УДК 578.087.1+004.383.8+612.087.1

Иванов А. И., Сомкин С. А., Андреев Д. Ю., Малыгина Е. А.

О МНОГООБРАЗИИ МЕТРИК, ПОЗВОЛЯЮЩИХ НАБЛЮДАТЬ РЕАЛЬНЫЕ СТАТИСТИКИ РАСПРЕДЕЛЕНИЯ БИОМЕТРИЧЕСКИХ ДАННЫХ «НЕЧЕТКИХ ЭКСТРАКТОРОВ» ПРИ ИХ ЗАЩИТЕ НАЛОЖЕНИЕМ ГАММЫ

Проведен анализ возможностей «нечетких экстракторов» и нейросетевых преобразователей биометрия-код. Показано, что для «нечетких экстракторов» переход в пространство метрики расстояний Хэмминга и/или использование метрики среднего значения показателей стабильности разрядов биокода приводит к автоматическому снятию защиты от наблюдения статистик распределения биоданных. Использование свойств нейросетевых преобразователей биометрия-код позволяет решить данные проблемы.

Ключевые слова: *метрика расстояний Хэмминга, метрика среднего показателя стабильности состояний разрядов биокода, биометрические данные, преобразователь биометрия-код.*

Ivanov A., Somkin S., Andreev D., Malygina E.

DIVERSITY METRICS TO WATCH ACTUAL BIOMETRIC DATA DISTRIBUTION STATISTICS «FUZZY EXTRACTORS» IN THEIR PROTECTION OF A RANGE

The analysis of the «fuzzy extractors» and converters biometrics-neural network code. It is shown, that for «fuzzy extractors» transition to the Hamming distance metrics and/or use the metric average indicators stability level bio-code will automatically remove the protection from bio-data distribution statistics. Use the properties of neural network converters biometrics-code allows you to solve these problems.

Keywords: *Hamming distance metric, metric, the average State-level bio-security, biometrics, converters biometrics-neural network code.*

1. Классификация преобразователей биометрия-код

Все преобразователи биометрии в код делятся на «нечеткие экстракторы» и нейросетевые преобразователи биометрия-код. Отличие между ними только в положении квантователя непрерывных биометрических данных. В «нечетких экстракторах» квантователь преобразует в код «сырые» биометрические данные, а далее эти данные исправляются за счет избыточности самокорректирующегося кода.

В нейросетевых преобразователях «сырые» биометрические данные первоначально обогащаются сумматорами нейронов, а далее уже обогащенные сигналы на выходах сумматоров квантуются выходным нелинейным элементом. Структурные схемы, отражающие положение квантователей в преобразователях биометрия-код, отображены на рис. 1.

В «нечетких экстракторах» может быть использован любой классический код, способный обнаруживать и исправлять ошибки. Обычно используются коды БЧХ (Боуза – Чоухуры – Хоквингема) с примерно 10 кратной избыточностью, способные править до 15% ошибок. То есть при 512 контролируемых биометрических параметров длина выходного кода «нечеткого экстрактора» составит 51 бит.

Нейронные сети осуществляют обогащение данных в непрерывной форме, и обычно для корректировки всех входных ошибок оказывается достаточно двукратной избыточности, то есть 512 входных биопараметров нейронная сеть преобразует в 256 бит выходного кода практически без ошибок.

С точки получения биометрических свойств нейросетевые преобразователи биометрия-код всегда лучше «нечетких экстракторов». Этот тезис никто не оспаривает. Это легко продемонстрировать на примере плохих биометрических данных, дающих ошибки в 50% и более разрядах биокода, скорректировать больше 50% ошибок классические самокорректирующиеся коды не способны, нейронные сети с этой проблемой справляются, если избыточность их становится трехкратной (входов в три раза больше, чем выходов).

2. Нечеткие экстракторы

Основным преимуществом «нечетких экстракторов» англоязычная криптографическая общественность считала их относитель-

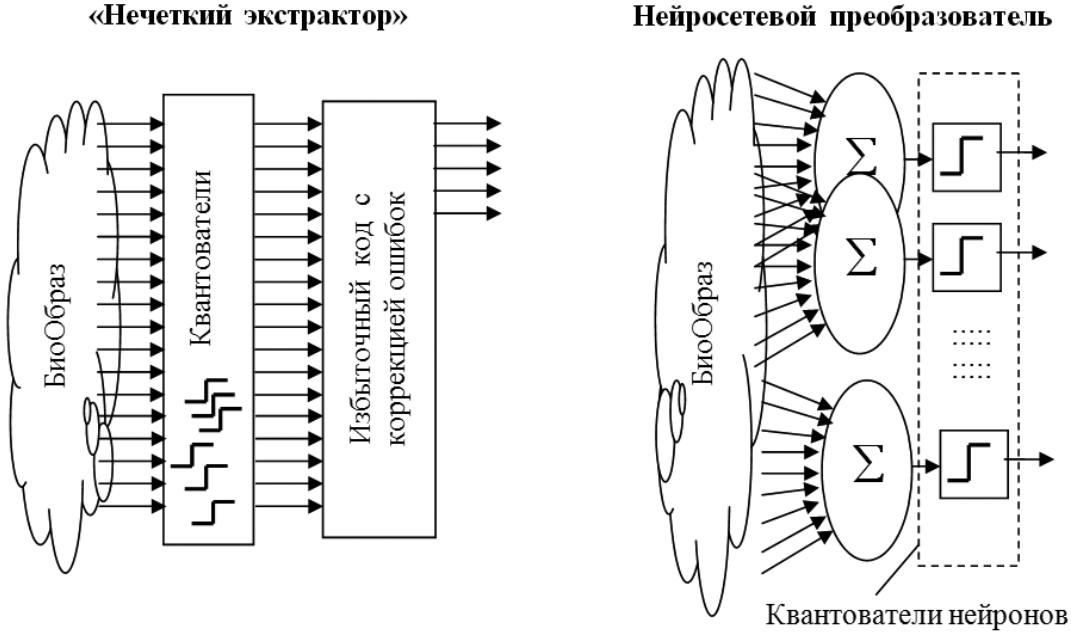

Рис. 1. «Нечеткие экстракторы» и нейросетевые преобразователи отличаются положением нелинейных элементов, квантующих непрерывные данные в код с конечным числом состояний

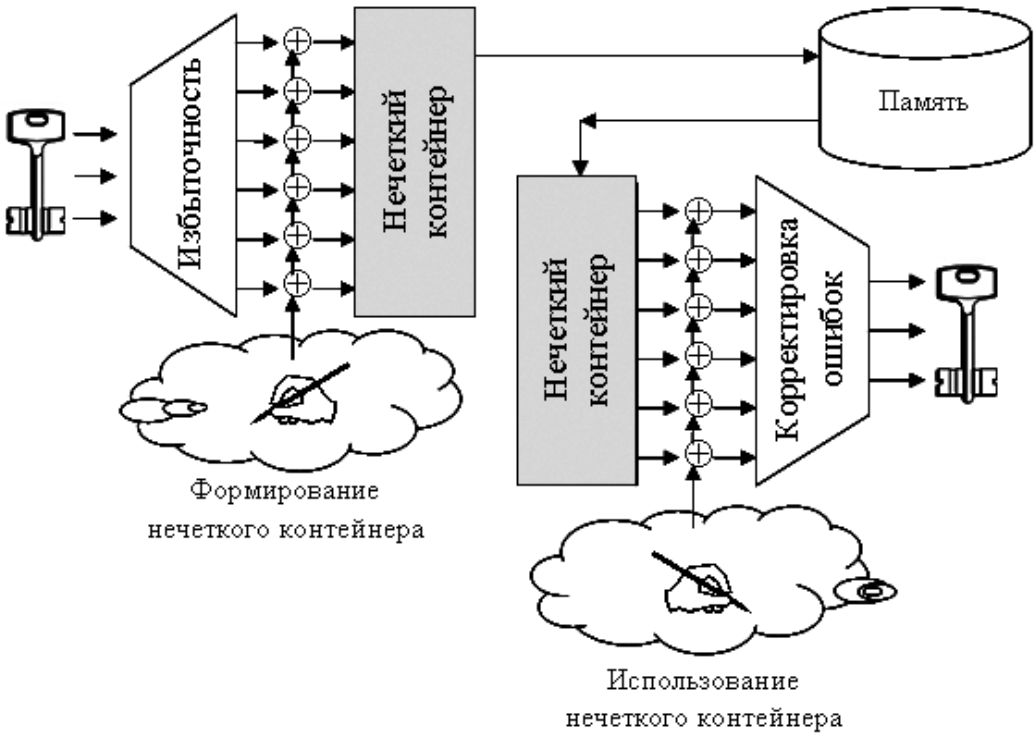

Рис. 2. Формирование и использование нечетких контейнеров

но высокий уровень защищенности и простоту (прозрачность) используемой защиты. Принцип защиты данных «нечетких экстракторов» иллюстрируется рис. 2.

Для защиты «сырых» биокодов используют секретный ключ. Этот ключ накрывают избыточным самокорректирующимся кодом (например, БЧХ), тем самым получают гамму в 10 раз длиннее кода секретного ключа. Далее накрывают «сырой» биокод гаммой, получая тем самым «нечеткий контейнер». «Нечеткий контейнер» хранят в памяти средств биометрической аутентификации. В США и странах НАТО такой способ считается относительно безопасным, и именно эта ветвь технологий за рубежом активно развивается. Конструкции «нечетких экстракторов» даны в ряде англоязычных публикаций[1-12], в работах[13-15] отражены усилия русскоязычных исследователей этого технологического направления.

В процессе аутентификации «нечеткий контейнер» извлекают из памяти и складывают его данные по модулю два с введенным и оцифрованным биометрическим образом. При этом восстанавливается избыточный самокорректирующийся код криптографического ключа, содержащий ошибки, унаследованные от двух биокодов (биокоде формирования нечеткого контейнера и биокоде аутентификации). Если таких ошибок меньше исправляющей способности самокорректирующегося кода, то они правятся.

3. Уязвимость защиты «нечетких контейнеров» в пространстве метрик расстояний Хэмминга

Если бы гаммой был защищен неизвестный текст, то такая защита сегодня считается надежной. Распространять этот тезис о безопасности на защиту «сырых» биокодов нельзя. Это иной объект защиты. Атакующий может подать на вход преобразователя 1000 образов «Чужой» и получить выборку из 1000 «сырых» биокодов «Чужой», накрытых одинаковой гаммой - "$\overline{g}$".

Если мы из пространства обычных кодов перейдем в пространство расстояний Хэмминга между кодами, то при вычислениях мы фактически снимаем одну и ту же гамму:

$$h("\overline{x}", "\overline{c}") = \sum_{i=1}^{512} "x_i" \oplus "c_i" = \sum_{i=1}^{512} ("g_i" \oplus "x_i") \oplus ("g_i" \oplus "c_i"), \qquad (1)$$

где "g_i" – состояния «0» или «1» *i*-го разряда, маскирующей биокод гаммы;

"c_i" – состояния «0» или «1» *i*-го разряда биокода «Свой»;

"x_i" – состояния «0» или «1» *i*-го разряда биокода «Чужой»,

512 – длина сравниваемых кодов.

В метрике расстояний Хэмминга защищающая биокод гамма неизвестного ключа исчезла, то есть биокоды «нечетких экстракторов» оказываются без защиты от наблюдения статистик распределений расстояний Хэмминга. Анализируя статистики распределения расстояний Хэмминга, удается выявить то направление, куда следует двигаться при направленном подборе биометрических данных неизвестного биометрического образа «Свой». Задача направленного подбора распределений биометрических данных образа «Свой» имеет полиномиальную сложность как для «нечетких экстракторов», так и для «нейросетевых контейнеров».

4. Уязвимость защиты «нечетких контейнеров» в пространстве метрики показателей стабильности био-кодов

Следует отметить, что метрика расстояний Хэмминга не единственная метрика, в которой защищающая «нечеткие контейнеры» гамма снимается. Точно такой же эффект снятия защитной гаммы наблюдается в метрике показателей стабильности разрядов биокода.

Показатель стабильности *i*-го разряда определяется через вероятности появления в разряде состояний «1» или «0»:

$$\gamma_i = 2 \cdot \left|0.5 - P("1_i")\right| = 2 \cdot \left|0.5 - P("0_i")\right|. \quad (2)$$

В предельной ситуации $P(«0_i»)=P(«1_i»)=0.5$ показатель стабильности оказывается нулевым. Если же значение разряда не меняется $P(«0_i»)=1$ или $P(«1_i»)=1$, то показатель стабильности оказывается единичным. Примеры гистограмм распределения показателей стабильности образа «Свой» и образов «Чужой» приведены на рис. 3.

Из данных рис. 3 видно, что по мере увеличения расстояния Хэмминга образа «Чужой» от образа «Свой» стабильность разрядов биокодов падает. Такой показатель, как средняя стабильность всех разрядов биокода – $E(\gamma)$, может использоваться для определения направления движения в сторону образа «Свой». При таком направленном переборе следует стремиться увеличивать по-

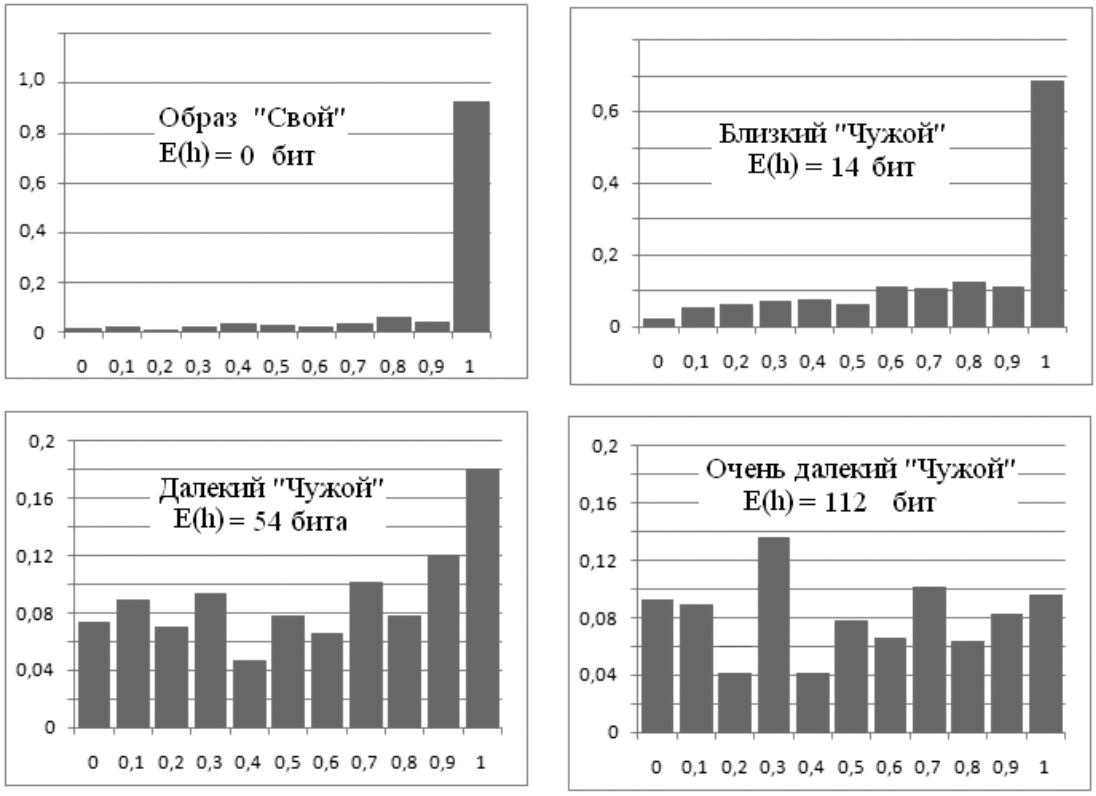

Рис. 3. Падение стабильности разрядов кодов «Чужой» по мере удаления образа «Чужой» от образа «Свой»

казатель стабильности разрядов при генетической селекции образов «Чужой», приближающихся к образу «Свой».

Очевидным является то, что гаммирование биокода не может повлиять на показатели стабильности разрядов кода. Гаммирование – это детерминированная операция, которая приводит только к смене наиболее вероятного состояния (состояние «0» может смениться на состояние «1»). Это никак не влияет на значение показателя стабильности (2), так как его можно вычислять и через вероятность $P(«0_i»)$, и через вероятность $P(«1_i»)$. Результат вычислений будет одинаковым.

5. Уязвимость защиты «нечетких контейнеров» в пространстве комбинированной метрики Хэмминга, взвешенной показателями стабильности разрядов био-кодов

Одним из важных теоретических моментов является то, что обе рассмотренные выше метрики могут быть объединены. Из них получается более сильная комбинированная метрика:

$$h("\overline{x}", "\overline{c}", \overline{\gamma}("x"), \overline{\gamma}("c")) = \sum_{i=1}^{512} ("x_i" \oplus "c_i") \cdot \gamma("x_i") \cdot \gamma("c_i") = \\ = \sum_{i=1}^{512} \left[("g_i" \oplus "x_i") \oplus ("g_i" \oplus "c_i")\right] \gamma("x_i") \cdot \gamma("c_i") \quad . \quad (3)$$

Эта метрика наследует лучшие свойства своих метрик-родителей, она одновременно учитывает и расхождения между разрядами сравниваемых биокодов «Чужой» и «Свой», и стабильность появления состояний в этих разрядах. Она учитывает то, что сравнивать между собой совершенно не стабильные разряды нет смысла. В этом случае расчет метрики Хэмминга полностью утрачивает физический смысл.

Практика применения взвешенной метрики Хэмминга показала, что она значительно увеличивает скорость направленного подбора биометрических данных. Комбинирование двух разных по содержанию метрик позволяет сократить адресное пространство направленного подбора от 20% до 40%.

6. Метрика среднего модуля коэффициентов корреляции и метрика энтропии биокодов

Еще одной метрикой, способной «видеть» реальные статистики данных, накрытых неизвестной гаммой, является метрика среднего модуля коэффициентов корреляции между случайно выбранными парами разрядов биокода. Очевидно, что для разрядов биокода «Свой» модуль коэффициентов корреляции оказывается близок к единице. Для кодов «Чужой» корреляционные связи всегда оказывается много меньше:

$$\left|r("c_i", "c_j")\right| \approx 1 \;\rangle\rangle\; \left|r("x_i", "x_j")\right|. \quad (4)$$

Это является следствием, того, что энтропия биокодов «Свой» оказывается практически нулевой, тогда как энтропия кодов «Чужой» оказывается крайне высокой:

$$H("\overline{c}") \approx 0 \langle\langle H("\overline{x}"). \quad (5)$$

Значение энтропии биокодов функционально связано с математическим ожиданием модулей коэффициентов парной корреляции (4), номограмма связи дана в монографии[16]. Легко показать, что наложение одной и той же гаммы на исследуемые биокоды не влияет на корреляционную и энтропийную метрики. В итоге получается, что мы можем наблюдать реальные статистики биокодов в 4 разных метриках (1), (2), (4), (5) и их комбинациях. Простое гаммирование данных «нечетких экстракторов» нельзя рассматривать как эффективную защиту биометрических данных. В пространствах описанных выше метрик легко строится атака направленного подбора данных биометрического образа «Свой».

7. Защита от атак направленного подбора данных образа «Свой»

Обычно атака направленного подбора данных неизвестного биометрического образа «Свой» ведется до момента подбора ключа аутентификации (рис. 2). При реализации атаки осуществляют подстановку образов «Чужой» из заранее созданной базы, осуществляя попутно их генетическую селекцию и скрещивание. При этом направление верного движения определяется тем вернее, чем длиннее оказываются наблюдаемые биокоды. Фактически атака направленного подбора выполня-

ется за счет возможности многомерной статистической обработки дынных в пространствах метрик (1), (2), (4), (5) и их комбинаций. Как было показано выше, для длинных биокодов нечетких экстракторов защита простым гаммированием не помогает. При этом исправить ситуацию для «нечетких экстракторов» нельзя, так как они нуждаются в длинных самокорректирующихся кодах, содержащих значительную избыточность. Пока биокод не исправлен, нельзя перемешивать его разряды, их последовательность должна полностью повторять заранее заданную структуру самокорректирующегося кода.

Совершенно иная ситуация возникает при использовании нейросетевых преобразователей биометрия-код. Их выходной биокод практически не содержит ошибок. То есть его можно защищать не только гаммированием, но и перемешиванием данных. Как только мы включаем в средство защиты механизмы перемешивания (механизмы размножения ошибок), наблюдать реальные статистики биометрических кодов в пространстве метрик (1), (2), (4), (5) уже не удается (структурные статистические связи разрушаются). Получается, что нейросетевые преобразователи биометрия-код вполне могут быть защищены от наблюдения реальных многомерных статистик в пространствах расстояний Хэмминга (1), средней стабильности разрядов (2), модулей корреляции (4), энтропии (5) и их комбинаций. Более того, ГОСТ Р 52633.3-201117 содержит прямые рекомендации того, как определить, в каком режиме находится «нейросетевой контейнер» (включен или нет защитный механизм размножения биометрических ошибок). То, что является непреодолимой угрозой для «нечетких экстракторов», достаточно просто отражается при использовании нейросетевых преобразователей биометрия-код[16].

Примечания

[1] Y. Dodis, L. Reyzin, A. Smith Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy, Data April 13, In EUROCRYPT, pages 523-540, 2004.

[2] F. Monrose, M. Reiter, Q. Li, S. Wetzel. Cryptographic key generation from voice. In Proc. IEEE Symp. on Security and Privacy, 2001.

[3] Arakala A., Jeffers J., Horadam K.J. Fuzzy Extractors for Minutiae-Based Fingerprint Authentication. // Advances in Biometrics (LNCS 4642), Springer, pp. 760-769, 2007.

[4] Balakirsky V.B., Ghazaryan A.R., Han Vinck A.J. Constructing Passwords from Biometrical Data. // Advances in Biometrics (LNCS 5558), Springer, pp. 889-898, 2009.

[5] Cauchie S., Brouard T., Cardot H. From features extraction to strong security in mobile environment: A new hybrid system. //On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops, Springer, pp. 489-498, 2006.

[6] Juels A., Wattenberg M. A Fuzzy Commitment Scheme // Proc. ACM Conf. Computer and Communications Security, 1999, p. 28–36.

[7] Juels A., Sudan M. A Fuzzy Vault Scheme // IEEE International Symposium on Information Theory, 2002.

[8] Kanade S., Petrovska-Delacretaz D., Dorizzi B. Multi-Biometrics Based Cryptographic Key Regeneration Scheme. //Proceedings of the 3rd IEEE international conference on Biometrics: Theory, applications and systems, p. 333-339, 2009.

[9] Lee Y.J., Bae K., Lee S.J., Park K.R., Kim J. Biometric Key Binding: Fuzzy Vault Based on Iris Images. // Proceedings of 2nd International Conference on Biometrics, p. 800–808, Seoul, South Korea, August 2007.

[10] Nandakumar K., Jain A.K., Pankanti S. Fingerprint-Based Fuzzy Vault: Implementation and Performance. //IEEE Transactions on Information Forensics and Security 2(4), pp. 744–757, 2007.

[11] Ramírez-Ruiz J., Pfeiffer C., Nolazco-Flores J. Cryptographic Keys Generation Using FingerCodes.// Advances in Artificial Intelligence - IBERAMIA-SBIA 2006 (LNCS 4140), p. 178-187, 2006.

[12] Yang S., Verbauwhede I. Automatic Secure Fingerprint Verification System Based on Fuzzy Vault Scheme // Proc. IEEE ICASSP 2005, p.609-612.

[13] Чморра А. Л. Маскировка ключа с помощью биометрии // Проблемы передачи информации. 2011. № 2(47). С. 128–143.

[14] Чморра А. Л., Уривский А. В. «Биометрическая система аутентификации», описание к патенту № RU2316120, 27.01.2008. Бюл. № 3.

[15] Ушмаев О. В., Кузнецов В. В. Алгоритмы защищенной верификации на основе бинарного представления топологии отпечатка пальцев // Информатика и ее применения. 2012. № 6(1). С. 132–140.

[16] Язов Ю. К. и др. Нейросетевая защита персональных биометрических данных. М.: Радиотехника. 2012. –160 с.

[17] ГОСТ Р 52633.3-2011 «Защита информации. Техника защиты информации. Тестирование стойкости средств высоконадежной биометрической защиты к атакам подбора». М.: Стандартинформ, 2012. – 16 с.

References

[1] Y. Dodis, L. Reyzin, A. Smith Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy, Data April 13, In EUROCRYPT, pages 523-540, 2004.

[2] F. Monrose, M. Reiter, Q. Li, S. Wetzel. Cryptographic key generation from voice. In Proc. IEEE Symp. on Security and Privacy, 2001.

[3] Arakala A., Jeffers J., Horadam K.J. Fuzzy Extractors for Minutiae-Based Fingerprint Authentication. // Advances in Biometrics (LNCS 4642), Springer, pp. 760-769, 2007.

[4] Balakirsky V.B., Ghazaryan A.R., Han Vinck A.J. Constructing Passwords from Biometrical Data. // Advances in Biometrics (LNCS 5558), Springer, pp. 889-898, 2009.

[5] Cauchie S., Brouard T., Cardot H. From features extraction to strong security in mobile environment: A new hybrid system. //On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops, Springer, pp. 489-498, 2006.

[6] Juels A., Wattenberg M. A Fuzzy Commitment Scheme // Proc. ACM Conf. Computer and Communications Security, 1999, p. 28–36.

[7] Juels A., Sudan M. A Fuzzy Vault Scheme // IEEE International Symposium on Information Theory, 2002.

[8] Kanade S., Petrovska-Delacretaz D., Dorizzi B. Multi-Biometrics Based Cryptographic Key Regeneration Scheme. //Proceedings of the 3rd IEEE international conference on Biometrics: Theory, applications and systems, p. 333-339, 2009.

[9] Lee Y.J., Bae K., Lee S.J., Park K.R., Kim J. Biometric Key Binding: Fuzzy Vault Based on Iris Images. // Proceedings of 2nd International Conference on Biometrics, p. 800–808, Seoul, South Korea, August 2007.

[10] Nandakumar K., Jain A.K., Pankanti S. Fingerprint-Based Fuzzy Vault: Implementation and Performance. //IEEE Transactions on Information Forensics and Security 2(4), pp. 744–757, 2007.

[11] Ramírez-Ruiz J., Pfeiffer C., Nolazco-Flores J. Cryptographic Keys Generation Using FingerCodes.// Advances in Artificial Intelligence - IBERAMIA-SBIA 2006 (LNCS 4140), p. 178-187, 2006.

[12] Yang S., Verbauwhede I. Automatic Secure Fingerprint Verification System Based on Fuzzy Vault Scheme // Proc. IEEE ICASSP 2005, p.609-612.

[13] Chmorra A.L. Maskirovka klyucha s pomoshch'yu biometrii [Concealement of the Key with the Help of Biometrics] // Problemy peredachi informatsii. 2011 No. 2(47). p. 128-143.

[14] Chmorra A.L., Urivskii A.V. «Biometrical System of Authentication», description to the patent №No. RU2316120, 27.01.2008. No. 3. (In Russ.)

[15] Ushmaev O.V., Kuznetsov V.V. Algoritmy zashchishchennoi verifikatsii na osnove binarnogo predstavleniya topologii otpechatka pal'tsev [Algorithms of Secure Verification on the Basis of Binary Representation of the Topology of Fingerprints] //Informatika i ee primeneniya. 2012. No. 6(1). p. 132-140.

[16] Yazov Yu.K. and others. Neirosetevaya zashchita personal'nykh biometricheskikh dannykh [Neural Network Security of Personal Biometrical Data]. Moscow: Radiotekhnika Publ.. 2012. – 160 p.

[17] All-Union State Standard R 52633.3-2011 «Information Security. Technology of Information Security. Testing of Survivability of Means of High-Reliable Biometrical Security to Brute-Force Attacks». Moscow: Standartinform Publ.. 2012. –16 p.

Иванов Александр Иванович, доктор технических наук, доцент, начальник лаборатории биометрических и нейросетевых технологий ОАО «Пензенский научно-исследовательский электротехнический институт». E-mail: ivan@pniei.penza.ru.

Сомкин Сергей Александрович, зам. начальника научно-исследовательского отдела ОАО «Пензенский научно-исследовательский электротехнический институт». E-mail: somkin@pniei.penza.ru.

Андреев Дмитрий Юрьевич, научный сотрудник лаборатории биометрических и нейросетевых технологий ОАО «Пензенский научно-исследовательский электротехнический институт». E-mail: mail.stray@gmail.com.

Малыгина Елена Александровна, аспирант кафедры «Информационная безопасность систем и технологий» ФБГОУ ВПО «Пензенский государственный университет». E-mail: mal890@yandex.ru.

Ivanov Alexander, doctor of technical sciences, Associate Professor, head of the laboratory of biometric and neural network technology «Penza research Electrotechnical Institute» E-mail: ivan@pniei.penza.ru

Somkin Sergei, Deputy Head of the Research Department, «Penza research Electrotechnical Institute». E-mail: somkin@pniei.penza.ru.

Andreev Dmitry, a researcher at the laboratory of biometric and neural network technology «Penza research Electrotechnical Institute». E-mail: mail.stray@gmail.com

Malygina Elena, graduate student «Security of information systems and technology» Penza State University. E-mail: mal890@yandex.ru