УДК 519.8;004

В.М. Сафонов[1]

МОДЕЛЬ КЛАВИАТУРНОГО ПОЧЕРКА В ЗАДАЧАХ ЗАЩИТЫ АВТОМАТИЗИРОВАННЫХ СИСТЕМ

Предлагается универсальная модель для описания клавиатурного почерка оператора автоматизированной системы, учитывающая информативность, коррелированность, устойчивость, изменчивость носителей признаков, а также ситуационные условия формирования образца клавиатурного почерка.

Ключевые слова: *клавиатурный почерк, модель клавиатурного почерка, информативность носителей признаков, коррелированность носителей признаков, изменчивость клавиатурного почерка, устойчивость носителей признаков, ситуационные условия формирования клавиатурного почерка.*

V.M. Safonov

MODEL OF KEYBOARD HANDWRITING IN THE PROTECTION OF AUTOMATED SYSTEMS

The author proposes a universal model to describe handwriting keyboard operator of the automated system, taking into account the information content correlated, stability, variability of media features, as well as the situational conditions of formation of the sample keyboard handwriting.

Keywords: *keyboard writing, handwriting keyboard model, the information content of media features, correlation of media features, the variability of handwriting keyboard, stability carriers signs, situational conditions of formation of a keyboard handwriting.*

Введение

В настоящее время с развитием средств обхода информационной защиты всё более важным становится вопрос идентификации личности оператора АРМ. Основной принцип построения современных средств защиты заключается в комплексном парировании воздействий на объект информатизации. Данные воздействия отличаются большим многообразием и хорошо описываются перечнем факторов, воздействующих на информацию, и определенных в ГОСТ Р 51275-99. Однако в настоящее время не все угрозы комплексно парируются как организационными, так и техническими средствами. В частности, такое ИТВ противника, как атака маскировкой под зарегистрированного пользователя, парируется только организационными мероприятиями.

Воздействие данного типа происходит по следующим сценариям: узнали пароль – вошли в систему; воспользовались отсутствием оператора после аутентификации – получили доступ к АРМу.

Анализ подходов к созданию средства защиты от данных типов воздействий привел к выводу о целесообразности использования поведенческих технологий для идентификации пользователя. В подтверждение данной мысли отметим, что на сегодняшний день разрабатывается большое число систем, идентифицирующих оператора по его поведению. Например, APPLE идентифицирует человека по особенностям колебания телефона при его походке [1].

Из всех технологий идентификации по динамическим характеристикам, которые возможно использовать для решения данной задачи, на сегодняшний день наибольшее развитие получила только технология идентификации по клавиатурному почерку (КП).

На данный момент существует несколько моделей клавиатурного почерка (Шарипов Р.Р. [3], Савинов А.Н. [4], Казарин М.Н. [5], Брюхомицкий Ю.А. [6]). Однако эти модели обладают множеством недостатков, среди них:

− модели учитывают только 3 носителя признака (НП) КП (длительность удержания конкретных клавиш, длительность пауз между удержаниями смежных в наборе клавиш, длительность возможного перекрытия в удержании смежных в наборе клавиш);

− не учитывается информативность носителей признаков КП;

− не учитывается коррелированность носителей признаков КП;

− не учитываются ситуационные условия идентификации оператора по КП.

[1] Адъюнкт кафедры математического и программного обеспечения Военно-космической академии им. А.Ф. Можайского.

Устранение данных недостатков позволит как повысить адекватность модели, так и точность идентификации по клавиатурному почерку.

Основная часть

При проведении исследований КП было зафиксировано достаточно большое количество НП КП, характеризующих индивидуальный стиль работы с клавиатурой. Приведем их список: период времени удержания клавиш, период времени между отпусканием предыдущей клавиши и нажатием следующей клавиши, период времени между двумя нажатиями, математическое ожидание периода времени удержания клавиши, среднеквадратическое отклонение периода времени удержания клавиши, математическое ожидание периода времени между отпусканием предыдущей клавиши и нажатием следующей клавиши, среднеквадратическое отклонение периода времени между отпусканием предыдущей клавиши и нажатием следующей клавиши, математическое ожидание периода времени между двумя нажатиями, среднеквадратическое отклонение периода времени между двумя нажатиями, математическое ожидание периода времени удержания N клавиш, среднеквадратическое отклонение периода времени удержания N клавиш, математическое ожидание периода времени между отпусканием предыдущей клавиши и нажатием следующей клавиши при нажатии N клавиш, среднеквадратическое отклонение периода времени между отпусканием предыдущей клавиши и нажатием следующей клавиши при нажатии N клавиш, математическое ожидание периода времени между двумя нажатиями при нажатии N клавиш, среднеквадратическое отклонение периода времени между двумя нажатиями при нажатии N клавиш, коэффициент корреляции между временем удержания клавиши и периодом времени между отпусканием предыдущей и нажатием следующей клавиши, количество нажатых горячих клавиш на N нажатых клавиш, период времени удержания отдельных клавиш, частота употребления слов предметной области, отношение времени между нажатиями клавиш к расстоянию между ними, частота возникновения ошибок при вводе текста на N символов, скорость ввода (количество символов на единицу времени) на родном языке, скорость ввода на иностранном языке, аритмичность набора, период времени между нажатиями клавиш, отношение нажатия табуляции к нажатию пробелов на N символов, использование вспомогательной цифровой клавиатуры (или отношение набора цифр на основной к набору цифр на вспомогательной клавиатуре), отношение использования клавиш стрелок к клавишам мыши (перемещение каретки мышью или стрелками), период времени пауз между «блоками набора» (характеризует манеру работы), отклонение от среднего периода времени между «блоками набора», наличие уникальных клавиш для идентификации факта работы на ноутбуке, типа клавиатуры, переключение между левой Alt+Shift или правой Alt+Shift (или Ctrl+Shift), большие перерывы в работе (их анализ для определения времени суток, обеденного перерыва и т.д.), период времени между нажатиями функциональных и рабочих клавиш.

Представляется, что НП КП классифицируются по следующим основаниям (рис. 1).

Рис. 1. Схема структуры классификации НП КП

Приведем пример классифицирования НП. Среднеквадратическое отклонение периода времени между отпусканием предыдущей клавиши и нажатием следующей клавиши при нажатии N клавиш – это НП 3 уровня, скалярный, характеризующийся непрерывной величиной, обобщенный, устойчивый, получаемый при использовании группы клавиш.

Как было указано выше, существует большое число НП КП и признаков КП. В рамках определения используемых признаков КП актуальна задача рассмотрения атрибутивных свойств НП КП. К моменту начала работы над статьей исследованию свойств НП должного внимания не уделялось.

Можно выделить следующие атрибутивные свойства НП КП:

– информативность НП КП;

– устойчивость НП КП в зависимости от объема используемых исходных данных;

– коррелированность НП КП;

– зависимость НП КП от ситуационных условий (психофизическое состояние оператора, особенности вводимой информации и т.п.).

Информативность признаков

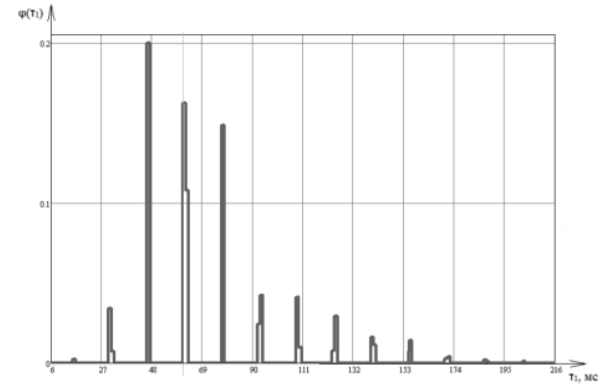
При принятии решения об использовании того или иного НП КП для идентификации пользователя необходимо учитывать, что каждый из НП потенциально может выделить неодинаковое число объектов или групп объектов из контрольной группы, то есть НП неодинаково информативны. На данный момент не предложено способа анализа информативности носителей признаков. Значения информативности НП КП могут использоваться для ранжировки НП КП по отношению превышения информативности. Кроме того, учет информативности позволит минимизировать количество НП КП, требуемых для идентификации, и тем самым уменьшить объем вычислений и используемой памяти.

Информативность НП КП – это свойство НП, или вектора НП, характеризующее число объектов (или групп), которые возможно выделить из контрольной группы на основании принятия ими различных значений этого признака.

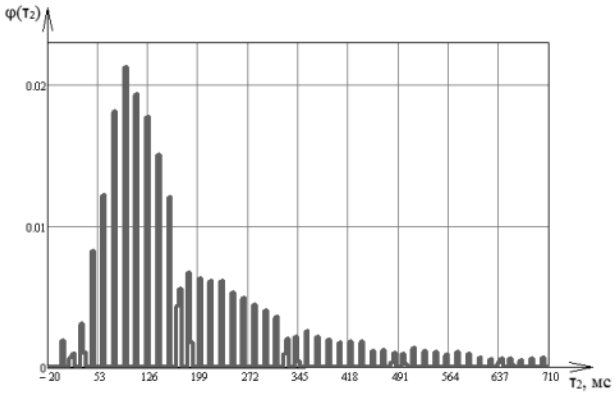
Предлагаем информативность НП определять как уровень энтропии НП. В этом случае информативность дискретного НП определяется выражением для определения энтропии h:

$$h = -\sum_{j=1}^{a} p_j \log_2 p_j , \qquad (1)$$

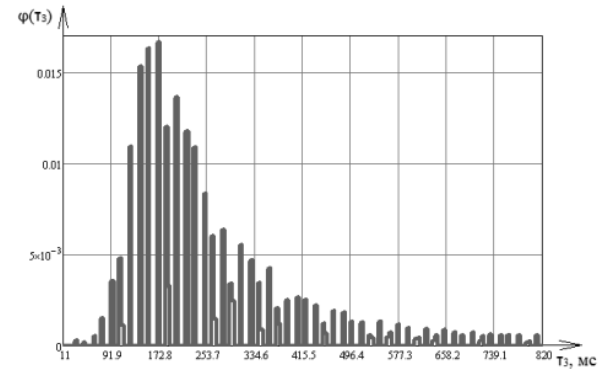
где p_j – вероятность j-го значения дискретной величины НП распределения; a – число значений дискретной величины. В рамках предлагаемого подхода: чем больше h, тем больше информативность НП. На рис. 2 приведены диаграммы вероятностей появления дискретных скалярных величин некоторых НП КП.



h = 3,45



h = 5,44



h = 5,53

Рис. 2. Диаграммы вероятностей появления дискретных скалярных величин, принимаемых (сверху вниз) периодами времени удержания клавиши, периодами времени между отпусканием предыдущей и нажатием следующей клавиши, периодами времени между нажатиями двух клавиш

Как следует из рисунка, для рассматриваемых признаков информативность периода времени между нажатиями двух клавиш самая высокая. Это обусловлено тем, что размах носителя

распределения данной величины самый большой, а следовательно, больше вариантов реализаций данной случайной величины. Поэтому, если для осуществления идентификации необходимо по тем или иным причинам выбрать один из носителей признаков, в качестве основного признака следует выбрать период времени между нажатиями двух клавиш.

Отметим, что информативность определяется таймером операционной системы и режимом её многозадачности. Увеличение интенсивности многозадачности будет увеличивать дискретность таймера, следовательно, дискретность моментов фиксации событий клавиатуры, следовательно, уменьшать информативность.

Однако существуют и непрерывные НП КП, например НП КП 3 уровня обработки, такие, как математическое ожидание периода времени между двумя нажатиями при нажатии n клавиш, среднеквадратическое отклонение периода времени между двумя нажатиями при нажатии n клавиш и т.д.

Расчет информативности непрерывного НП осуществляется в соответствии с выражением [7]:

$$h = -\int_{-\infty}^{\infty} \varphi_{\hat{x}}(x) \log_2\left(\varphi_{\hat{x}}(x)\right) dx. \qquad (2)$$

Ниже представлены графики плотности вероятности непрерывных скалярных значений НП КП: периода времени удержания клавиши и периодами времени между отпусканием предыдущей и нажатием следующей клавиши и результаты расчета их информативности (рис. 3).

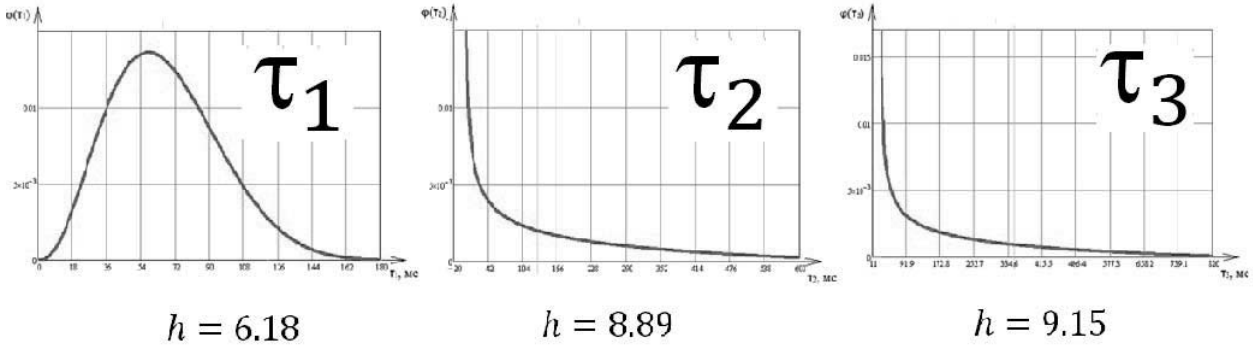


Рис. 3. График плотности вероятности непрерывных скалярных значений, принимаемых (слева направо) периодами времени удержания клавиши и периодами времени между отпусканием предыдущей и нажатием следующей клавиши

Таблица 1

Сравнение информативности непрерывных скалярных НП КП

НП КП	Значение информативности	
	по каждому нажатию	по скользящему среднему за 50
Период времени удержания клавиши	6,18	3,79
Период времени между отпусканием предыдущей и нажатием следующей клавиши	8,89	6,67
Период времени между нажатиями двух клавиш	9,15	5,48

Как следует из анализа информативности дискретных и непрерывных НП КП, во-первых, непрерывные НП более информативны, чем дискретные; во-вторых, чем шире носитель распределения НП КП, тем признак более информативен; в-третьих, чем более равномерное распределение НП КП при прочих равных условиях, тем больше информативность.

Приведем выражение для информативности двумерного векторного дискретного НП КП:

$$h = \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} \log_2 p_{ij}, \qquad (3)$$

где p_{ij} – вероятность попадания двумерного дискретного НП КП в i, j-е интервалы соответственно. Для трехмерного случая выражение будет иметь вид:

$$h = \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{s} p_{ijk} \log_2 p_{ijk}, \qquad (4)$$

для большей размерности по аналогии.

Выражение для расчёта информативности двумерного векторного непрерывного НП КП имеет вид:

$$h = -\int\int_{-\infty}^{\infty} \varphi_{\langle \hat{x},\hat{y} \rangle}(x,y) \log_2 \left(\varphi_{\langle \hat{x},\hat{y} \rangle}(x,y) \right) dxdy, \qquad (5)$$

где $\varphi_{\langle \hat{x},\hat{y} \rangle}(x, y)$ – плотность распределения НП КП $\langle \hat{x}, \hat{y} \rangle$. Для случая большей размерности выражение записывается аналогично.

Приведем некоторые соображения по ранжированию НП КП на основе отношения превышения информативности. Во-первых, чем выше значение информативности, тем выше ранг НП КП. Во-вторых, вероятности выпадения конкретного значения НП КП и плотности распределения вектора непрерывных НП КП рассчитываются на объединённой статистике всех рассматриваемых операторов.

Коррелированность носителей признаков клавиатурного почерка

Большое значение для выбора используемых НП КП при проведении идентификации имеет свойство коррелированности. Коррелированность характеризует зависимость случайных величин. Если коррелированность двух НП высокая, то рационально использовать только один из них. Степень выраженности свойства коррелированности НП КП характеризуется коэффициентом корреляции. Отметим, что коэффициент корреляции может быть рассчитан только для синхронных реализаций НП КП.

В табл. 2 в качестве примера зависимости признаков приведены рассчитанные значения коэффициентов корреляции для различных скалярных НП КП. Коэффициенты рассчитывались на основании выборок из 2500 символов, введенных одним из операторов.

Таблица 2

Коэффициенты корреляции НП КП

Обозначение НП КП	Название НП КП	$k_{\hat{\tau}_i \hat{\tau}_j}$		
		$\hat{\tau}_1$	$\hat{\tau}_2$	$\hat{\tau}_3$
$\hat{\tau}_1$	Продолжительность периода времени нажатия клавиши	1	–0,051	–0,024
$\hat{\tau}_2$	Продолжительность периода времени между отпусканием предыдущей клавиши и нажатием следующей	–0,051	1	0,866
$\hat{\tau}_3$	Продолжительность периода времени между нажатиями двух клавиш	–0,024	0,866	1

Продолжение таблицы

Как видно из приведенной таблицы, значение коэффициента корреляции между $\hat{\tau}_2$ и $\hat{\tau}_3$ достаточно высоко, что позволяет сделать вывод о том, что можно использовать только один из этих НП. Кроме того, значения коэффициентов корреляции случайных величин $\hat{\tau}_1$ и $\hat{\tau}_2$ близки к нулю, что свидетельствует об их независимости. Отметим, что для других операторов наблюдается аналогичная закономерность.

Устойчивость носителей признаков клавиатурного почерка

Для регистрации признаков КП необходимо сформировать эталонный КП операторов. Для этого оператором вводится некоторый текст. Можно задаться вопросом: какой объём текста необходим для формирования эталона КП оператора?

Кроме того, при работе СЗИ от НСД на базе КП надо определиться с числом символов, которое необходимо для идентификации КП.

Для ответа на сформулированные вопросы введем понятие устойчивости НП КП. Под устойчивостью признаков КП будем понимать способность системы идентификации КП обеспечивать заданную вероятность проявления признака КП.

Устойчивость характеризуется:

– устойчивостью средних значений периодов времени при нажатии n клавиш;

– устойчивостью закона распределения НП КП.

Для примера на рис. 4 приведены графики среднего значения периода времени нажатия клавиши и среднего значения периода времени между отпусканием предыдущей клавиши и нажатием следующей при увеличении числа набранных оператором символов. Как следует из приведенных графиков, при увеличении числа введенных символов более 1000 средние значения практически не увеличиваются, что указывает на объем текста, необходимого для создания эталона КП по указанному НП КП оператора.

На рис. 5 приведены графики аппроксимаций бета-распределением при различных объемах выборки при ее накоплении. В результате анализа изменяемости вида закона распределения выявлено, что при объеме выборки более 800 график закона практически не меняется.

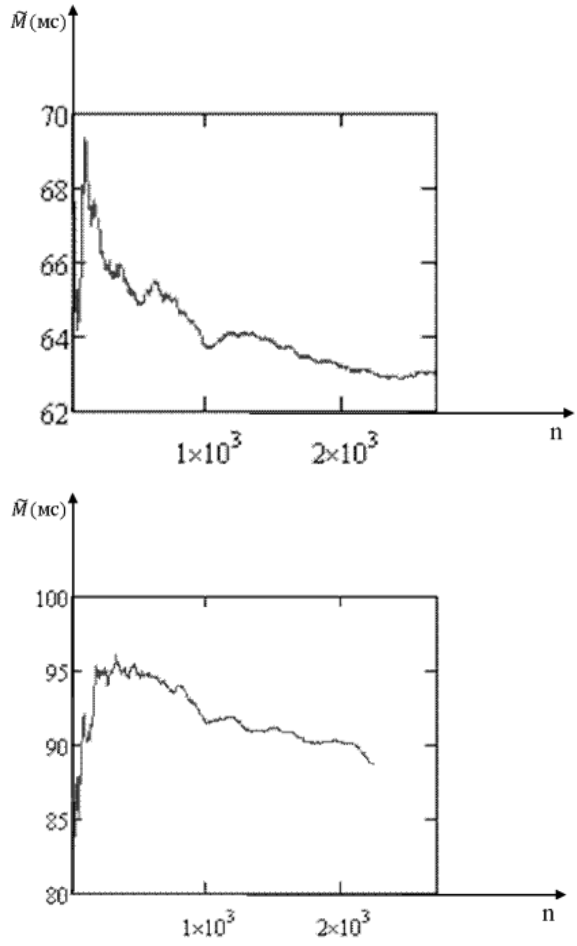


Рис. 4. Графики, характеризующие устойчивость среднего значения периода времени нажатия клавиши и среднего значения периода времени между отпусканием предыдущей клавиши и нажатием следующей клавиши (n – количество нажатий)

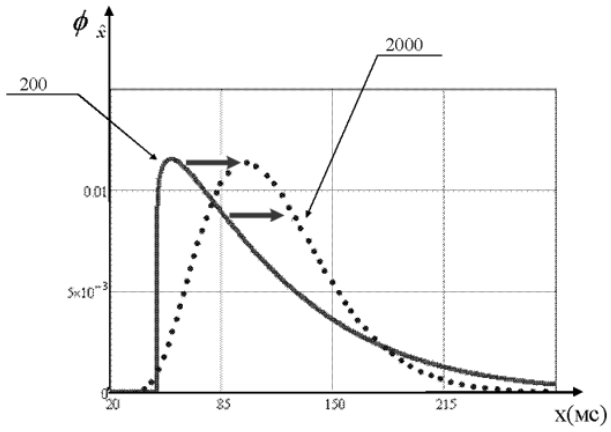


Рис. 5. Уточнение плотности распределения случайной величины продолжительности нажатия клавиши при фиксации 200 и 2000 нажатий

Общий профиль клавиатурного почерка

Текущие профили КП являются характеристикой индивидуальных особенностей работы пользователя с клавиатурой в рамках конкретной сессии. Представляется, что существуют некоторые характерные особенности поведения пользователя в рамках различных сессий. Эти особенности могут либо искажать реальный КП пользователя АС, либо влиять на изменение КП с течением продолжительных периодов времени, например в случае болезни, изменения психофизиологического состояния, возрастных изменений и т.д.

Поэтому правомерно использование понятия общего профиля (ОП) КП – описывающего основные черты текущего профиля (ТП) КП. Описание ОП КП возможно производить в рамках стохастического описания реализаций многомерного вектора случайных величин – атрибутов (носителей признаков) ТП КП. Если вектор $\hat{X}_{\langle n\rangle}$ – вектор атрибутов ТП КП, объединяющий все ТП КП, то ОП КП может быть описан законом распределения вектора случайных величин, например плотность распределения $\varphi_{\hat{X}}\left(\hat{X}_{\langle n\rangle}; h, A_{\langle m\rangle}\right)$ вектора $\hat{X}_{\langle n\rangle}$ случайных величин (где h – идентификатор пользователя, $A_{\langle m\rangle}$ – вектор характеристик условий «обобщения» ТП КП). Следует отметить, что условия обобщения могут характеризовать разные типы ОП КП, например ОП КП при разных психофизиологических состояниях или ОП КП пользователя в разное время суток и т.д.

Для описания законов распределения возможно использовать аппарат обобщенных функций. Примером такого использования является смесь законов распределений – закон распределения двумерного случайного вектора $\hat{X}_{\langle 2\rangle} = \langle \hat{x}, \hat{z} \rangle$, если один компонент $\hat{x}$ является непрерывной случайной величиной, а другой компонент $\hat{z}_j$ – дискретной случайной величиной, причем компоненты независимы.

Дискретная случайная величина $\hat{z}$ может задаваться известным рядом распределения.

Общая схема использования ОП КП имеет следующий вид, приведенный на рис. 6. Из данной схемы видно, каким образом связаны в задачах идентификации пользователя ТП КП и ОП КП.

Как видно из данной схемы, помимо формирования текущего клавиатурного почерка при осуществлении сессий работы с АРМ, происходит перманентное или периодическое уточнение КП пользователя АС и формирование базы данных ОП КП. Использование информации, накопленной в этой БД, позволит учитывать искажения и изменения КП и тем самым повысит адекватность моделируемому объекту.

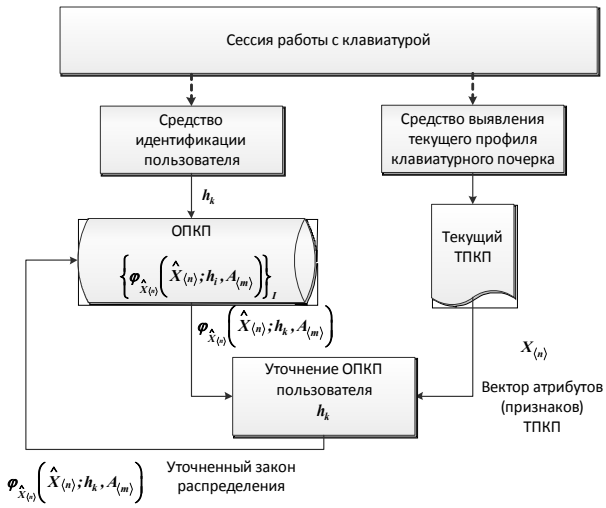


Рис. 6. Общая схема использования ОП КП

Таким образом, в данной главе сформулирована проблема искажения и изменения КП пользователя АС. Также предложен способ учета этих искажений и изменений в процессе работы со средствами идентификации по КП путем составления базы данных общего профиля клавиатурного почерка. Применение данной БД в соответствии с предложенной схемой позволит повысить адекватность модели КП, и как следствие – повысит точность идентификации пользователя АС.

Ситуационные условия формирования КП

Модель КП помимо адекватности моделируемому объекту, то есть непосредственно клавиатурному почерку пользователя, должна обладать адекватностью решаемым задачам. Как уже отмечалось, многие элементы модели КП окончательно формируются только с учетом ситуационных условий их формирования. Дадим определение. *Ситуационные условия формирования КП – это условия, фиксирующие доминирующие факторы формирования КП.*

Приведем классификацию ситуационных условий формирования КП:

– по режиму формирования: формирование эталона; формирование образца КП;

– по способу фиксации образцов КП: явное; скрытое;

– по принадлежности к задачам: защита АС; аудит безопасности АС;

– по характеру действий оператора: малое число вводимых символов; большое число вводимых символов;

– по способу ввода символов: однопальцевая печать; многопальцевая печать;

– по психофизиологическому состоянию оператора: в спокойном состоянии; в состоянии усталости; под действием веществ, изменяющих психофизиологическое состояние;

– по типу используемой клавиатуры: механическая; мембранная; электронная;

– по виду хранения эталона КП: в форме характеристик первого уровня; в форме характеристик второго уровня; в форме характеристик третьего уровня;

– по месту фиксации признаков КП: на АРМ оператора; на внешнем сервере;

– по требуемым затратам времени на идентификацию пользователя: высокооперативное; низкооперативное;

– по месту фиксации характеристик первого уровня: на уровне контроллера клавиатуры; на уровне событий операционной системы.

Таким образом, фиксация ситуационных условий и их учет при формировании КП обеспечивает предъявление требований реализации к СЗИ. Учет этих требований обеспечит модели КП адекватность решаемым задачам.

Заключение

Таким образом, в данной статье была предложена модель клавиатурного почерка, которая может обеспечить создание средств защиты информации для противодействия маскировке под зарегистрированного пользователя и позволяющая:

1) учесть большее количество признаков клавиатурного почерка;

2) учесть информативность и коррелированность признаков КП;

3) предъявить требования к тексту, который необходимо ввести пользователю для формирования эталона его КП;

4) учесть ситуационные условия формирования КП оператора;

5) учесть изменения клавиатурного почерка оператора во времени.

Всё это позволит повысить точность идентификации оператора по КП, а также снизить затраты вычислительных ресурсов на осуществление этой идентификации.

Литература

1. Болл Р.М., Коннел Дж.Х., Панканти Ш., Рахта Н.К., Сеньор Э.О Руководство по биометрии. – М. : Техносфера, 2007. – 386 с.

2. Apple узнает пользователя по походке [Электронный ресурс]. – Режим доступа: http://migom.by/news/apple_uznaet_polzovatelya_po_pohodke/

3. Шарипов Р.Р. Разработка полигауссовско-

го алгоритма аутентификации пользователей в телекоммуникационных системах и сетях по клавиатурному почерку: дис. … канд. техн. наук : 05.12.13 / Шарипов Рифат Рашатович. – Казань, 2006. – 137 с.

4. Савинов А.Н. Методы, модели и алгоритмы распознавания клавиатурного почерка в ключевых системах: дис. … канд. техн. наук : 05.13.19 / Савинов Александр Николаевич. – СПб., 2013. – 137 с.

5. Казарин М.Н. Разработка и исследование методов скрытого клавиатурного мониторинга: дис. … канд. техн. наук : 05.13.19 / Казарин Максим Николаевич. – Таганрог, 2006. – 142 с.

6. Брюхомицкий Ю.А. Клавиатурная идентификация личности. Исследование методов идентификации личности по клавиатурному почерку. – Saarbrücken : LAP LAMBERT Academic Publishing, 2012. – 138 с.

7. Энтропия непрерывной случайной величины [Электронный ресурс]. – Режим доступа: http://peredacha-informacii.ru/entropija-nepreryvnoj-sluchajnoj-velichiny.html