# § 6 КОДИРОВАНИЕ И ЗАЩИТА ИНФОРМАЦИИ

А.Н. Савинов

# АНАЛИЗ ПРИЧИН ВОЗНИКНОВЕНИЯ ОШИБОК ПЕРВОГО И ВТОРОГО РОДА В СИСТЕМАХ АВТОРИЗАЦИИ ОСНОВАННЫХ НА РАСПОЗНАВАНИИ КЛАВИАТУРНОГО ПОЧЕРКА

***Аннотация:****В работе рассматривается метод распознавания клавиатурного почерка, как способ аутентификации или идентификации операторов ключевой системы в процессе авторизации пользователя. Рассматриваются причины возникновения ошибок первого и второго рода в биометрических системах. Описывается понятие порога чувствительности подсистемы контроля доступа. Рассмотрены основные способы снижения вероятности возникновения данных ошибок. Предложена разработка биометрической системы контроля доступа с высокой степенью адаптивности. Предложено использовать самообучающуюся систему сравнения эталоного и текущего почерка. Предложено применять индивидуальные пороги обнаружения.*

***Ключевые слова:*** *биометрия, клавиатурный почерк, ошибки первого рода, ошибки второго рода,распознавание,идентификация, аутентификация, авторизация, порог доступа.*

В настоящее время всеобщей информатизации и автоматизации большое значение приобретают задачи защиты информации. Постоянно разрабатываются новые методы защиты, которые позволяют увеличивать надежность и стойкость систем, предназначенных для решения задач контроля и управления доступа к ключевым системам.

Среди задач защиты информации выделяются вопросы аутентификации (установление подлинности) пользователя ключевой системы. И одними из наиболее перспективных и активно развивающихся сейчас направлений являются методы биометрической аутентификации.

При рассмотрении любых систем принятия решений и (или) распознавания важнейшими показателями качества работы таких систем являются вероятности ошибок системы. Если система предназначена для разделения всех исследуемых объектов на два класса (а именно такое разделение осуществляют системы аутентификации пользователей — они должны разделить на два класса «свой-чужой» всех, кто пытается авторизоваться) то для нее актуальны два вида ошибок. Это так называемые ошибки первого рода, когда система принимает «своего» за «чужого». И ошибки второго рода, когда, наоборот, «чужого» система принимает за «своего».

Ошибки первого рода (англ. type I errors, α errors, false positives) и ошибки второго рода (англ. type II errors, β errors, false negatives) в математической статистике — это ключевые понятия задач проверки статистических ги-

потез. Тем не менее, данные понятия часто используются и в других областях, когда речь идёт о принятии «бинарного» решения (да/нет) на основе некоего критерия (теста, проверки, измерения), который с некоторой вероятностью может давать ложный результат.

Пусть дана выборка $\mathbf{X}=\{\mathbf{X}_1,\ldots,\mathbf{X}_n\}$ из неизвестного совместного распределения $\mathbf{P}^{\mathbf{X}}$, и поставлена бинарная задача проверки статистических гипотез: $\mathbf{H}_0$ — нулевая гипотеза, а $\mathbf{H}_1$ — альтернативная гипотеза. Допустим, что выборка соответствует клавиатурному почерку оператора, проходящего процесс аутентификации. Например, она представлена временем удержания оператором клавиш клавиатуры. Тогда нулевая гипотеза $\mathbf{H}_0$ будет соответствовать предположению, что аутентифицируемый пользователь действительно является зарегистрированным пользователем системы (именно тем, кем он представился системе) и его можно авторизовать. Альтернативная гипотеза $\mathbf{H}_1$ будет иметь противоположное значение: аутентифицируемый пользователь не является законным пользователем системы и должен получить отказ в авторизации.

Предположим, что задан статистический критерий (1) сопоставляющий каждой реализации выборки $\mathbf{X} = \mathbf{x}$ одну из имеющихся гипотез.

$$f : \mathbf{R}^n \rightarrow \{\mathbf{H}_0,\mathbf{H}_1\}, \quad (1)$$

Для примера с клавиатурным почерком в качестве статистического критерия возьмем меру Евклида (2) и стандартная непохожесть (порог чувствительности) **MaxN** подсистемы принятия решений, определяющая допустимое отклонение клавиатурного почерка от эталона, для принятия решения о том, что почерк тестируемого пользователя совпадает с эталонным, хранимым в базе.

s:=s+sqr((Times[i,0]-eTimes[i,0])/eTimes[i,0]), (2)

Здесь **s** – значение меры Евклида [6]. **Times[i,0]** – время удержания конкретной клавиши из выборки, соответствующей клавиатурному почерку тестируемого пользователя. **eTimes[i,0]** – время удержания конкретной клавиши, хранимое в эталонном образце клавиатурного почерка тестируемого пользователя. Согласно применению данного критерия возможны 2 случая:

Если **s < MaxN**, то отклонение характеристик почерка текущего оператора ключевой системы соответствует разрешенному диапазону. В этом случае принимается решение о том, что пользователь является законным и происходит процесс авторизации.

Если **s > MaxN**, то отклонение характеристик почерка текущего оператора ключевой системы не соответствует разрешенному диапазону. Значит, принимается решение о том, что пользователь не является законным и он получает отказ в авторизации.

Возможны следующие четыре ситуации (Таблица 1.):

1. Распределение $\mathbf{P}^{\mathbf{X}}$ выборки соответствует гипотезе $\mathbf{H}_0$, и она точно определена статистическим критерием, то есть $f(\mathbf{x}) = \mathbf{H}_0$. Значит, клавиатурный почерк пользователя совпадает с его эталонным почерком по заданному критерию. Пользователь является законным и успешно проходит авторизацию.
2. Распределение $\mathbf{P}^{\mathbf{X}}$ выборки соответствует гипотезе $\mathbf{H}_0$, но она неверно отвергнута статистическим критерием, то есть $f(\mathbf{x}) = \mathbf{H}_1$. Значит, клавиатурный почерк пользователя не совпадает с его эталонным почерком по заданному критерию. Пользователь является законным, но

система ошибочно принимает решение об отказе в авторизации.

3. Распределение $\mathbf{P^X}$ выборки соответствует гипотезе $\mathbf{H_1}$, и она точно определена статистическим критерием, то есть $\boldsymbol{f}(\mathbf{x}) = \mathbf{H_1}$. Значит, клавиатурный почерк пользователя не совпадает с его эталонным почерком по заданному критерию. Пользователь не является зарегистрированным пользователем системы и справедливо получает отказ в авторизации.
4. Распределение $\mathbf{P^X}$ выборки соответствует гипотезе $\mathbf{H_1}$, но она неверно отвергнута статистическим критерием, то есть $\boldsymbol{f}(\mathbf{x}) = \mathbf{H_0}$. Значит, клавиатурный почерк пользователя не совпадает с его эталонным почерком по заданному критерию. Пользователь не является законным, но ошибочно получает разрешение на авторизацию.

Во втором и четвертом случае говорят, что произошла статистическая ошибка, и её называют ошибкой первого и второго рода соответственно.

*Таблица 1. Возможные варианты исходов при применении гипотезы.*

| | | Верная гипотеза | |
|---|---|---|---|
| | | $\mathbf{H_0}$ (предположение о том, что аутентифицирующийся оператор является зарегистрированным пользователем системы) | $\mathbf{H_1}$ (предположение о том, что аутентифицирующийся оператор не является зарегистрированным пользователем системы) |
| Результат применения критерия | $\mathbf{H_0}$ (оператор проходит процесс авторизации успешно) | $\mathbf{H_0}$ верно принята (оператор является зарегистрированным пользователем системы и правомерно получает разрешение на авторизацию) | $\mathbf{H_0}$ неверно принята **(Ошибка второго рода)** (оператор не является зарегистрированным пользователем системы, но ошибочно получает разрешение на авторизацию) |
| | $\mathbf{H_1}$ (оператор получает отказ в авторизации) | $\mathbf{H_0}$ неверно отвергнута **(Ошибка первого рода)** (оператор является зарегистрированным пользователем системы, но ошибочно получает отказ в авторизации) | $\mathbf{H_0}$ верно отвергнута (оператор не является зарегистрированным пользователем системы справедливо получает отказ в авторизации) |

Пусть количество объектов в тестовом наборе равно **N**, из них **Np** – кол-во «положительных» (с меткой ‘1’) объектов, а **Nn** – кол-во объектов «отрицательных» (с меткой ‘-1’). Естественно, **N=Np+Nn**. Пусть количество

ложных пропусков **FN**, а ложных обнаружений **FP**, тогда несложно подсчитать количество верных пропусков (3) и верных обнаружений (true negatives, true positives) (4).

$$TP = Np - FN, \quad (3)$$

$$TN = Nn - FP, \quad (4)$$

Используя эти величины можно рассчитать нормированные уровни ошибок первого и второго рода (5), а также долю верно распознаваемых пропусков и обнаружений (6).

$$nFN=FN / Np * 100\%; \quad nFP = FP / Nn * 100\%, \quad (5)$$

$$nTN=TN / Nn * 100\%; \quad nTP = TP / Np * 100\%, \quad (6)$$

Такие величины более наглядно, в виде частоты (в процентах) встречаемости ошибок и верных обнаружений, отражают качество распознавания, поскольку не зависят (в явном виде) от количества объектов в тестовом наборе.

В биометрических системах выделяют следующие виды ошибок:

- **FAR** (False Acceptance Rate) – частота ложных приемов. Например, если из 100 проб входа в систему злоумышленником может произойти одна случайная идентификация его с законным пользователем, то **FAR**=0,01, что, в общем-то, многовато для статических (физиологических) систем и нормально для динамических (поведенческих);
- **FRR** (False Rejection Rate) – частота ложных отказов. Например, если на 100 аутентификаций, выполненных законным пользователем, произошло два неправомерных отказа, то **FRR**=0,02;
- **EER**, или **ERR** (Equal Error Rate, или ERror Rate) – частота ошибок. Это сложное понятие, которое формируется в связи с тем, что биометрическую систему обычно можно настраивать, варьировать ее параметры. Так вот **FAR** и **FRR** связаны между собой, и когда один показатель уменьшается, второй обязательно увеличивается. Если при каких-то настройках **FAR**=**FRR**, то это и есть значение **ERR**.

Если администратор системы занижает порог отказа в допуске, то система будет более «снисходительно» оценивать совпадение хранимого шаблона с данными пользователя, и, естественно, увеличится вероятность, что она по ошибке разрешит вход постороннему. Устанавливая порог слишком высоко, увеличивается вероятность того, что система будет отвергать вполне легитимных пользователей. С одной стороны, высокое значение **FRR** (вероятность ошибочного задержания «своего») может привести к дискредитации системы и снижению эффективности ее функционирования, так как при частых ложных срабатываниях персонал охраны практически перестает обращать внимание на задержания или отказы в доступе. С другой стороны, высокое значение **FAR** (вероятность ошибочного пропуска «чужого») увеличивает вероятность несанкционированного доступа. Учитывая зависимость **FAR**, **FRR** от установленных порогов обнаружения **A** (рисунок 1), следует отметить, что задача выбора порогов для администратора системы безопасности объекта чрезвычайно актуальна.

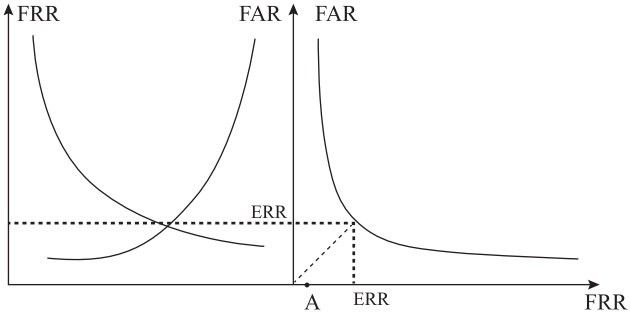


*Рисунок 1.*
*График зависимостей **FRR** и **FAR** от порога обнаружения*

Предложена разработка биометрической системы контроля доступа с высокой степенью адаптивности. Предложено обязатель-

ное наличие возможности изменения порогов обнаружения. Принцип контролируемости обеспечит в первую очередь наличием встроенных средств расчета **FAR** и **FRR**. При этом для расчета **FAR** целесообразно использовать предложенный аппарат сравнения методом «чужой» к «чужому» хранящихся в базе данных «эталонов» при вариации порога обнаружения. Предположим, что в базе хранятся **n** эталонов соответственно **n** операторов ключевой системы. Первый эталон клавиатурного почерка пользователя, хранящийся в базе данных, сравнивается со всеми остальными **n-1** эталонами клавиатурных почерков пользователей из этой же базы. Соответственно для первого эталона происходит **n-1** сравнений. Второй эталон уже сравнивался с первым эталоном. Значит, сравнение начнется с третьего эталона и всего для него произойдет **n-2** сравнений. Указанная процедура осуществляется до предпоследнего эталона базы. Это значит, что число возможных сравнений **VFAR** "чужой" к "чужому" в базе из **n** эталонов будет (9):

$$V\,FAR = \frac{n(n-1)}{2}, \qquad (9)$$

Для оценки **FRR** целесообразно использовать отношение количества отказов в доступе по критерию «биометрический контроль не пройден» к общему количеству попыток предъявления биометрических параметров (в упрощенном случае – к общему числу проходов). Таким образом, будет получена самообучающаяся система, которая устанавливает порог чувствительности в зависимости от вариаций хранящихся в базе почерков. Также система, после анализа почерков сотрудников-операторов ключевой системы, предложит администратору рекомендации по допустимым значениям порога с указанием вероятностей **FRR** и **FAR** при отклонении значения порога от нормальной величины. Естественно при первом запуске системы данные для определения **FRR** указанным методом или сравнением «свой» к «своему» (несколько образцов клавиатурного почерка одного и того же пользователя) в системе по понятным причинам отсутствуют. Таким образом, предлагается считать значение **FAR** обязательным для реализации системным параметром. А параметр **FRR** будем считать допустимым, если он соизмерим с вероятностью ошибки ложного срабатывания для систем контроля и управления доступом (СКУД), не имеющей биометрического контроля.

Фактически **FRR** определяется интенсивностью процессов авторизации операторов ключевой системы: если их мало, то **FRR** может быть относительно большим, а если много, то должен быть малым. Соответственно чем важнее статус охраняемой системы, тем меньше допущенных лиц (меньшее количество авторизаций) и, следовательно, может быть задано более высокое значение **FRR** при уменьшении **FAR**.

Предложено проводить оценку **FRR** дополнительно для каждого пользователя с целью выявления людей с явно выраженным влиянием субъективного фактора. В системе могут применяться наряду с общесистемными порогами обнаружения индивидуальные. Данные пороги могут действовать выборочно для определенных людей и задаваться администратором, а могут варьироваться (в требуемом диапазоне) автоматически для всего персонала в зависимости от индивидуальной **FRR**. Вариация порогов оказывается также полезной на этапе адаптации персонала к биометрической системе. При начальном вводе биометрических данных необходимо уделять внимание их качеству (в том числе и тому факту, что не все люди имеют хорошо выраженный клавиатурный

почерк), а, следовательно, в системе должны быть предусмотрены программные средства оценки качества эталона клавиатурного почерка и качества самого клавиатурного почерка. Следует отметить принципиальную важность автоматической подстройки хранящихся в базе данных эталонов с целью компенсации изменений биометрических параметров со временем ведь клавиатурный почерк – изменяющаяся со временем поведенческая биометрическая характеристика человека. Соответственно с изменением почерка увеличится количество ошибок **FRR**.

Таким образом, проведен анализ причин и последствий возникновения ошибок первого и второго рода в биометрических системах. Предложены способы снижения количества ошибок **FAR** и **FRR** в системах распознавания клавиатурного почерка, основанные на характеристиках и свойствах клавиатурного почерка операторов ключевых систем. Повышение качества функционирования подсистемы принятия решений о классификации авторизирующегося пользователя увеличивает интерес администраторов и служб, обеспечивающих безопасность ключевых систем, к подобного рода продуктам.

**Библиография:**


1. Иванов А.И. Нейросетевые алгоритмы биометрической идентификации личности.— Серия «Нейрокомпьютеры и их применение». Кн. 15. — М.: Радиотехника, 2004. – с. 22-50.
2. Рыбченко Д.Е., Критерии устойчивости и индивидуальности компьютерного почерка при вводе ключевых фраз.— Специальная техника средств связи. Серия «Системы, сети и технические средства конфиденциальной связи».— Пенза, ПНИЭИ, 1997, вып.№2 – с. 104-107.
3. Обзор технологий биометрической идентификации – 16.11.03. http://center.forever.kz/hard/other/f0003.htm
4. Ю.А. Брюхомицкий, М.Н. Казарин. Учебные биометрические системы контроля доступа по рукописному и клавиатурному почеркам.— Таганрог, ТРГУ, 2004 ( http://www.library.mephi.ru/data/scientific-sessions/2006/vnk13/0-1-12.doc)
5. Владимир Вежневец. Оценка качества работы классификаторов. Компьютерная графика и мультимедиа. Выпуск №4(1)/2006. http://cgm.computergraphics.ru/content/view/106
6. Методические указания по математическому анализу. Ч. 2. Курс лекций по математическому анализу (для студентов 2-го курса). Ч. 2. МФТИ. М., 2005. 213 с.
7. http://wapedia.mobi/ru/Мощность_критерия


**References (transliteration):**


1. Ivanov A.I. Neyrosetevye algoritmy biometricheskoy identifikatsii lichnosti.— Seriya «Neyrokomp'yutery i ikh primenenie». Kn. 15. — M.: Radiotekhnika, 2004. – s. 22-50.
2. Rybchenko D.E., Kriterii ustoychivosti i individual'nosti komp'yuternogo pocherka pri vvode klyuchevykh fraz.— Spetsial'naya tekhnika sredstv svyazi. Seriya «Sistemy, seti i tekhnicheskie sredstva konfidentsial'noy svyazi».— Penza, PNIEI, 1997, vyp.№2 – s. 104-107.
3. Obzor tekhnologiy biometricheskoy identifikatsii – 16.11.03. http://center.forever.kz/hard/other/f0003.htm

4. Yu.A. Bryukhomitskiy, M.N. Kazarin. Uchebnye biometricheskie sistemy kontrolya dostupa po rukopisnomu i klaviaturnomu pocherkam.— Taganrog, TRGU, 2004 ( http://www.library.mephi.ru/data/scientific-sessions/2006/vnk13/0-1-12.doc)
5. Vladimir Vezhnevets. Otsenka kachestva raboty klassifikatorov. Komp'yuternaya grafika i mul'timedia. Vypusk №4 (1)/2006. http://cgm.computergraphics.ru/content/view/106
6. Metodicheskie ukazaniya po matematicheskomu analizu. Ch. 2. Kurs lektsiy po matematicheskomu analizu (dlya studentov 2-go kursa). Ch. 2. MFTI. M., 2005. 213 s.
7. http://wapedia.mobi/ru/Moshchnost'_kriteriya

 