

Yuri A. Bryukhomitsky
Southern Federal University, 347928, Taganrog, St. Chekhov, 2,
e-mail: bya@tgn.sfedu.ru, ORCID iD is 0000-0002-7845-7487

Keyboard monitoring based upon the immunologic cloning

Keywords: biometric keyboard monitoring system, text-independent control and analysis keyboard script, immunologic cloning

The Biometric Keyboard Monitoring System is represented. It's intended for permanent text-independent control and analysis of automated information data systems users' keyboard script. It's suggested a keyboard monitoring method, which is combined the idea and advantages of threaded method of keyboard parameters representation and immunological approach to its realization, based upon the detectors cloning model. Suggested method potentially possesses a pinpoint accuracy, higher convergence rate of classification problems solving, ability to learn on only "own" class exemplars.

Ю.А. Брюхомицкий
Южный федеральный университет, 347928, г. Таганрог, ул. Чехова, 2,
e-mail: bya@tgn.sfedu.ru, ORCID iD is 0000-0002-7845-7487

КЛАВИАТУРНЫЙ МОНИТОРИНГ НА ОСНОВЕ ИММУНОЛОГИЧЕСКОГО КЛОНИРОВАНИЯ

Ключевые слова: биометрическая система клавиатурного мониторинга, текстонезависимый контроль и анализ клавиатурного почерка, иммунологическое клонирование

Рассматривается биометрическая система клавиатурного мониторинга, предназначенная для непрерывного текстонезависимого контроля и анализа клавиатурного почерка пользователей автоматизированных информационных систем. Предлагается метод клавиатурного мониторинга, сочетающий в себе идею и преимущества цепочного метода представления клавиатурных параметров и иммунологический подход к его реализации, основанный на модели клонирования детекторов. Предлагаемый метод потенциально обладает большей точностью, более высокой скоростью сходимости при решении задач классификации, способностью обучаться на экземплярах только одного «своего» класса.

В области информационной безопасности все более широкое применение находят биометрические технологии идентификации личности, в том числе, – основанные на анализе клавиатурного почерка. Особое место в этих технологиях занимает клавиатурный мониторинг пользователей автоматизированных систем [1].

Биометрическая система клавиатурного мониторинга предназначена для непрерывного текстонезависимого контроля и анализа клавиатурного почерка пользователей автоматизированных информационных систем. При необходимости клавиатурный мониторинг может проводиться скрытно (прозрачно) для пользователей. Системы клавиатурного мониторинга позволяют решать ряд важных задач в области информационной безопасности, которые трудно или невозможно решить другими методами: непрерывная (скрытная) верификация личности пользователей компьютерных систем; скрытное выявление пользователей-инсайдеров, совершающих действия, нарушающие информационную безопасность; обнаружение (при необходимости скрытное) отклонений психофизических характеристик пользователей от нормы и др.

В системах клавиатурного мониторинга, предназначенных для верификации личности работающего в информационной системе пользователя, первостепенное значение имеют точность и длительность верификации, которые, в свою очередь, определяются

КЛАВИАТУРНЫЙ МОНИТОРИНГ НА ОСНОВЕ ИММУНОЛОГИЧЕСКОГО КЛОНИРОВАНИЯ

способами представления и классификации клавиатурных биометрических параметров, а также, – подходами и методами, используемыми для реализации процедуры верификации.

Известны два основных подхода к представлению клавиатурных биометрических параметров. Первый подход основан на прямом измерении длительностей удержания и пауз между удержаниями одиночных клавиш. Недостатком такого подхода является низкая точность, обусловленная, недостаточной информативностью представления клавиатурных параметров. Вторым подход (цепочный) отличается от первого тем, что измеряются параметры клавиатурного набора не отдельных символов, а их сочетаний заданной длины (цепочек) [2]. Такой подход обладает более высокой точностью верификации, «оплачиваемой» дополнительным объемом вычислений.

Появление и интенсивное развитие нового направления искусственного интеллекта, – аппарата искусственных иммунных систем (ИИС) [3] – предопределило перспективу его использования, прежде всего, в области информационной безопасности. Это оправдывает исследования возможности и эффективности применения подхода ИИС для решения задач КМ. В работах [4] были показаны принципы построения систем КМ с использованием им иммунологической модели отрицательного отбора и строкового представления данных. Однако такие системы требуют значительных временных затрат на обучение, связанных со случайной генерацией и отбором эффективных для распознавания «чужих» детекторов.

Целью данной работы является разработка метода КМ, сочетающего в себе идею и преимущества цепочного метода представления клавиатурных параметров [2] и иммунологический подход к его реализации, основанный на модели клонирования детекторов [5, 6].

При использовании цепочного метода представления клавиатурных параметров множество всех возможных событий клавиатуры рассматривается как *алфавит* $\mathbf{A}$, содержащий два подмножества $\mathbf{A} = \mathbf{A}_y \cup \mathbf{A}_n$:

$\mathbf{A}_y \subset \mathbf{A}$ – события клавиатуры, состоящие в удержании одной из n клавиш;

$\mathbf{A}_n \subset \mathbf{A}$ – события клавиатуры, состоящие в наличии пауз между удержаниями очередных клавиш или перекрытий времени удержаний смежных при наборе клавиш.

Ограниченные последовательности событий клавиатуры множества $\mathbf{A}$, ориентированные слева направо, начинающиеся и оканчивающиеся событиями из подмножества $\mathbf{A}_y$, трактуются как *цепочки событий* $T_{i_1, i_2, \dots, i_p}$. Длиной r цепочки является общее число событий алфавита $\mathbf{A}$, входящих в эту цепочку: $|T_{i_1, i_2, \dots, i_p}| = r$, $i_1, i_2, \dots, i_p = 1, 2, \dots, n$. Поскольку при клавиатурном наборе события из множеств $\mathbf{A}_y$ и $\mathbf{A}_n$ строго чередуются, в каждой цепочке длины r будет содержаться p событий множества $\mathbf{A}_y$ и $q = p - 1$ событий множества $\mathbf{A}_n$. Длина цепочек всегда представлена нечетным целым числом и при $q = 1, 2, \dots$ равна $r = p + q = 2p - 1 = 2q + 1 = 3, 5, \dots$.

Таким образом, для заданного числа контролируемых клавиш n и заданной длины цепочек r суть цепочного метода представления клавиатурных параметров в терминах формальных грамматик состоит в формировании всех возможных цепочек событий алфавита $\mathbf{A}$ длины r в пространстве размерности $p = (r + 1)/2 = q + 1$.

Для представления клавиатурных параметров цепочным методом в поле действительных чисел P задается пространственная матрица размерности p

$$\mathbf{T}^p = \left\| T_{i_1, i_2, \dots, i_p} \right\|, \quad i_1, i_2, \dots, i_p = 1, 2, \dots, n, \quad p = 2, 3, \dots,$$

КЛАВИАТУРНЫЙ МОНИТОРИНГ НА ОСНОВЕ ИММУНОЛОГИЧЕСКОГО КЛОНИРОВАНИЯ

состоящая, в общем случае, из n^p элементов, представленных цепочками $T_{i_1, i_2, \dots, i_p}$, $i_1, i_2, \dots, i_p = 1, 2, \dots, n$, $p = 2, 3, \dots$.

При этом каждая цепочка длины r будет содержать r временных параметров из числового поля P :

- длительность τ_{i_1} – удержания клавиши i_1 ;
- длительность τ_{i_1, i_2} – паузы между удержаниями клавиш i_1 и i_2 ;
- длительность τ_{i_2} – удержания клавиши i_2 ;
- длительность τ_{i_2, i_3} – паузы между удержаниями клавиш i_2 и i_3 ;
-
- длительность τ_{i_p} – удержания клавиши i_p .

Для реальных языковых алфавитов $\mathbf{A}$, содержащих ограниченное число учитываемых при анализе символов, а также не используемых в языке сочетаний из более двух одинаковых последовательно идущих символов, общее число элементов матрицы $\mathbf{T}^p$ будет несколько меньшим, чем n^p .

Для описания и последующего использования пространственной матрицы $\mathbf{T}^p$, она представляется совокупностью своих сечений с фиксированным значением одного i_α или двух индексов i_α, i_β . В первом случае образуется совокупность *простых* сечений ориентации i_α , являющихся $(p-1)$ -мерными матрицами p -го порядка. Во втором случае образуется совокупность *двукратных* сечений ориентации i_α, i_β , являющихся $(p-2)$ -мерными матрицами p -го порядка.

При реализации цепочного метода длина цепочки $r = |T_{i_1, i_2, \dots, i_p}|$ выбирается фиксированной по формуле $r = 3 + 2l$, где $l = 0, 1, 2, \dots$.

Цепочный метод основан, по существу, на учете особенностей вхождения символов в контекст, когда идентифицирующие пользователя клавиатурные параметры определяются не одиночными событиями клавиатуры, а цепочками лингвистически связанных событий, характеризующихся более выраженной индивидуальностью их клавиатурного набора данным пользователем. Это обстоятельство хорошо согласуется с принципами представления и анализа информационных потоков, используемыми в ИИС.

Особенностью подхода ИИС является представление информационных потоков в виде последовательности информационных единиц фиксированного размера с последующей их децентрализованной обработкой. Используются два варианта представления информационных потоков: строковый и векторный. В первом случае за единицу информационного потока принимается строка фиксированной длины l , во втором – вектор в пространстве мерности l .

При строковом представлении анализ информационного потока осуществляется путем последовательного сопоставления строк потока с детекторами, имеющими аналогичный формат. Сопоставление осуществляется по принципу частичного совпадения строк в r смежных позициях. При векторном представлении анализ информационного потока осуществляется путем последовательного сопоставления векторов потока с векторами-детекторами с использованием различных мер близости – Евклида, Манхэттена и др. Степень близости единиц информационного потока и детекторов имитирует свойство аффинности иммунной системы, определяемое как степень связности чужеродного агента (антигена) и антителом, вырабатываемым иммунной системой.

Принцип иммунологического анализа клавиатурных параметров, представленных цепочным методом, заключается в том, что параметру аффинности содержательно и метрически предлагается поставить в соответствие длину цепочки $r = p + q$ лингвистически связанных событий клавиатуры (p – число событий из множества $\mathbf{A}_y$, $q = p - 1$ –

число событий из множества $\mathbf{A}_n$. Это позволяет использовать векторный способ иммунологического анализа клавиатурных параметров.

Для реализации предлагаемого метода задается r -мерное Евклидово пространство E^r . Мерность пространства r соответствует длине цепочки $|T_{i_1, i_2, \dots, i_p}|$, содержащей p последовательно следующих символов текста с q паузами между ними: $r = p + q$. Рабочее подпространство $E_p^r \subset E^r$ событий клавиатуры задается минимаксными значениями координат $x_1, x_2, \dots, x_r$, соответствующих предельным значениям временных параметров событий клавиатуры зарегистрированных в информационной системе пользователей.

Этапу обучения системы КМ предшествует создание набора $\mathbf{M}$ клавиатурных биометрических эталонов $\mathbf{M} = (\mathbf{M}_1^r, \mathbf{M}_2^r, \dots, \mathbf{M}_m^r)$, $\mathbf{M}_i^r \in \mathbf{M}^r$, $i = 1, 2, \dots, m$ всех m зарегистрированных в системе пользователей. Каждый эталон $\mathbf{M}_i^r$, формируется как элементы $T_{i_1, i_2, \dots, i_p}$ матрицы

$$\mathbf{T}^p = \|T_{i_1, i_2, \dots, i_p}\|, i_1, i_2, \dots, i_p = 1, 2, \dots, n, \quad p = 2, 3, \dots$$

и представлен совокупностью векторов $\mathbf{M}_i^r = (M_1^r, M_2^r, \dots, M_N^r)$, $M_j^r \in \mathbf{M}_i^r$, $i = 1, 2, \dots, m$, $j = 1, 2, \dots, N$, где $N = n^p$. Координаты векторов $M_j^r = (x_1, x_2, \dots, x_r)$ определяются значениями соответствующих временных параметров событий клавиатуры $\tau_1, \tau_2, \dots, \tau_r$ в $T_{i_1, i_2, \dots, i_p}$.

В отличие от традиционных схем распознавания образов, в которых эталоны представляют образцы «своих» и сопоставляются с неизвестными образцами, в иммунологической модели эталоны пользователей служат лишь основой для формирования собственно распознающих элементов – детекторов, которые моделируют антитела иммунной системы и предназначены для выявления чужеродных агентов (антигенов) путем установления степени связности с ними.

Для реализации такой схемы распознавания форматы входных данных, эталонов и детекторов выбираются одинаковыми. Информационный поток событий клавиатуры работающего пользователя рассматривается как последовательность цепочек $T_{i_1, i_2, \dots, i_p}$, представленных векторами $\mathbf{P}_k^r = (P_1^r, P_2^r, \dots)$, $P_k^r \in \mathbf{P}_k^r$, $k = 1, 2, \dots$. Координаты векторов $P_k^r = (x_1, x_2, \dots, x_r)$ определяются значениями соответствующих временных параметров текущих событий клавиатуры $\tau_1, \tau_2, \dots, \tau_r$.

Схему клавиатурного мониторинга рационально строить по принципу верификации, когда анализ клавиатурных параметров работающего пользователя осуществляется по отношению к эталону одного из множества зарегистрированных пользователей, который последним осуществил легальный вход в информационную систему. Такая схема сводится к сопоставлению параметров по принципу 1 : 1, что резко упрощает процедуру мониторинга. При смене одного легального пользователя другим достаточно сменить соответствующий биометрический эталон.

Формирование детекторов предлагается осуществить на основе иммунологической модели клональной селекции [5, 6] по алгоритму:

1°. Создается первоначальная популяция случайных детекторов $\mathbf{D}^r = (D_1^r, D_2^r, \dots, D_N^r)$, представленная N наблюдениями, $D_j^r \in \mathbf{D}^r$, $j = 1, 2, \dots, N$. Размер популяции $\mathbf{D}^r$ удобно принять равным числу векторов биометрического эталона $\mathbf{M}_i^r = (M_1^r, M_2^r, \dots, M_N^r)$, $j = 1, 2, \dots, N$.

КЛАВИАТУРНЫЙ МОНИТОРИНГ НА ОСНОВЕ ИММУНОЛОГИЧЕСКОГО КЛОНИРОВАНИЯ

2°. Для каждого детектора D_j^r из популяции $\mathbf{D}^r$ рассчитывается степень аффинности с векторами M_j^r эталона $\mathbf{M}_i^r$. Мерой аффинности является Евклидово расстояние между векторами M_j^r и D_j^r . Результатом будет матрица $\mathbf{A}$ (в общем случае пространственная), содержащая $N^2 = n^{2p}$ элементов аффинности:

$$\mathbf{A} \|d(x_{Mj}, x_{Dj})\|, \quad j = 1, 2, \dots, N,$$

где

$$d(x_{Mj}, x_{Dj}) = \sqrt{\sum_{j=1}^N (x_{Mj} - x_{Dj})^2}$$

3°. Из матрицы $\mathbf{A} \|d(x_{Mj}, x_{Dj})\|$ отбирается совокупность из l элементов, соответствующих наименьшей взаимной аффинности $d(x_{Mj}, x_{Dj})$ векторов M_j^r и D_j^r . Детекторы $\dot{D}_j^r$, соответствующие отобранным элементам матрицы $\mathbf{A} \|d(x_{Mi}, x_{Di})\|$ подвергается клонированию. При этом количество l образуемых клонов $\dot{D}_j^r$ пропорционально аффинности: $l = k_1 \cdot d(x_{Mj}, x_{Dj})$, где k_1 – коэффициент пропорциональности при клонировании. Это повышает вероятность покрытия области антигена.

4°. Образованные клоны $\dot{D}_j^r$ подвергаются мутации $\dot{D}_j^r \rightarrow \ddot{D}_j^r$ путем изменения произвольно выбранных координат детектора $(x_{D1}, x_{D2}, \dots, x_{Dr})$ на небольшую случайную величину δ : $\exists x \in x_{Dj} \quad x = x \pm \delta$. При этом степень мутации δ обратно пропорциональна аффинности: $\delta = k_2 / d(x_{Mj}, x_{Dj})$, где k_2 – коэффициент пропорциональности при мутации. Это сужает область поиска эффективных детекторов.

5°. Из мутированных клонов $\ddot{D}_j^r$ отбираются детекторы с максимальным уровнем аффинности $d(x_{Mj}, x_{Dj})$, которые заменяют детекторы исходной популяции $\mathbf{D}^r$. Элементы популяции с наиболее высоким уровнем аффинности переносятся в иммунную память $\bar{\mathbf{D}}^r$.

6°. Шаги 1-5 повторяются, пока не будет получена представительная популяция антигенов («чужих»).

Возможные критерии останова алгоритма обучения:

- время обучения;
- число итераций;
- размер популяции детекторов;
- предельно допустимое число не эффективных итераций, не добавляющих новых элементов в память.

Результатом работы алгоритма является сформированная иммунная память, представленная популяцией детекторов $\bar{\mathbf{D}}^r = (\bar{D}_1^r, \bar{D}_2^r, \dots, \bar{D}_N^r)$, которые выполняют роль лимфоцитов иммунной системы. Детекторы иммунной памяти «обязаны» не реагировать на события клавиатуры, порожденные работой «своего» легального пользователя, но должны распознавать клавиатурные параметры «чужого».

В рабочем режиме КМ последовательность событий клавиатуры пользователя, представленная векторами $\mathbf{P}_k^r = (P_1^r, P_2^r, \dots)$, $k = 1, 2, \dots$, сравнивается с детекторами иммунной памяти $\bar{D}_1^r, \bar{D}_2^r, \dots, \bar{D}_N^r$ с использованием меры близости Евклида $d(x_{Pk}, x_{Dj})$. Критический уровень степени близости $d(x_{Pk}, x_{Dj}) = d_0$ определяет границу для принятия системой КМ решения «свой/чужой» и задается, исходя из допустимых ошибок 1-го и 2-го рода. Срабатывание детектора $\bar{D}_j^r$ свидетельствует о том, что соответствующий вектор P_1^r контролируемого

КЛАВИАТУРНЫЙ МОНИТОРИНГ НА ОСНОВЕ ИММУНОЛОГИЧЕСКОГО КЛОНИРОВАНИЯ

информационного потока $\mathbf{P}_k^r$ событий клавиатуры работающего пользователя с высокой степенью вероятности принадлежит «чужому».

Особенностью идентификационных клавиатурных параметров является зависимость от психофизического состояния человека, что приводит к высокой степени вариации этих параметров для одной и той же личности. Это обстоятельство приводит к необходимости применения статистического подхода к процедуре принятия решения «свой/чужой», когда такое решение принимается не по одиночным срабатываниям детекторов.

Статистическую вероятность клавиатурного присутствия в системе «чужого» $\hat{P}^ч$ предлагается определять частотой срабатывания детекторов f :

$$\hat{P}^ч \approx f = \frac{n_D^+}{n_D},$$

где n_D^+ – число срабатываний детекторов в проведенных операциях проверки;

n_D – общее число проведенных операций проверки.

Принятие решения о наличии подмены «своего» пользователя «чужим» считается правомерным при превышении частоты f некоторого порогового значения f_{Π} :

$$\hat{P}^x \equiv \begin{cases} \hat{P}^c, & \text{если } f < f_{\Pi}; \\ \hat{P}^ч, & \text{если } f \geq f_{\Pi}. \end{cases}$$

Предложенный иммунологический метод КМ потенциально обладает рядом преимуществ. Учет индивидуальных корреляционных зависимостей между временными параметрами последовательно воспроизводимых символов текста позволяет с большей точностью идентифицировать пользователей, чем традиционные методы КМ на основе прямых измерений клавиатурных параметров. Используемая модель клональной селекции обладает высокой скоростью сходимости при решении задач классификации. В сочетании с векторным представлением клавиатурных данных она позволяет ускорить процесс обучения системы КМ и уменьшить время, необходимое для принятия достоверного решения о наличии «чужого». Еще одно преимущество модели – способность обучаться на экземплярах только одного «своего» класса.

СПИСОК ЛИТЕРАТУРЫ:

1. Брюхомицкий Ю.А. Клавиатурная идентификация личности. Lambert Academic Publishing, Saarbrücken, Germany, 2012, ISBN 978-3-8484-1119-1, – 140 с.
2. 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.
3. Dasgupta D., Artificial Immune Systems and Their Applications, Springer-Verlag, 1998.
4. Брюхомицкий Ю.А. Иммунологический подход к организации клавиатурного мониторинга / Известия ЮФУ. Технические науки. Тематический выпуск «Информационная безопасность». – Таганрог: Изд-во ИТА ЮФУ, 2014. – №2 (151). – С. 33-41.
5. De Castro L. N. The clonal selection algorithm with engineering applications / L. N. de Castro, F. J. von Zuben // Genetic and Evolutionary Computation Conference (GECCO'00): proceedings, Las Vegas, July 2000. – California : Morgan Kaufmann Publishers, 2000. – P. 36–37.
6. Зайцев С.А., Субботин С.А. Модели и методы автоматической классификации объектов по признакам на основе иммунотестирования // «Радиоэлектроника, информатика, управление». – Запорожье: Изд-во ЗНТУ, 2010, № 2. – с. 117-124.

REFERENCES:

1. Bryukhomitsky Yuri A. Klaviaturnaya identifikatsiya lichnosti. Lambert Academic Publishing, Saarbrücken, Germany, 2012, ISBN 978-3-8484-1119-1, – 140 с.
2. Bryukhomitsky Yuri A. Tsepchnyy metod klaviaturnogo monitoringa // «Izvestiya YUFU. Tekhnicheskiye nauki ». Tematicheskiy vypusk «Informatsionnaya bezopasnost'». - Taganrog: Izd-vo TTI YUFU, 2009. – №11. – P. 135-145.
3. Dasgupta D., Artificial Immune Systems and Their Applications, Springer-Verlag, 1998.
4. Bryukhomitsky Yuri A. Immunologicheskiy podkhod k organizatsii klaviaturnogo monitoringa / Izvestiya YUFU. Tekhnicheskiye nauki. Tematicheskiy vypusk «Informatsionnaya bezopasnost'». - Taganrog: Izd-vo ITA YUFU, 2014. – №2 (151). – P. 33-41.
5. De Castro L. N. The clonal selection algorithm with engineering applications / L. N. de Castro, F. J. von Zuben // Genetic and Evolutionary Computation Conference (GECCO'00): proceedings, Las Vegas, July 2000. – California : Morgan Kaufmann Publishers, 2000. – P. 36–37.
6. Zaytsev S.A., Subbotin S.A. Modeli i metody avtomaticheskoy klassifikatsii ob'yektov po priznakam na osnove immunokomp'yutinga // «Radioelektronika, informatika, upravleniye». - Zaporozh'ye: Izd-vo ZNTU, 2010, № 2. – P. 117-124.