

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РЕСПУБЛИКИ КАЗАХСТАН

Казахский национальный технический университет
имени К. И. Сатпаева

Б. С. Ахметов
А. И. Иванов
А. Ю. Малыгин
В. А. Фунтиков

ОСНОВЫ БИОМЕТРИЧЕСКОЙ АУТЕНТИФИКАЦИИ ЛИЧНОСТИ

Рекомендовано Научно-методическим советом университета
в качестве учебного пособия

Алматы 2014

УДК 004(075.8)
ББК 32.973-018.2я73
О-75

Рецензенты:

Джурунтаев Д. З. – д-р. техн. наук, проф. КазНТУ имени К. И. Сатпаева
Тукаев У. А. – д-р. техн. наук, проф. КазНУ имени аль-Фараби
Утепбергенов И. Т. – д-р. техн. наук, проф. АУЭС

О-75

Основы биометрической аутентификации личности: Учеб. пособие / Б. С. Ахметов, А. И. Иванов, А. Ю. Малыгин, В. А. Фунтиков. – Алматы: КазНТУ имени К. И. Сатпаева, 2014. – 151 с. Ил. 71. Табл. 4. Библиогр. – 98 назв.
ISBN 978-601-228-689-2

В пособии рассмотрена история становления и современного развития биометрических технологий. Изложены процедуры получения данных из биометрических параметров личности, достоинства и недостатки используемых статических и динамических биометрических образов.

Проанализированы нечеткие экстракторы биометрических параметров и нейросетевые преобразователи биометрия – код. Приведены особенности обучения искусственных нейронных сетей в соответствии с ГОСТ Р 52633.0–2006. Описываются интерфейс размещения данных обученной нейронной сети в открытом и защищенном нейросетевом контейнере, механизмы мультибиометрии как безопасного объединения нескольких нейросетевых контейнеров.

Рассмотрены вопросы тестирования по ГОСТ Р 52633.3–2011, формирования тестовой базы «Свой» и «Чужие» по ГОСТ Р 52633.1–2009 и генетический алгоритм размножения биометрических образов «Чужие» при увеличении размеров тестовых баз биометрических образов.

Учебное пособие подготовлено авторским коллективом Казахского национального технического университета имени К. И. Сатпаева и Пензенского государственного университета в соответствии с Меморандумом о сотрудничестве.

Может быть полезно бакалаврам, магистрантам и докторантам специальностей «Вычислительная техника и программное обеспечение», «Системы информационной безопасности», «Информационные системы», а также всем специалистам, интересующимся разработкой новых биометрических средств аутентификации личности и защиты информации.

УДК 004(075.8)
ББК 32.973-018.2я73

Печатается по плану издания Министерства образования и науки Республики Казахстан на 2014 год

ISBN 978-601-228-689-2

© Ахметов Б. С., Иванов А. И.,
Малыгин А. Ю., Фунтиков В. А., 2014
© КазНТУ, 2014

ВВЕДЕНИЕ

Защита личности человека и коллективная защита социумов всегда строились на биометрии. Ребенок всегда безошибочно определяет своих родителей. Мать и отец помнят образ своего ребенка и могут отличить его от множества иных детей. Образы матери, отца, сестер, братьев находятся в подсознании человека (в глубинах головного мозга) и позволяют людям иметь относительно безопасное детство.

Повзрослев, мы строим свою коллективную безопасность, с большим доверием относясь к знакомым людям: соседям, родственникам, знакомым, сослуживцам. Биометрия или умение нами распознавать знакомые образы людей начинает выполнять роль механизма обеспечения коллективной безопасности социумов.

Развитие общества происходит стремительно, его основой является обмен информацией между людьми. Первой информационной революцией, видимо, является создание людьми сложных высокоинформативных языков для передачи между людьми информации. Следующей информационной революцией является формализация (оцифровывание) устной речи через письменность. Буквы (знаки, иероглифы) на бумаге или на ином носителе информации – это уже цифровая форма кодирования устной речи человека.

Естественные языки людей, казахский, русский, английский способны описать как угодно сложные объекты, в том числе как угодно сложные биометрические образы того или иного человека. Однако естественные языки обычно дают очень сложное, громоздкое описание человека. Необходимость же в описании людей с целью государственного учета усиливается при усилении роли государства и городов. Государство всегда может восстановить родословную человека, опираясь на людей его знающих, но эта технология эффективна только для небольших поселений. Для учета прав жителей больших

районов и городов необходим паспорт – биометрический документ отражающий возраст, права, статус, биометрию гражданина. Необходимость в паспорте обычно возникает, когда человек из своей семьи из своего места, где его все знают, переезжает в другое место, где его никто не знает.

Учебное пособие посвящено тому, как правильно строить и применять биометрико-криптографическую защиту личной информации человека с использованием его личных криптографических ключей и тайных биометрических образов.

Учебное пособие подготовлено авторским коллективом Казахского национального технического университета имени К. И. Сатпаева и Пензенского государственного университета в соответствии с Меморандумом о сотрудничестве в ходе выполнения научного проекта: № 753. МОН.ГФ.13.13 «Исследование вариантов реализации и разработка действующего лабораторного образца ON-LINE системы биометрического обезличивания электронных историй болезней для медицинского учреждения».

1. ИСТОРИЯ ВОЗНИКНОВЕНИЯ И РАЗВИТИЯ БИОМЕТРИЧЕСКИХ ТЕХНОЛОГИЙ

Паспорта (удостоверения личности) начали массово выдаваться государствами в конце XIX и начале XX века. Современные паспорта могут содержать биометрические данные человека в виде образца его подписи, фотографии и иных биометрических параметров, таких как: пол, цвет волос, цвет глаз, рост и т. д. Бумажный паспорт как документ служит основанием для выполнения должностными лицами процедуры органолептической биометрической идентификации личности человека. Проверяющий должен прежде всего убедиться в подлинности бланка паспорта на бумажном носителе. Далее проверяющий обязан органолептически (с помощью своего зрения) сравнить фотографию в паспорте с лицом проверяемого. Эта процедура уже является достаточно сложной. Проблема состоит в том, что лицо человека меняется с возрастом, кроме того, фотография плоская, а лицо объемное. По этим причинам проверяющий может допускать ошибки, причиной которых является ограниченный объем биометрических данных о человеке в паспорте. Снять возникшую неопределенность можно несколькими способами. Во-первых, проверяющий может попросить проверяемого расписаться и сравнить автограф в паспорте с только что воспроизведенным автографом. Во вторых, проверяющий может опереться на показания нескольких людей, лично знающих проверяемого.

В целом органолептическая паспортная проверка личности является достаточно сложной процедурой, сопряженной со значительными вероятностями ошибок. Кроме того, наличие большого числа паспортов, выдаваемых более чем 100 различными государствами, делает проверку подлинности таких паспортов второй значительной проблемой. Все это привело к снижению доверия к международным паспортам в начале XXI века. События в Нью-Йорке 11 сентября 2001 г. обострили недоверие США к международным паспортам на бумажных носителях. США как явный информационный лидер еще в конце XX в. создает для себя ряд из 50 национальных биометрических стандартов, которые ориентированы на

дублирование органолептической биометрии автоматизированным биометрическим контролем личности.

После событий 11 сентября США выступают с инициативой создания нового поколения биометрических паспортов с радиочитаемой микросхеме, где должны находиться биометрические данные проверяемого человека. При этом предполагается, что подлинность биометрических данных в микросхеме паспорта гарантируются совпадением их номера с номером в паспорте и электронной цифровой подписью под данными. На данный момент считается, что подделать электронную цифровую подпись намного сложнее, чем создать фальшивый бланк паспорта с присутствующими в нем элементами защиты. По крайней мере, злоумышленник наладивший выпуск бланков поддельных паспортов, может рассчитывать на тиражирование своей атаки. Введение криптографической защиты, дополняющей полиграфическую защиту, лишает у злоумышленника возможность тиражирования поддельных паспортов. Взламывать электронную цифровую подпись под электронным содержанием микросхемы придется каждый раз заново. В этом отношении криптографическая защита целостности биометрических данных в микросхеме паспорта и других персональных данных значительно усиливает доверие к международным биометрическим паспортам нового поколения.

Криптографическая поддержка биометрических технологий является важнейшим моментом, обеспечивающим доверие к биометрическим паспортным данным или иным биометрическим данным. Форма, в которой осуществляется криптографическая поддержка биометрических технологий, может быть разной, однако без криптографической поддержки современная биометрия обойтись не может. Все это становится особенно очевидным, когда речь идет о применении биометрии в Интернете и иных открытых информационных пространствах. В открытых информационных пространствах криптография является единственным эффективным способом защиты, т. е. необходимо осуществлять эффективное связывание биометрии с криптографией. Например, это может быть сделано путем связывания открытого биометрического образа с открытым криптографическим ключом и тайного биометрического образа

человека с его тайным (личным) ключом. На рис. 1.1 показан нейросетевой вариант связывания открытого ключа асимметричной криптографии с геометрией лица человека и нейросетевой вариант связывания рукописного пароля человека с его личным криптографическим ключом.



Рис. 1.1. Структура открытого и закрытого электронного удостоверения, предназначенных для автоматов, осуществляющих биометрическую авторизацию личности пользователя

Если охватить и тот и другой варианты связывания биометрии с ключами электронной подписью, то получим два варианта открытого и закрытого электронного удостоверения личности, которыми можно пользоваться в Интернете и иных открытых информационных пространствах. Задача создания для среды Интернет безопасных биометрических паспортов и иных удостоверений личности стоит весьма и весьма актуально.

Всю существующую сегодня биометрию можно условно разделить на две ветви. Первая ветвь – это ветвь применения открытых биометрических образов человека. Как правило, это полицейская биометрия паспортно-визовых документов и баз биометрических данных о преступниках. Применяя открытый для всех биометрический образ человека, нельзя обеспечить человеку анонимность.

Вторая ветвь биометрии строится на тайне биометрического образа. При этом, как обеспечивается тайна биометрии, не имеет принципиального значения. Важно то, что биометрия тайных

образов человека оказывается намного сильнее обычной биометрии открытых образов. Происходит это из-за того, что злоумышленник не знает как атаковать биометрию тайных образов. Он вынужден создавать огромные базы из случайных биометрических образов, принадлежащих разным людям и далее пытаться их предъявлять биометрическому программному роботу, хранящему личный ключ своего хозяина. Открытые биометрические образы – это просто биометрия, тайные биометрические образы – это биометрия, умноженная на «тайну» рукописного пароля, голосового пароля, отпечатка пальца неизвестного лица.

Данное учебное пособие как раз и посвящено тому, как правильно строить и применять биометрико-криптографическую защиту личной информации человека с использованием его личных криптографических ключей и тайных биометрических образов.

1.1. Основные понятия и определения

1.1.1. Вероятности ошибок первого и второго рода

Всякая биометрическая защита строится на том, что она способна хорошо распознавать образ «Свой» и надежно выделять множество образов «Чужие» («Все Чужие»). Очевидно, что средство биометрической защиты (биометрической аутентификации) может ошибаться. Основной задачей (задачей № 1) для биометрии является обеспечение доступа донору биометрического образа «Свой». Ошибка при выполнении этой задачи рассматривается как ошибка первого рода. Основной характеристикой эффективности работы средства биометрической аутентификации является вероятность появления ошибок первого рода P_1 .

Второй задачей средства биометрической аутентификации является препятствовать доступу донору образа «Чужой». Второй важнейшей характеристикой биометрических средств является вероятность появления ошибок второго рода P_2 из-за возможных коллизий образов «Свой» и «Чужой» на

рассматриваемом множестве признаков (биометрических параметров).

Очевидно, что вероятность ошибок второго рода P_2 будет тем меньше, чем больше биометрических параметров принимает в расчет то или иное средство биометрической аутентификации. Высоконадежными можно считать только те биометрические средства, которые анализируют сотни или даже тысячи биометрических параметров. При этом атакующий не должен знать подбираемого биометрического образа, только в этом случае биометрия может рассматриваться как высоконадежная.

На сегодняшний день лучшие средства высоконадежной биометрической аутентификации обеспечивают вероятность ошибок второго рода на уровне одной миллиардной и меньше, т. е. злоумышленник, пытающийся преодолеть биометрическую защиту, должен предъявить миллиард разных биометрических образов (например, воспроизвести своей рукой миллиард рукописных паролей). Если на воспроизведение одного рукописного пароля уходит 10 секунд, то злоумышленнику потребуется 10 млрд секунд, что составит 321 год непрерывных усилий. Это много больше времени жизни одного человека. В одиночку простой подстановкой реальных биометрических образов нельзя преодолеть высоконадежную защиту тайным биометрическим образом.

1.1.2. Преобразование биометрических образов в код ключа доступа

Для того чтобы воспользоваться биометрической защитой, необходимо уметь преобразовывать многомерный континуум биометрических данных образа «Свой» в код криптографического ключа доступа. Для этого нужно отсканировать предъявленный биометрический образ и вычислить контролируемые биометрические параметры. Далее следует предъявить биометрические параметры на вход преобразователя биометрия-код, как это показано на рис. 1.2.

Сам преобразователь биометрии в код может быть выполнен разными технологиями, например, могут быть использованы так называемые «нечеткие экстракторы» и [1–3].

По пути создания «нечетких экстракторов» идут в основном англоязычные исследователи. В русскоязычной литературе публикаций по «нечетким экстракторам» мало [4, 5].

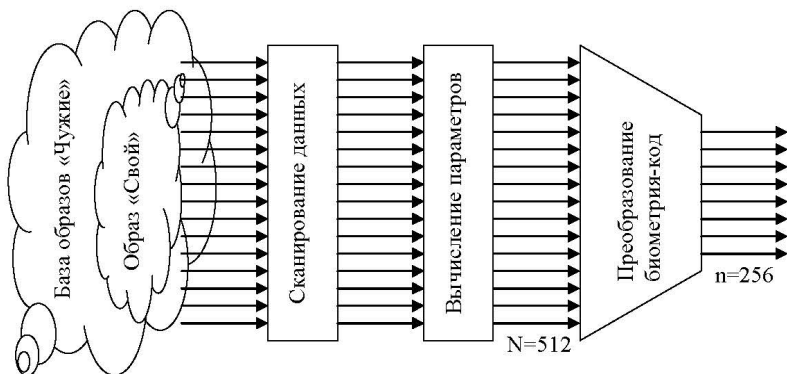


Рис. 1.2. Преобразование биометрических параметров в код ключа

В России и Казахстане преобразователи биометрия-код строятся с использованием искусственных нейронных сетей [6, 7]. Необходимо заранее выбрать структуру связей искусственных нейронов сети, а затем обучить ее [8–12] так, чтобы примеры биометрического образа «Свой» давали на выходе преобразователя код личного криптографического ключа гражданина России или Казахстана, а для примеров образов «Чужой» на выходе преобразователя получалась случайная кодовая комбинация.

Независимо от того, какая технология используется, преобразователь биометрия-код всегда имеет больше входных биометрических данных, чем число его выходов (чем длина выходного криптографического ключа). Так, на рис. 1.2 преобразователь имеет $N = 512$ входных биометрических параметров и только $n = 256^1$ выходных разрядов криптографического ключа. Условие $N > n$ всегда выполняется

¹ По российским стандартам на шифрование и формирование электронной цифровой подписи длина криптографического ключа составляет 256 бит.

из-за низкой информативности биометрических данных. Одного биометрического параметра, как правило, не хватает для получения одного бита криптографического ключа. Обязательно необходимо использовать избыточное число биометрических параметров для того, чтобы исправлять ошибки кодирования в «нечетких экстракторах» или «обогащать» входные биометрические данные нейронной сетью преобразователя.

*1.1.3. Представление многомерных континуумов
биометрических образов малым дискретным числом
примеров*

Одной из основных причин трудности биометрической аутентификации является высокая размерность задачи (приходится учитывать 512 и более «плохих» биометрических параметров). Именно по этой причине не удастся воспользоваться аппаратом классической линейной алгебры и многомерной статистики. Положение усложняется тем, что преобразователи биометрия-код приходится обучать (настраивать) на малом числе примеров в обучающей выборке.

Так, при обучении биометрических средств аутентификации пользователи готовы предъявить от 10 до 20 примеров биометрического образа «Свой». Однако если их попросить предъявить 100 или 200 примеров, то эта работа воспринимается пользователями негативно. На сегодня пользователи не готовы прилагать значительные усилия для обучения своего биометрического программного робота (преобразователя биометрия-код). Последнее означает, что 512-мерные распределения континуумов параметров образа «Свой» мы вынуждены представлять всего 20 примерами по каждому из параметров. Мы можем попытаться представить 512-мерный объем естественно-континуальной формы представления данного образа «Свой» через внутренний объем графа 20 примеров «Свой», однако это будет весьма и весьма слабым приближением. Эта ситуация отображена на рис. 1.3.

Пользуясь всего 20 примерами, невозможно точно вычислить математическое ожидание биометрических параметров $E(v_i)$, их

среднеквадратическое отклонение $\sigma(v_i)$ и коэффициенты корреляции между параметрами r_{ij} . На обучающей выборке из 20 примеров относительная ошибка в оценке математического ожидания может быть от 0 до $\pm 25\%$, относительная ошибка среднеквадратического отклонения может составлять от 0 до $\pm 50\%$, относительная ошибка оценки коэффициентов корреляции может быть от 0 до $\pm 100\%$. При столь неточных данных невозможно построить многомерную аналитическую модель распределения данных образа «Свой».

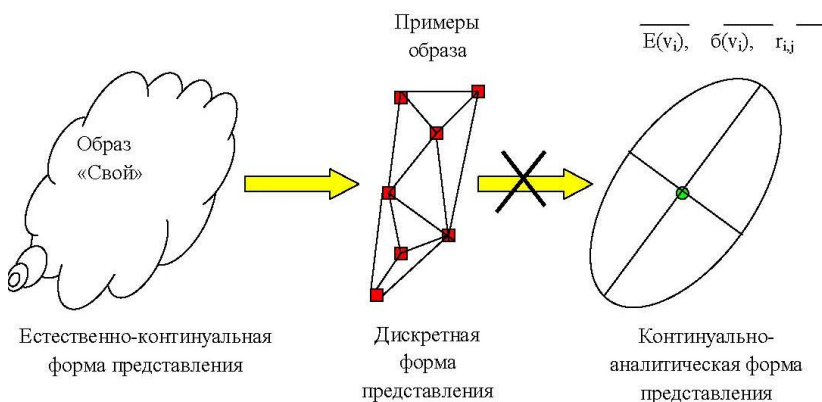


Рис. 1.3. Представление многомерного образа «Свой» малым числом примеров (7 примеров)

Даже, если придется работать в рамках гипотезы нормального нормированного распределения значений биометрических параметров, необходимо построить для них 512-мерную корреляционную матрицу. Придется обращаться 512-мерную матрицу² для того, чтобы вычислять поверхность гиперэллипса, разделяющую образ «Свой» от образов «Чужие», так как построить аналитическое описание разделяющей поверхности в

² Поверхность гиперэллипсов описывается квадратичными формами линейной алгебры, куда входит обратная ковариационная (корреляционная) матрица высокой размерности.

форме гиперэллипсов технически невозможно (правая стрелка рис. 1.3 зачеркнута).

Описанная выше ситуация получила название «проклятия размерности». Казалось бы, что учет дополнительных биометрических параметров должен приводить к более точным результатам, однако все происходит с точностью до наоборот. При попытках увеличивать размерность решаемой задачи она решается все хуже и хуже. Возникает эффект плохой обусловленности или накопления погрешностей вычислений. Именно из-за «проклятия» размерности в биометрии не удается пользоваться линейной алгеброй и классической многомерной статистикой. Однако применение искусственных нейронных сетей и «нечетких экстракторов» позволяет ослабить проблему «проклятия» размерности или вообще снять ее.

Как будет показано позднее, 20 примеров образа «Свой» вполне достаточно для качественного автоматического обучения большой искусственной нейронной сети с 512 входами и 256 выходами, если пользоваться алгоритмом автоматического обучения [8]. На данный момент пока только этот алгоритм справляется с шумами дискретизации биометрических данных, обусловленных тем, что естественная многомерно-континуальная форма представления образа «Свой» замещается приближением всего из 20 примеров. Всегда, когда непрерывная функция (континуум) представляется малым числом состояний (примеров), возникают очень большие погрешности дискретизации, которые способны накапливаться при вычислениях (при обучении нейронных сетей). Алгоритм обучения [8] является первым, но не единственным абсолютно устойчивым алгоритмом, который строится на усреднении многомерных данных (их интегрировании в многомерном пространстве³). Как следствие он работает тем лучше, чем выше размерность решаемой задачи.

³ Мы привыкли к интегрированию 512 отсчетов измерений во времени и согласны с тем, что операция интегрирования повышает точность вычислений. Для многомерных систем в место интегрирования по времени можно усреднять данные по 512 разным пространствам, это не менее эффективно давит естественные ошибки вычислений и делает их устойчивыми.

Возникает эффект, обратный эффекту «проклятия» размерности, который можно назвать «благодатью» высоких и сверхвысоких размерностей.

1.1.4. Переход в пространство расстояний Хэмминга

Входные биометрические данные, как правило, континуальны (непрерывны), а выходные коды преобразователя дискретны. Число возможных состояний кода всегда конечно. Любой преобразователь биометрия-код это одно из возможных отображений многомерных континуумов в конечное число состояний кода заданной длины. Соответственно для исследования его выходных кодов может быть использовано расстояние Хэмминга. Для вычисления расстояний Хэмминга для двух кодов одинаковой длины их сравнивают поразрядно, далее подсчитывают число не совпавших разрядов. Если представить произвольный выходной код в виде вектора $\bar{x}$ с множеством состояний разрядов «0» и «1», то расстояние Хэмминга до кода «Свой» $\bar{c}$ вычисляется следующим образом:

$$h = \sum_{i=1}^{256} x_i \oplus c_i, \quad (1.1)$$

где 256 – длина сравниваемых кодов;

i – номер сравниваемых разрядов;

$\oplus$ – операция сложения по модулю два.

Расстояние Хэмминга между выходными кодами «Чужие» и выходным кодом «Свой» измеряется в битах и оказалось очень удобной метрикой для статистических исследований [11–17]. Применение этой метрики к «нечетким экстракторам» и нейросетевым преобразователям биометрия-код дает разные распределения расстояний для образов «Свой» и «Все Чужие». «Нечеткие экстракторы» осуществляют квантование биометрических параметров образ до их обогащения. По этой причине коды образа «Свой» ведут себя нестабильно и имеют достаточно широкое распределение значений расстояний Хэмминга (смотри верхнюю часть рис. 1.4).

Из-за того, что нейросетевые преобразователи способны улучшать качество биометрических данных образа «Свой», распределение Хэмминга для кодов «Свой» сжимается (прижимается к «нулевой» оси системы координат в нижней части рис. 1.4).

Из верхней части рис. 1.4 видно, что, контролируя выходные коды «нечеткого экстрактора», необходимо принимать как коды «Свой» множество кодов с расстоянием Хэмминга до действительного кода «Свой» менее 40 бит. На выходе нейросетевого преобразователя биометрия-код коды «Свой» имеют много меньшую нестабильность до 7 бит.

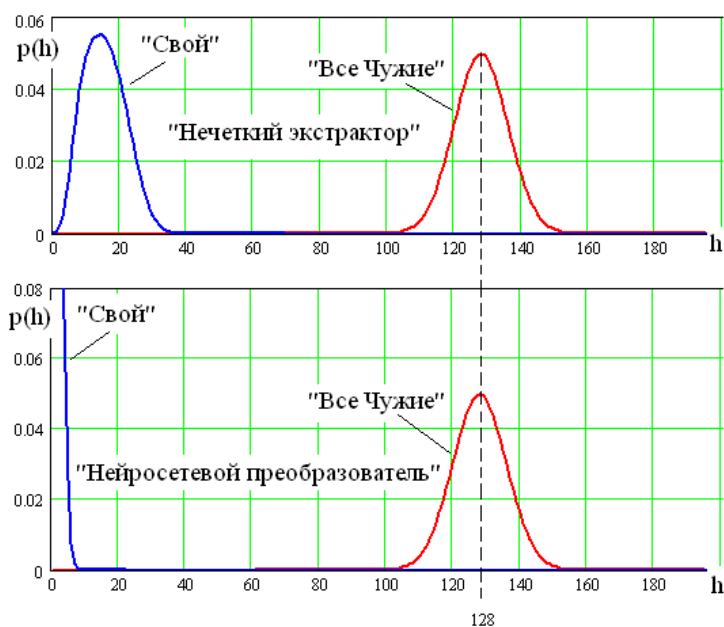


Рис. 1.4. Распределение расстояний Хэмминга для «нечетких экстракторов» нейросетевых преобразователей для кодов длины 256 бит

Важной особенностью распределения Хэмминга кодов «Все Чужие» является то, что оно очень хорошо описывается нормальным законом распределения значений. По этой причине

можно вычислить математическое ожидание распределения данных распределения $p(h(x))$ и его среднеквадратическое отклонение. По этим двум статистическим моментам удастся рассчитать ожидаемую вероятность ошибок второго рода как для нечетких экстракторов, так и для нейросетевых преобразователей:

$$P_2 \approx \frac{1}{\sigma(h(x)) \cdot \sqrt{2\pi}} \int_0^{\max(h(c))} \exp\left\{-\frac{(E(h(x)) - u)^2}{2 \cdot \sigma^2(h(x))}\right\} \cdot du, \quad (1.2)$$

где $\max(h(c))$ – максимально возможное значение расстояния Хэмминга кодов «Свой»;

$E(h(x))$ – математическое ожидание расстояний Хэмминга кодов «Чужие»;

$\sigma(h(x))$ – среднеквадратическое отклонение расстояний кодов «Чужие».

1.1.5. Информативность биометрических параметров

Выше по тексту постоянно шли упоминания о «хороших» и «плохих» биометрических параметрах. Для того, чтобы перейти от неточных языковых понятий «хороший» и «плохой», следует перейти к более точному показателю качества биометрического параметра – информативности. Для этой цели упростим ситуацию и будем рассматривать простейшее правило принятия решения «Свой» – «1» и «Чужой» – «0». При этом пороги квантования «Свой» разнесем на расстояние $\pm 3 \sigma(v)$ от математического ожидания $E(v)$ распределения значений «Свой». В этом случае вероятностью ошибок первого рода можно пренебречь, а вот вероятность ошибок второго рода всегда будет достаточно велика:

$$P_2(v) \approx \frac{1}{\sigma(\xi) \sqrt{2\pi}} \int_{\min(v)}^{\max(v)} \exp\left\{-\frac{(E(\xi) - u)^2}{2\sigma^2(\xi)}\right\} du, \quad (1.3)$$

где $E(\xi)$ – математическое ожидание биометрического параметра ξ образов «Все Чужие»;

$\sigma(\xi)$ – среднеквадратическое отклонение параметра ξ образов «Все Чужие»;

$\min(v)$ – минимальное значение параметра – v образа «Свой»;

$\max(v)$ – максимальное значение параметра – v образа «Свой»;

Информативность биометрического параметра – v оказывается тем выше, чем меньше вероятность ошибок второго рода при нулевой вероятности ошибок первого рода:

$$I(v) = -\log_2(P_2(v)). \quad (1.4)$$

Очевидно, что информативность биометрического параметра будет расти по мере стабилизации биометрического параметра (сжатия распределения «Свой»). Кроме того, на информативность влияет положение распределения данных образа «Свой» по отношению к распределению образов «Все Чужие». Эта ситуация иллюстрируется рис. 1.5.

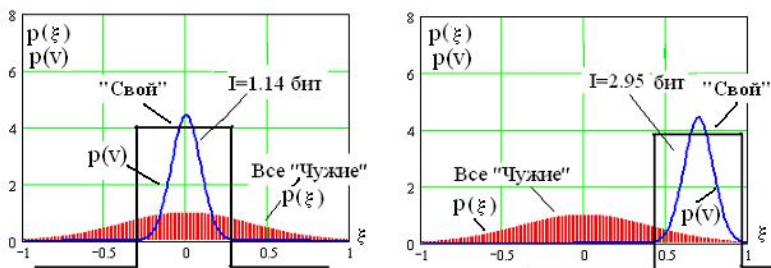


Рис. 1.5. Влияние положения распределения образа «Свой» $p(v)$ на информативность параметра v

Из рис. 1.5 видно, что смещение одного и того же распределения данных «Свой» на периферию распределения данных образа «Все Чужие» приводит к значительному росту информативности параметра. В первом приближении можно считать биометрический параметр «хорошим», если его информативность больше 1 бита. Соответственно «удовлетворительными» следует считать биометрические параметры с информативностью менее 1 бита.

Что считать «плохими» биометрическими параметрами сильно зависит от используемой технологии обработки данных. Для нечетких экстракторов «плохими» будут являться биометрические параметры с информативностью менее 0.5 бита. Для искусственных нейронных сетей «плохими» являются биометрические параметры, информативность которых менее 0.05 бита. Эффективность нейросетевой обработки биометрических данных как раз и обусловлена тем, что она позволяет использовать низко информативные данные, которые в принципе не могут обрабатываться нечеткими экстракторами.

1.1.6. Оценка энтропии преобразователя биометрия-код

Важнейшей характеристикой любого преобразователя биометрия-код является энтропия его выходных состояний, вычисленная при условии воздействия на преобразователь биометрия-код образами «Все Чужие». Важность этого показателя легко продемонстрировать на примере шифрования текстов. Текст на русском или казахском языках легко читается носителями этих языков, если он не зашифрован. Разряды кодов открытого естественного языка имеют значительные коэффициенты корреляции. Наличие высоких значений коэффициентов парной корреляции между разрядами кодов – это признак всех биометрических кодов, в том числе кодов коллективной биометрии русскоговорящих и казахов. Шифрование разрушает естественные корреляционные связи личных биометрических кодов и кодов парольных фраз, набранных на компьютере. После шифрования коды становятся «белым» шумом, состояния «0» и «1» в их разрядах становятся независимы, парные корреляции отсутствуют ($r_{ij} = 0.0$ для всех $i \neq j$).

Еще одной важной характеристикой «белого» шума является то, что энтропия каждого разряда таких кодов точно совпадает с одним битом, т. е. длина действительно случайного кода, представляющего «белый» шум, совпадает со значением его энтропии. Энтропия случайного кода длиной 256 бит точно равна 256 битам. Если же код оказывается не случайным, то его энтропия падает:

$$H(256) \langle 256 \text{ при } |r_{i,j}| \rangle > 0.0 \text{ хотя бы для одной пары } i \neq j. \quad (1.5)$$

Для всех кодов (с коррелированными и независимыми разрядами) энтропия связана с показателем стойкости к атакам подбора или с вероятностью ошибок второго рода. Если речь идет о симметричном шифровании на ключе длиной 256 бит, то вероятность расшифровывания текста с первой попытки случайной подстановки ключа составит

$$P_2 = 2^{256} \text{ или } H(256) = 256 = -\log_2(P_2). \quad (1.6)$$

В общем виде можно записать:

$$H(n) = -\log_2(P_2), \quad (1.7)$$

где n – длина биометрического кода;

P_2 – вероятность ошибок второго рода преобразователя биометрия-код.

Заметим, что ГОСТ Р 52633.0–2006 [6] накладывает ограничение на корреляционные связи в разрядах биометрических кодов «Чужой». Среднее значение модуля этих корреляционных связей не должно превышать 0.15. В случае если $E(|r_{i,j}|) = 0.15$, показатель вероятности ошибок второго рода такого преобразователя падает примерно на порядок $P_2 = 2^{25.6}$. В сравнении с шифрованием происходит потеря стойкости к атакам подбора на $2^{31.4}$ порядка. При больших значениях модулей коэффициентов парной корреляции происходит катастрофическое падение стойкости преобразователя биометрия-код к атакам подбора⁴. Если кому-то удастся сделать преобразователь биометрия-код, дающий 256-разрядный код с отсутствием в коде парных корреляций между разрядами, то такой преобразователь будет обладать стойкостью к атакам подбора, совпадающей со стойкостью защиты шифрованием с такой же длиной ключа.

⁴ Стойкость преобразователя биометрия-код обратно пропорциональна его вероятности ошибок второго рода.

Еще одним важным моментом является то, как следует вычислять энтропию выходных биометрических кодов. Пытаться использовать формулу Шеннона нельзя:

$$H(256) = - \sum_{i=1}^{2^{256}} P_i \cdot \log_2(P_i), \quad (1.8),$$

где P_i – вероятность появления i -го состояния био-кода.

Выполнить оценку энтропии по формуле (1.8) для преобразователей биометрия-код с 256 выходами (число состояний 2^{256}) технически невозможно. Алфавит возможных выходных состояний биометрических кодов слишком велик. В связи с этой проблемой вычислить энтропию можно только пользуясь формулой (1.7), предварительно оценив вероятность ошибок второго рода по формуле (2) в пространстве расстояний Хэмминга.

1.1.7. Оценка размерности решаемой биометрической задачи

Возможности интеллекта (искусственного и естественного) определяются тем, какой сложности (какой размерности) он может решать задачи. С точки зрения обычной математики, любую задачу можно записать как некоторую многомерную функцию $F(x_1, x_2, x_3, \dots, x_n)$, где n – число учитываемых переменных. Очевидно, что размерность задачи и ее сложность можно оценивать в штуках по числу переменных. Это весьма и весьма приземленный, но очень эффективный подход.

Если переменные сильно зависимы, то они создают не более чем видимость высокой размерности. Измерение числа входных переменных в штуках и числа выходных состояний преобразователя биометрия-код также в штуках почти ничего не говорит о размерности решаемой задачи. Действительная размерность решаемой задачи преобразования биометрия-код связана исключительно со значением выходной энтропии кодовых состояний преобразователя (1.8) и измеряется в битах.

Вычисляя энтропию выходных кодов по (1.8), мы фактически уменьшаем длину выходного кода до полноценной размерности задачи, рассматривая разницу между длиной кода в штуках и короткого кода в битах как некоторый бесполезный довесок с абсолютно коррелированными состояниями [10, 16–18]. Эта ситуация отображена на рис. 1.6.

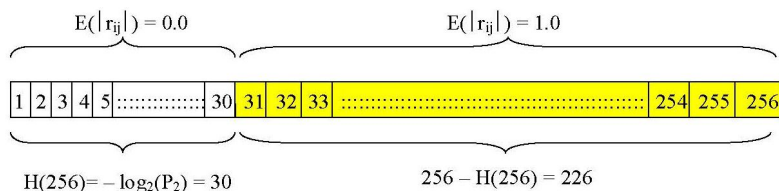


Рис. 1.6. Представление биокода длиной 256 бит со слабо коррелированными разрядами в виде 30 стойких некоррелированных разрядов и 226 бесполезных полностью коррелированных разрядов

При таком методе оценки размерность решаемой биометрической задачи может быть очень высокой даже при относительно малых длинах выходного кода, если его разряды слабо коррелированы. При росте корреляционных связей между разрядами кода размерность решенной задачи биометрической защиты падает. В случае, когда разряды кода полностью коррелированы при любой длине, его энтропия 1 бит, размерность также составляет 1 бит.

Число связей нейронов, слоев нейронов, входов у преобразователя биометрия-код и алгоритм обучения влияют косвенно на оценку размерности решаемой задачи.

Простое увеличение какого-то из перечисленных выше параметров без соответствующего роста энтропии выходных кодов преобразователя не приводит к росту размерности решенной задачи биометрической защиты информации [10, 16–18].

2. АНАЛИЗ СОВРЕМЕННЫХ БИОМЕТРИЧЕСКИХ ТЕХНОЛОГИЙ

2.1. Технология анализа рисунка отпечатка пальца

Самой распространенной в настоящий момент биометрической технологией является технология анализа папиллярных рисунков кожи пальцев (рисунков отпечатков пальцев) [19–22].

Серийно выпускаются и используются значительным спросом оптические сканеры рисунка отпечатка пальцев и емкостные сканеры. Для работы в офисных помещениях обычно применяют оптические сканеры, а для работы вне помещений (в банкоматах) применяются емкостные сканеры (рис. 2.1). Применение емкостных сканеров, например, в банкоматах связано с тем, что их технические условия допускают работу в более широком температурном диапазоне.



Оптический сканер



Емкостной сканер

Рис. 2.1. Оптический и емкостный сканеры рисунков отпечатков пальцев

При анализе папиллярного рисунка отпечатка пальца выделяют два типа признаков: глобальные признаки, локальные признаки. Глобальные признаки (центр папиллярного рисунка, дельта) не являются уникальными, но позволяют классифицировать тип папиллярного узора. Локальные признаки составляют точка окончания линии гребня и точка разветвления линии гребня. Взаимное расположение таких точек является уникальной информацией, по которой можно однозначно идентифицировать человека (рис. 2.2).

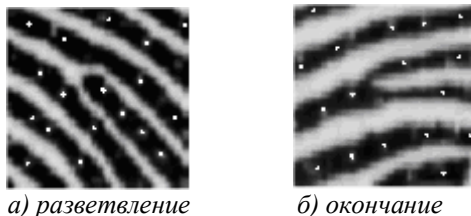
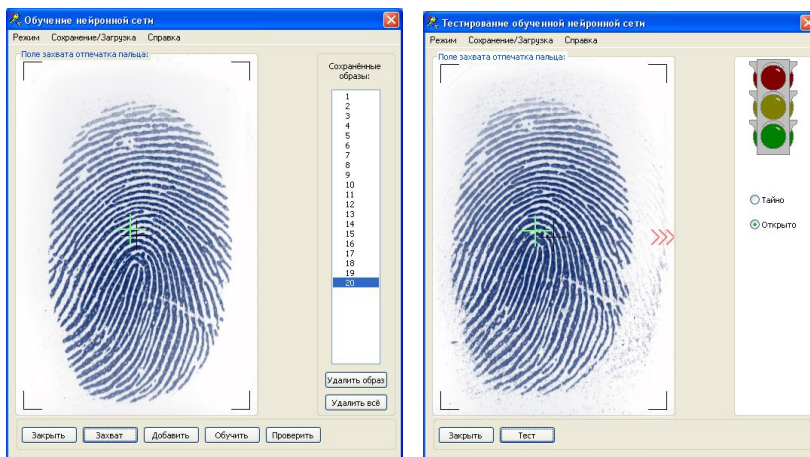


Рис. 2.2. Два типа локальных признаков особых точек на папиллярном рисунке отпечатка пальца (потовые железы видны как белые точки на темных гребнях кожи)

Обычно средство биометрико-нейросетевой аутентификации личности имеет два режима работы (рис. 2.3):

- режим обучения, требующий предъявления нескольких примеров рисунка отпечатка пальца;
- режим аутентификации, требующий предъявления верифицируемого рисунка.



а) обучение нейронной сети б) нейросетевая аутентификация

Рис. 2.3. Режимы работы программы действующего макета № 1

Если средство биометрического контроля рисунка отпечатка пальца построено на основе международного стандарта БиоAPI [23] и соответствует ГОСТ Р ИСО/МЭК 19784-1-2007 [24], а также

ГОСТ Р ИСО/МЭК 19784-2-2010 [25], то в качестве биометрического шаблона должен использоваться список особых точек с их взаимными координатами, сформированный по ГОСТ Р ИСО/МЭК 19794-2 [26]. Персональные биометрические данные человека в виде списка особых точек рисунка отпечатка пальца должны быть надежно защищены по классу 1 ФЗ РФ № 152 «О персональных данных» [27].

Если же эти же персональные биометрические данные будут размещены в обезличенном защищенном нейросетевом контейнере (как это сделано в действующем макете № 1), то особых требований к защите этой персональной биометрической информации ФЗ РФ № 152 не предъявляет.

Вероятность ошибок второго рода (пропуск «Чужого») для биометрических технологий этого типа составляет величины от 10^{-3} до 10^{-6} при использовании рисунка одного пальца. Вероятность ошибок второго рода сильно зависит от информативности рисунка используемого отпечатка пальца. Если особых точек на рисунке отпечатка мало, он обладает низкой информативностью и его стойкость к атакам подбора мала. Следует отметить, что информативность рисунка отпечатка пальца может быть значительно увеличена, если при обработке изображения выделять неравномерность положения самих потовых желез, расположенных на гребнях складок кожи относительно особенностей линий рисунка. Для этой цели необходимо существенно повысить разрешающую способность современных сканеров и скорректировать программное обеспечение предобработки биометрических параметров.

2.2. Технология анализа дерева кровеносных сосудов пальца и ладони

Практика использования сканеров рисунков отпечатков пальцев показала, что примерно для 8 % людей эта технология оказывается малоэффективной [19]. У части людей папиллярный рисунок кожи пальцев плохо выражен. Как правило, это женщины, дети и лица азиатского происхождения. В связи с этим в настоящее время японские фирмы ориентируются на технологию анализа рисунка кровеносных сосудов руки или

пальца. В Японии уже несколько лет работают сети банкоматов, построенные на анализе распределения кровеносных сосудов под кожей пальца на руке или кровеносных сосудов всей ладони. Корпорация «Хитачи» («Hitachi») специализируется на сканерах рисунка кровеносных сосудов под кожей пальца, корпорация «Фуджицу» («Fujitsu») производит сканеры и системы, анализирующие рисунок кровеносных сосудов всей ладони (рис. 2.4).



Рис 2.4. Примеры рисунков кровеносных сосудов пальца и руки

Вероятности ошибок второго рода обоих типов средств биометрической авторизации находится в интервале от 10^{-4} до 10^{-6} . Так как рисунок кровеносных сосудов изменить нельзя, оператор предоставления услуг платежной системы должен обеспечить хранение его биометрических параметров в конфиденциальной форме. Могут быть использованы защищенные биометрические контейнеры или биометрические хэш-функции.

Кроме того, базы этих персональных биометрических данных должны быть обезличены. Если используются обычные биометрические шаблоны, то защита базы биометрических шаблонов должна быть выполнена по классу 1 ФЗ РФ № 152.

Данная биометрическая технология может быть использована в банкоматах и офисных терминалах операторов платежных услуг, используемых для управления личным счетом.

2.3. Технология анализа геометрии ладони

Наряду с анализом кровеносных сосудов руки, банкоматы Японии дополнительно анализируют все иные геометрические параметры кисти руки. У каждого из нас разная длина и ширина пальцев, кроме того, кожа на руке имеет явно выраженные складки и морщины. На рис. 2.5 приведена фотография тыльной стороны руки. На рис. 2.5 приведена фотография тыльной стороны руки.



Рис. 2.5. Фотография тыльной стороны кисти руки, по которой можно оценить ее относительные геометрические параметры

Сегодня существуют два типа биометрических систем, построенных на анализе геометрии кисти руки. Системы первого типа построены на анализе изображений кисти руки или ладони руки, полученных в обычном свете. Системы второго типа строятся

на выделении расположения ветвлений дерева кровеносных сосудов на тыльной стороне кисти руки. Кровь, поступающая к кисти руки, обычно теплее самой руки и для выделения рисунка сосудов крови необходимо использовать оптические элементы, чувствительные к инфракрасной части спектра.

Весьма информативными являются параметры ладони руки, полученные путем выделения на ней характерных складок кожи (так называемых «линий жизни»). Геометрия расположения складок кожи на тыльной стороне кисти руки и на ладони являются весьма и весьма информативными биометрическими параметрами, которые пока не используются должным образом современными техническими средствами биометрического контроля.

Более подробные сведения об извлечении биометрических данных из параметров геометрии руки и их форматировании содержатся в стандарте ГОСТ Р ИСО/МЭК 19794–10–2010 [28].

2.4. Технология анализа радужной оболочки глаза и сосудов глазного дна

У каждого из людей параметры геометрии его тела уникальны, в том числе уникальными являются параметры глаз. На настоящий момент существуют два типа биометрических систем, построенных на измерении параметров глаза. К первому типу относятся системы, построенные на анализе рисунка радужной оболочки глаза. При этом по Даугмену [29] рисунки радужных оболочек глаз не зависят от генетики человека и различаются даже у однояйцовых близнецов. Ко второму типу относятся системы, построенные на анализе рисунка кровеносных сосудов глазного дна (сетчатки). Естественно, что оптические характеристики у первых и вторых систем оказываются принципиально разными. Примеры рисунков сетчатки глаза разных людей и одного рисунка кровеносных сосудов глазного дна приведены на рис. 2.6.

Анализ дерева кровеносных сосудов осуществляется на тех же принципах, что и анализ дерева кровеносных сосудов кисти руки или иных участков тела человека. При формировании шаблона находятся корни, стволы, ветви, далее они классифицируются по их длине, взаимному расположению и

диаметру. В правой нижней части рис. 2.6 хорошо видно, что все кровеносные сосуды имеют ярко выраженный общий корень. Этот корень должен рассматриваться как центр, относительно которого ведутся биометрические измерения геометрии стволов и ветвей разного уровня.

При анализе рисунка радужной оболочки глаза осуществляют ее сканирование по параллельным концентрическим кругам, расположенным вокруг зрачка. При этом учитывается, что зрачок глаза человека может менять свои размеры, адаптируясь к освещенности рассматриваемого предмета.

В связи с этим формирование биометрического шаблона радужной оболочки глаза и идентификацию человека по его рисунку целесообразно выполнять при одинаковой освещенности. В этом случае соотношение диаметра зрачка и внешнего диаметра радужной оболочки оказывается практически постоянным и не нуждается в пересчете.

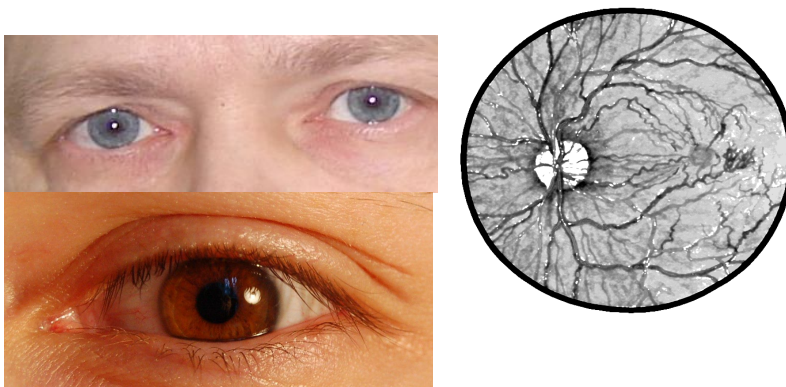


Рис. 2.6. Примеры рисунков радужной оболочки глаза разных людей и пример одного рисунка взаимного расположения кровеносных сосудов глазного дна

При сканировании радужной оболочки оценивают значение среднего уровня яркости изображения вдоль заранее заданной концентрической окружности. Далее концентрическая окружность делится на заданное число секторов, например 128 секторов, и вычисляется средняя яркость каждого сектора. Высокая яркость

и низкая яркость в секторе по отношению к средней яркости дают состояние «1» и «0» кода сканирования. Если использовать 8 концентрических окружностей сканирования, то, соответственно, конечный код сканирования будет иметь длину кода $128 \times 8 = 1024$ бита. Пример подобного кода приведен на рис. 2.7.

Следует отметить, что итоговая кодовая комбинация перепадов яркости радужной оболочки будет иметь сильно коррелированные (сильно зависимые) соседние отсчеты. Впрочем, значительная корреляция характерна для многих биометрических параметров и нужны специальные декорреляционные преобразования, которые позволяют сделать измеряемые данные достаточно независимыми.

Офтальмологические биометрические системы могут строиться не только на анализе рисунка радужной оболочки глаза и анализе рисунка кровеносных сосудов глазного дна. При хорошем разрешении используемой оптики на фотографиях глаза хорошо различимы кровеносные сосуды, снабжающие глазное яблоко кровью. Их рисунок уникален и вполне может быть использован для идентификации и аутентификации человека. Однако на сегодня подобных систем на биометрическом рынке пока нет.

Процедуры обработки офтальмологических биометрических данных радужной оболочки описаны в стандарте ГОСТ Р ИСО\МЕК 19794–6 «Автоматическая идентификация. Идентификация биометрическая. Форматы обмена биометрическими данными. Данные изображения радужной оболочки глаза».

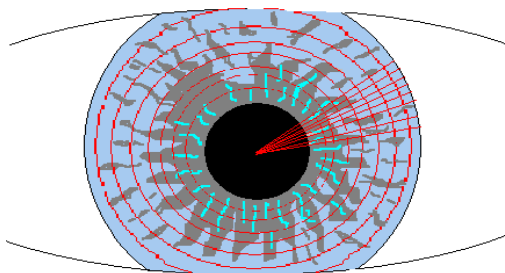


Рис. 2.7. Использование несколько концентрических окружностей при сканировании перепадов яркости радужной оболочки глаза

Видимо, наиболее приемлемой технология анализа радужной оболочки глаза является для авторизации пользователей банкоматами. Для банкоматов желательно использовать бесконтактный способ сканирования биометрических данных человека. На данный момент существуют технические средства видео- наблюдения, имеющие достаточно высокое разрешение при сканировании рисунка радужной оболочки глаза на расстоянии от 30 до 50 сантиметров.

Лидерами создания биометрических средств анализа радужной оболочки глаза являются англичане. Эта технология защищена патентами, срок которых в ближайшее время заканчивается. Действие патентов, защищающих эти биометрические технологии, на территорию России, Белоруссии и Казахстана не распространяется.

По данным англоязычных источников [29] вероятность ошибок второго рода при анализе рисунка радужной оболочки глаза составляет от 10^{-5} до 10^{-11} . Эти цифры явно завышены, так как ошибки оценивались в рамках гипотезы независимости разрядов кода, порождаемого рисунками радужной оболочки глаза. Для обеспечения свойства независимости отбрасывались разряды кода, имеющие корреляцию более 0.3.

Предположительно, учет остаточных коэффициентов корреляции кодов радужной оболочки глаза по методу четвертого раздела ГОСТ Р 52633.3–2011 должен привести к существенному снижению оценки вероятности ошибок второго рода.

Так как рисунок радужной оболочки глаза человека является статическим биометрическим образом, его эталонные параметры должны храниться в защищенной форме. В виде обычной хэш-функции кода или в виде биометрической хэш-функции, настроенной учитывать параметры рисунка радужной оболочки глаза конкретного человека [30]. Базы эталонных данных рисунков радужной оболочки глаза должны быть обезличены.

2.5. Двухмерная и трехмерная технологии анализа геометрии лица человека

Обычно люди узнают друг друга по особенностям черт лица. Удостоверения личности, паспорта, водительские права,

как правило, имеют фотографию их владельца. Попытки автоматизировать процедуру распознавания человека по особенностям черт лица [31–33] привели к появлению двух классов средств биометрической идентификации. Первый класс подобных средств анализирует плоские изображения лица человека (2D-портреты). Работа этих устройств иллюстрируется на рис. 2.8.

Двухмерные анализаторы плоских изображений находят на портрете человека «анфас» характерные точки (рис. 2.8) и расстояния:

- между центрами глаз;
- между линией глаз и кончиком носа;
- соответствующего наибольшей ширине носа;
- между дугами бровей и нижней точкой подбородка;
- между нижними точками мочек ушей;
- между кончиком носа и нижней точкой губ.



Рис 2.8. Выделение особых точек на плоских фотографиях человека с последующим вычислением расстояний между ними

При анализе двухмерных изображений портрета человека в профиль используются аналогичные хорошо выделяемые точки, при этом хорошо выделяемые точки на контуре лица могут быть дополнены расстояниями до хорошо выделяемых точек на

ушной раковине человека. Более подробная информация об обработке данных и их форматах содержится в ГОСТ Р ИСО/МЕК 19794–5 «Автоматическая идентификация. Идентификация биометрическая. Форматы обмена биометрическими данными. Данные изображения лица».

Двухмерные анализаторы взаимного расположения черт лица позволяют получить достаточно скудную биометрическую информацию, обеспечивая вероятности ошибок второго рода (пропуск «Чужого») на уровне 0.05. Переход к более сложному трехмерному анализу геометрии лица позволяет значительно увеличить объем получаемой биометрической информации. Для получения трехмерных изображений могут быть использованы две фотокамеры [34, 35] или одна фотокамера со специальной подсветкой. Например, это может быть подсветка системой продольных и поперечных полос света, образующих равномерную сетку при освещении плоскости. При такой подсветке на лице идентифицируемого человека получается сетка, ячейки которой деформированы пропорционально расстоянию от подсвечиваемых точек до фотокамеры. По значениям деформации ячеек сетки подсветки может быть восстановлен трехмерный биометрический образ поверхности кожи лица человека.

В России 3D-геометрией занимается компания ООО «Артек групп» (www.artec3D.com) г. Москва), которая специализируется на поставке относительно недорогих ручных сканеров для быстрого создания трехмерных моделей различных предметов. Трехмерное сканирование неподвижного предмета проводится при вращении вокруг него камеры и непрерывной съемкой со вспышкой, далее используется программное обеспечение, восстанавливающее по соседним снимкам трехмерное изображение предмета. У этой компании есть также вариант решения задачи 3D-сканирования лица человека по данным неподвижной камеры при подсветке лица человека сеткой лазерного излучения в невидимом для человека инфракрасном диапазоне.

Компания ООО «Наука Техника Безопасность» предлагает 3D-сканер человеческого лица с неподвижной камерой, расположенной на расстоянии от 60 до 120 см от исследуемого лица. Используется инфракрасная сетка подсветки, восстановление

3D-изображения ведется по технологии A4Vision, биометрический шаблон имеет размеры около 5 Кбайт.

По данным «Пензенского государственного университета», для систем 3D-зрения может быть исключена задача фокусировки двух камер за счет использования плоской лазерной подсветки лица человека. Предложенная в описании к патенту [36] измерительная схема сканирования представлена на рис. 2.9.

На рис. 2.9 обозначены C_1 , C_2 – HD-камеры; L – источник структурированной подсветки (шаблона); d – фокусное расстояние камер; r – расстояние от начала координат до камер; P – произвольная точка на объекте, принадлежащая плоскости структурированной подсветки.

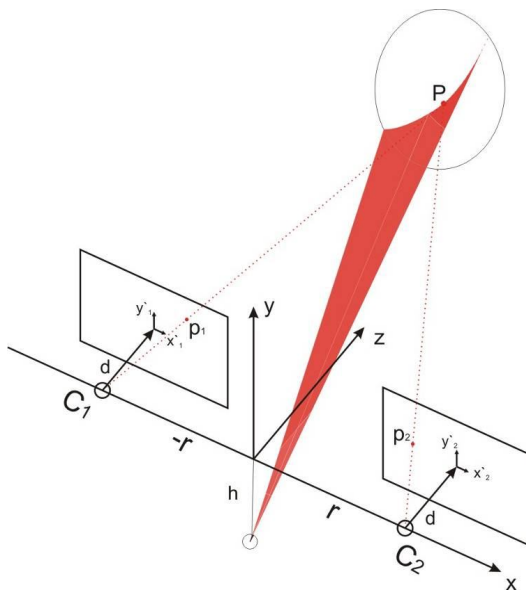


Рис. 2.9. Измерительная схема сканирования, исключающая необходимость автоматической фокусировки двух камер наблюдения

Преимуществом данного способа является отсутствие оптической системы с устройством автоматической фокусировки на объект, которая существенно усложняет конструкцию и

удорожает стоимость изделия; низкая стоимость и широкая распространенность источника структурированной подсветки, простота конструкции сканирующего модуля.

Задача восстановления 3D-модели заключается в определении координат точек, принадлежащих объекту по двум фотографиям.

Связь между координатами точки в пространстве и координатами её изображения можно описать с помощью оператора проективного отображения.

$$\begin{bmatrix} x'_1 \\ y'_1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & -r \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{d} & 1 \end{bmatrix} \times \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}; \quad (2.1)$$

$$\begin{bmatrix} x'_2 \\ y'_2 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & r \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{d} & 1 \end{bmatrix} \times \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}, \quad (2.2)$$

где x, y, z – координаты точки в евклидовом пространстве, d – фокусное расстояние камер;

r – расстояние от начала координат до камер, x'_1, y'_1 и x'_2 ;

y'_2 – координаты изображения точки на фотографии с камер $C1$ и $C2$, соответственно.

Решение системы из двух матричных уравнений (2.1) и (2.2) позволяет найти три координаты любой из точек – P на лице человека, находящихся в плоскости лазерной подсветки. Примеры восстановления трехмерной геометрии поверхности лица человека даны на рис. 2.10.

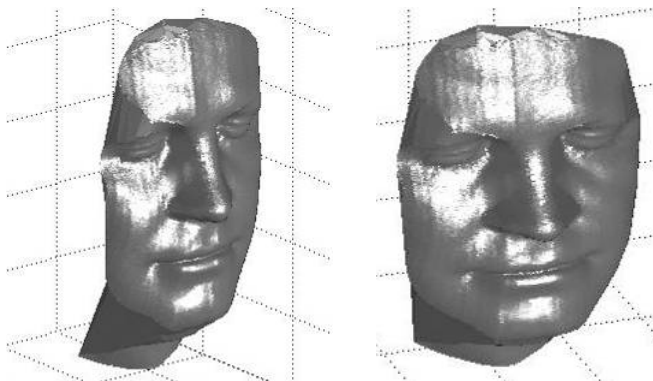


Рис. 2.10. Результат восстановления трехмерной поверхности лица человека

По данным исследований, проведенных в рамках работы [37] в Межотраслевой лаборатории тестирования биометрических устройств и технологий Пензенского государственного университета и научно-исследовательской лаборатории биометрических методов защиты информации КазНТУ имени К. И. Сатпаева, при восстановлении 512 центров круговых областей диаметром 10 мм, случайно выбранных на поверхности лица фрагментов, удается снизить вероятность ошибок второго рода до величины 10^{-5} . Этот показатель сопоставим с аналогичным показателем при контроле рисунков отпечатков пальца.

Анализ 3D-геометрии лица банкоматами следующего поколения является одной из реальных перспектив. Биометрические шаблоны 3D-геометрии лица, выполненные по биометрическим международным стандартам ISO/IEC JTC1 SC37, содержат хорошо воспринимаемую человеком информацию и потому должны быть надежно защищены. При размещении этих биометрических данных в обезличенном защищенном нейросетевом контейнере дополнительной защиты не требуется.

2.6. Технология анализа трехмерной геометрии ушных раковин

Создавая первую систему биометрической идентификации человека, Альфонс Бартильон большое внимание уделял

определению типа и формы ушных раковин человека. Причина столь большого внимания в том, что вариация положения и размеров ушных раковин у людей существенно выше, чем вариации размеров других частей тела. Кроме того, ушные раковины значительно отличаются по форме у разных людей.

На рис. 2.11 приведены фотографии ушных раковин пятерых случайно выбранных мужчин в масштабе, воспроизводящем реальные соотношения размеров ушных раковин.

Несмотря на то, что форма и размеры ушных раковин существенно варьируются, пока нет эффективных биометрических систем, использующих геометрию ушных раковин.

Видимо, это связано с проблемами трехмерного технического зрения. Пока средства искусственного зрения двумерны, они не могут создавать и запоминать сложную трехмерную форму ушных раковин. Извлечь же из двумерных образов ушных раковин достаточно большой объем биометрической информации не удастся.



Рис. 2.11. Различные формы ушных раковин пятерых случайно выбранных мужчин, приведенные в одном масштабе

Предпринимаются попытки использования оконтуривания изображения ушных раковин, однако, все программы оконтуривания оказываются очень чувствительными к боковой

подсветке и хорошо работают только на неподвижных объектах.

Предположительно интеллектуальные возможности биометрических систем будут постепенно усиливаться и, соответственно, наблюдение сложной геометрии ушных раковин станет доступным для практического применения, если перейти к анализу 3D-геометрии сложной формы ушных раковин.

Сделать геометрию ушных раковин действительно эффективной для защиты личной информации можно только обеспечив ее конфиденциальность через размещение в нейросетевых контейнерах и обеспечив дополнительно анонимность (обезличенность) владельца биометрии во время выполнения процедур аутентификации.

2.7. Одномерная и двухмерная технология анализа особенностей голоса

Слышать речь и пытаться говорить человек начинает практически со своего рождения, однако членораздельной (понятной для окружающих) речь ребенка становится только через два или три года, т. е. два или три года у каждого из нас ушло на то, чтобы научиться управлять нашими голосовыми связками, воспроизводя понятные окружающим звуки речи. Голосовой аппарат каждого из нас очень сложен (сложнее любого из искусственных музыкальных инструментов), овладев им однажды, мы на всю жизнь сохраняем свои индивидуальные особенности речи.

Речь человека состоит из звуков, которые при ее формализации могут быть отождествлены с буквами и, соответственно, устная речь человека может быть представлена в виде рукописного или печатного текста. Пример такого соответствия приведен на рис. 2.12.

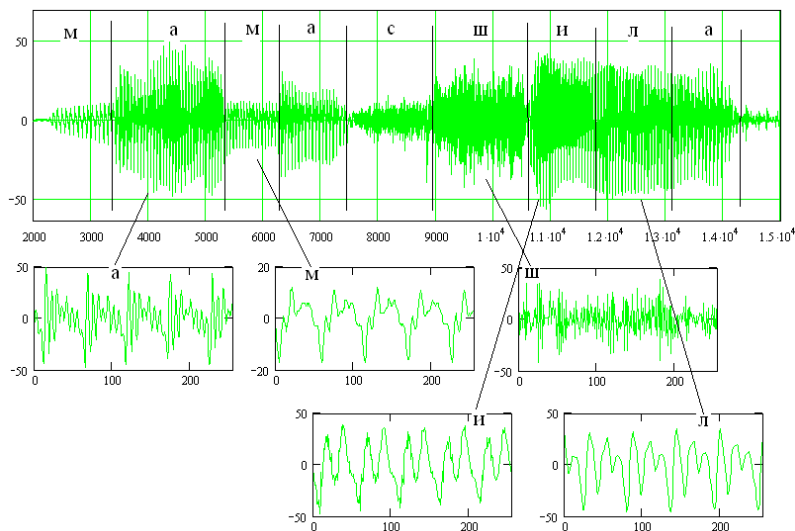


Рис. 2.12. Вариации давления звука человеком при голосовом произношении им фразы «мама шила»

Из данных рисунка видно, что каждый из звуков речи имеет свою, характерную для него форму колебаний давления. Часть звуков речи огласована (являются периодическими), а другая является шипящими (не периодическими). Для периодических звуков может быть найден основной период их следования (период основного тона). Как правило, период основного тона человека мало меняется. Его изменения обычно укладываются в интервал $\pm 20\%$. Тональные звуки, отображенные на рис. 2.12, имеют период основного тона примерно 70 отсчетов, снятых с частотой дискретизации 8 000 Гц. Это соответствует периоду основного тона 115 Гц.

В каждом периоде основного тона наблюдаются внутренние колебания звука (внутри периода основного тона). Наиболее высокочастотными являются внутренние колебания звука «а». Из данных рис. 2.12 видно, что на одно колебание основного тона звук «а» содержит 8 пиков внутренних колебаний, т. е. частота внутренних колебаний звука «а» в голосе говорящего в 8 раз выше, чем частота основного тона и составляет $115 \times 8 = 920$ Гц.

Минимальная частота внутренних колебаний наблюдается у звука «и» (два пика на один период основного тона), т. е. минимальная частота звуков внутренних колебаний голоса рис. 2.12 составляет примерно 230 Гц. Мы имеем ситуацию, когда наиболее важные в информативном отношении частоты заполнения периода основного тона изменяются в интервале от 200 до 1000 Гц.

2.8. Технология анализа динамики рукописного почерка

В настоящее время начинает активно использоваться органолептическая идентификация личности по рукописному автографу, особенно в платежных системах. Существующих средств нейросетевой автоматической идентификации личности человека достаточно для усиления надежности авторизации личности по динамике воспроизведения автографа человека, если сам процесс воспроизведения идет под контролем проверяющего. Наиболее эффективным является режим, в котором проверяемый несколько раз подписывает документ специальной ручкой, оставляющей след на мониторе с отображаемым документом.

Режим, когда проверяемый человек видит след от его пера на чувствительном экране, является принципиально важным, так как делает новую технологию аналогичной традиционной технологии формирования подписи под бумажным документом. Только в этом случае с человека снимается психологический барьер отторжения нового. На данный момент создание планшетных компьютеров с доверенной вычислительной средой и чувствительным экраном для воспроизведения на них рукописной подписи человека под электронным документом является технически выполнимой задачей. Эту задачу следует решить в рамках разработки типовых средств биометрической поддержки платежных операций. Пример планшетного доверенного вычислителя с чувствительным экраном и воспроизведением электронного документа на экране приведен на рис. 2.13.

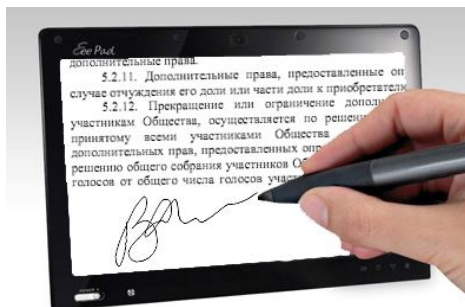


Рис. 2.13. Планшетный компьютер служащего с доверенной вычислительной средой, воспроизводящий формируемый электронный документ, который подписывается автографом на чувствительном экране

Надежность средств авторизации человека по его личному автографу может быть достаточно высока, если процедуру аутентификации делать мультиобразной (используется не менее двух, трех автографов, воспроизводимых человеком в ограниченный интервал времени). При использовании открытого биометрического образа человека в виде его автографа необходимо заставлять человека очень быстро воспроизводить свою подпись на подсознательном уровне. На данный момент в системах ограничения доступа по рукописному паролю или парольной фразе каких-либо ограничений на время воспроизведения подписи не накладывается. Технически ввести подобные ограничения не сложно.



Рис. 2.14. Стык графического планшета с ДВС в виде USB Био-токена для безопасного формирования ЭЦП плательщика

В режиме аутентификации личности по рукописному паролю отображать рукописный пароль уже не целесообразно. В связи с этим ДВС для выполнения процедур биометрической аутентификации может быть выполнена в виде биотокена, подключаемого к стандартному графическому планшету и вычислительному устройству доступа к обычно не доверенной вычислительной машине. Наиболее безопасным является выполнение всех операций по формированию цифровой подписи на личном ключе и операций по преобразованию тайного биометрического образа в личный ключ внутри доверенной вычислительной среды биотокена. Тайные биометрические образы рукописных паролей и личный ключ формирования цифровой подписи не должны покидать ДВС биотокена. В этом случае личный ключ формирования цифровой подписи и тайный биометрический образ нельзя скомпрометировать. Взаимодействие с недоверенной вычислительной средой через биотокен в форме USB-адаптера типового графического планшета изображено на рис. 2.14.

Авторизация по открытому рукописному автографу человека должна проводиться в банках и у иных операторов приема и выдачи платежей. Авторизацию по тайному биометрическому образу следует применять при дистанционном Интернете управлении банковским счетом. Во всех случаях должно быть обеспечено безопасное хранение персональных биометрических данных человека в защищенных самошифровании биометрических контейнерах [30] или в виде индивидуально настроенных на личную биометрию хэш-функций [38].

Биометрические данные о динамике открытых рукописных биометрических образов при их извлечении из доверенной вычислительной среды должны храниться в обезличенных защищенных нейросетевых контейнерах. Биометрические данные о динамике тайных рукописных биометрических образов при их извлечении из доверенной вычислительной среды могут храниться в личностных защищенных нейросетевых контейнерах. В настоящее время удается добиться вероятности ошибок второго рода на уровне от 10^{-2} до 10^{-4} для открытых рукописных образов в форме автографов и рукописного воспроизведения имени человека. В случае сохранения

рукописного парольного слова из пяти букв в тайне вероятность ошибок второго рода удастся снизить до величин от 10^{-4} до 10^{-8} . Переход к мультибиометрии (использованию нескольких разных рукописных слов) позволяет снизить степень вероятности появления ошибок пропорционально числу использованных биометрических образов.

Предположительно, что технология рукописной биометрической аутентификации личности будет использоваться только в офисных помещениях операторов платежных услуг и в домашних условиях пользователей услуг дистанционного электронного банкинга и услуг электронной Интернет торговли.

2.9. Перспективы использования мультибиометрических приложений

Объединение нескольких биометрических технологий, по-другому мультибиометрическое слияние, является одним из способов повышения уровня безопасности процедур биометрической авторизации личности при получении наличных денег или доступе к системе дистанционного управления личным счетом. Перспективы использования мультибиометрического слияния:

– Усиление стойкости биометрической системы к атакам подбора. Разница в стойкости между различными биометрическими технологиями особенно сильно проявляется при сравнении статической и динамической биометрии. Отдельные биометрические образы не всегда могут обеспечить необходимое качество и число биометрических параметров. Так, использование отпечатка пальцев в силу конечности числа особых точек и сложности рисунка, а также недостаточное качество изображения папиллярного рисунка, получаемого со сканера, в настоящее время не позволяет снизить вероятность ошибки второго рода до величины меньше 10^{-4} – 10^{-6} . Этого недостаточно для решения задачи гарантированного ограничения доступа только по биометрическому образу. Объединение нескольких «относительно слабых» видов биометрии позволит обеспечить требуемую вероятность ошибок второго рода:

– Снижение рисков компрометации биометрии. Однократная

компрометация биометрического образа исключает его из списка допустимых к использованию. Эта проблема актуальна для всех видов статической биометрии: отпечатка пальца, рисунка руки или ладони, рисунка вен, радужной оболочки глаза, геометрии лица и т. д. Решением задачи может стать мультибиометрическое объединение статической биометрии с динамической, т. е. изменяемой по желанию донора биометрии.

– Поддержка нескольких вариантов аутентификации. Применением нескольких альтернативных биометрических технологий аутентификации можно избежать потери авторизации в случае временной утраты способности ввода биометрии (например, в случае пореза пальца или «потери» голоса во время болезни). Запасной вариант аутентификации позволит выполнять операции в полном объеме без необходимости блокировки счета после компрометации части тайных биометрических образов или PIN-кодов. В этом случае клиенту банка достаточно заявить о потере возможности выполнения аутентификации одним или несколькими способами, а банк должен выполнить их блокировку. Альтернативные варианты аутентификации позволяют также реализовать совместное использование одного банковского счета супругами. В настоящий момент совместное использование карточного счета реализуется через выпуск нескольких идентичных копий пластиковых карт для одного счета.

– Поддержка нескольких участников. Ряд протоколов аутентификации требует использования нескольких ключей, хранимых отдельно друг от друга. Связь составных частей ключа с отдельными биометрическими образами позволяет использовать различные схемы объединения и контроля доступа со стороны третьих лиц или организаций. Важно отметить, что в случае биометрической авторизации передать полномочия на выполнение банковских операций намного сложнее, чем в случае с электронными носителями кодов доступа или PIN-кодов.

– Поддержка схем разделения секрета. При необходимости с помощью биометрии может быть выполнено связывание с частями секретов, распределенных между несколькими участниками процесса биометрической аутентификации.

– Сохранение требований безопасности к мультибиометри-

ческому преобразователю по сохранению тайны выходного кода и биометрических данных, требований по их уничтожению после завершения обучения, сложность организации атак с использованием хранимых параметров возможны только при выполнении требований пакета стандартов ГОСТ Р 52633 при разработке подсистемы биометрической авторизации участников платежных операций.

Процесс высоконадежной биометрической аутентификации в общем случае включает несколько стадий: ввод биометрии, ее предобработку, преобразование биометрия-код и принятие решения (рис. 2.15).

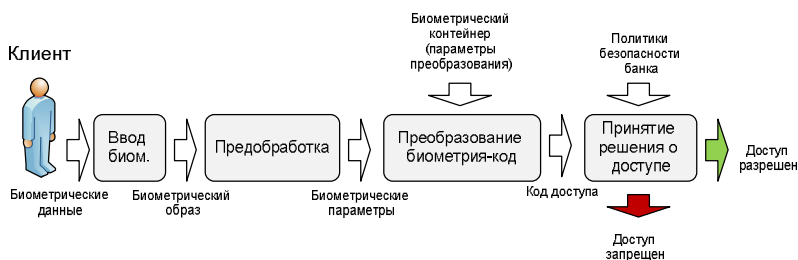


Рис. 2.15. Упрощенная схема процесса высоконадежной биометрической аутентификации

Во время ввода биометрии пользователь предъявляет свои биометрические данные физическому устройству съема биометрических характеристик, которые переводит возникающие при этом электрические или сигналы другой природы в цифровой формат представления (в биометрический образ).

Во время предобработки сформированный биометрический образ подвергается процедуре анализа и извлечения значимых биометрических параметров, которые имеют достаточное качество выделения образов «Свой» клиента из множества образов «Все Чужие», чтобы использовать их в процессе высоконадежной биометрической аутентификации.

Во время преобразования биометрия-код нестабильные в общем случае биометрические параметры преобразуются в некоторый выходной код доступа, используемый для аутентификации клиента банка, а также в значения индикаторов

близости предъявленной биометрии образу «Свой» клиента.

На этапе принятия решения банк осуществляет явное или неявное сравнение полученного кода доступа с эталонным значением, а также анализирует значения индикаторов и выносит решение о результате аутентификации: доступ разрешен, доступ запрещен. Правила принятия решения и порядок взаимодействия с клиентом во время аутентификации определяются политиками безопасности банка.

В зависимости от того, на какой стадии высоконадежной биометрической аутентификации выполняется слияние ГОСТ Р 52633.7, выделяют следующие виды мультибиометрии:

– Мультимодальная биометрическая аутентификация. Выполняется с использованием нескольких биометрических технологий, связанных с разными устройствами ввода. Примером служит биометрическая аутентификация по рукописному образу и отпечатку пальцев, выполняемая с помощью графического планшета и сканера отпечатка пальца.

– Мультиобразная биометрическая аутентификация. Выполняется с использованием несколько биометрических образов одной биометрической технологии. Примером служит биометрическая аутентификация с последовательным вводом нескольких рукописных слов-паролей или прикладыванием к одному сканеру нескольких пальцев правой и левой руки.

– Мультисенсорная биометрическая аутентификация. Выполняется с помощью нескольких устройств, предназначенных для ввода биометрического образа. Примером может служить ввод голосовой фразы с использованием нескольких микрофонов, отслеживающих положение аутентифицируемого в пространстве. Другой пример – аутентификация по объемному изображению лица, построенному с использованием двух камер наблюдения.

– Мультипримерная биометрическая аутентификация. Выполняется с помощью нескольких примеров одного биометрического образа. Используется в том случае, когда качество получаемых биометрических образов низкое и требуется их обогащение. Может выполняться незаметно для аутентифицируемого путем многократного съема биометрических данных (изображений частей тела, отпечатков пальцев) в течение

некоторого промежутка времени и последующей их обработки.

– Мультиалгоритмическая биометрическая аутентификация. По ней один биометрический образ обрабатывается несколькими разными алгоритмами с целью извлечения максимальной информации.

– Многопользовательская биометрическая аутентификация. Выполняется с участием нескольких доноров биометрических образов. Проводится в случае использования различных схем разбиения или разделения секрета.

– Многовариантная биометрическая аутентификация. Объединяет разные виды мультибиометрической аутентификации с помощью логической операции "или".

Все перечисленные виды мультибиометрической аутентификации могут использоваться совместно (смешанная мультибиометрическая аутентификация).

Любой из этих способов объединения нескольких биометрических технологий может использоваться для усиления информационной безопасности платежной системы на базе высоконадежных преобразователей биометрии-кода. При этом последующая модернизация системы путем добавления новых (альтернативных) технологий биометрической аутентификации при реализации хотя бы одной из них не нарушит общую схему биометрической аутентификации клиента.

Особенностью применения мультибиометрической аутентификации является необходимость поочередного или одновременного ввода нескольких биометрических образов. Работа по унификации этого процесса позволила описать процедуру взаимодействия в виде 4 шагов (рис. 2.16), которые могут быть реализованы во время дистанционной аутентификации клиента, аутентификации клиента через банкомат (терминал), аутентификации во время личного присутствия в отделении банка. Во время мультибиометрической аутентификации клиенту в соответствии с использованной схемой аутентификации будет последовательно предлагаться ввод примеров биометрических образов. Каждый введенный образ направляется в процессинговый центр банка. В нем анализируется текущая последовательность ввода биометрических образов на предмет достаточности для выработки кода аутентификации и соответствия их характеристик ожидаемым. При этом для

распознавания атаки на биометрию и коды аутентификации клиента используется биометрический аудит и оценка значений индикаторов качества предъявленной биометрии. На основе полученного результата принимается решение о продолжении процесса аутентификации или ее прекращении с положительным или отрицательным результатом, которое направляется клиенту. В зависимости от выбранной банком политики аутентификации может быть достигнут требуемый уровень безопасности и компромисс между снимаемыми угрозами информационной безопасности и удобством практического использования процедуры мультибиометрической аутентификации.

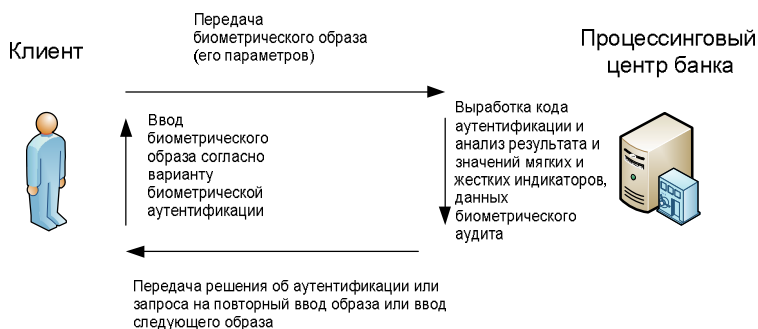


Рис. 2.16. Порядок предъявления биометрии во время мультибиометрической аутентификации

Переход к мультибиометрической аутентификации пользователей платежных систем существенно повышает надежность биометрической аутентификации личности с одной стороны и усложняет сам процесс аутентификации для пользователей с другой. Для того, чтобы облегчить пользователю его мультибиометрическую аутентификацию, она должна быть снабжена безопасными индикаторами близости предъявленных биометрических данных к эталонным. Безопасные индикаторы близости должны быть выполнены по требованиям ГОСТ Р 52633.6 – «Защита информации. Техника защиты информации. Требования к индикации близости предъявленных биометрических данных образу «Свой».

3. НЕЧЕТКИЕ ЭКСТРАКТОРЫ БИОМЕТРИЧЕСКИХ ПАРАМЕТРОВ

В настоящее время США и страны Евросоюза идут по пути связывания биометрических данных человека с его криптографическим ключом через применение так называемых «нечетких экстракторов» [39–51]. При этом «нечеткие экстракторы» рассматриваются как алгоритмы, выделяющие случайные, равномерно распределенные последовательности битов из биометрических данных в условиях зашумленности и способные компенсировать ошибки, возникающие из-за невозможности абсолютно точного повторного воспроизведения биометрических данных. «Нечеткие экстракторы» можно использовать для генерации ключей без необходимости их хранения в промежутках между обращениями к ним. По своей сути «нечеткие экстракторы» строятся путем бинарного квантования «сырых» биометрических данных и последующей корректировки ошибок квантования классическими избыточными кодами с обнаружением ошибок.

3.1. Простейший вариант квантования биометрических данных

Для биометрических технологий контролируемый биометрический параметр может быть представлен как некоторая функция, изменяющаяся во времени. Например, это могут быть изменение яркости участков радужной оболочки глаза при радиальном сканировании, изменение амплитуды колебаний звуковой волны, координаты колебаний пера при воспроизведении рукописного пароля. Пример одного из таких колебаний приведен на рис. 3.1.

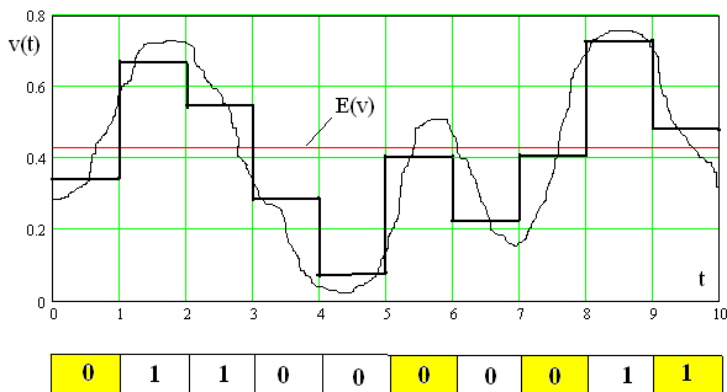


Рис. 3.1. Квантование кривой изменения биометрического параметра $v(t)$ относительно среднего значения

Простейший вариант квантования кривой изменения биометрических данных сводится к разбиению времени на заранее заданное число интервалов (на рис. 3.1 используется 10 интервалов). Далее в каждом из интервалов квантования подсчитывается среднее значение параметра $v(t)$ и сравнивается со средним значением по всем интервалам наблюдения $E(v)$. Если среднее значение параметра на интервале больше $E(v)$, то этому разряду биокода присваивается состояние «1». Если среднее значение параметра на интервале меньше $E(v)$, то этому разряду биокода присваивается состояние «0». В нижней части рис. 3.1 даны разряды получившегося биокода. Примерно половина разрядов биокода, сформированного таким образом оказывается нестабильными, так как попадают на участки квантования с примерно одинаковыми отрицательными и положительными фрагментами кривой $v(t)$. Нестабильные разряды биокода на рис. 3.1 отмечены темной заливкой. На практике большинство наиболее нестабильных разрядов биокода (до 80 %) маскируют, а оставшиеся 20 % нестабильных разрядов исправляют применением кодов, способных обнаруживать и корректировать ошибки.

3.2. «Нечеткий экстрактор» с сокрытием информации наложением гаммы

В том случае, если биокоды хранятся и обрабатываются в доверенной вычислительной среде, биометрическую аутентификацию личности можно выполнять простым сравнением шаблона биокода и биокода проверяемого человека. При такой технологии открытые шаблоны биокодов должны храниться в базе данных. По мере роста размеров базы биометрических шаблонов растет ее ценность, соответственно растет риск хищения базы биометрических кодов.

Одним из способов защиты биокодов человека является их гаммирование [52–59]. На сегодняшний день большинство «нечетких экстракторов» строится на гаммировании биокода личности с получением гаммы и снятием гаммы по схеме, представленной на рис. 3.2.

Из рис. 3.2 видно, что защитная гамма получается из неизвестного окружающим личного криптографического ключа человека. При формировании гаммы проводят вычисление синдромов ошибок выбранного самокорректирующегося кода. После добавления синдромов ошибок в избыточную часть самокорректирующегося кода и объединения ее с информационной частью самокорректирующегося кода длина полученной гаммы должна совпадать с длиной защищаемого биокода.

На рис. 3.2 показано, что длина личного ключа пользователя всегда меньше длины его биокода. Соотношение длины личного ключа и биокода выбирается из числа корректируемых в биокоде ошибок. Если применяют самокорректирующиеся коды со 100%-й избыточностью, то удастся исправить примерно 4 % ошибок в биокоде. В этом случае длина личного криптографического ключа должна составлять половину длины биокода или нечеткого контейнера.

При биометрической аутентификации происходит обратный процесс, только что сформированный биокод складывается по модулю два с нечетким контейнером. При этом происходит почти полное снятие, защищающее биокод гаммы. Гамма не снимается только в ошибочных битах биокода. Эти ошибочные биты биокода правятся самокорректирующимся кодом, и тем

самым восстанавливается личный ключ пользователя, который далее используется в протоколе криптографической аутентификации.

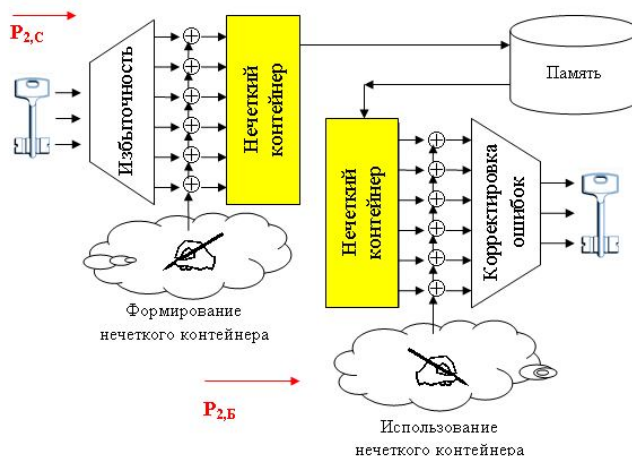


Рис. 3.2. Наложение гаммы при формировании нечеткого контейнера и ее снятие во время процедуры аутентификации

«Нечеткие экстракторы» интересны тем, что позволяют в явной форме разделить стойкость криптографической защиты биокода от наблюдения и стойкость биометрической защиты от атак подстановки случайных образов «Чужой». Очевидно, что стойкость криптографической защиты будет пропорциональна длине личного ключа (обратно пропорциональна вероятности ошибок второго рода $P_{2,C}$, обусловленных угадыванием личного ключа с первой попытки). Биометрическая стойкость защиты пропорциональна длине биокода и показателям стабильности его разрядов.

3.3. «Нечеткий экстрактор» с хэшированием биокода и наложением гаммы

Еще одним способом формирования защитной гаммы является применение предварительного хэширования биокода [55, 56]. Блок схема второго способа формирования гаммы приведена на рис. 3.3.

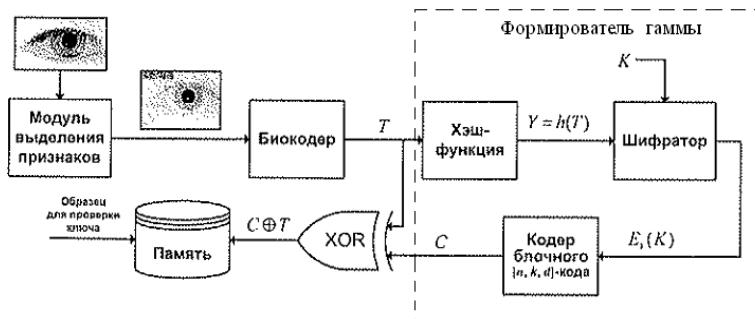


Рис. 3.3. Второй способ формирования гаммы «С» и создания нечеткого контейнера с использованием хэширования

По второму способу биокод после его хэширования шифруют на ключе K . Далее полученный шифротекст накрывают самокорректирующимся кодом и полученной гаммой накрывают биокод. Сформированный таким образом нечеткий контейнер можно безопасно хранить в базе данных.

При биометрической аутентификации осуществляют обратный процесс снятия гаммы, блок-схема которого приведена на рис. 3.4.

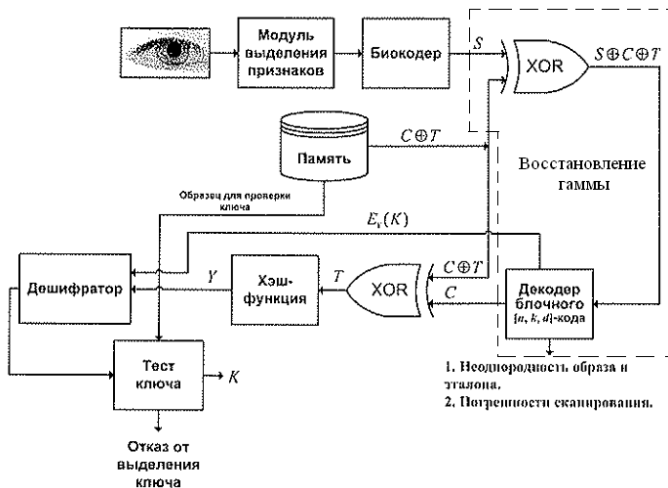


Рис. 3.4. Формирование восстанавливающей гаммы «С» перед повторным хэшированием и восстановлением ключа дешифратором

Из рис. 3.4 видно, что снятие гаммы происходит так же, как и у всех иных нечетких экстракторов. Нечеткий контейнер из базы данных складывают по модулю два с только что полученным биокодом. При этом происходит полное снятие гаммы в верных битах биокода. Ошибочные биты биокода корректируются за счет введенной ранее избыточности. Отличие первого способа (рис. 3.2) от второго состоит в дополнительной операции хэширования восстановленного биокода и более сложной операции восстановления ключа шифрования.

3.4. Выбор используемых в «нечетком экстракторе» самокорректирующихся кодов

Обычно в «нечетких экстракторах» стараются маскированием увеличить стабильность разрядов биокода до уровня от 5 до 10 % появления ошибок. Далее правят ошибки классическими кодами обнаружения и коррекции ошибок. Как правило, все классические самокорректирующиеся коды [57–59] без особых проблем правят до 10 % ошибок.

Теория помехоустойчивого кодирования достаточно сложна. Создано несколько десятков достаточно эффективных кодов, любой из которых может быть применен в нечетких экстракторах. Основным недостатком всех классических помехоустойчивых кодов является то, что они избыточны и требуют тратить часть биокода на заполнение избыточности классических самокорректирующихся кодов. Проиллюстрируем это на примере простейших блочных кодов с контролем четности. Принцип формирования информационной и избыточной части таких кодов иллюстрируется рис. 3.5.

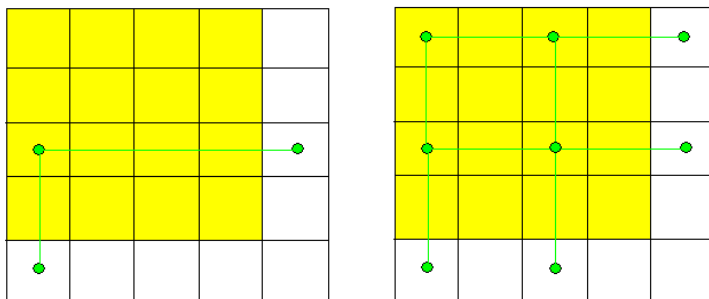


Рис. 3.5. Простейший блочный код с контролем четности столбцов и строк матрицы информационного поля

На рис. 3.5 заливкой отмечена информационная матрица блочного кода, а правая и нижняя строки матриц образуют избыточную часть кода. В полях правого столбца и нижней части матрицы кода записывают биты четности данных, соответственно, в строках и столбцах информационной матрицы. В правой нижней ячейке записывают бит четности информации в избыточной части блочного кода.

Проверку кода осуществляют сравнением показателя четности столбцов и матриц информационной составляющей кода с их битами четности в избыточной части кода. В левой части рис. 3.5 изображена ситуация, когда появилась одна ошибка в блоке из 25 бит. Эта ошибка легко обнаруживается и исправляется инвертированием информации на пересечении ошибочной строки и ошибочного столбца, т. е. рассматриваемый блочный код способен исправлять одну ошибку на блок, а обнаруживать порядка 8 ошибок в блоке. При избыточности кода в 56 % он с вероятностью $P = 1,0$ исправляет одиночные ошибки или примерно 4 % разнесенных по отношению друг к другу ошибок.

Если на каждый блок будет приходится по 2 ошибки (8 % ошибок), этот код правит все ошибки с вероятностью 0.5. Эта ситуация иллюстрируется правой частью рис. 3.5, из которой видно, что обнаруживаются 4 места, в которых возможны правки. Исправить же необходимо только 2 бита

блочного кода. Для того, чтобы однозначно править порядка 8–12 % ошибок, необходимо использовать более сложное кодирование (например, коды БЧХ или коды Рида-Соломона), имеющие избыточность от 100 до 200 %.

Заметим, что, используя классический помехоустойчивый код с 200 %-й избыточностью, фактически необходимо в информационной части оставлять только 1/3 разрядов. Попытки корректировать большее число ошибок в биокоде дают еще большие потери информации.

При исправлении 20 % ошибок в биокоде информационная составляющая исходного биокода снижается до 1/5 и даже до 1/10 от исходной длины.

Исключить эффект катастрофического сжатия информационной части био-кодов удастся, если перейти от контроля четности к контролю значения части разрядов хэш-функций, вычисленных по строкам и столбцам расширенной информационной матрицы 128x64 [60, 61], которая получается таблицей случайного выбора данных из меньшей информационной матрицы био-ключа 16x16. Технический эффект состоит в том, что младшие разряды хэш-функций столбцов и строк (1,- 2- или 3-го разрядов) можно хранить открыто из-за необратимости криптографических хэш-функций. Одиночные ошибки находятся так же, как и в простейших блочных кодах с контролем четности. Групповые ошибки находятся перебором состояний хэш-функций, соответствующих состояниям «верно» эталонных разрядов. Легко показать, что при увеличении контролируемых разрядов усеченной хэш-функции экспоненциально растет время, затрачиваемое на декодирование, однако при этом полиномиально растет число исправляемых ошибок [61].

4. НЕЙРОСЕТЕВЫЕ ПРЕОБРАЗОВАТЕЛИ БИОМЕТРИИ-КОДА

4.1. Итерационный алгоритм обучения одиночного нейрона (персептрона)

Информации, содержащейся в «плохих» биометрических параметрах, недостаточно для того, чтобы по одному биометрическому параметру принимать надежные решения. Именно в связи с этим «нечеткие экстракторы» рассматривают не один разряд биокода, а весь биокод и правят ошибки в нем. Более эффективным является иной технический прием, в соответствии с которым еще до квантования биометрических данных осуществляют их нейросетевое обогащение. При нейросетевом обогащении используют не все био-данные, а их группы. При этом, чем хуже данные, тем больше данных должно объединяться в одну группу, обогащаемую одним нейроном. Если подсчитать информативность анализируемых биометрических параметров, то можно приближенно оценить число нейронов, необходимых для обогащения (рис. 4.1).

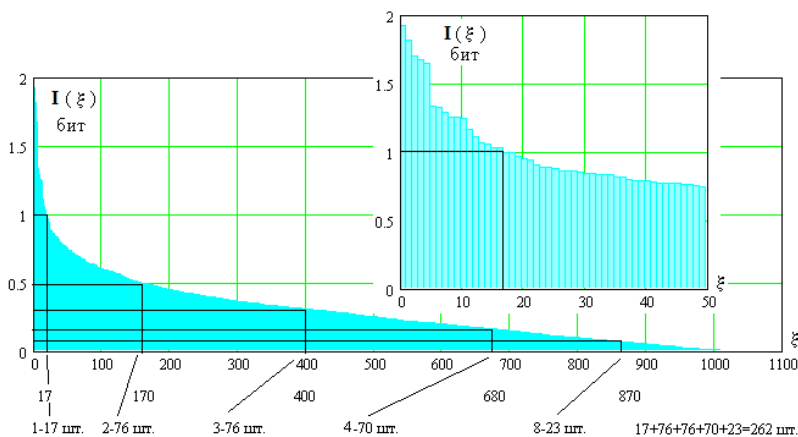


Рис. 4.1. Примерная оценка числа нейронов, необходимых для обогащения биометрических данных

Из распределения показателей информативности биометрических параметров (рис. 4.1) видно, что высокие показатели информативности имеют только 17 биометрических параметров, каждый из которых может самостоятельно использоваться после его квантования. Биопараметров «удовлетворительного» качества с информативностью от 1 до 0.5 бит – 152, они могут использоваться для создания 76 нейронов с двумя входами.

Столько же 76 нейронов получается с 3-мя входами, при этом информативность каждого из входов находится в интервале от 0.5 до 0.35 бит. Био-данные с еще меньшей информативностью приходится обогащать 70 нейронами с 4-мя входами и примерно 23 нейронами с 8-ми входами. Всего получается примерно 262 нейрона с неповторяющимися входами, при этом выходы каждого из нейронов дают по одному стабильному разряду биоключа. Следует иметь в виду, что приведенные выше оценки являются весьма и весьма приближенными, так как не учитывают сильные корреляционные связи между параметрами образа «Свой». В реальных биометрических системах число входов у нейронов должно быть примерно в два–три раза больше, чем следует из приведенных выше оценок.

Для того, чтобы воспользоваться нейросетевым обогащением био-данных, необходимо уметь обучать одиночные искусственные нейроны. По вопросам обучения искусственных нейронов и сетей из них существует обширная русскоязычная литература [62–66]. Существует несколько сотен алгоритмов обучения искусственных нейронов и множество вариантов исполнения функции возбуждения нейронов.

Применительно к решению задач биометрии [67–70] в основном у нейронов используется ступенчатая нечетная функция возбуждения (квантования) как у персептрона [70, 71]. Реже используется ступенчатая четная функция квантования, значительные перспективы имеют нейроны с немонотонными функциями квантования, имеющими 3 и более выходных состояний [66, 72]. Название «персептрон» закрепилось только за нейронами с простейшей функцией квантования выходных данных сумматора, в связи с этим далее по тексту термин

«персептрон» использоваться будет параллельно с более общим термином «нейрон».

Исторически сложилось так, что первыми были созданы алгоритмы обучения, построенные на итерационной минимизации среднеквадратической ошибки, принимаемых нейроном решений.

Отойдем от традиции и не будем рассматривать критерий минимума среднеквадратической ошибки, так как он в биометрии редко используется. Применительно к биометрии гораздо более эффективным является критерий так называемого качества разделения распределений «Свой» и «Все Чужие».

Основная функция нейросетевых приложений биометрии является выталкивание распределения образа «Свой» из области распределения образов «Все Чужие». Эта ситуация иллюстрируется рис. 4.2.

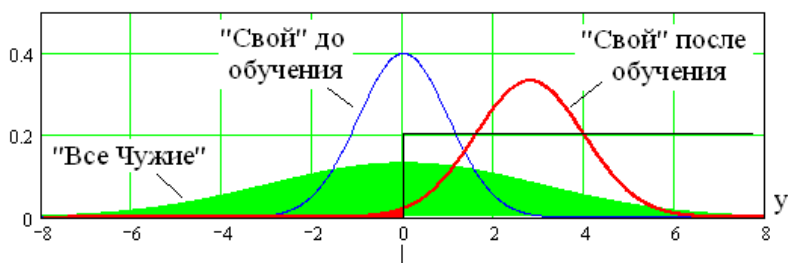


Рис. 4.2. Эффект выталкивания распределения данных образа «Свой» из области распределения данных «Все Чужие» на выходе сумматора нейрона

Перед обучением весовые коэффициенты сумматора нейрона задаются случайно, как следствие наиболее вероятной является ситуация, когда распределение состояний образа «Свой» на выходе сумматора нейрона оказывается в центре распределения образов «Все Чужие». Выходная функция квантования нейрона в биометрических приложениях всегда настраивается так, чтобы срабатывать в центре распределения образов «Все Чужие». Как следствие распределение образов «Свой» на выходе необученного нейрона оказывается нестабильным (дает примерно с равной вероятностью состояния «0» и «1»).

В том случае, когда нам нужно получить на выходе нейрона состояние «1» для образа «Свой», то необходимо обучать нейрон, выталкивая распределение образа «Свой» вправо из центра распределения образов «Все Чужие». То, насколько далеко мы вытолкнем распределение образа «Свой», зависит качество работы нейрона. Для контроля текущего качества обучения используется следующий показатель качества:

$$Q = \frac{|E(y_v) - E(y_\xi)|}{\sqrt{\sigma_v \cdot \sigma_\xi}}, \quad (4.1),$$

где $E(y_v)$ – математическое ожидание примеров «Свой»;

$E(y_\xi)$ – математическое ожидание примеров «Все Чужие»;

σ_v – среднееквадратическое отклонение примеров «Свой»;

σ_ξ – среднееквадратическое отклонение примеров «Все Чужие».

Показатель качества тем выше, чем больше разница между математическими ожиданиями разделяемых множеств. Показатель качества так же увеличивается, если при фиксированном расстоянии между центрами разделяемых множеств уменьшается среднее геометрическое среднееквадратических расстояний разделяемых множеств.

Пользуясь показателем качества (4.1), можно построить итерационную процедуру обучения нейрона (рис. 4.3) по критерию движения в сторону повышения качества. Для этой цели необходимо вычислять частные производные по качеству обучения $\frac{\partial Q}{\partial \mu_i}$. Если производная положительна, то следует

увеличивать значение регулируемого весового коэффициента μ_i . При отрицательной производной качества следует уменьшать регулируемый весовой коэффициент.

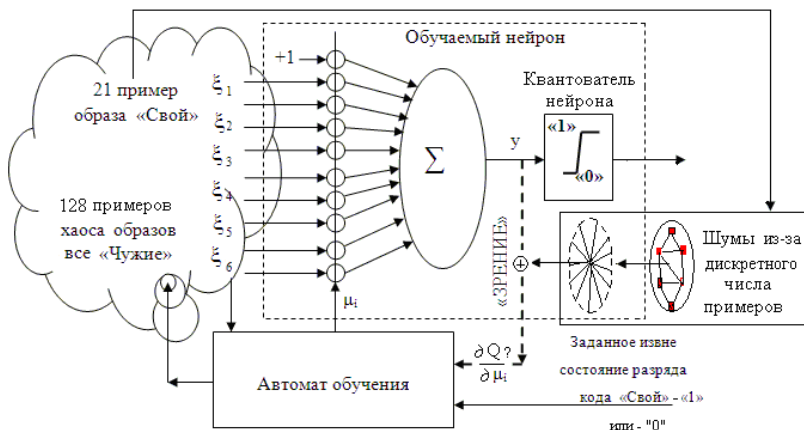


Рис. 4.3. Обучение одиночного нейрона итерационным алгоритмом поиска максимума качества обучения

Казалось бы, что вычисляя частные производные по качеству обучения, легко можно построить автомат итерационного обучения. Это действительно так, когда речь идет об обучении нейронов с малым числом входов при биометрических данных «хорошего» качества. Если это не так, то обучать нейроны с большим числом входных данных оказывается трудно. Проблема состоит во внутренних шумах процедуры обучения, которые обусловлены малым дискретным числом примеров обучения «Свой». Заранее создать достаточное число примеров «Все Чужие» технически не сложно. Намного сложнее заставить пользователей прикладывать свои усилия, создавая десятки или даже сотни примеров образа «Свой».

Вычислительная проблема обучения состоит в неустойчивости вычисления частных производных качества $\frac{\partial Q}{\partial \mu_i}$ при малом числе примеров «Свой». Погрешности $\Delta E(y_v)$, $\Delta \sigma(y_v)$, возникающие из-за малого числа примеров «Свой» усиливается при вычислении производных $\frac{\partial Q}{\partial \mu_i}$ и «забывает» полезную компоненту направления движения в сторону $\max\{Q\}$. Возникает хаотическое блуждание по многомерной поверхности

ошибок вместо направленного движения в сторону максимума качества обучения. Эта ситуация иллюстрируется рис. 4.4.

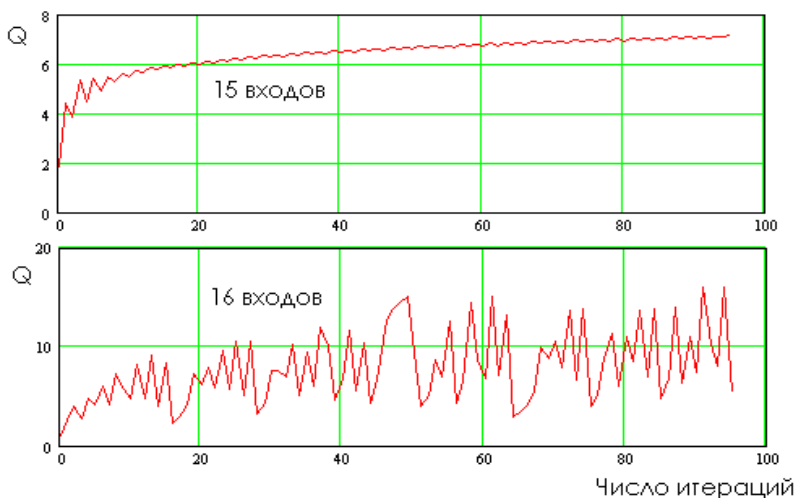


Рис. 4.4. Утрата устойчивости процедуры итерационного обучения одного нейрона

Из рис. 4.4 видно, что при 15 входах нейрона итерационное обучение происходит почти устойчиво, однако при увеличении числа входов нейрона до 16 устойчивость итерационного обучения утрачивается. Появляется множество так называемых локальных минимумов и максимумов качества обучения, являющихся не чем иным, как усилением собственного шума вычислительных процедур, обусловленного конечным числом примеров континуума данных образа «Свой».

4.2. Абсолютно устойчивый неитерационный алгоритм обучения нейрона

Проблема обучения нейронов биометрических приложений состоит в том, что человек не должен участвовать в этом процессе. Необходимо иметь абсолютно устойчивый автомат обучения, который гарантированно мог бы обучить нейрон. ГОСТ Р 52633.5–2011 рекомендует использовать неитерационные

алгоритмы обучения. Для этого типа алгоритмов нет необходимости вычислять частные производные и полностью размыкается петля обратной связи по ним (рис. 4.3).

Этот тип алгоритмов построен на «угадывании» ожидаемого решения через угадывание наиболее вероятных знаков при весовых коэффициентах и угадывании (предсказании) значений весовых коэффициентов. Естественно, что происходит определенное снижение качества обучения (нет оптимизации качества), однако возникает эффект высокой устойчивости вычислений. При этом устойчивость вычислений будет тем выше, чем больше входов у обучаемого нейрона. Вместо «проклятия» размерности появляется противоположный эффект «благодати» высоких и сверхвысоких размерностей.

Автомат угадывания знака весовых коэффициентов нейрона может быть построен на гипотезе вырождения сумматора до одного входа. Перед этим автомат обучения должен принять решение, в какую сторону он будет выталкивать образ «Свой» (вправо – R или влево – L). Будем считать, что нас интересует R-выталкивание образа «Свой» из множества образов «Все Чужие». Данная ситуация отображена на рис. 4.5 стрелкой.

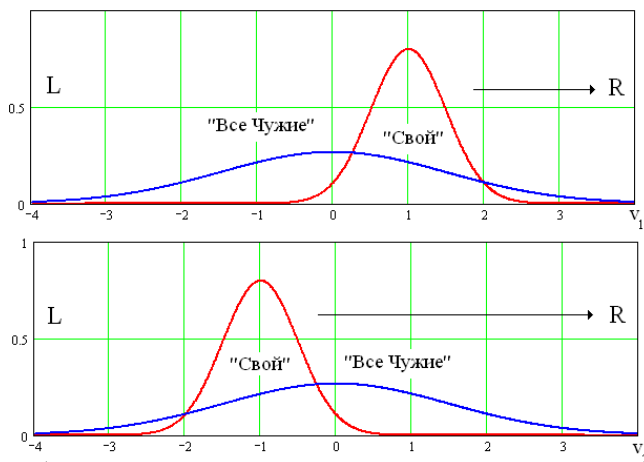


Рис. 4.5. Решение об R-выталкивании образов «Свой» из центра множества образов «Все Чужие»

Из рис. 4.5 видно, что заведомый отказ от решения многомерной задачи (у сумматора нейрона имеется только один вход) делает задачу выбора знака весового коэффициента одного входа тривиальной. Так, в верхней части рисунка знак весового коэффициента должен быть положительным, потому что математическое ожидание $E(v_1)$ положительно. В нижней части рисунка отображена обратная ситуация, математическое ожидание $E(v_2)$ отрицательно. Для эффективного R-выталкивания образа «Свой» необходимо сменить знак весового коэффициента второго входа сумматора. В итоге для R-выталкивания получается некоторая процедура предсказания знака весовых коэффициентов у обучаемого нейрона:

$$R - \text{предсказание } i\text{-го знака} \begin{cases} +\mu_i & \text{if } E(\xi_i) \leq E(v_i); \\ -\mu_i & \text{if } E(\xi_i) \geq E(v_i). \end{cases} \quad (4.2)$$

Очевидно, что если задаться обратной целью выталкивания образа «Свой» в левую сторону из множества образов «Все Чужие», то получится инверсия R правила (4.2):

$$L - \text{предсказание } i\text{-го знака} \begin{cases} +\mu_i & \text{if } E(\xi_i) \geq E(v_i); \\ -\mu_i & \text{if } E(\xi_i) \leq E(v_i). \end{cases} \quad (4.3)$$

На самом деле сложно точно указать тот или иной знак весового коэффициента обучаемого нейрона с множеством входов. Точно вычислить этот знак невозможно, и он в процессе итерационного обучения может несколько раз измениться, однако, примерное предсказание знака весового коэффициента нейрона через использование выражений (4.2) и (4.3) вполне возможно. Очевидно, что, чем выше вероятность верного предсказания знака весового коэффициента нейрона, тем устойчивее обучение нейрона.

Оценку эффективности предсказания знака весового коэффициента можно осуществить сравнивая знаки весовых коэффициентов обученного нейрона и знаки, полученные в результате вычислений по формулам (4.2) и (4.3). Численный эксперимент показывает, что вероятность ошибочного прогноза знака составляет от 0.11 до 0.27, т. е. процедуры прогноза знака весового коэффициента дают верный результат минимум в 73 %

случаев, максимум – в 89. Именно это обстоятельство и является предпосылкой для создания быстрых неитерационных алгоритмов обучения линейной части искусственных нейронов.

Если известно, что итерационные алгоритмы обучения нейрона неустойчивы и могут приводить как к переобучению нейрона, так и к недообучению нейрона, необходимо научиться предсказывать значения весовых коэффициентов нейрона опираясь на статистики входных данных «Свой» и «Все Чужие». Самым простым способом предсказания возможных значений весовых коэффициентов является учет нестабильности (с.к.о.) данных образа «Свой». Идея формирования такого прогноза состоит в ослаблении влияния нестабильных биометрических параметров через коэффициенты их влияния:

$$|\mu_i| = \frac{1}{\sigma(v_i)} . \quad (4.4)$$

При выборе коэффициентов по формуле (4.4) нестабильные коэффициенты с большим значением $\sigma(v_i)$ оказывают примерно в 2–3 раза меньшее влияние на результат по отношению к более стабильным коэффициентам с меньшей дисперсией.

Еще более эффективного выталкивания образа «Свой» из образов «Все Чужие» удастся достичь, если усилить влияние биометрических параметров, обладающих высокой уникальностью. Таким образом, весовые коэффициенты должны вычисляться по формуле

$$|\mu_i| = \frac{|E(\xi_i) - E(v_i)|}{\sigma(v_i)} . \quad (4.5)$$

Практика показывает, что вычисление весовых коэффициентов по формуле (4.5) позволяет достаточно быстро вытолкнуть образы «Свой» на периферию образов «Все Чужие».

Теоретически за счет увеличения числа учитываемых биометрических параметров $n \rightarrow \infty$ можно как угодно далеко вытолкнуть образ «Свой» на выходе обученного сумматора. Однако слишком сильное выталкивание образа «Свой» за пределы образов «Все Чужие» недопустимо из-за возможности осуществления атаки поиска биометрических близнецов.

4.3. Формирование и обучение сети нейронов. Нейросетевые контейнеры

Для того, чтобы получить био-код ключа доступа, необходимо создать сеть нейронов с числом выходов, равным длине ключа. Чем больше входов у нейронной сети и чем больше выходов у нейронной сети, тем выше качество принимаемых решений. Насколько сильна подобная связь видно из данных табл. 4.1.

Таблица 4.1
Рост качества решений в зависимости от числа входов и выходов искусственной нейронной сети

Число входов нейросети	Число выходов нейросети	Вероятность отказа «Своему»	Вероятность пропуска «Чужого»
5 входов	1 выход, 2 класса	$P_1 = 0,1$	$P_2 = 0,17$
48 входов	1 выход, 2 класса	$P_1 = 0,1$	$P_2 = 0,03$
480 входов	1 выход, 2 класса	$P_1 = 0,1$	$P_2 = 0,005$
480 входов	256 выходов, 2^{256} классов	$P_1 = 0,1$	$P_2 = 0,000000001$
Обучение нейросети с 480 входами и 256 выходами велось на 20 примерах образа «Свой» и 128 примеров образов «Чужой» алгоритмом ГОСТ Р 52633.5–2011			

Из табл. 4.1 видно, что простое увеличение числа входов (простое увеличение) числа учитываемых биометрических параметров не очень эффективно. Гораздо важнее параллельно с числом входов нейронной сети увеличивать число ее выходов. При одинаковом числе входов (480 входов) увеличение числа выходов с 1 до 256 дает выигрыш в качестве принимаемых нейронной сетью решений примерно в миллиард раз (дополнительно появляется шесть нулей). При этом время и иные затраты вычислительных ресурсов увеличиваются примерно в 100 раз, видна экспоненциальная связь размеров искусственного интеллекта и качества принимаемых им решений.

Одним из важнейших вопросов является выбор структуры, используемой нейронной сети. Обычно в литературе по искусственным нейронным сетям разделяют сети на одно-, двух- и трехслойные, а также сети с большим, чем три, числом нейронов. Столь широкое многообразие структур нейронных сетей для биометрии не актуально. ГОСТ Р 52633.5–2011 предусматривает либо однослойные, либо двухслойные нейронные сети. Для двухслойных нейронных сетей функции первого и второго слоев разделены. Нейроны первого слоя выполняют функцию обогащения биометрических данных и квантования обогащенных данных. Если качество обогащения оказалось недостаточно велико, то второй слой нейронов правит ошибки биокода нейронов первого слоя.

Следует отметить, что второй слой нейронов всегда может быть заменен обычным классическим кодом, обнаруживающим и исправляющим ошибки, однако нейросетевое корректирование ошибок выгоднее. Причина выгоды от применения нейросетевых корректоров состоит в том, что при обучении второго слоя на примерах биокодов «Свой» оценивают реальный показатель стабильности каждого из разрядов биокода.

Практика показывает, что подавляющее большинство разрядов биокода имеет высокую стабильность, нестабильными оказываются только отдельные разряды кода с точно известным положением.

Второй слой нейронов обучается корректировать нестабильные разряды и одновременно хэшировать (перемешивать) как стабильные, так и нестабильные разряды. Все классические коды с обнаружением и исправлением ошибок наоборот строятся в рамках гипотезы о равновероятном распределении ошибок между разрядами кода. Именно из-за этого классические самокорректирующиеся коды проигрывают нейросетевым корректорам ошибок, которые во время обучения учитывают реальное распределение показателей стабильности биокодов «Свой».

Кроме числа слоев нейронов сети необходимо выбирать число входов каждого нейрона и задать связи входов с номерами входов сети. Так, если вся нейросеть имеет 480 входов и средняя информативность входов составляет порядка

0.3 бита, то потребуется использовать нейроны с числом входов от 1 до 18 (в зависимости от качества используемых нейроном биометрических параметров и корреляционных связей между ними). Необходимое число входов может быть найдено только во время обучения нейрона. То есть изначально задают случайным выбором малое число входов далее, если качество решения не дотягивает до заданного, то увеличивают число входов нейрона. В конечном итоге получается однослойная сеть нейронов, причем каждый нейрон будет иметь свое число входов, подключенных случайно к входам всей сети. После обучения дополнительно получается таблица весовых коэффициентов для входных связей каждого из нейронов.

Формально обученная сеть описывается таблицами связей нейронов и таблицами весовых коэффициентов. Если сеть двухслойная, то таблицы номеров связей и таблицы весовых коэффициентов должны быть созданы для каждого из слоев нейронов. Слои нейронов обучаются последовательно. После обучения первого слоя нейронов проводят транслирование примеров образов «Свой» и «Все Чужие» с входа нейронной сети на выходы нейронов, получают примеры био-кодов и обучают на них нейроны второго слоя.

Таблицы связей нейронной сети и таблицы весовых коэффициентов, обученных нейронов, образуют так называемые нейросетевые контейнеры. В нейросетевом контейнере информации достаточно, чтобы в нужный момент воспроизвести программно обученную нейросеть и преобразовать биометрические данные человека в код его криптографического ключа доступа. Процедура аутентификации, построенная с использованием нейросетевого контейнера, приведена на рис. 4.6.

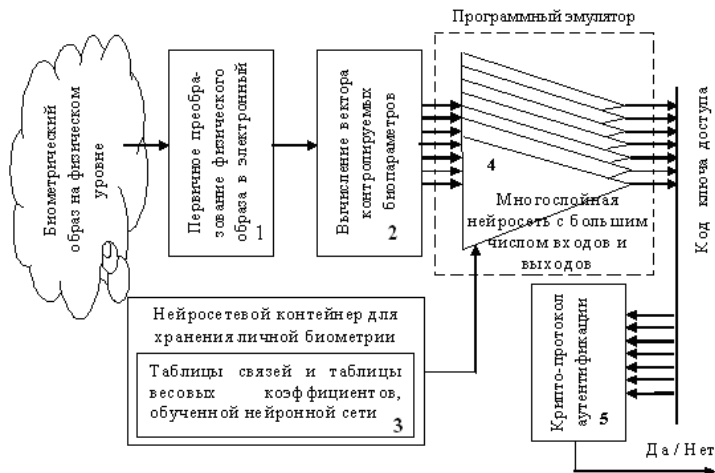


Рис.4.6. Блок-схема биометрической аутентификации с использованием нейросетевого контейнера

Формирование и использование нейросетевых контейнеров для хранения биометрии человека и интерфейсы взаимодействия с ними регламентируются ГОСТ Р 52633.4–2012 [73]. В случае, когда приходится использовать несколько разных биометрических технологий (мультибиометрическая аутентификация), взаимодействие нескольких нейросетевых контейнеров с биометрией осуществляется в соответствии с требованиями ГОСТ Р 52633.7–2013 [74].

4.4. Универсальность двухслойной сети искусственных нейронов

Выбор ГОСТ Р 52633.5–2011 двухслойной сети нейронов (персептронов) обусловлен тем, что эта нейросетевая структура универсальна и может решать любые задачи. М. Минский и С. Пейперт [75] показали, что одиночный персептрон не универсален. Персептрон осуществляет деление входного многомерного пространства биометрических параметров некоторой гиперплоскостью на две части, соответствующие выходному состоянию «0» и «1». Отобразить это для

многомерных пространств двумерной графикой рисунка в книге нельзя. Иллюстрировать можно только состояния двумерного персептрона (рис. 4.7).

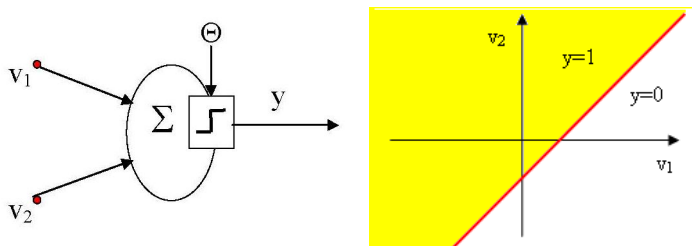


Рис. 4.7. Деление двумерным нейроном плоскости на две части «0» и «1»

Линейность персептрона не позволяет ему выполнять нелинейные функции, например, реализовывать достаточно простую функцию «исключающее ИЛИ». Ниже дана таблица состояний «исключающего ИЛИ» и графическое отображение этих состояний:

Состояния элемента XOR			элемента XOR	
v_1	v_2	y	v_1	y
-1	-1	«0»		
-1	+1	«1»		
+1	-1	«1»		
+1	+1	«0»		

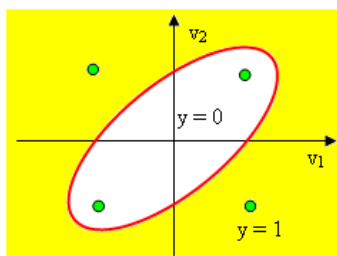


Рис. 4.8. Таблица состояний логической функции «исключающее ИЛИ» (XOR) и ее приближение радиальным «нейроном» с эллиптической пороговой функцией

Для того, чтобы убедиться в невозможности приближения персептроном нелинейной логической функции «исключающее ИЛИ», запишем ее состояния системой неравенств, соответствующей персептрону:

$$\begin{cases} -1 \cdot \mu_1 - 1 \cdot \mu_2 - \theta < 0; \\ -1 \cdot \mu_1 + 1 \cdot \mu_2 - \theta > 0; \\ +1 \cdot \mu_1 - 1 \cdot \mu_2 - \theta > 0; \\ +1 \cdot \mu_1 + 1 \cdot \mu_2 - \theta < 0. \end{cases} \quad (4.6)$$

Вычитая из первого уравнения системы (4.6) второе, а из третьего четвертое, получим несовместную систему

$$\begin{cases} -2 \cdot \mu_2 < 0; \\ -2 \cdot \mu_2 > 0. \end{cases} \quad (4.7)$$

Несовместность системы (4.7) означает, что персептрон нельзя обучить, так чтобы он был способен выполнять простейшую нелинейную функцию. В то же время, оставаясь в рамках линейной алгебры, мы можем воспользоваться квадратичной формой

$$e^2 = (E(\bar{v}) - \bar{v})^T \cdot R^{-1} \cdot (E(\bar{v}) - \bar{v}), \quad (4.8)$$

где R^{-1} – обратная корреляционная матрица биометрических параметров;

$\bar{v}$ – вектор нормированных по среднеквадратическим отклонениям биометрических параметров ($\sigma(v_i) = 1$).

Выражение (4.8) как раз и описывает эллипс в случае контроля двух биометрических параметров. Сравнивая вычисленное по (4.8) значение квадратичной формы с некоторым порогом Θ , получаем эллиптическую разделяющую функцию, принимающую состояние «0» и «1» (рис. 4.8).

Получается, что одиночный нейрон (персептрон) может выполнять крайне ограниченные функции линейного разделения, однако, если использовать двухслойную сеть персептронов, можно реализовать любую как угодно сложную разделяющую функцию. Например, если использовать сеть с двумя персептронами в первом слое сети и одним персептроном во втором, то можно получить приближение «исключающего ИЛИ» в виде двух параллельных прямых (рис. 4.9).

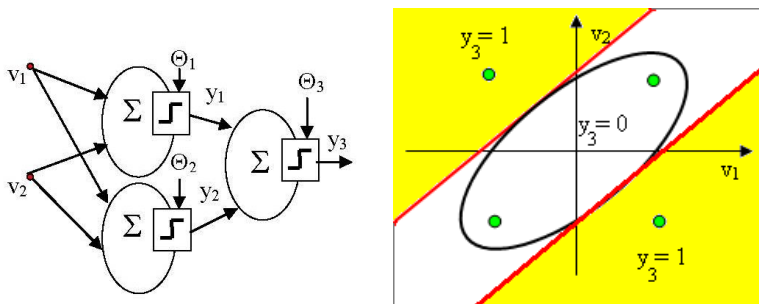


Рис. 4.9. Решение задачи «исключающее ИЛИ» двумя перцептронами в первом слое и одним перцептроном во втором

Из правой части рис. 4.9 видно, что две разделяющие плоскость прямые (двух перцептронов первого слоя) достаточно хорошо аппроксимируют функцию нечеткого «исключающее ИЛИ». Следует подчеркнуть, что приближение двухслойными перцептронами примитивных логических функций самостоятельного теоретического значения не имеет и является не более чем удобной иллюстрацией ограниченности единственного перцептрона. Это все дань традиции, заложенной М. Минским и С. Пейпертом [75]. Гораздо большее значение для теории и практики имеет приближение сетями перцептронов эллиптических разделяющих функций. На рис. 4.10 показаны два примера подобного кусочно-линейного приближения сетями с 4 и 8 перцептронами в первом слое.

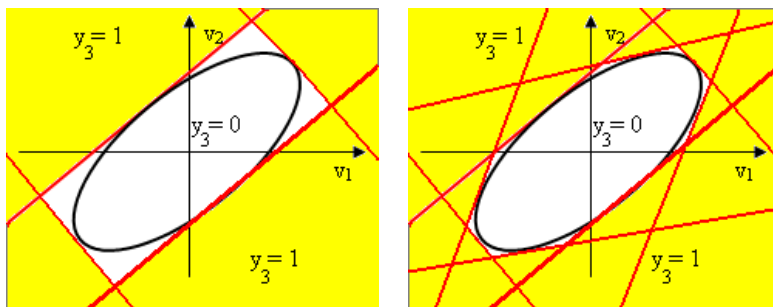


Рис. 4.10. Решение задачи приближения эллиптической решающей функции сетями с 4-мя и 8-ми перцептронами в первом слое

Чем больше персептронов в первом слое, тем более точной оказывается кусочно-линейная аппроксимация эллиптической разделяющей функции. Погрешность приближения любой нелинейной разделяющей функции оказывается пропорциональна числу нейронов в первом слое. Число воспроизводимых нелинейных разделяющих функций зависит от числа нейронов (персептронов) во втором слое. Размерность пространства приближения зависит от числа входов нейронов первого слоя.

Важнейшим преимуществом нейросетевых решающих правил является то, что воспользоваться на практике квадратичными формами высокой размерности (4.8) нельзя. Технически возможно вычислить корреляционную матрицу любой размерности по 20 примерам образа «Свой»:

$$R_2 = \begin{bmatrix} 1 & r_{1,2} \\ r_{2,1} & 1 \end{bmatrix}, R_3 = \begin{bmatrix} 1 & r_{1,2} & r_{1,3} \\ r_{2,1} & 1 & r_{2,3} \\ r_{3,1} & r_{3,2} & 1 \end{bmatrix}, \dots, \\ R_n = \begin{bmatrix} 1 & r_{1,2} & \dots & r_{1,n} \\ r_{2,1} & 1 & \dots & r_{2,n} \\ \dots & \dots & \dots & \dots \\ r_{n,1} & r_{n,2} & \dots & 1 \end{bmatrix}. \quad (4.9)$$

При этом корреляционные матрицы малых порядков от 2-го до 4-го достаточно хорошо обращаются и, пользуясь ими, можно создавать эллиптические разделяющие функции. Корреляционные матрицы (4.9) 5-го порядка и выше уже не обращаются из-за низкой точности вычисления коэффициентов парной корреляции на малой обучающей выборке. Линейная алгебра в лице квадратичных форм перестает работать.

Что касается нейросетевых приближений квадратичных форм, то они продолжают работать до любого порядка. Фактически обучение искусственных нейронных сетей по ГОСТ Р 52633.5–2011 при увеличении числа входов у нейронов осуществляется все устойчивее и устойчивее. Именно по этой

причине большие нейронные сети после их обучения работают надежнее классических алгоритмов многомерной статистики и линейной алгебры. Построить квадратичное решающее правило в виде 480-мерного гиперэллипса технически невозможно, тогда как построить аппроксимацию этого гиперэллипса 480-мерными параллелепипедами (4 персептрона с 480 входами в первом слое сети) вполне возможно.

Двухслойная сеть нейронов универсальна в совокупности со следующим за ней цифровым автоматом криптографической аутентификации. Эта совокупность позволяет создавать любые приложения биометрической защиты информации. Увеличение числа нейронов в слоях сети (в первом и во втором) является весьма и весьма эффективным техническим приемом.

4.5. Рациональный выбор длины биокода нейросетевых преобразователей

Если нам требуется для последующей криптографической аутентификации длина ключа 256 бит, то из этого условия однозначно вытекает, что нейросетевой преобразователь биометрия-код должен иметь 256 выходов. Если нейросетевой преобразователь однослойный, то уже первый слой должен иметь 256 нейронов.

Как отмечалось ранее, криптографическая защита всегда намного сильнее парольной защиты и биометрической защиты. Этот факт хорошо наблюдаем, когда изменяем число нейронов в первом слое сети. В первом приближении вероятность биометрических ошибок $P_{2,б}$ будет уменьшаться с ростом числа использованных нейронов, соответствующие кривые снижения отображены на рис. 4.11. Однако быстрый рост стойкости (быстрое снижение вероятности $P_{2,б}$) происходит не постоянно. Обычно линейное снижение вероятности $P_{2,б}$ наблюдается только на начальном участке роста числа нейронов. Далее происходит замедление роста показателя и, начиная с некоторого момента, вероятность ошибок второго рода вообще перестает снижаться.

Момент, когда происходит остановка снижения вероятности ошибок второго рода, зависит от информативности

биометрического образа «Свой». На рис. 4.11 представлены две кривые снижения вероятности ошибок биометрической составляющей аутентификации.

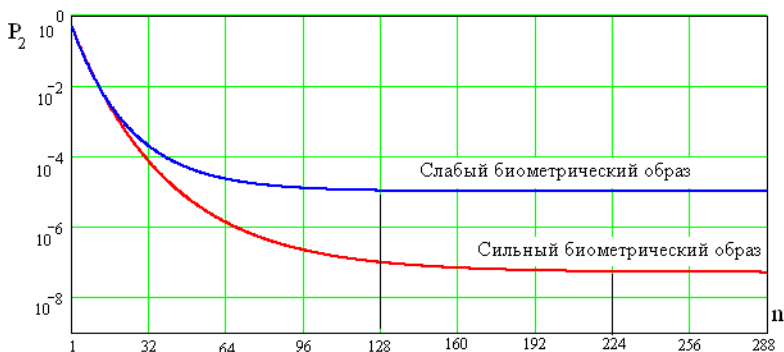


Рис. 4.11. Влияние числа выходов однослойной нейронной сети на значение вероятности ошибок второго рода

Из рис. 4.11 видно, что «слабый» мало информативный биометрический образ дает большее значение вероятности ошибок ($P_{2,б} = 10^{-4,7}$), насыщение для этого образа возникает при длине ключа 128 бит, т. е. при длине выходного кода 256 бит имеем 50 %-ю избыточность кода, которая безболезненно может расходоваться на обнаружение и исправление ошибок.

Более «сильный» биометрический образ имеет участок насыщения для кодов длиной более 224 разряда ($P_{2,б} = 10^{-7,1}$), т. е. для этого кода избыточность составит 12,5 %. Эта избыточность также может быть использована для обнаружения и исправления ошибок классическими самокорректирующимися кодами. Если этой избыточности не хватает, то следует увеличить длину выходного био-кода до необходимой.

В некоторых случаях (например, при использовании зарубежных криптографических алгоритмов) длина ключа может оказаться меньше длины до начала участка насыщения биометрико-нейросетевой защиты. Эта ситуация как раз соответствует второму случаю применения относительно «сильной» биометрии (нижняя кривая рис. 4.11) при необходимой длине ключа 128 бит. Формально можем

ограничиться 128 нейронами в первом слое нейронной сети, однако это приведет к определенным потерям показателей качества.

В случаях, когда технически выгодно иметь число нейронов больше длины выходного криптографического ключа (например, $224 > 128$), необходимо осуществить хэширование биокода. После хэширования следует обрезать хэш-функцию до нужной длины и использовать ее в качестве ключа.

В том случае, когда ключи изготавливаются вне биометрического приложения, привести значение хэш-функции к заданному значению удастся, если сложить ее по модулю два с дополнением. Если это необходимо, то хэширование биокода «Свой» выполняют с «солью»⁵. «Соль» и дополнение хэш-функции хранят открыто.

Заметим, что описанный выше прием годится не только для сокращения длины ключа, но и для увеличения его длины. Это означает, что для относительно «слабого» биометрического образа «Свой» (верхняя кривая рис. 4.11) можно обойтись сетью со 128 искусственными нейронами, увеличив далее длину биокода хэшированием со 128 до 256 бит. Подобный прием целесообразен для экономии вычислительных ресурсов. Криптографическое хэширование обычно выполняется примерно в 1000 раз быстрее, чем программное эмулирование искусственной нейронной сети.

4.6. Нейросетевая коррекция биокода вторым слоем нейронов

В том случае, если классические самокорректирующиеся коды дают слишком большие информационные потери (требуют слишком большой избыточности), ГОСТ Р 52633.5–2011 рекомендует использовать нейроны второго слоя для коррекции ошибок, допущенных нейронами первого слоя. Перед настройкой нейронов второго слоя стандарт рекомендует оценить стабильность каждого из разрядов биокода «Свой». Для

⁵ В качестве «соли» обычно используют случайное число, длина которого сопоставима с длиной применяемой хэш-функции. Применение «соли» затрудняет атакующему создание таблиц обращения хэш-функций.

этой цели подают тестовые примеры образа «Свой» и вычисляют вероятности ошибок первого рода по каждому из разрядов $P_{1,i}$. Далее вычисляют показатели стабильности каждого из разрядов:

$$\gamma_i = 2 \cdot |P_{1,i} - 0.5|, \quad (4.10)$$

где γ_i – показатель стабильности i -го разряда, принимающий значения 1.0 для абсолютно стабильных разрядов и значение 0.0 – для нестабильных разрядов с равновероятными состояниями «0» и «1».

При настройке нейрона второго слоя необходимо задавать его весовые коэффициенты пропорционально показателю стабильности (4.10). Знак весового коэффициента выбирается случайно. Входы корректирующего нейрона подключают к разрядам биокода случайно, при этом осуществляют преобразование значений разрядов биокода: состояние «0» преобразуют в состояние -1 , состоянию «1» присваивается значение $+1$. В результате перемножения случайных знаков весовых коэффициентов и случайных состояний разрядов биокода ± 1 на выходе сумматора нейрона появляется состояние, близкое к нулю.

Настройка нейрона сводится к случайным циклическим перестановкам знака у пар входов. Очевидно, что у нейрона с 64 входами максимально возможное выходное состояние составит близкое к величине $+64$, а минимально возможное значение будет близко к -64 . Если настраиваемый нейрон должен править 1 ошибку и давать состояние «1», то необходимо добиться максимального значения отклика на примеры «Свой», близкого к $+2$. Если настраиваемый нейрон должен править 1 ошибку и давать состояние «0», то необходимо добиться максимального значения отклика на примеры «Свой», близкого к -2 . Число ошибок, исправляемых нейроном, должно быть всегда примерно на единицу меньше модуля максимального отклика на примеры «Свой».

Следует отметить, что второй слой нейронов выполняет две функции. Во-первых, он правит ошибки биокода предыдущего слоя нейронов, во-вторых, сумматоры второго слоя нейронов (перемешивают) хэшируют данные кодов «Чужой».

4.7. Хэширование данных «Чужой», выполняемое нейронами второго слоя

Идеальный преобразователь биометрия-код должен полностью исключать неопределенность кодов «Свой» и максимально усиливать энтропию кодов «Чужой». Входные энтропии непрерывных данных примеров образа «Свой» и примеров образа «Чужой» сопоставимы:

$$H_{480}(\bar{v}) \approx H_{480}(\bar{\xi}). \quad (4.11)$$

После осуществления нейросетевого преобразования ситуация меняется:

$$\begin{cases} H_{256}(c) \approx 0; \\ H_{256}(x) \approx 256. \end{cases} \quad (4.12)$$

По сути дела, то, на сколько мала энтропия кодов «Свой» и на сколько близка энтропия кодов «Чужой» к предельному значению 256 бит, определяет близость реального преобразователя к идеальному.

Двухслойная нейросеть улучшает свои свойства от слоя к слою, т. е., обозначив коды первого слоя индексом 1, а коды второго – индексом 2, можно записать

$$\begin{cases} H_{256}(c_1) \rangle H_{256}(c_2) \approx 0; \\ H_{256}(x_1) \langle H_{256}(x_2) \approx 256. \end{cases} \quad (4.13)$$

Эта ситуация видна на соответствующих распределениях расстояний Хэмминга кодов «Свой» и кодов «Чужой» (рис. 4.12).

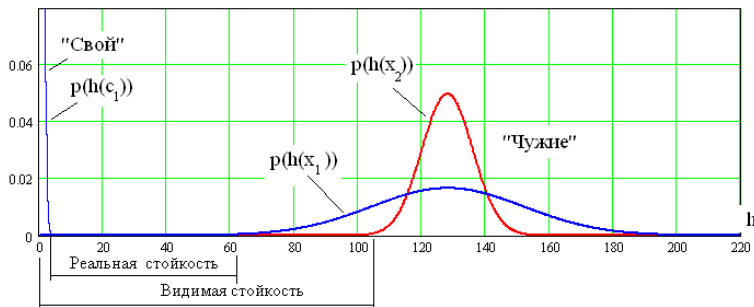


Рис. 4.12. Сжатие распределения расстояний Хэмминга на выходе нейронов второго слоя из-за хэширующих свойств нейронов

Из рис. 4.12 видно, что на выходе нейронов первого слоя биокод имеет стойкость к атакам подбора примерно 56 бит (расстояние между краями распределений расстояний Хэмминга «Свой» и «Чужие»). После корректировки ошибок нейронами второго слоя видимое расстояние между множествами составляет более чем 100 бит, однако это мнимая (завышенная) стойкость, обусловленная тем, что нейроны второго слоя не только правят биокод «Свой», но и хэшируют коды «Чужой». Хэширующие свойства второго слоя нейронов можно оценить как отношение энтропии кодов «Чужой» на выходе нейросети к энтропии биокодов на выходе нейронов первого слоя.

То, что нейроны второго слоя сети обладают достаточно сильными хэширующими свойствами, объясняется случайной расстановкой знаков весовых коэффициентов и выполняемой ими операцией округления результатов суммирования. Любая однонаправленная операция, приводящая к уменьшению длины кода, обладает некоторыми хэширующими свойствами, так как по коротким выходным кодам нельзя восстановить исходные данные. В нашем случае результаты суммирования исправляющего код нейрона могут меняться в интервале от -64 до $+64$, а его выходной код имеет только два значения «0» и «1». Присутствует операция усечения длины 9-разрядного кода до 1-го разряда.

5. ФОРМИРОВАНИЕ БАЗ БИОМЕТРИЧЕСКИХ ДАННЫХ И ТЕСТИРОВАНИЕ КАЧЕСТВА ОБУЧЕНИЯ СЕТИ НЕЙРОНОВ

5.1. Оценка вероятности возникновения ошибок первого рода на малых тестовых выборках

При тестировании преобразователей биометрия-код [76] ещё одной принципиально важной характеристикой является вероятность ошибок первого рода (ошибочный отказ в доступе донору образов "Свой"). В связи с этим базовый национальный стандарт ГОСТ Р 52633.0–2006 требует после каждого обучения осуществлять тестирование достигнутого качества принятия нейросетевых решений.

Для биометрических кодов образа "Свой" характерна предельная ситуация, соответствующая практически единичной корреляции парных разрядов биокода и нулевой энтропии:

$$\begin{cases} E(|r_{i,j}|) \approx 1.0; \\ H_{256}(c) \approx 0.0. \end{cases} \quad (5.1)$$

В этом предельном случае любимая всеми гипотеза независимости данных абсолютно не работает, все разряды выходного кода очень сильно коррелированы [77–81]. Если бы эта гипотеза работала, то мы бы имели биномиальный закон распределения значений биокодов. В связи с этим необходимо знать, к какому предельному распределению стремится биномиальный закон при выполнении условия (5.1).

Проведенные исследования показали, что предельным распределением биномиального закона является χ^2 -распределение с почти нулевым числом степеней свободы:

$$P(256, h, P_1 \rightarrow 0.0, |r| \rightarrow 1.0) \rightarrow \chi^2(\omega \approx 0), \quad (5.2)$$

где ω – число степеней свободы в первом приближении, описывающее нормированное расстояние Хэмминга, обнаруженное при первом отказе "Своему" в k -опытах:

$$\omega = \frac{h}{k}. \quad (5.3)$$

Если при проведении k опытов были обнаружены несколько ошибок с ненулевыми значениями расстояний Хэмминга, то число степеней свободы χ^2 -распределения вычисляется как среднее расстояние Хэмминга

$$\omega = E(h) \approx \frac{1}{k} \sum_{i=1}^k h_i. \quad (5.4)$$

В случае, когда в проведенной серии испытаний, состоящей из k -опытов, не обнаружен факт отказа в доступе, расчёт числа степеней свободы осуществляется исходя из гипотезы, что в следующем $(k + 1)$ испытании будет обнаружена минимальная ошибка в один бит:

$$\omega = \frac{1}{k + 1}. \quad (5.5)$$

Для того чтобы сократить время тестирования (объём тестовой выборки) от трех до 10 раз, оценку вероятности ошибок первого рода следует выполнять по следующей формуле:

$$P_1 \approx \int_1^{\infty} \frac{1}{2^{\frac{\omega}{2}} \cdot \Gamma\left(\frac{\omega}{2}\right)} \cdot x^{\frac{\omega}{2}-1} \cdot e^{-\frac{x}{2}} \cdot dx, \quad (5.6)$$

где $\Gamma(\cdot)$ – гамма функция.

Примеры оценки вероятности ошибок первого рода по формуле (5.6) приведены на рис. 5.1.

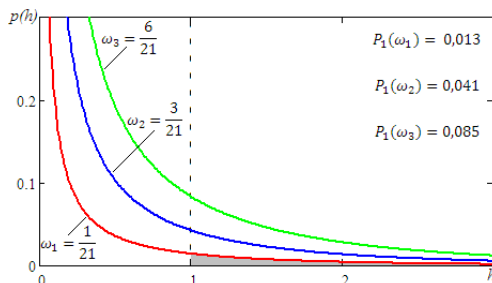


Рис. 5.1. Примеры оценки вероятностей ошибок первого рода по хи-квадрат-распределению с малыми показателями числа степеней свободы

Вычисление усредненного показателя корреляции между разрядами кодов "Свой" производится по следующей формуле:

$$r \approx \left(1 - \frac{2 \cdot E(h)}{n}\right). \quad (5.7)$$

То есть длина биометрического кода $n = 256$ начинает играть существенную роль только при оценке среднего модуля коэффициентов корреляции между биометрическими данными. Формула (5.7) является грубым приближением и работает только для достаточно больших значений модулей коэффициентов корреляции. При более точном вычислении необходимо дополнительно учитывать среднеквадратическое отклонение расстояний Хэмминга [77].

Следует подчеркнуть, что описанное выше приближение следует применять только при тестировании высокодоступных средств биометрической защиты, которые должны представлять доступ «Своему» с высокой вероятностью при первой попытке. Только в этом случае возникает проблема тестирования вероятности ошибок первого рода. Например, если задана вероятность доступа с первой попытки 0.999, то для ее корректной оценки требуется 10 000 тестовых примеров. Именно в этом случае следует использовать приближение распределения меры Хэмминга «Свой» хи-квадрат распределением с почти нулевым показателем числа степеней свободы.

Если вероятность ошибки первого рода велика (например, $P_1 \approx 0.1$), то для ее оценки вполне достаточно 20 тестовых примеров. При больших вероятностях ошибок первого рода проблемы тестирования их значений не существует. На данный момент эта ситуация характерна для большинства биометрических приложений аутентификации. Современные приложения биометрической аутентификации решают проблему высокой доступности за счет увеличения числа разрешенных попыток предъявления образа «Свой» во время одного сеанса. Так, если по политикам обеспечения информационной безопасности пользователю разрешено использовать три попытки предъявления биометрического образа «Свой», то вероятности получить доступ с 1, 2, 3 попыток составят $P_{1,1} = 0.1$, $P_{1,2} = 0.01$, $P_{1,3} = 0.001$. Предоставление пользователю трех попыток в место одной поднимает вероятность доступа с 0.9 до 0.999.

5.2. Проблемы оценки вероятности ошибок второго рода, необходимость формирования больших баз биометрических образов

После разработки и введения в действие стандарта ГОСТ Р 52633.5–2011 [82] открылась возможность производства разнотипных нейросетевых преобразователей биометрия-код разными участниками рынка средств защиты информации. При этом потребитель таких средств должен быть уверен в качестве того или иного средства биометрической защиты его личной информации. Если такое средство будет сопряжено с внешним средством криптографической защиты информации (СКЗИ), то потребитель должен быть уверен в высоком качестве самой биометрии, СКЗИ, а также их стыка.

Исследование кодов программной части и сертификация средств СКЗИ являются хорошо отработанными. Напротив, процедуры исследования кодов программной части и сертификация средств высоконадежной биометрической защиты информации на данный момент слабо изучены. Одной из основных причин такого положения является различная информативность биометрических образов «Свой». В СКЗИ только алгоритм шифрования и длина ключа определяют стойкость криптографической защиты к атакам подбора. Для высоконадежной биометрии стойкость к атакам подбора зависит от значительно большего числа факторов. Так, на нее будут оказывать влияние:

- показатель стабильности параметров биометрического образа «Свой»;
- показатель уникальности параметров биометрического образа «Свой»;
- число примеров образа «Свой», использованное при обучении;
- качество реализации алгоритма обучения;
- число биометрических параметров, учитываемых средством биометрической защиты информации (СБЗИ);
- длина выходного кода СБЗИ;

– алгоритм самошифрования или хэширования, использованный производителем для защиты параметров СБЗИ от внешнего наблюдателя.

Корректный аналитический учет столь большого числа взаимовлияющих факторов на сегодняшний день невозможен. В связи с этим возрастает роль организации практического тестирования уже обученного нейросетевого преобразователя биометрия-код. К сожалению, это направление исследований также оказывается сопряжено со значительными материальными затратами на подготовку тестовых баз биометрических образов больших размеров.

Проиллюстрируем вышеизложенное примером тестирования преобразователя биометрия-код со стойкостью к атакам подбора порядка 1 000 000 000 попыток. Для этой цели в идеальном случае нужно иметь тестовую базу, состоящую из 1 000 000 000 примеров случайных образов «Чужой» [83, 84]. Если преобразователь биометрия-код работает с рукописными образами, то и тестовая база должна иметь 10^9 рукописных образов «Чужой». На написание одного такого рукописного образа требуется порядка 5 секунд, т. е. для написания миллиарда рукописных образов одним человеком потребуется 5×10^9 с или 160 лет непрерывного письма. Время утроеится до 480 лет, если пишущий будет работать только по 8 ч в сутки. При привлечении к работе 1000 человек необходимую базу тестовых образов удастся собрать примерно за полугодие.

Непрерывная работа коллектива из 1000 человек в течение полугода требует слишком высоких материальных затрат. В связи с этим в ГОСТ Р 52633.2–2010 [85] рекомендуется дополнять базы естественных биометрических образов синтетическими биометрическими образами, полученными из естественных. Практика показывает допустимость 1000-кратного увеличения баз биометрических образов за счет добавления к ним синтетических образов, полученных морфингом пар естественных биометрических образов-родителей и мутациями.

Блок-схема подобного тестирования отображена на рис. 5.2.



Рис. 5.2. Блок-схема тестирования по классической схеме перебора биометрических образов «Чужой» с хранением полной базы из миллиарда образов

Казалось бы, что 1000-кратное снижение затрат на формирование баз биометрических образов должно решить проблему и осуществить тестирование по стандарту [76], однако это не так. На хранение 10^6 естественных биометрических образов потребуется 32 000 Гб памяти. На сегодняшний день сервер с такой памятью не является экзотикой, однако только процедура последовательного чтения всей этой памяти и последующая нейросетевая обработка извлеченной информации будут занимать порядка нескольких недель машинного времени. В связи с этим целесообразно предпринять специальные меры, позволяющие решать задачу тестирования с меньшими затратами вычислительных ресурсов и времени.

Необходимость в отказе от классических методов полного перебора «грубой силой» связана еще и с тем, что такой метод хорош только для дискретной криптографии и не гарантирует достоверной оценки для биометрии. Поле возможных состояний дискретного криптографического ключа упорядочено, конечно, и имеет равномерную плотность, т. е. при этих условиях корректно рассматривать задачу его полного перебора. Поле

возможных положений непрерывных биометрических образов «Свой» бесконечно и неоднородно по плотности расположения правдоподобных образов. Для того чтобы осуществить полный перебор поля, необходимо:

- упорядочить между собой биометрические образы «Чужой»;
- иметь гарантии того, что тестовая база не имеет аномального прореживания плотности расположения образов «Чужой» в районе отыскиваемого образа «Свой»;
- иметь надежные мягкие индикаторы, достоверно отражающие факт приближения образов «Чужой» на малое расстояние к образу «Свой».

Ни одна из перечисленных выше задач сегодня не решена и, как следствие, организовать надежный перебор полной базы биометрических образов «Чужой» сегодня нельзя. Применительно к криптографии для организации полного перебора ключа необходимо иметь только достаточно мощные вычислительные средства. Применительно к биометрии тайных образов «Свой» недостаточно только наличия вычислительных средств приемлемой мощности: необходимо иметь правильно сформированную и достаточно большую базу биометрических образов и иметь отсутствующие на данный момент хотя бы удовлетворительные решения перечисленных выше трех задач.

Следует отметить, что рано или поздно удовлетворительные решения перечисленных выше трех задач будут найдены, и в этом случае численный эксперимент по прямому «силовому» перебору полной тестовой базы может быть осуществлен. Естественно, что подобный численный эксперимент потребует достаточно больших материальных затрат и затрат времени, однако эти затраты в конечном итоге будут оправданны. Только прямой численный эксперимент может являться самым надежным подтверждением того, что биометрическая защита личной информации тайным образом «Свой» в исполнении той или иной фирмы действительно является высоконадежной. Только прямой «силовой» перебор полной тестовой базы биометрических образов «Чужой» может подтвердить или опровергнуть те или иные упрощенные методики оценки стойкости преобразователей биометрия-код к атакам подбора.

5.3. Размножение примеров биометрических образов через их скрещивание

Так как создать базу естественных биометрических образов достаточно больших размеров технически сложно, необходимо иметь технологию, позволяющую формировать некоторое ядро базы естественных биометрических образов и при необходимости размножать его до нужных размеров. Далее будем понимать под размножением процедуру линейного копирования параметров двух биометрических образов-родителей в параметры биометрических образов-потомков. Такой подход заимствован из биологии и получил условное название «генетические алгоритмы» [77, 85]. На рис. 5.3 иллюстрируется технология генетического размножения биометрических образов-потомков путем скрещивания двух образов-родителей, представленных векторами примеров биометрических параметров $\bar{v}_i, \bar{\xi}_i$.

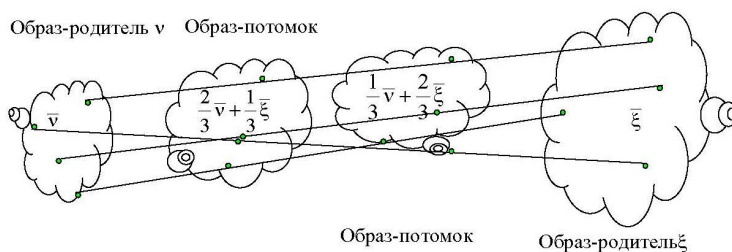


Рис. 5.3. Размножение морфингом примеров двух биометрических образов-родителей

Предположим, что изначально имеется по 20 примеров образа-родителя v , представленного векторами k биометрических параметров $\bar{v}_1, \bar{v}_2, \bar{v}_3, \dots, \bar{v}_{20}$ и примерно столько же примеров образа-родителя ξ , представленного векторами тех же k биометрических параметров $\bar{\xi}_1, \bar{\xi}_2, \bar{\xi}_3, \dots, \bar{\xi}_{21}$. Необходимо получить от двух биометрических образов-родителей по два образа-потомка. Если синтезировать равноудаленные между

собой и ближайшим родителем образы-потомки, то при синтезе первого образа-потомка необходимо использовать 2/3 части примера образа v и 1/3 часть примера образа ξ . То есть биометрические параметры примеров образа-потомка, наиболее похожего на первый образ-родитель, вычисляются следующим образом:

$$v'_{i,k} = \frac{2}{3}v_{i,k} + \frac{1}{3}\xi_{j,k}, \quad (5.8)$$

где i – номер примера образа родителя v и одновременно номер примера образа-потомка;

k – номер учитываемого биометрического параметра;

j – номер примера образа родителя ξ .

Аналогично вычисляются параметры примеров векторов второго образа-потомка, наиболее близкого к образу-родителю ξ :

$$\xi'_{j,k} = \frac{2}{3}\xi_{j,k} + \frac{1}{3}v_{i,k}. \quad (5.9)$$

Если недостаточно синтеза двух образов-потомков, то можно увеличить их количество, разбив расстояние между параметрами образов-родителей на большее число частей. На рис. 5.4 приведен пример деления расстояний между биометрическими параметрами двух образов на 11 частей для получения 11 образов-потомков.

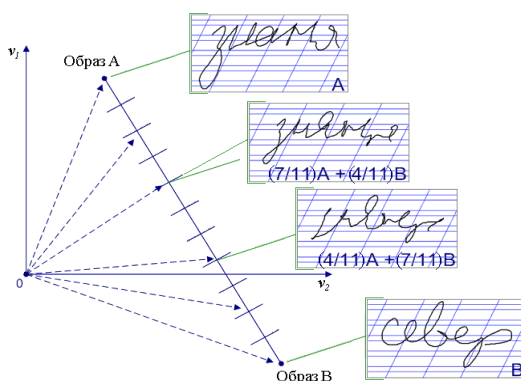


Рис. 5.4. Пример размножения морфингом двух экземпляров биометрических рукописных образов «Знамя» и «Север»

Описанным выше способом можно синтезировать очень большое число примеров образов-потомков. Каждая из возможных пар примеров образов-родителей будет давать разные примеры образов потомков. В частности, 20 примеров образа родителя v порождают по 20 примеров каждого из образов-потомков при использовании всего одного примера образа родителя ξ (всего имеем 21 пример образа родителя ξ), т. е. при необходимости можно получить до $20 * 21 = 420$ примеров каждого двух образов-потомков.

Принципиально важным свойством морфинг-размножения образов-потомков через размножение их примеров является то, что они наследуют от своих образов-родителей все статистические моменты. В частности, происходит наследование векторов математических ожиданий, векторов среднеквадратических отклонений и корреляционных матриц взаимосвязей между биометрическими параметрами. Все эти статистические моменты являются крайне важными характеристиками и без их воспроизведения нельзя быть уверенным в корректности синтезированных биометрических образов.

Технологический прием размножения скрещиванием примеров биометрических образов выгоден не только при формировании образов-потомков. Этот прием оказывается весьма эффективным и при обучении искусственных нейронных сетей на малых обучающих выборках. Обычно пользователи биометрических систем во время их обучения должны предъявить от 12 до 16 примеров биометрического образа «Свой». Это позволяет устойчиво обучать нейроны, имеющие от 12 до 16 входов. Для того чтобы устойчиво обучать нейроны с большим числом входов, необходимо увеличить размер обучающей выборки.

Если пользователь отсутствует или негативно воспринимает необходимость воспроизводить дополнительные биометрические образы, то вполне возможно использование процедуры размножения примеров биометрического образа «Свой» через их морфинг.

Иллюстрация процедуры размножения примеров образа «Свой» дана на рис. 5.5, где кружками отмечены четыре

исходных биометрических примера, а треугольниками помечены положения синтезированных примеров. Алгоритм синтеза дополнительных примеров образа «Свой» состоит в поиске двух примеров с номерами j и m , имеющих между собой максимальное Евклидово расстояние между векторами их биометрических параметров:

$$\max \sqrt{\sum_{i=1}^k (v_{j,i} - v_{m,i})^2} \text{ для } j \neq m. \quad (5.10)$$

Из этого рисунка видно, что малое число неизвестных параметров подбирается достаточно быстро: всего за несколько минут машинного времени. Увеличение числа

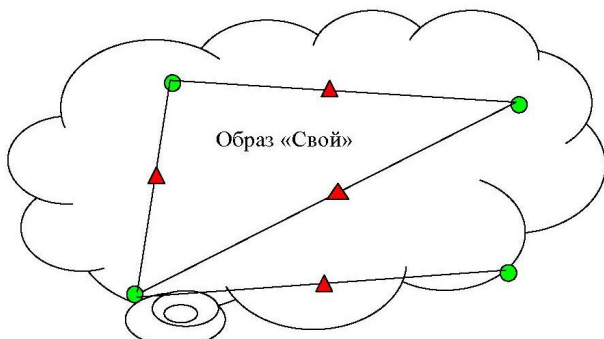


Рис. 5.5. Размножение морфингом примеров образа «Свой»

подбираемых входов приводит к росту времени подбора и числу необходимых на подбор примера образа «Свой», дополняющего уже имевшиеся ранее примеры. Для каждой дополненной новым примером выборки вновь находится пара наиболее непохожих друг на друга примеров, и эти примеры-родители скрещиваются в одинаковой пропорции.

5.4. Синтез базы биометрических образов, равномерно заполняющей многомерное пространство «Все Чужие»

При формировании баз естественных биометрических образов обычно используются те данные, которые есть (которые удалось собрать). Если нужно дополнить базу, то проще всего это делать, случайно выбирая людей и получая от них случайные образы (это вполне возможно в случае сбора биометрии голосовых паролей или рукописных паролей). Вполне естественно, что гарантировать равномерное заполнение многомерного пространства «Все Чужие» случайно полученными биометрическими образами «Чужой» нельзя. То есть перед тем, как пользоваться базой для тестирования, необходимо как минимум выровнять многомерную базу, заполнив ее пустоты. На рис. 5.6 дана двухмерная проекция «плохо» заполненной базы образов «Чужие»:

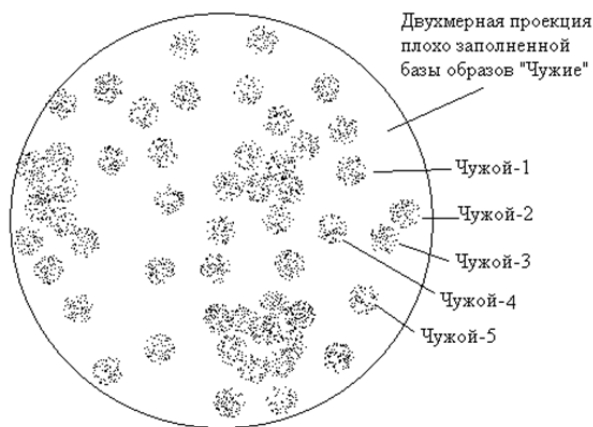


Рис. 5.6. Неравномерно заполненная тестовая база биометрических образов «Чужие»

Для того, чтобы оценить неравномерность заполнения базы, достаточно взять любой обученный нейросетевой преобразователь биометрия-код «Свой», далее следует подать

на его входы все образы выравниваемой базы. При этом хорошо заполненные фрагменты многомерного пространства «Чужие» должны давать равновероятные состояния «0» и «1» в соответствующих разрядах выходного кода. Напротив, участки многомерного пространства с плохой балансировкой биометрических данных должны давать существенную неравномерность состояний «0» и «1» в соответствующих разрядах биокода. На рис. 5.7 приведен пример распределения вероятности появления состояния «1» в разрядах биокода, полученный просмотром всех образов базы «Чужой».

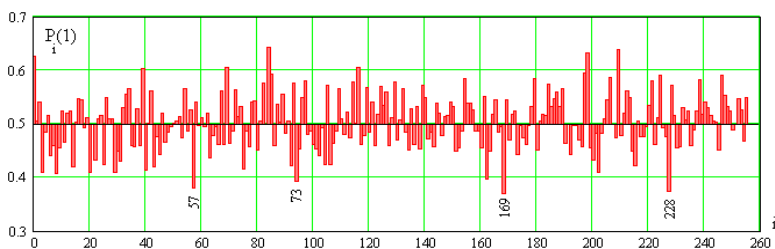


Рис. 5.7. Распределение вероятности появления состояния «1» в разрядах биокода, полученного просмотром всех образов базы «Чужой»

Из рис. 5.7 видно, что вероятности появления состояния «1» в разрядах 57, 73, 169, 228 наименьшие. Это является следствием нехватки биометрического образа, который бы давал в этих разрядах состояние «1». Для этого необходимо найти образы родители, дающие во всех перечисленных разрядах недостающее состояние «1». Далее необходимо выбрать два таких образа родителей и получить от них образ потомок. Образов родителей, обладающих нужными свойствами, достаточно много (1/32 от всех образов «Чужой»), проблем с их поиском не возникает. Синтезированный образ-потомок попадает в разреженное многомерное пространство базы «Чужие».

Очевидно, что последовательно повторяя поиск образов родителей для недостающих данных и получая от них образы-потомки, мы постепенно будем заполнять разреженности в базе образов «Чужие». Описанную выше процедуру следует проводить

до момента, пока база тестовых биометрических образов «Чужие» не будет иметь нужные размеры. Для того что бы исключить возможные неоднородности, порождаемые выбранной нейронной сетью «Свой», необходимо при заполнении базы использовать несколько обученных искусственных нейронных сетей «Свой-1», «Свой-2», «Свой-3»,...

Если требуется создавать очень большие базы тестовых биометрических образов, то равномерности состояний «0» и «1» в каждом разряде оказывается недостаточно. В этом случае необходимо переходить к контролю равномерности появления состояний в произвольно выбранных парах разрядов. Следует контролировать равновероятность появления состояний «00», «01», «10», «11». При этом размножать нужно родителей, являющихся носителями недостающих состояний.

5.5. Быстрая оценка стойкости к атакам подбора по усредненным модулям парных корреляционных связей

Вполне очевидным является то, что полная независимость (некоррелированность) биометрических кодов «Чужой» дает последовательности кодов «белым шумом». Такие последовательности будут иметь энтропию, точно совпадающую с длиной биометрического кода:

В ином предельном случае, когда все разряды кодов абсолютно коррелированы, энтропия кодов «Чужой» оказывается единичной, независимо от длины кодов:

$$\begin{cases} E(|r_{ij}|) = 1; \\ H(n) = 1. \end{cases} \quad (5.11)$$

В общем случае энтропия выходных кодов «Чужой» описывается трехмерной функцией

$$H(n, E(|r|), E(P_{2,i})) = (\beta(n, E(|r|), E(P_{2,i})) \cdot (n-1) + 1) \cdot H(1), \quad (5.12)$$

где $H(1)$ – среднее значение одномерной энтропии всех разрядов выходного кода;

$P_{2,i}$ – вероятность ошибки второго рода для i -го разряда кода «Чужой».

Форма описания (5.12) удобна тем, что содержит почти мультипликативный трехмерный коэффициент связи многомерной энтропии и среднего значения одномерных частных энтропий по каждому из выходных разрядов биометрического кода. Для наиболее важного случая $E(P_{2,i}) = 0.5$ удалось построить номограмму двухмерного почти мультипликативного коэффициента связи [77, 86]. Однако этот подход оказался не продуктивен, так как кривые подобных номограмм описываются очень сложными аналитическими зависимостями.

Возможен еще один путь описания многомерной энтропии, которая строится, опираясь на так называемые многомерные корреляционные функции. Этот тип функций вводится через отношение многомерных функций энтропии:

$$R(n) = \left\{ 1 - \frac{H(n)}{n \cdot H(1)} \right\} = \left\{ 1 - \frac{H(n)}{\sum_{i=1}^n H_i(1)} \right\}, \quad (5.13)$$

где $H_i(1)$ – одномерная энтропия i -го разряда.

Обычная двухмерная корреляция r совпадает с $R(2)$. Для более высоких значений размерности n и $R(n)$ связаны между собой номограммой, приведенной на рис. 5.8. Номограмма построена по данным, приведенным в табл. 5.1.

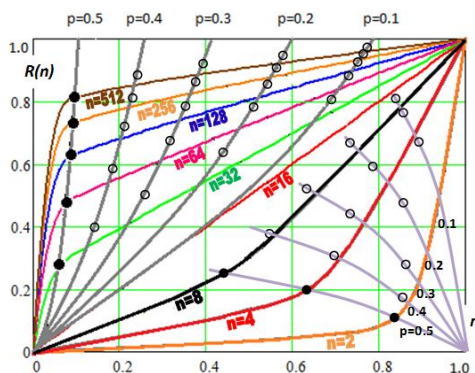


Рис. 5.8. Номограмма связи значений высокоразмерных корреляций $R(n)$ с усредненными значениями модулей обычных корреляционных функций r

Таблица 5.1
Значения кривых номограммы рис. 5.8 для параметра $P = 0.5$

Значения многомерной корреляции $R(n)$												
		Двухмерная корреляция r										
		0,00	0,10	0,20	0,30	0,40	0,50	0,60	0,70	0,80	0,90	1,00
Размерность n	2	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,09	0,30	1,00
	4	0,00	0,03	0,05	0,08	0,11	0,13	0,18	0,32	0,53	0,76	1,00
	8	0,00	0,06	0,10	0,15	0,23	0,33	0,43	0,57	0,71	0,86	1,00
	16	0,00	0,10	0,21	0,32	0,42	0,51	0,62	0,71	0,82	0,90	1,00
	32	0,00	0,32	0,39	0,47	0,55	0,62	0,70	0,77	0,85	0,92	1,00
	64	0,00	0,50	0,56	0,62	0,67	0,73	0,78	0,84	0,89	0,95	1,00
	128	0,00	0,63	0,68	0,72	0,77	0,81	0,84	0,88	0,92	0,96	1,00
	256	0,00	0,74	0,77	0,81	0,84	0,86	0,89	0,92	0,95	0,97	1,00
	512	0,00	0,81	0,84	0,87	0,89	0,91	0,93	0,95	0,96	0,98	1,00

Из рис. 5.8 видно, что кривые связи высокоразмерных корреляций $R(n)$ с усредненными значениями модулей обычных корреляционных функций r имеют значительные почти линейные участки. Это связано со смысловой однородностью связываемых второй номограммой функций. Причина сложности номограммы энтропия-корреляция в том, что она связывает разнородные параметры [77]. Из рис. 5.8 видно, что значительные участки кривых фактически образованы почти линейными функциями (кривые хорошо описываются кусочно-линейным приближением со сплайн-стыковкой линейных фрагментов). Крайне удобным является так же то, что сплайн-сшивка линейных участков находится на параболе (точки отмечены темной заливкой на рис. 5.8). Для каждого из дополнительных условий $P = 0.4$, $P = 0.3$, $P = 0.2$, $P = 0.1$ получаются другие параболы и иные точки сплайн-сшивки. Эти данные приведены в табл. 5.2.

Таблица 5.2

Координаты точек сплайн сшивки линейных участков,
находящиеся на разных параболах

Координаты (r ; R(n))						
		Параметр р				
		0.5	0.4	0.3	0.2	0.1
Размерность n	2	(0.84;0.12)	(0.86;0.18)	(0.85;0.28)	(0.86;0.48)	(0.88;0.68)
	4	(0.63;0.21)	(0.71;0.32)	(0.73;0.43)	(0.78;0.61)	(0.86;0.77)
	8	(0.44;0.25)	(0.54;0.38)	(0.63;0.52)	(0.73;0.66)	(0.85;0.82)
	32	(0.05;0.28)	(0.14;0.40)	(0.26;0.50)	(0.43;0.63)	(0.68;0.80)
	64	(0.06;0.48)	(0.18;0.59)	(0.32;0.67)	(0.51;0.78)	(0.72;0.89)
	128	(0.07;0.63)	(0.22;0.71)	(0.35;0.78)	(0.54;0.84)	(0.75;0.92)
	256	(0.08;0.73)	(0.23;0.81)	(0.38;0.86)	(0.56;0.91)	(0.76;0.96)
	512	(0.09;0.82)	(0.24;0.89)	(0.39;0.92)	(0.57;0.95)	(0.77;0.98)

Следует отметить, что все координаты точек сшивки лежат на двух типах парабол. Для размерностей выше 16 параболы строятся из общей точки, расположенной в левом нижнем углу рис. 5.8 (точка $R(n) = 0, r = 0$). Для низких размерностей параболы имеют общую точку в правом нижнем углу рис. 5.8 (точка $R(n) = 0, r = 1$). Данные первой и второй парабол сведены в табл. 5.3.

Таблица 5.3

Параболы, описывающие положение точек сшивки линейных
участков

Полиномы многомерной корреляции R(n)		
Параметр р	Двухмерная корреляция r	
	Размерность n=32, 64, 128, 256, 512	Размерность n=2, 4, 8
p=0.5	$56.19r^2 + 1.59r$	$-0.42r^2 + 0.15r + 0.27$
0.4	$5.86r^2 + 2.12r$	$-1.19r^2 + 0.98r + 0.20$
0.3	$2.54r^2 + 1.01r$	$-1.02r^2 - 0.11r + 1.13$
0.2	$1.47r^2 + 0.58r$	$-4.17r^2 + 4.51r - 0.33$
0.1	$0.83r^2 + 0.59r$	$-62.5r^2 + 112.5r - 50$

Из всего вышеизложенного следует, что размерность $n = 16$ играет особую роль, она имеет линейную связь между 16-мерной и обычной 2-мерной корреляцией:

$$R(16) = r. \quad (5.14)$$

При размерностях до и после 16-я энтропия и корреляции имеют связь по разному вогнутые (выпуклые). Размерность $n = 16$ – граница, разделяющая просто многомерные преобразования и высокоразмерные преобразования (точное значение границы $n = 15.471$).

Преимущество изложенного выше подхода состоит в значительном ускорении времени вычисления многомерной энтропии выходных биометрических кодов. Кроме того, этот метод многомерных корреляций может быть использован для оценки энтропии естественных языков (русского, английского и казахского) [77, 87]. В частности, данный метод позволяет оценивать энтропию до 256 букв, связанных (зависимых) длинных кодов текстов естественных языков. При использовании метода удастся добиться ускорения вычислений в 10^{104} раз (десять в 104-й степени) при вычислении энтропии русского языка для последовательностей из 256 рядом стоящих букв (знаков). Это следствие того, что нет необходимости ждать появления очень редких событий. Номограммы, представленные выше, позволяют предсказывать появление редких событий, анализируя малые объемы кодов. Для вычисления среднего значения модулей корреляции и показателя вероятности угадывания достаточно малых выборок, содержащих от 200 до 10 000 кодов.

6. УГРОЗЫ СРЕДСТВАМ БИОМЕТРИЧЕСКОЙ ЗАЩИТЫ ДОСТУПА К ИНФОРМАЦИИ, НЕ ИМЕЮЩИМ КРИПТОГРАФИЧЕСКОЙ ЗАЩИТЫ БИО-ДАННЫХ И АТАКИ, РЕАЛИЗУЮЩИЕ ЭТИ УГРОЗЫ

6.1. Угроза модификации программы защиты информации атакой на «последний» бит

Обычно средства не криптографической защиты информации делятся на программные и аппаратные. Аппаратные средства не криптографической защиты информации считаются более надежными, так как программное обеспечение, размещенное в них, недоступно для исследования и модификации.

Совершенно иная ситуация возникает, когда средства защиты доступа к информации, в том числе средства биометрической защиты, выполнены программно. Программная защита может быть скопирована и далее возможно ее исследование и модификация. Хакеры атакуют программную защиту обычно только в том случае, когда ею пользуется множество людей и появляется возможность тиражировать успех после взлома программной защиты. Самым простым способом взлома программной защиты является поиск «последнего» бита решающего правила. Пример реализации биометрико-нейросетевой защиты с «последним» битом решающего правила приведен на рис. 6.1.

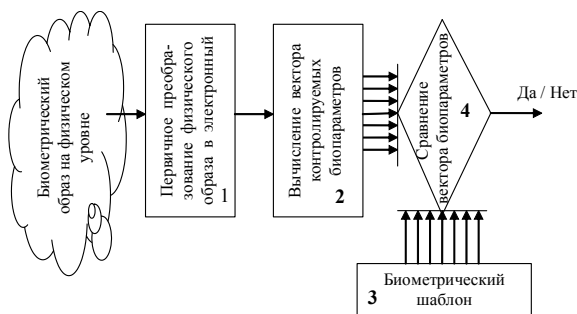


Рис. 6.1. Биометрико-нейросетевая защита, имеющая «последний» бит решающего правила

Очевидно, что достаточно инвертировать последний бит решающего правила и биометрическая защита будет выполнять обратную функцию (будет пропускать «Чужих» и останавливать «Своего»). Возможны и более изощренные варианты атаки модификации, когда подменяется биометрический шаблон или расширяются границы допустимых изменений биометрических параметров. Все эти атаки приводят к успеху и самое неприятное – все они легко тиражируются. Именно из-за этого программная защита считается более слабой, чем аппаратная.

Защита от модификаций биометрических программ выполняется двумя способами. Во-первых, может быть использована доверенная вычислительная среда, контролирующая перед запуском программы ее целостность. Во-вторых, вместо одного «последнего» бита может быть использовано 256 «последних» бит. В этом случае перебрать все возможные состояния 256 бит вычислительно сложно.

6.2. Угроза поиска биометрических близнецов атакой перебора большой базы образов «Чужой»

Еще одной эффективной угрозой является поиск биометрических близнецов по базе образов «Чужие». Когда мы с вами видим близнецов, мы их путаем, настолько они похожи. Совершенно в такой же ситуации оказываются биометрические автоматы, которые оказываются не способны различать биометрических близнецов. Этим могут воспользоваться злоумышленники. Зная биометрию атакуемого, они могут воспользоваться большой базой образов «Чужой» и найти атакуемому его близнеца. Эта ситуация отображена на рис. 6.2, где даны распределения образов «Чужие» (левая часть экранной формы) и распределение примеров образа «Свой» (левая часть рисунка).

Данные рисунка отражают ситуацию, характерную для обучения всех нейронных сетей классическими алгоритмами разделения двух классов. Из рисунка видно, что пороговая функция срабатывает на границе распределения «Свой», далеко от центра распределения образов «Чужие». По этой причине у атакующего появляется возможность найти в большой базе

«Чужие» образы, очень близкие к образу «Свой», являющиеся «почти» близнецами или «полными» близнецами. После выявления биометрического близнеца его можно посылать в место «настоящего» донора биометрии. Биометрический автомат не увидит разницы между биометрическими близнецами.

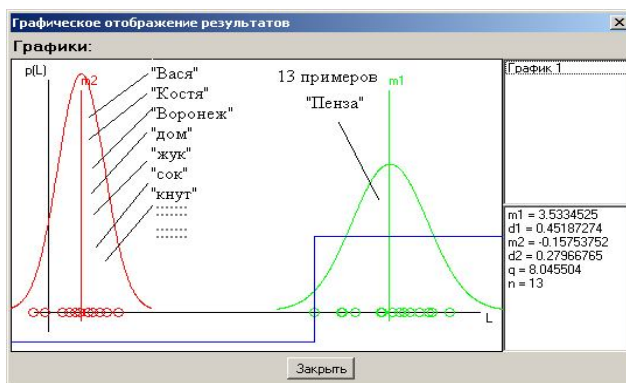


Рис. 6.2. Экранная форма вывода результатов обучения нейрона в среде моделирования «НейроПреподаватель»

Самым эффективным противодействием атаке поиска близнецов является расположение решающего правила в центре распределения образов «Чужие» в соответствии с требованиями ГОСТ Р 52633.0–2006 и ГОСТ Р 52633.5–2011. При этом атакующий, не зная кода ключа доступа, не может найти близнеца. Только если атакующий знает код ключа доступа или точно знает биометрический образ атакуемого, он сможет найти биометрического близнеца. Именно применение множества нейронов и расположение порога переключения функций квантования всех нейронов нейросети в центре множества «Чужие» является предпосылкой создания эффективной защиты от атак поиска биометрических близнецов.

6.3. Угроза перебора возможных входных и выходных состояний автомата биометрической защиты

В случае использования длинных выходных биокодов (например, длиной 256 бит) число возможных состояний кода становится велико 2^{256} . Перебирать столь большое число состояний нет смысла. Как правило, хакеры, столкнувшись с задачами криптографической сложности, вообще не атакуют защиту, а пытаются найти какую-либо иную уязвимость.

То же самое еще в большей мере относится к попыткам атаковать биометрическую защиту со стороны первичных датчиков сканирования биометрической информации. Покажем это на примере сканера рисунка отпечатка пальца, который имеет разрешение 320×480 (500 dpi) с возможностью учета 256 градаций яркости (8 разрядов). Число возможных комбинаций такого датчика составит $2^{1\,228\,800}$. Перебирать столь значительное число возможных состояний нет смысла.

На этапе преобработки при выполнении алгоритма выделения особых точек на рисунке отпечатков пальцев получается вектор с дискретными значениями 1 и 0, имеющий размерность 153 600. В этом случае общее число состояний биометрических данных уже составит $2^{153\,600}$. Число состояний значительно снизилось, однако остается недоступным для простого перебора.

Последующее двумерное преобразование дискретных данных в контролируемые непрерывные параметры сокращает их размерность до 256. Множество допустимых значений непрерывного параметра на входе искусственной нейронной сети описывается континуумом и ограничено только точностью его представления. В том случае, если на кодировку одного непрерывного биометрического параметра используется 4 бита (он может быть представлен 16 состояниями), то число комбинаций значений биометрических параметров на входе преобразователя биометрия-код составит 2^{1024} . Это значительно больше, чем декларируемая производителями биометрии стойкость к атакам подбора 10^8 или 2^{26} . Получается, что атаковать биометрию наиболее выгодно, подставляя на вход преобразователя биометрия-код 10^8 или 2^{26} биометрических образов «Чужой». Эта тактика много выгоднее перебора

состояний выходного криптографического ключа или перебора возможного состояния входных биометрических данных.

Совершенно такая же ситуация возникает и при использовании иных биометрических технологий. Например, типовое средство голосовой биометрической защиты использует парольную фразу длиной 4 с (фраза «Невероятно сильный мороз»). При её оцифровывании с частотой 8 кГц 8-разрядным АЦП получается поток из 32 000 8-битных данных. Число возможных комбинаций значений исходных данных составит $2^{256\,000}$. При извлечении 512 коэффициентов одномерного преобразования Фурье и округлении их до 4 разрядов число возможных комбинаций сократится до 2^{2048} . Это число состояний намного больше заявляемой производителем стойкости к атакам подбора 10^6 или 2^{20} . И в этом случае атаковать выгоднее, подставляя на вход преобразователя биометрия-код голосовые парольные фразы, а не случайные битовые последовательности.

6.4. Угроза перехвата биометрического образа, атака изготовления муляжа

Перехват биометрического образа с последующим изготовлением его муляжа является самой эффективной из атак. Противостоять этой атаке могут только организационно-технические мероприятия. Так пользователь, защищающий свои персональные данные рисунком отпечатка пальца, оставляет свой потожировой след на стекле сканера. Если этот след не стереть, то следующий подошедший за легальным пользователем получит его биометрические данные. Изготовление муляжа с нужным рисунком отпечатка пальца не является сложной технической задачей.

Противостоять подобному перехвату можно только заставив пользователя протирать за собой сканер папиллярного рисунка кожи пальца или приобретая сканеры с автоматическим протиранием чувствительного поля.

Еще одной организационно технической мерой является обезличивание защищаемой информации. Отпечатки пальцев вполне могут являться серьезной биометрической защитой, если атакующий не будет знать имя человека и то, какой

биометрической технологией он пользуется. Это вполне реализуемо, например, при электронных обезличенных платежах, когда провайдер скрывает IP-адрес платильщика и осуществляет транзакции со своего IP-адреса. В этом случае атакующему нужно узнать IP-адрес и проникнуть в квартиру платильщика для того, чтобы снять с его личных вещей его отпечатки пальцев. Более того, если атака не удалась нападающему, нужно поставить подслушивающее устройство для перехвата голосового пароля биометрического доступа или видеокамеру, перехватывающую графику рукописной парольной фразы.

6.5. Угроза частичной компрометации биокода из-за наличия избыточных связей в нейросети, атака Маршалко

Искусственные нейронные сети отличаются от «нечетких экстракторов» тем, что они способны обогащать исходно бедные биометрические данные, и уже обогатив их, осуществлять квантование. Объединяя по-разному одни и те же биометрические данные, можно получить огромное число нейронов. При этом часть нейронов будет иметь общие входные связи с соседними нейронами. Подобная избыточность (перекрывание входных связей у нейронов) может использоваться для компрометации биокода. Впервые этот тип атак описан Г. Б. Маршалко [88]. Суть атаки Маршалко иллюстрируется рис. 6.3.

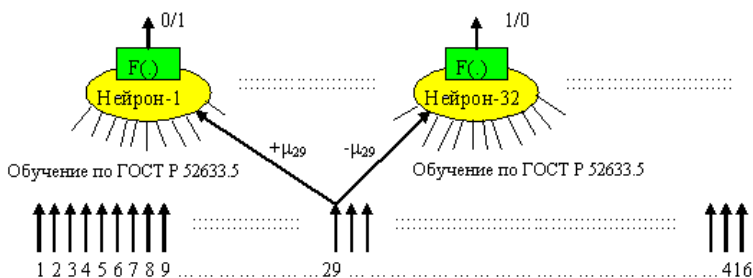


Рис. 6.3. Атака Маршалко, направленная на избыточность искусственных нейронных сетей

Из рис. 6.3 видно, что у нейрона-1 и нейрона-32 обнаружена общая входная связь (29 – вход всей сети нейронов). Так как ГОСТ Р 52633.5 дает четкий алгоритм обучения на ряду с положительными эффектами (рост устойчивости вычислений и сокращение времени обучения), возникает негативный эффект, в соответствии с которым весовые коэффициенты нейрона-1 и нейрона-32 будут одинаковы по модулю. Более того, сравнивая знаки этих весовых коэффициентов, можно сделать вывод о состояниях 1 и 32 разрядов биокода. Если знаки весовых коэффициентов одинаковы, то одинаковы и состояния 1 и 32 выходных разрядов. Если знаки весовых коэффициентов разные, то 1 и 32 разряды биокода связаны инверсией.

Численное моделирование показывает, что эффективность атаки Маршалко высока, реализуя ее, удастся связать порядка 200 бит из общего числа 256 бит ключа. То есть остаточная стойкость биокода составляет всего лишь – 56 сильно коррелированных бит. Противостоять столь эффективной атаке можно тремя способами [89]:

1. Полностью исключить перекрытие входных связей нейронов (первый Российский вариант противодействия атаке Маршалко);
2. Защитить самошифрованием данных нейронов с перекрестными связями (второй Российский вариант противодействия атаке Маршалко);
3. Частично вернуться к итерационному алгоритму обучения нейронов с общими связями и найти иные качественные нейросетевые решения, дающие знак весовых коэффициентов, обратный алгоритму ГОСТ Р 52633.5–2011 для 50 % выявленных общих входных связей нейронов (третий вариант противодействия атаке Маршалко, разработанный в КазНТУ имени К. И. Сатпаева).

6.6. Угроза частичной компрометации образа «Свой» атакой обращения матрицы нейросетевых функционалов при знании выходного кода ключа

Линейная алгебра является хорошо проверенным фундаментом. Если попытаться воспользоваться привычными

формами линейной алгебры, то можно записать следующие матричные уравнения для векторов биометрических параметров i -тых примеров образа «Свой» $\bar{v}_i$ и для биометрических параметров i -тых примеров образа «Чужой» $\bar{\xi}_i$:

$$\begin{bmatrix} \text{ПБК} \\ N \times n \end{bmatrix} \cdot \bar{v}_i = "c", \quad (6.1)$$

$$\begin{bmatrix} \text{ПБК} \\ N \times n \end{bmatrix} \cdot \bar{\xi}_i = "x_i". \quad (6.2)$$

В левой и правой частях матричных уравнений (6.1) и (6.2) вектора имеют разную природу. В правой части каждая компонента векторов – это разряды кодов, в левой части уравнения вектора имеют континуальные (непрерывные в некотором интервале) компоненты. Связывает правую и левую части матрица нелинейных нейросетевых функционалов преобразователя биометрия-код (ПБК).

Для каждого преобразователя биометрия-код матрица нейросетевых функционалов всегда конкретна, в частности, для однослойной нейронной сети функционалы этой матрицы можно записать в виде функций, выполняемых отдельными нейронами. Естественно, что при этом связи, существующие внутри нейронной сети, могут быть очень сложными, и они должны описываться взаимодействием множества таблиц (для описания каждого из нейронов требуется как минимум две таблицы: таблица весов входов нейрона и таблица связей входов нейрона).

О матрице нейросетевых функционалов мы знаем то, что ее настройка (обучение) намного сложнее ее применения. Мы можем оценить качество решений принимаемых той или иной матрицей обученных нейросетевых функционалов, однако во многом остальном свойства подобных нелинейных матриц оказываются плохо изученными. В частности, остается открытым вопрос о том, возможно ли осуществлять обратное преобразование матричных нелинейных уравнений (6.1) и (6.2). Можно ли, зная код «Свой» и располагая большой базой

тестовых образов «Чужие», восстановить образ «Свой» по матрице нейросетевых функционалов ПБК.

Как показали проведенные исследования, эта задача имеет не только значительный академический интерес, но также важна для практики тестирования и оценки стойкости нейросетевых преобразователей биометрия-код. В принципе решение этой задачи можно строить на решении линейной системы неравенств вида (4.6), построенных для каждого из нейронов нейросети, однако такой подход оказывается малоэффективен. Методы линейного программирования имеют свои ограничения и не работают для больших и сверхбольших искусственных нейронных сетей. Более эффективными оказались генетические алгоритмы отбора в пространстве расстояний Хэмминга и последующего скрещивания образов родителей. Общий смысл процедур направленной генетической селекции иллюстрируется рис. 6.4.

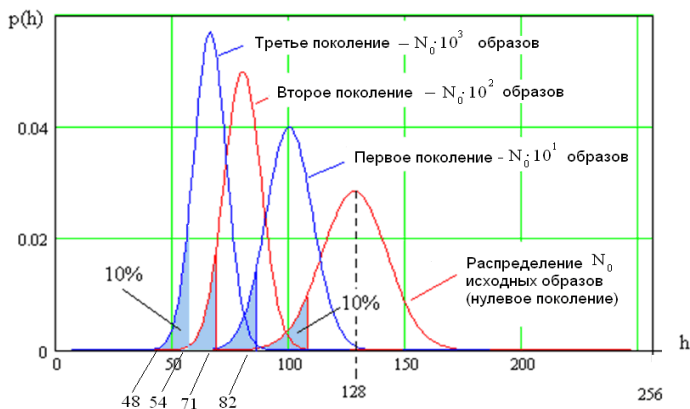


Рис. 6.4. Генетические алгоритмы направленной селекции образов-родителей наиболее близких к биометрическому образу «Свой»

Работа алгоритма начинается с того, что исходную базу образов «Чужой» рассматривают как нулевое исходное поколение селекционного материала. Далее подают на исследуемый преобразователь биометрия-код примеры образов «Чужие» и получают биокоды их откликов. Далее эти биокоды

«Чужой» сравнивают с кодом «Свой» и получают распределение расстояний Хэмминга для нулевого поколения. После этого выбирают порядка 10 % образов «Чужой», наиболее близких по расстоянию Хэмминга к образу «Свой» (на рис. 6.4 наиболее близкие расстояния отмечены темной заливкой).

Для того чтобы получить следующее поколение образов, скрещивают между собой выделенные образы, применяя мутации (вводя случайный шум в биометрические параметры скрещиваемых образов родителей). Размножение осуществляют до момента, пока численность первого поколения не достигнет первоначальной численности нулевого поколения. Далее повторяют процедуру селекции образов родителей, наиболее близких к образу «Свой» в новом поколении. Процедуру повторяют многократно, постепенно двигаясь в сторону восстановления параметров неизвестного образа «Свой». Практика показала, что подобные вычисления за несколько минут машинного времени позволяют скомпрометировать вектор биометрических параметров образа «Свой» от 80 до 90 %. То есть последнее поколение наиболее близких биометрических образов будет давать коды, совпадающие с биокодом образа «Свой» от 80 до 90 %.

Защита от этого типа атак строится путем встраивания в нейросетевую преобразователь биометрия-код механизма размножения ошибок, например, такой механизм может быть построен шифрованием данных следующих нейронов на коде ключа, сформированном из выходных состояний предшествующих нейронов. Включение механизма размножения ошибок делает наблюдатели расстояний Хэмминга «слепыми», так как любая ошибка в биокоде неверно расшифровывает данные следующих нейронов (делает все следующие нейроны не обученными).

6.7. Угроза частичной компрометации образа «Свой» атакой обращения матрицы нейросетевых функционалов путем генетического подбора и ориентира по параметру стабильности разрядов биокода

Угроза состоит в том, что выходные разряды кода «Свой» очень стабильны. По мере того, как образ «Чужой» удаляется от образа «Свой», стабильность разрядов кодов «Чужой» падает. Эта ситуация отображена на рис. 6.5.

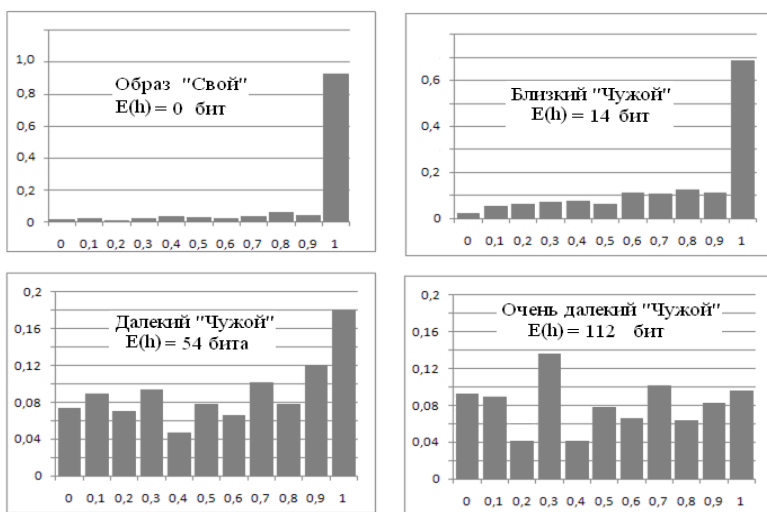


Рис. 6.5. Падение стабильности разрядов кодов «Чужой» по мере удаления образа «Чужой» от образа «Свой»

Из рис. 6.5 видно, что по гистограмме распределения показателей стабильности разрядов выходных кодов можно оценивать расстояние Хэмминга до образа «Свой». Это свойство кодов «Чужой» позволяет косвенно оценивать их расстояние Хэмминга без знания выходного биокода «Свой». В свою очередь это позволяет организовать атаку генетической селекции по частичной компрометации образа «Свой»,

ориентируясь на показатели стабильности. На рис. 6.6 дана блок-схема организации такой атаки.

В соответствии с блок-схемой рис. 6.6 подаются данные базы образов «Чужие» на исследуемый ПБК, при этом появляются выходные биокоды. Для каждого из образов «Чужой» вычисляют показатель стабильности и прогнозируют по этому показателю расстояние Хэмминга. Это открывает возможность запуска алгоритма генетического обращения матрица ПБК и частичного восстановления биометрического образа «Свой», описанного в предыдущем параграфе.

Принципиальным отличием двух алгоритмов обращения матрицы ПБК является то, что метрика стабильности разрядов биокода и метрика расстояний Хэмминга имеют разную природу. По этой причине алгоритм минимизации расстояния Хэмминга (рис. 6.4) построен на выборе данных, находящихся в левой части нормального распределения $p(h)$. Алгоритм максимизации стабильности предусматривает выбор наиболее стабильных образов-родителей в правой части нормального распределения $p(\gamma)$ (рис. 6.6).

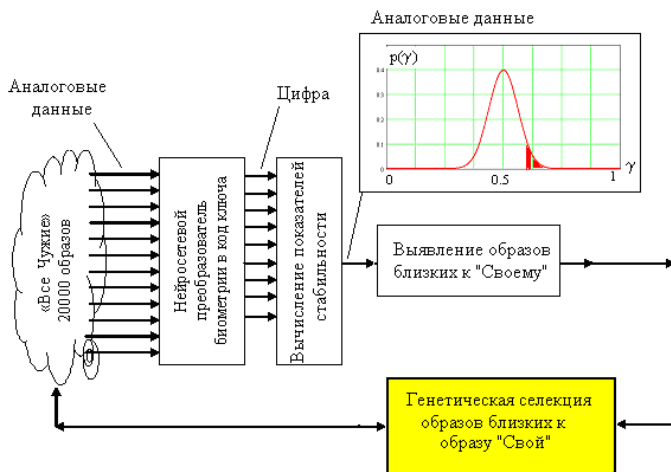


Рис. 6.6. Организация атаки генетического обращения матрицы нейросетевых функционалов и частичной компрометации образа «Свой» путем наблюдения показателей стабильности кодов «Чужой»

Еще одной особенностью генетических алгоритмов обращения матриц нейросетевых функционалов является то, что решений у этой задачи два. Восстанавливаются образы, близкие к образу «Свой» и биометрические образы, близкие к инверсии⁶ образа «Свой». В итоге после 64-х поколений генетической селекции получаются образы с высокой стабильностью, у которых плотность распределения расстояний Хэмминга имеют два всплеска, как это показано на рис. 6.7.

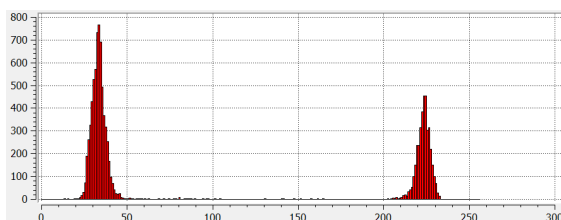


Рис. 6.7. Итоговое распределение мер Хэмминга стабильных образов, полученных на 64-м поколении генетической селекции по критерию максимизации показателей стабильности

Из рис. 6.7 видно, что во время проведения селекции по показателю стабильности необходимо сортировать образы родителей по двум группам с близкими по отношению друг к другу расстояниями Хэмминга. Все это является следствием того факта, что бинарные нейроны обладают симметрией по отношению к образу «Свой» и его инверсии (рис. 6.8).

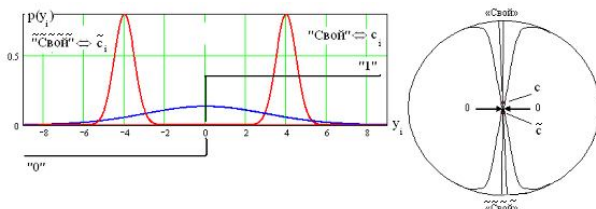


Рис. 6.8. Гиперсфера высоких значений энтропии выходных кодов ПБК

⁶ Инвертированные образы «Свой» имеют инвертированные входные биометрические данные и инвертированный выходной код.

Из-за наличия подобной симметрии гиперсфера высоких значений энтропии биокодов имеет для провала почти до нулевого значения энтропии (правая часть рис. 6.8). Наличие прямого и инверсного провалов энтропии кодов является очевидной уязвимостью, которая может быть использована для организации атаки частичного восстановления кода ключа доступа.

Защита от описанной выше атаки направленного подбора образа «Свой» и его инверсии может быть выполнена введением в состав нейросетевого преобразователя биометрия-код механизма размножения ошибок. В этом случае монотонность связи расстояния Хэмминга и показателя стабильности разрядов разрушается. При включении механизма размножения ошибок прямой и инверсный провалы в гиперсфере высоких значений энтропии превращаются в очень узкие проколы (правая часть рис. 6.8).

6.8. Угроза частичной компрометации выходного биокода и его инверсии, обладающих минимальной энтропией

В связи с существованием двух симметричных проколов в сфере высоких значений энтропии выходных кодов ПБК (правая часть рис. 6.8) возможна организация атаки по извлечению из нейросетевых контейнеров биокода ключа доступа. Подобная атака строится на том, что на входы нейросетевого преобразователя биометрия-код подаются случайные данные от генератора белого нормального шума с нулевым математическим ожиданием и среднеквадратическим отклонением, таким же как у множества «Все Чужие». На эти случайные воздействия на выходах ПБК появляются коды. Создают базу из полученных кодов случайный «Чужой» и далее вычисляют энтропию каждого из кодов по отношению к алфавиту всех других кодов.

Вычисление энтропии кодов нельзя вести по Шеннону, так как это очень затратная в вычислительном отношении процедура. Вычисление энтропии следует осуществлять быстрыми процедурами в пространстве расстояний Хэмминга.

При этих вычислениях каждый раз код, чья энтропия оценивается, должен рассматриваться как код «Свой» и относительно него должно строиться распределение расстояний Хэмминга.

В конечном итоге получаются значения распределения энтропии кодов как некоторая функция расстояний Хэмминга до действительного значения кода «Свой» и его инверсии, соответствующие данные отображены на рис. 6.9.

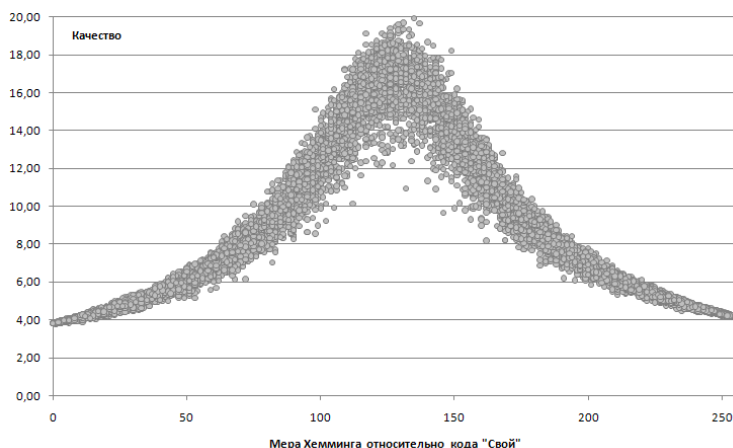


Рис. 6.9. Распределение показателей качества (показателей энтропии) для кодов, полученных случайным воздействием на входы ПБК

Из рис. 6.9 видно, что выбрав коды с минимальным значением энтропии (показателя качества), мы вновь получаем две группы кодов. Одна группа близка к коду «Свой», а вторая – к коду инверсия «Свой». Для восстановления биокода ключа доступа достаточно разделить эти коды на две группы, проинвертировать значения кодов одной из групп и вычислить наиболее вероятное значение состояний «0» или «1» в каждом разряде кода.

При этом часть разрядов будет восстанавливаться с высокой достоверностью, а часть – с меньшей. Для того чтобы найти код доступа полностью, следует перебирать разряды кода, восстановленные с низкой достоверностью. Подобная атака

позволяет восстановить (скомпрометировать) до 85 % кода «Свой», т. е. при длине биокода 256 бит перебирать потребуется примерно 40 бит. Такая остаточная стойкость почти полностью скомпрометированного биокода ключа сопоставима со стойкостью нейросетевых преобразователей биометрия-код, заявляемой ведущими производителями.

Защита от атак извлечения биокода «Свой» из нейросетевых контейнеров осуществляется введением в них механизма размножения ошибок, который подробнее будет описан в следующей главе.

6.9. Угроза частичной компрометации образа «Свой» атакой обращения матрицы нейросетевых функционалов путем подбора минимизацией взвешенной метрики Хэмминга

Следует отметить, что выделение атак по принципам их построения – это искусственный прием в виде некоторой декомпозиции задачи защиты от этих атак. В жизни все происходит иначе, атакующий использует сразу несколько вариантов атаки одновременно и идет по пути наименьшего сопротивления. Таким образом, необходимо рассматривать эффективные комбинации описанных выше атак.

Одной из наиболее эффективной является атака, строящаяся одновременно и на метрике Хэмминга и на метрике стабильности [90, 91]. Идея комбинированной метрики Андреева состоит в том, что метрика Хэмминга хорошо работает для детерминированных кодов с показателем стабильности, близким к единице ($E(\gamma) \approx 1.0$). Более того, в биометрических приложениях метрика стабильности и метрика Хэмминга дополняют друг друга. Например, для кодов «Свой» расстояние Хэмминга минимально ($h = 0.0$), а стабильность максимальна ($E(\gamma) \approx 1.0$). Для далеких кодов «Чужой» все наоборот, расстояние Хэмминга максимально ($E(h) \approx 128$), а показатели стабильности минимальны ($E(\gamma) \approx 0.0$). Пользуясь этими свойствами, можно построить меру Хэмминга, взвешенную по показателям стабильности ее разрядов. При взвешивании удобно пользоваться нормированной мерой Хэмминга

$$h_{\text{Re}} = \frac{1}{n} \sum_{i=1}^n c_i \oplus x_i \quad (6.3)$$

и ее дополнением

$$h_{\text{Im}} = \sqrt{(1 - h_{\text{Re}}^2)} = \frac{1}{n} \sum_{i=1}^n \tilde{c}_i \oplus x_i = \frac{1}{n} \sum_{i=1}^n c_i \oplus \tilde{x}_i. \quad (6.4)$$

Соответственно взвешенная метрика Хэмминга определяется как

$$\begin{cases} \gamma h_{\text{Re}} = \frac{1}{n} \sum_{i=1}^n \gamma(c_i) \cdot c_i \oplus \gamma(x_i) \cdot x_i; \\ \gamma h_{\text{Im}} = \frac{1}{n} \sum_{i=1}^n \gamma(c_i) \cdot c_i \oplus \gamma(x_i) \cdot \tilde{x}_i; \end{cases} \quad (6.5)$$

Необходимость во введении мнимой части метрики (6.4) обусловлена тем, что взвешенная метрика (6.5) должна учитывать показатели стабильности не только не совпадающих разрядов кодов, но и совпадающих разрядов сравниваемых кодов.

Практика применения взвешенной метрики Хэмминга показала ее высокую эффективность. При переходе к ней удастся примерно в полтора раза снизить длину, не поддающуюся компрометации части кода.

Защита от атак, строящихся на базе взвешенной метрики Хэмминга, создается введением в нейросетевой преобразователь биометрия-код механизма размножения ошибок.

7. КРИПТОГРАФИЧЕСКАЯ ЗАЩИТА СОДЕРЖАНИЯ НЕЙРОСЕТЕВОГО КОНТЕЙНЕРА

7.1. Классическая защита персональных данных хэшированием 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 рукой пользователя, на графическом планшете начинает играть роль уникальность почерка человека. Каждый из нас пишет буквы по-своему, при этом то, как мы пишем, сильно зависит от того, какая буква была предыдущей и какая будет следующей. Играют роль двойки и тройки букв. Если биометрический пароль всего из 5 рукописных букв скомпрометирован, то его стойкость к атакам подбора мала от 10^2 до 10^3 . Нескомпрометированный короткий рукописный пароль из 5 букв обеспечивает стойкость к атакам подбора от 10^5 до 10^{12} , все зависит от стабильности почерка и его уникальности. Сохранение в тайне биометрического пароля по своему влиянию намного 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 биокодов доступа считается безопасным при использовании специально сконструированных криптографических хэш-функций. Именно криптографические хэш-функции можно рассматривать как необратимые, т. е. их прямое вычисление занимает мало времени, а обратная задача поиска по значению хэш-функции исходных данных является практически невыполнимой (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 2^{128} попыток подбора, подобной защиты вполне достаточно для паролей и биокодов со стойкостью к атакам подбора 2^{40} . Подобная практика сегодня является общепринятой, т. е. используемая для защиты криптографическая хэш-функция должна обеспечивать показатель стойкости примерно в три раза больше, чем пароли или биокоды по отдельности или в сумме (при двухфакторной аутентификации). Если вместо криптографического хэширования используется шифрование, то соотношение между длиной ключа шифрования и стойкостью биометрии должно оставаться примерно таким же.

7.2. Криптографическая 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) [92–94]. Эта ситуация отображена на рис. 7.1.

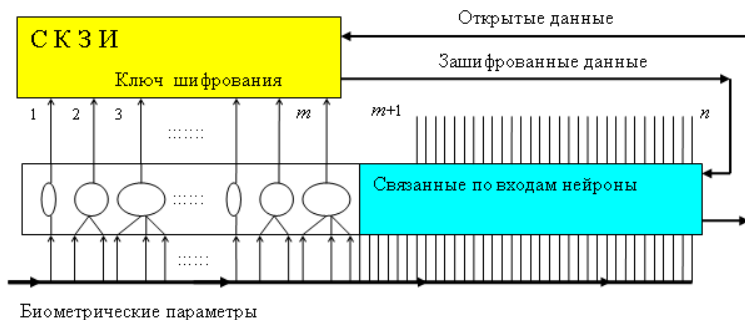


Рис. 7.1. Защита данных нейросетевого контейнера шифрованием избыточных перекрестных связей

Из рис. 7.1 видно, что данные первых m -нейронов хранятся открыто, а данные последующих зашифрованы на ключе длиной m во время подготовки защищенного контейнера. При использовании защищенного контейнера пользователь предъявляет свою биометрию, которая преобразуется в биокод длиной m и далее на этом коде проводится расшифровывание ранее защищенных шифрованием данных последующих нейронов.

Возникает естественный вопрос о том, какой 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 ключа пропорционально трехкратному 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 биопараметра 0.2 бита нам потребуется иметь в среднем по 5 входов у каждого нейрона – получается 600 входов для 600 биопараметров. Если число входов у сети нейронов только 480, то оставшиеся 220 будут повторяться у открытой части нейронов. Каждый нейрон будет связан со своими соседями примерно 2-мя входами. Два входа, перекрывающихся с соседями у каждого нейрона, вполне

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 рис. 7.2 это отображено постепенным усилением заливки (защиты данных). Открытая часть нейронов заливки не имеет.

Получается, что одним длинным ключом нельзя 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 криптографической защиты важен выбор длины дискрета приращенния 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 (смотри номограмму параграфа 5.5.) 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.

7.4. Защита данных нейросетевого контейнера с использованием криптографических ключей растущей длины от 1 бита до 256 бит

Рассмотрим предельный случай многократного блочного шифрования защищаемых данных обученной нейронной сети с ростом длины ключа в 1 бит. Открытым в такой сети остается только один (первый) нейрон, данные всех остальных нейронов оказываются защищены. Возникает иллюзия, что защита на 16 линейно растущих ключах одинакова по своей эффективности защиты на 16-битном ключе. Это действительно так, если

рассматривать только перебор состояний бинарного и 16-ричного дерева. Для биометрии это ложный результат. Если учесть, что во время атаки нужно параллельно создавать большую равномерно заполненную базу «Чужие», то выясняются преимущества «однобитного» шифрования.

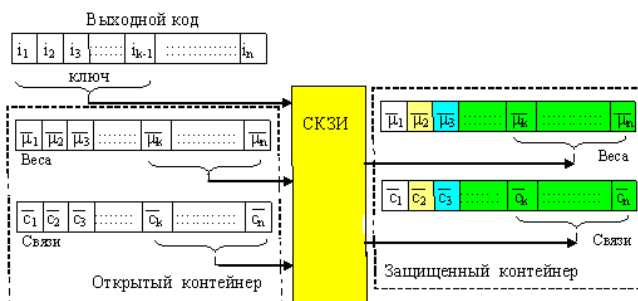


Рис. 7.3. Преобразование открытого контейнера в защищенный с шифрованием на ключах 1, 2, 3, ..., 256 бит

Для доказательства приведенного выше утверждения рассмотрим формирование бинарного дерева по возможным состояниям выходного кода защищенного контейнера и сортировке базы «Чужие» по возможным состояниям этого дерева. Пример такого бинарного дерева приведен на рис. 7.4.

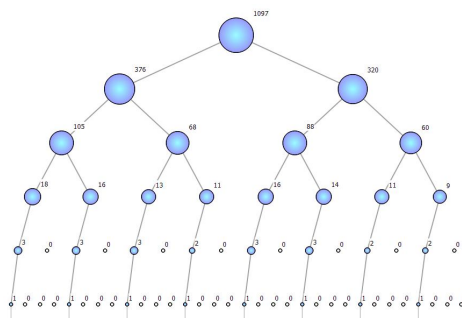


Рис. 7.4. Бинарное дерево сортировки базы образов «Чужие» по уровню стабильности 0.7 с использованием защищенного контейнера рукописного пароля «Слон»

Из рис. 7.4 видно, что, оставляя в базе «Чужой» только стабильные образы (уровень стабильности 0.7 и выше), исходная база из 1097 образов «Чужой» имеет нулевые вершины уже на третьем уровне. Далее сортировать базу и одновременно считать ее достаточно заполненной невозможно. Уже на пятом уровне нужно дополнять отсутствующие примеры образов «Чужой», синтезируя их из образов родителей (см. параграф 5.4).

Если не отбрасывать нестабильные биометрические образы (порог стабильности установим 0.3), то бинарное дерево будет гораздо более заполненным, что отображено на рис. 7.5.

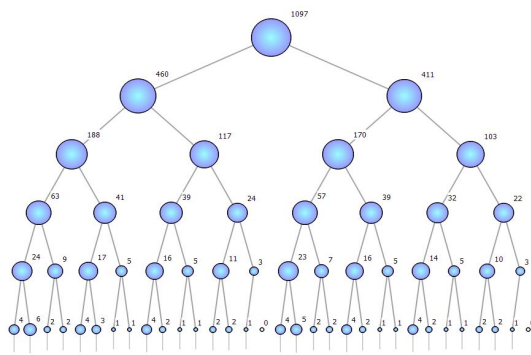


Рис. 7.5. Бинарное дерево сортировки базы образов «Чужие» по уровню стабильности 0.3 с использованием защищенного контейнера рукописного пароля «Слон»

Из рис. 7.5 видно, что сортировка нестабильных биометрических образов дает гораздо более густое бинарное дерево. Тем не менее оно также нуждается в привлечении дополнительной информации путем синтеза нужного числа образов-потомков. Важнейшим свойством приведенных выше бинарных деревьев является то, что они оказываются достаточно равномерно заполнены до некоторого уровня. Нет эффекта стягивания данных к ветви биокода «Свой». Соседние ветви (близкие к коду ветви «Свой» на расстояниях от 0 до 7) не отличаются по своим статистикам от произвольной ветви бинарного дерева.

Доказательство более высокого уровня защиты при «однобитном» шифровании сводится к простому сравнению предполагаемых деревьев сортировки большой базы биометрических образов. Если известно, что стойкость нейросетевого контейнера составляет 2^{32} и у атакующего имеется база биометрических образов 2^{32} , то атакующий при защите «однобитным» шифрованием, сортируя базу, проходит 15 лишних раз. Дополнительные затраты на простое чтение базы «Чужие» оказываются выше примерно в два раза при «однобитном шифровании». Эта ситуация отображена на рис. 7.6.

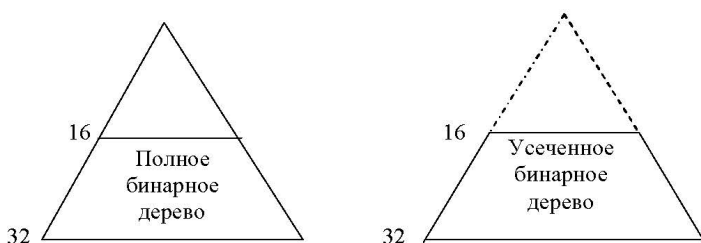


Рис. 7.6. Полное бинарное дерево (левая часть рисунка), разворачиваемое атакующим, не знающим параметров следующих защищенных 15 нейронов

Из рис. 7.6 видно, что, упорядочивая (сортируя базу Чужие») по отношению к защищенному нейросетевому контейнеру, атакующий вынужден дополнительно 15 раз осуществлять сортировку базы, восстанавливая возможные значения параметров обученных нейронов. Если считать, что чтение записи в большой базе данных наиболее длительная операция, то «однобитное» 32-кратное шифрование данных с монотонным ростом ключа от 1 бита до 32 бит потребует вдвое большего времени, чем однократное защитное шифрование на ключе 16 бит. Однако это только малозначимая линейная часть усложнения задачи.

Наиболее существенный выигрыш в уровне защищенности появляется вследствие того, что полное разворачивание 32 уровней бинарного дерева возможных состояний потребует 2^{32+1} операций проверки, сортировки и, возможно, размножения

недостающих данных. Если же мы будем иметь усеченное сверху дерево по уровню 2^{16} состояний, то злоумышленник может попытаться обойти необходимость развертывания нижележащих 16 слов бинарного дерева. Он может подобрать код неизвестного 16-битного ключа, затратив на это 2^{16} операций. При этом могут использоваться короткие проверки, например, наличие перекрестных связей у нейронов. В итоге получается, что 32-кратная криптографическая защита данных нейросети с линейно возрастающей длиной ключа выгоднее в 2^{16+1} раз однократного шифрования на ключе в 16 бит 16 уровней усеченного бинарного дерева сортировки базы образов «Чужой». Эта ситуация является переломной, так как дальнейшее увеличение длины ключа однократного шифрования 17, 18, 19, ..., 32 и более бит должно приводить к появлению интегральных эффектов высокой размерности и тем самым еще больше ослабить защиту данных обученной нейронной сети.

Получается, что для злоумышленника, атакующего биометрическую защиту, полученную 256-кратным шифрованием ключами 1,2,3,...,256 бит несколько сложнее, чем защита, полученная 32-кратным шифрованием на ключах 16, 32, 48,....., 256 бит. Рост сложности атаки должен падать экспоненциально, так как стойкость биометрии не выше, чем 2^{40} , т. е. длина ключа однократного шифрования, начиная с длины 40 бит, должна значительно повысить выигрыш от применения многократного шифрования с линейным ростом длины ключа на один бит.

7.5. Казахстанский вариант усиления стойкости шифрования за счет частичного возврата к итерационному обучению нейронов с перекрывающимися связями

Одной из самых эффективных атак на нейросетевые контейнеры сегодня является атака Маршалко. Атака строится на том, что быстрые алгоритмы обучения больших искусственных нейронных сетей строятся на почти детерминированных вычислениях весовых коэффициентов нейронов – $\mu(E(v_i), \sigma(v_i), \sigma(\xi_i))$. Именно по этой причине алгоритмы

обучения оказываются быстрыми и по этой же причине становится возможна атака Маршалко.

Очевидно, что если мы изменим тактику обучения больших нейронных сетей и после предписанных ГОСТ Р 52633.5–2011 вычислений предпримем их итерационное уточнение, то, находясь в районе стандартных решений, сможем найти множество близких решений лучше и хуже стандартного. При этом опасность заикливания решений вместо отрицательной роли будет играть положительную. Из-за внутренних шумов малого числа примеров «Свой» все итерационные алгоритмы обучения будут флюктуировать по некоторой многомерной поверхности, минимизируя влияние собственных ошибок, допущенных при вычислении производных показателя качества по каждому из регулируемых параметров. При итерационном обучении и большом числе входов у нейрона возникает естественный процесс зашумления весовых коэффициентов нейронов (рис. 4.4).

Исследования, проведенные в научно-исследовательской лаборатории биометрических методов защиты информации КазНТУ имени К. И. Сатпаева показали, что все найденные итерационным алгоритмом обучения нейрона решения не имеют точно совпадающих между собой весовых коэффициентов. Более того, от 1 до 15 % весовых коэффициентов этих решений будут иметь разные знаки. В этом случае атака Маршалко оказывается неработоспособна, т. е. можно путем нескольких запусков итерационного обучения и дообучения найти решения, у которых весовые коэффициенты общих связей нейронов будут иметь знаки, противоречащие знакам алгоритма обучения по ГОСТ Р 52633.5–2011.

Естественно, что алгоритм частичного итерационного обучения или алгоритм полного итерационного обучения нейронов сети существенно замедляют процесс обучения, однако, при этом возникает эффективная защита от атаки на избыточные связи нейронов.

Еще одним алгоритмом, разработанным казахстанскими учеными, является быстрый алгоритм принудительного искажения данных. По этому алгоритму 50 % общих связей нейронов делаются небольшими, и у них меняется знак. Далее

эти искаженные связи не меняются при обучении. Это обеспечивает получение нужного результата после первого же обучения нейрона. При этом для того, чтобы итерационный алгоритм обучения работал, изменяют случайным образом веса всех других входов нейрона до 15 % (без изменения знака), и только после этого запускается итерационный алгоритм обучения. Практика показала, что, пользуясь итерационным алгоритмом обучения, удастся вернуть ранее утраченное качество обучения примерно за 100–200 итераций. Более того, примерно 60 % новых решений оказываются лучше, чем решения, получаемые неитерационно по ГОСТ Р 52633.5–2011.

Таким образом, гибридный вариант неитерационного обучения нейросетевого преобразователя биометрия-код с его незначительным искажением и дообучением дает решения более высокого качества, абсолютно устойчивые к атакам Маршалко.

8. ЗАЩИТА ПЕРСОНАЛЬНЫХ БИОМЕТРИЧЕСКИХ ДАННЫХ НЕЙРОНАМИ С МНОГОПороГОВЫМИ ФУНКЦИЯМИ КВАНТОВАНИЯ

8.1. Переход к использованию нейронов с квантованием по двум и более порогам

В случае, когда используется квантование относительно бедных биометрических данных без их предварительного нейросетевого обогащения, большую часть биометрических данных приходится отбрасывать, так как они дают решения слишком низкого качества (для них злоумышленник легко решает обратную задачу). Только незначительная часть необогащенных биометрических данных после их бинарного квантования дает высокий показатель информативности.

В том случае, если мы до квантования осуществляем нейросетевое обогащение, ситуация меняется. Увеличивая число входов, можно существенно поднять информативность выходных данных за счет сжатия распределения данных образа «Свой». Например, можно добиться, чтобы информативность выделения образа «Свой» четной функцией составляла два или даже три бита (рис. 8.1).

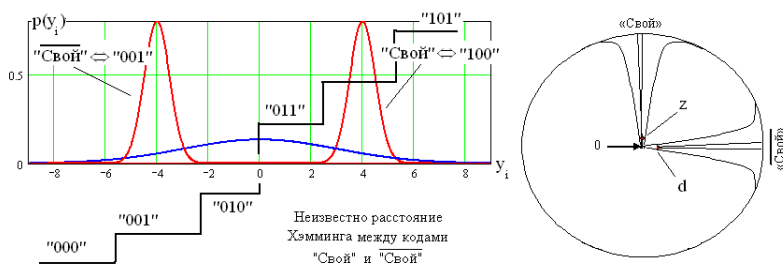


Рис. 8.1. Использование квантования выходного сигнала сумматора нейрона по 5 порогам с шестью выходными дискретными состояниями

Из рис. 8.1 видно, что диаметр распределения данных образа «Свой», оказывается примерно в шесть раз уже диаметра образов «Все Чужие». Другими словами, в большом диаметре

«Все Чужие» можно разместить шесть распределений «Свой» с меньшими диаметрами, при этом сторонний наблюдатель (после удаления данных обучения) не будет знать, в каком именно месте распределения «Чужие» будет находиться действительное распределение данных «Свой» [94–96].

Квантование по 5 порогам (по 6 выходным уровням) может осуществляться разными пороговыми функциями. В частности, на рис. 8.1 приведена простейшая монотонная функция квантования. Может быть использована гораздо более сложная не монотонная функция квантования, как это показано на рис. 8.2.

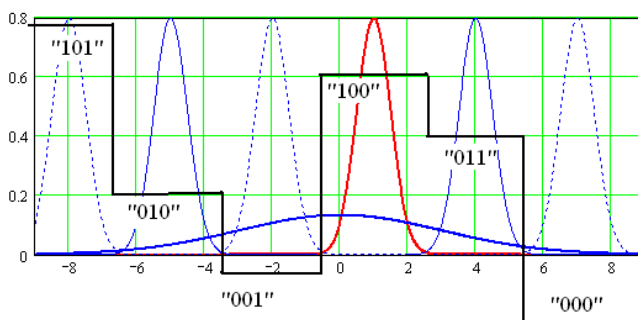


Рис. 8.2. Возможность размещения 6 распределений данных образов «Свой» внутри распределения «Все Чужие» большего диаметра

Следует отметить то, что переход от обычных бинарных квантователей к t -арным квантователям, с точки зрения защиты информации, крайне выгоден. В частности, такой переход приводит к смещению «инверсного прокола» сферы высокого уровня энтропии в некоторое отличающееся от инверсного кода положение. Это происходит из-за того, что состояний у одного нейрона становится более чем два. В частности, если бинарный квантователь для образа «Свой» будет давать состояние «1» (для образа «инверсия Свой», он будет давать состояние «0»), то для 6-арного монотонного квантователя образ «Свой» будет соответствовать трем битам ключа «100» (образ «инверсия Свой» будет соответствовать состоянию «001»). Между этими кодами нет связи инверсии. Это повышает уровень защиты информации. Атаки, созданные под бинарные нейронные сети (см. раздел 6),

оказываются неэффективными при защите данных нейронными сетями с более сложными функциями квантования.

В том случае, если образ «инверсия Свой» точно не попадает в центр одного из интервалов квантования, второй прокол сферы высоких значений энтропии вообще исчезает (рис. 8.3). Это также усиливает уровень нейросетевой защиты биометрических данных.

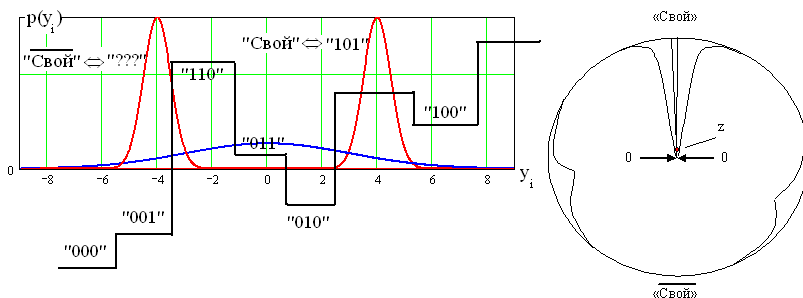


Рис. 8.3. Пример квантования выходных данных сумматора нейрона немонотонной многопороговой функцией

Из рис. 8.3 видно, что квантующая пороговая функция может быть немонотонной. Подобная немонотонность нелинейного квантующего преобразования легко реализуема на практике через применение монотонной функции (рис. 8.1) и индивидуальных таблиц случайных перестановок. Прием применения подобных таблиц усиливает уровень защищенности за счет роста хэширующих свойств преобразователя биометрия-код. Фактически проводится дополнительное табличное хэширование данных на выходе каждого из нейронов. Хэширование данных можно выполнять разными способами, в т. ч. могут использоваться хэш-таблицы случайных цифр.

Качество табличного выполнения хэширования зависит от того, сколько состояний имеет таблица перестановок и того на сколько эти перестановки случайны. Как обычно наибольший хэширующий эффект по разрушению корреляционных связей данных образов «Чужие» наблюдается для хэширующих таблиц с малым числом состояний. По мере роста числа состояний хэширующих таблиц влияние числа их состояний на

эффективность хэширования ослабляется, так как существовавшие на выходе сумматора нейрона корреляционные связи данных образов «Чужой», уже оказываются почти нулевыми.

8.2. Требования к обучению нейронов с многоуровневым квантованием выходного сигнала сумматора

Переход от использования бинарных нейронов к нейронам с многоуровневым квантованием данных [94–98] существенно меняет идеологию обучения искусственных нейронных сетей. В место одной подзадачи обучения линейной части нейрона (настройки его сумматора) возникает дополнительная задача настройки порогов многоуровневого квантователя. Эта ситуация отражена на рис. 8.4 в виде появления двух разных автоматов обучения линейной и нелинейной части нейрона.

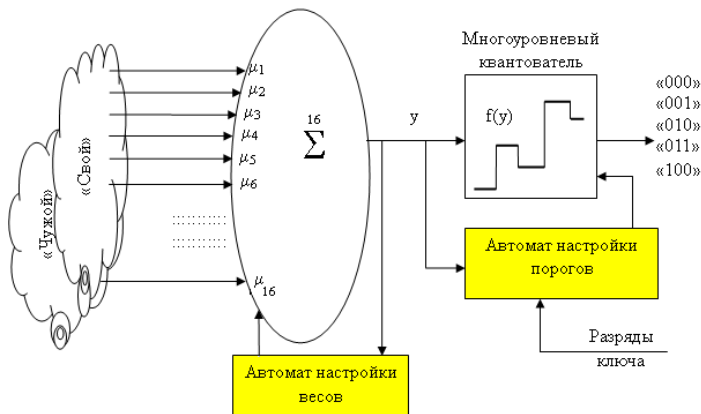


Рис. 8.4. Раздельное обучение линейной и нелинейной части нейрона

Задача обучения распадается на две независимые подзадачи. Сначала необходимо обучить линейную часть многоуровневого нейрона по некоторому достаточно устойчивому алгоритму автоматического обучения. Далее необходимо перейти ко второй подзадаче обучения нелинейной части нейрона или расстановки порогов срабатывания нелинейной функции квантования. Обе эти задачи должны решаться независимо,

только в случае разделения функции обучения нейрона на две подфункции удастся построить два достаточно устойчивых обучающих автомата.

8.3. Обучение линейной части многопорогового нейрона по критерию повышения стабильности распределения «Свой»

Ранее отмечалось, что пороговая функция делит многомерную сферу образов «Все Чужие» на две области «0» и «1», соответствующей гиперплоскостью. При обучении по ГОСТ Р 52633.5 разделяющая гиперплоскость проходит через центр области «Все Чужие». При настройке весовых коэффициентов по ГОСТ Р 52633.5–2011 происходит эффект некоторого увеличения размеров области «Свой» и эффект ее выталкивания на периферию области «Все Чужие». Эта ситуация отображена в левой части рис. 8.5. Подобная тактика обучения приводит к тому, что распределение данных образов «Свой» рано или поздно перестает пересекаться разделяющей гиперплоскостью, и нейрон начинает надежно узнавать примеры образа «Свой».

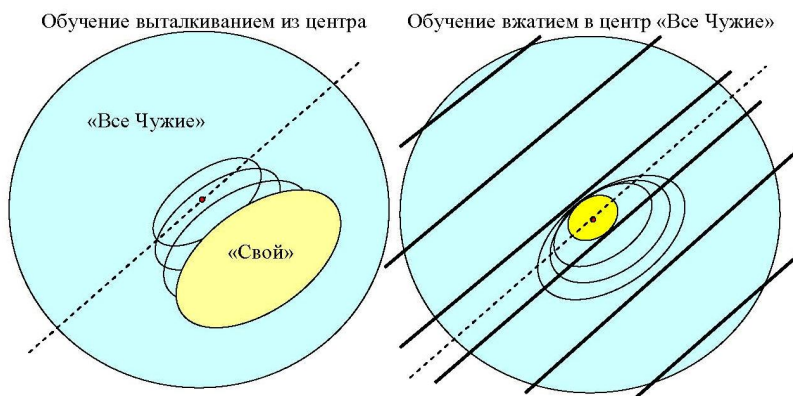


Рис. 8.5. Две стратегии обучения линейной части нейрона

Если изменить критерий качества обучений и потребовать, чтобы алгоритм обучения как можно сильнее сжимал распределение данных «Свой» по критерию:

$$Q_c = \frac{\sigma(y_\xi)}{\sigma(y_v)}, \quad (8.1)$$

то ситуация меняется коренным образом.

Все алгоритмы обучения, повышающие показатель стабильности (8.1), работают сжимая диаметр распределения данных «Свой» и смещая его центр в район центра множества образов «Все Чужие». Эта ситуация отображена в правой части рис. 8.1. При этом, как осуществляется максимизация критерия (8.1), не имеет значения. Могут использоваться как итерационные алгоритмы поиска максимума, построенные на вычислении частных производных $\frac{\partial Q_c}{\partial \mu_i}$, так и иные неитерационные алгоритмы

вычисления весовых коэффициентов.

Например, неитерационные алгоритмы быстрого обучения линейной части нейронов могут быть построены на базе процедуры связывания пар биометрических данных, через замещение одной сложной многомерной задачи несколькими более простыми двухмерными задачами. Подобная декомпозиция задачи отражена на рис. 8.6.

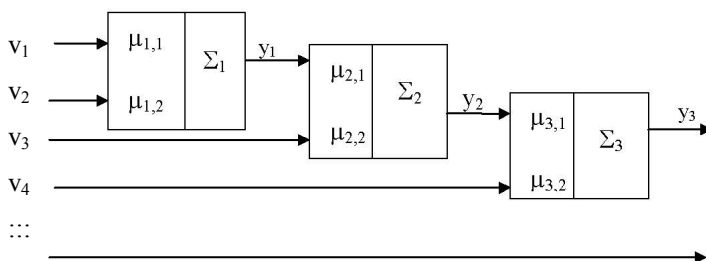


Рис. 8.6. Декомпозиция многомерной задачи на последовательное решение двухмерных задач

Вначале находятся весовые коэффициенты сумматора с двумя входами v_1 , v_2 и выходом y_1 путем следующих вычислений:

$$\left\{ \begin{array}{l} \mu_{1,1} = \frac{1}{\sigma(v_1)}; \\ \mu_{1,2} = -\text{sign}\{r(v_1, v_2)\} \frac{1}{\sigma(v_2)}; \\ r(v_1, v_2) = \frac{1}{N} \sum_{i=1}^N \frac{(E(v_{1i}) - v_{1i}) \cdot (E(v_{2i}) - v_{2i})}{\sigma(v_1) \cdot \sigma(v_2)}. \end{array} \right. \quad (8.2)$$

После того, как весовые коэффициенты первого из двухвходовых сумматоров вычисляют весовые коэффициенты второго сумматора:

$$\left\{ \begin{array}{l} \mu_{2,1} = \frac{1}{\sigma(y_1)}; \\ \mu_{2,2} = -\text{sign}\{r(y_1, v_3)\} \frac{1}{\sigma(v_3)}; \\ r(y_1, v_3) = \frac{1}{N} \sum_{i=1}^N \frac{(E(y_{1i}) - y_{1i}) \cdot (E(v_{3i}) - v_{3i})}{\sigma(y_1) \cdot \sigma(v_3)}. \end{array} \right. \quad (8.3)$$

Сравнивая выражения (8.2) и (8.1), нетрудно заметить, что при вычислениях весовых коэффициентов каждого из сумматоров нужно знать среднеквадратические отклонения его входных биометрических параметров и знак коэффициента корреляции между этими параметрами. Иначе говоря, вычисления можно продолжать как угодно долго, обучая таким алгоритмом нейрон с очень большим числом входов. При этом устойчивость вычислений оказывается достаточно высокой, так как при больших значениях коэффициентов парной корреляции ошибиться с выбором знака весовых коэффициентов практически невозможно. При малых значениях коэффициента парной корреляции вероятность ошибочного определения знака велика, однако в этих случаях знак коэффициента корреляции

практически не оказывает влияния на рост показателя качества. Если биометрические данные некоррелированы, то знаки весовых коэффициентов могут выбираться произвольно.

В случае, когда весовые коэффициенты k входных двухвходовых сумматоров вычислены, все частные сумматоры могут быть свернуты в один общий сумматор нейрона с k входами. При этом общие входные весовые коэффициенты будут вычисляться как произведения частных весовых коэффициентов:

$$\begin{cases} \mu_1 = \prod_{i=1}^k \mu_{i,1} \\ \mu_2 = \prod_{i=1}^{k-1} \mu_{i,2} \\ \dots\dots\dots \\ \mu_k = \mu_{k,k} \end{cases} \quad (8.4)$$

8.4. Сборка двух разных алгоритмов обучения линейной части нейрона

После получения двух решений для линейной части нейрона (сжатием и выталкиванием образа «Свой»), комбинируя их можно получить множество линейных комбинаций этих решений. Эта ситуация отображена на рис. 8.7.

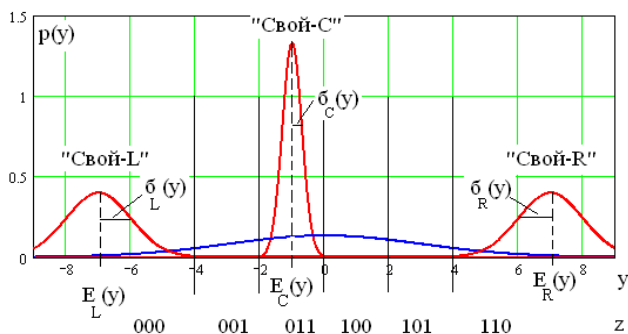


Рис. 8.7. Получение из двух линейных решений третьего решения

Это означает, что, изменяя регулируемый параметр «а» от 0 до 1 в уравнении морфинга

$$\mu_i(a, C, R) = a \cdot \mu_i(R) + (1 - a) \cdot \mu_i(C), \quad (8.5)$$

можно плавно перемещать математическое ожидание распределения «Свой-С» в сторону математического ожидания распределения «Свой-Р». При этом с.к.о. $\sigma_C(y)$ будет плавно увеличиваться до $\sigma_R(y)$.

Более того, известно, что путем простого инвертирования нейросетевого решения $\mu_i(R)$ легко получить инверсное решение $\mu_i(L)$. Последнее означает, что, используя весовые коэффициенты найденных трех нейросетевых решений $\mu_i(R)$, $\mu_i(L)$, $\mu_i(C)$, можно поместить центр распределения «Свой» в любую точку распределения «все Чужие». Далее необходимо найти границы этого распределения и соответственно определить число состояний дискретизирующего выходные данные нелинейного элемента нейрона. В зависимости от выбранного положения математического ожидания откликов «Свой» число устойчивых состояний нелинейности $f(y)$ находится как отношение

$$k \approx \frac{\sigma(y_\xi)}{\sigma(y_v)}. \quad (8.6)$$

Для того чтобы скрыть действительное положение границ «Свой», достаточно ввести справа и/или слева дополнительные ложные границы. На рис. 8.7 дан пример ложных границ, поставленных справа и слева от действительного положения распределения «Свой-С».

Очевидно, что ложные границы должны быть правдоподобными, т. е. не выходить за края распределения «все Чужие», а также интервалы между ними должны расширяться по мере движения в сторону периферии распределения «Все Чужие». Гарантеей тайны положения действительных границ «Свой» является правдоподобное расположение ложных границ. Правдоподобие расположения ложных границ обусловлено тем, что параметров действительных распределений «Свой-С», «Свой-Р», «Свой-Л» сторонний наблюдатель не знает. Промежуточные данные о настройках нелинейного элемента уничтожаются после обучения каждого нейрона.

8.5. Настройка порогов функции многоуровневого квантования

Второй частью обучения нейрона является выбор числа порогов функции квантования и настройка положения порогов квантования. Самым простым способом оценки сверху числа порогов является вычисление отношения среднеквадратических отклонений образов «Чужие» и образа «Свой», после его сжатия

$$\max(n) \approx \frac{\sigma(y_{\xi})}{\sigma(y_{\text{сжатие}})} . \quad (8.7)$$

Оцененное по формуле (8.7) число порогов является избыточным, если воспользоваться столь большим числом порогов, то получим ситуацию, отображенную на рис. 8.8.

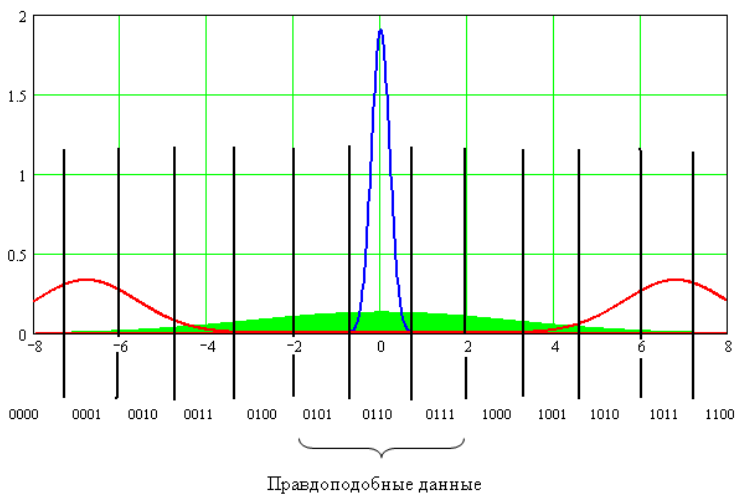


Рис. 8.8. Максимально возможное (избыточное) число порогов функции квантования

Из рис. 8.8 видно, что атакующий, наблюдая слишком частую сетку порогов, будет рассматривать как правдоподобные только пороги, находящиеся в центре распределения «Все Чужие». Предположительно атакующему будет известна связь

статистики ширины между соседними порогами с положением центра распределения образа «Свой». Другими словами, введение слишком большого числа порогов с одинаковыми расстояниями неминуемо приводит к гипотезе о расположении данных образа «Свой» в центре распределения «Все Чужие».

В связи с этим более рациональной является тактика, когда настройка положения порогов с одной стороны ведется, исходя из положения образа данных «Свой», а с другой должна учитывать эффект расширения порогов при движении их на периферию данных образов «Все Чужие». Также необходимо, чтобы при движении автомата расстановки порогов в центр образов «Все Чужие» расстояние между порогами сжималось. В этом случае получаются правдоподобные положения порогов (рис. 8.9).

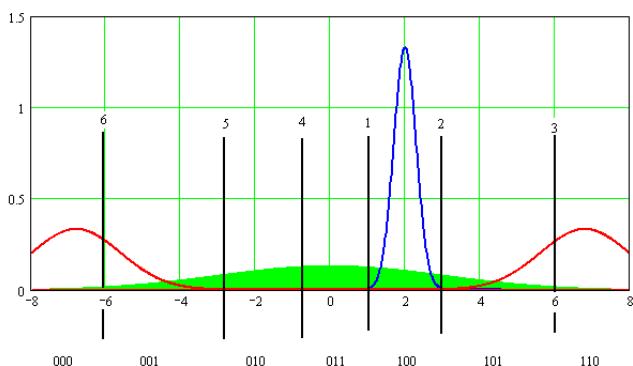


Рис. 8.9. Правдоподобная процедура расстановки порогов

Очевидно, что первыми должны выставляться пороги вокруг выделяемого распределения образа «Свой» (на рис. 8.9 это пороги с номерами 1 и 2). Зная действительные пороги выделения образа «Свой», можно найти вероятность ошибок второго рода для настраиваемого i -го нейрона:

$$P_{2,i} = \frac{1}{\sigma(y_{\xi})\sqrt{2\pi}} \int_{\min(y_{vi})}^{\max(y_{vi})} \exp\left\{-\frac{(E(y_{\xi}) - y_{\xi})^2}{2\sigma^2(y_{\xi})}\right\} dy_{\xi}. \quad (8.8)$$

При правдоподобной расстановке порогов в идеале для всех пороговых интервалов должна быть одна и та же вероятность ошибок второго рода (8.8). Опираясь на это, нетрудно расставить пороги 3, 4, 5, 6, как это показано на рис. 8.9. В том, что пороги получаются правдоподобными, легко убедиться перемещая в центр между ними центр распределения данных образа «Свой».

Если теперь сравнить число порогов рис. 8.8 и 8.9, то легко убедиться в том, что правдоподобное число порогов всегда оказывается примерно в два раза больше, чем их максимальная оценка.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Feng Hao, Ross Anderson, and John Daugman.* Crypto with Biometrics Effectively, IEEE TRANSACTIONS ON COMPUTERS. – VOL. 55. – № 9. – SEPTEMBER 2006.
2. *Y. Dodis, L. Reyzin, A. Smith.* Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy, Data April 13, In EUROCRYPT, pages 523–540, 2004.
3. *Arakala A., Jeffers J., Horadam K. J.* Fuzzy Extractors for Minutiae-Based Fingerprint Authentication. // Advances in Biometrics (LNCS 4642), Springer. – Pp. 760–769, 2007.
4. *Чморра А. Л.* Маскировка ключа с помощью биометрии «Проблемы передачи информации». – 2011. – № 2(47). – С. 128–143.
5. *Чморра А. Л., Уривский А. В.* Биометрическая система аутентификации / Патент № RU2316120. – Оpubл. 27.01.2008. – Бюл. № 3. – Приоритет от 12.05.2004.
6. ГОСТ Р 52633.0–2006 «Защита информации. Техника защиты информации. Требования к средствам высоконадежной биометрической аутентификации». – М.: Стандартинформ, 2006.
7. *Язов Ю. К. и др.* Нейросетевая защита персональных биометрических данных / Ю. К. Язов (редактор и автор) / В. И. Волчихин, А. И. Иванов и др. – М.: Радиотехника, 2012. – 157 с.
8. ГОСТ Р 52633.5–2011 «Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия-код доступа». – М.: Стандартинформ, 2012.
9. *Волчихин В. И., Иванов А. И., Фунтиков В. А.* Быстрые алгоритмы обучения нейросетевых механизмов биометрико-криптографической защиты информации // Монография. – Пенза: Изд-во Пензенского государственного университета, 2005. – 273 с.
10. *Ахметов Б. С., Иванов А. И. и др.* О связи высокоразмерной энтропии и высокоразмерной корреляции с математическим ожиданием модулей коэффициентов парной корреляции. – Усть-Каменогорск: ВКГТУ, 2013. – С. 240–244.
11. *Ахметов Б. С., Иванов А. И. и др.* Оценка вероятностей появления ошибок нейросетевых преобразователей биометрия-

код на основе малых выборок. Высокие технологии – залог устойчивого развития // Мат. II Межд. науч. конф. – Т.1. – Алматы, 2013. – С. 234–237.

12. *Ахметов Б. С., Иванов А.И. и др.* Синхронизация процедур нейросетевого обучения особенностям рукописного почерка по точкам смены направления движения пера. – Алматы: КазНТУ // Сб. тр. II Межд. науч.-прак. конф. «Информационно-инновационные технологии: интеграция науки, образования и бизнеса». – 2011. – Т. 2. – С.118–123.

13. ГОСТ Р 52633.3–2011 «Защита информации. Техника защиты информации. Тестирование стойкости средств высоконадежной биометрической защиты к атакам подбора». – М.: Стандартинформ, 2012.

14. *Ахметов Б. С., Волчихин В. И. и др.* Алгоритмы тестирования биометрико-нейросетевых механизмов защиты информации // Монография. – Алматы: LEM. – 2013. – 152 с.

15. *Ахметов Б. С., Захаров О. С. и др.* Метод оценки вероятностей появления ошибок нейросетевых преобразователей биометрия-код, использующий очень малые тестовые выборки // Вестн. КазНТУ. – 2013. – № 3.

16. *Ахметов Б. С., Иванов А. И. и др.* Энтропийно-корреляционный подход к расчету вероятности совместного появления большого числа зависимых событий // Вестн. КБТУ. – 2013. № 2(25) – С. 55–59.

17. *Ахметов Б. С., Иванов А. И. и др.* Биометрическая аутентификация граждан в открытых информационных пространствах // Тр. I Межд. науч.-прак. конф. «Интеллектуальные информационные и коммуникационные технологии – средство осуществления третьей индустриальной революции в свете стратегии "Казахстан-2050"». – Астана: ЕНУ им. Л. Н. Гумилева, 2013. – С. 458-459.

18. *Малыгин А. Ю., Ахметов Б. С. и др.* Учет влияния корреляционных связей на результаты тестирования преобразователей биометрия-код // Информационные и телекоммуникационные технологии: образование, наука, практика: Сб. тр. Межд. науч.-практ. конф. – Алматы: КазНТУ, 2012. – С. 34–37.

19. *Болл Руд и др.* Руководство по биометрии / Болл Руд,

Коннел Джонатан Х. и др. – М.: Техносфера, 2007. – 368 с.

20. Ларин П. З., Реввер Е. И. Обведем вокруг пальца? Обман биометрических систем доступа, использующих дактилоскопическую идентификацию личности // Информационная безопасность – Information Security». – Groteck. – 2004. – № 3.

21. Arakala A., Jeffers J., Horadam K. J. Fuzzy Extractors for Minutiae-Based Fingerprint Authentication // Advances in Biometrics (LNCS 4642). Springer. – 2007. – Pp. 760–769.

22. Ratha N. K., Chikkerur S. – Generating cancelable fingerprint templates // IEEE Trans. PAMI, 29(4). – 2007. – P. 561–572.

23. ISO/IEC 19784–1:2006 "BioAPI – Biometric Application Programming Interface – Part 1: BioAPI Specification", <http://www.bioapi.org/>

24. ГОСТ Р ИСО/МЭК 19784–1–2007 «Автоматическая идентификация. Идентификация биометрическая. Биометрический программный интерфейс». – Ч. 1. Спецификация биометрического программного интерфейса. – М.: Стандартинформ, 2007.

25. ГОСТ Р ИСО/МЭК 19784–2–2010 «Автоматическая идентификация. Идентификация биометрическая. Биометрический программный интерфейс». – Ч. 2. Интерфейс поставщика биометрической функции архива. – М.: Стандартинформ, 2011.

26. ГОСТ Р ИСО/МЭК 19794–2–2005 «Автоматическая идентификация. Идентификация биометрическая. Форматы обмена биометрическими данными». – Ч. 2. Данные изображения отпечатка пальца – контрольные точки. – М.: Стандартинформ, 2007.

27. Федеральный закон Российской Федерации от 27 июля 2006 г. № 152–ФЗ «О персональных данных». Режим доступа: <http://www.rg.ru/>

28. ГОСТ Р ИСО/МЭК 19794–10–2010 «Автоматическая идентификация. Идентификация биометрическая. Форматы обмена биометрическими данными. Данные геометрии контура кисти руки». – М.: Стандартинформ, 2011.

29. John Daugman «Probing the Uniqueness and Randomness of Iris Codes: Results From 200 Billion Iris Pair Comparisons» Proceedings of the IEEE | Vol.

30. Язов Ю. К. и др. Нейросетевая защита персональных

биометрических данных. – М.: Радиотехника. – 2012. – 160 с.

31. Белоцерковский О. М., Глазунов А. С., Щенников В. В. Компьютерное распознавание человеческих лиц // Зарубежная радиоэлектроника. Успехи современной радиоэлектроники. – 1997. – № 8. – С. 3–14.

32. Глазунов А. Компьютерное распознавание человеческих лиц // Открытые системы. – 2000. – № 3, www.osp.ru/os/2000/03/043.htm.

33. Пентланд А., Чаудхари Т. Распознавание лиц для интеллектуальных сред // Открытые системы. – 2000. – № 3, www.osp.ru/os/2000/03/028.htm.

34. Шатило Л., Стокман Дж. Компьютерное зрение. – М.: Бином, 2009. – 752 с.

35. Дуда Р., Харп П. Распознавание образов и анализ сцен. – М.: Мир. 1976.

36. А.с. № RU 112 992 «Устройство бесконтактного измерения геометрических параметров лица человека» / Захаров С. М., Митрохин М. А. Опубл. 27.01.12.

37. ОКР «Разработка и подготовка производства телекоммуникационного оборудования, разработка программного сетевого, прикладного и специального обеспечения для создания цифровых сетей связи с персонализированным доступом». – Пенза: ОАО «НПП «Рубин», ФГБОУ ВПО «ПГУ», 2012.

38. А. с. RU 2451409 С2. Способ однозначного хэширования неоднозначных биометрических данных / А. И. Иванов, С. М. Иванов. Заявлено 26.01.2010; Опубл. 20.05.2012, Бюл. № 14.

39. Uludag U., Pankanti S., Prabhakar S., Jain A. K. Biometric cryptosystems: issues and challenges // Proceedings of the IEEE. – 2004. – Vol. 92. – № 6. – P. 948–960.

40. Juels A., Sudan M. A. Fuzzy Vault Scheme // Proceedings of IEEE International Symposium on Information Theory. Lausanne, Switzerland, 2002. – P. 408.

41. Dodis Y., Reyzin L., Smith A. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data // Advances in Cryptology – EUROCRYPT 2004. Christian Cachin

and Jan Camenisch, ed. – Vol. 3027 of Lecture Notes in Computer Science. Springer – Verlag, 2004. – P. 79–100.

42. *Sahai A., Waters B.* Fuzzy identity-based encryption // Proceedings of EUROCRYPT 2005, LNCS 3494. Springer – Verlag, 2005. – P. 457–473.

43. *Baek J., Susilo W., Zhou J.* New construction of fuzzy identity-based encryption // Proceedings of the 2nd ACM Symposium on Information, Computer and Communications Security. ACM New York. NY, USA, 2007. – P. 368–370.

44. *Fang L. M., Wang J. D., Ren Y. J., Xia J. Y., Bian S. Z.* Chosen – Ciphertext Secure Fuzzy Identity – Based Key Encapsulation without ROM. Cryptology ePrint Archive: Report 2008/139, 2008. URL: <http://eprint.iacr.org/2008/139>.

45. *Fang L., Xia J.* Full Security: Fuzzy Identity Based Encryption. Cryptology ePrint Archive: Report 2008/307. URL: <http://eprint.iacr.org/2008/307>.

46. *Yang P., Cao Z., Dong X. I.* Fuzzy Identity Based Signature. Cryptology ePrint Archive: Report 2008/002. URL: <http://eprint.iacr.org/2008/002>.

47. *Burnett A., Duffy A., Dowling T. A.* Biometric Identity Based Signature Scheme. Cryptology ePrint Archive: Report 2004/176. URL: <http://eprint.iacr.org/2004/176>.

48. *Dodis Y., Katz J., Reyzin L., Smith A.* Robust fuzzy extractors and authenticated key agreement from close secrets // Advances in Cryptology – CRYPTO 2006. Cynthia Dwork, ed. Vol. 4117 of Lecture Notes in Computer Science. Springer – Verlag. – 20–24 August 2006. – P. 232–250.

49. *Kanukurthi B., Reyzin L.* Key Agreement from Close Secrets over Unsecured Channels. Cryptology ePrint Archive: Report 2008/494. URL: <http://eprint.iacr.org/2008/494>.

50. *Boyen X.* Reusable cryptographic fuzzy extractors // Eleventh ACM Conference on Computer and Communication Security. ACM. – October 25–29 2004. – P. 82–91.

51. *Boyen X., Dodis Y., Katz J., Ostrovsky R., Smith A.* Secure remote authentication using biometric data // Advances in Cryptology – EUROCRYPT 2005. R. Cramer, ed. Springer – Verlag, 2005. – Vol. 3494 of LNCS. – P. 147–163.

52. *John Daugman*. Probing the Uniqueness and Randomness of Iris Codes: Results From 200 Billion Iris Pair Comparisons» Proceedings of the IEEE. – Vol. 94. – № 11. – November 2006. – Pp. 1928–1935.

53. *Daugman John G*. Biometric personal identification system based on iris analysis, US Patent Number: 5,291,560 Data of Patent: Mar.1, 1994.

54. *Ушмаев О. С., Кузнецов В. В.* Получение устойчивого криптографического ключа из биометрической характеристики изображения отпечатков пальцев. – The 22nd International Conference on Computer Graphics and Vision, Russia, Moscow, October 01–05, 2012. – Pp. 125–127.

55. *Смит Ричард Э.* Аутентификация: от паролей до открытых ключей. – М.: Вильямс, 2002. – 424 с.

56. А. с. RU 2316120 С2 Биометрическая система аутентификации / *Чморра А. Л., Уривский А. В.* Заявлено 05.2014; Опубл. 27.01.2008, Бюл. № 3.

57. *Морелос-Сарагоса Р.* Искусство помехоустойчивого кодирования. – М.: Техносфера, 2007. – 320 с.

58. *Питерсон У., Уэлдон Э.* Коды, исправляющие ошибки // Монография / Под ред. *Добрушиной Р. Л., Самойленко С. И.* – М.: МИР, 1976. – 364 с.

59. *Самойленко С. И.* Помехоустойчивое кодирование. – М.: Радио и связь, 1968. – 240 с.

60. *Безяев А. В.* Нейросетевой преобразователь биометрии в самокорректирующийся код, совершенно не обладающий избыточностью // Нейрокомпьютеры: разработка и применение. – 2012. – № 3. – С. 52–55.

61. *Безяев А. В., Фунтикова Ю. В.* Оптимизация параметров хэш-кодов, осуществляющих обнаружение и корректирование ошибок в биометрических преобразователях // Тр. науч.-техн. конф. кластера пензенских предприятий, обеспечивающих безопасность информационных технологий. – Т 8. – Пенза, 2012. – С. 76–78 (<http://пниэи.рф/activity/science/BIT/T8-p76.pdf>).

62. *Круглов В. В., Дли М. И., Голунов Р. Ю.* Нечёткая логика и искусственные нейронные сети. – М.: Физматлит, 2001. – 221 с.

63. Рутковская Д., Пилиньский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткие системы / Пер. с польского И. Д. Рудинского. – М.: Горячая линия – Телеком, 2004.
64. Саймон Хайкин. Нейронные сети: Полный курс. – М.: Вильямс, 2006. – С. 1104.
65. Уоссермен Ф. Нейрокомпьютерная техника: теория и практика. – М.: Мир, 1992. – 240 с.
66. Иванов А. И. Биометрическая идентификация личности по динамике подсознательных движений // Монография. – Пенза: Изд-во ПГУ, 2000.
67. Волчихин В. И., Иванов А. И., Фунтиков В. А. Быстрые алгоритмы обучения нейросетевых механизмов биометрико-криптографической защиты информации // Монография. – Пенза: Изд-во ПГУ, 2005. – 273 с.
68. Иванов А. И. Нейросетевые алгоритмы биометрической идентификации личности. Сер. «Нейрокомпьютеры и их применение». – М.: Радиотехника, 2004. – Вып. 15. – 144 с.
69. Иванов А. И., Кисляев С. Е., Гелахвили П. А. Искусственные нейронные сети в биометрии, медицине, здравоохранении. – Самара: ООО «Офорт», 2004. – 236 с.
70. Минский М., Пейперт С. Перцептроны. – М.: Мир, 1971. – 261 с.
71. Розенблатт Ф. Принципы нейродинамики: Перцептроны и теория механизмов мозга. – М.: Мир, 1965. – 480 с.
72. Куликов С. В., Секретов М. В. и др. Учет «тяжелых» хвостов ненормального закона распределения биометрических параметров все «Чужие» при настройке нелинейного элемента нейрона с несколькими дискретными состояниями // Нейрокомпьютеры: разработка, применение. – 2012. – № 3. – С. 56–59.
73. ГОСТ Р 52633.4–2012 «Защита информации. Техника защиты информации. Интерфейсы взаимодействия с нейросетевыми преобразователями биометрия-код». – М.: Стандартинформ, 2012.
74. ГОСТ Р 52633.7–2013 «Защита информации. Техника защиты информации. Высоконадежная мультибиометрическая аутентификация». – М.: Стандартинформ, 2013.

75. Минский М., Пейперт С. Персептроны. – М.: Мир, 1971.

76. ГОСТ Р 52633.3–2011 «Защита информации. Техника защиты информации. Тестирование стойкости средств высоконадежной биометрической защиты к атакам подбора». – М.: Стандартинформ, 2011.

77. Язов Ю. К., Волчихин В. И. и др. Нейросетевая защита персональных биометрических данных // Монография. – М.: Радиотехника, 2012. – 160 с.

78. 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.

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

80. Ахметов Б. С., Волчихин В. И. и др. Моделирование длинных биометрических кодов, воспроизводящих корреляционные связи выходных данных нейросетевого преобразователя // Нейрокомпьютеры: разработка, применение. – М.: Радиотехника. – 2012. – № 3. – С. 40–43.

81. Ахметов Б. С., Волчихин В. И. и др. Учет влияния корреляционных связей на результаты тестирования преобразователей биометрия-код // Сб. тр. Межд. науч.-прак. конф. «Информационные и телекоммуникационные технологии: образование, наука, практика». – Алматы: КазНТУ, 2012. – С. 34 – 37.

82. ГОСТ Р 52633.5–2011 «Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия-код доступа». – М.: Стандартинформ, 2011.

83. Ахметов Б. С., Безяев В. С. Сравнительная оценка уровня безопасности средств биометрической аутентификации // Сб. тр. II Межд. науч.-прак. конф. «Информационно-инновационные технологии: интеграция науки, образования и бизнеса». – Алматы: КазНТУ, 2011. – Т. 1. – С. 382–386.

84. *B. Akhmetov, A. Ivanov, V. Funtikov, I. Urnev* Evaluation of Multidimensional Entropy on Short Strings of Biometric Codes with Dependent Bits. «Progress in Electromagnetics Research Symposium» PIERS Proceedings, August 19–23, Moscow, RUSSIA, 2012. – P. 66–69.

85. ГОСТ Р 52633.2–2010 «Защита информации. Техника защиты информации. Требования к формированию синтетических биометрических образов, предназначенных для тестирования средств высоконадежной биометрической аутентификации». – М.: Стандартинформ, 2011. 24 с.

86. *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 Proceedings. – August 19–23. – Moscow, RUSSIA, 2012. – P. 61–65.

87. *Иванов А. И., Фунтиков В. А.* Моделирование кодовых последовательностей с энтропией естественных и искусственных биометрических языков «Инфокоммуникационные технологии». – 2010. – Т. 8. – № 4. – С. 75–79.

88. *Маршалко Г. Б.* Вопросы оценки стойкости нейросетевой системы биометрической аутентификации // Мат. конф. «РусКрипто–2013»

http://www.ruscrypto.ru/netcat_files/File/ruscrypto.2013.051.zip

89. *Ахметов Б. С., Иванов А. И., Картбаев Т. С.* Нейрожелілердің артық байланыстары болуынан биометриялық-кодтың ішінара әшкерелену қаупі мен оларға қарсы тұру әдісі // ҚазҰТУ хабаршысы. – 2014. – № 2 (102). – Б. 147–151.

90. *Андреев Д. Ю.* Оценка эффективности применения обычной и взвешенной метрики Хэмминга при упорядочивании баз естественных биометрических образов // Нейрокомпьютеры: разработка, применение. – 2012. – № 3. – С. 31–34.

91. *Андреев Д. Ю.* Сокращение затрат на нечеткую адресацию биометрических образов через использование взвешенной меры Хэмминга. Труды научно-технической конференции кластера пензенских предприятий, обеспечивающих Безопасность информационных технологий. – 2012. – Т. 8. – Пенза. – С. 32–36 (<http://www.pniei.penza.ru/RV-conf/T8/C32.pdf>)

92. А. с. RU 2346397 С1. Способ защиты персональных данных биометрической идентификации и аутентификации / *Иванов А. И., Фунтиков В. А., Ефимов О. В.* Заявлено 26.06.2007; Опубл. 10.02.2009, Бюл № 4.

93. *Фунтиков В. А., Назаров И. Г., Бурушкин А. А.* Национальные стандарты России: конфиденциальность персональных биометрических данных // Стандарты и качество. – 2010. – № 7. – С. 28–33.

94. *Язов Ю. К. и др.* Нейросетевая защита персональных биометрических данных / *Ю. К. Язов* (редактор и автор), соавторы *В.И. Волчихин, А.И. Иванов, В.А. Фунтиков, И.Г. Назаров.* – М.: Радиотехника, 2012. – 157 с.

95. *Иванов А. И.* Биометрическая идентификация личности по динамике подсознательных движений. – Пенза: Изд-во ПГУ, 2000. – 188 с.

96. *Волчихин В. И., Иванов А. И., Фунтиков В. А.* Быстрые алгоритмы обучения нейросетевых механизмов биометрико-криптографической защиты информации // Монография. – Пенза: ПГУ, 2005. – 273 с.

97. *Куликов С. В., Секретов М. В. и др.* Учет «тяжелых» хвостов ненормального закона распределения биометрических параметров все «Чужие» при настройке нелинейного элемента нейрона с несколькими дискретными состояниями // Нейрокомпьютеры: разработка, применение. – 2012. – № 3. – С. 56–59.

98. А. с. RU 2451409 С2. Способ однозначного хэширования неоднозначных биометрических данных / *А. И. Иванов, С. М. Иванов.* Заявлено 26.01.2010; Опубл. 20.05.2012, Бюл. № 14.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	3
1. ИСТОРИЯ ВОЗНИКНОВЕНИЯ И РАЗВИТИЯ БИОМЕТРИЧЕСКИХ ТЕХНОЛОГИЙ.....	5
1.1. Основные понятия и определения.....	8
1.1.1. Вероятности ошибок первого и второго рода.....	8
1.1.2. Преобразование биометрических образов в код ключа доступа.....	9
1.1.3. Представление многомерных континуумов биометрических образов малым дискретным числом примеров.....	11
1.1.4. Переход в пространство расстояний Хэмминга.....	14
1.1.5. Информативность биометрических параметров.....	16
1.1.6. Оценка энтропии преобразователя биометрия-код....	18
1.1.7. Оценка размерности решаемой биометрической задачи.	20
2. АНАЛИЗ СОВРЕМЕННЫХ БИОМЕТРИЧЕСКИХ ТЕХНОЛОГИЙ.....	22
2.1. Технология анализа рисунка отпечатка пальца.....	22
2.2. Технология анализа дерева кровеносных сосудов пальца и ладони.....	24
2.3. Технология анализа геометрии ладони.....	26
2.4. Технология анализа радужной оболочки глаза и сосудов глазного дна.....	27
2.5. Двухмерная и трехмерная технологии анализа геометрии лица человека.....	30
2.6. Технология анализа трехмерной геометрии ушных раковин.....	35
2.7. Одномерная и двухмерная технология анализа особенностей голоса.....	37
2.8. Технология анализа динамики рукописного почерка.	39
2.9. Перспективы использования мультибиометрических приложений.....	42
3. НЕЧЕТКИЕ ЭКСТРАКТОРЫ БИОМЕТРИЧЕСКИХ ПАРАМЕТРОВ.....	48
3.1. Простейший вариант квантования биометрических данных.....	48

3.2.	«Нечеткий экстрактор» с сокрытием информации наложением гаммы.....	50
3.3.	«Нечеткий экстрактор» с хэшированием биокода и наложением гаммы.....	51
3.4.	Выбор используемых в «нечетком экстракторе» самокорректирующихся кодов.....	53
4.	НЕЙРОСЕТЕВЫЕ ПРЕОБРАЗОВАТЕЛИ БИОМЕТРИЯ-КОД.....	56
4.1.	Итерационный алгоритм обучения одиночного нейрона (персептрона).....	56
4.2.	Абсолютно устойчивый неитерационный алгоритм обучения нейрона.....	61
4.3.	Формирование и обучение сети нейронов. Нейросетевые контейнеры.....	65
4.4.	Универсальность двухслойной сети искусственных нейронов.....	68
4.5.	Рациональный выбор длины биокода нейросетевых преобразователей.....	73
4.6.	Нейросетевая коррекция био-кода вторым слоем нейронов.....	75
4.7.	Хэширование данных «Чужой», выполняемое нейронами второго слоя.....	77
5.	ФОРМИРОВАНИЕ БАЗ БИОМЕТРИЧЕСКИХ ДАННЫХ И ТЕСТИРОВАНИЕ КАЧЕСТВА ОБУЧЕНИЯ СЕТИ НЕЙРОНОВ.....	79
5.1.	Оценка вероятности возникновения ошибок первого рода на малых тестовых выборках.....	79
5.2.	Проблемы оценки вероятности ошибок второго рода, необходимость формирования больших баз биометрических образов.....	82
5.3.	Размножение примеров биометрических образов через их скрещивание.....	86
5.4.	Синтез базы биометрических образов, равномерно заполняющей многомерное пространство «Все Чужие».	90
5.5.	Быстрая оценка стойкости к атакам подбора по усредненным модулям парных корреляционных связей.	92

6.	УГРОЗЫ СРЕДСТВАМ БИОМЕТРИЧЕСКОЙ ЗАЩИТЫ ДОСТУПА К ИНФОРМАЦИИ, НЕ ИМЕЮЩИМ КРИПТОГРАФИЧЕСКОЙ ЗАЩИТЫ БИОДАННЫХ И АТАКИ, РЕАЛИЗУЮЩИЕ ЭТИ УГРОЗЫ.....	97
6.1.	Угроза модификации программы защиты информации атакой на «последний» бит.....	97
6.2.	Угроза поиска биометрических близнецов атакой перебора большой базы образов «Чужой».....	98
6.3.	Угроза перебора возможных входных и выходных состояний автомата биометрической защиты.....	100
6.4.	Угроза перехвата биометрического образа, атака изготовления муляжа.....	101
6.5.	Угроза частичной компрометации биокода из-за наличия избыточных связей в нейросети, атака Маршалко.....	102
6.6.	Угроза частичной компрометации образа «Свой» атакой обращения матрицы нейросетевых функционалов при знании выходного кода ключа.....	103
6.7.	Угроза частичной компрометации образа «Свой» атакой обращения матрицы нейросетевых функционалов путем генетического подбора и ориентира по параметру стабильности разрядов биокода.....	107
6.8.	Угроза частичной компрометации выходного биокода и его инверсии, обладающих минимальной энтропией.....	110
6.9.	Угроза частичной компрометации образа «Свой» атакой обращения матрицы нейросетевых функционалов путем подбора минимизацией взвешенной метрики Хэмминга.....	112
7.	КРИПТОГРАФИЧЕСКАЯ ЗАЩИТА СОДЕРЖАНИЯ НЕЙРОСЕТЕВОГО КОНТЕЙНЕРА.....	114
7.1.	Классическая защита персональных данных хэшированием паролей и биометрических кодов доступа.....	114
7.2.	Криптографическая защита данных нейросетевого контейнера простым шифрованием.....	115

7.3.	Защита нейронов блочным шифрованием с растущей длинной криптографического ключа.....	117
7.4.	Защита данных нейросетевого контейнера с использованием криптографических ключей растущей длины от 1 бита до 256 бит.....	118
7.5.	Казахстанский вариант усиления стойкости шифрования за счет частичного возврата к итерационному обучению нейронов с перекрывающимися связями....	122
8.	ЗАЩИТА ПЕРСОНАЛЬНЫХ БИОМЕТРИЧЕСКИХ ДАННЫХ НЕЙРОНАМИ С МНОГОПороГОВЫМИ ФУНКЦИЯМИ КВАНТОВАНИЯ.....	125
8.1.	Переход к использованию нейронов с квантованием по двум и более порогам.....	125
8.2.	Требования к обучению нейронов с многоуровневым квантованием выходного сигнала сумматора.....	128
8.3.	Обучение линейной части многопорогового нейрона по критерию повышения стабильности распределения «Свой».....	129
8.4.	Сборка двух разных алгоритмов обучения линейной части нейрона.....	132
8.5.	Настройка порогов функции многоуровневого квантования.....	134
	Библиографический список.....	137

Учебное издание

Бахытжан Сражатдинович Ахметов
Александр Иванович Иванов
Александр Юрьевич Малыгин
Вячеслав Александрович Фунтиков

ОСНОВЫ БИОМЕТРИЧЕСКОЙ
АУТЕНТИФИКАЦИИ ЛИЧНОСТИ

Учебное пособие

Нач. РО УИЦ	<i>З. А. Губайдулина</i>
Редактор	<i>З. А. Губайдулина</i>
Компьютерная верстка	<i>А.Б. Аришова</i>

Подписано в печать 24.10.2014 г.

Тираж 300 экз. Формат 60х84х 1/16. Бумага типогр. № 1.

Уч.-изд.л. 9,5. Усл. п.л. 8,7. Заказ № 48. Цена договорная.

Издание Казахского национального технического университета

им. К. И. Сатпаева

Учебно-издательский центр КазНТУ,

г. Алматы, ул. Сатпаева, 22

ISBN 978-601-228-689-2



9