УДК 004.056

И.А. Ходашинский, М.В. Савчук, И.В. Горбунов, Р.В. Мещеряков

Технология усиленной аутентификации пользователей информационных процессов

Целью работы является разработка технологии, реализующей систему усиленной аутентификации, на базе динамической аутентификации по клавиатурному почерку пользователя. В качестве объекта усиления была выбрана парольная аутентификация. В процессе работы были проанализированы существующие методы классификации пользователей при использовании метода биометрической аутентификации по клавиатурному почерку. Также был рассмотрен подход к выявлению ключевых параметров характеристики клавиатурного почерка пользователя. Полученные знания были использованы для реализации библиотеки. Проведено исследование с целью определения оптимального значения коэффициента Стьюдента, отвечающего за уменьшение ошибки первого рода.

Ключевые слова: аутентификация, пользователь, биометрические методы, параметрическое обучение, ковариационная матрица.

Одним из главных элементов любой системы защиты от несанкционированного доступа (НСД) является элемент, обеспечивающий контроль доступа к ЭВС и контроль работы в них. Этот элемент защиты выполняет свои функции с помощью процедур идентификации и аутентификации пользователей. Эти процедуры важны, потому что любой системе защиты от НСД для выполнения своей задачи необходимо, чтобы все легальные пользователи были идентифицированы и гарантировалось бы соответствие между пользователями и их идентификаторами, так как все остальные элементы системы защиты работают с идентифицированными субъектами.

Система защиты выполняет аутентификацию на основе определенной уникальной информации, которая характеризует конкретного пользователя системы [1]. Такая информация называется аутентификационной. В зависимости от ее типа методы аутентификации классифицируют на следующие группы: парольная, имущественная и биометрическая.

Простыми методами аутентификации являются методы, основанные на использовании уникальной информации, которая известна пользователю и которую он предъявляет в процессе аутентификации, так называемая парольная система аутентификации. Данные методы получили наибольшее распространение благодаря простоте реализации, внедрению и использованию. Но все они имеют огромный недостаток, который делает их непригодными для применения в серьезных системах защиты. Этот недостаток состоит в том, что при нарушении конфиденциальности пароля пользователя сразу нарушается защита всей информации, к которой он (владелиц) имеет доступ.

Анализ динамической аутентификации по клавиатурному почерку пользователя по биометрическим характеристикам

Биометрия представляет собой аутентификацию людей на основе их физиологических и поведенческих характеристик [2]. К числу физиологических характеристик принадлежат особенности отпечатков пальцев, сетчатки и роговицы глаз, геометрия руки и лица и т.п. К поведенческим характеристикам относятся динамика ручной подписи, стиль работы с клавиатурой, тембр голоса [3, 4]. На стыке физиологии и поведения находятся особенности голоса и распознавание речи.

Методы биометрической аутентификации делятся на две большие группы:

– статические методы, которые основываются на физиологической (статической) характеристике человека, т.е. уникальном свойстве, данном ему от рождения и неотъемлемом от него;

– динамические методы, которые основываются на поведенческой (динамической) характеристике человека – особенностях, характерных для подсознательных движений в процессе воспроизведения какого-либо действия (подписи, речи, динамики клавиатурного набора) [5].

Биометрическая аутентификация не определяет пользователя с абсолютной точностью. С паролем все просто: он либо эквивалентен эталону, либо нет. Системы биометрической аутентификации опознают пользователя с определенной вероятностью, так как биометрическая система может не

узнать легального пользователя или, что еще хуже, принять чужого за своего. Поэтому все системы биометрической аутентификации оцениваются по двум характеристикам:

– отказ в доступе (ошибка первого рода – FRR, false rejection rate) – с какой вероятностью система не узнает зарегистрированного пользователя;

– ложный доступ (ошибка второго рода – FAR, false access rate) – вероятность ошибочного допуска нелегального пользователя.

Описание механизма динамической аутентификации

В динамических биометрических системах аутентификации, реализующих аутентификацию по рукописному и клавиатурному почерку, соответствующие биометрические характеристики личности преобразуются обычно в некоторый вектор **V** информативных биометрических параметров, представляемый в *N*-мерной ортогональной системе координат.

Компоненты вектора **V** формируются в результате прямых измерений параметров почерка (клавиатурного) [6] или путем вычисления линейных ортогональных функционалов Фурье, Уолша, Хаара по полной реализации вводимой пользователем парольной фразы [4]. Компоненты вектора V в общем случае обладают внутренней корреляцией.

В основе механизма аутентификации лежит огромная сложность задачи управления движениями человека. Формально эту задачу можно представить в виде некоторой упрощенной модели с тремя выходами и множеством входов управления приводами-мышцами.

В первом приближении относительная сложность задачи может быть оценена по числу входов модели или по числу мышц, участвующих в том или ином типе движения. Так, во время письма ручкой оказываются задействованы мышцы большинства пальцев руки и мышцы предплечья [7]. Всего может быть задействовано более 50 мышц, но, видимо, наиболее существенное влияние будут оказывать примерно 10 мышц. То есть при управлении рукой во время письма человеку приходится в реальном времени решать как минимум десятимерную задачу управления.

При работе с клавиатурой (слепое печатание текста) дополнительно включаются еще примерно 20 мышц плеча и плечевого пояса каждой руки, т.е. при слепой печати двумя руками потенциально могут быть задействованы 140 мышц. Исходя из предположения, что наибольшее влияние оказывают только 20% от общего числа мышц, получим $n = 28$ или примерно 28-мерную задачу управления.

При воспроизведении речи участвуют 44 мышцы груди, 9 мышц живота и брюшной полости, 28 мышц лица и челюстей, 12 мышц языка, 9 мышц глотки, 6 мышц мягкого неба. В общей сложности оказывается задействовано порядка 110 мышц, что при гипотезе 20% группы наибольшего влияния дает 22-мерную задачу управления.

Следует обратить внимание на то, что во всех рассмотренных случаях задача управления, во-первых, оказывается многомерной (как минимум 10-мерной) и, во-вторых, число управляющих входов всегда существенно превышает число выходов.

Задача большой размерности позволяет сделать предположение об уникальности клавиатурного почерка пользователя и утверждать о целесообразности использования динамических методов идентификации пользователя на практике, в частности, метода идентификации пользователя по особенностям клавиатурного почерка.

Анализ метода аутентификации по клавиатурному почерку

Описание параметров пользователя для аутентификации

В задаче аутентификации пользователя по клавиатурному почерку важным этапом является обработка первичных данных. В результате этой обработки входной поток данных разделяется на ряд признаков, характеризующих те или иные качества аутентифицируемой личности. В дальнейшем эти признаки, подвергаясь обработке, позволяют получить ряд эталонных характеристик пользователя.

Начальный этап обработки данных – фильтрация. На этом этапе из потока данных удаляется информация о «служебных» клавишах – клавишах управления курсором, функциональных клавишах и т.п.

Затем выделяется информация, относящаяся к характеристикам пользователя:

– количество ошибок при наборе;

– интервалы между нажатиями клавиш;

– время удержания клавиш;

– число перекрытий между клавишами;

– степень аритмичности при наборе;

– скорость набора.

После статистической обработки этих данных рассчитанные эталонные характеристики пользователя сохраняются в базе данных.

Методы классификации пользователей

В процессе аутентификации, по существу, решается задача классификации неизвестного пользователя, предъявившего свои биометрические параметры в виде вектора **V**, на «своего» и «чужого». При этом применяются различные методы классификации:

– геометрические, использующие различные меры близости предъявляемого вектора **V** к биометрическому эталону $\mathbf{V}_Э$ (мера Хэмминга, Евклидова мера и др.);

– методы на основе применения искусственных нейронных сетей (ИНС);

– параметрического обучения классификатора.

Задача построения классификатора, реализующего разделение входных биометрических векторов **V** на «своих» и «чужих», в общем случае сводится к выбору какого-либо метода построения дискриминантной функции $g(\mathbf{V})$, реализующей указанное разделение. В свою очередь, выбор метода построения $g(\mathbf{V})$ зависит от характера классифицируемых объектов.

Если параметры классов объектов изначально не известны, то дискриминантная функция $g(\mathbf{V})$ задается в явном виде с использованием m действительных параметров $w_1, w_2, \ldots, w_m$, называемых весами. Из множества различных функции $g(\mathrm{V})$ выделяют определенные классы функций: линейные, кусочно-линейные, квадратичные и др. [8]. Для построения классификатора из каких-либо практических соображений (точность, быстродействие, простота и т.п.) выбирается один из классов функций. Последующее обучение классификатора в выбранном классе функций $g(\mathbf{V})$ сводится к процедуре подбора весов $w_1, w_2, \ldots, w_m$. К такому методу построения дискриминантной функции $g(\mathbf{V})$ относятся геометрические методы, методы, использующие обучаемые ИНС.

В том случае, если параметры классов объектов изначально известны, можно воспользоваться другими − параметрическими методами обучения, когда дискриминантную функцию получают, используя обучающее множество объектов для оценок самих величин параметров класса.

Геометрические методы и методы, основанные на ИНС, обладают следующими недостатками:

– снижение точности классификации вследствие грубой аппроксимации областей решения (геометрические методы);

– неопределенно долгий процесс обучения, возможность возникновения тупиков и т.п. (методы на основе ИНС);

– проблема обучения на неопределенно широкий класс возможных «чужих» пользователей (методы на основе ИНС) [9].

Параметрический метод обучения классификатора лишен указанных недостатков. Поэтому именно он был выбран для использования в разрабатываемой системе аутентификации.

Особенности аутентификации по клавиатурному почерку

Временные интервалы между нажатием клавиш на клавиатуре и время удержания (нажатия) клавиш позволяют достаточно однозначно охарактеризовать почерк работы пользователя на клавиатуре, что подтверждается рядом экспериментов [4]. При этом временные интервалы между нажатием клавиш характеризуют темп работы, а время удержания клавиш характеризует стиль работы с клавиатурой (резкий удар или плавное нажатие).

Однако существует ряд ограничений на применение данного способа на практике. Применение способа идентификации по клавиатурному почерку целесообразно только по отношению к пользователям с достаточно длительным опытом работы с компьютером и сформировавшимся почерком работы на клавиатуре, т.е. программистам, секретарям-референтам, машинисткам и т.п. В противном случае вероятность неправильного опознавания «легального» пользователя существенно возрастает и делает непригодным данный способ идентификации на практике.

Идентификация пользователя по клавиатурному почерку возможна [10] по набору ключевой фразы и произвольного текста.

Принципиальное отличие этих двух способов заключается в том, что в первом случае ключевая фраза задается пользователем в момент регистрации его в системе (пароль), а во втором случае используются ключевые фразы, генерируемые системой каждый раз в момент идентификации пользователя. Оба способа подразумевают два режима работы – обучение и идентификацию.

На этапе обучения пользователь вводит некоторое число раз предлагаемые ему тестовые фразы. При этом рассчитываются и запоминаются эталонные характеристики данного пользователя. На этапе идентификации рассчитанные оценки сравниваются с эталонными, на основании чего делается вывод о совпадении или несовпадении параметров клавиатурного почерка. Выбор текста, на котором выполняется обучение системы, – достаточно важный этап для нормального функционирования системы. Предлагаемые пользователю фразы необходимо подбирать таким образом, чтобы используемые в них символы полностью и равномерно покрывали рабочее поле клавиатуры.

Эталонные характеристики пользователя, полученные на этапе обучения системы, позволяют сделать выводы о степени стабильности клавиатурного почерка пользователя и определить доверительный интервал разброса параметров для последующей идентификации пользователя. Во избежание дискредитации работы системы можно отсеивать пользователей, клавиатурный почерк которых не обладает необходимой стабильностью (табл. 1) [4].

Таблица 1

Характеристики стабильности клавиатурного почерка пользователей

Ошибки, %	Аритмичность, %	Скорость, зн./мин	Характеристика перекрытий		Оценка
			Число перекрытий, %	Используемое число пальцев	
Менее 2	Менее 10	Более 200	Более 50	Все	Отлично
Менее 4	Менее 15	Более 150	Более 30	Большинство	Хорошо
Менее 8	Менее 20	Более 100	Более 10	Несколько	Удовл.
Более 8	более 20	Менее 100	Менее 10	По одному	Неуд.

Технология выявления ключевых параметров характеристики клавиатурного почерка пользователя

Анализ и построение модели клавиатурного почерка пользователя

В общем виде, функция $\vartheta(t)$, описывающая процесс набора текста пользователем на клавиатуре, может быть представлена следующим образом:

$$\vartheta(t)=\gamma(t)+\theta(t)+\lambda(t), \quad (1)$$

где $\gamma(t)$ – составляющая, характеризующая подсознательные процессы мышления при наборе текста; $\theta(t)$ – составляющая сознательных процессов мышления; $\lambda(t)$ – механические характеристики клавиатуры, влияющие на процесс набора текста.

Основной задачей системы биометрической идентификации пользователя по особенностям клавиатурного почерка является задача выделения и последующей идентификации компоненты $\gamma(t)$ из функции $\vartheta(t)$, которая определяет исходные данные для системы идентификации после измерения временных характеристик клавиатурного почерка пользователя. Для этого соответственно необходимо выделить компоненты $\theta(t)$ и $\lambda(t)$ из исходной функции $\vartheta(t)$. Очевидно, что из-за невозможности построения механической модели движений человека при наборе текста единственным приемлемым решением является сбор статистических данных о клавиатурном почерке большого числа пользователей и построение эмпирических зависимостей $\theta(t)$ и $\lambda(t)$.

Вероятность аутентификации пользователя по времени удержания клавиш в зависимости от длины ключевой фразы является значительно более стабильной характеристикой клавиатурного почерка пользователя, чем время между нажатиями клавиш (пауз), которое и растет с ростом длины ключевой фразы. Это объясняется тем, что процесс нажатия клавиши на клавиатуре является истинно подсознательным процессом мышления. Характер данной функции практически не меняется для широкого круга пользователей независимо от их квалификации и опыта работы с клавиатурой. Отсюда следует, что составляющая $\gamma(t)$ наиболее точно характеризуется временем удержания клавиш при вводе текста пользователем.

Время между нажатиями клавиш является менее стабильной характеристикой клавиатурного почерка пользователя, чем время удержания клавиш. Функция вероятности идентификации от пауз: между нажатиями клавиш имеет максимум своего значения при длине ключевой фразы порядка 8–10 символов. Это объясняется тем, что ключевые фразы небольшой длины, состоящие из одного, максимум двух слов, пользователь набирает подсознательно. Подсознательные движения стабильны до тех пор, пока в них не вмешивается более высокий сознательный уровень мышления, что приво-

дит к появлению эффекта «сороконожки», сбивающейся при попытке понять, как же она ходит. Проявление данного эффекта объясняет уменьшение вероятности аутентификации пользователя при превышении длины ключевой фразы некоторого критического уровня. Следует отметить, что значение данного порога достаточно сильно варьируется для пользователей с различным опытом работы с клавиатурой и может колебаться от 6 до 30 символов. После этого предела даже у квалифицированных машинисток наблюдается эффект включения сознательного мышления и остановок в наборе текста для принятия решения. В соответствии с изложенными выводами можно говорить о том, что в системах аутентификации пользователя по особенностям клавиатурного почерка не рекомендуется использовать слишком длинные выражения в качестве ключевой фразы, так как это приводит к тому, что пользователь начинает «осмысленно» выполнять набор текста, что может привести к снижению качества его аутентификации. Таким образом, можно утверждать, что время между нажатиями клавиш наиболее полно передает характер составляющей $\theta(t)$ функции клавиатурного почерка $\vartheta(t)$. Исключение составляющей $\theta(t)$ из функции $\vartheta(t)$ осуществляется введением корреляционной функции $\Delta(t)$, эмпирически получаемой на основе анализа имеющихся статистических данных. Так как характер корреляционной функции $\Delta(t)$ должен выбираться в зависимости от опыта работы пользователя с клавиатурой, согласно табл. 1, в нее необходимо ввести параметр L, который определяет соответствие корреляционной функции и составляющей $\theta(t)$.

Исследования параметров клавиатурного почерка одного пользователя на разных клавиатурах дают разброс вероятности идентификации примерно в 0,5%, что вполне можно объяснить случайными причинами. Полученный результат позволяет утверждать, что механические характеристики клавиатуры практически не влияют на особенности клавиатурного почерка пользователя, и, соответственно, пренебречь составляющей $\lambda(t)$ в выражении (1).

С учетом вышесказанного можно записать функцию $\vartheta(t,L)$ следующим образом:

$$\vartheta(t,L)=\gamma(t)+\Delta(t,L). \tag{2}$$

Очевидно, что с течением времени клавиатурный почерк пользователя может претерпевать некоторые изменения. Это может выражаться в том, что при выполнении i-й процедуры идентификации мы получим вероятность идентификации ρ_i, отличающуюся на некоторую величину ξ от математического ожидания m_ρ вероятностей, полученных на предыдущих операциях идентификации пользователя.

С учетом этого можно записать выражение (2) в следующем виде:

$$\vartheta(t,L)=\big(\gamma(t)+\Delta(t,L)\big)T(t,L), \tag{3}$$

где $T(t,L)$ – корреляционная функция, учитывающая изменение параметров клавиатурного почерка пользователя с течением времени.

Формирование вектора входных параметров

Пусть n – длина ключевой фразы, применяемой для идентификации пользователя; t_i^{down}, t_i^{up} – время нажатия и время отпускания i-й клавиши; c_i – ASCII код i-й клавиши ($i=\overline{1,n}$), тогда

$$\begin{aligned} t_i^{\text{press}} &= t_i^{\text{up}} - t_i^{\text{down}},\ i=\overline{1,n};\\ t_i^{\text{pause}} &= t_{i+1}^{\text{down}} - t_i^{\text{up}},\ i=\overline{1,n-1}, \end{aligned} \tag{4}$$

где t_i^{press} – время удержания i-й клавиши в нажатом положении; t_i^{puse} – длительность паузы между отпусканием i-й клавиши и нажатием $(i+1)$-й клавиши.

Аритмичность скорости набора α и математическое ожидание $\overline{m}^{\text{pause}}$ вычисляются следующим образом:

$$\overline{m}^{\text{pause}}=\frac{\sum\limits_{i=1}^{n-1}\dfrac{t_i^{\text{pause}}}{t_{\max}^{\text{pause}}}}{(n-1)};\quad \alpha=\sqrt{\frac{\sum\limits_{i=1}^{n-1}\left(\dfrac{t_i^{\text{pause}}}{t_{\max}^{\text{pause}}}-\overline{m}^{\text{pause}}\right)^2}{n-2}}. \tag{5}$$

Аритмичность времени удержания клавиш β и математическое ожидание $\overline{m}^{\text{press}}$ вычисляются следующим образом:

$$\overline{m}^{\text{press}} = \frac{\sum_{i=1}^{n} t_i^{\text{press}}}{n};\ \beta = \sqrt{\frac{\sum_{i=1}^{n}\left(\frac{t_i^{\text{press}}}{t_{\max}^{\text{press}}} - \overline{m}^{\text{press}}\right)^2}{n-1}}. \tag{6}$$

Скорость набора $\bar{s}$:

$$\bar{s} = \frac{\left(t_n^{\text{up}} - t_1^{\text{down}}\right)}{60} s_{\max}, \tag{7}$$

где $s_{\max}$ – максимальная скорость набора текста, что соответствует 900 знакам/мин [11].

Наличие факта перекрытия между i-й и (i+1)-й клавишами определяется, если выполняется условие $t_i^{\text{down}} \le t_{i+1}^{\text{up}}$ и $t_{i+1}^{\text{down}} \le t_i^{\text{up}}$. В противном случае факт перекрытия клавиш отсутствует. Для клавиш, у которых наблюдается факт перекрытия, итерационно выполняется следующий процесс:

$$n_c = n_c + 1;$$

$$t_c = t_c + t_i^{\text{up}} - t_{i+1}^{\text{down}},$$

т.е. подсчитывается число клавиш с перекрытиями и суммарное время перекрытий.

Далее рассчитывается нормированное среднее время перекрытий $\overline{t_c}$ и среднеквадратичное отклонение $\overline{t_c}$:

$$\overline{t_c} = \frac{t_c}{n_c t_{\max}^{\text{press}}};\ d_c = \sqrt{\frac{\sum_{i}\left(\frac{t_i^{\text{up}} - t_{i+1}^{\text{down}}}{t_{\max}^{\text{press}}} - \overline{t_c}\right)^2}{n_c - 1}}. \tag{8}$$

В результате получаем вектор входных параметров $\mathbf{v}$, используемый для идентификации пользователя по клавиатурному почерку при фиксированной длине ключевой фразы, и его длину l:

$$\begin{aligned} &\mathbf{v} = \left\{t_1^{\text{press}},\ldots,t_n^{\text{press}},t_1^{\text{pause}},\ldots,t_{n-1}^{\text{pause}},\overline{c_1},\ldots\overline{c_n},\overline{m}^{\text{pause}},\alpha,\overline{m}^{\text{press}},\beta,\bar{s},\overline{t_c},d_c\right\}; \\ &l = 3n - 6 = 3(n-2), \end{aligned} \tag{9}$$

где $\overline{c_i}$ – нормированное значение ASCII кодов клавиш:

$$\overline{c_i} = \frac{c_i - 32}{223}; \tag{10}$$

В соответствии с (1) необходимо разбить сформированный вектор входных параметров $\mathbf{v}$ на составляющие $\mathbf{v}_\gamma$ и $\mathbf{v}_\theta$, характеризующие подсознательные и сознательные процессы соответственно:

$$\begin{aligned} &\mathbf{v}_\gamma = \left\{t_1^{\text{press}},\ldots,t_n^{\text{press}},\overline{m}^{\text{press}},\beta,\overline{t_c},d_c\right\}; \\ &l_\gamma = n + 4; \\ &\mathbf{v}_\theta = \left\{t_1^{\text{pause}},\ldots,t_{n-1}^{\text{pause}},\overline{m}^{\text{pause}},\alpha,\bar{s}\right\}; \\ &l_\theta = n + 2; \\ &\mathbf{v} = \mathbf{v}_\gamma + \mathbf{v}_\theta + \left\{\overline{c_1},\ldots\overline{c_n}\right\}; \\ &l = l_\gamma + l_\theta + n. \end{aligned} \tag{11}$$

В сформированный вектор входных параметров включаются не только величины, отражающие особенности клавиатурного почерка пользователя, но и сама ключевая фраза, а именно ASCII коды нажатых клавиш, что позволяет еще более повысить надежность работы системы в режиме аутентификации пользователя. Эта надежность достигается тем, что по этим ASCII кодам будет проводиться парольная аутентификация и только в случае ее успешного прохождения будет производиться аутентификация по клавиатурному почерку, которая будет давать окончательное решение по процедуре аутентификации.

Технология биометрической аутентификации пользователя на основе параметрического обучения классификатора

Для работы системы динамической аутентификации по клавиатурному почерку авторизованный пользователь должен предъявить L своих подписей, т.е. пройти этап регистрации (обучения). Эти **L** подписей пользователя соответствуют L реализациям вектора биометрических параметров $\mathbf{V}=\{\mathbf{V}_1, \mathbf{V}_1, \ldots \mathbf{V}_L\}$.

В динамических биометрических системах аутентификации, использующих характеристики клавиатурного почерка, распределение векторов биометрических параметров $\mathbf{V}_i$ в N-мерном пространстве в большинстве случаев можно считать близким к нормальному [4], поэтому векторы $\mathbf{V}_i$, $i=\overline{1,L}$ можно задать в виде функции плотности нормального распределения векторов $\mathbf{V}_i$ с неизвестными средними, которая при $L\to\infty$ в ортогональной системе координат описывается гиперэллипсоидом рассеивания. Причем, в общем случае компоненты биометрических векторов $\mathbf{V}_i$, $i=\overline{1,L}$ коррелированы между собой, т.е. главные оси гиперэллипсоида рассеивания не параллельны осям координат. Следовательно, получив формулу такого гиперэллипсоида, аутентификацию пользователя можно осуществлять путем контроля попадания вектора его биометрических параметров **V** внутрь N-мерной области, описываемой гиперэллипсоидом рассеивания [9].

Тогда параметрический метод обучения классификатора будет состоять из трех последовательных этапов:

Этап 1. Устанавливается в явном виде зависимость дискриминантных функций $g(\mathbf{V})$ от параметров, характеризующих функцию плотности нормального распределения векторов $\mathbf{V}_i$.

Этап 2. По обучающему множеству векторов $\mathbf{V}_i$ оцениваются величины этих параметров.

Этап 3. Предполагается, что указанные оценки являются истинными значениями параметров, и они подставляются в выражение для $g(V)$, полученное на этапе 1.

Процедуру аутентификации пользователя, предъявившего свои биометрические параметры в виде вектора **V**, можно рассматривать как задачу классификации вектора **V** на (M+1) классов – по числу M зарегистрированных в системе пользователей («свои»), плюс один класс для всех остальных не зарегистрированных в системе пользователей («чужие»). Однако перед прохождением процедуры аутентификации пользователь должен вначале аутентифицировать себя одним из общепринятых способов, в нашем случае это будет парольная система аутентификации. При этом предварительно зарегистрированные в системе биометрические параметры данного пользователя могут использоваться как параметры классификатора. Это обстоятельство позволяет без нарушения общности свести задачу классификации вектора **V** на (M + 1) классов к задаче классификации только на два класса: «свой» – вектор $\mathbf{V}_C$ и «чужой» – вектор $\mathbf{V}_\text{Ч}$. При наличии в системе многих зарегистрированных пользователей, после аутентификации каждого из них достаточно поочередно внести в классификатор в качестве эталона соответствующие ему параметры (один из M зарегистрированных векторов $\mathbf{V}_C$).

Для классификации пользователей на два класса «свой» (вектор $\mathbf{V}_C$) и «чужой» (вектор $\mathbf{V}_\text{Ч}$) параметрический классификатор можно реализовать с использованием только одной дискриминантной функции $g(\mathbf{V})$, знак которой будет определять принадлежность предъявленного вектора **V** к одному из двух классов: $\mathbf{V}_C$ и $\mathbf{V}_\text{Ч}$. При этом области распределения биометрических параметров всевозможных «чужих» в совокупности можно рассматривать как интегральную область «все чужие», расположенную вокруг компактной области «свой» [9].

Пусть в общем случае область распределения биометрических параметров «своего» пользователя задана множеством образцов $\mathbf{\Psi}_C$, состоящим из L векторов $\mathbf{V}_{Ci}$, i = 1...L, нормально распределенных в N-мерном пространстве ортогональной системы координат, а каждый вектор $\mathbf{V}_{Ci}$, i = 1…L представлен своими N компонентами

$$\mathbf{V}_{Ci}=\left\{\nu_1,\nu_1,\ldots,\nu_j,\ldots,\nu_N\right\},\ j=\overline{1,N}.$$

Центр распределения векторов $\mathbf{V}_{Ci}$ находится в точке $(\xi_1,\xi_2,\ldots,\xi_N)$, которая определяется N математическими ожиданиями $m_{\nu 1}=\xi_1$, $m_{\nu 2}=\xi_2$, …, $m_{\nu N}=\xi_N$. Центральные моменты второго порядка распределения векторов $\mathbf{V}_{Ci}$, образуют квадратную матрицу моментов (ковариационную матрицу)

$$\mathbf{Q} = \left\| \lambda_{jk} \right\| = \begin{Vmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1N} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{2N} \\ \dots & \dots & \dots & \dots \\ \lambda_{N1} & \lambda_{N2} & \dots & \lambda_{NN} \end{Vmatrix},$$

где

$$\lambda_{jk} = \lambda_{kj} = M\left(\nu_j - \xi_j\right)\left(\nu_k - \xi_k\right) = \begin{cases} \sigma_{ij}^2 & \text{при } j = k, \\ \operatorname{cov}\left\{\nu_j, \nu_k\right\} & \text{при } j \neq k. \end{cases} \quad j,k = \overline{1,N}.$$

Для нормального закона распределения векторов $\mathbf{V}_{Ci}$, i=1…L функция плотности распределения имеет вид

$$f\left(\nu_1, \nu_2, \dots, \nu_N\right) = \frac{1}{\sqrt{(2\pi)^N \det\left\| \lambda_{jk} \right\|}} \exp\left[-\frac{1}{2} \sum_{j=1}^{N} \sum_{k=1}^{N} \Lambda_{jk} \left(\nu_j - \xi_j\right)\left(\nu_k - \xi_k\right) \right], \tag{12}$$

где $\det\left\| \lambda_{jk} \right\|$ – определитель ковариационной матрицы $\mathbf{Q} = \left\| \lambda_{jk} \right\|$.

Коэффициенты Λ_{jk} составляют матрицу $\mathbf{\Lambda} = \left\| \Lambda_{jk} \right\|$, обратную ковариационной матрице $\mathbf{Q} = \left\| \lambda_{jk} \right\|$.

Для вычисления коэффициентов Λ_{jk} используется стандартная формула

$$\Lambda_{jk} = (-1)^{j+k} \frac{M_{jk}}{\det\left\| \lambda_{jk} \right\|}, \tag{13}$$

где M_{jk} – минор определителя $\det\left\| \lambda_{jk} \right\|$, получаемый из него вычеркиванием j-й строки и k-го столбца.

Выражение, фигурирующее в показателе экспоненты функции плотности нормального распределения векторов $\mathbf{V}_{Ci}$, является положительно-определенной квадратичной формой. Поверхности, на которых эта квадратичная форма постоянна

$$\frac{1}{2} \sum_{j=1}^{N} \sum_{k=1}^{N} \Lambda_{jk} \left(\nu_j - \xi_j\right)\left(\nu_k - \xi_k\right) = \text{const}, \tag{14}$$

являются поверхностями равных плотностей вероятностей в N-мерном пространстве и представляют собой гиперэллипсоиды, которые группируются вокруг точки $(\xi_1, \xi_2, \dots, \xi_N)$.

Обозначая константу в правой части выражения (14) через k^2, получим

$$\frac{1}{2} \sum_{j=1}^{N} \sum_{k=1}^{N} \Lambda_{jk} \left(\nu_j - \xi_j\right)\left(\nu_k - \xi_k\right) = k^2. \tag{15}$$

Константа k задает коэффициент пропорциональности между длинами a_j главных полуосей гиперэллипсоида и соответствующими среднеквадратическими отклонениями σ_j:

$$a_1 = k \cdot a_1;\ a_2 = k \cdot a_2;\ \dots,\ a_N = k \cdot a_N.$$

Для оптимального решения задачи классификации из всех поверхностей равных плотностей вероятностей целесообразно выбрать ту, которая характеризует рассеивание векторов $\mathbf{V}_{Ci}$ относительно точки $(\xi_1, \xi_2, \dots, \xi_N)$. Эта поверхность соответствует так называемому единичному гиперэллипсоиду, у которого главные полуоси равны соответствующим среднеквадратическим отклонениям $\sigma_1, \sigma_2, \dots, \sigma_N$, т.е. для единичного гиперэллипсоида k = 1, и выражение (15) преобразуется к виду

$$\frac{1}{2} \sum_{j=1}^{N} \sum_{k=1}^{N} \Lambda_{jk} \left(\nu_j - \xi_j\right)\left(\nu_k - \xi_k\right) = 1. \tag{16}$$

Из-за ограниченной статистики биометрических образцов, предъявляемых на стадии регистрации «своим» пользователем, всегда остается вероятность того, что образец, предъявленный этим же пользователем при аутентификации, выйдет за пределы зафиксированного в эталоне диапазона. Для уменьшения этой вероятности дополнительно задается величина допуска между областями «свой» и

«все чужие» в виде коэффициента Стьюдента $C\left[L_1,(1-P_1)\right]$ [4], исходя из заданной ошибки первого рода (вероятность P_1 ложного отказа «своему» пользователю) и числа L предъявленных на стадии регистрации образцов. Введение указанного допуска в уравнение (16) приводит его к виду

$$\frac{1}{2}\sum_{j=1}^{N}\sum_{k=1}^{N}\Lambda_{jk}\left(\nu_j-\xi_j\right)\left(\nu_k-\xi_k\right)=\left\{C\left[L,(1-P_1)\right]\right\}^2. \qquad (17)$$

Значения коэффициента Стьюдента при заданных P_1 и L приведены в табл. 2.

Иллюстрация метода для двухмерного пространства ($N = 2$) показана на рис. 1.

Таблица 2

Значения коэффициента Стьюдента

Число образцов L	Вероятность ошибки первого рода − P_1 (вероятность отказа в допуске подлинному пользователю)								
	0,1	0,05	0,03	0,025	0,02	0,015	0,01	0,005	0,0025
2	3,07	6,31	10,56	12,5	15,9	21,21	31,82	63,7	127,3
3	1,88	2,92	3,89	4,3	4,85	5,64	6,97	9,92	14,1
4	1,63	2,35	2,95	3,18	3,48	3,82	4,54	5,84	7,54
5	1,53	2,13	2,60	2,78	2,99	3,25	3,75	4,60	5,60
6	1,47	2,01	2,44	2,57	2,75	3,01	3,37	4,03	4,77
7	1,43	1,94	2,31	2,45	2,61	2,83	3,14	3,71	4,32
8	1,41	1,89	2,24	2,36	2,51	2,72	3,00	3,50	4,03
9	1,39	1,86	2,19	2,37	2,45	2,63	2,90	3,36	3,83
10	1,38	1,83	2,15	2,26	2.40	2,57	2,82	3,25	3,69
12	1,36	1,80	2,09	2,16	2.33	2,49	2,72	3,11	3,49
14	1,35	1,77	2,06	2,14	2,28	2,43	2,65	3,01	3,37
16	1,34	1,75	2,03	2,12	2,24	2,39	2,60	2,95	3,28
18	1,33	1,74	2,01	2,10	2,22	2,36	2,57	2,9	3,22
21	1,33	1,73	1,99	2,09	2,19	2,33	2,53	2,85	3,15
26	1,32	1,71	1,97	2,06	2,16	2,30	2,49	2,79	3,07
31	1,31	1,70	1,95	2,04	2,14	2,27	2,46	2,75	3,03
41	1,30	1,68	1,93	2,02	2,12	2,25	2,42	2,7	2,97
∞	1,28	1,65	1,89	1,96	2,06	2,18	2,33	2,58	2,81

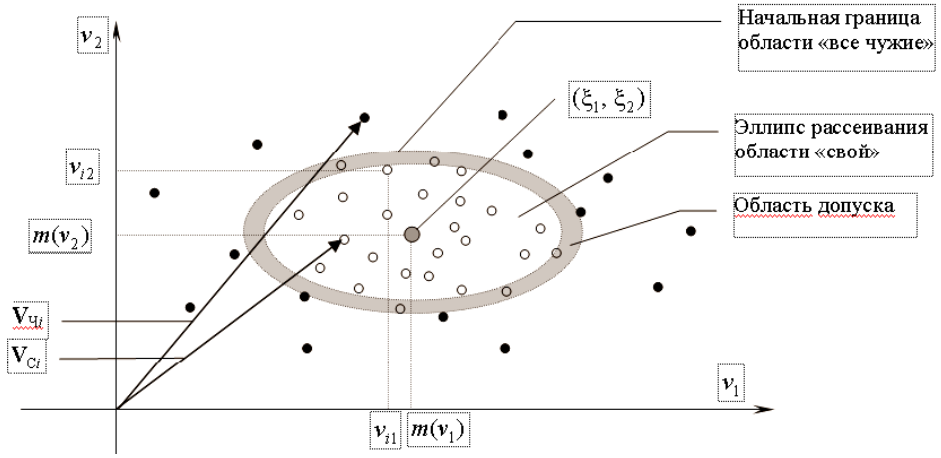

Рис. 1. Иллюстрация метода биометрической аутентификации пользователя на основе параметрического обучения классификатора

Используем полученное выражение (17) для формирования дискриминантной функции $g(V)$, разделяющей области «свой» и «все чужие»:

$$g(V)=\frac{1}{2}\sum_{j=1}^{N}\sum_{k=1}^{N}\Lambda_{jk}\left(\nu_j-\xi_j\right)\left(\nu_k-\xi_k\right)-\left\{C\left[L,(1-P_1)\right]\right\}^2. \qquad (18)$$

Уравнение $g(\mathbf{V}) = 0$ в этом случае будет определять искомую разделяющую поверхность, а знак функции $g(\mathbf{V})$ – принадлежность входного вектора $\mathbf{V}$ к одному из двух классов: «свой» или «чужой» (попадание в область «все чужие»):

$$\begin{array}{lll} g(V)<0, & \text{если} & V \in V_C, \\ g(V)>0, & \text{если} & V \in V_Ч. \end{array}$$

Таким образом, процедура аутентификации сводится теперь к проверке: попадает ли предъявленный пользователем вектор биометрических параметров $\mathbf{V}$ в область, описываемую, выражением (18).

Алгоритм работы метода биометрической аутентификации на основе параметрического обучения классификатора представлен на рис. 2.

Реализация системы усиленной аутентификации

В качестве основы для усиления был взят метод парольной аутентификации, который обладает рядом положительных качеств, описанных ранее. В качестве «усилителя» был выбран механизм аутентификации по клавиатурному почерку, являющийся одним из методов динамической аутентификации по биометрическим данным.

Модуль усиления аутентификации целесообразно выполнить в качестве самостоятельной библиотеки, которая будет полностью выполнять аутентификацию по клавиатурному почерку, методом параметрического обучения классификатора, в нее только нужно подать необходимые данные из пользовательского приложения, так называемый вектор входных параметров.

Для использования возможностей библиотеки нужно в пользовательское приложение внедрить модуль подключения библиотеки, состоящий из процедур сбора данных и двух функций, одна из которых будет передавать данные для регистрации нового пользователя, а вторая – для непосредственной аутентификации.

Само пользовательское приложение может содержать свой механизм парольной аутентификации и использовать тот, который был реализован в процессе выполнения данной работы. Данная возможность библиотеки позволяет как усиливать уже существующие системы парольной аутентификации, так и внедрять системы аутентификации в приложения, которые до этого не имели такой системы.

Рис. 2. Алгоритм работы метода биометрической аутентификации на основе параметрического обучения классификатор

Для тестирования данного приложения в системе аутентификации было зарегистрировано три пользователя. Для регистрации каждый из пользователей ввел 5 раз свой логин и пароль. На следующем этапе каждый из пользователей ввел 50 раз свои входные данные (для определения ошибки первого рода) и по 50 раз данные других двух пользователей (для определения ошибки второго ро-

да). По полученным результатам аутентификации пользователей были определены ошибки первого и второго рода у Пользователя1, Пользователя2 и Пользователя3, которые представлены в виде графиков на рис. 3–5 соответственно.

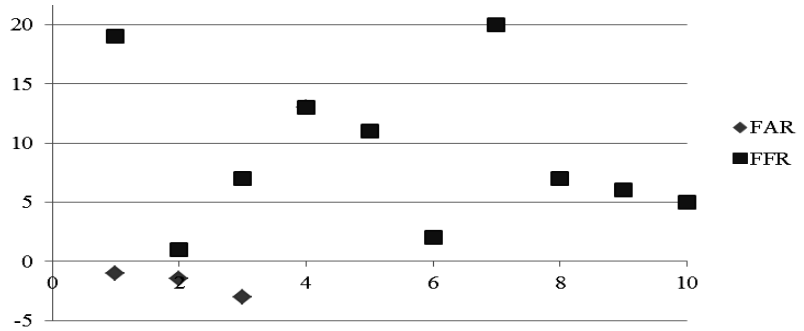

Рис. 3. Ошибки первого (FRR) и второго (FAR) рода у Пользователя1

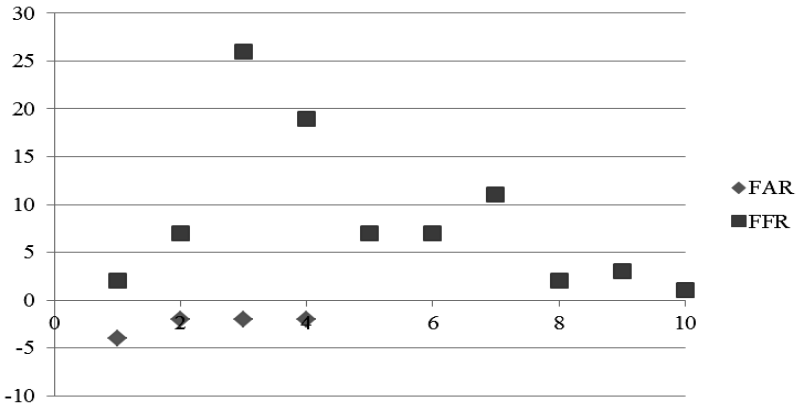

Рис. 4. Ошибки первого (FRR) и второго (FAR) рода у Пользователя2

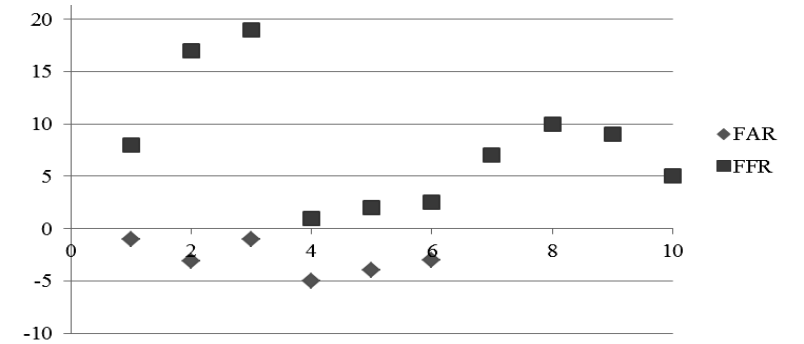

Рис. 5. Ошибки первого (FRR) и второго (FAR) рода у Пользователя3

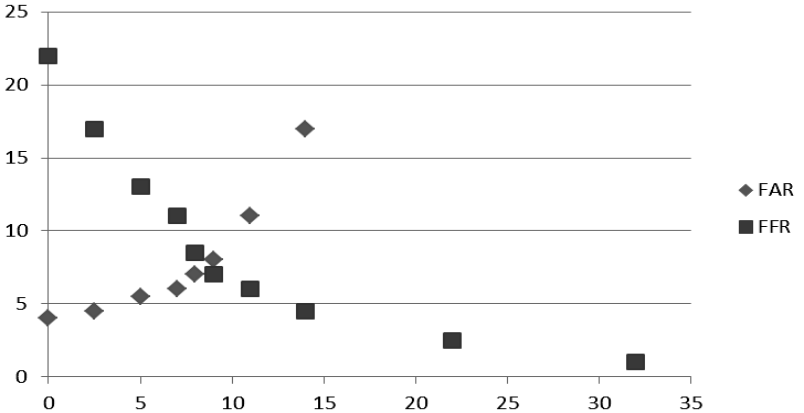

Рис. 6. Ошибки первого (FRR) и второго (FAR) рода для различных коэффициентов Стьюдента

Из представленных графиков видно, что среднее количество ошибок первого рода (FRR) равно 10,7, а второго рода (FAR) – 4,3. Тогда вероятность этих ошибок можно найти по следующей формуле:

$$P_{FRR/FAR} = \frac{m_{FRR/FAR} \cdot 100}{n}, \quad (19)$$

где $m_{FRR/FAR}$ – количество ошибок аутентификации; n – общее количество попыток аутентификации.

По формуле (19) были найдены вероятности ошибок:

$$P_{FRR} = \frac{m_{FRR} \cdot 100}{n} = \frac{10{,}7 \cdot 100}{50} = 21{,}4; \quad P_{FAR} = \frac{m_{FAR} \cdot 100}{n} = \frac{4{,}3 \cdot 100}{100} = 4{,}3.$$

Данные результаты являются предварительными, так как были получены без использования в программе корректирующего коэффициента (Стьюдента), который уменьшает ошибку первого рода. По полученным данным можно сделать вывод, что значение вероятности ошибки первого рода очень велико и действительно есть необходимость использовать коэффициент Стьюдента. Значение ошибки второго рода является вполне нормальным для разработанного тестового варианта, так как не были использованы никакие дополнительные методы обработки ключевых параметров пользователя.

Исследование выбора оптимального значения коэффициента Стьюдента, отвечающего за уменьшение ошибки первого рода

Для исследования выбора оптимального значения коэффициента Стьюдента, отвечающего за уменьшение ошибки первого рода, в системе аутентификации было зарегистрировано три пользователя. Для регистрации каждый из пользователей ввел 5 раз свой логин и пароль. На следующем этапе каждый из пользователей ввел 100 раз свои входные данные (для определения ошибки первого рода) и по 50 раз данные других двух пользователей (для определения ошибки второго рода).

Выполнение указанных этапов производилось для каждого значения коэффициента Стьюдента, взятого из табл. 2 и возведенного в квадрат с учетом того, что количество образцов для создания эталонной характеристики клавиатурного почерка пользователя равняется пяти. Полученные результаты представлены в виде графика на рис. 6.

Из графика видно, что увеличение значения коэффициента Стьюдента уменьшает ошибку первого рода, так как область наборов легального пользователя увеличивается на значение коэффициента, и те наборы параметров клавиатурного почерка, которые не соответствовали эталонной характеристике пользователя, теперь оказались в области наборов легального пользователя. Но в то же время в расширяющую область наборов легального пользователя начинают попадать чужие наборы, вследствие чего вероятность ошибки второго рода начинает расти. Оптимальным значением коэффициента Стьюдента, с точки зрения равновесия ошибок первого и второго рода, является точка пересечения линий тренда графиков этих ошибок. По графику этой точке соответствует значение 7,7284, т.е. оптимальным является значение коэффициента Стьюдента, равное 2,78. При этом значении коэффициента количество ошибок первого рода (FRR) равно 8,671, а второго рода (FAR) – 7,037.

По формуле (19) были найдены вероятности ошибок:

$$P_{FRR} = \frac{m_{FRR} \cdot 100}{n} = \frac{8{,}681 \cdot 100}{100} \approx 8; \quad P_{FAR} = \frac{m_{FAR} \cdot 100}{n} = \frac{7{,}037 \cdot 100}{100} \approx 7.$$

При сравнении с результатами, полученными ранее, видно, что использование коэффициента Стьюдента позволило значительно снизить вероятность ошибки первого рода при небольшом увеличении вероятности ошибки второго рода.

Заключение

Таким образом, реализована технология усиленной аутентификации пользователей. Приведенный пример тестирования данной технологии подтверждает корректность работы и ее применимость для решения поставленных задач. Кроме того, в работе исследовалась применимость различных подходов к анализу ключевых параметров клавиатурного почерка пользователя, были рассмотрены подходы к выявлению ключевых параметров. Было проведено исследование выбора оптимального значения коэффициента Стьюдента, при котором ошибки первого и второго рода минимальны.

Литература

1. Сарбуков А.Е. Аутентификация в компьютерных системах / А.Е. Сарбуков, А.А. Грушо // Системы безопасности. – 2003. – № 5(53). – С. 118–122.

2. Задорожный В. Обзор биометрических технологий // Защита информации. Конфидент. – 2003. – № 5. – С. 26–29.

3. Фор А. Восприятие и распознавание образов. – М.: Машиностроение, 1989. – 103 с.

4. Иванов А.И. Биометрическая идентификация личности по динамике подсознательных движений. – Пенза: Изд-во Пенз. гос. ун-та, 2000. – 188 с.

5. Горелик А.Л. Методы распознавания / А.Л. Горелик, В.А. Скрипкин. – М.: Высшая школа, 1984. – 80 с.

6. Широчин В.П. Динамическая аутентификация на основе анализа клавиатурного почерка / В.П. Широчин, А.В. Кулик, В.В. Марченко // Вестник Национального технического университета Украины «Информатика, управление и вычислительная техника». – 1999. – № 32. – С. 3–16.

7. Анатомия человека / Борзяк Э.И., Бочаров В.Я., Волкова Л.И. и др.; под ред. М.Л. Сапина. – М.: Медицина, 1987. – Т. 1. – 109 с.

8. Нильсон Н. Обучающиеся машины / пер. с англ. – М.: Мир, 1967. – 180 с.

9. Брюхомицкий Ю.А. Метод обучения нейросетевых биометрических систем на основе копирования областей / Ю.А. Брюхомицкий, М.Н. Казарин // Перспективные информационные технологии и интеллектуальные системы (Электронный журнал). – 2003. – № 3 (15). – С. 17–23.

10. Расторгуев С.П. Программные методы защиты информации в компьютерах и сетях. – М.: Изд-во агентства «Яхтсмен», 1993. – 120 с.

11. Корнеева А.П. Машинопись и основы современного делопроизводства / А.П. Корнеева, А.М. Амелина, А.П. Загребельный. – М.: Просвещение, 1979. – 212 с.

Ходашинский Илья Александрович
Д-р техн. наук, профессор каф. комплексной информационной безопасности электронно-вычислительных систем (КИБЭВС) ТУСУРа
Тел.: (382-2) 41-34-26
Эл. почта: his@keva.tusur.ru

Савчук Михаил Владиславович
Ст. аудитор Департамента внутреннего аудита по ИТ ООО «ЕвразХолдинг»
Эл. почта: mikhail.v.savchuk@gmail.com

Горбунов Иван Викторович
Аспирант каф. КИБЭВС ТУСУРа
Тел.: +7-923-405-60-61
Эл. почта: noby.Ardor@gmail.com

Мещеряков Роман Валерьевич
Канд. техн. наук, доцент каф. КИБЭВС ТУСУРа
Тел.: (382-2) 41-34-26
Эл. почта: mrv@keva.tusur.ru

Hodashinsky I.A., Savchuk M.V., Gorbunov I.V., Meshcheryakov R.V.
Strong authentication technology of the users of information processes

The aim of the paper is to show the development of technology which implements the strong authentication system, based on dynamic authentication on the user's keyboard handwriting. Password authentication is chosen as the object of additional security control. In the process of the research we analyzed the existing methods of classifying the users by means of biometric authentication method, based on keyboard handwriting. We considered the approach of identification the key parameters of keyboard handwriting of a user. The gained knowledge was used to implement the library. We also conducted the research to determine the optimal values of Student's coefficient, which is responsible for the reduction of Type I error.
Keywords: authentication, user, biometric techniques, parameter learning, covariance matrix.