Всероссийская научно-техническая конференция, посвященная 100-летию со дня рождения Б.И. Рамеева: «Информационно-управляющие и телекоммуникационные системы специального назначения». СЕКЦИЯ: «Биометрическая поддержка криптовалют и блокчейн реестров». Доклад состоится 16 мая, 2018 года, с 13^{30}-13^{45}, конференц-зал Технопарка «РАМЕЕВ», ул. Центральная, 1, г. Пенза.

Майоров А.В., Сомкин С.А., Юнин А.П., Акмаев А.Ж.

Оценка стойкости защищенных нейросетевых преобразователей биометрия-код с использованием больших баз синтетических биометрических образов

Аннотация.

Актуальность и цели. Целью работы является тестирование стойкости к атакам подбора нейросетевых контейнеров, в которых таблицы связей нейронов и таблицы их весовых коэффициентов защищены гаммированием.

Материалы и методы. Используется база естественных биометрических образов «Чужой». Далее проверяется сколько первых бит ключа подбирается на базе естественных биометрических образов. После этого размер базы увеличивают путем скрещивания образов-родителей и получения образов-потомков алгоритмом ГОСТ Р 52633.2. Контролируют размеры монотонно увеличивающейся тестовой базы и число подобранных на ней бит ключа.

Результаты. В логарифмической шкале размер тестовой базы связан линейно с длиной ключа, подбираемого при численном эксперименте. Пользуясь этим свойством, удается прогнозировать время, которое займет подбор ключа длиной 256 или 512 бит по данным реального подбора длины ключа от 40 до 64 бит.

Выводы. Защита таблиц связей нейронов гаммированием позволяет реализовать механизм размножения ошибок. Одной ошибки образа «Чужой» на одном нейроне достаточно, что бы все последующие таблицы связей нейронов восстановились неправильно. Возникает эффект размножения ошибок.

Ключевые слова: нейросетевой преобразователь биометрия-код, гаммирование таблиц связей нейронов, размножение биометрических образов.

Совместное использование биометрии и криптографии

Переход к цифровой экономике приводит к необходимости создания механизмов повышения доверия к электронным документам, находящимся в открытом информационном пространстве (например, хранящимся на облачных сервисах). Традиционные методы криптографической защиты информации (шифрование, формирование цифровой подписи) обладают низкой эргономичностью. Обычный пользователь не может запомнить длинный пароль из случайных знаков или свой криптографический ключ. В связи с этим в России [1, 2] и за рубежом [3, 4, 5] активно развиваются методы преобразования личной биометрии человека в его криптографический ключ.

На рисунке 1 приведена схема, так называемых, «нечетких экстракторов» и схема нейросетевых преобразователей биометрия-код. Из рисунка видно, что выходной код криптографического ключа «нечетких экстракторов» короткий. Это происходит из-за того, что «нечеткие экстракторы» используют классические коды с высокой избыточностью, способные обнаруживать и корректировать ошибки. Так Даугман [5] при корректировке кода, получаемого из рисунка радужной оболочки глаза, использует код с 20-ти кратной избыточностью. То есть при расчете длины выходного ключа число входных биометрических параметров следует уменьшить в 20 раз для «нечетких экстракторов». Как следствие, «нечеткие экстракторы» нельзя применять совместно с сильными криптографическими схемами, ориентированными на использование длинных криптографических ключей.

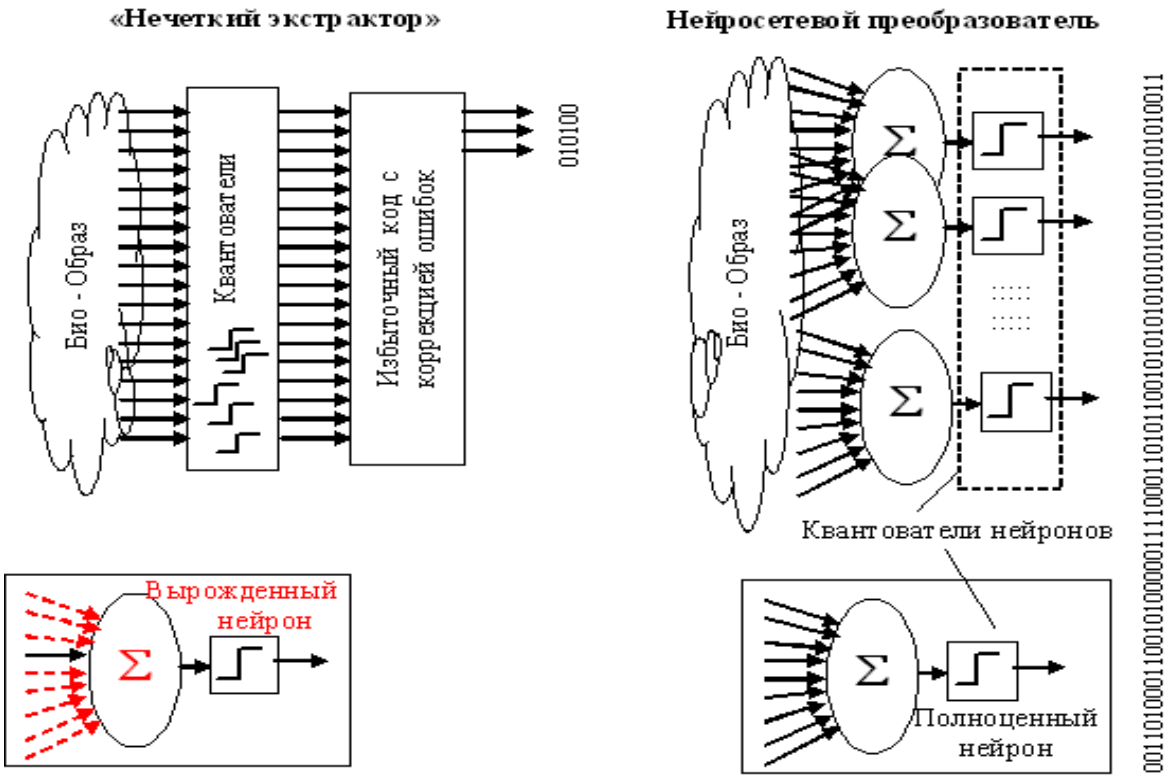

Рис. 1. Схемы организации «нечетких экстракторов» и нейросетевых преобразователей биометрия-код

Ситуация меняется, когда используются нейросетевые преобразователи биометрия-код [1, 2]. В таких конструкциях за каждый бит ключа отвечает один нейрон, число нейронов может быть любым. То есть отечественные нейросетевые преобразователи биометрия-код могут работать с любыми криптографическими схемами защиты информации.

Классические коды с обнаружением и исправлением ошибок хорошо исследованы, по этой причине проектирование «нечетких экстракторов» является простой инженерной задачей. Иначе обстоит дело с нейросетевыми преобразователями биометрия-код. Искусственные нейронные сети плохо учатся, кроме того до конца неизвестно, как их защищать, какова стойкость биометрико-нейросетевой защиты информации.

Для решения задачи обучения искусственных нейронных сетей в России разработан и введен в действие национальный стандарт ГОСТ Р 52633.5 [6]. Для защиты данных обученных нейронных сетей ТК 026 «Криптографическая защита» в настоящее время создает техническую спецификацию [7].

После обучения каждый нейрон преобразователя биометрия-код имеет таблицу связей с биометрическими параметрами и таблицу весовых коэффициентов, как это показано на рисунке 2. Известно техническое решение по патенту RU2346397 [8], по которому осуществляют шифрование таблиц связей и таблиц весовых коэффициентов. Это техническое решение является слишком сложным, так как ориентировано на применения криптографических алгоритмов шифрования, например, выполненных по ГОСТ Р 34.12 [9]. Практическая реализация средств криптографической защиты информации упрощается, если в место полноценного шифрования используется криптографическая хэш-функция, например, выполненная по ГОСТ Р 34.11 [10].

Техническая спецификация [7] строится на том, что таблицы первого нейрона преобразователя биометрия-код накрыты гаммой - Γ_1, которая получается хэшированием пароля доступа:

$$\Gamma_1 = hesh("пароль") \qquad (1).$$

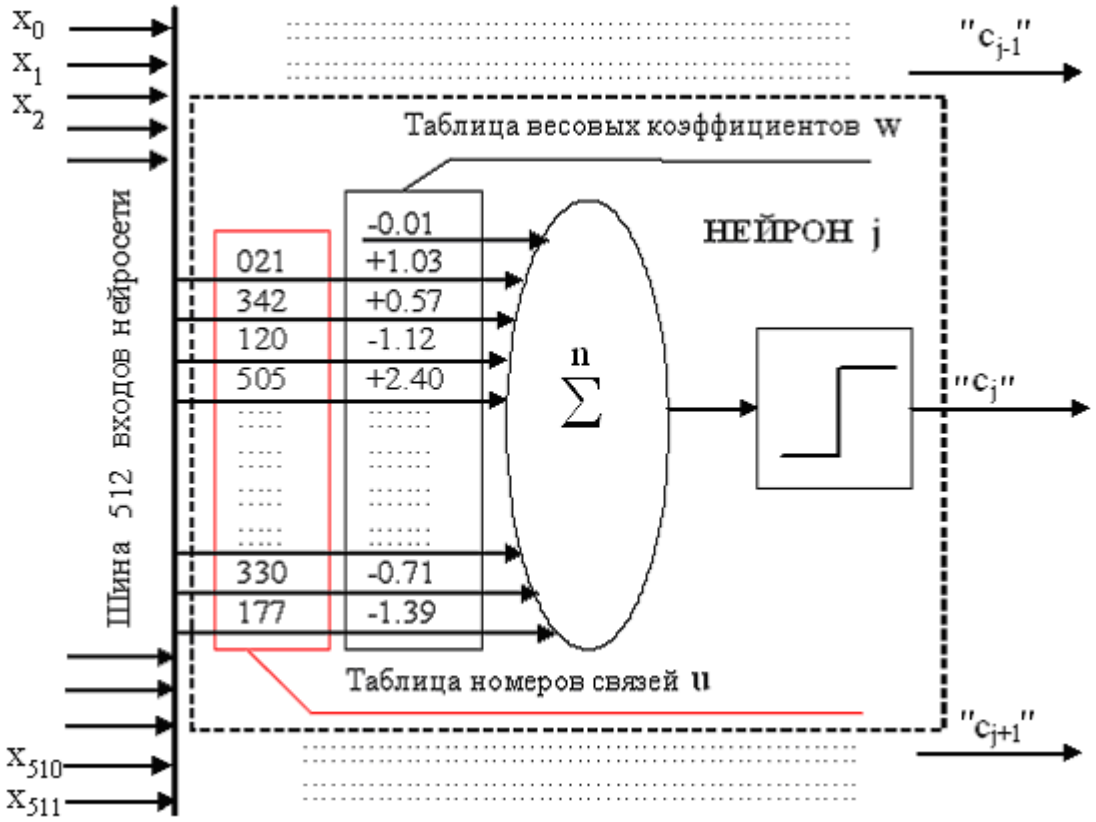

Рис. 2. Каждый нейрон преобразователя биометрия-код имеет собственную таблицу входных связей и собственную таблицу весовых коэффициентов.

Таблицы связей второго нейрона накрываются гаммой, полученной хэшированием таблицы связей гаммы первого нейрона в конкатенации с выходным состоянием первого нейрона:

$$\Gamma_2 = hesh("\Gamma_1","c_1") \qquad (2).$$

Получается рекурсивная схема, на каждом шаге которой следующая гамма, зависит от предыдущей гаммы и состояния предшествующего нейрона:

$$\Gamma_k = hesh("\Gamma_{k-1}","c_{k-1}") \qquad (3).$$

Таблицы связей и значения разрядов кода на выходах нейронов известны при обучении нейронной сети. По этой причине после обучения и назначения пароля доступа все гаммы для всех таблиц нейронов могут быть вычислены. После того как таблицы связей накрыты гаммами, получается защищенный нейросетевой контейнер, который может храниться с привлечением облачного сервиса.

При использовании защищенного контейнера пользователь набирает свой пароль доступа и предъявляет свой биометрический образ. Биометрический образ должен порождать на выходах нейросетевого преобразователя верный выходной код $\{"c_1","c_2","c_3",...,"c_{256}"\}$. Появляется возможность повторно рекуррентно (3) восстановить все гаммы, которыми ранее были накрыты таблицы связей нейронов. Если возникает хотя бы одна ошибка в коде пароля доступа или в коде выходных состояний нейронов, восстановить все гаммы нельзя. Возникает эффект размножения ошибок.

Известно, что стойкость кода биометрико-нейросетевого преобразования к атакам подбора много ниже, чем стойкость криптографического ключа такой же длины [11]. Более того, каждый биометрический преобразователь, обученный на свой биометрический образ, будет обладать своей стойкостью к атакам подбора. То есть, после каждой

процедуры обучения и каждой процедуры защиты данных таблиц нейронов необходимо выполнить процедуру тестирования.

Для тестирования следует заранее собрать базу образов «Чужой», например, состоящую из 10 000 образов. Далее пользователь должен ввести свой пароль и начать подставлять на входы защищенного преобразователя данные 10 000 образов «Чужой». При этом, на выходе преобразователя появится 10 000 кодов. О стойкости защищенного нейросетевого преобразователя следует судить, анализируя младшие разряды кодов.

Проведенный численный эксперимент показал, что для рукописного образа «slovo_01» база данных из 10 000 образов дает множество совпадение до 40 младших разрядов кода. Можно утверждать, что число попыток подбора, позволяющих найти часть ключа в 40 бит, составляет 10 000 образов «Чужой». Формальная запись этого утверждения:

$$П_{40} = 10\,000 \qquad (4).$$

Ограниченный размер тестовой базы обусловлен высоким уровнем трудоемкости ее формирования, если придерживаться требований ГОСТ Р 52633.1 [12]. Для того, что бы увеличить размер тестовой базы необходимо скрещивать пары образов-родителей по ГОСТ Р 52633.2 [13], получая образы-потомки. Примеры подобного скрещивания иллюстрируются рисунком 3.

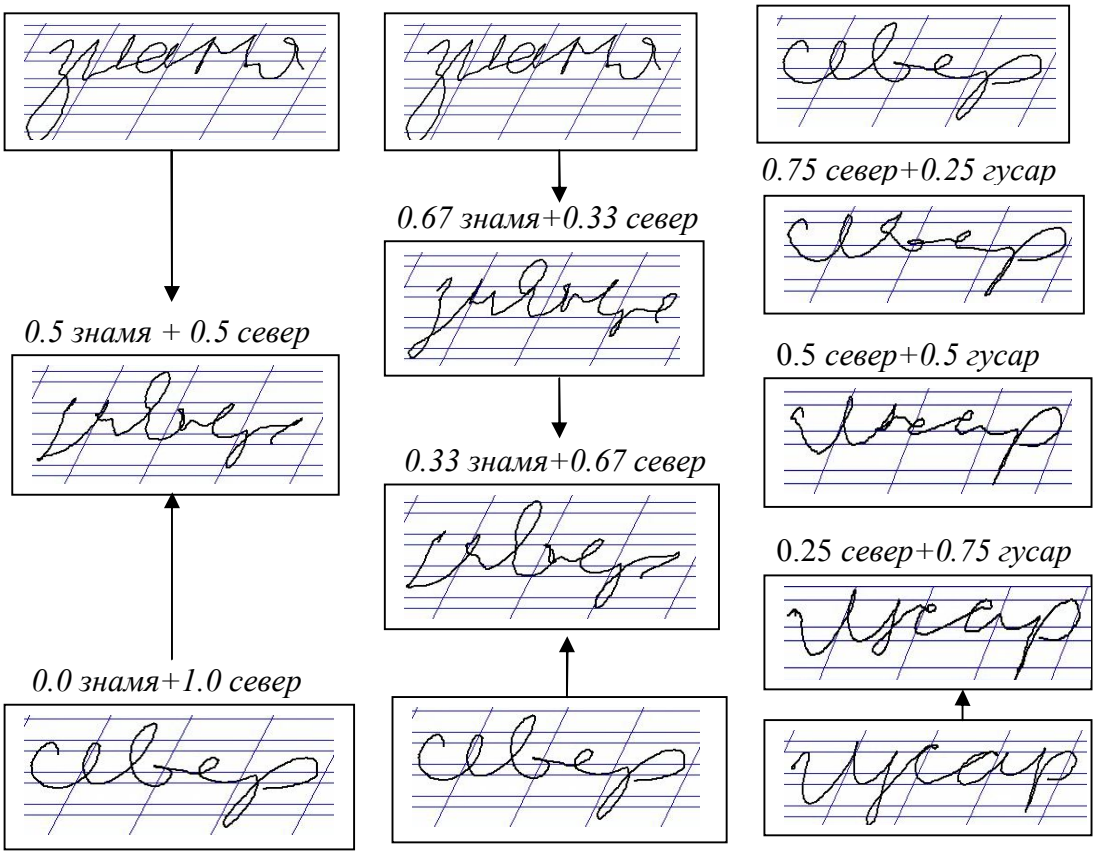

Рис. 3. Получение от образов-родителей 1, 2, 3 образов-потомков линейным морфингом по ГОСТ Р 52633.2

Увеличив размеры тестовой базы, например, в 1000 раз мы вновь можем проверить сколько на этой базе удается подобрать бит ключа. В случае если будет обнаружен факт совпадения первых 44 бит, можно записать:

$$П_{44} = 10\,000\,000 \qquad (5).$$

Очевидно, что постепенно увеличивая размеры тестовой базы, мы будем подбирать все больше и больше бит ключа. На рисунке 4 приведена прямая, отражающая результаты подбора.

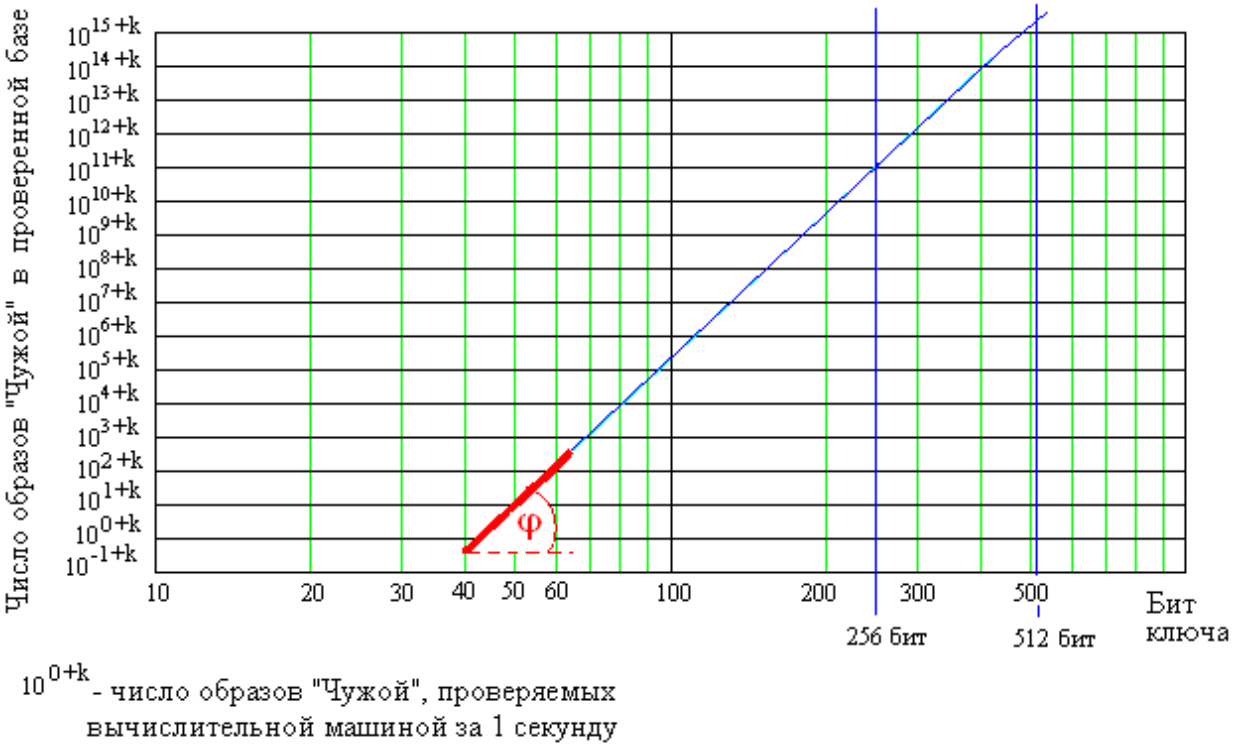

Рис. 4. Прогноз размеров тестовой базы образов «Чужой», необходимой для подбора 256 бит ключа на выходе нейросетевого преобразователя биометрия-код, обученного на биометрическом образе «slovo_01» (данные из приложения)

По результатам численного эксперимента получено время 0.47 секунды перебора первой базы естественных биометрических образов $П_{40}$. После того как выполнено многократное размножение биометрических данных удалось подобрать первых 64 бита ключа за время 6 минут 51 секунду.

При реализации численного эксперимента производился последовательный подбор 40, 42, 44, …, 64 бит ключа с контролем размеров базы образов «Чужой» $П_{40}$, $П_{42}$, $П_{44}$, …., $П_{64}$. Как показано на рисунке 4 промежуточные данные сливаются в одну прямую (утолщенная линия с углом наклонна - φ). Опираясь на результаты численного эксперимента, удается предсказать размер базы тестовых образов «Чужой», необходимой для подбора 256 бит ключа. Необходимые расчеты описываются, приведенной ниже системой из трех уравнений:

$$\begin{cases} tg(\varphi) = \dfrac{\log_{10}(П_{64}) - \log(П_{40})}{64 - 40} \\ D = \log_{10}(П_{64}) + tg(\varphi) \cdot (256 - 64) \\ П_{256} \approx 10^{D} \end{cases} \qquad (6)$$

Из данных рисунка 4 следует, что на подбор 256 бит ключа должно уйти 10^{11} секунд, что составит примерно 3 000 лет непрерывной работы вычислительной машины, на которой проводился численный эксперимент. Нами использовалась вычислительная машина с процессором Intel core i5-2500 (3.3 ГГц) ОЗУ 4Гб. При использовании компьютера большей мощности время на подбор сократится, однако оно остается достаточно большим для обычных настольных компьютеров. Размер базы образов «Чужой» инвариантен к мощности вычислительной машины.

ЛИТЕРАТУРА:

1. ГОСТ Р 52633.0-2006 «Защита информации. Техника защиты информации. Требования к средствам высоконадежной биометрической аутентификации».

2. Язов Ю.К. и др. Нейросетевая защита персональных биометрических данных. //Ю.К.Язов (редактор и автор), соавторы В.И. Волчихин, А.И. Иванов, В.А. Фунтиков, И.Г. Назаров // М.: Радиотехника, 2012 г. 157 с. IBSN 978-5-88070-044-8.

3. Juels A., Wattenberg M. A Fuzzy Commitment Scheme // Proc. ACM Conf. Computer and Communications Security, Singapore — November 01 – 04, 1999, p. 28–36.

4. 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

5. Feng Hao, Ross Anderson, and John Daugman. Crypto with Biometrics Effectively, IEEE TRANSACTIONS ON COMPUTERS, VOL. 55, NO. 9, SEPTEMBER 2006.

6. ГОСТ Р 52633.5-2011 «Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия-код доступа».

7. Техническая спецификация (проект, публичное обсуждение начато с 01.02.2017 членами ТК 26 «Криптографическая защита информации») ЗАЩИТА НЕЙРОСЕТЕВЫХ БИОМЕТРИЧЕСКИХ КОНТЕЙНЕРОВ С ИСПОЛЬЗОВАНИЕМ КРИПТОГРАФИЧЕСКИХ АЛГОРИТМОВ

8. Патент RU 2346397 «Способ защиты персональных данных биометрической идентификации и аутентификации», приоритет от 26.06.07, авторы Иванов А.И., Фунтиков В.А., Ефимов О.В., владельцы опубликовано 10.02.2009 Бюл. №4.

9. ГОСТ Р 34.12 -2015 «Информационная технология. Криптографическая защита информации. Режимы работы блочных шифров.»

10. ГОСТ Р 34.11-2012 «Информационная технология. Криптографическая защита информации. Функция хэширования.»

11. ГОСТ Р 52633.3-2011 «Защита информации. Техника защиты информации. Тестирование стойкости средств высоконадежной биометрической защиты к атакам подбора».

12. ГОСТ Р 52633.1-2009 «Защита информации. Техника защиты информации. Требования к формированию баз естественных биометрических образов, предназначенных для тестирования средств высоконадежной биометрической аутентификации»

13. ГОСТ Р 52633.2-2010 «Защита информации. Техника защиты информации. Требования к формированию синтетических биометрических образов, предназначенных для тестирования средств высоконадежной биометрической аутентификации»

Сведения об авторах:

Майоров Александр Викторович, к.т.н., старший научный сотрудник лаборатории биометрических и нейростевых технологий АО «ПНИЭИ», входит в группу специалистов, разрабатывающих техническую спецификацию [7], тл. р. (841-2) 59-33-10.

Сомкин Сергей Александрович, зам. генерального директора АО «ПНИЭИ», входит в группу специалистов, разрабатывающих техническую спецификацию [7], тл. р. (841-2) 59-33-35.

Юнин Алексей Петрович - ведущий специалист АО «ПНИЭИ», входит в группу специалистов, разрабатывающих техническую спецификацию [7], тл. р. (841-2) 59-33-20.

Акмаев Артур Жиганшеевич - ведущий специалист АО «ПНИЭИ», входит в группу специалистов, разрабатывающих техническую спецификацию [7], тл. р. (841-2) 59-33-30.