ПАРОЛЬНАЯ И НЕПРЕРЫВНАЯ АУТЕНТИФИКАЦИЯ ПО КЛАВИАТУРНОМУ ПОЧЕРКУ СРЕДСТВАМИ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Крутохвостов Д.С.[1], Хиценко В.Е.[2]

Все известные методы аутентификации пользователя на основе клавиатурного почерка не позволяют регулировать обе вероятности ошибок типа пропуск «чужого» и отказ «своему». В ГОСТ Р 52633.0–2006, где прописаны требования к средствам высоконадежной биометрической аутентификации, приводятся лишь рекомендации к организации выборочной проверки этих вероятностей. Из рекомендуемой там же стойкости к атакам подбора клавиатурного ключа следует вероятность пропуска чужого равная 0,01, что представляется довольно опасным. В работе проведен краткий обзор существующих подходов к аутентификации по клавиатурному почерку. Отмечается относительная сложность подходов и исключительно эмпирические способы оценки вероятностей ошибок. Предложен и исследован статистический метод парольной аутентификации на основе ранговых корреляций и непараметрических критериев выявления различий в параметрах положения и масштаба образцов набора пароля в виде последовательностей интервалов удержания клавиш и между нажатиями клавиш. Для задачи непрерывной аутентификации кроме выявления качественных отличий почерка предложен критерий Колмогорова–Смирнова для постоянного сопоставления эталонной и наблюдаемой эмпирических функций распределения интервалов удержания наиболее употребительных клавиш. А также вычисление ранговых корреляций между последовательностями этих интервалов. Экспериментальная проверка показала близость относительных частот ошибок с заранее установленными теоретическими вероятностями. Все предложенные методы отличаются алгоритмической простотой и математически обоснованными возможностями регулирования требуемых вероятностей ошибок обоих типов.

Ключевые слова: *информационная безопасность, клавиатурный почерк,парольная аутентификация, непрерывная аутентификация, непараметрическиеметодыстатистики, критерий Колмогорова–Смирнова, коэффициент ранговой корреляции Спирмена.*

DOI: 10.21681/2311-3456-2017-5-91-99

Введение

Важнейшим аспектом безопасности любой информационной системы является чёткое разграничение доступа к ресурсам системы и к её управлению. На сегодняшний момент наиболее актуальными методами разграничения являются биометрические методы контроля. Они имеют ряд преимуществ: биометрические характеристики индивидуальны для каждого человека и их трудно подделать, невозможно забыть или потерять. Одним из направлений биометрии является аутентификация пользователя по его клавиатурному почерку. Область применения обширна: от компьютерных систем и сетей небольших предприятий до систем, обеспечивающих выполнение критически важных процессов. Достоинства такого аутентификатора это прежде всего скрытностьи экономичность. Нужны лишь устройство ввода (клавиатура) и специальное программное обеспечение, позволяющее проводить анализ почерка.

В рамках исследования были кратко описаны существующие подходы к анализу клавиатурного почерка, представлены обоснованные ограничения к их применению в решении поставленных задач, проведен вероятностный анализ задач парольной и непрерывной аутентификации, предложены методы решения, позволяющие заранее установить приемлемые вероятности ошибок обоих типов и проведена экспериментальная проверка близости относительных частот ошибок и теоретических вероятностей.

Обзор состояния вопроса

Аутентификация по клавиатурному почерку может быть разделена на 2 вида: статическую (парольную) и динамическую (непрерывную). Статической аутентификацией является проверка и сопоставление характеристик почерка при наборе определённой текстовой последовательности, например, логина и пароля пользователя. Непрерывная аутентификация направлена на постоян-

1 Крутохвостов Дмитрий Сергеевич, аспирант, Новосибирский государственный технический университет, Новосибирск, Россия. E-mail: nero_92@mail.ru

2 Хиценко Владимир Евгеньевич, кандидат технических наук, доцент, Новосибирский государственный технический университет, Новосибирск, Россия. E-mail: xicenko@corp.nstu.ru

ный анализ почерка пользователя при работе за клавиатурой с целью выявления и сопоставления субъективных особенностей почерка.

Научное издательство Hindawi провело исследование подходов к распознаванию клавиатурных почерков[1].Здесь мы кратко резюмируем результаты этого исследования, не приводя ссылок на все используемые источники.

Основными инструментами являются группа статистических методов и группа методов машинного обучения, использующих нейронные сети, генетические алгоритмы и иные методы анализа данных, не относящиеся к математической статистике.

В начале исследований в данной области было популярно использование статистических методов. Задача сводилась к проверке гипотез о сходстве таких характеристик интервалов клавиатурного набора как среднее, медиана и стандартное отклонение. Для их сравнения используют, в частности, t-тест Стьюдента,F-статистику. Это предполагает нормальность распределения интервалов, что не всегда соответствует действительности.

Несомненным преимуществом подхода, использующего эти статистические процедуры, является возможность точно установить вероятность ошибки, но только типа отказ «своему».

Используются также такие вероятностные подходы к принятию решений по аутентификации, как байесовское оценивание, скрытые марковские модели, взвешенная вероятность. Здесь уже отсутствуют возможности предварительной установки приемлемых вероятностей ошибок.

Распространены методы машинного обучения, одним из которых является кластерный анализ. Он предназначен для объединения образцов почерка, расположенных в метрическом пространстве их параметров в близкие группы, кластеры. Из методов кластерного анализа используются метод k-средних и метод нечёткой кластеризации (fuzzy k-means) [2].

К методам машинного обучения также относится использование нейронных сетей, деревьев принятия решений, нечёткой логики и генетических алгоритмов.

Рассмотрим подробнее нейронную сеть, которая имитирует функции нейронов при обработке информации.

Классическая структура нейронной сети состоит из входного слоя, выходного слоя, и, по меньшей мере, одного скрытого слоя. Выборочные данные итеративно подаются в сеть, чтобы получить некоторые результаты, основанные на текущем состоянии начальных, заранее определенных, весов данной сети. Результаты работы сети сравниваются с истинными,и вычисляется значение ошибки. Это значение затем возвращается обратно в сеть, где идет коррекция весов в каждом скрытом слое. Эта обратная связь работает до тех пор, пока значение ошибки не оказывается ниже заданного порога.

Нейронные сети, как утверждается, способны производить лучший результат, чем статистические методы [6]. Тем не менее, классификаторы требуют не только эталонов легальных пользователей, но и образцы почерков «нелегальных» пользователей для обучения сети. Это является непрактичным на начальном этапе обучения. Кроме того, любое добавление, удаление или обновление профиля (эталона) пользователя в системе требует переобучения всей сети и, таким образом, время обработки увеличивается. Разделение баз данных и повышение квалификации сети в течение периода ожидания системы было предложено в качестве попытки разрешить эту проблему. При использовании нейронных сетей широко распространёнными являются следующие методы: сеть радиально-базисных функций, векторная квантизация при обучении, многослойный персептрон, самоорганизующиеся карты Кохонена.

Дерево принятия решений является примером распознавания образов в задачах с небольшим набором данных легальных пользователей. Это, как правило, требует меньших вычислительных мощностей по сравнению с нейронной сетью. Основная концепция - рекурсивно разделить обучающие данные таким образом, чтобы коэффициент увеличения информации максимизировался на каждом уровне иерархического дерева. Это продолжается до тех пор, пока не выйдем на вершины с одним классом или получим исчерпывающую информацию [13].

Подход на основе нечеткой логики использует многозначную логику для моделирования неоднозначных данных. Основная идея состоит в построении ограниченной области решения на основе обучающих данных с помощью функций принадлежности и нечетких правил. После того, как пространство признаков определено, степень истинности определяется на основе вычисления значений принадлежности. Применение нечеткой логики при аутентификации клавиатурного почерка исследовано в статье [2].

Генетический алгоритм [8], оптимизация роя частиц и оптимизация колонии муравьев - эти методы применяются, чтобы выбрать наиболее

оптимизированные образцы почерка для классификации, тем самым повышая ее точность.

Другой известный классификатор [1] выявляет злоумышленника, отделяя образцы легальных пользователей от неизвестных. Это метод опорных векторов (SVM), который создает область наименьшую из возможных, которая охватывает основные векторные данные почерка, относящихся к конкретному пользователю. SVM отображает вектор в многомерном пространстве признаков с помощью функции ядра (например, линейной, полиномиальной, сигмовидной или радиальной базисной). Алгоритм ищет разделительную функцию, которая инкапсулирует основные образцы, содержащиеся во входном векторе и векторе вне этой области. В результате, разделительная функция способна создавать более сложные границы, чтобы лучше определить, какие данные принадлежит новому шаблону. SVM, как утверждается в [15], имеет большую производительность по сравнению с нейронной сетью и требует меньше вычислительной мощности. Тем не менее, производительность его остается под вопросом, когда набор функций ядра слишком велик. Говоря о методах машинного обучения,важно понимать, что здесь мы можем только эмпирически определить относительные частоты ошибок обоих типов.

В данной работе предложены статистические методы парольной и непрерывной аутентификации, выгодно отличающиеся алгоритмической простотой, инвариантностью к закону распределения параметров почерка и, главное, возможностью устанавливать и регулировать вероятности ошибок обоих родов на основе сопоставления возможных потерь от этих ошибок.

Парольная аутентификация

Главный недостаток известных способов парольной аутентификации пользователя на основе сравнения клавиатурных почерков, как и других биометрических признаков – отсутствие математически обоснованных критериев выявления различий и, следовательно, возможности заранее установить приемлемые вероятности ошибок обоих родов. Те из методов, где статистические критерии все же привлекаются, не учитывают зависимость эталонной и текущей попыток набора, обусловленных геометрией клавиатуры и возможностями человеческой руки (ситуация связных выборок [5,7]), необоснованно предполагают нормальность временных интервалов при наборе пароля и относительно сложны.

Образец набора пароля можно представить последовательностью длительностей удержания клавиш $\tau_к$ и интервалов $\tau_и$ между отпусканием предыдущей и нажатием следующей клавиши. При этом $\tau_и$ может принимать отрицательные значения на близко расположенных клавишах. Если пароль состоит из n символов, то получим n значений $\tau_к$ и n-1 значений $\tau_и$. Рассмотрим несколько подходящих к ситуации статистических критериев различия сохраненной в памяти эталонной попытки и текущей попытки набора пароля. При этом желательно иметь возможность регулировать допустимые вероятности ошибок обоих родов: допуска чужого пользователя и ошибочного отказа своему. Сопоставляя последствия этих двух ошибок, можно выбрать приемлемое соотношение их вероятностей.

Задача сравнения двух образцов почерка относится к ситуации зависимых выборок. Дело в том, что интервалы $\tau_и$ определяются расположением символов на клавиатуре, а также темпом, возможностями руки, манерой и стабильностью почерка, т.е. индивидуальным почерком. А интервалы $\tau_к$ определяются практически только почерком. Именно геометрия клавиатуры приводит к тому, что интервалы между близкими клавишами в среднем будут меньше. В сущности, мы сравниваем одну и ту же выборку интервалов, но полученную в разных условиях, точнее, разными пользователями. В этом проявляется зависимость и это предопределяет выбор критериев различия.

Строго говоря, образец почерка это $2n$-1 значений различных и зависимых случайных величин. Попытка поиска одномерного закона распределения этих величин в каком-либо параметрическом семействе здесь совершенно несостоятельна. Также сомнительно полезной и неразрешимой практически представляется попытка оценивания $2n$-1-мерного закона совместного распределения зависимых интервалов. В этой ситуации единственно возможными являются непараметрические критерии выявления сходства и различия.

Первым этапом парольной аутентификации должен быть корреляционный анализ сходства последовательностей интервалов. Наиболее простым непараметрическим аналогом парного коэффициента корреляции является ранговый коэффициент Спирмена $\rho_з$. При проверке значимости корреляции имеем две гипотезы H_0: $\rho_з \le 0$, положительной корреляции с эталоном нет – «чужой» и одностороннюю альтернативную гипотезу H_1: $\rho_з > 0$, корреляция положительна – «свой». При этом именно зависимость интервалов $\tau_и$ от геометрии

Таблица 1.

Критические значения $\rho_{кр,\alpha}$ при односторонней альтернативе

n	Уровень значимости α							
	0,008	**0,007**	**0,006**	**0,005**	**0,004**	**0,003**	**0,002**	**0,001**
5	0,953	0,957	0,962	0,966	0,971	0,976	0,982	0,988
6	0,910	0,916	0,923	0,931	0,939	0,948	0,958	0,971
7	0,865	0,874	0,883	0,893	0,903	0,915	0,929	0,948
8	0,823	0,833	0,843	0,854	0,867	0,882	0,899	0,922
9	0,785	0,796	0,806	0,818	0,832	0,848	0,868	0,896
10	0,750	0,760	0,772	0,784	0,799	0,817	0,839	0,870
11	0,718	0,729	0,741	0,754	0,769	0,787	0,810	0,844

клавиатуры дает смещение корреляции в пользу H_1, что подтверждает эксперимент. Для объективности первого этапа следует проверять ρ_s только между последовательностями интервалов $\tau_к$.

Известно [6], что при справедливости H_0 верхние α %-е точки статистики Имана – Коновера

$$J = \frac{\rho_s}{2}\left(\sqrt{n-1} + \sqrt{\frac{n-2}{1-\rho_s^2}}\right) \quad (1)$$

могут быть найдены как $j(\alpha, n) = (z_\alpha + t_{\alpha,n-2})/2$, где z_α и $t_{\alpha,n-2}$ – верхние α%-е точки стандартного нормального распределения и распределения Стьюдента соответственно. Так были рассчитаны критические значения $\rho_{кр,\alpha}$ для разных уровней значимости α (Табл.1,[6]). Напомним, что здесь мы можем регулировать α, то есть, вероятность пропустить «чужого», как более опасную ошибку.

Однако высокая корреляция свидетельствует только о близости последовательностей интервалов по форме огибающих (сходная ритмика набора). При этом они могут существенно отличаться по параметру положения (средний темп) или в степени рассеяния относительно него, то есть, по стандартному отклонению (нестабильность почерка). Образцы почерка, показанные на рис.1, дают значимую корреляцию, но явно различны по медиане и по стандартному отклонению.

Очевидна необходимость последующей проверки попытки, прошедшей корреляционный этап, на близость к эталону по параметрам распределения. Опыт подтверждает, что сравнения параметров положения и масштаба нужно проводить отдельно для $\tau_к$ и $\tau_и$ как разных особенностей почерка. Замечено также, что эти проверки не являются независимыми. Следовательно, попытка, похожая на эталон по форме, должна пройти еще четыре проверки, чтобы доказать свое сходство с эталоном по среднему темпу и по степени нестабильности.

В этих критериях нуль-гипотеза формулируется иначе, чем в корреляционном анализе, а именно,

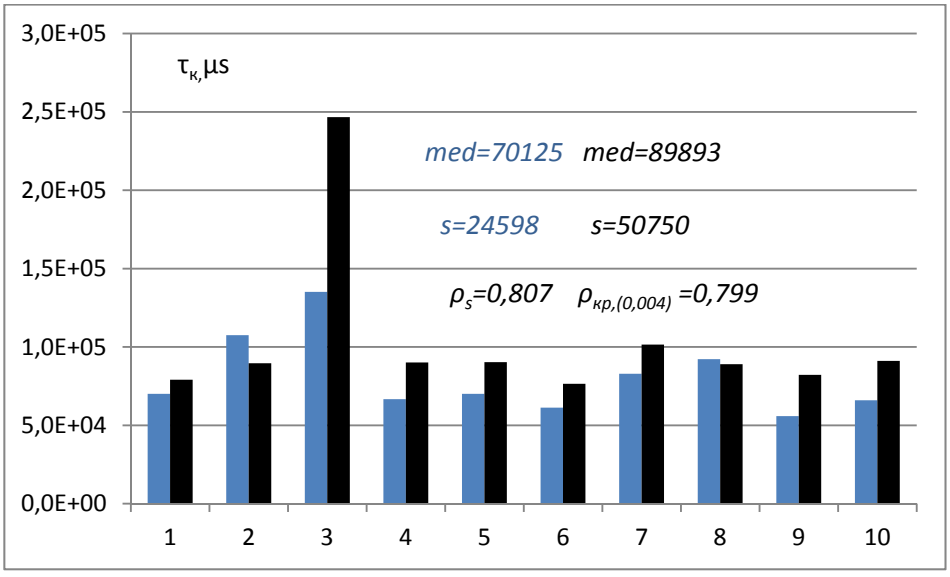

***Рис.1**. Примеры почерков близких по форме, но отличающихся по параметрам распределения.*

H_0: различия в параметре нет – «свой» при двусторонней альтернативе H_1: различие есть – «чужой». Таким образом, здесь мы можем регулировать вероятность ошибочно не пустить «своего».

Из большого арсенала средств непараметрической статистики, предназначенных для сравнения двух зависимых выборок по параметру положения, иначе говоря, выявляющих сдвиг между ними, был выбран знаково-ранговый критерий Уилкоксона. Этот критерий достаточно прост, но главное, таблицы критических значений статистики Уилкоксона, будут использоваться и для сравнения параметров масштаба, что упрощает аутентификатор.

Проверка по критерию Уилкоксона реализуется следующим образом. Имеем пары измерений (x_i, y_i), $i=1,\ldots,l$, гдех-эталон, у- проверяемый пользователь. Находим сдвиги $z_i = y_i - x_i$ и ранги модулей этих сдвигов $rank|z_i|$. Вычисляем суммы рангов, соответствующие отрицательным сдвигам T^- и положительным – T^+. Статистика Уилкоксона равна минимальной из этих сумм $T_{эмп} = min\{T^-, T^+\}$. Если $T_{эмп}$ окажется не более критического значения $T_{кр(\alpha)}$, H_0 отклоняем на принятом уровне значимости α, равном вероятности ошибочно отказать в доступе «своему».

Этому сравнению с эталоном подвергаем последовательно интервалы $\tau_к$ $(l=n)$ и $\tau_и$ $(l=n-1)$и при первом же отклонении H_0 отказываем в доступе.

Следующие два этапа – проверка различий в параметре масштаба тех же интервалов $\tau_к$ и $\tau_и$. Для этой задачи в ситуации связных выборок известны два непараметрических критерия: Сендвика-Олссона и Краута-Линерта [4]. Преимущество первого в простоте вычислений, а второй менее чувствителен к различию в параметрах положения сравниваемых выборок. Но поскольку к этому этапу пройдут лишь попытки, не имеющие сдвига по критерию Уилкоксона, то выбираем первый.

Критерий Сендвика-Олссона работает с теми же парами (x_i, y_i), $i=1,\ldots,l$. Находим медианы каждой из выборок и вычисляем сдвиг модулей отклонения от медиан $v_i = |y_i - medy_i| - |x_i - medx_i|$. Находим ранги модулей сдвигов $rank|v_i|$ и подсчитываем суммы рангов T^-, соответствующих отрицательным сдвигам, и положительным - T^+. Затем проверяем симметрию v_i по тому же критерию Уилкоксона, т.е. сравниваем статистику $T_{эмп} = min\{T^-, T^+\}$ с критическим значением из той же таблицы. Понятно, что и эту проверку различия в параметре масштаба проводим раздельно для интервалов $\tau_к$и $\tau_и$.

Алгоритм работы предлагаемого парольного аутентификатора следующий:

1. Вычисляем коэффициент ранговой корреляции между эталонной и фактической последовательностью интервалов. Если ρ_s менее критического значения, взятого из табл.1 для принятого уровня значимости α (вероятность пустить «чужого»), то в доступе отказываем.

2. Если отказа нет, то проводим еще 4 проверки. Сначала сравниваются $\tau_к$ по критерию знаковых рангов Уилкоксона, затем этой же проверке подвергаются $\tau_и$. Затем сравниваем параметры масштаба интервалов $\tau_к$ *и* $\tau_и$ по критерию Сэндвика-Олссона. И при первом же случае, когда соответствующая статистика оказывается не более табличного $T_{кр(\alpha)}$, отказываем в доступе на принятом уровне значимости α, равном вероятности ошибочно отказать в доступе «своему».

Этот алгоритм был реализован в программе, на которую было получено свидетельство о государственной регистрации программы для ЭВМ № 2014618829.

Была осуществлена экспериментальная проверка парольного аутентификатора. Участвовали 23 человека разного возраста, имеющих навыки работы с клавиатурой и установившийся почерк. Каждый делал 5-6 попыток набора пароля из десяти букв. В ходе эксперимента установлено, что при формировании эталона необходимо заставить пользователя достичь с помощью аутентификатора однородности почерка по всем названным критериям, обучиться стабильной ритмике набора. Так, чтобы эталон, полученный усреднением по попыткам и сохраненный в памяти, не отличался от последующих попыток, то есть, чтобы попытки проходили все 5 проверок с требуемой вероятностью. Пять участников эксперимента, выступая в качестве «своего», добивались указанной однородности почерка. В реальной работе можно предусмотреть адаптацию эталона к изменениям почерка с течением времени.

В результате из 2920 попыток при установленном α=0,001 было получено 2 ошибки прохождения первого корреляционного порога, т.е. относительная частота пропуска «чужого» равна 0,0007. Причем эти две попытки не прошли последующие четыре проверки на различие параметров сдвига и масштаба при установленном α=0,01.

Если обозначить вероятности ошибочно отклонить H_0 в каждом из четырех ранговых критериев различия $\alpha_i = P\{T_{эмпi} \le T_{кpi} | H_0\}, i = 1\ldots,4$ то вероятность отказать «своему» при независимости результатов проверок равна $1-(1-\alpha_1)(1-\alpha_2)$

$(1-\alpha_3)(1-\alpha_4)$.Отсюда можно найти и установить вероятность отказать «своему». Принимая эти вероятности равными α=0,01, получаем вероятность этой ошибки примерно 0,04. Из 115 попыток пройти свой эталон неудачными оказались 4, то есть фактическая относительная частота отказа «своему» составила 0,035.

Однако предположение о независимости проверок сомнительно, и проверка гипотезы о независимости $\tau_к$ и $\tau_и$ по критерию Кендэлла [5] показала, что примерно у половины участников зависимость проявляется в виде отрицательной корреляции. Это можно объяснить эффектом «перекрытия», когда при переходе на близко расположенную клавишу мы удерживаем нажатой предыдущую и отпускаем после нажатия на следующую, так что $\tau_к$ увеличивается, а $\tau_и$ становится отрицательной. Понятно, что при этом статистики $T_{эш}$ первой и второй проверок при сопоставлении различных в указанном смысле почерков будут уменьшаться и фактическая вероятность отказать своему будет менее установленной. Эта особенность почерка представляет интерес для непрерывной аутентификации.

Непрерывная аутентификация

Этот метод скрытого мониторинга должен применяться совместно с парольной аутентификацией для снижения риска, связанного с возможной подменой пользователя.

На основе литературы [1,3,9-17] можно выделить два этапа верификации пользователя: качественный и количественный. Первый выявляет нарушения заранее выявленных индивидуальных особенностей работы с клавиатурой легального пользователя. Это использование альтернативных служебных клавиш, (например, клавиши Backspace и Delete; использованию дополнительных ключей в клавиатуре, например, числа с numpad; при письме заглавных букв (CapsLock или правый/левый Shift)).Эти особенности проявляются на подсознательном уровне, и попытки контролировать их неизбежно отразятся на изменении динамики почерка.

Характерной особенностью почерка может служить время и частота пауз в наборе. Наличие вышеназванных «перекрытий», выявление отклонений от типичных для пользователя маршрутов в сети также служит качественным отличием почерка и основанием для тревоги. Для принятия решений о различии качественных особенностей почерка можно непрерывно обновлять и сравнивать частотные гистограммы употребления альтернативных клавиш и других названных особенностей на основе статистики χ^2 и проверять гипотезу о сходстве с установленной вероятностью отказа «своему».

Второй этап подразумевает проверку статистических гипотез о сходстве законов распределения $\tau_к$ и $\tau_и$ или близости числовых характеристик этих законов для наиболее часто встречающихся в текстах букв и интервалов между парами букв.

Представляется уместным использование последовательного анализа на основе статистик типа Колмогорова-Смирнова [18,19] для непрерывного сравнения эталонных функций распределения с наблюдаемыми, формируемыми по мере увеличения числа интервалов для часто встречающихся клавиш. Таким образом, задача непрерывной аутентификации принципиально отличается тем, что в ходе работы пользователи за клавиатурой идёт постоянное накопление данных почерка и обновление значений используемых статистик.

Мы непрерывно сравниваем эмпирические функции распределения легального пользователя (эталонные $F_m^{эт}(\tau_k)$), построенные для нескольких часто встречающихся букв с функциями распределения проверяемого пользователя, построенными на основе собранной к текущему моменту времени информации о длительностях удержания этих же клавиш $F_n(\tau_k)$.Здесь m и n - количество нажатий клавиши. Для проверки гипотезы $H_0: F_m^{эт}(\tau_k)) = F_n(\tau_k)$ против двусторонней альтернативы $H_1: F_m^{эт}(\tau_k)) \neq F_n(\tau_k)$ вычисляется статистика Колмогорова-Смирнова $D_{m,n} = max|F_m^{эт}(\tau_k)) - F_n(\tau_k)|$. Видно, что $D_{m,n}$ представляет собой максимальный перепад высот между двумя ступенчатыми эмпирическими функциями распределения. Затем находим $\Lambda = \sqrt{(mn/(m+n))}D_{m,n}$ и сравниваем с табличным $\lambda(\alpha, m, n)$. Если $\Lambda \geq \lambda(\alpha, m, n)$, мы можем отклонить гипотезу H_0, т.е. подозревать несанкционированный вход с вероятностью ошибочно отказать «своему» не более α. Таблицы критических значений $\lambda(\alpha, m, n)$ доступны в [5] и, поскольку в нашей задаче m фиксировано, объём необходимых для аутентификатора таблиц невелик.

Этому сравнению мы подвергаем эталонные функции распределения $\tau_и$ для каждой из заранее составленного списка наиболее распространённых букв и обновляемые с ростом n функции распределения проверяемого пользователя. Для повышения надежности можно проверять аналогично близости функций распределения интервалов $\tau_и$ между несколькимипарами часто

встречающихся букв. Примечательно, что критерий Колмогорова-Смирнова реагирует на любые различия параметров законов распределения, то есть, выявляет сдвиг в средних значениях и в нестабильности сравниваемых интервалов.

Открытым вопросом остаётся выбор и обоснование объёма*т*и режима обновления эталонной выборки, а также выбор минимального значения n как момента оптимальной остановки процесса наблюдения в случае, если достигнутая значимость окажется менее заранее установленного уровня α. При этом происходит отказ в продолжение работы или объявление тревоги.

Чтобы установить приемлемую величину ошибки пропуска «чужого», мы можем опять же использовать коэффициент ранговой корреляции Спирмена. Для этого мы должны хранить в качестве эталона усреднённые последовательности $\tau_к$ для часто встречающихся при работе легального пользователя слов. Когда в тексте проверяемого пользователя возникают эти слова, мы вычисляем коэффициент Спирмена и сравниваем его значение с критическим (табл.1) с приемлемой для нас вероятностью пропуска чужого также как при парольной аутентификации.

Таким образом, идет непрерывный мониторинг качественных и количественных отличий от эталонов с остановкой этого процесса при достижении критических значений используемых статистик, при этом происходит отказ в доступе.

Алгоритм работы предлагаемого непрерывного аутентификатора следующий:

1. После некоторого интервала накопления статистических данных сопоставляем эталонную и текущую частотные гистограммы встречаемости качественных особенностей почерка по статистике χ^2. При достижении критического значения $\chi^2_{кр(\alpha)}$ прекращаем доступ с принятой вероятностью ошибки типа отказа «своему» не более α. Иначе продолжаем накапливать данные.

2. Одновременно сравниваем эталонную и текущую эмпирические функции распределения интервалов удержания каждой из списка наиболее употребляемых клавиш. Сравниваем вычисленную статистику Колмогорова-Смирнова с критическим значением и при первом же случае, когда статистика λ оказывается не менее табличного $\lambda(\alpha, m, n)$, отказываем в доступе на принятом уровне значимости α, равном вероятности ошибочно отказать в доступе. Иначе продолжаем накапливать данные.

3. Если отказа нет и уже имеются типичные слова, вычисляем коэффициент Спирмена между последовательностями интервалов удержания клавиш $\tau_к$ в эталонном и проверяемом слове и сравниваем этот коэффициент с критическим значением на уровне, приемлемом для ошибки пропуска «чужого».

4. Если отказ в доступе еще не произошел, продолжаем наблюдение, повторяя проверки при каждом увеличении *n* и появлении качественных особенностей и типичных слов.

В ходе экспериментальной проверки количественного этапа непрерывной аутентификации были получены следующие результаты. Из 126 попыток сравнения почерка легального пользователя со своим эталоном в 3 попытках была объявлена «ложная» тревога при уровне значимости 0,05. В качестве легального пользователя выступали разные люди. Таким образом, относительная частота этой ошибки оказалась равной0,024.

Экспериментальная проверка сходства часто встречающихся слов на основе коэффициента Спирмена при заданной вероятности пропуска «чужого» 0,001 дала относительную частоту этой ошибки 0,0008, что практически совпадает с результатом проверки корреляционного этапа парольной аутентификации.

Выводы

Предложенный метод аутентификации по клавиатурному почерку демонстрирует свою эффективность, поскольку эмпирически найденные значения относительных частот ошибок обоих родов не превышают теоретических вероятностей, установленных на основе математической статистики. Алгоритмы проверки клавиатурного почерка просты и экономичны в вычислительном отношении, что обеспечивает возможность их использования в мобильных средствах коммуникации и терминальных устройствах доступа. Наконец, они могут применяться как составная часть комплексных биометрических средств защиты от несанкционированного доступа.

***Рецензент**: Пестунов Андрей Игоревич, кандидат физико-математических наук, доцент, Новосибирский государственный университет экономики и управления, г. Новосибирск, Россия. E-mail: a.i.pestunov@nsuem.ru*

Литература

1. PinShenTeh, Andrew Beng Jin Teoh, and ShigangYue, «A Survey of Keystroke Dynamics Biometrics», The Scientific World Journal, vol. 2013, Article ID 408280, 24 pages, 2013. doi:10.1155/2013/408280
2. S. Mandujano and R. Soto, «Deterring password sharing: user authentication via fuzzy c-means clustering applied to keystroke biometric data», in Proceedings of the 5th Mexican International Conference in Computer Science (ENC '04), pp. 181–187, September 2004.
3. Казанцев И. С. Анализ клавиатурного почерка в процессах аутентификации, идентификации и обнаружения подмены оператора // Молодой ученый. — 2016. — №9. — С. 167-169.
4. Хиценко В.Е. Непараметрическая статистика в задачах защиты информации. Конспект лекций, Новосибирск: Изд-во НГТУ, 2012, 196с.
5. Холлендер М., Вулф Д.А. Непараметрические методы статистики. М.: Финансыистатистика, 1983. – 518 с.
6. Хиценко В.Е., Крутохвостов Д.С. Статистическая аутентификация по клавиатурному почерку. Труды V-ой Межрегиональной научно-практическая конференции «Информационная безопасность и защита персональных данных. Проблемы и пути их решения», Брянск, 2013, С.97-102.
7. KhitcenkoV. E. Increasing the reliability of autentification keyboard handwriting / V. E. Khitcenko, D. S. Krutokhvostov // Актуальные проблемы электронного приборостроения (АПЭП–2014) = Actual problems of electronic instrument engineering (APEIE–2014) : тр. 12 междунар. конф., Новосибирск, 2–4 окт. 2014 г.: в 7 т. – Новосибирск : Изд-во НГТУ, 2014. – Т. 1. – С. 262–265. – 250 экз. – ISBN 978-1-4799-6019-4, ISBN 978-5-7782-2506-0.
8. K. Sung and S. Cho, «GA SVM wrapper ensemble for keystroke dynamics authentication», in Advances in Biometrics, D. Zhang and A. Jain, Eds., vol. 3832, pp. 654–660, Springer, Berlin, Germany, 2005.
9. Messerman,T. Mustafiґc, S. A. Camtepe, and S. Albayrak, «Continuous and non-intrusive identity verification in realtimeenvironments based on free-text keystroke dynamics», in Proceedings of the International Joint Conference on Biometrics(IJCB '11), pp. 1–8, October 2011.
10. F. Monrose and A. D. Rubin, «Keystroke dynamics as a biometric for authentication», Future Generation Computer Systems, vol. 16, no. 4, pp. 351–359, 2000.
11. D.Gunetti and C. Picardi, «Keystroke analysis of free text», ACM Transactions on Information and System Security, vol. 8, no. 3, pp. 312–347, 2005.
12. T. Shimshon, R.Moskovitch, L. Rokach, andY. Elovici, «Clustering di-graphs for continuously verifying users according to their typing patterns», in Proceedings of the IEEE 26th Convention of Electrical and Electronics Engineers in Israel (IEEEI '10), pp. 445–449, November 2010.
13. M. Pusara, An Examination of User Behavior for User Re-Authentication, Purdue University, West Lafayette, Ind, USA,2007.
14. J. C. Stewart, J. V. Monaco, S.-H. Cha, and C. C. Tappert, «An investigation of keystroke and stylometry traits for authenticating online test takers», in Proceedings of the International Joint Conference on Biometrics (IJCB '11), pp. 1–7, October 2011.
15. E. Yu and S. Cho, «Keystroke dynamics identity verification—itsproblems and practical solutions», Computers and Security, vol.23, no. 5, pp. 428–440, 2004
16. H. Davoudi and E. Kabir, «A new distance measure for free text keystroke authentication», in Proceedings of the 14th International CSI Computer Conference (CSICC '09), pp. 570–575, October 2009.
17. T. Shimshon, R.Moskovitch, L. Rokach, and Y. Elovici, «Continuous verification using keystroke dynamics», in Proceedings ofthe International Conference on Computational Intelligence andSecurity (CIS '10), pp. 411–415, December 2010.
18. Ширяев А.И. Обнаружение спонтанно возникающих эффектов// ДАН СССР. -1961. т.138, № 4. - С.794-801.
19. Обнаружение изменения свойств сигналов и динамических систем. Подред. М. БассвильиА. Банвенист. М.: «Мир», 1989. – 278 с.

PASSWORD AUTHENTICATION AND CONTINUOUS AUTHENTICATION BY KEYSTROKE DYNAMICS USING MATHEMATICAL STATISTICS

Krutohvostov D.[3], Khitsenko V.[4]

Abstract. *Almost all known methods of user authentication based on the keystroke dynamics don't allow you to adjust the required error probabilities of «insider skipping» and «false alarm». In GOST R 52633.0–2006, where the requirements for highly reliable biometric authentication are prescribed, only recommendations are given for organizing a selective verification of these probabilities.Out of the recommended resistance to attacking the selection of a keyboard key, there is a probability of missing someone else's equal to 0.01, which seems quite dangerous.In this paper, we briefly review and analyze the existing approaches to authentication based on the keystroke dynamics. There is a relative complexity and alone empirical ways of estimating the*

3 Dmitry Krutohvostov, graduatestudent, NovosibirskStateTechnical University, Novosibirsk, Russia. E-mail: nero_92@mail.ru

4 Vladimir Khitsenko, Ph. D.,NovosibirskStateTechnical University, Novosibirsk, Russia.E-mail: xicenko@corp.nstu.ru

probabilities of errors. A statistical method of password authentication based on rank correlations and nonparametric criteria for detecting differences in the position and scale parameters of the password set samples in the form of sequences of key hold intervals and between keystrokes is proposed and investigated. For the problem of continuous authentication, in addition to identifying qualitative differences in keystroke dynamics, the Kolmogorov–Smirnov test is proposed for a constant comparison of the empirical distribution functions of the holding intervals of the most commonly used keys. And also the calculation of rank correlations between intervals when typing frequently used terms. The experimental verification showed the closeness of the relative error frequencies with the theoretically established. All the proposed methods differ in algorithmic simplicity and mathematically justified possibilities of controlling the required error probabilities of both types.

Keywords:*Information security, password authentication, continuous authentication, keystroke dynamics, Kolmogorov–Smirnov test, Spearman rank correlation coefficient.*

References

1. Pin Shen Teh, Andrew Beng Jin Teoh, and Shigang Yue, «A Survey of Keystroke Dynamics Biometrics», The Scientific World Journal, vol. 2013, Article ID 408280, 24 pages, 2013. doi:10.1155/2013/408280
2. S. Mandujano and R. Soto, «Deterring password sharing: user authentication via fuzzy c-means clustering applied to keystroke biometric data», in Proceedings of the 5th Mexican International Conference in Computer Science (ENC '04), pp. 181–187, September 2004.
3. Kazantsev I. S. Analiz klaviaturnogo pocherka v protsesse autentifikatsii, identifikatsii i obnaruzheniya podmeny operatora // Molodoy uchenyy. - 2016. - №9. - S. 167-169.
4. KhitsenkoV.Ye. Neparametricheskaya statistika v zadachakh zashchity informatsii. Konspekt lektsiy, izd-vo NSTU,2012g. 196s.
5. Khollender M., Vulf D.A. Neparametricheskiye metody statistiki. M .: Finansy i statistika, 1983. - 518 s..
6. KhitsenkoV.Ye., Krutokhvostov D.S. Statisticheskaya autentifikatsiya po klaviaturnomu pocherku. Trudy V-oy Mezhregional'naya nauchno-prakticheskay konferentsiya «Informatsionnaya bezopasnost' i zashchita personal'nykh dannykh. Problemy i putiikh resheniya », Bryansk, 2013, S.97-102.
7. KhitsenkoV. E. Increasing the reliability of autentification keyboard handwriting / V. E. Khitcenko, D. S. Krutokhvostov // Актуальные проблемы электронного приборостроения (АПЭП–2014) = Actual problems of electronic instrument engineering (APEIE–2014): тр. 12 междунар.конф.,Новосибирск, 2–4 окт. 2014 г.:в 7 т. – Новосибирск :Изд-во НГТУ, 2014. – Т. 1. – С. 262–265. – 250 экз. – ISBN 978-1-4799-6019-4, ISBN 978-5-7782-2506-0.
8. K. Sung and S. Cho, «GA SVM wrapper ensemble for keystroke dynamics authentication», in Advances in Biometrics, D. Zhang and A. Jain, Eds., vol. 3832, pp. 654–660, Springer, Berlin, Germany, 2005.
9. Messerman, T. Mustafiŕc, S. A. Camtepe, and S. Albayrak, «Continuous and non-intrusive identity verification in realtime environments based on free-text keystroke dynamics», in Proceedings of the International Joint Conference on Biometrics(IJCB '11), pp. 1–8, October 2011.
10. F. Monrose and A. D. Rubin, «Keystroke dynamics as a biometric for authentication», Future Generation Computer Systems, vol. 16, no. 4, pp. 351–359, 2000.
11. D.Gunetti and C. Picardi, «Keystroke analysis of free text», ACM Transactions on Information and System Security, vol. 8, no. 3, pp. 312–347, 2005.
12. T. Shimshon, R.Moskovitch, L. Rokach, andY. Elovici, «Clustering di-graphs for continuously verifying users according to their typing patterns», in Proceedings of the IEEE 26th Convention of Electrical and Electronics Engineers in Israel (IEEEI '10), pp. 445–449, November 2010.
13. M. Pusara, An Examination of User Behavior for User Re-Authentication, Purdue University, West Lafayette, Ind, USA,2007.
14. J. C. Stewart, J. V. Monaco, S.-H. Cha, and C. C. Tappert, «An investigation of keystroke and stylometry traits for authenticating online test takers», in Proceedings of the International Joint Conference on Biometrics (IJCB '11), pp. 1–7, October 2011.
15. E. Yu and S. Cho, «Keystroke dynamics identity verification—its problems and practical solutions», Computers and Security, vol.23, no. 5, pp. 428–440, 2004
16. H. Davoudi and E. Kabir, «A new distance measure for free text keystroke authentication», in Proceedings of the 14th International CSI Computer Conference (CSICC '09), pp. 570–575, October 2009.
17. T. Shimshon, R.Moskovitch, L. Rokach, and Y. Elovici, «Continuous verification using keystroke dynamics», in Proceedings of the International Conference on Computational Intelligence and Security (CIS '10), pp. 411–415, December 2010.
18. Shiryayev A.I. Obnaruzheniye spontanno voznikayushchikh posledstviy // DAN USSR. -1961. T.138, № 4. - S.794-801.
19. Obnaruzheniye izmeneniya svoystv signalov i dinamicheskikh sistem. Pod red. M. Bassvil' i A. Banvenist. M.: «Mir», 1989. – 278 s.