ЕВРАЗИЙСКОЕ НАУЧНОЕ ОБЪЕДИНЕНИЕ
EURASIAN SCIENTIFIC ASSOCIATION
www.esa-conference.ru

Текстонезависимая биометрическая идентификация личности на основе иммунной модели

Брюхомицкий Юрий Анатольевич, кандидат технических наук, доцент
Южный Федеральный университет (г. Таганрог)

Предлагается иммунологический подход к решению задачи распознавания сигналов динамической биометрии, базирующийся на принципах массово-параллельной децентрализованной обработки данных, использующихся в искусственных иммунных системах. Особенностью подхода является представление сигналов динамической биометрии последовательностями информационных единиц определенного формата, с последующей обработкой в темпе их поступления на основе иммунологической модели клональной селекции с положительным отбором. Предлагаемый подход позволяет обобщить существенно различные методы идентификации личности по динамическим биометрическим параметрам разной модальности.

Ключевые слова: *искусственные иммунные системы; системы динамической биометрической идентификации личности; иммунологическая модель клональной селекции.*

Введение. Динамические системы биометрической идентификации личности (динамическая биометрия) основаны на анализе индивидуальных особенностей хорошо заученных подсознательных движений человека. Практическое применение в настоящее время получили системы анализа голоса [1], рукописи [2] и клавиатурного почерка [3, 4]. Динамическая биометрия (ДБ) используется преимущественно как средство идентификации личности в информационных и телекоммуникационных системах.

Идентификация личности с помощью систем ДБ может осуществляться по фиксированной (парольной) фразе. Такие системы относительно просты в реализации, но весьма уязвимы для атак воспроизведения. Существуют также системы ДБ, которые обеспечивают идентификацию личности при воспроизведении любого текста на любом языке (текстонезависимая ДБ). Они существенно сложнее в реализации и требуют больше времени для проведения процедуры идентификации, но позволяет решать существенно более широкий круг задач информационной безопасности:

- непрерывная скрытная верификация личности пользователей информационных систем;
- скрытное выявление инсайдеров;
- скрытное выявление психофизических отклонений личности от нормы;
- аудит безопасности информационных систем на основе интерактивного взаимодействия с пользователями (аналог детектора лжи) и др.

В системах текстонезависимой ДБ эталоны личности строятся на основе достаточно больших образцов текста соответствующей модальности. При этом возникает ряд принципиальных проблем, связанных с необходимостью использования биометрических эталонов большой размерности, трудностью их формирования, анализа и сопоставления с образцами.

Постановка задачи. Общей особенностью текстонезависимой ДБ является представление исходных данных сигналами (функциями времени) структура которых содержит все необходимые для идентификации индивидуальные особенности личности. Для решения задачи распознавания таких сигналов предложено множество методов, которые сводятся к переводу их в какое-либо статическое представление, после чего задача распознавания решается уже в формате их статического представления.

В данной работе предлагается подход к решению задачи распознавания сигналов ДБ, базирующийся на принципах массово-параллельной децентрализованной обработки данных, используемых в искусственных иммунных системах (ИИС) [5]. Особенностью подхода является представление сигналов ДБ в виде последовательности информационных иммунологических единиц определенного формата, с их последующей децентрализованной обработкой в темпе поступления на основе иммунологической модели клональной селекции [6].

Решение поставленной задачи. Воспроизведение произвольного текста средствами ДБ любой модальности реализуется совокупностью заученных подсознательных движений, которые преобразуются в электрические сигналы (функции времени) В общем случае эти сигналы являются многомерными: $\boldsymbol{x}(t) = x_1(t), x_2(t), \dots, x_n(t)$.

На этапе предварительной обработки сигнал $\boldsymbol{x}(t)$ оцифровывается $\boldsymbol{x}(t) \to \boldsymbol{x}(t_i) = \boldsymbol{x}(i), i = 1, 2, \dots$, масштабируется, из него исключаются длительные паузы, не обусловленные индивидуальными особенностями воспроизведения текста. В голосовой биометрии исключаются также неинформативные (с точки зрения распознавания голоса) фонемы шипящих звуков.

Отсчеты сигнала $\boldsymbol{x}(i), i = 1, 2, \dots$ можно рассматривать как точки метрического пространства E^n, представленные векторами признаков $\boldsymbol{x_i} = x_{1i}, x_{2i}, \dots, x_{ni}$, а сам сигнал $\boldsymbol{x}(i)$, – как последовательность $\mathbf{X}_i = \boldsymbol{x}_1, \boldsymbol{x}_2, \dots$ элементов, представленных векторами признаков: $\boldsymbol{x_i}$. В математическом смысле последовательность $\mathbf{X}_i$ «пробегает» конечное множество Ψ_x векторов признаков $\boldsymbol{x_i}$ биометрии данной личности.

Исследования в области ДБ [4, 7], показывают, что индивидуальные особенности личности в наибольшей степени проявляются при воспроизведении не одиночных символов текста, а синтаксически связанных фрагментов текста. Использование этого феномена при анализе позволяет строить системы биометрической идентификации личности с существенно более высокими характеристиками по точности.

Следуя указанному подходу последовательность $\mathbf{X}_i$ расчленяется на фрагменты одинакового размера по r отсчетов $\boldsymbol{x_i}$ в каждом фрагменте. Результатом будет новая последовательность $\mathbf{Y}_j = \boldsymbol{y}_1, \boldsymbol{y}_2, \dots, j = 1, 2, \dots$, каждый элемент $\boldsymbol{y}_j$ которой содержит r n-мерных векторов $\boldsymbol{x_i}$ исходной последовательности $\mathbf{X}_i$:

ЕВРАЗИЙСКОЕ НАУЧНОЕ ОБЪЕДИНЕНИЕ
EURASIAN SCIENTIFIC ASSOCIATION
www.esa-conference.ru

$$\boldsymbol{y}_j = \boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_r, i = 1, 2, \dots, r, j = 1, 2, \dots .$$

Совокупность векторов $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_r$ каждого фрагмента $\boldsymbol{y}_j$ можно представить как один s-мерный вектор $\boldsymbol{y}_j$, содержащий $s = n \times r$ компонент:

$$\boldsymbol{y}_j = \begin{bmatrix} y_{11} & y_{12} \dots & y_{1r} \\ y_{21} & y_{22} \dots & y_{2r} \\ \dots & \dots \dots & \dots \\ y_{n1} & y_{n2} \dots & y_{nr} \end{bmatrix}.$$

В итоге, образ ДБ личности будет представлен последовательностью $\mathbf{Y}_j$ s-мерных векторов признаков $\boldsymbol{y}_j$ в пространстве признаков E^s.

Последовательность $\mathbf{Y}_j$, ограниченная N_y элементами

$\bar{\mathbf{Y}}_j = \boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_{N_y}, j = 1, 2, \dots N_y,$

можно трактовать как биометрический эталон данной личности. При распознавании данных ДБ в режиме верификации последовательность $\bar{\mathbf{Y}}_j$, выступает в качестве эталона «своего». При этом каждая информационная единица $\boldsymbol{y}_j$ последовательности $\bar{\mathbf{Y}}_j$ отражает представление нескольких последовательно воспроизводимых (голосом, клавиатурой или ГП) символов текста. Общий характер распределения биометрических данных определенной личности представлен множеством областей, каждая из которых отражает распределение одного определенного фрагмента $\boldsymbol{y}_j$ контекстно-связанных символов. Число областей соответствует числу типов фрагментов $\boldsymbol{y}_j$, встречающихся в эталонной последовательности $\mathbf{Y}_j$.

В фазе обучения ИИС создается начальная популяция детекторов в метрике векторов $\boldsymbol{y}_j$. Затем по принципу положительного отбора выявляются детекторы начальной популяции, которые в пространстве E^s наиболее близки к областям распределения векторов $\boldsymbol{y}_j$. Близость векторов в пространстве E^s моделирует свойство аффинности клеток иммунной системы. Выявленные детекторы на основе итерационной процедуры подвергаются клонированию, гипермутации и отбора. Останов процедуры обучения осуществляется по определенным признакам, свидетельствующим о достижении достаточной степени покрытия детекторами области распределения биометрических признаков в рабочем пространстве E^s.

На этапе распознавания контролируемые биометрические данные личности $\mathbf{Y}_j$ сравниваются с детекторами по принципу близости. Соотношение числа срабатывающих детекторов к их общему числу дает оценку вероятности для принятия системой решения «свой–чужой».

Алгоритм обучения ИИС.

1. Создание последовательности $\bar{\mathbf{Y}}_j, j = 1, 2, \dots N_y$, представляющей биометрический эталон «своего».

2. Создание путем случайной генерации (с равномерным законом распределения) начальной популяции детекторов $\mathbf{D}_k^\lambda = \boldsymbol{d}_1, \boldsymbol{d}_1, \dots, \boldsymbol{d}_{N_d}, \lambda = 0, k = 1, 2, \dots, N_d$ представленных векторами в формате векторов $\boldsymbol{y}_j$.

3. Для каждой пары $\boldsymbol{d}_k \in \mathbf{D}_k^\lambda$ и $\boldsymbol{y}_j \in \bar{\mathbf{Y}}_j$ вычисляется степень их взаимной аффинности. В качестве меры аффинности a_{kj} используется Евклидово расстояние между векторами $\boldsymbol{d}_k$ и $\boldsymbol{y}_j$:

$$a_{kj}(\boldsymbol{d}_k, \boldsymbol{y}_j) = \sqrt{\sum_{p=1}^{N} (d_{kp} - y_{jp})^2}, p = 1, 2, \dots, N, N = N_d \cdot N_y.$$

Результатом вычислений будет матрица взаимной аффинности $\mathbf{A}$, содержащая $N = N_d \cdot N_y$ элементов a_{kj}:

$$\mathbf{A}\|a_{kj}(\boldsymbol{d}_k, \boldsymbol{y}_j)\| = \left\| \begin{matrix} a_{11} & a_{12} & \dots & a_{1N_y} \\ a_{21} & a_{22} & \dots & a_{2N_y} \\ \dots & \dots & \dots & \dots \\ a_{N_d1} & a_{N_d2} & \dots & a_{N_dN_y} \end{matrix} \right\|, k = 1, 2, \dots, N_d \; j = 1, 2, \dots, N_y .$$

4. Из каждого столбца матрицы $\mathbf{A}$ отбирается l из N_d детекторов $\boldsymbol{d}_k$, соответствующих наибольшей взаимной аффинности $a_{kj}(\boldsymbol{d}_k, \boldsymbol{y}_j), k = 1, 2, \dots, l, j = 1, 2, \dots, N_y$ и подвергаются операции клонирования C:

$$C[\boldsymbol{d}_k] = \boldsymbol{d}_k^c, k = 1, 2, \dots, l \; c = 1, 2, \dots, q$$

Количество образуемых клонов q_k каждого из l детекторов $\boldsymbol{d}_k^c$ пропорционально взаимной аффинности $a_{kj}(\boldsymbol{d}_k, \boldsymbol{y}_j)$:

$q_k \propto k_c \cdot a_{kj}(\boldsymbol{d}_k, \boldsymbol{y}_j),$

где – k_c коэффициент пропорциональности при клонировании.

При этом общее количество образованных клонов должно оставаться равным N_d:

$$\sum_{k=1}^{l} q_k = \left\lfloor \sum_{k=1}^{l} k_c \cdot a_{kj}(\boldsymbol{d}_k, \boldsymbol{y}_j) \right\rfloor = N_d.$$

То есть все детекторы популяции $\mathbf{D}_k^\lambda$ заменяются клонами $\boldsymbol{d}_k^c$: $\mathbf{D}_k^\lambda \to \mathbf{D}_k^{\lambda c}$. Очевидно, что для выполнения этого условия

$$k_c = {N_d}/{\sum_{k=1}^{l} a_{kj}(\boldsymbol{d}_k, \boldsymbol{y}_j)}.$$

Операция клонирования повышает вероятность покрытия детекторами областей распределения соответствующих биометрических признаков $\boldsymbol{y}_j$.

5. Все клоны $\boldsymbol{d}_k^c$ популяции $\mathbf{D}_k^{\lambda c}$ подвергаются операции гипермутации G:

$G[\boldsymbol{d}_k^c] = \boldsymbol{d}_k^{cG}, k = 1, 2, \dots, l \; c = 1, 2, \dots, q.$

Операцию гипермутации G клонов предлагается реализовать путем изменения на случайные величины $0 < \xi < \delta$ некоторого числа m компонент векторов детекторов $\boldsymbol{d}_k^c$. При этом гипермутация G клонов $\boldsymbol{d}_k^c$ осуществляется обратно

www.esa-conference.ru

пропорционально взаимной аффинности $a_{kj}(\boldsymbol{d}_k, \boldsymbol{y}_j)$:

$$G \propto k_m \,/\, a_{kj}(\boldsymbol{d}_k, \boldsymbol{y}_j),$$

где $-k_m$ – коэффициент гипермутации клонов $\boldsymbol{d}_k^{cG}$ определяемый из условия: $m = 1$ при $\max\limits_{k=1,2,\dots,l} a_{kj}(\boldsymbol{d}_k, \boldsymbol{y}_j)$

Операция гипермутации сужает область поиска новых эффективных детекторов.

Детекторы $\boldsymbol{d}_k^{cGm}$ заменяют популяцию $\mathbf{D}_k^{\lambda c}$ на новую $\mathbf{D}_k^{\lambda cG}$.

6. Для каждой пары $\boldsymbol{d}_k^{cG} \in \mathbf{D}_k^{\lambda cG}$ и $\boldsymbol{y}_j \in \overline{\mathbf{Y}}_j$ вычисляется степень взаимной аффинности.

$$a_{kj}(\boldsymbol{d}_k^{cG}, \boldsymbol{y}_j) = \sqrt{\sum_{p=1}^{N} \left(d_{kp}^{cG} - y_{jp}\right)^2}, p = 1, 2, \dots, N.$$

Результатом будет матрица взаимной аффинности $\mathbf{A}$, содержащая $N = N_d \cdot N_y$ элементов a_{ij}:

$\mathbf{A}\left\|a_{kj}(\boldsymbol{d}_k^{cG}, \boldsymbol{y}_j)\right\|, k = 1, 2, \dots, N_d \; j = 1, 2, \dots, N_y.$

7. Из каждого столбца матрицы $\mathbf{A}$ отбирается совокупность из l детекторов $\boldsymbol{d}_k^{cG}$, соответствующих наибольшей взаимной аффинности $a_{kj}(\boldsymbol{d}_k^{cG}, \boldsymbol{y}_j), k = 1, 2, \dots, l, j = 1, 2, \dots, N_y$. Полученные детекторы $\boldsymbol{d}_k^{cGm}$ образуют популяцию детекторов памяти $\mathbf{D}^{\mathrm{M}}$.

8. Проверка условия останова: при невыполнении условия, – переход на следующий шаг, иначе на шаг 10.

9. $(N_d - l)$ детекторов популяции $\mathbf{D}_k^{\lambda cG}$, обладающих наименьшей аффинностью $\boldsymbol{a_{kj}(d_k, y_j)}$ заменяются новыми, путем их случайной генерации (с равномерным законом распределения) новой популяции детекторов $\mathbf{D}_k^{\lambda} = \boldsymbol{d}_1, \boldsymbol{d}_1, \dots, \boldsymbol{d}_{N_d}, \lambda = \lambda + 1, k = 1, 2, \dots, (N_d - l)$ представленных векторами в формате векторов $\boldsymbol{y}_j$.

10. Останов, конец алгоритма.

Условием останова алгоритма является достижение заданного максимального размера популяции детекторов памяти $\mathbf{D}^M = \mathbf{D}_{\max}^M$, образующейся при $k = N_M$.

В фазе распознавания элементы $\boldsymbol{y}_j$ анализируемой последовательности биометрических признаков $\mathbf{Y}_j$ сопоставляются с детекторами $\boldsymbol{d}_k^M$ популяции памяти $\mathbf{D}^M, k = 1, 2, \dots, N_M$ с использованием меры близости Евклида между векторами $\boldsymbol{y}_j$ и $\boldsymbol{d}_k^M$:

$$\nabla(\boldsymbol{y}_j, \boldsymbol{d}_k^M) = \sqrt{\sum_{\upsilon=1}^{s} \left(y_{j\upsilon} - d_{k\upsilon}\right)^2}.$$

Критический уровень близости $\nabla(\boldsymbol{y}_j, \boldsymbol{d}_k^M) = \nabla^*$ определяет границу для принятия системой решения «свой/чужой» и задается, исходя из допустимых ошибок первого рода. Если для некоторой пары $\boldsymbol{y}_j$ и $\boldsymbol{d}_k^M$ $\nabla(\boldsymbol{y}_j, \boldsymbol{d}_k^M) > \nabla^*$, то считается, что элемент $\boldsymbol{y}_j$ анализируемой биометрии $\mathbf{Y}_j$, принадлежит «чужому».

Существенные вариации параметров динамической биометрии в последовательностях $\mathbf{Y}_j$ и значительные размеры самих последовательностей $\mathbf{Y}_j$ определяют целесообразность применения статистического подхода для принятия ИИС решения «свой»-«чужой» [8, 9]. При таком подходе контролируется частота f выполнения условия $\nabla(\boldsymbol{y}_j, \boldsymbol{d}_k^M) > \nabla^*$, которая определяет статистическую вероятность принадлежности анализируемой биометрии «чужому»:

$$\hat{P}^{\text{ч}} \approx f = n_{\nabla}^{+}/n_{\nabla},$$

где n_{∇}^{+} число случаев выполнения условия $\nabla(\boldsymbol{y}_j, \boldsymbol{d}_k^M) > \nabla^*$ в n_{∇} проведенных операциях сопоставлений $\boldsymbol{y}_j$ с $\boldsymbol{d}_k^M$.

Принятие решения о принадлежности анализируемой биометрии «чужому» считается обоснованным, при превышении частоты f заданного порогового значения $f_{\text{п}}$:

$$\mathbf{Y}_j \equiv \begin{cases} \mathbf{Y}_j^{\text{с}}, \text{если } f < f_{\text{п}}; \\ \mathbf{Y}_j^{\text{ч}}, \text{если } f \geq f_{\text{п}}, \end{cases}$$

где $\mathbf{Y}_j^{\text{с}}$ – последовательность биометрических признаков «своего»;

$\mathbf{Y}_j^{\text{ч}}$ – последовательность векторов признаков «чужого»;

Заключение. Предлагаемый подход в рамках иммунологического представления позволяет обобщить существенно различные существующие методы идентификации личности по динамическим биометрическим параметрам разной модальности – голоса, рукописи и клавиатурного набора.

Отличиями предлагаемого подхода являются:

- возможность текстонезависимого анализа ДБ различной модальности, произвольного объема и содержания;
- замена интегральной оценки результатов анализа биометрических данных за фиксированный период времени, применяемой в традиционных подходах к их непрерывной оценке в темпе их поступления, с возможностью своевременного принятия решения о присутствии «чужого»;
- применение иммунологической МКС, которая хорошо согласуется с большинством задач динамической идентификации личности, использующих режим верификации, что позволяет существенно уменьшить число детекторов, необходимых для эффективного распознавания биометрических данных.

Литература:

1. Матвеев Ю.Н. Технологии биометрической идентификации личности по голосу и другим модальностям. Вестник МГТУ им. Н.Э. Баумана, серия Приборостроение. – 2012. – № 2. – С. 46-61.

2. Брюхомицкий Ю.А., Казарин М.Н. Система аутентификации личности по почерку // Сборник трудов научно-практической конференции с международным участием «Информационная безопасность». – Таганрог: Изд-во ТРТУ, 2002. – С. 22-29.

3. Брюхомицкий Ю.А., Казарин М.Н. Метод биометрической идентификации пользователя по клавиатурному почерку на основе разложения Хаара и меры близости Хэмминга // Известия ТРТУ. – Таганрог: Изд-во ТРТУ, 2003. – № 4(33). – С. 141-149.

4. Брюхомицкий Ю.А. Цепочный метод клавиатурного мониторинга // «Известия ЮФУ. Технические науки». Тематический выпуск «Информационная безопасность». – Таганрог: Изд-во ТТИ ЮФУ, 2009. – №11. – С. 135-145.

5. Dasgupta D. Artificial Immune Systems and Their Applications, Ed., Springer-Verlag. – 1999.

6. De Castro L.N., Von Zuben F.J. The Clonal Selection Algorithm with Engineering Applications, submitted to GEC-CO'00. – 2000. – P. 36–37.

7. Брюхомицкий Ю.А., Казарин М.Н. Методы многосвязного представления клавиатурного почерка / Материалы III Международной конференции «Нелокальные краевые задачи и родственные проблемы математической биологии, информатики и физики. – Нальчик, 5-8 декабря 2006 г. – С. 68-69.

8. Брюхомицкий Ю.А. Иммунологичекий метод верификации рукописи с использованием векторного представления данных // Известия ЮФУ. Технические науки. – Ростов-на-Дону: Изд-во ЮФУ, 2016. – №9 (182). – С. 50-57.

9. Брюхомицкий Ю.А. Клавиатурный мониторинг на основе иммунологического клонирования // Безопасность информационных технологий. – М.: Изд-во МИФИ, 2016. – № 4 (40). – С. 5-11.