ЗАЩИТА КОРПОРАТИВНОЙ ИНФОРМАЦИОННОЙ СИСТЕМЫ НА БАЗЕ РАСПОЗНАВАНИЯ КЛАВИАТУРНОГО ПОЧЕРКА ПОЛЬЗОВАТЕЛЕЙ

С.С. Махонченко, Р.П. Затеев,
Е.А. Кочегурова
Томский политехнический университет
makhonchenkoss@tpu.ru

Введение

Современный мир не может существовать без компьютерных информационных систем, между которыми происходит непрерывное взаимодействие с целью передачи различной информации. Зачастую для получения этой информации пользователям требуется иметь персональные данные, с помощью которых происходит взаимодействие пользователя с какой-либо информационной системой. Где существует защита персональной информации, там всегда есть и кражи, утечки и взломы с целью получения этой информации.

По данным агенства InfoWatch в 2017 г. зарегистрировано 254 случая утечки конфиденциальной информации из коммерческих и некоммерческих компаний, а также государственных организаций, работающих в России, и 2131 случай в мире. В результате утечек скомпрометировано 5.8 млн записей, относящихся к персональным данным [1]. Анализ статистики по годам, показывает увеличение количества случаев утечек персональной информации с каждым годом, что говорит о том, что исследование методов и алгоритмов защиты информации является не только актуальной задачей, но и самой приоритетной в цифровом мире.

В рамках данного исследования предлагается усилить парольную аутентификацию пользователей дополнительным методом распознавания клавиатурного почерка для создания системы, обеспечивающей наиболее безопасное хранение персональной информации.

Анализ существующих методов аутентификации

Аутентификация (англ. authentication, от греч. — реальный, истинный) — процесс проверки принадлежности субъекту прав доступа к информационным ресурсам системы в соответствии с предъявленным им идентификатором.

Существуют различные методы аутентификации:

1. парольная;
2. биометрическая;
3. двухфакторная (многофакторная);
4. аппаратная (аутентификация при помощи технических средств).

Парольная аутентификация является самым распространенным методом из-за простоты реализации и логической ясности принципов функционирования. Несмотря на существование множества угроз для данной схемы авторизации, она используется в большинстве информационных систем.

Биометрическая аутентификация использует для удостоверения личности людей их биометрические данные. Биометрические данные человека подразделяются на *физиологические* и *поведенческие*.

Двухфакторная аутентификация — это метод идентификации пользователя в каком-либо сервисе при помощи запроса аутентификационных данных двух разных типов.

Под *аппаратной* аутентификацией принято понимать аппаратно-программные системы идентификации и аутентификации или устройства ввода идентификационных признаков.

Одним из главных инструментов управления процессами, происходящими в организациях, является корпоративная информационная система, которая включает в себя инфраструктуру и информационные сервисы. Проблема контроля доступа к информации и разграничения полномочий пользователей – одна из основных проблем при работе с сетевой инфраструктурой в организациях, где человеческий фактор играет самую важную роль при несоблюдении предписанных правил хранения и доступа к персональной информации. В большинстве случаев в политике безопасности организации прописана информация о методах аутентификации и авторизации пользователей, а также распределены их полномочия согласно месту работы в организации и занимаемой должности.

Анализ методов распознавания клавиатурного почерка

Клавиатурный почерк (КП) – набор динамических характеристик работы на клавиатуре, которые определяются следующими параметрами, благодаря которым можно верифицировать законного оператора [2]:

- скорость ввода - количество введенных символов, разделенное на время печатания;
- динамика ввода - характеризуется временем между нажатиями клавиш и временем их удержания;
- частота возникновение ошибок при вводе;
- использование клавиш - например, какие функциональные клавиши нажимаются для ввода заглавных букв.

На Рис. 1 изображены основные параметры клавиатурного почерка: время удержания клавиши (ВУК) и пауза между нажатиями клавиш.

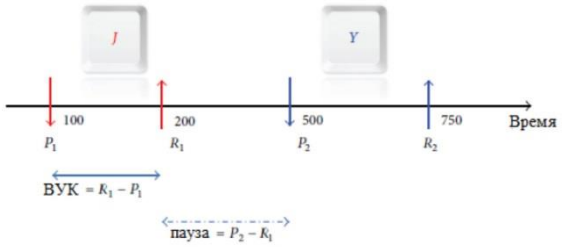

Рис. 1. Время удержания клавиш и пауза

Основные сложности при анализе клавиатурного почерка:

- разброс параметров клавиатурного почерка в зависимости от психофизического состояния пользователя;
- разброс параметров клавиатурного почерка в зависимости от используемой клавиатуры;
- необходимость сбора большого количества статистических данных для каждого исследования КП, отсутствие готовых баз данных с образцами КП.

Подходы к распознаванию пользователей на основе КП можно разделить на 3 основные группы [3]:

- на основе метрических расстояний;
- статистические методы;
- методы машинного обучения.

Самым популярным методом распознавания является оценка метрического расстояния и по частоте применения составляет 23% [4].

Использование статистических методов в задаче распознавания клавиатурного почерка актуально и сейчас. В эту группу входят Марковские модели, байесовские методы и методы, основанные на функции гауссовой плотности и взвешенной вероятности.

В 16% используется метод искусственных нейронных сетей, который относится к группе распознавания образов на основе машинного обучения. Считается, что данный метод более эффективен, чем статистические методы, но сложность использования метода, как классификатора КП, заключается в необходимости обучать метод на сетях, образцах подлинных и неподлинных пользователей.

Также, к этой группе распознавания КП принято относить следующие известные алгоритмы: дерево принятия решений, нечеткая логика и эволюционные вычисления.

Дерево принятия решений достаточно популярный метод из-за низкой вычислительной сложности. Однако, вследствие рекурсивности процедуры распознавания, эффективность метода достигается только при небольших множествах подлинных и неподлинных пользователей.

Нечеткая логика используется для моделирования задач с неоднозначными данными при помощи многозначной логики. Метод базируется на построении границ области принятия решений исходя из данных обучения с функциями принадлежности и нечеткими правилами. После того, как выделится пространство признаков, а также после вычисления значений принадлежности, необходимо произвести идентификацию категории, к которой относится исследуемый объект (шаблон).

Эволюционные вычислительные методы, построенные на идее естественного отбора, в задаче идентификации КП используют группу известных методов: генетические алгоритмы (ГА), алгоритмы роевого интеллекта и другие. Перечисленные алгоритмы осуществляют направленный поиск максимального совпадения анализируемых клавиатурных шаблонов повышая точность распознавания.

Заключение

КП является одним из перспективных способов аутентификации, поскольку не существует людей с идентичным компьютерным ритмом. К тому же это самый низкозатратный способ аутентификации.

Для создания системы, идентифицирующей пользователя по КП, необходимо оценить методы распознавания в применении к данной задаче. Для сбора информации разработан вариант программного приложения с клиент-серверной технологией. Приложение протестировано, получены вполне удовлетворительные результаты по сбору и обработке первичной информации о клавиатурной динамике [3]. Сейчас ведутся работы по расширению функционала программного приложения для его использования внутри определенного сегмента корпоративной локальной сети Томского политехнического университета.

Список использованных источников

1. Утечки данных. Россия. 2017 год. [Электронный ресурс] / Аналитический центр InfoWatch. - URL: www.infowatch.ru/analytics (дата обращения 16.11.2018).
2. Васильев В.И., Ложников П.С., Сулавко А.Е., Еременко А.В. Технологии скрытой биометрической идентификации пользователей компьютерных систем (обзор) // Вопросы защиты информации. 2015. № 3 (110). с. 37-47.
3. Kochegurova E. A., Gorokhova E.S., Mozgaleva A. I. Development of the Keystroke Dynamics Recognition System // Journal of Physics: Conference Series. 2017. v. 803 № 1. pp. 1-6.
4. Teh P. S., Teoh A.B., Yue S. A Survey of Keystroke Dynamics Biometrics // The Scientific World Journal. 2013. v. 2013. pp. 1-24.