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

Иммунологическая реализация клавиатурного мониторинга пользователей компьютерных систем

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

Введение. В системах клавиатурного мониторинга (КМ) решается задача непрерывной аутентификации пользователей компьютерных систем, по индивидуальным особенностям их клавиатурной работы на произвольных текстах, направленная на выявление факта подмены легального пользователя («своего») нелегальным («чужим»). Решающее значение при этом являются точность и время обнаружения подмены. Построение систем КМ, удовлетворяющих этим требованиям, тесно связано со способами представления клавиатурных биометрических параметров пользователей и методами их классификации [1].

Наиболее простым и распространенным способом представления клавиатурных биометрических параметров пользователя является прямое измерение временных характеристик клавиатурного ввода. Как правило, контролируются два типа временных параметров элементарных событий клавиатуры: длительность удержания клавиш и длительность пауз между удержаниями очередных клавиш. Возможные перекрытия времен удержаний смежных при наборе клавиш, интерпретируется отрицательным значением длительности паузы. На этапе аутентификации текущие значения клавиатурных параметров сравниваются с эталонными значениями, предварительно сформированными для данного пользователя в виде статистических оценок вариаций каждого параметра. На основании итогового баланса произведенных сравнений принимается соответствующее аутентификационное решение. Недостатком такой реализации КМ является низкая точность, обусловленная малой информативностью принятого представления клавиатурных параметров.

Последующие исследования в этой области показали, что для конкретного пользователя временные параметры одних и тех же событий клавиатуры, но встречающиеся в различных сочетаниях, заметно отличаются, что свидетельствует о наличии устойчивых индивидуальных корреляционных зависимостей между временными параметрами последовательно воспроизводимых символов текста. В результате был предложен цепочный метод представления клавиатурных параметров [1], обладающий существенно более высокой точностью аутентификации пользователя, но и требующий значительно больших затрат вычислительных ресурсов.

Перспективным подходом к организации КМ является использование принципов искусственных иммунных систем (ИИС) [2]. Иммунологический подход к реализации КМ приобретает ряд положительных качеств: устойчивость к вариации клавиатурных параметров, более высокую скорость обнаружения «чужого». Предложенная в [3] реализация такого подхода с использованием строкового представления клавиатурных параметров требует формирования и использования большого числа распознающих элементов – детекторов, что приводит к значительным затратам вычислительных ресурсов.

Постановка задачи. Целью данной работы является решение задачи КМ, сочетающее в себе идею и преимущества цепочного метода представления клавиатурных параметров, и иммунологический подход к его реализации, основанный на использовании МОО с числовым представлением данных.

Решение поставленной задачи. Множество всех возможных событий клавиатуры можно рассматривать как алфавит $\mathbf{A}$, содержащий два подмножества $\mathbf{A} = \mathbf{A}_у \cup \mathbf{A}_п$:

- $\mathbf{A}_у$ – события клавиатуры, состоящие в удержании одной из n клавиш;
- $\mathbf{A}_п$ – события клавиатуры, состоящие в наличии пауз между удержаниями очередных клавиш.

Ограниченные последовательности событий клавиатуры из множества $\mathbf{A}$, ориентированные слева направо, начинающиеся и оканчивающиеся событиями из подмножества $\mathbf{A}_у$, в терминологии формальных грамматик, можно трактовать, как цепочки событий $T_{i_1,i_2,\ldots,i_p}$. Длиной r цепочки является общее число событий алфавита $\mathbf{A}$, входящих в эту цепочку: $\left|T_{i_1,i_2,\ldots,i_p}\right| = r,\ \ i_1, i_2, \ldots, i_p = 1, 2, \ldots, n.$ Поскольку при клавиатурном наборе события из множеств $\mathbf{A}_у$ и $\mathbf{A}_п$ строго чередуются, в каждой цепочке длины r будет содержаться p событий множества $\mathbf{A}_у$ и $q = p - 1$ событий множества $\mathbf{A}_п$. Длина цепочек при $q = 1, 2, \ldots$ будет равна $r = p + q = 2p - 1 = 2q + 1 = 3, 5, \ldots$. 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 цепочного метода представления и обработки клавиатурных параметров удобно применить аппарат матриц.

При независимой регистрации всех событий клавиатуры длительность удержания всех контролируемых клавиш $i = 1, 2, \ldots, n$ отображается одномерной матрицей-строкой

$$\mathbf{T}^1 = [\tau_i],\ \ i = 1, 2, \ldots, n,$$

а результат регистрации длительности пауз между удержаниями всех парных сочетаний клавиш $i = 1, 2, \ldots, n$ отображается двумерной квадратной матрицей

www.esa-conference.ru

$$\mathbf{T}^2 = \left\| T_{ij} \right\|, \qquad i,j = 1,2,\dots,n.$$

В минимальном варианте реализации цепочного метода ($r = 3$, $p = 2$, $q = 1$) в поле действительных чисел P необходимо задать квадратную матрицу $\mathbf{T}^2 = \left\| T_{i1,i2} \right\|$, каждый элемент которой будет представлен цепочкой $T_{i_1,i_2,\dots,i_p}$ длины $r = 3$, содержащей три временных параметра из числового поля P:

– длительность τ_{i1} – удержания клавиши i_1;

– длительность $\tau_{i1,i2}$ – паузы между удержаниями клавиш;

– длительность τ_{i2} – удержания клавиши i_2.

В общем случае для реализации цепочного метода в поле действительных чисел P необходимо задать пространственную матрицу размерности p

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

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

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

Нетрудно видеть, что цепочный метод фактически использует многомерное представление связанных между собой событий клавиатуры, при котором каждой цепочке событий соответствует точка (ξ_1, ξ_2, ..., ξ_p) в пространстве мерности p с координатами ξ_k, $k = 1, 2, ..., p$, определяемыми событиями $i = 1, 2, ... n$ из множества $\mathbf{A}_y$.

Экспериментальные исследования подтвердили увеличение точности клавиатурной аутентификации пользователя с переходом от методов анализа усредненных одиночных событий клавиатуры к цепочному методу анализа. В то же время непосредственная реализация цепочного метода требует на этапе аутентификации больших объемов вычислений, связанных с обработкой пространственных матриц. Это обстоятельство стимулирует попытку реализации цепочного метода на иммунологических принципах.

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

Основной моделью распознавания «чужих» в ИИС является модель отрицательного отбора (МОО), которая приближенно воспроизводит механизм микробиологической защиты организма, реализуемый живой иммунной системой [4]. МОО содержит две фазы функционирования: обучение и распознавание. В фазе обучения осуществляется случайная генерации и последующий отбор информационных единиц – детекторов, отражающих аномальную структуру единиц исследуемых данных. В фазе распознавания исследуемые дынные сопоставляются с детекторами по принципу частичного соответствия. Положительные отклики детекторов свидетельствуют о наличии аномальных единиц в исследуемых данных.

Предлагается следующая иммунологическая реализация КМ.

В r-мерном Евклидовом пространств E^r задается рабочее подпространство $E_p^r \subset E^r$ событий клавиатуры ограниченное минимаксными значениями координат x_1, x_2, ..., x_r, которым соответствуют предельные значения вариации временных параметров τ_i и $\tau_{i,j}$, $i,j = 1,2,\dots,n$ событий клавиатуры p и q.

В фазе обучения ИИС КМ создаются числовые детекторы, необходимые для обнаружения отклонения клавиатурного почерка фактически работающего в компьютерной системе пользователя от почерка «своего» легального пользователя. Создание детекторов осуществляется на основе прототипа типичного процесса клавиатурного набора, свойственного «своему» пользователю, представленного его клавиатурным эталоном $\mathbf{T}^C$.

Клавиатурный эталон $\mathbf{T}^C$ пользователя строится на основе p-мерной матрицы, временных параметров событий клавиатуры, показанной выше, и представлен совокупностью информационных единиц в виде точек ξ_1, ξ_2, ..., ξ_N в пространстве E_p^r. Каждая точка $\xi_i = (x_1, x_2, ..., x_r)$, $i = 1, 2, ..., N$ соответствует цепочке символов длины $r = p + q$, а ее координаты задают значения временных параметров двух видов событий клавиатуры: $\mathbf{A}_y$ и $\mathbf{A}_n$. Параметр N определяется размерами текста, необходимого для формирования представительного эталона $\mathbf{T}^C$ на стадии обучения ИИС КМ.

Числовые детекторы представляют собой гиперсферы в пространстве E_p^r, которые характеризуются координатами своих центров и радиусами r_d. Простейший способ их формирования – случайная генерация векторов в пространстве E_p^r, концы которых соответствуют центрам гиперсфер. При этом сформированные гиперсферы детекторов должны с равномерной плотностью покрывать пространство E_p^r без пересечения с точками $\xi_1, \xi_2, ..., \xi_N$ клавиатурного эталона $\mathbf{T}^C$ «своего» пользователя.

Сопоставление признаков анализируемых клавиатурных данных с детекторами осуществляется на основе меры близости между соответствующими векторами в пространстве признаков. Принимается, что детектор обнаруживает элемент анализируемых данных, отличный от «своего», если расстояние от центра детектора до элемента данных меньше радиуса детектора. В качестве меры расстояния чаще всего используют Евклидово расстояние или расстояние Манхэттена.

При использовании способа случайного формирования детекторов их число, необходимое для качественного КМ, может оказаться чрезмерно большим, что увеличивает вычислительные затраты как в фазе обучения ИИС КМ, так и в рабочей фазе КМ. Для сокращения вычислительных затрат можно использовать т. н. V-детекторы (variable-sized detectors) [5], которые представлены гиперсферами с варьируемыми значениями своих радиусов.

Процесс формирования каждого V-детектора содержит две фазы. В первой фазе производится случайная генерация точки в пространстве E_p^r описанным выше способом. Во второй фазе сформированный детектор увеличивает свой объем, заполняя свободное от точек эталона $\mathbf{T}^C$ пространство E_p^r. Для этого первоначально сгенерированная точка $x_j = (x_1, x_2, ..., x_r)$ принимается в качестве центра $a_j = (a_1, a_2, ..., a_r)$ гиперсферы S^r, которая равномерно увеличивает свой радиус R. Процесс осуществляется в пошаговом режиме, в дискретные моменты времени:

$$[x_1(t_i) - a_1]^2 + [x_2(t_i) - a_2]^2 + \cdots + [x_r(t_i) - a_r]^2 = R(t_i)^2;$$

$$R(t_i) = R(t_{i-1}) + \Delta R;$$

$$x_j(t_i) = x_j(t_{i-1}) + \Delta R,\ j = 1, 2, \ldots, r,$$

$$t_i = t_{i-1} + \Delta t,\ i = 1, 2, \ldots$$

Процесс увеличения гиперсферы:

$$S^r(t_i) < S^r(t_{i+1}) < S^r(t_{i+2}), \qquad t_i = t_{i-1} + \Delta t, \qquad i = 1, 2, \ldots$$

продолжается до тех пор, пока в некоторый момент времени $t = t_k$ не произойдет ее пересечения с ближайшей точкой эталона «своего» пользователя $\xi_l = (x_{1l}, x_{2l}, \ldots, x_{rl}) \in \mathbf{T}^C$:

$$[x_{1l}(t_k) - a_1]^2 + [x_{2l}(t_k) - a_2]^2 + \cdots + [x_{rl}(t_k) - a_r]^2 \le R(t_k)^2. \quad (*)$$

Гиперсфера $S^r(t_k)$, зафиксированная в момент времени $t = t_k$, является прототипом детектора. Принимая во внимание ошибки первого рода, конечный объем детектора целесообразно принять равным тому, который был получен на предыдущем шаге итерационного процесса (*). Детальная процедура формирования детекторов изложена в работе [6].

Распределение гиперсфер детекторов различных радиусов в пространстве E_p^r зависит от распределения точек $\xi_1, \xi_2, ..., \xi_N$ эталона $\mathbf{T}^C$. В общем случае в процессе генерации детекторов, по мере покрытия ими пространства E_p^r, частота выполнения условия (*) и, следовательно, – отбраковки генерируемых детекторов будет повышаться. Повышение частоты отбраковки будет происходить также по мере уменьшения конечных радиусов R гиперсфер S^r. Эта тенденция является следствием постепенного заполнения рабочего пространства E_p^r детекторами, и ее можно использовать в качестве условий останова процедуры формирования детекторов:

– сравнение частоты f отбраковки детекторов с заданным пороговым значением частоты f_0;

– сравнение усредненного значения текущего радиуса $R_{ср}$ детекторов, сформированных за несколько последних шагов, с заданным пороговым значением R_0.

Это позволит определить эффективное количество детекторов, необходимых для распознавания «чужого». Величины f_0 и R_0 целесообразно увязать с ошибками первого и второго рода при проведении КМ.

При использовании V-детекторов пространство E_p^r будет покрываться существенно меньшим числом детекторов. При этом вычислительные затраты по созданию V-детекторов в фазе обучения ИИС КМ несколько возрастут, а в рабочей фазе КМ существенно уменьшатся. Учитывая то обстоятельство, что обучение ИИС КМ проводится гораздо реже, чем непосредственно рабочий режим распознавания клавиатурных данных, общие вычислительные затраты на иммунологическую реализацию КМ существенно снизятся.

В рабочей фазе КМ, в процессе клавиатурной работы контролируемого пользователя формируется некоторая последовательность событий клавиатуры, представленная точками $\xi_1, \xi_2, ..., \xi_N$, где $\xi_i = (x_1, x_2, ..., x_r)$, $i =$

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

1, 2, ..., *N*. Процесс распознавания «свой-чужой» реализуется путем пошаговой проверки каждой очередной точки ξ_j последовательности на попадание ее координат x_1, x_2, ..., x_r в гиперсферы S^r сформированных ранее детекторов. Срабатывание детектора свидетельствует о том, что данная точка в определенной степени отличается от точек эталона $\mathbf{T}^{\mathrm{C}}$ «своего» пользователя и должна рассматриваться как элемент данных «чужого».

Принятие решения о наличии подмены легитимного пользователя «чужим» целесообразно реализовать на основе «мягкого» решающего правила, основанного на статистической вероятности $p[P^{\mathrm{ч}}(t)]$ присутствия в КС «чужого», которое представлено пороговой частотой $f_{\mathrm{п}}$ срабатывания детекторов [7].

Заключение. По сравнению с общепринятым методом КМ, основанным на прямых измерениях временных интервалов событий клавиатуры, цепочный метод КМ обладает большей точностью аутентификации контролируемого пользователя, за счет учета индивидуальных корреляционных зависимостей между временными параметрами последовательно воспроизводимых символов текста. По сравнению с методом КМ, реализованным на принципах ИИС со строковым представлением данных [8], предложенный метод обладает как большей точностью аутентификации контролируемого пользователя, так и меньшим временем для принятия решения о клавиатурной работе «чужого». Замена строковых детекторов – числовыми с варьируемым пространственным объемом, позволяет снизить вычислительные затраты на реализацию непосредственно рабочего режима распознавания клавиатурных данных.

Литература:

1.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.

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

3.Брюхомицкий Ю.А. Иммунологические принципы организации клавиатурного мониторинга пользователей компьютерных систем / Материалы XII Международной научно-практической конференции «Информационная безопасность». Ч. I. – Таганрог: Изд-во ТТИ ЮФУ, 2012. – С. 10-19.

4.Forrest S., Perelson A. S., Allen L., Cherukuri R., Self-nonself discrimination in a computer, Proc. of 1994 IEEE Symposium on Research in Security and Privacy, 1994 г., с. 202-212

5.Ji Z., Dasgupta D. Real-valued negative selection algorithm with variable-sized detectors // Genetic and Evolutionary Computation (GECCO 2004): Proceedings. – Berlin–Heidelberg: Springer-Verlag, 2004. – Ser. LNCS 3102. – Part I. – P. 287–298.

6.Брюхомицкий Ю.А. Иммунологический подход к идентификации личности по динамическим биометрическим параметрам / Известия ЮФУ. Технические науки. – Ростов-на-Дону: Изд-во ЮФУ, 2017. – №5 (190). – С. 56-66.

7.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.

8.Брюхомицкий Ю.А. Иммунологические принципы организации клавиатурного мониторинга пользователей компьютерных систем / Материалы XII Международной научно-практической конференции «Информационная безопасность». Ч. I. – Таганрог: Изд-во ТТИ ЮФУ, 2012. – С. 10-19.