http://rosoperator.ru/rameev100/
Всероссийская научно-техническая конференция, посвященная 100-летию со дня рождения одного из основоположников советской вычислительной техники Б.И. Рамеева, на тему: «Информационно-управляющие и телекоммуникационные системы специального назначения». Секция «Биометрическая поддержка криптовалют, блокчейн реестров, облачных сервисов», проводится в конференц-зале Технопарка высоких технологий «Рамеев», доклад 16 мая 2018 года с 13^{45} до 14^{00} (ул. Центральная, д. 1), г. Пенза.

УДК 519.7+621

РЕШЕНИЕ ПРОБЛЕМЫ ФОРМИРОВАНИЕ БАЗ БИОМЕТРИЧЕСКИХ ОБРАЗОВ ДЛЯ ТЕСТИРОВАНИЯ НЕЙРОСЕТЕВЫХ ПРЕОБРАЗОВАТЕЛЕЙ БИОМЕТРИЯ-КОД
(Положительный опыт 2004-2017 гг.)

Малыгин А.Ю.

Показано, что для тестирования средств высоконадежной биометрии обязательно необходимо создавать сбалансированные обезличенные базы биометрических образов с использованием естественных и синтезированных на их основе синтетических биометрических образов.

В связи с тем, что средства высоконадежной биометрической аутентификации [1, 2] в соответствии с национальным стандартом [3] имеют высокую стойкость к атакам подбора, то для полноты их тестирования возникает необходимость создания баз реальных биометрических образов сопоставимых по своим размерам со стойкостью тестируемых средств [4 – 6]. При решении этой непростой и трудоемкой задачи возникает целый ряд вопросов без решения, которых невозможно корректное решение самой задачи. Данные вопросы, на наш взгляд, можно выстроить в следующей последовательности:

1. Требования к испытуемым, при сборе биометрических образов.
2. Требования к аппаратуре преобразования физического образа испытуемого в электронный образ и программному обеспечению для обработки реальных биометрических образов.
3. Требования к создаваемым тестовым базам реальных биометрических образов.

Указанные выше вопросы неразрывно связаны между собой и некорректное решение одного из них может значительно исказить конечный результат.

Исследования, проводимые в межотраслевой лаборатории тестирования биометрических устройств и технологий ПГУ показали, что тестовая машина дает стойкость биометрической защиты к атакам подбора кода ключа порядка 10^{16}, что вполне достаточно для ряда практически значимых приложений, которые в настоящее или ближайшее время могут поступить на рынок. Как следствие, для тестирования такого средства прямой подстановкой потребуется база биометрических образов, содержащая на два - три порядка больше биометрических образов, чем заявленная производителем стойкость биометрической защиты [3].

Оценивая затраты времени и людских ресурсов на формирование подобных баз случайных биометрических образов [6] приходится учитывать то, что испытуемый при всей своей лояльности к средствам биометрической аутентификации, формируя данные должен выполнить ряд операций: осознать то, что он должен написать; написать рукописное слово (фразу); проконтролировать корректность рукописного ввода; ввести биометрический образ. Естественно, что все эти операции занимают определенное время. Хронометраж затрат времени показывает, что специально подготовленный человек (время предварительной подготовки занимает порядка 40 минут) вводит рукописные образы длиной по 5-6 символов за время порядка 10 секунд. Затраты времени на похожие

операции ввода рукописных символов могут существенно различаться для людей разного темперамента и разного рода профессий. При этом рассматриваются только «идеальные» условия: удобные рабочие места с хорошей освещенностью, отсутствие внешних раздражителей и нормальное психофизиологическое состояние доноров биометрии. При этом, естественно, что люди, сталкивающиеся в своей профессиональной деятельности с необходимостью рукописных записей, вводят в ПЭВМ биометрические рукописные образы быстрее.

Несложные расчеты показывают, что для формирования базы из 10^{19} образов потребуется порядка 10^{14} лет работы одного донора биометрии. Естественно, что такие затраты времени и людских ресурсов не могут быть осуществлены.

Отметим, что при формировании биометрических голосовых образов затраты времени и ресурсов оказываются примерно такими же. Голосовая информация вводится быстрее, однако ее приходится вводить больше. Голос обладает несколько меньшей информативностью, если сравнивать между собой одинаковые слова-пароли (парольные фразы).

Наряду с вышесказанным отметим, что в результате проведенных исследований при сборе биометрических данных выявлено то, что все реальные биометрические образы обладают некоторой нестабильностью. Именно неоднозначность, нестабильность, размытость биометрических образов является основной проблемой при их преобразовании в однозначный, четкий криптографический ключ [7]. Пример естественных вариаций нестабильности рукописного образа приведен на рисунке 1.

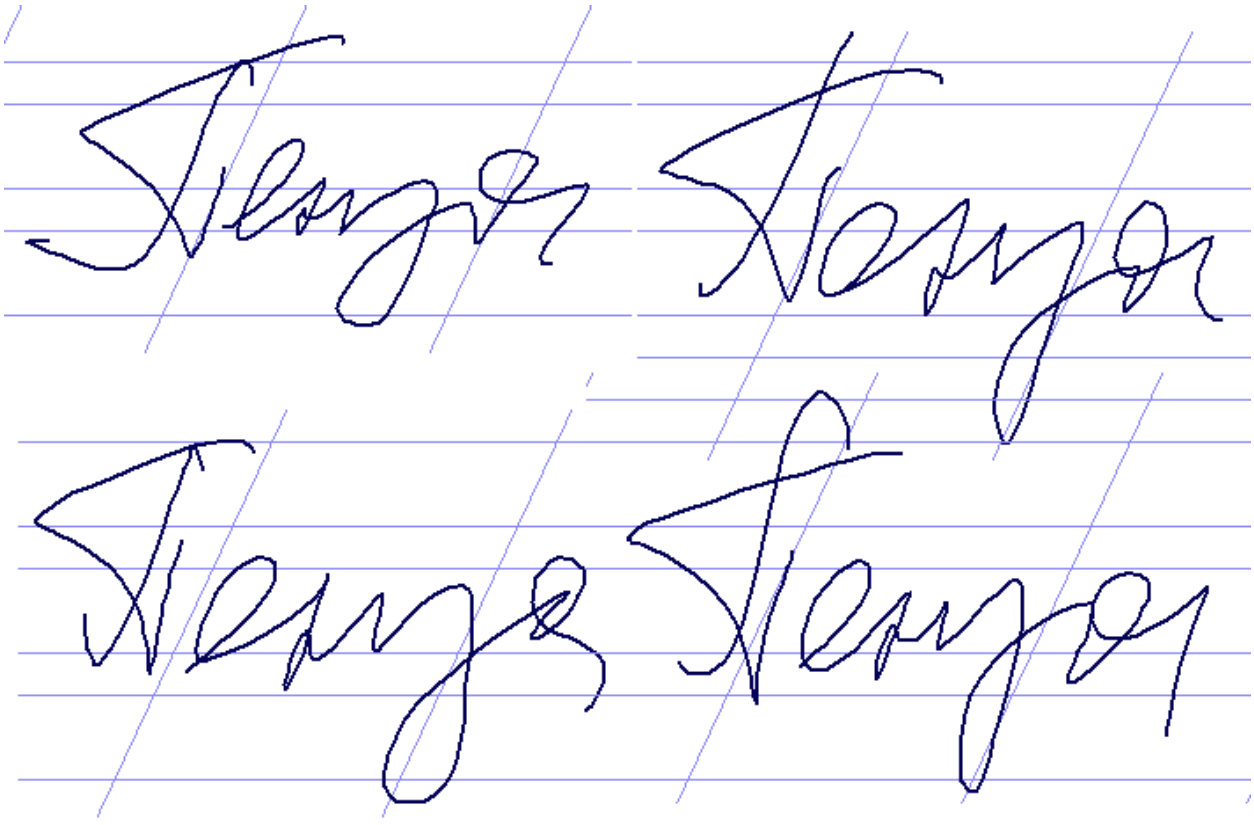

Рис. 1. Примеры естественной нестабильности биометрических образов

Как следствие этого явилась необходимость контролировать нестабильность воспроизведения пользователем его биометрического образа и классифицировать пользователей по их нестабильности.

Для того, чтобы оценить стабильность биометрического образ в целом, необходимо найти математическое ожидание всех дисперсий для каждого из контролируемых биометрических параметров. Сравнивая между собой математические ожидания

дисперсий разных биометрических образов $m(б_1)$ и $m(б_2)$ можно сравнить стабильность воспроизведения разных биометрических образов.

Следует подчеркнуть, что стабильность воспроизведения биометрических образов является относительной величиной. Она определяется в первую очередь вектором контролируемых биометрических параметров. Во-вторую очередь она зависит от умения пользователя стабильно воспроизводить свой биометрический образ. Если это касается воспроизведения рукописных образов, то прежде чем тренироваться стабильно воспроизводить слово-пароль желательно попытаться изменить это слово, отыскивая наиболее стабильные для конкретного почерка сочетания рукописных букв.

Все пользователи для каждой конкретной биометрической системы могут быть классифицированы по стабильности воспроизведения ими их биометрических образов. Для формирования классификации необходимо оценить стабильность нескольких сотен пользователей и построить их нормированное распределение показателя стабильности. Эксперименты показали, что реальная гистограмма распределения пользователей по классам смещена в сторону низко стабильных пользователей с повышенными значениями дисперсий. Аппроксимация гистограммы нормальным законом распределения значений дает достаточно хорошую, точность, однако не учитывает естественную асимметрию реальных гистограмм.

Классификация пользователей по стабильности осуществляется делением распределения «Все Свои» пользователи на интервалы равные среднеквадратическому отклонению, так что бы в центре центрального интервала оказывалось математическое ожидание.

Очевидно, что стабильность воспроизведения биометрических образов является крайне важным статистическим показателем. Естественно требовать от формируемых баз биометрических образов того, чтобы они хорошо отражали реальную действительность по процентному содержанию биометрических образов, принадлежащих к разным классам стабильности.

При экспериментальном синтезе классифицирующего распределения пользователей по их стабильности следует обратить внимание на то, что оно должно строиться только для людей способных пользоваться биометрической защитой). Как правило, малые дети, старики, люди с нервными расстройствами имеют очень нестабильный рукописный почерк. Встроенный в средство аутентификации автомат, учитывающий большую нестабильность ввода рукописных образов, должен предупреждать таких пользователей о невозможности обеспечения надежной биометрической защиты.

Еще одним важным фактором применением классификации по стабильности является выявление случаев саботажа (сговора), когда часть пользователей намеренно ослабляет свою биометрическую защиту. Такая ситуация редка при защите пользователями своих интересов, но достаточно часто встречается при защите корпоративных данных. Для прекращения саботажа служащих, как правило, достаточно самого факта его выявления и объяснения служащему негативных для него последствий его же действий (несоответствие занимаемому служебному положению, по квалификации, психофизиологическому состоянию, нелояльность к корпоративной биометрии).

Не менее важна для систем биометрической защиты степень уникальность биометрического образа «Свой». Очевидно, что злоумышленник всегда будет стараться выбирать при атаках подбора наиболее вероятные состояния входов биометрической защиты. То есть, злоумышленник имеет модель среднестатистического «Своего» (или модель «Все Чужие») и будет стараться использовать ее при организации атак. Естественно, что чем больше образ «Свой» будет отличаться от среднестатистического образа, тем выше степень биометрической защиты.

Интуитивно понятно, что уникальность одного биометрического параметра можно оценить через вероятность случайного попадания в интервал «Свой» при эмуляции

данных «Все Чужие». Эта вероятность в рамках гипотезы нормальности законов распределения значений будет описываться следующим выражением:

$$P_c = \frac{1}{\sqrt{2\pi}\sigma_B} \int_{m_c-3\sigma_c}^{m_c+3\sigma_c} \exp(\frac{(m_B - x)^2}{2\sigma_B^2})dx \quad , \qquad (1)$$

где: m_c, $б_c$ – математическое ожидание и дисперсия распределения параметра «Свой»;
m_B, $б_B$ – математическое ожидание и дисперсия распределения параметра «ВСЕ Чужие».

Очевидно, что вероятность (1) будет уменьшаться при снижении дисперсии «Своего», а также при вытеснении центра множества «Свой» на периферию распределения «Все Чужие».

Уникальность конкретного биометрического параметра будет описываться обратной величиной вероятности (1). Так как анализируемых биометрических параметров много необходимо оценивать среднюю уникальность этих параметров или меру уникальности всего биометрического образа:

$$U = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{P_{C,i}}$$

Уникальность, как и любой иной значимый статистический параметр, может быть использована для классификации биометрических образов. Необходимо отметить, что различные биометрические образы обладают разной информативностью (разной сложностью). Очень простые биометрические образы легко подбираются и не могут обладать необходимой стойкостью. Выделяют статические – неизменные биометрические образы, данные человеку от рождения [1]. К ним относятся рисунки отпечатков пальцев, дерево сосудов глазного дна, радужная оболочка глаз, геометрия ладони, геометрия лица. Как правило, человек не может по своему желанию изменять (усложнять) свой статический биометрический образ. Как следствие статический образ легко компрометируется и обладает ограниченной информативностью.

Свои динамические биометрические образы, напротив, человек легко может изменить. Например, рукописный образ слова пароля, легко может быть изменен сменой самого пароля [1]. Динамические биометрические образы могут быть изменены (усложнены) при необходимости. Их можно сохранять в тайне и за счет усложнения повышать их стойкость к атакам подбора. Для динамических биометрических образов (рукописный и голосовой почерк, клавиатурный почерк) в [3] приводятся только ограничения снизу на длину эквивалентного ключа. Ограничений с верху на этот параметр для этих технологий нет, однако сложность самого биометрического образа и эффективная длина ключа связаны. Чем сложнее биометрический образ, тем сложнее его подбор. Это непреложное правило справедливое для любых (статических или динамических) биометрических образов.

В силу того, что стойкость биометрических образов прямо зависим от их сложности и эта сложность может быть достаточно просто оценена, необходимо при формировании больших баз биометрических образов их балансировать по сложности биометрических образов. То есть базы рукописных и голосовых биометрических образов должны содержать число слов из 5 букв (число отпечатков пальцев с 22 особенностями) в процентном отношении столько же, сколько их содержится в естественном языке (в естественном распределении рисунков отпечатков пальцев).

Рассмотрим требования к качеству преобразователей биометрических образов физического уровня в биометрические электронные образы.

Разнотипные высоконадежные распознаватели во многом сходны. Обобщенная схема устройств высоконадежной биометрической аутентификации личности человека приведена на рисунке 2.

В структурной схеме рисунка 2, блок-1 осуществляет преобразование физического нечеткого биометрического образа человека в электронный биометрический нечеткий образ через первичные преобразователи физических величин в электронные цифровые данные. Блок-2 осуществляет нормировку электронных образов и вычисление вектора биометрических параметров, например, в виде коэффициентов Фурье, в средствах аутентификации по динамике воспроизведения рукописного пароля. Блок-3 осуществляет преобразование вектора биометрических параметров в код ключа (пароля) для последующей криптографической аутентификации. Блок-4 осуществляет криптографическую аутентификацию пользователя по его ключу или паролю, выдавая на выход решение «ДА/НЕТ».

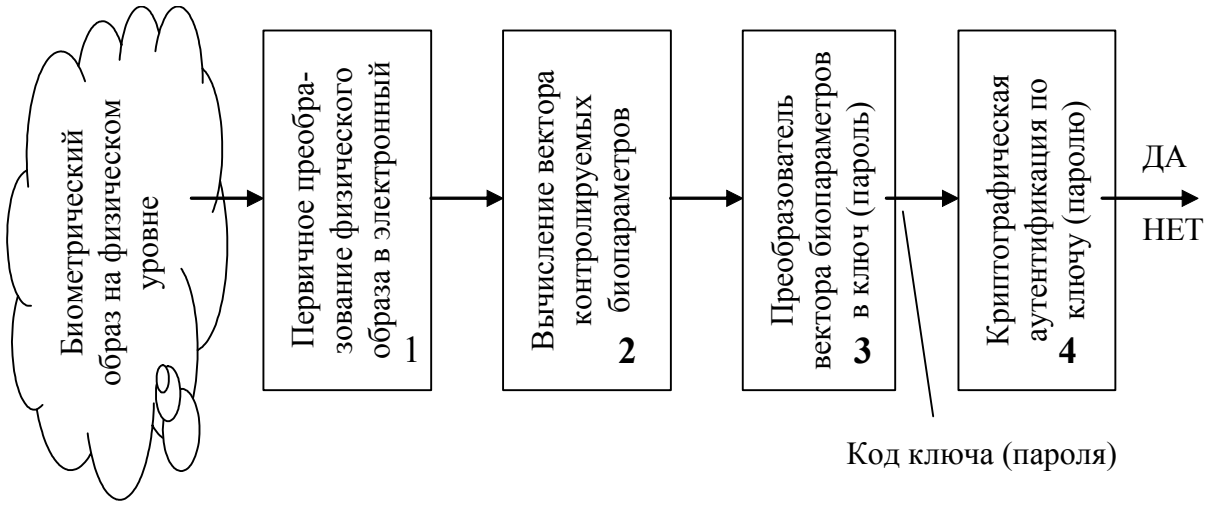

Рис. 2. Структурная схема типового высоконадежного аутентификатора личности человека

Очевидно, что при формировании баз биометрических образов рукописных, голосовых, отпечатков пальцев, …. необходимо использовать именно те датчики, которые использует конкретная система биометрической защиты. Однако такой подход может быть использован только при тестировании относительно «слабой» биометрической защиты. Для «слабой» биометрии достаточно формировать тестовые базы, состоящие из малого числа биометрических образов. Как следствие всегда можно повторить работу по формированию малых баз биометрических образов заново для любого датчика.

При формировании больших баз биометрических образов такой подход неприемлем. Из-за большой трудоемкости этой работы базы должны быть универсальны. То есть при их формировании необходимо использовать наиболее точные на текущий момент средства ввода рукописной графики и звука. Это делается из-за того, что может возникнуть необходимость тестирования очень точных систем аутентификации, тогда собранная информация будет использована в том виде, в каком хранится. Если потребуется снизить точность преобразований физических биометрических образов в электронные биометрические образы, то это всегда можно сделать с помощью специально написанного программного конвертора. Из хороших данных путем ввода «шума» всегда можно сделать плохие данные. Обратная операция просто невозможна.

Таким образом, при формировании больших и универсальных баз биометрических образов требуется использовать как можно более точные и как можно более информативные преобразователи. Для формирования рукописных образов на сегодня подходят для этой цели практически все графические планшеты способные воспринимать 1024 градаций степени нажатия пера и имеющие разрешение 4064 линий на дюйм при рабочем поле 5,5х4 дюйма. Требований к шумам и линейности преобразователей не предъявляется.

При формировании голосовых баз биометрических образов должна использоваться звуковая карта способная оцифровывать звук с частотой 44 КГц, в режиме стереозаписи при дискретизации АЦП по уровню не менее 10 разрядов. Должен использоваться широкополосный микрофон, закрепленный на гарнитуре и второй широкополосный микрофон, закрепленный перед пользователем. При записи голосовых парольных фраз должен быть обеспечен низкий уровень посторонних шумов. Очевидно, что из-за требования к низкому уровню сторонних шумов формировать голосовые базы данных сложнее, так как потребуются помещения подобные студиям звукозаписи.

Немаловажным аспектом при формировании «эталонных» баз биометрических данных являются требования к программному обеспечению автоматизированного формирования базы биометрических тестовых образов.

В связи с высокой трудоемкостью формирования больших баз биометрических образов необходимо стремиться к всемерной экономии затрачиваемых ресурсов. При формировании малых баз возможно привлечение квалифицированного инструктора, под руководством которого каждый пользователь будет выполнять свою работу. Такой подход при формировании больших бах биометрических образов, неприемлем, так как существенно увеличивает их стоимость.

Как показали проведенные исследования при формировании больших баз биометрических образов допустимо привлечение квалифицированного инструктора на этапе обучения тестируемого (40 минутные инструктаж и пробная работа). Далее программное обеспечение ввода биометрических образов должно заменить тестируемого инструктора. Оно должно быть максимально доступно для понимания тестируемым, иметь минимум режимов управления и контролировать все действия тестируемого. При отклонениях тестируемого от заданного оптимального режима программное обеспечение должно автоматически сообщать тестируемому о его ошибке. Например, говорите чуть громче или пишите крупнее и быстрее. Сообщения тестируемому должны выдаваться в голосовой и графической формах.

При этом особые требования предъявляются также и к программным средствам формирования парольных слов (фраз). Тестируемый должен воспроизводить только заданные ему образы (рукописные и голосовые). Генераторы парольных слов и фраз программного обеспечения должны иметь возможность взаимной синхронизации из центра. Эти генераторы должны быть способны воспроизводить заданную из центра последовательность слов (фраз) или формировать независимую последовательность случайных слов (фраз).

При формировании рукописных и голосовых парольных образов допускается использование словарей, допускается усиление слов словарей их вариациями, принятыми в языке тестируемого, допускается их усиление числами в форме, принятой при написании или голосовом воспроизведении.

Все данные о поведении тестируемого и формировании его биометрического образа обязательно документируются. Программное обеспечение должно вести автоматический журнал регистрации полученной биометрической информации и времени ее получения [4].

При формировании баз биометрических образов должна быть обеспечена анонимность тестируемых с тем, чтобы их биометрическая информация в соответствии с ФЗ №152 "О персональных данных" [5] не могла быть использована кем-либо против них в настоящем или будущем времени.

Хочется отметить, что ценность естественных баз биометрических образов состоит в том, что они хорошо отражают реальное распределение биометрических признаков среди людей. Обычно предполагают, что чем больше база биометрических образов, тем она точнее отражает действительность. К сожалению, эксперименты, проведенные в Лаборатории при выполнении целого ряда работ, показали, что это не всегда так. Вернее, это может быть так, если представители различных возрастных групп людей, различных профессий и различных темпераментов представлены в тестовых группах в тех же

пропорциях, что и в обществе. Добиться подобной представительности крайне сложно. В связи с этим необходимо проводить специальные исследования, позволяющие оценить представительность полученной тестовой выборки. Под каждый тип биометрических образов должны быть сформированы свои критерии представительности тестовой выборки.

Если критерии представительности тестовой выборки построены, то представительность самой выборки может не зависеть от ее размеров. В частности, может быть искусственно сформирована малая по размерам тестовая выборка, хорошо удовлетворяющая по ее представительности вектору критериев представительности. Эта выборка должна строиться таким образом, чтобы с одной стороны она состояла из реальных или правдоподобно синтезированных биометрических образов [9 – 11], а с другой стороны она должна верно отражать (отображать с заданной погрешностью) вектор выбранных критериев представительности.

В качестве критериев представительности могут выступать:

– статистические характеристики среднестатистического пользователя по некоторому биометрическому параметру (математическое ожидание, среднеквадратическое отклонение, коэффициенты корреляции и другие статистические моменты);

– статистические характеристики среднего по некоторой группе пользователя, выделенной по некоторому биометрическому параметру (математическое ожидание, среднеквадратическое отклонение, коэффициенты корреляции и другие статистические моменты);

– представительность по численности групп пользователей, классифицированных по некоторому биометрическому параметру.

В качестве биометрических параметров могут быть использованы любые параметры, например, стабильность, уникальность, стойкость к атакам подбора.

Заключение

Таким образом, для формирования естественных тестовых баз данных для тестирования средств высоконадежной биометрической аутентификации необходимы большие материальные, людские и временные затраты. Однако без их создания при тестировании нельзя получить полную характеристику стойкости исследуемого устройства. Выходом из создавшегося тупика может стать параллельная работа по созданию реальных и синтезированных на основе реальных искусственных баз биометрических образов, отражающих основные характеристики реальных биометрических образов.

Литература:

1. *Ахметов Б.С.* Основы биометрической аутентификации личности/ Б.С. Ахметов, А.И. Иванов, А.Ю. Малыгин, В.А. Фунтиков/ учебное пособие /Казахстан, Алматы, изд-во КазНТУ им. К.И.Сатпаева. 2014 г., 151 с. ISBN 978-601-228-689-2.

2. *Волчихин В.И.* Быстрые алгоритмы обучения нейросетевых механизмов биометрико-криптографической защиты информации / монография/ В.И. Волчихин, А.И. Иванов, В.А. Фунтиков. Пенза: ПГУ - 2005 г. - 276 с.

3. ГОСТ Р 52633.0 - 2006 «Защита информации. Техника защиты информации. Требования к высоконадежным средствам биометрической аутентификации». - М.: Стандартинформ, 2007.

4. *Малыгин А.Ю.* Формирование больших и сверхбольших баз биометрических образов для тестирования средств высоконадежной биометрической защиты /А.Ю. Малыгин/ Вопросы защиты информации № 4(79) 2007. - С. 17-21.

5. *Волчихин В.И.* О проблеме ресурсов при тестировании стойкости высоконадежных биометрических технологий. /В.И. Волчихин, А.Ю. Малыгин, М.Ю. Лупанов, А.В. Семенов. /Вопросы защиты информации. №4, 2005. М.: Изд-во ВИНИ, С.15 - 16.

6. *Малыгин А.Ю.* Оценка размеров технически реализуемых баз биометрических образов, необходимых для корректного тестирования высоконадежных нейросетевых преобразователей / А.Ю. Малыгин, В.И. Волчихин, В.В. Федулаев, А.В. Безяев /Нейрокомпьютеры: разработка, применение. М.: Радиотехника, №12, 2007. - С.52-54.

7. Основы теории безопасного преобразования биометрических данных в код личного ключа доступа: учеб. пособие [Электронный ресурс]. – Электрон. текстовые, граф. данные (47 Мб, CDR диск) / Ю.К. Язов, О.А. Остапенко, А.И. Иванов, В.А. Фунтиков. Воронеж: ФГБОУ ВПО «Воронежский государственный технический университет», 2014.

8. Федеральный закон от 27.07.2006 N 152-ФЗ (ред. от 29.07.2017) "О персональных данных" Режим доступа: https://fzakon.ru/laws/federalnyy-zakon-ot-27.07.2006-n-152-fz/?yclid=539543931311628356

9. Малыгин А.Ю. Требования к синтетическим базам биометрических образов и генераторам для их формирования / А.Ю. Малыгин, В.В. Федулаев, Д.Н. Надеев /Нейрокомпьютеры: разработка, применение. М.: Радиотехника, №12, 2007. - С.60-64.

10. *Ахметов Б.С.* Оценка рисков высоконадежной биометрии /Б.С. Ахметов, Д.Н. Надеев, В.А. Фунтиков, А.И. Иванов, А.Ю. Малыгин /Монография. Алматы: Из-во КазНТУ им. К.И. Сатпаева, 2014 г.- 108 с.

11. ГОСТ Р 52633.2-2010 «Защита информации. Техника защиты информации. Требования к формированию синтетических биометрических образов, предназначенных для тестирования средств высоконадежной биометрической аутентификации. - М.: Стандартинформ, 2011.

Малыгин Александр Юрьевич – д.т.н., профессор, начальник межотраслевой лаборатории тестирования биометрических устройств и технологий» ФГБОУ ВО «Пензенский государственный университет», mal890@yandex.ru