Нечеткий экстрактор для генерации ключей шифрования на основе параметров клавиатурного почерка[1]

А.Е. Сулавко, А.В. Еременко, Е.В. Толкачева, С.С. Жумажанова

Аннотация. Статья посвящена проблеме генерации ключевых последовательностей на основе клавиатурного почерка пользователей компьютерных систем. В рамках проведенного исследования предложено несколько вариаций нечетких экстракторов для решения выдвинутых задач. Проведена серия вычислительных экспериментов по оценке эффективности предложенных методик, определены оптимальные параметры нечетких экстракторов.

Ключевые слова: нечеткие экстракторы, генерация криптографических ключей на основе биометрических данных, распознавание образов, клавиатурный почерк, идентифицирующие признаки

Введение

В последние годы с повестки дня мировых СМИ не сходят новости об утечках конфиденциальной информации на уровне правительственных и общественных организаций. Анализ ситуации показывает, что злоумышленники, в основном, используют методы социальной инженерии, в то время как владельцы конфиденциальной информации пренебрегают шифрованием данных или используют слабые пароли для доступа к ним. По данным Министерства предпринимательства, инноваций и ремесел Великобритании и PWC 76% сетевых атак на компании стали возможны из-за ненадежных или украденных паролей. Согласно имеющейся статистике, число происшествий, связанных с хищением информации, растет с каждым годом. По данным глобального исследования информационной безопасности предприятий, проведенного PricewaterhouseCoopers, в 2014 году количество инцидентов в сфере безопасности увеличилось на 25%. При этом средний уровень финансовых убытков вырос на 18% [1]. По данным аналитических исследований компании Zecurion 2015 года удельный ущерб (на каждую крупную утечку) составил 25,29 млн. долл. В сумме потери мировой экономики от утечек конфиденциальной информации вследствие несанкционированного доступа исчисляются десятками миллиардов долларов в год [2]. Указанные данные в существенной степени подтверждаются другой известной крупной компанией InfoWatch [3]. Все источники говорят об одном - требуется повысить защищенность от угрозы неавторизованного доступа со стороны внутренних и внешних злоумышленников.

Традиционным способом защиты информации от несанкционированных воздействий является шифрование. Современные алгоритмы шифрования предоставляют достаточно высокий уровень защищенности. Однако вопросы

[1]Работа выполнена при финансовой поддержке РФФИ (грант №15-07-09053)

выбора ключей для асимметричного и симметричного шифрования, а также их защиты во время хранения и передачи отнюдь не тривиальны, их проработка требует внушительных финансовых затрат. Если найти устойчивые преобразования для осуществления однозначной и неотъемлемой «привязки» ключа для шифрования к биометрическим характеристикам каждой конкретной личности, данные вопросы можно будет считать закрытыми. Настоящая работа направлена на разработку методов «привязки» клавиатурного почерка человека к криптографическим ключам и посвящена поиску таких преобразований. В отличие от статических признаков, для хищения и изготовления «муляжа» которых на сегодняшний день существует множество способов, составить образ клавиатурного почерка и фальсифицировать портрет работы пользователя на компьютере если и возможно, то очень проблематично. Парольное слово (или фразу) можно сохранить в тайне, а также изменить, что невозможно при использовании открытых статических биометрических образов (радужная оболочка глаза, геометрия ладони) либо варианты замены ограничиваются количеством органов субъекта (отпечаток пальца). Данная особенность усиливает защитные свойства биометрических систем такого класса. Существенным недостатком является низкая надежность принимаемых биометрической системой решений вследствие низкой стабильности динамических признаков и их изменчивости со временем.

Формирование базы признаков клавиатурного почерка

Привлекательность клавиатурного почерка обусловлена простотой реализации технологий генерации ключа, идентификации и аутентификации на его основе. Данные процедуры не требуют специального оборудования и являются привычными для пользователя. В работе [4] предложено использовать в качестве признаков клавиатурного почерка временные интервалы между нажатием клавиш, характеризующие темп работы с клавиатурой, и временные интервалы удержания клавиш, характеризующие стиль работы с клавиатурой. Информативность парольной фразы определяется ее длиной. Временные интервалы содержат информацию о субъекте, имеющем выработанный клавиатурный почерк. В процессе обучения оператор подбирает удачные решения задачи набора текста на клавиатуре и запоминает их путем многократных повторений. Программы управления мышцами запоминаются в подсознательной сфере субъекта и реализуются автоматически. Временные интервалы характеризуются нормальным законом распределения [5-7]. Известно, что время между нажатием клавиш является информативным признаком в том случае, если клавиши достаточно удалены друг от друга [5]. Данную особенность можно объяснить на основании закона Фиттса [8]: чем дальше или точнее выполняется движение, тем больше коррекции необходимо для его выполнения, и соответственно, больше времени требуется для внесения этой коррекции, при внесении коррекции движений проявляются индивидуальные особенности человека.

Для сбора данных клавиатурного почерка привлечено 80 испытуемых. Каждым испытуемым не менее чем по 50 раз были введены следующие парольные фразы: «разрешить доступ к информационной системе», «прошу предоставить доступ к информации», «авторизация пользователя компьютерной системы», а также одну фразу, выбранную субъектом самостоятельно. Фраза должна состоять не менее чем из 21, но не более 42 символов (включая пробелы), данная длина парольной фразы предложена в [5] как рекомендуемая (слишком длинные парольные фразы сложно запоминаемы и воспроизводимы, велика вероятность ошибки при наборе фразы на клавиатуре). Некорректные реализации фраз (введенные с явными отклонениями) были исключены при помощи способа, предложенного в работе [9] и адаптированного для клавиатурного почерка. Таким образом, было собрано 12000 реализаций известных парольных фраз и 4000 реализаций тайных парольных фраз. Кроме того, каждым субъектом был произведен непрерывный ввод произвольного текста, состоящего из 4500 символов. При вводе регистрировались коды клавиш, времена удержания и паузы между нажатием клавиш. Во всех случаях использовались

Рис. 1. Реализация клавиатурного почерка на основе парольной фразы

обычные традиционные клавиатуры для офисной работы с механическими переключателями для замыкания соответствующих участков электронной цепи. Каждый субъект при вводе паролей и текста использовал одну клавиатуру.

Каждую парольную фразу решено представить в виде массива времен удержания и пауз между нажатием фактически нажатых клавиш (Рис. 1). Время удержания каждой отдельной клавиши незначительно зависит от последовательности нажимаемых клавиш в отличие от пауз между нажатиями клавиш, где эталонное описание требуется создавать для всех пар клавиш либо для наиболее информативных пар. При непрерывном вводе текста использовать паузы между нажатием сложно, т.к. требуется слишком долгая процедура обучения (пользователь должен напечатать большой объем текстов для создания эталона всех информативных сочетаний клавиш). Поэтому образцы клавиатурного почерка при вводе текста решено преобразовать в массив времен удержания клавиш (удалив информацию о паузах между нажатием клавиш). Таким образом, планируется проверить 2 подхода к генерации ключей на основе клавиатурного почерка: с использованием фиксированных парольных фраз (тайных и отрытых) и произвольного фрагмента текста.

Метод нечетких экстракторов и помехоустойчивое кодирование

«Нечетким экстрактором» называют метод (или общий алгоритм), выделяющий случайные, равномерно распределенные последовательности битов из биометрических данных в условиях зашумленности [10]. «Нечеткие экстракторы» способны компенсировать ошибки, возникающие вследствие технической невозможности получения одинаковых значений биометрических характеристик при их повторном вводе субъектом. Такие алгоритмы базируются на теории информации и помехоустойчивом кодировании и обычно используются для генерации криптографических ключей без необходимости их хранения в промежутках между обращениями к ним [10].

Общая концепция построения такого генератора заключается в следующем. Изначально случайным образом генерируется битовая последовательность, которая кодируется помехоустойчивым кодом [11]. В качестве кодов, исправляющих ошибки, могут использоваться коды Хемминга, Адамара, Боуза — Чоудхури — Хоквингема (БЧХ-коды), Рида-Соломона (являются частным случаем БЧХ) [12]. Сгенерированная битовая последовательность может быть предназначена для идентификации, аутентификации или генерации криптографических ключей шифрования. Данная последовательность объединяется с эталонными характеристиками биометрических признаков субъекта (биометрическим эталоном), в качестве которых в настоящей работе предполагается использовать вектор средних значений (математических ожиданий) выбранных признаков. Способы объединения могут быть различными – от простого сложения по модулю 2 до использования нетривиальных алгоритмов, учитывающих конфигурацию распределений зна-

чений признаков. Результатом объединения является открытая строка, которая может храниться на общедоступном сервере. Чтобы получить сгенерированную ранее последовательность (ключевой материал – ключ шифрования или ключ доступа, в зависимости от назначения экстрактора) субъект вводит новую реализацию биометрических признаков, которая обрабатывается соответствующим образом и «вычитается» из открытой строки (по принципу, обратному способу объединения) для «отсоединения» биометрических данных. После «отсоединения» полученная битовая последовательность будет изменена вследствие отличия предъявленных биометрических данных от эталонных. После применения кода, исправляющего ошибки к полученной строке, в случае высокой степени «схожести» предъявленного биометрического образа и эталонного (т.е., если количество несовпадающих бит эталонных и предъявленных значений признаков не превысит исправляющую способность кода), будет найдена исходная последовательность битов, которая и является ключевым материалом [13]. Описанный метод иллюстрируется на Рис. 2. Если предъявленные биометрические данные будут достаточно близки к эталонным, то исходная случайная строка и восстановленная строка будут равны (S1=S9, Рис. 2) и будет сгенерирован верный ключ. На практике можно осуществлять хеширование восстанавливаемой строки, если требуется получать на выходе строку фиксированной длины, однако стойкость генерируемого ключа определяется длиной кодируемой (восстановленной) строки в исходном виде. Под близостью подразумевается количество допущенных ошибок, которое может исправить помехоустойчивый код (количество отличающихся бит, байт, значений признаков – это зависит от используемого алгоритма кодирования и декодирования, т.е. процедур 3.1 и 3.2, Рис. 2). Например, в метрике Хемминга, если каждый признак кодируется 8 битами, близость образов определяется по формуле: $HD=E \oplus R/(n \cdot 8)$, где E – вектор эталонных значений признаков, R – вектор предъявляемых значений признаков.

Представленная типовая структура нечеткого экстрактора позволяет производить модификации по трем основным направлениям:

1. Способы представления эталонных и предъявляемых значений признаков (процедуры 1.1 и 1.2 на Рис. 2).
2. Способы «объединения» и «разделения» битовых последовательностей и закодированной строки (процедуры 2.1 и 2.2 на Рис. 2).
3. Способы помехоустойчивого кодирования и декодирования генерируемых случайных строк (процедуры 3.1 и 3.2 на Рис. 2).

Первое направление подразумевает разработку способов округления, трансляции, кодирования или иных преобразований над значениями признаков, результатом которых будет битовая последовательность, зависящая от биометрических характеристик. Очевидно, что не все биты, которыми представлено значение произвольного признака, являются информативными и значимыми. Изначально все значения признаков представлены десятичными числами с различной областью значений. Каждый

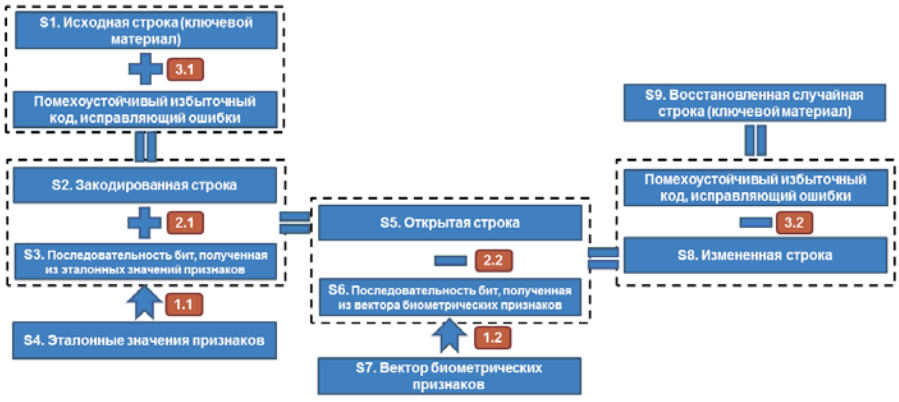

Рис. 2. Типовая схема нечеткого экстрактора

раз при вводе данных значение признака у каждого субъекта меняется (значительно или нет – все зависит от информативности данного признака для данного субъекта). Целесообразно минимизировать неинформативные биты или признаки при их преобразовании в битовую последовательность. В работе [14] описаны уязвимости нечетких экстракторов, которые можно кратко сформулировать следующим образом: нечеткие экстракторы, осуществляющие гаммирование «сырых» биометрических данных, оказываются без защиты от наблюдения реальных многомерных статистик в пространствах расстояний Хэмминга, средней стабильности разрядов, модулей корреляции, энтропии и их комбинаций. Если брать значения признаков в исходном виде, меняя лишь их представление на двоичное, а потом наложить гамму (сложить по модулю 2 битовый вектор со случайной строкой) для формирования открытой строки, то атаку перебора значений признаков можно существенно (многократно) ускорить, используя различные техники. Поэтому очень важно преобразовывать «сырые» значения признаков перед гаммированием таким образом, чтобы количество ошибок при повторном вводе данных субъектом в битовом векторе было минимальным и было очень сложно выявить нестабильные области в последовательности бит, извлеченной из «сырых» биометрических данных.

Решено кодировать каждое исходное значение признака одним байтом. Для этого предложено 3 способа, которые планируется протестировать в сочетании с различными алгоритмами кодирования. Каждый способ представляет собой отображение области значений признака на множество Y: $y=f(x)$, где $x \in X$, X – множество возможных значений признака, $y \in Y$. Различие способов определяются областью преобразованных значений признака. Для первого способа $y \in \{0, 1, 3, 7, 15, 31, 63, 127\}$, для второго – $y \in \{0, 1, 3, 7, 15, 31, 63, 127, 255, 254, 252, 248, 240, 224, 192, 128\}$, для третьего – $y \in [0;255]$. Выходное значение y представляется в двоичном виде. Предполагается, что область значений признаков может являться общедоступной информацией.

В работе [15], показана связь эффективности коррекции ошибок с методами группирования битов с разной вероятностью единичной ошибки. Предлагается процедура индивидуальной оценки информативности признаков с учетом относительной частоты появления единичных (или нулевых) бит в преобразованном значении признака. Под информативностью подразумевается интегральный показатель стабильности бит в преобразованном значении признака. Предложенную процедуру можно использовать совместно с любым способом отображения (преобразования) исходных значений признаков в битовую последовательность. Для каждого признака по всем отобранным для создания эталона преобразованным реализациям вычисляется относительная частота появления единичных (или нулевых) бит. Например, пусть введено 10 реализаций, необходимо определить показатель информативности для определенного признака по его преобразованным значениям: 00011111, 00111111, 00011111, 00001111, 00011111, 01111111, 00111111, 00011111, 11111110, 01111111. Относительная частота появления единичного бита в разрядах преобразованного значения составляет: 0,1; 0,3; 0,5; 0,9; 1; 1; 1; 0,9. Далее определяется интегральная вероятность появления единичного бита во всех разрядах, при этом относительная частота берется как вероятность, а вероятности равные 0 или 1 преобразуются в некоторое число, приближенное по значению к 0 и 1, но не равное им (чтобы общее произведение не стало равным нулю). Для рассматриваемого случая можно взять значения 0,01 и 0,99.

$0,1 \cdot (1-0,1) \cdot 0,3 \cdot (1-0,3) \cdot 0,5 \cdot (1-0,5) \cdot 0,9 \cdot (1-0,9) \cdot 0,99 \cdot (1-0,99) \cdot 0,99 \cdot (1-0,99) \cdot 0,99 \cdot (1-0,99) \cdot 0,9 \cdot (1-0,9) =$
$0,1 \cdot (0,9) \cdot 0,3 \cdot (0,7) \cdot 0,5 \cdot (0,5) \cdot 0,9 \cdot (0,1) \cdot 0,99 \cdot (0,01) \cdot 0,99 \cdot (0,01) \cdot 0,99 \cdot (0,01) \cdot 0,9 \cdot (0,1) =$
$0,000000000371357684775$.

Чем больше разрядов будет иметь частоты, близкие к 0 или 1, тем меньше получится итоговое произведение и тем выше интегральная оценка стабильности (информативности) признака для субъекта. Далее все признаки ранжируются по информативности (изменяется их порядок – от самого информативного к самому малоинформативному) и отбирается определенное количество признаков, остальные от-

брасываются. Оптимальное количество признаков, при котором вероятности ошибок 1-ого и 2-ого рода будут наименьшими – параметр экстрактора, который для каждой задачи будет различным. В настоящем исследовании данный будет определяться посредством серии вычислительных экспериментов для каждого исходного набора признаков. Количество информативных признаков и их последовательность требуется хранить на отдельном носителе или на выделенном сервере.

Способы «объединения» и «разъединения» битовых последовательностей (процедуры 2.1 и 2.2 на Рис. 2) обычно сводятся к операции сложения по модулю 2 (xor). Изначально планировалось разработать модифицированные способы на основе правил нечеткой логики [16-17], адаптировав для этого один из алгоритмов нечеткого вывода (Tsukamoto, Sugeno, Mamdani, Larsen и др.) [18-19]. Нечеткий вывод планировалось осуществлять как отображение функции принадлежности значения признака к субъекту (функция принадлежности должна определяться на основе функции плотности распределения признака или быть равной ей) на закодированную битовую последовательность (закодированную строку). Однако при таком подходе требуется хранить вспомогательную информацию о функциях принадлежности признаков, фактически требуется хранить эталон либо подробную информацию для его восстановления. Такой подход сводит на нет основное преимущество нечетких экстракторов – отсутствие необходимости в хранении эталона. Поэтому от разработки модификаций данных процедур пришлось отказаться.

Процедуры 3.1 и 3.2 (Рис. 2) определяются выбранными алгоритмами помехоустойчивого кодирования и декодирования. Существуют коды, способные исправлять одиночные и групповые ошибки, т.е. несколько ошибок за раз. Количество ошибок, которое может исправить код, называют исправляющей способностью кода. В нечетких экстракторах находят применение коды Хемминга, Адамара, Рида-Соломона и д.р. На данном уровне требуется обеспечить высокую степень восстановления ключевого материала из измененной строки (S8 на Рис. 2) при наименьшей избыточности кода, исправляющего ошибки. Требуется найти наиболее эффективный алгоритм и оптимальное соотношение исправляющей способности кода и других параметров, при котором вероятности ошибок 1-ого и 2-ого рода и избыточность помехоустойчивого кода будут наименьшими для заданного пространства признаков. Нужно учесть, что кодируемая последовательность (являющаяся кличем, S1 на Рис. 2) и закодированная строка (S2 на Рис. 2) не совпадают по длине, количество бит первой меньше, чем количество бит второй (S2>S1). При этом стойкость генерируемых ключей определяется длиной кодируемой строки (S1) и чем больше избыточность кода, тем длинней закодированная строка (S2) и тем больше требуется биометрических признаков, чтобы «покрыть» ее целиком при «объединении».

Коды Хемминга — простейшие линейные коды, способные исправить одну ошибку. Вследствие низкой исправляющей способности битовое представление кодируемую последовательность придется разбивать на составляющие, каждую из которых потребуется кодировать отдельно. При этом разбиение по 8 бит может оказаться недостаточным. Если поделить кодируемую строку на порции из 4-х бит, каждую из которых требуется кодировать отдельно (каждый признак будет представлять собой 2 кодируемых сообщения по 4 бита), то даже если в одном из признаков не совпадут более 2 бит – генерируемый ключ будет отличаться. При этом избыточность такого кода велика, она составляет 75%, т.е. на каждые 4 бита информации требуется еще 3 корректирующих бита. Поэтому применение кодов Хемминга для рассматриваемых задач априорно можно считать неэффективным.

Коды Адамара, обладая большим кодовым расстоянием, позволяют соответственно исправить и большое количество ошибок. Это достигается ценой высокой избыточности. Иногда такая цена приемлема. Для реализации кода Адамара необходимо построить матрицу Адамара соответствующего порядка. Для любого целого $n > 0$ квадратная матрица $H = (h_{ij})$ порядка n называется матрицей Адамара, если $h_{ij} \in \{+1, -1\}\, \forall i, j$ и $HH^t = nI$, где I — единичная матрица. Пример:

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

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

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

Образуем код из всех строк матрицы H и их отрицаний. Такой код (или его $\{0,1\}$-соответствие) называется кодом Адамара. Выберем в коде Адамара длины n кодовые слова, начинающиеся с $+1$, отбросим первые координаты и переведем в $\{0,1\}$-код. Результатом будет двоичный код длины $n-1$, размера n, являющийся эквидистантным с расстоянием $n/2$. Построенный код называется укороченным кодом Адамара. Нечеткий экстрактор на основе кода Адамара будет кодировать битовое представление случайной строки (секретного ключа), принимая в качестве параметра размер блока. От размера блока зависит исправляющая способность кода, оптимальный размер блока может быть найден в процессе эксперимента.

Коды БХЧ (Боуза — Чоудхури — Хоквингема) – это широкий класс циклических кодов, применяемых для защиты информации от ошибок при ее передаче по каналам связи. Код БЧХ отличается возможностью построения кода с заранее определёнными корректирующими свойствами, а именно, минимальным кодовым расстоянием. Этот код включен в формат POCSAG систем поискового радиовызова. Большинство циклических кодов используют один алгоритм построения помехоустойчивых кодовых комбинаций, а отличаются лишь методикой выбора образующего многочлена. В БЧХ-коде построение образующего многочлена, в основном, зависит от двух параметров: от длины кодового слова n и от числа исправляемых ошибок s. Особенностью кода является то, что для исправления числа ошибок $s >= 2$ еще недостаточно условия, что между комбинациями кода минимальное кодовое расстояние $d_{min} = 2 * s + 1$. Необходимо также, чтобы длина кода n удовлетворяла условию $n = 2h - 1$, где h - любое целое число. При этом n всегда будет нечетным числом и принимать значения: 1, 3, 7, 15, 31, 63, 127.. и т.д, т.е не все m могут быть заданы исследователем, как параметр. Кодирующий многочлен $g(x)$ для БЧХ-кода, длина кодовых слов которого n, строится так. Находится примитивный многочлен минимальной степени q такой, что $n \leq 2^q - 1$ или $q \geq \log_2(n+1)$. Пусть α - корень этого многочлена, тогда рассмотрим кодирующий многочлен $g(x) = НОК(m_1(x), \dots, m_{d-1}(x))$, где $m_1(x), \dots, m_{d-1}(x)$ - многочлены минимальной степени, имеющие корнями соответственно $\alpha, \alpha^2, \dots, \alpha^{d-1}$. Для декодирования могут применяться алгоритм Берлекемпа — Мэсси, Евклидов алгоритм, алгоритм Питерсона — Горенстейна — Цирлера (ПГЦ). Последний использовался в работе [13] и его было решено использовать в настоящем исследовании. Нечеткий экстрактор на основе кодов БЧХ кодирует битовое представление случайной строки целиком и принимает в качестве параметра исправляющую способность, оптимальное значение которой для каждого набора признаков будет вычисляться в процессе экспериментов. Важным и широко используемым подмножеством кодов БЧХ являются коды Рида-Соломона. Это такие коды БЧХ, у которых мультипликативный порядок алфавита символов кодового слова делится на длину кода. Далее решено ограничиться общей реализацией кодов БЧХ при создании экстракторов.

Определение наиболее эффективной методики генерации ключа на основе предложенных вариантов

Имеющиеся биометрические данные использовались для имитации процесса генерации секретных ключей, привязанных к субъекту. Структурная схема имитационной модели для проведения эксперимента представлена на Рис. 3. На первом этапе по равномерному закону генерируются сами ключевые последовательности (секретные ключи). Биометрические данные подаются в некотором количестве на вход экстрактора с заданными параметрами вместе с секретным ключом субъекта. На выходе экстрактора будет получена открытая строка. Оптимальное количество реализаций, необходимых для формирования открытой строки определяется в процессе эксперимента (в [20] указано, что для

Рис. 3. Реализация клавиатурного почерка на основе парольной фразы

создания эталона по клавиатурному почерку целесообразно использовать от 26 реализаций). Параметрами являются:

- способ округления;
- количество учитываемых (наиболее информативных) признаков;
- алгоритм кодирования/декодирования;
- размер блока (для кодов Адамара);
- исправляющая способность кода (для кодов БЧХ).

Таким образом, для каждого субъекта формируется открытая строка.

На втором этапе производится генерация секретных ключей, т.е. восстановление исходных сгенерированных секретных ключей из открытой строки и их сравнения с первоначальными ключами. За ошибку 1-ого рода (FRR, false reject rate) принимается ситуация, при которой система генерирует нехарактерное для субъекта (несовпадающее с оригиналом) значение ключа ($K^*_i \neq K_i$, Рис. 3), т.е. чем ниже вероятность ошибки 1-ого рода, тем выше стабильность выработки ключа. При такой ошибке на практике пользователь услуги не сможет расшифровать данные или правильно подписать информационный ресурс при помощи ЭЦП. За ошибку 2-ого рода (FAR, false acceptance rate) принимается ситуация, при которой ключи, полученные из биометрических данных двух различных субъектов, совпадают ($K^*_j = K_i$, Рис. 3). При такой ошибке может быть осуществлен несанкционированный доступ к зашифрованным данным, либо подделана ЭЦП. Коэффициентом равновероятной ошибки EER называется общий процент (вероятность) ошибочных решений, если EER=FRR=FAR.

Подаваемые на втором этапе данные субъектов могут быть получены непосредственно от самого субъекта либо сгенерированы при помощи метода Монте-Карло (либо на основе преобразования Бокса — Мюллера) под параметры закона распределения признаков для данного субъекта. Генерация значений обычно используется, когда недостаточно реальных данных для получения приемлемой статистической достоверности. Метод Монте-Карло не учитывает корреляционную зависимость между признаками и возникновение сбоев, поэтому при искусственной генерации реализаций субъектов, получаемые результаты могут оказаться несколько выше, чем на практике. В настоящем исследовании для обеспечения высокой достоверности результатов (свыше 0,99) на вход алгоритму генерации ключей подавались как реальные, так и сгенерированные реализации (40% от общего числа реализаций) для случая выработки ключей на основе непрерывного ввода текста. Для случая генерации ключей на основе парольных фраз использовались только реальные данные, полученные от субъектов (без генерации случайных величин). Процесс генерации ключей повторялся с различными сочетаниями параметров экстрактора для всех реализаций всех субъектов, имеющихся в базе. При оценке FAR для случая тайных биометрических образов в качестве ошибки принималась ситуация случайного совпадения генерируемого ключа i-ого субъекта с ключом j-ого субъекта

Табл. 1. Основные результаты по генерации ключевых последовательностей на основе клавиатурного почерка (достоверность результатов не менее 0,99)

Признаки/база образцов	СО	КР	ДК	КОД	FRR	FAR_1	FAR_2	ДИ
Парольная фраза (время нажатия и между нажатиями клавиш, 8000 образцов)	4	30	48	БЧХ	0,104	0,009	0,021	0,005
2 реализации фразы (время нажатия и между нажатием клавиш, 8000 образцов)	4	30	48	БЧХ	0,064	0,01	0,025	0,005
Непрерывный ввод текста 1500 символов (время нажатия клавиш, 400 образцов)	4	30	192	БЧХ	0,061	0,023	-	0,02
Ввод произвольного текста (1400 образцов, 26 субъектов) [21]					≈0,12		≈0,12	0,05
Парольная фраза [22]			12		≈0,484			

СО – способ округления
КР – количество реализаций при формировании открытой строки
ДК – длина генерируемого ключа в битах
КОД – название кода, исправляющего ошибки
FAR_1 – вероятность ошибки 2-ого рода для тайных биометрических образов;
FAR_2 – вероятность ошибки 2-ого рода для известных биометрических образов (подделок)
ДО(ДИ) – достоверность и доверительный интервал (последнее указывается в скобках)

при использовании в эксперименте реализаций различных для каждого субъекта парольных фраз (которые испытуемый выбирал самостоятельно). При оценке FAR для случая открытых (известных) биометрических образов в качестве ошибки принималась ситуация совпадения генерируемого ключа *i*-ого субъекта с ключом *j*-ого субъекта при использовании в эксперименте реализаций фиксированных парольных фраз («разрешить доступ к информационной системе», «прошу предоставить доступ к информации», «авторизация пользователя компьютерной системы»). При генерации ключей на основе текста случай тайных образов не рассматривается. Наилучшие результаты эксперимента и их сравнение с достигнутыми ранее можно видеть в Табл. 1.

Заключение

Генерацию ключа приемлемой длины при достаточной надежности возможно осуществить только на основе ввода непрерывного текста (не менее 1500 символов). Разработан способ генерации ключа длиной 192 бита на основе клавиатурного почерка субъекта, регистрируемого при непрерывном вводе текста (от 1500 символов), на базе кодов БЧХ с применением процедуры оценки признаков по информативности индивидуально для каждого субъекта с вероятностью ошибок генерации 1-ого и 2-ого рода 0,061 и 0,023, достоверность результата 0,99 при доверительных интервалах 0,005.

Разработан способ генерации ключа длиной 48 бит на основе клавиатурного почерка субъекта, регистрируемого при вводе парольной фразы (либо 2-х реализаций фразы), на базе кодов БЧХ с применением процедуры оценки признаков по информативности индивидуально для каждого субъекта, вероятности ошибок генерации 1-ого и 2-ого рода составили:

- FRR=0,104 (из 2-х реализаций – FRR=0,064);
- если пароль неизвестен – FAR=0,009 (из 2-х реализаций – FAR=0,01);
- если пароль известен – FAR=0,021 (из 2-х реализаций – FAR=0,025).

Достоверность результата 0,99 при доверительных интервалах 0,02.

Литература

1. Управление киберрисками во взаимосвязанном мире. Основные результаты Глобального исследования по вопросам обеспечения информационной безопасности. Перспективы на 2015 год. PricewaterhouseCoopers. Режим доступа: http://www.pwc.ru/ru_RU/ru/riskassurance/publications/assets/managing-cyberrisks.pdf.
2. Утечки конфиденциальной информации. Предварительные итоги 2014 года. Zecurion Analytics. 2015 г. – Режим доступа: http://www.zecurion.ru/upload/iblock/fe3/Zecurion_Data_leaks_2015.pdf. – (дата обращения: 16.03.2015).
3. Глобальное исследование утечек конфиденциальной информации в 2014 году. Аналитический центр InfoWatch. 2015 г. – Режим доступа: http://www.infowatch.ru/sites/default/files/report/analytics/russ/InfoWatch_global_report_2014.pdf. – (дата обращения: 20.03.2015).
4. Брюхомицкий Ю.А., Казарин М.Н. Учебно-методическое пособие к циклу лабораторных работ «Исследование биометрических систем динамической аутентификации пользователей ПК по рукописному и клавиатурному почеркам» по курсу: «Защита информационных процессов в компьютерных системах». − Таганрог: Изд-во ТРТУ, 2004. – 38с.
5. Иванов А. И. Биометрическая идентификация личности по динамике подсознательных движений / А. И. Иванов. – Пенза : Изд-во Пенз. гос. ун-та, 2000. – 188 с.
6. Ложников П.С., Сулавко А.Е. Технология идентификации пользователей компьютерных систем по динамике подсознательных движений // Автоматизация и современные технологии / Машиностроение. - Москва: 2015, №5, С. 31-36.
7. Епифанцев Б.Н., Ложников П.С., Сулавко А.Е., Борисов Р.В. Комплексированная система идентификации личности по динамике подсознательных движений // Безопасность информационных технологий / ФГУП «ВИМИ» - Москва : 2011, № 4. С. 97-102.
8. Раскин Д. Интерфейс: новые направления в проектировании компьютерных систем. — СПб: Символ-плюс, 2010. — 272 с.
9. Сулавко А.Е., Еременко А.В., Самотуга А.Е. Исключение искаженных биометрических данных из эталона субъекта в системах идентификации // Информационные технологии и вычислительные системы / ЛЕНАНД. - Москва: 2013, № 3. С. 96-101.
10. Dodis, Y., Reyzin, L., Smith, A. Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy Data // Proceedings from Advances in Cryptology. EuroCrypt. – 2004. – P. 79-100.
11. Robert H Morelos-Zaragoza. The art of error correcting coding. John Wiley & Sons, 2006. — 320 p.
12. Соловьева Ф. И. Введение в теорию кодирования: Учебное пособие / Новосиб. гос. ун-т. Новосибирск, 2006. 127 с.
13. Еременко А.В., Сулавко А.Е. Исследование алгоритма генерации криптографических ключей из биометрической информации пользователей компьютерных систем // Информационные технологии / «Новые технологии» - Москва: 2013, №11. – С. 47–51.
14. Иванов А. И., Сомкин С. А., Андреев Д. Ю., Малыгина Е. А. О многообразии метрик, позволяющих наблюдать реальные статистики распределения биометрических данных «нечетких экстракторов» при их защите наложением гаммы // «Вестник Уральского федерального округа. Безопасность в информационной сфере». 2014. № 2(12). С. 16–23.
15. Scotti, F., Cimato, S., Gamassi, M., Piuri, V., Sassi, R. Privacy-aware Biometrics: Design and Implementation of a Multimodal Verification System // 2008 Annual Computer Security Applications Conference, IEEE. – 2008. – P. 130-139.
16. Lozhnikov P.S., Sulavko, A.E., Volkov D.A. Application of noise tolerant code to biometric data to verify the authenticity of transmitting information / Control and Communications (SIBCON), 21-23 May 2015, Omsk, Russia – p.1-3. ISBN 978-1-4799-7102-2, DOI: 10.1109/SIBCON.2015.7147126.
17. Еременко А.В., Сулавко А.Е. Способ двухфакторной аутентификации пользователей компьютерных систем на удаленном сервере с использованием клавиатурного почерка // Прикладная информатика / НОУ ВПО «МФПУ «Синергия», Москва, 2015, №6.
18. Tsoukalas L. H. Fuzzy and Neural Approaches in Engineering / L. H. Tsoukalas, R. E. Uhrig. – New York: John Wiley&Sons.Inc, 1997. – 587 p.
19. Круглов В. В. Нечеткая логика и искусственные нейронные сети : учеб. пособие / В. В. Круглов, М. И. Дли, Л. Ю. Голунов. – М.: Физматлит, 2001. – 224 с.
20. Сулавко А.Е., Еременко А.В. Метод сжатия собственных областей классов образов в пространстве малоинформативных признаков // Искусственный интеллект и принятие решений / ЛЕНАНД. - Москва: 2014, № 2. С. 102-109.
21. Харин Е.А. Построение систем биометрической аутентификации с использованием генератора ключевых последовательностей на основе нечетких данных / Е.А. Харин, С.М. Гончаров, П.Н. Корнюшин // Матер. 50-й Всерос. межвуз. науч.-техн. конф. – Владивосток: ТОВМИ, 2007. – С. 112–115.
22. F. Monrose, M. K. Reiter and R. Wetzel. Password hardening based on keystroke dynamics. //Proceedings of sixth ACM Conference on Computer and Communications Security. — CCCS, 1999.

Сулавко Алексей Евгеньевич. Старший преподаватель ФГБОУ ВО «Омский государственный технический университет». Окончил. СибАДИ в 2009 году. Кандидат технических наук. Количество печатных работ: 40. Область научных интересов: распознавание образов, биометрия, искусственный интеллект. E-mail: sulavich@mail.ru

Еременко Александр Валериевич. Доцент ФГБОУ ВПО «Омский государственный университет путей сообщения». Окончил СибАДИ в 2006 году. Кандидат технических наук. Количество печатных работ: 35. Область научных интересов: распознавание образов, биометрия, искусственный интеллект, криптографические системы защиты информации.
E-mail: nexus-@mail.ru

Толкачева Елена Викторовна. Доцент ФГБОУ ВПО «Омский государственный университет путей сообщения». Окончила Омский государственный университет в 2003 году. Кандидат технических наук. Количество печатных работ: 15. Область научных интересов: распознавание образов, биометрия, искусственный интеллект, криптографические системы защиты информации. E-mail: tolkacheva_ev@mail.ru.

Жумажанова Самал Сагидулловна. Инженер ООО «Инновационные ВЕБ-технологии». Аспирант ФГБОУ ВПО ОмГУПС. Окончила. СибАДИ в 2015 году. Количество печатных работ: 3. Область научных интересов: распознавание образов, биометрия. E-mail: samal_shumashanova@mail.ru

Fuzzy extractor for generating the encryption key based on the parameters of keystroke dynamics

A.E. Sulavko, A.V. Eremenko, E.V. Tolkacheva, S. S. Zhumazhanova

Abstract. The article is devoted to the generation of key sequences on the basis of keystroke dynamics of users of computer systems. As part of the study suggested several variations of fuzzy extractors to solve the problems put forward. A series of numerical experiments to assess the effectiveness of the proposed methods have been done, the optimal parameters of fuzzy extractors have been defined.
Keywords: fuzzy extractors, generate cryptographic keys based on biometrics, pattern recognition, keystroke dynamics, identifying features.

Sulavko Alexey E. Senior Lecturer of OmSTU. PhD in Technical Sciences. Graduated from the SibADI in 2009. Number of publications: 40. Area of scientific interests: pattern recognition, biometrics, artificial intelligence. E-mail: sulavich@mail.ru

Eremenko Alexander V. Assistant professor of OSTU. PhD in Technical Sciences. Graduated from the SibADI in 2006. Number of publications: 35. Area of scientific interests: pattern recognition, biometrics, artificial intelligence, cryptographic information protection system. E-mail: nexus-@mail.ru

Tolkacheva Elena V. Assistant professor of OSTU. PhD in Technical Sciences. Graduated from the OmSU in 2003. Number of publications: 15. Area of scientific interests: Data Mining, machine learning, fuzzy logic. E-mail: tolkacheva_ev@mail.ru

Zhumazhanova Samal S. Engineer of LLC "Innovative Web Technologies". Graduated from the SibADI in 2015. Number of publications: 3. Area of scientific interests: pattern recognition. E-mail: samal_shumashanova@mail.ru