Труды научно-технической конференции кластера пензенских предприятий,
обеспечивающих **БЕЗОПАСНОСТЬ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ.**
Том 9, Пенза-2014, с. 7-8 (http://www.pniei.penza.ru/RV-conf/T9/C7)

ГИПОТЕЗА χ^2 РАСПРЕДЕЛЕНИЯ РАССТОЯНИЙ ХЭММИНГА ДЛЯ КОДОВ БИОМЕТРИЧЕСКОЙ АУТЕНТИФИКАЦИИ ПРИМЕРОВ ОБРАЗА «СВОЙ»

Фунтикова Ю.В., Иванов А.И., Захаров О.С. (г. Пенза)
Пензенский государственный университет, ОАО «ПНИЭИ»

В настоящее время активно идут процессы информатизации современного общества, создаются порталы электронных правительств и иных значимых для граждан информационных интернет-ресурсов. Для повышения доверия к интернет-ресурсам необходимо создавать электронные интернет-паспорта и удостоверения личности, безопасно используемые в открытых информационных пространствах при биометрико-криптографической аутентификации личности. В США и странах Евросоюза для этой цели используют «нечеткие экстракторы» [1], Россия развивает технологию нейросетевого преобразования биометрических данных личности в код аутентификации [2].

Как для российской технологии, так и для технологии западных стран важно иметь достаточно точную аналитическую модель описания операций преобразования биометрия-код. Однослойный нейросетевой преобразователь биометрия-код описывается двумя нелинейными матричными уравнениями:

(1) (2)

$$\overline{L}\left\{\begin{bmatrix}\mu_{1,1} & \mu_{1,2} & \cdots & \mu_{1,N}\\ \mu_{1,2} & \mu_{2,2} & \cdots & \mu_{2,N}\\ \cdots & \cdots & \cdots & \cdots\\ \mu_{1,n} & \mu_{2,n} & \cdots & \mu_{n,N}\end{bmatrix}\times\begin{bmatrix}v_1\\ v_2\\ v_3\\ \cdots\\ v_N\end{bmatrix}\right\}=\begin{bmatrix}"c_1"\\ "c_2"\\ \cdots\\ "c_n"\end{bmatrix};\quad \overline{L}\left\{\begin{bmatrix}\mu_{1,1} & \mu_{1,2} & \cdots & \mu_{1,N}\\ \mu_{1,2} & \mu_{2,2} & \cdots & \mu_{2,N}\\ \cdots & \cdots & \cdots & \cdots\\ \mu_{1,n} & \mu_{2,n} & \cdots & \mu_{n,N}\end{bmatrix}\times\begin{bmatrix}\xi_1\\ \xi_2\\ \xi_3\\ \cdots\\ \xi_N\end{bmatrix}\right\}=\begin{bmatrix}"x_1"\\ "x_2"\\ \cdots\\ "x_n"\end{bmatrix},$$

где $\mu_{i,j}$ - весовые коэффициенты сумматора i-го нейрона, преобразующего вектор входных непрерывных биометрических параметров в дискретные состояния; v_j – входной биометрический параметр образа «Свой»; ξ_j – входной биометрический параметр образа «Чужой»; $"c_i"$ - разряд выходного кода «Свой»; $"x_i"$ - разряд выходного кода «Чужой»; $\overline{L}\{:\}$ - вектор операторов квантования выходного сигнала сумматоров нейронов сети.

Необходимость описания одного и того же объекта двумя похожими нелинейными уравнениями обусловлена тем, что исследуемый объект ведет себя для данных образа «Свой» и данных образа «Чужой» совершенно по-разному. Неопределенность примеров данных образа «Свой» нейросетевой преобразователь сводит к нулю, откликаясь на каждый k-тый пример вектора биометрических параметров - $\overline{v}_k$ одним и тем же вектором дискретных состояний $"\overline{c}"$. Фактически энтропия входных биометрических данных полностью уничтожается нейросетевым преобразователем биометрия-код.

Совершенно иная ситуация возникает, когда нейросетевому преобразователю биометрических данных в код предъявляются вектора биометрических параметров образа «Чужой» - $\overline{\xi}_k$. В этом случае каждый из k-тых примеров одного и того же биометрического образа «Чужой» дает разные вектора дискретных выходных состояний кода нейросети -$"\overline{x}_k"$. Для данных биометрического образа «Чужой» происходит

хэширование (перемешивание) кодов во время нейросетевого преобразования. Возникает эффект усиления входной энтропии биометрического образа «Чужой».

Идеальный нейросетевой преобразователь биометрия-код должен обеспечивать нулевую энтропию вектора состояний кода "$\overline{c}$" и достаточно большую энтропию вектора состояний кодов "$\overline{x}_k$". Формально это может быть записано следующим образом:

$$H("\overline{c}") = 0.0 \qquad (3); \qquad H("\overline{x}_k") \approx n = 256 \qquad (4),$$

где n- длина выходного кода аутентификации.

Для реального преобразователя биометрия-код условия (3) и (4) выполнены быть не могут в силу того, что вероятность ошибок первого рода (отказ в аутентификации «Своему») не является нулевой:

$$P_1 \neq 0.0 \qquad (5).$$

Кроме того, вероятность ошибок второго рода (вероятность коллизии образов «Свой» и «Чужой») всегда намного больше вероятности ошибки последующей криптографической аутентификации:

$$P_2 \rangle\rangle P_A = 2^{-n} = 2^{-256} \qquad (6).$$

Следствием условия (5) является то, что среднее значение модулей коэффициентов корреляции выходных разрядов кода «Свой» оказывается чуть меньше единицы:

$$E(\left|r("c_i","c_j")\right|) \approx 0.99...\langle 1.0 \qquad (7);$$

А следствием условия (6) - среднее значение модулей корреляции разрядов кодов «Чужой» всегда больше нуля:

$$E(\left|r("x_i","x_j")\right|) \approx 0.05...\rangle 0.0 \qquad (8).$$

Среднее значение модулей коэффициентов корреляции между разрядами биокодов играет значительную роль. При описании статистических свойств преобразователей биометрия-код нельзя пренебрегать показателями корреляционных связей или реальным значением энтропии.

Простейшим способом учета корреляционных связей является приведение биометрических данных к «равнокоррелированным». Если умножить независимые данные (полученные от генератора случайных чисел) на симметричную связывающую матрицу [a] с единичной диагональю, то данные становятся одинаково коррелированными [2]:

$$\begin{bmatrix} 1 & a & \cdots & a \\ a & 1 & \cdots & a \\ \cdots & \cdots & \cdots & \cdots \\ a & a & \cdots & 1 \end{bmatrix} \times \begin{bmatrix} \xi_{1,i} \\ \xi_{2,i} \\ \cdots \\ \xi_{n,i} \end{bmatrix} = \begin{bmatrix} y_{1,i} \\ y_{2,i} \\ \cdots \\ y_{n,i} \end{bmatrix} \quad \Rightarrow \quad R = \begin{bmatrix} 1 & r & \cdots & r \\ r & 1 & \cdots & r \\ \cdots & \cdots & \cdots & \cdots \\ r & r & \cdots & 1 \end{bmatrix} \qquad (9).$$

Этот технический прием удобен тем, что для получения нужных статистических параметров описания объекта исследования достаточно подобрать значение всего одного регулируемого параметра – a. Параметр - a следует изменять до того момента, пока статистические данные модели не совпадут со статистическими данными исследуемого преобразователя биометрия-код. На сегодняшний день в качестве статистических данных стало общепринятым использование расстояний Хэмминга [1, 2] между исследуемыми кодами и кодом «Свой». Типичное распределение расстояний Хэмминга приведено на рисунке 1.

Из рисунка 1 видно, что корректно построенный преобразователь биометрия-код имеет нормальный закон распределения расстояний Хэмминга между кодом «Свой» и кодами «Чужие». Все преобразователи биометрия-код описываются биномиальным законом распределения, который при большом числе степеней свободы хорошо нормализуется [2].

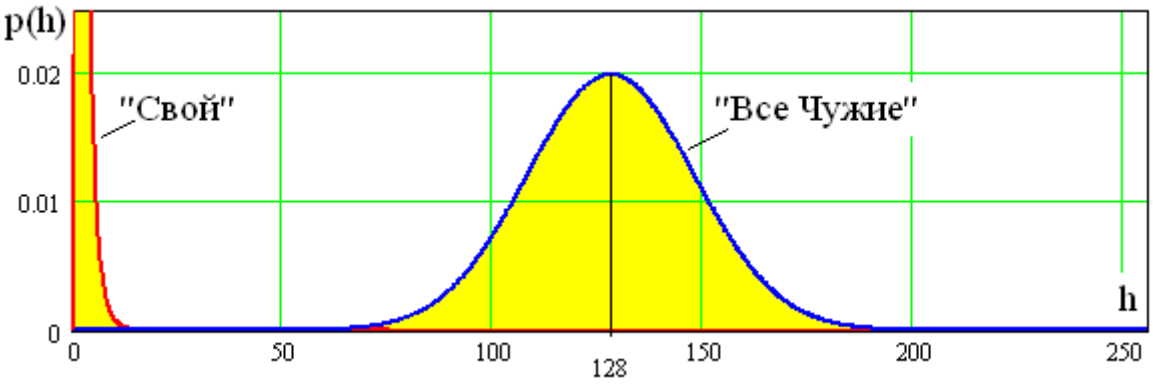

Рис. 1. Распределения расстояний Хэмминга между кодом «Свой» и иными выходными кодами длиной 256 бит

Распределение расстояний Хэмминга между примерами образа «Свой» так же описывается биномиальным распределением зависимых данных. Из практики известно, что при малом числе степеней свободы биномиальное распределение хорошо приближается хи-квадрат распределением [3].

Предположим, что в результате тестирования мы получили значение математического ожидания расстояний Хэмминга $E(h) = 3.5$ для кодов «Свой». То есть остаточное число степеней свободы нейросетевого преобразователя составит m=3.5. Для того чтобы получить распределение $\chi^2(m=3.5, r \neq 0.0)$ зависимых биометрических данных, воспользуемся моделью (9) и получим плотности распределения χ^2 для трех степеней свободы и разного уровня коррелированности. Соответствующие кривые приведены в левой части рисунка 2.

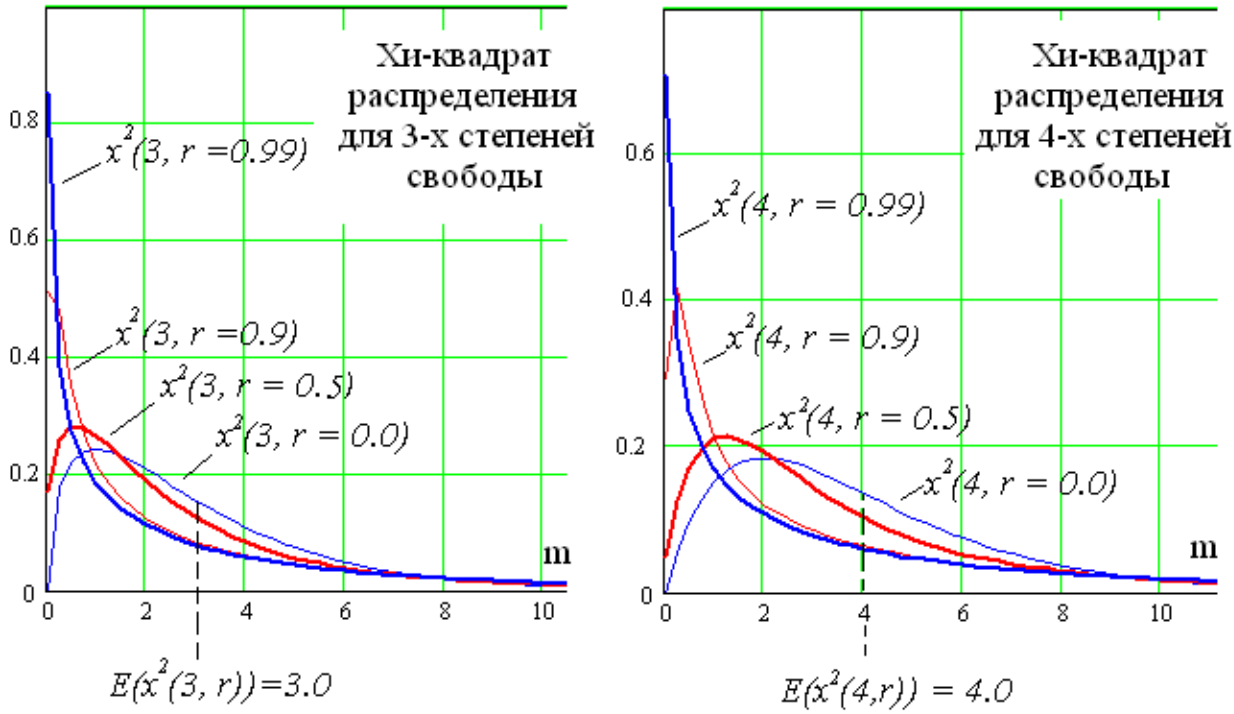

Рис. 2. Кривые хи-квадрат распределений для трех и четырех степеней свободы при разных значениях коррелированности биометрических данных

Особенностью всех χ^2 распределений (как зависимых, так и независимых) является то, что их математическое ожидание всегда совпадает с числом степеней свободы. Пользуясь этим свойством, можно перейти от показателей распределений хи-квадрат с

целым числом степеней свободы к показателям с дробным числом степеней свободы. Для этого необходимо в нужной пропорции сложить плотности распределения ближайших целых степеней свободы. В частности для получения распределения с числом степеней свободы m = 3.5 для r = 0.99 следует усреднить. Соответствующие распределения для 3 и 4 степеней свободы:

$$\chi^2(3.5, r) = \frac{1}{2}\chi^2(3, r) + \frac{1}{2}\chi^2(4, r) \qquad (10).$$

Весовые в уравнении (10) совпадают с расстояниями до ближайших целых значений степеней свободы зависимых хи-квадрат распределений.

Очевидно, что линейной комбинацией хи-квадрат распределений целых степеней свободы можно получить значение хи-квадрат распределения для любого дробного числа степеней свободы, зависимых данных.

ВЫВОД

Технический прием перехода к «равнокоррелированным» данным позволяет без особых вычислительных трудностей переходить к вычислению хи-квадрат распределений любой коррелированности и любого дробного значения показателя числа степеней свободы. Все это позволяет надеяться, что тестирование высокодоступных нейросетевых преобразователей биометрия-код можно будет осуществлять на достаточно малых выборках примеров образа «Свой». Если процедуры оценки вероятностей ошибок второго рода, рекомендуемые ГОСТ Р 52633.3-2011, позволяют сократить объем данных в миллионы раз, то аналогичный выигрыш может быть получен и при оценке вероятностей ошибок первого рода в случае подтверждения гипотезы хи-квадрат распределения расстояний Хэмминга для кодов «Свой».

Литература:

1. Feng Hao, Ross Anderson, and John Daugman. Crypto with Biometrics Effectively, IEEE TRANSACTIONS ON COMPUTERS, VOL. 55, NO. 9, SEPTEMBER 2006.
2. Язов Ю.К. и др. Нейросетевая защита персональных биометрических данных. М.: Радиотехника. 2012 г., 160 с.
3. Кобзарь А.И. Прикладная математическая статистика для инженеров и научных работников. М. ФИЗМАТЛИТ, 2006 г., 816 с.

Статья поступила 03.04.2013 года, публикуется по положительной рецензии д.т.н. проф. Малыгина А.Ю.