http://rosoperator.ru/rameev100/

Всероссийская научно-техническая конференция, посвященная 100-летию со дня рождения одного из основоположников советской вычислительной техники Б.И. Рамеева, на тему: «Информационно-управляющие и телекоммуникационные системы специального назначения». Секция «Биометрическая поддержка криптовалют, блокчейн реестров, облачных сервисов», проводится в конференц-зале Технопарка высоких технологий «Рамеев», доклад 16 мая 2018 года с 13^{00} до 13^{15} (ул. Центральная, д. 1), г. Пенза.

УДК: 519.2, 612.087, 621.319.7

Корнеев О.В. (ФАУ «ГНИИИ ПТЗИ ФСТЭК России»)

Обезличивание медицинских электронных документов при их хранении с привлечением облачных сервисов (технология SafeNet, разработка национальной биометрической платформы)

Аннотация.

Показано, что один из эффективных способов защиты персональных данных при их размещении в облачных хранилищах является обезличивание электронных документов, например, медицинских электронных историй болезни. При этом злоупотребления своим обезличиванием со стороны пациентов исключается за счет использования биометрической аутентификации личности. В этом случае пациент не может подменить себя другим человеком с другой биометрией, а врач не имеет технических возможностей дезавуировать обезличивание электронных медицинских документов. Защита больших объемов персональной медицинской информации строится на том, что злоумышленник, получивший незаконный доступ к облачному хранилищу не может узнать, какому человеку принадлежит та или иная история болезни.

Технология обезличивания электронных медицинских документов на текущий момент наиболее глубоко проработана. Она может быть использована как образец облачной организации национальной биометрической платформы (технология SafeNet)**.** Ожидается, что следующее поколение электронных паспортов граждан РФ должно храниться в облаках в обезличенной форме при этом владелец электронного биометрического паспорта знает его параметры хранения и легко может им воспользоваться. Злоумышленник же напротив для того, что бы добраться до электронного паспорта нужного ему человека будет вынужден перебирать очень большое число возможных вариантов, вскрывая защиту каждого из проверяемых вариантов.

Введение

В настоящее время активно идут процессы информатизации современного общества, как итог, наша персональная информация постепенно перемещается в Интернет облака. Ярким примером общего вектора развития являются медицинские информационные системы. В 2006 году в России был введен в действие отечественный стандарт, регламентирующий требования к электронной истории болезни [1], это позволило разработать типовую медицинскую информационную систему [2]. Далее, встал вопрос о переходе к использованию типового электронного места врача [3]. В итоге, информационная технология уже позволяет создавать интегрированные электронные медицинские карты [4], которые могут размещаться как на локальном сервере медицинской информационной системы, так и на Интернет серверах поставщиков облачных услуг.

Естественно, что эта общая тенденция порождает новые угрозы информационной безопасности, которые должны быть устранены с учетом уже сложившейся технической практики [5] и национального законодательства [6]. За рубежом проблема решается через биометрическую аутентификацию личности человека с использованием, так называемых, «нечетких экстракторов» [7]. В России для этих же целей используются искусственные

нейронные сети [8], применяя которые можно осуществлять обезличивание [9] медицинских электронных документов, в случае их размещения в облачных хранилищах.

Очевидным является так же то, что массовое использование биометрического обезличивания персональной информации облачных сервисов подпадает под юрисдикцию государственных регуляторов рынка средств информационной безопасности. От производителей потребитель в праве потребовать сертификаты ФСБ России на соответствующие криптографические модули защиты, а так же собственник облачного сервера в праве потребовать у производителя сертификат ФСТЭК России на соответствие нейросетевых преобразователей биометрии в код пароля доступа (если таковые используются в продукте).

В первом случае, экспертная организация ФСБ России (испытательная лаборатория) должна оценить корректность программной реализации модулей криптографической защиты информации, использованных для защиты облачных хранилищей. Во втором случае экспертная организация (испытательная лаборатория, аккредитованная ФСТЭК России) должна оценить корректность реализации нейросетевых преобразований на соответствие требованиям пакета национальных стандартов ГОСТ Р 52633.xx и, в том числе, стойкость к атакам подбора нейросетевой биометрической защиты по алгоритмам тестирования ГОСТ Р 52633.3-2011 [10].

Важнейшим технологически-правовым моментом всех облачных сервисов является то, что облачные услуги не подпадают под действие ФЗ № 152 «О персональных данных» [6], если персональные данные зашифрованы (имеется сертификат ФСБ России на средство шифрования) или персональные данные надежно обезличены (имеется сертификат ФСТЭК России на средство биометрической аутентификации, исключающей подмену обезличенного пользователя).

То есть, владелец облачного сервиса не является «оператором» персональных данных и не осуществляет «обработку» персональных данных. От владельца облачного сервиса требуется не более, чем физическое размещение его оборудования (серверов и дата центров) на территории Российской Федерации.

Оператором персональных данных, в нашем случае является, медицинская организация, владеющая медицинской информационной системой (МИС). Именно медицинская организация должна уведомить Уполномоченный орган Роскомнадзора по защите прав субъектов персональных данных для последующей регистрации в качестве оператора по обработке персональных данных. Именно медицинская организация должна подготовить техническое задание по созданию системы защиты персональных данных и затем убедить Уполномоченный орган Роскомнадзора о достаточном уровне защиты персональных данных в медицинской информационной системе, опираясь на наличие сертификатов ФСБ России и/или ФСТЭК России.

Актуальная модель угроз для персональных данных пациентов медицинской информационной системы

В прошлом веке регистратура поликлиник имела стеллажи, заполненные медицинскими картами пациентов. К врачу можно было попасть только после того, как в регистратуре найдут твою медицинскую карту. Похитить несколько тонн бумаги незаметно было технически невозможно. Сегодня технология изменилась, личная электронная медицинская карта хранится на сервере медицинской информационной системы. Системный администратор медицинской информационной системы, увольняясь с очередного места работы, легко может прихватить с собой архив медицинских карт всех пациентов с актуальной информацией за несколько лет.

Более того, в рамках наиболее продвинутых медицинских учреждений США отмечены случаи, когда в их информационной системе появлялись вирусы, уничтожающие всю актуальную информацию и параллельно шифрующие архивы на новом ключе. Администрация любого респектабельного лечебного учреждения, как

правило, не стремиться к поиску и публичному наказанию виновных. Для нее куда более важным является восстановление медицинского технологического процесса.

Следует подчеркнуть, что столь радикальные технические решения, как тотальное шифрование не всегда оправдано. Даже в специализированных медицинских учреждениях, ориентированных на лечение социально-значимых заболеваний, диагнозы и методы лечения повторяются. Вполне достаточно изъять из всех электронных медицинских документов персональные данные пациентов, позволяющие найти человека помимо его воли [11].

Злоумышленник, похитивший архив электронных медицинских документов, сможет узнать цену всех медицинских услуг, оказанных пациенту, он будет иметь полную информацию о ходе лечения и его результатах. Однако, он не сможет узнать самого главного – кто является пациентом, какой номер страхового медицинского полиса у пациента. Этих данных нет в медицинской информационной системе, они хранятся отдельно.

Для технологии обезличивания особой роли не играет, где будут размещены данные (обезличенные архивы). После обезличивания электронные истории болезни могут открыто храниться, как на локальном сервере медицинской информационной системе без доступа к сети Интернет, так и на сервере провайдера облачных услуг. Эти две ситуации отображены на рисунке 1.

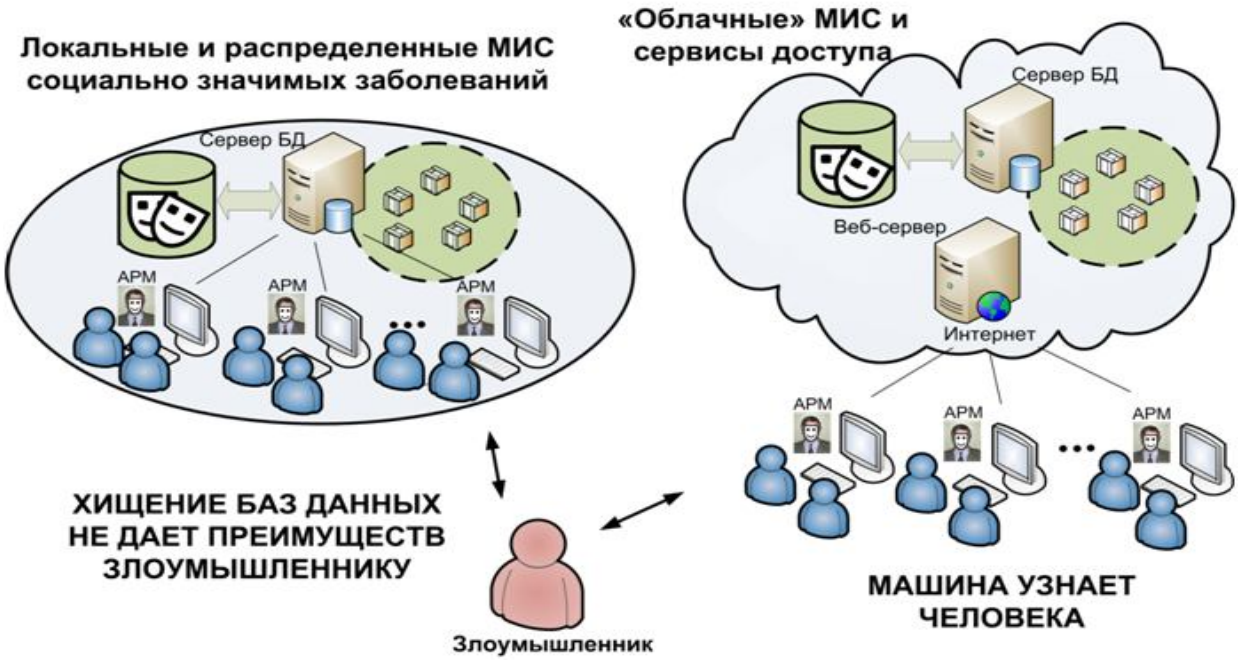

Рис. 1. Два варианта размещения обезличенных электронных медицинских документов на локальном сервере медицинской информационной системы и в облачном хранилище

И в том и в другом случае, злоумышленник не имеет возможности добраться до информации о полноценных персональных данных человека. Чем чаще встречается заболевание, тем надежнее оказывается защита наших персональных данных через обезличивание электронных медицинских документов.

Фактически, обезличивание информации медицинских электронных документов является защитой от злоупотреблений со стороны персонала, обслуживающего пациентов (системного администратора МИС, провизора аптеки, техников-лаборантов при исследовании биологических проб, группы лечащих врачей по нескольким медицинским специализациям). При этом, опираться на «клятву Гиппократа» можно только рассматривая категорию – «врачи», иной медико-технический персонал будет связан не

более чем некоторыми обязательствами о неразглашении сведений «Для служебного пользования», принятыми по отношению к работодателю при трудоустройстве.

Новая информационная технология позволяет перейти от обычных организационных мероприятий по защите персональных данных к отсутствию технических возможностей у всего медико-технического персонала медицинской организации. Естественно, что в этой новой ситуации возникают и новые угрозы.

Основной угрозой, порождаемой обезличиванием электронных медицинских документов является злоупотребления со стороны пациентов. Пациент может быть заинтересован в фальсификации объективных данных о своем здоровье. Кому-то хочется быть здоровым и такие люди могут попытаться подменить себя здоровым человеком. Возможна противоположная ситуация, когда здоровый человек, по каким-то причинам хочет казаться больным.

Биометрическая поддержка обезличивания медицинских электронных документов

Люди способны практически безошибочно узнавать друг друга. Эту способность долгое время принято было рассматривать как прерогативу естественного интеллекта. Однако, начиная с конца прошлого века по всему миру активно стали развиваться биометрические технологии [12, 13]. Очевидным лидером этих процессов были США [12], поставившие цель создания и продвижение нового поколения биометрических паспортно-визовых документов. Сформулированная США цель была достигнута и сегодня большинство стран имеет биометрическое усиление своих международных паспортов. Российский международный паспорт, как и паспорта других стран, имеет встроенный радио-читаемый процессор с биометрическими данными своего владельца.

Одной из неприятностей биометрических технологий США, стандартизированных ISO/IEC JTC1 SC 37 (Biometrics), является полное отсутствие гарантий информационной безопасности. Так называемые «биометрические шаблоны», хранящиеся в радио-читамых процессорах био-паспортов, полностью повторяют идеологию полицейской биометрии, ориентированную на поиски преступников. По этой причине открытые «биометрические шаблоны» ISO/IEC JTC1 SC 37 (Biometrics) нельзя размещать на Интернет серверах облачного хранилища.

В связи с этим, появилось специальное направление исследований, которое занимается защищенным вариантом исполнения средств биометрической аутентификации [7, 8, 13]. Защищенный вариант в форме зарубежных «нечетких экстракторов» [7] или отечественный вариант нейросетевых преобразователей биометрия-код [8, 13] изначально ориентированы под безопасное хранение персональных биометрических данных человека на сервере МИС или в облаках. На рисунке 2 иллюстрируется технология нейросетевого преобразования того или иного биометрического образа в код, используемый при последующей криптографической аутентификации.

Каждый из биометрических образов перед его использованием должен быть отсканирован, далее, из биометрического образа должны быть извлечены контролируемые биометрические параметры. Эти вычисленные параметры подаются на входы заранее обученной искусственной нейронной сети, которая для любого примера образа «Свой» порождает на своих выходах длинный код аутентификации (код личного криптографического ключа или длинного пароля доступа). Если злоумышленник - «Чужой» будет подставлять случайные биометрические образы, то на выходах нейронной сети он будет получать случайные выходные коды.

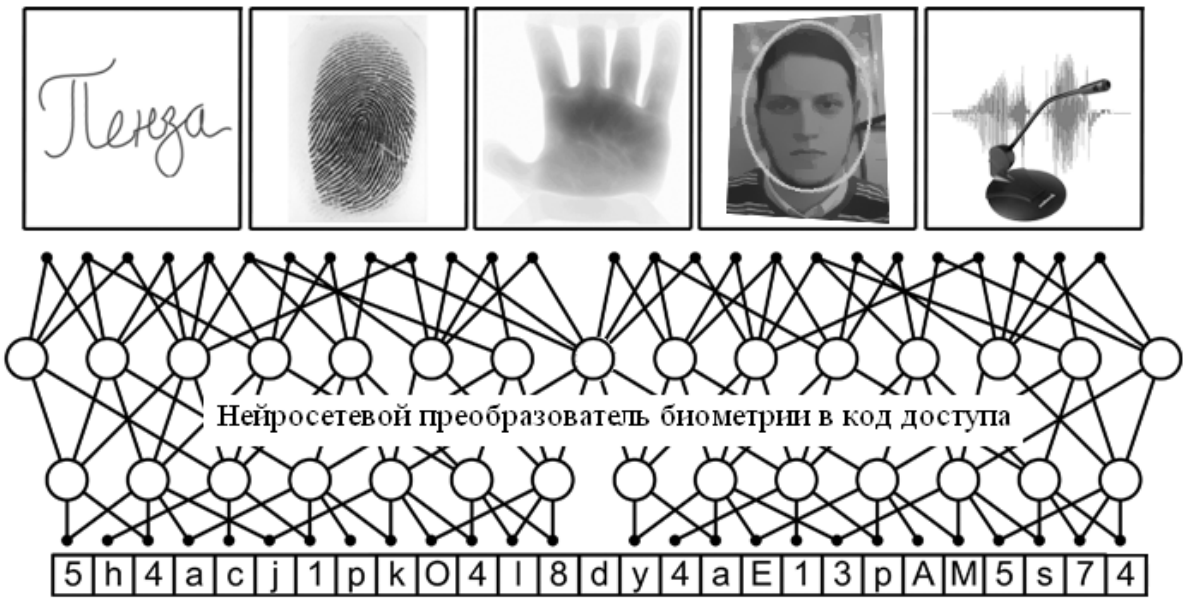

Рис. 2. Биометрические образы человека, преобразуемые искусственной нейронной сетью в код криптографического ключа или длинного пароля доступа

Основная технологическая функция биометрической поддержки обезличивания состоит в том, что в место медицинского персонала машина узнает человека. Незнакомый врачу пациент своей биометрией подтверждает свои полномочия без предъявления паспорта и объявления своего подлинного имени. Предъявив медицинской информационной системе идентификационную карту (аналог радио-читаемой карты, например, используемой в метро) и рисунок своего отпечатка пальца пациент подтверждает то, что он ранее официально зарегистрирован. Как минимум, существует бумажный договор, где указаны действительные персональные данные пациента в связке с идентификационным номером его идентификационной карты.

Пациент не может обмануть информационную систему, отдав свой идентификатор другому человеку. Врач не может принести вред пациенту через разглашение тайны его диагноза, так как он не знает подлинных персональных данных пациента.

Технология извлечения знаний из преобразователей биометрия-код

Большинству специалистов кажется, что размещение персональных биометрических данных в зарубежном «нечетком экстракторе» или в отечественной нейронной сети, является достаточно надежной защитой. На самом деле это не так, времена Шеннона давно прошли, каждый из нас имеет вычислительную машину, а хакеры имеют возможность использования сотен зомбированных серверов. Это позволяет хакерам организовывать достаточно сложные атаки на биометрию.

Целью атак является попытка извлечения знаний из «нечеткого экстрактора» или нейросетевого контейнера. Оба типа этих преобразователей по своему определению должны иметь область примеров образов «Свой» с нулевой энтропией выходных кодовых состояний. В связи с этим, мы можем заранее создать базу из 1024 образов «Чужой», подавать эти образы на вход преобразователя биометрия-код. Далее, мы можем рассчитать энтропию каждого их полученных кодов по отношению ко всем иным кодам. Результат таких вычислений приведен на рисунке 3.

Из рисунка 3 видно, что энтропия всех кодов образов «Чужой» находится в интервале от 4 бит до 19 бит (данные отложены по вертикальной оси). При этом расстояние Хэмминга между выходными кодами меняется от 0 до 256 бит (данные горизонтальной оси). С право и с лево находятся коды с минимальной энтропией, наиболее близкие к коду образа «Свой» и его инверсии.

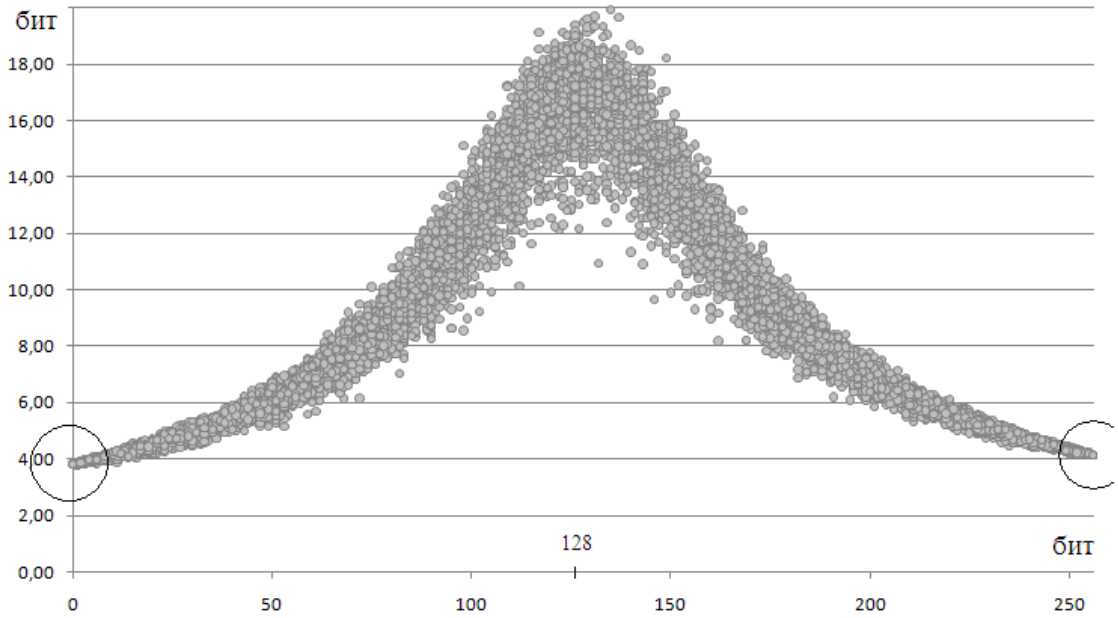

Рис. 3. Распределение частных значений энтропии выходных кодов 1024 биометрических образов «Чужой» для нейросетевого преобразователя биометрия-код

Мы не знаем, какой код дает образ «Свой», однако мы можем выделить две группы кодов-претендентов и, соответствующие им группы образов «Чужой». Обычно в первой и второй группах оставляют по 20 образов, что соответствует 2% от исходной базы. Этот процесс можно повторить, восстановив численность образов «Чужой» до первоначальной численности. Восстановление численности выполняется скрещиванием двух образов-родителей и получением от них образов-потомков по ГОСТ Р 52633.2 [14].

Процедуры скрещивания и селекции следует повторять в нескольких поколениях, что позволяет извлечь порядка 97% достоверной информации об образе «Свой» и коде, им порождаемом. При этом для извлечения знаний из «нечеткого экстрактора» потребуется 5 поколений (10 минут машинного времени обычного компьютера), а для извлечения знаний из нейронной сети потребуется 50 поколений (50 минут машинного времени).

Гарантии безопасности хранения личной биометрии пациентов на облачном сервере

Очевидно, что 10 или 50 минут машинного времени для злоумышленников не является сколько-нибудь существенным препятствием. То есть, «нечеткие экстракторы» и нейросетевые преобразователи нельзя хранить на облачных серверах. «Нечеткие экстракторы» и нейросетевые преобразователи следует рассматривать как аналоги обычных паролей, вводимых с клавиатуры. В вычислительной машине или на сервере нельзя хранить пароли – это опасно. В вычислительной машине и на сервере обычно хранят хэш-функции от пароля. То есть, безопасно хранить на сервере мы можем некоторую последовательность хэш-функций от нейросетевого преобразователя биометрия-код.

Для пояснения технологии защиты на рисунке 4 приведена обобщенная структура нейросетевого преобразователя биометрия-код.

Каждый нейрон обученного преобразователя имеет две таблицы. Первой является таблица связей входов нейрона со входами нейронной сети в целом. Второй

является таблица весовых коэффициентов сумматора для каждого из нейронов. Этих двух таблиц достаточно.

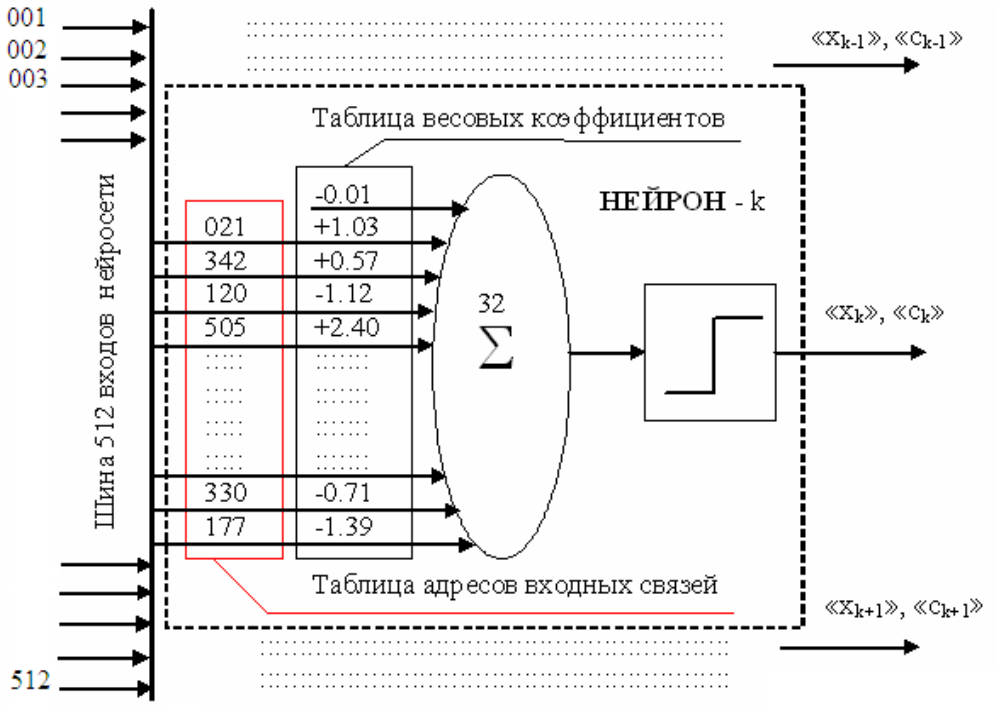

Рис. 4. Обобщенная структура нейросетевых преобразователей

Стандартный алгоритм обучения нейронной сети ГОСТ Р 52633.5-2011 [15] предполагает задание таблиц выходных связей от генератора псевдослучайных чисел. Для того, что бы защитить нейросетевые контейнеры, достаточно выработать гамму, накрывающую первую и вторую таблицы. Для защиты таблиц первого нейрона гамма вырабатывается хэшированием конкатенации обычных данных аутентификации:

$$\Gamma_1 = \text{HASH}(\text{соль}|\text{пароль}) \qquad (1).$$

Этого достаточно для снятия защиты с таблиц первого нейрона. После открытия таблиц может быть выполнена операция по преобразованию биометрического образа первым нейроном и получено его выходное состояние «c_1».

Гамма для таблиц второго нейрона вычисляется с учетом выходного состояния первого нейрона:

$$\Gamma_2 = \text{HASH}(\text{соль}|\text{пароль}|c_1) \qquad (2).$$

На каждой итерации число учитываемых выходных состояний нейронов увеличивается. Так для i-го нейрона, используемая гамма будет строиться хэшированием конкатенации (i-1) состояния предшествующих нейронов:

$$\Gamma_i = \text{HASH}(\text{соль}|\text{пароль}|c_1|c_2|\ldots|c_{i-1}) \qquad (3).$$

Очевидно, что любая ошибка при вводе пароля или ошибочная подстановка другого биометрического образа будет приводить к тому, что таблицы обученной нейронной сети восстановить не удастся. Как следствие. не удастся восстановить и выходной код нейросетевой биометрической аутентификации, поддерживающей обезличивание.

То, что таблицы нейросетевого контейнера защищены гаммированием, а гаммы создаются вызовом полноценных криптографических хэш-функций, является гарантией

безопасного хранения биометрических данных на сервере провайдера облачных услуг. Из-за защиты таблиц обученных нейронных сетей гаммированием, хакерам уже не удается наблюдать неоднородности энтропии кодов нейросетевого преобразователя (рисунок 3). Энтропия выходных кодов перестает иметь какие либо особенности, атакующий вынужден решать вычислительно сложную задачу по восстановлению неизвестных данных после их хэширования (1), (2), (3)

Типовая структура МИС, использующая защиту электронных медицинских документов через их обезличивание

В целом структура медицинской информационной системы изменяется незначительно. Ее безопасность строится на применении асимметричной криптографии (поддерживается инфраструктура открытых ключей), что иллюстрируется рисунком 5.

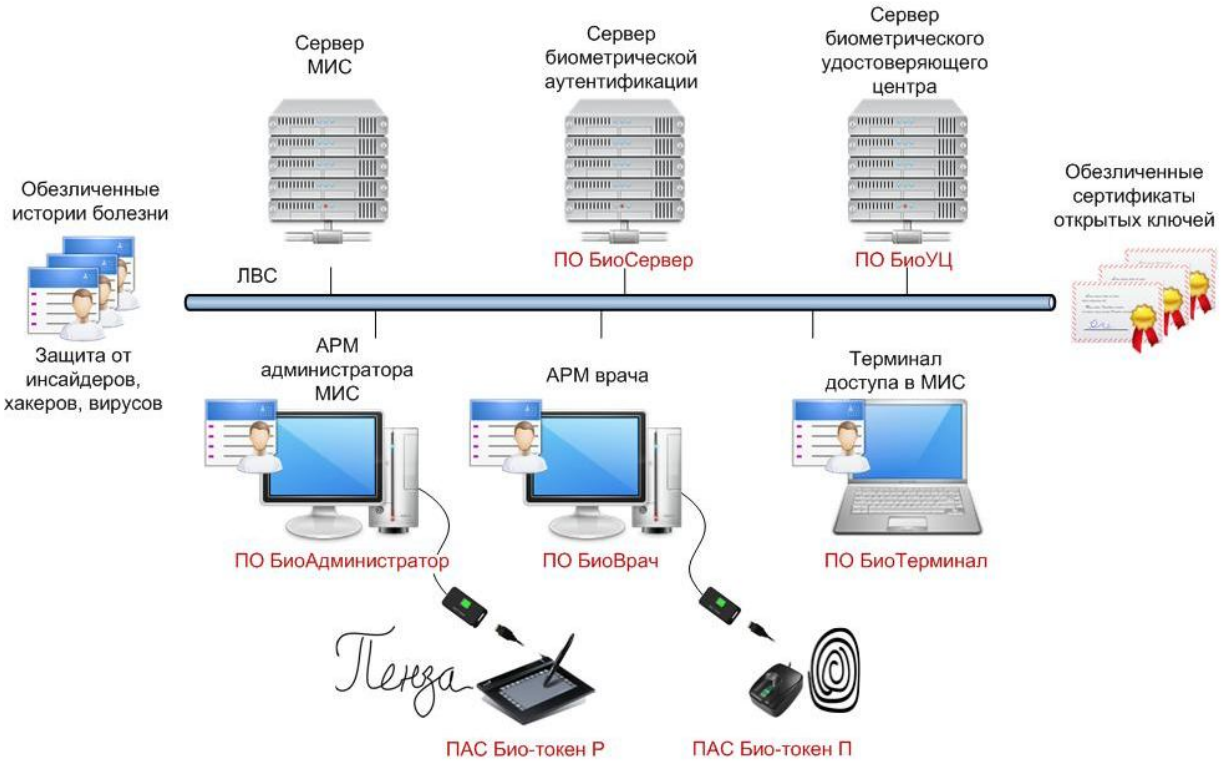

Рис. 5. Структура медицинской информационной системы, поддерживающей обезличенные электронные документы (обезличенные сертификаты открытых ключей)

По сравнению с традиционными техническими решениями в структуре МИС появляется сервер биометрической аутентификации и сервер биометрического удостоверяющего центра. В место обычных ЛИЧНОСТНЫХ сертификатов открытых ключей в МИС используются обезличенные сертификаты. Во всем остальном применяются типовые аппаратно-программные решения. Существующей сегодня продукт [11], пока ориентирован под две биометрические технологии (рукописный почерк и анализ рисунка отпечатка пальца). Датчики съема биометрической информации стандартные, сканер биометрии к вычислительной машине подключается через доверенную вычислительную среду USB «БиоТокен». В доверенной вычислительной среде USB «БиоТокен» выполняются все потенциально опасные операции:

- распаковка защищенного нейросетевого контейнера;
- обработка биометрических данных;
- формирование и проверка цифровой подписи.

Хищение обезличенных историй болезни и иных электронных документов, заражение вирусами МИС для описанного выше технического решения не опасны. МИС, построенные на подобных принципах могут использовать типовое рабочее место врача [2, 3] с интегрированной электронной медицинской картой [4], хранящейся как на локальном сервере, так и на облачных серверах.

Технология SafeNet, разработка национальной биометрической платформы

Общей тенденцией развития информационного общества является создание больших облачных хранилищ данных. Облачный сервис удобен тем, что пользователь может обращаться к своим данным, оставаясь мобильным. Одним из перспективных направлений развития является переход к использованию электронных биометрических паспортов и удостоверений личности, хранящихся, например, на облачном сервере ФМС России.

У любого гражданина РФ появляется возможность не носить с собой паспорт или удостоверение личности. При необходимости гражданин РФ может доказать наличие у него прав, скачав свой электронный паспорт с облачного сервиса ФМС России. При этом электронные документы граждан должны храниться в обезличенной форме, что является одной из дополнительных степеней защиты информации. Настоящий владелец электронного биометрического паспорта знает его параметры (знает параметры его хранения) и легко может им воспользоваться. Злоумышленник же, напротив, для того, что бы добраться до электронного паспорта нужного ему человека будет вынужден перебирать очень большое число возможных вариантов, вскрывая защиту каждого из проверяемых вариантов.

Естественно, что система обезличивание электронных паспортов может быть многоуровневой и иметь динамическую подсистему смены адресов хранения документов после каждого авторизованного обращения к облачному хранилищу. Принципиально важным является то, что обезличивание электронных документов и регулярная смена их адресов хранения в системе является достаточно эффективной дополнительной степенью защиты персональных данных.

Опытно-конструкторская работа «Лекарь» [11] (выполненная в 2014-2015 г.г.) показала, что владелец электронного хранилища за счет обезличивания электронных документов многократно снижает требования к криптографическим механизмам защиты каждого отдельного электронного документа. Формально по ФЗ № 152 «О персональных данных» собственник облачного хранилища юридически не должен получать от пользователей их согласия на обработку персональных данных, содержащихся в обезличенных и криптографически защищенных нейросетевых контейнерах. В частности ФМС России при обезличенном Интернет хранении электронных паспортов, защищенных отечественной криптографией перестает играть роль «оператора персональных данных» по ФЗ № 152. Оборудование, на котором развернуто облачное хранилище не нуждается в лицензировании.

ЛИТЕРАТУРА:

1. ГОСТ Р 52636-2006. Электронная история болезни. Общие положения.
2. Федеральная типовая медицинская информационная система (ФТМИС). Разработчик «Крокус Консалдинг» 2008 г., государственный контракт по ФЦП «Электронная Россия (2002-2010 годы).
3. Электронное рабочее место врача. Руководство пользователя. Москва-2014 г.
4. Зингерман Б.В., Шкловский-Корди Н.Е., Карп В.П., Воробъев А.И. Интегрированная электронная медицинская карта: задачи и проблемы. //Врач и информационные технологии. №1, 2015 г., с. 24-27.

5. Костков Д. Защита облачных вычислений: общие международные подходы. //Первая миля. №8, 2015 г.
6. Федеральный закон «О персональных данных» от 27.07.2006 № 152.
7. Dodis Y., Reyzin L., Smith A. Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy // Proc. EUROCRYPT, April 13, pages 523-540, 2004.
8. ГОСТ Р 52633.0-2006 «Защита информации. Техника защиты информации. Требования к средствам высоконадежной биометрической аутентификации».
9. Методические рекомендации по применению приказа Роскомнадзора от 5 сентября 2013 г. № 996 «Об утверждении требований и методов по обезличиванию персональных данных».
10. ГОСТ Р 52633.3-2011 «Защита информации. Техника защиты информации. Тестирование стойкости средств высоконадежной биометрической защиты к атакам подбора».
11. Разработка аппаратно-программного комплекса обезличивания биометрических данных больных социально-значимыми заболеваниями, содержащихся в цифровых медицинских документах, обрабатываемых в медицинских информационных системах России и Беларуси. Разработчик – ФАУ «ГНИИИ ПТЗИ ФСТЭК России» (2014-2015 годы), государственный контракт по Программе Союзного государства «Совершенствование системы защиты общих информационных ресурсов Беларуси и России на основе высоких технологий» на 2011-2015 годы.
12. Болл Руд и др. Руководство по биометрии. / Болл Руд, Коннел Джонатан Х., Панканти Шарат, Ратха Налини К., Сеньор Эндрю У. // Москва: Техносфера, 2007. -368 с., (перевод с английского).
13. 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 г. 157 с. IBSN 978-5-88070-044-8.
14. ГОСТ Р 52633.2-2010 «Защита информации. Техника защиты информации. Требования к формированию синтетических биометрических образов, предназначенных для тестирования средств высоконадежной биометрической аутентификации»
15. ГОСТ Р 52633.5-2011 «Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия-код доступа».

Корнеев Олег Владимирович – старший научный сотрудник ФАУ «ГНИИИ ПТЗИ ФСТЭК России», г. Воронеж, E-mail: visualfire@rambler.ru.