

С. М. Диденко, В. А. Шапцев

Методика отображения информационного почерка пользователя

Изложен подход к идентификации пользователей в системах с ограниченным доступом, основанный на анализе параметров информационного почерка пользователя во время работы в системе. Рассмотрены особенности параметров и предложен метод распознавания информационного почерка.

1. Понятие информационного почерка

Под информационным почерком (ИнП) конкретного пользователя будем понимать область пространства параметров, отражающих особенности работы пользователя с различными устройствами ввода (клавиатура, компьютерная мышь, графический планшет, джойстик и т. п.). Ожидается, что все пользователи имеют различающиеся ИнП в некоторой метрике этого пространства. Отметим, что понятие «информационный почерк» встречено нами только в работах [1, 2].

В настоящей работе предполагается, что случайный процесс взаимодействия пользователя с устройствами ввода компьютера — мышью и клавиатурой — стационарен на временном интервале анализа.

2. Выбор множества параметров для идентификации информационного почерка

Параметры траекторий компьютерной мыши

В [3] определено множество параметров траектории курсора при манипулировании мышью. Предложено каждую траекторию отображать вектором пяти параметров:

$$m = (T, L, U_p, d, a), \quad (1)$$

где T — время движения манипулятора до его остановки; L — длина траектории; U_p — скорость разгона; d — время между остановкой указателя и подтверждающим нажатием кнопки манипулятора; a — угол направления начального движения. Большинство из этих параметров определяет скоростные характеристики (динамику) работы пользователя с компьютерной мышью.

В результате измерений множества траекторий может быть сформирована статистическая выборка из N векторов, представленная в виде матрицы эксперимента размерностью $N \times 5$. Эта матрица является базой для разнообразной обработки с целью формирования уникального компьютерного образа пользователя. Например:

- вектор средних (простейший случай);
- 10 координат вершин гиперпараллелепипеда, формируемых минимальными и максимальными значениями в столбцах матрицы эксперимента;
- 5-мерная гистограмма — аналог вероятностной меры, с 5-мерным множеством интервалов (ячеек пространства параметров).

Проблема нестационарности почерка может быть решена, в частности, регулярным/нерегулярным повторением акта «обучения» системы, формирующей и идентифицирующей ИнП.

Параметры клавиатурного почерка

Исследования показали [4, 5], что клавиатурный почерк конкретного пользователя обладает стабильностью. Это позволяет с большой достоверностью идентифицировать пользователя, работающего с клавиатурой. В качестве параметров динамики клавиатуры авторы указанных работ предлагают средние значения временных интервалов t_{Hi} между нажатием клавиш и длительностей времени $t_{удj}$ их удержания. При этом временные интервалы между нажатием клавиш характеризуют темп работы, а время удержания клавиш — стиль работы с клавиатурой: резкий удар или плавное нажатие [4].

К этим двум характеристикам добавляют среднюю скорость набора текста, измеренную в различных эпизодах работы с клавиатурой [4]:

$$U_n = \frac{\sum_{i=1}^n u_i}{n},$$

где u_i — средняя скорость набора в i -м эпизоде.

При этом за некоторое время измерения появляется M значений 3-мерного вектора, характеризующих M эпизодов взаимодействия пользователя с клавиатурой, — матрица эксперимента $M \times 3$. Ее обработка также

может быть сведена к трем вышеперечисленным вариантам, если считать процесс работы пользователя с клавиатурой стационарным.

Таким образом, для идентификации информационного почерка пользователя необходимо отслеживать значения как минимум восьмимерного вектора:

$$P = (T, L, U_p, d, a, U_n, t_{уд}, t_n). \quad (2)$$

Исследование взаимосвязей между параметрами информационного почерка на выборке объемом более 10 000 у нескольких пользователей показывает, что статистическая связь параметров клавиатурного почерка и параметров траекторий компьютерной мыши слаба (табл. 1). Это говорит о самоценности каждого из компонентов вектора (2) параметров ИнП. Из корреляционной матрицы следует также гипотеза об относительной автономии каждого из параметров и отсутствии избыточности в совокупности выбранных характеристик ИнП.

Таблица 1

Корреляционная матрица по Пирсону параметров ИнП

Параметры	L	U_p	T	d	a	t_n	$t_{уд}$	U_n
L	1	0,27	0,68	0,11	0,49	0,23	-0,1	-0,05
U_p		1	0,04	-0,11	-0,07	0,28	0,03	0,07
T			1	0,2	0,45	0,13	0,14	0,05
d				1	0,27	0,41	0,25	0,47
a					1	0,01	0,06	0,03
t_n						1	0,51	0,41
$t_{уд}$							1	0,54
U_n								1

3. Две процедуры формирования эталонов информационного почерка

В основе большинства алгоритмов распознавания лежит гипотеза компактности, состоящая в том, что реализации одного образа (матрица $N \times 8$) в 8-мерном геометрическом пространстве образуют «компактные сгустки». Процесс идентификации ИнП состоит из двух основных этапов: обучения (отображения) ИнП каждого пользователя и принятия решения о принадлежности наблюдаемого процесса взаимодействия пользователя с ПК одному из эталонов (отображений) ИнП.

Процедура формирования образа ИнП в реальном времени на этапе «обучения» может быть разбита на три повторяющиеся стадии:

1) измерение значений параметров событий и процессов, связанных с указанными выше устройствами во время работы пользователя с компьютером.

Формирование матрицы эксперимента. При этом необходимо нормирование значений параметров;

2) формирование образа ИнП одним из перечисленных способов;

3) шифрование результата.

Пополнение обучающей выборки может происходить, например, через каждые 5 мин. Программное обеспечение для осуществления этих измерений создано и апробировано. Высокая производительность современных компьютеров делает этот процесс совершенно незаметным для пользователя.

Определяющей на этом, первом этапе является стадия 2. Учитывая, что вектор P случаен, необходимы методы математической статистики для формирования отображений ИнП. При этом нет проблемы репрезентативности выборки, так как получение количества измерений в несколько тысяч не составляет труда.

Первая процедура основана на отображении ИнП точкой в 8-мерном пространстве — $ИнП_c = (p_1, \dots, p_8)$, где p_i , ($i = 1, \dots, 8$), — средние значения столбцов матрицы эксперимента.

Вторая процедура сводится к редукции матрицы эксперимента в матрицу 2×8 , первой строкой которой являются минимальные значения, второй — максимальные значения соответствующих столбцов матрицы эксперимента. В этом случае образ ИнП — это многомерный параллелепипед ($ИнП_n$):

$$\Delta P = (p_{11} \leq p_1 \leq p_{12}, p_{21} \leq p_2 \leq p_{22}, \dots, p_{81} \leq p_8 \leq p_{82}). \quad (3)$$

4. Алгоритмы идентификации пользователя

Правила формирования $ИнП_c$ и $ИнП_n$ определяют процедуры идентификации ИнП работающего за

компьютером пользователя. Напомним, что рассматриваем пока только стационарный процесс взаимодействия пользователя с компьютером.

Алгоритм 1. Уместно рассмотреть простейший вариант сопоставления евклидовых расстояний между текущим образом ИнП (текущей точкой) и каждым из созданных образов допущенных к системе пользователей. Существенным ограничением этого метода является невозможность абсолютной идентификации пользователя. Правда, эту проблему можно решить в процессе длительного использования метода и создания d -окрестностей ИнП_c. Для этого необходимо отслеживать флуктуации положения конкретной точки и определять эту окрестность в виде гиперпараллелепипеда, объем которого

$$V_n = (p_{1, \max} - p_{1, \min}) \cdot (p_{2, \max} - p_{2, \min}) \cdot \dots \cdot (p_{n, \max} - p_{n, \min}).$$

Границы гиперпараллелепипеда рассматриваются в виде минимальных и максимальных значений координат либо в статистических доверительных интервалах для координат. Последнее потребует как минимум проверки гипотезы на нормальность распределений координат.

Алгоритм 2 базируется на представлении ИнП в виде гиперпараллелепипеда — ИнП_n. В частности, можно использовать метод «дробящихся эталонов» — распознавание с использованием покрытия обучающих выборок каждого образа простыми фигурами, усложняющимися по мере необходимости [6]. Принятие решения о сходстве или различии двух ИнП_n при пересечении гиперпараллелепипедов осуществляется следующим образом. При пересечении двух образов область пересечения делится на дополнительные образы следующего поколения, которые в свою очередь при пересечении также могут быть разбиты (рис.). Использование в качестве покрывающих фигур гиперпараллелепипедов позволит упростить процесс вычисления объемов возможных пересечений разных образов — ИнП_n разных пользователей. Этот метод может дать экономию времени расчетов до 30 % [6]. Решение о сходстве/различии ИнП принимается по результатам сравнения объемов всех пересечений ИнП_n.

Для абсолютной идентификации пользователя требуется существенно большая статистика об ИнП_n конкретного пользователя. Далее определяется минимальное значение объема пересечения текущего и эталонного ИнП_n и принимается решение об их идентичности.

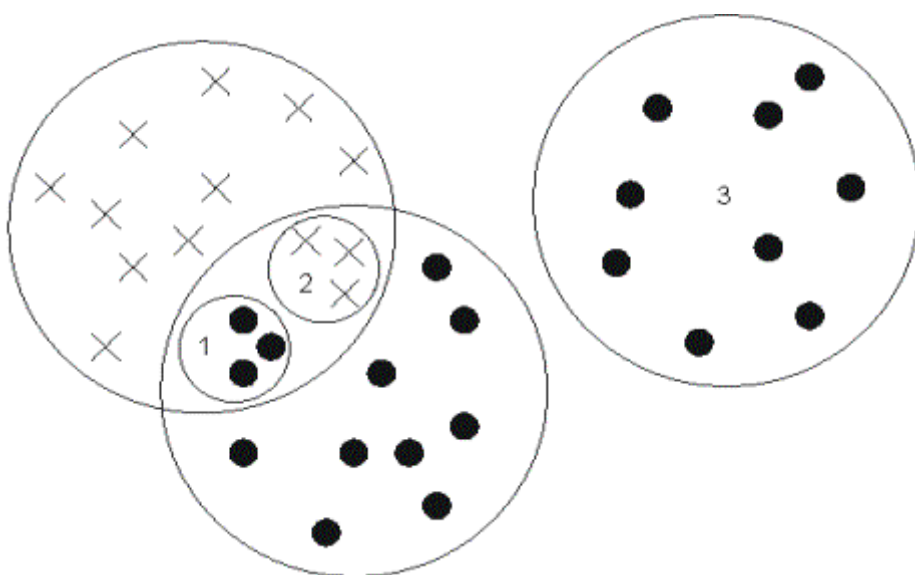


Рис. Последовательное деление пересечений гиперпараллелепипедов ИнП_n на подобласти

Заключение

В статье введено понятие информационного почерка более общее, чем «клавиатурный почерк», и дано его отображение в компьютере в виде точки и/или гиперпараллелепипеда в 8-мерном евклидовом пространстве. Предлагается перевод термина «информационный почерк» (ИнП) на английский язык как InfoHandWorking (IHW). В словарях и справочниках [7–9], Интернете (Rambler, Yahoo) прецедентов не обнаружено.

Представлена процедура формирования в компьютере образа ИнП в реальном времени. Программы для мониторинга динамики мыши и клавиатуры разработаны и используются авторами для продолжения исследований по оценке эффективности алгоритмов распознавания.

Литература

1. Бушуев С. И., Авраменко В. С. Аутентификация пользователей в автоматизированных системах на основе информационного почерка // Сб. тр. 1-й Международной науч.-практ. конф. «Проблемы современной геополитики. Продление НАТО на Восток — проблемы безопасности России и стран СНГ». 2002 г. // www.chuvashi.narod.ru.
2. Власов А. Н. Способ представления координатной составляющей информационного почерка пользователя // Материалы м/у науч. конф. по мягким вычислениям. СПб., 2003. Т. 1. С. 116–119.
3. Диденко С. М., Шапцев В. А. Исследование динамики работы пользователя с манипулятором мышь // Математическое и информационное моделирование. Тюмень: Изд-во Тюм. ун-та, 2004. С. 295–304.
4. Гузик В. Ф., Галуев Г. А., Десятерик М. Н. Биометрическая нейросетевая система идентификации пользователя по особенностям клавиатурного почерка // Нейрокомпьютеры. Разработка, применение. 2001. № 7–8. С. 104–118.
5. Расторуев С. П. Программные методы защиты информации в компьютерах и сетях. М.: Изд-во Агентства «Яхтсмен», 1993. 188 с.
6. Загоруйко Н. Г. Прикладные методы анализа данных и знаний. Новосибирск: Изд-во Ин-та математики, 1999. 270 с.
7. Бобылев В. Н. Краткий этимологический словарь научно-технических терминов. М.: Логос, 2004. 95 с.
8. Кузьмин Ю. А., Владимиров В. А., Гельман Я. Л. и др. Краткий англо-русский технический словарь. 2-е изд., испр. М.: ЧеРо, 1998. 416 с.
9. Ожегов С. И. Толковый словарь русского языка. М.: АЗЪ, 1995. 928 с.

S. M. Didenko, V. A. Shaptsev

IDENTIFICATION METHODS REGARDING
USER'S INFORMATION HAND

The work suggests an approach to identification of users in the systems with a limited access, basing on parameters of user's information hand under his operation in the system. The article describes parameters of information hand suggesting a method for its decoding.