

ОБ ОДНОМ ПОДХОДЕ К СОЗДАНИЮ КОМПЛЕКСНОЙ МОДЕЛИ ПОЛЬЗОВАТЕЛЯ ПЕРСОНАЛЬНОГО КОМПЬЮТЕРА В РАМКАХ СИСТЕМЫ КОНТРОЛЯ И УПРАВЛЕНИЯ ДЕЯТЕЛЬНОСТЬЮ ИТ-КОМПАНИИ

Никонов В. С.

ФГБОУ ВПО «Пермский государственный национально-исследовательский университет», Пермь, Россия (614990, г. Пермь, ул. Букирева, 15), e-mail: Comrade.NF@gmail.com

В статье предлагается подход к созданию комплексной модели пользователя персонального компьютера, основанной на совокупности различных особенностей поведения. Описаны общие принципы работы разрабатываемой системы контроля и управления деятельностью ИТ-компаний на основе комплексных моделей сотрудников. Указано место модели в рамках данной системы. Подробно описывается модель: ее составляющие и формат представления данных в ней. Предлагаемый подход позволяет объединить в себе ряд важных биометрических характеристик, которые отражают поведение пользователя за персональным компьютером: клавиатурный почерк, характер работы с манипулятором «мышь», сетевая активность, активность в рамках операционной системы. В целом, модель описывает пользователя с различных точек зрения и затрагивает различные стороны его взаимодействия с компьютером. Разработанная модель позволяет достичь существенных результатов при ее использовании, что подтверждается результатами, представленными в ранее опубликованных работах автора.

Ключевые слова: моделирование пользователя персонального компьютера, биометрия.

AN APPROACH TO THE CREATION OF A COMPLEX MODEL OF THE PERSONAL COMPUTER USER IN THE CONTROL AND MANAGEMENT SYSTEM FOR IT-COMPANIES

Nikonov V. S.

Perm State National Research University, Perm, Russia (614990, Perm, Bukireva str., 15), e-mail: Comrade.NF@gmail.com

The article offers an approach to the creation of a complex model of the PC user, based on the aggregation of various features of user's behavior. The general principles of control and management system for IT-companies based on complex employees models are described. Model's place within this system is specified. Components of the model and the it's data presentation format are described in details. The proposed approach allows to combine a number of important biometrics which reflect the user's behavior over a PC, such as keystroke dynamics, the nature of interaction with «mouse» manipulator, user's network activity, user's activity whitin the operating system. In general this model describes the user from different points of view and address the different aspects of his interaction with the PC. Model allows to achieve significant results. It is confirmed by the results presented in previously published articles.

Key words: PC user modeling, biometrics.

Введение

В настоящее время существует определенная потребность в системах контроля и управления деятельностью компаний, работающих в сфере информационных технологий. В таких компаниях большую часть работы сотрудники выполняют за персональными компьютерами. В первую очередь эти потребности вызваны возросшим количеством потоков информации, которые требуют все более эффективного управления. Кроме того, для повышения эффективности работы компании необходимы постоянный дистанционный

контроль за ее сотрудниками и предоставление руководящему персоналу отчетной информации об эффективности работы персонала в наглядной, понятной даже для неспециалиста форме.

Под контролем понимается наблюдение за работой сотрудника, сбор информации о нем, непрерывная обработка поступающих данных и сохранение их в долговременном хранилище, обеспечение безопасного доступа к корпоративной информации. Под управлением понимается предоставление руководящему персоналу компании информации об эффективности работы сотрудников: результаты расчета показателей эффективности для персонала, результаты расчета математических моделей персонала. На сегодняшний день не существует комплексных подходов, которые бы могли в полной мере решать поставленные задачи.

Целью данной статьи является описание одного из функциональных блоков разрабатываемой системы в рамках создания комплексной модели пользователя на основе его поведения за компьютером. Данные, предоставляемые этим блоком, предназначены для оценки того, насколько качественно и эффективно сотрудник выполняет свои должностные обязанности. Оценка этого производится путем сопоставления полученной модели пользователя с эталонной, составленной на основании экспертных мнений и эмпирических данных. Сопоставление, в данном случае – распознавание пользователя, процесс его идентификации с большим набором выходных данных, который будет использоваться для дальнейшей оценки. На данный момент не существует широко известных методов моделирования, которые бы основывались на поведении пользователей во время работы за компьютером. Такие методы в свою очередь базируются на биометрических параметрах каждого пользователя.

Биометрические параметры

Основное применение биометрических технологий на практике – системы автоматического распознавания пользователя на основе его уникальных физических или поведенческих параметров. Биометрические параметры человека могут быть разделены на два больших класса:

- Физиологические признаки, к которым можно отнести различные особенности тела человека.
- Поведенческие признаки. К ним относят поведение человека при работе на клавиатуре, особенности разговора, подпись и т.п.

Большинство биометрических систем работает по следующему принципу: в базе данных хранится цифровой образец отпечатка пальца, снимок радужной оболочки и т.п. При сканировании личной информации человека, полученные данные сравниваются с образцом и проверяются на идентичность.

Системы биометрической идентификации выполняют следующие функции [1]:

- **Верификация.** Производится проверка пользователя с использованием его имени, смарт-карты, идентификатора. Полученные биометрические данные сравниваются только с одним определённым эталоном из базы данных.
- **Идентификация.** Производится проверка пользователя без использования какой-либо дополнительной информации. Полученные данные сравниваются со всеми эталонами из базы данных.

Перечислим некоторые из методов биометрического распознавания:

- По отпечаткам пальцев.
- По форме руки.
- По форме лица.
- По рукописному почерку.
- По клавиатурному почерку.
- По голосу.
- По сетчатке глаза.

Возможно сравнение методов биометрической идентификации по следующим признакам:

- **Универсальность.** Каждый пользователь должен иметь этот признак.
- **Уникальность.** Насколько хорошо можно отличить одного пользователя от другого.
- **Простота.** Насколько просто реализуем алгоритм распознавания по признаку.
- **Эффективность.**

Модель пользователя

Для построения модели необходимо правильно подобрать набор параметров. Данный набор должен содержать в себе наиболее важные параметры, которые в полной форме отражают поведение пользователя за компьютером. Кроме того, необходимо определить оптимальное представление данных, полученных по набору выбранных признаков.

Для построения модели нами были выделены основные источники данных о поведении пользователя за компьютером. Так как такие характеристики (биометрический параметры)

пользователя, как компьютерный почерк, особенности работы с приложениями, сетевая активность требуют отдельного моделирования, суммарная модель пользователя рассматривается как комплексная, то есть представляющая собой организованную совокупность отдельных подмоделей.

Модель пользователя включает в себя ряд подмоделей:

- подмодель клавиатурного почерка;
- подмодель оперирования «мышью»;
- подмодель активности пользователя в рамках операционной системы (ОС);
- подмодель сетевой активности пользователя;
- метамодель, отображающая основные данные об основной модели, о возможных изменениях в ней;
- ряд других параметров: уникальный идентификатор, пароль, цифровые подписи оборудования.

С помощью данного набора подмоделей можно наиболее полно описать характер работы пользователя за персональным компьютером и его взаимодействие с внешней средой.

Подмодель представляет собой составную часть комплексной модели, которая только содержит информацию о пользователе, которая была собрана с компьютера. То есть подмодель – это контейнер, который может содержать как ложные, так и истинные данные.

Основным способом представления собранных данных является так называемый идентификационный вектор – вектор в многомерном пространстве, который является уникальным идентификатором пользователя относительно выбранной подмодели. Для различных подмоделей формат векторов может быть различным, но независимо от этого каждый его элемент содержит:

- уникальный в рамках данного вектора ключ элемента;
- значение, соответствующее ключу элемента.

Ключи и соответствующие им значения могут быть различных типов, начиная от простых типов, заканчивая сложными структурами данных. Количество элементов, которое может храниться в одном вектора, может быть любым.

Разделение векторов для различных подмоделей существенно сокращает размерность пространства поиска. При разделении для каждой подмодели создается свое пространство поиска, что позволяет с большей точностью производить идентификацию с использованием каждого из методов распознавания.

Для построения подмодели клавиатурного почерка изначально производится сбор необходимых данных. Сбор производится за определенные интервалы времени, подбор которых может быть произведен эмпирически [4].

Собранные данные содержат:

- набор символов, который был набран пользователем;
- временные промежутки между нажатиями клавиш;
- временные промежутки удержания каждой клавиши.

Используя эти данные, можно определить основные характеристики работы пользователя на клавиатуре и построить идентификационные векторы, которые будут входить в окончательную модель пользователя [3]. Подмодель клавиатурного почерка включает в себя усредненные величины, определяющие характер работы человека за компьютером.

Нами был построен специальный расширяемый словарь, который содержит в себе:

- наиболее часто используемые двух- и трехбуквенные сочетания в русском и английском языках;
- коды основных клавиш клавиатуры.

Основная информация для построения была взята из частотных словарей русского и английского языков. Основываясь на этой информации, мы имеем возможность построить идентификационные векторы, которые определяют среднее время набора буквенных сочетаний, долю использования клавиш при работе на клавиатуре, среднее время их удержания.

В итоге, подмодель клавиатурного почерка включает в себя:

- Идентификационный вектор пользователя, составляющими которого являются временные промежутки набора двухбуквенных сочетаний из словаря.
- Идентификационный вектор пользователя, составляющими которого являются временные промежутки набора трехбуквенных сочетаний из словаря.
- Различные усредненные показатели работы пользователя за клавиатурой: процент использования символьных клавиш, процент использования управляющих и модифицирующих клавиш, среднее время набора каждой клавиши, частоты набора каждой клавиши.

Процесс построения модели оперирования «мышью» практически ничем не отличается от построения модели клавиатурного почерка, за исключением измеряемых величин.

Собираемые данные представляют собой:

- координаты движения курсора «мыши»;
- координаты нажатий клавиш «мыши»;
- временные координаты нажатий клавиш «мыши».

Полученные данные после обработки также представляются в виде идентификационных векторов, которые в данном случае состоят из усредненных величин. Эти величины рассчитываются на основе данных, собранных за определенный интервал времени.

После обработки данных будет получена следующая информация:

- Средняя длина кривых, проведенных курсором «мыши»;
- Средняя скорость движения курсора «мыши»;
- Среднее ускорение при движении курсором «мыши»;
- Частоты нажатия клавиш «мыши» в различных областях монитора.

Кривая, проведенная курсором «мыши», – путь, по которой прошел курсор. Каждая кривая отделяется от другой кривой определенным промежутком времени, который может быть установлен опытным путем. Кроме того, длина пути не может быть меньше некоторой заранее установленной величины, в ином случае такой путь не является кривой.

Для построения подмодели активности пользователя в рамках ОС производится сбор следующих данных:

- события запуска приложений;
- события завершения работы приложений.

Обработанная информация помещается в идентификационный вектор, который определяет частоты использования различных приложений, установленных на компьютере пользователя.

Данные о сетевой активности пользователя собираются аналогичным образом. Все данные, проходящие через сетевые соединения компьютера, просматриваются и обрабатываются. Из всех обработанных данных система может получить только необходимые для построения модели, а именно перехваченные ей запросы по протоколу HTTP.

Этих данных достаточно для определения характера сетевой активности пользователя: часто используемые сетевые сервисы и приложения, основной тип пересылаемых данных, чаще всего используемые веб-браузеры.

Кроме основных подмоделей, в рамках разработанной модели используются метамодель, содержащая метаинформацию о самой модели, о возможных изменениях в ней. Она содержит данные об изменениях, произошедших в системе: установка нового программного обеспечения, установка или замена внешних устройств и т.п. С помощью этих данных, система, использующая модель, может принять решение об изменении основной модели, о пересмотре некоторых из ее составных частей, о переобучении алгоритмов распознавания и т.п.

Рассмотрение взаимодействия пользователя с компьютером с различных точек зрения дает большее количество важной информации, необходимой для идентификации и в то же время увеличивает надежность распознавания.

Заключение

В статье представлен подход к созданию комплексной модели пользователя персонального компьютера для дальнейшего использования ее в системе контроля и управления деятельностью ИТ-компании. Модель будет использована для оценки качества и эффективности работы сотрудников. Входные данные для оценки – выходные данные процесса сопоставления полученной и эталонной моделей (распознавание, идентификация).

Разработанная модель наиболее полно отражает поведение пользователя за персональным компьютером. Она сочетает в себе основные биометрические параметры, которые играют значительную роль в формировании характера поведения пользователя. Такой подход дает большее количество важной информации, необходимой для идентификации.

Эффективность данного подхода подтверждается результатами, представленными в ранее опубликованных работах автора [2].

Список литературы

1. Загоруйко Н. Г. Методы распознавания и их применение. М.: Сов. Радио, 1972. 247 с.
2. Никонов В. С. Nixa 3.0: Система распознавания пользователя персонального компьютера // Материалы III Студенческого регионального конкурса инновационных проектов по программе У.М.Н.И.К. 30-31 марта 2011, г. Пермь, 2011. С. 86 – 88.
3. Никонов В. С., Юрков К. А. Современная система идентификации личности по клавиатурному почерку // Конференция «Технологии Microsoft в теории и практике». Сборник трудов. Челябинск: Изд-во ЮУрГУ, 2009. С. 102 – 104.
4. Никонов В. С., Юрков К. А. Nixa 3.0: Система идентификации пользователя на основе нейронных сетей // Современные проблемы математики и ее прикладные аспекты: сб. ст (по материалам практ. конф. молодых ученых). Пермь: Перм. гос. ун-т, 2010. С. 104 – 109.

5. Hebb D. O. The Organization of Behavior. John Wiley & Sons: 1949.

Рецензенты:

Андрианов Дмитрий Леонидович, доктор физико-математических наук, профессор, заведующий кафедрой информационных систем и математических методов в экономике Пермского государственного национального исследовательского университета, генеральный директор ЗАО «ПРОГНОЗ», г. Пермь.

Максимов Владимир Петрович, доктор физико-математических наук, профессор кафедры информационных систем и математических методов в экономике Пермского государственного национального исследовательского университета, г. Пермь.