise without weighing them. *Results.* The sum of the modules of the first 7 samples of the autocorrelation function of the analyzed noise is a powerful criterion for testing the hypothesis of independence of discrete data with a sample of 256 bits. This criterion has a low almost linear computational complexity and at the same time gives a high level of linear separability of dependent and independent data. The power of this new statistical test is higher than the power of similar statistical tests based on calculating Hamming distances. *Conclusions.* The proposed statistical criterion can be used when testing biometric authentication codes in a small-sized trusted computing environment with low bit depth, low power consumption, and a small amount of long-term and random access memory.

**Keywords**: evaluation of white noise quality, statistical criteria for testing the hypothesis of independence, low computational complexity

**For citation**: Volchikhin V.I., Ivanov A.I., Ivanov A.P., Strokov A.V. Correlation test for code sequence randomness with almost linear computational complexity. *Izvestiya vysshikh uchebnykh zavedeniy. Povolzhskiy region. Tekhnicheskie nauki = University proceedings. Volga region. Engineering sciences.* 2021;(3):25–33. (In Russ.). doi:10.21685/2072-3059-2021-3-3

## Введение

Начиная с конца прошлого века во всех странах активно ведутся работы по использованию объединения преимуществ биометрии и криптографии. Биометрия удобна, но ненадежна, криптография сильна, но требует бережного обращения с длинными ключами, которые человек не может запомнить.

В связи с этим в России разработан пакет из 7 национальных стандартов, регламентирующий требования к нейросетевым преобразователям особенностей биометрии в код личного криптографического ключа пользователя. Очевидным является также то, что действительно высокий уровень информационной безопасности может быть обеспечен только в том случае, если все биометрические и криптографические преобразования выполняются в доверенной вычислительной среде. Стоимость такой доверенной вычислительной среды должна быть низка, а ее применение должно быть массовым.

В идеале утеря, кража и поломка доверенной вычислительной среды пользователя не должны приводить к катастрофам для пользователя. Одним из важнейших функционалов такой доверенной вычислительной среды является поддержка синтеза качественных криптографических ключей пользователя из шумоподобной (неповторяемой по воле пользователя) компоненте его биометрических данных [1, 2]. Эту ситуацию поясняет рис. 1, где отображены два рукописных слова «*Пенза*», написанных почерком пользователя.

В силу того что человек не может физически точно повторить свои движения, нестабильная, неповторяемая, случайная компонента достаточно легко выделима из биометрических данных обучения его нейросетевого преобразователя биометрия-код.

Для того чтобы усилить шумовые свойства нестабильной части биометрических данных, над ними необходимо выполнить хэширование. В слу-

чае если доверенная вычислительная среда («БиоТокен») построена на базе достаточно мощного процессора, то хэширование может быть криптографическим [3]. Если использован процессор с малым потреблением энергии, с малым числом разрядов, с малым объемом оперативной и постоянной памяти (например, процессор SIM-карты, идентификационной карты), то приходится использовать ослабленные процедуры защиты информации. В связи с этим на рис. 1 показано использование в качестве простейшего хэширования процедура CRC-4 подсчета контрольной суммы [4].

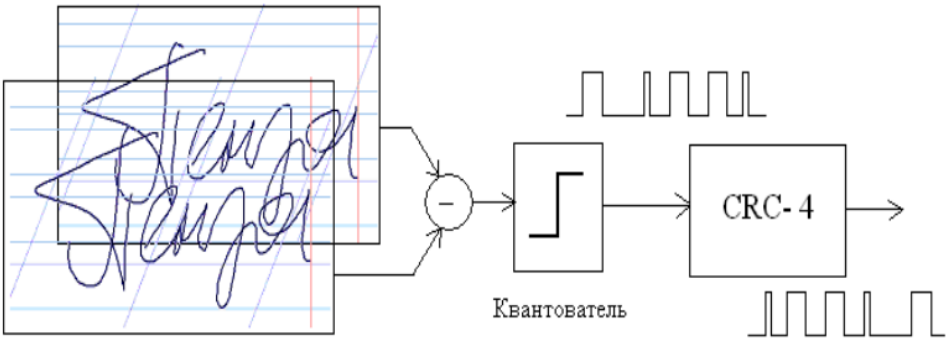


Рис. 1. Выделение нестабильной части биометрических данных, их оцифровывание и хэширование

Таким образом, из нестабильной части биометрических данных пользователя могут быть получены достаточно большие объемы нестабильных, неповторяющихся компонент биометрических данных, хэширование которых (криптографическое или некриптографическое) способно давать «сырой» ключевой материал достаточного объема. Далее для того чтобы получить ключи достаточно высокого качества, необходимо выполнить внутри доверенной вычислительной среды их тестирование [5]. Например, для этой цели могут быть использованы тесты NIST (Национального института стандартизации и технологий США) либо иные популярные тесты. К сожалению, большинство известных тестов ориентировано на применение мощных вычислителей, так как они обладают экспоненциальной вычислительной сложностью. Естественно, что такие тесты нельзя применять в доверенной вычислительной среде с малопотребляющим процессором.

В связи с эти начали активно развиваться системы тестирования с почти линейной вычислительной сложностью [5–8], построенные на понижении требований к вычислительным ресурсам за счет перехода от анализа обычных кодов к расстояниям Хэмминга. При таком подходе снимается проблема вычислительной сложности, реализации тестов, однако остро встает вопрос о том, насколько была утрачена мощность самих процедур тестирования.

В данной статье мы попытаемся показать, что наряду с системой тестов в пространствах расстояний Хэмминга могут существовать и другие тесты, с одной стороны, обладающие низкой вычислительной сложностью, а с другой – демонстрирующие существенно более высокую мощность разделения коррелированных и некоррелированных данных.

### Линейное связывание случайных данных

Очевидно, что достаточно случайные данные могут быть получены от любого из программных генераторов. Эти данные в первом приближении мы можем рассматривать как эталонные данные «белого» шума. Для сравнения с ними желательно иметь некоторый эталон зависимых данных. Его можно легко получить, воспользовавшись скользящим окном шириной в 8 случайных отсчетов:

$$\tilde{x}_j = \frac{1}{8}\sum_{i=0}^{7} x_{j+i} \,. \tag{1}$$

Если мы вычислим коэффициенты автокорреляции

$$r(x_j, x_{j+1}) = \frac{E\left\{(x_j - E(x)\cdot(x_{j+i} - E(x)\right\}}{\sigma(x)\cdot\sigma(x)} \tag{2}$$

для эталонного шума и эталонных зависимых данных, то мы получим соотношения, хорошо соответствующие классике. Данные имитационного моделирования для последовательностей в 256 бит представлены на рис. 2.

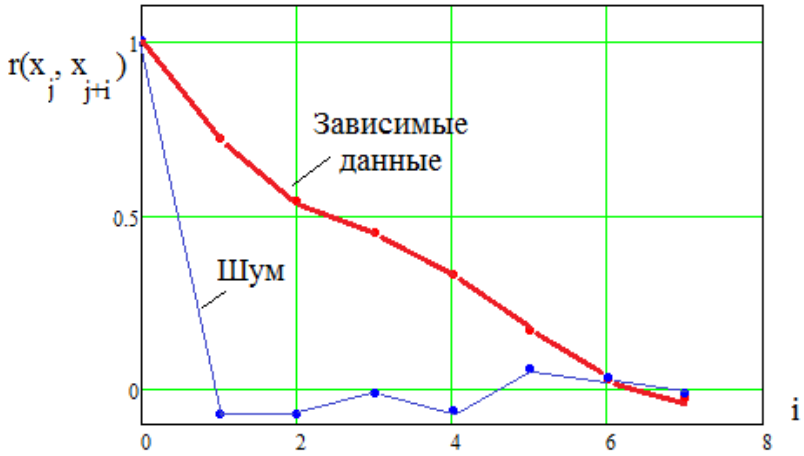


Рис. 2. Хорошее совпадение значений коэффициентов автокорреляции шума (данные программного генератора случайных чисел) и зависимых данных

Заметим, что наклон автокорреляционной функции зависимых данных всегда будет связан с шириной окна сглаживания (1). Чем больше ширина окна, тем медленнее будет снижение функции автокорреляции зависимых данных. Автокорреляционная функция шума при любом сдвиге должна давать малые случайные значения, изменяющиеся вблизи нулевого состояния.

### Плохая линейная разделимость зависимых и независимых данных при их анализе в пространстве расстояний Хэмминга

Одной из проблем тестирования больших нейронных сетей с 256 выходами является необходимость применения очень больших тестовых баз. Проблема усугубляется тем, что сбор баз «Чужих» биометрических образов зако-

нодательно ограничен практически во всех развитых странах. Без согласия на обработку персональных данных нельзя собирать и хранить «Чужие» биометрические образы.

Выход из этого технологического тупика дает отечественный стандарт (ГОСТ Р 52633.3–2011. Защита информации. Техника защиты информации. Тестирование стойкости средств высоконадежной биометрической защиты к атакам подбора). Этот стандарт рекомендует отказаться от статистического анализа большого числа выходных кодовых состояний. Вместо прямого статистического анализа поля возможных кодовых состояний стандарт рекомендует перейти в пространство расстояний Хэмминга до кода образа «Свой», на который была обучена нейросеть:

$$"h" = \sum_{i=1}^{256} ("c_i") \oplus ("x_i"), \tag{3}$$

где $"c_i"$ – состояние $i$-го разряда кода «Свой»; $"x_i"$ – состояние $i$-го разряда, анализируемого кода «Чужой»; $\oplus$ – операция сложения по модулю 2 бинарных разрядов двоичного числа.

В результате вычисления свертки Хэмминга (3) при суммировании большого числа случайных переменных происходит нормализация задачи и ее экспоненциальное упрощение [9, 10]. На рис. 3 представлены распределения расстояний Хэмминга для шума независимых данных и случайных данных с существенной зависимостью.

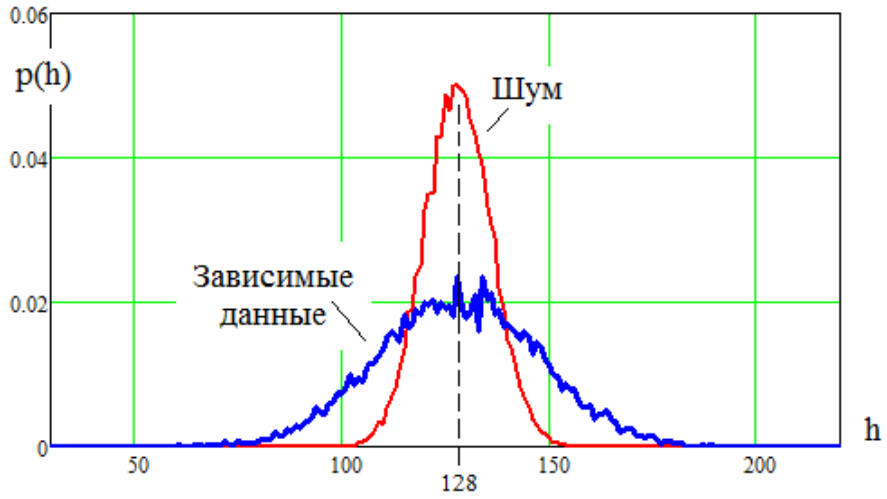


Рис. 3. Отсутствие линейной разделимости шума
и зависимых данных в пространстве расстояний Хэмминга

Очевидно, что для вычисления математического ожидания и стандартного отклонения вполне достаточно 30 опытов. То есть для достаточно надежного быстрого тестирования нейросети в рамках гипотезы нормального распределения расстояний Хэмминга вполне достаточно малых тестовых выборок. При этом вероятность ошибок второго рода (принятия «Чужого» как пример образа «Свой») оценивается по следующей формуле:

$$P_2 \approx \frac{1}{\sigma(h)\cdot\sqrt{2\cdot\pi}}\int_0^1 \exp\left\{\frac{-\left(u-E(h)\right)^2}{2\cdot\left(\sigma(h)\right)^2}\right\}du\,. \qquad (4)$$

Формально при попадании расстояния Хэмминга в точку «$h$» ≡ «0» злоумышленник угадывает код «Свой» подбором. Если мы имеем дело с настоящим «белым» шумом, то угадывание кода «Свой» маловероятно. В этом случае математическое ожидание равно $E(h)$ = 128 бит, а стандартное отклонение должно составлять $\sigma(h)$ = 8 бит. В теории идеальный «белый» шум должен иметь энтропию 256 бит. Формально многомерная энтропия зависимых кодов «Чужой» может быть оценена следующим образом:

$$H("x_1, x_2, ...., x_{256}") \approx -\log_2(P_2)\,. \qquad (5)$$

Чем больше корреляционная сцепленность кодов «Чужой», тем больше стандартное отклонение распределений расстояний Хэмминга и тем меньше их энтропия (5). С одной стороны, на этом могут быть построены соответствующие статистические критерии проверки гипотезы независимости данных [6–8], а с другой – плотность распределения значений расстояний Хэмминга (см. рис. 3) находится внутри плотностей распределения значений зависимых данных. Мы наблюдаем плохую линейную разделимость данных «белого» шума и зависимых данных. Это косвенно свидетельствует об относительно низкой мощности системы статистических критериев, построенных на вычислении расстояний Хэмминга по разным модулям [6–8].

### Модуль-автокорреляционный статистический критерий проверки гипотезы независимости дискретных выборок длинной в 256 бит

Заметим, что автокорреляционная функция случайного шума мала (см. рис 2), она значительно меньше значений автокорреляционной функции зависимых данных. В связи с этим обстоятельством мы можем построить модуль-автокорреляционный статистический критерий оценки корреляционной сцепленности длинных кодов:

$$\Sigma|r| = \sum_{i=1}^{7}\left|\frac{E\left\{(x_j - E(x)\cdot(x_{j+i} - E(x)\right\}}{\sigma(x)\cdot\sigma(x)}\right|. \qquad (6)$$

На рис. 4 приведены результаты численного моделирования эффективности нового статистического критерия (6).

Из рис. 4 видно, что в пространстве нового статистического критерия распределение «белого» шума и распределение зависимых данных оказались легко разделимы. Это говорит о значительном повышении мощности нового статистического критерия в сравнении с ранее построенными статистическими критериями пространства множества расстояний Хэмминга.

Если предположить, что квадратичным решающим правилом удастся разделить данные белого шума и зависимые данные (рис. 3) с вероятностями ошибок первого и второго рода $P_1 \approx P_2 \approx P_{EE} \approx 0{,}2$, то мы получим примерно 1000-кратный выигрыш по мощности для нового статистического критерия $P_1 \approx P_2 \approx P_{EE} \approx 0{,}0002$. Столь высокий выигрыш обусловлен ситуацией, отображенной на рис. 4, где расстояние от центра распределения «белого» шума

до порога принятия решения близко к четырем стандартным отклонениям распределения «белого» шума.

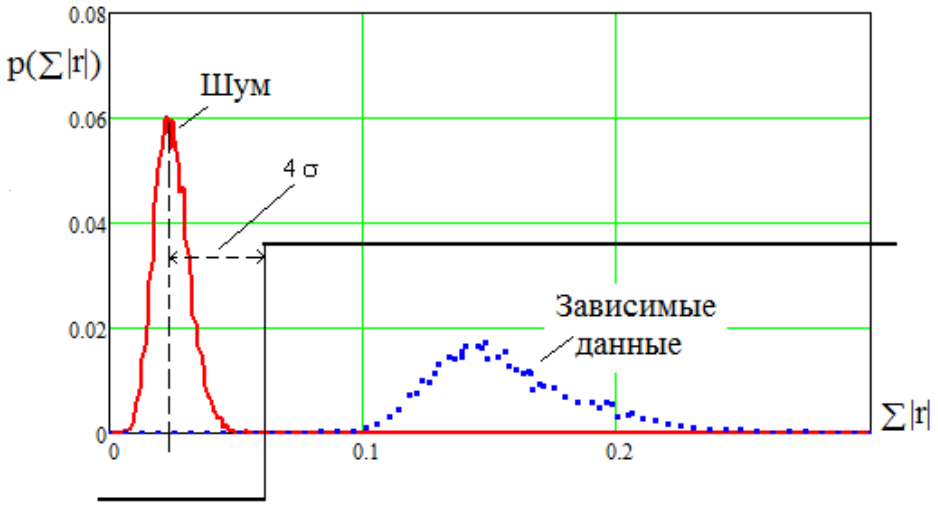


Рис. 4. Высокая линейная разделимость шума и зависимых данных в пространстве автокорреляционных сверток длинных кодов

## Заключение

Задача вычисления автокорреляционных функций имеет квадратичную сложность, и в этом отношении новый класс статистических критериев хуже ранее исследованных критериев, построенных в пространствах сверток Хэмминга [8]. Тем не менее высокая потенциальная мощность нового класса статистических критериев дает возможность заменить одним сильным критерием десятки более слабых статистических критериев. В этом отношении сегодня нельзя сделать однозначный вывод о предпочтении статистических критериев, ориентируясь только на их возможность реализации в доверенной вычислительной среде с мало потребляющим процессором. Необходимо продолжить исследования, оптимизирую потребляемую вычислительную мощность, а также число применяемых статистических критериев и итоговое качество их совместного применения.

## Список литературы

1. Строков А. В., Казанцев Е. И. Программное средство создания действительно случайных криптографических ключей из неоднозначной компоненты биометрических данных динамики рукописного почерка пользователя // Безопасность информационных технологий : труды I Всерос. науч.-техн. конф. Пенза : Изд-во ПГУ, 2019. С. 139–143.
2. Юнин А. П., Иванов А. И., Строков А. В., Махсудов С. Р. Нейросетевое обобщение трех стандартных тестов контроля качества «белого шума», получаемого хешированием случайной части биометрических данных // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. Пенза : Изд-во ПГУ, 2020. С. 49–56.
3. Криптографическая хеш-функция. URL: https://ru.wikipedia.org/wiki/Криптографическая_хеш-функция (дата обращения: 02.04.2021).
4. Циклический избыточный код. URL: https://ru.wikipedia.org/wiki/Циклический избыточный_код (дата обращения: 02.04.2021).

5. Григорьев А. Ю. Методы тестирования генераторов случайных и псевдослучайных последовательностей // Ученые записки УлГУ. Серия: Математика и информационные технологии. 2017. № 1. С. 22–28.
6. Волчихин В. И., Иванов А. И., Юнин А. П., Малыгина Е. А. Многомерный портрет цифровых последовательностей идеального «белого шума» в свертках Хэмминга // Известия высших учебных заведений. Поволжский регион. Технические науки. 2017. № 4. С. 4–13. doi:10.21685/2072-3059-2017-4-1
7. Юнин А. П., Иванов А. И., Ратников К. А., Кольчугина Е. А. Оценка качества «белого» шума: реализация теста «стаи обезьян» через множество сверток Хэмминга, построенных для разных систем счисления // Известия высших учебных заведений. Поволжский регион. Технические науки. 2018. № 4. С. 54–64. doi:10.21685/2072-3059-2018-4-5.
8. Иванов А. И., Юнин А. П. Эмбрион искусственного интеллекта: компактная нейросетевая проверка качества случайных последовательностей, полученных из биометрических данных : препринт. Пенза : Изд-во ПГУ, 2021. 68 с.
9. Иванов А. И., Кубасов И. А., Самокутяев А. М. Тестирование больших нейронных сетей на малых выборках // Надежность и качество сложных систем. 2020. № 3. С. 72–79. doi:10.21685/2307-4205-2020-3-9
10. Иванов А. И. Искусственный интеллект высокого доверия. Ускорение вычислений и экономия памяти при тестировании больших сетей искусственных нейронов на малых выборках // Системы безопасности. 2020. № 5. С. 60–62.

## References

1. Strokov A.V., Kazantsev E.I. A software tool for generating truly random cryptographic keys from an ambiguous biometric component of a user's handwriting dynamics. *Bezopasnost' in-formatsionnykh tekhnologiy: trudy I Vseros. nauch.-tekhn. konf. = Information technology security: proceedings of the 1st All-Russian scientific and engineering conference*. Penza: Izd-vo PGU, 2019:139–143. (In Russ.)
2. Yunin A.P., Ivanov A.I., Strokov A.V., Makhsudov C.P. Neural network generalization of three standard tests of quality control of “white noise” obtained by hashing a random part of biometric data. *Bezopasnost' informatsionnykh tekhnologiy: sb. nauch. st. po materialam II Vseros. nauch.-tekhn. konf. = Information technology security: proceedings of the 2nd All-Russian scientific and engineering conference*. Penza: Izd-vo PGU, 2020:49–56. (In Russ.)
3. *Kriptograficheskaya khesh-funktsiya = Cryptographic hash function*. (In Russ.). Available at: https://ru.wikipedia.org/wiki/Kriptografi-cheskaya_khesh-funktsiya (accessed 02.04.2021).
4. *Tsiklicheskiy izbytochnyy kod = Cyclic redundancy code*. (In Russ.). Available at: https://ru.wikipedia.org/wiki/Tsiklicheskiy izbytochnyy_kod (accessed 02.04.2021).
5. Grigor'ev A.Yu. Testing methods for generators of random and pseudo-random sequences. *Uchenye zapiski UlGU. Seriya: Matematika i informatsionnye tekhnologii = Proceedings of Ulyanovsk State University. Series: Mathematics and informational technologies*. 2017;(1):22–28. (In Russ.)
6. Volchikhin V.I., Ivanov A.I., Yunin A.P., Malygina E.A. Multidimensional portrait of digital sequences of ideal “white noise” in Hamming convolutions. *Izvestiya vysshikh uchebnykh zavedeniy. Povolzhskiy region. Tekhnicheskie nauki = University proceedings. Volga region. Engineering sciences*. 2017;(4):4–13. (In Russ.). doi:10.21685/2072-3059-2017-4-1
7. Yunin A.P., Ivanov A.I., Ratnikov K.A., Kol'chugina E.A. The quality assessment of “white noise”: implementation of the “flock of monkeys” test through a set of Hamming convolutions built for different number systems. *Izvestiya vysshikh uchebnykh zavedeniy. Povolzhskiy region. Tekhnicheskie nauki = University proceedings. Volga region. Engineering sciences*. 2018;(4):54–64. (In Russ.). doi:10.21685/2072-3059-2018-4-5

8. Ivanov A.I., Yunin A.P. *Embrion iskusstvennogo intellekta: kompaktnaya neyrosetevaya proverka kachestva sluchaynykh posledovatel'nostey, poluchennykh iz biometricheskikh dannykh: preprint = Artificial intelligence embryo: compact neural network quality check of random sequences obtained from biometric data: preprint*. Penza: Izd-vo PGU, 2021:68. (In Russ.)
9. Ivanov A.I., Kubasov I.A., Samokutyaev A.M. Testing large neural networks on small samples. *Nadezhnost' i kachestvo slozhnykh system = Reliability and quality of complex systems*. 2020;(3):72–79. (In Russ.). doi:10.21685/2307-4205-2020-3-9
10. Ivanov A.I. Iskusstvennyy intellekt vysokogo doveriya. Uskorenie vychisleniy i ekonomiya pamyati pri testirovanii bol'shikh setey iskusstvennykh neyronov na malykh vyborkakh. *Sistemy bezopasnosti = Security systems*. 2020;(5):60–62. (In Russ.)

**Информация об авторах / Information about the authors**

***Владимир Иванович Волчихин***
доктор технических наук, профессор, президент Пензенского государственного университета (Россия, г. Пенза, ул. Красная, 40)

E-mail: president@pnzgu.ru

***Vladimir I. Volchikhin***
Doctor of engineering sciences, professor, president of Penza State University (40 Krasnaya street, Penza, Russia)

***Александр Иванович Иванов***
доктор технических наук, доцент, научный консультант, Пензенский научно-исследовательский электротехнический институт (Россия, г. Пенза, ул. Советская, 9)

E-mail: ivan@pniei.penza.ru

***Aleksandr I. Ivanov***
Doctor of engineering sciences, associate professor, scientific adviser, Penza Research Institute of Electrical Engineering (9 Sovetskaya street, Penza, Russia)

***Алексей Петрович Иванов***
кандидат технических наук, доцент, заведующий кафедрой технических средств информационной безопасности, Пензенский государственный университет (Россия, г. Пенза, ул. Красная, 40)

E-mail: ap_ivanov@pnzgu.ru

***Aleksey P. Ivanov***
Candidate of engineering sciences, associate professor, head of the sub-department of technical facilities of information security, Penza State University (40 Krasnaya street, Penza, Russia)

***Алексей Валерьевич Строков***
заместитель начальника управления режима и защиты информации, ООО «Организационно-технические решения» (Россия, г. Москва, Дмитровское шоссе, 60А)

E-mail: strokov.alexey@otr.ru

***Aleksey V. Strokov***
Deputy head of the office of access and information security, Organizational & Technological Solutions 2000 LLC (60A Dmitrovskoye highway, Moscow, Russia)

**Авторы заявляют об отсутствии конфликта интересов / The authors declare no conflicts of interests**.

**Поступила в редакцию / Received** 09.08.2021

**Поступила после рецензирования и доработки / Revised** 20.09.2021

**Принята к публикации / Accepted** 05.10.2021