ЗАЩИТА ИНФОРМАЦИИ

УДК 681.324

ЯДЕРНАЯ МОДЕЛЬ КЛАССИФИКАТОРА БИОМЕТРИЧЕСКИХ ПАРАМЕТРОВ ЛИЧНОСТИ*

Ю.А. Брюхомицкий

Рассматривается метод построения и результаты экспериментального исследования вероятностного классификатора, предназначенного для биометрических систем контроля доступа по рукописному и клавиатурному почеркам. Принцип работы классификатора состоит в построении решающего правила доступа на основе оценки суммарной функции плотности распределения биометрических параметров «своего» пользователя.

Ключевые слова: *биометрические параметры личности, вероятностный классификатор, ядерная модель, суммарная плотность распределения, аппроксимация областей.*

Key words: *biometric identity settings, probabilistic classifier, a kernel model, the cumulated density distribution, trendline areas.*

Динамические системы биометрической идентификации (ДСБИ) личности по рукописному и клавиатурному почеркам основаны на анализе индивидуальных особенностей динамики подсознательных движений, которые в общем случае могут быть представлены N-мерными векторами биометрических признаков $\mathbf{V} \in R^N$. В математической постановке такие системы решают задачу классификации векторов $\mathbf{V}$ на классы: «свой» – $\mathbf{V}^+$ и «чужой» – $\mathbf{V}^-$. Решающее правило классификации основано на сравнении векторов $\mathbf{V}$ с эталонными векторами $\mathbf{V}_Э^k$, $k = \overline{1,M}$, созданными для M зарегистрированных в системе («своих») пользователей.

В большинстве случаев хорошим приближением для аппроксимации распределения векторов $\mathbf{V}^+$ является гауссово распределение. Такая аппроксимация эффективно используется в параметрических методах классификации [2, с. 6–13]. Однако в ДСБИ часто наблюдаются флуктуации контролируемых параметров, обусловленные суточными биоритмами, психофизическим состоянием и другими факторами [4, с. 3]. Это приводит к тому, что с течением времени распределение векторов $\mathbf{V}^+$, возможно, даже оставаясь в рамках нормального, изменяет свои числовые характеристики. Хорошим приближением для аппроксимации распределения биометрических данных в этом случае является смешанное гауссово распределение с несколькими центрами. Ситуацию для одномерного распределения иллюстрирует рис. 1.

В примере на рис. 1 для формирования эталона $\mathbf{V}_Э$ на этапе обучения параметрическим методом будет использовано единственное распределение с плотностью $p_{t1}(v)$ и математическим ожиданием $m_{t1}(v)$, поскольку именно оно было присуще данному пользователю в момент обучения t_1. Впоследствии при флуктуациях биометрического параметра v во времени данный пользователь будет получать отказ в допуске чаще нормы, определяемой ошибкой первого рода данной ДСБИ.

* Работа выполнена при поддержке гранта РФФИ № 08-07-00117-а.

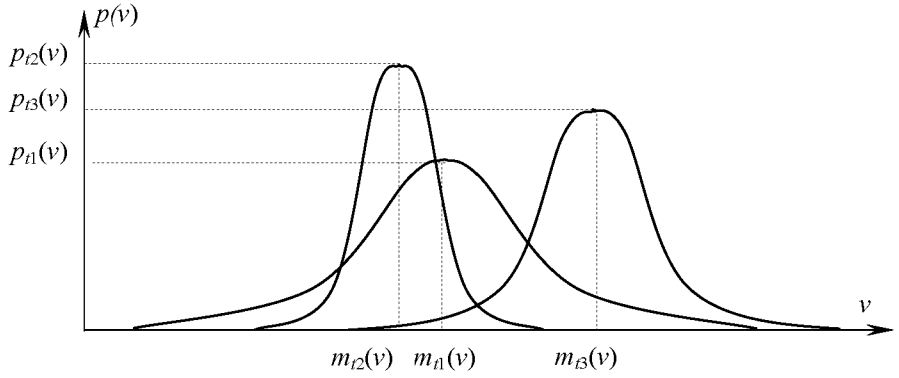

Рис. 1. Флуктуации контролируемого биометрического параметра v во времени

Для компенсации флуктуаций характеристик распределения можно пытаться отслеживать изменения биометрических параметров во времени, создавая для каждого «своего» пользователя серию эталонов и используя в текущем сеансе аутентификации на каком-то разумном основании подходящий эталон. Однако очевидно, что такой подход значительно усложнит процессы регистрации и аутентификации пользователей ДСБИ.

В работе [1, с. 147–154] предложен иной подход, который позволяет изначально рассматривать распределения векторов $\mathbf{V}^+$ как результат нестационарного случайного процесса воспроизведения обучающей выборки $\mathbf{\Psi}^+ = \{\mathbf{V}_i^+\}$, $i = \overline{1, L}$. Суть подхода состоит в аппроксимации области распределения биометрических параметров произвольной конфигурации набором ядерных функций, моделирующих отдельные образцы обучающей выборки.

Область распределения биометрических параметров «своего» пользователя задается множеством учебных данных $\mathbf{\Psi}^+ = \{\mathbf{V}_i^+\}, i = \overline{1, L}$:

$$\mathbf{V}_i^+ = (v_1^+, v_2^+, ..., v_N^+), \quad j = \overline{1, N}, \quad i = \overline{1, L}.$$

Для оценки плотности распределения каждого учебного вектора $\mathbf{V}_i^+$ используется ядерная функция, задающая конфигурацию каждой подобласти, в виде упрощенной формы функции Гаусса с центром в точке $\mathbf{V}_i^+$:

$$p(\mathbf{V}_i^+) = \varphi_i(\mathbf{V}) = \exp\left(-\frac{\left\|\mathbf{V} - \mathbf{V}_i^+\right\|^2}{2\mu^2}\right), \quad i = \overline{1, L}, \tag{1}$$

где μ – параметр, задающий ширину ядерной функции $\varphi(\mathbf{V})$.

Оценка функции плотности распределения для всего множества $\mathbf{\Psi}^+ = \{\mathbf{V}_i^+\}, i = \overline{1, L}$ получается суммированием функций (1):

$$\Phi(\mathbf{V}) = \sum_{i=1}^{L} \varphi(\mathbf{V}) = \sum_{i=1}^{L} \exp\left(-\frac{\left\|\mathbf{V} - \mathbf{V}_i^+\right\|^2}{2\mu^2}\right). \tag{2}$$

Входные данные классификатора нормализуются к единичной длине:

$$(v_j)' = v_j \Big/ \sqrt{\sum_{j=1}^{N} v_j^2}, \quad (v_{ij}^+)' = v_{ij} \Big/ \sqrt{\sum_{j=1}^{N} (v_{ij}^+)^2}.$$

В результате несложных преобразований оценка функции плотности распределения для всего множества $\mathbf{\Psi}^+ = \{\mathbf{V}_i^+\}, i = \overline{1, L}$ при нормализованных входных данных в покомпонентном представлении принимает вид

$$\Phi(\mathbf{V}) = \sum_{i-1}^{L} \exp\left(\frac{1}{\mu^2} \sum_{j=1}^{N} (v_j)' \cdot (v_{ij}^+)' - 1\right). \tag{3}$$

Задача построения решающего правила для классификации входных векторов **V** на «свой» ($\mathbf{V}^+$) и «чужой» ($\mathbf{V}^-$) решается путем построения разделяющей гиперповерхности G. В условиях наличия только обучающего множества $\mathbf{\Psi}^+ = \{\mathbf{V}_i^+\}, i = \overline{1, L}$ гиперповерхность G может быть задана в виде порогового значения суммарной плотности вероятности $\Phi_{\text{п}}(\mathbf{V})$, являющегося минимально допустимым для «своего» пользователя.

Значение $\Phi_{\text{п}}(\mathbf{V})$ образуется следующим способом. Неизвестный образец в выражении (3) заменяется на один из учебных (k-образец, $k = \overline{1, L}$), после чего вычисляются евклидовы расстояния от k-образца до остальных (L-1) учебных образцов. С помощью ядерной функции эти расстояния преобразуются в конечном итоге в значение суммарной плотности вероятности $\Phi(\mathbf{V}_k)$. Процедура повторяется последовательно для всех L учебных векторов $\mathbf{V}_k^+, k = \overline{1, L}$, $i \neq k$. Выражение для этих вычислений имеет вид:

$$\Phi(\mathbf{V}_k) = \sum_{i-1}^{L-1} \exp\left(\frac{1}{\mu^2} \sum_{j=1}^{N} (v_k)' \cdot (v_{ij}^+)' - 1\right), \quad k = \overline{1, L}, i = \overline{1, L-1}, i \neq k. \tag{4}$$

Далее полученные значения плотности вероятности $\Phi(\mathbf{V}_k)$, $k = \overline{1, L}$ ранжируются по величине, и выбирается минимальное из них, которое трактуется как пороговое значение суммарной плотности распределения $\Phi_{\text{п}}(\mathbf{V})$, являющееся минимально допустимым для «своего» пользователя:

$$\Phi_{\text{п}}(\mathbf{V}) = \min\,[\Phi(\mathbf{V}_k),\ k = \overline{1, L}, i = \overline{1, L-1}, i \neq k.]$$

Величина $\Phi_{\text{п}}(\mathbf{V})$ определяет искомую гиперповерхность G, отделяющую область распределения биометрических параметров «свой» от остального пространства.

Решающее правило для классификации векторов $\mathbf{V}^+$ и $\mathbf{V}^-$ имеет вид

$$\mathbf{V} \in \begin{cases} \mathbf{V}^+, \text{ если } \operatorname{sign}[\Phi(\mathbf{V}] - \Phi_{\text{п}}(\mathbf{V})] = 1; \\ \mathbf{V}^-, \text{ если } \operatorname{sign}[\Phi(\mathbf{V}] - \Phi_{\text{п}}(\mathbf{V})] = 0. \end{cases} \tag{5}$$

Главной проблемой построения эффективных ядерных моделей является оптимальный подбор ширины μ ядерной функции. Если параметр μ будет слишком мал, то в результирующей оценке плотности будет много пиков, отслеживающих положение отдельных образцов данных. Наоборот, если параметр μ будет чрезмерно большим, то область расположения данных будет большой и сильно сглаженной. При этом информация о структуре данных полностью теряется. Задача состоит в отыскании оптимальной величины μ, при которой результирующая оценка представляла бы собой хорошую реконструкцию истинной плотности распределения данных.

На рис. 2 показаны графики двумерной плотности распределения $\Phi_{\text{п}}(\mathbf{V})$, формирующей разделяющую поверхность G при различных значениях ширины μ ядерной функции.

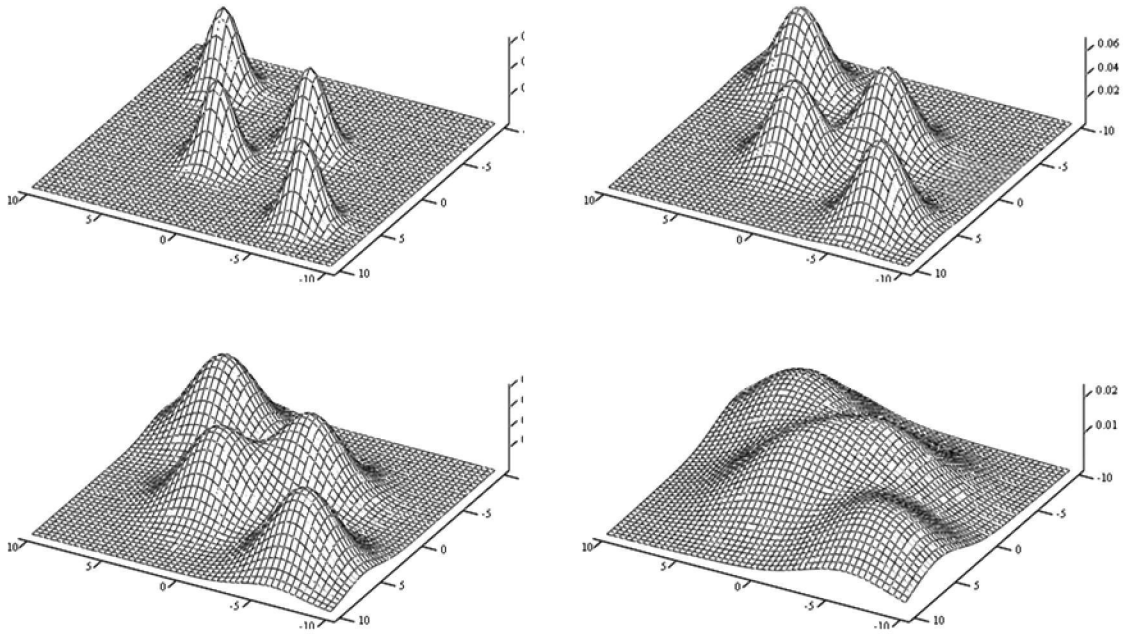

Рис. 2. Графики двумерной плотности распределения $\Phi_{\text{п}}(\mathbf{V})$, при различных значениях ширины μ ядерной функции

Ширина μ ядерной функции непосредственно влияет и на классифицирующие способности ДСБИ, поскольку выступает аргументом функций ошибок первого рода $P_1(\mu)$ (вероятность не допуска «своего») и второго рода $P_2(\mu)$ (вероятность пропуска «чужого»). В ДСБИ зависимости $P_1(\mu)$ и $P_2(\mu)$ имеют взаимно обратный характер, а точка их пересечения дает величину равновероятной ошибки $P_1(\mu) = P_2(\mu)$, которая обычно и выступает в качестве характеристики точности ДСБИ. Поэтому точка $P_1(\mu) = P_2(\mu)$ задает оптимальное значение ширины μ ядерной функции (рис. 3).

Для определения оптимального значения μ и эффективности метода в целом была разработана его программная модель, которая позволяла контролировать уровень ошибок классификатора в зависимости от величины μ. Входными данными модели выступали биометрические характеристики реальных пользователей, полученные ранее в ДСБИ BioKey в ходе выполнения лабораторных работ студентами Таганрогского технологического института Южного федерального университета. Массив входных данных содержал 1000 векторов $\mathbf{V}_{ij}$ размерностью $N = 32$, которые принадлежали 50 пользователям ($M = 50$), причем каждый пользователь был представлен своими 20 образцами ($L = 20$).

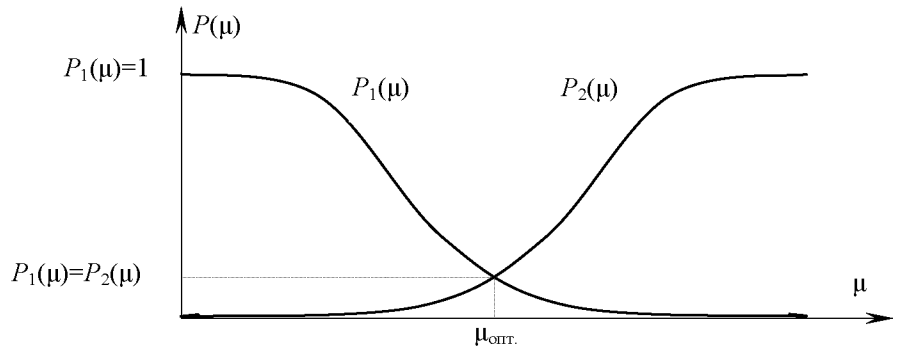

Рис. 3. График зависимостей $P_1(\mu)$ и $P_2(\mu)$

Достоверное вычисление ошибки P_1 на имеющемся ограниченном массиве данных не представлялось возможным, поскольку для формирования интегральной области «свой» использовались все 20 образцов данного пользователя и предъявление любого образца этого пользователя заведомо предопределяло его попадание в область, т.е. всегда $P_1 = 0$.

Для вычисления ошибки P_2 в качестве «чужих» выступали все образцы за исключением 20 образцов, использованных для формирования интегральной области «свой». Погрешность вычислений при этом составляла 0,1 %.

Вначале расчет ошибок P_2 был проведен для всего контингента пользователей при фиксированном значении $\mu = 0{,}15$. Результаты в виде гистограммы приведены на рис. 4 (для отсутствующих в гистограмме отсчетов $P_2 = 0$). Среднее значение ошибки $(P_2)_{ср.} = 0{,}175$, причем для разных пользователей ошибка P_2 имеет очень широкий разброс – от 0,001 до 1,0. Это является следствием того, что часть пользователей не обладала выраженным клавиатурным почерком и распределение их биометрических параметров сопровождалось большой дисперсией, а значит, и большой ошибкой P_2.

На рис. 5 в качестве примера показаны двухмерные проекции (по первым двум компонентам v_1 и v_2) областей распределения биометрических параметров для трех пользователей с номерами 1, 10 и 15. Пример демонстрирует существенную разницу в характеристиках распределения биометрических параметров между пользователем 10 (большая дисперсия) и пользователями 1,15 (небольшая дисперсия).

Общепринятым [3] считается, что привлечение для регистрации в клавиатурных ДСБИ пользователей с плохим клавиатурным почерком недопустимо. Поэтому были вычислены дисперсии распределений биометрических параметров для каждого пользователя и для упрощения проведено их усреднение по всем компонентам векторов $\mathbf{V}_{ij}$. Это позволило из 50 пользователей отобрать 29, у которых средняя дисперсия не превышала значения 0,5. Сравнение номеров отобранных пользователей с номерами тех, у которых ошибка второго рода была менее 10 % (рис. 4), показало, что они практически полностью совпадают. В результате по обоим показателям в группу отобранных пользователей вошло 32 человека.

Гистограмма распределения ошибки P_2 для отобранной группы пользователей при фиксированном значении $\mu = 0{,}15$ приведена на рис. 6 (для отсутствующих отсчетов $P_2 = 0$). Среднее значение $(P_2)_{ср.} = 0{,}0182$.

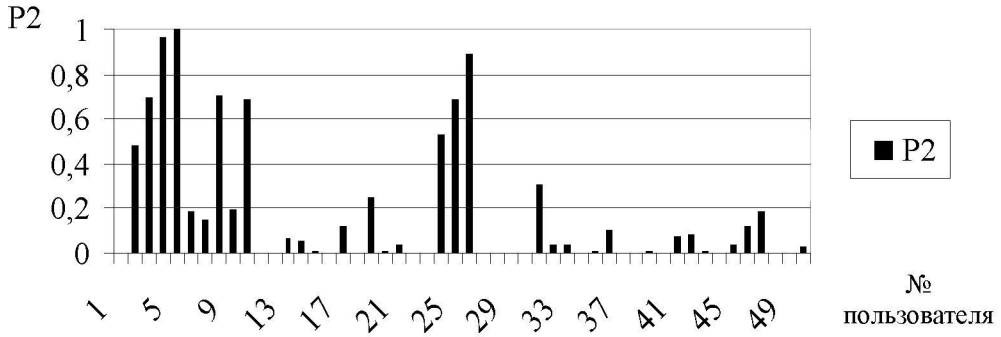

Рис. 4. Гистограмма распределения ошибок P_1 и P_2 у всех пользователей

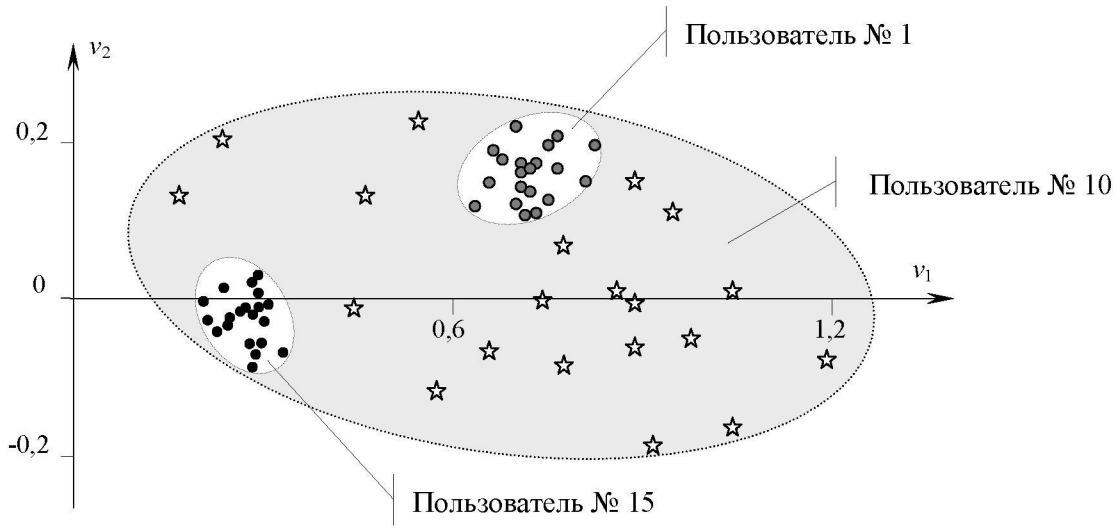

Рис. 5. Двухмерные проекции областей распределения биометрических параметров для трех пользователей

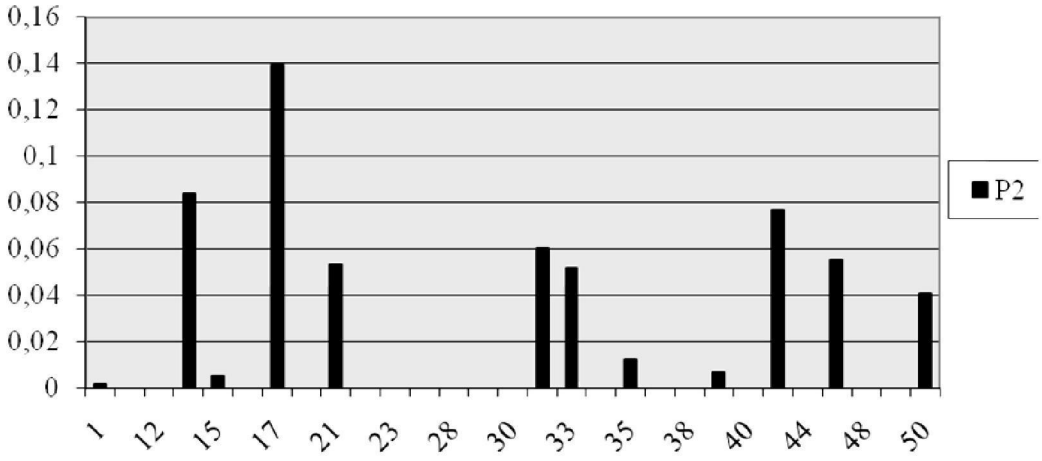

Рис. 6. Гистограмма распределения ошибок P_1 и P_2 для отобранной группы пользователей

Таким образом, среднее значение ошибки $(P_2)_{ср}$ для отобранной группы пользователей по сравнению с полным контингентом уменьшилось в 10 раз, а диапазон разброса значений P_2 сократился с (0÷1) до (0÷0,14).

На основе правила (5) была вычислена также зависимость P_2 (μ) для отобранной группы и всего контингента пользователей (рис. 7).

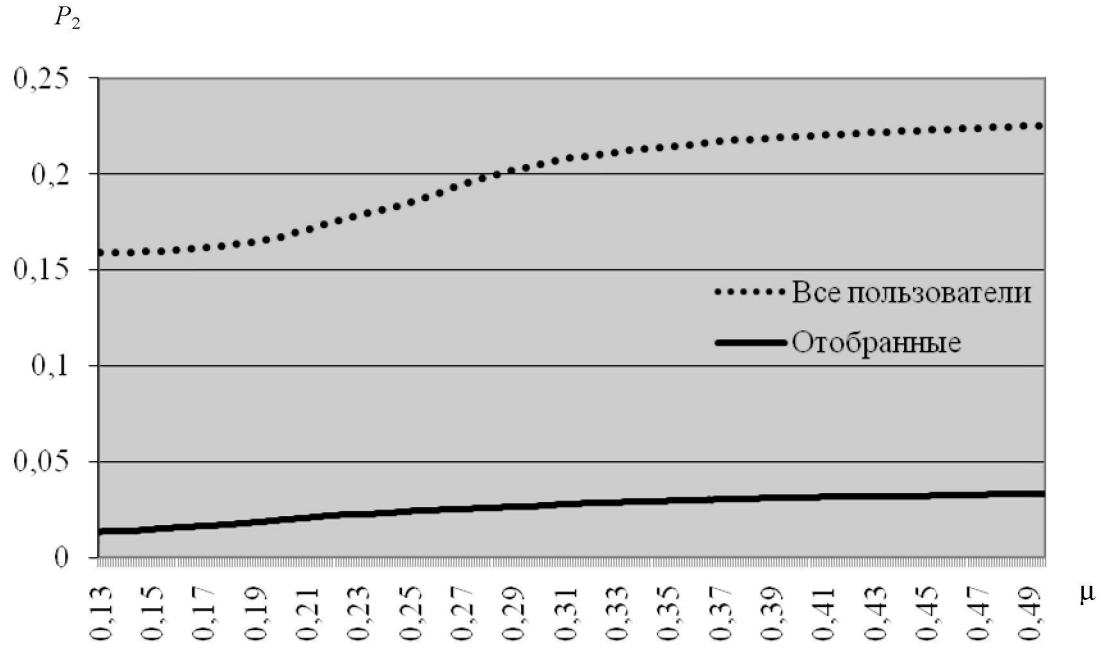

Рис. 7. График зависимости $P_2(\mu)$ для всего контингента пользователей и отобранной группы

Используемые для экспериментов реальные биометрические данные, к сожалению, не позволили непосредственно построить зависимости для ошибки P_1. Для этого нужно было бы иметь около 1000 образцов одного «своего» пользователя. Однако полученные данные для ошибок $P_2(\mu)$ в совокупности с косвенными расчетами $P_1(\mu)$ позволяют с уверенностью предположить, что точка равновероятной ошибки $P_1(\mu) = P_2(\mu)$ при изменении значений $P_1(\mu)$, $P_2(\mu)$ в диапазоне 0,01÷0,05 находится внутри диапазона 0,1÷1,0 значений ширины μ ядерной функции. Более точный подбор параметра μ целесообразно выполнить на этапе реализации ДСБИ, исходя из конкретных требований для решаемой прикладной задачи.

Проведенные эксперименты прямо или косвенно подтвердили работоспособность и эффективность предложенного метода классификации биометрических параметров на основе ядерной модели. Метод позволяет в условиях использования нестабильных эталонов на ограниченных наборах учебных данных строить простые ДСБИ с точностью классификации, соответствующей достигнутому уровню для динамической биометрии.

Библиографический список

1. ***Брюхомицкий, Ю. А.*** Классификация нестационарных вероятностных биометрических параметров личности / Ю. А. Брюхомицкий // Известия ЮФУ. – 2008. – № 8. – С. 147–154. – (Сер. Технические науки).

2. ***Брюхомицкий, Ю. А.*** Параметрическое обучение биометрических систем контроля доступа / Ю. А. Брюхомицкий, М. Н. Казарин // Вестник компьютерных и информационных технологий. – М. : Изд-во «Машиностроение», 2006. – № 2 (20). – С. 6–13.

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

4. ***Широчин, В. П.*** Динамическая аутентификация на основе анализа клавиатурного почерка / В. П. Широчин, А. В. Кулик, В. В. Марченко. – Режим доступа: http://www.masters.donntu.edu.ua, свободный. – Заглавие с экрана. – Яз. рус.

УДК 004.056.55

ВЫБОР СРЕДСТВ ЗАЩИТЫ ДЛЯ ПОСТРОЕНИЯ СИСТЕМЫ ЗАЩИЩЕННОГО ИНФОРМАЦИОННОГО ОБМЕНА

Р.Ю. Волик

Рассматриваются вопросы выбора средств защиты информации при построении виртуальных частных сетей (VPN) с целью организации систем защищенного информационного взаимодействия между распределенными сетями и вопросы обеспечения безопасности при передаче информации с использованием сетей общего доступа.

Ключевые слова: *виртуальная частная сеть, межсетевые экраны.*

Key words: *virtual private network, firewall.*

В настоящее время широкое распространение получает электронный документооборот как один из видов электронного взаимодействия между сторонами в виде информационного обмена.

Информационные системы становятся сегодня одним из главных инструментов управления бизнесом, важнейшим средством производства современного предприятия.

Одновременно с внедрением средств обеспечения электронного документооборота и, как следствие, повышением производительности предприятия возникает вероятность реализации угроз информационной безопасности, которым могут быть подвержены информационные ресурсы предприятия. Поэтому вопрос обеспечения безопасности информации при осуществлении документооборота встает наиболее остро.

Защита информации в процессе ее передачи по открытым каналам основана на построении защищенных виртуальных частных каналов связи, называемых туннелями VPN. Туннель VPN обладает свойствами защищенной выделенной линии, причем эта защищенная выделенная линия развертывается в рамках общедоступной сети, например интернет-сети. Туннель VPN представляет собой соединение, проведенное через открытую сеть, по которому передаются криптографически защищенные пакеты виртуальной сети.

VPN создает виртуальные защищенные «туннели» в открытых TCP/IP сетях типа интернет-сети. Наличие «туннеля» позволяет решить двойную задачу:

- исключить перехват проходящей по «туннелю» информации;
- исключить подключение незарегистрированного компьютера к VPN, изменение информации и любые сетевые атаки.

VPN-продукты позволяют организовывать защищенные туннели между офисами компании. При этом абсолютно неважно, через какого провайдера конкретная рабочая станция подключится к защищенным ресурсам предприятия. По оценке CNews, относительная доля VPN на российском рынке ИБ уже достигает 30 % [3, 4].

VPN технология обеспечивает:

- защиту (конфиденциальность, подлинность и целостность) передаваемой по сетям информации;
- контроль доступа в защищаемый сектор сети;
- безопасный доступ пользователей VPN к ресурсам сетей общего пользования;
- централизованное управление политикой корпоративной сетевой безопасности.