DOI: 10.24892/RIJIE/20230402

Метод квантования биометрических данных рукописного почерка в нечетких экстракторах*

Баянов Б.И.
Казанский национальный исследовательский технический университет
им. А.Н. Туполева - КАИ
г. Казань, Российская Федерация
bayanov_bulat@mail.ru

Аннотация. **В организации информационных систем предприятий немаловажным аспектом является обеспечение информационной безопасности конфиденциальных данных. Для этого внедряются автоматизированные системы аутентификации, электронной цифровой подписи, системы шифрования. Нами предлагается совместное использование биометрических и криптографических систем защиты. Для этого проанализированы текущие современные биометрические криптосистемы и выделены наиболее подходящие из них. В экспериментальной части статьи описывается применение метода нечетких экстракторов в задаче формирования криптографического ключа длиной 210 бит на основе биометрических данных рукописных подписей публичной базы данных SVC 2004. Данные представляют отличные друг от друга рукописные подписи в виде значений координат положения пера на графическом планшете. Рассматривается внедрение предложенного нами метода квантования с линейным преобразованием в алгоритм, основанный на нечетких экстракторах. Алгоритм характеризуется сниженными оценками ошибок первого и второго рода в принимаемых решениях. Точность распознавания подлинных пользователей выстроенной системы составляет 93,5%, при этом точность распознавания неподлинных пользователей составила 99,75%.**

Ключевые слова: **аутентификация, верификация, защита информации, криптографический ключ, динамический рукописный почерк, нечеткие экстракторы, преобразователь Биометрия-код.**

Введение

Одной из основных целей автоматизированных систем защиты информации в информационных системах организаций или предприятий является снижение рисков угроз информационной безопасности. Реализация таких угроз как неправомерное копирование, уничтожение, фальсификация конфиденциальной информации в свою очередь может нанести колоссальный ущерб бизнес-процессам организаций, что требует значительных денежных затрат на восстановление. Таким образом, сотрудничество руководства предприятий со специалистами по защите информации и финансовая поддержка систем информационной безопасности компьютерных систем и систем управления позволит существенно снизить лишние затраты и обезопасить делопроизводство предприятий [1].

Специалисты по защите информации выделяют несколько направлений, по которым ведется внедрение автоматизированной системы защиты информации: организационная (правила приема сотрудников предприятия на работу, система контроля и управления допуском сотрудников предприятия, политика информационной безопасности), инженерно-техническая (настройка телекоммуникационных систем, программно-аппаратное обеспечение информационной безопасности, установка программ от вирусов и угроз, периодическое обновление парольных фраз пользователей), правовая (требования государственных стандартов, федеральных законов по защите информации) [2, 3].

Переход на электронный документооборот и на дистанционный формат взаимодействия сотрудников предприятий создают запросы на внедрение новых современных средств по защите конфиденциальных данных. Например, при таком переходе многие организации столкнулись с проблемой подтверждения документов при помощи печатей и подписей, которые требуют личного присутствия подписантов. Для частичного решения такой проблемы существуют электронные цифровые подписи, где личность подтверждается секретным криптографическим ключом, хранящимся на внешнем диске или токене (специальный защищенный USB флеш-накопитель). Также информация, идентифицирующая пользователей информационных систем, криптографические ключи могут содержаться в смарт-картах, которые обычно используются в контрольно-пропускных пунктах предприятий.

Предоставление парольных фраз или секретной информации на смарт-картах в системах аутентификации или управления доступом не подтверждает подлинность зарегистрированных пользователей, т.к. парольная фраза или секретная информация может быть предоставлена злоумышленником. Нами предлагается внедрение биометрических данных рукописного почерка подлинного пользователя в секретную парольную фразу, что подтвердит его подлинность. Средства защиты информации, сочетающие биометрические и криптографические системы, могут быть применены в информационных системах предприятий следующим образом: подтверждение личности подписанта электронных документов, подтверждение личности в системах аутентификации, шифрование электронных документов на основе биометрических признаков личности.

Анализ современных био-криптографических систем

В теории по биометрической защите информации рукописного почерка выделяют три типа подделок: случайная подделка (злоумышленник ничего не знает о рукопис-

* Статья публикуется по рекомендации программного комитета Всероссийской научно-технической конференции "Пром-Инжиниринг", https://icie-rus.org

ной записи), простая подделка (злоумышленник знает о рукописной записи в печатном виде, например, имя пользователя), умелая подделка (злоумышленник наблюдал за процессом ввода рукописной записи). В зависимости от типа подделки разработчиками биометрических систем защиты информации строится алгоритм, преобразующий исходные данные в конечный результат: либо криптографический ключ (КК), либо принятие решения о допуске пользователя. В ходе многочисленных экспериментов мы считаем, что для умелых подделок обязательным является включение в обучающую выборку случайных рукописных записей. При этом для умелых подделок целесообразным является применение нейросетевых моделей (преобразователь Биометрия-код) [4, 5]. Если же парольная рукописная фраза держится пользователем в секрете, то включение в обучающую выборку случайных рукописных записей является необязательным. На наш взгляд здесь достаточным является применение подходов, близких к методу нечетких экстракторов (НЭ) [6].

Метод НЭ активно используется в исследованиях ученых за рубежом, также ими рассматриваются схожие методы, зачастую включающие те же принципы помехоустойчивого кодирования и хеширования. Одним из таких методов является метод Fuzzy Commitment и Secure Sketch [7, 8]. Отличительными особенностями выделяется метод нечеткого хранилища (Fuzzy Vault). В его основе лежит метод полиномиальных вычислений. Для защиты набора биометрических признаков ВА в количестве n случайным образом генерируется пользовательский ключ КВ длиной M бит. Код, исправляющий ошибки (например, код Рида-Соломона [9]) применяется к КВ, и выявленный избыточный код конкатенируется с КВ, получая закодированный ключ КС длиной N бит (N>M). Затем строится полином P степени L (L<n) с использованием КС в качестве коэффициентов. Проекция полинома P(A) вычисляется для каждого элемента A, получая набор подлинных точек G = P(BAi). Чтобы скрыть настоящие точки, случайным образом генерируются точки, которые не пересекаются ни с полиномом P, ни с набором ВА. Наконец, нечеткое хранилище Fuzzy Vault состоит из объединения набора G и ложных точек. В процессе аутентификации предоставляется тестовый биометрический образец B′A для декодирования хранилища и восстановления КВ. Если образец B′A существенно совпадает с ВА, то B′A может идентифицировать подлинные точки из нечеткого хранилища Fuzzy Vault [10]. В теории по биометрической защите информации применение такого метода целесообразно в условиях наличия малого количества биометрических признаков (БП), идентифицирующих биометрический образ, и зачастую рассматривается как повышение уровня защиты биометрических данных в системах верификации пользователей. Также одним из недостатков является уязвимость к атаке подбора возможных значений ключа (атака brute-force).

В отечественной научной литературе пользуется популярностью нейросетевой преобразователь Биометрия-код (НПБК). Согласно серии стандартов ГОСТ Р 52633 [11] на вход искусственной нейронной сети подаются от 11 до 21 биометрических образцов подлинного пользователя и 64 биометрических образца случайных подписей неподлинных пользователей. Образцы подписей предоставляются в виде последовательностей значений БП. Выходными значениями для подлинных образцов подписи является определенная заранее последовательность значений, идентифицирующая КК. Для неподлинных образцов подписей выходными значениями являются последовательности случайных значений, несовпадающих с КК. При этом архитектура обученной искусственной нейронной сети хранится в базе данных и затем используется для восстановления КК из предоставляемых тестовых образцов подписи подлинного пользователя. Основным недостатком такого подхода является необходимость в наличии биометрических образцов случайных подписей.

Для решения нашей задачи наиболее подходящим методом является метод нечетких экстракторов. Принцип работы метода описывается следующим образом. В методе НЭ формируется битовая последовательность биометрических данных, затем применяется операция XOR с битовой последовательностью случайных значений, которая является КК. Ко второй битовой последовательности добавляется избыточный код, исправляющий ошибки. Результат «суммирования» операцией XOR как открытая строка и код, исправляющий ошибки, хранятся в базе данных и не обладают полезной для злоумышленника информацией. Для восстановления КК из открытой строки «вычитается» битовая последовательность тестового образца биометрических данных и результат «вычитания» корректируется избыточным кодом, исправляющим ошибки. В итоге подлинному пользователю предоставляется КК [12].

Результаты

Вышеописанный метод НЭ имеет высокие требования к качеству сформированных БП [13] и выбору метода квантования исходных данных. Если же на вход алгоритма НЭ подаются значения, имеющие широкий разброс возможных значений, то в битовом представлении тестовые значения БП будут иметь кодовые расстояния, неприемлемые для исправляющей способности соответствующего кода алгоритма. Для снижения кодового расстояния в методе НЭ значения БП описываются кодами Грея. Также для частичного устранения этой проблемы нами предлагается применение метода квантования с линейным преобразованием.

$$k_{Ri} = \frac{1}{x_{MAXi} - x_{MINi}}, \quad (1)$$

$$l_{Ri} = \lfloor k_{Ri} \cdot x_{MAXi} \rfloor - k_{Ri} \cdot x_{MAXi} + \frac{1}{2}, \quad (2)$$

где k_{Ri} – угловой коэффициент; x_{MAXi} и x_{MINi} – верхняя и нижняя границы возможных значений i-го БП соответственно; l_{Ri} – коэффициент сдвига; $\lfloor x \rfloor$ – знак округления в меньшую сторону числа x.

Воспользовавшись формулами (1) и (2), значения исходных данных, например, попадающие в интервал в [50; 60], с коэффициентами k_{Ri} = 0,1 и l_{Ri} = 0,5 линейным преобразованием $k_{Ri}*x_{Ri}+l_{Ri}$ попадают в интервал [5.5; 6.5]. При округлении получившихся значений к ближайшему целому БП описывается цифрой 6. При этом для других интервалов [50; 60], [30; 40] или [120; 130] имеются те же значения коэффициентов k_{Ri} = 0,1 и l_{Ri} = 0,5. Это говорит о

том, что наличие коэффициентов не раскрывает информацию об исходных данных. Такой способ квантования значительно снизит кодовые расстояния нечетких числовых значений БП.

В ходе экспериментов с применением вышеописанного метода квантования нами было выявлено, что при среднем количестве несовпадающих БП от 1 до 3 для подлинного пользователя среднее количество несовпадающих БП для неподлинных пользователей составляет около 40. Логично предположить, что здесь требуется методы корректировки отклонений значений БП. Нами принято решение использовать метод НЭ, включающий метод частичной корректировки ошибок в кодовой комбинации. Для этого мы использовали готовый пакет данных fuzzy_extractor языка программирования Python [14]. В этом пакете есть возможность сформировать КК длины в 30 символов в кодировке utf-8, что составляет 210 бит (каждый символ описывается 7 битами) полезной информации, идентифицирующей уникальную кодовую комбинацию. Кодирование исходных БП производилось согласно коду Грея, где в каждой группе из 7 бит помещалась информация о трех БП. Первый и второй БП описывается двумя битами, третий БП описывается тремя битами. Итого, при формировании КК могло быть учтено 90 наилучших БП. Решение описывать БП двумя или тремя битами было принято из-за того, что подавляющее большинство конечных числовых значений, описывающих БП, варьируются в пределах от 1 до 4. Если же значение БП составляет 14 и описывается тремя битами, то берется остаток от деления на 8 и значение БП приравнивается к 6 (по коду Грея в битовом представлении это 111).

Для проверки выстроенного алгоритма формирования КК на основе биометрических данных рукописного почерка рассмотрены значения координат положения пера на графическом планшете рукописных подписей публичной базы данных SVC 2004 [15], где рассматривались только случайные подделки (сравнение подлинной подписи со случайными подписями других пользователей). В этой публичной базе данных имеется по 20 образцов подлинных подписей 40 пользователей. В наших экспериментах для каждого пользователя в обучающую выборку включено 16 образцов подлинных подписей, в тестовую выборку включено 4 образца подлинных подписей и 780 образцов неподлинных подписей (по 20 образцов неподлинных подписей 39 пользователей). Для оценки точности биометрической системы защиты использовались оценки: ошибки первого рода (False Rejection Rate – ложный отказ в доступе), ошибки второго рода (False Acceptance Rate – ложное предоставление доступа) и равного уровня ошибок (Equal Error Rate). Равный уровень ошибок (EER) количественно определяется при равенстве ошибок первого и второго рода (FRR=FAR). Нами получены следующие результаты: при использовании метода НЭ FRR ≈ 5%, FAR ≈ 0,25%, EER ≈ 2,5%; при исключении метода НЭ FRR ≈ 6,5%, FAR ≈ 0,25%, EER ≈ 3,15%. Результаты продемонстрировали, что при указанных условиях метод НЭ, основанный на коде, исправляющем ошибки, улучшает точность работы построенных алгоритмов биометрических систем защиты информации. Предложенные подходы могут быть полезны также в задачах верификации и идентификации пользователей.

Заключение

Информатизация всевозможных сфер жизнедеятельности человека и всеобщий переход на ведение электронного документооборота вынуждает нас переходить на более современные и эффективные средства защиты информации. Одним из таких средств являются биометрические криптосистемы, которые можно применить в системах аутентификации пользователей, электронной цифровой подписи и в системах шифрования. В ходе сравнения современных методов в биометрических системах защиты выделен метод нечетких экстракторов. В экспериментальной части работы нами предложен метод квантования с линейным преобразованием исходных данных и совместное использование его с методом нечетких экстракторов. Результаты экспериментов, проведенных на основе исходных данных образцов рукописных подписей публичной базы данных SVC 2004, доказали целесообразность применения предложенного подхода. Точность представленного алгоритма формирования криптографических ключей по оценке равного уровня ошибок составила 97,5%. Работа может быть полезна для специалистов по информационной безопасности, внедряющих автоматизированные средства защиты информации в информационных системах предприятий.

Литература

1. Гришина Н.В. Основы информационной безопасности предприятия. Учебное пособие. – М.: НИЦ ИНФРА-М, 2021. – 216 с.

2. Родичев Ю.А. Нормативная база и стандарты в области информационной безопасности. Учебное пособие. – СПб.: Питер, 2021. – 256 с.

3. Вострецова Е.В. Основы информационной безопасности: учеб. пособие для студентов вузов. – Екатеренбург: Изд-во Уральского университета, 2019. – 204 с.

4. ГОСТ Р 52633.5-2011. Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия-код доступа. – М.: Стандартинформ, 2018. – 15 с.

5. Сулавко А.Е. Перспективные нейросетевые алгоритмы распознавания динамических биометрических образов в пространстве взаимозависимых признаков / А.Е. Сулавко, С.С. Жумажанова, Г.А. Фофанов // Динамика систем, механизмов и машин. – 2018. – Т. 6, №4. – С. 130-145.

6. Dodis Y. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data / Y. Dodis, R. Ostrovsky, L. Reyzin, A. Smith // SIAM Journal on Computing. – 2008. – is. 1(38). – P. 97-139.

7. Shi S. Fingerprint Recognition Strategies Based on a Fuzzy Commitment for Cloud-Assisted IoT: A Minutiae-Based Sector Coding Approach / S. Sha, C. Jia, Z. Xin-Li, L. Yang, G. Jing-Liang, W. Yun-Jiang // IEEE ACCESS. – 2019. – Vol. 7. – P. 44803-44812.

8. Кузнецов В.В. Новый метод получения устойчивого ключа из динамической биометрической подписи // Системы и средства информатики. – 2015. – Т. 25, №2. – С. 85-95.

9. Todd K. Moon Error Correction Coding: Mathematical Methods and Algorithms. – John Wiley & Sons, 2005. – 804 p.

10. Ponce-Hernandez W. Fuzzy Vault Scheme Based on Fixed-Length Templates Applied to Dynamic Signature Verification / W. Ponce-Hernandez, R. Blanco-Gonzalo, J. Liu-Jimenez, R. Sanchez-Reillo // IEEE ACCESS. – 2020. – Vol. 8. – P. 11152-11164.

11. ГОСТ Р 52633.3-2011. Защита информации. Техника защиты информации. Тестирование стойкости средств высоконадежной биометрической защиты к атакам подбора. – М.: Стандартинформ, 2018. – 16 с.

12. Сулавко А.Е. Нечеткий экстрактор для генерации ключей шифрования на основе параметров клавиатурного почерка / А.Е. Сулавко, А.В. Еременко, Е.В. Толкачева, С.С. Жумажанова // Информационные технологии и вычислительные системы. – 2016. – №4. – С. 69-79.

13. Баянов Б.И. Сравнительный анализ биометрических параметров в задаче формирования криптографического ключа на основе рукописного почерка / Б.И. Баянов, И.И. Исмагилов // Математические методы в технологиях и технике. – 2021. – №6. – С. 55-58.

14. https://pypi.org/project/fuzzy-extractor/ (дата обращения: 11.01.2023).

15. https://cse.hkust.edu.hk/svc2004 (дата обращения: 11.01.2023).

DOI: 10.24892/RIJIE/20230402

The Quantization Method of Biometric Data of Handwriting in the Fuzzy Extractors

Bayanov B.
Kazan National Research Technical University named after A.N. Tupolev
Kazan, Russian Federation
bayanov_bulat@mail.ru

Abstract. **The important aspect of the development of enterprise information systems is to ensure the information security of confidential data. For this, automatic authentication systems, digital signatures, and encryption systems are implemented into enterprise information systems. We propose a hybrid system that includes biometric and cryptographic security methods. We analyzed modern biometric cryptosystems and identified the most suitable of them. The experimental part of the article describes the application of the fuzzy extractor method in the problem of generating a 210-bit cryptographic key based on the biometric data of handwritten signatures of the SVC 2004 public database. The initial data are random forgeries of handwritten signatures. We consider pen position coordinates on a graphics tablet. The proposed quantization method with linear transformation is implemented into the algorithm based on fuzzy extractors. The algorithm demonstrates low values of False Rejection Rate and False Acceptance Rate. The recognition accuracy of the protection system of enrolled users refers to 93.5%, the recognition accuracy of the protection system of unenrolled users is 99.75%.**

Keywords: **authentication, verification, information protection, cryptographic key, dynamic handwriting, fuzzy extractors, Biometric-code converter.**

Библиографическое описание статьи

Баянов Б.И. Метод квантования биометрических данных рукописного почерка в нечетких экстракторах // Машиностроение: сетевой электронный научный журнал. – 2023. – Т.10, №4. – С. 8-11. DOI: 10.24892/RIJIE/20230402

Reference to article

Bayanov B. The quantization method of biometric data of handwriting in the fuzzy extractors, *Russian Internet Journal of Industrial Engineering*, 2023, vol.10, no.4, pp. 8-11. DOI: 10.24892/RIJIE/20230402
