

РАЗРАБОТКА СИСТЕМЫ РАСПОЗНАВАНИЯ ПОЛЬЗОВАТЕЛЯ ПО КЛАВИАТУРНОМУ ПОЧЕРКУ

Горохова Е.С., Соколова Е.В., Демиденко Л.Р.

Научный руководитель: Кочегурова Е.А.

Томский политехнический университет

GorokhovaES@mail.ru

Введение

В настоящее время важность информации сложно переоценить. Необходимо разграничивать доступ для различных пользователей. При этом стандартные средства аутентификации, такие как логин и пароль, не в состоянии уже обеспечить необходимую степень защиты, поскольку пароль может быть взломан или украден. По этой причине биометрические методы аутентификации, использующие биологические характеристики конкретного индивида, становятся всё более популярны [1]. В том числе к ним относится метод распознавания пользователя по клавиатурному почерку. Этот метод удобен для пользователя и службы безопасности, поскольку мониторинг нажатий на клавиши проводится в скрытом режиме и не требует от пользователя никаких дополнительных действий. Одно из главных достоинств данного метода в дешёвизне необходимого оборудования, поскольку почти каждый компьютер и так имеет клавиатуру. Поэтому методам распознавания пользователя по клавиатурному почерку должно быть уделено особое внимание.

Описание алгоритма

Клавиатурный почерк – это набор динамических характеристик работы пользователя на клавиатуре [2]. Наиболее распространёнными его характеристиками являются: время удержания клавиши (ВУК), пауза между нажатиями, наличие наложений.

Образцы клавиатурного почерка можно сравнивать различными способами, например, с использованием вероятностно-статистических методов или алгоритмов на основе нейронных сетей.

В упрощённой форме, алгоритм распознавания клавиатурного почерка может быть разделён на 2 стадии. На первой стадии программа накапливает статистическую выборку характеристик клавиатурных почерков пользователей. Вторая стадия предназначена для сравнения характеристик полученного образца клавиатурного почерка с эталонными. Сравнение может происходить с помощью любой меры близости. В работе использовано Евклидово расстояние, которое может быть рассчитано следующим образом:

$$P = \sqrt{\sum_{i=1}^N (t_{et} - t_{cur})^2} \quad (1)$$

где N – количество различных символов,

t_{et} – эталонное время удержания клавиши;

t_{cur} – текущее время удержания клавиши.

В данной работе отдельно рассматривается время удержания в случае наличия или отсутствия наложений. Учитываются только символы русского и английского алфавита.

На этой стадии обучения происходит накопление статистики характеристик клавиатурных почерков различных пользователей. Сюда включается название клавиши, описание события клавиатуры – клавиша может быть нажата или отпущена – и время события.

Отдельно подсчитывается количество ошибок ввода. Его полагаем равным количеству нажатий на клавиши “Delete” и “Backspace”.

Указанные данные собираются клиентским приложением и отправляются на сервер для дальнейшего анализа.

Затем подсчитывается среднее время удержания для каждой клавиши и определяется наличие наложений.

После этого, если пользователь работает в системе в первый раз, образец его клавиатурного почерка сохраняется для нового пользователя с выбранным именем. Иначе образец почерка добавляется к списку образцов этого пользователя. Если список получается слишком длинным – более 10 – самый старый образец почерка удаляется системой.

Такой подход позволяет принять во внимание, что клавиатурный почерк пользователя может изменяться с течением времени и в зависимости от психо-эмоционального состояния. Например, исследования показывают [3], что уставшие люди печатают медленнее и совершают больше ошибок.

Программная реализация

Серверный компонент разработанного программного приложения отвечает за хранение и обработку данных о клавиатурных почерках различных пользователей. Клиентский компонент используется для сбора необходимой информации с клавиатуры. Данные собираются, когда пользователь печатает на клавиатуре. Кроме того, клиентский компонент может представить результат для пользователя в графическом или табличном виде. Необходимые данные передаются между сервером и клиентами через TCP-сокеты. Ещё одной особенностью созданной системы является её возможность вести журнал событий, указывая, когда какой пользователь работал за

компьютером, на основании данных о клавиатурном почерке.

Анализ результатов

Сейчас разработанная система распознавания клавиатурного почерка находится на стадии тестирования. Были использованы образцы клавиатурного почерка одиннадцати пользователей. Их клавиатурные почерки были сопоставлены и проанализированы. Во время эксперимента пользователи работали в обычном режиме в любом программном приложении и не замечали дополнительного мониторинга. Программа собирала статистические данные и сохраняла в базе данных. Результаты эксперимента представлены ниже.

На рисунке 1 представлен график среднего времени удержания для каждой клавиши - ВУК. Также показаны верхняя (ВУК+) и нижняя граница доверительного интервала (ВУК-).

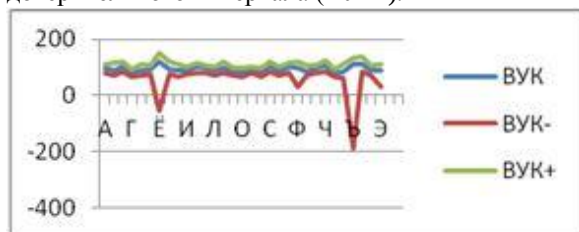


Рис. 1. Среднее время удержания для каждой клавиши

На рисунке 2 показано среднее время удержания клавиш для разных пользователей. При этом было рассмотрено три группы букв по частоте появления в текстах (с частотой >3%, 1-3%, <1%) в соответствии с рекомендациями Национального комитета русского языка [4]. Представлены результаты для наиболее часто встречающейся первой группы букв. Имена пользователей приведены в легенде.

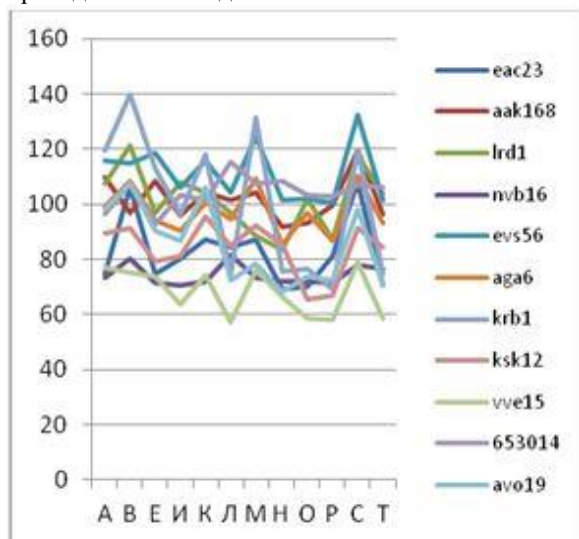


Рис. 2. Среднее время удержания клавиш для разных пользователей

На рисунке 3 показано среднее время

удержания для разных образцов клавиатурного почерка одного пользователя. Результаты представлены также для первой группы букв.

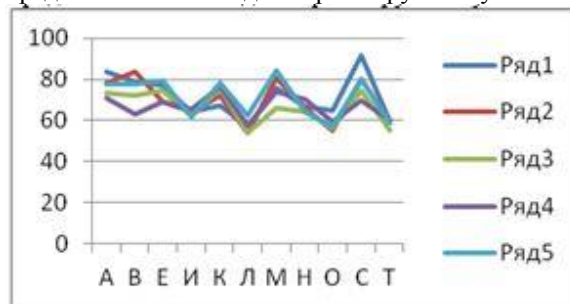


Рис. 3. Среднее время удержания для образцов почерка одного пользователя

Результаты тестирования позволяют сделать вывод, что каждому человеку присущ свой клавиатурный почерк. В то же время, клавиатурные почерки разных пользователей различаются более чем на 10% и являются статистически различимыми. Поэтому аутентификацию на основе клавиатурного почерка можно считать эффективной.

Заключение

Целью данной работы является разработка алгоритмического и программного обеспечения системы распознавания клавиатурного почерка. Был разработан алгоритм на основе вероятностно-статистического метода. Он позволяет сохранять образцы клавиатурного почерка пользователя и сравнивать их с целью аутентификации.

Система была протестирована с участием нескольких пользователей. Анализ результатов показал эффективность алгоритма распознавания клавиатурного почерка как инструмента аутентификации

Список использованных источников

1. Jain A, Flynn P and Ross A A 2008 Handbook of Biometrics Springer US pp. 1-2
2. Moskovitch R, Feher C, Messerman A, Kirschnick N, Mustafić T, Camtepe A, Löhlein B, Heister U, Möller S, Rokach L and Elovici Y 2009 Identity Theft, Computers and Behavioral Biometrics Proc. of the 2009 IEEE Int. Conf. on Intelligence and security informatics IEEE Press Piscataway, NJ, USA pp.155-160
3. Nazmul H N Alam J M Mahmud H and Hasan K 2014 Identifying emotion by keystroke dynamics and text pattern analysis Behaviour and Information Technology Taylor & Francis, Inc. Vol 33 Issue 9 pp.987-996
4. Русский алфавит. Частотность букв русского языка (по НКРЯ). [Электронный ресурс] / Инженерный справочник. – URL: <http://www.dpva.ru/Guide/GuideUnitsAlphabets/Alphabets/FrequencyRuLetters/>