

**Ахметов Б.С., Иванов А.И., Фунтиков В.А.,
Безяев А.В., Малыгина Е.А.**

**ТЕХНОЛОГИЯ ИСПОЛЬЗОВАНИЯ
БОЛЬШИХ НЕЙРОННЫХ СЕТЕЙ
ДЛЯ ПРЕОБРАЗОВАНИЯ НЕЧЕТКИХ
БИОМЕТРИЧЕСКИХ ДАННЫХ
В КОД КЛЮЧА ДОСТУПА**

Монография

Алматы 2014 г.

УДК 004.4
ББК 32.973.26-02
Т38

Рецензенты:

д.т.н., профессор Тукаев У.А., д.т.н., профессор Джурунтаев Д.З.,
д.т. н., профессор Утепбергенов И.Т.

Ахметов Б.С., Иванов А.И., Фунтиков В.А., Безяев А.В., Малыгина Е.А.

Технология использования больших нейронных сетей для преобразования нечетких биометрических данных в код ключа доступа: монография.

– Алматы: ТОО «Издательство LEM», 2014 – 144 с.

ISBN 978-601-239-327-9

В коллективной монографии сотрудников Казахского национального технического университета им. К.И. Сатпаева, Пензенского научно-исследовательского электротехнического института и Пензенского государственного университета изложена технология использования больших нейронных сетей для преобразования нечетких биометрических данных в код ключа доступа.

Рассмотрены процедуры обучения и тестирования высоконадежных механизмов биометрико-нейросетевой защиты информации и персональных данных, созданных в соответствии с ГОСТ Р 52633.0-2006 «Защита информации. Техника защиты информации. Требования к средствам высоконадежной биометрической аутентификации».

В работе приведены примеры использования нейросетевой биометрии при анонимной аутентификации личности пациента медицинского учреждения и дистанционной аутентификации личности в сети Интернет.

Рассмотрена классическая защита персональных данных хешированием паролей и биометрических кодов доступа, защита данных нейросетевого контейнера простым шифрованием и защита нейронов блочным шифрованием с растущей длиной ключа.

Монография подготовлена в рамках выполнения комплексного проекта по теме: «Разработка и подготовка производства телекоммуникационного оборудования, разработка программного сетевого, прикладного и специального обеспечения для создания цифровых сетей связи с персонализированным доступом» в соответствии с Постановлением Правительства РФ №218 от 09.04.2010 г. (Договор № 13G25.31.0039 от 07.09.2010 г.) и научного проекта по теме: №753. МОН.ГФ.13.13 «Исследование вариантов реализации и разработка действующего лабораторного образца ON-LINE системы биометрического обесличивания электронных историй болезней для медицинского учреждения».

Для специалистов по искусственным нейронным сетям, защите информации в системах медицинского электронного документооборота, созданию электронных карт и паспортов с автоматической биометрико-нейросетевой аутентификации личности.

Печатается по решению Ученого Совета КазНТУ имени К.И.Сатпаева

ISBN 78-601-239-327-9

© Б.С. Ахметов, А.И. Иванов, А.В.Фунтиков

© В.А. Безяев, Е.А. Малыгина

© КазНТУ им. К.И. Сатпаева, 2014

СОДЕРЖАНИЕ

Термины и определения	6
Введение	11
<i>1 Алгоритмы обучения нейросетевых преобразователей биометрия-код.....</i>	<i>15</i>
1.1. Особенности автоматов обучения нейросетевых преобразователей «биометрия-код», реализующих критерий максимума различия собственной энтропии кодов «Свой» и кодов «Чужие».....	15
1.2. Оценка «нано» и «пико» вероятностей появления редких событий биометрической аутентификации на малых тестовых выборках.....	23
1.3. Сверхбыстрый алгоритм оценки энтропии существенно зависимых дискретных выходных состояний нейросетевых преобразователей биометрия-код	30
1.3.1 Ограниченные возможности классического метода вычисления энтропии	30
1.3.2 Сверхбыстрые алгоритмы вычисления энтропии в пространстве кодов Хэмминга.....	34
1.3.3 Связь статистик энтропии Хэмминга и классической энтропии	37
1.4. Связь энтропии выходных состояний нейросетевых преобразователей биометрия-код с коэффициентами парной корреляции	42
1.4.1 Принятые упрощения	43
1.4.2 Модификация схемы Бернулли при проведении численного эксперимента	46
1.5. Генетический алгоритм итерационного обращения матриц нейросетевых функционалов путем минимизации расстояний Хэмминга	48
1.6. Энтропийно-корреляционная оценка хэширующих свойств нейросетевого преобразователя «биометрия-код доступа»	53

1.7. Аппроксимация биномиального зависимого закона композициями нормального, равномерного, арксинусного распределений значений	62
1.8. Моделирование длинных биометрических кодов, воспроизводящих корреляционные связи выходных данных нейросетевого преобразователя.....	68
1.8.1 Аналитическая связь коэффициентов равной корреляции и параметра связывающей матрицы.....	69
1.8.2 Эффект размывания параметра «а=0» дополнительным псевдослучайным сигналом.....	70
1.8.3 Эффект размывания параметра «а» и «-а» дополнительным псевдослучайным сигналом.....	72
1.9. Учет «тяжелых хвостов» ненормального закона распределения биометрических параметров все «Чужие» при настройке нелинейного элемента нейрона с несколькими дискретными состояниями.....	73
1.9.1 Настойка параметров нелинейной части нейрона	74
1.9.2 Регулирование формы нелинейного элемента под заданное из вне значение фрагмента кода «Свой»	77
 <i>2. Алгоритмы тестирования, практическая реализация и защита механизмов биометрико-нейросетевой аутентификации личности</i>	 80
2.1. Быстрые алгоритмы тестирования средств высоко-надежной биометрической аутентификации личности.....	80
2.2. Особенности тестирования средств биометрико-нейросетевой аутентификации личности на малых тестовых базах.....	92
2.2.1 Формирование тестовой базы биометрических образов «Чужой».....	93
2.2.2 Общая схема тестирования с использованием малой тестовой базы.....	93
2.2.3 Оценка стойкости средства биометрико-нейросетевой аутентификации к атакам подбора при использовании малых тестовых баз.....	95
2.3. Биометрико-нейросетевой классификатор папиллярных рисунков отпечатков пальцев личности.....	96

2.3.1 Синтез нейросетевого классификатора по типу узора отпечатка пальца.....	97
2.3.2 Выбор анализируемых данных и обучение нейронов	101
2.3.3 Результаты обучения.....	102
2.4. Биометрико-нейросетевое обезличивание персональных данных личности при обращении в медицинское учреждение	104
2.5. Использование нейросетевой биометрической аутентификации личности в сети интернет.....	108
2.6. Классический вариант защиты персональных данных хешированием паролей и биометрических кодов доступа ...	116
2.7. Высокостойкая защита данных нейросетевого онтейнера простым шифрованием	117
2.8. Защита нейронов блочным шифрованием с растущей длиной ключа	119
2.9. Сравнительный анализ биометрических и специальных защитных свойств «нечетких экстракторов» (технология США, Канады и Евросоюза) с нейросетевыми преобразователями биометрия-код, выполненными по требованиям пакета российских стандартов ГОСТ Р 52633.xx	120
2.9.1 Обеспечение высокого уровня защитных свойств нейросетевыми технологиями и технологиями «нечетких экстракторов»	120
2.9.2 Обеспечение высокого уровня биометрических свойств нейросетевыми технологиями и технологиями «нечетких экстракторов»	130
2.9.3 Обеспечение стандартами нейросетевых технологий и технологий «нечетких экстракторов»	132
Заключение.....	135
Список использованных источников.....	137

ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ

ГОСТ Р 52633.0-2006

Биометрическая аутентификация – аутентификация пользователя, осуществляемая путем предъявления им своего биометрического образа.

Биометрические данные – данные с выходов первичных измерительных преобразователей физических величин, совокупность которых образует биометрический образ конкретного человека.

Биометрическая идентификация – преобразование совокупности примеров биометрических образов человека, позволяющее описать их стационарную и случайную составляющие, например, в виде математического ожидания и дисперсий контролируемых параметров или, например, в виде параметров обученной сети искусственных нейронов.

Дискретный биометрический параметр – биометрический параметр, множество допустимых значений которого описывается конечным множеством.

Непрерывный биометрический параметр – биометрический параметр, множество допустимых значений которого описывается континуумом значений.

Вектор биометрических параметров (ВБП) – нумерованный набор биометрических параметров, являющихся отображением одного биометрического образа, имеющих одну и ту же интерпретацию и формат представления. Один биометрический образ может описываться множеством биометрических параметров.

Биометрический образ – образ человека, полученный с выходов первичных измерительных преобразователей физических величин, подвергающийся далее масштабированию и иной первичной обработке с целью извлечения из него контролируемых биометрических параметров человека.

Биометрический образ «Свой» – биометрический образ легального пользователя.

Биометрический образ «Чужой» – биометрический образ

злоумышленника, пытающегося преодолеть биометрическую защиту.

Биометрические образы «Все чужие» – совокупность множества биометрических образов «Чужой», верно отражающая статистику попыток подбора злоумышленниками образов «Свой».

Биометрический пример – совокупность биометрических данных, полученная с выхода первичного преобразователя при однократном предъявлении человеком своего биометрического образа.

Биометрический механизм – механизм преобразования физического биометрического образа в вектор биометрических параметров или код ключа (пароля). Биометрический механизм является функционально неполной частью системы защиты или средства защиты информации.

Биометрические параметры – параметры, полученные после предварительной обработки биометрических данных. Параметрами могут являться, например, коэффициенты Фурье кривых колебаний пера при воспроизведении человеком рукописного пароля.

Вероятность ошибки первого рода – вероятность ошибочного отказа «Своему» пользователю в биометрической аутентификации.

Вероятность ошибки второго рода – вероятность ошибочной аутентификации «Чужого» как «Своего» (ошибочная аутентификация).

Высоконадежная биометрическая аутентификация – биометрическая аутентификация с приемлемой вероятностью ошибок первого рода и гарантированно малой вероятностью ошибок второго рода, сопоставимой по своему значению с вероятностью случайного подбора кода неизвестного криптографического ключа при малом числе попыток подбора.

Динамический биометрический образ – биометрический образ, изменяемый человеком по своему желанию, например, рукописный образ слова-пароля.

Статический биометрический образ – образ, данный человеку от рождения, неизменяемый по воле человека, например, рисунок отпечатка пальца.

Тайный биометрический образ – биометрический образ, сохраняемый пользователем в тайне. Для сохранения в тайне динамического биометрического образа необходимо сохранить в тайне пароль, порождающий его; для сохранения в тайне статического биометрического образа необходимо обеспечить анонимность пользователя.

Открытый биометрический образ – биометрический образ человека, общедоступный для наблюдения. Обычно открытые биометрические образы являются статическими, однако и динамические биометрические образы могут быть открытыми, например, рукописный автограф человека.

Средство биометрической аутентификации – средство, способное принимать аутентификационное решение с определенной вероятностью и имеющее в своем составе: биометрические механизмы преобразования биометрических данных в вектора биометрических параметров большой размерности, преобразователь биометрия-код ключа (пароля), механизм криптографической аутентификации.

Механизм биометрической аутентификации – функционально неполный фрагмент средства биометрической аутентификации, преобразующий биометрические данные, но не способный принимать аутентификационные решения из-за отсутствия механизма криптографической аутентификации.

Нейросетевой преобразователь биометрия-код – заранее обученная искусственная нейронная сеть с большим числом входов и выходов, преобразующая частично случайный вектор входных биометрических параметров «Свой» в однозначный код криптографического ключа (длинного пароля) и преобразующая любой иной случайный вектор входных данных в случайный выходной код.

Нейрон – сумматор нескольких биометрических параметров, на выходе которого подключена нелинейная пороговая функция с двумя выходными состояниями «0» и «1».

Нейронная сеть – множество нейронов, объединенных в сеть путем соединения входов нейронов одного слоя с выходами нейронов другого слоя, причем, входы нейронов первого слоя являются входами всей нейронной сети, а выходы нейронов последнего слоя являются выходами нейронной сети.

Преобразователь биометрия-код – преобразователь, способный преобразовывать вектор нечетких, неоднозначных биометрических параметров «Свой» в четкий однозначный код ключа (пароля). Преобразователь, откликающийся случайным выходным кодом на воздействие случайного входного вектора, не принадлежащего множеству образов «Свой».

Исходный ключ – последовательность, получаемая на выходе нейросетевого преобразователя биометрия-код при выполнении нейросетевого преобразования.

Сформированный ключ – последовательность, разряды которой зависят от разрядов всех исходных биометрических параметров, использованных для его получения.

Критерий Хемминга – мера сравнения двух последовательностей одинаковой длины, вычисляемая путем подсчета различающихся разрядов.

Отклик процедуры индикации – значение, выдаваемое процедурой жесткой или мягкой индикации в качестве результата.

Обучение биометрического средства – обучение биометрического средства аутентифицировать человека с заданными вероятностями ошибок первого и второго рода на одном или нескольких примерах биометрических образов «Свой».

Донор биометрических образов – лицо, добровольно участвующее в формировании базы естественных биометрических образов путем предоставления своих собственных биометрических образов для преобразования их в цифровую форму.

Естественный биометрический образ – биометрический образ донора, полученный в виде выходных биометрических данных первичного преобразователя и представленный одним или несколькими примерами.

Синтетический биометрический образ – биометрический образ, полученный путем имитационного моделирования естественных биометрических образов и представленный одним или несколькими примерами.

Злоумышленник – лицо, заинтересованное в получении возможности несанкционированного доступа к конфиденциальной информации, представляющей промышленную и коммерческую тайну, предпринимающее попытку такого доступа или совершившее его.

Показатель стабильности биометрического параметра – характеристика i -го контролируемого биометрического параметра, вычисляемая по формуле:

$$c(v_i) = \frac{\sigma_{\text{Чужой}}(v_i)}{\sigma_{\text{Свой}}(v_i)} \quad (\text{П.1})$$

где $\sigma_{\text{Чужой}}(v_i)$ – стандартное отклонение i -го биометрического параметра множества образов «Чужой»;

$\sigma_{\text{Свой}}(v_i)$ – стандартное отклонение i -го биометрического параметра множества образов «Свой».

Показатель стабильности разряда исходного ключа – показатель, изменяющийся в пределах от 0.0 (разряд абсолютно нестабилен) до 1.0 (разряд полностью стабилен), вычисляемый по следующей формуле:

$$\omega_i = 2 \cdot |0,5 - P_{0,i}| = 2 \cdot |0,5 - P_{1,i}| \quad (\text{П.2})$$

где $P_{0,i}$ – вероятность появления состояния «0» в контролируемом i -том разряде;

$P_{1,i}$ – вероятность появления состояния «1» в контролируемом i -том разряде.

Для образов «Свой» разряды выходного кода обычно стабильны, то есть для них выполняется условие $0,5 \leq \omega_i \leq 1,0$.

Для образов «Чужой» большинство разрядов выходного кода нестабильны, то есть для них выполняется условие $0 \leq \omega_i \leq 0,5$.

ВВЕДЕНИЕ

Сегодня все мы без исключения являемся свидетелями процессов активной информатизации нашего общества. Без мобильной связи, Интернета, компьютеров современный человек обойтись уже не может. К сожалению, эти процессы приводят не только к положительным изменениям нашей жизни. Наряду с повышением производительности труда, повышением качества товаров и услуг информатизация порождает ряд серьезных проблем. Одними из таких проблем является проблема высоконадежной идентификации и аутентификации личности пользователя инфокоммуникационных систем и проблема защиты его персональных данных в открытом информационном пространстве.

Одно из наиболее перспективных направлений защиты подобных систем от несанкционированных воздействий – биометрические методы распознавания пользователя по его неповторимым биометрическим данным.

Анализ зарубежных разработок в области биометрии показал, что существующие сегодня биометрические средства идентификации личности, разработанные по стандарту BioAPI и использующие обычные шаблоны биометрических образов не способны обеспечить как конфиденциальность хранимой биометрической информации, так и высоконадежную идентификацию личности пользователя.

Большинство из существующих технологий используют статические образы, неизменяемые по воле человека и данные ему от рождения: рисунки сетчатки и радужной оболочки глаза, рисунки отпечатков пальцев, форму лица, ушей и т.д. Данные технологии были разработаны для «полицейских» функций – поиска террористов, преступников, незаконных мигрантов.

После трагических сентябрьских событий, произошедших в 2001 г. в США, данные технологии биометрической идентификации получили свое применение в паспортно-визовых документах нового поколения. При этом контроль над работой средства локальной идентификации должен в обязательном порядке осуществляться специальным персоналом. Для осуществления дистанционной идентификации данные средства не пригодны.

Сегодняшние простейшие биометрические устройства должны рано или поздно уступить свое место более сложным и более интеллектуальным биометрическим устройствам следующего поколения, позволяющим осуществлять высоконадежную автоматическую как локальную, так и дистанционную аутентификацию личности пользователя. Как нельзя лучше для этого подходят механизмы с использованием искусственных нейронных сетей большого и сверхбольшого размера. Долгое время сдерживающим фактором их использования было отсутствие быстрых алгоритмов их обучения.

Разработка быстрых алгоритмов обучения и тестирования механизмов обучения больших искусственных нейронных сетей и их применение позволило решить задачу увеличения входного массива данных для обучения нейронной сети и тем самым значительно повысить качество выходного решения, при минимальных временных затратах на сам процесс обучения искусственной нейронной сети.

Остановимся более подробно на особенностях исследуемой области:

- во-первых высоконадежная нейросетевая биометрия, в отличие от зарубежной биометрии, только начинает показывать о своих преимуществах. Технология целиком и полностью базируется на российских национальных стандартах серии ГОСТ Р 52633 хх. Базовым стандартом этой серии является ГОСТ Р 52633.0-2006 «Защита информации. Техника защиты информации. Требования к средствам высоконадежной биометрической аутентификации». Примечательно, что первый Российский национальный стандарт по защите биометрических данных вступил в действие на 5 лет раньше, соответствующих ему международных аналогов.

- во-вторых, если использовать нейросетевой преобразователь биометрия-код доступа к информационному ресурсу, то биометрический шаблон человека исчезает. Вместо шаблона биометрического образа человека биометрические устройства следующего поколения хранят в биометрических контейнерах только таблицы связей и таблицы значений весовых коэффициентов обученной искусственной нейронной сети. Однозначно восстановить по этим данным исходный биометрический образ человека и его код

доступа невозможно. Этим самым обеспечивается выполнение требований ФЗ РФ №152 «О персональных данных».

Свойство нейросетевых технологий обеспечивать конфиденциальность (анонимность) биометрических данных распознаваемых людей и их кодов доступа имеет первостепенное значение. Только обеспечивая конфиденциальность, анонимность, обезличенность биометрии граждан и их кодов доступа мы можем построить действительно цифровое демократическое общество. В настоящее время зарубежные биометрические средства не гарантируют полную сохранность биометрического шаблона человека;

- в-третьих, биометрический контейнер с данными можно безопасно хранить на удаленных серверах (облачных серверах) и передавать по незащищенным каналам передачи данных;

- в-четвертых, высоконадежная нейросетевая биометрия позволяет полностью автоматизировать (исключить полностью человеческий фактор) в процессе локальной или дистанционной аутентификации личности в том числе при регистрации и аутентификации на облачных серверах и службах. Другими словами искусственный интеллект нейросетевого подсознания роботов по своим возможностям, оказывается сопоставим с естественным интеллектом людей-экспертов, что подтверждено пока только для задач биометрико-нейросетевой аутентификации;

- в-пятых, высоконадежная нейросетевая биометрия исключает передачу биометрического идентификатора другим лицам и организацию атаки подмены шаблона в процессе аутентификации. При использовании в качестве биометрического идентификатора голосовых фраз-паролей или рукописных паролей, учитывающих динамику и временные параметры их ввода, высоконадежная нейросетевая биометрия позволяет учитывать психофизиологическое состояние пользователя, и не позволяет злоумышленнику силой или при помощи психотропных средств заставить пользователя пройти аутентификацию.

- в-шестых, самостоятельная смена биометрического идентификатора, путем переучивания нейросетевого преобразователя биометрия-код с индикацией качества обучения;

- в-седьмых, сравнительная простота использования средства аутентификации. Практический опыт использования разработан-

ных макетов для обучения и формирования тестовых баз студентами вуза показал быстроту освоения, дружелюбность интерфейса и легкость пользования средствами аутентификации.

Представляемая монография является совместным трудом казахстанских и российских ученых в области исследований высоконадежной биометрико-нейросетевой аутентификации. В рамках этих исследований выполняется ряд совместных НИР, готовится к изданию учебное пособие и курс лабораторных работ по высоконадежной биометрико-нейросетевой аутентификации, которые не имеют мировых аналогов. Опубликован ряд статей в ведущих зарубежных научных журналах. Сделаны доклады на международных конференциях и симпозиумах в Германии, Индии, Испании, Казахстане, России, Сингапуре, Южной Кореи. Готовятся и защищаются кандидатские диссертации в России и диссертации PhD доктора в Казахстане по исследуемой тематике.

Монография может быть полезна для специалистов, занимающихся приложением искусственных нейронных сетей при решении задач защиты информации в платежных системах, системах электронного документооборота, в системах электронных паспортов и удостоверений личности с автоматической биометрической идентификацией.

1 АЛГОРИТМЫ ОБУЧЕНИЯ НЕЙРОСЕТЕВЫХ ПРЕОБРАЗОВАТЕЛЕЙ БИОМЕТРИЯ-КОД

1.1 Особенности автоматов обучения нейросетевых преобразователей «биометрия-код», реализующих критерий максимума различия собственной энтропии кодов «Свой» и кодов «Чужие»

В настоящее время в рамках технического комитета по стандартизации РФ «Защита информации» (ТК №362) активно ведутся работы по созданию пакета национальных стандартов, регламентирующих требования к средствам биометрической аутентификации личности, обеспечивающих конфиденциальность, анонимность и обезличенность массового оборота персональных биометрических данных. Одним из наиболее важных стандартов является ГОСТ Р 52633.5-2011 [1], регламентирующий автоматическое обучение нейросетевых преобразователей «биометрия-код». Значимость этого стандарта обусловлена необходимостью регламентации процедуры обучения нейросетевого преобразователя «биометрия-код», построенной на знании кода «Свой» и использовании нескольких примеров биометрического образа «Свой», компрометация которых приводит к тому, что биометрическое средство аутентификации перестает быть высоконадежным и характеризуется только остаточной стойкостью к атакам подбора скомпрометированного биометрического образа. В связи с этой угрозой обучение не должно проводиться в присутствии третьих лиц. По требованиям базового национального стандарта ГОСТ Р 52633.0-2006 [2] нейронная сеть преобразователя «биометрия-код» должна обучаться автоматически, при этом после обучения исходные данные (код «Свой» и примеры образа «Свой») должны уничтожаться. В данной статье рассматриваются некоторые особенности автоматов обучения нейросетевых преобразователей «биометрия-код».

По требованиям ГОСТ Р 52633.0-2006 [2] обучение нейросетевого преобразователя «биометрия-код» (ПБК) должно выполняться таким образом, чтобы каждый i -ый пример образа «Свой»

после его преобразования в вектор контролируемых биометрических параметров $\vec{v}_i$ давал на выходах ПБК одинаковые коды $\vec{c}$:

$$\begin{bmatrix} \text{ПБК} \\ N \times n \end{bmatrix} \cdot \vec{v}_i = \vec{c} \quad (1.1)$$

где $\begin{bmatrix} \text{ПБК} \\ N \times n \end{bmatrix}$ – матрица нелинейных нейросетевых функциона-

лов ПБК размерности $(N \times n)$, осуществляющих нелинейную нейросетевую обработку вектора из N непрерывных биометрических параметров $\vec{v}_i$ по их преобразованию в код $\vec{c}$ – вектор, состоящий из n разрядов.

В том случае, когда на входы нейросетевого ПБК будут подаваться случайные вектора биометрических образов «Чужой» – $\vec{\xi}$, то каждый i -ый вектор ПБК должен откликаться случайным кодом $\vec{z}$:

$$\begin{bmatrix} \text{ПБК} \\ N \times n \end{bmatrix} \cdot \vec{\xi}_i = \vec{z}_i \quad (1.2)$$

Фактически нейросетевой ПБК осуществляет сжатие практически до нуля собственной энтропии биометрического образа «Свой», а для биометрических образов «Чужой» нейросетевой ПБК существенно усиливает энтропию до величины, которая примерно в 5 раз меньше ее предельного значения. Для того чтобы одновременно выполнить условия (1.1) и (1.2), необходимо строить специальные автоматы обучения нейронов нейронных сетей ПБК. На рис. 1.1 приведена блок-схема работы автомата обучения одного нейрона.

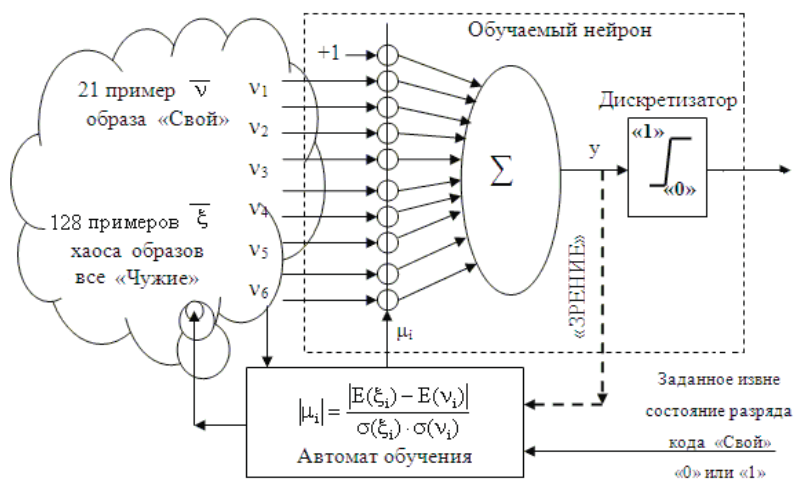


Рис. 1.1. Блок-схема автоматического обучения одного нейрона нейросети ПБК

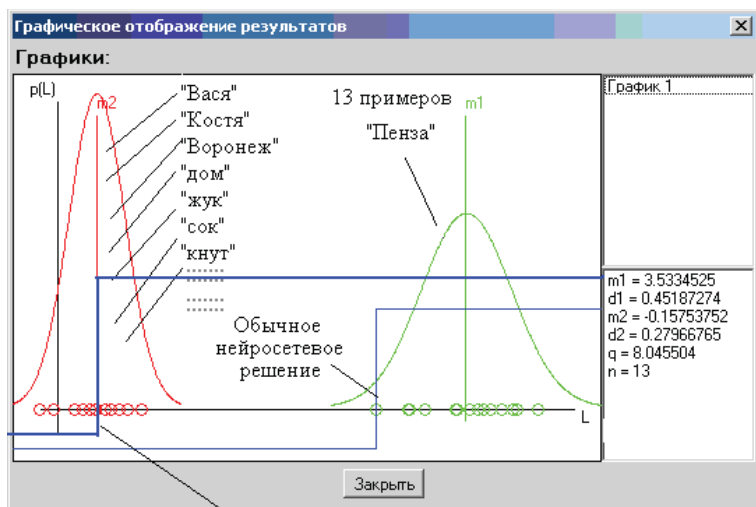
Одной из основных особенностей алгоритма обучения, рекомендуемого ГОСТ Р 52633.5, является то, что нелинейная функция нейрона осуществляет переключение состояния «0» в состояние «1» в точке математического ожидания откликов образов все «Чужие» – $E(y(\xi))$. Это нейросетевое решение показано на рис. 1.2 утолщенной линией. В отличие от обычных нейросетевых решений (тонкая линия на рис. 1.2) оно позволяет сделать атаку «поиска близнецов» неэффективной. Атака заключается в поиске и использовании данных похожего человека в базе персональных данных (например, удостоверений личности с электронными фотографиями) людей соответствующего пола и возраста, при этом выделяются тип лица человека и характерные его признаки. Если размеры базы позволяют, то похожего человека можно найти достаточно быстро. Далее самозванец может пользоваться чужим электронным удостоверением личности, при этом, чтобы научиться похоже расписываться, самозванец затратит лишь несколько дней.

Из рис. 1.2 видно, что обычное нейросетевое решение дает состояние «1» для образов «Свой» и состояние «0» для образов

«Чужие». ГОСТ Р 52633.5 требует отказаться от такого подхода к решению задачи и добиваться равновероятного состояния «0» и «1» для множества образов «Все Чужие».

Казалось бы, что происходит катастрофическое ухудшение качества принимаемых нейронной сетью решений, так как каждый разряд выходного кода ПБК ошибается с вероятностью 0.5 (при этом возникает ошибка второго рода, когда биометрический образ «Чужой» принимается за образ «Свой»). На самом деле катастрофы не происходит из-за выполнения условия (1.2), по которому выходные коды ПБК должны быть слабо коррелированными. При полном отсутствии корреляционных связей выходных кодов длиной 256 бит имеет место фантастически малое значение вероятностей ошибок второго рода:

$$P_2 = 0.5^{256} \approx 10^{-77} \quad (1.3)$$



Нейросетевое решение, исключающее атаку "поиска близнецов"

Рис. 1.2. Два варианта нейросетевых решений (обычное решение – тонкая линия) и решение по ГОСТ Р 52633.5 (утолщенная линия)

К сожалению, полного уничтожения корреляционных связей между разрядами выходных кодов $\vec{z}_i$ добиться не удастся. В свя-

зи с этим ГОСТ Р 52633.0-2006 [2] допускает наличие парных корреляционных связей между разрядами выходных кодов ПБК (допустимо наличие корреляционных связей, математическое ожидание модулей которых не превышает 0.15). В предельном случае, когда $E(|r|) = 0.15$, в показателе выражения (1.3) появляется поправочный коэффициент $\beta \approx 0.1$:

$$P_2 = 0.5^{\beta 256} \approx 0.5^{0.1 \cdot 256} \approx 0.5^{26} \approx 10^{-8} \quad (1.4)$$

Наличие корреляционных связей в выходных кодах «Чужой» ПБК, несомненно, ослабляет биометрическую защиту, однако ее стойкость к атакам подбора на уровне 10^8 попыток оказывается вполне приемлемой для многих приложений.

Еще одной важной особенностью автоматов обучения в соответствии с ГОСТ Р 52633.5-2011 [1] является то, что в них не рекомендуется использовать итерационные алгоритмы обучения искусственной нейронной сети ПБК. На сегодняшний день известно несколько сотен итерационных алгоритмов обучения нейронных сетей, однако все они теряют устойчивость при увеличении числа входов у нейронов или при снижении качества входных биометрических данных. Пример утраты устойчивости на основе оценки параметра качества обучения Q для сумматора нейрона, определяемого по аналогии с параметром качества обучения каждого нейрона, введенного в стандарте ГОСТ Р 52633.1 [3], как

$$Q = \frac{|E_{\text{Чужой}}(y) - E_{\text{Свой}}(y)|}{\sigma_{\text{Чужой}}(y) + \sigma_{\text{Свой}}(y)} \quad (1.5)$$

где E и σ математическое ожидание и среднеквадратическое отклонения сигнала y сумматора, приведен на рисунке 1.3.

Кривые, приведенные на рисунке 1.3, построены по результатам обучения нейрона с 15 и 16 входами на одних и тех же 13 примерах образа «Свой» и 128 примерах разных образов «Чужие». Обучение велось в среде имитационного моделирования «НейроУчитель».

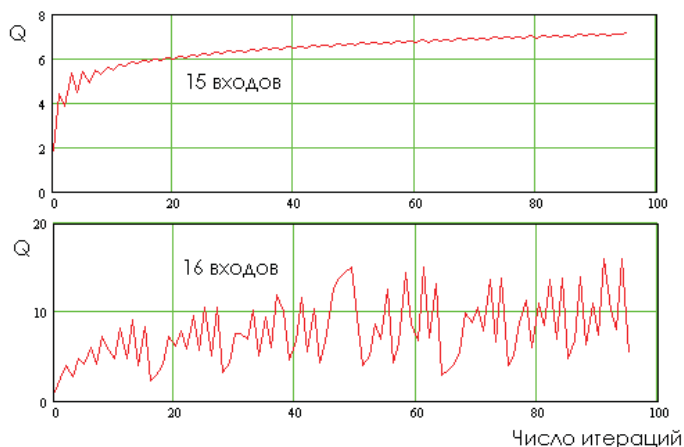


Рис. 1.3. Утрата устойчивости автоматом итерационного обучения при увеличении числа входов, обучаемого нейрона

Из данных рис. 1.3 видно, что при 15 входах у обучаемого нейрона качество его выходных решений практически монотонно увеличивается на всех 100 итерациях. При увеличении числа входов у обучаемого нейрона до 16 происходит утрата монотонности итерационного обучения (нижняя часть рисунка 1.3). Фактически процесс обучения теряет устойчивость, и, соответственно, у разработчиков автоматов обучения нейронов возникают значительные трудности.

Обычно для итерационных алгоритмов обучения число примеров в обучающей выборке должно примерно совпадать с числом входов у обучаемого нейрона. Если это условие не выполняется, то автомат итерационного обучения начинает оптимизировать ошибки представления биометрического образа «Свой» конечной выборкой примеров этого образа. Повысить устойчивость автомата итерационного обучения (вернуться к монотонному росту качества обучения на каждой следующей итерации) всегда можно через увеличение примеров образа «Свой» в обучающей выборке.

Еще одной причиной утраты устойчивости процедуры обучения является снижение качества входных биометрических данных (показатель качества биометрических данных определен в

ГОСТ Р 52633.1-2009 [3]). При плохом качестве входных данных срыв устойчивости происходит при меньшем числе входов у обучаемого нейрона. Эта ситуация отображена на рис. 1.4. В верхней части рисунка 1.4 отображена ситуация, соответствующая процессам, показанным на рисунке 1.3. При хорошем входном качестве 13-ти примеров образа «Свой» итерационное обучение является устойчивым для 1, 2, ..., 13 входов обучаемого нейрона.

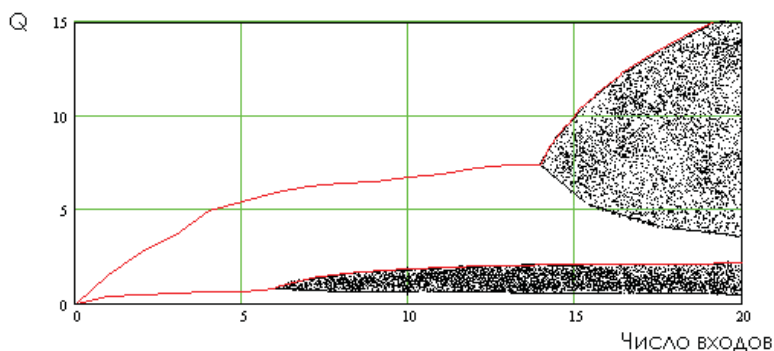


Рис. 1.4. Связь положения точки утраты устойчивости обучения (точки бифуркации) с числом входов нейрона и качеством входных (выходных) биометрических данных

Однако уже при 14 входах происходит утрата устойчивости обучения. При 15 входах нейрона ложные экстремумы функционала качества уже хорошо наблюдаемы (показаны в верхней части рисунка 1.3). В нижней части рисунка 1.4 отображена ситуация обучения нейрона на 13 примерах образа «Свой» с биометрическими данными плохого качества. Из рисунка следует, что качество обучения нейрона на плохих данных оказывается хуже. При этом срыв устойчивости обучения происходит уже при 6 входах обучаемого нейрона.

В связи с отмеченной выше закономерностью в соответствии с ГОСТ Р 52633.5-2011 [1] требуется отказаться от использования итерационных алгоритмов обучения (на рисунка 1.1 обратная связь с выхода сумматора нейрона на вход автомата обучения показана пунктиром).

По новому стандарту рекомендуется использовать прямое вычисление модулей весовых коэффициентов через математические ожидания и среднеквадратические отклонения биометрических параметров:

$$|\mu_i| = \frac{|E(\xi_i) - E(v_i)|}{\sigma(\xi_i) \cdot \sigma(v_i)} \quad (1.6)$$

Знак при весовом коэффициенте (1.4) выбирается автоматически, исходя из соотношения $E(\xi_i)$ $E(v_i)$ и требуемого выходного состояния «0» или «1» при воздействии на нейрон образами «Свой». Из-за того, что алгоритм обучения по ГОСТ Р 52633.5 исключает итерационный поиск весовых коэффициентов нейронов, процедуры обучения нейронных сетей становятся рекордно быстрыми и устойчивыми. Снимается также проблема обучения многослойных нейронных сетей (число слоев нейронов у обучаемых нейронных сетей может быть любым), однако стандарт рассматривает только однослойные и двухслойные нейронные сети. Нейронные сети с тремя и более слоями нейронов являются избыточными, и для их применения необходимо специальное обоснование.

В конечном итоге алгоритм обучения по ГОСТ Р 52633.5-2011 [1] позволяет экспоненциально снижать вероятности ошибок второго рода, увеличивая число учитываемых биометрических параметров. Графики связи вероятности ошибок второго рода и числа учитываемых биометрических параметров приведены на рисунке 1.5.

Из рисунка 1.5 видно, что для малого числа хороших биометрических параметров рекомендуемый стандартом алгоритм обучения проигрывает обычным итерационным алгоритмам. При большом числе биометрических параметров ситуация становится обратной, становится выгодным использовать новые алгоритмы обучения. Более того, при неограниченном увеличении числа входов у нейросети качество принимаемых ею решений может быть как угодно велико. Неограниченное увеличение числа входов у нейросетевого преобразователя может быть получено только при использовании мультибиометрии, так как информа-



Рис. 1.5. Зависимости роста качества принимаемого решения (снижения вероятности ошибок второго рода – P_2), как функция длины вектора анализируемых параметров

тивность одного биометрического образа конечна [4-6] и быстро исчерпывается по мере увеличения числа входов преобразователя биометрия-код.

Таким образом, предложенные в соответствии с принятыми стандартами по высоконадежной биометрии решения по обучению нейронной сети с большим числом выходов и входов позволяют добиться устойчивого процесса обучения и полностью его автоматизировать, что крайне важно для широкого практического применения высоконадежной биометрической аутентификации личности, обеспечивающей конфиденциальность, анонимность и обезличенность массового оборота персональных биометрических данных.

1.2 Оценка «нано» и «пико» вероятностей появления редких событий биометрической аутентификации на малых тестовых выборках

Тестирование нейросетевых преобразователей биометрия-код (ПБК) является достаточно сложной научно-технической задачей из-за высоких значений стойкости ПБК к атакам подбора. Процедуры тестирования по ГОСТ Р 52633.3-2011, с одной стороны

позволяют получать достоверные численные оценки стойкости к атакам подбора ПБК, а с другой стороны отказаться от использования больших и сверхбольших баз биометрических образов «Чужой».

При решении задачи тестирования ПБК предполагается, что выходные коды нейросетевого ПБК в пространстве расстояний Хэмминга от кода «Свой» могут быть описаны биномиальным законом зависимых (связанных между собой) событий.

Известен классический закон распределения плотностей вероятности независимых событий соответствующий численному эксперименту по схеме Бернулли. Применительно к ПБК в рамках гипотезы независимости разрядов выходного кода длиной – n, злоумышленник сможет угадать k разрядов кода «Свой» с вероятностью описываемой классическим биномиальным законом:

$$P(k, n, P(.)) = \left[\frac{n!}{k!(n-k)!} \right] \cdot \{P(.)\}^k \cdot \{1 - P(.)\}^{n-k} \quad (1.7)$$

где $P(k)$ – вероятность угадывания k разрядов выходного кода длиной – n;

$P(.)$ – показатель асимметрии использованной в опытах Бернулли единственной монеты. $P(.)$ совпадает с вероятностью появления состояний «0», если состояние «0» для асимметричной монеты наиболее вероятно, либо совпадает с вероятностью появления состояний «1», если состояние «1» для асимметричной монеты наиболее вероятно. Параметр $P(.)$ отражает вероятность появления наиболее вероятного состояния разряда кода.

Очевидно, что можно усложнить схему Бернулли и последовательно бросать некоторую последовательность асимметричных меченых номером монет (каждая монета имеет свой номер последовательного бросания). Кроме того, каждая из монет может обладать собственной асимметрией. Однако последовательность поочередного бросания монет должна быть всегда одна и та же. Эта ситуация соответствует разной стабильности состояний в каждом из разрядов биометрического кода и приводит к следующей модификации классической записи:

$$P(k, n, P(.)) = \left[\frac{n!}{k!(n-k)!} \right] \cdot \left\{ \prod_{i=1}^k P_i(.) \right\} \cdot \prod_{j=1}^{n-k} \{1 - P_j(.)\} \quad (1.8)$$

где $P_i(.)$ – показатель асимметрии одного из угаданных разрядов;

$P_j(.)$ – показатель асимметрии одного из не угаданных разрядов кода.

Для того чтобы вернуться к привычной степенной форме записи биномиального закона, следует использовать среднее геометрическое соответствующих показателей асимметрии:

$$P(k, n, P(.)) = \left[\frac{n!}{k!(n-k)!} \right] \cdot \{\Theta(P_i(.))\}^k \cdot \{\Theta(1 - P_j(.))\}^{n-k} \quad (1.9)$$

где $\Theta(.)$ - оператор вычисления среднего геометрического.

Модифицированная схема Бернулли с разными монетами усложняется еще сильнее, если бросать монеты не независимо, а зависимо (горстью). Тогда их состояния после падения оказываются зависимыми (коррелированными). В этом случае функция распределения значений оказывается четырехмерной, причем четвертой переменной оказывается значение математического ожидания модулей парных коэффициентов корреляции - $E(|r|)$ между случайно выбранными разрядами выходных кодов ПБК. В общем виде зависимость от корреляционных связей может быть описана следующим образом:

$$P(k, n, \Theta(.), E(|r|)) = \left[\frac{n! f_1(E(|r|))}{k!(n-k)!} \right] \cdot \{\Theta(P_i(.))\}^{k \cdot f_2(E(|r|))} \cdot \{\Theta(1 - P_j(.))\}^{(n-k) \cdot f_3(E(|r|))} \quad (1.10)$$

где $f_1(.), f_2(.), f_3(.)$ – три одномерные функции переменной - $E(|r|)$.

Нейросетевые преобразователи биометрия-код, видимо, являются первым искусственным объектом, для описания которого нельзя пользоваться упрощенной схемой Бернулли, приводящей

к появлению биномиального закона с независимыми разрядами кода (1.7). Однако точного аналитического описания более сложного биномиального закона зависимых данных с разной асимметрией в каждом разряде биометрического кода (1.10) пока нет. На данный момент имеются только результаты имитационного моделирования распределения расстояний Хэмминга для различных значений одинаковых коэффициентов парной корреляции – r [7-10].

Основной проблемой при создании ГОСТ Р 52633.3-2011 являлось то, что аналитическое четырехмерное описание биномиального закона зависимых данных (1.10) пока отсутствует и необходимо создать процедуры неявной оценки вероятностей ошибок второго рода ПБК. Для этой цели новый стандарт рекомендует использовать базу из 10 000 тестовых биометрических образов «Чужой». Такой размер тестовой базы примерно в 1 000 000 раз меньше, чем требуется для примитивного тестирования последовательной подстановкой всех имеющихся тестовых образов.

Для того чтобы достоверно оценить «нано» и «пико» вероятности редких зависимых (коррелированных) событий биометрической аутентификации, ГОСТ Р 52633.3-2011 предлагает использовать генетический алгоритм подбора биометрических образов «Чужой», наиболее похожих на образ «Свой». Генетический алгоритм подбора иллюстрируется рисунком 1.6. В нулевом поколении выбираются наиболее близкие образы «Чужой» к образу «Свой». Далее выбранные образы скрещиваются между собой и из образов родителей получаются образы – потомки. Процедура скрещивания двух рукописных образов родителей «знамя» и «Север» иллюстрируется рисунком 1.7. При скрещивании образов-родителей используются процедуры морфинга, определенные в стандарте ГОСТ Р 52633.2-2010.

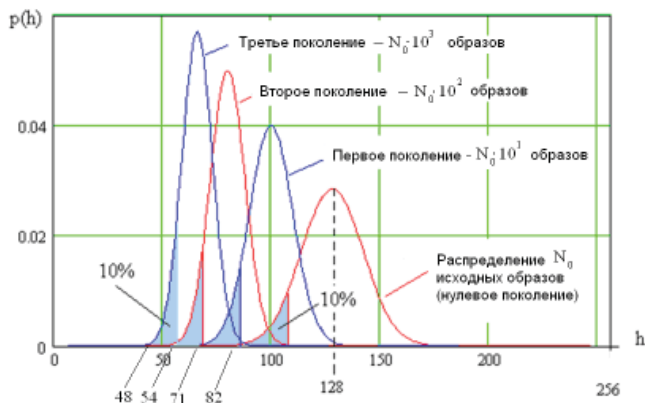


Рис. 1.6. Изменение плотности распределения расстояний Хэмминга для четырех поколений генетического подбора

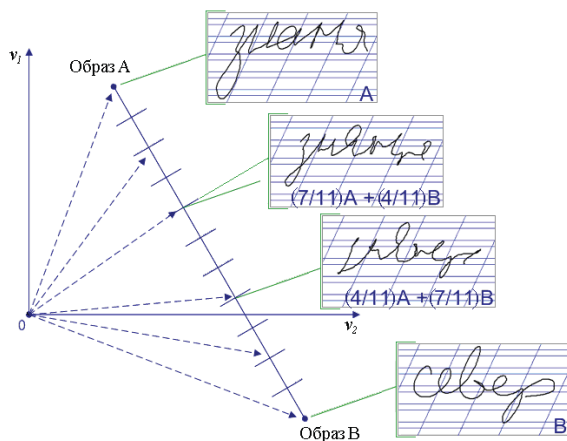


Рис. 1.7. Скрещивание двух образов-родителей и получение двух образов-потомков путем морфинга по ГОСТ Р 52633.2-2010

Описанной выше процедурой генетического подбора удастся подобрать (синтезировать) биометрические образы «Чужой», дающие коды отклики ПБК, совпадающие с кодом «Свой» от 70% до 90%. Фактически имеется эффективная процедура компрометации образа «Свой», если выходной код ПБК для образа «Свой» известен.

Следующим этапом тестирования ГОСТ Р 52633.3-2011 рекомендует осуществлять случайный подбор уже почти полностью скомпрометированного биометрического образа «Свой». Случайный подбор осуществляется по блок-схеме, приведенной на рисунке 1.8.

При этом типе тестирования большинство биометрических параметров ($100\%-a\%$) заимствуется от образа «Свой», а меньшее число биометрических параметров ($a\%$) заимствуется от генератора псевдослучайных данных, имитирующего множество образов «Все Чужие». Тестирование по блок-схеме рисунка 1.8 повторяют многократно, случайно выбирая скомпрометированные и неизвестные параметры биометрического образа «Свой». Результаты тестирования усредняют. Пример остаточного подбора почти скомпрометированного биометрического образа дается на рисунке 1.9.

Оценку стойкости обученного нейросетевого ПБК далее осуществляют путем умножения числа примеров образов «Чужой», потраченных на частичную компрометацию биометрического образа «Свой», на среднее число примеров образов «Чужой», потраченных на подбор уже почти полностью скомпрометированного образа «Свой» по блок-схеме рисунка 1.9.

В конечном итоге по процедурам ГОСТ Р 52633.3-2011 удастся оценивать «нано» и «пико» вероятности зависимых событий биометрической аутентификации, находящиеся в интервале вероятностей от $P_2 \approx 10^{-9}$ до $P_2 \approx 10^{-12}$ на тестовых базах, состоящих всего из 10 000 естественных биометрических образов «Чужой». При этом используются обычные средства вычислительной техники с обычной производительностью и обычными размерами памяти. Изложенные в стандарте алгоритмы тестирования позволяют на несколько порядков ускорить время тестирования и на несколько порядков снизить размеры базы тестовых биометрических образов.

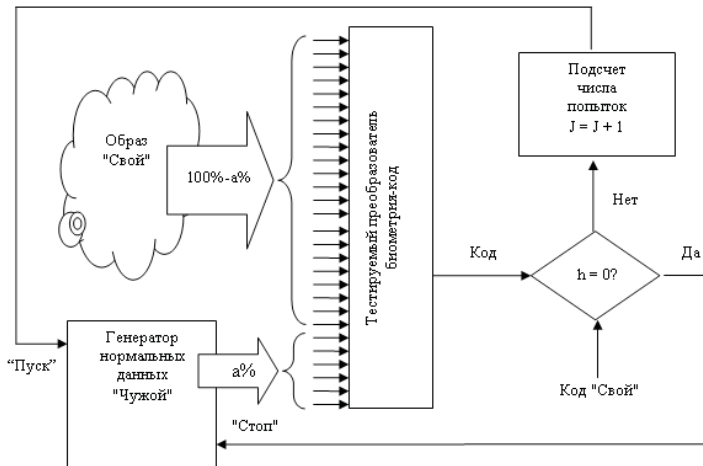


Рис. 1.8. Блок-схема тестирования остаточной стойкости почти полностью скомпрометированного биометрического образа

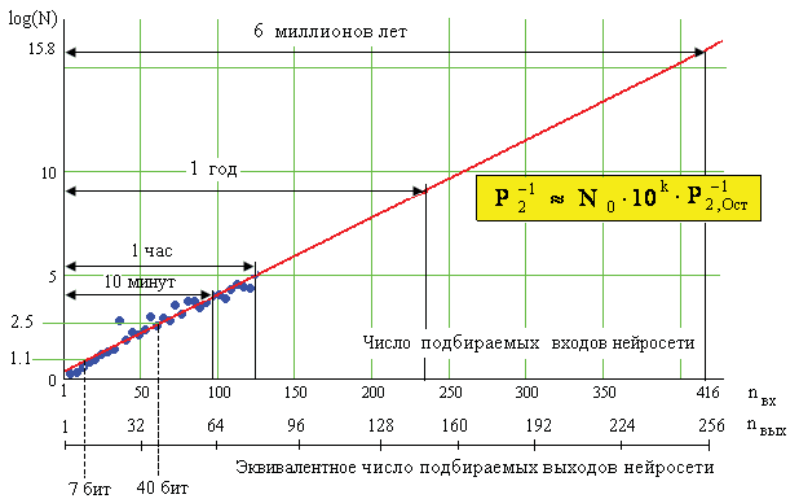


Рис. 1.9. Линейный рост числа попыток подбора в логарифмических шкалах по мере роста числа неизвестных разрядов биометрического кода

1.3 Сверхбыстрый алгоритм оценки энтропии существенно зависимых дискретных выходных состояний нейросетевых преобразователей биометрия-код

1.3.1 Ограниченные возможности классического метода вычисления энтропии

Процедуры вычисления многомерной энтропии зависимых кодов исследованы достаточно хорошо применительно к лингвистике [11-13]. Запись речи (например, на русском языке кириллицей в ASCII кодах) можно рассматривать как некоторую последовательность сильно коррелированных 8 битных кодов, отражающих особенности русского языка. По Шеннону для некоторой конечной системы кодов (например, кодов букв языка) может быть вычислена энтропия этой системы через вероятности появления того или иного кода в анализируемой последовательности. Энтропию одиночных кодов обычно оценивают через вероятности их появления:

$$H(x) = - \sum_{i=1}^S P(x_i) \cdot \log_2(P(x_i)) \quad (1.11)$$

где x – кодировка одиночной буквы (знака) сообщения;
 i – порядковый номер одиночной буквы (знака) сообщения;
 S – общее число букв (знаков) в языке сообщения,
 $P(x_i)$ – вероятность появления кода x_i .

Выражение (1.11) описывает энтропию появления одиночных символов (кодов букв). Это описание не может учитывать корреляционные связи, характерные для разных букв в текстах русского языка. Для более точного описания энтропии системы (например, естественного языка) необходимо учитывать энтропию появления пар букв:

$$H(x_1, \tilde{\delta}_2) = - \sum_{1i=1}^S \sum_{2i=1}^S P(x_{1i}, x_{2i}) \cdot \log_2(P(x_{1i}, x_{2i})) \quad (1.12)$$

где x_1 – кодировка первой буквы (знака) в паре знаков сообщения;

x_2 – кодировка второй буквы (знака) в паре знаков сообщения;

По аналогии могут быть построены выражения для энтропии троек букв, четверок букв и так далее. В итоге для вычисления энтропии группы из «п» символов имеет место следующее выражение:

$$H(x_1, x_2, \dots, x_k) = - \sum_{i=1}^S \sum_{2i=1}^S \dots \sum_{ki=1}^S P(x_{i_1}, x_{2i_1}, \dots, x_{ki_1}) \times \log_2(P(x_{i_1}, x_{2i_1}, \dots, x_{ki_1})) \quad (1.13)$$

Фактически, с использованием выражения (1.13) вычисляется k-мерная энтропия языка или нейросетевого преобразователя «биометрия-код» при анализе его кодов.

Для каждой из систем кодов (систем кодирования текстов языка) существует некоторое значение «к», которое достаточно полно позволяет описать ее энтропию. Например, для оценки энтропии русского языка необходимо учитывать группы из 17 знаков, а для оценки английского языка необходимо принимать во внимание группы из 22 знаков [14]. То есть, расчет реальной энтропии русского языка является 17-мерной задачей, а расчет энтропии английского языка является 22-мерной задачей. На рисунке 1.10 даны графики изменения энтропии русского, английского, татарского языков в зависимости от числа учитываемых букв при восьми битной кодировке каждой из букв.

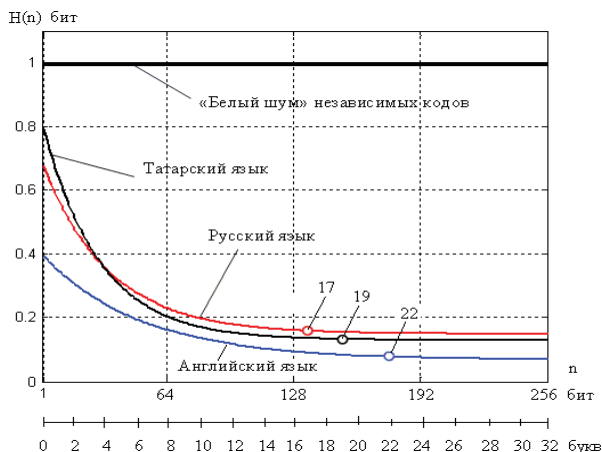


Рис. 1.10. Нормированная энтропия естественных языков (приходящаяся на 1 бит при 8 битной кодировке знаков текстов) для разного числа учитываемых знаков

Из данных рисунка 1.10 видно, что энтропия естественных языков падает при учете групп из нескольких рядом стоящих букв, однако этот процесс останавливается. Далее увеличивать размерность вычисляемой энтропии языка нет смысла.

Описанный выше классический подход к расчету многомерной энтропии, к сожалению, технически нереализуем при исследовании реальных языков и преобразователей «биометрия-код». Основной недостаток классического многомерного вычисления энтропии состоит в том, что требуются огромные размеры исходных данных. Как правило, требуется массив исходных данных, превышающий число возможных состояний исследуемого кода. Например, для оценки энтропии одиночных знаков (букв) русскоязычных текстов (кириллица – 32 буквы, 64 буквы с учетом регистра, 70 символов с учетом пробела и знаков препинания) требуется исходный текст, примерно, из 500 знаков.

Для вычисления одномерной энтропии по формуле (1.11) размер статистически исследуемой представительной выборки должен быть примерно в 10 раз больше, чем размер алфавита (число кодировок состояний одного знака – S). В том случае, если необходимо вычислить двумерную энтропию по формуле (1.12), то устойчивые результаты получаются, когда используются тексты русского языка, состоящие из 5 000 и более знаков. Для вычисления трехмерной энтропии (учитываются тройки допустимых сочетаний знаков) требуется использовать текст, состоящий из 50 000 и более знаков.

Получается, что представительная выборка должна всегда иметь размер больше, чем число существующих сочетаний, букв, пар букв, троек букв, ..., k -грамм знаков. В частности для естественных языков размеры достаточно представительных выборок хорошо описываются следующим соотношением:

$$N(k) \approx S^{\frac{k}{2}+1} \quad (1.14)$$

Размерность задачи n прямо входит в показатель степени, то есть размеры обучающей выборки растут по показательному закону с ростом размерности оцениваемой энтропии системы. Это

все приводит к тому, что прямой расчет энтропии естественных языков становится вычислительно неразрешимой задачей. Сегодня существуют реально вычисленные значения только для 7-грамм (7 сочетаний кодов знаков). Попытки прямого вычисления большей размерности сочетаний знаков через классические многомерные вычисления (1.14) приводят к тому, что перестает хватать вычислительных ресурсов общедоступной вычислительной техники. Перестает хватать размеров доступных оцифрованных текстов того или иного естественного языка (нужна память на миллионы терабайт), кроме того, обычные ПЭВМ вынуждены будут непрерывно работать над подобной задачей десятки лет.

Получается, что классический многомерный энтропийный анализ невозможно применить даже для решения общественно значимой задачи энтропийного описания естественных языков. Естественно, что этот же инструмент оказывается непригоден для описания нейросетевых преобразователей «биометрия-код», имеющих еще большую размерность кодов выходного биометрического языка. Если при энтропийном анализе английского языка достаточно учитывать последовательности только из 22 букв (см. кривую рис. 1.10) в 8 битной кодировке (176 бит), то при анализе преобразователей «биометрия-код» требуется учитывать уже 256, 512, 1024 битные коды. Прямые методы вычисления высокоразмерной энтропии по формуле (1.13) становятся технически нереализуемыми.

Шеннон при оценке предельных значений энтропии английского языка пользовался косвенными вычислениями, построенными на привлечении людей – носителей английского языка [13] (они восстанавливали часть замазанных в тексте знаков). Прямых вычислений многомерной энтропии нельзя было осуществить 70 лет назад, нельзя их осуществить и сейчас. Тем не менее, задача вычисления оценок энтропии высоких размерностей остается актуальной. Более того, актуальность этой задачи растет в связи с необходимостью оценивать стойкость к атакам подбора нейросетевых преобразователей «биометрия-код».

1.3.2 Сверхбыстрые алгоритмы вычисления энтропии в пространстве кодов Хэмминга

Так как классический метод вычисления многомерной энтропии требует огромных вычислительных затрат и огромных размеров исходных биометрических данных, необходимо предпринять усилия по созданию более экономичных вычислительных процедур. Одним из вариантов значительного снижения объемов вычислений является переход из поля обычных кодов в поле кодов расстояний Хэмминга [15].

Предположим, что известен код «Свой» на 64 выходах нейросетевого преобразователя, тогда вместо статистического исследования поля выходных кодов, имеющего 2^{64} состояний, можно получить распределение значений меры Хэмминга, имеющего всего 65 состояний. Пример такого распределения приведен на рисунке 1.11.

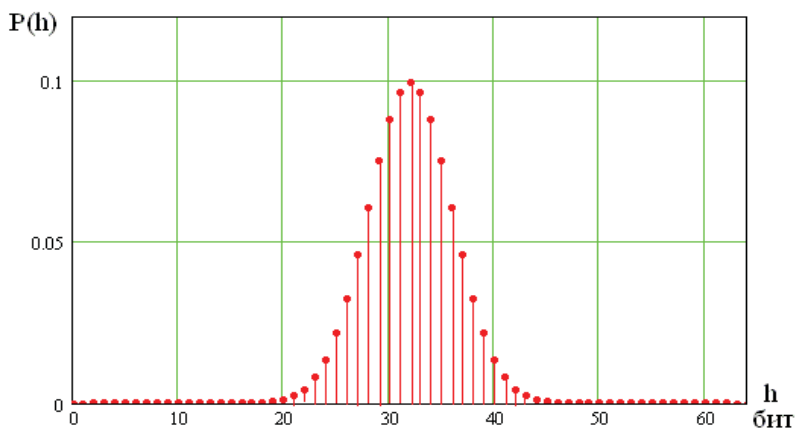


Рис. 1.11. Распределение значений меры Хэмминга для 64-разрядного выходного кода типа «белый шум»

Очевидно, что вычисление по полю 2^{64} возможных состояний и вычисление по полю кодов из 65 возможных состояний принципиально отличаются по затрачиваемым вычислительным ресурсам. То есть переход от вычисления классической 64-мерной

энтропии $H(64)$ к вычислению 64-мерной энтропии Хэмминга $\tilde{H}(64)$ эквивалентен переходу от 64-мерных вычислений к одномерным вычислениям. При этом энтропия Хэмминга будет вычисляться следующим образом:

$$\tilde{H}(n) = - \sum_{i=0}^n P(h_i) \cdot \log_2(P(h_i)) \quad (1.15)$$

где n – размерность задачи (длина исследуемых кодов);

$P(h)$ – вероятность появления значения расстояния Хэмминга – h относительно выходного кода «Свой».

Переход от обычной классической энтропии к энтропии Хэмминга является нелинейным преобразованием. На рис. 1.12 даны кривые, отражающие связи значений энтропии Хэмминга с входной размерностью исследуемых кодов при разном значении коэффициентов парной корреляции между разрядами этих кодов. Эти кривые рассчитаны для симметричных распределений значений меры Хэмминга, расположенных в центре динамического диапазона возможных значений (рис. 1.11).

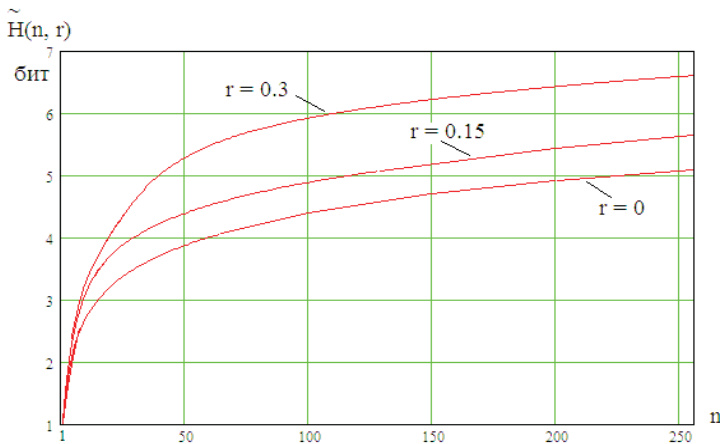


Рис. 1.12. Кривые роста энтропии кодов Хэмминга высоких размерностей при разных значениях коэффициентов парной корреляции разрядов, исследуемых кодовых последовательностей

Из данных рисунка 1.12 видно, что энтропия Хэмминга является существенно нелинейной функцией при малых размерно-

стях (n менее 100). При более высоких размерностях энтропия Хэмминга практически линейно зависит от размерности, как и классическая энтропия. То есть, энтропию кодов Хэмминга вполне можно использовать для статистических исследований выходных кодов преобразователей «биометрия-код». При этом получается существенное преимущество в снижении размеров требуемой тестовой выборки.

Вычисляя энтропию Хэмминга, удастся пересчитывать ее в классическую энтропию с использованием нелинейной функции преобразования:

$$\dot{I}(\bar{n}, E(r|)) = F(\tilde{H}(n, E(r|))) \quad (1.16)$$

Важнейшим преимуществом перехода к вычислению энтропии Хэмминга является то, что резко снижаются требуемые размеры исходных данных. На рис. 1.13 показаны кривые снижения ошибки вычисления классической энтропии при увеличении числа используемых кодов.

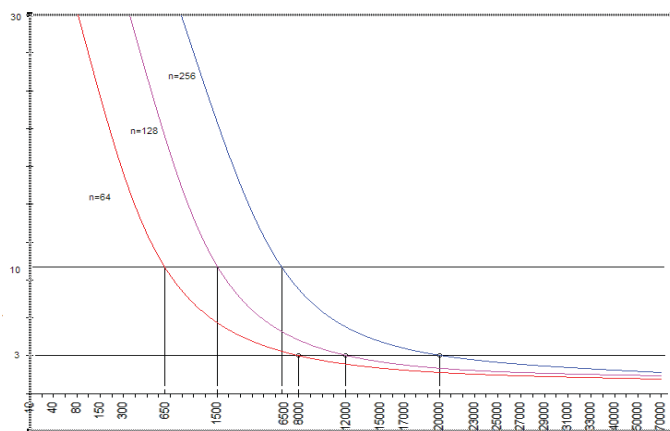


Рис. 1.13. Кривые снижения относительной ошибки оценки классической энтропии Хэмминга при увеличении объема обучающей (представительной) выборки

Из данных рисунка 1.13 видно, что относительная ошибка превышает 3% при объемах представительной выборки 8000, 1200 и 20000 примеров кодов-откликов длиной 64, 124, 256 бит.

То есть, размер представительной выборки для оценки энтропии Хэмминга линейно зависит от размерности задачи:

$$\tilde{N}(n) \approx 4100 + 62.5 \cdot n \quad (1.17)$$

где n – длина статистически исследуемых кодов.

Если допустимые ограничения ошибки энтропии составляют 10%, то абсолютные размеры представительных выборок значительно сокращаются (до 650, 1500, 6500 примеров кодов длиной 64, 128, 256 бит), однако связь их с размерностью решаемой задачи становится квадратичной:

$$\tilde{N}(n) \approx 210 + 1.2 \cdot n + 0.072 \cdot n^2 \quad (1.18)$$

Для нас важно то, что линейная связь (1.17) и квадратичная связь (1.18) размеров представительной выборки приводят к технически решаемым задачам для n от 256 и выше. Это принципиально важное отличие от показательной зависимости, которая не позволяет решать задачу даже при сравнительно низких размерностях от 32 до 48 бит.

Только переход в пространство энтропии Хэмминга позволяет вычислять энтропии высоких и сверхвысоких размерностей. В частности, удастся вычислять энтропию естественных языков на общедоступной вычислительной технике на относительно малых размерах используемых для вычисления статистик текстов. При этом использовались тексты трех языков, содержащие порядка 500 000 знаков, но сокращение текстов до 10000 знаков не приводит к ощутимому изменению результатов.

1.3.3 Связь статистик энтропии Хэмминга и классической энтропии

Приближенная оценка вероятности ошибок второго рода преобразователя «биометрия-код» [16; 17] может быть построена на использовании гипотезы нормального распределения значений:

$$P_2 \approx \frac{1}{\sqrt{2\pi}\sigma(h)} \int_{-\infty}^0 \exp\left\{-\frac{(E(h) - u)^2}{2\sigma^2(h)}\right\} du \quad (1.19)$$

Как показала практика, выражение (1.19) дает достаточно хорошую точность прогноза, однако в нем отсутствует физический смысл. Интегрирование в выражении (1.19) осуществляется по отрицательным значениям расстояний Хэмминга, а это противоречит определению расстояний Хэмминга. Такое несоответствие определенным образом компрометирует создаваемую теорию. Попытаемся уйти от этого несоответствия и более точно определить связь статистик Хэмминга с энтропией исследуемого преобразователя «биометрия-код».

В работах [14] показано, что энтропия Хэмминга по своей сути является усредненной информацией по всем возможным событиям $h_0=0, h_1=1, \dots, h_{n-1}=(n-1), h_n=n$. Формально это утверждение можно записать следующим образом:

$$\tilde{H}(h) = E(I(h_i)) \quad (1.20)$$

где $I(h_i) = -\log_2(P(h_i))$ - информация о появлении события h_i .

С другой стороны классическая энтропия является не чем иным как усредненной информацией о появлении всех допустимых значений кодов:

$$H(n) = E(I(\bar{x}_i)) \quad (1.21)$$

где $I(\bar{x}_i) = -\log_2(P(\bar{x}_i))$ - информация о появлении i -го кода - $\bar{x}_i$ соответствующего алфавита кодов длиной - n .

Проведенные исследования показали наличие четкой связи классической энтропии исследуемой последовательности кодов с информацией о событии $h_0=0$. Сформулируем эту связь в виде основной теоремы:

$$H(n) = I(h_0) \quad (1.22)$$

Полная формулировка основной теоремы сводится к тому, что для всех выходных кодов с зависимыми или независимыми разрядами, с равновероятными состояниями разрядов или иными вероятностями их состояний справедливо соотношение (1.22).

Подчеркнем, что справедливость соотношения (1.22) для классического биномиального закона распределения значений (коды независимы и вероятности состояния «0», «1» в каждом разряде кода одинаковы) легко доказывается алгебраическими преобразованиями.

В том случае, если простейшая схема Бернулли не работает (разряды кодов зависимы или вероятности состояния «0», «1» в каждом разряде различаются) доказательство основной теоремы (1.22) было осуществлено средствами численного моделирования. Наиболее просто численное моделирование может быть осуществлено для независимых кодов, так как для их описания можно воспользоваться биномиальным законом распределения значений. Примеры результатов проведенного численного эксперимента в среде MatCad приведены в таблице 1.1.

Таблица 1.1

Численное моделирование с использованием биномиального независимого закона распределения значений для равновероятных состояний $p=0.5$ и $p=0.25$ для $n=1, 2, 3, 4, 5$

	$p=0.5$	$p=0.25$
$n=1$	$i := 0..1$ $h_i := \text{dbinom}(i, 1, 0.5)$ $H := \sum_{i=0}^1 -h_i \cdot \log(h_i, 2) \quad h = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$ $H = 1 \quad -\log(h_0, 2) = 1$	$i := 0..1$ $h_i := \text{dbinom}(i, 1, 0.25)$ $H := \sum_{i=0}^1 -h_i \cdot \log(h_i, 2) \quad h = \begin{pmatrix} 0.75 \\ 0.25 \end{pmatrix}$ $H = 0.811 \quad -\log(h_0, 2) = 0.415$
$n=2$	$i := 0..2$ $h_i := \text{dbinom}(i, 2, 0.5)$ $H := \sum_{i=0}^2 -h_i \cdot \log(h_i, 2) \quad h = \begin{pmatrix} 0.25 \\ 0.5 \\ 0.25 \end{pmatrix}$ $H = 1.5 \quad -\log(h_0, 2) = 2$	$i := 0..2$ $h_i := \text{dbinom}(i, 2, 0.25)$ $H := \sum_{i=0}^2 -h_i \cdot \log(h_i, 2) \quad h = \begin{pmatrix} 0.562 \\ 0.375 \\ 0.062 \end{pmatrix}$ $H = 1.248 \quad -\log(h_0, 2) = 0.83$

n=3	$i := 0..3$ $h_i := \text{dbinom}(i, 3, 0.5)$ $H := \sum_{i=0}^3 -h_i \cdot \log(h_i, 2)$ $H = 1.811$ $-\log(h_0, 2) = 3$ $h = \begin{pmatrix} 0.125 \\ 0.375 \\ 0.375 \\ 0.125 \end{pmatrix}$	$i := 0..3$ $h_i := \text{dbinom}(i, 3, 0.25)$ $H := \sum_{i=0}^3 -h_i \cdot \log(h_i, 2)$ $H = 1.542$ $-\log(h_0, 2) = 1.245$ $h = \begin{pmatrix} 0.422 \\ 0.422 \\ 0.141 \\ 0.016 \end{pmatrix}$
n=4	$i := 0..4$ $h_i := \text{dbinom}(i, 4, 0.5)$ $H := \sum_{i=0}^4 -h_i \cdot \log(h_i, 2)$ $H = 2.031$ $-\log(h_0, 2) = 4$ $h = \begin{pmatrix} 0.062 \\ 0.25 \\ 0.375 \\ 0.25 \\ 0.062 \end{pmatrix}$	$i := 0..4$ $h_i := \text{dbinom}(i, 4, 0.25)$ $H := \sum_{i=0}^4 -h_i \cdot \log(h_i, 2)$ $H = 1.762$ $-\log(h_0, 2) = 1.66$ $h = \begin{pmatrix} 0.316 \\ 0.422 \\ 0.211 \\ 0.047 \\ 3.906 \times 10^{-3} \end{pmatrix}$
n=5	$i := 0..5$ $h_i := \text{dbinom}(i, 5, 0.5)$ $H := \sum_{i=0}^5 -h_i \cdot \log(h_i, 2)$ $H = 2.198$ $-\log(h_0, 2) = 5$ $h = \begin{pmatrix} 0.031 \\ 0.156 \\ 0.313 \\ 0.313 \\ 0.156 \\ 0.031 \end{pmatrix}$	$i := 0..5$ $h_i := \text{dbinom}(i, 5, 0.25)$ $H := \sum_{i=0}^5 -h_i \cdot \log(h_i, 2)$ $H = 1.936$ $-\log(h_0, 2) = 2.075$ $h = \begin{pmatrix} 0.237 \\ 0.396 \\ 0.264 \\ 0.088 \\ 0.015 \\ 9.766 \times 10^{-4} \end{pmatrix}$

При анализе содержания таблицы 1.1 следует иметь в виду, что для биномиального закона независимых данных схемы Бернулли точное описание энтропии соответствует выражению:

$$H(n, p) = -\log_2 \{(1-p)^n\} \quad (1.23)$$

Для независимых событий, полученных не по схеме Бернулли и для зависимых событий вывести столь простое соотношение как (1.23) невозможно, однако проведенное численное моделирование показывает, что основная теорема (1.22) для них выполняется. При моделировании влияния ненулевых корреляционных связей наиболее просто убедиться в справедливости основной

теоремы (1.22) для частного случая значения 0.5 всех парных коэффициентов корреляции [7; 18]. В этом случае распределение Хэмминга становится равномерным, а энтропия точно вычисляется по следующей формуле:

$$H(n, p = 0.5, E(r) = 0.5) = -\log_2 \left\{ p(h_0) = \frac{1}{n+1} \right\} \quad (1.24)$$

Одним из важнейших следствий основной теоремы (1.22) является то, что для оценки энтропии исследуемого преобразователя «биометрия-код» нужно знать только вероятность появления события $h_0=0$.

Все остальные вероятности событий $h_1=1, h_2=2, \dots, h_n=n$ необходимо знать только, если не известно значение $p(h_0)$.

Вторым важнейшим следствием основной теоремы (1.22) является возможность уточнения прогноза вероятностей ошибок второго рода, исследуемой в пространстве кодов Хэмминга. Если считать, что:

$$P_2 = \frac{1}{2^{H(n)}} \quad (1.25)$$

то выражение (1.23) может быть уточнено:

$$P_2 \approx \frac{1}{\sqrt{2\pi} \cdot \sigma(h)} \int_0^1 \exp \left\{ \frac{-(E(h) - \tilde{h})^2}{2 \cdot \sigma^2(h)} \right\} d\tilde{h} \quad (1.26)$$

где $\tilde{h}$ – псевдонепрерывная переменная.

Вычисления по формуле (1.25) иллюстрируются на рис. 1.14. Практика показала, что прогнозирование вероятности ошибок второго рода по формуле (1.26) существенно точнее прогнозирования по формуле (1.19).

Таким образом, если мы отказываемся вычислять многомерную энтропию длинных кодов по Шеннону и переходим к ее вычислениям в пространстве расстояний Хэмминга нам уже не нужны огромные вычислительные затраты для вычисления вероятностей очень редких событий. Предложенный в данной работе метод позволяет пользоваться выборками всего из 10 000 кодов

длинной 256, 512, 1024 бит. То есть удастся понизить размеры представительной выборки на десятки порядков. Именно из-за этого предложенный метод оказывается сверхбыстрым при вполне приемлемой относительной погрешности оценки энтропии, находящейся в пределах от 2% до 5%.

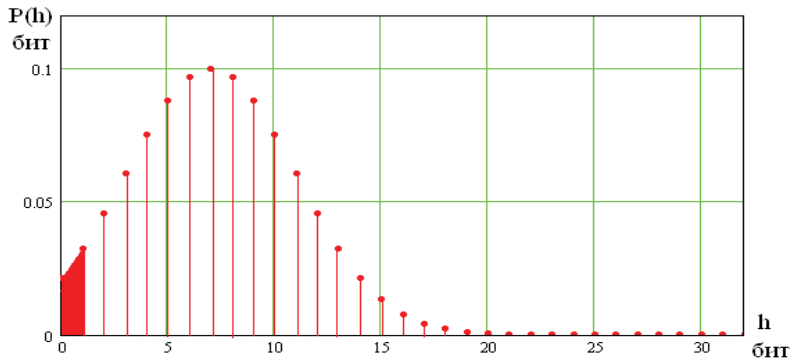


Рис. 1.14. Прогнозирование вероятности ошибок второго рода в рамках гипотезы нормального распределения кодов Хэмминга через вычисление вероятности попадания непрерывной переменной h в интервал от 0.0 до 1.0

1.4 Связь энтропии выходных состояний нейросетевых преобразователей биометрия-код с коэффициентами парной корреляции

Одной из важнейших научных задач является разработка методов оценки энтропии длинных кодов с существенно зависимыми состояниями разрядов. Если бы состояния разрядов 256-ти битных выходных кодов нейросетевого преобразователя оказались полностью независимыми при воздействии на нейропреобразователь случайными образами «Чужой», то мы получили бы идеальный белый шум с энтропией 256 бит. К сожалению, столь идеальной картины наблюдать не удастся. Парные корреляционные связи выходов преобразователя биометрия-код оказываются весьма и весьма существенными. На рисунке 1.15 приведена гистограмма распределения парных корреляционных связей обученного преобразователя биометрия-код для образов все «Чужие».

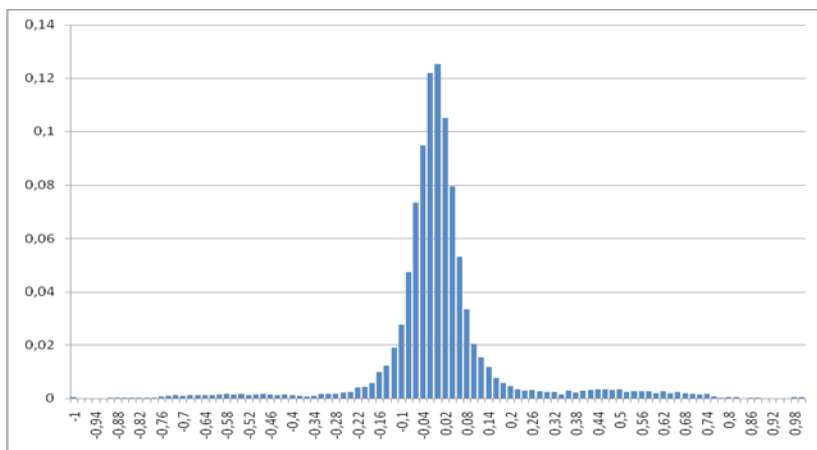


Рис. 1.15. Типичное распределение парных коэффициентов корреляций на выходе преобразователя биометрия-код

Если пойти по классическому пути Клода Шеннона и попытаться прямыми вычислениями рассчитать энтропию выходных зависимых кодов длиной 256 бит, то потребуются огромные вычислительные затраты, а так же потребуются огромная база случайных тестовых образов «Чужие». В связи с этим возникает задача расчета гистограммы распределения парной корреляции вычисления среднего от модуля значений коэффициентов парной корреляции и предсказанию по этому значению ожидаемой энтропии выходных кодов.

1.4.1 Принятые упрощения

Упростить задачу, предположив, что все выходы хорошо сбалансированы и имеют одинаковую входную энтропию $H_i = 1$ бит. В этом случае можно средствами имитационного моделирования [19-21] воспроизвести ситуацию. Мы можем сделать коэффициенты парной корреляции одинаковыми и плавно изменять их от $r = 0$ до $r = 1$. Параллельно мы можем рассчитывать энтропию одним из сверхбыстрых алгоритмов. В итоге мы получим номо-

грамму связи выходной многомерной энтропии как зависимость числа биометрических входов и коэффициента корреляции – r . На рисунке 1.16 приведена эта номограмма значений энтропии для выходных биометрических длиной $n = 2, 3, 4, 5$ бит.

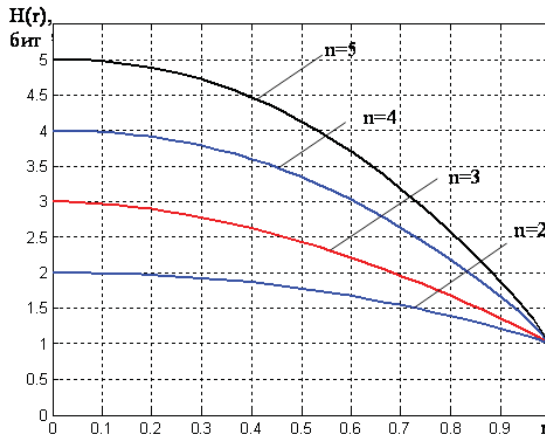


Рис. 1.16. Номограмма изменений значений энтропии для нескольких входов- n и разных значений взаимной корреляции между входами

Из рис. 1.16 видно, что при $r = 0.0$ происходит сложение энтропии нескольких входов, однако с ростом коэффициента корреляции $r \rightarrow 1$ происходит уменьшение значений многомерной корреляции. В предельном случае $r = 1$ многомерная энтропия и энтропия одного выхода полностью совпадают.

Для описания отыскиваемой функциональной связи целесообразно зависимость значений многомерной энтропии представить в виде соответствующего разложения с некоторой мультипликативной функцией $\beta(n, r)$, независимой от значения энтропии одного выходного разряда $H(x_1)$:

$$H(x_1, x_2, \dots, x_n) = (\beta(n, r) \cdot (n - 1) + 1) \cdot H(x_1) \quad (1.27)$$

Номограмма значений мультипликативной функции связи приведена на рис. 1.17

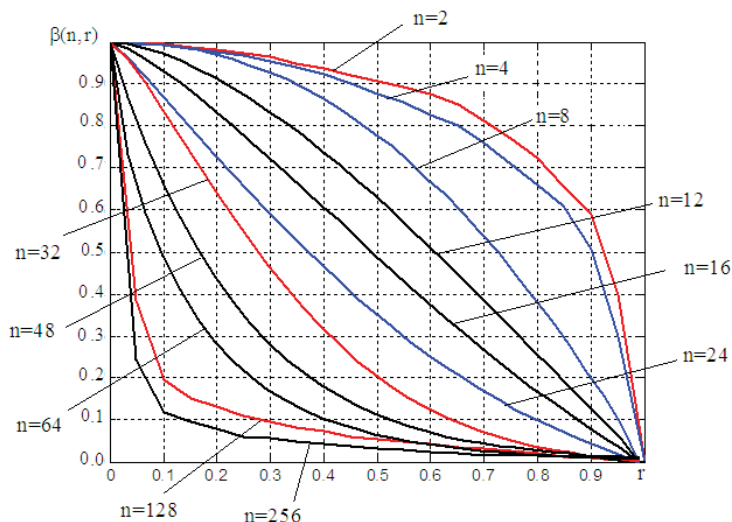


Рис. 1.17. Номограмма значений мультипликативной функции связи энтропии выходных кодов с коэффициентом равной коррелированности разрядов.

Таким образом, используя номограмму рисунку 1.17, всегда можно по вычисленному значению математического ожидания модуля коэффициента - $E(|r|) \approx r$ оценить многомерную энтропию всех учитываемых при аутентификации разрядов биометрического кода. Для этого достаточно воспользоваться выражением (1.27).

Из рисунка 1.17 видно, что при $n \approx 16$ мультипликативная функция связи $\beta(n, r)$ близка к линейной. Фактически размерность $n \approx 16$ можно считать границей между низкоразмерными биометрическими системами аутентификации и высокоразмерными системами аутентификации. В низкоразмерных системах слабые корреляционные связи между биометрическими параметрами или разрядами выходного кода слабо влияют на результирующую многомерную энтропию (можно пренебрегать влиянием слабых корреляционных связей). На энтропию низкоразмерных систем значительно влияют только значительные корреляционные свя-

зи. Для высокоразмерных систем биометрической аутентификации ситуация становится обратной. Сильное влияние оказывают слабые корреляционные связи, а корреляционные связи близкие к единице практически не оказывают влияние на значение многомерной энтропии.

Необходимо подчеркнуть, что деление биометрических систем на низкоразмерные (до 16 выходных разрядов) и высокоразмерные (много больше 16 разрядов) является принципиальным. Те приемы статистической обработки, которые хорошо себя зарекомендовали для низкоразмерных систем, нельзя переносить на высокоразмерные системы. Высокоразмерная энтропия очень чувствительна даже к малым коэффициентам корреляции, именно из-за пренебрежения малыми корреляционными связями могут появиться замечательные в рекламном отношении результаты. Однако эти результаты, видимо, окажутся куда более скромными, если преобразователь «биометрия-код» с завышенными корреляционными связями тестировать по требованиям ГОСТ Р 52633.3-2011.

Необходимо подчеркнуть, что номограмма рисунка 1.17 построена применительно к идеальным условиям (коэффициенты корреляционных связей одинаковы и энтропии всех биометрических параметров совпадают между собой). При реальных расчетах необходимо пользоваться средним значением энтропии всех биометрических параметров и усредненным значением модуля коэффициентов парной корреляции:

$$H(x_1, x_2, \dots, x_n) = (\beta(n, r) \cdot (n - 1) + 1) \cdot E\{H(x_i)\} \quad (1.28)$$

где $E\{H(x_i)\}$ математическое ожидание энтропий каждого из n выходных разрядов зависимых кодов.

1.4.2 Модификация схемы Бернулли при проведении численного эксперимента

Использование искусственных нейронных сетей для осуществления преобразования биометрических параметров в код является не единственной технологией. За рубежом (в США и Канаде) ведутся исследования [22-24] по применению так называемых

«нечетких экстракторов». То есть можно ожидать, что в будущем на рынке России и Казахстана появятся преобразователи биометрия-код, созданные на основе нечеткой логики. Из работы [24] следует, что эти преобразователи не гарантируют выполнения условия равной вероятности появления состояния «0» и «1» в каждом разряде кода. То есть каждый разряд выходного кода таких преобразователей будет иметь свое значения вероятности появления состояния «0». Как следствие необходимо модифицировать соотношение (1.28) под ситуацию возможных вариаций значений $P(\llcorner 0 \rrcorner)$ для каждого из разрядов исследуемых выходных кодов.

Для достижения объявленной выше цели необходимо отказаться от классической схемы Бернулли, по которой было выведено соотношение (1.28). По схеме Бернулли бросается одна монета, имеющая заданное соотношение вероятности появления состояния «0» и «1». Изменим схему и будем использовать 256 монет, для которых вероятность появления состояния «0» плавно увеличивается от нуля до единицы. То есть мы имеем ряд не совпадающих между собой вероятностей $P_i(\llcorner 0 \rrcorner)$. Численный эксперимент показывает, что наличие разных вероятностей $P_i(\llcorner 0 \rrcorner)$ приводит к существенному изменению энтропии, причем эта связь линейна с математическим ожиданием модуля отклонения вероятности от значения 0.5. Результаты численного эксперимента приведены на рисунке 1.18.

Так как связь энтропии и вероятности появления состояний оказывается крайне простой (линейной) модификация соотношения (1.28) сводится к простому изменению масштаба используемой при расчетах номограммы рисунка (1.16):

$$H(x_1, x_2, \dots, x_n) = (\beta(n, r) \cdot (n-1) + 1) \cdot E\{H(x_i)\} \cdot 2 \cdot (0.5 - E\{|0.5 - P_i(\llcorner 0 \rrcorner)|\}) \quad (1.29)$$

где $E\{|0.5 - P_i(\llcorner 0 \rrcorner)|\})$ - математическое ожидание модуля отклонения вероятности появления состояния «0» от значения 0.5.

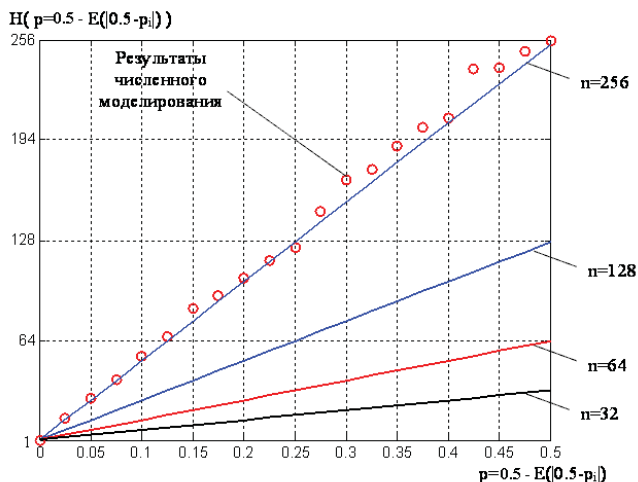


Рис. 1.18. Линейная связь значения энтропии выходных независимых кодов с математическим ожиданием модуля отклонения вероятностей $P(\text{«0»})$ от значения 0.5

Таким образом, если мы располагаем достаточно точной номограммой связи энтропии со средним значением модулей парных коэффициентов корреляции разрядов выходных кодов (рис. 1.15), то мы можем вычислять не только энтропию хорошо сбалансированных преобразователей биометрия-код, но и менее качественных преобразователей, имеющих значительный небаланс состояний «0» и «1» в разрядах контролируемых выходных кодов. Для этого достаточно простого усреднения модулей отклонений вероятностей $P(\text{«0»})$ от значения 0.5. Значение 0.5 для вероятностей $P_i(\text{«0»})$ является особой точкой симметризации вычислений.

1.5 Генетический алгоритм итерационного обращения матриц нейросетевых функционалов путем минимизации расстояний Хэмминга

По требованиям ГОСТ Р 52633.0-2006 [2] обучение нейросетевого преобразователя биометрия-код (ПБК) должно выполняться таким образом, чтобы каждый i -ый пример образа «Свой» после

его преобразования в вектор контролируемых биометрических параметров - $\bar{V}_i$, давал на выходах ПБК одинаковые коды - $\bar{C}$:

$$\begin{bmatrix} \text{ПБК} \\ N \times n \end{bmatrix} \cdot \bar{V}_i = \bar{C} \quad (1.30)$$

где $\begin{bmatrix} \text{ПБК} \\ N \times n \end{bmatrix}$ - матрица нелинейных неросетевых функционалов

ПБК размерности $(N \times n)$, осуществляющих нелинейную нейросетевую обработку вектора из N непрерывных биометрических параметров - $\bar{V}_i$ по их преобразованию в код $\bar{C}$, состоящий из n разрядов.

Подчеркнем, что при автоматическом синтезе ПБК и его обучении по ГОСТ Р 52633.5-2011 [1] может быть использована однослойная сеть нейронов или двухслойная сеть нейронов, причем каждый из нейронов сети имеет выходную переключательную функцию с двумя возможными ее состояниями «0» и «1». Вполне возможно пойти далее рекомендаций ГОСТ Р 52633.5-2011 [1] и строить ПБК на основе нейронных сетей с большим числом слоев. То есть нейросетевые функционалы матрица

$\begin{bmatrix} \text{ПБК} \\ N \times n \end{bmatrix}$ связывающие входы и выходы преобразователя

могут соответствовать нейронным сетям с 1, 2, 3 и более числом слоев. Матричная форма записи уравнения (1.30) не накладывает, каких либо ограничений на число слоев искусственных нейронов преобразователя биометрия-код.

Заметим, что форма записи уравнения (1.30) похожа на обычное матричное уравнение линейной алгебры. Известно, что в линейной алгебре существует операция обращения матриц, позволяющая по выходному вектору восстанавливать входной вектор. Значительный научно-практический интерес имеет синтез аналогичной процедуры восстановления неизвестных входных биометрических данных в рамках существенно нелинейной алгебры матриц нейросетевых функционалов.

Рассмотрим процедуру восстановления неизвестного входного воздействия, использующую тестовую базу, содержащую порядка

10 000 примеров случайных образов «Чужие», сформированную по требованиям ГОСТ Р 52633.1-2009 [3]. Блок-схема реализации процедуры итерационного восстановления неизвестного вектора входных биометрических параметров «Свой» отображена на рисунке 1.19.

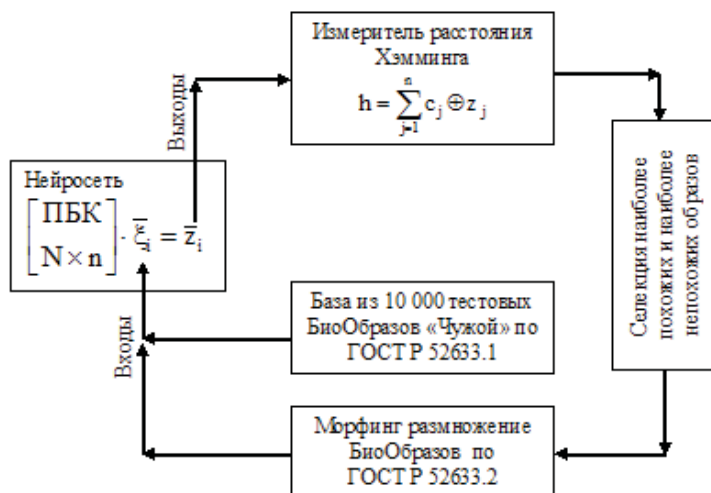


Рис. 1.19. Блок-схема процедуры восстановления вектора неизвестных входных воздействий «Свой» на нейросеть по ее известному выходному коду $\bar{c}$

В том случае, если тестовая база биометрических образов сформирована из действительно случайных образов «Чужой», нейросеть откликается на каждый i -тый вектор «Чужой» - $\bar{\xi}_i$ своим случайным кодом - $\bar{z}_i$:

$$\begin{bmatrix} \text{ПБК} \\ N \times n \end{bmatrix} \cdot \bar{\xi}_i = \bar{z}_i \quad (1.31)$$

Если теперь для каждого i -го кода найти расстояние Хэмминга до кода «Свой»:

$$h = \sum_{j=1}^n c_j \oplus z_j \quad (1.32)$$

то мы получим распределение кодов Хэмминга нулевого поколения. Для длинных кодов распределение расстояний Хэмминга хорошо описывается нормальным законом распределения значений. Пример такого распределения дан на рисунке 1.20. Исходное нулевое поколение естественных биометрических образов хорошо обученного преобразователя биометрия-код имеет математическое ожидание равное половине длины кодов (для рисунка 1.20 оно составляет 128 бит). Как видно из рисунка 1.20 наиболее близкие образы «Чужой» к образу «Свой» будут иметь расстояния Хэмминга лежащие в интервале от 92 до 100 бит (этот фрагмент распределения залит темным цветом). Наиболее не похожие образы «Чужой» будут иметь расстояния Хэмминга от 158 до 165 бит. Заметим, что все относительно наиболее непохожие образы будут наиболее похожими к некоторому образу «Чужой» являющемуся инверсией образу «Свой» и дающему отклик в виде инверсии кодов «Свой».

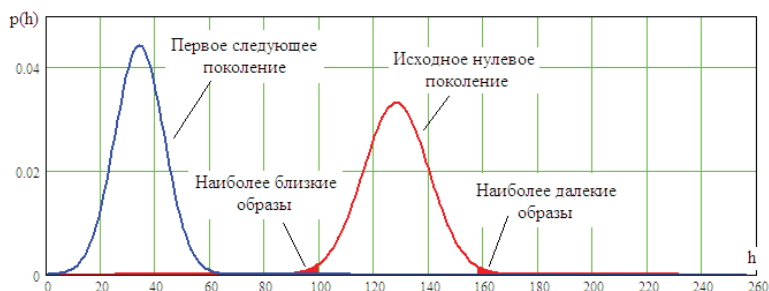


Рис. 1.20. Распределения расстояний Хэмминга для исходного и последующего поколений биометрических образов, подаваемых на нейронную сеть с 256 выходами.

Следует подчеркнуть то, что если выбрать только самые похожие образы с расстоянием Хэмминга от 92 до 100 бит и размножать их по процедурам ГОСТ Р 52633.3-2011 [25], то невозможно итерационно восстановить параметры отыскиваемого входного вектора «Свой». По мере движения к точке $h=0$ происходит существенное сужение дисперсий распределений расстояний Хэмминга следующих поколений. С каждым следующим поколением

происходит все большее и большее сужение дисперсий распределений расстояний Хэмминга.

Для того, чтобы избежать этого необходимо осуществлять морфинг между парами биометрических образов принадлежащими, выявленным наиболее близким и наиболее удаленным образам. При этом, если руководствоваться рекомендациями ГОСТ Р 52633.2-2009 [26], то мы всегда будем попадать в интервал распределения расстояний Хэмминга только первого поколения. Для того, что двигаться в сторону точки нулевого расстояния Хэмминга необходимо данные наиболее непохожих образов брать с отрицательным знаком. Эта ситуация отображена на рисунке 1.21.

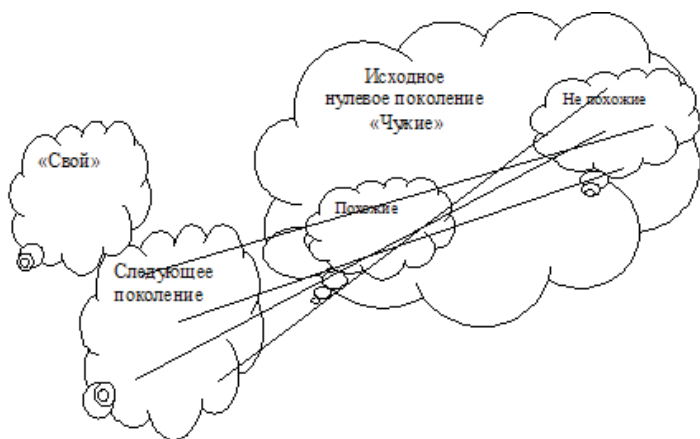


Рис. 1.21. Вынос синтетических образов за пределы исходного поколения при морфинге наиболее похожих образов с «отрицательными» непохожими образами

Из данных рисунка 1.21 видно, что в исходном нулевом поколении «Чужие» выделены два подмножества наиболее «похожих» и наиболее «не похожих» биометрических образов. Именно пары «похожих» и «не похожих» биометрических образов используются далее как образы-родители для получения искусственных образов-потомков.

Морфинг преобразование пар образов-родителей осуществляется следующим образом:

$$\xi_{1,i} = (1 - a) \cdot \xi_{0Б,i} - a \cdot \xi_{0Д,i} \quad (1.33)$$

где a - регулируемый при морфине параметр; $\xi_{1,i}$ - i -тый параметр синтезированного образа первого поколения; $\xi_{0Б,i}$ - i -тый параметр одного из наиболее близких образов-родителей нулевого поколения; $\xi_{0Д,i}$ - i -тый параметр одного из наиболее далеких образов-родителей нулевого поколения.

Особенностью рассматриваемой операции синтеза неизвестного биометрического образа является то, что морфинг (1.33) выполняют постепенно меняя регулируемый параметр от 0 до 3, наблюдая за расстоянием Хэмминга синтезируемого образа. Регулируемый параметр увеличивают до момента пока расстояние Хэмминга уменьшается. Если осуществлять эти операции для различных пар близких и далеких образов, то мы получим ядро из нескольких десятков родителей следующего первого поколения. Далее число примеров образов «Чужой» в новом поколении должно быть доведено до числа образов в исходном нулевом поколении процедурами мутации и скрещивания, выполняемыми по ГОСТ Р 52633.2-2009. Как видно из рисунка 1.20 распределение расстояний Хэмминга для следующего поколения располагается много ближе к точке $h=0$, чем распределение исходного нулевого поколения. Операцию по синтезу новых поколений следует повторять до момента обнаружения в примера биометрического образа «Свой» с нулевым расстояние Хэмминга.

1.6 Энтропийно-корреляционная оценка хэширующих свойств нейросетевого преобразователя «биометрия-код доступа»

Разработанные в России и в Казахстане методы высоконадежной биометрии [4-10] основаны на использовании искусственных нейронных сетей больших размеров для обработки и одновременно безопасного применения персональных биометрических данных. Формальная запись применения нейронной сети имеет следующую форму:

$$[\text{НСП}] \cdot \vec{\xi}_N = \vec{x}_n \quad (1.34)$$

где [НСП] - матрица, описывающая обученную искусственную нейронную сеть преобразователя «биометрия-код»; $\vec{\xi}_N$ - вектор из N непрерывных входных биометрических параметров (число N может находиться в пределах от 512 для рукописного почерка до 2048 в случае рисунка сетчатки глаза); $\vec{x}_n$ - двоичный код отклика искусственной нейронной сети длиной n (обычно $n > N$).

Связывание нейросетевым преобразователем вектора непрерывных биометрических данных (континуумов) с вектором дискретных данных выходного кода иллюстрируется рисунком 1.22.

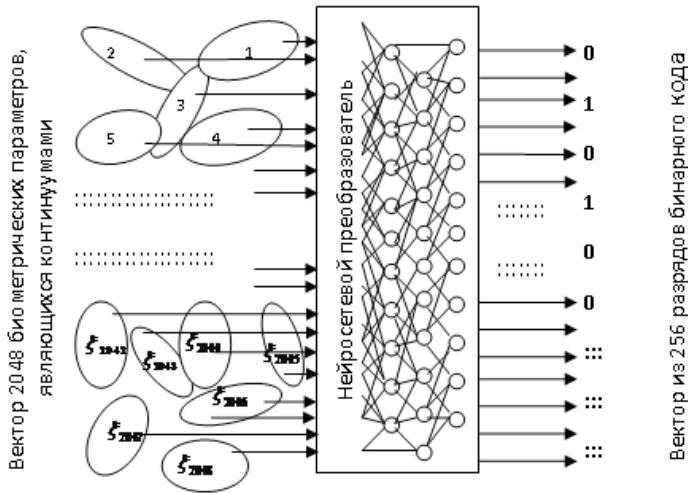


Рис. 1.22. Нейросетевое связывание вектора непрерывных биометрических параметров (вектора континуумов) с дискретным вектором выходного кода доступа

Необходимо отметить, что нейросетевое преобразование [НСП] не может быть представлено в виде классической линейной матрицы с постоянными коэффициентами. В типовом случае оно состоит из N независимых элементарных преобразований, описывающих

нелинейную свертку части пространства элементов вектора входных биометрических параметров $\bar{\xi}_N$.

В случае однослойной нейронной сети эта свертка отождествляется с преобразованием, выполняемым единичным нейроном (персептроном со ступенчатой функцией возбуждения). В случае двухслойной нейронной сети каждое элементарное нейросетевое преобразование является двойной сверткой, представимой первым слоем сверток (персептроном первого слоя) и одной последующей обобщающей пространственной сверткой (персептроном второго слоя).

В соответствии с [2] вектор входных биометрических параметров «Свой» - $\bar{\xi}_{N, \text{«Свой»}}$ должен приводить к появлению кода «Свой» на выходе нейросетевого преобразователя с высокой вероятностью (при этом вероятность ошибок первого рода, то есть вероятность ошибочного принятия образа «Свой» за образ «Чужой», составляет величину ≈ 0.05). В случае, когда биометрические параметры предъявляет «Чужой» в виде вектора - $\bar{\xi}_{N, \text{«Чужой»}}$, то вероятность ошибок второго рода, то есть вероятность ошибочного принятия образа «Чужой» за образ «Свой» должна быть сопоставима с ошибками парольных систем и составлять величину $\approx 10^{-10}$.

Для того чтобы добиться столь низкой вероятности ошибок второго рода в [2] рекомендуется использовать достаточно большое число выходов у нейросетевого преобразователя ($n \approx 256$). Кроме того, при воздействии случайными биометрическими образами «Чужой» состояния «0» и «1» каждого разряда выходного кода должны быть равновероятны. Другим важнейшим требованием к нейросетевому преобразователю является независимость (некоррелированность) разных разрядов выходного кода при воздействии на нейросетевую преобразователь случайными биометрическими векторами - $\bar{\xi}_{N, \text{«Чужой»}}$.

Нейросетевой преобразователь непрерывных биометрических данных в код для образа «Свой» фактически ведет себя как усилитель корреляционных связей биометрических данных. Для биометрических данных образа «Чужой» нейросетевой преобразователь «биометрия-код» осуществляет их хэширование (перемешивание), тем самым осуществляется ослабление его собственных

корреляционных связей. Для эффективно функционирующего нейросетевого преобразователя из-за отсутствия процедуры хэширования биометрических данных образа «Свой» значительные исходные корреляционные связи образа «Свой» преобразуются в выходные коды «Свой» практически с единичной корреляцией, то есть происходит снижение исходной энтропии до нуля и корреляционные связи увеличиваются до своего максимума (коэффициент корреляции составляет величину 1). Для образа «Чужой» хэширование наоборот приводит к увеличению энтропии.

Наличие корреляции между параметрами для образа «Свой» приводит к возможности реализации атак подбора биометрических параметров. Это значит, что для оценки стойкости нейросетевого преобразования к атакам подбора необходимо оценивать остаточные корреляционные связи между парами разрядов выходного кода. Именно эта задача решается разработанными стандартами ГОСТ Р 52633.1-2009 [3], ГОСТ Р 52633.2 [26] и ГОСТ Р 52633.3 [25]. По сути, решение этой задачи сводится к энтропийно-корреляционной оценке хэширующих свойств преобразователя «биометрия-код»

При этом могут быть использованы два показателя: коэффициент усиления энтропии или коэффициент ослабления парных корреляционных связей. Для образа «Свой» происходит снижение энтропии и усиление парных корреляционных связей, то есть коэффициенты усиления энтропии и ослабления парных корреляционных связей меньше единицы (в логарифмической шкале имеют отрицательные значения). Для образа «Чужой», наоборот, происходит увеличение энтропии и ослабление корреляционных связей, при этом показатели больше единицы (в логарифмической шкале имеют положительные значения). Идеальное хэширование для образа «Чужой» имеет место, когда корреляции получаются нулевые, а энтропия максимально возможная.

Коэффициент корреляции, с помощью которого могут быть оценены входные корреляционные связи биометрического образа «Свой» для разных пар биометрических параметров, определяются следующим образом:

$$r(\xi_i, \xi_j) = \frac{1}{\tilde{N}} \sum_{i=1}^{\tilde{N}} \frac{(\xi_i^{(n)} - E(\xi_i^{(n)})) \cdot (\xi_j^{(n)} - E(\xi_j^{(n)}))}{\sigma(\xi_i^{(n)}) \times \sigma(\xi_j^{(n)})} \quad (1.35)$$

где $E(.)$ - оператор вычисления математического ожидания;
 $\sigma(.)$ - оператор вычисления среднеквадратического отклонения;
 $x_i^{(n)}, x_j^{(n)}$ - значения i -го и j -го биометрических параметров в n -м эксперименте;

$\tilde{N}$ - число независимых экспериментов, обычно $\tilde{N}$ удается задать от 60 до 100.

Идентифицировать статистические свойства нейросетевого преобразования удается, вычисляя среднее значение этого показателя, при этом коэффициент ослабления парных корреляционных связей оценивается как «коэффициент усиления среднего значения» модуля коэффициента корреляции образов «Свой»:

$$K_{\text{Свой}}(r) = \frac{1}{E(|r(\xi_i, \xi_j)|)} \quad (1.36)$$

где $E(|r(\xi_i, \xi_j)|)$ - математическое ожидание модуля коэффициентов парной корреляции, которое вычисляется для несовпадающих между собой индексов j и i , при этом эти индексы выбираются случайно в интервале от 1 до N .

Эффект хэширования для образа «Чужой» описывается соответствующим коэффициентом ослабления:

$$K_{\text{Чужой}}(r) = \frac{E(|r(x_i, x_j)|)}{E(|r(\xi_i, \xi_j)|)} \quad (1.37)$$

Отметим, что в числителе выражения (1.37) находится математическое ожидание модулей коэффициентов корреляционных связей разрядов кода «Чужой». При этом вычисление коэффициентов корреляционных связей дискретных параметров осуществляется несколько иначе, чем непрерывных параметров по формуле (1.35).

В общем случае коэффициент корреляции двух дискретных разрядов выходного кода - x_1, x_2 вычисляется по следующей формуле:

$$r(x_1, x_2) = r(\tilde{x}_1, \tilde{x}_2) = \frac{1}{\tilde{N}} \sum_{i=0}^{\tilde{N}} \frac{(\tilde{x}_{1,i} - E(\tilde{x}_1)) \cdot (\tilde{x}_{2,i} - E(\tilde{x}_2))}{\sigma(\tilde{x}_1) \cdot \sigma(\tilde{x}_2)} \quad (1.38)$$

где $\tilde{x}_1 = 1$ при $x_1 = 1$ и $\tilde{x}_1 = -1$ при $x_1 = 0$;

$\tilde{x}_2 = 1$ при $x_2 = 1$ и $\tilde{x}_2 = -1$ при $x_2 = 0$ (в данном случае проводится преобразование в новую систему координат $x \rightarrow \tilde{x}$).

Следует отметить, что для правильно обученного нейросетевого преобразователя при переходе в новую систему координат математические ожидания состояния каждого их разрядов выходного кода должны оказаться нулевыми ($E(\tilde{x})=0$), а среднеквадратические отклонения – единичными ($\sigma(\tilde{x})=1$). Это существенно упрощает формулу (1.38).

Вычисление парных корреляций вида (1.38) не является вычислительно сложной задачей. В случае, если численный эксперимент надежно показывает отсутствие корреляционных связей между произвольно выбранными парами разрядов, стойкость нейросетевого преобразования к атакам подбора будет сопоставима со стойкостью криптографического ключа длиной n . То есть при $n = 256$ эта стойкость будет составлять 2^{256} , что является слишком оптимистической оценкой. Практика показывает, что при обучении искусственных нейронных сетей большого размера полностью устранить корреляцию между парами разрядов не удастся, и именно по этой причине происходит значительная утрата стойкости к атакам подбора. Очень малая вероятность ошибок второго рода, составляющая величину $\approx 2^{-256}$, характерная для полностью независимых выходных кодов, увеличивается до величины $\approx 2^{-33}$, если появляются корреляционные связи между разрядами со средним значением их модуля более 0,15.

По аналогии с коэффициентом ослабления корреляционных связей (1.37) можно было бы использовать коэффициент усиления энтропии, однако вычислить его оказывается очень сложно, так как, во-первых, сегодня практически невозможно рассчитать высокоразмерную энтропию существенно зависимых непрерывных входных биометрических данных, а во-вторых, нужны специальные процедуры по оценке почти нулевого значения энтропии выходных кодов для образа «Свой». В этом отношении коэффициент ослабления парных корреляций нейронной сетью является более удобным для практики показателем, он, по сути, является некоторым аналогом коэффициента усиления энтропии биометрического образа «Свой».

Необходимо отметить, что обычный многослойный нейросетевой преобразователь «биометрия-код» не обеспечивает высокого уровня хэширования. Для того чтобы поднять качество хэширования, могут быть использованы различные механизмы размножения ошибок. Например, может быть использован механизм сложения по модулю 2 части выходного кода «Свой» и записанных в дискретной форме параметров обученной нейронной сети. То есть после обучения искусственной нейронной сети производят самошифрование части ее параметров (весовых коэффициентов нейронов и номеров связей между нейронами) сложением по модулю два с частью разрядов выходного кода. Блок-схема этой процедуры приведена на рис. 1.23.

Параметры первого нейрона хранятся открытыми, а параметры следующего нейрона уже оказываются накрыты суммированием по модулю два с выходным кодом первого нейрона. Параметры третьего нейрона накрывают суммой по модулю два выходных данных двух первых нейронов.

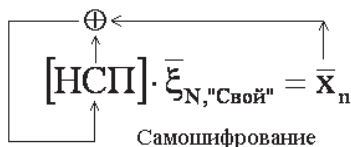


Рис. 1.23. Шифрование параметров нейронной сети на фрагментах выходного кода

Шифрование ведут циклически, накрывая суммированием по модулю два выходными данными уже сработавших нейронов параметры (веса и номера связей) следующих еще не задействованных нейронов.

Из-за того, что образ «Свой» дает один и тот же выходной код, процедура расшифрования связей нейронов оказывается симметричной. Ее условная блок-схема отображена на рис. 1.24.



Рис. 1.24. Обратная процедура расшифрования параметров нейронной сети на фрагментах выходного кода

Так как для образа «Свой» выходные коды при самошифровании и саморасшифровании одинаковы, описанная выше процедура размножения ошибок не мешает положительной биометрической аутентификации. Иначе обстоит дело для образов «Чужой». Они дают другой выходной код (этот код не совпадает с кодом «Свой», использованным ранее для самошифрования). В итоге запускается механизм размножения ошибок для биометрических образов «Чужой», что приводит к практически идеальному хешированию оставшейся части выходного кода. Эта ситуация отображается блок-схемой, приведенной на рисунке 1.25.

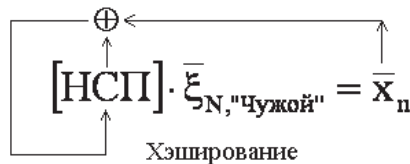


Рис. 1.25. Схема хэширования выходных данных нейросетевого преобразователя при предъявлении ему случайного вектора биометрических параметров «Чужой» с последующим эффектом размножения ошибок

Механизм безопасного объединения биометрической аутентификации и аутентификации на дополнительном пароле строится на основе сложения по модулю два дополнительного пароля с выходными кодами защищенного самошифрования нейросетевого преобразователя (рисунок 1.26).

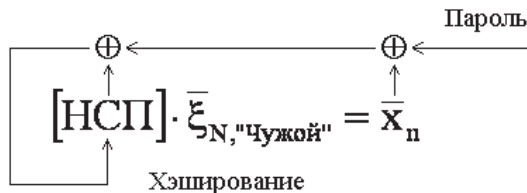


Рис. 1.26. Безопасное объединение нейросетевой биометрической аутентификации с внешним паролем или иной биометрической процедурой нейросетевой аутентификации

Подчеркнем, что хэширующие свойства нейронной сети, реализованные в соответствии с мнемоническими схемами рисунков 1.25 и 1.26, уже сложно оценить с использованием коэффициента ослабления (усиления) корреляционных связей, так как после хэширования парные корреляции между разрядами становятся практически нулевыми и корреляционные оценки оказываются малоэффективными. В этом случае для описания хэширующих свойств нейросети с самошифрованием можно использовать только коэффициент усиления энтропии [28]. Однако, как было отмечено выше, сегодня практически невозможно рассчитать высокоразмерную энтропию существенно зависимых непрерывных входных биометрических данных, поэтому применяется прием, когда сравнивается энтропия выходных кодов преобразователя, реализованного в соответствии с мнемоническими схемами рисунков 1.25 и 1.26, с энтропией незащищенного преобразователя, то есть определяется «коэффициент усиления энтропии» в виде:

$$K_{\text{Чужой}}^{(H)} = \frac{H_{\text{Самошифр}}}{H_{\text{Незащищ}}} = \frac{n}{H_{\text{Незащищ}}} \quad (1.39)$$

где $H_{\text{Самошифр}}$ - энтропия выходных кодов преобразователя «биометрия-код» с самошифрованием (совпадает с длиной выходного кода n);

$H_{\text{Незащищ}}$ - энтропия незащищенного преобразователя «биометрия-код».

Экспериментальные исследования и расчеты показали, например, что при среднем значении модулей корреляции (и соответственно коэффициенте ослабления корреляционных связей), равном 0.15, коды преобразователя «биометрия-код» с 256 выходами имеют энтропию порядка 25 бит. То есть переход к самошифрованию данных нейросетевого преобразователя дает коэффициент усиления энтропии (1.39), равный примерно 0.1 от максимально возможного значения энтропии кодов с некоррелированными разрядами и близкий к значению коэффициента ослабления корреляционных связей. В этом отношении коэффициенты усиления (ослабления) корреляций и энтропии дополняют друг друга для случаев, когда можно использовать лишь один из указанных показателей.

Таким образом, энтропийная или корреляционная оценка хэширующих свойств нейросетевого преобразователя «биометрия-код доступа» путем расчета коэффициентов усиления энтропии и ослабления парных корреляционных связей позволяет на количественной основе оценивать качество его функционирования и сравнивать стойкость таких преобразователей к атакам подбора биометрических параметров.

1.7 Аппроксимация биномиального зависимого закона композициями нормального, равномерного, арксинусного распределений значений

В настоящее время активно развивается технология разграничения доступа к информационным ресурсам с использованием механизмов аутентификации пользователя по его уникальным биометрическим данным: рисунку отпечатка пальца, почерку, голосу, сетчатке и радужной оболочке глаза, 2-D и 3-D лица т.д.

При этом зарубежные исследователи получают код аутентификации пользователя из входного биометрического образа [22; 23].

Впервые в России в рамках комплексного проекта «Разработка и подготовка производства телекоммуникационного оборудования, разработка программного сетевого, прикладного и специального обеспечения для создания цифровых сетей связи с персонализированным доступом», непосредственными исполнителями которого являлись и все авторы настоящей монографии, созданы промышленные образцы средств высоконадежной биометрико-нейросетевой аутентификации личности. В соответствии с ГОСТ Р 52633.0-2006 эти средства строятся на основе нейросетевых преобразователей биометрия-код [2]. В данных преобразователях происходит смешивание данных биометрического образа с кодом аутентификации пользователя. Обучение преобразователей биометрия-код должно производиться в соответствии с требованиями ГОСТ Р 52633.5 [1]. Входными данными обучения являются параметры биометрических образов пользователей «Свой» и «Чужие». На выходах обученного преобразователя биометрия-код

формируются код-отклик пользователя «Свой» при подаче на входы преобразователя образов «Свой». При подаче на входы преобразователя образов «Чужие» на его выходах формируются случайные коды-отклики «Чужие». При этом, в соответствии с ГОСТ Р [2], после процедуры обучения обязательно проводится процедура тестирования с использованием встроенной в средство аутентификации тестовой базы.

Вероятность ложной аутентификации личности при использовании преобразователя биометрия-код с 416 входами и 256 выходами составляет примерно 10-9 (данные для рукописного слова-пароля из пяти знаков) [1].

Процедуры тестирования средств высоконадежной биометрической аутентификации выполняются исходя из оценки статистического распределения вероятностей ложной аутентификации пользователей «Свой» и «Чужие» [29].

В работе [19] показано, что на практике выходные данные нейросетевых преобразователей биометрия-код являются коррелированными. Одной из причин являются наличие корреляционных связей между входными параметрами преобразователя. Другая причина состоит в структурной организации связей между нейронами преобразователя. Моделирование корреляционных связей выходных данных биометрико-нейросетевого преобразователя выполнено в работе [18]. В данной работе показано, что распределение вероятностей ошибок ложной аутентификации пользователей «Чужие» может быть описано биномиальным зависимым законом распределения значений. Вероятности ложной аутентификации пользователя «Свой» оценивается через дробный показатель степеней свободы закона распределения значений хи-квадрат [30].

Рост корреляционных связей между разрядами выходных кодов приводит к росту среднеквадратического отклонения значений расстояний Хэмминга между кодом «Свой» и случайными выходными кодами «Чужие». Монотонный рост корреляции между выходными разрядами кодов «Чужие» приводит к соответствующему росту среднеквадратического отклонения расстояний Хэмминга. Соответствующие распределения расстояний Хэмминга приведены на рисунке 1.27.

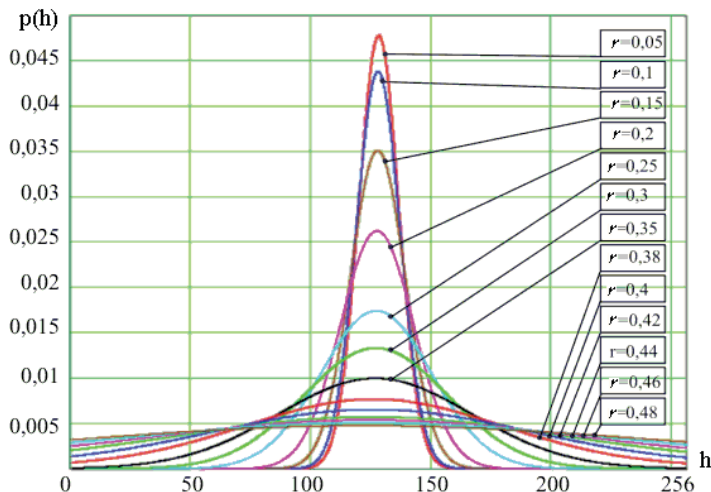


Рис. 1.27. Эволюция плотности распределения значений меры Хэмминга при монотонном росте модуля корреляционных связей между выходными разрядами кодов преобразователя

Получается, что при монотонном росте модулей корреляционной связи в интервале от $r=0.0$ до $r=0.63$ форма закона распределения значений меры Хэмминга эволюционирует, проходя три разных закона распределения: нормальный закон, равномерный закон и закон арксинуса. Формы всех трех этих закона приведены на рисунке 1.28.

К сожалению, сегодня отсутствует точное аналитическое описание эволюции плотности распределения значений меры Хэмминга, соответствующей плотности распределения значений биномиального зависимого закона. На сегодняшний день имеются только таблицы аппроксимации этих плотностей, построенные для разных значений коэффициентов корреляции или среднеквадратического отклонения.

Аппроксимировать плотности распределения значений биномиального зависимого закона удобно в нормированной метрике Хэмминга, изменяющейся в пределах от 0 до 1.

Тогда приближающее распределение для интервала модулей корреляции от 0.0 до 0.37 записывается следующим образом:

$$p(h, E(h), \sigma(h)) = \frac{1}{\sqrt{2\pi} \cdot \sigma(h)} \exp \left\{ \frac{-(E(h) - h)^2}{2 \cdot \{\sigma(h)\}^2} \right\} \quad (1.40)$$

где $p(h, E(h), \sigma(h))$ – псевдо-непрерывная плотность распределения значений меры Хэмминга – h , которая на самом деле является огибающей дискретного распределения;

$\sigma(h)$ – среднее квадратическое отклонение нормированной меры Хэмминга;

$E(h)$ – математическое ожидание значений нормированной меры Хэмминга.

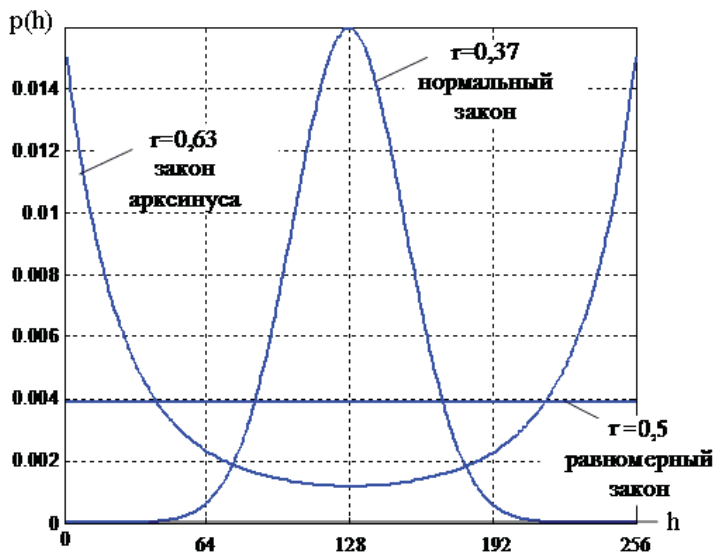


Рис. 1.28. Формы трех классических законов, являющихся этапами эволюции биномиального закона распределения значений зависимых данных

В интервале корреляции разрядов 256-битных кодов от 0.35 до 0.5 распределение расстояний Хэмминга хорошо описывается следующим выражением:

$$p(h) \approx a(r) \cdot \frac{1}{257} + \frac{(1-a(r)) \cdot \exp\left(\frac{-(128-h)^2}{2\sigma^2(h) \cdot (1-a^2(r))}\right)}{\sqrt{2\pi} \cdot \sigma(h) \cdot (1-a(r))} \quad (1.41)$$

где

$$a(r) = \frac{r-0.35}{0.15} \quad (1.42)$$

Эволюция плотности распределения значений меры Хэмминга при монотонном росте модуля корреляционных связей в пределах от $r=0.5$ до $r=0.92$ приведена на рисунке 1.29.

В интервале коррелированности разрядов 256-битных кодов от $r=0.5$ до $r=0.87$ распределение расстояний Хэмминга хорошо описывается следующим приближением:

$$p(h) \approx a(r) \cdot \frac{1}{257} + \frac{1-a(r)}{\pi \sqrt{h(1-h)}} \quad (1.43)$$

где

$$a(r) = \frac{r-0.87}{0.37} \quad (1.44)$$

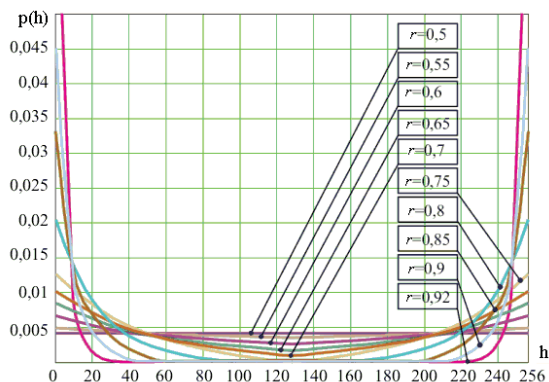


Рис. 1.29. Эволюция плотности распределения значений меры Хэмминга при монотонном росте модуля корреляционных связей в пределах от $r=0.5$ до $r=0.92$

При изменении модулей коэффициентов корреляции в интервале от 0.92 до 1.0 хорошая аппроксимация плотности распределения значений биномиального зависимого закона получается при приближении правой и левой ветви хи-квадрат распределениями, имеющими малые значения дробных показатели числа степеней свободы [30]:

$$p(h, E(h), \sigma(h)) = \frac{1}{2^{\frac{m}{2}+1} \Gamma\left(\frac{m}{2}\right)} \left\{ h^{\frac{m}{2}-1} \cdot \exp\left(\frac{-h}{2}\right) + |1-h|^{\frac{m}{2}-1} \cdot \exp\left(\frac{-|1-h|}{2}\right) \right\} \quad (1.45)$$

где m – малое значение дробного показателя числа степеней свободы (число m может быть менее 1.0, но всегда больше 0.0), подбираемое при аппроксимации плотности распределения; $\Gamma(.)$ – гамма функция.

По выражениям (1.40-1.45) может быть оценена вероятность ложной аутентификации «Чужого» обученного нейросетевого преобразователя биометрия-код с 256 выходами сбалансированного по вероятности появления «0» в разряде кода. Распределения, получаемые по этим выражениям, являются симметричными относительно математического ожидания нормированной меры Хэмминга.

Эти распределения может использоваться в методиках тестирования сбалансированных биометрико-нейросетевых преобразователей изложенных в [1]. Ошибка оценки вероятности ложной аутентификации равна 0,03% от 256 бит. При несбалансированности по вероятности появления «0» в разряде кода необходимо синтезировать таблицы биномиального зависимого закона распределения значений с учетом небаланса по этому параметру. Синтезировать таблицы вероятностей ложной аутентификации пользователей «Все Чужие» необходимо на основе результатов моделирования биномиального зависимого закона распределения значений полученных в [18]. Она учитывает небаланс распределения нормированной меры Хэмминга относительно ее математического ожидания вероятностей появления «0» в бите кода и эффект больших хвостов при значениях корреляции стремящейся к единице. Исходя из требований [2] шаг по корреляции необходимо выбирать равным 0,01. Исходя из наиболее часто встречающихся на практике ситуаций значений параметров тестируемых нейросетевого преобразователя биометрия-код шаг по вероятности появления «0» в бите кода нужно выбирать равным 0,01.

1.8 Моделирование длинных биометрических кодов, воспроизводящих корреляционные связи выходных данных нейросетевого преобразователя

Нейросетевые преобразователи биометрия-код при воздействии на них образами «Чужой-1», «Чужой-2», ..., «Чужой-N» дают псевдослучайные коды длиной 256 бит с существенно коррелированными разрядами. Возникает необходимость в моделировании таких кодовых последовательностей для оценки связи их энтропии с коэффициентами парной корреляции.

Моделирование псевдослучайных кодов с полностью независимыми разрядами является хорошо изученной задачей. Обычно независимые коды получают через хеширование данных или путем использования рекуррентных последовательностей [31; 32]. Однако при попытке воспроизведения последовательности с зависимыми (коррелированными) разрядами возникают существенные трудности. В частности, если для этой цели использовать специальные матричные преобразования [33], то невозможно создавать зависимые коды большой длины. Методом, изложенным в [33], нельзя получить вектор из 256 псевдослучайных величин с заданной корреляционной матрицей. В связи с этим в работах [14; 29] предложено использовать упрощенный вариант связывающих матриц. Предложено использовать связывающие матрицы с единичной диагональю и параметром «а» вне диагонали:

$$\begin{bmatrix} 1 & a & \dots & a \\ a & 1 & \dots & a \\ \vdots & \vdots & \ddots & \vdots \\ a & a & \dots & 1 \end{bmatrix} \times \begin{bmatrix} x_{1,i} \\ x_{2,i} \\ \vdots \\ x_{n,i} \end{bmatrix} = \begin{bmatrix} y_{1,i} \\ y_{2,i} \\ \vdots \\ y_{n,i} \end{bmatrix} \Rightarrow R = \begin{bmatrix} 1 & r & \dots & r \\ r & 1 & \dots & r \\ \vdots & \vdots & \ddots & \vdots \\ r & r & \dots & 1 \end{bmatrix} \quad (1.46)$$

Для таких матриц исходный вектор псевдослучайных данных независимых данных - $\bar{X}_i$ с одинаковыми распределениями превращается в вектор псевдослучайных зависимых данных - $\bar{Y}_i$, так же имеющий одинаковые нормальные распределения значений.

При этом корреляционная матрица зависимости таких данных оказывается крайне простой (все коэффициенты парной корреляции оказываются одинаковыми).

Для получения данных со случайно изменяющимися знаками равных по модулю коэффициентов корреляции, достаточно умножить на «-1» половину столбцов или строк связывающей матрицы (1.46).

1.8.1 Аналитическая связь коэффициентов равной корреляции и параметра связывающей матрицы

При воспроизведении длинных векторов псевдослучайных данных можно пользоваться различными генераторами вектора исходных независимых случайных величин $\bar{X}_i$. Наиболее часто в биометрических приложениях используют генераторы псевдослучайных данных с нормальным законом распределения значений или с равномерным законом распределения значений.

Если исходные псевдослучайные данные имеют симметричный относительно математического ожидания закон распределения значений, то для этого случая существует аналитическая связь между коэффициентами равной корреляции и регулируемым параметром:

$$r = \frac{2 \cdot a + (n - 2) \cdot a^2}{1 + (n - 1) \cdot a^2} \quad (1.47)$$

Столь простая аналитическая зависимость (1.47) делает использование генераторов псевдослучайных данных с симметричными плотностями распределения значений предпочтительным. На рисунке 1.30 приведена номограмма связи регулируемого параметра с одинаковым модулем элементов корреляционной матрицы разной размерности.

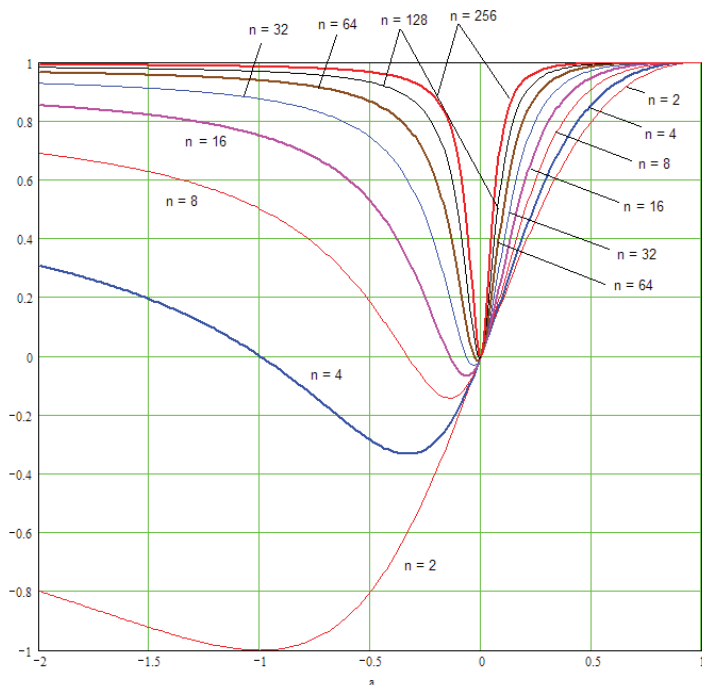


Рис. 1.30. Номограмма связи регулируемого параметра «а» связывающей матрицы со значением модуля парных корреляционных связей выходных данных.

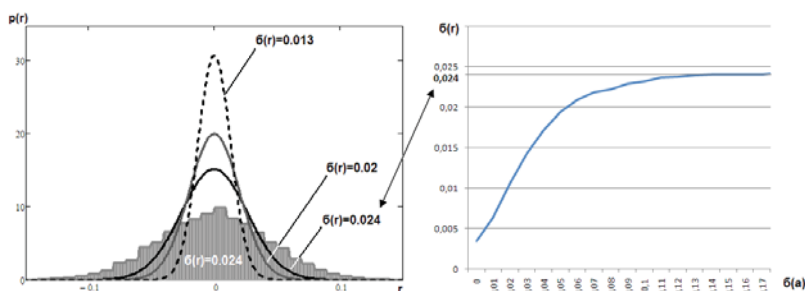
В случае если исходные данные имеют несимметричный закон распределения, значений простой аналитической зависимости вида (1.47) получить не удастся. В связи с этим приходится восстанавливать зависимость $r(a, n)$ численно, что несколько увеличивает ошибки моделирования.

1.8.2 Эффект размывания параметра «а=0» дополнительным псевдослучайным сигналом

Измерение параметров распределения коэффициентов парных корреляций выходных разрядов биометрия-код дает соответствующее распределение их значений. Математическое ожидание этого распределения нулевое. Для воспроизведения реальных значения параметров наблюдаемого распределения значений ко-

эффицентов парной корреляции необходимо иметь процедуры, способные регулировать среднеквадратическое отклонение моделируемых данных.

Наиболее простым способом моделирования является подстановка в качестве параметров связывающей матрицы случайных коэффициентов с нулевым математическим ожиданием ($E(a)=0$) и регулируемой дисперсией. В этом случае мы получаем данные с распределениями коэффициентов корреляции, отображенными в левой части рисунка 1.31.



(правая часть рисунка) и график связи среднеквадратических отклонений размываемого параметра «а» с этим же параметром распределения значений парных коэффициентов корреляции.

Из рисунка 1.31 видно, что при малых значениях размываемых данных распределение значений парных коэффициентов корреляции является нормальным, однако с ростом среднеквадратического отклонения $b(r)$ происходит отход от нормального закона распределения значений. В предельном случае $b(r)=0.024$ распределение становится существенно отличным от нормального. В левой части рисунка 1.31 полученное распределение значений отображено гистограммой, соответствующий этой гистограмме нормальный закон отображен утолщенной линией.

В связи с тем, что среднеквадратические отклонения моделируемых параметров и входных данных связаны кривой с явно выраженным насыщением (правая часть рисунка 1.31), регулировать среднеквадратические отклонения распределения коэффициентов парной корреляции в широких пределах не удастся.

1.8.3 Эффект размывания параметра «а» и «-а» дополнительным псевдослучайным сигналом

Попытка расширить динамический диапазон коэффициентов парных корреляции неминуемо приводят к тому, что приходится изменять знаки на противоположные у половины строк (столбцов) связывающей матрицы (1.46) и использовать не нулевое математическое ожидание регулируемого параметра «а». В этом случае распределение значений коэффициентов корреляции становится двумодальным, как это показано на рисунке 1.32.

Модуль математического ожидания входного размываемого параметра «а» близок к статическому значению этого параметра вычисленному через преобразование (1.47). Из рисунка 1.32 видно, что по мере увеличения математического параметра «а» растет параметр $\hat{b}(r)$, однако при этом увеличивается провал распределения в точке $r=0$.

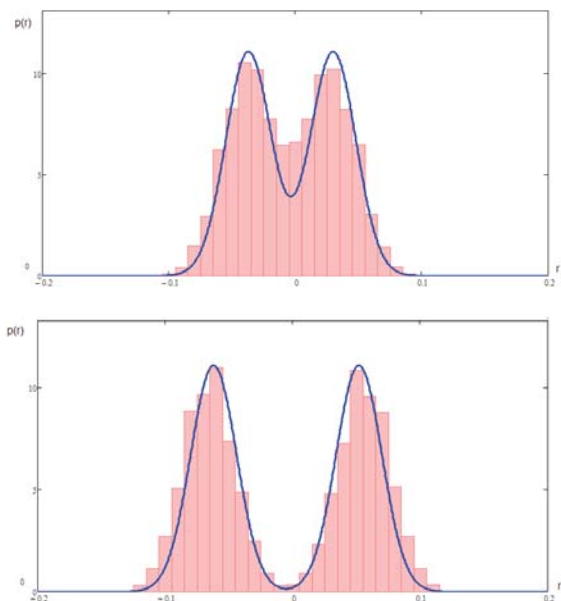


Рис. 1.32. Примеры двумодальных распределений значений коэффициентов корреляции

Таким образом, метод упрощенного моделирования многомерных данных через привлечение высокоразмерных связывающих матриц вида (1.46) не позволяет получать распределения значений парных коэффициентов корреляции с нормальным законом и большим среднеквадратическим отклонением. Попытки увеличения среднеквадратического отклонения моделируемых распределений до величины более 0.024 неминуемо приводят к появлению двумодальности. Воспроизводить нормальный (одномодальный) закон распределения значений столь простым способом не удастся. С одной стороны, связывающие корреляционные матрицы вида (1.46) позволяют моделировать коррелированные данные очень высоких размерностей, однако при этом воспроизводить распределения значений близкие к нормальным для высоких значений $b(r)$ не удастся.

1.9 Учет «тяжелых хвостов» ненормального закона распределения биометрических параметров все «Чужие» при настройке нелинейного элемента нейрона с несколькими дискретными состояниями

По стандарту ГОСТ Р 52633.5 [1] первый слой нейронов преобразователей биометрия-код одновременно осуществляет обогащение биометрических данных и их дискретизацию, при этом качество обогащения биометрических данных ограничивается. Возможны ситуации, когда используется предварительное обогащение данных в виде так называемых «нечетких» экстракторов [22; 24]. Возможна и другая ситуация, в соответствии с которой при обучении искусственных нейронов сняты ограничения на уровень обогащения биометрических данных.

При высоком уровне первичного обогащения данных обыкновенные нейроны (персептроны) оказываются малоэффективны. Более эффективными оказываются искусственные нейроны с несколькими выходными дискретными состояниями [34]. Пример такого нейрона с нелинейным элементом, дающим пять устойчивых состояния, приведен на рисунке 1.33.

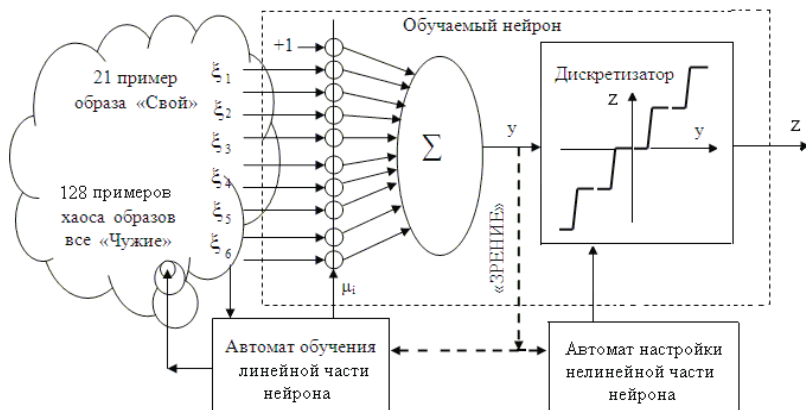


Рис. 1.33. Обучение искусственного нейрона с несколькими выходными дискретными состояниями двумя автоматами

Эффективность применения нейронов с несколькими выходными дискретными состояниями обусловлена тем, что один нейрон дает не два, а более возможных состояний кода (3, 4, 5, ... выходных состояний), что на тех же данных дает возможность получить ключ большей длины [35].

1.9.1 Настойка параметров нелинейной части нейрона

Необходимость в специальной настройке параметров нелинейности нейрона обусловлена тем, что предварительное обогащение исходной информации всегда дает конкретное распределение множеств «Свой» и все «Чужие». Пример такого распределения приведен на рисунке 1.34.

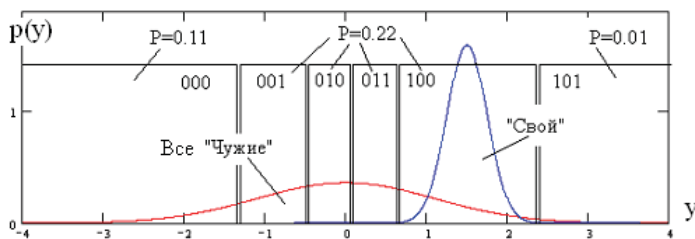


Рис. 1.34. Пример плотностей распределения значений «Свой» и все «Чужие» на выходе хорошо обученной линейной части нейрона

Из теории известно, что энтропия дискретных выходных состояний нейрона с несколькими дискретными состояниями будет максимальна, если каждое из этих дискретных состояний при воздействии образами все «Чужие» будет появляться с равными вероятностями. Из рисунка 1.34 видно, что динамический диапазон изменения выходных состояний линейной части нейрона разбит на шесть поддиапазонов, в пятом поддиапазоне находится множество «Свой». В пятый поддиапазон множества «Свой» образы все «Чужие» попадают с вероятностью $P_5=0.22$. То есть, стремясь к максимальному значению энтропии выходных решений нейрона, мы должны стремиться сделать вероятность попаданий образов все «Чужие» $P_1=P_2=P_3=P_4=P_6=0.22$. Для ситуации, изображенной на рисунке 1.34, эта задача имеет решение только для $P_2=P_3=P_4=0.22$. Крайние динамические диапазоны $P_1=0.11 \neq 0.22$, $P_6=0.01 \neq 0.22$ не могут давать нужную вероятность.

Таким образом, автомат настройки параметров нелинейного элемента нейрона должен быть способен вычислять оптимальное число устойчивых состояний функции дискретизации и уметь вычислять оптимальную ширину нелинейных участков выходной функции дискретизации.

Следует подчеркнуть, что расчет ширины интервалов дискретизации и их количества крайне просто осуществляется, если множество образов все «Чужие» описывается нормальным законом распределения значений. Однако, практика показала, что распределение все «Чужие» существенно ненормально (имеет «тяжелые хвосты»). Пример реального распределения все «Чужие» на выходе сумматора обученного нейрона приведен на рисунке 1.35.

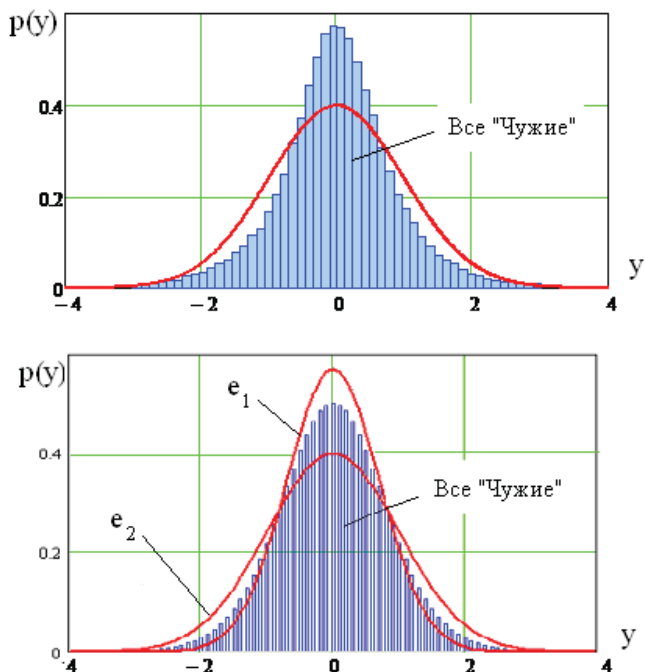


Рис. 1.35. Распределение все «Чужие» на выходе обученного сумматора, существенно отличающееся от действительного нормального

Из верхней части рисунка 1.35 видно, что реальное распределение данных все «Чужие» на выходе обученного нейрона существенно отличается от нормального. Несмотря на то, что сумматор имеет 24 входа, полной нормализации входных данных не происходит и приходится при расчетах использовать смесь двух нормальных законов (нижняя часть рисунка 1.35).

При аппроксимации реального распределения все «Чужие» используется процедура подбора трех параметров:

$$p(y) = \frac{e_1}{\sigma(e_1)\sqrt{2\pi}} \exp\left(\frac{-y^2}{2\sigma^2(e_1)}\right) + \frac{(1-e_1)}{\sigma(e_2)\sqrt{2\pi}} \exp\left(\frac{-y^2}{2\sigma^2(e_2)}\right) \quad (1.48)$$

где e_1 – вероятность появления в смеси данных первого нор-

мального закона распределения значений; $\sigma(e_1)$ - среднее квадратическое отклонение первого закона распределения значений; $e_2 = (1 - e_1)$ - вероятность появления в смеси данных второго нормального закона распределения значений; $\sigma(e_2)$ - среднее квадратическое отклонение второго закона распределения значений.

Подбор параметров аппроксимации осуществляется за счет минимизации интегрального значения модуля разности между выражением (1.48) и реальным распределением значений все «Чужие». Подбор осуществляется пропорциональным увеличением (уменьшением) среднее квадратических отклонений первого и второго распределения смеси:

$$\begin{cases} \sigma(e_1) = \sigma(e) \cdot (1 - \alpha); \\ \sigma(e_2) = \sigma(e) \cdot (1 + \alpha) \end{cases} \quad (1.49)$$

где $\sigma(e)$ - среднее квадратическое отклонение реального распределения все «Чужие»;

α - параметр подбора.

Принятое при аппроксимации условие симметрии (1.49) позволяет снизить число подбираемых параметров до двух (подбираются только параметры e_1 и α), что делает решаемую задачу устойчивой.

В итоге мы получаем аналитическое описание имеющихся ненормальных данных все «Чужие», пользуясь которым мы можем численно найти ширину всех соседних интервалов к интервалу «Свой», добиваясь заданной вероятности попадания в эти интервалы множества все «Чужие».

1.9.2 Регулирование формы нелинейного элемента под заданное из вне значение фрагмента кода «Свой»

Следует отметить, что преобразователи биометрия-код развиваются в двух направлениях. Отечественные нейросетевые преобразователи по ГОСТ Р 52633.0-2006 [2] при обучении нейронной сети получают личный ключ пользователя из вне. Зарубежные исследователи создают преобразователи биометрия-код с ключом, который извлекается из уникальности того или иного

биометрического образа [24; 34]. То есть по зарубежным технологиям до преобразования биометрия-код ключ неизвестен и после его определения он нуждается в проверке на соответствие требованиям, предъявляемым к ключам.

Изложенный выше подход к настройке параметров монотонно-ступенчатой нелинейности нейрона является аналогом того, что предлагают специалисты США и стран НАТО. В соответствии с этим подходом нейрон со сложным элементом (рисунок 1.33) будет давать трехбитный фрагмент ключа «Свой», имеющий двоичную форму «100», как это показано на рисунке (1.34).

Для того, чтобы выполнить требования базового национального стандарта ГОСТ Р 52633.0-2006 [2] и иметь возможность задавать ключ из вне, необходимо отказаться от применения простейших одинаковых монотонных ступенчатых нелинейных элементов дискретизации. Так если нам в место трех бит ключа «Свой» - «100», требуется другая комбинация, например, «000», то мы можем циклической перестановкой кодов рисунка 1.34 добиться нужного выходного кода при воздействии на нейрон образами «Свой». Подобная настройка не меняет положение и соотношение ранее найденных поддиапазонов дискретизации. Как показано в левой части рисунка 1.36, циклический сдвиг кодов эквивалентен некоторой резке нелинейного элемента и смещению его исходных фрагментов вверх или в низ со сдвигом.

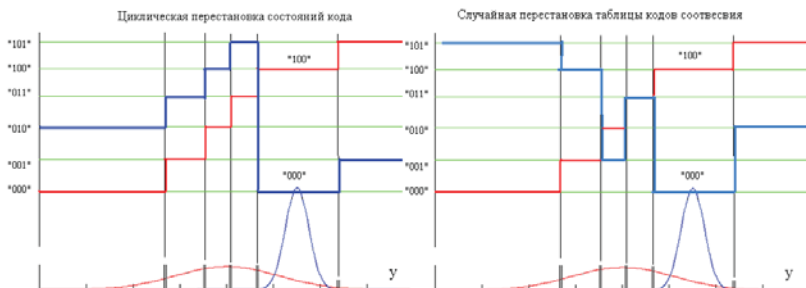


Рис. 1.36. Два способа синтеза нелинейных элементов дискретизации с учетом заданного значения выходного кода «Свой»

Практика показала, что использование операции циклического сдвига при настройке формы нелинейного элемента нейрона не является оптимальной. Более качественные результаты получаются, если использовать случайную перестановку кодов, соответствующих участкам нелинейности вне участка «Свой». При таком способе изменения форму нелинейного элемента усиливаются хеширующие свойства обученного нейрона по отношению к образам все «Чужие» [20]. Нелинейные элементы с длительными монотонными участками хуже хешируют (перемешивают) данные все «Чужие».

2. АЛГОРИТМЫ ТЕСТИРОВАНИЯ, ПРАКТИЧЕСКАЯ РЕАЛИЗАЦИЯ И ЗАЩИТА МЕХАНИЗМОВ БИОМЕТРИКО-НЕЙРОСЕТЕВОЙ АУТЕНТИФИКАЦИИ ЛИЧНОСТИ

2.1 Быстрые алгоритмы тестирования средств высоконадежной биометрической аутентификации личности

Для обеспечения высоконадежной идентификации человека необходимо привлекать биометрические технологии, способные безопасно взаимодействовать с криптографическими механизмами. В настоящее время практически все страны, имеющие значимый национальный научно-технический потенциал пытаются решать эти задачи. Две страны Россия и США, являясь лидерами технологий защиты информации, открыто публикуют результаты своих исследований. США идут по пути использования нечеткой математики [36], ученые этой страны предлагают мировому сообществу специализированные «fuzzy» обогатители (экстракторы) превращающие бедную неоднозначную размытую биометрическую информацию в сильный личный ключ пользователя.

Россия предлагает мировому сообществу иной путь использования больших и сверхбольших искусственных нейронных сетей, которые заранее обучаются преобразовывать размытые биометрические данные пользователя в его личный криптографический ключ. Теория создания подобных преобразователей биометрия/код [5; 36] позволяет надеяться на их высокую стойкость по отношению к атакам подбора и попыткам изучения. Правильно построенный преобразователь биометрия/код ведет себя как классическая необратимая хэш-функция. При этом достаточно сложная нейронная сеть с 256 выходами позволяет обеспечивать стойкость к атакам случайного подбора на уровне 10^{22} (22 степень) попыток. Для того, чтобы проверить качество российских биометрико-нейросетевых преобразователей, потребуется сформировать огромные базы случайных биометрических образов.

Исследования, проведенные российскими и казахскими учеными на базе Межотраслевой лаборатории тестирования био-

метрических устройств и технологий ПГУ и научно-исследовательской лаборатории по созданию и тестированию средств для биометрии КазНТУ, показали, что на воспроизведение одного случайного биометрического образа, например: рукописного пароля, у человека уходит порядка 10 секунд. То есть на формирование необходимой базы случайных биометрических образов, предназначенной для тестирования высоконадежной биометрии, силами одного добровольца уйдет 10^{17} лет. Даже если привлечь к работе все население Пензенской области (1,42 млн. жителей) сроки создания базы сократятся до 10^{11} лет, что так же неприемлемо. Из подобных достаточно простых расчетов вытекает бесперспективность всех попыток оценки стойкости современных биометрико-нейросетевых преобразователей прямым численным экспериментом. Эта задача сопоставима по своей сложности с задачей подбора криптографического ключа.

В России разработан и с 1 апреля 2007 года вступил в силу ГОСТ Р 52633 – 2006 «Защита информации. Техника защиты информации. Требования к высоконадежным биометрическим средствам аутентификации». Одно из главных требований ГОСТа, является наличие средств встроенного контроля вероятности ошибок средств защиты или вероятности удачи атаки подбора. Проблема состоит в том, что биометрический пароль (рукописный или голосовой) пользователь должен сохранить в тайне от всех. Пользователю предлагается удобная цифровая комбинация, слово, фразу. При этом ему нельзя доверять среднестатистическим характеристикам, заявленным производителем.

Тайный биометрический образ может оказаться слабым и обеспечивать низкую стойкость защиты к атакам подбора. Чтобы убедиться в стойкости нейросетевой защиты на конкретном биометрическом образе после обучения нейросети необходимо протестировать стойкость преобразователя. Для этого необходимо использовать специальные методы ускоренного тестирования стойкости биометрико-нейросетевой защиты. Возникает целый комплекс вопросов, связанный с ускоренным тестированием средств биометрико-нейросетевой защиты. Кроме того, к этому комплексу примыкают вопросы сертификации и полного (неускоренного) тестирования средств защиты самим производи-

телем или некоторым независимым органом сертификации (испытаний). В свою очередь испытательный центр (лаборатория) должна иметь соответствующие методики испытаний и большие базы случайных биометрических образов, верно отражающих реальную статистику распределения биометрических параметров «Своих» пользователей и наиболее вероятную статистику организации потенциальным злоумышленником атак подбора.

Как было сказано выше, для высоконадежной биометрии обычной является ситуация когда ожидаемая вероятность ошибки второго рода «Пропуск Чужого» составляет величину близкую к 10^{-17} . Это означает, что для прямого вычисления столь малой величины потребуется использовать не менее 10^{18} тестовых примеров. Очевидно, что такой тест выполнить нереально.

Гораздо более реальным является воспроизведение всего 70 случайно выбранных слов, например, с разных страниц из книги. На рукописное воспроизведение уйдет порядка 10...12 минут. При такой тестовой выборке удастся вычислить математическое ожидание откликов с относительной ошибкой 2,7% ($m = 127,4 \pm 3,4$). Дисперсию удастся вычислить с существенно большей относительной ошибкой 21% ($\sigma = 13,6 \pm 2,9$). Это означает, что в наихудшем случае мы можем оценить снизу значение показателя степени стойкости защиты с относительной ошибкой 28%.

Для ошибок вычисления математического ожидания наихудшими является те, которые смещают распределение значений множества «Чужие» в левую сторону (к нулю). Соответственно, наихудшее для системы биометрической защиты значение математического ожидания составит $m = 127,4 - 3,4 = 124$. Наихудшее значение дисперсии будет наибольшим, то есть $\sigma = 13,6 + 2,9 = 16,5$.

Исходное распределение значений (без учета конечности тестовой выборки) в рамках гипотезы нормального закона распределения дает оценку стойкости биометрической защиты 1016,84. После сдвига влево и расширения дисперсии (левое множество примеров рисунка 2.1) его распределение дает существенно меньшую оценку стойкости $-10^{13,93}$.

То есть знание закона распределения [37-39] контролируемых данных на выходе нейронной сети позволяет сократить размеры тестовой выборки на 16 порядков. При этом мы гарантируем от-

носительную ошибку вычисления показателя степени стойкости к атакам подбора - 28%. По сравнению с изначальной оценкой стойкость снизилась на 2 порядка, но при этом мы имеем инженерную гарантию, обусловленную знанием закона распределения значения. Если гарантий нормального закона распределения значений на выходе нейросети нет, то, соответственно, исчезают и гарантии снижения стойкости только на два порядка в сравнении с начальной оценкой.

Касаясь размеров тестовой выборки, то ее десятикратное увеличение до 700 случайных образов приведет к почти десятикратному снижению ошибки оценки показателя степени стойкости к атакам подбора. То есть, при наличии знания о виде закона распределения значения контролируемой величины при измерении показателя стойкости защиты погрешность измерения может быть уменьшена до величины 2,8% и ниже. При этом, каких либо сверх требований к размерам баз тестовых образов не возникает.

Как показала практика, тестирование на нормальность биометрических нейросетевых механизмов требует достаточно высокой квалификации экспериментаторов и значительных размеров баз независимых тестовых биометрических образов. В ряде случаев подобное тестирование непосильно для малых фирм. В связи с этим желательно проводить подобное тестирование специализированными тестовыми лабораториями, обладающими кадрами подтвержденной высокой квалификации и заранее подготовленными большими базами независимых биометрических образов. Чем выше требования к справедливости гипотезы нормальности, тем больше затрат времени и иных ресурсов требует задача проверки этой гипотезы с высоким уровнем достоверности. Чем выше заявляемая производителем достоверность принимаемых его нейросетевым механизмом решений, тем труднее доказать справедливость гипотезы нормальности.

Даже в том случае, если мы точно знаем, что законы распределения «Свой», «Чужой» имеют нормальное распределение значений, мы все равно не гарантированы от ошибок, связанных с конечностью тестовой выборки [40]. Очевидно, что чем больше тестовая выборка, тем меньше модуль вычисления погрешности математического ожидания или при $n \rightarrow \infty$, $|\Delta m_t| \rightarrow 0$. Пользу-

ясь этим можно решать и обратную задачу. Если известно, что тестовые воздействия на нейросеть статистически независимы и число их известно, то, пользуясь распределением Стьюдента [39] мы можем оценить значение погрешности вычисления математического ожидания, предварительно задавшись доверительной вероятностью.

Аналогом таблиц Стьюдента является приведенная ниже на рисунке 2.1 номограмма.

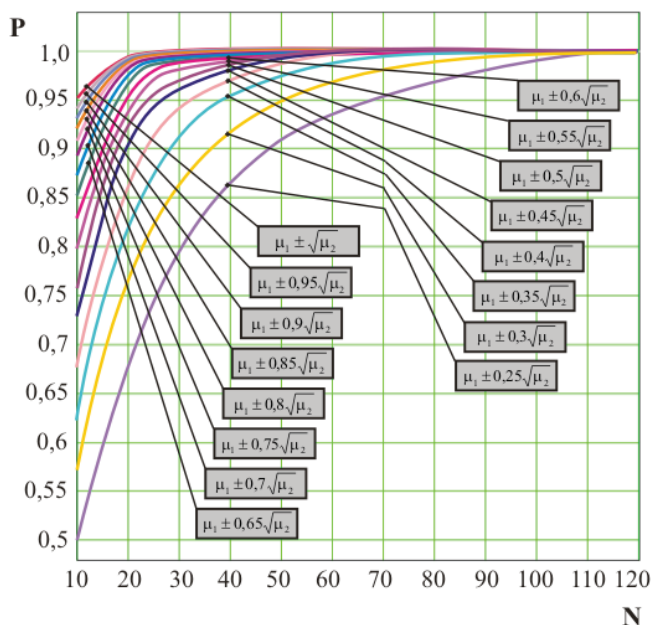


Рис. 2.1. Номограмма связи ошибки вычисления первого момента $\mu_1 = t_i$ с расчетным значением второго статистического момента $\mu_2 = \sigma_i^2$ через заданное значение доверительной вероятности – P

Из номограммы рисунка 2.1 видно, что при уже при 30 примерах ошибка вычисления математического ожидания попадает в интервал $\pm \sigma_i$ с вероятностью близкой к единице. Это обусловлено тем, что ошибка вычисления математического ожидания независимых отсчетов пропорциональна σ_i и падает пропорционально $1/\sqrt{N}$. Для $N = 30$ $\sqrt{30} \approx 5.47$. Задавая интервал $\pm \sigma_i$ при $N =$

30, мы с высокой вероятностью оказываемся в интервале $\pm 5.47\sigma_m$, где σ_m – собственная дисперсия случайной величины являющейся математическим ожиданием - m_t , полученной усреднением 30 независимых измерений.

Номограмма интересна тем, что показывает явную статистическую связь ошибки вычисления первого момента $\mu_1 = m_t$ со значение второго момента $\mu_2 = \sigma_t^2$. По индукции следует ожидать наличие подобной связи между ошибкой вычисления второго момента (дисперсии) и значением третьего статистического момента – μ_3 (асимметрии). На рисунке 2.2 приведена номограмма, связывающая ошибку вычисления среднеквадратического отклонения $\sqrt{\mu_2} = \sigma$ со значением $\sqrt[3]{\mu_3}$.

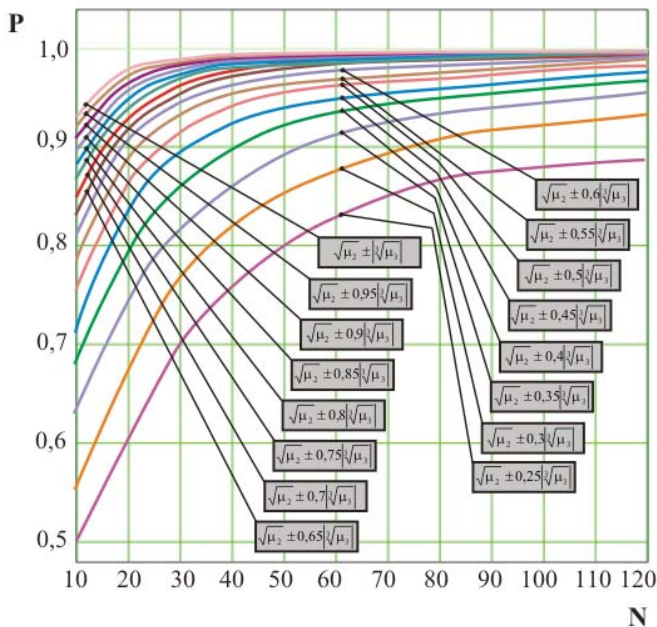


Рис. 2.2. Номограмма, связывающая вероятность попадания ошибки вычисления среднеквадратического отклонения $\sqrt{\mu_2} = \sigma$ в интервал, определяемый статистическим моментом третьего порядка - μ_3

Очевидно, что пользуясь номограммой рисунка 2.2 мы всегда можем оценить ошибку $\Delta\sigma$, задавшись доверительной вероятностью P и зная число независимых тестовых испытаний – N . В свою очередь, зная модуль ошибки вычисления математического ожидания $|\Delta m|$ и модуль ошибки вычисления среднеквадратического отклонения $|\Delta\sigma|$ мы всегда можем найти ошибку оценки вероятности ΔP в рамках гипотезы нормального закона распределения значений данных на линейном выходе последнего нейрона.

Из приведенных выше материалов следует, что при абсолютно достоверном «знании» закона распределения значений, мы всегда можем указать погрешность вычисленных статистических оценок на относительно небольших тестовых выборках. Если мы можем указать погрешность статистических оценок, то мы уже имеем не статистические оценки, а статистические измерения [37].

Проводя аналогию ускоренных статистических испытаний [37] с измерениями можно утверждать, что идентификация закона распределения значений контролируемого параметра - это эквивалент формирования эталонной меры плотности распределения значений контролируемой величины. При обычных измерениях физических величин в измерительном приборе обязательно присутствует эталон измеряемой физической величины (вольта, ампера, секунды,...). При статистических измерениях на неограниченно больших тестовых выборках эталона вероятности в измерительной системе может не быть в явной форме, однако в этом случае высокоточные измерения потребуют огромных затрат времени на формирование достаточно большой тестовой выборки. При этом, как правило, возникают физические, временные, экономические, организационные барьеры, ограничивающие размеры тестовой выборки и тем самым ограничивающие точность статистических измерений.

Попытки обойти эти физические, временные, экономические, организационные барьеры всегда связаны с необходимостью иметь «знание» о контролируемой величине, знание о плотности распределения значений контролируемого параметра. Знание эталона измеряемой величины в форме ее эталона в измерительном приборе и знание закона распределения значений при статисти-

стических измерениях это одинаковые сущности, реализующие экономически оправданные измерения.

Очевидно, что высокоточная идентификация закона распределения значений контролируемого параметров в некоторой физической системе это не что иное, как формирование условий (эталонов) для последующих быстрых измерений. То, насколько точно нам удастся построить эталон, определяет погрешности будущих измерений (обычных и статистических).

В этом плане создание больших и сверхбольших баз независимых биометрических образов и их использование для идентификации закона распределения значений контролируемой величины – это эквивалент формирования статистического эталона. Распространение знаний о законах распределения выходных данных, характерных для биометрико-нейросетевых преобразователей, а также распространение хорошо сбалансированных (представительных) малых баз тестовых образов – это тиражирование эталонов статистических измерений стойкости биометрико-нейросетевых преобразователей.

Так же как создание эталонов для физических измерений – это государственная функция, создание предпосылок для статистических измерений параметров защиты информации также является функцией государства. Государству невыгодно, что бы его граждан обманывали при биометрической аутентификации и для противодействия этому будет создана система государственной сертификации средств безопасной биометрической аутентификации. В свою очередь для сертификации высоконадежной биометрии именно государству придется создавать статистические эталоны и добывать знания о законах распределения значений контролируемых параметров.

Если подходить к задаче идентификации закона распределения значений контролируемой статистической величины, как к элементу формирования высокоточного эталона, то становятся оправданными те огромные затраты, которые могут потребоваться. Более того, затраты на создание высокоточного измерительного эталона не являются разовыми. На протяжении всего жизненного цикла высокоточного измерительного эталона требуются непрерывные усилия на его поддержку в работоспособном состоянии.

Применительно к рассмотренной выше задаче подтверждения гипотезы нормального закона распределения значений, формирование статистического эталона есть не что иное, как формирование достаточно большой базы реальных биометрических образов подтверждающих гипотезу с заданной погрешностью.

На каждый конкретный момент времени должно быть известно, что гипотеза нормальности кем-то (например, межведомственной лабораторией тестирования биометрических устройств и технологий Пензенского государственного университета) подтверждена с заданной относительной погрешностью тестирования $\Delta\%=0.0000001$.

Это означает, что все малые производители средств биометрической аутентификации, выполнив требования стандарта [37], могут тестировать свои устройства, опираясь на гипотезу нормальности закона распределения значений. Очевидно, что подобное тестирование может быть осуществлено только, если его результат будет хуже как минимум на один порядок. Погрешность эталона должна быть на порядок выше погрешности «средства измерения». Однако как только гипотеза нормальности становится обоснованной, проблемы того, как измерять, как подсчитывать результат снимаются сами собой в силу того, что все нюансы этого определены, соответствующими российскими [5; 38] и международными стандартами.

Проведение предварительных исследований показывает, что подобный подход вполне может быть реализован для относительно «слабой» биометрии. Можно указать и подтвердить пределы, в которых гипотеза нормальности работает [37-39].

Высоконадежные биометрические устройства со стойкостью 10-12 и выше проще создать, чем доказательно проверить их стойкость прямым численным экспериментом. Для разрешения этого вопроса необходимо создавать и исследовать реальные базы биометрических образов достаточно больших размеров, для чего необходимы людские, финансовые, ресурсы времени и вычислительных машин; искусственными синтезированными базами не обойтись. Исследования, проведенные в лаборатории тестирования биометрических и нейросетевых технологий КазНТУ и межотраслевой лаборатории тестирования биометрических

устройств и технологий ПГУ, показали, что для больших баз биометрических образов, собранных по специальным методикам [5], закон распределения близок к нормальному [37].

Из теории математической статистики известно, что «идеальный» нормальный закон определяется двумя первыми моментами – μ_1 и μ_2 . Зная их, нетрудно рассчитать необходимый минимальный размер баз биометрических образов при условии, что в генеральной совокупности рассматриваемый признак действительно распределен по нормальному закону [38].

Пример на рисунке 2.3 показывает изменение параметров закона из-за ошибок в расчете μ_1 и μ_2 .

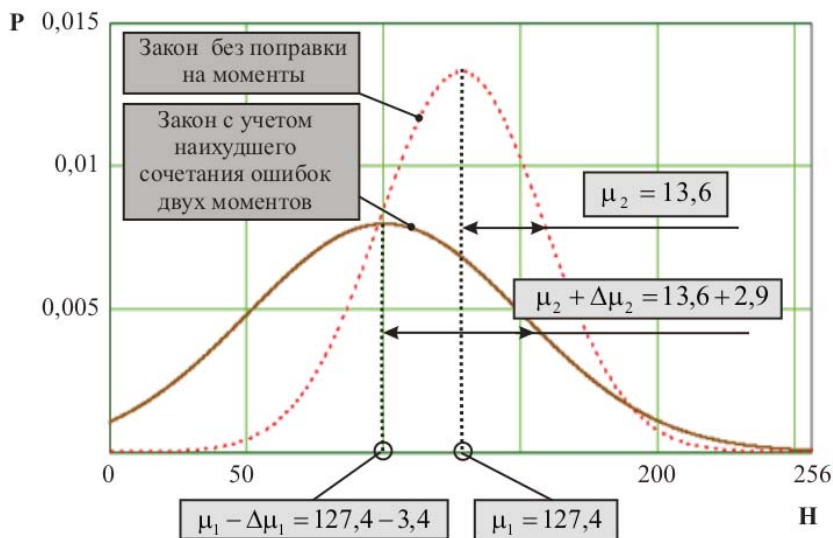


Рис. 2.3. Смещение закона, описывающего распределение выходных состояний преобразователя, полученное добавления ошибок определения двух первых моментов

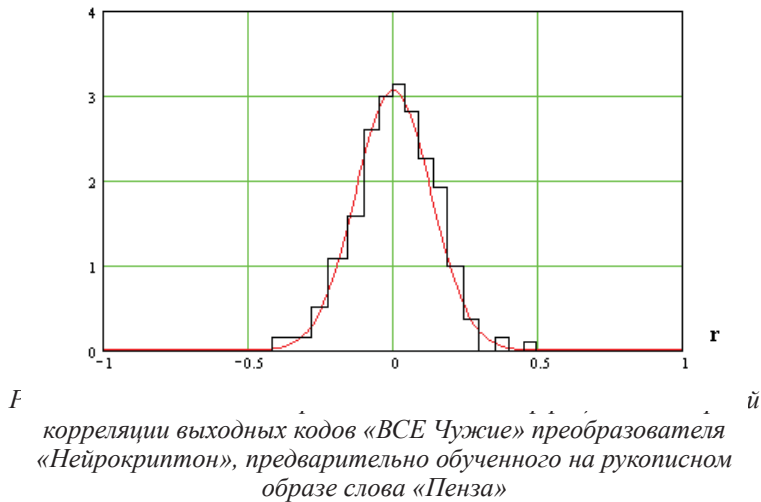
Допустим, что у нас есть биометрических 70 примеров, и мы знаем, что распределение нормальное с моментами $\mu_1=127.4$, $\mu_2=13.6$. Приведенная ошибка вычисления первого момента $\Delta\mu_1=0,027$. Функция распределения должна быть смещена влево (изменение стойкости преобразователя в сторону занижения) –

$\mu_1 = 127,4 \pm 3,4$. Приведенная ошибка определения второго момента с вероятностью, близкой к единице, $\Delta \mu_2 = 0,21$, он находится в пределах $\mu_2 = 13,6 \pm 2,9$. С учетом наименее благоприятной ситуации кривая нормального закона смещается влево (рисунок 2.4). Из-за этого исходная расчетная стойкость биометрического механизма – 1016,84 к атакам подбора (правая кривая распределения значений рисунка 2.3) должна быть снижена до величины 1013,93, приведенная относительная ошибка в определении показателя стойкости не превышает 0,28. Даже малое количество примеров атаки (70 штук) при точном знании закона распределения дают вполне приемлемые результаты для практики. Это позволяет надеяться на то, что для получения соответствующих гарантий по стойкости биометрических средств к атакам подбора гораздо важнее знание закона распределения значения контролируемых параметров, чем объем тестовой выборки.

Намного сложнее ситуация складывается для высоконадежных средств биометрической аутентификации. Как показали расчеты, для них гипотеза нормальности не работает. В этом случае необходимо осуществлять вид закона распределения и статистически доказывать его адекватность действительности [39; 40].

По теории у идеальных преобразователей биометрия/код корреляция (парная и многомерная) выходных кодов «Все Чужие» должна быть нулевой. Однако этого никогда не удастся достичь для реальных преобразователей биометрия/код. Реальные коэффициенты корреляции имеют вполне ощутимые значения. В качестве примера на рисунке 2.4 дана типичная гистограмма значений коэффициентов парной корреляции выходных кодов «ВСЕ Чужие» преобразователя «Нейрокриптон», предварительно обученного на рукописном образе слова «Пенза». На практике имеется существенное отличие от нуля значений коэффициентов корреляции (см. рисунок 2.4). Это означает, что многомерные нейросетевые преобразователи биометрия/код относятся к новому классу статистических объектов, которые должны описываться некоторой модификацией классического биномиального закона распределения значений меры Хемминга. Для терминологического выделения этой модификации далее будем ее называть биномиальным зависимым законом распределения как альтерна-

тиву классического биномиального независимого законно распределения (независимого в точке параметр $p = 0.5$).



В связи с тем, что нам необходимы высокоуровневые гарантии стойкости биометрико-нейросетевых преобразователей к атакам подбора, мы должны иметь высокоточное (численное или аналитическое) описание нового биномиального зависимого закона распределения значений. То есть необходимо проделать значительную работу по аналитическому или численному статистическому описанию нового неклассического объекта - биометрико-нейросетевых преобразователей [40; 41].

Кроме решения задачи идентификации нового неклассического закона распределения необходимо будет доказать его эквивалентность реальному закону распределения выходных кодов преобразователей биометрия/код. При этом возникает еще одна трудность. Дело в том, что классическое распределение χ^2 выведено для независимых опытов (так же как и классический биномиальный закон). Задачей второго уровня является корректировка таблиц классического дискретного закона распределения χ^2 для независимых отсчетов под реальную ситуацию проверки существенно зависимых данных.

2.2 Особенности тестирования средств биометрико-нейросетевой аутентификации личности на малых тестовых базах

В соответствии с ГОСТ Р 52633.0 – 2006 [2], тестирование средств биометрико-нейросетевой аутентификации должно проводиться с использованием баз биометрических образов «Свой» и «Чужой», размеры которых являются достаточными для подтверждения характеристик тестируемых средств аутентификации. При этом создание баз биометрических образов «Свой» не вызывает особых трудностей и требует минимальных временных затрат, однако в соответствии с требованиями [2; 3], размеры используемых баз биометрических образов «Чужой» должны быть на порядок больше заявленной производителем стойкости средства биометрической аутентификации. Поэтому наиболее сложной частью задачи явилось тестирование стойкости обученных преобразователей биометрия-код к атакам подбора. В общем случае задача тестирования стойкости преобразователя биометрия код требует формирования очень больших баз тестовых биометрических образов и применения вычислительных средств с большим объемом памяти, а также с высокой производительностью.

Как показали проведенные исследования в рамках выполнения комплексного проекта, создать достаточно большой объем естественных биометрических образов по требованиям ГОСТ 52633.1 – 2009 [3] практически невозможно. Одним из выходов из создавшегося положения может стать многократное увеличение размеров тестовой базы биометрических образов за счет синтетических биометрических образов, созданных по ГОСТ 52633.2 – 2010 [26]. Однако к настоящему времени данные механизмы находятся в стадии разработки и при тестировании были использованы ранее разработанные методики и положения, изложенные в ГОСТ Р 52633.4 [25].

Тестирование средства высоконадежной биометрической аутентификации проводилось при условии высокого уровня взаимного доверия, которое возникает в случае, когда донор биометрии является владельцем средства биометрической аутентификации со встроенным блоком самотестирования.

Тогда донор биометрии способен после каждого этапа обучения нейросетевого преобразователя биометрия-код самостоятельно протестировать его уровень стойкости к атакам подбора.

Обязательным условием при этом является то, что донор биометрии должен иметь сертификат качества работы встроенного в средство биометрической аутентификации блока самотестирования [25].

2.2.1 Формирование тестовой базы биометрических образов «Чужой»

Для тестирования обученного преобразователя биометрии-код была сформирована тестовая биометрическая база рукописных образов слова-пароля [3], включающая в себя 10000 примеров различных естественных биометрических рукописных образов «Чужой», которые ранее не были использованы при обучении тестируемого нейросетевого преобразователя биометрия-код.

Контроль уровня независимости образов «Чужой» был осуществлен взаимным сравнением их кодов на выходе тестируемого преобразователя по критерию Хэмминга. Критерий Хэмминга вычислялся между каждым из кодов-откликов «Чужой». При этом велся учет того, что в малой тестовой базе не должно быть пар биометрических образов, чьи выходные коды по критерию Хэмминга выпадают из интервала.

$$\left[\frac{n - \sqrt{n}}{2}, \frac{n + \sqrt{n}}{2} \right] \quad (2.1)$$

где n - число разрядов в выходном коде тестируемого преобразователя биометрия-код.

2.2.2 Общая схема тестирования с использованием малой тестовой базы

При выполнении процедуры тестирования преобразователя биометрия-код биометрические образы малой тестовой базы «Чужой» подаются на входы преобразователя однократно в произвольном порядке.

Блок-схема тестирования приведена на рисунке 2.5. В про-

цессе тестирования кодовые отклики преобразователя биометрия-код сравниваются с кодом «Свой», и строится распределение значений показателей критерия меры Хэмминга - $p(h)$.

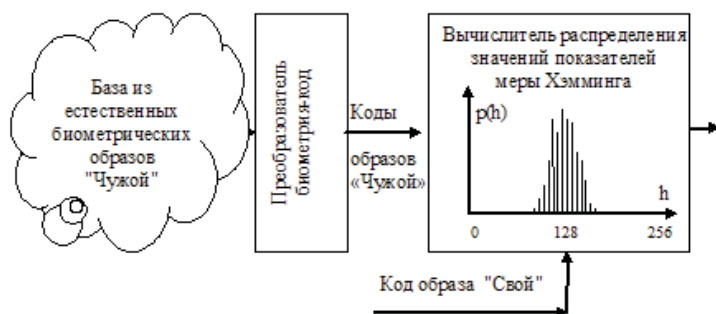


Рис. 2.5. Блок-схема тестирования с использованием базы естественных биометрических образов «Чужой»

По форме и параметрам распределения значений показателя критерия Хэмминга осуществляют прогнозирование появления нулевого значения. Вероятность появления нулевого значения критерия Хэмминга соответствует коллизии - ситуации появления оклодового отклика «Свой» при предъявлении одного из образов «Чужой» (Рисунок 2.6).

Распределение значений показателя Хэмминга является дискретным, однако при больших значениях длины выходного кода, например 256 бит и более, эффектом дискретности можно пренебречь и рассматривать описывающую функцию дискретного распределения как непрерывное распределение с некоторыми статистическими моментами [25; 29; 42; 43].

По полученной выборке кодов-откликов вычисляют математическое ожидание значений показателя критерия Хэмминга - $E(h)$ и среднеквадратическое отклонение показателей критерия Хэмминга - $\sigma(h)$.

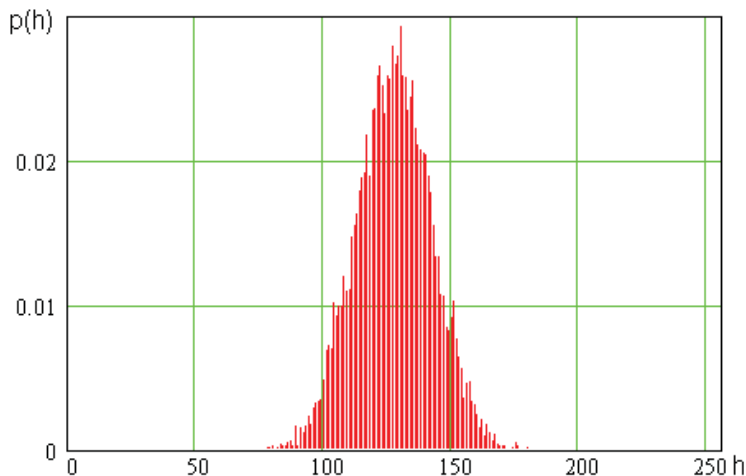


Рис. 2.6. Распределение значений показателя критерия Хэмминга на выходе преобразователя биометрия-код при подаче образов «Чужой»

2.2.3 Оценка стойкости средства биометрико-нейросетевой аутентификации к атакам подбора при использовании малых тестовых баз

При тестировании приближенная оценка вероятности ошибки второго рода – P_2 (пропуск «Чужого») тестируемого преобразователя биометрия-код оценивалась в рамках гипотезы нормальности закона распределения значений показателей критерия Хэмминга [25; 29]:

$$P_2 \approx \frac{1}{2} - \frac{1}{\sqrt{2\pi}} \cdot \int_0^{\frac{E(h)}{\sigma(h)}} \exp(-t^2/2) dt \quad (2.2)$$

Стойкость к атакам подбора оценивалась как обратная величина – $1/P_2$.

Как показали результаты экспериментов в ходе выполнения комплексного проекта по тестированию с различными размерами баз естественных биометрических образов «Чужой» реальное распределение значений показателей критерия Хэмминга отличается от нормального закона распределения значений.

В связи с этим прогноз, вычисленный по формуле (2.2), является приближенным и должен быть скорректирован соответствующей таблицей поправок, учитывающих показатели средней стабильности, уникальности, качества биометрических параметров тестируемого биометрического образа, а также среднее значение модуля парных корреляций [18; 19; 25; 29; 42-44].

2.3 Биометрико-нейросетевой классификатор папиллярных рисунков отпечатков пальцев личности

Уникальный биометрический идентификатор, каковым является отпечаток пальца, служит ключом, который невозможно потерять. Идентификация папиллярных узоров была исторически одной из первых и исключительно важных для практики задач в области распознавания образов и обработки изображений. Английский антрополог Фрэнсис Гальтон в 1892 году определил, что в основе метода идентификации личности по отпечаткам пальцев лежит факт их уникальности [45]. Он же создал первую систему классификации признаков. Изучая большое количество узоров, Гальтон определил, что есть четыре основных типа рисунка отпечатка пальца, от которых происходят все остальные. Он выделял интегральные особенности – треугольные образования из папиллярных линий. Находил положения и выявил зависимости их появления. Это явилось базой для создания системы дактилоскопической регистрации. Используя систему классификации Гальтона, Эдвард Генри создал дактилоскопическую формулу, за которой в криминалистике утвердилось название Гальтона-Генри [45]. Согласно формуле, количество различных комбинаций типов рисунка составило 1024. Что равняется числу разделов картотек. Такая система широко использовалась для поиска преступников, без вести пропавших. Основным недостатком такой системы является необходимость отпечатков всех десяти пальцев человека. Отсутствие хотя бы одного, требовало расчета вероятностных формул. Поэтому в 1957 г. был найден новый подход идентификации. В отличие от формульных систем он был назван кодовым [46]. Сейчас его чаще называют классиче-

ским методом, который означает, что существуют особые точки в строении папиллярного узора, уникальные для одного пальца. С развитием вычислительных технологий такой метод обрел высокую скорость сравнения, однако, не может работать с нечеткими папиллярными узорами. Современными автоматизированными дактилоскопическими информационно-поисковыми системами (АДИС) используются классический метод и метод на основе интегральных признаков, что позволяет «обнаруживать» преступников по следам пальцев. Однако у систем АДИС очень узкая специализация. Задача системы заключается не в том, чтобы просто принять решение «Свой-Чужой», а установить по отпечатку пальца или его фрагменту личность владельца, либо предложить несколько возможных вариантов, то есть идентифицировать [47]. На сегодняшний день системы контроля и управления доступом (СКУД) все больше завоевывают рынок биометрических услуг в России. Биометрические СКУД не могут конкурировать с системами контроля доступа на основе магнитных карт, поскольку не позволяют автоматизировать процесс регистрации и аутентификации большого потока людей. Иностранные разработчики заявляют о 100% идентификации по отпечаткам пальцев, умалчивая при этом, что при тестировании они не учитывают «плохие отпечатки». Поэтому требуется разработка наиболее быстрых алгоритмов сравнения отпечатков для того чтобы устранить проблему, связанную с потерей времени в процессе регистрации и аутентификации большого потока людей. В данной статье будет рассмотрен алгоритм, который позволяет классифицировать рисунок отпечатка пальца. Алгоритм основан на нейронных сетях, где используется «нечеткая логика».

2.3.1 Синтез нейросетевого классификатора по типу узора отпечатка пальца

За основу в АДИС берется «дельтическая» классификация. Эта система различает рисунок отпечатка пальца по количеству и положению дельтообразных участков папиллярных линий.

В зависимости от положения дельт рисунок подразделяется на 4 класса: дуговые, петлевые левые, петлевые правые и завитковые (круговые). Таким образом, для определения типа рисунка

система выполняет поиск дельт на отпечатке, что является сложной задачей обработки изображений, особенно если эти изображения плохого качества.

Дельты образуются тремя потоками папиллярных линий и располагаются в нижней части отпечатка по краям внутреннего рисунка. Как правило, центр отпечатка локализован в середине. Левая и правая петля «тянется» от центра и заканчивается к низу. Дуговой рисунок имеет пиковую точку в центре. Завитковый узор сосредоточен также в центре рисунка. Из этого следует, что рисунок отпечатка, находящийся выше его центра, не имеет макроособенностей и не несет в себе уникальной информации. Поэтому для определения класса необходимо анализировать область отпечатка ниже его центра.

Для поиска дельт некоторые системы используют маску, с которой сравнивают значения всего поля. Это требует значительно времени.

Если охарактеризовать поле направлений папиллярных линий у рисунка отпечатков различных классов, то можно выявить особенности присущие каждому классу. В разрабатываемом классификаторе будем использовать непосредственно особенности направления или расположения папиллярного узора. Полученное изображение с «живого» сканера отпечатка пальца, делится на несколько ячеек с фиксированной длиной и шириной. Размер ячеек зависит от требуемой точности. Расположение линий в каждой ячейке описывается параметром α (рисунок 2.7).

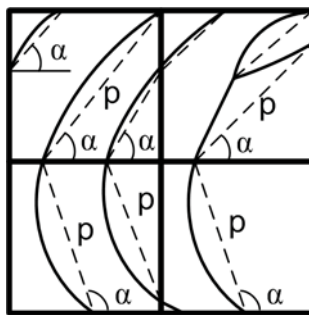


Рис. 2.7. Определение биометрического параметра

Папиллярные линии, располагаемые в ячейках, имеют начало и конец. Отрезок p , соединяющий точки начала и конца, образует угол с горизонтальной линией ячейки, который определяется из условия (2.3):

$$\alpha = \begin{cases} \alpha, & \alpha \in [0; \pi/2] \\ \alpha - \pi & \alpha \in [\pi/2; \pi] \end{cases} \quad (2.3)$$

Угол считается положительным при значениях α от 0 до $\pi/2$ и отрицательным при значениях от $\pi/2$ до π . В одну ячейку обычно попадает несколько папиллярных линий, поэтому результирующий угол находится как среднее значение образованных углов.

Ячейки рассматриваются поочередно и по каждой из них, формируется биометрический параметр. В данной статье приведен пример получения биометрического параметра из рисунка отпечатка, полученного со сканера разрешением 320 x 480. Ячейки размером 40 x 40 пикселей образуют матрицу из 96 значений (рисунок 2.8).

Точность биометрического параметра зависит от длины p , чем она больше, тем точнее получается результат. Максимальное значение p составляет 56 пикселей. Пороговое значение величины p находится исходя из качества изображения, полученного с «живого» сканера. По характеру изменения значений α в выделенных ячейках можно сделать определенный вывод о классе типа узора отпечатка (рисунок 2.9).

X	X	X	X	X	X	X	X
X	X	26	13	-5	X	X	X
X	44	36	16	-5	22	X	X
X	53	43	20	-15	-26	-40	X
X	63	57	12	-39	-50	-60	X
X	70	69	69	-69	-66	-64	X
X	73	75	81	85	-80	-72	X
X	49	65	74	79	84	84	X
X	36	37	51	65	71	72	X
X	22	27	28	41	57	50	X
X	-11	14	22	-21	20	30	X
X	-9	-11	13	9	X	X	X

Рис. 2.8. Матрица биометрических параметров

X	X	X	X	X	X	X	X
X	X	26	13	-5	X	X	X
X	44	36	16	-5	22	X	X
X	53	43	20	-15	-26	-40	X
X	63	57	12	-39	-50	-60	X
X	70	69	69	-69	-66	-64	X
X	73	75	81	85	-80	-72	X
X	49	65	74	79	84	84	X
X	36	37	51	65	71	72	X
X	22	27	28	41	57	50	X
X	-11	14	22	-21	20	30	X
X	-9	-11	13	9	X	X	X

Рис. 2.9. Стабильные ячейки биометрических параметров

Уменьшение α по диагонали в направлении справа налево свидетельствует о принадлежности рисунка к классу «левая петля».

Уменьшение α по диагонали в направлении слева направо свидетельствует о принадлежности рисунка к классу «правая петля».

Уменьшение α по диагонали в направлении справа налево и слева направо свидетельствует о принадлежности рисунка к классу «дуга». При этом параметры отличаются по знаку. Завитковый класс рисунка отпечатка пальца характеризуется приблизительно равными по модулю значениями α ячеек, располагающихся вдоль середины.

2.3.2 Выбор анализируемых данных и обучение нейронов

Таким образом, значение α в используемых ячейках зависит от класса рисунка отпечатка. В нейросетевом классификаторе будем использовать однослойную нейронную сеть с 4 нейронами. На вход каждого следует подать 12 значений α , из ячеек с наибольшей стабильностью параметра [42]. Обучим нейронную сеть на 20 примерах «Свои» и 256 примерах «Чужие». В результате обучения, первый и второй нейрон на своем выходе выдают по 4 бита, третий нейрон – 5 бит, четвертый нейрон – 3 бита. Общая разрядность кода составляет 16 бит. На рисунке 2.10 представлен график распределения «Свои» и «Чужие» первого нейрона.

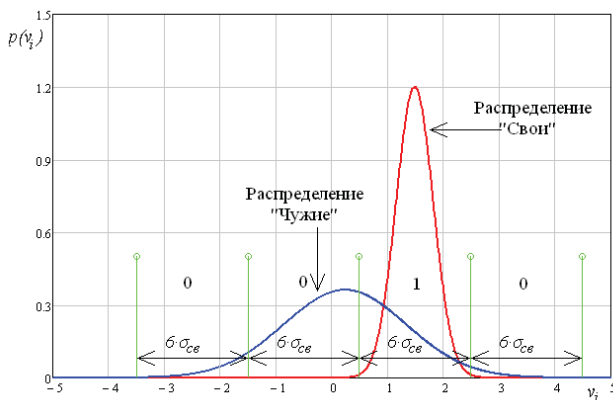


Рис. 2.10. Распределение первого нейрона

Граница разрядов определяется исходя из правила «трех сигм». Практически все значения нормального распределения «Свои» лежат в интервале $[\bar{M}_{CB.} - 3 \cdot \sigma_{CB.}; \bar{M}_{CB.} + 3 \cdot \sigma_{CB.}]$, где $\bar{M}_{CB.}$ – математическое ожидание распределения «Свои»; $\sigma_{CB.}$ – среднее квадратическое отклонение распределения «Свои». То есть, ширина равна $6 \cdot \sigma_{CB.}$. Общий алгоритм определения числа разрядов:

– считается число разрядов каждого нейрона

$$N_{P.H.i}, N_{P.H.i+1} \dots N_{P.H.i+n};$$

– находится максимальное количество разрядов

$$N_{P.H.\max} = \max(N_{P.H.i}, N_{P.H.i+1} \dots N_{P.H.i+n});$$

– определяется общая разрядность кода $N_{ОБЩ.}$ по формуле (2.4):

$$N_{ОБЩ.} = \begin{cases} n \cdot \text{floor}\left(\frac{\sigma_{ЧЖ.}}{\sigma_{CB.}}\right), & N_{P.H.\max} < \text{floor}\left(\frac{\sigma_{ЧЖ.}}{\sigma_{CB.}}\right) \\ n \cdot N_{P.H.\max} & N_{P.H.\max} \geq \text{floor}\left(\frac{\sigma_{ЧЖ.}}{\sigma_{CB.}}\right) \end{cases} \quad (2.4)$$

где n – число нейронов, floor – функция нахождения наибольшего целого, $\sigma_{ЧЖ.}$, $\sigma_{CB.}$ – среднее квадратическое отклонение приуров «Чужие» и «Свои» соответственно.

2.3.3 Результаты обучения

При появлении на входе нейронов значений биометрического параметра α , полученных при распознавании своего образа на выходах формируется кодовая последовательность в соответствии с таблицей 2.1.

Таблица 2.1

Код на выходах нейронов при предъявлении образов «Свои»

№ Нейрона	Код на выходе
1	0010
2	0010
3	00010
4	010

Полученный код составляет часть ключа [41]. Предъявление случайного чужого биометрического образа дает случайный выходной код. Дальнейший поиск микроособенностей отпечатка пальца не имеет смысла. Обучение нейронной сети с примерами другого класса аналогично.

На рисунке 2.11 приведен папиллярный узор среднестатистического «Чужого», полученный нахождением средней величины направления в каждой ячейке.

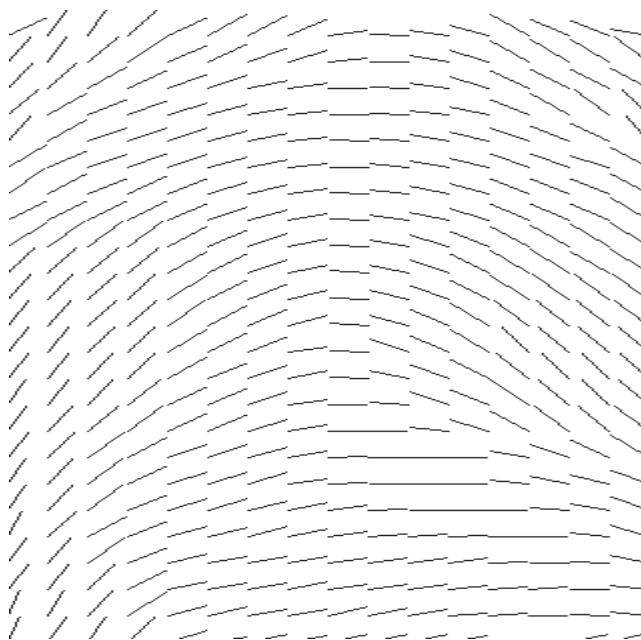


*Рис. 2.11. Среднестатистический отпечаток пальца
распределения «Чужие»*

На рисунке четко определяется центр среднестатистического отпечатка пальца. Направление узора выше центра доказывает отсутствие здесь макроособенностей или наличие уникальной информации.

Таким образом, нейронная сеть выполняет «простую» задачу классификации, для того чтобы не терять времени на «сложной»

задаче классического метода. Предъявление случайного «чужого» биометрического образа дает случайный выходной код, что говорит о принадлежности отпечатка другому классу. При тестировании нейронной сети в 90 % случаев предъявление случайных биометрических образов «Чужие» дает случайный код на выходе для среднестатистического узора. Для уникальных узоров случайный выходной код появляется в большем числе случаев. Крайне важным является то, что выходной код «Свой» классификатора может быть задан технологией [44; 48]. Это позволяет использовать 16-ти разрядный код нейросетевого классификатора как фрагмент кода «Свой».

2.4 Биометрико-нейросетевое обезличивание персональных данных личности при обращении в медицинское учреждение

В настоящее время биометрические технологии представляют основу безопасности в системах аутентификации, где защищенность от несанкционированного доступа к объектам или ресурсам имеет исключительную важность для общества. В связи с развитием электронного документооборота биометрические технологии должны давать наибольшую гарантию людям.

В настоящее время в России и Казахстане взят курс на замену паспорта и всех социальных документов универсальной электронной картой (УЭК), хотя электронный документооборот требует защиты от возможного нецелевого использования и распространения личной информации.

Законы гарантируют всем нам доступ к информации о своем здоровье и одновременно сохранение врачебной тайны. Совместить свободный доступ к информации с сохранением врачебной тайны далеко не всегда возможно. Эти требования противоречивы, возникает классическое противоречие между свойствами доступности и конфиденциальности информации [49; 50].

Несмотря на очевидные успехи информатизации медицины, в частности ее переход на электронный документооборот, проблема обеспечения права граждан на конфиденциальность их личной медицинской информации (на врачебную тайну) только

усугубляется. Защита персональных данных больных позволит людям больше доверять медицинским учреждениям и активнее взаимодействовать с органами здравоохранения.

Защита персональных данных и анонимность пациентов необходима для эффективной борьбы общества с социально значимыми заболеваниями. Зачастую люди просто боятся сдать анализы (с чего и начинается лечение), что препятствует обнаружению болезни на ранних стадиях развития, а в худшем случае – спасению человека. Одним из путей решения этих проблем является обезличивание пациентов (обеспечение анонимности пациентов) [51].

В России разработана серия стандартов ГОСТ Р 52633, которая позволяет осуществить обезличивание пациентов, используя нейросетевой анализ персональных биометрических данных [2]. Новые технологии сводятся к тому, что используется большая сеть искусственных нейронов, которая автоматически обучается преобразовывать биометрический образ человека, чем служит рисунок отпечатка пальца, в личный ключ (рисунок 2.12).

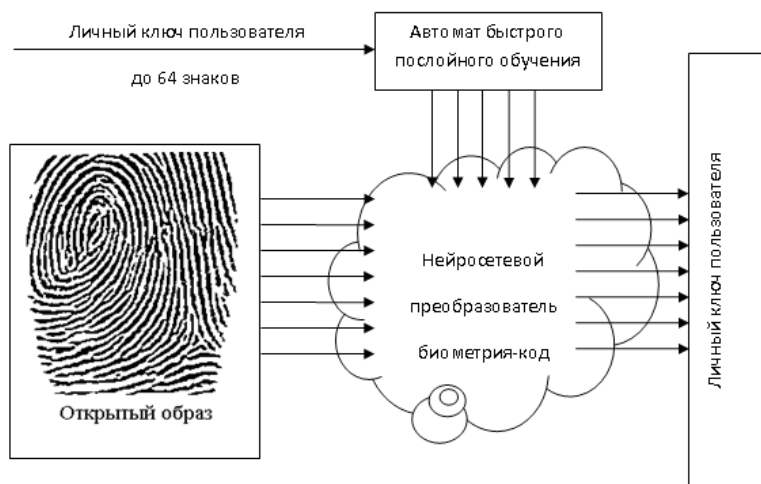


Рис. 2.12. Анонимная биометрическая идентификация больного с сокрытием его биометрического образа в параметрах нейросетевого преобразователя

Допустим, пациенту необходимо зарегистрироваться в медицинском учреждении для сдачи тестовых анализов. Регистрация заключается в создании двух ключей: открытого и личного [52] (рисунок 2.13).

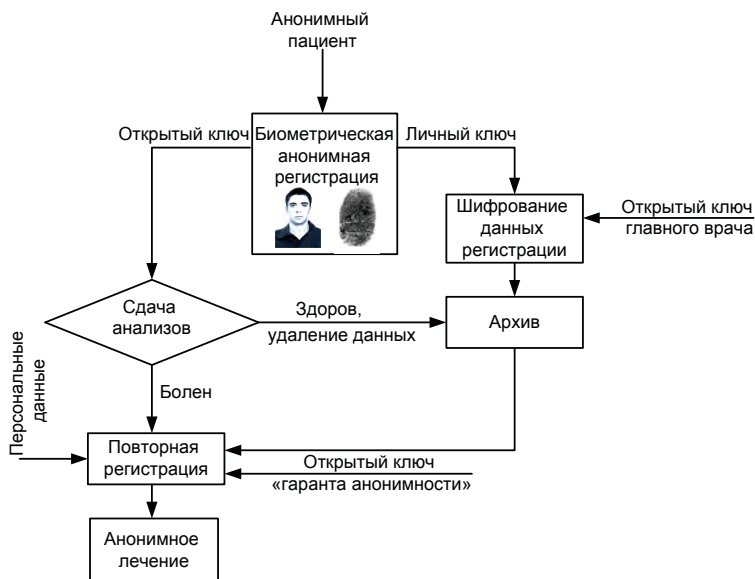


Рис. 2.13. Нейросетевое обезличивание пациентов при первичном анонимном обращении в медицинское учреждение

Процедура регистрации проходит на терминале, оборудованном биометрическим сканером и веб-камерой. Для создания открытого ключа пациенту требуется обучить нейронную сеть, прикладывая, например средний палец правой руки к сканеру. В результате при появлении на входах нейронной сети биометрических параметров отпечатка пальца она будет формировать открытый ключ пациента. Личный ключ пациент создает генератором случайных чисел. Веб-камера фиксирует момент регистрации пользователя. Видеофрагмент регистрации и отпечатки пальца пациента хранятся до момента определения у него отрицательного результата анализа. В случае положительного результата анализа видеофрагмент регистрации и отпечатки пальца сохраняются. Все данные регистрации шифруются на общем ключе,

который формируется из личного ключа пациента и открытого ключа главного врача. Зашифрованные данные располагаются в архиве медучреждения.

При сдаче анализов пациент использует свой открытый ключ, личные данные человека никаким образом не связаны с его биологической пробой. При получении результатов медицинского освидетельствования пациенту следует доказать свою идентичность, то есть получить на выходе нейронной сети свой открытый ключ.

Например, из ста человек, сдавших анализы двое имеют положительный результат, то есть больны. 98 человек с отрицательным результатом могут раскрыть свою анонимность и получить справку о своем здоровье на свое подлинное имя. Все данные регистрации здоровых пациентов удаляются. Поскольку двум больным требуется лечение, их анонимность должна быть раскрыта. Для того чтобы больной мог продолжить анонимное лечение, ему следует выбрать гаранта своей анонимности (адвоката, нотариуса, центр сохранения анонимности). Больной имея свой личный ключ и открытый ключ «гаранта анонимности» может сформировать общий секретный ключ (рисунок 2.14).



Рис. 2.14. Опора пациента на открытый ключ «гаранта анонимности» для сокрытия своих персональных данных, хранящихся в зашифрованной форме неопределенно долго медицинским учреждением

Паспортные данные больного, полученные после повторной регистрации (фамилия, имя, отчество, дата рождения, место рождения, адрес проживания, серия и номер паспорта, номер страхового полиса), шифруют на общем секретном ключе «гаранта анонимности» и больного. Шифротекст хранится только в медицинском учреждении. В случаях, предусмотренных законом, при снятии анонимности больного, шифротекст передается «гаранту анонимности», который, имея открытый ключ больного и свой личный ключ, способен выработать общий ключ и расшифровать персональные данные.

Таким образом, полная анонимность пациента будет гарантирована при получении отрицательного результата анализа, поскольку никакие персональные данные человека не сохраняются в процессе его регистрации. Описанная процедура биометрической анонимной регистрации, сохраняет высокий уровень конфиденциальности персональных данных человека и обеспечивает его полную анонимность [53]. Однако в силу действующего законодательства медучреждение имеет возможность деавуировать анонимность больного.

2.5 Использование нейросетевой биометрической аутентификации личности в сети интернет

В настоящее время Интернет вырос до огромной публичной общемировой сети сервисов и информационных хранилищ. Возможности прямой передачи сообщений, чтения почты и новостей, пересылки файлов с данными, совершать покупки в электронных магазинах, удаленные консультации, управлять ценными бумагами привлекают в сеть миллионы людей. Однако, несмотря на все те преимущества, которые дают глобальные сети ее пользователям, очевидна, прямая угроза информационной безопасности пользователей. Передаваемая пользователем в сеть в открытом виде, персональная информация (например, пароли, имена, личные данные, паспортные сведения, номера кредитных карточек) может быть выделена злоумышленником из потока данных. Ничто не помешает злоумышленнику при желании выдавать себя за другого пользователя и проводить операции от его имени.

Таким образом, очевидная необходимость защиты конфиденциальных данных заставляет обратить внимание на следующие аспекты процесса клиент-серверного взаимодействия: передача данных в недоверенной среде (например, в глобальной сети, Интернет); надежная аутентификация клиента сервером; безопасное хранение конфиденциальной информации в недоверенной среде сервера.

В настоящее время наиболее популярным методом защиты передаваемой по сети Интернет конфиденциальной информации, является применение защищенных сеансов между клиентом и сервером, организованных при помощи протоколов SSL/TLS. Стандарт протокола SSL версии 3.0 описан в [54]. Стандарт IETF протокола TLS описан в RFC 2246 [55] в январе 1999 года; в RFC 4346 [56] определена его версия 1.1. Применяемая в SSL/TLS криптографическая защита пересылаемых пакетов данных и механизмы взаимной аутентификации клиента и сервера позволяют предотвратить пассивный перехват, обнаружить искажение и фальсификацию конфиденциальной информации. В существующих реализациях определено использование следующих криптографических систем на открытом ключе - RSA, Diffie-Hellman, DSA, Fortezza; алгоритмов симметричной криптографии - RC2, RC4, IDEA, DES, AES, Camellia; однонаправленных функции хеширования - MD2, MD4, MD5, SHA. Кроме того, возможно расширение реализаций алгоритмами, соответствующих российским криптографическим стандартам ГОСТ 28147-89, ГОСТ Р 34.10-94 и ГОСТ Р 34.11-94, что делает возможным выполнение модификации протокола с учетом национальных требований к безопасности передаваемой информации.

Согласно спецификации протоколов TLS/SSL, аутентификация клиента является необязательной, и может быть проведена сервером с использованием любой информации, хранимой пользователем в тайне: имени пользователя и пароля; почтового адреса и пароля; информации о номерах счетов кредитных карточек; ответов на контрольные вопросы; одноразовых паролей. В настоящей статье рассматривается процедура персональной биометрической аутентификации пользователя (в дальнейшем «Клиента»), ее варианты и области применения. Кроме того, затрагивается

вопрос безопасного хранения аутентификационных данных на стороне сервера.

Требования к высоконадежным средствам биометрической аутентификации определены в ГОСТ Р 52633-2006 [2]. Согласно стандарту, нейропреобразователи биометрия-код могут обеспечить надежную аутентификацию человека по его биометрическим характеристикам, таким как голос, рукописный подчерк, отпечатки пальцев и т.д. Высокая надежность средств биометрической аутентификации позволяет использовать их для дистанционной аутентификации. Дешевым и перспективным для массового применения является средство биометрической аутентификации человека по рукописному образу. Для их использования пользователю персонального компьютера достаточно приобрести планшет. В устройствах со встроенной возможностью рукописного ввода использование механизма биометрической аутентификации «бесплатно». В КПК, смартфонах и ноутбуках с сенсорными экранами ввод рукописных образов представляется совершенно естественным и удобным. В случае биометрической аутентификации вводимый пользователем личный тайный рукописный образ преобразуется в некоторый тайный длинный код аутентификации. Сравнение значения хеш-функции длинного кода аутентификации с ее эталонным значением позволяет выполнить однозначную идентификацию «Клиента».

По месту проведения преобразования биометрия-код может быть рассмотрено 2 варианта протокола дистанционной биометрической аутентификации «Клиента». В первом случае обучаемая сеть и процедуры извлечения из нее длинного кода аутентификации хранятся и исполняются на сервере, а «Клиент» передает для аутентификации свой личный биометрический образ. Во втором случае алгоритмы и данные для преобразователя хранятся на сервере, но высылаются «Клиенту» в начале процедуры аутентификации [57]. Преобразование биометрия-код, т.е. выделение длинного кода аутентификации, проводится на стороне «Клиента». После этого вычисляется хеш-функция от кода и ее значение отправляется серверу для завершения процедуры аутентификации.

В связи с тем, что при применении первой схемы организации процедуры биометрической аутентификации может быть обеспе-

чена более простая защита от атак подбора биометрических образов и разработаны надежные механизмы биометрического аудита на стороне сервера, он рассматривается как базовый.

Рассмотрим схему дистанционной биометрической аутентификации пользователя в защищенном HTTPS соединении, представленную на рисунке 2.15, с преобразованием биометрия-код на стороне сервера.

На первом этапе работы, в соответствии с процедурами протоколов SSL/TLS, устанавливается соединение между машиной «Клиента» и сервера, согласуются криптографические алгоритмы на открытом ключе, на симметричных ключах, алгоритмы хеширования. Клиент и сервер обмениваются «приветствиями» и ключами. Сервер отправляет клиенту сертификат для подтверждения своей подлинности; клиентом производится аутентификация сервера. После все передаваемые между машиной «Клиента» и сервера пакеты шифруются на симметричном ключе, т.е. дальнейший обмен данными проводится по защищенному каналу.

На втором этапе проводится биометрическая аутентификация Клиента. На шаге 1, показанном на рисунке 2.15, Клиент вводит личный тайный биометрический образ и имя (или, например, ФИО). Биометрический образ и имя передается по защищенному каналу серверу, который выполняет процедуру аутентификации. На шаге 2 личный биометрический образ «Клиента», с указанным именем, при помощи нейросетевого преобразователя биометрия-код конвертируется в длинный код аутентификации «Клиента». На 3-м шаге выполняется сличение полученного хеш-функции длинного кода с ее эталоном, хранимый сервером в тайне. В случае совпадения кодов, аутентификация считается выполненной успешно, и сервер отправляет «Клиенту» подтверждение успешной аутентификации. Обмен данными продолжается в авторизованном режиме. В случае несовпадения хеш-функции длинного кода аутентификации и ее эталона, аутентификация считается не пройденной и, в зависимости от решения, принимаемого биометрическим аудитором могут рассмотрены следующие варианты: «Клиенту» предлагается ввести личный образ еще раз; в случае обнаружения атаки, происходит разрыв соединения.

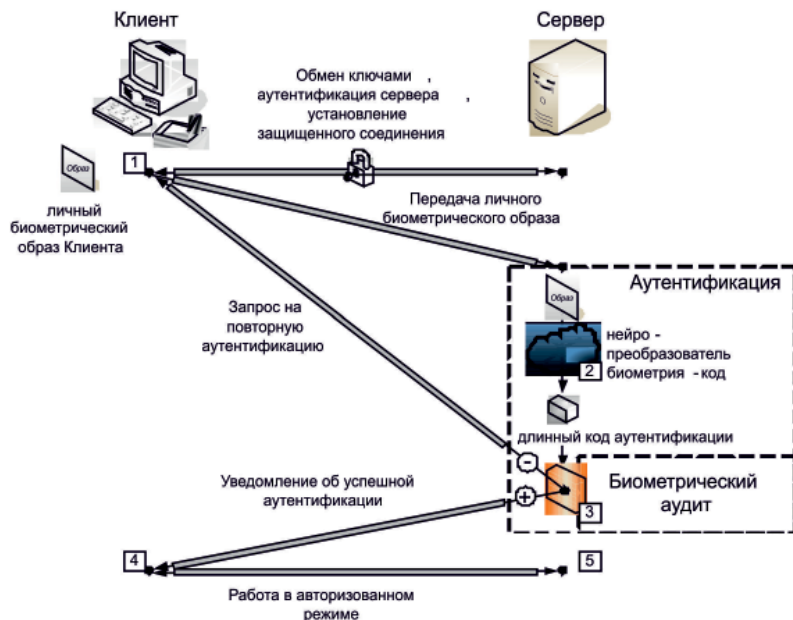


Рис. 2.15. Биометрическая аутентификация пользователя

Биометрический аудит является неотъемлемой частью механизма аутентификации «Клиента» и предполагает обнаружение и предотвращение атак сервера через биометрию. Разные типы атак биометрической информации описаны в [34]. Подобные атаки могут быть легко обнаружены и отражены, если сервер оснащен соответствующими средствами защиты и обнаружения атак.

В стандарте ГОСТ Р ИСО/МЕК 19784-1 [58] дается спецификация биометрического программного интерфейса (BioAPI), определяющего общую высокоуровневую модель биометрического распознавания, пригодную для любой биометрической технологии. Вопросы связанные с надежным хранением биометрических шаблонов, используемых для верификации и идентификации пользователя в нем не определены. Однако, в связи с существующими реальными внешними и внутренними угрозами компрометации личных шаблонов пользователя, важным становится вопрос их надежного хранения на сервере. Хищение баз

данных биометрических шаблонов, например, с особыми точками отпечатков пальцев пользователей приводит к нежелательности их использования в других системах из-за их компрометации! Очевидно, что системы, использующие биометрическую аутентификацию пользователя должны поддерживать и механизмы надежного хранения биометрических шаблонов пользователя даже в доверенной среде сервера.

В этой связи важно отметить, что в предлагаемом методе биометрической аутентификации «Клиента» по рукописному образу личная биометрия пользователя хранится на сервере в нейросетевом контейнере и не может быть извлечена из него злоумышленником. Кроме того, злоумышленник не может установить связь нейросетевого контейнера с конкретным пользователем (обеспечена анонимность биометрической записи). При смене кода аутентификации полностью меняется содержание нейросетевого контейнера даже для одного и того же биометрического образа [59]. Сильная биометрическая защита в соответствии с идеологией ГОСТ Р 52633-2006 может быть обеспечена только при сохранении в тайне самого биометрического образа и длинного кода аутентификации. Обеспечивать сохранение тайны параметров биометрической аутентификации необходимо не только самому пользователю, но и системе в целом. В этом плане спецификация BioAPI (ГОСТ Р ИСО/МЕК 19784-1) является ущербной. Использование ГОСТ Р ИСО/МЕК 19784-1 без учета требований ГОСТ Р 52633-2006 допустимо только в слабозащищенных системах полицейского контроля, не предъявляющих особых требований к обеспечению анонимности проверяемого и сохранению в тайне его биометрии.

Такие факторы как наличие защищенного канала между клиентом и сервером, простота изменения личного биометрического образа новым позволяют «Клиенту» изменять свой биометрический образ аутентификации дистанционно. Последовательность операций процесса удаленной замены биометрического образа представлена на рисунке 2.16.

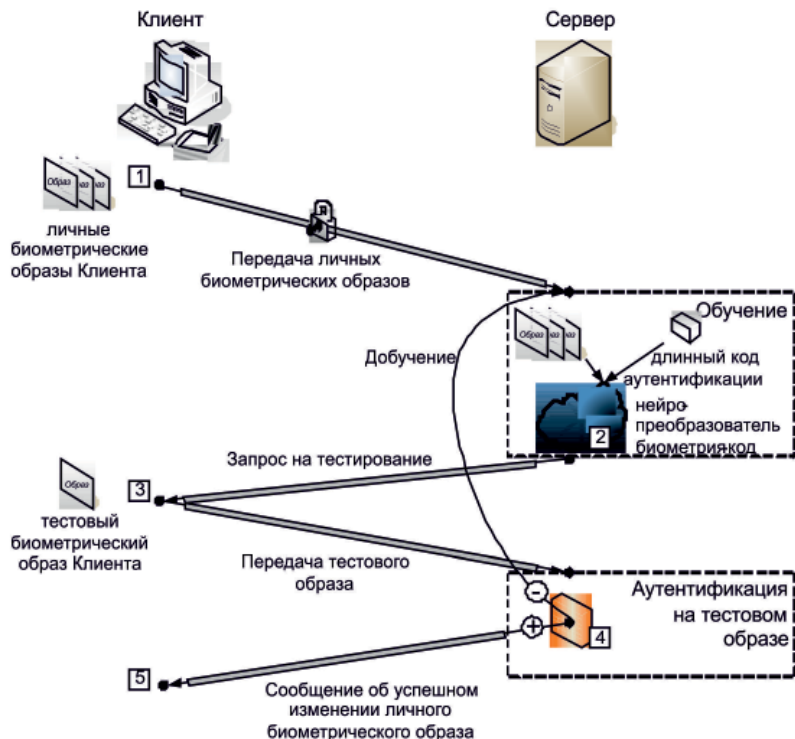


Рис. 2.16. Обучение на личных биометрических образах

Замена личного биометрического образа «Клиента» производится только при наличии установленного защищенного соединения после обязательного завершения процедуры биометрической аутентификации (или без нее при первичной регистрации «Клиента»). На первом этапе устанавливается защищенное SSL/TLS соединение между клиентской машиной и сервером, «Клиентом» производится аутентификация сервера по высылаемому им сертификату. На втором этапе, как было описано выше, выполняется биометрическая аутентификация «Клиента», в ходе которой определяется его длинный личный секретный код аутентификации.

Третий этап состоит из двух шагов. На первом шаге «Клиент» вводит N новых личных биометрических образов. Количество

личных образов выбирается, исходя из общей статистики обучения людей и личной статистики обучения «Клиента», если такая имеется. Введенные образы отправляются серверу. На стороне сервера на шаге 2 происходит обучение нейропреобразователя биометрия-код проводить операцию извлечения длинного кода аутентификации при предъявлении ему биометрического образа «Клиента».

Четвертый этап заключается в контрольной проверке нейропреобразователя и, при необходимости, в его дообучении на личных образах «Клиента». Для этого «Клиенту» отправляется запрос на ввод тестового биометрического образа. «Клиент» вводит биометрический образ, и он передается серверу. Сервер на шаге 4 выполняет тестовую аутентификацию «Клиента». Если тестовая аутентификация оказывается успешной, обучение считается завершенным и обученный преобразователь биометрия-код начинает использоваться для аутентификации клиента. Если контрольная аутентификация неуспешна, то нейропреобразователь биометрия-код дообучается на тестовом образе, и процедуры четвертого этапа повторяются. «Клиент» в любой момент может прервать обучение; в этом случае новый личный биометрический образ отвергается. Таким образом, на всем протяжении процедуры смены личного биометрического образа гарантируется подлинность использующего ее пользователя и действительность проводимых им изменений.

Предложение Netscape защищенных протоколов SSL 3.0 в 1996 году [54] было серьезной подвижкой в плане обеспечения безопасности Интернет бизнес-процессов, электронной торговли, Интернет банкинга. Однако на тот момент отсутствовали технологические возможности высоконадежной дистанционной биометрической аутентификации пользователя. Появление национального общероссийского стандарта [2] резко меняет ситуацию, наконец-то появились технологические возможности надежно идентифицировать не только машину, но и человека, который обращается с этой машины к доверенному серверу. Все это позволяет надеяться на появление в ближайшее время модификаций существующих RFC стандартов, учитывающих новые возможности дистанционной биометрической аутентификации.

2.6 Классический вариант защиты персональных данных хешированием паролей и биометрических кодов доступа

Классическая парольная аутентификация личности строится на том, что сам пароль в операционной системе не хранится, хранится криптографическая хэш-функция пароля в ячейках памяти адрес которых связан с открытым именем пользователя (логин). К сожалению, парольная защита для части приложений оказывается недостаточна, люди не хотят запоминать длинные парольные фразы из случайных знаков двух регистров. Для запоминания удобны короткие осмысленные парольные фразы, которые не являются стойкими к атакам подбора. Созданы словари удобных людям коротких парольных фраз с помощью которых системный администратор обнаруживает слабые пароли.

Биометрия удобна тем, что очень большие объемы информации человек может получать на простых парольных фразах. Стойкость биометрической защиты не зависит от того есть ли смысл в парольной фразе или его нет. Когда парольное слово воспроизводится рукою пользователя на графическом планшете, то начинает играть роль уникальность почерка человека [10; 37]. Каждый из нас пишет буквы по-своему, при этом то, как мы пишем сильно зависит от того какая буква была предыдущей и какая буква будет следующей. Играть роль двойки и тройки букв. Если биометрический пароль всего из 5 рукописных букв скомпрометирован, то его стойкость к атакам подбора мала от 102 до 103. Не скомпрометированный короткий рукописный пароль из 5 букв обеспечивает стойкость к атакам подбора от 105 до 1012, все зависит от стабильности почерка и его уникальности. Сохранение в тайне биометрического пароля по своему влиянию много сильнее биометрии как таковой.

Естественно, что биометрический код доступа в операционной системе нельзя хранить открыто. Так же как мы храним хэш-функцию пароля, мы должны хранить хэш-функцию биометрического кода. Если предполагается совместная парольно-биометрическая аутентификация личности, то безопасным является совместное хранение хэш-функции конкатенации пароля и биометрического кода.

Хранение хэш-функций паролей и био-кодов доступа считается безопасным при использовании специально сконструированных криптографических хэш-функций. Именно криптографические хэш-функции можно рассматривать как необратимые. То есть их прямое вычисление занимает мало времени, а обратная задача поиска по значению хэш-функции исходных данных является практически невыполнимой задачей (если длина хэшируемых данных больше или равна длине хэш-функции).

Практика применения хэширования паролей в операционных системах семейства Windows строится на применении хэш-функции MD-5 с длиной выходного кода 128 бит. Условно можно говорить о сложности обратной задачи на уровне 2128 попыток подбора, подобной защиты вполне достаточно для паролей и био-кодов со стойкостью к атакам подбора 240. Подобная практика сегодня является общепринятой, то есть используемая для защиты криптографическая хэш-функция должна обеспечивать показатель стойкости примерно в три раза больше, чем пароли или био-коды по отдельности и ли в сумме (при двухфакторной аутентификации). Если в место криптографического хэширования используется шифрование, то соотношение между длиной ключа шифрования и стойкостью биометрии должно оставаться примерно таким же.

2.7 Высокостойкая защита данных нейросетевого контейнера простым шифрованием

Одним из наиболее простых и понятных способов защиты является получение с выходов первых (открытых) нейронов ключа и шифрование на нем данных других нейронов (защищенных) [36; 60; 61]. Эта ситуация отображена на рисунке 2.17.

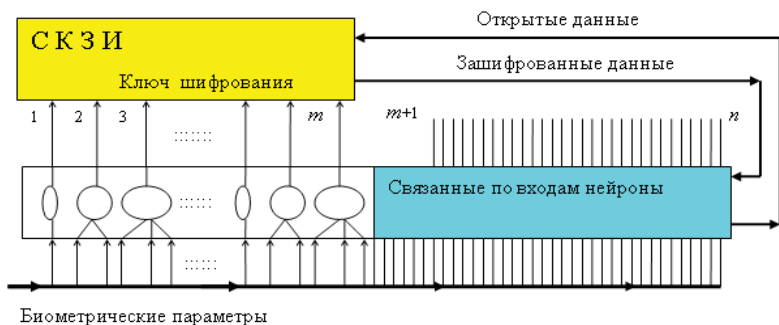


Рис. 2.17. Защита данных нейросетевого контейнера шифрованием избыточных перекрестных связей

Из рисунка 2.17 видно, что данные первых m нейронов хранятся открыто, а данные последующих нейронов зашифрованы на ключа длиной m во время подготовки защищенного контейнера. При использовании защищенного контейнера пользователь предъявляет свою биометрию, которая преобразуется в био-код длиной m и далее на этом коде производится расшифровывание ранее защищенных шифрованием данных последующих нейронов.

Возникает естественный вопрос о том, какова должна быть длина кода ключа шифрования – m . Очевидно, что длина ключа шифрования должна выбираться исходя из ряда криптографических и биометрических ограничений. Самым простым вариантом является выбор длины кода пропорционально трех кратному показателю стойкости к атакам подбора (при стойкости биометрии 2^{40} стойкость шифрования должна составить 2^{120}). То есть при длине ключа шифрования 120 бит потребуется хранить открыто данные 120 первых нейронов (таблицы весов и связей для 120 нейронов общедоступны).

При средней информативности одного био-параметра 0.2 бита нем потребуется иметь в среднем по 5 входов у каждого нейрона получается 600 входов для 600 био-параметров. Если число входов у сети нейронов только 480, то оставшиеся 220 входов будут повторяться у открытой части нейронов. Каждый нейрон будет связан со своими соседями примерно 2 входами. Два входа перекрывающихся с соседями у каждого нейрона вполне достаточ-

но, для того, чтобы была успешно организована атака Маршалко. Так же успешно на длинные серии открытых данных обученных нейронов могут быть применены атаки на стабильность разрядов, атаки на извлечения кода ключа «Свой» через наблюдение энтропии. Все эти атаки приводят к сокращению длины кода от 80% до 90%, то есть в место 120 бит иллюзорных (мнимых) мы получим шифрование на 12 или 24 реальных битах. Стойкость такого шифрования оказывается ниже объявленной стойкости биометрической защиты во много раз (2^{12} - 2^{24} много меньше 2^{40}).

2.8 Защита нейронов блочным шифрованием с растущей длиной ключа

Получается, что понятный всем вариант шифрования на одном длинном ключе реализовать не удастся из-за эффективности существующих атак на биометрию. Все эти атаки строятся на наблюдении большого числа выходов у незащищенных нейронов для эффективной защиты от всех интегральных атак на большое число открытых нейронов. Для того, чтобы снизить эффективность интегральных атак на биометрию следует отказаться от создания одного длинного ключа и использовать множество ключей увеличивающейся длины. Пример подобного блочного шифрования приведен на рисунке 2.18.

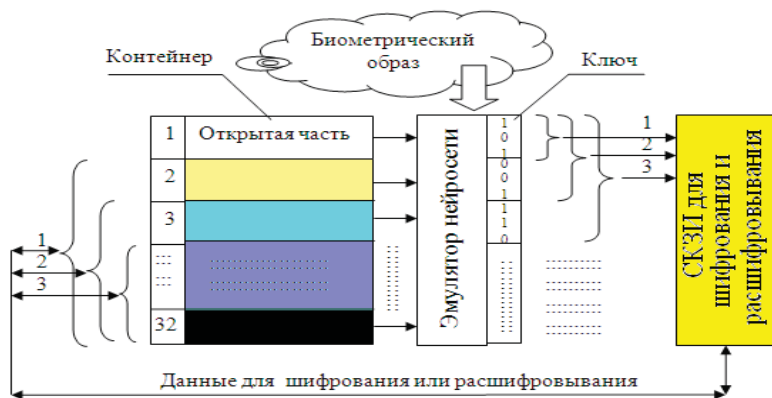


Рис. 2.18. Защита данных блочным шифрованием с растущей длиной ключа 3, 6, 12, 15..., 256, бит

Все малые ключи объединяются друг с другом и многократно шифруют данные всех последующих обученных нейронов в процессе подготовки защищенного нейросетевого контейнера. При использовании защищенного нейросетевого контейнера идет обратный процесс, ключ полученный на открытых нейронах расшифровывает данные всех последующих нейронов. Становится открытой еще одна часть ключа, которая восстанавливается, объединяется с первой и расшифровывает следующие данные. На рисунке 2.18 это отображено постепенным усилением заливки (защиты данных). Открытая часть нейронов заливки не имеет.

Получается, что одним длинным ключом мы не можем пользоваться из-за наличия нескольких интегральных атак на открытую биометрию, обязательно нужна схема блочного шифрования с монотонно увеличивающейся длиной ключа и уменьшающейся длиной шифротекста. В этой конструкции защиты важен выбор длины дискрет приращения ключа шифрования.

Длина дискреты роста ключа шифрования может быть любой в интервале от 1 до 16 бит. Этот интервал соответствует монотонному росту вычислительной сложности работы с непрерывными данными. В точке $n = 16$ многомерная энтропия состояний преобразователей биометрия-код меняет свои свойства и начинают работать описанные в предыдущей главе интегральные атаки. Интегральные атаки становятся тем эффективнее, чем длиннее биометрический код доступный для статистического исследования.

2.9 Сравнительный анализ биометрических и специальных защитных свойств «нечетких экстракторов» (технология США, Канады и Евросоюза) с нейросетевыми преобразователями биометрия-код, выполненными по требованиям пакета российских стандартов ГОСТ Р 52633.xx

2.9.1 Обеспечение высокого уровня защитных свойств нейросетевыми технологиями и технологиями «нечетких экстракторов»

Одной из важнейших проблем развития современного информационного общества является проблема безопасной биометрии-

ческой аутентификации личности в сети Интернет, например, при обращении граждан к электронному правительству. При авторизованном доступе к значимым для граждан информационным ресурсам нельзя пользоваться обычными паспортами и идентификационными картами. В связи с этим во всех развитых странах ведутся работы по решению проблемы биометрико-криптографической аутентификации личности в открытых информационных пространствах.

В частности США, Канада и страны Евросоюза активно вкладывают свои ресурсы в развитие так называемых «нечетких экстракторов» [62]. Технология строится на преобразовании биометрических данных в код и исправлении примерно 25% ошибок в разрядах био-кода. Корректировка ошибок осуществляется классическими кодами с обнаружением и исправлением ошибок.

Для того чтобы защитить биометрические данные при транспорте и хранении их накрывают гаммой, полученной из ключа аутентификации. Схема формирования и применения «нечетких контейнеров» приведена на рисунке 2.19.

При формировании «нечеткого контейнера» используют биометрический код, полученный с некоторым числом ошибок. Для того чтобы защитить этот код используют гамму, которую получают из ключа аутентификации, размещая его в информативных разрядах какого-либо классического самокорректирующегося кода. В избыточной части кода размещают, вычисленные показатели синдромов ошибок самокорректирующегося кода [62]. Сложение гаммы из избыточного самокорректирующегося кода и био-кода образует «нечеткий контейнер», который может храниться в базе данных.

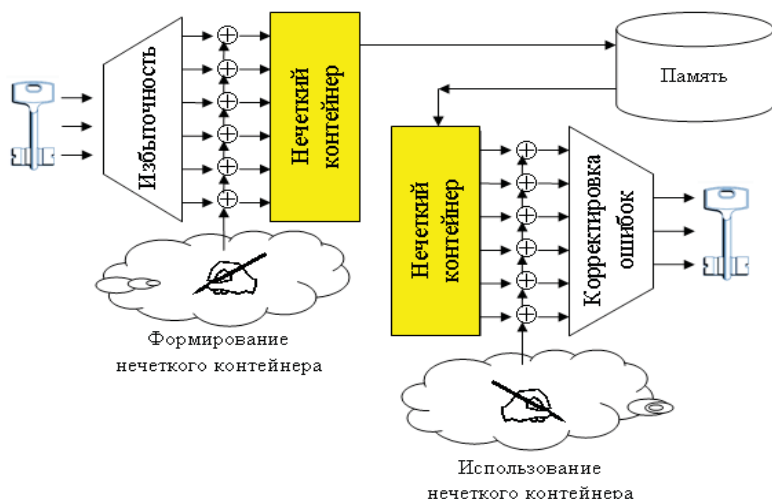


Рис. 2.19. Формирование и использование «нечетких контейнеров»

При использовании «нечеткого контейнера» его извлекают из базы данных и складывают по модулю два с проверяемым био-кодом. Сложение по модулю два снимает защитную гамму и дает некоторый самокорректирующийся код ключа аутентификации. Состояния разрядов этого кода восстанавливают, пользуясь его корректирующей способностью, в итоге мы получаем восстановленный (скорректированный) код ключа аутентификации.

Российские и казахские ученые предложили иной подход к биометрической защите информации, который построен на использовании больших искусственных нейронных сетей. В соответствии с базовым стандартом ГОСТ Р 52633.0-2006 обучают нейронную сеть таким образом, чтобы при воздействии на нее примерами образа «Свой» на выходах нейронной сети возникал ключ аутентификации. При этом искусственная нейронная сеть должна учиться автоматически (автомат обучения должен реализовываться по ГОСТ Р 52633.5-2011). Далее параметры обученной нейронной сети должны размещаться в таблицах нейросетевого контейнера, выполненного по ГОСТ Р 52633.4-2011. Нейросетевой контейнер может храниться в памяти компьютера или

базе данных. На рисунке 2.20 приведена схема формирования и применения нейросетевых контейнеров.

При применении нейросетевого контейнера его данные передают программному эмулятору, который воссоздает обученную искусственную нейронную сеть. При предъявлении на вход нейросети примера образа «Свой» на выходах нейросети появляется ключ аутентификации. Если злоумышленник предъявит случайный пример образа «Чужой», то на выходах нейронной сети появится случайный код.

Защита персональных данных личности осуществляется тем, что по данным обученной нейронной сети сложно восстановить параметры биометрического образа. В случае если процедура аутентификации проводится в доверенной вычислительной среде подобной защиты достаточно.

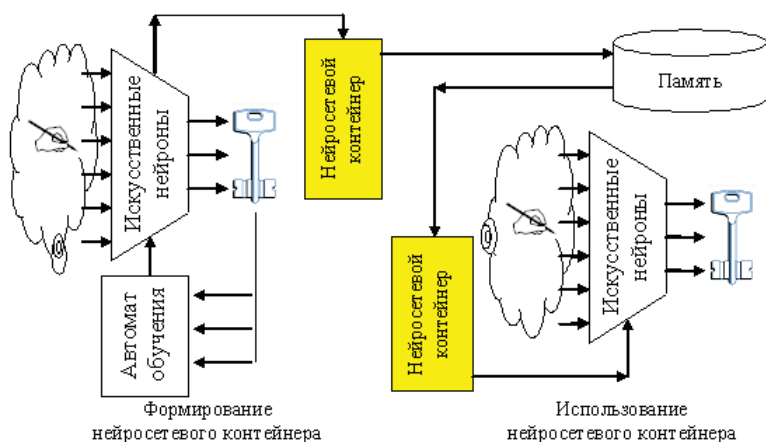


Рис. 2.20. Формирование и применение нейросетевого контейнера

В более сложных случаях хранения данных в открытых базах или при транспорте данных по открытым каналам связи используют защищенные нейросетевые контейнеры. Защита нейросетевых контейнеров осуществляется их самошифрованием. При самошифровании используется часть ключа аутентификации, на которой шифруются данные еще не использованных нейронов. Схема самошифрования [36] иллюстрируется рисунком 2.21.

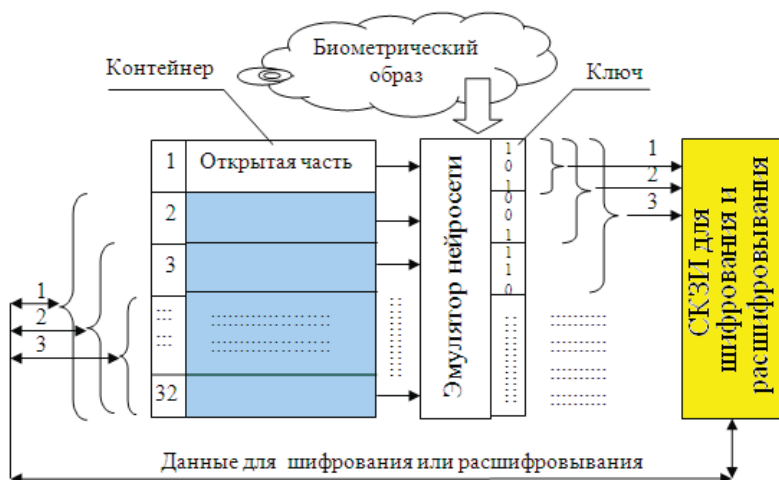


Рис. 2.21. Шифрование данных еще не использованных нейронов блоками с возрастающей длиной ключа 3 бита, 6 бит, 9 бит, 12 бит и так далее

Защищенный нейросетевой контейнер всегда имеет открытую часть данных (на рисунке 2.21 нет заливки темным цветом). Например, это могут быть данные первых трех нейронов обученной нейронной сети. Если во время биометрической аутентификации будет предъявлен образ «Свой», то первые три нейрона сработают верно, в первых трех разрядах возникнет верный код, который расшифрует данные всех последующих нейронов. В частности будут расшифрованы данные трех следующих нейронов. Возникает итерационный процесс постепенного расшифровывания групп еще не использованных нейронов заранее обученной искусственной нейронной сети. Процесс шифрования данных контейнера и процесс их расшифровывания для образов «Свой» оказывается полностью обратимым.

Совершенно иная ситуация возникает, когда на вход эмулятора нейросети предъявляются данные биометрического образа «Чужой». В этом случае после первого неверно угаданного бита верное расшифровывание прекращается, возникает эффект размножения биометрических ошибок. После первого неверно

угаданного бита «Чужим» происходит хэширование всех последующих данных из-за неверного расшифровывания параметров обученной нейросети.

Если попытаться тестировать защищенный нейросетевой преобразователь биометрия-код, то он будет показывать очень высокие показатели защиты. Примеры выходных статистик расстояний Хэмминга между кодом «Свой» и кодами «Все Чужие» для открытого и защищенного нейросетевого преобразователя дан на рисунке 2.22 (рисунок заимствован из приложения В ГОСТ Р 52633.3-2011).

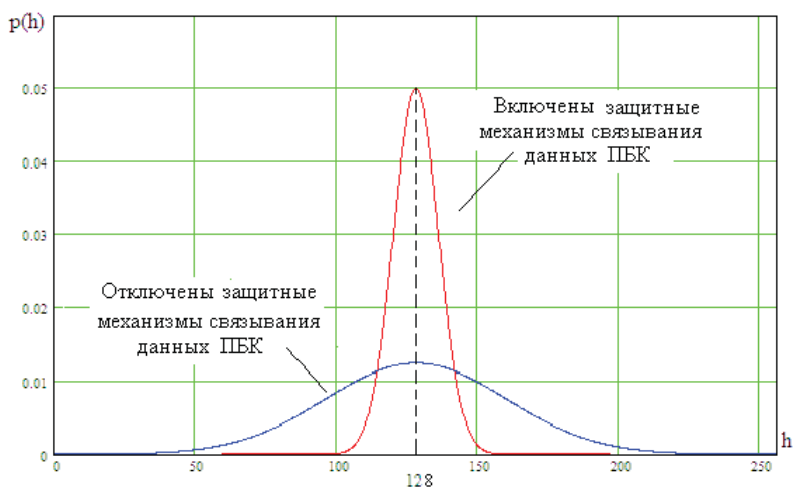


Рис. 2.22. Распределения расстояний Хэмминга «Свой»/«Чужие» для открытого нейросетевого контейнера и защищенного самошифрованием нейросетевого контейнера

Из рисунка 2.22 видно, что выходные статистики кодов открытого и защищенного нейросетевых контейнеров, выполненных по требованиям пакета стандартов ГОСТ Р 52633.xx с высокой точностью описываются нормальным законом распределения. Именно по этой причине ГОСТ Р 52633.3-2011 оценку стойкости нейросетевых преобразователей биометрия-код (ПБК) рекомендует выполнять через предсказание вероятности попадания расстояния Хэмминга в интервал от 0 до 1. Подобный способ

вычисления дает значения вероятности ошибок второго рода (вероятность пропуска «Чужого») порядка 0.0000001 и менее. Если этот же нейросетевой контейнер защитить самошифрованием (включить механизм связывания данных обученного ПБК), то вероятность ошибок второго рода резко падает до величины $P_2 \approx 2 \cdot 256$ при условии, что нейросетевой преобразователь биометрия-код имеет 256 выходов. Столь значительное падение наблюдаемой вероятности ошибок второго рода является следствием того, что наблюдение реальных статистик расстояний Хэмминга становится недоступным из-за самошифрования (размножения ошибок). Возникает иллюзия очень высокой стойкости биометрической защиты. Для того чтобы такой иллюзии не возникало ГОСТ Р 52633.3-2011 требует оценивать стойкость к атакам подбора открытого нейросетевого контейнера, так как после защиты шифрованием данных нейросетевого контейнера уже ничего нельзя увидеть.

Совершенно иная ситуация возникает в «нечетких экстракторах», предлагаемых к использованию фирмами США, Канады, Евросоюза. Проблема «нечетких экстракторов» состоит в том, что простое гаммирование сравниваемых кодов не приводит к изменению расстояний Хэмминга между гаммированным кодом «Свой» и гаммированным кодом «Чужой»:

$$HD = \sum_{i=1}^n c_i \oplus x_i = \sum_{i=1}^n (g_i \oplus c_i) \oplus (x_i \oplus g_i) = \sum_{i=1}^n (g_i \oplus (g_i \oplus (c_i \oplus x_i))) \quad (2.5)$$

где c_i – разряды био-кода «Свой», x_i – разряды био-кода «Чужой», g_i – разряды защищающей био-коды гаммы.

При вычислении расстояний Хэмминга гаммирование сравниваемых кодов снимается (2.5), так как операция сложения по модулю два коммутативна (последовательность операций сложения по модулю два можно выполнять в любом порядке). Последнее означает, что все известные на данный момент атаки, построенные на вычислении расстояний Хэмминга могут быть применены к восстановлению кода «Свой» из данных, размещенных в «нечетком контейнере». Все эти неприятности являются следствием возможности наблюдать статистики распределений расстояний

Хэмминга между кодом «Свой» и кодами «Чужие» (2.5), защитить эти статистики простое гаммирование био-кодов не может. В этом отношении нейросетевые преобразователи биометрия-код оказываются гораздо более эффективными, так как их защита самошифрованием перемешивает данные и защищает от наблюдения статистики распределения расстояний Хэмминга (сжимает распределение расстояний Хэмминга как это показано на рисунке 2.22).

Отсутствием защиты распределений расстояний Хэмминга фактически пользуется Даугманом [62] оценивая энтропии «нечеткого экстрактора», осуществляющего преобразование рисунка радужной оболочки глаза в био-код аутентификации. На рисунке 2.23 приведены данные распределения расстояний Хэмминга между кодами «Свой» и «Чужие», полученные численным экспериментом [62].

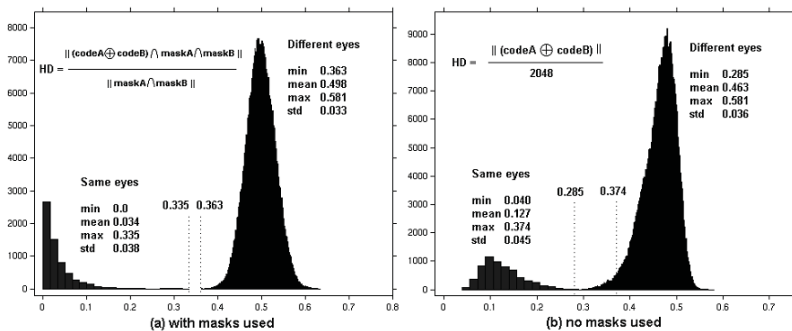


Рис. 2.23. Распределение нормированных расстояний Хэмминга «нечеткого экстрактора» по Даугману [62]

В правой части рисунков 2.23 (a) и 2.23 (b) даны распределения расстояний Хэмминга для био-кодов «Свой» и био-кодов «Чужой», в левой части рисунков 2.23 (a) и 2.23 (b) приведены распределения расстояний Хэмминга между био-кодами «Свой». Стойкость скорректированного био-кода к атакам подбора оценивается расстояниями между краями двух распределений расстояний Хэмминга. На рисунке 2.23 эти расстояния отмечены пунктирными линиями.

Фактически применение в «нечетких экстракторах» кодов с обнаружением и исправлением ошибок приводит к тому, что все коды «Свой» корректируются до одного состояния. Это эквивалентно смещению начала оси координат и изменению ее шкалы. В частности, для левой части рисунка Даугмана [62], соответствующие деформации иллюстрируются рисунком 6, который построен исходя из гипотезы о длине выходного кода 256 бит (по аналогии с рисунком 2.22).

Сравнивая рисунок 2.22 и правую часть рисунка 2.24, легко убедиться в отсутствии у «нечетких экстракторов» механизма размножения ошибок. Распределение расстояний Хэмминга био-кодов «Свой» и «Чужой» для «нечетких экстракторов» будет лежать в интервале от 25 бит до 230 бит при использованной Даугманом [62] базы из 700 примеров образов «Чужой». Все это позволяет выделить примерно 10 наиболее близких образов «Чужой» с расстояниями Хэмминга от 20 до 50 и наиболее далеких образов «Чужой» с расстоянием Хэмминга от 200 до 230 бит. Положение наиболее близких и наиболее далеких образов «Чужой» на рисунке 6 отмечено овалами на правом и левом «хвостах» нормального распределения.

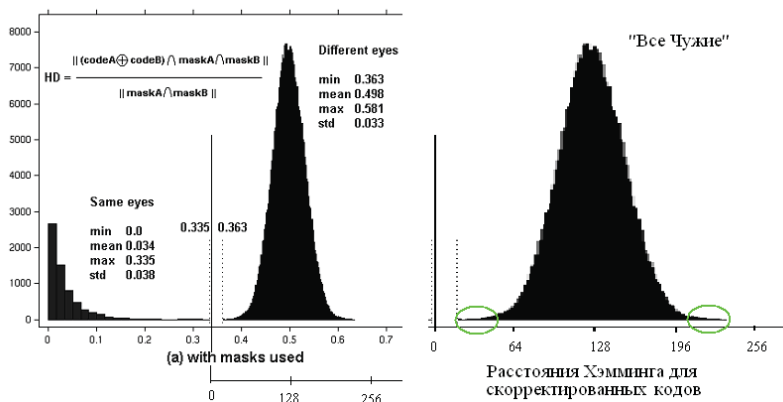


Рис. 2.24. Учет эффекта правки кодов «Свой» и «Чужой» самокорректирующимся кодом (распределение расстояний Хэмминга кодов «Свой» сжимается в точку)

Если мы умеем выделять наиболее близкие биометрические образы «Чужой», то мы можем использовать их как генетический материал и создать скрещиванием следующее поколение образов, которое будет состоять из гораздо более похожих «Чужих» на образ «Свой». Скрещивать похожие образы и селективировать их можно в течении нескольких поколений, постепенно компрометируя неизвестный образ «Свой». Эта процедура подробно описана в стандарте по тестированию ГОСТ Р 52633.3-2011. На рисунке 2.25 иллюстрируется движение распределения кодов «Чужие» в сторону кодов «Свой» при смене поколений искусственно синтезируемых образов [36].

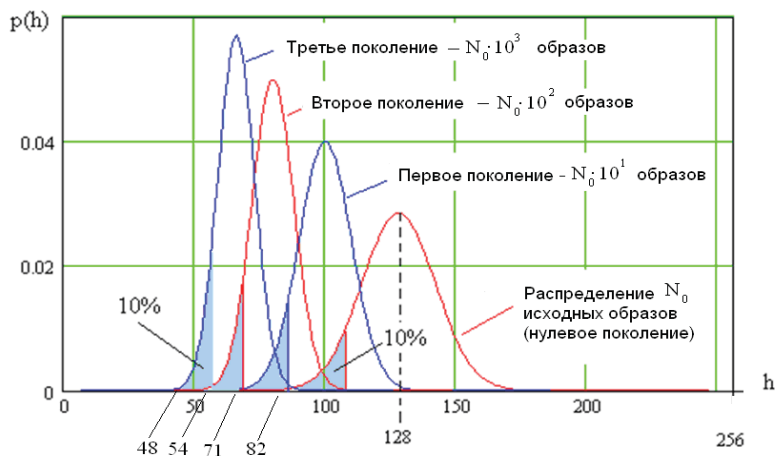


Рис. 2.25. Смещение распределений расстояний Хэмминга при последовательной селекции наиболее близких образов к образу «Свой» в нескольких поколениях

Практика показывает, что через 7 поколений удастся скомпрометировать до 90% биометрического образа «Свой» и кода ключа «Свой», располагая исходной базой из 10 000 биометрических образов «Чужой». Однако столь эффективный способ компрометации персональных аутентификационных данных перестает работать, если включен механизм размножения ошибок (рисунок 2.21). Самошифрование нейросетевых контейнеров эквивалентно гаммированию с обратной связью. При таком способе защиты

первая же ошибка в биометрических данных приводит к изменению синтезируемой далее гаммы, механизм снятия гаммы (2.5) при вычислении расстояний Хэмминга перестает работать.

Комбинация кодов образов «Чужой» наиболее близких и наиболее далеких используется в том случае, когда необходимо восстановить неизвестное значение кода «Свой» (или гаммированного значения кода «Свой»). Для этой цели усредняют значения состояний разрядов близких кодов и инверсии далеких кодов (чем дальше код от кода «Свой», тем он ближе к инверсному коду «Свой»). Этот прием позволяет проявить разряды кода «Свой» или его варианта, накрытого неизвестной гаммой при гаммировании без обратной связи.

Все выше приведенные аргументы свидетельствуют о том, что защиты биометрических данных простым гаммированием недостаточно. Для получения действительно надежной защиты персональных биометрических данных простое гаммирование должно быть усилено механизмом перемешивания данных, как это сделано в алгоритмах стойкого симметричного шифрования DES, AES, ГОСТ Р 28147-86.

Вывод 1. Нейросетевые преобразователи биометрия-код после защиты их данных самошифрованием оказываются эффективнее чем «нечеткие экстракторы», так как они осуществляют защиту данных одновременно и гаммированием и перемешиванием. Применять «нечеткие экстракторы» в России, Казахстане и Белоруссии нецелесообразно.

2.9.2 Обеспечение высокого уровня биометрических свойств нейросетевыми технологиями и технологиями «нечетких экстракторов»

Важнейшими показателями технологий биометрической аутентификации является обеспечиваемые ими вероятности ошибок первого и второго рода. Ошибки первого рода возникают когда средство биометрической аутентификации отказывает в доступе «Своему». В технологии «нечетких экстракторов» ошибки био-кода «Свой» правятся классическими кодами с обнаружением и исправлением ошибок. Проблема применения классических самокорректирующихся кодов состоит в том, что они избыточны.

При необходимости править большое число ошибок приходится использовать коды с очень большой избыточностью. Коды БЧХ способны править [36]:

- до 4% ошибок при 50% избыточности;
- до 6 % ошибок при 100% избыточности;
- до 8 % ошибок при 200% избыточности;
- до 10% ошибок при 400% избыточности;
- до 12 % ошибок при 800% избыточности;
- до 14 % ошибок при 1600% избыточности.

Классических кодов, способных исправлять 50% ошибок не существует, так как такие коды имеют огромную избыточность и пренебрежимо малую информационную часть. Половина ошибок – это теоретический предел для любых самокорректирующихся кодов, построенных в рамках гипотезы равновероятного распределения ошибок по их разрядам.

Для нейронных сетей 50% предела корректируемых ошибок нет. Нейронные сети (второй слой нейронов сети) при их обучении по ГОСТ Р 52633.5-2011 могут править коды с ошибками, число которых выше 50%. Это принципиально новое свойство возникает у искусственных нейронных сетей из-за того, что они обучаются. При обучении нейронные сети учитывают вероятности появления ошибок в каждом из анализируемых разрядов био-кода. Учет неравномерности распределения ошибок в разрядах био-кода является принципиально важным преимуществом нейросетевых корректоров био-кода. Именно из-за этого нейронные сети всегда дают более низкие значения ошибок первого и второго рода в сравнении с «нечеткими экстракторами».

Нейронные сети потеряют свое преимущество по сравнению с «нечеткими экстракторами», только в том случае, если математики смогут синтезировать самокорректирующиеся коды с учетом конкретного распределения ошибок в разрядах восстанавливаемого био-кода.

Поясним сказанное на конкретных примерах. Так Даугман [62] получает из рисунка радужной оболочки глаза 2048 битный био-код. Из левой части рисунка 2.23 б) следует, что ошибки в коде могут составлять до 25%. Если мы будем править 14%, то получим 16-ти кратную избыточность. То есть длина ключа (эн-

тропия ключа) составит $2048/16=128$ бит, что хорошо согласуется с данными самого Даугмана [62].

Следует заметить, что рисунок радужной оболочки глаза – это наиболее качественный биометрический образ из всех известных. Если пытаться применять биометрические данные рисунка отпечатка пальца, то данные получаются много хуже. Так, если из рисунка отпечатка пальца извлекать 256 контролируемых параметров и применять 12 кратную избыточность кода БЧХ, то получим $256/12=21$ бит, что так же хорошо согласуется с данными отечественных авторов [63].

Предположительно, что отказ от классических самокорректирующихся кодов и применение искусственных нейронных сетей позволит от полтора до двух раз увеличить энтропию выходных кодов. Для радужной оболочки глаза она может составить 256 бит, для рисунка отпечатка пальца энтропия может составить порядка 42 бит.

Вывод 2. Биометрические свойства нейросетевых преобразователей существенно выше, чем аналогичные свойства «нечетких экстракторов» в силу того, что нейронные сети используют обучение под каждый реальный не равномерный закон распределения ошибок в разрядах корректируемого био-кода. Применять «нечеткие экстракторы» в России, Казахстане и Белоруссии нецелесообразно.

2.9.3 Обеспечение стандартами нейросетевых технологий и технологий «нечетких экстракторов»

Одним из наиболее важных аргументов в пользу применения той или иной технологии является то, насколько эта технология стандартизована. В этом отношении безусловное преимущество принадлежит технологии применения нейросетевых преобразователей биометрия-код аутентификации. В период с 2005 по 2012 год под нейросетевую технологию в России создано 8 стандартов и введено в действие 7 стандартов, номера и названия этих стандартов даны в таблице 2.2.

Таблица 2.2

№	Номер и полное название стандарта
1	ГОСТ Р 52633.0-2006 «Защита информации. Техника защиты информации. Требования к средствам высоконадежной биометрической аутентификации»
2	ГОСТ Р 52633.1-2009 «Защита информации. Техника защиты информации. Требования к формированию баз естественных биометрических образов, предназначенных для тестирования средств высоконадежной биометрической аутентификации»
3	ГОСТ Р 52633.2-2010 «Защита информации. Техника защиты информации. Требования к формированию синтетических биометрических образов, предназначенных для тестирования средств высоконадежной биометрической аутентификации».
4	ГОСТ Р 52633.3-2011 «Защита информации. Техника защиты информации. Тестирование стойкости средств высоконадежной биометрической защиты к атакам подбора»
5	ГОСТ Р 52633.4-2012 «Защита информации. Техника защиты информации. Интерфейсы взаимодействия с нейросетевыми преобразователями биометрия-код»
6	ГОСТ Р 52633.5-2011 «Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия-код доступа»
7	ГОСТ Р 52633.6-2012 «Защита информации. Техника защиты информации. Требования к индикации близости предъявленных биометрических данных образу «Свой»
8	ГОСТ Р 52633.7-201х «Защита информации. Техника защиты информации. Высоконадежная мультибиометрическая аутентификация»

Применительно к использованию «нечетких экстракторов» на сегодняшний день созданы только стандарты самого верхнего уровня. По уровню стандартизации эти технологии существенно уступают отечественным технологиям биометрико-нейросетевой аутентификации. Названия и номера соответствующих международных стандартов приведены в таблице 2.3.

Таблица 2.3

№	Номер и полное название стандарта
1	ISO/IEC 24745:2011 Information technology -- Security techniques -- Biometric information protection
2	ISO/IEC 24761:2009 Information technology -- Security techniques -- Authentication context for biometrics
3	ISO/IEC 19792:2009 Information technology. Security techniques. Security evaluation of biometrics.

Вывод 3. На текущий момент нейросетевые преобразователи биометрия-код имеют более высокий уровень стандартизации в сравнении с «нечеткими экстракторами». Для технологии нечетких экстракторов пока созданы только базовые стандарты высокого уровня (эквивалент ГОСТ Р 52633.0-2006). Процесс стандартизации технологии «нечетких экстракторов» идет с задержкой от 3 до 5 лет в сравнении с процессом стандартизации технологии нейросетевых преобразователей биометрия-код.

ЗАКЛЮЧЕНИЕ

В коллективной монографии коллектива российских и казахских ученых рассмотрены вопросы обучения и тестирования средств биометрико-нейросетевой аутентификации, достоинства и недостатки биометрических средств идентификации и аутентификации, разработанных по требованиям международного стандарта BioAPI и российского стандарта серии 52633.XX (7 стандартов). Отмечено, что международный стандарт BioAPI не рассматривает вопросы дистанционной защищенной идентификации и определяет требования только для полицейских средств локальной идентификации. При этом необходима ручная настройка средства идентификации и постоянный контроль специального персонала.

Показано, что достоинствами средств биометрико-нейросетевой аутентификации, разработанных по требованиям ГОСТ Р 52633, являются: автоматический процесс обучения нейросетевого преобразователя биометрия-код самостоятельно самим пользователем (без вмешательства персонала); экспресс-контроль стойкости средства после процесса обучения, гарантированное удаление биометрических данных из вычислительной среды после экспресс-контроля; невозможность восстановления шаблона биометрических образов из сформированного после процесса обучения биометрико-нейросетевого контейнера; гарантированная анонимность, конфиденциальность и обезличенность персональных данных, находящихся в биометрико-нейросетевом контейнере; безопасное хранение в больших базах данных (в том числе удаленных (облачных)), обработка и передача в системах обработки данных; возможность как локальной, так и дистанционной высоконадежной аутентификации личности.

В качестве биометрического идентификатора выбран – папиллярный рисунок отпечатка пальца. Основная причина данного выбора – наиболее доступный и приемлемый для каждого гражданина. Доля мирового рынка биометрических средств, использующих данный идентификатор составляет, по отдельным зарубежным оценкам, более 52 %, что дает определенные надежды

на использование разрабатываемых средств аутентификации не только на рынке Республики Казахстан, но и за ее пределами.

Монография подготовлена в рамках выполнения комплексного проекта по теме: «Разработка и подготовка производства телекоммуникационного оборудования, разработка программного сетевого, прикладного и специального обеспечения для создания цифровых сетей связи с персонализированным доступом» в соответствии с Постановлением Правительства РФ №218 от 09.04.2010 г. (Договор № 13G25.31.0039 от 07.09.2010 г.) и научного проекта по теме: №753. МОН.ГФ.13.13 «Исследование вариантов реализации и разработка действующего лабораторного образца ON-LINE системы биометрического обезличивания электронных историй болезней для медицинского учреждения».

Материал монографии может быть полезен для специалистов по искусственным нейронным сетям, защите информации в системах медицинского электронного документооборота, создания электронных карт и паспортов с автоматической биометрической аутентификации личности.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1 ГОСТ Р 52633.5-2011. Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия-код доступа. – М.: Стандартинформ, 2011. – 20 с.

2 ГОСТ Р 52633.0-2006. Защита информации. Техника защиты информации. Требования к средствам высоконадежной биометрической аутентификации. – М.: Стандартинформ, 2006. – 24 с.

3 ГОСТ Р 52633.1-2009. Защита информации. Техника защиты информации. Требования к формированию баз естественных биометрических образов, предназначенных для тестирования средств высоконадежной биометрической аутентификации. – М.: Стандартинформ, 2009. – 24 с.

4 Ахметов Б.С., Волчихин В.И., Иванов А.И., Малыгин А.Ю. Алгоритмы тестирования биометрико-нейросетевых механизмов защиты информации (Монография). – Алматы: Изд-во КазНТУ им. К.И. Сатпаева, 2013 - 152 с. ISBN 978-101-228-586-4, <http://portal.kazntu.kz/files/publicate/2014-01-04-11940.pdf>

5 Ахметов, Б.С., Иванов А.И. Нейросетевая мультибиометрическая аутентификация личности гражданина в системе электронного правительства /Б.С. Ахметов, В.А. Алисов, С.Е. Вятчанин //В сборнике трудов Международного симпозиума «Надежность и качество – 2012». – Пенза: Изд-во ПГУ, 2012. – Том 1. – С. 227-229.

6 B. Akhmetov, V. Volchihin, A. Malygin, A. Ivanov, T. Kartbayev. Highly reliable multi-biometric authentication of human-being personality to support citizens interaction with E-government and E-busines //3rd world conference on information technology. University of Barcelona, Faculty of Library and Information Science, Spain, Barcelon, 14-16 november, 2012. – P. 74-81.

7 Надеев Д.Н. Синтез таблиц вероятности ошибок первого и второго рода для неидеальных биометрико-нейросетевых преобразователей с 256 выходами «Нейрокомпьютеры: разработка, применение» – 2007. – №12. – С. 30-31

8 V. Volchikhin, A. Ivanov, A. Malygin, I. Urnev. Information-telecommunication system with multibiometric protection of user's

personal data «Progress in Electromagnetics Research Symposium» (PIERS 2012). Published by The Electromagnetics Academy 777 Concord Avenue, Suite 207 Cambridge. – P. 70-73

9 Ахметов Б. С., Иванов А.И., Малыгин А.Ю., Елфимов А.В. Синхронизация процедур нейросетевого обучения особенностям рукописного почерка по точкам смены направления движения пера. //В сб. трудов II Международной НПК «Информационно-инновационные технологии: интеграция науки, образования и бизнеса». Республика Казахстан, Алматы: Изд-во КазНТУ им. К.И. Сатпаева, 2011. Том 2. – С. 118-123.

10 B. Akhmetov, A. Ivanov, A. Malygin, I. Urnev, V. Funtikov Statistical description of output states of the neural network «biometrics-code» transformers «Progress in Electromagnetics Research Symposium» (PIERS 2012). Published by The Electromagnetics Academy 777 Concord Avenue, Suite 207 Cambridge P.61-65

11 Стратанович Р.Л. Теория информации. М.: Советское радио, 1975 г., 424 с.

12 Кульбак С. Теория информации и статистика. М.: Наука, 1967 г.

13 Яглом А.М., Яглом И.М. Вероятность и информация. М.: Дом Книги, 2007 г., 512с.

14 Иванов А.И., Фунтиков В.А., Майоров А.В., Надеев Д.Н. Моделирование кодовых последовательностей с энтропией естественных и искусственных биометрических языков. Инфокоммуникационные технологии Том 8, № 4 , 2010 г., с. 75-79, <http://ikt.psuti.ru>.

15 Фунтиков В.А. Энтропийный подход к тестированию стойкости нейросетевого преобразователя биометрия-код к атаке подбора/ «Нейрокомпьютеры: разработка, применение» 2009. – №6. – С. 25-29.

16 Иванов А.И., Малыгин А.Ю., Безяев А.В. Проверка гипотезы нормального распределения статистических значений выходного кода преобразователя биометрия-код по критерию Пирсона χ^2 //В сб. мат. Всероссийск. научно-технической конф. «Информационные и телекоммуникационные технологии» - Рязань: Изд-во РВВКУС, 2009. – С. 80-86.

17 Малыгин А.Ю., Иванов А.И., Безяев А.В., Зюзин Ю.М.

Синтез таблиц вероятности ошибок первого и второго рода для нейросетевых преобразователей биометрия-код //В сборнике материалов Международного симпозиума «Надежность и качество - 2010». - Пенза: Изд-во ПГУ, 2010. – С. 100-103.

18 Надеев Д.Н. Моделирование биномиального зависимого закона распределения значений вероятностей ошибок нейросетевых преобразователей для высоконадежной биометрической защиты «Вопросы защиты информации», 2008, № 3, с. 31-35.

19 Малыгин А.Ю., Надеев Д.Н., Иванов А.И. Две причины не идеальности нейросетевых преобразователей биометрия-код по выходным случайным состояниям кодов «Чужие» /Вопросы защиты информации № 2(81) 2008., с.19-21

20 Иванов А.И., Майоров А.В., Язов Ю.К. Энтропийно-корреляционная оценка хэширующих свойств нейросетевого преобразователя биометрия-код доступа/ Нейрокомпьютеры: разработка, применение» М.: Радиотехника. №3, 2012. С. 35-40.

21 Ахметов Б.С., Волчихин В.И., Куликов С.В., Малыгина Е.А. Моделирование длинных биометрических кодов, воспроизводящих корреляционные связи выходных данных нейросетевого преобразователя/ Нейрокомпьютеры: разработка, применение» М.: Радиотехника. №3, 2012. С. 40-43.

22 Y. Dodis, L. Reyzin, A. Smith Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy, Data April 13, 2004.

23 A. Cavoukian, A. Stoianov Biometric Encryption: A Positive-Sum Technology that Achieves Strong Authentication, Security AND Privacy, March 2007, <http://www.ipc.on.ca>.

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

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

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

27 Ю.К.Язов, И.Г.Назаров, А.И.Иванов, О.В.Ефимов. Нейросетевая технология защиты личных биометрических данных// Нейрокомпьютеры: разработка и применение. 2009. №9. С. 5-10.

28 В.А.Фунтиков. Энтропийный подход к тестированию стойкости нейросетевого преобразователя биометрия-код к атакам подбора// Нейрокомпьютеры: разработка и применение. 2009. №9. С. 56-58.

29 Малыгин А.Ю., Волчихин В.И., Иванов А.И., Фунтиков В.А. Быстрые алгоритмы тестирования нейросетевых механизмов биометрико-криптографической защиты информации /Пенза-2006 г., Издательство Пензенского госуниверситета., 161 с.

30 Захаров О.С., Иванов А.И. Учет корреляционных связей биометрических данных через дробный показатель степеней свободы закона распределения значений хи-квадрат. «Инфокоммуникационные технологии» Том 6, № 1, 2008 г., с. 12-15.

31 Кнут Д. Искусство программирования. Т. 2. М.: «Наука» 1972 г.

32 Брюс Шнайер Прикладная криптография: протоколы, алгоритмы, исходные тексты на языке Си». М.: «Триумф» 2002 г. <http://uritix.com/r-and-d/lectures>

33 Шалыгин А.С., Палагин Ю.И. Прикладные методы статистического моделирования. Л.: Машиностроение, 1986 г. – 320 с.

34 Иванов А.И. Биометрическая идентификация личности по динамике подсознательных движений. – Пенза: Изд-во Пенз. гос. ун-та, 2000. – 188 с.

35 Положительное решение по заявке на изобретение № 201012606/09(003581) «Способ однозначного хэширования неоднозначных биометрических данных», авторы: Иванов А. И., Иванов С. М., Назаров И. Г., Фунтиков В. А., Ефимов О. В. Майоров А. В., Язов Ю. К.

36 Волчихин В.И., Иванов А.И., Назаров И.Г., Фунтиков В.А., Язов Ю.К. Нейросетевая защита персональных биометрических данных М: Изд-во «Радиотехника», 2012 г. – С 157-160.

37 Ахметов Б.С., Безяев В.С., Иванов А.И., Малыгин А.Ю., Сериков И.В. Сравнительная оценка уровня безопасности средств биометрической аутентификации //В сб. трудов II Международной НПК «Информационно-инновационные технологии: инте-

грация науки, образования и бизнеса». Республика Казахстан, Алматы: Изд-во КазНТУ им. К.И. Сатпаева, 2011. Том 1 – С.382 – 386.

38 Цветков Э. И. Методические погрешности статистических измерений – Л.: Энергоатомиздат, 1984 г., - 190 с.

39 ГОСТ Р 50779.21-2004 «Статистические методы. Правила определения и методы расчета статистических характеристик по выборочным данным. Часть 1. Нормальное распределение».

40 Боровиков В. STATISTICA: искусство анализа данных на компьютере. Для профессионалов – СПб.: Питер, 2001, - 656 с.

41 Волчихин В.И., Иванов А.И., Фунтиков В.А., Малыгин А.Ю. и др. Нейросетевые преобразователи биометрических образов человека в код его личного криптографического ключа (под редакцией А.Ю. Малыгина) М: Изд-во Радиотехника. Научная серия «Нейрокомпьютеры и их применение» Книга 29. 2008– 88 стр.

42 Волчихин В.И., Иванов А.И., Фунтиков В.А. Быстрые алгоритмы обучения нейросетевых механизмов биометрико-криптографической защиты информации. Монография. Пенза-2005 г. Издательство Пензенского государственного университета, 273 с.

43 Ахметов Б.С., Волчихин В.И., Малыгин А.Ю., Урнев И.В. Учет влияния корреляционных связей на результаты тестирования преобразователей биометрия-код В сб. трудов Международной научно-практической конф. «Информационные и телекоммуникационные технологии: образование, наука, практика», Республика Казахстан, Алматы: Изд-во Казахского национального технического ун-та, 2012, С.34 – 37.

44 Волчихин В.И., Иванов А.И., Фунтиков В.А., Малыгина Е.А. Перспективы использования искусственных нейронных сетей с многоуровневыми квантователями в технологии биометрико-нейросетевой аутентификации // Известия высших учебных заведений. Поволжский регион. Технические науки. 2013, №4(28) С. 88 – 99.

45 Э. Локар. Руководство по криминалистике. -М: Юридическое издательство НКЮ СССР, 1941. -544 с.

46 Статистическая дактилоскопия: Методические проблемы / Под ред. Л.Г. Эджубова. -М.:Городец, 1999 .-184 с.

47 Болл Руд и др. Руководство по биометрии. / Болл Руд, Коннел Джонатан Х., Панканти Шарат, Ратха Налини К., Сеньор Эндрю У. // -М.: Техносфера, 2007. -368 с.

48 Куликов С.В., Секретов М.В., Захаров О.С., Иванов А.И. Учет «тяжелых хвостов» ненормального закона распределения биометрических параметров все «Чужие» при настройке нелинейного элемента нейрона с несколькими дискретными состояниями/ «Нейрокомпьютеры, разработка, применение» № 3, 2012.с. 56-60.

49 Рыбалкин С.Б., Иванов А.И. Технология биометрической идентификации, обеспечивающая анонимность больных при ведении электронных историй социально значимых заболеваний /Современные технологии безопасности 2006 г., № 3,4 (18,19), с.55-57.

50 Ашенбрер И.В., Иванов А.И., Рыбалкин С.Б. Нейросетевое обезличивание медицинского документооборота – эффективный прием биометрической защиты персональных данных. «Нейрокомпьютеры: разработка, применение» №12, 2007, с.80-82.

51 Патент РФ 2473125 08.12.2011. Способ классификации рисунков отпечатка пальца при аутентификации личности /Волчихин В.И., Фунтиков В.А., Иванов А.И., Секретов М.В., Майоров А.В., Малыгин А.Ю. <http://elibrary.ru/item.asp?id=18859157>.

52 Патент РФ №2371765. Способ анонимной биометрической регистрации человека / Иванов А.И.– №2008101520/09; Заявл. 14.01.2008 // Описание изобретения к патенту.–№30.

53 Ahmetov A.A., Doszhanova B.S., Ivanov A.I.,A. Malygin, Kartbaev T.S. Biometric depersonalization electronic medical patients illness stories //The 6th International Conference on Public Safety, Republic of Poland, Poznan, 2013.

54 DRAFT: SSL 3.0 specification, 1996, <http://wp.netscape.com/eng/ssl3/draft302.txt>;

55 RFC 2246: «The TLS Protocol Version 1.0», 1999, <http://tools.ietf.org/html/rfc2246>;

56 RFC 4346: «The Transport Layer Security (TLS) Protocol Version 1.1», 2006, <http://tools.ietf.org/html/rfc4346>;

57 RU 2273877 – «Способ распределения ключей в большой территориально разнесенной системе» патент на изобретение,

дата приоритета 16.08.2004, авторы: Ефимов О.В., Иванов А.И., Фунтиков В.А., патентообладатель ФГУП «ПНИЭИ» (RU).

58 ГОСТ Р ИСО/МЕК 19784-1 «Автоматическая идентификация. Идентификация биометрическая. Биометрический программный интерфейс. Спецификация биометрического программного интерфейса».

59 Akhmetov B.S., Doszhanova A.A., Ivanov A.I., A. Malygin, Kartbaev T.S. Dynamic biometric authentication of personality in open information spaces Word Academy of science, engineering and technology (WASET). ISSUE 79 July 2013, Singapore. P. 129 – 138.

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

61 Фунтиков В.А., Назаров И.Г., Бурушкин А.А. Национальные стандарты России: конфиденциальность персональных биометрических данных. «Стандарты и качество» № 7, 2010 г. с. 28-33.

62 Feng Hao, Ross Anderson, and John Daugman. Crypto with Biometrics Effectively, IEEE Transactions on Computers, vol. 55, No. 9, September 2006.

63 Ушмаев О.В., Кузнецов В.В. Алгоритмы защищенной верификации на основе бинарного представления топологии отпечатка пальцев. //Информатика и ее применения. 6(1), 132-140.

Подписано в печать 25.06.2014.
Формат 60x84^{1/16}. Бумага офсетная. Гарнитура «Times».
Объем 9 печ. л. Тираж 500 экз. Заказ № 908.

Отпечатано в ТОО «Издательство LEM»
050008, г. Алматы, ул. Ауэзова, 82, уг. ул. Мынбаева, тел./факс 375-51-33.