ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«БЕЛГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ»
(НИУ «БелГУ»)

ИНСТИТУТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И ПРИКЛАДНОЙ МАТЕМАТИКИ
КАФЕДРА МАТЕМАТИЧЕСКОГО И ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ИНФОРМАЦИОННЫХ СИСТЕМ

АУТЕНТИФИКАЦИЯ ПОЛЬЗОВАТЕЛЯ ПО ИНДИВИДУАЛЬНОМУ ЭЛЕКТРОННОМУ ПОЧЕРКУ

Магистерская диссертация
обучающегося по направлению подготовки 02.04.01 «Математика и компьютерные науки» очной формы обучения, группы 07001531
Кирушевой Анастасии Игоревны

Научный руководитель
к.т.н, доцент Михелев Владимир Михайлович

Рецензент
к.т.н., доцент Жихарев Александр Геннадьевич

БЕЛГОРОД 2017

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

С развитием информационных технологий и сети Интернет возрастает потребность в обеспечении аутентичности данных, передаваемых по Сети. Фальсификация личности на сегодняшний день представляет большую опасность в отношении наносимого финансового ущерба. По оценкам Zecurion Anflytics за 2013 и 2014 гг., совокупные потери мировой экономики от подобных атак составили более 42 млрд. долл. Традиционные процедуры аутентификации основаны на проверке пароля, аппаратного идентификатора или биометрических данных пользователя.

Слабое звено паролей – человеческий фактор. Даже стойкий пароль, удовлетворяющий современным требованиям безопасности, не является гарантией надежной защиты, так как пользователь сам может сообщить его злоумышленнику, либо хранить пароли в ненадежном месте. Аппаратный идентификатор можно украсть или потерять. Последний способ (использование биометрии) является наиболее надежным, однако также не лишен недостатков. Физиологические признаки человека находятся «на виду», и существует множество способов их хищения незаметно для владельца.

В настоящее время активно идут процессы информатизации общества. Появляется все больше вед- сервисов. Многие государства стремятся создать электронное правительство для оказания услуг гражданам. Доверие к таким веб-службам со стороны пользователей должно быть наивысшим

Однако настоящее время разработаны технологии изготовления муляжей отпечатков пальцев, радужки, изображения лица и других биометрических признаков. Использование злоумышленником этих

технологий для совершения криминальных преступлений является вполне вероятностным событием и вопросом соответствующей ситуации.

Одним из основных факторов, определяющих состояние защищенности той или иной ключевой системы информационной инфраструктуры, является эффективность функционирования подсистемы управления доступом и защиты информации. Парольные и атрибутные методы идентификации и аутентификации имеют ряд существенных недостатков. Главный из них - возможность обмана системы защиты, путем кражи или имитации атрибута или взлома пароля. Второй недостаток данных методов идентификации и аутентификации - невозможность обнаружения подмены законного авторизированного пользователя. В данном случае злоумышленник может нанести вред обрабатываемой информации, когда оператор оставляет без присмотра систему с пройденной процедурой авторизации.

Методы аутентификации по биометрическим параметрам личности, в том числе и по клавиатурному почерку, способны обеспечить повышенную, по сравнению с другими способами проверки соответствия, точность и удобство для операторов автоматизированных систем. Методы постоянного клавиатурного мониторинга позволяют обнаруживать подмену законного оператора и блокировать ключевую систему от вторжения злоумышленника. Таким образом, задача исследования моделей, методов и алгоритмов распознавания клавиатурного почерка операторов ключевых систем является актуальной на данный момент.

Огромный интерес к биометрии обусловлен рядом объективных причин. В классических парольных системах кража приводит к компрометации всей системы. Более того, законный пользователь, потеряв или испортив карту, теряет возможность доступа к системе. Системы на основе биометрии практически лишены этих недостатков — идентификатор неразрывно связан с самим пользователем, поэтому потеря или изменение идентификатора

возможны только в чрезвычайных происшествиях, а современные сканеры биометрических данных позволяют обнаруживать попытки использования муляжей.

Все биометрические характеристики человека можно разделить на два класса — статические и динамические [25]. Статические — неизменяемые в течение продолжительного времени характеристики личности, данные ей от рождения. Примером статических характеристик могут быть рисунок радужной оболочки глаза, форма лица, рисунок папиллярных узоров пальца. Динамические характеристики отражают особенности, характерные для подсознательных движений в процессе воспроизведения какого-либо действия. Примерами подобных характеристик могут служить голос, почерк человека [7].

В системах биометрического сканирования ошибки первого и второго рода являются большой проблемой, использующих распознавание радужной оболочки или сетчатки глаза, черт лица и т.д. Такие сканирующие системы могут ошибочно отождествить кого-то с другим, «известным» системе человеком, информация о котором хранится в базе данных. Противоположной ошибкой будет неспособность системы распознать легитимного зарегистрированного пользователя, или опознать подозреваемого в преступлении.

- Ошибка первого рода (False Rejection Rate, FRR) — вероятность того, что легитимный пользователь может быть не распознан системой. Приемлемым уровнем ошибок первого рода в современных биометрических системах является 1% [12].
- Ошибка второго рода (False Acceptence Rate, FAR) — вероятность ошибочной аутентификации (идентификации) нелегитимного пользователя. Современные системы биометрической аутентификации позволяют достигать уровней ошибок второго рода менее, чем 0.00001% [12].

Ошибки первого и второго рода в первую очередь связаны с технической невозможностью получения всегда одинаковых цифровых образов данной биометрической характеристики при каждом ее сканировании. Шумы в датчике, различное положение частей тела человека при сканировании, искажения самих характеристик (мимика, ожоги или порезы пальцев, световые блики и т.п.) — все это отражается на формируемом цифровом образе [13].

Современные системы биометрической аутентификации включают в себя два модуля — модуль регистрации и модель аутентификации (идентификации) пользователей. При регистрации происходит многократное считывание выбранной характеристики, вычисление некоторого среднего его значения и запись этого значения в базу данных системы. В дальнейшем, при прохождении процедуры аутентификации пользователь вновь предъявляет свои биометрические данных. Модуль аутентификации производит сравнение полученного образа с тем, что хранится в базе и на основании сходства делает вывод об успешно/неуспешно пройденной процедуре.

В таких системах после выбора биометрической характеристики определяются функция расстояния между двумя образами и некоторый коэффициент сходства двух образов. Если расстояние между двумя образами не превышает этот коэффициент, то система рассматривает их, как образы, принадлежащие одному пользователю. В противном случае делается вывод о принадлежности образов разным пользователям. Подобные системы имеют ряд существенных недостатков, связанных с тем, что сами образы хранятся в базе данных:

- Две различные системы, использующие одну и ту же биометрическую характеристику, используют идентичную ключевую информацию о пользователе.
- В этих системах невозможна анонимная аутентификация пользователей.

- Существует возможность похищения базы данных эталонов злоумышленником. Шифрование базы данных лишь увеличит время доступа к самим данным. Более того, злоумышленником может быть администратор этой базы данных [13].

Возможность применения хэш-функций к цифровым образам позволила бы решить указанные проблемы. Однако использование хэш-функций затрудняется нечеткостью самих биометрических данных. Таким образом, одной из основных проблем, стоящих перед разработчиками биометрических систем, является проблема выработки уникальных фиксированных битовых строк из биометрических данных при каждой операции сканирования.

Первые результаты в этом направлении были получены в 2003 году, когда группа ученых из США предложили общие подходы к генерации ключевых последовательностей из нечетких данных. Суть процедуры, предложенной ими, заключается в использовании помехоустойчивого кодирования с целью устранения в определенном смысле незначительных искажений цифровых образов, получающихся при каждом сканировании биометрических данных человека [31].

Целью настоящей дипломной работы является разработка автоматизированной системы генератора ключевых последовательностей на основе биометрических данных пользователей.

Для достижения указанной цели были поставлены следующие задачи:

1) Разработка и реализация помехоустойчивого кода для евклидового расстояния;
2) Разработка алгоритма генерации ключевых последовательностей на основе клавиатурного почерка и его реализация в виде программного комплекса;
3) Экспериментальная проверка полученных результатов на образцах клавиатурных почерков реальных пользователей.

В первой главе настоящей дипломной работы даются основные понятия биометрии и описываются основные недостатки, присущие всем современным биометрических системам, а также общий подход к генерации ключевых последовательностей из нечетких данных.

Во второй главе рассматривается проблема генерации ключевых последовательностей из клавиатурного почерка пользователей, описываются алгоритм устранения грубых ошибок из наблюдений, помехоустойчивый код для евклидового расстояния.

В третьей главе описана реализация модели системы биометрической аутентификации, и приводятся результаты экспериментов, проведенных на этой модели.

В тексте выпускной квалификационной работы присутствуют 2 таблицы, 19 рисунков, 2 таблицы. В конце работы прилагается 1 приложение. Общий объем работы 72 страницы без учета приложений.

В завершении выпускной квалификационной следуют заключение и выводы по результатам проделанной работы, а также выводы по каждой из частей.

ГЛАВА 1. БИОМЕТРИЧЕСКИЕ ТЕХНОЛОГИИ

1.1. Биометрия

Биометрические методы распознавания применяются человечеством на протяжении всей его истории. Действительно, чаще всего мы узнаем знакомых людей именно с их помощью: по лицу, голосу или походке. Первые шаги массового применения биометрических методов можно связать с Альфонсом Бертильоном, который, в 90-е годы XIX века, работая в

картотеке парижской полиции, пришел к выводу, что для сочетания 14 единиц измерения (рост, окружность и длина головы, длина верха части туловища и т.п.) взрослого человека шанс совпадения по теории вероятности бесконечно мал (порядка $3{,}5*10^{-9}$) и если производить тщательный обмер каждого преступника и аккуратно заносить результаты в личные карточки, станет возможна безошибочная идентификация.

На смену идентификации по сумме измерений пришла дактилоскопия. В начале XX века англичанином Эдвардом Генри был предложен способ, благодаря которому идентификация по отпечаткам занимала несколько минут. В стройную научную дисциплину биометрию по отпечаткам пальцев превратил английский математик Карл Пирсон в 1903-1905 годы. Через 10лет система отпечатков стала практиковаться во всей Европе.

Как видно из исторических примеров, биометрические технологии, в первую очередь дактилоскопические, давно применяются в криминалистике, а с конца прошлого века в связи с развитием вычислительной техники и информационных технологий возникла возможность формализовать алгоритмы распознавания человека по его внешнему виду или особенностям поведения и применять для этого автоматизированные системы.

Применение биометрических характеристик для аутентификации в настоящее время переживает период бурного развития. В связи с этим все большее значение приобретает идентификация личности человека – потребителя информации в огромном количестве процессов, что вызывает высокий интерес общества к биометрическим технологиям в последние годы. Они являются весьма привлекательными для организации контроля любого доступа, так как обеспечивают высокий уровень надежности идентификации, могут быть интегрированы в любые системы контроля доступа одновременно с различными ключами и паролями.

Биометрические системы являются весьма удобными и дружелюбными для пользователей. В отличие от паролей и носителей информации для

систем контроля доступа (ключей, карт и т.д.), которые могут быть потеряны, украдены, скопированы, они основаны на биометрических параметрах (отпечатки пальцев, форм рук, лица и т.д.), которые всегда с нами и проблема их сохранности, как правило решается автоматически. Потерять их гораздо сложнее. Основными преимуществами биометрических технологий и систем безопасности являются следующие факторы:

- Избавление пользователей от проблем, связанных с потерей ключей и удостоверений личности;
- Уникальность биометрических характеристик каждого человека делает невозможным их использование третьими лицами;
- Процесс распознавания, благодаря интуитивности программного и аппаратного интерфейса, понятен и доступен людям любого возраста и не знает языковых барьеров;
- В случае каждого обращения к системе можно доказать авторство того или иного действия.

Биоидентификация основана на уникальности характеристик человеческого тела – не существует двух людей с одинаковыми биометрическими признаками. Под биометрикой понимают область науки, изучающую методы измерения физических характеристик и поведенческих черт человека для последующей идентификацией и аутентификаций личности.

Биометрической характеристикой человека (БХЧ) или биометрическим параметром называется его измеренная физическая характеристика или персональная поведенческая черта, в процессе сравнения которой с аналогичной ранее зарегистрированной БХЧ реализуется процедура идентификации. Основными источниками БХЧ являются отпечатки пальцев, радужная оболочка, голос, лицо, манера работы на клавиатуре компьютера, подпись и др.

Обычно при классификации биометрических технологий выделяют две группы систем по типу используемых биометрических параметров:

- Первая группа систем использует статические биометрические параметры: отпечатки пальцев, геометрия руки, сетчатка глаза и т. п.
- Вторая группа систем использует для идентификации динамические параметры: динамика воспроизведения подписи или рукописного ключевого слова, голос и т. П [25].

Все биометрические системы работают практически по одинаковой схеме. Во-первых, система запоминает образец биометрической характеристики (это и называется процессом записи). Во время записи некоторые биометрические системы могут попросить сделать несколько образцов для того, чтобы составить наиболее точное изображение биометрической характеристики [33]. Затем полученная информация обрабатывается и преобразовывается в математический код.

Кроме того, система может попросить произвести ещё некоторые действия для того, чтобы «приписать» биометрический образец к определённому человеку. Например, персональный идентификационный номер (PIN) прикрепляется к определённому образцу, либо смарт-карта, содержащая образец, вставляется в считывающее устройство. В таком случае снова делается образец биометрической характеристики и сравнивается с представленным образцом.

Биометрика решает вопросы верификации и идентификации. В первом случае задача состоит в том, что полученная биометрическая характеристика соответствует ранее взятой. Верификация использует для проверки того, что субъект является именно тек, за кого себя выдает.

Идентификация решает вопрос поиска для получаемой биометрической характеристики наиболее подходящий из ранее взятых. Это последовательное осуществление сравнений полученной характеристики со

всеми имеющимися в базе данных. При этом в качестве результата будет выбрана более похожая ранее взятая характеристика или не будет вообще никакого результата, если степень похожести оказалась меньше заданной для вех сравнений.

Идентификация по любой биометрической системе проходит четыре стадии:

- Запись — физический или поведенческий образец запоминается системой;
- Выделение — уникальная информация выносится из образца и составляется биометрический образец;
- Сравнение — сохранённый образец сравнивается с представленным;
- Совпадение/несовпадение — система решает, совпадают ли биометрические образцы, и выносит решение [17].

Смысл биометрических систем безопасности, во- первых, состоит в том, чтобы доказать, что вы-это вы, и если посторонний может выдать себя за вас – систем малопригодна. Такой результат называется ошибочной позитивной идентификацией. Во-вторых, необходимо исключить возможность того, что система примет вас за другого человека. То есть она должна доказать, что вы-это вы, а не кто-любо другой, и если вы не сможете убедить в этом систему, значит она опять-таки не очень хороша. Такой вариант называется ошибочной негативной идентификацией.

Таим образом, оценка эффективности биометрических технологий, помимо стоимостных показателей и удобства использования, основывается на использовании двух основных вероятностных параметров – Ошибка ложного допуск (FAR) и ошибки ложного отказа (FRR). Их так же именуют ошибками первого и второго рода соответственно.

Необходимо учитывать взаимосвязь этих показателей: искусственно снижая уровень «требовательности» системы (FAR), мы, как правило, уменьшаем процент ошибок FRR, и наоборот [12].

На сегодняшний день все биометрические технологии являются вероятностными, ни одна из них не способна гарантировать полное отсутствие ошибок FAR/FRR, и нередко данное обстоятельство служит основой для не слишком корректной критики биометрии.

Очевидно, что ошибка ложного допуска более опасна с точки зрения безопасности, а точка ложного отказа приводит к уменьшению удобства пользования системой, которая не иногда не распознает человека с первого раза.

Эти две вероятности взаимосвязаны, попытка уменьшения одной приводит к увеличению второй, поэтому н практике в зависимости от требований к системе выбирается определенный компромисс.

Биометрические технологии активно применяются во многих областях, связанных с обеспечением безопасности доступа к информации и материальным объектам, а также в задачах уникальной идентификации личности.

1.1.1. Динамические методы биометрической аутентификации

Методы этой группы основываются на поведенческой (динамической) характеристике человека, то есть построены на особенностях, характерных для подсознательных движений в процессе воспроизведения какого-либо действия. Рассмотрим методы аутентификации этой группы [29]:

По рукописному почерку. Подпись — один из классических способов идентификации, применяемый уже несколько столетий в юридической

практике, банковском деле и торговле. Автор придумывает себе факсимиле и отрабатывает его тренировками. Желательно, чтобы факсимиле не повторяло обычное написание букв и имело дополнительные элементы (росчерки, наложения букв и т.д.). Есть два независимых способа идентификации по подписи:

1. Идентификация по рисунку подписи на документе.
2. Идентификация по динамике подписи, вводимой в компьютер.

В первом способе сравниваются два изображения. С этим лучше справляется человек. Во втором способе есть данные о колебаниях пера при воспроизведении подписи в трёхмерном пространстве (X, Y — координаты и Z — давление на планшет). С этим может справиться только компьютер. Вероятность ошибки 1 рода составляет около 0,01 — вполне приемлемо. Однако для ошибок второго рода это чрезвычайно много. Для снижения вероятности используют ключевое слово.

Среди достоинств можно отметить невысокую стоимость и относительную привычность для человека.

К недостаткам относятся высокий уровень ошибок 1 и 2 рода, необходимость приучения к работе с планшетом перед регистрацией (значительно отличается от письма обычной ручкой из-за размеров пера, других масштабов и невозможности наблюдения за процессом и результатом одновременно), продолжительное время регистрации пользователя (более 2 минут).

По клавиатурному почерку. Метод в целом аналогичен вышеописанному, но вместо росписи используется некая ключевая фраза и не нужно никакого специального оборудования, кроме стандартной клавиатуры. Такую аутентификацию легко сделать двухфакторной, если в качестве ключевой фразы использовать личный пароль пользователя. Основной характеристикой по которой строится свертка для идентификации

— динамика набора кодового слова. Встречаются системы, в которых используются специальные клавиатуры, позволяющие регистрировать силу давления на клавиши, что значительно снижает вероятности ошибок первого и второго родов.

По голосу. Идентификация человека по голосу — один из традиционных способов распознавания, применяемый повсеместно. Можно легко узнать собеседника по телефону, не видя его. Также можно определить психологическое состояние по эмоциональной окраске голоса. Т. к. голосовая идентификация бесконтактна и не требует от человека особых усилий, ведутся работы по созданию голосовых замков и систем ограничения доступа к информации. Интерес в этой области связан ещё и с прогнозами повсеместного внедрения голосовых интерфейсов — предполагается их широкое использование в построении «интеллектуальных зданий».

Основная проблема низкой надежности метода заключается в большом разнообразии проявления голоса одного человек: голос может меняться в зависимости от настроения, состояния здоровья, возраста и т.д. Это разнообразие определяет серьезные трудности при выделении индивидуальных свойств человека. Кроме того, учет шумовой компоненты является еще одной серьезной и не до конца решенной проблемой в практическом применении идентификации по голосу.

1.1.2. Статические методы биометрической аутентификации

Методы данной группы основываются на физиологической (статической) характеристике человека, то есть уникальной характеристике, данной ему от рождения и неотъемлемой от него. Большинство статических характеристик невозможно изменить без глубокого хирургического вмешательства, а некоторые из них в принципе не поддаются модификации

современной медициной. Рассмотрим некоторые методы аутентификации этой группы [3]:

По отпечатку пальца. Наиболее популярный в настоящее время метод, который занимает устойчивое доминирующее положение – 59% по данным 2011 г.

Из достоинств этого метода стоит отметить очень хорошее отношение цена/качество, а также небольшие размеры сканеров, что позволяет размещать их в портативных устройствах.

К недостаткам стоит прежде всего отнести негативное отношение со стороны пользователей, возможность легкого изготовления муляжей, а также зависимость от чистоты пальца, а также от его повреждений.

По форме кисти. Данный метод, построен на геометрии кисти руки. С помощью специального устройства, состоящего из камеры и нескольких подсвечивающих диодов (включаясь по очереди, они дают разные проекции ладони), строится трехмерный образ кисти руки по которому формируется свертка и распознается человек. Есть два подхода при использовании геометрии руки.

К серьезным недостаткам относятся громоздкость сканирующих устройств (за некоторым исключением), а также невысокая сложность изготовления муляжа для устройств первого типа (использующих только геометрические характеристики).

По сетчатке глаза. В основе методов — уникальность рисунка кровеносных сосудов глазного дна. Для того, чтобы этот рисунок стал виден, человеку нужно посмотреть на удаленную световую точку, и таким образом подсвеченное глазное дно сканируется специальной камерой.

Подобные методы обеспечивают одни из самых низких уровней ошибок первого и второго рода (уступают только анализу ДНК), однако применение систем сильно ограничено высокой стоимостью сканеров,

неудобством пользования (необходимо снимать очки, прикладывать глаз к окуляру и в течение некоторого времени испытывать негативные ощущения от сфокусированного свечения сканера), а также невозможностью использования пользователями с различными заболеваниями глазного яблока.

По радужной оболочке глаза. Рисунок радужной оболочки глаза является уникальной для каждого человека биометрической характеристикой. Он формируется в первые полтора года жизни и остаётся практически без изменений до самой смерти (изменения радужной оболочки связаны с болезнями). Изменения могут выражаться в виде изменения цвета, появления пигментных пятен, линий, кругов, изменений обвода оболочки, деформации зрачка и т.д. Эти особенности используются для диагностики заболеваний и предрасположенности к ним.

К достоинствам данных методов относится высокая степень распознавания, низкая вероятность ошибок первого и второго рода, бесконтактное сканирование, а также безразличие к наличию очков и большинству распространенных болезней глаз. Из недостатков можно выделить лишь необходимость «привыкания» к системе.

По форме лица. Технологии распознавания по лицу ненавязчивы (распознавание происходит на расстояние, без задержек и отвлечения внимания), как правило, пассивны (не требуют каких-либо действий со стороны человека), не ограничивают пользователя в свободе перемещений и относительно недороги. По лицу человека можно узнать его историю, симпатии и антипатии, болезни, эмоциональное состояние, чувства и намерения по отношению к окружающим. Всё это представляет особый интерес для автоматического распознавания лиц (например, для выявления потенциальных преступников).

Чаще всего используются три метода распознавание по форме лица:

1. Корреляционный (метод согласованной фильтрации).
2. Метод на основе преобразований Карунена-Лоэва и понятии «собственных лиц» (“EigenFace”).
3. Метод на основе линейного дискриминантного анализа и понятия “Fisherface” (от имени учёного Роберта Фишера) [4].

Системы идентификации по лицу часто используются в аэропортах при прохождении паспортного контроля, в людных местах для поиска преступников, в казино для выявления мошенников и на фейс-контроле для недопуска лиц из «чёрного списка». Преимущества подобных методов прежде всего заключаются в отсутствии необходимости каких-либо действия со стороны пользователей. К недостаткам же стоит отнести невозможность различать близнецов.

По термограмме лица. Информационными признаками являются рисунки вен и артерий, которые повторить в муляже практически невозможно. Для получения термографического образа используются камеры, улавливающие инфракрасное излучение. Они могут работать в полной темноте. Информационные признаки же не зависят ни от температуры лица, ни от пластических операций, ни от старения человека. Инфракрасная камера позволяет получать образ даже на значительном удалении от человека. Однако, цена таких камер на 1—2 порядка выше, чем у камер на обычных матрицах. Такие системы можно использовать в системах контроля доступа с высокой степенью ответственности и надёжности. Они обеспечивают очень высокую точность распознавания. Исследования показывают, что однояйцовые близнецы имеют свои уникальные термограммы. К недостаткам можно отнести необходимость положения лица строго в анфас (в камеру).

По ДНК. Теоретически, данные методы позволяют однозначно идентифицировать пользователя, устранив тем самым ошибки первого и второго рода. Преимущества данного способы очевидны, однако

используемые в настоящее время методы получения и обработки ДНК работают настолько долго и, что такие системы используются только для специализированных экспертиз. Кроме того, стоимость проведения подобного анализа также не способствует его широкому распространению.

Статическая и динамическая биометрии — две взаимно дополняющие друг друга ветви. Основное преимущество статической биометрии — относительная независимость от психологического состояния пользователей, малые затраты их усилий и, следовательно, возможность организации биометрической идентификации больших потоков людей. Преимуществом же динамических методов является простота и низкая стоимость их технической реализации.

Для использования в системах аутентификации, биометрическая характеристика должна быть уникальной, постоянной и собираемой, кроме того она должна быть устойчива к подделке.

Уникальность означает, что не должно существовать двух людей с идентичными характеристиками. Постоянство (перманентность) — характеристика не должна изменяться со временем. Собираемость (измеримость) — возможность быстро и легко получить и детализировать характеристику от индивидуума.

Как видно из экспертной оценки свойств биометрических характеристик человека (Табл. 1.1), ни одна из существующих характеристик не удовлетворяет всем требованиям в полной мере. Кроме того, общей характеристикой, используемой для сравнения различных методов и способов биометрической идентификации, являются статистические показатели — ошибки первого и второго рода.

Таблица 1.1

Оценка свойств БХ

Характеристика	Уникальность	Перманентность	Измеримость	Устойчивость к подделке
Рукописный почерк	+	+	+++++	+
Клавиатурный почерк	++	+	+++++	++
Голос	+++	++	++++	+
Отпечаток пальца	++++	+++++	+++	+++
Форма кисти рук	+++	+++	+++	+
Радужная оболочка глаза	+++++	+++++	+++	++++
Форма лица	+++	+++	+++	+
Термограмма лица	++++	++++	+++	++++
ДНК	+++++	+++++	+	+++++

Ошибка первого рода возникает в случае отказа в доступе легитимному пользователю. Ошибка второго рода возникает в случае предоставления доступа злоумышленнику. Предоставление корректного пароля в системе аутентификации по паролю, всегда дает корректный результат о подтверждении подлинности. Но если в биометрическую систему аутентификации представлены легитимные биометрические характеристики, это, тем не менее, не гарантирует корректной аутентификации.

В первую очередь это связано с технической невозможностью получения абсолютно одинаковых цифровых образов при каждом сканировании биометрической характеристики одного пользователя. Причинами тому являются «шумы», различное положение и угол наклона частей тела человека при сканировании [4]. Ограничения методов обработки полученных образов также накладывают свой отпечаток на качество распознавания.

Во вторую очередь следует отметить изменчивость самих биометрических характеристик. Так, например, любые динамические характеристики очень сильно зависят от эмоционального и физического состояния личности. Усталость, раздражение, состояние алкогольного опьянения и т.п. могут очень сильно изменить предоставляемые человеком динамические биометрические характеристики. Статические характеристики, как уже упоминалось, обладают гораздо большей устойчивость по отношению к состоянию индивида, однако, и они подвержены изменению с течением времени, в связи с окружающей обстановкой и ситуациями, выйденными за рамки обыденной жизни — пользователь может обжечь палец, что приведет к повреждению рисунка папиллярных линий; ушиб кисти может на некоторое время привести к значительному изменению формы; а некоторые заболевания глазного яблока препятствуют получению легитимного образа рисунка кровеносных сосудов глазного дна.

Есть также вероятность, что может быть подтверждена подлинность человека, выдающего себя за другого, легитимного пользователя [5]. Это связано с тем, что биометрические характеристики разных людей могут быть в достаточной степени схожи друг с другом. Так, например, у близнецов наблюдается поразительное сходство очень многих биометрических характеристик. В большинстве биометрических систем имеется возможность уменьшать вероятность одной из ошибок за счет увеличения вероятности другой — так называемая «чувствительность» системы. Полностью избавиться от одной из ошибок можно только увеличив вероятность другой до 100%, что, разумеется, неприемлемо.

Логически любая система биометрической аутентификации (идентификации) может быть разделена на два модуля: модуль регистрации и модуль аутентификации (идентификации) [6]. Модуль регистрации отвечает за «обучение» системы. На этапе регистрации биометрические датчики сканируют физиогномику человека для того, чтобы создать цифровое

представление. Специальный модуль обрабатывает это представление с тем, чтобы выделить характерные особенности и сгенерировать более компактное и выразительное представление, называемое шаблоном. Для изображения лица такими характерными особенностями могут стать размер и относительное расположение глаз, носа и рта; для клавиатурного почерка — временные интервалы между нажатиями клавиш, для радужной оболочки глаза — так называемый ирис-код [4].

Поскольку при каждом сканировании характеристики получаются несколько различные значения, для более уверенного распознавания в дальнейшем необходимо найти некий усредненный шаблон для каждого пользователя. Это достигается путем многократного сканирования, после чего необходимо тем или иным методом исключить из рассмотрения те шаблоны, в которых содержатся некоторые недопустимые случайны погрешности, например, размазанные изображения, сильные световые блики и т.п., и найти средний образ. В большинстве современных систем биометрической аутентификации шаблон для каждого пользователя храниться в базе данных системы. Эта база данных может быть централизованной или распределенной — когда шаблон каждого пользователя сохраняется на смарт-карте или другом цифровом носителе и передается пользователю [7].

Модуль аутентификации отвечает за распознавание человека. На этапе аутентификации биометрический датчик снимает характеристики человека, аутентификация которого проводится и преобразует эти характеристики в тот же цифровой формат, в котором храниться шаблон. Полученный шаблон сравнивается с хранимым шаблоном с тем, чтобы определить, соответствуют ли эти шаблоны друг другу. Подобные системы могут работать в двух режимах — идентификации, когда происходит поиск наиболее подходящего шаблона из всей базы данных, и аутентификации, когда происходит сравнение с конкретным указанным шаблоном. В таких системах изначально

задается определенный коэффициент сходства. Если разница между предоставленным образом и хранимым эталоном не превышает этот коэффициент, то пользователь успешно проходит процедуру аутентификации, в противном случае система принимает решение отказать в доступе. Подобные системы имею ряд существенных недостатков [7]:

1) Невозможность обеспечить анонимность пользователей. Более того, в различных системах, использующих одни и те же биометрические показатели, ключевая информация с некоторой погрешностью будет совпадать.
2) Невозможность использования самих биометрических данных для построения криптографических ключей. Это связано с нечеткостью самих данных и погрешностями при «сканировании».
3) Возможность похищения базы данных эталонов злоумышленником. В частности, злоумышленником может быть сам администратор базы данных.

Шифрование базы данных в данном случае лишь увеличит время доступа к данным. Получение из нечетких биометрических данных уникальных фиксированных битовых строк дала бы возможность построения систем, лишенных указанных выше недостатков.

1.2. Процедура генерации ключевых последовательностей на основе нечетких данных

Рассмотрим произвольное конечное метрическое пространство N, точки которого представляют собой цифровые отпечатки выбранной биометрической характеристики личностей. При этом предполагается, что достаточно близкие точки этого пространства представляют собой цифровые

образы биометрической характеристики одного пользователя из системы, в то время, как удаленные точки представляют собой цифровые отпечатки биометрической характеристики различных пользователей. Для каждой из биометрических характеристик удобно использовать свою метрику. Так, например, для ирис-кода предоставляется возможным использовать в качестве расстояния дистанцию Хэмминга, а для геометрии лица или клавиатурного почерка — обыкновенное евклидово расстояние [15].

1.2.1. Представление исходных данных

Сообщения, передаваемые по каналу связи, представляют собой кодовые комбинации N точек A_i многомерного метрического пространства с координатами $x(i) = \big(x_1(i), x_2(i), \dots, x_m(i)\big), 1 \le i \le N$ [19]. Рассматривается случай N=2^n точек для безызбыточного равномерного двоичного кодирования которых потребуется n-битные кодовые комбинации. В качестве меры, определяющей расстояние между парой течек A_i и A_j, например может рассматриваться евклидово расстояние (1.1):

$$d_{i,j} = \sqrt{\sum_{l=1}^{m}(x_l(i) - x_l(j))^2},$$

$$d_{i.j} = d_{j.i}\,, d_{i.j} = 0, d_{i.j} > 0 \qquad (1.1)$$

при i≠j.

1.3. Описание генетического алгоритма кодирования

На первом шаге работы алгоритма производится поиск начальной точки нулевой категории, которая является центром тяжести массива всех исходных точек. Для каждой точки A_i вычисляется сума ее евклидовых расстояний (1.2)

$$D_i = \sum_{j=1}^{N} d_{i.j} \qquad (1.2)$$

до остальных точек массива, тогда начальной точкой A_i будет та, у которой это значение минимально (1.3):

$$i_0 = argmin\{D_i, 1 \le i \le N\}. \qquad (1.3)$$

Она кодируется n-битной комбинацией, состоящей из одних нулей.

На втором шаге производится поиск течек первой категории – это группа из n точек, в целом наиболее близкая к начальной точке A_i. Их кодируют любой из перестановок n-битных кодовых комбинаций, содержащих одну единицу: (00..001),(00…010),(00…100) и так далее. Для этого находим следующие значения (1.4):

$$i_{1,1} = argmin\{d_{i_0,j};\ 1 \le i \le N, j \ne i_0\},$$

$$i_{1,2} = argmin\{d_{i_0,j};\ 1 \le i \le N, j \ne i_0,\ j \ne i_{1,1}\}, \qquad (1.4)$$

…

$$i_{1,n} = argmin\{d_{i_0,j};\ 1 \le i \le N, j \ne i_0,\ \ j \ne i_{1,1},\ \ j \ne i_{1,2}, \dots,\ \ j \ne i_{1,n-1}\},$$

тогда точки $A_{i_{1,1}}, A_{i_{1,2}}, \dots, A_{i_{1,n}}$ – это точки первой категории с множеством индексов (1.5)

$$\Theta_1 = \{i_{1,1}, i_{1,2}, \dots, i_{1,n}\}. \qquad (1.5)$$

На третьем шаге производится поиск точек, которые в совокупности наименее удалены от точек первой категории – это точки второй категории, кодовые комбинации которых будут отличаться одной дополнительной единицей от кодовых комбинаций точек первой категории. Для удобства введем множество индексов точек (1.6), свободных для рассмотрения на этом шаге, то есть тех, которые не являются ни начальной точкой, ни точкой первой категории:

$$\Psi_1 = \{k;\ 1 \le k \le N, k \notin \Theta_1, k \ne i_0\}. \qquad (1.6)$$

Дальнейший анализ будет проводиться на основе двумерного массива (1.7)

$$\{d_{i,j}(k), d_{i,j}(k) = d_{i,k} + d_{k,j}\} \qquad (1.7)$$

который содержит значения суммы расстояний от каждой из «свободных» точек $A_k, k \in \Psi_1$ до каждой из $n_2 = C_n^2$ пар точек первой категории $A_{i_{1,n}}, A_{j_{1,n}}$, i,j $\in \Theta_1$. Поиск точек второй категории и наименее удаленных от них точек первой категории будет определяться следующим соотношениями (1.8):

$$k_{2,1}(i_{1,p_1}, j_{1,q_1}) = argmin\{d_{i,j}(k);\ k \in \Psi_1,\ i,j \in \Theta_1, i \ne j\},$$

$$k_{2,2}(i_{1,p_2}, j_{1,q_2}) = argmin\begin{Bmatrix} d_{i,j}(k);\ k \in \Psi_1,\ k \ne k_{2,1},\ i,j \in \Theta_1, \\ i \ne j, \\ (i,j) \ne (i_{p_1}, j_{q_1}) \end{Bmatrix},$$

… (1.8)

$$k_{2,m}(i_{1,p_m}, j_{1,q_m}) = argmin\begin{Bmatrix} d_{i,j}(k);\ k \in \Psi_1,\ k \ne k_{2,1}, k \ne k_{2,2}, \dots, \\ k \ne k_{2,m-1}\ i,j \in \Theta_1, i \ne j, \\ (i,j) \ne (i_{p_1}, j_{q_1}), \\ (i,j) \ne (i_{p_2}, j_{q_2}), \dots, (i,j) \ne (i_{p_{m-1}}, j_{q_{m-1}}) \end{Bmatrix},$$

Тогда точки $A_{k_{2,1}}, A_{k_{2,2}}, \dots, A_{k_{2,m}}$ - это точки второй категории с множеством индексов $\Theta_2 = \{k_{2,1}, k_{2,2}, \dots, k_{2,m}\}$. Присвоение кодовых

комбинаций точками второй категории осуществляется следующем образом: если, например, точке A_{1,p_2} соответствует кодовая комбинация (000…010), а точке A_{1,q_2} – (000…110), то ближайшей к ним по исходному расстоянию точке второй категории $A_{k_{2,2}}$ присваивается кодовая комбинация (000…110). Расстояние Хэмминга между этой точкой и двумя предыдущими равно единице, а между кодовыми комбинациями самих порождающих точек A_{1,p_2} и A_{1,q_2} равно двум.

На четвертом шаге аналогичным образом рассматриваются тройки точек (I,j,k), i,j,k $\in\Theta_2$ второй категории, удовлетворяющие следующему условию: расстояние Хэмминга между ним попарно должно быть равно двум, т.е. r(i,j) =r(i,k)= r(j,k) =2 [10]. Эти тройки впоследствии и образуют $n_3 = C_n^3$ точек третьей категории, чьи кодовые комбинации будут содержать уже три единицы. Например, если тройке точек второй категории соответствовали кодовые комбинации (0011…0), (1001…0) и (1010…0), тогда ближайшей к ним по исходному расстоянию точке третьей категории будет присвоена комбинация (1011…0).

По той же схеме происходит поиск и кодирование четвертой и последующих категорий. В итоге, на последнем шаге останется одна свободная точка, которой будет присвоена кодовая комбинация, состоящая из одних единиц.

1.4. CRC- коды

Циклические избыточные коды (Cyclic redundancy check, CRC) используется для проверки целостности передаваемых данных или блоков данных, когда передача осуществляется в пакетных режимах. Принцип их работы заключается в передаваемом блоку данных дополнительной

последовательности бит, таким образом, что результирующий блок будет делиться без остатка на порождающий полином CRC- кода, т.е. в качестве контрольной суммы берется остаток от деления значения порождающего полинома [7]. Использование этого кода в технологии LTE делает его анализ особенно актуальным для сетей связи 4-ого поколения и передачи данных в Интернете.

Пусть $p(x)$ – примитивный многочлен степени m, тогда порождающий многочлен CRC-кода $G_m(x)$ можно записать в виде произведения (1.9):

$$G_m(x) = (1 + x)p(x). \qquad (1.9)$$

С помощью порождающего многочлена $G_m(x)$ может быть построен циклический CRC (n,k) – код с параметрами $n = 2^m - 1, k = 2^m - m - 2$, имеющий m+1 проверочных символов и $d_{min} = 4$. CRC-коды обладают следующими важными свойствами:

- Все ошибки кратности 3 или меньше обнаруживаются;
- Все ошибки нечетной кратности обнаруживаются;
- Все пакеты ошибок длины m+1 или меньше обнаруживаются;
- Доля не обнаружимых пакетов ошибок длины m+2 составляет 2^{-m};
- Доля не обнаружимых пакетов ошибок длины большей m+2 составляет $2^{-(m-1)}$;

где n- длина кодового слова, k- длина информационного блока, d_{min} – минимальное расстояние между кодовыми комбинациями.

Использование CRC- кодов совместно с предварительным генетическим кодированием метризированного источника сообщений позволяет не только обнаружить указанные ранее различные виды ошибок, если таковые присутствуют в принятых блоках данных, но и значительно

снизить их влияние на данные непосредственно при последующем декодировании [7].

В подобной системе каждому доверенному пользователю выдается цифровой пропуск, содержащий значение хэш-функции от сгенерированного ключа и соответствующую открытую строку V. Всю информацию необходимо подписать, используя один из алгоритмов цифровой подписи.

При прохождении процедуры аутентификации пользователь предъявляет цифровой пропуск и свои биометрические данные. Если значение хэш-функции от полученного ключа совпадает со значением, записанным на цифровом пропуске, то делается вывод об успешной аутентификации.

Применение генерации ключевой последовательности нечетких данных позволяет не только получать уникальные фиксированные битовые строки из биометрических данных пользователей, но также дает возможность использования различной ключевой информации в каждой из систем, в которых работает пользователь.

Построение систем аутентификации с использованием генерации ключевой последовательности нечетких данных включает в себя следующие основные этапы [4]:

1) Выбор биометрической характеристики для использования.
2) Выбор доверительного интервала — той максимальной разницы между двумя образами, при которой система считает их принадлежащими одному пользователю.
3) Выбор наиболее подходящей метрики для данной биометрической характеристики.
4) Построение кода, исправляющего ошибки для выбранной метрики. При этом исправляющая способность кода однозначно определяется выбранным доверительным интервалом.

Основные преимущества, которые дает применение генератора ключевых последовательностей в биометрических системах, можно описать так [4]:

1) Возможность избавиться от необходимости хранения эталонных значений биометрических данных пользователей. Вместо этого в системах аутентификации достаточно будет хранить значения хэш-функций от сгенерированных ключевых строк.
2) Невозможность восстановления биометрического образа из ключевой строки.
3) Невозможность использования найденного или украденного цифрового пропуска (смарт-карты).
4) Возможности получения различных и независимых между собой ключей во всех системах на основе одной и той же биометрической характеристики.
5) Возможность анонимной аутентификации в группе доверенных пользователей.
6) Возможность неразрывно связывать пользователя и его цифровую подпись. Невозможно подписать документ, имея только одну из составляющих.

1.5. Вывод

В последние годы во всем мире наблюдается все возрастающий интерес к методам распознавания и идентификации личности. Основный пути и способы решения этих задач лежат в области разработки биометрических систем. В биометрических системах для распознавания человека используется совокупность биометрических характеристик, основанных на биологических особенностях человеческого тела.

Биометрическая защита более эффективна в сравнении с такими методами, как использование паролей, PIN-кодов, смарт-карт, жетонов, поскольку биометрия позволяет идентифицировать именно конкретного человека, а не устройство. Традиционные методы защиты не исключают возможности потери или кражи информации, вследствие чего она становится доступной незаконным пользователям. Биометрическая система безопасности позволяет отказаться от парольной защиты либо служит для ее усиления.

Одной из основных причин которые существенно повысили значимость автоматической обработки и анализа биометрической информации, явилось повышение требований к функциональным возможностям автоматических систем безопасности, расположенных в общественных местах, связанных с необходимостью в реальном времени выполнять необходимые действия по установлению личности присутствующих на контрольной территории людей, причем, не только бесконтактно, но и без специального сотрудничества со стороны идентифицируемые персон.

В настоящий момент общее состояние биометрических технологий в мире еще нельзя признать удовлетворительным. Скорее можно говорить о биометрии как о быстро развивающейся области исследований и приложений, в которой еще не удалось достичь требуемых показателей. Целый ряд серьезных проверок, проведенных в последнее время, показал недостаточную надежность таких систем.

ГЛАВА 2. ПРОЕКТИРОВАНИЕ И РЕАЛИЗАЦИЯ

2.1. Технологии

Oracle ADF

Oracle ADF (Application Development Framework) основан на шаблоне проектирования Model-View-Controller (MVC) (рис.2.1). Приложения MVC разделяются на следующие уровни[1]:

1) уровень модели, который взаимодействует с источниками данных и обрабатывает бизнес-логику приложения;
2) визуальный уровень, который представляет собой пользовательский интерфейс приложения;
3) уровень контроллера, который управляет потоком приложения и выступает в качестве интерфейса между уровнем модели и визуальным уровнем.

Разделение приложения на эти три уровня упрощает техническое обслуживание и повторное использование компонентов в различных приложениях. Независимость каждого уровня от других приводит к ориентированной на сервисы архитектуре (SOA).

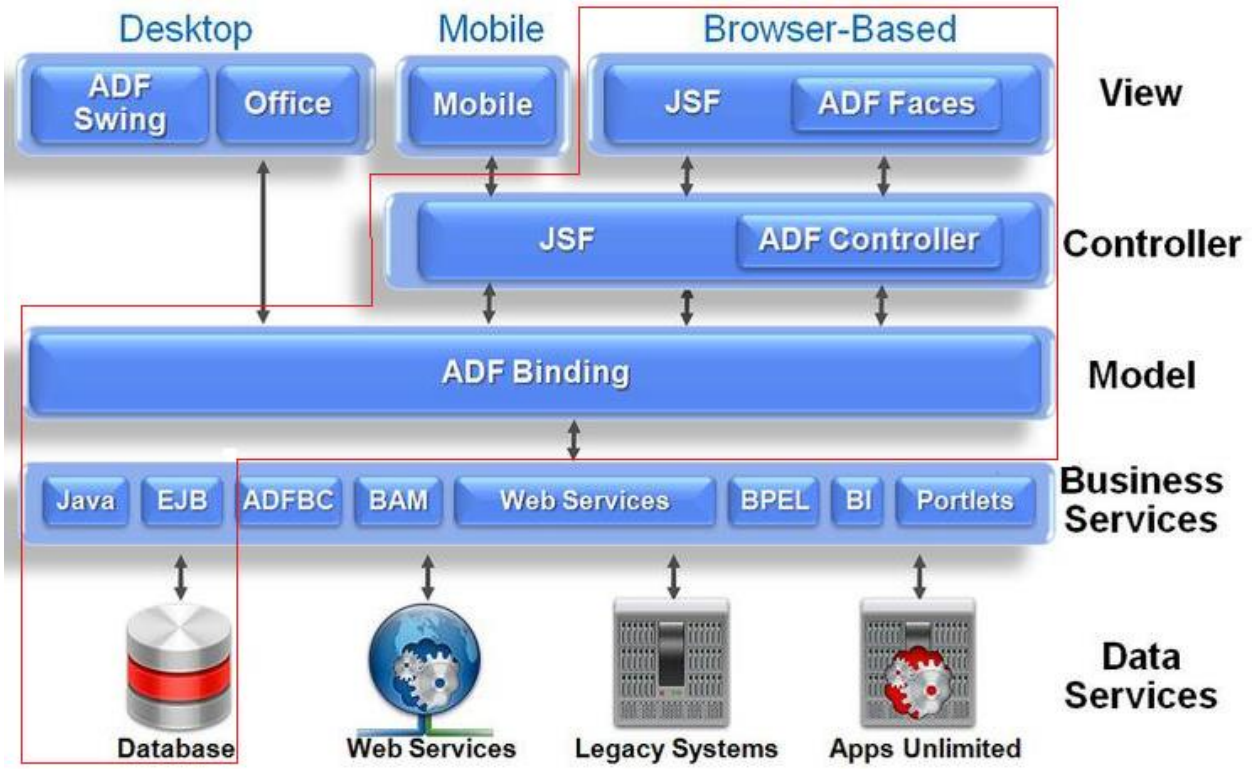

Рис.2.1. Архитектура ADF

На данной схеме выделена часть, необходимая для разработки системы генерации ключевой последовательности на основе биометрических данных пользователей с WEB-интерфейсом. На рисунке различаются четыре слоя.

Визуальный уровень представляет пользовательский интерфейс приложения. Визуальный уровень может быть основан на HTML, JSP, Java Server Faces (JSF), визуальных Java компонентах или XML для отрисовки пользовательского интерфейса. В нашем случае, в качестве UI (user interface) слоя web-приложения, выступают JSF страницы. ADF Faces, в свою очередь, предоставляет набор готовых UI компонентов.

Слой контроллера управляет потоком приложений и обрабатывает пользовательский ввод. Под ADF Controller понимаются Task Flows. Это специальные компоненты, декларативно описывающие последовательность действий, для выполнения какой-либо задачи.

ADF Binding (привязка данных) совместно с Data Controls (элементы управления данными) являют собой слой модели, содержащий в себе бизнес

логику, предоставляющий единый интерфейс, что в свою очередь позволяет не задумываться о том, какой внешний источник используется. Слой модели взаимодействует с объектами бизнес сервисов, которые используются другими слоями. Data Controls абстрагируют детали реализации бизнес сервисов, Data Bindings определяют атрибуты способы управления данными и связывают их с пользовательским интерфейсом, обеспечивая чистое разделение визуального уровня и уровня модели.

Слой Business Services обеспечивает доступ к данным из различных источников. Этот слой обеспечивает управление транзакциями, а также выполнение бизнес-логики.

Java Server Faces

При разработке Web-интерейса , интерфейс системы был создан на JSF страницах. JavaServer Faces (JSF) — это фреймворк для веб-приложений, написанный на Java. Он служит для того, чтобы облегчать разработку пользовательских интерфейсов для Java EE-приложений [1]. В отличие от прочих MVC - фреймворков, которые управляются запросами, подход JSF основывается на использовании компонентов. Состояние компонентов пользовательского интерфейса сохраняется когда пользователь запрашивает новую страницу и затем восстанавливается, если запрос повторяется. Для отображения данных обычно используется JSP, Facelets, но JSF можно приспособить и под другие технологии, например XUL.

Созданная быть гибкой, технология Jav aServer Faces усиливает существующие стандартные концепции пользовательского интерфейса (UI) и концепции Web-уровня без привязки разработчика к конкретному языку разметки, протоколу или клиентскому устройству [2]. Классы компонентов пользовательского интерфейса, поставляемые вместе с технологией Java Server Faces, содержат функциональность компонент, а не специфичное для клиента отображение.

Язык программирования Java

Т.к. разработка велась на Oracle ADF, программирование бизнес-логики и программной логики выполнялось на Java. Java — сильно типизированный объектно-ориентированный язык программирования, разработанный компанией Sun Microsystems (в последующем приобретённой компанией Oracle). Приложения Java обычно транслируются в специальный байт-код, поэтому они могут работать на любой компьютерной архитектуре, с помощью виртуальной Java-машины [2].

Язык разметки XML

Т.к. информационная система разрабатывается с web-интерфейсом, пользовательский интерфейс разработан на XML [29]. XML (англ. eXtensible Markup Language — расширяемый язык разметки) — рекомендованный Консорциумом Всемирной паутины язык разметки, фактически представляющий собой свод общих синтаксических правил. XML — текстовый формат, предназначенный для хранения структурированных данных (взамен существующих файлов баз данных), для обмена информацией между программами, а также для создания на его основе более специализированных языков разметки (например, XHTML). XML является упрощённым подмножеством языка SGML[39].

ДостоинстваXML:

− язык разметки, позволяющий стандартизировать вид файлов-данных, используемых компьютерными программами, в виде текста, понятного человеку;

− поддерживает юникод;

− в формате xml могут быть описаны такие структуры данных, как записи, списки и деревья;

− это самодокументируемый формат, который описывает структуру и имена полей так же как и значения полей;

− имеет строго определённый синтаксис и требования к анализу, что позволяет ему оставаться простым, эффективным и непротиворечивым. Одновременно с этим, разные разработчики не ограничены в выборе

экспрессивных методов (например, можно моделировать данные, помещая значения в параметры тегов или в тело тегов, можно использовать различные языки и нотации для именования тегов и т. д.);

− формат, основанный на международных стандартах;

− иерархическая структура xml подходит для описания практически любых типов документов, кроме аудио и видео мультимедийных потоков, растровых изображений, сетевых структур данных и двоичных данных;

− представляет собой простой текст, свободный от лицензирования и каких-либо ограничений;

− не зависит от платформы;

− является подмножеством sgml (который используется с 1986 года) [39]. Уже накоплен большой опыт работы с языком и созданы специализированные приложения;

− не накладывает требований на порядок расположения атрибутов в элементе и вложенных элементов разных типов, что существенно облегчает выполнение требований обратной совместимости;

− в отличие от бинарных форматов, xml содержит метаданные об именах, типах и классах описываемых объектов, по которым приложение может обработать документ неизвестной структуры (например, для динамического построения интерфейсов);

− имеет реализации парсеров для всех современных языков программирования;

− существует стандартный механизм преобразования xslt, реализации которого встроены в браузеры, операционные системы, web-серверы;

− поддерживается на низком аппаратном, микропрограммном и программном уровнях в современных аппаратных решениях[39].

PL/SQL Developer

При разработке пакетов процедур и функций использовалось такое средство разработки как PL/SQL Developer. PL/SQL Developer – интегрированная среда разработки для СУБД Oracle, которую создала и распространяет фирма Allround Automations, основанная в голландском городе Энсхеде [29]. Oracle SQL Developer изначально поддерживает работу с Oracle Database, существуют плагины, обеспечивающие подключение из среды к другим системам управления базами данных, в частности, реализован доступ к IBM DB2, Microsoft Access, Microsoft SQL Server, MySQL, Sybase ASE, Teradata Database. Oracle SQL Developer поставляется с OWA (Oracle Web Agent, или mod_plsql) — модулем расширения для веб-сервера Apache, помогающем в создании динамических веб-страниц с использованием PL/SQL с Oracle SQL Developer [4]. Особенности данного средства разработки:

- многофункциональный pl/sql редактор;
- встроенный отладчик;
- построитель запросов;
- sql окно;
- командное окно.

Язык PL/SQL

Для организации взаимодействия с СУБД Oracle использовался язык запросов PL/SQL. PL/SQL (Procedural Language / Structured Query Language) — язык программирования, процедурное расширение языка SQL, разработанное корпорацией Oracle [3].

PL/SQL даёт возможность использовать переменные, операторы, массивы, курсоры и исключения. PL/SQL позволяет разработчикам обрабатывать данные в реляционной базе, используя императивный стиль программирования. Операторы SQL могут быть легко вызваны

непосредственно из PL/SQL-процедуры, функции или из триггера (иногда с некоторыми ограничениями) [4].

В PL/SQL допускается включать готовые SQL-выражения непосредственно в код. В таком случае проверка выражения на корректность осуществляется уже при компиляции кода. Так, например, если используемая в запросе таблица не существует, то ошибка будет выдана уже на этапе компиляции.

Для большей гибкости часто статические запросы заменяются запросами, формируемыми динамически. Недостаток динамического SQL в том, что динамические запросы, разумеется, не могут быть проверены на этапе компиляции [29]. Если, например, используемой в запросе таблицы не существует, то при работе выполнении операции OPEN будет выброшено исключение. PL/SQL платформонезависим.

Model

При разработке приложения использовались такие компоненты ADF (Рис.2.3) как:

- сущности (entity objects);
- представления (view objects);
- ассоциации (association);
- ссылки (viewlink);
- модули приложений (application modules);
- тестеры бизнес-компонентов [23].

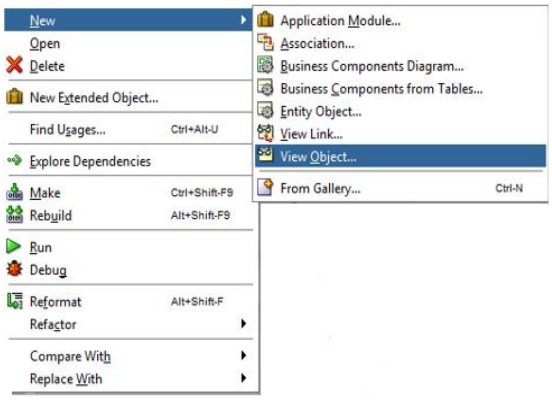

Рис.2.3 Создание новых компонентов ADF

При стандартном подходе Entity Object представляет собой строку в таблице базы данных (Рис.2.4) и упрощает изменение его данных путем обработки всех операций [23]. Он может инкапсулировать бизнес-логику, что бы гарантировать, применение индивидуальных бизнес- правил. Entity Object может быть основан на таблице, представлении или на специально написанном пакете.

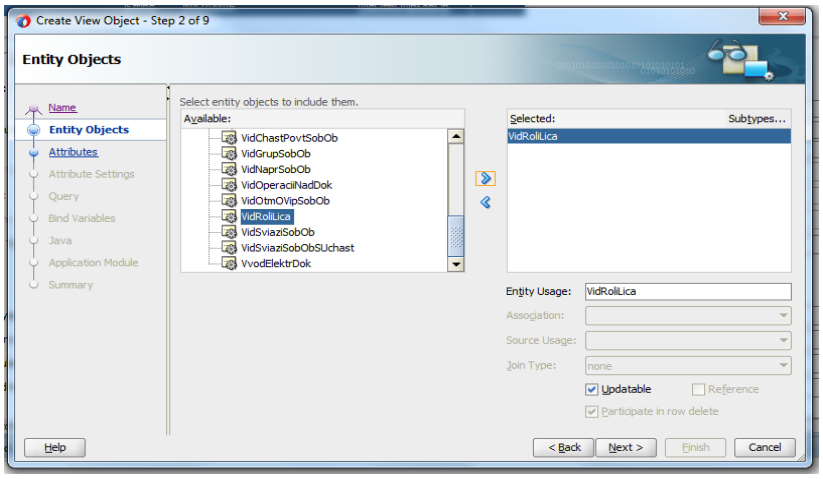

Рис.2.4 Создание Entity Objects

Представление можно понимать, как некоторую структуру данных, связанную с одной или несколькими сущностями. View Object представляет собой SQL-запрос. Для проведения DML-операций View Object должен быть построен на Entity Object, иначе этот View Object способен только отображать данные без возможности их редактирования [23]. Один VO может быть построен на нескольких EO в зависимости от сложности запроса,

таким образом при добавлении записи во VO ADF с помощью Entity Objects сам генерирует необходимые DML-операции для каждой таблицы.

Associations описывают отношения между Entity объектами, представляя собой внешние ключи таблиц. View Links – указывают на взаимосвязь между View Objects, могут базироваться на ассоциациях [4].

Ассоциации описывают отношения между сущностями и могут рассматриваться как реализации внешних ключей в таблицах базы данных. Модуль приложения является транзакционным компонентом, который клиент использует для работы с приложением. Он определяет обновляемую модель данных наряду с процедурами и функциями, относящиеся к логической единице работы, связанной с задачей конечного пользователя верхнего уровня. Модель данных (Data Model) состоит из экземпляров (instance) View Objects и View Links [29]. Один View Object может быть включен в модель данных несколько раз, при этом каждый отдельный экземпляр будет независим. Application Module с помощью класса DBTransaction управляет транзакциями (Рис.2.5).

Controller

В Controller обрабатываются потоки страниц веб- приложений. Для программирования логики переходов в приложении используется технология ADF Task Flow, где вы можете передать контроль приложений между различными видами деятельности, такими как переходы со страницы на страницу, методы на управляемых компонентами, потоками задач. Каждый Task Flow содержит часть навигационного графа приложения [24]. Узлы Task Flow являются действием, а узел активности (activity) представляет собой простую логическую операцию, как отображение страницы, выполнения логики приложения, или вызов другого Task Flow.

Существует два вида Task Flow (Рис.2.6):

- Unbounded Task Flow - Task Flow, страницы которого служат для авторизации пользователя и могут быть вызваны извне. Unbounded Task Flow состоит из всех видов деятельности и потоков управления в приложении, которые входит в не ограниченном потоке задач.
- Bounded Task Flow – специализированная форма потока задач, что в отличие от неограниченного потока задач, имеет единственную точку входа и ноль или более точек выхода [23].

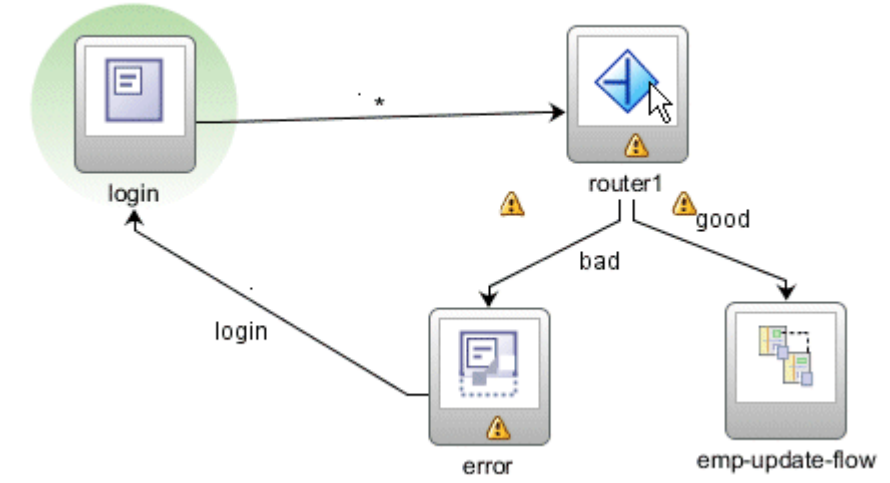

Рис.2.6 Пример Task Flow

View

Для создания страниц была использована технология JSF. Для связи визуальных компонентов с данными используются bindings (Рис.2.7а, 2.7б), которые прописываются в page definition этой страницы.

Рис.2.7а Представление страницы

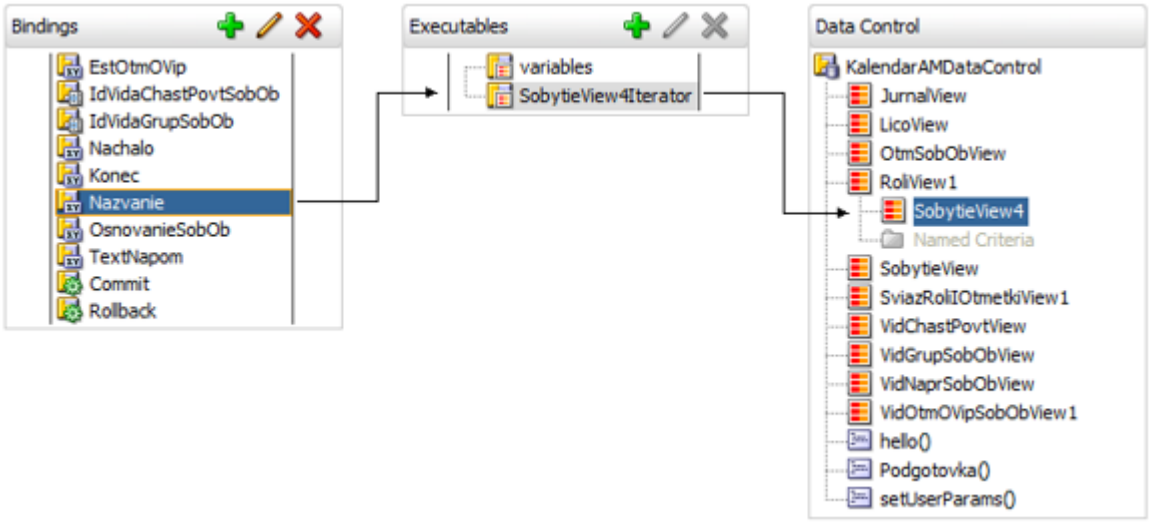

Рис.2.7б Bindings представленной страницы

Отличительная особенность Oracle ADF заключается в автоматическом конструировании форм на основе ранее созданных объектов просмотра (View Objects) [24]. Перенеся с помощью drag-n-drop ярко красный элемент - Iterator (Рис.2.8), можно получить на странице форму.

Oracle ADF поддерживает использование декларативного парадигмы программирования на протяжении всего процесса разработки, чтобы пользователи могли сосредоточиться на логике создания приложения без необходимости вдаваться в подробности реализации.

Процесс разработки веб-приложения на высоком уровне, как правило, включает в себя следующее [30]:

- Создание рабочего пространства приложения: с помощью мастера JDeveloper автоматически добавляет библиотеки и конфигурацию, необходимую для технологий, которые вы выбираете, и структуру вашего приложения в проекты с пакетами и каталогами.
- Моделирование объектов базы данных: вы можете создать автономную копию любой базы данных, а также использовать JDeveloper редактор и diagrammers изменять определения и схему обновления.

- Проектирование контроля приложений и навигации: вы можете использовать diagrammers визуально определить поток управления приложением и навигацию. JDeveloper создает базовый XML для вас.
- Определение общих ресурсов: вы можете использовать библиотеку ресурсов, который позволяет просматривать и использовать импортированные библиотеки, просто перетащив их в приложение.
- Создание бизнес-компоненты для доступа к данным: из таблиц базы данных, создавать объекты сущностей с помощью мастера или диалоговые окна.
- Реализация пользовательского интерфейса с помощью JSF: панель управления данных JDeveloper содержит представление объектов просмотра для вашего приложения. Создание пользовательского интерфейса так же просто, как перетаскивание объекта на странице и выбрав компонент пользовательского интерфейса, который вы хотите отобразить исходные данные. Для компонентов пользовательского интерфейса, которые не DataBound, вы используете палитры компонентов перетащить компоненты. JDeveloper создает весь код страницы для вас.
- Связывание компонентов пользовательского интерфейса для данных, используя слой ADF Model: При перетаскивании объекта из панели управления данных, JDeveloper автоматически создает привязки между страницей и моделями данных.
- Включение проверки и обработки ошибок: после того, как приложение будет создан использовать редакторы для добавления дополнительной проверки и определить обработку ошибок.
- Тестирование и отладка: JDeveloper включает в себя интегрированный сервер приложений, который позволяет полностью протестировать приложение без необходимости упаковать его и развернуть его [29].

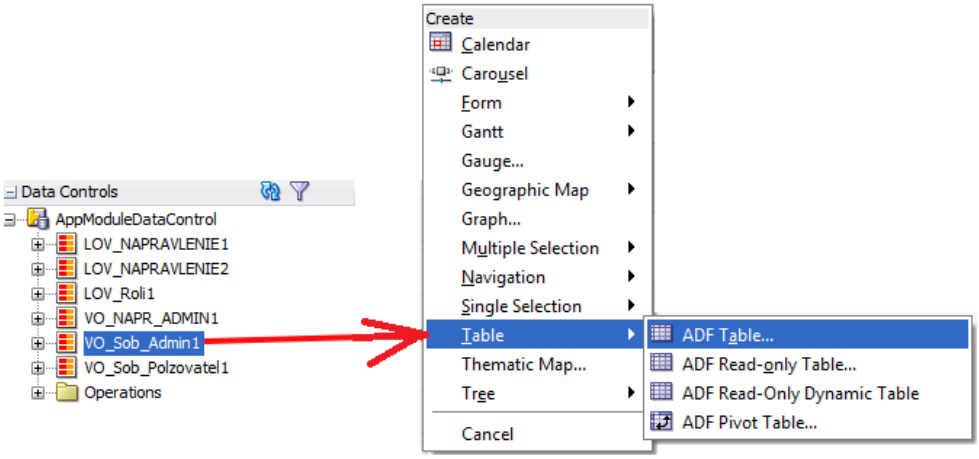

Рис.2.8 Создание элементов web-приложения

Data Controls предоставляет стандартный набор операций для управления данными. Эти операции выполняются с использованием функциональных возможностей Business Components.

2.2. Эксплуатируемая БД

Система разрабатывается на основе базы данных предприятия ООО «Газпром трансгаз Ухта».

Была изучена данная база данных Общества, выявлены данные для дальнейшего использования в проекте, спроектирована логическая модель (Рис.2.10) с учётом сущностей уже существующих в базе данных предприятия.

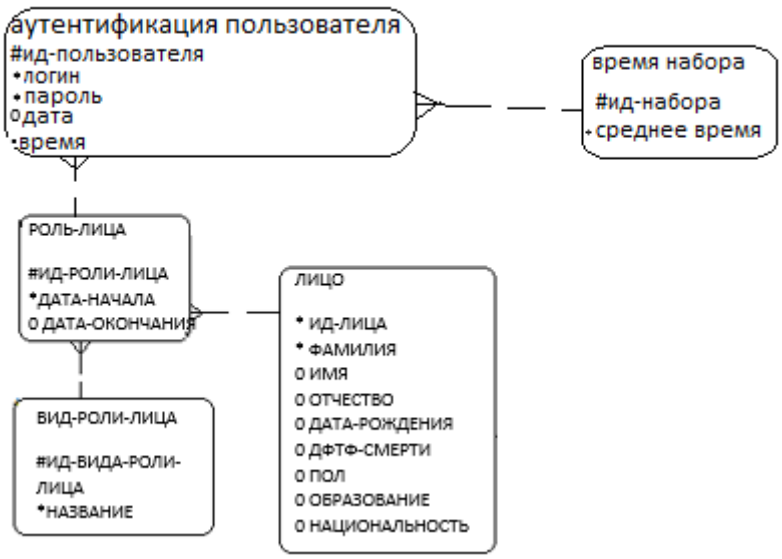

Рис. 2.10 Логическая модель

2.3. Генератор ключевых последовательностей на основе клавиатурного почерка пользователей

Как уже было сказано ранее, в основе методов аутентификации пользователей по клавиатурному почерку лежит уникальность динамики набора пользователем ключевой фразы. В системах аутентификации пользователей на основе данной биометрической характеристики могут использоваться следующие элементы [3]:

1. Время между нажатиями клавиш, соответствующих соседним буквам ключевой фразы;
2. Время удержания клавиши;
3. Сила давления на клавиши.

Последнее возможно только при наличии специальных клавиатур. Кроме названных характеристик большое значение для правильной идентификации (аутентификации) пользователя имеет так называемый рисунок почерка. Под этим термином подразумевается последовательность

значений, представляющих собой разность между двумя соседними временными интервалами — своего рода «производная» по почерку.

Рисунок почерка показывает относительные замедления или ускорения при наборе ключевой фразы или отдельных слов [6]. Характеристика эта также достаточно индивидуальна, что подтверждено рядом экспериментов. В настоящей работе в качестве биометрической характеристики используется время между нажатиями клавиш, соответствующих соседним буквам ключевой фразы.

Система биометрической аутентификации (идентификации) пользователей по клавиатурному почерку может быть реализована двумя способами [13]:

1) по вводу заранее определенной ключевой фразы
2) по набору «свободного» текста

В свою очередь система на основе ввода заранее определенной ключевой фразы может либо устанавливать единую фразу для всех пользователей, либо использовать секретные ключевые фразы для каждого пользователя. В последнем случае аутентификация является уже двухфакторной. При сохранении парольной фразы в секрете вероятность ошибки второго рода значительно снижается — по сути получается классическая парольная система с дополнительной биометрической аутентификацией.

В настоящей работе для экспериментов была выбрана единая фраза для всех пользователей, что позволяет более точно оценить именно биометрическую составляющую. В практических системах рекомендуется использовать различные фразы, тем более, что с технической точки зрения обе системы реализуются с одинаковыми затратами. Системы биометрической аутентификации пользователей по клавиатурному почерку реализуют три основных функции:

1) сбор информации;

2) обработка информации (механизмы сравнений с эталонными значениями);
3) принятие решений по результатам аутентификации.

Первая и третья функции реализуются алгоритмически одинаково (различие составляют некоторые коэффициенты), а вот вторая функция — обработка информации или механизмы сравнений с эталонными значениями — принципиально отличается. Сравнение вновь полученных значений времен удержаний клавиш с эталонными значениями в системах производятся по аддитивной характеристике. Одним из самых больших недостатков данной биометрической характеристики является ее большая изменчивость. У людей, которые постоянно работают с текстом (секретари, программисты и т.п.) и имеют уже большой опыт набора, клавиатурный почерк является достаточно стабильной характеристикой. В то же время люди, недавно начавшие обращаться с клавиатурой или пользующиеся ей нечасто, имеют крайне нестабильный почерк. В ходе многочисленных экспериментов было показано также, что, хотя в течение некоторого промежутка времени — от нескольких десятков минут до нескольких часов — почерк пользователя остается достаточно стабильным, однако, в течение всего рабочего дня величина математического ожидания временных интервалов между нажатиями соседних букв может значительно меняться.

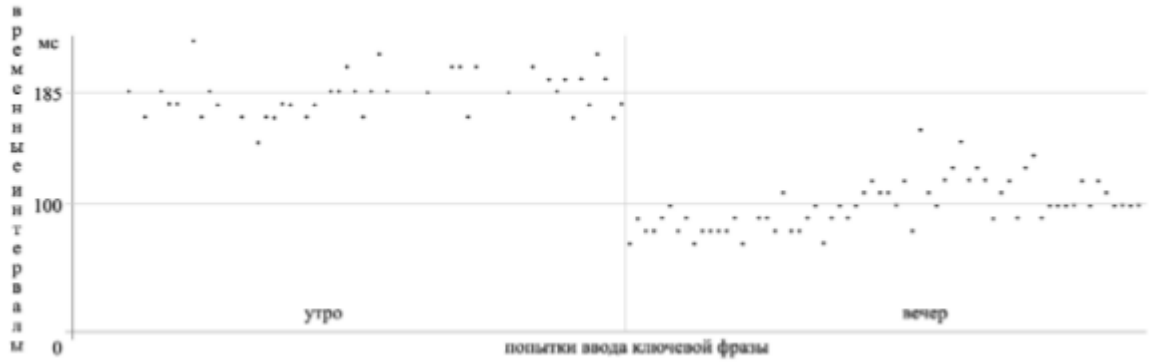

Рис.2.9 Разница математических ожиданий временных интервалов между нажатиями клавиш, соответствующих двум соседним буквам утром и вечером

Таким образом, для того, чтобы более качественно аутентифицировать пользователя, необходимо либо значительно увеличить доверительный интервал, что приведет к значительному увеличению вероятности ошибки второго рода, либо произвести многократную регистрацию пользователя в разное время суток и при аутентификации принимать во внимание время прохождения процедуры.

В качестве исходных данных предполагается, что временные интервалы между нажатиями соседних букв ключевой фразы, как правило, подчиняются нормальному закону распределения. Однако, данный факт необходимо проверять всякий раз при вводе ключевой фразы. Данное предположение позволяет применять при решении задачи аутентификации хорошо разработанные алгоритмы для построения доверительных интервалов, а также для исключения грубых ошибок в наблюдениях. Для решения задачи выработки ключевой строки для пользователя по его клавиатурному почерку необходимо предварительно в режиме регистрации выполнить следующие действия:

1) Выбрать ключевую фразу, желательно так, чтобы буквы этой фразы были равномерно расположены на клавиатуре.
2) Заставить пользователя несколько раз ввести эту фразу.
3) Исключить из набора грубые ошибки, которые могут возникнуть, например, когда внимание пользователя в процессе набора было отвлечено.
4) Найти средние значения временных интервалом между нажатиями соседних букв.
5) Определить доверительные интервалы и вычислить исправляющую способность кода.
6) На основании полученных данных выработать ключевую строку U и соответствующую ей «открытую» строку V.

2.4. Выбор ключевой фразы

Как уже было сказано выше, в настоящей работе используется единая ключевая фраза для всех пользователей с целью оценки качества именно биометрической составляющей системы. При выборе фразы мы руководствовались следующими, принципами:

1) Ключевая фраза должна быть осмысленным текстом, поскольку думать о таком тексте намного проще, нежели о случайном наборе символов — пользователь меньше отвлекается на обдумывание следующей буквы, а, следовательно, временные интервала в большей степени зависят именно от владения пользователем клавиатурой.
2) Буквы фразы должны быть достаточно равномерно распределены по клавиатуре. При неслучайном распределении букв фразы на клавиатуре появится достаточно сильная зависимость между ними, а зависимость временных интервалов от реального клавиатурного почерка значительно ослабнет.
3) Фраза не должна быть слишком короткой или слишком длинной. На короткой фразе очень трудно будет определить почерк пользователя, значительно возрастет вероятность ошибки второго рода. Слишком длинная фраза может привести к появлению таких проблем, как частые ошибки в наборе, изменения временных интервалов. Это повлечет за собой значительно увеличение вероятности ошибки первого рода. Более того, при неоднократном неуспешном вводе ключевой фразы, пользователь может начать нервничать, что приведет к искажению его почерка.

2.5. Алгоритм исключения грубых ошибок

Грубая погрешность, или промах – это погрешность результата отдельного измерения, входящего в ряд измерений, которая для данных условий резко отличается от остальных результатов этого ряда. Источником грубых погрешностей нередко бывают резкие изменения условий измерения и ошибки, допущенные оператором. К ним можно отнести:

- неправильный отсчет по шкале измерительного прибора, происходящий из-за неверного учета цены малых делений шкалы;
- неправильная запись результата наблюдений, значений отдельных мер использованного набора, например гирь;
- хаотические изменения параметров питающего СИ напряжения, например его амплитуды или частоты.

Корректная статистическая обработка выборки возможна только при ее однородности, т.е. в том случае, когда все ее члены принадлежат к одной и той же генеральной совокупности [3]. В противном случае обработка данных бессмысленна. "Чужие" отсчеты по своим значениям могут существенно не отличаться от "своих" отсчетов. Их можно обнаружить только по виду гистограмм или дифференциальных законов распределения.

Если «свои» и «чужие» отсчеты различаются по значениям, то их исключают из выборки (рис.2.10,а). Особую неприятность доставляют отсчеты, которые хотя и не входят в компактную группу основной массы отсчетов выборки, но и не удалены от нее на значительное расстояние, – так называемые предполагаемые промахи (рис.2.10,6) [4].

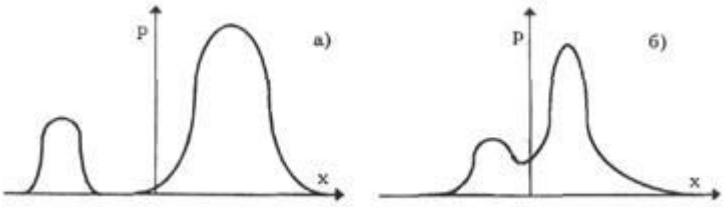

Рис.2.10 Проявление промахов на дифференциальном законе распределения вероятности

При однократных измерениях обнаружить промах не представляется возможным. Для уменьшения вероятности появления промахов измерения проводят два-три раза и за результат принимают среднее арифметическое полученных отсчетов. При многократных измерениях для обнаружения промахов используют статистические критерии, предварительно определив, какому виду распределения соответствует результат измерений.

Пусть X — случайна величина, которая подчиняется нормальному закону распределения и представляет собой временные интервалы между нажатиями клавиш на клавиатуре при наборе ключевой фразы.

X= $(x_1, x_2, \dots, x_n)$ — временные интервалы между вводом одних и тех же символов, полученные из n повторов (2.1) ввода ключевой фразы.

$$x_1, x_2, \dots, x_n (x_1 < x_2 < \dots < x_n) \qquad (2.1)$$

Критерий Диксона (2.2) определяется как

$$K_Д = \frac{x_n - x_{n-1}}{x_n - x_1} \qquad (2.2)$$

Если $K_Д$ больше критического значения Z_q (Таб. 2) при заданном уровне значимости q $(q = 1 - P)$, то результат x_j считают промахом [6].

Таблица 2. Критические значения критериев Диксона

n	Z_q при уровне значимости q, равном			
	0,1	0,05	0,05	0,01
4	0,68	0,76	0,85	0,89
5	0,56	0,64	0,73	0,78
6	0,48	0,56	0,64	0,70
7	0,43	0,51	0,6	0,64
8	0,4	0,47	0,54	0,59
9	0,37	0,44	0,51	0,56
10	0,35	0,41	0,48	0,53
12	0,32	0,38	0,44	0,48
14	0,29	0,35	0,41	0,45
16	0,28	0,33	0,39	0,43
18	0,26	0,31	0,37	0,41
20	0,26	0,30	0,36	0,39
25	0,23	0,28	0,33	0,36
30	0,22	0,26	0,31	0,34

После исключения грубых погрешностей переходят к решению второй задачи - исключению известных систематических погрешностей введением поправки или поправочного множителя в результат измерений.

Итак, поправка численно равна значению систематической погрешности, противоположна ей по знаку (2.3) и алгебраически суммируется с результатом измерения:

$$q = -\Delta_C \qquad (2.3)$$

Поправку определяют экспериментально по результатам поверки СИ или в результате специальных исследований, которые проводят для определения погрешностей.

2.6. Вывод

В данной главе был спроектирован и реализован прототип системы генерации ключевой последовательности на основе клавиатурного почерка, была выбрана архитектура системы, разработан прототип интерфейса системы, рассмотрена логическая модель базы данных, представлена схема работы приложения и пользовательский интерфейс.

Выбранная трехзвенная клиент-серверная архитектура позволяет перенести вычисления и логику обработки данных на сервер приложений. Сервер может быть представлен кластером серверов. Для работы с системой, пользователю необходим только браузер.

Была изучена существующая база данных. Разработаны специальные представления, которые упрощают проектирование приложения в Oracle ADF. Были разработаны прототипы интерфейса системы: формы ввода данных.

3. РЕЗУЛЬТАТЫ ЭСПЕРИМЕНТОВ

Исходя из указанных принципов была выбрана ключевая фраза — «парольная фраза». Во время эксперимента при прохождении процедуры регистрации пользователям предлагалось вводить ключевую фразу по несколько раз в различное время суток. Уровень владения клавиатурой большинства пользователей оценивался как «хороший», остальных — «отличный». Некоторые из пользователей владеют слепым десятипальцевым методом набора. Эксперимент показал (рис.3.1), что величина разброса временных интервалом между нажатиями одинаковых пар соседних букв фразы у пользователей, владеющих слепым десятипальцевым методом набора, и у пользователей, не владеющих этим методом, но имеющим достаточно большой опыт набора и регулярно пользующихся клавиатурой,

практически не отличаются друг от друга. Количество попыток ввода у разных пользователей варьировалось в пределах от 38 до 170.

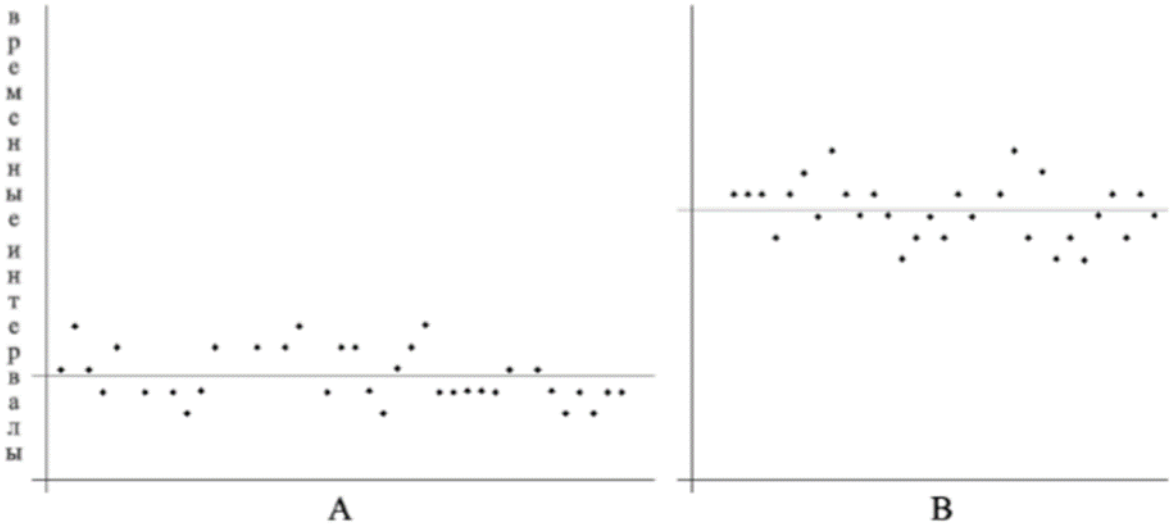

Рис. 3.1. Разброс временных интервалов между нажатиями двух одинаковых букв ключевой фразы у пользователя, владеющего слепым десятипальцевым методом набора (A), и пользователя, не владеющего данным методом (B)

В настоящей работе использовались выборки достаточно большой длины (от 30 до 170 элементов). На рисунках 3.2 и 3.3 показаны выборки до и после обработки алгоритмом исключения грубых ошибок.

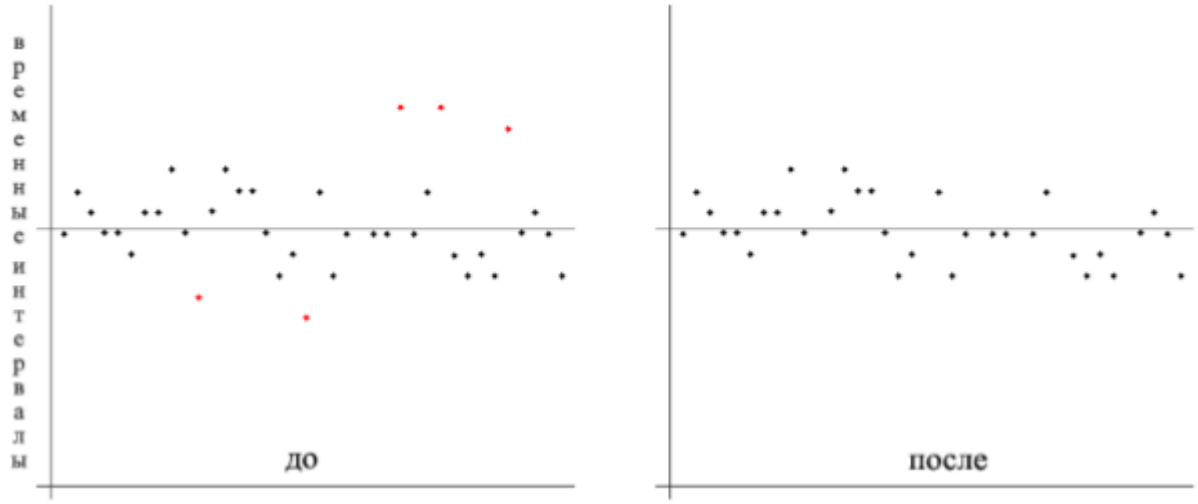

Рис. 3.2. Выборки временных интервалов между нажатиями двух соседних букв до и после обработки алгоритмом исключения грубых ошибок (пользователь 1)

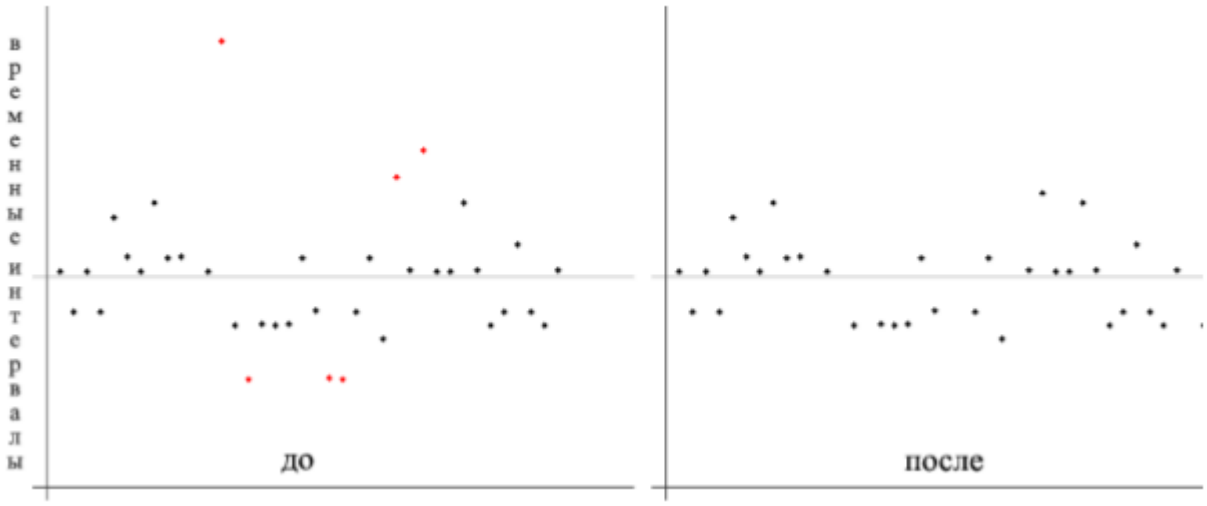

Рис. 3.3. Выборки временных интервалов между нажатиями двух соседних букв до и после обработки алгоритмом исключения грубых ошибок (пользователь 2)

После обработки алгоритмом исключения грубых ошибок можно определить математические ожидания временных задержек и получить тем самым усредненный цифровой образ биометрических данных пользователей. Таким образом, цифровой образ клавиатурного почерка пользователя при данном подходе представляет из себя вектор. Примеры таких образов можно увидеть на рисунках 3.4 и 3.5.

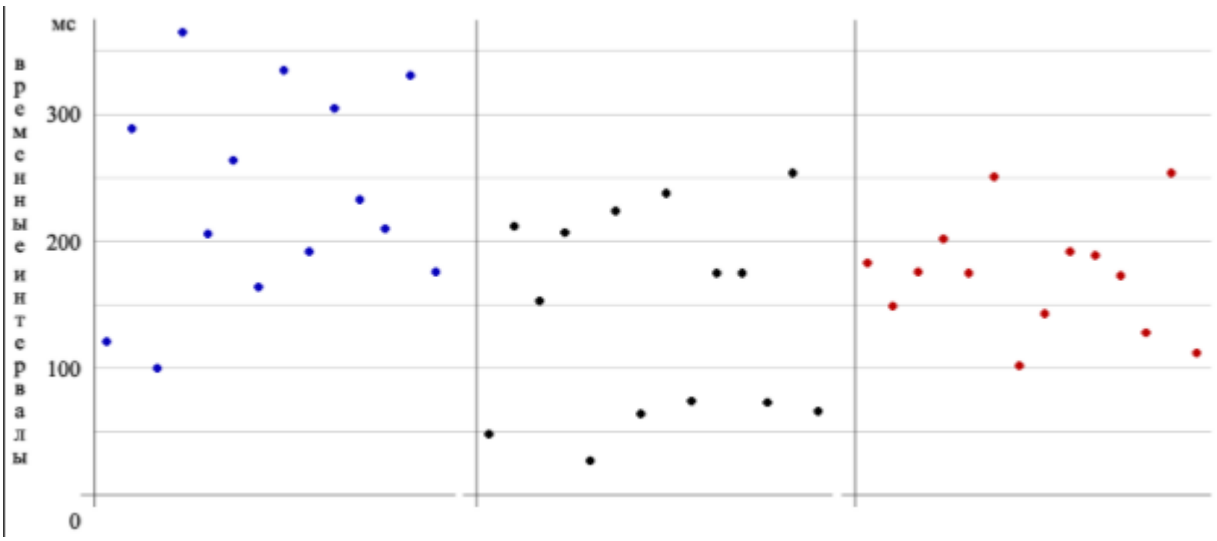

Рис. 3.4. Примеры цифровых образов клавиатурного почерка различных пользователей

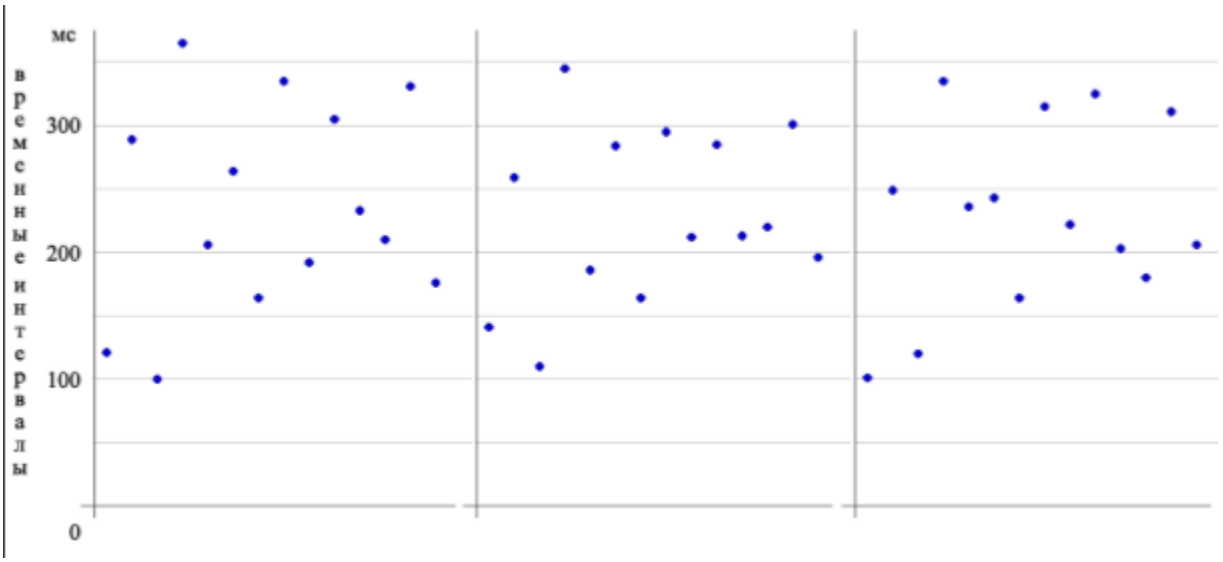

Рис. 3.5. Примеры цифровых образов клавиатурного почерка одного пользователя

Помимо средних значений временных интервалов следует также определить и средний разброс этих значений для каждого интервала отдельно. Стоит отметить, что при аутентификации каждый временной интервал оценивается независимо от остальных. Из полученных значений разбросов определяются доверительные интервалы для каждого отдельного временного интервала. Здесь существует две возможности — либо определить для каждого временного промежутка единый доверительный интервал для всех пользователей системы, либо для каждого пользователя

хранить персональные доверительные интервалы. Второй подход позволяет значительно снизить вероятность ошибки второго рода, при этом практически не изменив вероятности ошибки первого рода. В данной работе мы придерживаемся второго подхода. Примеры доверительных интервалов представлены на рисунке 3.6.

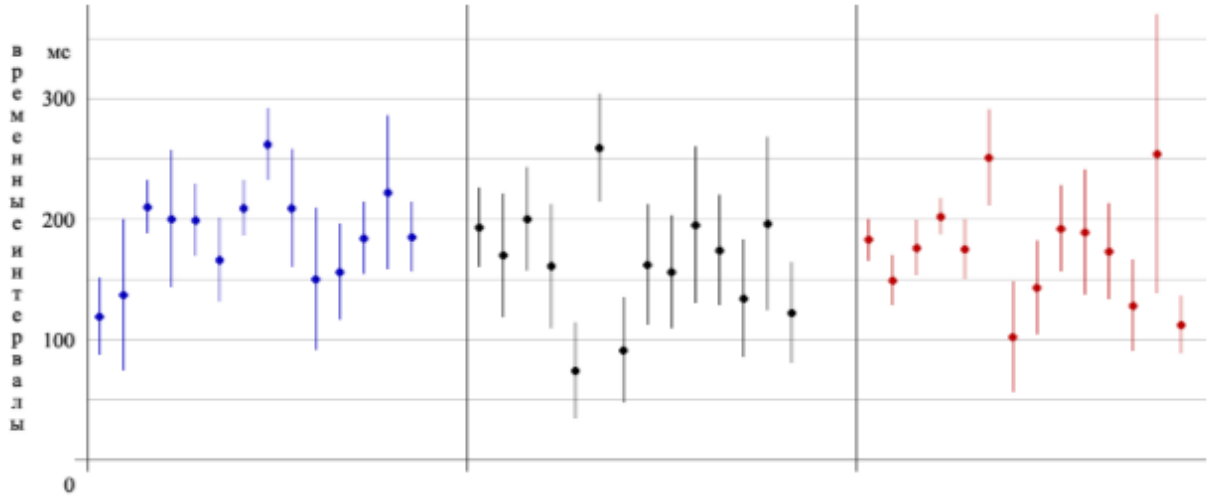

Рис. 3.6. Доверительные интервалы для элементов клавиатурного почерка различных пользователей

После нахождения цифрового образа и доверительного интервала при помощи генератора ключевых последовательностей необходимо выработать ключевую информацию. Как было сказано выше, для этого необходимо определить наиболее подходящую функцию расстояния между двумя образами и использовать помехоустойчивое кодирование для этого расстояния.

В случае клавиатурного почерка наиболее подходящей метрикой является обыкновенное евклидово расстояние между отдельными координатами цифрового образа. Таким образом, в данном случае необходим код, способный исправлять ошибки изменения числовой величины на некоторое значение.

3.1. Помехоустойчивое кодирование для евклидового расстояния

В общем случае задачу можно поставить таким образом: пусть есть информационный канал, допускающий искажения типа изменения числового значения менее, чем на R по абсолютной величине. Необходимо организовать безопасную передачу информации по такому каналу. Предлагается следующий алгоритм (Приложение 1): Пусть M — исходное сообщение, представляющее из себя беззнаковое целое число. Тогда кодированное сообщение C может быть вычислено следующим способом (3.1):

$$C = M * 2R + R = R * (2M + 1) \qquad (3.1)$$

В процессе передачи сообщения могут возникнуть помехи, которые приведут к тому, что адресат получит искаженное сообщение (3.2):

$$C' = C + d = M * 2R + R + d = R * (2M + 1) + d\ , \qquad (3.2)$$

где $|d| < R$

Получив кодированное сообщение C', абонент вычисляет синдром q (3.3) как остаток от деления C на R:

$$q = C'(mod\ R) \qquad (3.3)$$

Равенство синдрома q нулю означает то, что в процессе передачи сообщения искажений не произошло. В этом случае исходное сообщение вычисляется следующим образом (3.4):

$$M = \frac{C'-R}{2R} \qquad (3.4)$$

Если же при передачи возникли ошибки, то синдром q будет отличен от нуля. Легко видеть, что, если произошло увеличение значения, т.е. $d > 0$, то частное от деления полученного сообщения должно быть числом нечетным (3.5), т.е.

$$d > 0 \Rightarrow \left[\frac{C'}{R}\right] = 2M + 1 \text{ и } q = d, \quad M = \frac{C' - q - R}{2R} = \frac{C' - q}{2R} - \frac{1}{2} \qquad (3.5)$$

В противном случае частное окажется числом четным (3.6) и

$$d < 0 \Rightarrow \left[\frac{C'}{R}\right] = 2M \quad \text{и} \quad q = R + d, \quad M = \frac{C' - q}{2R} \qquad (3.6)$$

В общем случае исходное сообщение можно найти следующим образом (3.7):

$$M = \left[\frac{C' - q}{2R}\right] \qquad (3.7)$$

Описанный код может быть использован в генераторах ключевых последовательностей на основе таких биометрических характеристик, как клавиатурный почерк, геометрия руки, трехмерное изображение лица и др.

3.2. Модель системы аутентификации с использованием генератора ключевых последовательностей

Опираясь на все описанные выше процедуры, можно построить модель генерации ключевых последовательностей на основе клавиатурного почерка пользователей: в процессе регистрации пользователя в системе, собираются и обрабатываются несколько цифровых образов его клавиатурного почерка [4]. После исключения грубых ошибок из наблюдений происходит вычисление вектора средних значений временных задержек между нажатиями клавиш, соответствующих соседним буквам ключевой фразы M, а также вектора доверительных интервалов для каждого отдельного временного интервала D. При этом, если длина ключевой фразы равна n + 1, то длина векторов M и D (3.8) составит n.

$$M = (m_1, m_2, \dots, m_n) \qquad (3.8)$$

$$D = (d_1, d_2, \dots, d_n)$$

Вектор исправляющих способностей помехоустойчивого кода R = (r_1, r_2, …, r_n) однозначно определяется из вектора доверительных интервалов D. Обозначим через C_e (m, r) функцию кодирования, а через C_d (c, r) — функцию декодирования описанного выше помехоустойчивого кода. m, c и r — кодируемое и кодированное сообщения и исправляющая способность кода соответственно. Каждая координата вектора ключевых значений U = (u_1, u_2, …, u_n) генерируется случайным образом, а координаты вектора «открытых» строк V = (v_1, v_2, …, v_n) определяются как (3.9)

$$v_i = C_e(u_i, r_i) + m_i \qquad (3.9)$$

Собственно ключ в данной схеме может быть получен из вектора ключевых значений. В настоящей работе ключ вырабатывался как значение хэш-функции md5 от строки (3.10), полученное путем конкатенации координат вектора ключевых значений, т.е.

$$K = md5(u_1 u_2 \dots u_n) \qquad (3.10)$$

После проведения всех необходимых вычислений, пользователю выдается электронный ключ (смарт-карта или другой физический носитель), на котором записана следующая информация:

1) Вектор «открытых» значений V
2) Вектор значений исправляющих способностей R
3) Значение хэш-функции от ключа H(K)
4) Цифровую подпись, подтверждающую остальные данные — S (V, R, H(K))

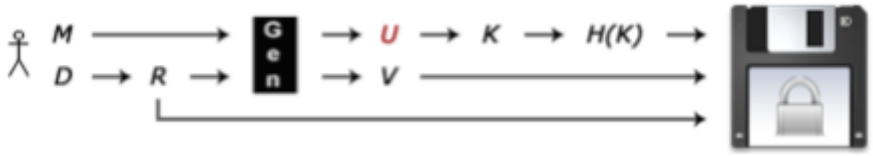

Рис.3.7. Регистрация пользователя в системе и создание электронного пропуска пользователя

При прохождении процедуры аутентификации пользователь предъявляет свой электронный ключ и вводит ключевую фразу. На основании полученных данных M', V и R система вычисляет ключевой вектор U', находит значение хэшфункции от этого вектора H(U') и сравнивает с хранимым на карте. Если значение совпадают, система делает вывод об успешной аутентификации, в противном случае происходит отказ в доступе, регистрация неудачной попытки и предложение пользователю снова ввести ключевую фразу.

3.3. Результат экспериментов

Всего для эксперимента были собраны более 1400 образцов клавиатурного почерка от 26 пользователей, уровень владения клавиатурой которых оценивался как «хороший» или «отличный». Некоторые из пользователей владели слепым десятипальцевым методом набора. Образцы почерков от большинства пользователей собирались в течение всего рабочего дня.

На основании полученных данных для каждого пользователя были рассчитаны векторы доверительных интервалов и, соответственно, векторы исправляющих способностей помехоустойчивого кода. Однако, величины исправляющих способностей в реализованной системе можно регулировать путем умножения на определенный коэффициент — это позволяет оценить

влияние чувствительности системы на величины ошибок первого и второго рода (Рис. 3.8).

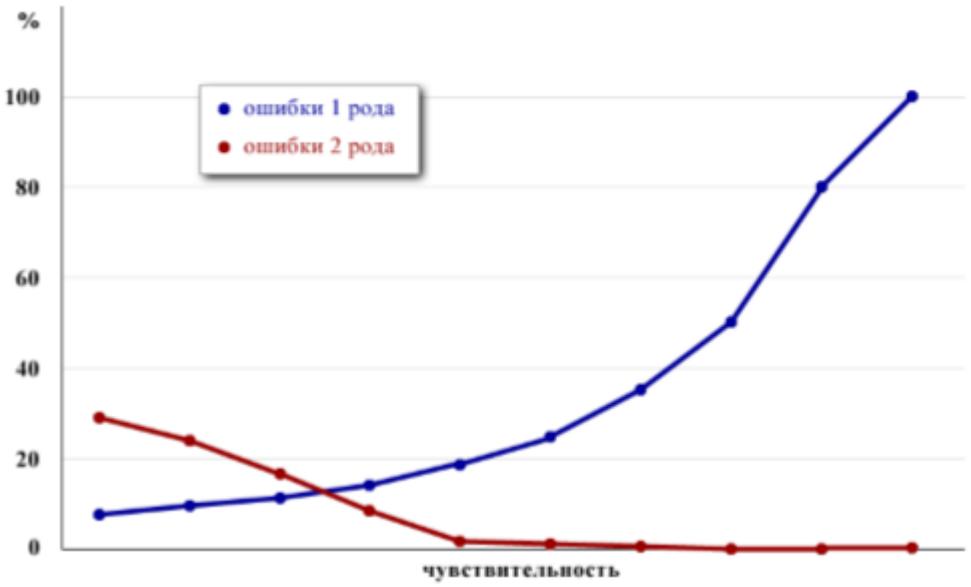

Рис. 3.8. Зависимость ошибок первого и второго рода от чувствительности системы

Как видно из результатов, клавиатурный почерк пользователей пригоден для использования в системах, где не требуется высокий уровень безопасности. Кроме того, его можно использовать в качестве дополнительной меры защиты вкупе с другими характеристиками.

3.4. Вывод

Таким образом, рассмотренный метод может применяться в комплексе с другими механизмами для решения задач различных задач. Во-первых, это может быть повышение защищенности информационных ресурсов в организациях с высокими требованиями к защите информации. Во-вторых, благодаря анализу психофизического состояния, данные методы могут применяться в организациях, в которых необходимо обеспечить высокий уровень концентрации внимания сотрудников во время работы. Основным

достоинством метода является отсутствие необходимости использования дополнительного оборудования, что позволяет создавать гибкие настраиваемые подсистемы аутентификации и мониторинга действий оператора информационной системы. Однако, несмотря на свои достоинства, данная область мало изучена, и, на мой взгляд, имеет огромный потенциал.

ЗАКЛЮЧЕНИЕ

В работе рассмотрен общий теоретический подход, позволяющий генерировать ключевую информацию на основе нечетких данных. Этот подход лег в основу разработки и реализации двух генераций ключевых последовательностей с использованием различных биометрических характеристик пользователей. Для реализации генерации ключевой последовательности на основе клавиатурного почерка был разработан код, позволяющий исправлять ошибки типа изменения значения по абсолютной величине.

В процессе экспериментов было выявлено множество особенностей данной биометрических характеристик человека. В частности, было обнаружено, что иногда в процессе набора ключевой фразы внимание пользователя было отвлечено, что приводило к значительным изменениям временных интервалов по сравнению с их среднеквадратичными отклонениями. Для того, чтобы подобные значения не учитывались в дальнейших расчетах, был разработан алгоритм исключения грубых ошибок.

Алгоритм позволяет значительно повысить качество усредненного шаблона при регистрации пользователя. Генерация ключевой последовательности на основе клавиатурного почерка был реализован в виде программного комплекса. Для экспериментов были собраны 1400 образцов клавиатурного почерка от 26 реальных пользователей. Уровень пользования клавиатурой всех пользователей оценивался как «хороший» или «отличный», некоторые из них владели слепым десятипальцевым набором. В результате экспериментов была также получена зависимость величин ошибок I и II рода от чувствительности системы. Было обнаружено, что уровень ошибки II рода может достигать 0 при вероятности ошибки I рода 50%. Это позволяет говорить о том, что клавиатурный почерк может быть использован в качестве основы для построения ключевой последовательности.

Генераторы ключевых последовательностей на основе биометрических данных пользователей не только позволяют избавится от большинства

недостатков, присущих классическим системам биометрической аутентификации, но и дают возможность применения биометрических технологий в криптографии.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1) Брэдли Д. Браун. Oracle Database. Создание Web-приложений. – М.: Издательство «Лори», 2010. – 722 с.
2) Мартин К. Соломон, Нирва Мориссо-Леруа, Джули Басу «Oracle. Программирование на языке Java» -М.: Издательство «Лори», 2010. – 342с.
3) Стивен Фейерштейн, Билл Прибыл «Oracle PL/SQL для профессионалов. 6-е издание.» -Спб.: Питер, 2015. – 682с.
4) Аруп Нанда, Стивен Фейерштейн «Oracle PL/SQL для администраторов баз данных» -М.: Символ-Плюс, 2008. – 105с.
5) Харин Е. А. Генерация ключевой информации на основе биометрических данных пользователей. //Труды XLV международной научной студенческой конференции. — Новосибирск: НГУ, 2007. — С. 181—187
6) Корнюшин П. Н., Гончаров С. М., Харин Е. А. Создание системы аутентификации на основе клавиатурного почерка пользователей с использованием процедуры генерации ключевых последовательностей из нечетких данных. //Сборник материалов IV Международной научнопрактической конференции «Интеллектуальные технологии в образовании, экономике и управлении — 2007». — Воронеж: ВИЭСУ, 2007.
7) Защита информации. Техника защиты информации. Требования к средствам высоконадежной биометрической аутентификации: ГОСТ Р 52633-2006 от 01.04.2007 — 24 с.
8) Гузик В.Ф., Десятерик М.Н. Биометрический метод аутентификации пользователя. // «Известия ТРТУ» №2(16), ТРТУ, Таганрог, 2000.
9) Задорожный В. Обзор биометрических технологий // Защита информации. Конфидент. –2003. – № 5.
10) Фор А. Восприятие и распознавание образов. – М.: Машиностроение, 1989.

11) Синева И.С., Баталов А.Э., Фенчук М.М. Повышение помехоустойчивости кодов CRC при помощи предварительного генетического кодирования моторизованного источника сообщений. Фундаментальные проблемы радиоэлектронного приборостроения/ Материалы Международной научно-технической конференции “INTERMATIC- 2013”, 2-6 декабря 2013 г., Москва/ Под ред. Академика РАН А.С. Сигова – М.: Энергоатомиздат, часть 4,2013. С.65-70.

12) Золотарев В.В., Овечкин Г.В. Помехоустойчивое кодирование. Методы и алгоритмы: Справочник / Под. ред. чл.-кор. РАН Ю.Б. Зубарева. –М.: Горячая линия –Телеком,2004

13) Зюко А.Г., Фалько А.И., Панфилов И.П., Банкет В.Л., Иващенко П.В. Помехоустойчивость и эффективность систем передачи информации / Под ред. Зюко А.Г. М.: Радио и связь, 1985. - 272 с.

14) Савинов, А.Н. Анализ решения проблем возникновения ошибок первого и второго рода в системах распознавания клавиатурного почерка / А.Н. Савинов, В.И. Иванов // Вестник Волжского университета имени В.Н. Татищева: науч.-теор. журнал. Серия «Информатика». - Тольятти: Волжский университет им. В.Н. Татищева, 2011. - Вып. 18. - С. 115-119.

15) Савинов, А.Н. Использование биометрической системы распознавания клавиатурного почерка при выполнении технологических требований обеспечения безопасности ключевой системы / А.Н. Савинов // Инновационные ресурсы и национальная безопасность в эпоху глобальных трансформаций. Пятнадцатые Вавиловские чтения»; постоянно действующая Всерос. междисципл. науч. конф. с междунар. участием: в 2 ч. / редкол.: В.П. Шалаев и др. - Ч. 2. - Йошкар-Ола: МарГТУ, 2012. - С. 282-284.

16) Круглов В.В. Нечеткая лоика и искуственные нейронные сети: учеб. пособие. М.: Физматлит, 2001. -224с.

17) Сулавко А.Е., Еременко А.В., Самотуга А.Е. Искючение искаженных биометрических даных из эталона субъекта в системах идентификации// Информационные технологии и вычислительные системы. 2013. №3. С.96-101.

18) X. Boyen. Reusable cryptographic fuzzy extractors. //Report 2004/358, http://eprint.iacr.org. — 2004.

19) Иванов А.И. Биометрическая идентификация личности по динамике подсознательных движений. – Пенза: ПГУ, 2000.

20) Задорожный В. Обзор биометрических технологий // Защита информации. Конфидент. –2003. – № 5.

21) Петров А.А. Компьютерная безопасность. Криптографические методы защиты. М.: ДМК Пресс, 2008. -448 с.:ил.

22) Лебеденко Ю.И. Биометрические системы безопасности. Тула: изд-во ТулГУ, 2012. 160с.

23) Himanshu Marathe “Oracle Fusion Middleware Developing Applications with Oracle ADF Desktop Integration 12*c* (12.1.2)” - USA: McGraw-Hill, 2013. – 324стр. , ил.(Серия "Osborne ORACLE Press Series")

24) Jenny Gelhausen, Penny Avril, Willie Hardie «Oracle Database 12c Product Family» - USA: Redwood Shores, 2014. – 20 с.

25) Y. Dodis, L. Reyzin, and A. Smith. Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy Data. //Proceedings from Advances in Cryptology EuroCrypt. — 2004.

26) U. Uludag, S. Pankanti, P. S., and A. Jain. Biometric crptosystems: issues and challenges. //Proceedings of the IEEE 92. — 2004. — pp. 948—960.

27) http://www.intuit.ru/studies/courses/3580/822/lecture/30592?page=2

28) http://www.intuit.ru/studies/courses/76/76/lecture/27940?page=2

29) https://ru.wikipedia.org/wiki

30) http://denisorlovmusic.ru/articles/java/JDeveloper_11g_R2_Developing_Rich_Web_Applications_With_Oracle_ADF_Rus/Part2.htm
31) http://allthingsoracle.com/oracle-database-12c-new-features-part-i/
32) https://people.eecs.berkeley.edu/~dawnsong/
33) http://www.cs.cmu.edu/~pvenable/publications.html
34) http://infoprotect.net/varia/klaviaturnyiy_pocherk
35) http://www.securitylab.ru/blog/personal/aguryanov/29985.php
36) http://tekhnosfera.com/razrabotka-poligaussovogo-algoritma-autentifikatsii-polzovateley-v-telekommunikatsionnyh-sistemah-i-setyah-po-klaviaturno
37) http://tekhnosfera.com/metody-modeli-i-algoritmy-raspoznavaniya-klaviaturnogo-pocherka-v-klyuchevyh-sistemah#ixzz4j9oGuVZP
38) http://allthingsoracle.com/oracle-database-12c-new-features-part-3/
39) http://www.oracle.com/ru/database/overview/index.html
40) https://djbook.ru/ch12s03.html

ПРИЛОЖЕНИЕ 1

Алгоритм помехоустойчивого кодирования для евклидового расстояния