Ложников П.С.

БИОМЕТРИЧЕСКАЯ ЗАЩИТА ГИБРИДНОГО ДОКУМЕНТООБОРОТА

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

ОМСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

П. С. Ложников

БИОМЕТРИЧЕСКАЯ ЗАЩИТА ГИБРИДНОГО ДОКУМЕНТООБОРОТА

Монография

НОВОСИБИРСК
ИЗДАТЕЛЬСТВО СИБИРСКОГО ОТДЕЛЕНИЯ
РОССИЙСКОЙ АКАДЕМИИ НАУК
2017

УДК 004.056
ББК 32.97
Л71

Ложников П. С.

Биометрическая защита гибридного документооборота: монография / П. С. Ложников; М-во обр. и науки РФ, ФГБОУ ВО «ОмГТУ». — Новосибирск: Изд-во СО РАН, 2017. — 130 с.

Очерчен круг актуальных проблем для систем защиты смешанного документооборота. Предложено перейти к концепции гибридного документооборота, ключевым отличием которой является использование биометрических признаков при формировании закрытого (секретного) ключа ЭП (ЭЦП).

Разработаны модель и технология защиты гибридного документооборота на основе биометрических данных рукописных образов, клавиатурного почерка и лица. Проанализированы указанные признаки, оценена их информативность.

Рассмотрены современные алгоритмы формирования решений при распознавании субъектов и генерации ключевых последовательностей на основе биометрических данных (нечеткие экстракторы, нейросетевые преобразователи биометрия—код на основе персептронов и алгоритма обучения по ГОСТ Р 52633.5-2011, сети квадратичных форм, многомерных функционалов Байеса и других функционалов). Определены оптимальные алгоритмы для решения поставленных задач. Выявлен и экспериментально подтвержден ряд важных тезисов. Подготовка монографии осуществлена при поддержке РФФИ (гранты № 16-07-01204, 16-37-50049).

Рецензенты:
Р. В. Мещеряков, доктор техн. наук,
М. Ю. Хачай, доктор физ.-мат. наук

Утверждено к печати научно-техническим советом
Омского государственного технического университета
Протокол № 8 от 05.09.2017

ISBN 978-5-7692-1561-2

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

В основу данной монографии положены результаты НИОКР, полученные автором и руководимым им коллективом. Работая с биометрической подписью, мы увидели перспективность этой технологии. В настоящее время такой привычный способ подтверждения аутентичности документов заменяется электронной подписью. Вследствие этого обстоятельства утрачивается первоначальный смысл использования автографа. Предлагается принципиально новая технология использования биометрической подписи, и в связи с этим возникает понятие гибридного документа [1—4], существующего на бумажном и электронном носителях и защищенного с помощью электронной подписи, которая формируется с использованием биометрических данных его владельца.

За последние десятилетия электронные системы управления документами стали необходимым средством оснащения офиса любой компании. Подавляющее большинство офисных документов сегодня создается с помощью информационных технологий. Однако в России полный переход в обозримом будущем на «безбумажные технологии» не произойдет по нескольким причинам. Во-первых, существуют требования законодательства и нормативных актов к оформлению наиболее значимых в деловой деятельности документов исключительно на бумаге (уставные документы, лицензии, кадровое делопроизводство и т. п.). Во-вторых, большинству современных руководителей при принятии решений удобнее традиционно работать с бумагой. Еще не выросло то поколение управленцев, которое использовало электронные средства коммуникации со школьной скамьи. Поэтому на настоящий момент реально можно говорить лишь о переходе от бумажного документооборота к смешанному.

Традиционными методами защиты электронных документов являются шифрование и электронная подпись (ЭП, ранее использовалось устаревшее на данный момент идентичное понятие электронно-цифровой подписи, ЭЦП). Современные алгоритмы шифрования надежны при условии использования длинных случайных ключей-паролей. Однако в силу «человеческого фактора» либо требования к генерации паролей не

выдерживаются на практике, либо генерируемые пароли и ключи хранятся ненадлежащим образом. Вопросы безопасного хранения и передачи ключей (паролей) по каналам связи являются нетривиальными, их проработка требует внушительных финансовых затрат. В результате нередко данным вопросам уделяется недостаточно внимания и аутентификаторы субъектов теряются или попадают к третьим лицам.

Существенный недостаток ЭП — она может быть передана другому лицу, т. е. в отличие от традиционной подписи она является отчуждаемой от своего владельца. При этом будет сохраняться юридическая значимость документов, подписанных ЭП посторонним лицом, что может привести к катастрофическим последствиям для бизнеса и репутации человека. Именно поэтому в рамках Национальной технологической инициативы (НТИ)[1] ставятся задачи по разработке и внедрению систем электронной подписи с биометрической активацией. Такие системы позволят избежать фальсификаций юридически значимых решений. Именно созданию технологии ЭП с биометрической активацией и ее использованию для защиты документов на любых видах носителей посвящена настоящая работа.

Квалифицированный «инсайдер», имеющий легитимный доступ к ЭП руководителя или получивший его в результате несанкционированных действий, воспользовавшись привилегиями своего положения в компании, — вот на данный момент портрет наиболее опасного потенциального злоумышленника. Многие руководители пока не готовы и не считают нужным регулярно сидеть за компьютером и подписывать документы, если они не имеют большого значения для деловой деятельности. Данное обстоятельство является началом большинства злоупотреблений в этой сфере.

Для подтверждения актуальности выбранного направления обратимся к мировой статистике компьютерных преступлений. По данным глобального ежегодного исследования PricewaterhouseCoopers (PwC), число инцидентов, связанных с нарушением информационной безопасности (ИБ), и размеры причиняемого ими ущерба неуклонно растут [5, 6]. Согласно этим исследованиям, в 2009 г. PwC зафиксировано 3,4 млн инцидентов, в 2010 — 9,4, в 2011 — 22,7, в 2012 — 24,9, в 2013 — 28,9, в 2014 — 42,8 млн [5]. При этом оценки среднего ущерба от одного инцидента ИБ в зависимости от масштабов бизнеса компании исчисляются от 0,41 до 5,9 млн долларов США [5]. По другим данным, в 2015 г. заре-

[1] НТИ — это долгосрочная комплексная программа по созданию условий для обеспечения лидерства российских компаний на новых высокотехнологичных рынках, которые будут определять структуру мировой экономики в ближайшие 15—20 лет.

гистрирован рекордный ущерб в мире от утечек информации — более 29 млрд долларов (ранее эта цифра не поднималась выше 25,11 млрд) [7]. Российская картина преступлений в сфере информационной безопасности также показывает необходимость внедрения новых мер защиты. По данным InfoWatch Россия оказалась на 2 месте (после США) в мире по числу утечек конфиденциальной информации [8]. В целом к 2016 г. количество крупных инцидентов в мире увеличилось на 38 %, в России — в 2,5 раза [6]. Таким образом, данные на 2016 г. говорят о ежегодном увеличении убытков от нарушений информационной безопасности. Следует признать, что либо средства защиты документов недостаточно эффективны, либо они неэффективно применяются. Сформировалась точка зрения: традиционные методы защиты информации не позволяют эффективно бороться с некоторыми видами угроз, так как не решают проблему «человеческого фактора» [9]. Учитывая ситуацию, все больше крупных компаний (Yahoo!, Google) разрабатывают альтернативные способы аутентификации, в том числе биометрические [10]. По данным PwC, на октябрь 2016 г. 57 % организаций используют биометрические системы аутентификации [11]. Тем не менее, в работах по защите документооборота уделяется недостаточно внимания биометрическим методам [12—14].

Конечно, решить проблему «человеческого фактора» полностью нельзя, так как в противном случае человечеству придется отказаться от свободы воли. Любая попытка ее кардинального решения является утопией. В реальной жизни человек контролирует машины, поэтому всегда есть вероятность совершения преднамеренных и непреднамеренных действий в обход установленных правил (по этой причине, например, нельзя гарантировать конфиденциальность информации, размещенной на бумаге). Тем не менее, можно существенно (многократно) уменьшить масштаб проблемы, если исключить следующие основные факторы, которые ее усиливают:

1. Отчуждаемые от владельца аутентификаторы всегда могут быть переданы третьим лицам (умышленно или нет). По этой причине ключ (пароль) можно забыть, потерять, подменить, украсть.
2. Известно, что защищенность системы определяется самым слабым звеном в системе безопасности, поэтому нужно стремиться к обеспечению примерно равного уровня защиты для документов на любых видах носителей, чего на практике не достигается.

Можно сказать, что эти два тезиса являются основанием для проведения исследований, изложенных в настоящей монографии.

Первая проблема решается, если найти устойчивые и абсолютно надежные преобразования для осуществления однозначной и неотъемлемой «привязки» ключа-пароля (секретного ключа ЭП) к биометрическим

характеристикам каждой конкретной личности, тогда вопросы хранения и распространения всех видов аутентификаторов можно будет также считать закрытыми. Вторая — разработкой соответствующей модели защиты документооборота.

Подводя черту под всем сказанным, заключим: на сегодня реальная практика требует внедрения эффективных биометрических систем защиты документов на электронных и бумажных носителях от угроз нарушения конфиденциальности (по крайней мере, для электронных документов) и целостности, с обеспечением их аутентичности. В настоящей монографии описываются результаты исследований, опираясь на которые можно реализовать техническое решение по защите документов в соответствии с указанными требованиями. В рамках проведенных исследований разработана технология биометрической защиты гибридного документооборота.

1. МОДЕЛЬ ЗАЩИТЫ ГИБРИДНОГО ДОКУМЕНТООБОРОТА

1.1. Нормативно-правовая и терминологическая база в области защиты документов биометрическими методами

Документ возник для того, чтобы зафиксировать информацию и придать ей юридическую силу. Со временем наиболее распространенным материальным носителем стала бумага, соответственно, документооборот, использующий информацию на бумажных носителях, называют «бумажным». В ходе развития науки и технологии человечество вступило на путь информатизации. Информацию все чаще стали хранить на компьютере, т. е. в электронном, или цифровом, виде. Постепенно сформировались понятия электронного документа и электронного носителя информации.

Существуют несколько определений для термина «документ», в частности, в ГОСТ Р 51141-98 (на момент написания настоящей работы стандарт более не являлся действующим): «Документ (документированная информация) — зафиксированная на материальном носителе информация с реквизитами, позволяющими ее идентифицировать». Позднее в ГОСТ Р ИСО 15489-1-2007 [15] дана более подробная формулировка, где документ определяется как «зафиксированная на материальном носителе идентифицируемая информация, созданная, полученная и сохраняемая организацией или физическим лицом в качестве доказательства при подтверждении правовых обязательств или деловой деятельности». На смену ГОСТ Р 51141-98 введен ГОСТ Р 7.0.8-2013, где отдельно выделяются понятия «документ» и «документированная информация». В общем случае «документ» определяется как «зафиксированная на носителе информация с реквизитами, позволяющими ее идентифицировать» [16]. Также в новом стандарте вводятся понятия: официального, архивного, электронного документа, подлинника и дубликата документа, юридической силы и значимости документа, а также даны многие другие определения из данной предметной области. Приведем основные и наименее очевидные из тех, которые могут затрагиваться в настоящем исследовании [16]:

- *носитель информации*: материальный объект, предназначенный для закрепления, хранения (и воспроизведения) речевой, звуковой или изобразительной информации;
- *юридическая значимость документа*: свойство документа выступать в качестве подтверждения деловой деятельности либо событий личного характера;
- *юридическая сила документа*: свойство официального документа вызывать правовые последствия;
- *аутентичность*: свойство электронного документа, гарантирующее, что электронный документ идентичен заявленному;
- *целостность*: состояние электронного документа, в который после его создания не вносились никакие изменения;
- *гриф ограничения доступа*: реквизит, свидетельствующий об особом характере информации документа и ограничивающий доступ к нему;
- *документооборот*: движение документов в организации с момента их создания или получения до завершения исполнения или отправки;
- *конвертирование*, или *конвертация*: процесс перемещения электронных документов с одного носителя на другой или из одного формата в другой;
- *миграция*: перемещение электронных документов из одной информационной системы в другую с сохранением аутентичности, целостности, достоверности документов и их пригодности для использования;
- *подпись*: реквизит, содержащий собственноручную роспись должностного или физического лица.

Последний термин по смыслу полностью совпадает понятием *автографа*, которое является более предпочтительным, так как подчеркивает, что речь идет именно о рукописном открытом биометрическом образе. Помимо открытых образов существуют секретные (тайные) образы — *рукописные пароли*. Далее под *подписью*, или *рукописным образом*, будем подразумевать совокупность нескольких рукописных символов либо слово, которые воспроизводятся рукой подписанта, а также все данные об этом процессе, которые регистрируются с использованием устройства ввода. Понятия *автограф* и *рукописный пароль* будем использовать для явного обозначения открытого и тайного рукописного образа.

Согласно Общероссийскому классификатору управленческой документации (ОКУД), *аналоговые* (*бумажные*) документы по виду делятся на: подлинник (первый или единичный экземпляр документа), дубликат (повторный экземпляр подлинника документа, имеющий юридическую

силу), копия (документ, полностью воспроизводящий информацию подлинного документа и все его внешние признаки или часть их, не имеющий юридической силы), заверенная копия (копия документа, на которой в соответствии с установленным порядком проставляются реквизиты, придающие ей юридическую значимость и силу), выписка (копия части документа, оформленная в установленном порядке). Многие из этих понятий дублируют определения, данные в ГОСТ Р 7.0.8-2013 [16], однако в них явным образом указывается, какой тип документа имеет юридическую силу, т. е. может вызывать правовые последствия.

Несложно распространить понятия *миграции* и *конвертации* также на бумажные версии документа.

В федеральном законе № 149 (ФЗ) «Об информации, информационных технологиях и о защите информации» имеются более детальные определения некоторых важных понятий [17]:

- *конфиденциальность информации*: обязательное для выполнения лицом, получившим доступ к определенной информации, требование не передавать такую информацию третьим лицам без согласия ее обладателя;
- *оператор информационной системы*: гражданин или юридическое лицо, осуществляющие деятельность по эксплуатации информационной системы, в том числе по обработке информации, содержащейся в ее базах данных;
- *электронный документ:* документированная информация, представленная в электронной форме, т. е. в виде, пригодном для восприятия человеком с использованием электронных вычислительных машин, а также для передачи по информационно-телекоммуникационным сетям или обработки в информационных системах.

В связи с простотой создания копий электронных документов введено понятие «реализации электронного документа», что следует понимать как отдельный элемент множества, представляющего электронный документ, существующий или который может существовать в части электронной или цифровой среды, предоставляющей документу определенные возможности и накладывает на него определенные ограничения [18]. У электронного документа может существовать неограниченное множество эквивалентных реализаций, имеющих юридическую силу. Для документов многократного действия (законы, руководящие документы, приказы и т. д.) достаточно доказать существование множества его реализаций, что возможно на основе единственной, имеющей соответствующие атрибуты. Для документов однократного или кратного действия (билеты, абонементы и т. д.) необходимо дополнительно гарантировать кратность применения. Например, ограничить срок действия до-

кумента и логически отслеживать кратность регистрации его реализаций в течение срока его действия [18, 19].

Электронный документ признается равнозначным документу на бумажном носителе, если он подписан электронной подписью или иным аналогом собственноручной подписи [17]. В ФЗ № 63 «Об электронной подписи» [20] дано схожее с ГОСТ Р 7.0.8-2013 определение этого термина:

электронная подпись — информация в электронной форме, которая присоединена к другой информации в электронной форме (подписываемой информации) или иным образом связана с такой информацией и которая используется для определения лица, подписывающего информацию.

Федеральным законом № 63 предусмотрены два типа электронных подписей: простая и усиленная. Простая ЭП является комбинацией из логина и пароля. С ее помощью можно подтвердить, что электронное сообщение отправлено определенным лицом. Усиленная ЭП имеет две формы: квалифицированная и неквалифицированная. Неквалифицированная ЭП не только идентифицирует отправителя, но и подтверждает, что с момента подписания документ не менялся. Сообщение с простой или неквалифицированной ЭП может быть приравнено к бумажному документу, подписанному собственноручно, но только в предусмотренных ФЗ № 63 случаях и по предварительной договоренности сторон. Квалифицированная электронная подпись подтверждается сертификатом от аккредитованного удостоверяющего центра и во всех случаях приравнивается к бумажному документу с «живой» подписью.

Отметим, что ранее в соответствии с устаревшим на данный момент ФЗ № 1 (ФЗ № 63 заменил ФЗ № 1) использовалось практически эквивалентное понятие электронной цифровой подписи (ЭЦП). Для простоты будем считать понятия ЭЦП и усиленной ЭП равнозначными, а под ЭП будет подразумеваться именно усиленная электронная подпись (без уточнения — квалифицированная или неквалифицированная), так как нас в рамках монографии интересует понятие электронной подписи прежде всего как криптографического механизма. Приведем другие важные определения, касающиеся ЭП [20]:

- *ключ электронной подписи*: уникальная последовательность символов, предназначенная для создания электронной подписи;
- *ключ проверки электронной подписи*: уникальная последовательность символов, однозначно связанная с ключом электронной подписи и предназначенная для проверки подлинности электронной подписи.

Как можно видеть, понятия *ключа электронной подписи* и *ключа проверки электронной подписи* идентичны понятиям *открытого* и *закрытого* (*секретного*) ключей для асимметричного шифрования. В рамках настоящей работы будем пользоваться терминами из криптографии.

Несмотря на очевидные преимущества электронных документов, на сегодня по множеству причин (существуют требования законодательства к оформлению наиболее значимых документов на бумаге, большинству руководителей при принятии юридически значимых решений удобнее работать с бумагой) почти на любом предприятии в том или ином виде реализуется смешанный документооборот. На данный момент нормативная и законодательная база в отношении обеспечения безопасности документооборота включает ФЗ № 152 «О персональных данных», ФЗ № 149 [17], ФЗ № 63 [20], ГОСТ Р 7.0.8-2013 [16], ГОСТ Р ИСО 15489-1-2007 [15], а также ряд постановлений, приказов, требований и рекомендаций со стороны Правительства РФ, Центрального банка РФ, ФСТЭК, Министерства связи и массовых коммуникаций РФ (приказы № 17, 32, 41, 81, 92, 120, 151, 187, 288, 320, 321, 360, 533, 1214-р, постановление «О видах электронной подписи, использование которых допускается при обращении за получением государственных муниципальных услуг») и др.

Рассмотрим также семейство отечественных стандартов ГОСТ Р 52633, которые устанавливают требования к процедурам обработки биометрической информации и преобразователям нечетких биометрических образов субъекта в его длинный пароль или ключ, используемый для аутентификации. Базовым стандартом семейства является ГОСТ Р 52633.0-2006 [21], где приводятся принципиально важные для настоящей работы определения, многие из которых необходимо раскрыть в настоящей монографии:

- *биометрический образ*: образ человека, полученный с выходов первичных измерительных преобразователей физических величин, подвергающийся далее масштабированию и иной первичной обработке с целью извлечения из него контролируемых биометрических параметров человека;
- *биометрические параметры*: параметры, полученные после предварительной обработки биометрических данных;
- *преобразователь «биометрия—код»* (*ПБК*): преобразователь, способный преобразовывать вектор нечетких, неоднозначных биометрических параметров «Свой» в четкий однозначный код ключа (пароля). Преобразователь, откликающийся случайным выходным кодом на воздействие случайного входного вектора, не принадлежащего множеству образов «Свой»;
- *нейросетевой преобразователь «биометрия—код»* (*НПБК*): заранее обученная искусственная нейронная сеть с большим числом входов и выходов, преобразующая частично случайный вектор входных биометрических параметров «Свой» в однозначный код криптографического ключа (длинного пароля) и преобразующая

любой иной случайный вектор входных данных в случайный выходной код;

- *динамический биометрический образ*: биометрический образ, изменяемый человеком по своему желанию, например, рукописный образ слова-пароля;
- *статический биометрический образ*: образ, данный человеку от рождения, не изменяемый по воле человека, например, рисунок отпечатка пальца;
- *тайный биометрический образ*: биометрический образ, сохраняемый пользователем в тайне;
- *открытый биометрический образ*: биометрический образ человека, общедоступный для наблюдения;
- *физический муляж*: муляж, выполненный на физическом уровне, исходя из знания физического эффекта, на котором работает датчик считывания биометрического средства защиты и знания индивидуальных особенностей подделываемого на физическом уровне биометрического образа;
- *электронный муляж*: электронные данные, имитирующие биометрические данные пользователя при тестировании или попытках обхода системы защиты;
- *вероятность ошибки первого рода*: вероятность ошибочного отказа «Своему» пользователю в биометрической аутентификации;
- *вероятность ошибки второго рода*: вероятность ошибочной аутентификации «Чужого» как «Своего» (ошибочная аутентификация).

Приведем также несколько ключевых определений из ГОСТ Р 52633.5-2011 [22], требующих дальнейших пояснений:

- *биометрический пример*: совокупность биометрических данных, полученная с выхода первичного преобразователя при однократном предъявлении человеком своего биометрического образа;
- *естественный биометрический образ*: биометрический образ донора, полученный в виде выходных биометрических данных первичного преобразователя и представленный одним или несколькими примерами;
- *синтетический биометрический образ*: биометрический образ, полученный путем имитационного моделирования естественных биометрических образов и представленный одним или несколькими примерами;
- *нейросетевой биометрический контейнер:* организованный по определенным правилам блок данных, содержащий параметры обученного нейросетевого преобразователя биометрия—код доступа;

- *защищенный нейросетевой биометрический контейнер:* контейнер, в котором некоторые части скрыты от непосредственного изучения путем использования обратимого или необратимого преобразования.

Исходя из приведенных определений, термины *биометрический пример* и *биометрический образ* обозначают соответственно единичный образец биометрических данных и их совокупность. Однако в литературе часто путают данные понятия. В настоящей работе может употребляться также термин *образец* (подписи, рукописного пароля, клавиатурного почерка), идентичный по значению понятию *примера*. Также в линейке стандартов определены образ легального пользователя информационной системы (образ «Свой») и злоумышленника, пытающегося преодолеть биометрическую защиту (образ «Чужой»).

Во многих источниках понятие *биометрического параметра* путается с понятием *признака*, которое является идентичным с точки зрения байесовской классификации. В настоящем исследовании под биометрическим признаком и параметром подразумевается одно и то же — некоторая величина, обладающая определенным физическим смыслом и характеризующая субъекта.

Также часто в литературе встречается понятие *биометрического эталона*, или *эталонного описания субъектов*, — некоторого отображения биометрического образа «Свой», описываемого усредненными (обобщенными) значениями биометрических признаков с учетом их вариабельности. Например, эталоном можно назвать параметры распределения значений биометрических признаков или веса искусственной нейронной сети после обучения. Процесс обучения направлен на формирование эталона субъекта.

Продолжая разбираться в терминологии, отметим, что изначально сложилось два основных подхода к реализации ПБК: НПБК [22] и «нечеткие экстракторы» [23]. «*Нечетким экстрактором*» называют метод (или общий алгоритм), выделяющий случайные, равномерно распределенные последовательности битов из биометрических данных в условиях зашумленности. Данный подход активно развивался за рубежом и основан на использовании кодов, исправляющих ошибки, применяемых к «сырым», не обогащенным биометрическим данным для корректировки нестабильных бит генерируемого ключа-пароля. Данный подход иногда рассматривается как частный случай НПБК (сеть с вырожденными нейронами, имеющими по одному входу) [24]. Известны схожие версии изложения данного подхода: Fuzzy Vault («нечеткое хранилище») [25], Fuzzy Commitment [26] и т. д. Некоторые из них обладают бóльшим числом недостатков, чем классический «нечеткий экстрактор», который яв-

ляется общей схемой генерации, построенной на использовании классических самокорректирующихся кодов. Далее объединим все указанные и аналогичные схемы общим названием — нечеткий экстрактор.

Отметим, что реализация методов «нечеткого экстрактора» или НПБК не зависит от того, что они генерируют: ключ симметричного шифрования, закрытый ключ асимметричного шифрования (секретный ключ ЭП) или пароль. Используемый генератор настраивают на выдачу заданной последовательности бит — *ключевой последовательности*.

При равенстве вероятностей ошибок 1-го (FRR, false reject rate) и 2-го (FAR, false acceptance rate) рода (FRR = FAR) говорят о коэффициенте равной вероятности (EER, equal error rate). Иногда FRR, FAR и EER измеряются в процентах. Под ошибочной аутентификацией подразумевается авторизация злоумышленника, пытающегося преодолеть биометрическую защиту. Если перейти от задачи аутентификации к формулированию задачи генерации секретного ключа ЭП или пароля из биометрических данных, то за ошибки можно принять следующие ситуации. Если генерируется не характерный для субъекта ключ, происходит ошибка 1-го рода, а за ошибку 2-го рода принимается ситуация, при которой ключ, полученный из биометрических данных субъекта, равен или близок (в определенной метрике расстояний) к ключу другого субъекта настолько, что принимается за чужой ключ.

На момент написания монографии серия ГОСТ Р 52633 насчитывала семь принятых (утвержденных) стандартов, и еще один находился на рассмотрении. Они регламентируют требования к средствам высоконадежной биометрической аутентификации, к формированию баз естественных и синтетических (искусственных) биометрических образов, предназначенных для тестирования средств высоконадежной биометрической аутентификации, к процессу их тестирования на стойкость к атакам подбора, к процессу их обучения, к определению близости предъявляемого биометрического образца к эталону, а также определяют интерфейсы взаимодействия с НПБК.

1.2. Трудности выполнения требований «равной защиты» документов на электронных и бумажных носителях

В работах [1—4] описана концепция гибридного документооборота, подразумевающая использование одинаковых средств защиты документа как в электронном, так и в бумажном виде (в отличие от смешанного документооборота), в том числе:

1) использование автографа для сохранения юридической значимости документа;

2) использование электронной подписи для проверки целостности и аутентичности документа.

Обеспечить сравнимый уровень защищенности документов, используя для этого схожие инструменты, одновременно в аналоговой (бумажный носитель) и электронной среде весьма проблематично. Для защиты электронных документов предусмотрено больше возможностей, но нельзя сказать, что в реальной практике они защищены лучше. Бесполезно требовать от специалистов по управлению документами и информационной безопасностью найти общее инженерное решение возникающих проблем с документами обоих видов, так как существует ряд принципиальных проблем выполнения требования «равной защиты»:

1. Электронно-цифровую подпись ЭП невозможно применить к документу (реализации документа) на бумажном носителе.

2. ЭП является отчуждаемой от владельца, последствия данного факта часто являются значительными с точки зрения финансовых потерь [5—8], а нередко катастрофическими.

3. Изображение автографа, который применяется в бумажном документе для сохранения и подтверждения юридической значимости, может быть скопировано с целью последующей фальсификации других документов (как электронных, так и бумажных).

4. Применение изображения автографа при работе с электронными документами не гарантирует, что сам автограф воспроизведен подписантом. Нет быстрого способа автоматизированной проверки аутентичности автографа. Почерковедческая экспертиза — дорогостоящая и длительная процедура.

5. Миграцию документов на бумажном носителе за пределы организации сложно отследить или предотвратить, «человеческий фактор» всегда будет влиять на безопасность систем. Однако действия по созданию реализаций бумажных документов контролируются информационными технологиями. Если документ конфиденциального содержания выводится на печать, это действие не должно оставаться без внимания, требуется подтвердить личность субъекта, инициировавшего данное действие. Процедуру подтверждения личности требуется производить непрерывно в процессе работы субъекта с документом.

Ключом к решению описанных проблем является надежная привязка всех аутентификаторов субъекта (паролей, ключей шифрования и ЭП, кодов доступа и т. д.) к его биометрическим характеристикам. При этом нужно учитывать следующее:

1. Процедура ввода биометрических данных должна быть ненавязчивой, не должна вызывать отторжения у пользователя, усложнять процесс работы, нарушать или усложнять существующие бизнес-процессы в организации.

2. Внедрение новых методов биометрической защиты документов должно быть экономически обоснованным. Плюсом в сложившейся ситуации будет являться возможность реализации средств защиты на стандартном оборудовании компьютерных систем. Закупка нового оборудования не должна быть затратной.

3. Связанный с аутентификатором биометрический образ субъекта должен быть тайным либо его копирование или воспроизведение другими лицами на практике должно быть неосуществимым (очень маловероятным).

1.3. Понятие гибридного документа

Под *гибридным документом* в настоящем исследовании подразумевается юридически значимый документ, который может находиться на электронном или бумажном носителе, содержащий изображение автографа, защищенный с помощью электронной подписи, а также тайных или открытых биометрических образов, благодаря чему можно быстро (за приемлемое время) проверить его целостность и аутентичность независимо от типа носителя. Дополнительно он может содержать гриф ограничения доступа для защиты конфиденциальности при работе с его электронными копиями (реализациями).

Особенностью гибридного документа является наличие следующих атрибутов: электронно-цифровая подпись (ЭП), при формировании которой используются не только секретный ключ, но и биометрические данные пользователя, а также хэш-функция документа и рукописная подпись субъекта [1—4]. Использование биометрических параметров для защиты документов при формировании ЭП является ключевым отличием гибридного документооборота от смешанного.

В работе [4] предложен один из возможных вариантов построения гибридного документа с использованием изображения автографа и двух электронных подписей (одной охватывается текст документа, второй — весь документ вместе с изображением автографа). К недостатку такого решения можно отнести ограниченность биометрической защиты — секретный ключ ЭП не генерируется непосредственно из биометрических данных субъекта. По этой причине требуется применять две электронные подписи. В настоящем исследовании предлагаются другой способ построения гибридного документа и усовершенствованная модель защиты гибридного документооборота с использованием процедуры генерации секретного ключа ЭП из биометрических данных и всего одной ЭП.

1.4. Общие аспекты построения модели защиты гибридного документооборота

Для создания эффективной системы защиты гибридных документов (гибридного документооборота) требуется детально проработать ее модель. Модель есть отображение качественных характеристик явлений мира объектов, описываемого физическими законами, в количественные показатели виртуального мира, описываемого логическими законами. При разработке модели гибридного документооборота прежде всего следует исходить из наличия *двух* кардинально разных сред существования документа: электронной и аналоговой [27].

Переходя к вопросу выбора биометрических образов, применяемых для защиты документов, отметим, что биометрические системы не лишены недостатков: статические образы (отпечатка пальца, сетчатки или радужки глаза и др.) не являются секретными, поэтому их можно скопировать, изготовив физический или электронный муляж (второй вариант нужен для удаленной аутентификации). Тайные биометрические образы могут быть основаны только на динамических биометрических признаках, таких как особенности воспроизведения подписи и рукописных паролей, особенности голоса и клавиатурного почерка. На сегодня динамические признаки дают более высокую долю ошибочных решений при аутентификации, чем статические. Но потенциал тайных динамических образов значительно выше, так как они могут содержать секрет, а их длина неограниченна [28].

Учитывая приведенные ранее доводы, для защиты документов на этапе хранения и в процессе перемещения наилучшим образом подходят рукописные образы. Автограф субъекта уже давно используется для подтверждения аутентичности бумажных реализаций документов и является реквизитом, придающим им юридическую значимость и силу, однако его изображение не является секретным. В секрете содержатся динамические параметры его воспроизведения (параметры давления, скорости перемещения пера и другие особенности). Поэтому на практике для подтверждения аутентичности наиболее важных документов автограф рекомендуется использовать совместно с рукописным паролем или короткой парольной фразой. Автограф можно также рассматривать как дополнительный идентификатор субъекта (помимо логина или ФИО), а рукописный пароль — как аутентификатор.

Как уже было указано, организовать процесс непрерывной проверки того, кто работает с бумажной реализацией документа, невозможно (как и победить полностью проблему «человеческого фактора»). Однако это можно реализовать для электронных реализаций гибридных документов.

В некоторых случаях целесообразно дополнительно обеспечить скрытость этой операции [9, 29]. Скрытый мониторинг позволяет выявить потенциального нарушителя с точностью до рабочего места, а при наличии видеонаблюдения — определить личность внутреннего нарушителя. При регистрации несанкционированного доступа к определенным данным могут быть применены методики дезинформации, чтобы задержать злоумышленника и дать время службе безопасности на его поимку «с поличным» [30]. При этом конфиденциальная информация должна быть зашифрована на криптографическом ключе, зависящем от биометрических данных человека, имеющего к ней доступ в соответствии с его полномочиями (чтобы нельзя было пойти «в обход» системы защиты). Для этой цели могут быть использованы биометрические параметры лица, клавиатурного почерка. Эти характеристики можно зарегистрировать на любом современном штатном компьютере посредством стандартного оборудования (клавиатура, веб-камера).

Сформулируем основные этапы работы с гибридным документом в соответствии с предлагаемой моделью его защиты (рис. 1.1):

1. Обучение биометрической системы генерации секретных ключей ЭП. Включает обучение и настройку преобразователя биометрия—код, по окончании которых формируются эталоны, используемые для дальнейшей аутентификации субъектов. Производится сохранение полученных эталонных описаний субъектов на сервере.

2. Идентификация субъекта, генерация секретного ключа, формирование защищенного гибридного документа. На сервер отправляются ло-

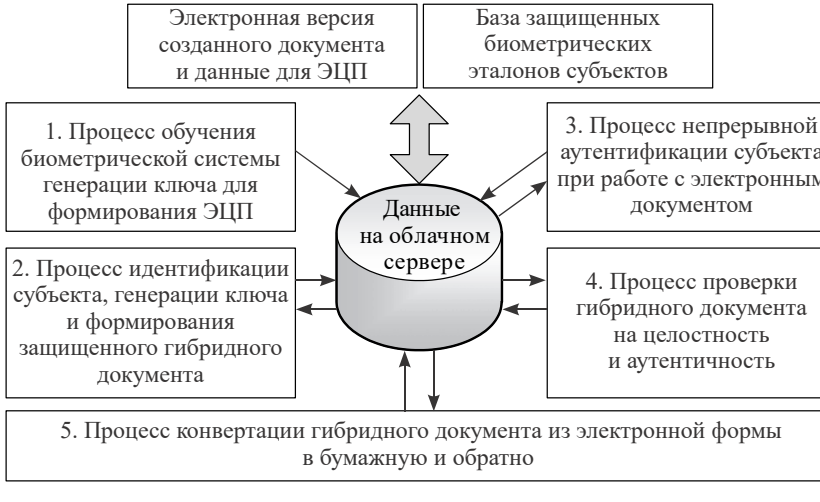

Рис. 1.1. Структурная схема системы защищенного гибридного документооборота.

гин, автограф субъекта (вместе с секретным рукописным образом) и текст документа, после формирования сервер возвращает субъекту готовый к использованию гибридный документ с информацией для проверки целостности и аутентичности.

3. Непрерывная аутентификация субъекта в процессе его работы с электронной версией документа для защиты конфиденциальности информации, содержащейся в нем. Осуществляется контроль биометрических параметров оператора в реальном времени, на основе которых генерируется ключ для расшифровки конфиденциальной информации.

4. Проверка гибридного документа на целостность и аутентичность. Включает сканирование прикрепленной информации, текста документа, сравнение с данными на сервере. При обнаружении нарушения целостности субъект имеет возможность восстановить содержание документа.

5. Перевод гибридного документа из электронной среды в аналоговую и обратно.

1.5. Связывание биометрических характеристик с секретным ключом ЭП владельца и формирование гибридного документа

Основное отличие преобразователей биометрия—код от методов обычной биометрической аутентификации состоит в том, что каждый образец биометрических данных предварительно преобразуется в битовую (ключевую) последовательность, которую возможно использовать в целях аутентификации субъекта или криптографической защиты документов (в качестве кода доступа, ключа шифрования и т. д.). При этом эталон субъекта должен храниться в виде вспомогательной информации, не позволяющей восстановить из нее биометрические характеристики субъекта (рис. 1.2). Требования к защите биометрического эталона при разработке систем высоконадежной биометрической аутентификации прописаны в ГОСТ Р 52633.0-2006 [21] (пункты 5.2—5.3 стандарта):

1. Для усиления стойкости биометрической защиты к атакам изучения и модификации программного обеспечения (ПО) высоконадежные варианты ее технической реализации не должны содержать примеров биометрических образов пользователя, биометрического эталона образов пользователя и кода ключа (пароля) пользователя. Эта информация является конфиденциальной и должна быть защищена при хранении. Кроме того, следы этой конфиденциальной информации должны быть гарантированно уничтожены после выполнения каждой конкретной процедуры аутентификации.

2. Для средств высоконадежной биометрической аутентификации допустимо сокрытие конфиденциальной информации о коде ключа (па-

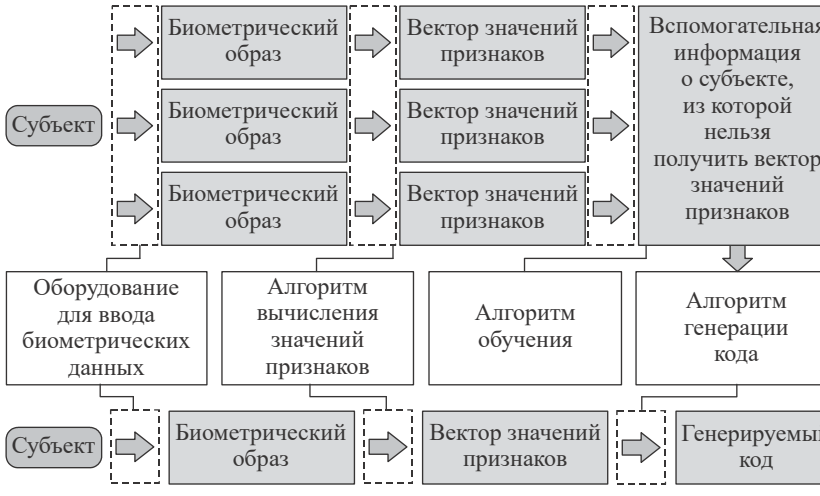

Рис. 1.2. Общая схема преобразователя биометрия—код.

роля) пользователя и его биометрических образах в таблицах параметров и связей нейросетевого преобразователя биометрических параметров в ключ (пароль). Кроме того, допустимо применение и иных способов сокрытия этой информации, например, в форме таблиц преобразователя вектора биометрических параметров в ключ (пароль), использующего нечеткую математическую обработку биометрических данных.

На этапе обучения субъект вводит обучающие примеры рукописных образов, а также образов лица и клавиатурного почерка в процессе непрерывной работы субъекта на компьютере.

В качестве рукописного образа используется изображение автографа в сочетании с его биометрическими характеристиками, что усложняет возможность подделки. Биометрические признаки могут вычисляться из образа автографа (в данном случае биометрический образ будет открытым) или из дополнительно вводимого рукописного пароля (тайного образа).

Для создания эталонных описаний лица и клавиатурного почерка пользователю может быть предложено решить ряд простых задач на компьютере, общей длительностью порядка двух минут. За это время субъект произведет достаточное количество нажатий на клавиши, образ клавиатурного почерка субъекта будет сформирован, будут зафиксированы параметры его лица.

Вопросам вычисления биометрических признаков посвящена вторая глава данной монографии.

На втором этапе (см. рис. 1.1 и 1.3) выполняются идентификация и последующая аутентификация субъекта по параметрам рукописного образа, что позже позволит использовать как изображение автографа, так и биометрические признаки идентифицированного человека для формирования гибридного документа. Субъект вводит логин, воспроизводит рукописный образ (на графическом планшете или планшетном компьютере), после чего полученные данные подписи отправляются на сервер, затем выполняется генерация закрытого ключа ЭП подписанта из предъявленных биометрических данных. Эта процедура является аналогом аутентификации, так как документ не будет аутентичным, если ключ будет сгенерирован неверно. Проверить правильность ключа до формирования ЭП можно следующими способами:

1. Провести шифрование произвольных данных на сгенерированном закрытом ключе и их дешифрование на открытом ключе. Если расшифрованные данные не равны исходным — пользователя можно считать не авторизованным. Данный способ является предпочтительным, так как не требует знания правильного ключа и, следовательно, его хранения на сервере.

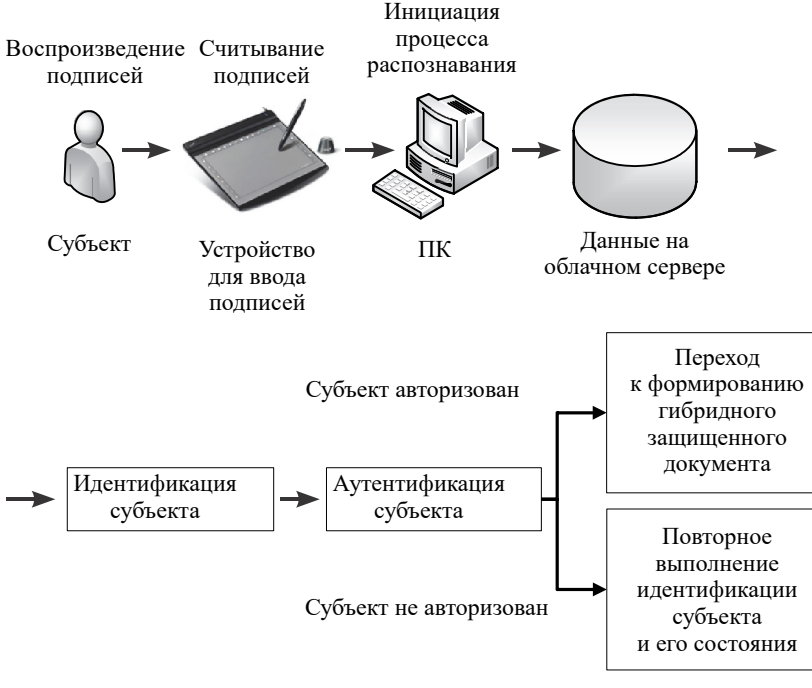

Рис. 1.3. Структурная схема процедуры аутентификации субъекта перед созданием гибридного документа.

2. Сравнить сгенерированный закрытый ключ с оригиналом (правильным ключом). При использовании дополнительной меры близости ключей можно выдавать верный закрытый ключ, если сгенерированный ключ является достаточно близким к нему (отличается на некоторое количество бит). Данный способ проверки противоречит требованиям ГОСТ Р 52633.0-2006.

Идентифицировать субъекта можно не только посредством логина, но и с помощью биометрической процедуры. Одним из вариантов такой процедуры может быть идентификация подписанта по особенностям автографа (открытого образа) с помощью процедуры последовательного применения формулы гипотез Байеса [31—33], которая на данный момент дает вероятность ошибок 1-го и 2-го рода 0,01 и 0,0033 (в сочетании с клавиатурным почерком) [31].

Если субъект авторизован (сгенерированный секретный ключ ЭП является верным), возможен переход к следующему шагу данного этапа — формированию гибридного документа. Иначе необходимо пройти процедуру идентификации с последующей аутентификацией повторно.

При создании гибридного документа (рис. 1.4) формируется электронно-цифровая подпись из хэш-свертки текстовой составляющей документа и закрытого ключа. К документу добавляется изображение автографа. Электронная версия оригинала документа и сформированная ЭП сохраняются на сервере в привязке к защищенному аккаунту их владельца. Формируется ссылка на оригинал документа для его резервного восстановления. Сформированная ссылка на сервер, указывающая на документ, добавляется к файлу электронной версии документа, а при его переводе в бумажную среду заносится в матричный код, называемый QR-кодом [34] (англ. quick response — быстрый отклик). QR-коды разрабо-

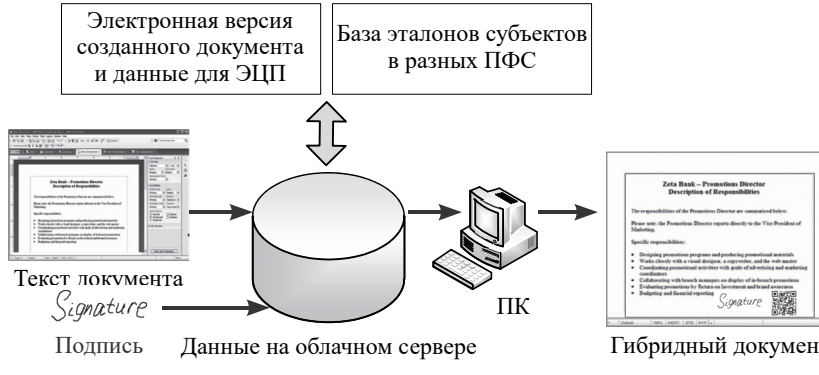

Рис. 1.4. Структурная схема процедуры формирования гибридного документа.

таны в 1994 г. японской корпорацией Denso-Wave для кодирования информации и изначально использовались в промышленности для маркировки деталей при транспортировке. В настоящее время они широко распространены среди пользователей мобильных приложений. По фотографии QR-кода можно моментально получить закодированную информацию о любом предмете, на который он нанесен (вероятность ошибки чтения информации ничтожно мала). В настоящей работе предлагается использовать байтовое кодирование при помощи алгоритма, реализованного в библиотеке QrCode.Net, которая соответствует ГОСТ Р ИСО/МЭК 18004-2015 [34].

После формирования гибридного документа его владелец может ограничить доступ других лиц к произвольным частям электронных реализаций этого документа (или документу целиком), а также запретить определенные действия, совершаемые с документом (печать, редактирование и т. д.), либо всем пользователям, либо всем кроме определенных. При этом содержание каждой из этих частей документа будет зашифровано на открытом ключе того субъекта, которому предоставляется доступ. Если к одной из частей документа имеют доступ более одного субъекта, создается несколько копий этой части, каждая из которых шифруется на соответствующем открытом ключе. К документу будут прикреплены соответствующие грифы ограничения доступа.

На третьем этапе производится непрерывный мониторинг биометрических данных субъекта, осуществляющего работу с ним. Предлагается реализовать данный этап, производя мониторинг (при необходимости скрытый) пользователя в реальном времени, пока он работает с документом, регистрируя при этом параметры его лица и клавиатурного почерка.

Вводимые характеристики будут использоваться для генерации закрытого ключа, применяемого в дальнейшем для расшифровки частей документа, доступ к которым ограничен (рис. 1.5). Техника активной защиты документа от угрозы нарушения конфиденциальности позволит работать с ним только лицам, допущенным к содержащейся в нем информации, и выполнять только те действия с его содержимым, которые разрешил производить над документом его владелец. При фиксации изменений биометрических характеристик субъекта, регистрируемых в процессе работы, документ временно

Рис. 1.5. Иллюстрация процесса получения доступа к фрагменту электронной реализации гибридного документа.

«изменяет» либо «скрывает» свое содержимое целиком или полностью, блокирует часть функций по его редактированию. Пользователь не сможет пойти «в обход» программы, так как вся конфиденциальная информация в документе будет зашифрована на соответствующих криптографических ключах. Блоки конфиденциальной информации дублируются, каждый дубликат содержит копию определенной информации, зашифрованной на открытом ключе определенного субъекта, имеющего к ней доступ. Общедоступная для всех субъектов информация (если таковая имеется в документе) не шифруется.

При наличии на компьютере веб-камеры описанная процедура избирательного доступа к фрагментам документа может задействовать только параметры лица и не требовать от пользователя практически никаких специальных действий, если расположить камеру над монитором.

Отметим, что для каждого документа или типа биометрического образа может быть создана отдельная пара «открытый ключ—закрытый ключ».

Чтобы выполнить проверку целостности и аутентичности документа (этап 4, рис. 1.1 и 1.6), следует извлечь из гибридного документа ссылку на документ, данные о субъекте и связанную с ним информацию об ЭП. В случае обнаружения соответствующей записи на сервере (указывающей на существование документа, подписанного этим субъектом), на клиентский компьютер отправляются данные об автографе и открытом ключе ЭП владельца документа. Происходит считывание текстовой части документа, вычисление его хэш-свертки, ее сравнение с хэш-сверткой, полученной путем расшифровки ЭП открытым ключом. Если целостность не нарушена, документ имеет юридическую силу, в противном случае, используя ссылку на оригинал, можно восстановить истинное содержание документа, сформированного изначально.

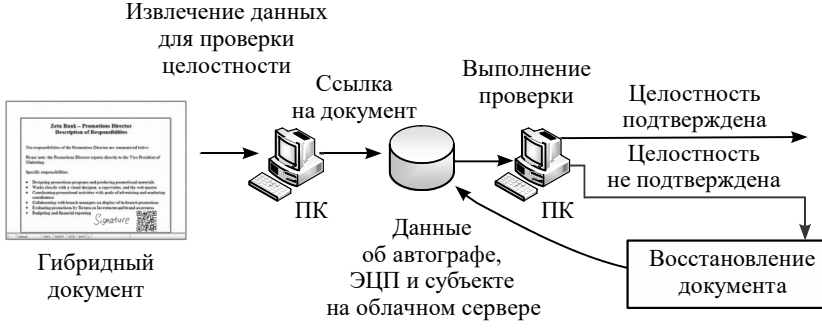

Рис. 1.6. Структурная схема процедуры проверки целостности гибридного документа.

1.6. Достигнутые результаты по надежности генерации ключевых последовательностей и распознаванию субъектов на основе биометрических данных

Эффективность описанных в предыдущем разделе процедур защиты зависит от вероятностей ошибочных решений при генерации секретных ключей ЭП на основе биометрических данных. Экспериментальная часть собственных исследований изложена в пятой главе монографии. В настоящем разделе приведем краткую справку относительно ранее достигнутых результатов по рассматриваемому направлению.

Имеются результаты по генерации ключей (паролей) на основе подписей [35—39]. В [35] проведено исследование на подписях 126 испытуемых (работа поддержана испанским министерством науки и технологий — MCYT TIC2003-08382-C05-01 и европейской комиссией по науке и технологиям — IST-2002-507634 Biosecure NoE projects), получен следующий результат: FRR = 57,30 % при FAR = 1,18 % для профессиональных подделок и FAR = 0,32 % для подделок, выполненных без наблюдения за воспроизведением оригинальной подписи.

В работе [36] получены FRR = 7,05 % и FAR, близкая к нулю (в рамках эксперимента не было ошибок 2-го рода). Однако испытуемых, принимавших участие в эксперименте, слишком мало (всего 11), чтобы говорить о высокой достоверности. Также известен следующий результат: EER ≈ 9 % при длине генерируемого ключа-пароля до 100 бит [37].

В работе [38] получен результат для метода замены скомпрометированных ключей-паролей биометрическими хэшами, получаемыми из рукописных образов субъектов. Значение равной ошибки EER для этого метода составило не более 6,7 % для 40 испытуемых.

В [39] разработан метод выработки ключевой последовательности из тайного рукописного образа (пароля) со следующими показателями ошибок: FRR = 28 %, FAR = 1,2 %. Данный результат был получен без использования кодов, исправляющих ошибки (т. е. нечетких экстракторов).

Известны также следующие результаты по верификации подписей (с предварительной генерацией ключевой последовательности и без):

FAR = 1,07 %, FRR = 6 % [40];

FAR = 7,4 %, FRR = 6,4 % [41];

EER = 3,68 % [42].

Известны результаты по генерации ключевых последовательностей на основе параметров двумерного изображение лица: FRR = 7,99 % при FAR≈0,11 % [43], трехмерного изображения лица: FRR = 22 % при FAR≈0,25 %, длина генерируемой последовательности 155 бит [44].

В [45] приводится результат по распознаванию субъектов по лицу, в соответствии с которым доля верных решений при тестировании на двух базах данных составила 99,6 и 99,5 %. Первый набор данных содержит видеофрагменты, на каждом из которых запечатлен один из 20 испытуемых, по 75 видеофрагментов на каждого субъекта. Второй набор данных содержит по 30 видеофрагментов с каждым из 15 испытуемых (разрешение видео 640 × 480, скорость 15 кадров в секунду, длительность 15 с или более). Оба набора данных были записаны в помещении с неизменным освещением. Другой результат получен в работе [46] и составляет 99 % верных идентификационных решений при использовании базы лиц 214 испытуемых (CMU Face in Action (FIA), разрешение видео 640 × 480, скорость 30 кадров в секунду). Запись испытуемых осуществлялась из различных точек и в различное время (с интервалами в месяц и более). Известен результат из работы [47]: вероятность правильных решений составила 0,988, эксперимент проводился на базе данных из 500 видеоклипов, на которых запечатлено 24 субъекта. Указанные результаты получены на основе анализа видеофрагментов, результаты идентификации по статическому изображению более скромные (так как в видеофайле больше информации и имеется возможность использовать признаки движения). Полученные результаты могли бы устроить потенциального потребителя, если не брать во внимание, что декларируемая надежность достигается далеко не при любых условиях, существуют мешающие воздействия, которые повышают число ошибок до неприемлемых значений (например, маскировка лица субъекта [48]).

Существуют отечественные промышленные решения для идентификации человека по изображению лица. Например, система идентификации ПАПИЛОН-Полифейс компании АО «ПАПИЛОН» (российская производственная и IT-компания, занимающаяся комплексной разработкой и внедрением высокотехнологичных автоматизированных программно-технических систем для государства и бизнеса). Компанией ЗАО «Институт Информационных Технологий» (ЗАО ИИТ) также реализованы программные компоненты для идентификации или верификации личности человека по цифровому изображению. Результаты испытаний систем идентификации личности по изображению лица обеими компаниями в открытой литературе не опубликованы, что оставляет открытым вопрос о показателях надежности работы данных систем.

ПБК на базе клавиатурного почерка характеризуются следующими показателями надежности:

EER = 12 % для 26 испытуемых [49];

FRR = 9 % при FAR = 8 % для 24 испытуемых [50];

FRR = 48,4 % при длине ключа—пароля 12 бит [51];

EER = 6 % при участии в эксперименте 12 испытуемых [52].

В 2004 г. продекларирована вероятность правильной идентификации субъектов по клавиатурному почерку на основе нечеткой логики и метода нечеткой кластеризации С-средних, равная 0,98 (авторы: Mandujano, Soto) [53]. Количество проведенных опытов было равным 3375 при наличии в базе данных 15 эталонов субъектов. В 2005 г. автором Araujo был получен следующий результат по аутентификации пользователей компьютерных систем на основе статистических алгоритмов: FAR = = 1,89 %, FFR = 1,45 % при количестве испытуемых 30 и количестве проведенных опытов 553 [53]. В 2006 г. авторы Lv и Wang достигли показателя EER = 1,41 % при аутентификации пользователей с помощью специальных клавиатур, регистрирующих давление пальцев на клавиши [53], количество испытуемых составило 100, количество опытов — 5000. В 2007 г. привлекают внимание два результата. Первый получен автором Meszaros на основе метода главных компонент и меры Евклида: EER = 10 % при числе испытуемых 25 и 2400 проведенных опытах. Второй получен при помощи метода численной оптимизации (Particle swarm optimization, PSO): FAR = = 0,41 %, FFR = 2,07 % (EER > 1 %) при количестве испытуемых 24 и 4200 опытах (авторы: G. Azevedo, G. Cavalcanti, and E. C. Filho) [53]. В 2009 г. можно также выделить два результата по аутентификации субъектов: EER = 10 % (25 испытуемых, 4500 опытов, автор: Hwang) [53] и аналогичный результат EER = 1 % (10 испытуемых, 300 опытов, авторы: Saevanee, Bhattarakosol), который приводится в работах [53, 54].

Как можно видеть, до 2007 г. исследователям удавалось постепенно повышать надежность распознавания, но за последние годы дальнейшего снижения вероятностей ошибочных решений не обнаружено. Из новых результатов (2015 г.) можно привести работу [55], где декларируется вероятность ошибок около 0,013 (EER = 1,34 %).

Наивысшие известные достоверно подтвержденные результаты, полученные на основе рассматриваемых биометрических образов: признаки — EER, лицо — 0,4 % [45], клавиатурный почерк — 1 % [53], подпись — 3,68 % [42]. То есть достигнутые результаты по надежности распознавания субъектов в пространстве выявленных признаков в целом недостаточно высоки для внедрения на практике, за исключением параметров лица. Однако имеется ряд факторов, негативно влияющих на точность выделения признаков лиц [48]:

- размер и ориентация лица в пространстве (возникает необходимость в нормализации изображения относительно наклона и поворота головы);
- низкое качество изображения, недостаточная яркость или контрастность (применяются специальные методы нормализации);

- макияж и маскировка (борода, усы, очки) существеннее всего влияют на выделение признаков, эффективным способом борьбы с данными мешающими воздействиями является только комплексирование с другими независимыми группами признаков, не подверженных данным мешающим воздействиям.

Причина того что вероятность ошибок распознавания субъекта по подписи значительно выше, вероятнее всего обусловлена накопительной природой признаков лица и клавиатурного почерка. Для распознавания подписанта используется один рукописный образ. Для распознавания оператора по клавиатурному почерку и/или лицу в реальном времени на вход алгоритма принятия решений поступает множество векторов значений признаков: чем дольше время мониторинга, тем больше информации (кадров, на которых зафиксировано лицо, и нажатий клавиш).

В заключение для сравнения приведем известные результаты по генерации ключевых последовательностей на основе статических биометрических образов с последующим распознаванием субъекта:

- на основе отпечатка пальца — FRR = 0,9 % при FAR ≈ 0 %, длина генерируемой последовательности — 296 бит [56];
- на основе радужки (изображения радужной оболочки глаза) — FRR = 0,47 % при FAR≈0 %, длина генерируемой последовательности — 140 бит [57].

Эти показатели могут быть ориентиром при разработке систем генерации секретного ключа ЭП и аутентификации субъектов в пространстве биометрических признаков, так как полностью (с запасом) удовлетворяют потребителя. Настоящая работа была направлена на улучшение показателей, приведенных выше.

1.7. Перевод документа из аналоговой среды в электронную и обратно

Опишем последний этап работы с гибридным документом — конвертацию из одной среды существования в другую. Перенос документа на бумажный носитель происходит при его печати, в процессе чего на бумагу добавляется матричный код, содержащий информацию, которая позже позволит проверить целостность и аутентичность бумажной версии документа, и ссылку на оригинальный электронный документ. При переводе документа из «бумажной формы» в электронную выполняется сканирование текстового содержания документа, затем чтение информации из матричного кода. Таким образом, независимо от формы представления, гибридный документ содержит информацию, достаточную для проверки его целостности и аутентичности. В случае если целостность

гибридного документа нарушена (как умышленно, так и неумышленно), дополнительные данные (ссылка на сервер), добавляемые к нему при его формировании, позволяют восстановить его истинное содержимое.

Распознавание основной текстовой информации документа может проводиться одним из существующих алгоритмов. В настоящее время существует несколько библиотек функций распознавания печатного текста (OCR-библиотек), реализованных в виде готовых программных модулей с открытым исходным кодом. Одна из наиболее точных (поддерживающая русский язык) — Tesseract. В 1995 г. она стала одной из лучших по точности распознавания текста в конкурсе «The Fourth Annual Test of OCR Accuracy» [58]. По данным независимого тестирования [59] на фрагментах текста (размер кегля от 14), доля ошибочных решений обновленной системы Tesseract составила от 0 до 0,57 %. Для снижения количества ошибок может использоваться шрифт Arial (наиболее разборчивый из распространенных шрифтов, число ошибок снижается до 0—0,13 %). Если текст бумажного документа набран исключительно заглавными буквами жирным шрифтом, вероятность ошибки стремится к нулевой отметке. При вычислении хэш-свертки следует игнорировать пунктуацию и регистр символов для компенсации основной доли ошибок. Также хэширование текстовой составляющей документа целесообразно проводить для каждой страницы печатного текста в отдельности, соответственно размещая на них отдельный матричный код. В качестве хэш-функции в простейшем случае можно использовать MD5 (или более современный алгоритм хэширования). Исправление ошибочных разрядов в битовом представлении распознанного текста можно исправить одним из существующих алгоритмов помехоустойчивого кодирования: коды Адамара, БЧХ-коды (в частности Рида—Соломона). Классические коды, исправляющие ошибки, при 50 % избыточности позволяют компенсировать 4 % ошибок (более не требуется) [60]. При кодировании текстовой составляющей гибридного документа будет сформирована битовая строка, содержащая синдромы ошибок и информативную часть кода и которая может храниться на сервере вместе с остальной информацией о документе.

Переходя в практическую плоскость, можно выделить несколько уровней защиты гибридного документа в зависимости от его юридической значимости. Каждому уровню соответствуют определенные параметры отображения текстовой части документа при переводе на бумажный носитель, что влияет на максимально возможное число исправляемых ошибок:

1. Обычный уровень (шрифты Arial, Times New Roman, до 0,57 % ошибок).

2. Повышенный уровень (шрифт Arial, до 0,13 % ошибок).

3. Наивысший уровень (шрифт Arial, заглавные буквы, жирный шрифт, 0 % ошибок).

1.8. Выводы

Резюмируем основной смысл настоящей главы монографии.

Итак, разработана модель защиты гибридного документооборота, которая, во-первых позволяет реализовать требование «равной защиты» документов, находящихся как на электронном, так и на бумажном носителе, что является ключевой проблемой смешанного документооборота, а во-вторых дает возможность запретить передачу секретного ключа ЭП третьим лицам. При создании гибридного документа субъект вводит автограф, тайный или открытый рукописный образ для доступа к секретному ключу ЭП. Модель позволяет защитить документ от подделки, оперативно проверить его целостность и аутентичность и восстановить оригинал, если документ поврежден. Также модель позволяет защитить электронную реализацию документа от угрозы нарушения конфиденциальности содержащейся в нем информации. Для этого можно использовать рукописные образы либо образы лица и/или клавиатурного почерка. Предлагаемая модель не нарушает существующих бизнес-процессов организации и позволяет реализовать защиту с использованием стандартного оборудования компьютерных систем (веб-камера, клавиатура, смартфон или планшет).

Эффективность предложенной модели напрямую зависит от надежности способа генерации секретного (закрытого) ключа ЭП из биометрических данных. Поиску оптимальных способов связи «человек—ключ» и повышению надежности их работы посвящены следующие главы настоящей монографии.

2. БИОМЕТРИЧЕСКИЕ ПРИЗНАКИ ДЛЯ ЗАЩИТЫ ГИБРИДНЫХ ДОКУМЕНТОВ

Прежде всего, необходимо сформировать репрезентативную выборку биометрических образов субъектов. Выборка должна насчитывать достаточное количество биометрических образцов, чтобы в дальнейшем исследовании делать достоверные заключения. Образцы должны быть проанализированы с целью выделения информативных признаков, которые позволяют распознать субъекта и могут использоваться для выработки личного секретного ключа ЭП. Полученные от испытуемых биометрические примеры в дальнейшем использовались для проведения вычислительных экспериментов по оценке надежности разрабатываемых способов генерации ключевых последовательностей и верификации образов испытуемых. Для этого выборка делилась на обучающую и тестовую составляющие.

2.1. Формирование базы биометрических признаков для анализа, обучения автоматов распознавания и проведения экспериментов

Проведен эксперимент по сбору образцов биометрических данных. Сформирована база данных автографов 65 субъектов, для ввода которых испытуемые пользовались графическим планшетом Wacom. В качестве рукописного образа субъекты использовали автограф (примерно в 75 % случаев) или рукописный пароль (25 % случаев). Каждым испытуемым было введено более 50 образцов подписи.

Также сформирована база данных непрерывного мониторинга 100 испытуемых. Все испытуемые решали тестовые задания на компьютере, при этом осуществлялся мониторинг их действий с использованием разработанного программного модуля. Веб-камера была направлена по отноше-

нию к испытуемому таким образом, чтобы было возможно осуществить локализацию лица субъекта (пользователь был повернут лицом к камере либо с незначительным углом отклонения от нее). Тестовые задания создавали необходимость ввода текста на клавиатуре, демонстрировались изображения, вследствие чего субъект вынужден был смотреть на экран (в направлении камеры). Частота съемки составила 15 кадров в секунду, с разрешением видео 800 × 600. Длительность тестового задания составляла 1 ч. Программный модуль скрыто регистрировал следующие данные:

- при обнаружении лица (и основных ключевых особенностей лица) в кадре осуществлялась видеосъемка, разложение видеоданных на последовательность кадров, содержащих только изображение лица;
- при нажатии на клавишу регистрировалось время удержания клавиш и пауза между нажатием этой и предыдущей клавиши.

Собранные данные подвергались статистической обработке, в результате из каждого образца вычислялся вектор значений признаков — величин, характеризующих испытуемых, предоставивших свои биометрические данные.

2.2. Параметры воспроизведения рукописных паролей и автографа субъекта

Рассмотрим биометрические признаки, которые были использованы при разработке технологии защиты гибридного документооборота. Начнем с параметров подписей.

В компьютерном представлении подпись может состоять из функций положения пера на планшете $x(t)$, $y(t)$ и давления пера на планшет $p(t)$, где t — это время в дискретной форме. Будем обозначать значения этих функций через x_i, y_i, p_i. Необходимо определить признаки, характеризующие владельца подписи. Далее использовались признаки из работ [61, 62] (табл. 2.1).

Образцы подписи различаются по продолжительности (количеству отсчетов). Первоначально необходимо привести их к единой продолжительности, выполнив операцию нормирования, состоящую из следующих этапов:

1. Исключаются все отчеты с нулевым давлением в начале и конце подписи.

2. Производится одномерное преобразование Фурье для $x(t)$, $y(t)$ и $p(t)$.

3. Производится обратное преобразование Фурье для указанных функций, учитывая, что размерность на выходе должна соответствовать числу, которое является ближайшим меньшим кратным степени 2.

Таблица 2.1

Признаки рукописных образов, рассмотренные в работе

Краткое описание признаков	Количество признаков
Расстояния между некоторыми точками подписи, нормированные по длине подписи, в трехмерном пространстве (третье измерение — это давление $p(t)$). Точки выбираются равномерно с некоторым шагом, далее находятся расстояния между всеми парами этих точек [61]	120
Нормированные по энергии амплитуды первых 16 (наиболее низкочастотных) гармоник функции скорости перемещения пера на планшете $v_{xy}(t)$ и функции давления пера на планшет $p(t)$. Первые 16 гармоник содержат более 95 % энергии сигналов $v_{xy}(t)$ и $p(t)$, что характерно для всех испытуемых [61, 62]	32
Некоторые характеристики статического изображения подписи: отношение длины подписи к ее ширине, центр подписи, угол наклона подписи, угол наклона между центрами половин подписи [61]	5
Коэффициенты корреляции между всеми парами функций подписи $x(t)$, $y(t)$, $p(t)$ и их производными $x'(t)$, $y'(t)$, $p'(t)$ [61, 62]	15
Некоторые значения функций $x(t)$, $y(t)$ и $p(t)$, а также функции скорости перемещения пера на планшете $v_{xy}(t)$. Точки выбираются равномерно с некоторым шагом [61]	64

Часть пространства признаков формировалась посредством построения матрицы расстояний между отчетами подписи. Элементы r_{ij} (расстояние между i-й и j-й координатами) матрицы в 3-мерном пространстве (давление — третье измерение) вычисляются по формуле

$$r_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (p_i - p_j)^2}. \qquad (2.1)$$

Аналогичным образом вычисляется матрица расстояний в 2-мерном пространстве (без учета давления).

Поскольку при расчете получается слишком много элементов, что требует слишком значительных вычислительных ресурсов, необходимо производить вычисления расстояний с некоторым шагом. Далее производится нормирование полученной матрицы по длине подписи: $r'_{ij} = r_{ij}/r_{12} + r_{23} + \cdots + r_{(n-1)n}$. Нормированные элементы r'_{ij} полученной матрицы являются биометрическими признаками.

Вычисляются некоторые признаки, характеризующие внешний вид подписи:

1. Отношение длины подписи к ее ширине.

2. Центр подписи, описываемый координатами C_X, C_Y, C_P.

3. Угол наклона подписи. Под углом подписи понимается косинус среднего угла наклона ломаной траектории подписи к оси абсцисс:

$$\theta = \frac{1}{N-1}\sum_{i=1}^{N}\frac{x_{i+1}-x_i}{\sqrt{(x_{i+1}-x_i)^2+(y_{i+1}-y_i)^2}}.$$

Угол наклона между центрами половин подписи. После того как был найден центр подписи, разобьем множество $(X, Y, Z) = \{(x_i, y_i, p_i)\}$ на два подмножества $L = \{(x_i, y_i, p_i) | x_i \le C_x\}$ и $R = \{(x_i, y_i, p_i) | x_i > C_x\}$ и найдем центры этих подмножеств:

$$C_{X_L} = \frac{1}{|L|}\sum_{x_i \in L}^{N} x_i, \quad C_{Y_L} = \frac{1}{|L|}\sum_{y_i \in L}^{N} y_i, \quad C_{P_L} = \frac{1}{|L|}\sum_{p_i \in L}^{N} p_i,$$

$$C_{X_R} = \frac{1}{|R|}\sum_{x_i \in R}^{N} x_i, \quad C_{Y_R} = \frac{1}{|R|}\sum_{y_i \in R}^{N} y_i, \quad C_{P_R} = \frac{1}{|R|}\sum_{p_i \in R}^{N} p_i.$$

Следующая категория признаков основана на использовании преобразования Фурье. Давление на планшет и функция скорости пера на планшете $v(t)$ подвергаются разложению по формуле

$$X_k = \sum_{i=0}^{N-1} x_i e^{\frac{-j2\pi ki}{N}}, \tag{2.2}$$

где X_k — k-я гармоника в комплексной форме $\mathrm{Re}_k + j\mathrm{Im}_k$, x_i — i-е значение функции, N — количество отсчетов в дискретном сигнале.

Значения $v(t)$ вычисляются по формуле

$$v_i = \sqrt{(x_{i+1}-x_i)^2+(y_{i+1}-y_i)^2}. \tag{2.3}$$

Исходный дискретный сигнал представляется в виде суммы функций:

$$f(t_i) = \sum_{k=0}^{N-1}\left[\frac{\mathrm{Re}_k}{N}\cos\left(\frac{2\pi k t_i}{T}\right) - \frac{\mathrm{Im}_k}{N}\sin\left(\frac{2\pi k t_i}{T}\right)\right] =$$

$$= \sum_{k=0}^{N-1} A_k \cos\left(2\pi t_n / T_k + \phi_k\right) = \sum_{k=0}^{N-1} A_k \cos\left(2\pi t_i v_k + \phi_k\right) = \sum_{k=0}^{N-1} G_k(t_i).$$

При использовании $v(t)$ исчезает зависимость от того, под каким углом расположен планшет относительно руки подписанта. Можно воспользоваться быстрым (БПФ) или обычным дискретным (ДПФ) преобразованием Фурье. В отличие от ДПФ, которое имеет сложность порядка $O(N^2)$, БПФ имеет сложность $O(N\log_2 N)$.

Далее будем функцию $G_k(t) = A_k \cos(2\pi t_n / T_k + \phi_k)$ называть k-й гармоникой. Амплитуды вычисляются в соответствии с формулой

$$A_k = \frac{1}{N}\sqrt{\mathrm{Re}_k^2 + \mathrm{Im}_k^2}. \tag{2.4}$$

Далее производится расчет энергии функции по формуле

$$E_p = \int_{-\infty}^{\infty} A^2(t)\,dt. \tag{2.5}$$

На следующем шаге производится деление амплитуды каждой гармоники на значение энергии сигнала. Эта операция называется нормированием амплитуд по энергии и осуществляется с целью приведения различных реализаций подписи к одному масштабу. Использовалось 16 нормированных амплитуд первых наиболее низкочастотных гармоник функции давления и функции скорости пера на планшете в качестве признаков по аналогии с работой [62] (16 гармоник содержат более 95 % энергии сигнала и соответствуют частоте колебаний руки подписанта).

Помимо описанных характеристик признаками в настоящей работе являются коэффициенты парной корреляции между функциями $x(t)$, $y(t)$ и $p(t)$ (и их производными). Установлено, что данные коэффициенты корреляции для каждого рукописного образа подписи субъекта близки по значениям и для рукописных образов подписей различных субъектов существенно различаются [63]. Все указанные признаки имеют распределение значений, близкое к нормальному, что проверялось критерием хи-квадрат Пирсона. Общее число признаков 236.

2.3. Параметры лица, регистрируемые в процессе непрерывного мониторинга субъекта

Технологии идентификации по лицу можно разделить на две группы: анализирующие статические изображения [64, 65] и анализирующие видеопоток [66—68]. Во втором случае в качестве признаков могут использоваться микроколебания головы и особенности мимики лица [69—71]. В настоящем исследовании анализировался видеопоток.

Можно выделить следующие категории признаков лица:

1. Физиологические параметры лица (форма контуров лица, бровей, глаз, носа, губ, ушей, подбородка, а также расстояния между ними и их взаимное расположение) [64]. Выделение многих особенностей часто базируется на взаимном расположении глаз и губ. Также используется признак асимметрии лица. Согласно некоторым исследованиям, наиболее узнаваемыми являются люди с асимметричными лицами [64].

2. Параметры подсознательных движений [72]: особенности мимики субъекта, траектории движений головы в процессе ввода пароля, параметры микроколебаний головы (тремора) [67, 68]. Признаками могут быть коэффициенты фурье-разложения (часть амплитудного спектра) функций координат фиксированной точки на лице, например, глаз или носа. Анализ мимики лица может использоваться для защиты от предъявления распечатанного изображения лица. Наличие изменений на лице можно определить по изменению корреляции областей лица на различных кадрах.

3. Изображение радужной оболочки глаза может быть использовано в качестве признака или совокупности признаков. Для получения изображения глаза приемлемого качества необходимо, чтобы пользователь находился на определенном расстоянии от камеры без движения [73], что затрудняет использование данного признака для скрытой идентификации, либо нужна камера с высокой разрешающей способностью (требуется специальное оборудование). При наличии только стандартной веб-камеры удается зафиксировать только общие параметры глаз, например, их цвет.

4. Иные параметры (цвет кожи, плотность излучения оптического потока, коэффициенты вейвлет-разложений фрагментов изображений, например, габор-коэффициенты и т. п.) [74, 75].

Преимущество методов идентификации по лицу в видимом диапазоне — относительная стабильность признаков (низкая по сравнению со статическими биометрическими признаками, в частности отпечатками пальцев [76], но высокая в сравнении с динамическими признаками клавиатурного почерка и подписи). Со временем физиологические параметры лица меняются незначительно (для заметного изменения у взрослого человека должно пройти несколько лет).

В настоящей работе использовались следующие группы признаков лица:

1. Расстояния между глазами, правым (левым) глазом и центром лица, правым (левым) глазом и кончиком носа, правым (левым) глазом и центром рта, центром рта и центром лица, кончиком носа и центром рта, центром рта и кончиком носа (в пикселях, значения нормировались по диагонали лица в кадре).

2. Площади глаз, носа, рта (значения нормировались по площади лица).

3. Коэффициенты корреляции яркости и цветовых составляющих пикселей (в соответствии с моделью RGB) между всеми парами следующих областей лица: правый глаз, левый глаз, нос, рот. Данные признаки характеризуют мимику, асимметрию и непроизвольные движения лица субъектов.

4. Коэффициенты корреляции яркости и цветовых составляющих пикселей (в соответствии с моделью RGB) следующих областей лица,

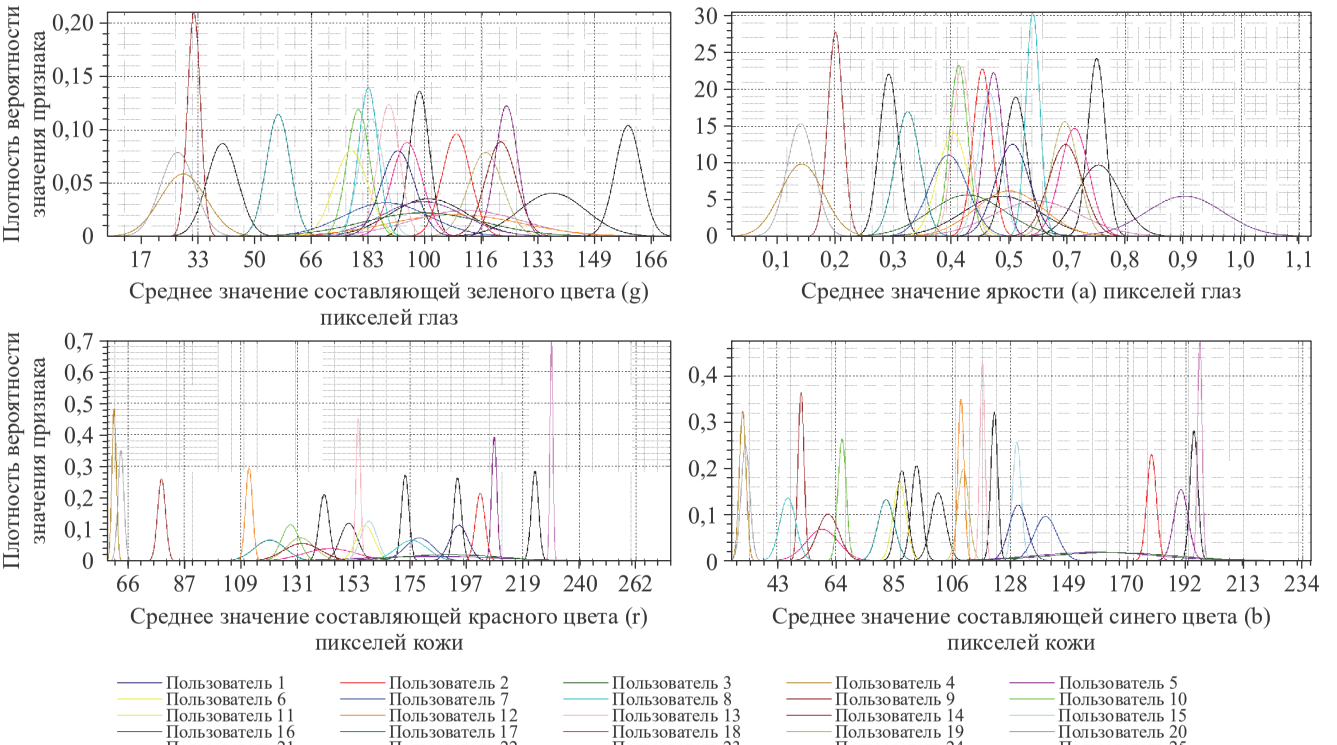

Рис. 2.1. Плотности распределения значений четырех из наиболее информативных признаков лица, характеризующих 25 испытуемых.

выделяемых на соседних кадрах: правый глаз, левый глаз, нос, рот. Данные признаки характеризуют мимику и непроизвольные движения лица.

5. Средние показатели интенсивности яркости, а также красной (R), зеленой (G) и синей (B) составляющих пикселей, характеризующих цвет глаз и кожи. Данные признаки относятся к наиболее информативным из рассматриваемых в настоящей работе, так как обладают наименьшими площадями пересечения функций плотности вероятности собственных значений (рис. 2.1).

Для выделения лица, глаз, носа, рта использовался метод Виолы—Джонса, позволяющий детектировать объекты на изображениях в реальном времени [77]. Данный метод распознавания объектов на изображении является одним из лучших по эффективности распознавания и скорости работы (обладает крайне низкой вероятностью ошибок и распознает черты лица под углом до 30°) [78, 79]. Для разделения области радужки и зрачка производится поиск внутренней границы радужки (внешней границы зрачка) при помощи алгоритма обнаружения окружностей на основе преобразования Хафа [80].

Имеется погрешность вычисления описанных 62 признаков, связанная с точностью работы метода Виолы—Джонса и преобразования Хафа, условиями съемки и особенностями видеозаписей. Все признаки имеют распределение, достаточно близкое к нормальному, что проверялось критерием хи-квадрат.

2.4. Параметры клавиатурного почерка, регистрируемые в процессе непрерывного мониторинга субъекта

Для формирования пространства признаков клавиатурного почерка субъектов обычно используются временные характеристики нажатий клавиш. В работе [81] указано пять типовых характеристик клавиатурного почерка:

- время между моментами, в которые определенная клавиша нажата и отпущена (dwell time);
- время между моментами, в которые одна из клавиш нажата, а другая отпущена;
- время между моментами, в которые одна из клавиш отпущена и следующая клавиша нажата (flight time);
- время между моментами, в которые одна из клавиш нажата и следующая клавиша нажата (паузы между нажатием клавиш);
- время между моментами, в которые очередная клавиша отпущена и следующая клавиша отпущена.

Эти данные могут быть использованы непосредственно в качестве признаков, либо по этим данным может формироваться временна́я функция в процессе ввода парольной фразы [82] с последующим разложением ее при помощи вейвлет-преобразований (например, по ортогональному базису функций Хаара [83]), коэффициенты которых используются в качестве признаков. Иногда используются только наиболее информативные или часто встречаемые в словах сочетания символов [84]. Указанные в работе [81] временны́е характеристики достаточно нестабильны и могут меняться со временем, в том числе с течением дня, а также зависят от навыков работы субъекта с клавиатурой [80]. Известны работы, в которых помимо указанных характеристик анализировалась сила нажатия (давление) на клавиши [85, 86], которую можно измерить специальными малогабаритными датчиками (их необходимо встроить в клавиатуру), а также вибрация клавиатуры при нажатии на клавиши [87].

В последнее время наблюдается рост числа работ по исследованию клавиатурного почерка пользователей мобильных устройств [80, 86, 88]. В работе [80] в качестве признаков также указывается площадь области соприкосновения пальца с сенсорным дисплеем (и другие производные от данной величины признаки).

Авторами работы [89] создана клавиатура с автономным питанием (копится заряд от нажатий клавиш), которая способна не только идентифицировать хозяина, но и определить некоторые личностные характеристики печатающего субъекта. В отличие от традиционных клавиатур, использующих для генерации сигналов о нажатии клавиш механические переключатели для замыкания соответствующих участков электронной цепи, принцип действия клавиатуры [89] основан на трибоэлектрическом эффекте. Трибоэлектрический эффект — это процесс перетекания электрического заряда с одного материала (в данном случае кожи пальца) на поверхность другого (его роль играет полиэтилентерефталат (ПЭТ) — термопластик) при их контакте друг с другом. Этот эффект является одним из видов так называемой контактной электрификации. Формирующиеся при нажатии клавиш сигналы характеризуются напряжением и силой тока, которые являются функциями времени. Эти сигналы коррелируют с особенностями динамики ввода текста, размерами подушечек пальцев и их биоэлектрическим потенциалом. Они не только характеризуют время нажатия, но и количественно описывают конкретные динамические изменения в процессе набора текста. Спектральный анализ измеряемых электрических сигналов показал, что рассматриваемые сигналы у разных людей различаются, проводились эксперименты в группе из 104 человек в возрасте от 14 до 69 лет [89]. Появление работ такого рода указывает на то, что идентификационные возможности клавиатурного почерка еще полностью не исследованы.

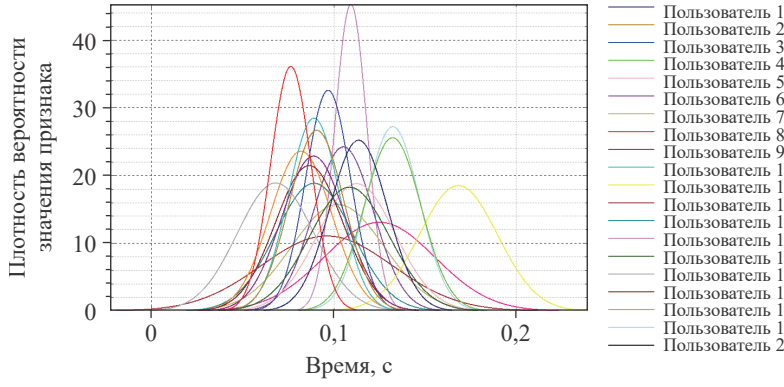

Рис. 2.2. Плотности распределения времен удержания клавиши «ф» 20-ти испытуемых пользователей.

Базовыми признаками являются временны́е интервалы между нажатием клавиш, которые характеризуют темп работы, а также время удержания клавиш, характеризующее стиль работы с клавиатурой. Известно, что значения данных характеристик имеют распределение, близкое к нормальному [90] (рис. 2.2). В качестве признаков клавиатурного почерка для непрерывной скрытой идентификации субъекта в процессе профессиональной деятельности в настоящей работе решено использовать данные временны́е характеристики, так как их можно зарегистрировать с помощью стандартной клавиатуры.

2.5. Оценка информативности и взаимной корреляционной зависимости признаков

Информативность группы признаков определяется как совокупность оценок, полученных по двум основным критериям:

1. Средняя величина и разброс площадей пересечения функций плотностей вероятности признаков, характеризующих различных испытуемых. Данный показатель целесообразно подсчитывать по каждому признаку отдельно для всех возможных пар испытуемых. Часто под информативностью подразумевают именно этот показатель, но измеряемый в битах информации (площади переводятся в биты по формуле Хартли или Шеннона) [24].

2. Средняя величина и разброс значений коэффициентов парной корреляции между признаками (т. е. сечениями признаков, так как значения признаков по сути — это случайные величины).

В зависимости от данных показателей целесообразность использования тех или иных методик принятия решений может меняться (об этом

речь пойдет в следующих двух главах). Вычислим данные показатели и приведем соответствующие графики (рис. 2.3—2.8). Для первого показателя также определим математическое ожидание и среднеквадратичное отклонение M_x и S_x по каждому признаку, чтобы сравнить их информативность (кроме клавиатурного почерка, так как информативность каждого признака этой группы в среднем примерно одинакова, но различается для разных субъектов). Чем больше показатель S_x, тем больше разброс оценок информативности признака для различных субъектов.

По представленным рисункам видно, что признаков клавиатурного почерка, описывающихся очень малой площадью пересечения (от 0 до 0,15) плотностей вероятности для двух произвольных субъектов, заметно меньше по сравнению с признаками лица, а значит, они хуже характеризуют субъектов. Однако они обладают меньшей взаимной корреляционной зависимостью (более 90 % признаков имеют коэффициент парной

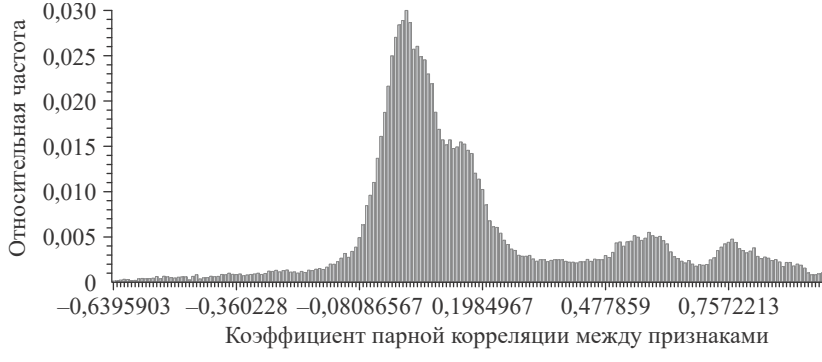

Рис. 2.3. Коэффициенты корреляции между парами признаков подписей.

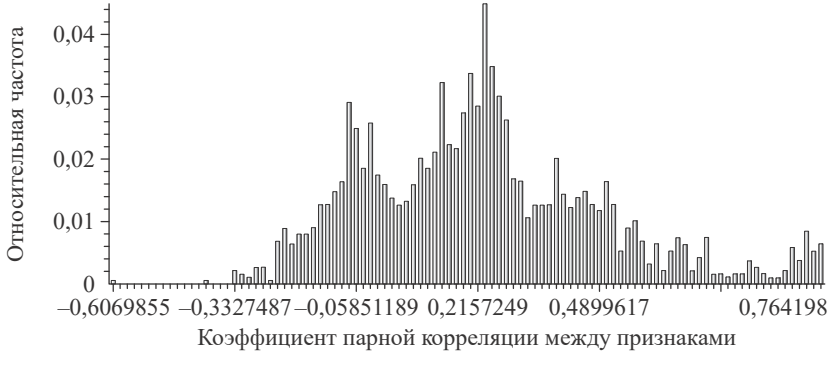

Рис. 2.4. Коэффициенты корреляции между парами признаков лица.

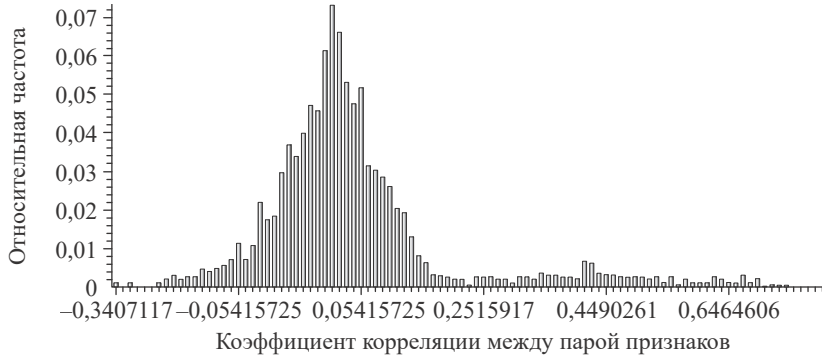

Рис. 2.5. Коэффициенты корреляции между парами признаков клавиатурного почерка.

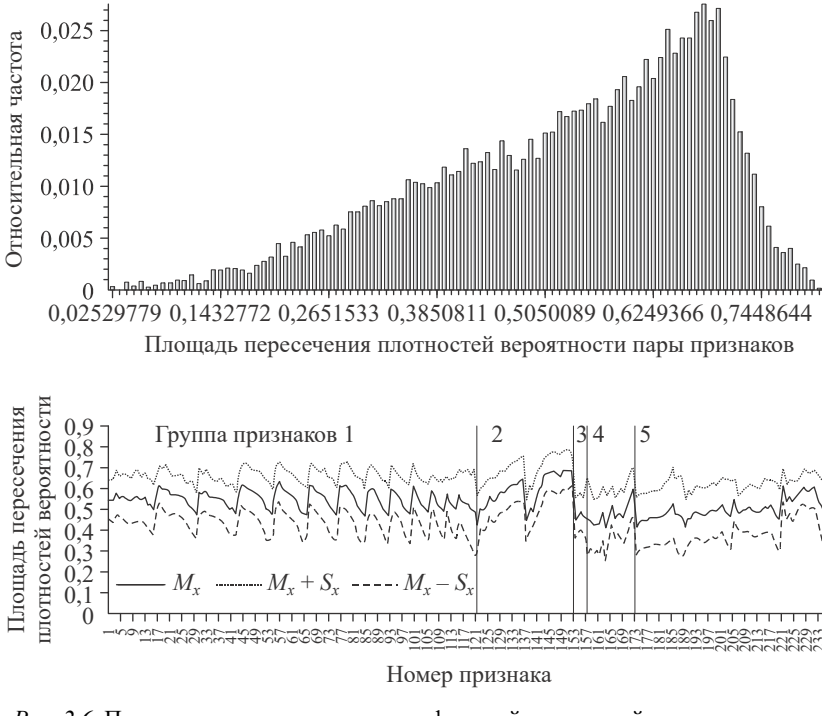

Рис. 2.6. Площади пересечения всех пар функций плотностей вероятности каждого признака подписей, характеризующих различных испытуемых.

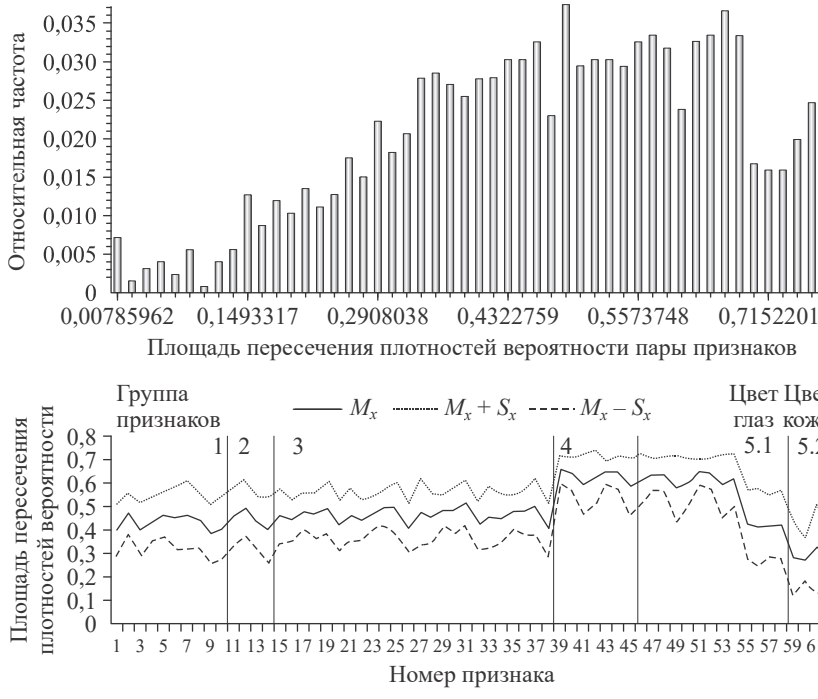

Рис. 2.7. Площади пересечения всех пар функций плотностей вероятности каждого признака лица, характеризующих различных испытуемых.

Рис. 2.8. Площади пересечения всех пар функций плотностей вероятности каждого признака клавиатурного почерка, характеризующих различных испытуемых.

корреляции менее 0,3), что позволяет эффективнее использовать для распознавания по этим признакам функционалы квадратичных форм [91]. Высокий коэффициент корреляции (более 0,9) признаков лица дают красная (r) цветовая составляющая и яркость пикселей. Признаки подписи по интегральным показателям информативности занимают позицию между признаками лица и клавиатурного почерка. Наиболее информативными из рассмотренных признаков являются те, что характеризуют цвет кожи, а также цвет глаз. Как видно из рис. 2.7, для некоторых субъектов площадь пересечения плотности вероятности этих признаков меньше 0,1 (более 15 % субъектов имеют такую площадь, исходя из гипотезы о нормальном распределении данных площадей, которая была принята после проверки критерием хи-квадрат).

Волнообразные линии на рис. 2.6, относящиеся к информативности расстояний между отчетами подписи, говорят о следующем: чем дальше разнесены во времени отчеты, тем менее информативен признак для субъекта. Каждая «волна» относится к нескольким признакам, описывающим расстояние между некоторой точкой подписи и некоторыми другими точками.

Также в качестве признаков рассматривались особенности траекторий перемещения курсора мыши из работ [9, 92]. Однако площади пересечения плотностей вероятности для данных признаков оказались слишком значительны (более 80 % площадей превысило 0,7) и от данных признаков решено отказаться.

2.6. Выводы

В настоящей главе была собрана база для исследований биометрических признаков достаточного объема. Предложено пространство признаков рукописных образов (паролей или автографов), клавиатурного почерка и лица. Признаки рассмотрены с различных сторон. Приводятся оценки взаимной корреляционной зависимости выявленных признаков и площадей пересечения плотностей вероятности их значений. Предложено использовать признаки, характеризующие цвет кожи и цвет глаз, которые являются самыми информативными из всех рассмотренных в настоящей работе. Определение этих характеристик происходит почти безошибочно, при этом данные параметры лица менее всего уязвимы перед таким мешающим воздействием, как маскировка (для удачной маскировки требуется применять грим и надевать линзы).

Распознавание пользователей по лицу уже давно используется на практике. Достигнутую надежность в некоторых случаях можно счи-

тать удовлетворительной для потребителя. Однако имеющийся существенный недостаток — уязвимость перед маскировкой распознаваемых субъектов может быть устранен только при комплексировании данной категории признаков с другими. Для решения поставленных задач логичнее всего объединить параметры лица с параметрами клавиатурного почерка. Анализ клавиатурного почерка — перспективный метод скрытого распознавания пользователей компьютерных систем, так как полностью реализуется на стандартном периферийном оборудовании.

Желательным является достижение надежности генерации секретного ключа ЭП, сравнимой с показателями систем на базе статических биометрических признаков. Результаты предыдущих исследований позволяют сделать осторожный вывод, что в задаче непрерывной аутентификации к этим показателям приблизиться возможно [93—95], превысив достигнутый уровень по каждому направлению биометрических технологий отдельно, выявленный в результате представленного аналитического исследования [48].

3. СПОСОБЫ ГЕНЕРАЦИИ СЕКРЕТНЫХ КЛЮЧЕЙ ЭП НА ОСНОВЕ БИОМЕТРИЧЕСКИХ ДАННЫХ ДЛЯ ЗАЩИТЫ ГИБРИДНЫХ ДОКУМЕНТОВ

Перейдем к рассмотрению методов генерации ключа (пароля) из биометрических данных. Ранее упоминалось (см. раздел 1.1), что на данный момент имеется два подхода для реализации процедур генерации секретного ключа ЭП с возможностью защиты биометрического эталона (этого требует ГОСТ Р 52633.0-2006 [21]): на основе «нечетких экстракторов» и искусственных нейронных сетей. Рассмотрим эти подходы и их модификации подробней и проведем эксперименты по сравнению их эффективности при решении поставленных задач.

3.1. Классическая модель нечеткого экстрактора

«Нечеткие экстракторы» способны компенсировать ошибки, возникающие вследствие технической невозможности получения одинаковых значений биометрических характеристик при их повторном вводе субъектом. Такие алгоритмы базируются на теории информации и помехоустойчивом кодировании и обычно используются для генерации криптографических ключей без необходимости их хранения в промежутках между обращениями к ним [23].

Для генерации секретного ключа ЭП нечетким экстрактором необходимы биометрические данные и дополнительная общедоступная информация, хранящаяся на сервере (носителе), из которой нельзя восстановить эталон (во всяком случае, не существует простого способа этого сделать). Данная информация называется открытой строкой. Сначала генерируется случайная равномерно распределенная битовая последовательность, которая является ключом ЭП. Далее осуществляется помехоустойчивое кодирование ключа (к битовой последовательности добавля-

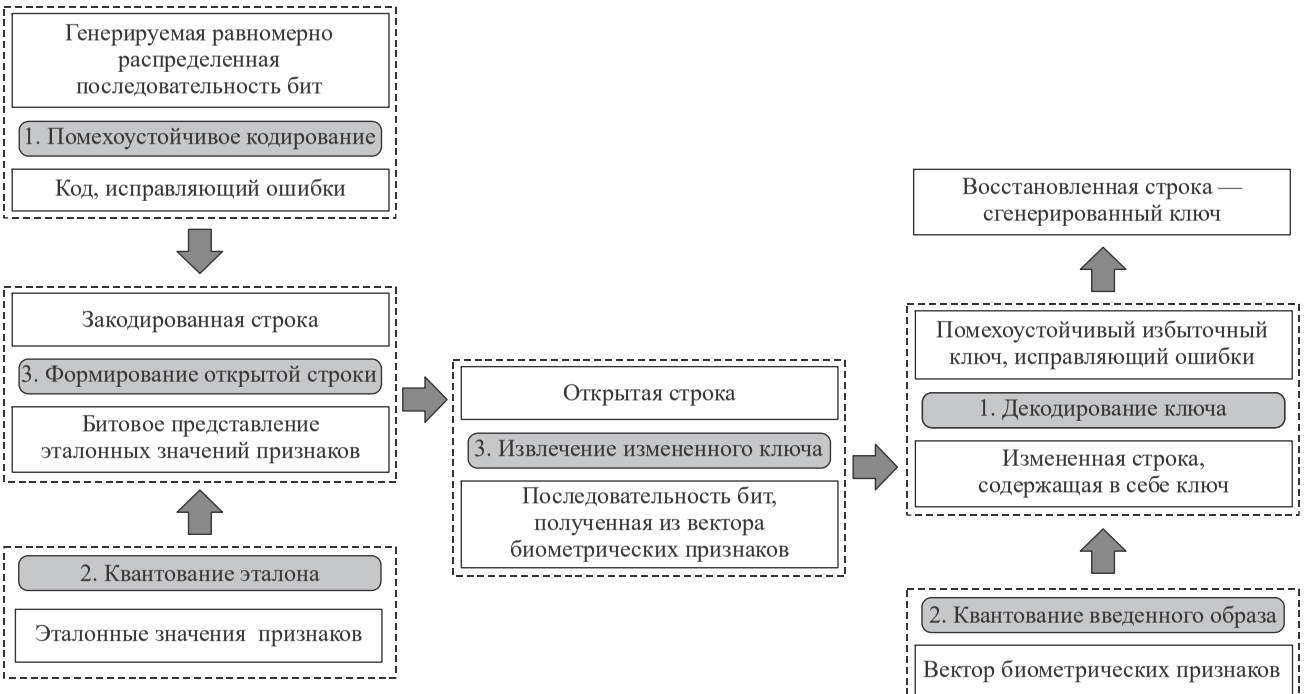

Рис. 3.1. Типовая схема нечеткого экстрактора.

ются синдромы ошибок помехоустойчивых кодов). На полученную избыточную строку накладывается гамма в виде битового представления эталонных биометрических данных (в качестве эталона обычно берется вектор средних значений признаков). Для получения битового представления биометрические данные квантуются. На выходе получается открытая строка, которую можно хранить на сервере [23]. Чтобы получить сгенерированную ранее последовательность (ключ ЭП), субъект вводит новую реализацию признаков, которая обрабатывается аналогичным образом (вычисляются значения признаков, которые далее подвергаются квантованию), «вычитается» от открытой строки, к результату этой операции применяются выбранные корректирующие коды. Это необходимо, так как после «отсоединения» полученная битовая последовательность будет изменена, вследствие отличия предъявленных биометрических данных от эталонных. Если расстояние Хемминга между введенным вектором признаков и эталонным вектором (который накладывался на избыточную строку) не превышает исправляющей способности кода, то после декодирования будет восстановлен исходный ключ, в противном случае ключ будет другой. Длина ключа будет тем меньше, чем больше исправляющая способность кода.

Описанный метод представлен на рис. 3.1.

Таким образом, хранение непосредственно эталонных характеристик (эталона субъекта) не требуется, нужно хранить лишь открытую строку. Если предъявленные биометрические данные будут достаточно близки к эталонным (схожи с эталоном субъекта), то будет сгенерирован верный ключ (исходная случайная строка и восстановленная будут равны, см. рис. 3.1).

Типовая структура (см. рис. 3.1) нечеткого экстрактора позволяет производить модификации трех видов:

1) разработка и модернизация способов помехоустойчивого кодирования и декодирования;

2) разработка и модернизация способов квантования биометрических данных;

3) разработка и модернизация способов «объединения» и «расформирования» битовых последовательностей (обычно сводятся к операции сложения по модулю 2).

3.2. Классические коды, исправляющие ошибки

Корректирующие коды — коды, служащие для обнаружения или исправления ошибок, возникающих при передаче информации под влиянием помех, а также при ее хранении. Для этого при записи (передаче)

в полезные данные добавляют специальным образом структурированную избыточную информацию (контрольное число), а при чтении (приеме) ее используют для того, чтобы обнаружить или исправить ошибки. Естественно, что число ошибок, которое можно исправить, ограниченно и зависит от конкретного применяемого кода. С кодами, исправляющими ошибки, тесно связаны коды обнаружения ошибок. В отличие от первых, последние могут только установить факт наличия ошибки в переданных данных, но не исправить ее. Код, исправляющий ошибки, может быть также использован для обнаружения ошибок.

По способу работы с данными коды, исправляющие ошибки, можно разделить на следующие:

- блоковые (блочные) коды, дробящие информацию на фрагменты постоянной длины и обрабатывающие каждый из них в отдельности;
- сверточные коды, работающие с данными как с непрерывным потоком.

Существуют коды, способные исправлять одиночные и групповые ошибки, т. е. несколько ошибок за раз. Количество ошибок, которое может исправить код, называют исправляющей способностью кода.

В нечетких экстракторах находят применение коды Хемминга, Адамара, Боуза—Чоудхури—Хоквингема (БЧХ-коды), Рида—Соломона (являются частным случаем БЧХ) [96, 97]. Рассмотрим кратко данные коды и определим те, которые целесообразно апробировать в настоящем исследовании.

Коды Хемминга — простейшие линейные коды, способные исправить одну ошибку. Код Хемминга может быть представлен в таком виде, что синдром ошибок будет равен номеру позиции, в которой произошла ошибка. Это свойство позволяет сделать декодирование очень простым. Однако вследствие низкой исправляющей способности битовое представление кодируемой последовательности придется разбивать на составляющие, каждую из которых потребуется кодировать отдельно. Вследствие этого в сумме получается большая избыточность. Например, разбиение по 8 бит является недостаточным, несовпадение битовых представлений признаков субъекта от реализации к реализации более чем в один бит на один признак является почти стопроцентной ситуацией при используемом пространстве признаков. Поделим кодируемую строку на порции из 4 бит, каждую из которых требуется кодировать отдельно. Избыточность такого кода велика, она составляет 75 %, т. е. на каждые 4 бита информации требуется еще 3 корректирующих бита. При этом даже если в одном из признаков не совпадут более 2 бит — генерируемый ключ ЭП будет отличаться. Такой подход, вероятно, может использоваться при работе со статическими биометрическими признаками, напри-

мер отпечатками пальцев, для исправления единичных ошибок аппаратуры при сканировании. Но применение данного подхода для рассматриваемых задач априорно можно считать неэффективным. Поэтому данный код далее рассматриваться не будет.

В кодах Адамара расстояние между любыми двумя кодовыми словами одинаково и поэтому совпадает с кодовым расстоянием. Подобные коды называют эквидистантными. Коды Адамара, обладая большим кодовым расстоянием, позволяют соответственно исправить и большое количество ошибок. Это достигается ценой высокой избыточности.

Для построения и реализации кода Адамара той или иной длины необходимо построить сначала матрицу Адамара соответствующего порядка.

Для любого целого $n > 0$ квадратная матрица $H = (h_{ij})$ порядка n называется матрицей Адамара, если $h_{ij} \in \{+1, -1\} \forall i, j$ и $HH^t = nI$, где I — единичная матрица.

Пример:

$$H_2 = \begin{pmatrix} +1 & +1 \\ +1 & -1 \end{pmatrix}.$$

Для матриц Адамара H_n при $n = 2m$ очевидна следующая конструкция:

$$H_{2^{m+1}} = \begin{pmatrix} +H_{2^m} & +H_{2^m} \\ +H_{2^m} & -H_{2^m} \end{pmatrix}.$$

Образуем код из всех строк матрицы H и их отрицаний. Такой код (или его $\{0, 1\}$-соответствие) называется кодом Адамара. Проведем теперь следующие операции. Выберем в коде Адамара длины n кодовые слова, начинающиеся с $+1$, отбросим первые координаты и переведем в $\{0, 1\}$-код. Результатом будет двоичный код длины $n - 1$, размера n, являющийся эквидистантным с расстоянием $n/2$. Построенный таким образом код называется укороченным кодом Адамара.

Нечеткий экстрактор на основе кода Адамара кодирует битовое представление случайной строки целиком, принимая в качестве параметра размер блока. От размера блока зависит исправляющая способность кода, оптимальный размер блока может быть найден в процессе эксперимента.

Коды БЧХ (Боуза—Чоудхури—Хоквингема) — это широкий класс циклических кодов, применяемых для защиты информации от ошибок при ее передаче по каналам связи. Код БЧХ отличается возможностью построения кода с заранее определенными корректирующими свой-

ствами, а именно, минимальным кодовым расстоянием. Этот код включен в формат POCSAG систем поискового радиовызова.

Большинство циклических кодов используют один алгоритм построения помехоустойчивых кодовых комбинаций, а различаются лишь методикой выбора образующего многочлена. В БЧХ-коде построение образующего многочлена в основном зависит от двух параметров: от длины кодового слова n и от числа исправляемых ошибок s. Особенностью кода является то, что для исправления числа ошибок $s \geq 2$ еще недостаточно условия, что между комбинациями кода минимальное кодовое расстояние $d_{\min} = 2s + 1$. Необходимо также, чтобы длина кода n удовлетворяла условию $n = 2h - 1$, где h — любое целое число. При этом n всегда будет нечетным числом и принимать значения: 1, 3, 7, 15, 31, 63, 127 и т. д, т. е. не все m могут быть заданы исследователем как параметр. Вариант построения кода БЧХ по шагам описан в работе [98].

Для декодирования могут применяться те же алгоритмы, что и для других помехоустойчивых кодов, но есть и более оптимальные алгоритмы, разработанные специально для БЧХ-кодов. Среди возможных алгоритмов декодирования алгоритм Берлекемпа—Мэсси, евклидов алгоритм, алгоритм Питерсона—Горенстейна—Цирлера (ПГЦ).

Нечеткий экстрактор на основе кодов БЧХ кодирует битовое представление случайной строки целиком и принимает в качестве параметра исправляющую способность, оптимальное значение которой для каждого набора признаков может быть вычислено в процессе экспериментов.

Широко используемым подмножеством кодов БЧХ являются коды Рида—Соломона. Это такие коды БЧХ, у которых мультипликативный порядок алфавита символов кодового слова делится на длину кода. Таким образом, $m = 1$ и поле символов $GF(q)$ совпадает с полем локаторов ошибок $GF(q^m)$. Поскольку поле символов и поле локаторов ошибок совпадают, все минимальные многочлены линейны. В коде Рида—Соломона, исправляющем t ошибок, обычно порождающий многочлен записывается в виде

$$g(x) = (x - \alpha^{l_0})(x - \alpha^{l_0+1})\ldots(x - \alpha^{l_0+d-2}),$$

где l_0 — некоторое целое число (в том числе 0 и 1), с помощью которого иногда удается упростить кодирование. Обычно полагается $l_0 = 1$. Степень многочлена $g(x)$ равна $d - 1$. Длина полученного кода n, минимальное расстояние d. Код содержит $r = d - 1$ проверочных символов, число информационных символов $k = n - r = n - d + 1$. Таким образом, $d = n - k + 1$ и код Рида—Соломона является разделимым кодом с максимальным расстоянием. Код Рида—Соломона над $GF(q^m)$, исправляющий t ошибок, требует $2t$ проверочных символов, и с его помощью исправля-

ются произвольные пакеты ошибок длиной t и меньше. Согласно теореме о границе Рейгера, коды Рида—Соломона являются оптимальными с точки зрения соотношения длины пакета и возможности исправления ошибок. Код Рида—Соломона является одним из наиболее мощных кодов, исправляющих многократные пакеты ошибок. Применяется в каналах, где пакеты ошибок могут образовываться столь часто, что их уже нельзя исправлять с помощью кодов, исправляющих одиночные ошибки.

3.3. Недостатки нечетких экстракторов

К принципиальным недостаткам нечетких экстракторов относятся:

1. Все классические коды вносят избыточность. Чем больше исправляющая способность кода, тем больше избыточности и меньше длина генерируемого ключа-пароля. В нечетком экстракторе длина ключа жестко зависит от исправляющей способности кода. В этом несложно убедиться. При исправлении 14 % ошибок посредством наиболее эффективных на сегодня корректирующих кодов избыточная часть открытой строки (дополнительной информации, хранимой в открытом виде, необходимой для генерации ключа) составит 1600 % [60], длина ключа при этом будет в 17 раз меньше длины открытой строки.

2. Классические коды не могут исправить большое количество ошибок, поэтому их невозможно использовать вместе с малоинформативными биометрическими характеристиками. К примеру, не существует кодов, способных исправлять 50 % ошибок, поскольку такие коды имеют огромную избыточность и пренебрежимо малую информационную часть [60, 99].

3. В [100] описаны уязвимости нечетких экстракторов, позволяющие ускорить перебор значений биометрических параметров с целью фальсификации генерируемого ключа (пароля). Считается, что наложение на биометрические данные гаммы в виде строки битов является надежной защитой для обеих составляющих только в случае равновероятной единичной ошибки в битовом представлении вектора биометрических признаков, чего на практике не наблюдается. В работе [101] показана связь эффективности коррекции ошибок с методами группирования битов с разной вероятностью единичной ошибки. Несмотря на предпринятые в [101] усилия, единого подхода для решения этого вопроса не выработано.

4. Нечеткие экстракторы квантуют «сырые» биометрические данные и не учитывают параметры распределения значений признаков, в результате они должны давать более высокую долю ошибок по сравнению с нейросетевыми преобразователями биометрия—код, которые в свою очередь располагают этими данными, кодируя их весовыми коэффици-

ентами нейронов [60, 99, 100]. Нечеткие экстракторы способны работать только с очень информативными признаками (с площадями пересечения функций плотностей вероятности признаков порядка 0—0,3).

Последняя проблема фактически не решается. Однако для решения остальных проблем А. В. Безяевым, А. И. Ивановым и Ю. В. Фунтиковой специально для биометрии предложен альтернативный метод помехоустойчивого кодирования [102]. Метод позволяет создавать самоорганизующиеся коды, которые не поглощают биометрические данные своей избыточностью, в отличие от классических самокорректирующих кодов. Они безопасно хранят синдромы ошибок в виде усеченных хэш-функций отдельно в виде дополнительной информации. Предлагаемые коды можно использовать как при реализации схемы нечеткого экстрактора, так и по отношению к кодам, генерируемым при помощи нейронной сети. Второй подход является предпочтительным, так как данные предварительно обогащаются в континуальной форме нейронами первого слоя.

Идея создания самоорганизующихся кодов состоит в отказе от траты части разрядов битового представления (квантованных) биометрических данных (далее *биокода*) на заполнение избыточной части. В качестве информации эквивалентной избыточности используются три последних разряда хэш-функции биокода или его фрагмента. В работе [102] для этого используется функция MD5. Три последних разряда хэш-функции могут использоваться для корректировки от одной до четырех ошибок биокода. При большей исправляющей способности возрастает время на перебор значений хэш-функции, и соответственно на генерацию ключевой последовательности [102]. В зависимости от количества фрагментов, на которые разбивается биокод и которые хэшируются по отдельности, изменяется исправляющая способность кода. При этом, оценив статистику выходных значений первого слоя нейронов, можно разбить обогащенный биокод (код на выходе первого слоя нейронов) таким образом, чтобы учесть неравномерное распределение единичной ошибки.

В [102] рекомендуется применять следующую схему кодирования. На первом шаге обработки данных вычисляется хэш-функция от первого фрагмента обогащенного биокода субъекта. На втором шаге вычисляется хэш-функция от первых двух фрагментов обогащенного биокода. На каждом следующем шаге длина хэшируемого фрагмента биокода увеличивается. На последнем шаге хэшируется весь биокод. При такой схеме формирования хэш-остатков исправляющая способность хэш-кода остается прежней, а число перебираемых злоумышленником состояний оказывается экспоненциально связано с числом фрагментов биокода [102], поэтому независимо от числа фрагментов данная схема формирования хэш-остатков предпочтительнее.

3.4. Модификации классической модели нечеткого экстрактора

Очевидно, что «сырые» биометрические данные по большей части состоят из неинформативных частей. На эффективность нечеткого экстрактора влияет способ предварительного квантования «сырых» биометрических данных. В рамках эксперимента решено дискретизировать значения признака в соответствии с преобразованием: $y = f(x)$, где x — значение признака, а y принадлежит множеству {0, 1, 3, 7, 15, 31, 63, 127, 255, 254, 252, 248, 240, 224, 192, 128}, аналогично тому, как это выполнялось в работе [103]. Значения y представляются в двоичном 8-битном виде. Суть операции преобразования иллюстрирует рис. 3.2. Данное преобразование существенно снижает количество единичных ошибок на этапе квантования, что сказывается на вероятностях ошибок 1-го и 2-го рода (они существенно снижаются) [103]. Также возрастает формальная длина генерируемого ключа (пароля). Однако энтропия квантованных данных сильно падает, что, конечно, отрицательно сказывается на защитных свойствах экстрактора (секретный ключ или пароль и биометрию субъекта более нельзя считать надежно защищенными, если хранить их в открытой строке). Поэтому будем использовать данный способ квантования (см. рис. 3.2) только в целях тестирования эффективности «нечетких экстракторов». Также этот способ требует знания границ области значений признаков, т. е. экстрактор нужно обучить на образцах данных «Чужой». Более «честный» способ квантования использовался в работах [52, 62]. Но вероятности ошибок также оказались значительны даже при

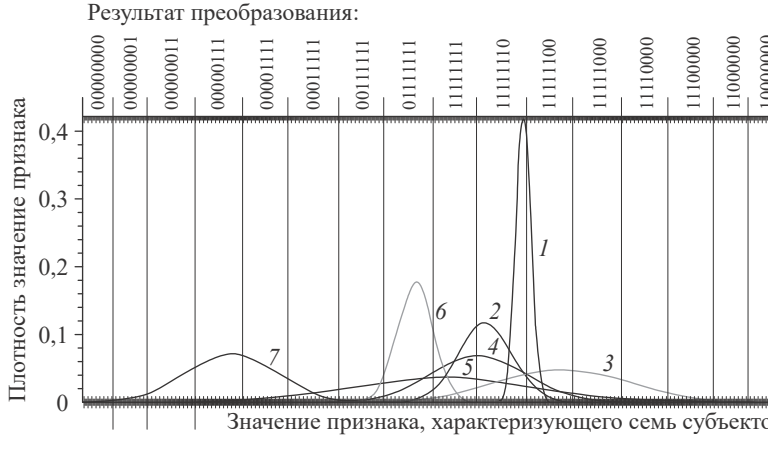

Рис. 3.2. Квантование «сырых» биометрических данных.

низкой достоверности полученных результатов (количество испытуемых в работах [52, 62] составило 12—14).

Попробуем частично компенсировать отсутствие обогащения данных в классической модели за счет применения методики оценки стабильности битового представления признаков по формуле (3.1), предложенной в [22, 100]. Для каждого субъекта выбирается определенное количество признаков, для которых произведения вычисляемых по формуле

$$\omega_i = 2|0{,}5 - P_{0,i}| = 2|0{,}5 - P_{1,i}| \qquad (3.1)$$

величин будет наивысшим, где $P_{0,i}$ — вероятность (относительная частота) появления нуля в i-м разряде кода, $P_{1,i}$ — вероятность (относительная частота) появления единицы в i-м разряде кода.

Процедура оценки информативности производится только на этапе формирования открытой строки (непосредственно перед объединением битовых последовательностей). Далее эта информация сохраняется и используется на этапе генерации секретного ключа (т. е. при получении и «отсоединении» битового представления вектора значений признаков от открытой строки). Битовые представления наиболее стабильных признаков склеиваются для получения итоговой битовой последовательности, которая в свою очередь «объединяется»/«отсоединяется» от закодированного помехоустойчивым кодом секретного ключа.

Непосредственно после оценки стабильности признаков, перед конкатенацией их битовых представлений производится их перемешивание для того, чтобы сделать распределение единичных ошибок более равномерным. Позиции битовых представлений признаков внутри результирующей последовательности бит запоминаются.

На этапе генерации секретного ключа информация о позициях признаков субъекта используется для осуществления выборки наиболее информативных из них. Признаки преобразуются в частные битовые последовательности, которые объединяются в результирующую последовательность бит с учетом позиций признаков. Получаемая последовательность вычитается из открытой строки, к результату этой операции применяется код, исправляющий ошибки.

При использовании описанной модификации нужно хранить дополнительную информацию о номерах стабильных признаков и их последовательности. Данную информацию целесообразно держать в секрете, т. е. требуется отдельный сервер или носитель. В работе [103] используется схожая методика оценки стабильности признаков, но вместо (3.1) применяется другой оператор. Полученные в [103] результаты можно назвать весьма высокими в плане вероятностей ошибочных решений. Однако необходимость хранения дополнительной информации и используемый

способ квантования создают уязвимости. Но даже при учете всех модификаций нечеткий экстрактор работает хуже нейронных сетей по надежности генерации ключевых последовательностей, как можно убедиться далее.

3.5. Нейросетевые преобразователи биометрия—код и их особенности

Сдерживающим фактором в применении нейронных сетей является сложный процесс их обучения. Небольшие нейронные сети быстро обучаются (на малом количестве примеров), но принимают низкокачественные решения. По мере увеличения размеров (количества слоев, нейронов и их входов) решения становятся более достоверными (на уровне людей-экспертов или лучше), но при этом растет сложность обучения нейросети, появляются проблемы «тупиков» и «зацикливания обучения», в результате этот процесс становится неприемлемо долгим либо неосуществимым [28, 104]. Для биометрии требуются сверхбыстрые алгоритмы обучения (выполняемые за несколько секунд на обычном персональном компьютере) [28]. Для данной цели не могут быть использованы итерационные алгоритмы, поскольку они теряют устойчивость при увеличении числа входов нейронов или при снижении качества биометрических данных [60]. Решение проблемы предложено в стандарте ГОСТ Р 52633.5-2011 [22]. Рекомендуется использовать прямое вычисление модулей весовых коэффициентов через математические ожидания и среднеквадратические отклонения биометрических параметров «свой» и «чужой». Благодаря этому процедуры обучения становятся рекордно быстрыми и устойчивыми [60]. Для обучения сети требуется не менее 21 образца данных «Свой» и 64 независимых образца данных «Чужой» (образцы от разных субъектов).

Высоконадежные биометрические устройства с вероятностью ошибочных решений 10^{-12} и выше проще создать, чем доказательно проверить эту вероятность прямым численным экспериментом [60]. Для ПБК со сверхнизкими показателями ошибок не годится схема Бернулли [60], так как для атак прямого подбора необходимы объемные базы данных биометрических признаков, собрать которые невозможно по причине нехватки населения Земли. Для тестирования НПБК предложены процедуры морфинга, определенные в ГОСТ Р 52633.2-2010, используя которые удается оценить «нано» и «пико» вероятности ошибок второго рода биометрической аутентификации на тестовых базах, состоящих из 10000 естественных биометрических образов [60].

Преимуществом НПБК по сравнению с «нечетким экстрактором» является процедура обогащения. Обогащение позволяет работать с «пло-

хими биометрическими данными» и восстанавливать до 50 % ошибок исходных данных [102]. При высоком уровне первичного обогащения данных обыкновенные нейроны (персептроны) оказываются малоэффективны. Более эффективными оказываются искусственные нейроны с несколькими выходными дискретными состояниями [60]. Практика показала, что использование операции циклического сдвига при настройке формы нелинейного элемента нейрона не является оптимальной. Более качественные результаты получаются, если определенным участкам области значений сумматора нейрона (вне интервала «Свой») задавать выходные коды случайным образом. При таком способе усиливаются хэширующие свойства обученного нейрона по отношению к образам все «Чужие» (энтропия бит генерируемого кода на основе примера «Чужой» близка к максимально возможной для его длины), нелинейные элементы с длительными монотонными участками хуже перемешивают данные [60]. В [104] предложено использовать трид-нейроны с двумя выходными состояниями, которые имеют два порога квантования. Использование трид-нейронов позволяет повысить длину генерируемого кода в 2 раза, а энтропию кодов — в полтора, если квантователь выходных значений не является монотонной дискретной функцией. Позже в работе [24] предложена более общая концепция z-арных нейронов, т. е. нейронов с z бинарными выходами.

По требованиям ГОСТ Р 52633.0-2006 [21] при поступлении на вход НПБК образца данных «Чужой» вероятности значений «0» и «1» разрядов выходного кода должны быть равными (допускается разница в количестве различных значений разрядов не более 10 %). Для того чтобы поднять качество хэширования, могут быть использованы различные механизмы размножения ошибок, например, сложение по модулю 2 части выходного кода «Свой» от изолированного потока нейронов и записанных в дискретной форме параметров обученной нейронной сети остальных нейронов (весовых коэффициентов нейронов и номеров связей между нейронами) [60]. Шифрование параметров нейронов на выходах других нейронов также предлагается использовать для защиты таблиц нейросетевых функционалов от анализа с целью восстановления эталона субъекта [60, 99]. Данный принцип защиты называется защищенным нейросетевым биометрическим контейнером [99]. Размер ключа для шифрования параметров нейронов выбирается по конструктивным особенностям и возможностям нейронной сети. Данная схема уязвима к атаке Г. Б. Маршалко, которая строится на наблюдении большого числа выходов у незащищенных нейронов [99]. Чтобы снизить эффективность таких атак на биометрию, следует отказаться от создания одного длинного ключа шифрования параметров нейронов и использовать множество ключей

увеличивающейся длинны [60]. Тем не менее, на сегодня защищенные нейросетевые контейнеры являются наиболее эффективным средством защиты оцифрованной биометрии на этапе хранения [99].

Нужно отметить, что длина эффективного кода (аутентификатора) зависит от количества информативных признаков, простое увеличение количества нейронов не ведет к аналогичному росту эффективного кода, так как энтропия генерируемого кода не соответствует его длине [99]. Вместе с тем, чтобы снизить вероятность ошибок второго рода до уровня парольной защиты ($\sim 10^{-8}$), необходимо использовать достаточно большое число выходов у НПБК.

Для обучения НПБК необходимо, чтобы биометрические данные имели закон распределения, близкий к нормальному, для контроля за этим используется критерий Пирсона, а также его модификации [105, 106].

Достоинством биометрических сетей является то, что биометрический шаблон человека не хранится более в памяти компьютера, вместо него хранятся весовые коэффициенты между нейронами (не существует эффективного способа восстановления параметров распределения биометрических признаков из нейросетевого биометрического контейнера) [60, 99]. Защитные свойства усиливаются при использовании защищенных нейросетевых контейнеров.

3.6. Модель нейросетевого преобразователя биометрия—код в соответствии с ГОСТ Р 52633.5-2011

В ГОСТ Р 52633.5-2011 [22] рекомендуется использовать однослойные или двухслойные нейронные сети (сети с большим количеством слоев являются избыточными и для их применения необходимо специальное обоснование [60, 99]). Первый слой осуществляет обогащение данных, второй играет роль кодов, исправляющих ошибки [22, 60, 99]. Алгоритм из ГОСТ Р 52633.5-2011 служит для послойного обучения сети нейронов: сначала осуществляется обучение первого слоя, далее эти же обучающие данные подаются на вход второго слоя сети и вычисляются весовые коэффициенты нейронов второго слоя. Модули весов нейронов первого и второго слоя вычисляются детерминированно по формулам [22]

$$\mu_i = |E_{\text{ч}}(x_i) - E_{\text{с}}(x_i)|/\sigma_{\text{ч}}(x_i)\sigma_{\text{с}}(x_i), \qquad (3.2)$$

где $E_{\text{с}}(x_i)$ — математическое ожидание (среднее значение) значений признака для образа «Свой», $\sigma_{\text{с}}(x_i)$ — среднеквадратичное отклонение значений признака для образа «Свой», $E_{\text{ч}}(x_i)$ и $\sigma_{\text{ч}}(x_i)$ — аналогичные показатели для образа «Чужой». Знак весового коэффициента, при условии, что нейрон должен выдавать единицу («1»), выбирается исходя из правила:

«+», если $E_ч(x_i) < E_с(x_i)$, иначе «–». Если нейрон должен выдавать нуль («0»), знаки весовых коэффициентов инвертируются;

$$\mu = a_2\omega_i/E(\omega_i), \tag{3.3}$$

где a_2 — стабилизирующий коэффициент для нейронов второго слоя, экспериментально подбираемый для каждой задачи выработки ключевой последовательности, ω_i — показатель стабильности i-го разряда выходного кода нейронов первого слоя, вычисляемый по формуле (3.1) [22, 100], $E(\omega_i)$ — математическое ожидание (среднее значение) показателей стабильности разрядов выходного кода нейронов первого слоя.

Алгоритм обучения позволяет настроить сеть на выдачу заданного ключа (пароля, кода доступа) и случайной битовой последовательности при поступлении образа неизвестного пользователя. Для обучения требуется не менее 21 реализации биометрических данных «Свой» и не менее 64 реализаций биометрических данных «Чужой».

При использовании второго слоя необходимо перейти от промежуточных кодов «0» и «1» к эквивалентным «–1» и «1». Число входов нейронов второго слоя рекомендуется выбирать от 0,2 до 0,8 от числа нейронов первого слоя (или от числа нейронов первого слоя, так как каждый нейрон по ГОСТ Р 52633.5-2011 имеет один выход). Рекомендации по выбору количества нейронов первого и второго слоев аналогичные и описаны в стандарте [22]. Связи нейронов первого слоя с нейронами второго слоя задаются случайно. Обработчики признаков связывают с нейронами первого слоя сначала последовательно, а при превышении номера нейрона над числом признаков случайно. Далее осуществляется корректировка знаков весовых коэффициентов, которая носит эмпирический характер, с целью добиться желаемой вероятности ошибок аутентификации [22]. Операции по настройке слоев нейросети подробно описаны в работах [60, 99].

Выход сумматора нейрона любого слоя на этапе принятия решений определяется по формуле

$$y = \sum_{i=1}^{m} \mu_i v_i + \mu_0, \tag{3.4}$$

где v_i — i-й вход нейрона, m — число входов, μ_i — весовой коэффициент i-го входа, μ_0 — нулевой вес, отвечающий за переключатель квантования нейрона.

Следует отметить, что обучение сети нейронов стандартным алгоритмом [22] обладает рекордной устойчивостью и имеет линейную вычислительную сложность. Стандартизованный алгоритм [22] абсолютно устойчив, так как не является итерационным (в нем исключен направлен-

ный итерационный поиск весовых коэффициентов нейронов). Фактически стандартный алгоритм обучения является полностью детерминированным, так как он однозначно вычисляет знаки и значения весовых коэффициентов нейронов, опираясь на знание вектора математических ожиданий и стандартных отклонений биометрических параметров образа «Свой», а также вектора стандартных отклонений образов «Чужие».

3.7. Квадратичные формы и иные функционалы

Если при решении задачи биометрической аутентификации идти по пути классической линейной алгебры, то решающее правило придется строить на основе квадратичных форм (см. ниже формулу (3.5)), что подразумевает обращение корреляционных (ковариационных) матриц очень высокой размерности (100—1000 и более).

$$y(\overline{\nu}) = (E(\overline{\nu}) - \overline{\nu})^T \left[R\right]^{-1} (E(\overline{\nu}) - \overline{\nu}), \qquad (3.5)$$

где $\overline{\nu}$ — вектор нормированных биометрических параметров с единичными стандартными отклонениями.

Для биометрических параметров корреляционные матрицы 2-го и 3-го порядков могут быть обращены. Корреляционные матрицы более высоких порядков не обращаются, так как данная задача относится к плохо обусловленным.

Проблема плохой обусловленности большинства вычислительных алгоритмов является одной из важнейших в области современной науки и техники. Иногда эту проблему образно называют «проклятием размерности». Наиболее глубоко эта проблема изучена применительно к задаче решения систем линейных уравнений или обращения матриц [107, 108]. Введен и широко используется специальный параметр контроля «качества матрицы» (число обусловленности — cond[A]). Параметр «качества матрицы» или cond[A] может изменяться в пределах от 1 до ∞ и по своей сути является коэффициентом усиления погрешности исходных данных. При cond[A] = 1 погрешность исходных данных не усиливается. При cond[A] > 1 погрешность конечного результата увеличивается пропорционально числу ее обусловленности.

Обычно перед решением системы линейных уравнений $\left[A\right] \cdot \overline{x} = \overline{y}$ оценивают число обусловленности матрицы [A] и вектор ошибок входных данных — $\Delta\overline{x}$. Это позволяет заранее оценить значение вектора ошибок конечного результата $\Delta\overline{y} = \text{cond}[A] \cdot \Delta\overline{x}$.

Если результирующая погрешность $\Delta\overline{y}$ оказывается велика, то существует два пути возможного ее понижения. Во-первых, можно снизить погрешность исходных данных за счет их накопления (увеличения объ-

ема выборки в k раз). Применив метод наименьших квадратов [109], мы можем снизить результирующую ошибку в $\sqrt{k}$ раз. Во-вторых, мы можем регуляризовать задачу по Тихонову [110], добившись необходимых значений вектора ошибок вычислений.

Следует отметить, что приведенные выше два метода повышения стабильности результатов матричных вычислений актуальны для любых вычислений. Любые вычисления можно улучшить, вводя избыточность (увеличивая выборку) или применив какой-либо метод регуляризации вычислений.

В связи с этим часто используются сети из квадратичных форм, не учитывающих корреляционные связи.

Метрик, ориентированных на отсутствие в обрабатываемых данных корреляционных связей, множество. Самой распространенной является метрика хи-квадрат Пирсона

$$\chi = \sum_{i=1}^{m} \frac{(E(\nu_i) - \nu_i)^2}{\sigma(\nu_i)^2}, \tag{3.6}$$

где ν_i — значение i-го признака (входа нейрона), $E(\nu_i)$ — математическое ожидание (среднее значение) i-го признака (входа нейрона), $\sigma(\nu_i)$ — среднеквадратичное отклонение i-го признака (входа нейрона).

Метрику (3.6) следует называть метрикой Пирсона, так как при независимых данных ее плотность распределения значений точно совпадает с распределением хи-квадрат [111]:

$$p(x^2) = \frac{1}{2^{\frac{m}{2}}\Gamma\left(\frac{m}{2}\right)} x^{\left(\frac{m}{2}-1\right)} \exp\left(\frac{-x^2}{2}\right),$$

где $\Gamma(\ )$ — гамма-функция, m — число степеней свободы или число учитываемых биометрических параметров.

Следует отметить, что в случае применения метрики Пирсона (3.6) для зависимых биометрических данных ее мощность падает, число степеней свободы m становится дробным (фрактальным) числом.

Другой рассматриваемой квадратичной формой является метрика Евклида, вычисляемая по формуле

$$\varepsilon = \sqrt{\sum_{i=1}^{m} (E(\nu_i) - \nu_i)^2}, \tag{3.7}$$

где ν_i — i-й вход нейрона, $E(\nu_i)$ — математическое ожидание i-го входа нейрона. Данная метрика является более слабой, так как не учитывает среднеквадратичное отклонение биометрического параметра.

3.8. Формирование сетей квадратичных форм

Следует отметить, что основной проблемой искусственного интеллекта является проблема его обучения или его настройки. Очевидно, что независимо от используемой технологии искусственный интеллект работать будет тем лучше, чем больше параметров будет учтено при (обучении или настройке) решающего правила. Размерность решающего правила принципиально важна. Чем выше размерность решаемой задачи, тем эффективнее работает искусственный интеллект.

Основное отличие сети квадратичных форм (или иных функционалов) заключается в строении искусственного нейрона. Классический нейрон и нейрон стандарта ГОСТ Р 52633.5-2011 состоят из сумматора и пороговой функции на выходе нейрона, которая трансформирует полученную сумму обработанных параметров от каждого синапса (входа нейрона) в бинарное значение «0» или «1». Нейрон квадратичной формы может быть основан на метрике Евклида, Пирсона, Махаланобиса и др. [72]. К преимуществам квадратичных форм можно отнести отсутствие необходимости обучения на образцах «Чужой» и возможность нелинейного разделения собственных областей эталонов в пространстве признаков [72] (персептрон осуществляет линейное разделение).

При формировании сети функционалов (3.6) и (3.7) обработчики признаков нужно соединять с нейронами случайным образом, избегая повторов.

Сеть квадратичных форм можно реализовать с одним слоем нейронов или двумя слоями нейронов. Первый слой состоит из нейронов, рассчитывающих выход по одной из указанных выше формул, от этого зависит тип сети: Пирсона—Хемминга (3.6), Евклида—Хемминга (3.7) и т. д. Независимо от типа нейрона, значение на выходе его функционала сравнивается с пороговым. Для каждого нейрона существует оптимальный порог срабатывания, который вычисляется (кроме персептронов) исходя из произведения $\theta = y_{\max} h$, где $y_{\max}$ — это максимальное значение функционала при поступлении на вход обучающих реализаций образа «Свой», h — стабилизирующий коэффициент, экспериментально подбираемый для каждого пространства признаков по минимальной сумме вероятностей ошибок 1-го и 2-го рода. Если порог превышен, нейрон выдает единицу («1»), иначе нуль («0»). Настройка сети на нужный выходной код производится инвертированием выходных значений отдельных нейронов. Так как нейроны выдают бинарные значения, их сети называют по имени метрики и Хемминга.

Второй слой нейронной сети можно полностью скопировать из стандарта ГОСТ Р 52633.5-2011. Второй слой играет роль кодов, исправляю-

щих ошибки, его можно применить к любой нейросети, будь то сеть квадратичных форм, персептрон или его модификация. В качестве альтернативы второго слоя можно применить схему восстановления ошибочных бит генерируемого ключа, предложенную в [102] (коды Безяева). Последний вариант является предпочтительным, так как позволяет исправить заданное число ошибок.

Особенностью функционалов (3.6) и (3.7) является необходимость хранения параметров распределения признаков. На практике этот недостаток устраняется путем создания сети из различных видов нейронов с изолированным потоком персептронов (3.4), на выходах которых будут шифроваться параметры нейронов иных функционалов по принципу за-

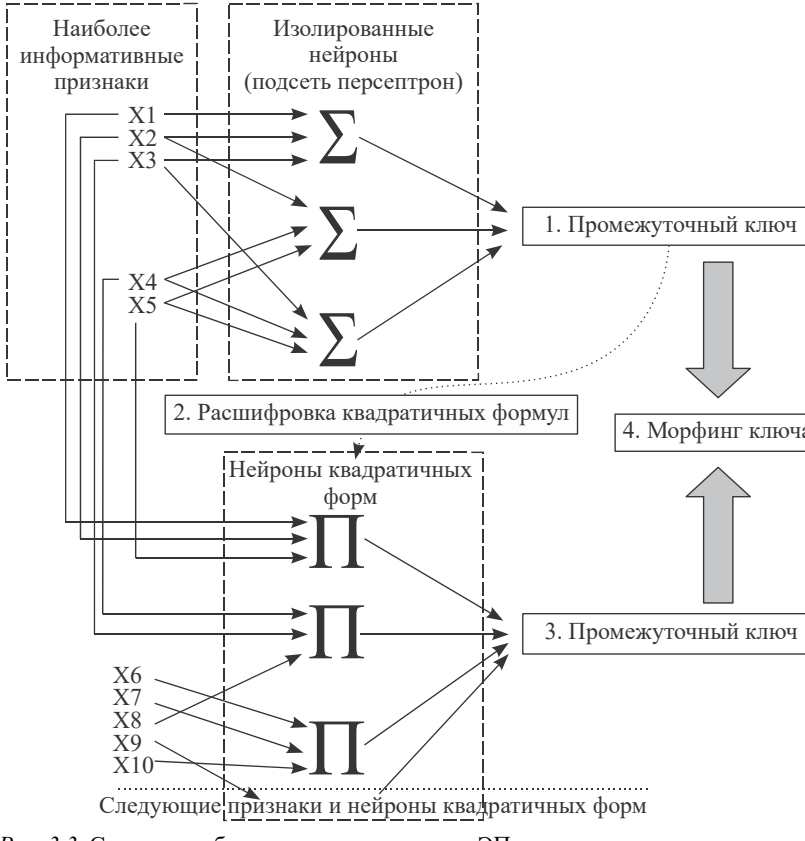

Рис. 3.3. Схема выработки секретного ключа ЭП или пароля с использованием сети на базе различных видов нейронов и принципа защищенного нейросетевого контейнера.

щищенного нейросетевого контейнера [60, 99]. Этот принцип иллюстрируется на рис. 3.3. Необходимо подготовить такие изолированные нейроны, на входы которых будут подаваться значения наиболее информативных признаков. При верной выдаче фрагмента генерируемого ключа изолированными нейронами параметры нейронов квадратичных форм будут расшифрованы правильно. В результате будет формироваться оставшаяся часть ключа. В противном случае сеть должна генерировать случайный шум, так как расшифрованные значения весовых коэффициентов будут некорректны (либо не будут соответствовать эталону субъекта). Можно определить несколько потоков изолированных нейронов, чтобы шифровать данные многократно, каждый раз осуществляя морфинг нового промежуточного ключа на основе предыдущего и генерируемого очередным потоком, что усилит защиту весовых коэффициентов [60, 99]. Однако в настоящей работе решено ограничиться лишь предложением по реализации такой сети. Оценку эффективности сетей нужно производить без использования принципа защищенного нейросетевого контейнера [60, 99].

3.9. Влияние качества эталонов на результаты генерации секретных ключей ЭП на основе подписей

Подпись субъекта, как и клавиатурный почерк, являются динамическими биометрическими образами, меняющимися в течение жизни. Изменения могут быть закономерные, связанные с изменением характера ввода подписи с течением времени. Для учета данных изменений применяются методики обновления эталона при верной аутентификации, а также рекомендуется периодически создавать новый эталон (период задается в зависимости от характера этих изменений). Однако существуют и спонтанные изменения, связанные с неточностью оборудования и неточностями ввода подписи субъектом. Часто при воспроизведении очередного образца подписи субъект допускает существенные изменения, причины могут быть различны: рука подписанта дрогнула, субъект отвлекся на внешние факторы и др. В результате многие признаки после обработки таких некорректных подписей имеют не характерные для субъекта значения. Поэтому при формировании эталонных значений признаков важно, чтобы учитывались только корректно введенные образцы подписи.

Простейшим способом была бы оценка корреляционных связей подписей после процедуры нормирования по длительности. Однако исходные функции $x(t)$, $y(t)$, a также их производные (функции скорости по координатам x и y) зависимы от угла наклона подписи. Спонтанные движения руки подписанта при воспроизведении подписи, как правило, отра-

жаются на функции скорости и давления пера на планшет, что должно приводить к снижению корреляционных связей между соответствующими функциями некорректной и корректных подписей субъекта. Однако исследования показали, что у некоторых испытуемых даже коэффициенты корреляции между корректно введенными реализациями этих функций имеют низкие значения (менее 0,3, табл. 3.2). Аналогичным образом дело обстоит с функцией давления на планшет (табл. 3.3).

Такие корреляционные связи являются слабыми (по шкале Чеддока [112], табл. 3.4), не пригодными для оценки степени некорректности подписи. Поэтому при поиске некорректных реализаций предложено производить оценку корреляционных связей между векторами значений признаков, полученных из исходных образцов подписи. Данные коэффициенты корреляции являются более стабильной величиной для всех испытуемых (табл. 3.5).

Предлагаемая методика поиска некорректных реализаций при формировании открытой строки применяется после преобразования введенных пользователем подписей в векторы значений признаков, но перед созданием эталонного описания образа субъекта, и заключается в следующем:

1. Строится матрица R коэффициентов корреляции $r(x_i, x_j)$ между i-ми и j-ми реализациями биометрических данных — векторами значений признаков для k введенных образцов подписи. В данном случае оценка взаимной зависимости идет не для признаков, а для образцов подписи.

Таблица 3.2

Низкая корреляционная зависимость функций скорости пера на планшете

Номер образца подписи	1	2	3	4	5	6
	Испытуемый 1					
1	1	0,1958	–0,14	0,2321	0,0445	0,0263
2	0,1958	1	0,3052	0,1644	0,2662	–0,323
3	–0,140	0,3052	1	0,3611	0,5941	0,1769
4	0,2321	0,1644	0,3611	1	–0,068	0,2878
5	0,0445	0,2662	0,5941	–0,068	1	0,1843
6	0,0263	–0,323	0,1769	0,2878	0,1843	1
	Испытуемый 2					
1	1	–0,03	–0,373	–0,167	–0,036	–0,3881
2	–0,03	1	0,2261	0,2339	0,2246	0,1095
3	–0,373	0,2261	1	0,4161	0,0241	0,1681
4	–0,167	0,2339	0,4161	1	0,7196	0,62
5	–0,036	0,2246	0,0241	0,7196	1	0,3798
6	–0,388	0,1095	0,1681	0,62	0,3798	1

Таблица 3.3

Низкая корреляционная зависимость функций давления пера на планшет

Номер образца подписи	1	2	3	4	5	6
	Испытуемый 1					
1	1	0,1757	0,2098	0,3311	0,0065	0,6785
2	0,1757	1	0,5028	0,4952	0,299	0,5286
3	0,2098	0,5028	1	0,0928	0,4808	0,3574
4	0,3311	0,4952	0,0928	1	0,1816	0,6677
5	0,0065	0,299	0,4808	0,1816	1	–0,075
6	0,6785	0,5286	0,3574	0,6677	–0,075	1
	Испытуемый 2					
1	1	0,5053	0,0101	0,4088	0,278	0,1633
2	0,5053	1	0,1638	0,7963	0,8215	0,3897
3	0,0101	0,1638	1	0,6181	0,4945	0,7722
4	0,4088	0,7963	0,6181	1	0,9384	0,7229
5	0,278	0,8215	0,4945	0,9384	1	0,7576
6	0,1633	0,3897	0,7722	0,7229	0,7576	1

2. Вычисляется коэффициент множественной корреляции $c_{i;\,1,\,2,\,\ldots,\,i-1,\,\ldots,\,i+1,\,\ldots,\,k}$ для каждого i-го вектора с остальными векторами значений признаков, полученных на предыдущем шаге. Как известно, коэффициент множественной корреляции может быть вычислен по формуле

$$c_{i;\,1,\,2,\,\ldots,\,k} = \sqrt{1 - \frac{|R_k|}{R_{ii}}}, \tag{3.8}$$

где $|R_k|$ — определитель матрицы R_k коэффициентов корреляции k векторов значений признаков, R_{ij} — алгебраическое дополнение элемента q_{ij} матрицы R_k.

Так как задача обращения высокоразмерных корреляционных матриц между образцами биометрических данных также является плохо обусловленной (число cond[R] является высоким при достаточном для по-

Таблица 3.4

Шкала Чеддока [112]

Количественная мера тесноты связи (модуль коэффициента корреляции)	Качественная характеристика силы связи
0,1—0,3	Слабая
0,3—0,5	Умеренная
0,5—0,7	Заметная
0,7—0,9	Высокая
0,9—0,99	Весьма высокая

Таблица 3.5

Корреляционная зависимость между векторами значений признаков, получаемых из подписей

Номер образца подписи	1	2	3	4	5	6
	Испытуемый 1					
1	1	0,8319	0,6850	0,8222	0,5331	0,6465
2	0,8319	1	0,7592	0,8023	0,5028	0,6491
3	0,685	0,7592	1	0,6768	0,6114	0,5291
4	0,8222	0,8023	0,6768	1	0,5029	0,7802
5	0,5331	0,5028	0,6114	0,5029	1	0,518
6	0,6465	0,6491	0,5291	0,7802	0,518	1
	Испытуемый 2					
1	1	0,4637	0,4048	0,5066	0,5451	0,4091
2	0,4637	1	0,5985	0,884	0,9278	0,6636
3	0,4048	0,5985	1	0,7344	0,6813	0,8983
4	0,5066	0,884	0,7344	1	0,9556	0,7148
5	0,5451	0,9278	0,6813	0,9556	1	0,7508
6	0,4091	0,6636	0,8983	0,7148	0,7508	1

строения эталона количестве реализаций), то размерность матриц решено выбирать небольшой. Обучающая выборка делится на фрагменты по несколько реализаций подписи (от 3 до 5, в зависимости от числа cond[R]), по каждому фрагменту строится корреляционная матрица и вычисляется коэффициент множественной корреляции. Таким образом, производится оценка корреляционной зависимости каждой реализации подписи от других введенных подписей.

3. Вычисляемые коэффициенты корреляции сравниваются с некоторым пороговым значением. Если полученный коэффициент меньше, то соответствующая данному коэффициенту корреляции подпись считается некорректной и не учитывается при формировании открытой строки. Вместо нее субъекту предлагается ввести другую подпись, для которой процедура повторяется.

Эмпирически установлено, что оптимальное пороговое значение множественного коэффициента корреляции в рассматриваемом пространстве признаков для каждого субъекта может различаться, но его значение, как правило, соответствует высокой корреляционной зависимости.

Проведена оценка эффективности предложенной методики исключения некорректных образцов подписи на примере модифицированных нечетких экстракторов на базе кодов Адамара.

Каждая подпись была преобразована в битовую последовательность (с учетом оценки стабильности бит значений признаков). Для формирования битовых последовательностей на разных этапах эксперимента использовалось различное количество признаков. Оптимальное количество признаков и размер блока для кода Адамара определялись на основании вычислительного эксперимента.

В ходе вычислительного эксперимента часть введенных испытуемыми образцов (по 21 от каждого субъекта, именно такое количество является минимальным для обучений НПБК в соответствии с ГОСТ Р 52633.5-2011, решено не отступать от данного правила) была использована для формирования открытых строк. Остальные образцы были задействованы для генерации секретных ключей ЭП. Эксперимент повторялся несколько раз, при этом обучающая и экспериментальная выборки подписей изменялись случайным образом. Основные результаты эксперимента представлены на рис. 3.4—3.8. Достоверность (доверительная вероятность) результатов составила более 0,99 при доверительном интервале 0,03.

Как видно из рисунков, оптимальным размером блока для кодов Адамара является 6 бит (по крайней мере, в задаче выработки секретного ключа ЭП из подписи). При меньшем размере блока вероятности ошибок

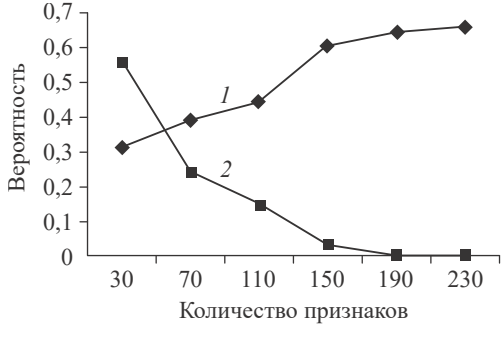

Рис. 3.4. Вероятности ошибок генерации ключа ЭП до использования методики исключения некорректных реализаций при размере блока 5 бит.

Здесь и на рис. 3.5—3.8: *1* — FRR, *2* — FAR.

Рис. 3.5. Вероятности ошибок генерации ключа ЭП до использования методики исключения некорректных реализаций при размере блока 6 бит.

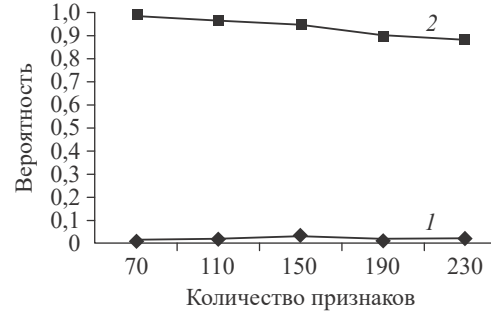

Рис. 3.6. Вероятности ошибок генерации ключа ЭП до использования методики исключения некорректных реализаций при размере блока 7 бит.

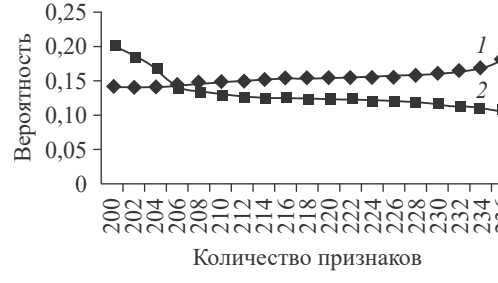

Рис. 3.7. Вероятности ошибок генерации ключа ЭП до использования методики исключения некорректных реализаций при размере блока 6 бит (детализация результата).

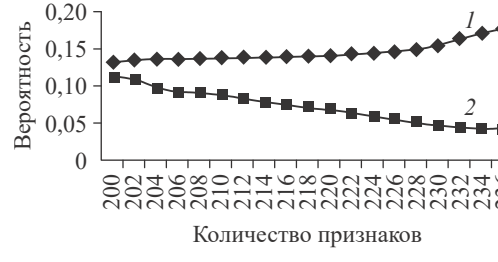

Рис. 3.8. Вероятности ошибок генерации ключа ЭП после применения методики исключения некорректных реализаций при размере блока 6 бит.

становятся выше, а длина генерируемого ключа ниже. При большем размере длина ключа увеличивается, однако, вероятности ошибок генерации слишком значительны.

Таким образом, исключение некорректных биометрических образцов является важной частью метода распознавания субъектов, если для этого используются динамические образы. По результатам эксперимента, предложенная методика в среднем позволяет снизить количество ошибок первого рода на 6,5 %, второго рода на — 46 %. Наилучшим результатом по генерации секретных ключей ЭП на основе кодов Адамара в рассмотренном пространстве признаков можно считать следующий: FRR = 0,148, FAR = 0,05 при длине ключа 304 бита (для этого использо-

валось 228 признаков). Результат был получен с исключением некорректных реализаций при формировании открытой строки и с размером блока кодируемого сообщения 6 бит. При данных параметрах сумма ошибок 1-го и 2-го рода была наименьшей. Преимуществом экстракторов на основе кода Адамара, по сравнению с БЧХ, является высокая скорость работы.

3.10. Экспериментальное сравнение нечетких экстракторов, сетей квадратичных форм и персептронов при генерации ключевых последовательностей на основе динамических биометрических образов

Требования к нейросетевым преобразователям биометрия—код изложены в семействе отечественных стандартов ГОСТ Р 52633, число которых существенно превышает число зарубежных аналогичных стандартов для нечетких экстракторов (ISO/IEC 24745:2011, ISO/IEC 24761:2009, ISO/IEC 19792:2009) [60, 99], т. е. данный подход лучше стандартизован. Сопоставительные данные о преимуществах и недостатках рассмотренных подходов можно видеть в табл. 3.6. Сравним описанные подходы экспериментально.

Хорошие (стабильные) биообразы должны давать неплохие результаты и для нечетких экстракторов, и для нейронных сетей. Очень хорошие (высокоинформативные) образы вообще не нуждаются в обогащении. Поэтому сравнение рассмотренных в настоящей главе методов генерации проведем на примере малоинформативных образов подписи и клавиатурного почерка. Для сравнения нечетких экстракторов с НПБК испытуемым было предложено ввести по 100 реализаций длинной парольной фразы (не более 50 символов).

Проведено натурное моделирование — вычислительный эксперимент с реальными биометрическими данными подписей и клавиатурного почерка субъектов, как при натурном, только информация подавалась на вход алгоритмов в автоматическом режиме. Часть биометрических данных использовалась для обучения, остальные — для экспериментальной оценки надежности выработки ключа доступа и последующей аутентификации. Количество образцов обучающей выборки решено сделать идентичным для нейронных сетей, сетей квадратичных форм и нечетких экстракторов: 21 реализация образа «Свой» и 64 реализации образа «Чужой» для персептронов (по одной реализации на каждого другого испытуемого). Вероятности ошибок 1-го и 2-го рода подсчитывались следующим образом: $FRR = er_1/ex_1$, $FAR = er_2/ex_2$, где er — количество ошибок соответствующего рода, ex — количество опытов для выявления ошибки соответствующего рода. Также подсчитывалась сумма ошибок 1-го и

Таблица 3.6

Преимущества и недостатки преобразователей биометрия—код

Подход	Преимущества	Недостатки
Персептроны ГОСТ Р 52633.5-2011 и их модификации	1. Обогащает данные 2. Хорошо стандартизован 3. Маскирует биометрический эталон 4. Возможность создания защищенного нейросетевого контейнера	1. Требуется обучать сеть на образцах данных «Чужой» (других субъектов)
Нечеткий экстрактор	1. Не требуется обучать сеть на образцах данных «Чужой» (образцах других субъектов) 2. Не требуется хранить эталон (защищает эталон гаммой) 3. Простота реализации на практике	1. Не учитывает параметры распределения признаков 2. Длина выходного ключа зависит от исправляющей способности кода 3. Помехоустойчивые коды крайне избыточны, поэтому длина генерируемого ключа оказывается низкой 4. Возможно ускорить перебор биометрических данных для фальсификации ключа (пароля) [100]
Сети квадратичных форм и иных функционалов	1. Не требуется обучать сеть на образцах данных «Чужой» 2. Обогащают данные эффективней, чем персептроны 3. Возможность создания защищенного нейросетевого контейнера	1. Возникает необходимость хранить параметры распределения значений признаков и фрагменты генерируемого ключа, как следствие, нужно формировать защищенный нейросетевой контейнер

2-го рода *ErrorRate* как площадь пересечения функций плотностей вероятности расстояний Хемминга от генерируемых кодов реализациями образов «Свой» до ожидаемого (идеального) кода и от генерируемых кодов реализациями образов «Чужой». Указанные плотности аппроксимировались нормальным законом распределения для кодов «Чужой» и бета-распределением для кодов «Свой».

Коды Рида—Соломона целесообразно применять с максимально возможной исправляющей способностью (рис. 3.9).

Все рассматриваемые в эксперименте сети имели один слой нейронов. Длина выходного кода нейросети во всех случаях была равной количеству нейронов. В соответствии с рекомендациями, данными в [91, 113], количество входов нейрона для сетей квадратичных форм в задаче вери-

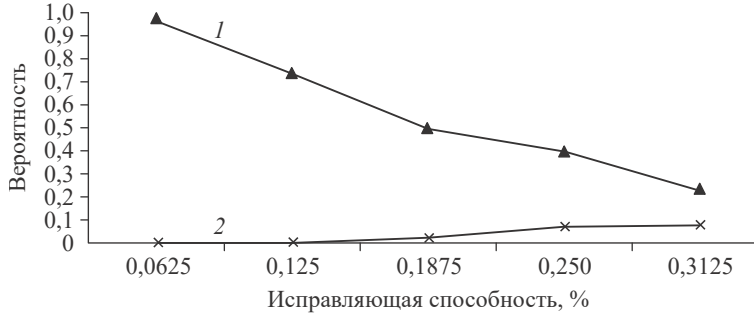

Рис. 3.9. Вероятности ошибок выработки ключа нечетким экстрактором на основе кодов Рида—Соломона при использовании 236 признаков подписи.

1 — FRR, *2* — FAR.

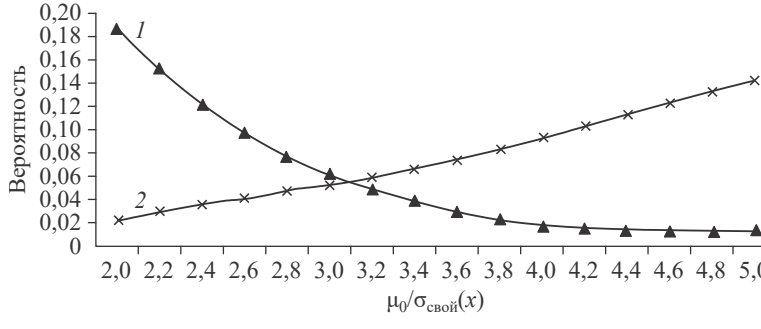

Рис. 3.10. Вероятности ошибок выработки ключа нейронной сетью по ГОСТ Р52633.5 с одним слоем при использовании 236 признаков подписи.

1 — FRR, *2* — FAR.

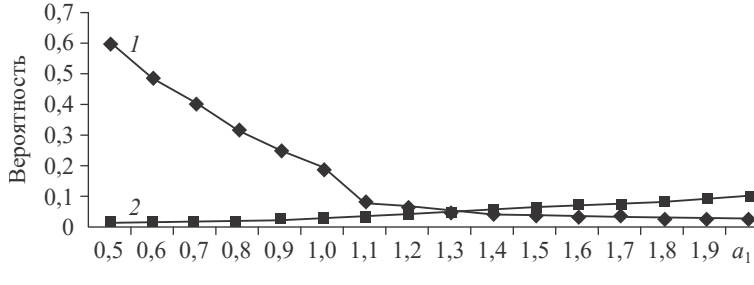

Рис. 3.11. Вероятности ошибок генерации ключа на основе признаков подписи сетью Пирсона—Хемминга при наличии 118 входов у нейронов.

1 — FRR, *2* — FAR.

Таблица 3.7

Сравнение НПБК с нечеткими экстракторами с использованием признаков подписи

Способ генерации, количество признаков	FRR	FAR	Длина ключа
Модифицированный нечеткий экстрактор (коды Адамара), 228	0,148	0,05	304 бит
Нечеткий экстрактор (коды Рида—Соломона), 236	0,228	0,076	360 бит
Модифицированный нечеткий экстрактор (коды Рида—Соломона), 90	0,191	0,033	150 бит
НПБК (1 слой), 236	0,029	0,074	236 бит
НПБК (2 слоя), 236	0,045	0,051	236 бит

Таблица 3.8

Сравнение НПБК с нечеткими экстракторами с использованием признаков клавиатурного почерка

Способ генерации	FRR	FAR	Длина ключа
Модифицированный нечеткий экстрактор (коды БЧХ)	0,104	0,021	46 бит
НПБК (1 слой)	0,033	0,031	63 бит
НПБК (2 слоя)	0,022	0,03	112 бит

Таблица 3.9

Сравнение НПБК с сетями квадратичных форм с использованием признаков подписи

Способ генерации, количество признаков	FRR	FAR	*ErrorRate*	Число входов нейрона
Сеть Пирсона—Хемминга	0,044	0,046	0,057	59
Сеть Евклида—Хемминга	0,097	0,118	0,302	59
Персептроны (1 слой) ГОСТ Р 52633.5-2011	0,028	0,076	0,067	59
Сеть Пирсона—Хемминга	0,041	0,054	0,058	118
Сеть Евклида—Хемминга	0,084	0,155	0,314	118
Персептроны (1 слой) ГОСТ Р 52633.5-2011	0,029	0,074	0,068	118
Сеть Пирсона—Хемминга	0,032	0,066	0,059	177
Сеть Евклида—Хемминга	0,066	0,211	0,320	177
Персептроны (1 слой) ГОСТ Р 52633.5-2011	0,02	0,1	0,067	177

фикации подписей целесообразно делать в 2 раза меньше количества признаков. Для признаков подписей количество входов определялось, опираясь на это требование, решено повторить все вычисления в трех вариантах: число входов нейрона равно четверти количества признаков, половине количества признаков и трем четвертям количества признаков. Число нейронов решено сделать равным количеству признаков: 236, так как общей рекомендации для сетей квадратичных форм нет.

Тестирование нейронных сетей проводилось без построения защищенного нейросетевого контейнера. Данное требование формулировалось в работах [60, 99] по отношению к стандартизованным персептронам. Лучшие результаты (по наименьшей сумме FFR и FAR) проведенного натурного моделирования приведены в табл. 3.7—3.9.

3.11. Выводы

Проведенные исследования показали, что нейронные сети в значительной степени превосходят нечеткие экстракторы по надежности генерации ключевых последовательностей, что согласуется с данными научной литературы. В первую очередь это обусловлено тем, что классический нечеткий экстрактор (и аналогичные схемы, основанные на помехоустойчивом кодировании «сырых» биометрических данных) не учитывает законы распределения значений признаков. Квантование необогащенных данных также является неэффективным. Можно сделать вывод, что нечеткие экстракторы целесообразно использовать только в совокупности с высокоинформативными признаками статических биометрических образов. Для динамических параметров подсознательных движений нечеткие экстракторы неприменимы, так как ошибки оказываются неприемлемо высокими, а длина ключа (пароля) низкой. При использовании дополнительных усовершенствований с целью учета параметров распределения битовых представлений признаков и искусственного увеличения длины биокода показатели надежности, сравнимые с нейронными сетями, все равно не достигаются. При этом теряется простота реализации классической схемы нечеткого экстрактора и ослабевают его защитные свойства. Скорость работы нечетких экстракторов (особенно на базе кодов БЧХ) существенно ниже, чем у нейронных сетей. А проблема долгого обучения нейросети решается стандартом ГОСТ Р 52633.5-2011.

По данным экспериментальных оценок, сети квадратичных форм Пирсона—Хемминга превосходят персептроны из ГОСТ Р 52633.5-2011 по надежности выработки ключевой последовательности. Сеть Евклида—Хемминга работает значительно хуже сети Пирсона—Хемминга.

При создании НПБК требуется выполнить требования ГОСТ по защите биометрического эталона, поэтому при использовании нейронов на базе иных функционалов (не персептронов), на практике данное требование можно реализовать посредством принципа защищенного нейросетевого контейнера. Целесообразно задействовать классические нейроны (персептроны), на вход которых подавать только наиболее информативные признаки с невысокой взаимной корреляцией (например, характеризующие цвет глаз и кожи). На выходах этих нейронов можно осуществить шифрование параметров классических квадратичных форм или других функционалов. Результаты проведенных экспериментов также отражены в работах [61, 114, 115].

4. ГЕНЕРАЦИЯ КЛЮЧЕВЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ С УЧЕТОМ ЗАВИСИМЫХ БИОМЕТРИЧЕСКИХ ДАННЫХ В СИСТЕМЕ ЗАЩИТЫ ГИБРИДНОГО ДОКУМЕНТООБОРОТА

4.1. Многомерные функционалы Байеса и их связь с многомерными корреляционными функциями

Исследования показывают, что корреляционная зависимость между значениями многих биометрических признаков значительна (см. рис. 2.3, 2.5, 2.7). Метрика Пирсона, к сожалению, теряет мощность при росте коррелированности данных. Ее поведение отличается от того, как ведет себя алгоритм последовательного применения двухмерного правила Байеса [31], при использовании которого влияние сильных корреляционных связей между признаками, а также эталонными описаниями образов не столь существенно сказывается на результатах идентификации. В связи с этим необходимо создать метрику, которая работала бы эффективно как с независимыми данными, так и с зависимыми.

Двухмерный вариант правила Байеса был опубликован в 1763 г. Введенное Байесом понятие совместной вероятности появления зависимых событий оказалось очень продуктивным. Оценивать уровень зависимости двух переменных посредством коэффициента корреляции стали только в XIX в. (примерно через 100 лет). Фактически Байес первым предложил эффективный инструмент учета взаимного влияния двух событий, что эквивалентно переходу к использованию двухмерных вероятностей:

$$P(\nu_1, \nu_2) = P(\nu_1 / \nu_2)P(\nu_2) = P(\nu_2 / \nu_1)P(\nu_1). \qquad (4.1)$$

В биометрии недостаточно учитывать вероятности появления пар событий. При реализации решающего правила [31, 116, 117] приходится многократно применять двухмерное правило (4.1). В связи с тем что задачи биометрической идентификации многомерны, необходимо повы-

шать размерность правила Байеса. В трехмерном варианте правило Байеса выглядит следующим образом:

$$P(\nu_1, \nu_2, \nu_3) = P((\nu_1, \nu_2)/\nu_3)P(\nu_3) = P(\nu_1/(\nu_2, \nu_3))P(\nu_2, \nu_3) = \ldots \quad (4.2)$$

Заметим, что двухмерный вариант формулы Байеса (4.1) имеет только две возможных формы записи условных вероятностей. Трехмерный вариант имеет уже 3! варианта возможных условных вероятностей. В общем случае *n*-мерный вариант формулы Байеса допускает *n*! форм записи условных вероятностей:

$$\begin{aligned} P(\nu_1, \nu_2, \ldots, \nu_n) &= P((\nu_1, \nu_2, \ldots, \nu_{n-1})/\nu_n)P(\nu_n) = \\ &= P((\nu_1, \nu_2, \ldots, \nu_{n-2})/(\nu_{n-1}, \nu_n))P(\nu_{n-1}, \nu_n) = \ldots \end{aligned} \quad (4.3)$$

Таким образом, можно получить правило Байеса для задач любой размерности. Запись многомерного варианта формулы Байеса (4.3) — формальна. На практике часто происходит последовательное применение простейшей (4.1) или модифицированной двумерной формулы Байеса при исследовании очень большого числа пар событий [31, 116, 117]. Однако исследования показывают, что многомерный вычислительный элемент Байеса работает тем лучше, чем выше его размерность и выше значения модулей коэффициентов его равной коррелированности [118].

Заметим, что биометрические данные непрерывны, и, соответственно, для непрерывных данных мы можем записать формулы Байеса применительно к многомерным плотностям распределения значений. Для этого достаточно в формулах (4.1)—(4.3) заменить большую букву «*P*», обозначающую вероятность, на малую букву «*p*», обозначающую плотность распределения вероятности. После подобного перехода можно перейти к непрерывной энтропии многомерных распределений биометрических данных:

$$\begin{cases} h(\nu_1) = -\int_{-\infty}^{+\infty} p(\nu_1)\log_2(p(\nu_1))d\nu_1, \\ h(\nu_1, \nu_2) = -\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p(\nu_1, \nu_2)\log_2(p(\nu_1, \nu_2))d\nu_1 d\nu_2, \\ \ldots \end{cases} \quad (4.4)$$

Уже в пространствах многомерной непрерывной энтропии формулы Байеса будут иметь следующий вид:

$$\begin{cases} h(\nu_1, \nu_2) = h(\nu_1/\nu_2) + h(\nu_2) = h(\nu_2/\nu_1) + h(\nu_1), \\ h(\nu_1, \nu_2, \nu_3) = h((\nu_1, \nu_2)/\nu_3) + h(\nu_3) = h(\nu_1/(\nu_2, \nu_3)) + h(\nu_2, \nu_3), \\ \ldots \end{cases} \quad (4.5)$$

В итоге получается, что мы имеем три варианта многомерных статистических описаний зависимых событий (объекта биометрической идентификации). Это все можно было бы отнести к бесполезной схоластике, если из этих многомерных описаний не следует конструктивных выводов.

Для нас конструктивным является получение связи многомерных описаний зависимых данных с коэффициентами корреляции. Оказалось, что подобная связь достаточно проста, если использовать пространства многомерных энтропий. В частности, модуль обычного коэффициента двухмерной корреляции связан следующим образом с непрерывными энтропиями:

$$\left|r(\nu_1, \nu_2)\right| = 1 - \frac{h(\nu_1, \nu_2)}{h(\nu_1) + h(\nu_2)} = 1 - \frac{h(\nu_1 / \nu_2) + h(\nu_2)}{h(\nu_1) + h(\nu_2)}. \qquad (4.6)$$

Фактически мы получили корреляционно-энтропийный вариант теоремы Байеса, который связывает коэффициент корреляции с непрерывными энтропиями двух биометрических параметров. Очевидно, что обычный двухмерный вариант легко обобщается под модуль коэффициента корреляции любой размерности:

$$\left|r(\nu_1, ..., \nu_n)\right| = 1 - \frac{h(\nu_1, \nu_2, ..., \nu_n)}{h(\nu_1) + h(\nu_2) + ... + h(\nu_n)} =$$

$$= 1 - \frac{h((\nu_1, ..., \nu_{n-1}) / \nu_n) + h(\nu_n)}{h(\nu_1) + h(\nu_2) + ... + h(\nu_n)}. \qquad (4.7)$$

При переходе в практическую плоскость реализации n-мерного правила Байеса для первого биометрического параметра получим соотношение

$$y_{k,j} = \sum_{i=1}^{m} \left| \frac{|E(a_k) - a_{k,j}|}{\sigma(a_k)} - \frac{|E(a_i) - a_{i,j}|}{\sigma(a_i)} \right|, \qquad (4.8)$$

где $a_{i,j}$ — значение i-го признака (входа нейрона) с высоким значением модуля корреляции $|r_{i,k}|$ по отношению к k-му биометрическому признаку $a_{k,j}$ $(i \neq k)$, j — номер биометрического образца образа «Свой», для которого вычисляется функционал, $E(a_i)$; $\sigma(a_i)$ — математическое ожидание и среднеквадратичное отклонение i-го признака (входа нейрона).

Нетрудно убедиться, что при функциональной зависимости признаков значение метрики (4.8) будет иметь нулевое значение для любого образца «Свой». По мере снижения коэффициентов корреляции значение функционала (4.8) возрастает, растет также его стандартное отклонение.

4.2. Метрика Байеса—Пирсона

Обычные процедуры последовательного применения двухмерных правил Байеса [31, 116, 117] достаточно трудно балансировать. Это свя-

зано с проблемами разложения многомерных функций (вероятности, плотности вероятности, непрерывной энтропии) на множество двухмерных функций. Снять проблему сшивки двухмерных функций в нужную многомерную функцию удается, если использовать специальную метрику Байеса—Пирсона [113]:

$$\chi = \sum_{j=1}^{m}\sum_{i=1}^{m}\left|\frac{|E(\nu_i)-\nu_i|}{\sigma(\nu_i)} - \frac{|E(\nu_j)-\nu_j|}{\sigma(\nu_j)}\right|, \qquad (4.9)$$

где ν_i — значение i-го признака (входа нейрона), $E(\nu_i)$ — математическое ожидание (среднее значение) i-го признака (входа нейрона), $\sigma(\nu_i)$ — среднеквадратичное отклонение i-го признака (входа нейрона).

Метрика Байеса—Пирсона (4.9), так же как и метрика Пирсона (3.6), не содержит в явной форме вычислительных операций с коэффициентами корреляции, однако коэффициенты m-мерной корреляции биометрических данных (4.7) сильно влияют на нее. Механизм влияния корреляционных связей на метрику (4.9) поясняется рис. 4.1.

На рис. 4.1 дано распределение биометрических параметров всех образов, и на его фоне — распределения значений двух сильно зависимых (сильно коррелированных) биометрических параметров ν_1 и ν_k. Будем исходить из того, что наблюдаемый параметр $\nu_1 = -5$. Тогда в зависимости от знака коэффициента корреляции между биометрическими параметрами возможно два интервала, куда могут попасть ожидаемые априорно значения распределения параметра ν_k. Мы заранее можем вычислить знак коэффициента корреляции и интервал, куда должен попасть k-й биомет-

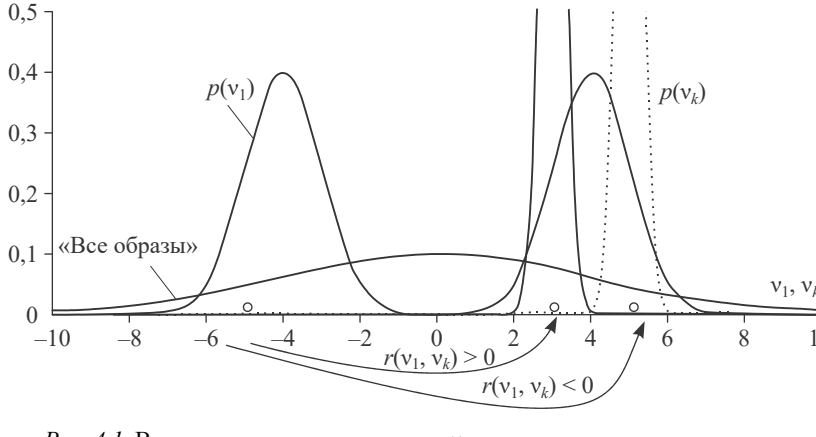

Рис. 4.1. Влияние состояния параметра ν_1 на значение сильно зависимого параметра ν_k.

рический параметр. Чем выше корреляция, тем у́же интервал, куда должен попасть k-й биометрический параметр. На рис. 4.1 ожидаемые интервалы даны узкими распределениями значений остаточной случайной составляющей, пики которых выходят за пределы поля рис. 4.1.

К сожалению, указать достаточно точно границы ожидаемых интервалов априорного попадания k-го контролируемого биометрического параметра мы не можем. Проблема состоит в том, что на малых обучающих выборках все статистические параметры $E(\nu_i)$, $\sigma(\nu_i)$, $E(\nu_k)$, $\sigma(\nu_k)$, $r(\nu_i, \nu_k)$ мы оцениваем со значительными ошибками. Самыми низкоточными являются оценки коэффициентов корреляции. На рис. 4.2 даны распределения ошибочных оценок значений коэффициентов корреляции при обучающих выборках в 9, 16, 32, 64 примеров биометрического образа «Свой».

Из рис. 4.2 видно, что при использовании 16 примеров биометрического образа «Свой» ошибки в оценках коэффициентов корреляции ±0,6 для данных со слабой зависимостью (центр рисунка) и ±0,2 для сильно зависимых данных (края рисунка). При таких значениях ошибок предсказывать границы ожидаемых интервалов для зависимых данных сложно. В этом как раз и состоит проблема настройки решающих правил с многократным применением двухмерного правила Байеса (4.1).

Основным преимуществом метрики Байеса—Пирсона (4.9) является то, что она не предусматривает выставление порогов на каждый контролируемый параметр. Применение этой метрики предполагает выставление одного порога на все m контролируемых биометрических параметров. Этот единственный порог должен выставляться таким, чтобы все

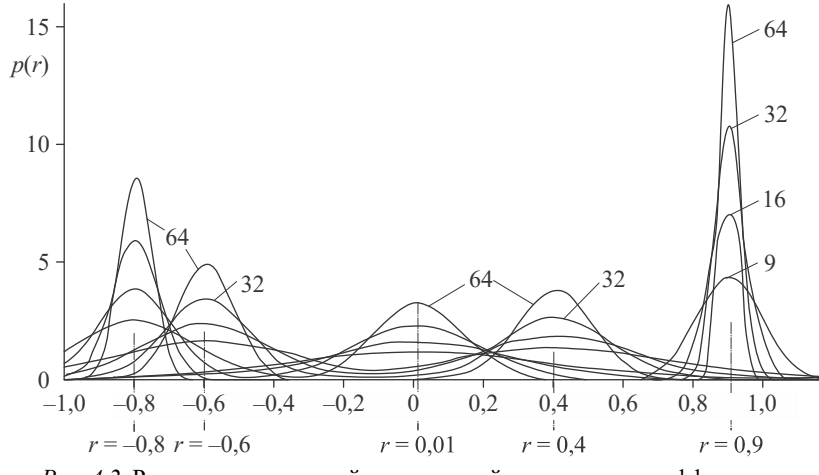

Рис. 4.2. Распределения случайных значений вероятности коэффициентов корреляции, обусловленных конечной выборкой наблюдений.

обучающие примеры образа «Свой» давали бы значение метрики (4.9) менее порога. Тогда все образы, чья метрика Байеса—Пирсона (4.9) выше порога, следует признавать «Чужими». Если ошибки в вычислениях параметров $E(\nu_i)$, $\sigma(\nu_i)$ случайны, они усредняются при вычислениях. Метрика Байеса—Пирсона (4.9) подавляет случайные ошибки оценок статистических моментов $E(\nu_i)$, $\sigma(\nu_i)$. Чем выше размерность метрики m, тем сильнее осуществляется подавление случайных ошибок $\Delta E(\nu_i)$, $\Delta\sigma(\nu_i)$.

Сеть Байеса—Пирсона—Хемминга на основе функционалов (4.9) формируется аналогично сети Пирсона—Хемминга.

4.3. Особенности формирования сетей Байеса—Хемминга

Правило Байеса построено на учете априорной информации о том, что параметры зависимы. Опираясь на значение одного параметра, можно сузить интервал допустимых значений другого параметра (см. рис. 4.2). Математическое ожидание i-го зависимого параметра для центрированных и нормированных биометрических параметров совпадает до знака с уже определенным параметром ν_c:

$$\begin{cases} E(\nu_i) = \nu_c, & \text{если } r(\nu_i, \nu_c) \geq 0, \\ E(\nu_i) = -\nu_c, & \text{если } r(\nu_i, \nu_c) \leq 0. \end{cases} \tag{4.10}$$

Для оценки значения коэффициента сжатия необходимо использовать только положительно коррелированные данные, далее средствами имитационного моделирования зависимых данных [119, 120] следует получить плотности распределения модуля разности значений связанных биометрических параметров:

$$\Delta = |\nu_i - \nu_c|. \tag{4.11}$$

Численный эксперимент дает разные плотности распределения значений $p(\Delta)$ для разных значений их взаимной коррелированности r. Кривые полученных плотностей распределения значений приведены на рис. 4.3. Он иллюстрирует интуитивно понятную связь: чем выше коррелированность данных, тем уже интервал, куда должен попасть второй

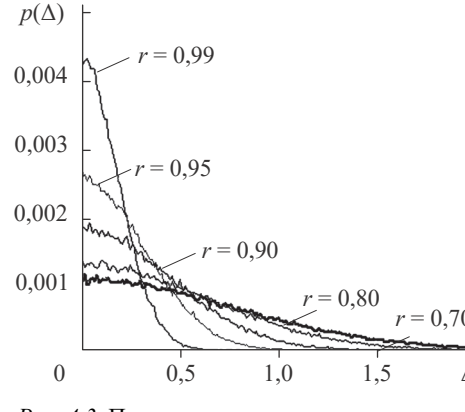

Рис. 4.3. Плотности вероятности интервалов ожидания появления зависимых событий.

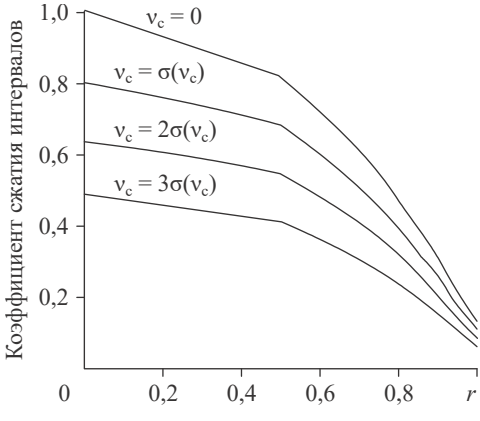

Рис. 4.4. Кривые коэффициента сжатия интервалов ожидаемого появления зависимого параметра.

контролируемый биометрический параметр. Пользуясь выявленными статистическими зависимостями, можно рассчитать коэффициент сжатия интервала вокруг математического ожидания (4.10). Численное моделирование показывает, что на коэффициент сжатия влияет не только коррелированность данных, но и то, насколько математическое ожидание (4.10) второго параметра отклоняется от центра. Данные моделирования отражены на рис. 4.4.

Из приведенных выше данных следует, что решающее правило Байеса будет работать тем лучше, чем сильнее будут коррелированы используемые биометрические параметры. Также требуется, чтобы корреляция была приблизительно равной. Поэтому при формировании сети функционалов (4.8) изначально вычислялись коэффициенты парной корреляции r между всеми сечениями (совокупностями значений) всех признаков обучающей выборки. Обработчики признаков соединялись с входами нейронов исходя из модуля равной коррелированности, под которой подразумевается, что разница $|r|$ для признаков не превышает τ (значение τ задавалось как параметр). Сначала формируется K_1 нейронов для первого признака, потом K_2 — для второго и так далее. С первым из K_1 нейронов соединяются первый признак и признаки, которые имеют модуль коэффициента корреляции с ним в интервале $[1; 1 - \tau]$, если их число не менее двух (минимальная размерность). Далее производится поиск признаков, имеющих модуль коэффициента корреляции с первым признаком в интервале $[1 - \tau; 1 - 2\tau]$, процедура повторяется многократно, пока $1 - l\tau > \phi$, где ϕ — максимальный модуль коэффициента корреляции между признаками, менее которого признаки не учитывались, l — номер итерации поиска признаков с равной корреляцией. Аналогичным образом формируются по K_j нейронов на признак, с каждым из которых связан j-й признак и i-е признаки, если $1 - (l - 1)\tau < |r_{i,j}| < 1 - l\tau$. Общее количество нейронов сети равно сумме коэффициентов K_j, т. е. число нейронов для каждой сети (для каждого из испытуемых) не является фиксированным, как и количество входов нейронов.

4.4. Снижение требований к размеру обучающей выборки при переходе к использованию многомерных корреляционных функционалов Байеса

Применение искусственных нейронных сетей вместо классических квадратичных форм [121—123] следует рассматривать как один из методов топологической регуляризации нерешаемой задачи. Производится топологическая замена очень трудной (технически не решаемой) задачи обращения корреляционных матриц высокой размерности на независимое обучение множества искусственных нейронов. Классический коэффициент парной корреляции двух массивов (биометрических параметров — сечений значений двух разных признаков, или реализаций — двух векторов значений признаков) вычисляется по следующей формуле:

$$r(x_1, x_2) = \frac{1}{n}\sum_{i=1}^{n}\frac{(E(x_1) - x_{1,i})(E(x_2) - x_{2,i})}{\sigma(x_1)\sigma(x_2)}, \qquad (4.12)$$

где $E(.)$ — операция вычисления математического ожидания, $\sigma(.)$ — операция вычисления стандартного отклонения, n — размер массива.

На малых выборках биометрических данных возникают значительные ошибки вычисления математических ожиданий $\Delta E(x_1)$, $\Delta E(x_2)$ и стандартных отклонений $\Delta\sigma(x_1)$, $\Delta\sigma(x_2)$. Совершенно так же, как и при решении систем линейных уравнений, возникает эффект накопления ошибок. Оценить возможные значения ошибок вычисления коэффициентов парной корреляции можно по распределениям, приведенным на рис. 4.5.

Из рис. 4.5 видно, что при размере выборки $n = 4$ значения коэффициента корреляции попадают в интервал от 0,85 до 0,95 с вероятностью 0,35. Однако если размер тестовой выборки увеличить до 8, 16, 32, 64, происходит рост вероятности попадания вычисляемых значений в заданный интервал до величин 0,59, 0,77, 0,89, 0,96 соответственно.

Из теории известно [120, 123], что многомерные биометрические данные «Свой» всегда можно симметризовать, т. е. заменить иными многомерными данными с той же самой энтропией. После симметризации корреляционная матрица биометрических данных вне диагонали будет иметь одинаковые коэффициенты парных корреляционных связей. Если биометрические данные сильно коррелированы, то мы получим вне диагонали коэффициенты парной корреляции близкими к единице. Ситуация, когда все коэффициенты корреляционной матрицы вне диагонали близки к единице, приводит к очень плохой обусловленности матриц квадратичных форм. В частности для данных, соответствующих рис. 4.5, пятимерная корреляционная матрица будет иметь большое число обусловленности:

$$\operatorname{cond}\begin{bmatrix} 1 & 0,9 & 0,9 & 0,9 & 0,9 \\ 0,9 & 1 & 0,9 & 0,9 & 0,9 \\ 0,9 & 0,9 & 1 & 0,9 & 0,9 \\ 0,9 & 0,9 & 0,9 & 1 & 0,9 \\ 0,9 & 0,9 & 0,9 & 0,9 & 1 \end{bmatrix} = 46, \tag{4.13}$$

$$\overline{\lambda} = \begin{bmatrix} 4,6 \\ 0,1 \\ 0,1 \\ 0,1 \\ 0,1 \end{bmatrix}. \tag{4.14}$$

То есть ошибки, возникающие при вычислении коэффициентов корреляции при попытках применения квадратичных форм (3.5), будут усиливаться примерно в 46 раз, что недопустимо для практики.

Неприятности, порождаемые симметричными корреляционными матрицами, связаны с тем, что число обусловленности определяется как отношение максимального и минимального собственных чисел λ_i используемой матрицы:

$$\operatorname{cond}[R] = \frac{\max(\lambda_i)}{\min(\lambda_i)}. \tag{4.15}$$

Как видно из выражения (4.15), собственные числа симметричной корреляционной матрицы (4.13) имеют одно большое значение $\lambda_1 = 4,6$ и четыре одинаковых малых значений собственных чисел $\lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = 0,1$. Симметричные корреляционные матрицы являются наихудшим вариантом для квадратичных форм (3.5), так как они не по-

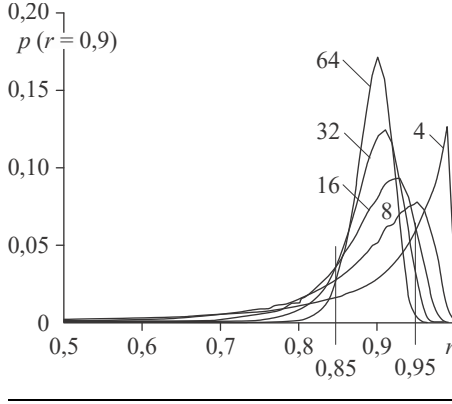

Рис. 4.5. Плотности распределения значений коэффициентов парной корреляции для сильно зависимых данных $E(r) = 0,9$ при $n = 4, 8, 16, 32, 64$ примеров.

зволяют выделить и отбросить пару наиболее зависимых биометрических параметров.

Полная симметрия корреляционных связей, являющаяся наихудшей для квадратичных форм, оказывается наилучшим соотношением данных для функционалов наибольшего правдоподобия Байеса [91]. Следует отметить, что классический двухмерный коэффициент корреляции (4.12) по своей сути есть не что иное, как одна из форм записи двумерного правила Байеса (4.1). В этом легко убедиться, рассматривая предельные значения коэффициентов корреляции:

$$\begin{cases} P(x_1) = P(x_2) \quad \text{или} \quad P\left(x_1/x_2\right) = P\left(x_2/x_1\right) = 1 \quad \text{при} \quad r(x_1, x_2) = 1; \\ P(x_1, x_2) = P(x_1)P(x_2) \quad \text{при} \quad r(x_1, x_2) = 0. \end{cases} \tag{4.16}$$

Для интервала корреляционных связей от 0,7 до 1,0 корреляционный вариант формулы Байеса можно записать следующим образом:

$$r\left(x_2, x_1\right)P(x_1) \approx r(x_1, x_2)P(x_2). \tag{4.17}$$

Приближение (4.17) становился точным равенством только при предельно высоких значениях коэффициентов парной корреляции. В принципе можно построить некоторые корреляционные функции $F(r(x_1, x_2), P(x_2))$, $F(r(x_1, x_2), P(x_1))$, которые приближение (4.17) сделают равенством при всех больших значениях модулей коэффициентов корреляции. Важно то, что коэффициенты парных корреляций функционально связаны с правилом Байеса и, соответственно, на их базе может быть создано множество решающих правил Байеса разной размерности. Одним из самых простых правил является использование равнокоррелированных биометрических параметров с суммированием близких значений, вычисленных на реальных данных коэффициентов парной корреляции. В частности, может быть использован симметричный корреляционный функционал Байеса третьего порядка:

$$R(x_1, x_2, x_3) = \frac{r(x_1, x_2) + r(x_1, x_3) + r(x_2, x_3)}{3}. \tag{4.18}$$

Решающее правило для такого функционала строится в виде правого и левого порогов для его допустимых значений от $\min(R(x_1, x_2, x_3))$ до $\max(R(x_1, x_2, x_3))$. Определение на реальных биометрических данных образа «Свой» допустимых порогов является настройкой (обучением) трехмерных решающих правил Байеса.

Очевидно, что по аналогии с трехмерным корреляционным функционалом Байеса может быть использован четырехмерный симметричный корреляционный функционал Байеса. В общем случае m-мерный симмет-

ричный корреляционный функционал Байеса имеет следующее формальное описание:

$$R(x_1, x_2, ..., x_n) = \{r(x_1, x_2) + r(x_1, x_2) + \ldots + r(x_1, x_n) + r(x_2, x_3) + \ldots$$
$$+ r(x_2, x_n) + r(x_{n-1}, x_n)\} \frac{1}{(n-1) + (n-2) + \ldots + (n-n)}. \quad (4.19)$$

Нетрудно заметить, что корреляционные функционалы Байеса третьего порядка построены на усреднении трех коэффициентов парной корреляции, каждый из которых имеет случайную ошибку, обусловленную малой выборкой биометрических данных [124]. Так как происходит усреднение трех коэффициентов парной корреляции, случайные составляющие ошибок каждого из частных коэффициентов корреляции ослабляются в $\sqrt{3}$ раз в значениях результирующего трехмерного корреляционного функционала (4.18). Четырехмерный корреляционный функционал подавляет случайную составляющую в $\sqrt{6}$ раз. При применении m-мерного корреляционного функционала происходит подавление случайных составляющих ошибки примерно в $\sqrt{(m^2 - m)/2}$ раз. С ростом размерности решающих правил мы наблюдаем почти линейный рост подавления случайных ошибок, возникающих из-за малого размера обучающей (тестовой) выборки.

Эффект подавления случайных составляющих погрешности с ростом размерности корреляционных функционалов Байеса можно оценить численным моделированием. Для этой цели необходимо использовать равнокоррелированные данные, получаемые умножением вектора псевдослучайных независимых чисел на симметричную связывающую матрицу [120, 123] в соответствии с формулой (3.19). В итоге мы получаем выходные данные с близкими коэффициентами корреляции (рис. 4.6).

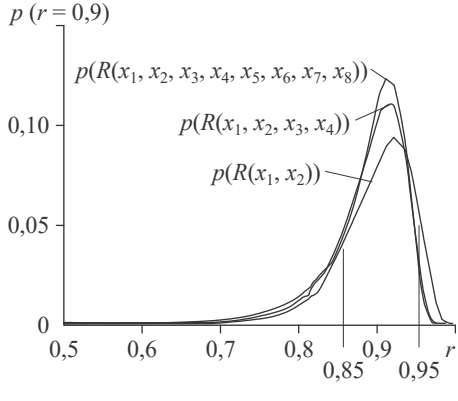

Рис. 4.6. Распределения значений классических двухмерных коэффициентов корреляции, а также четырехмерных и восьмимерных корреляционных функционалов Байеса, вычисленных на выборках из 16 примеров.

Из рис. 4.6 видно, что по мере увеличения размерности корреляционных функционалов Байеса монотонно падает их стандартное отклонение, что приводит к росту вероятности попадания вычисленного функционала в интервал значений от 0,85 до 0,95. Так, для классического двухмерного коэффициента корреляции выборка из 16 примеров дает вероятность 0,77 попадания значений в заданный интервал. Переход к использованию четырехмерных корреляционных функционалов позволяет повысить вероятность попадания в интервал от 0,85 до 0,95 до величины 0,84. Это эквивалентно увеличению тестовой выборки с 16 до 24 примеров (рост объема выборки на 50 %). Еще больший рост объема выборки будет наблюдаться при переходе от двухмерных коэффициентов корреляции к восьмимерным корреляционным функционалам. В этом случае в заданный интервал значения функционалов попадают с вероятностью 0,88, что эквивалентно росту обучающей выборки с 16 до 31 примера (рост объема выборки на 94 %). При дальнейшем увеличении роста размерности вычисляемых корреляционных функционалов монотонно будет падать случайная составляющая погрешности вычислений. Это эквивалентно некоторой топологической регуляризации вычислений, осуществляемых сетями Байеса—Хемминга [91].

4.5. Экспериментальное сравнение способов генерации ключевых последовательностей на основе подписей субъектов

Проведен вычислительный эксперимент. Подписи были преобразованы в реализации. Для обучения использовалось по 21 реализации образа «Свой» (а также по одной реализации каждого образа для обучения персептронов на данных «Чужой»). Далее проводились серии опытов по верификации субъектов сетями персептронов, Байеса—Пирсона—Хемминга, Пирсона—Хемминга и Байеса—Хемминга. В процессе эксперимента изменялись значения параметров N, m, τ и ϕ. Подсчитывалось общее число ошибок 1-го и 2-го рода, FRR и FAR вычислялись как отношения количества ошибок соответствующего рода к числу проведенных опытов с использованием соответственно реализаций «Свой» и «Чужой». Рисунок 4.7 демонстрирует графики вероятностей ошибки верификации любого рода испытуемых при оптимальных значениях порога (определяется по минимуму FRR + FAR в каждой серии испытаний).

При использовании только «плохих», сильно коррелирующих признаков ($0{,}5 \le \phi < 0{,}9$) вероятность ошибки генерации сетями Байеса—Хемминга (рис. 4.7) ниже, чем в случае, если использовать также признаки со слабой корреляцией ($\phi < 0{,}5$). Это подтверждает тезис о том, что много-

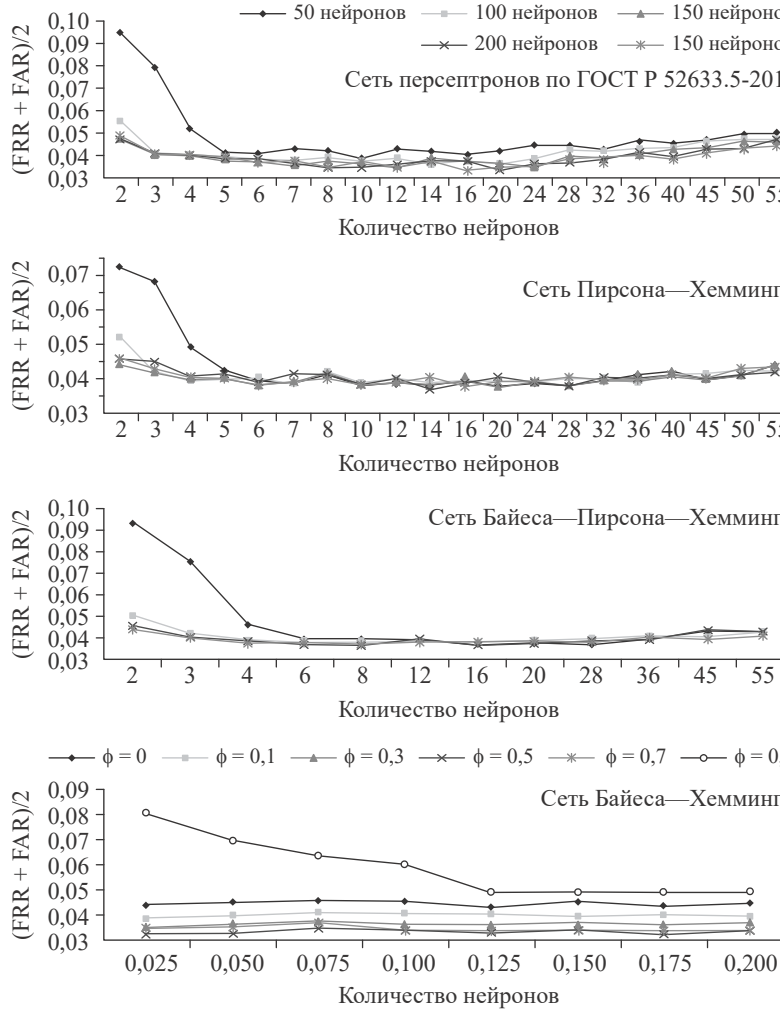

Рис. 4.7. Вероятности ошибок верификации 65 испытуемых (при пороге $H = 0$).

мерный функционал Байеса работает тем лучше, чем выше корреляция между признаками (при работе с «хорошими» признаками функционал с большей вероятностью выдает неверный бит кода). Число признаков с высокой корреляцией невелико (см. рис. 2.4), поэтому при $\phi = 0{,}9$ размерность сети низкая и ошибок много, но при увеличении значения τ их количество интенсивно снижается. В остальных случаях этот параметр влияет на вероятность ошибок незначительно. С ростом значения τ ин-

тервал равной корреляции расширяется, т. е. в один функционал попадают признаки со все более различной взаимной корреляцией. Одновременно с этим размерность функционалов m повышается. Это не противоречит тезису о том, что многомерный функционал Байеса повышает работоспособность при увеличении размерности, если признаки равно коррелированы, т. е. снижение вероятности может не происходить по причине неравной корреляции. Однако при $\phi = 0{,}9$ почти все признаки коррелируют одинаково, так как τ не может превысить 0,1 (при $\phi = 0{,}9$, $\tau > 0{,}1$ ничего не меняется, см. рис. 4.7).

Из рис. 4.7 видно, что повышение размерности функционалов снижает вероятность ошибок до определенного момента — участка насыщения. Увеличение N не влияет существенным образом на вероятность ошибки, если не корректировать коды, генерируемые сетью. Корректировка ключевой последовательности может быть осуществлена вторым слоем нейронов, реализованным по ГОСТ Р 52633.5-2011 [22], или методом помехоустойчивого кодирования, предложенным в [102] и разработанным специально для биометрии. Второй вариант предпочтительней, так как позволяет скорректировать фиксированное количество бит кода.

Рисунок 4.8 иллюстрирует следующее: если повысить пороговое значение расстояния Хемминга H от генерируемого до верного кода, то можно получить выигрыш по сумме FRR + FAR. Это и позволяет получить метод помехоустойчивого кодирования из [102] без необходимости в хранении эталонного (верного) кода, на который настраивается сеть, для вычисления H. После анализа результатов эксперимента были найдены оптимальные пороговые значения H, при которых сумма ошибок 1-го и 2-го рода была минимальной. Уменьшение ошибок (ΔEr — коэффициент снижения FRR + FAR) может достигаться, пока корреляция выходных значений нейронов невысока. В этом отношении разные сети имеют различия (рис. 4.9). Наиболее показательными являются результаты верификации сетями Байеса—Хемминга, даже с учетом, что при $\phi = 0{,}9$ вероятность ошибки не меняется (см. рис. 4.7 и 4.10).

Полученные результаты (табл. 4.1) могут быть улучшены при добавлении множества менее информативных признаков (двумерные и трехмерные амплитуды гармоник при синусах и косинусах произведения функций $x(t)\,y(t)$ и $x(t)\,y(t)$, $p(t)$, коэффициенты корреляции между коэффициентами одно-, двух- и трехмерного ряда Фурье, а также вейвлет-преобразований). Рассмотренная система из 236 признаков имеет максимум по r (при $r \approx 0$, см. рис. 2.4). В интервал [0;0,3] попадает порядка 50 % признаков, их следует анализировать сетями Пирсона—Хемминга. Более 30 % имеет модуль корреляции в интервале [0,3; 0,7], данные признаки следует обрабатывать сетями персептронов, обученными по ГОСТ Р 52633.5. По-

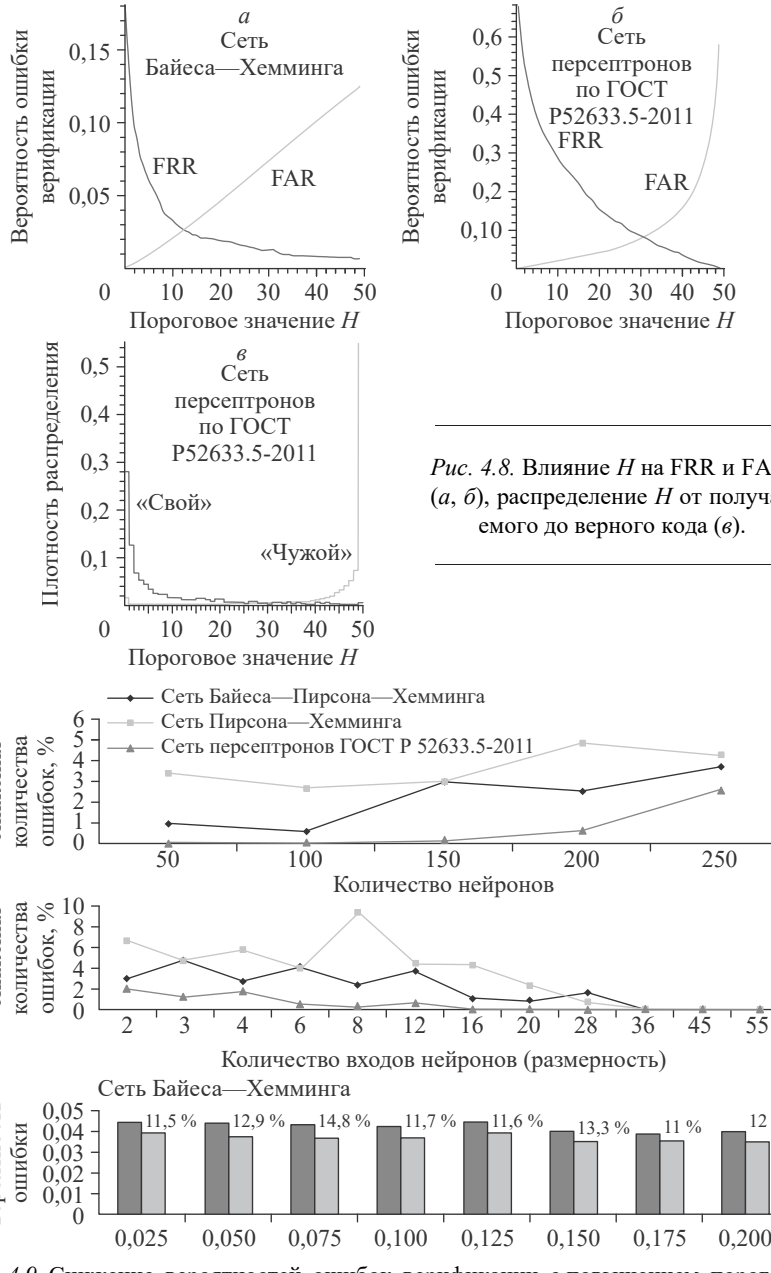

Рис. 4.8. Влияние *H* на FRR и FAR (*а*, *б*), распределение *H* от получаемого до верного кода (*в*).

Рис. 4.9. Снижение вероятностей ошибок верификации с повышением порогового значения *H*.

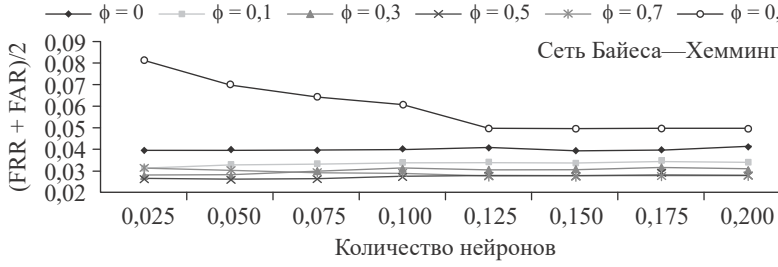

Рис. 4.10. Вероятности ошибок верификации 65 испытуемых (при пороге $H > 0$).

Таблица 4.1

Наилучшие полученные параметры надежности верификации субъектов по подписи

Тип сети	FRR, $H = 0$	FAR, $H = 0$	Параметры сети, $H = 0$	FRR, $H > 0$	FAR, $H > 0$	Параметры сети, $H > 0$	Среднее ΔEr, %
Байеса—Пирсона—Хемминга	0,0421	0,0308	$m = 8$; $N = 200$	0,0407	0,029	$m = 4$; $N = 250$	2,08
Пирсона—Хемминга	0,0317	0,0421	$m = 14$; $N = 200$	0,0236	0,0459	$m = 5$; $N = 50$	3,59
НПБК ГОСТ Р 52633.5	0,0307	0,0361	$m = 16$; $N = 250$	0,0307	0,0361	$m = 16$; $N = 250$	0,6
Байеса—Хемминга	0,034	0,0312	$\tau = 0{,}05$; $\phi = 0{,}5$	0,0288	0,0232	$\tau = 0{,}05$; $\phi = 0{,}5$	12,3

рядка 30 % признаков с $|r| > 0{,}5$—0,7 следует обрабатывать сетями многомерных корреляционных функционалов Байеса—Хемминга. Создав сеть из нескольких видов нейронов, можно снизить FRR и FAR.

4.6. Экспериментальное сравнение способов генерации ключевых последовательностей на основе данных непрерывного мониторинга пользователей компьютерных систем

Проведен аналогичный вычислительный эксперимент (изменялись значения параметров N, m, τ и ϕ — максимальный модуль коэффициента корреляции между признаками, менее которого признаки не учитывались сетью Байеса—Хемминга). При верификации генерируемого ключа-пароля решение принималось исходя из расстояния Хемминга H. После выработки ключа-пароля сетью ($H = 0$) производилась корректировка его

ошибочных бит при помощи кодов из работы [102] ($H > 0$). Эксперимент повторялся в трех вариантах: на входы сетей подавались векторы значений признаков, вычисляемые по данным мониторинга длительностью 30, 60 и 150 с (синтезированы усредненные естественные биометрические образы, количество тестовых образцов каждого вида составило соответственно 12000, 6000 и 2400). При этом предварительно вычислялись средние значения дублирующихся признаков. Это позволяет обеспечить повышение стабильности значения признака — снизить среднеквадратичное отклонение (рис. 4.11). За счет этого при увеличении длительности мониторинга и накоплении большего количества значений признаков удается добиться снижения числа ошибок (табл. 4.2). При увеличении времени мониторинга вероятность ошибок существенно снижается.

При времени мониторинга 150 с вероятность ошибок верификации субъектов по любым признакам любыми сетями, кроме персептронов, близка к нулю (ошибок не было зафиксировано, поэтому можно указать приблизительное равенство ≈0 либо, если быть точнее, исходя из количества проведенных опытов, можно указать, что получена вероятность ошибки менее 0,0005). Дополнительная корректировка ошибок в ключепароле позволяет достичь существенно более низкой суммы вероятностей ошибок 1-го и 2-го рода (в этом можно убедиться, по графикам ошибок, приведенным в Приложении А).

Наилучший результат в разных случаях достигается различными сетями. Сеть Байеса—Хемминга не дает наилучших результатов, из-за малого количества сильно коррелирующих признаков (см. рис. 2.6 и 2.8).

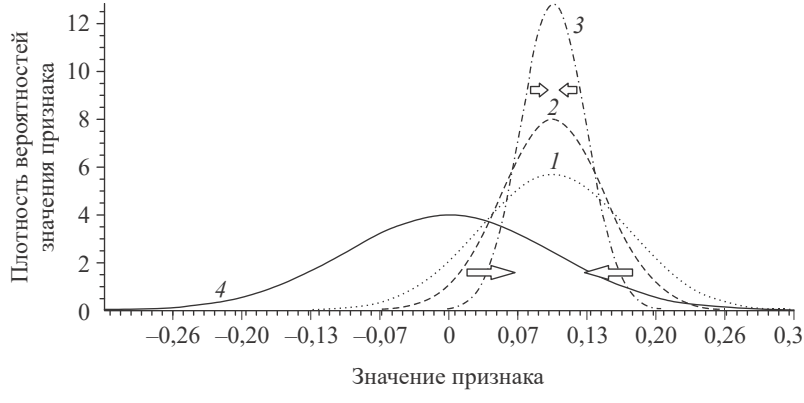

Рис. 4.11. Уменьшение среднеквадратичного отклонения (отмечено стрелками) средних значений признаков при увеличении времени мониторинга.

1 — «Свой» 30 с, *2* — «Свой» 60 с, *3* — «Свой» 150 с, *4* — все «Чужие».

Таблица 4.2

Наилучшие полученные параметры надежности верификации субъектов по лицу и клавиатурному почерку

Признаки	Время мониторинга	Тип сети	Параметры сети	FRR	FAR
Лицо	30 с	Пирсона—Хемминга	$N = 50$, $m = 5$	0,0014	0,002
Клавиатурный почерк	30 с	Персептроны	$N = 80$, $m = 10$	0,058	0,0636
Лицо + клавиатурный почерк	30 с	Байеса—Пирсона—Хемминга	$N = 120$, $m = 5$	0,002	0,0036
Лицо	60 с	Пирсона—Хемминга	$N = 80$, $m = 15$	≈0	0,0002
Клавиатурный почерк	60 с	Персептроны	$N = 200$, $m = 50$	0,034	0,043
Лицо + клавиатурный почерк	60 с	Персептроны	$N = 60$, $m = 25$	0,002	0,0009
Лицо	150 с	Любой	Различные	≈0	≈0
Клавиатурный почерк	150 с	Любой (кроме персептронов)	Различные	≈0	≈0
Лицо + клавиатурный почерк	150 с	Любой	Различные	≈0	≈0

При поступлении в сеть Байеса—Хемминга слабо коррелирующих признаков совместно с сильно коррелирующими вероятности FRR и FAR становятся выше, т. е. многомерный функционал Байеса работает тем лучше, чем сильней корреляция между признаками. Детальные результаты можно видеть в Приложении А.

4.7. Выводы

Сформулируем ключевые выводы. В процессе работы экспериментально подтвержден тезис о том, что многомерный функционал Байеса работает тем лучше, чем выше коэффициент равной коррелированности признаков и выше его размерность. Функционал Байеса (4.8) отличается от квадратичных форм (3.5) и (3.6), которые утрачивают свою работоспособность при росте корреляционных связей в биометрических данных в отличие от функционалов Байеса, которые, наоборот, улучшают свою работоспособность. Повышение размерности функционалов позволяет снизить вероятность ошибок до попадания на участок насыщения, даль-

нейшее повышение размерности не дает преимуществ. При реализации решающих правил нет смысла экономить на вычислительных ресурсах. Сформировав сеть из нескольких решающих правил с близкими вероятностями ошибок и объединив их результаты, удается получить существенный выигрыш в показателях интегральной ошибки распознавания субъектов. Наивысшим потенциалом по снижению ошибок таким способом обладает сеть Байеса—Хемминга (12,38 %), наименьшим — персептроны (< 1 %). При генерации кода на основе биометрических данных аналогичного эффекта можно добиться корректировкой нескольких неверных бит на выходе первого слоя сети нейронов [102]. Увеличение количества решающих правил целесообразно проводить, пока правила ошибаются по-разному, т. е. не полностью коррелированы.

На малых тестовых выборках коэффициенты корреляции биометрических данных имеют значительную погрешность. Это препятствует их использованию при обучении (настройке) классических квадратичных форм и сетей Байеса. Как следует из материалов данной главы, увеличение размерности корреляционных функционалов Байеса эквивалентно формальному увеличению тестовой выборки. Применение четырехмерных корреляционных функционалов Байеса эквивалентно увеличению на 50 % размеров тестовой выборки. Этот эффект можно рассматривать как значительное снижение числа обусловленности или как значительное повышение устойчивости вычислений по отношению к случайным ошибкам, порождаемым малыми объемами тестовых выборок.

Отмеченный выше эффект повышения устойчивости вычислений наблюдается для любых биометрических данных, однако для получения этого эффекта необходимо использовать только одинаково коррелированные по модулю биометрические данные. Главное — это близость модулей коэффициентов корреляции. Многомерные корреляционные функционалы Байеса (4.8) легко могут быть модифицированы под использование одинаково коррелированных по модулю биометрических данных.

Крайне важным является то обстоятельство, что сетями Байеса—Хемминга можно дополнить и сети квадратичных форм, и нейронные сети. Сети Байеса ведут себя противоположно иным технологиям. Увеличение коррелированности биометрических данных приводит к улучшению работы сетей Байеса и упрощению их настройки. Для всех других рассмотренных технологий эффект от увеличения корреляционных связей биометрических данных обратный. В связи с этим сети Байеса—Хемминга следует использовать в совокупности с рассматриваемыми здесь другими технологиями для создания нейронной сети. При создании сети из различных видов нейронов требуется выполнить требования ГОСТ по защите биометрического эталона, что можно реализовать по-

средством принципа защищенного нейросетевого контейнера. Поэтому целесообразно задействовать классические нейроны (персептроны), на вход которых подавать только наиболее информативные и низко или средне коррелирующие признаки (с коэффициентом корреляции до 0,7). На выходах этих нейронов можно осуществить шифрование параметров классических квадратичных форм, нейронов Байеса или Пирсона. На входы квадратичных форм (нейронов Пирсона) можно подавать данные с малой корреляцией (коэффициент корреляции до 0,3).

С увеличением времени мониторинга стандартного оборудования компьютера удается существенно снизить количество ошибочных решений при верификации субъекта (генерации ключевых последовательностей) на основе клавиатурного почерка и параметров лица. При времени мониторинга 150 с вероятность ошибок снижается почти до нуля при использовании признаков лица и клавиатурного почерка в отдельности и совместно. При комплексировании двух независимых образов резко усложняется возможность фальсификации системы. Вероятности ошибок составили:

- 30 с — FRR = 0,002, FAR = 0,0036;
- 60 с — FRR = 0,002, FAR = 0,0009;
- 150 с — FRR < 0,0005, FAR < 0,0005.

Для различных признаков могут быть найдены функционалы, наилучшим образом работающие с ними. Подбирать функционалы целесообразно исходя из взаимной корреляции между признаками и площадей пересечения плотностей вероятности их значений. Для признаков подписи этот функционал является байесовским. Наилучший результат по генерации ключевых последовательностей на основе подписей был получен сетью Байеса—Хемминга и составил: FRR = 0,0288; FAR = 0,0232.

ЗАКЛЮЧЕНИЕ

Основные результаты данной работы можно сформулировать следующим образом:

1. Разработаны модель и технология защиты гибридного документооборота на основе тайных или открытых рукописных образов, позволяющая использовать равные средства защиты электронных и бумажных версий (реализаций) документов, осуществлять быструю проверку их целостности и аутентичности, запрет на передачу секретного ключа ЭП третьим лицам, восстановление оригинала документа при нарушении его целостности.

2. Предложена методика вычисления информативных биометрических признаков, основанных на анализе клавиатурного почерка и видеоданных, получаемых от веб-камеры, включающая локализацию лица субъекта при работе пользователя компьютерной системы с электронными документами в реальном времени. Предложен способ создания эталона пользователя в пространстве признаков лица и клавиатурного почерка.

3. Предложен способ формирования эталонов подписантов из 21 реализации открытого или тайного рукописного образа (автографа или пароля), основанный на спектральном и корреляционном анализе функций координат и давления пера на устройство ввода, вычислении матрицы расстояний между ключевыми точками рукописного образа в трехмерном пространстве (давление — третье измерение) и определения некоторых параметров изображения рукописного образа. Предложена методика исключения некорректных реализаций рукописного образа (введенных субъектом с явными отклонениями) при формировании эталона, снижающая долю ошибок последующей генерации нечеткими экстракторами секретных ключей ЭП из вновь введенных субъектом реализаций данного образа. По результатам эксперимента зарегистрировано до 6,5 %

меньше ошибок 1-го рода и до 46 % меньше ошибок 2-го рода при использовании разработанной методики.

4. Получены результаты вычислительных экспериментов по оценке надежности генерации ключевых последовательностей на основе выявленных признаков с использованием нечетких экстракторов на базе различных кодов, исправляющих ошибки (Адамара, БЧХ, в частности Рида—Соломона) и сетей искусственных нейронов на базе различных функционалов (персептронов, метрик Пирсона, Байеса—Пирсона, Евклида и многомерных функционалов Байеса) и параметров их сетей (количество нейронов, их входов и др.) с последующей корректировкой нестабильных бит специальными самокорректирующимися кодами.

5. Разработана модель модифицированного нечеткого экстрактора с применением методики оценки стабильности битовых (квантованных) представлений признаков индивидуально для каждого субъекта, существенно снижающей вероятности ошибок 1-го и 2-го рода при генерации ключевых последовательностей.

6. Экспериментально подтверждено, что нечеткие экстракторы существенно уступают нейронным сетям по надежности генерации ключевых последовательностей (вероятностям ошибок и длине получаемого ключа шифрования, ЭП или пароля).

7. Экспериментально подтверждено, что многомерный функционал Байеса совершает меньшее количество ошибок, если возрастают коэффициент равной коррелированности признаков и его размерность.

8. Показано, что многомерные корреляционные функционалы Байеса могут быть модифицированы под использование одинаково коррелированных по модулю биометрических данных. На малых тестовых выборках коэффициенты корреляции биометрических данных имеют значительную погрешность. Это препятствует их использованию при обучении (настройке) классических квадратичных форм и сетей Байеса. Предложено воспользоваться приемом симметризации корреляционных связей. Доказано, что в этом случае требования к объему биометрических данных существенно снижаются. Как следствие, настройка (обучение) квадратичных форм и сетей наибольшего правдоподобия Байеса становятся гораздо более устойчивыми задачами. Последнее эквивалентно многократному снижению требований к размерам обучающей выборки примеров биометрического образа «Свой».

9. Подтверждено, что повышение количества функционалов и их размерности позволяет снизить вероятность ошибок до попадания на участок насыщения, после чего вероятность ошибок не снижается. Наивысшим потенциалом по снижению ошибок за счет повышения количества нейронов в рамках проведенных экспериментов обладает сеть

Байеса—Хемминга (12,38 %), наименьшим — персептроны (< 1 %). Для различных признаков могут быть найдены функционалы, наилучшим образом работающие с ними. Подбирать функционалы целесообразно исходя из взаимной корреляции между признаками и площадей пересечения плотностей вероятности их значений.

10. В серии экспериментов определены следующие оптимальные способы генерации ключевых последовательностей (которые можно использовать в качестве ключей для шифрования или формирования ЭП, если задействовать при генерации принцип защищенного нейросетевого контейнера) и последующей аутентификации субъектов:

- способ генерации ключа (пароля) в реальном времени на основе признаков лица с вероятностями ошибочных решений, зависящими от времени мониторинга действий субъекта при активной его работе на компьютере: 30 с — FRR = 0,0014, FAR = 0,002; 60 с — FRR < 0,0001, FAR = 0,0002; 150 с — FRR < 0,0005, FAR < < 0,0005;
- способ генерации ключа (пароля) в реальном времени на основе признаков клавиатурного почерка с вероятностями ошибочных решений, зависящими от времени мониторинга действий субъекта при активной его работе на компьютере: 30 с — FRR = 0,058, FAR = 0,0636; 60 с — FRR = 0,034, FAR = 0,043; 150 с — FRR < < 0,0005, FAR < 0,0005;
- способ генерации ключа (пароля) в реальном времени на основе признаков лица и клавиатурного почерка с вероятностями ошибочных решений, зависящими от времени мониторинга действий субъекта при активной его работе на компьютере: 30 с — FRR = = 0,002, FAR = 0,0036; 60 с — FRR = 0,002, FAR = 0,0009; 150 с — FRR < 0,0005, FAR < 0,0005;
- способ генерации ключа (пароля) на основе подписей субъектов с использованием сети многомерных функционалов Байеса, специальных неклассических кодов, исправляющих ошибки, предназначенных для биометрии и принципа защищенного нейросетевого биометрического контейнера с вероятностью ошибок 1-го и 2-го рода 0,0288 и 0,0232 соответственно.

Некоторые из полученных результатов использовались при разработке программного комплекса SignToLogin для аутентификации пользователей компьютерных систем по подписи [125].

ПРИЛОЖЕНИЕ А
ГРАФИКИ ВЕРОЯТНОСТЕЙ ОШИБОЧНЫХ РЕШЕНИЙ ПРИ ГЕНЕРАЦИИ КЛЮЧЕВЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ НА ОСНОВЕ ДАННЫХ НЕПРЕРЫВНОГО МОНИТОРИНГА

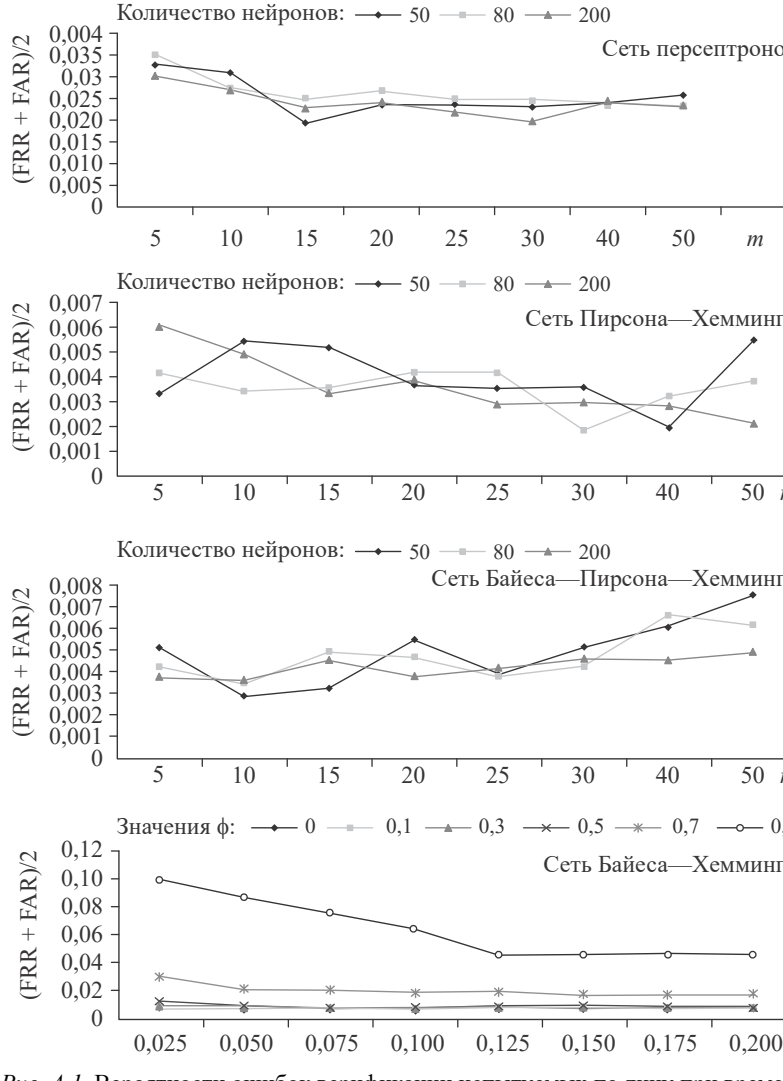

Рис. А.1. Вероятности ошибок верификации испытуемых по лицу при времени мониторинга стандартного оборудования 30 с (при пороге $H = 0$).

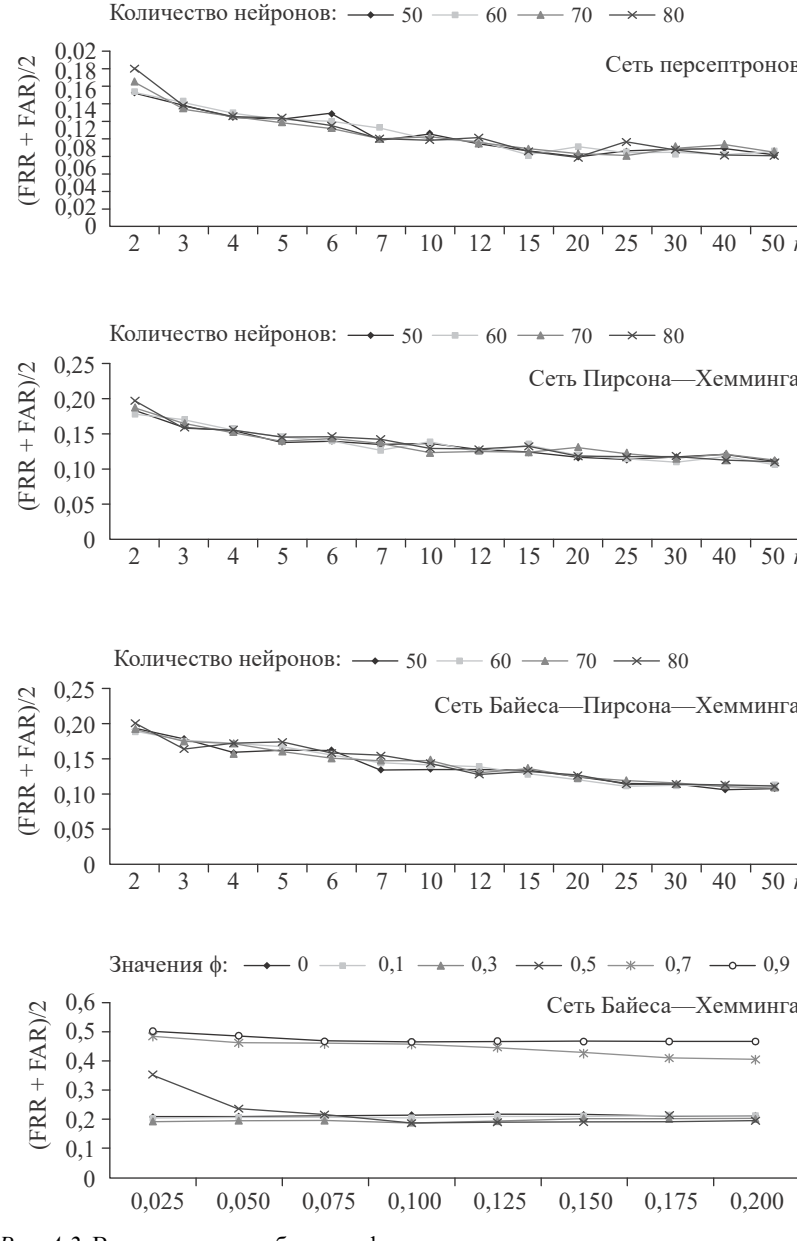

Рис. А.2. Вероятности ошибок верификации испытуемых по клавиатурному почерку при времени мониторинга стандартного оборудования 30 с (при пороге $H = 0$).

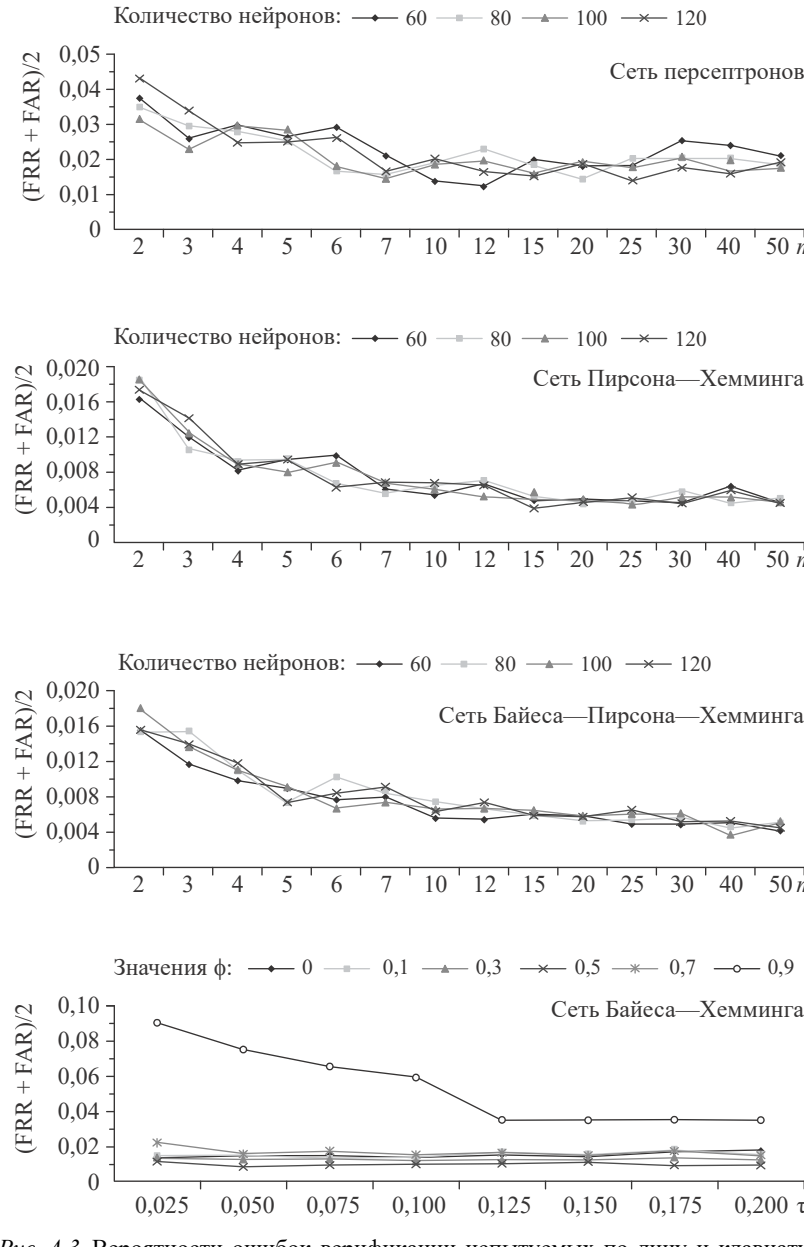

Рис. А.3. Вероятности ошибок верификации испытуемых по лицу и клавиатурному почерку при времени мониторинга стандартного оборудования 30 с (при пороге $H = 0$).

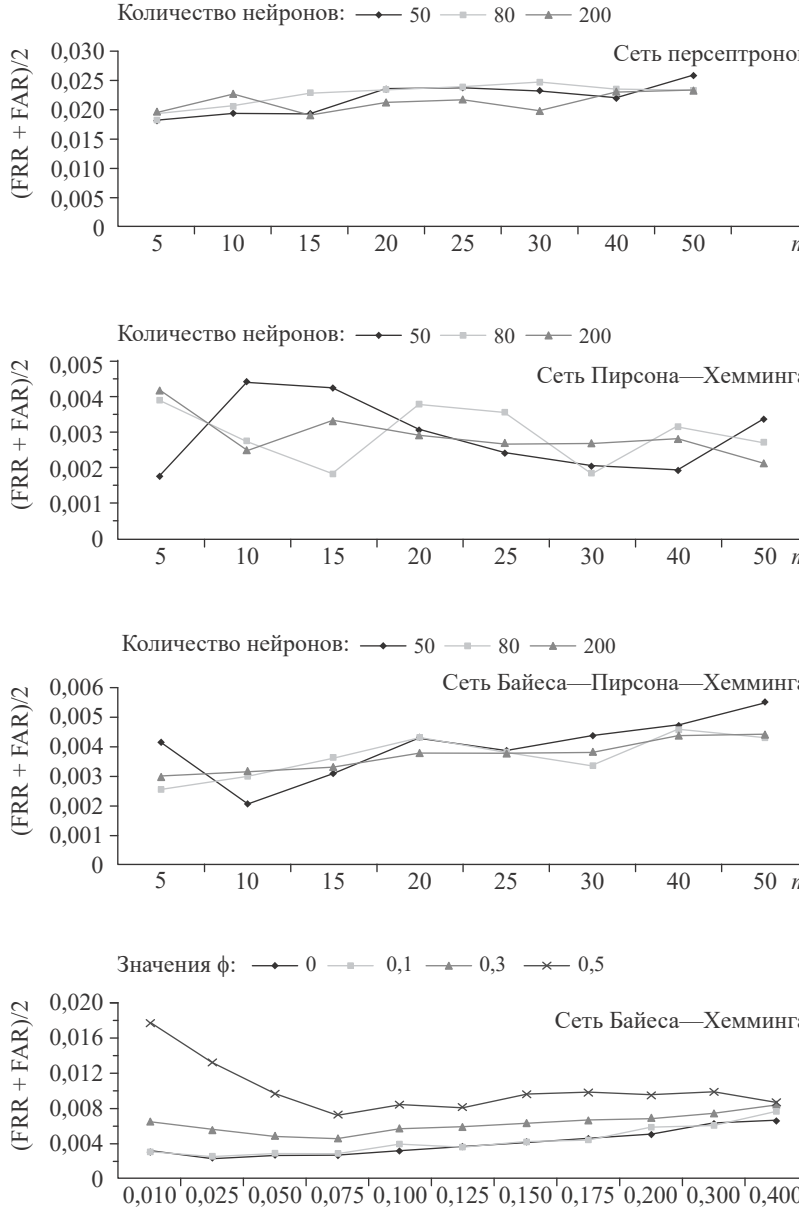

Рис. А.4. Вероятности ошибок верификации испытуемых по лицу при времени мониторинга стандартного оборудования 30 с (при пороге $H > 0$).

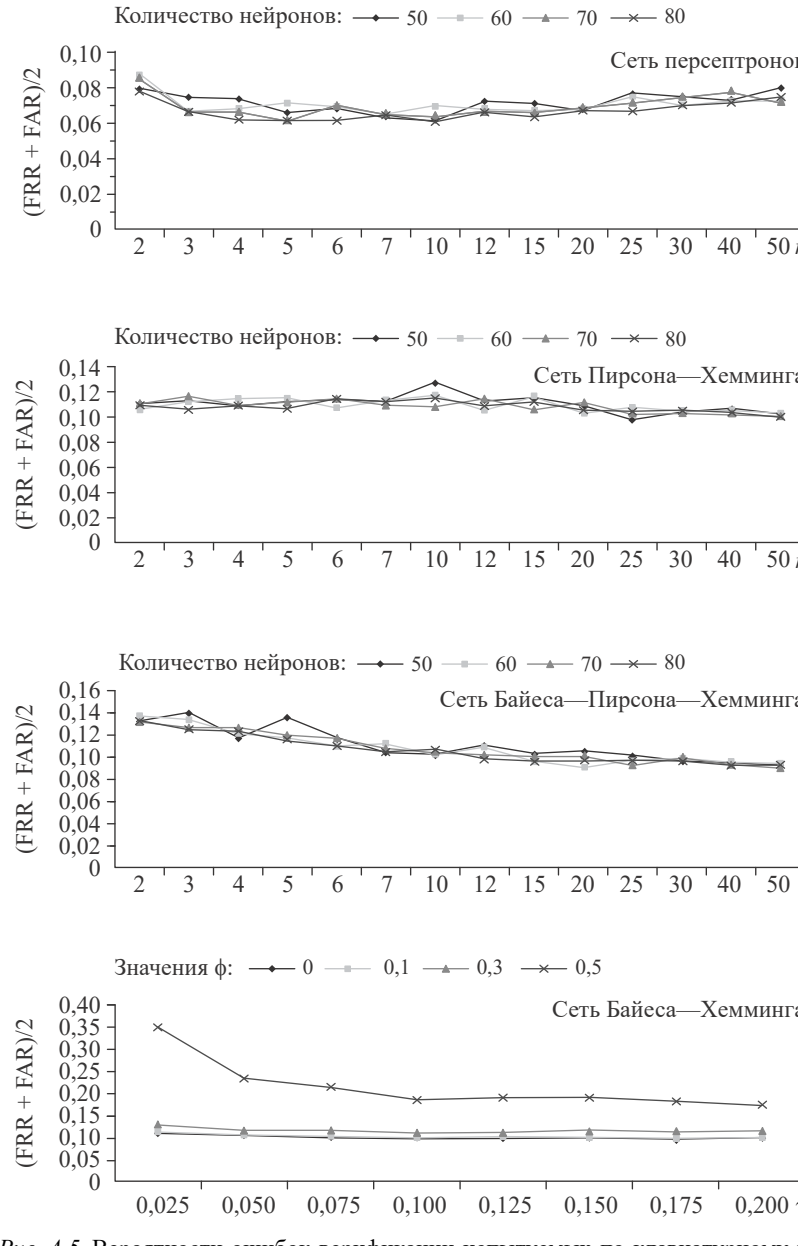

Рис. А.5. Вероятности ошибок верификации испытуемых по клавиатурному почерку при времени мониторинга стандартного оборудования 30 с (при пороге $H > 0$).

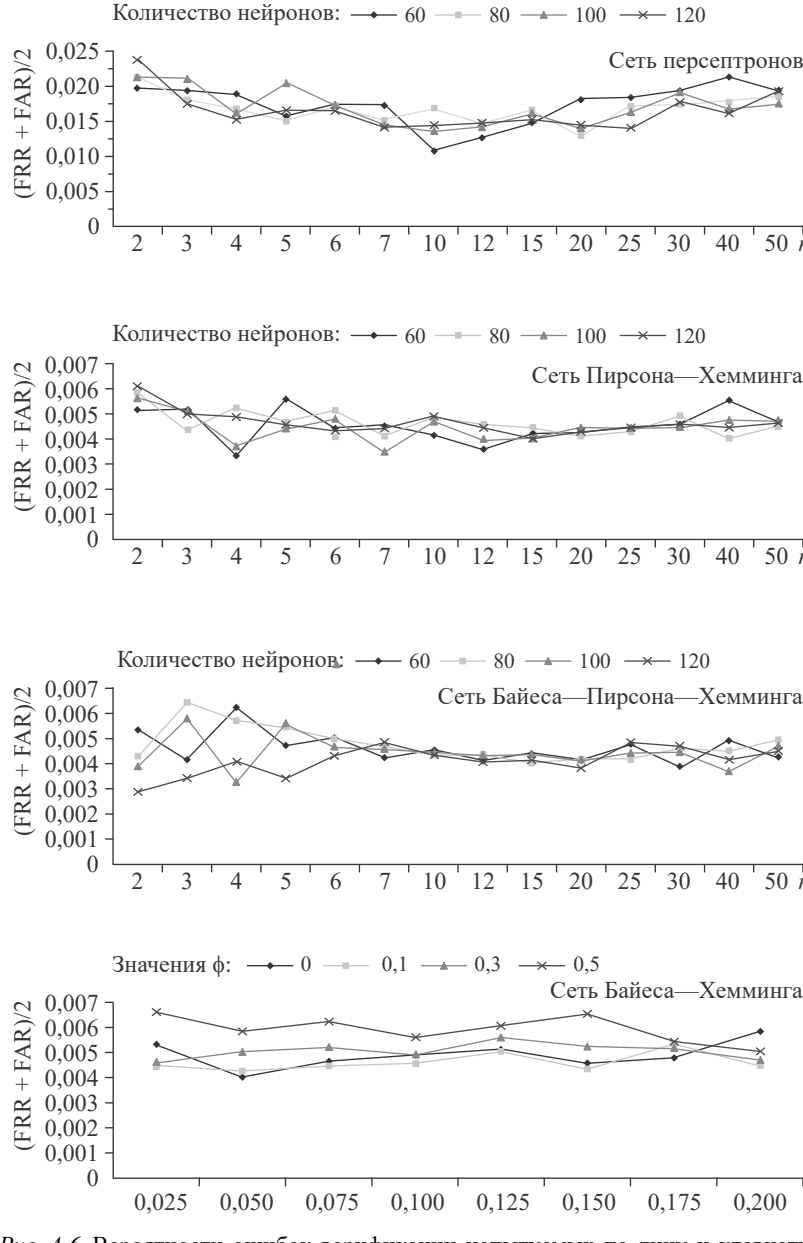

Рис. А.6. Вероятности ошибок верификации испытуемых по лицу и клавиатурному почерку при времени мониторинга стандартного оборудования 30 с (при пороге $H > 0$).

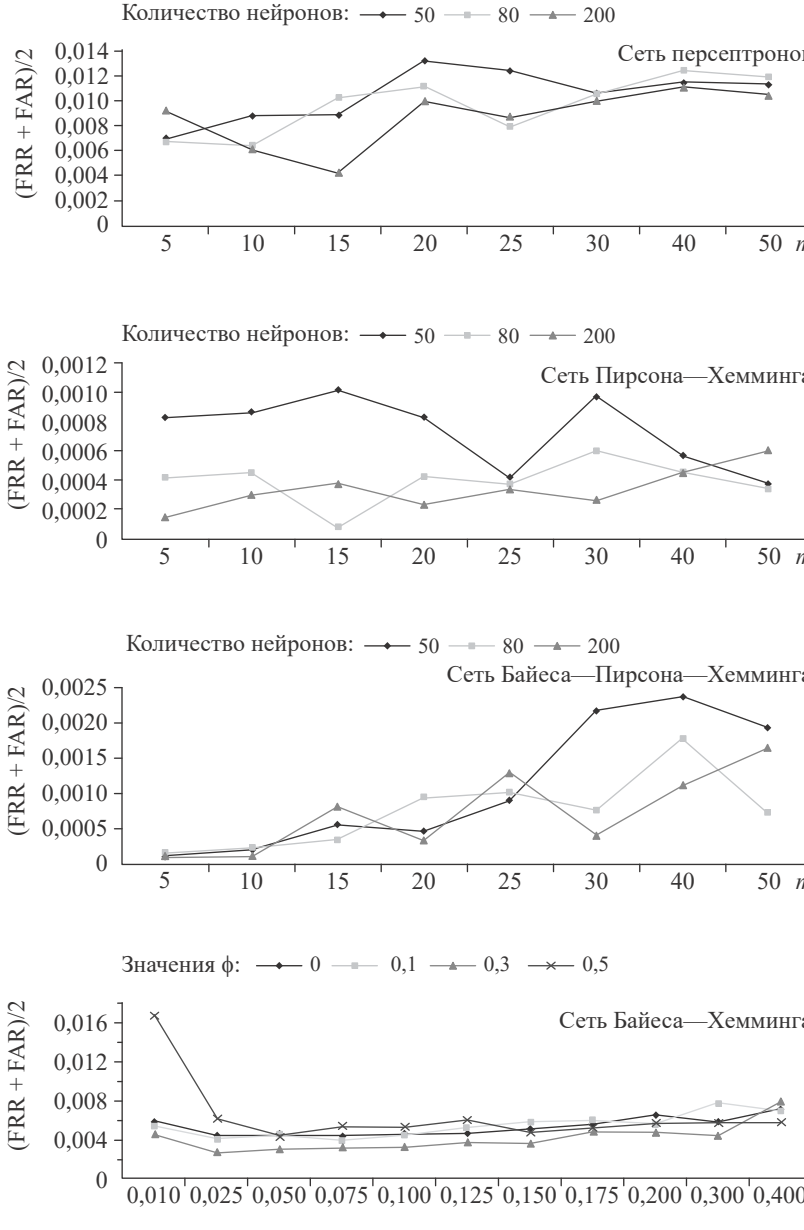

Рис. А.7. Вероятности ошибок верификации испытуемых по лицу при времени мониторинга стандартного оборудования 60 с (при пороге $H > 0$).

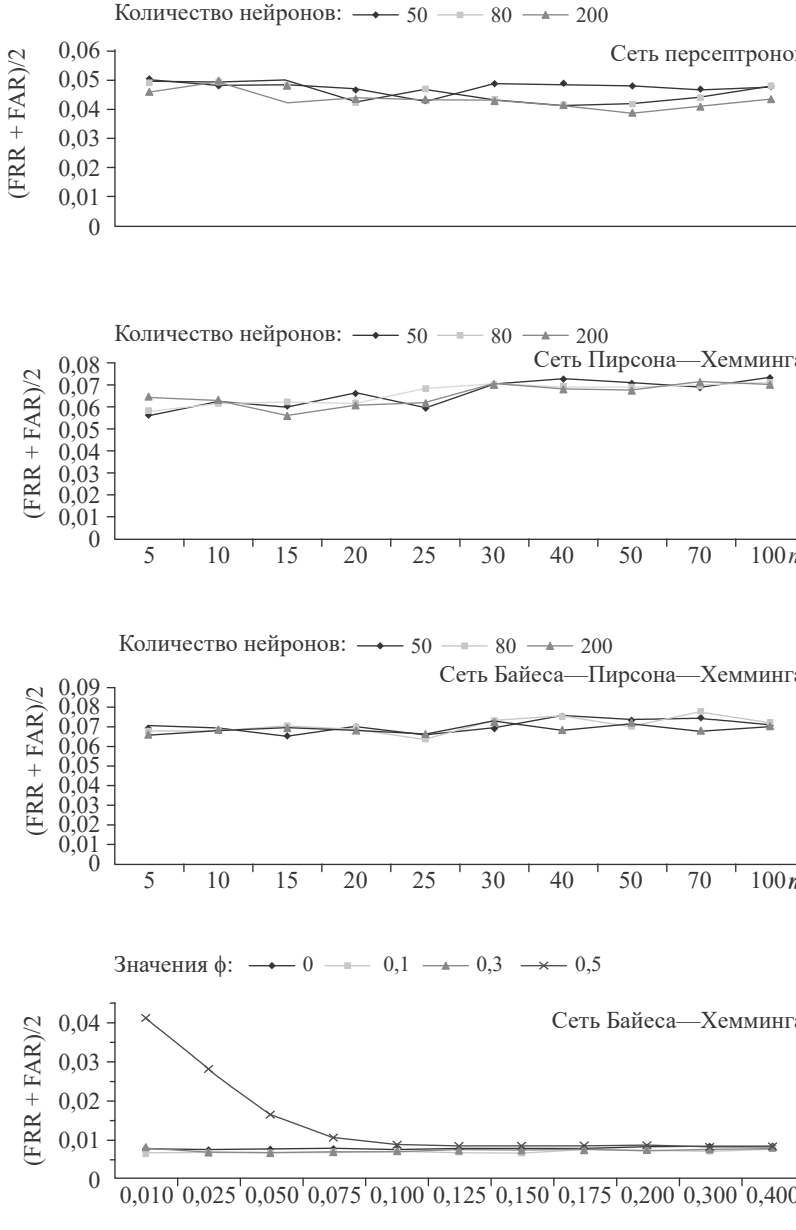

Рис. А.8. Вероятности ошибок верификации испытуемых по клавиатурному почерку при времени мониторинга стандартного оборудования 60 с (при пороге $H > 0$).

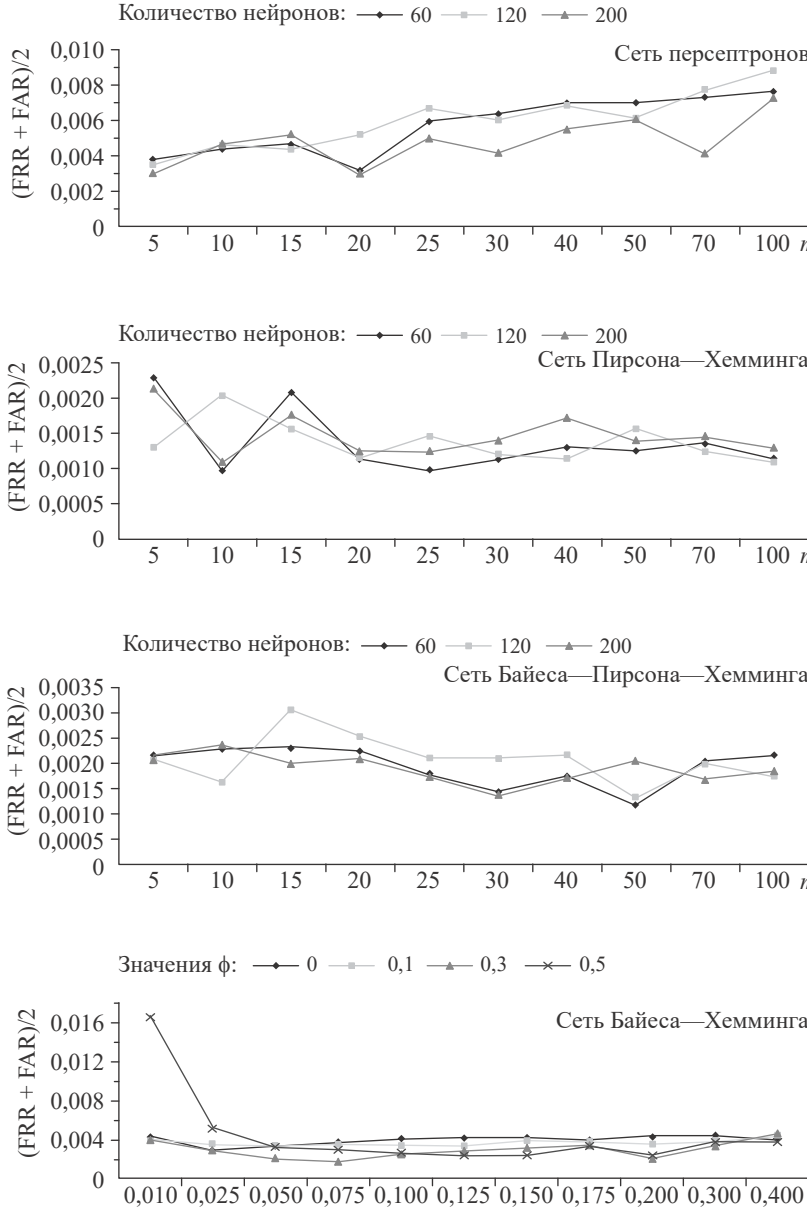

Рис. А.9. Вероятности ошибок верификации испытуемых по лицу и клавиатурному почерку при времени мониторинга стандартного оборудования 60 с (при пороге $H > 0$).

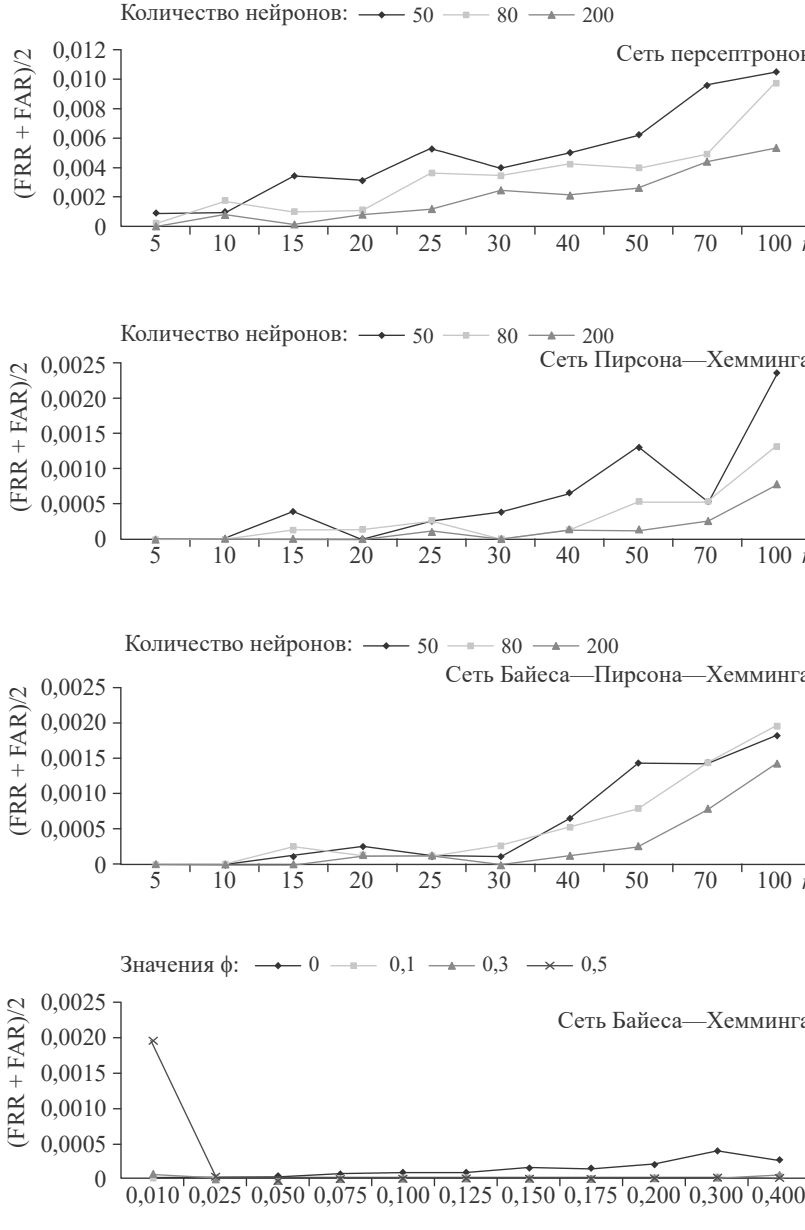

Рис. А.10. Вероятности ошибок верификации испытуемых по клавиатурному почерку при времени мониторинга стандартного оборудования 150 с (при пороге $H > 0$).

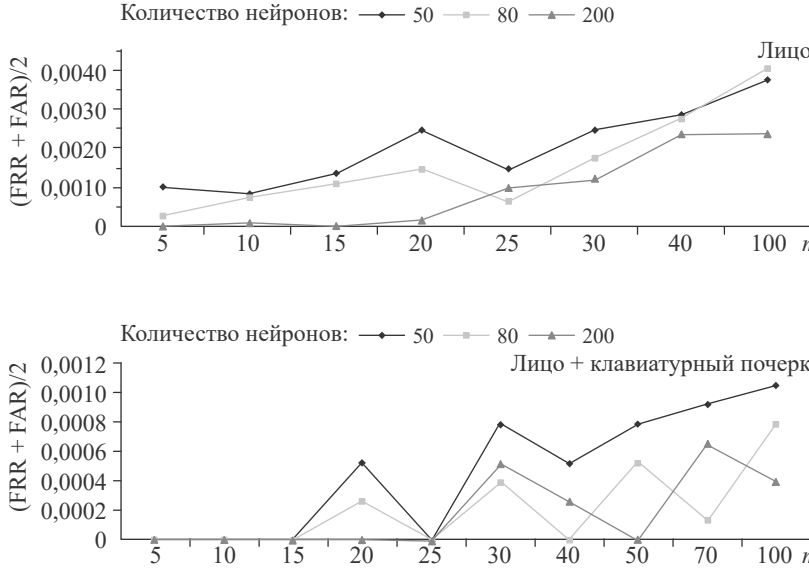

Рис. А.11. Вероятности ошибок верификации испытуемых сетями персептронов при времени мониторинга стандартного оборудования 150 с (при пороге $H > 0$).

ПРИЛОЖЕНИЕ Б
МЕЖДУНАРОДНЫЕ БИОМЕТРИЧЕСКИЕ СТАНДАРТЫ ИСО/МЭК СТК 1/ПК 37 (ISO/IEC JTC1 SC37), ДЕЙСТВУЮЩИЕ НА ТЕРРИТОРИИ РФ, ЗАКРЕПЛЕННЫЕ ЗА ТК 098

Обозначение и название стандарта на английском языке	Обозначение и название стандарта на русском языке
ISO/IEC 2382-37:2012 Information technology — Vocabulary — Part 37: Biometrics	ГОСТ ISO/IEC 2382-37-2016 Информационные технологии. Словарь. Часть 37. Биометрия
ISO/IEC 19794-1:2006 Information technology — Biometric data interchange formats — Part 1: Framework	ГОСТ ISO/IEC 19794-1-2015 Информационные технологии. Биометрия. Форматы обмена биометрическими данными. Часть 1. Структура
ISO/IEC 19794-2:2005 Information technology — Biometric data interchange formats — Part 2: Finger minutiae data	ГОСТ Р ИСО/МЭК 19794-2-2013 Информационные технологии. Биометрия. Форматы обмена биометрическими данными. Часть 2. Данные изображения отпечатка пальца — контрольные точки
ISO/IEC 19794-3:2006 Information technology — Biometric data interchange formats — Part 3: Finger pattern spectral data	ГОСТ Р ИСО/МЭК 19794-3-2009 Автоматическая идентификация. Идентификация биометрическая. Форматы обмена биометрическими данными. Часть 3. Спектральные данные изображения отпечатка пальца
ISO/IEC 19794-4:2005 Information technology — Biometric data interchange formats — Part 4: Finger image data	ГОСТ Р ИСО/МЭК 19794-4-2014 Информационные технологии. Биометрия. Форматы обмена биометрическими данными. Часть 4. Данные изображения отпечатка пальца
ISO/IEC 19794-5:2005 Information technology — Biometric data interchange formats — Part 5: Face image data	ГОСТ Р ИСО/МЭК 19794-5-2013 Информационные технологии. Биометрия. Форматы обмена биометрическими данными. Часть 5. Данные изображения лица
ISO/IEC 19794-6:2005 Information technology — Biometric data interchange formats — Part 6: Iris image data	ГОСТ Р ИСО/МЭК 19794-6-2014 Информационные технологии. Биометрия. Форматы обмена биометрическими данными. Часть 6. Данные изображения радужной оболочки глаза

Обозначение и название стандарта на английском языке	Обозначение и название стандарта на русском языке
ISO/IEC 19794-7:2007 Information technology — Biometric data interchange formats — Part 7: Signature/sign time series data	ГОСТ Р ИСО/МЭК 19794-7-2009 Автоматическая идентификация. Идентификация биометрическая. Форматы обмена биометрическими данными. Часть 7. Данные динамики подписи
ISO/IEC 19794-8:2006 Information technology — Biometric data interchange formats — Part 8: Finger pattern skeletal data	ГОСТ Р ИСО/МЭК 19794-8-2015 Информационные технологии. Биометрия. Форматы обмена биометрическими данными. Часть 8. Данные изображения отпечатка пальца — остов
ISO/IEC 19794-9:2007 Information technology — Biometric data interchange formats — Part 9: Vascular image data	ГОСТ Р ИСО/МЭК 19794-9-2015 Информационные технологии. Биометрия. Форматы обмена биометрическими данными. Часть 9. Данные изображения сосудистого русла
ISO/IEC 19794-10:2007 Information technology — Biometric data interchange formats — Part 10: Hand geometry silhouette data	ГОСТ Р ИСО/МЭК 19794-10-2010 Автоматическая идентификация. Идентификация биометрическая. Форматы обмена биометрическими данными. Часть 10. Данные геометрии контура кисти руки
ISO/IEC 19794-11:2013 Information technology — Biometric data interchange formats — Part 11: Signature/sign processed dynamic data	ГОСТ Р ИСО/МЭК 19794-11-2015 Информационные технологии. Биометрия. Форматы обмена биометрическими данными. Часть 11. Обрабатываемые данные динамики подписи
ISO/IEC 19795-1:2006 Information technology — Biometric performance testing and reporting — Part 1: Principles and framework	ГОСТ Р ИСО/МЭК 19795-1-2007 Автоматическая идентификация. Идентификация биометрическая. Эксплуатационные испытания и протоколы испытаний в биометрии. Часть 1. Принципы и структура
ISO/IEC 19795-2:2007 Information technology — Biometric performance testing and reporting — Part 2: Testing methodologies for technology and scenario evaluation	ГОСТ Р ИСО/МЭК 19795-2-2008 Автоматическая идентификация. Идентификация биометрическая. Эксплуатационные испытания и протоколы испытаний в биометрии. Часть 2. Методы проведения технологического и сценарного испытаний
ISO/IEC TR 19795-3:2007 Information technology — Biometric performance testing and reporting — Part 3: Modality-specific testing	ГОСТ Р ИСО/МЭК ТО 19795-3-2009 Автоматическая идентификация. Идентификация биометрическая. Эксплуатационные испытания и протоколы испытаний в био-

Обозначение и название стандарта на английском языке	Обозначение и название стандарта на русском языке
	метрии. Часть 3. Особенности проведения испытаний при различных биометрических модальностях
ISO/IEC 19795-4:2008 Information technology — Biometric performance testing and reporting — Part 4: Interoperability performance testing	ГОСТ Р ИСО/МЭК 19795-4-2011 Информационные технологии. Биометрия. Эксплуатационные испытания и протоколы испытаний в биометрии. Часть 4. Испытания на совместимость
ISO/IEC 19795-6:2012 Information technology — Biometric performance testing and reporting — Part 6: Testing methodologies for operational evaluation	ГОСТ Р ИСО/МЭК 19795-6-2015 Информационные технологии. Биометрия. Эксплуатационные испытания и протоколы испытаний в биометрии. Часть 6. Методология проведения оперативных испытаний
ISO/IEC 19784-1:2006 Information technology — Biometric application programming interface — Part 1: BioAPI specification	ГОСТ Р ИСО/МЭК 19784-1-2007 Автоматическая идентификация. Идентификация биометрическая. Биометрический программный интерфейс. Часть 1. Спецификация биометрического программного интерфейса
ISO/IEC 19784-2:2007 Information technology — Biometric application programming interface — Part 2: Biometric archive function provider interface	ГОСТ Р ИСО/МЭК 19784-2-2010 Автоматическая идентификация. Идентификация биометрическая. Биометрический программный интерфейс. Часть 2. Интерфейс поставщика биометрической функции архива
ISO/IEC 19784-4:2011 Information technology — Biometric application programming interface — Part 4: Biometric sensor function provider interface	ГОСТ Р ИСО/МЭК 19784-4-2014 Информационные технологии. Биометрия. Биометрический программный интерфейс. Часть 4. Интерфейс поставщика функции биометрического датчика
ISO/IEC 19785-1:2015 Information technology — Common Biometric Exchange Formats Framework — Part 1: Data element specification	ГОСТ Р ИСО/МЭК 19785-1-2008 Автоматическая идентификация. Идентификация биометрическая. Единая структура форматов обмена биометрическими данными. Часть 1. Спецификация элементов данных
ISO/IEC 19785-2:2006 Information technology — Common Biometric Exchange Formats Framework — Part 2: Procedures for the operation of the Biometric Registration Authority	ГОСТ Р ИСО/МЭК 19785-2-2008 Автоматическая идентификация. Идентификация биометрическая. Единая структура форматов обмена биометрическими данными. Часть 2. Процедуры действий регистрационного органа в области биометрии

Обозначение и название стандарта на английском языке	Обозначение и название стандарта на русском языке
ISO/IEC 19785-4:2010 Information technology — Common Biometric Exchange Formats Framework — Part 4: Security block format specifications	ГОСТ Р ИСО/МЭК 19785-4-2012 Информационные технологии. Биометрия. Единая структура форматов обмена биометрическими данными. Часть 4. Спецификация формата блока защиты информации
ISO/IEC 24708:2008 Information technology — Biometrics — BioAPI Interworking Protocol	ГОСТ Р ИСО/МЭК 24708-2013 Информационные технологии. Биометрия. Протокол межсетевого обмена БиоАПИ
ISO/IEC 24709-1:2007 Information technology — Conformance testing for the biometric application programming interface (BioAPI) — Part 1: Methods and procedures	ГОСТ Р ИСО/МЭК 24709-1-2009 Автоматическая идентификация. Идентификация биометрическая. Испытания на соответствие биометрическому программному интерфейсу (БиоАПИ). Часть 1. Методы и процедуры
ISO/IEC 24709-2:2007 Information technology — Conformance testing for the biometric application programming interface (BioAPI) — Part 2: Test assertions for biometric service providers	ГОСТ Р ИСО/МЭК 24709-2-2011 Информационные технологии. Биометрия. Испытания на соответствие биометрическому программному интерфейсу (БиоАПИ). Часть 2. Тестовые утверждения для поставщиков биометрических услуг
ISO/IEC 24709-3:2011 Information technology — Conformance testing for the biometric application programming interface (BioAPI) — Part 3: Test assertions for BioAPI frameworks	ГОСТ Р ИСО/МЭК 24709-3-2013 Информационные технологии. Биометрия. Испытания на соответствие биометрическому программному интерфейсу (БиоАПИ). Часть 3. Тестовые утверждения для инфраструктур БиоАПИ
ISO/IEC 24713-1:2008 Information technology — Biometric profiles for interoperability and data interchange — Part 1: Overview of biometric systems and biometric profiles	ГОСТ ISO/IEC 24713-1-2013 Информационные технологии. Биометрические профили для взаимодействия и обмена данными. Часть 1. Общая архитектура биометрической системы и биометрические профили
ISO/IEC 24713-2:2008 Information technology — Biometric profiles for interoperability and data interchange — Part 2: Physical access control for employees at airports	ГОСТ Р ИСО/МЭК 24713-2-2011 Информационные технологии. Биометрия. Биометрические профили для взаимодействия и обмена данными. Часть 2. Физический контроль доступа сотрудников аэропорта
ISO/IEC 24713-3:2009 Information technology — Biometric profiles for interoperability and	ГОСТ Р ИСО/МЭК 24713-3-2016 Информационные технологии. Биометрия. Биометрические профили для взаимодей-

Обозначение и название стандарта на английском языке	Обозначение и название стандарта на русском языке
data interchange — Part 3: Biometrics-based verification and identification of seafarers	ствия и обмена данными. Часть 3. Биометрическая верификация и идентификация моряков
ISO/IEC 29109-1:2009 Information technology — Biometrics — Conformance testing methodology for biometric data interchange formats defined in ISO/IEC 19794 — Part 1: Generalized conformance testing methodology	ГОСТ Р ИСО/МЭК 29109-1-2012 Информационные технологии. Биометрия. Методология испытаний на соответствие форматам обмена биометрическими данными, определенным в комплексе стандартов ИСО/МЭК 19794. Часть 1. Обобщенная методология испытаний на соответствие
ISO/IEC 29109-4:2010 Information technology — Conformance testing methodology for biometric data interchange formats defined in ISO/IEC 19794 — Part 4: Finger image data	ГОСТ Р ИСО/МЭК 29109-4-2015 Информационные технологии. Биометрия. Методология испытаний на соответствие форматам обмена биометрическими данными, определенным в комплексе стандартов ИСО/МЭК 19794. Часть 4. Данные изображения отпечатка пальца
ISO/IEC 29109-5:2014 Information technology — Conformance testing methodology for biometric data interchange formats defined in ISO/IEC 19794 — Part 5: Face image data	ГОСТ Р ИСО/МЭК 29109-5-2013 Информационные технологии. Биометрия. Методология испытаний на соответствие форматам обмена биометрическими данными, определенным в комплексе стандартов ИСО/МЭК 19794. Часть 5. Данные изображения лица
ISO/IEC 29109-6:2011 Information technology — Conformance testing methodology for biometric data interchange formats defined in ISO/IEC 19794 — Part 6: Iris image data	ГОСТ Р ИСО/МЭК 29109-6-2016 Информационные технологии. Биометрия. Методология испытаний на соответствие форматам обмена биометрическими данными, определенным в комплексе стандартов ИСО/МЭК 19794. Часть 6. Данные изображения радужной оболочки глаза
ISO/IEC 29109-7:2011 Information technology — Conformance testing methodology for biometric data interchange formats defined in ISO/IEC 19794 — Part 7: Signature/sign time series data	ГОСТ Р ИСО/МЭК 29109-7-2016 Информационные технологии. Биометрия. Методология испытаний на соответствие форматам обмена биометрическими данными, определенным в комплексе стандартов ИСО/МЭК 19794. Часть 7. Данные динамики подписи
ISO/IEC 29109-8:2011 Information technology — Conformance testing methodology for bio-	ГОСТ Р ИСО/МЭК 29109-8-2016 Информационные технологии. Биометрия. Методология испытаний на соответствие

Обозначение и название стандарта на английском языке	Обозначение и название стандарта на русском языке
metric data interchange formats defined in ISO/IEC 19794 — Part 8: Finger pattern skeletal data	форматам обмена биометрическими данными, определенным в комплексе стандартов ИСО/МЭК 19794. Часть 8. Данные изображения отпечатка пальца — остов
ISO/IEC 29794-1:2016 Information technology — Biometric sample quality — Part 1: Framework	ГОСТ Р ИСО/МЭК 29794-1-2012 Информационные технологии. Биометрия. Качество биометрических образцов. Часть 1. Структура
ISO/IEC 29141:2009 Information technology — Biometrics — Tenprint capture using biometric application programming interface (BioAPI)	ГОСТ Р ИСО/МЭК 29141-2012 Информационные технологии. Биометрия. Одновременное получение изображений отпечатков десяти пальцев с помощью БиоАПИ
ISO/IEC TR 24741:2007 Information technology — Biometrics tutorial	ГОСТ Р 54412-2011/ISO/IEC TR 24741:2007 Информационные технологии. Биометрия. Обучающая программа по биометрии
ISO/IEC TR 24722:2015 Information technology — Biometrics — Multimodal and other multibiometric fusion	ГОСТ Р 54411-2011/ISO/IEC TR 24722:2007 Информационные технологии. Биометрия. Мультимодальные и другие мультибиометрические технологии
ISO/IEC 7816-11:2004 Identification cards — Integrated circuit cards — Part 11: Personal verification through biometric methods	ГОСТ Р ИСО/МЭК 7816-11-2013 Карты идентификационные. Карты на интегральных схемах. Часть 11. Верификация личности биометрическими методами
ISO/IEC 11694-6:2014 Identification cards — Optical memory cards — Linear recording method — Part 6: Use of biometrics on an optical memory card	ГОСТ Р ИСО/МЭК 11694-6-2011 Карты идентификационные. Карты с оптической памятью. Метод линейной записи данных. Часть 6. Использование биометрических данных на картах с оптической памятью

ПРИЛОЖЕНИЕ В
НАЦИОНАЛЬНЫЕ СТАНДАРТЫ НЕЙРОСЕТЕВОЙ БИОМЕТРИИ, ЗАКРЕПЛЕННЫЕ ЗА ТК 362 (СЕМЕЙСТВО ГОСТ Р 52633)

Обозначение	Название стандарта
ГОСТ Р 52633.0-2006	Защита информации. Техника защиты информации. Требования к средствам высоконадежной биометрической аутентификации
ГОСТ Р 52633.1-2009	Защита информации. Техника защиты информации. Требования к формированию баз естественных биометрических образов, предназначенных для тестирования средств высоконадежной биометрической аутентификации
ГОСТ Р 52633.2-2010	Защита информации. Техника защиты информации. Требования к формированию синтетических биометрических образов, предназначенных для тестирования средств высоконадежной биометрической аутентификации
ГОСТ Р 52633.3-2011	Защита информации. Техника защиты информации. Тестирование стойкости средств высоконадежной биометрической защиты к атакам подбора
ГОСТ Р 52633.4-2012	Защита информации. Техника защиты информации. Интерфейсы взаимодействия с нейросетевыми преобразователями биометрия—код
ГОСТ Р 52633.5-2011	Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия—код доступа
ГОСТ Р 52633.6-2013	Защита информации. Техника защиты информации. Требования к индикации близости предъявленных биометрических данных образу «Свой»
ГОСТ Р 52633.7-2017	Защита информации. Техника защиты информации. Высоконадежная мультибиометрическая аутентификация
ГОСТ Р 52633.6-2013	Защита информации. Техника защиты информации. Требования к индикации близости предъявленных биометрических данных образу «Свой»

ПРИЛОЖЕНИЕ Г

НАЦИОНАЛЬНЫЕ КРИПТОГРАФИЧЕСКИЕ СТАНДАРТЫ, КОТОРЫЕ ДОЛЖНЫ ИСПОЛЬЗОВАТЬСЯ ПРИ РЕАЛИЗАЦИИ БИОМЕТРИЧЕСКОЙ ЗАЩИТЫ, ЗАКРЕПЛЕННЫЕ ЗА ТК 026

Обозначение	Название стандарта
ГОСТ Р 34.10-2012	Информационная технология. Криптографическая защита информации. Процессы формирования и проверки электронной цифровой подписи
ГОСТ Р 34.11-2012	Информационная технология. Криптографическая защита информации. Функция хэширования
ГОСТ Р 34.12-2015	Информационная технология. Криптографическая защита информации. Блочные шифры
ГОСТ Р 34.13-2015	Информационная технология. Криптографическая защита информации. Режимы работы блочных шифров
	Системы обработки информации. Защита криптографическая. Техническая спецификация (проект). Защита нейросетевых биометрических контейнеров с использованием криптографических алгоритмов

ПРИЛОЖЕНИЕ Д
МЕЖДУНАРОДНЫЕ СТАНДАРТЫ ПО ЗАЩИТЕ БИОМЕТРИЧЕСКИХ ДАННЫХ ISO/IEC JTC1 SC27

Обозначение	Название стандарта
ISO/IEC 24745:2011	Information technology — Security techniques — Biometric information protection
ISO/IEC 24761:2009	Information technology — Security techniques — Authentication context for biometrics
ISO/IEC 19792:2009	Information technology. Security techniques. Security evaluation of biometrics

СПИСОК ЛИТЕРАТУРЫ

1. *Иванов А. И., Ложников П. С., Самотуга А. Е.* Технология формирования гибридных документов // Кибернетика и системный анализ. 2014. Т. 50, № 6. С. 152—156.
2. *Ложников П. С., Самотуга А. Е.* Технология проверки целостности и аутентичности документов в гибридном документообороте // Изв. ТулГУ. Технические науки. Тула: Изд-во ТулГУ, 2013. Вып. 3. С. 402—408.
3. *Ложников П. С., Самотуга А. Е.* Способ формирования гибридных документов с использованием биометрической подписи // Электронные средства и системы управления: Матер. докл. X Международной научно-практической конференции (12—14 ноября 2014 г.): в 2 ч. Ч. 2. Томск: В-Спектр, 2014. С. 79—83.
4. *Ложников П. С., Иванов А. И.* Способ формирования электронного документа и его копий. Патент на изобретение № 2543928 от 03.02.2015. М.: Федеральная служба по интеллектуальной собственности, патентам и товарным знакам.
5. *Управление* киберрисками во взаимосвязанном мире. Основные результаты Глобального исследования по вопросам обеспечения информационной безопасности. Перспективы на 2015 год. PriceWaterHouseCoopers. (Режим доступа: http://www.pwc.ru/ru_RU/ru/ riskassurance/publications/assets/managing-cyberrisks.pdf, дата обращения: 22.08.2016).
6. *The Global* State of Information Security® Survey 2016. PricewaterhouseCoopers. (Режим доступа: http://www.pwc.com/gx/en/issues/ cyber-security/information-security-survey/download.html, дата обращения: 27.06.2016).
7. *Утечки* конфиденциальной информации в России и в мире. Итоги 2016 года. Zecurion Analytics. (Режим доступа: http://www.zecurion.ru/upload/iblock/1e5/Zecurion_Data_Leaks_2016_full.pdf, дата обращения: 06.07.2016).
8. *Утечки* данных. Банки. Мир, Россия. 2015 год. InfoWatch. (Режим доступа: https://www.infowatch.ru/analytics/reports, дата обращения: 06.07.2016).
9. *Еременко А. В., Левитская Е. А., Сулавко А. Е., Самотуга А. Е.* Разграничение доступа к информации на основе скрытого мониторинга действий пользователей в информационных системах: непрерывная идентификация // Вестник СибАДИ. Омск: СибАДИ, 2014. № 6(40). С. 92—102.
10. *Biometrics* for Mobile Devices: Global Market Analysis and Forecasts. Research and Markets report, February 2015 (Режим доступа: http://www.researchandmarkets.com/research/chx7q6/biometrics_for., дата обращения: 10.06.2015).

11. *Moving* forward with cybersecurity and privacy. (Режим доступа: http://www.pwc.ru/ru/riskassurance/publications/assets/gsiss-report_2017_eng.pdf, дата обращения: 11.12.2016).
12. *Алиев А. Т.* Разработка моделей, методов и алгоритмов перспективных средств защиты информации в системах электронного документооборота на базе современных технологий скрытой связи: Автореф. дисс. … канд. техн. наук. Ростов-на-Дону, 2008. 20 с.
13. *Шелепина Е. А.* Правовые аспекты электронного документооборота: Автореф. дисс. … канд. юр. наук. М., 2007. 29 с.
14. *Шишаева Е. Ю.* Правовое регулирование использования электронного документа в предпринимательской деятельности: Автореф. дис. … канд. юр. наук. М., 2005. 26 с.
15. *ГОСТ* Р ИСО 15489-1-2007. Национальный стандарт Российской Федерации. Система стандартов по информации, библиотечному и издательскому делу. Управление документами. Общие требования. М.: Стандартинформ, 2007. 23 с.
16. *ГОСТ* Р 7.0.8-2013. Система стандартов по информации, библиотечному и издательскому делу. Делопроизводство и архивное дело. Термины и определения. М.: Стандартинформ, 2014. 16 с.
17. *Федеральный* закон от 27.07.2006 № 149-ФЗ. Об информации, информационных технологиях и о защите информации. 2006. 20 с.
18. *Конявский В. А.* Основы понимания феномена электронного обмена информацией / Гадасин В. А. Минск: Беллитфонд, 2004. 282 с.
19. *Конявский В. А.* Методы и аппаратные средства защиты информационных технологий электронного документооборота: Автореф. дисс. … канд. техн. наук. М., 2005. 47 с.
20. *Федеральный* закон от 06.04.2011 № 63-Ф. Об электронной подписи. 2011. 11 с.
21. *ГОСТ* Р 52633.0-2006. Защита информации. Техника защиты информации. Требования к средствам высоконадежной биометрической аутентификации. М.: Стандартинформ, 2006. 24 с.
22. *ГОСТ* Р 52633.5-2011. Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия—код доступа. М.: Стандартинформ, 2011. 20 с.
23. *Dodis Y., Reyzin L., Smith A.* Fuzzy extractors: How to generate strong keys from biometrics and other noisy // EUROCRYPT. 2004. April. P. 523—540.
24. *Иванов А. И.* Многомерная нейросетевая обработка биометрических данных с программным воспроизведением эффектов квантовой суперпозиции. Пенза: ПНИЭИ, 2016. 133 с.
25. *Juels A., Sudan M.* A Fuzzy vault scheme // Designs, Codes and Cryptography. 2006. Vol. 38, Iss. 2. P. 237—257. doi: 10.1007/s10623-005-6343-z.
26. *Juels A., Wattenberg M.* A Fuzzy commitment scheme // Proc. ACM Conf. Computer and Communications Security. 1999. P. 28—36.
27. *Гадасин В. А., Конявский В. А.* От документа — к электронному документу. Системные основы. М.: РФК-Имидж Лаб, 2001. 192 с.
28. *Иванов А. И.* Нейросетевые алгоритмы биометрической идентификации личности / Под ред. А. И. Галушкина. М.: Радиотехника, 2004. 144 с. (Научная серия «Нейрокомпьютеры и их применение»; № 15).

29. *Сулавко А. Е., Еременко А. В., Левитская Е. А.* Разграничение доступа к информации на основе скрытого мониторинга действий пользователей в информационных системах: портрет нелояльного сотрудника // Изв. Транссиба. 2015. № 1(21). С. 80—89.
30. *Епифанцев Б. Н., Ложников П. С., Сулавко А. Е.* Альтернативные сценарии авторизации при идентификации пользователей по динамике подсознательных движений // Вопросы защиты информации. 2013. № 2. С. 28—35.
31. *Епифанцев Б. Н., Ложников П. С., Сулавко А. Е.* Алгоритм идентификации гипотез в пространстве малоинформативных признаков на основе последовательного применения формулы Байеса // Межотраслевая информационная служба. 2013. № 2 (163). С. 57—62.
32. *Lozhnikov P. S., Sulavko A. E., Samotuga A. E.* Personal identification and the assessment of the psychophysiological state while writing a signature // Information. 2015. N 6. P. 454—466.
33. *Епифанцев Б. Н., Ложников П. С., Сулавко А. Е., Жумажанова С. С.* Идентификационный потенциал рукописных паролей в процессе их воспроизведения // Автометрия. 2016. № 3. С. 28—36.
34. *ГОСТ* Р ИСО/МЭК 18004-2015. Информационные технологии. Технологии автоматической идентификации и сбора данных. Спецификация символики штрихового кода QR Code.
35. *Santos M. F., Aguilar J. F., Garcia J. O.* Cryptographic key generation using handwritten signature // Proceedings of SPIE. 2006. Vol. 6202. P. 225—231.
36. *Vielhauer C., Steinmetz R., Mayerhöfer A.* Biometric hash based on statistical features of online signatures // ICPR'02: Proceedings of the 16 th International Conference on Pattern Recognition (ICPR'02). 2002. Vol. 1. P. 100—123.
37. *Maiorana E., Campisi P.* Fuzzy commitment for function based signature template protection // IEEE Signal Processing Letters. 2010. N 17. P. 249—252.
38. *Yip K. W., Goh A., Ling D. N. C., Jin A. T. B.* Generation of replaceable cryptographic keys from dynamic handwritten signatures // Proc. ICB, Lecture Notes in Computer Science. 2006. N 3832. P. 509—515.
39. *Hao F., Chan C. W.* Private key generation from on-line handwritten signatures // Information Management & Computer Security. 2002. N 2. Iss. 10. P. 159—164.
40. *Zhan Enqi, Guo Jinxu, Zheng Jianbin, Ma Chan, Wang Linjuan.* On-line handwritten signature verification based on two levels back propagation neural network // Proc. of the 2009 International Symposium on Intelligent Ubiquitous Computing and Education. 2009. P. 202—2005.
41. *Iranmanesh V., Ahmad Sh. M. S., Adnan W. A. W., Yussof S., Arigbabu O. A., Malallah F. L.* Online handwritten signature verification using neural network classifier based on principal component analysis // The Scientific World Journal. 2014. Vol. 2014. P. 1—8. doi: 10.1155/2014/381469
42. *Hao Chang, Huaizhong Bao, Yutao Sun, Sulin Wei.* Online signature verification based on feature combination and classifier fusion // Journal of Information & Computational Science. 2013. Vol. 10, N 6. P. 1613—1621.
43. *Ao M., Li S. Z.* Near infrared face based biometric key binding // Proc. of the 3rd International Conference on Biometrics 2009 (ICB'09) LNCS: 5558. 2009. P. 376—385.

44. *Kelkboom E. J. C., Zhou X., Breebaart J., Veldhuis R. N. S., Busch C.* Multi-algorithm fusion with template protection // Proc. of the 3rd IEEE Int. Conf. on Biometrics: Theory, Applications and Systems (BTAS'09). 2009. P. 1—7.
45. *Mian A.* Unsupervised learning from local features for video-based face recognition // Proc. 2008 IEEE International Conference on Automatic Face and Gesture Recognition. 2008. P. 6.
46. *Park U., Jain A., Ross A.* Face recognition in video: Adaptive fusion of multiple matchers // Proc. 2007 IEEE Conference on Computer Vision and Pattern Recognition. 2007. P. 1.
47. *Liu X., Cheng T.* Video-based face recognition using adaptive hidden Markov models // Proc. 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2003. Vol. 1. P. 340—345.
48. *Васильев В. И., Ложников П. С., Сулавко А. Е., Еременко А. В.* Технологии скрытой биометрической идентификации пользователей компьютерных систем // Вопросы защиты информации. 2015. № 3. С. 37—47.
49. *Харин Е. А., Гончаров С. М., Корнюшин П. Н.* Построение систем биометрической аутентификации с использованием генератора ключевых последовательностей на основе нечетких данных // Матер. 50-й Всерос. межвуз. науч.-техн. конф. Владивосток: ТОВМИ, 2007. С. 112—115.
50. *Bleha S., Obaidat M.* Computer users verification using the perceptron algorithm // IEEE Transactions on Systems, Man and Cybernetics. 1993. N 23(3). P. 900—902.
51. *Monrose F., Reiter M. K., Wetzel R.* Password hardening based on keystroke dynamics // Proc. of Sixth ACM Conference on Computer and Communications Security. CCCS, 1999. P. 73—82.
52. *Еременко А. В., Сулавко А. Е.* Двухфакторная аутентификация пользователей компьютерных систем на удаленном сервере по клавиатурному почерку // Прикладная информатика. 2015. № 6. С. 48—59.
53. *Banerjee S. P., Woodard D. L.* Biometric authentication and identification using keystroke dynamics: A survey // Journal of Pattern Recognition Research. 2012. N 7. P. 116—139.
54. *Antal M., Zsolt Szabó László, László I.* Keystroke dynamics on android platform // 8th International Conference Interdisciplinarity in Engineering, INTER-ENG 2014, 9—10 October 2014. Tirgu Mures, Romania. 2014. P. 1—7.
55. *Jun Chen, Guang Zhu, Jin Yang, Qingshen Jing, Peng Bai, Weiqing Yang, Xuewei Qi, Yuanjie Su, Zhong Lin Wang.* Personalized keystroke dynamics for self-powered human machine interfacing // ACS Nano. 2015. N 9 (1). P. 105—116. doi:10.1021/nn506832w.
56. *Teoh A., Kim J.* Secure biometric template protection in fuzzy commitment scheme // IEICE Electron. Express. 2007. N 4(23). P. 724—730.
57. *Hao F., Anderson R., Daugman J.* Combining cryptography with biometrics effectively // IEEE Transactions on Computers. 2006. N 55(9). P. 1081—1088.
58. *Rice R. S., Jenkins R. F., Nartker T. A.* The Fourth Annual Test of OCR Accuracy. (Режим доступа: http://stephenvrice.com/images/AT-1995.pdf, Дата обращения: 02.09.2012).

59. *Фан Нгок Хоанг.* Алгоритмы обработки и анализа символов вейвлет-преобразованием, методом главных компонент и нейронными сетями: Автореф. дисс. … канд. техн. наук. Томск, 2014. 20 с.
60. *Ахметов Б. С., Иванов А. И., Фунтиков В. А., Безяев А. В., Малыгина Е. А.* Технология использования больших нейронных сетей для преобразования нечетких биометрических данных в код ключа доступа. Алматы: ТОО «Издательство LEM», 2014. 144 с.
61. *Ложников П. С., Сулавко А. Е., Еременко А. В., Волков Д. А.* Экспериментальная оценка надежности верификации подписи сетями квадратичных форм, нечеткими экстракторами и персептронами // Информационно-управляющие системы. 2016. № 5. С. 73—85.
62. *Lozhnikov P. S., Sulavko A. E., Volkov D. A.* Application of noise tolerant code to biometric data to verify the authenticity of transmitting information / Control and Communications (SIBCON), 21—23 May 2015. Omsk, Russia, 2015. P. 1—3.
63. *Еременко А. В.* Повышение надежности идентификации пользователей компьютерных систем по динамике написания паролей: Автореф. дисс. … канд. техн. наук. Омск, 2011. 20 с.
64. *Sinha P.* Face recognition by humans: nineteen results all computer vision researchers should know about // Proc. of the IEEE. 2006. Vol. 94, N 11. P. 1948—1962.
65. *Mishra S., Dubey A.* Face recognition approaches: a survey // International Journal of Computing and Business Research (IJCBR). 2015. N 6. (Режим доступа: http://www.researchmanuscripts.com/January2015/1.pdf, дата обращения 27.10.2017).
66. *Chen Y.-C., Patel V. M., Phillips P. J., Chellappa R.* Dictionary-based face recognition from video // European conference computer vision (ECCV). Vol. 7577. Lecture notes in computer science. Springer, Berlin. P. 766—779. doi: 10.1007/978-3-642-33783-3_55.
67. *Barr J. R., Bowyer K. W., Flynn P. J., Biswas S.* Face recognition from video: a review // International Journal of Pattern Recognition and Artificial Intelligence. 2012. April 20. P. 1—56.
68. *Patil S. A., Deore P. J.* Video-based face recognition: a survey // World Journal of Science and Technology. 2012. N 2(4). P. 136—139.
69. *Heseltine T., Pears N., Austin J.* Three-dimensional face recognition: An eigensurface approach // International Conference on Image Processing (ICIP). 2004. P. 1421—1424.
70. *Medioni G., Choi J., Kuo C.-H., Fidaleo D.* Identifying noncooperative subjects at a distance using face images and inferred three-dimensional face models // IEEE Transactions on Systems, Man, and Cybernetics — Part a: Systems and Humans. 2009. Vol. 39, N 1. P. 12—24.
71. *Songkun Li, Jing Luo, Chunbo Xiu*. An overview of three dimensional face recognition // Journal of Engineering. 2013. Vol. 1, N 4. P. 57—62.
72. *Иванов А. И.* Биометрическая идентификация личности по динамике подсознательных движений. Пенза: Изд-во Пенз. гос. ун-та, 2000. 188 с.
73. *Jafri R., Arabnia H. R.* A Survey of face recognition techniques // Journal of Information Processing Systems. 2009. Vol. 5, N 2. 28 p.
74. *Delac K., Grgic M., Bartlett M. S.* Recent advances in face recognition. 2008. 236 p.

75. *Ketki Kalamkar, Prakash Mohod.* A review on face recognition using different techniques // International Journal of Advanced Research in Computer Science and Software Engineering. 2015. N 5(1). P. 99—102.
76. *Руководство* по биометрии / Р. М. Болл, Дж. Х. Коннел, Панканти и др. М.: Техносфера, 2007. 368 с.
77. *Viola P., Jones M. J.* Rapid object detection using a boosted cascade of simple features // Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition. 2001. doi: 10.1109/CVPR.2001.990517
78. *Cho H., Hwang S. Y. J.* High-performance on-road vehicle detection with non-biased cascade classifier by weight-balanced training // EURASIP Journal on Image and Video Processing. 2015. P. 16. doi:10.1186/s13640-015-0074-5.
79. *Srinivasa K. G., Gosukonda S.* Continuous multimodal user authentication: coupling hard and soft biometrics with support vector machines to attenuate noise // CSI Transactions on ICT. 2014. Vol. 2, Iss. 2. P. 129—140. doi:10.1007/s40012-014-0054-4.
80. *Hough P. V. C.* A method and means for recognizing complex patterns. U. S. Patent. 1962. No. 3.069.654.
81. *Pisani P. H., Lorena A. C.* A systematic review on keystroke dynamics // Journal of the Brazilian Computer Society. 2013. N 19(4). doi:10.1007/s13173-013-0117-7.
82. *Брюхомицкий Ю. А., Казарин М. Н.* Учебно-методическое пособие к циклу лабораторных работ «Исследование биометрических систем динамической аутентификации пользователей ПК по рукописному и клавиатурному почеркам» по курсу: «Защита информационных процессов в компьютерных системах». Таганрог: Изд-во ТРТУ, 2004. 38 с.
83. *Ахмед Н., Рао К. Р.* Ортогональные преобразования при обработке цифровых сигналов / Пер. с англ. / Под ред. И. Б. Фоменко. М.: Связь, 1980. 248 с.
84. *Чалая Л. Э.* Модель идентификации пользователей по клавиатурному почерку // Искусственный интеллект. 2004. № 4. С. 811—817.
85. *Chang T. Y., Tsai C. J., Lin J. H.* A graphical-based password keystroke dynamic authentication system for touch screen handheld mobile devices // J. Syst. Softw. 2012. Vol. 85, N 5. P. 1157—1165.
86. *Lv H.-R., Wang W.-Y.* Biologic verification based on pressure sensor keyboards and classifier fusion techniques // IEEE Transactions on Consumer Electronics. 2006. Vol. 52, N 3. P. 1057—1063.
87. *Hidetoshi Nonaka, Masahito Kurihara.* Sensing pressure for authentication system using keystroke dynamics // International Journal of Computer, Control, Quantum and Information Engineering. 2007. Vo 1, N 1 (Режим доступа: http://waset.org/ publications/2995/sensing-pressure-for-authentication-system-using-keystroke-dynamics, дата обращения: 30.03.2015).
88. *Babaeizadeh M., Bakhtiari M., Maarof M. A.* Keystroke dynamic authentication in mobile cloud computing // International Journal of Computer Applications. 2014. Vol. 90, N 1. P. 29—36.
89. *Jun Chen, Guang Zhu, Jin Yang, Qingshen Jing, Peng Bai, Weiqing Yang, Xuewei Qi, Yuanjie Su, Zhong Lin Wang*. Personalized Keystroke Dynamics for Self-Powered Human Machine Interfacing // ACS Nano. 2015. Vol. 9, N 1. P. 105—116. doi:10.1021/nn506832w.

90. *Епифанцев Б. Н., Ложников П. С., Сулавко А. Е., Борисов Р. В.* Комплексированная система идентификации личности по динамике подсознательных движений // Безопасность информационных технологий. 2011. № 4. С. 97—102.
91. *Иванов А. И., Ложников П. С., Качайкин Е. И.* Идентификация подлинности рукописных автографов сетями Байеса—Хэмминга и сетями квадратичных форм // Вопросы защиты информации. 2015. № 2. С. 28—34.
92. *Борисов Р. В., Зверев Д. Н., Сулавко А. Е., Писаренко В. Ю.* Оценка идентификационных возможностей особенностей работы пользователя с компьютерной мышью // Вестник Сибирской государственной автомобильно-дорожной академии. 2015. № 5(45). С. 106—113.
93. *Васильев В. И., Ложников П. С., Сулавко А. Е., Жумажанова С. С.* Оценка идентификационных возможностей биометрических признаков от стандартного периферийного оборудования // Вопросы защиты информации. 2016. № 1. С. 12—20.
94. *Ложников П. С., Сулавко А. Е., Бурая Е. В., Писаренко В. Ю.* Непрерывная биометрическая аутентификация пользователей в процессе работы на компьютере // Вопросы кибербезопасности. 2017. № 3. С. 24—34.
95. *Васильев В. И., Еременко А. В., Сулавко А. Е., Жумажанова С. С.* Идентификационный потенциал стандартного оборудования в задачах скрытого распознавания пользователей компьютерных систем // Матер. X Международной IEEE научно-технической конференции «Динамика систем, механизмов и машин». Омск, Россия, 15–17 ноября, 2016. С. 236—242.
96. *Morelos-Zaragoza R. H.* The art of error correcting coding. John Wiley & Sons. 2006. 320 p.
97. *Соловьева Ф. И.* Введение в теорию кодирования: Учеб. пособие. Новосибирск: НГУ, 2006. 127 с.
98. *Еременко А. В., Сулавко А. Е.* Исследование алгоритма генерации криптографических ключей из биометрической информации пользователей компьютерных систем // Информационные технологии. 2013. № 11. С. 47—51.
99. *Иванов А. И.* Нейросетевая защита конфиденциальных биометрических образов гражданина и его личных криптографических ключей. Пенза: ПНИЭИ, 2014. 57 с.
100. *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 // Вестник УрФО. Безопасность в информационной сфере. 2014. № 2(12). С. 16—23.
101. *Scotti F., Cimato S., Gamassi M., Piuri V., Sassi R.* Privacy-aware Biometrics: Design and Implementation of a Multimodal Verification System // 2008 Annual Computer Security Applications Conference, IEEE. 2008. P. 130—139.
102. *Безяев А. В., Иванов А. И., Фунтикова Ю. В.* Оптимизация структуры самокорректирующегося биокода, хранящего синдромы ошибок в виде фрагментов хэш-функций // Вестник УрФО. Безопасность в информационной сфере. 2014. № 3(13). С. 4—13.

103. *Lozhnikov P. S., Sulavko A. E., Eremenko A. V., Volkov D. A.* Methods of generating key sequences based on parameters of handwritten passwords and signatures // Information. 2016. № 7(4).P. 59; doi:10.3390/info7040059.
104. *Волчихин В. И., Иванов А. И., Фунтиков В. А., Малыгина Е. А.* Перспективы использования искусственных нейронных сетей с многоуровневыми квантователями в технологии биометрико-нейросетевой аутентификации // Изв. вузов. Поволжский регион. 2013. № 4(28). С. 86—96.
105. *Иванов А. И., Ахметов Б. Б., Серикова Ю. И.* Усиление мощности хи-квадрат критерия при десятикратном увеличении числа степеней свободы статистических вычислений на малых тестовых выборках // Надежность и качество сложных систем. 2016. № 4 (16). С. 121—127. doi: 10.21685/2307-4205-2016-4-17.
106. *Иванов А. И., Газин А. И., Вятчанин С. Е., Перфилов К. А.* Сравнение мощности хи-квадрат критерия и критерия Крамера — фон Мизеса для малых тестовых выборок биометрических данных // Надежность и качество сложных систем. 2016. № 2(14). С. 67—73.
107. *Форсайт Дж., Молер К.* Численное решение систем линейных алгебраических уравнений. М.: Мир, 1969. 168 с.
108. *Райс Дж.* Матричные вычисления и математическое обеспечение. М.: Мир, 1984. 412 с.
109. *Лоусен Ч., Хенсон Р.* Численное решение задач методом наименьших квадратов. М.: Наука, 1966. 230 с.
110. *Тихонов А. Н., Арсенин В. Я.* Методы решения некорректных задач. М.: Наука, 1979. 248 с.
111. *Кобзарь А. И.* Прикладная математическая статистика. Для инженеров и научных работников. М.: Физматлит, 2006. 816 с.
112. Статистика: учеб. / Под ред. И. И. Елисеевой. М.: ТК Велби; Проспект, 2006. 448 с.
113. *Ложников П. С., Иванов А. И., Качайкин Е. И., Сулавко А. Е.* Биометрическая идентификация рукописных образов с использованием корреляционного аналога правила Байеса // Вопросы защиты информации. ФГУП. 2015. № 3. С. 48—54.
114. *Lozhnikov P. S., Sulavko A. E., Eremenko A. V., Buraya E. V.* Methods of generating key sequences based on keystroke dynamics // X International IEEE Scientific and Technical Conference «Dynamics of Systems, Mechanisms and Machines» (Dynamics), 15—17 November, 2016. Omsk, Russia, 2016. P. 1—5. doi: 10.1109/Dynamics.2016.7819038.
115. *Lozhnikov P. S., Sulavko A. E., Eremenko A. V., Volkov D. A.* Method of protecting paper and electronic text documents through a hidden biometric identifier based on a signature // X International IEEE Scientific and Technical Conference «Dynamics of Systems, Mechanisms and Machines» (Dynamics), 15—17 November, 2016, Omsk, Russia. P. 1—5. doi: 10.1109/Dynamics.2016.7819037.
116. *Епифанцев Б. Н., Ложников П. С., Сулавко А. Е.* Сравнение алгоритмов комплексирования признаков в задачах распознавания образов // Вопросы защиты информации. 2012. № 1. С. 60—66.

117. *Epifantsev B. N., Lozhnikov P. S., Sulavko A. E., Zhumazhanova S. S.* Identification potential of online handwritten signature verification // Optoelectronics, Instrumentation and Data Processing. 2016, № 3(52). P. 238—244. doi: 10.3103/S8756699016030043.
118. *Иванов А. И., Ложников П. С., Серикова Ю. И.* Снижение размеров достаточной для обучения выборки за счет симметризации корреляционных связей биометрических данных // Кибернетика и системный анализ. 2016. Т. 52, № 3. С. 49—56.
119. *Ахметов Б. С., Волчихин В. И., Иванов А. И., Малыгин А. Ю.* Алгоритмы тестирования биометрико-нейросетевых механизмов защиты информации Казахстан, Алматы, КазНТУ им. Сатпаева, 2013. 152 с. (Режим доступа: http://portal.kaznu.kz/ files/publicate/2014-01-04-11940.pdf, дата обращения: 24.05.2015).
120. *Ахметов Б. С., Надеев Д. Н., Фунтиков В. А., Иванов А. И., Малыгин А. Ю.* Оценка рисков высоконадежной биометрии. Алматы: Изд-во КазНТУ им. К. И. Сатпаева, 2014. 108 с.
121. *Качайкин Е. И., Иванов А. И.* Идентификация авторства рукописных образов с использованием нейросетевого эмулятора квадратичных форм высокой размерности // Вопросы кибербезопасности. 2015. № 4(12). С. 42—47.
122. *Волчихин В. И., Ахметов Б. Б., Иванов А. И.* Быстрый алгоритм симметризации корреляционных связей биометрических данных высокой размерности // Изв. вузов. Поволжский регион. Технические науки. 2016. № 1. С. 3—7.
123. *Нейросетевая* защита персональных биометрических данных // В. И. Волчихин, А. И. Иванов, В. А. Фунтиков, И. Г. Назаров. М.: Радиотехника, 2012. 157 с.
124. *Волчихин В. И., Иванов А. И., Серикова Ю. И.* Компенсация методических погрешностей вычисления стандартных отклонений и коэффициентов корреляции, возникающих из-за малого объема выборок // Изв. вузов. Поволжский регион. Технические науки. 2016. № 1. С. 45—49.
125. *Lozhnikov P., Sulavko A.* Cloud biometrical system identification through handwriting dynamics «SignToLogin». Certificate of registration № ТХ 7-640-429. 18.12.2012.

Научное издание

Ложников Павел Сергеевич

БИОМЕТРИЧЕСКАЯ ЗАЩИТА

ГИБРИДНОГО ДОКУМЕНТООБОРОТА

Монография

Редактор *Н. А. Лившиц*
Корректор *Н. В. Счастнева*
Технический редактор *Н. В. Бутакова*

Подписано в печать 30.12.2017
Уч.-изд. л. 8,2. Усл. печ. л. 8,13. Формат 60×90/16
Тираж 500 экз. Заказ № 256

Издательство СО РАН
630090, Новосибирск, Морской просп., 2
E-mail: psb@sibran.ru
тел. (383) 330-80-50
Отпечатано в Издательстве СО РАН
Интернет-магазин Издательства СО РАН
http://www.sibran.ru