УДК 004:7.056:519.71 *МРНТИ 50.37.23, 27.47.15*

А. И. Иванов, д.т.н., ***А. Ю. Малыгин,*** д.т.н.,
Д. Н. Надев, к.т.н., ***К. Т. Сауанова****, к.т.н.

Пензенский государственный университет
Алматинский университет энергетики и связи*

МОДИФИКАЦИЯ КЛАССИЧЕСКОГО БИНОМИАЛЬНОГО ЗАКОНА РАСПРЕДЕЛЕНИЯ ЗНАЧЕНИЙ, УЧИТЫВАЮЩАЯ СУЩЕСТВЕННУЮ ЗАВИСИМОСТЬ БИОМЕТРИЧЕСКИХ КОДОВ[1]

В работе показано, что аппроксимацию биномиального зависимого закона можно получить, используя композицию нормального, равномерного, арксинусного распределения значений. Полученная аппроксимация может использоваться при симметрии распределения меры Хэмминга выходного кора преобразователя биометрия-код относительно математического ожидания. Ошибка аппроксимации составляет 0,03 % от 256 бит выходного кода преобразователя биометрия-код.
Ключевые слова: биометрия, преобразователь биометрия-код, энтропия зависимых данных, энтропия распределения расстояний меры Хэмминга.

Берілген жұмыста мәндердің нормальды, біркелкі, арксинусты таралу композициясын қолдану арқылы биномиальді тәуелді аппроксимация заңын алуға болатыны көрсетілген. Алынған аппроксимацияны математикалық күтумен салыстырмалы алғанда биометрия-код өңдеушісін Хэммингтің шығыс қабықшасының симметриялық таралу шамасы негізінде қолдануы мүмкін. Аппроксимация қателігі биометрия-код өңдеушісінің шығыс кодындағы 256 биттің 0,03 % құрайды.
Түйінді сөздер: биометрия, биометрия-код өңдеуші, тәуелді деректердің энтропиясы, таралу қашықтығындағы Хэмминг өлшемнің энтропиясы.

[1]Статья подготовлена в рамках выполнения комплексного проекта «Разработка и подготовка производства телекоммуникационного оборудования, разработка программного сетевого, прикладного и специального обеспечения для создания цифровых сетей связи с персонализированным доступом» в соответствии с Постановлением Правительства № 218 от 09.04.2010 г.

The work shows that approximation of the binomial dependent law can be obtained using a composition of normal uniform arc sin distributions of values. The received approximation can be used at symmetry of distribution of the Hamming distance of the outgoing code of the biometry-code converter non-dimensionalized to mathematical expectation. The approximation error makes 0.03 % from 256 bits of the outgoing code of the biometry-code converter.
Key words: a biometry, biometry-cod converter, entropy of the dependent data, entropy of distribution of distances of the Hamming measure.

В настоящее время защиту информации в открытых системах выполняют с помощью биометрико-криптографических механизмов аутентификации пользователя. Зарубежные исследователи получают код аутентификации из входного биометрического образа [1,2]. В соответствии с российским ГОСТ Р 52633.0-2006 [3] средства высоконадежной биометрической аутентификации рекомендуется строить на основе нейросетевых преобразователей биометрия-код. Такие преобразователи смешивают биометрический образ с кодом аутентификации пользователя. Обучение преобразователей биометрия-код производится по требованиям, изложенным в ГОСТ Р. Входными данными обучения являются параметры биометрических образов пользователей «Свой» и «Все Чужие». На выходах обученного преобразователя биометрия-код формируется код-отклик пользователя «Свой» при подаче на входы преобразователя образов «Свой». При подаче на входы преобразователя образов «Все Чужие» на его выходах формируются случайные коды-отклики «Все Чужие». Вероятность ложной аутентификации преобразователя с 256 выходами и 416 входами составляет 10^{-9} [4]. Методики тестирования средств высоконадежной биометрической аутентификации выполняются исходя из оценки статистического распределения вероятностей ложной аутентификации пользователей «Свой» и «Все Чужие» [5].

В работе [6] показано, что на практике выходные данные нейросетевых преобразователей биометрия-код являются коррелированными. Одна из причин - наличие корреляционных связей между входными параметрами преобразователя. Другая причина состоит в структурной организации связей между ней-

ронами преобразователя. Моделирование корреляционных связей выходных данных биометрико-нейросетевого преобразователя выполнено в [7]. Показано, что распределение вероятностей ошибок ложной аутентификации пользователей «Все Чужие» может быть описано биномиальным зависимым законом распределения значений. Вероятности ложной аутентификации пользователя «Свой» оцениваются через дробный показатель степеней свободы закона распределения значений хи-квадрат [8].

Рост корреляционных связей между разрядами выходных кодов приводит к росту среднеквадратического отклонения значений расстояний Хэмминга между кодом «Свой» и случайными выходными кодами «Чужие». Монотонный рост корреляции между выходными разрядами кодов «Чужие» приводит к соответствующему росту среднеквадратического отклонения расстояний Хэмминга. Соответствующие распределения расстояний Хэмминга приведены на рис. 1.

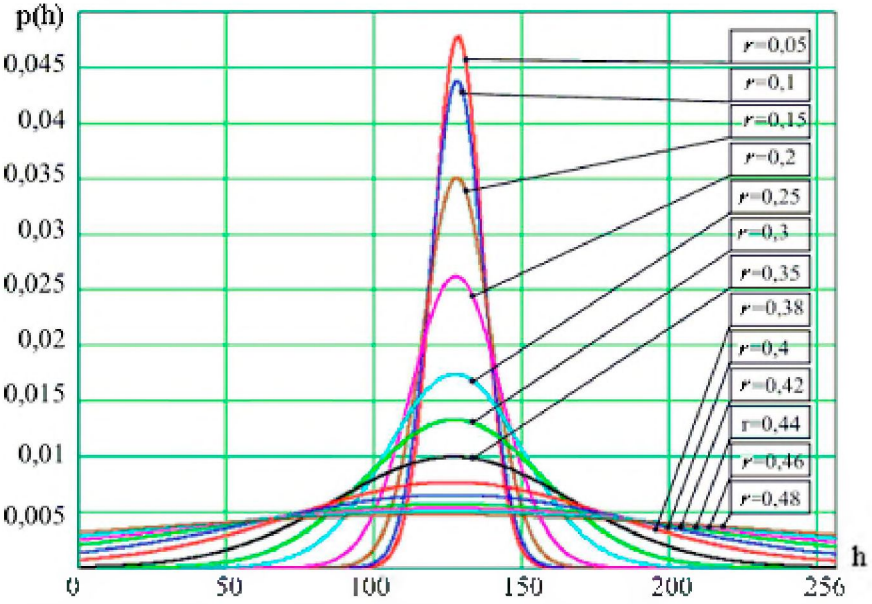

Рис. 1. Эволюция плотности распределения значений меры Хэмминга при монотонном росте модуля корреляционных связей между выходными разрядами кодов преобразователя

Получается, что при монотонном росте модулей корреляционной связи в интервале от r=0.0 до r=0.63 форма закона распределения значений меры Хэмминга эволюционирует, проходя 3 разных закона распределения: нормальный закон, равномерный закон и закон арксинуса (рис. 2).

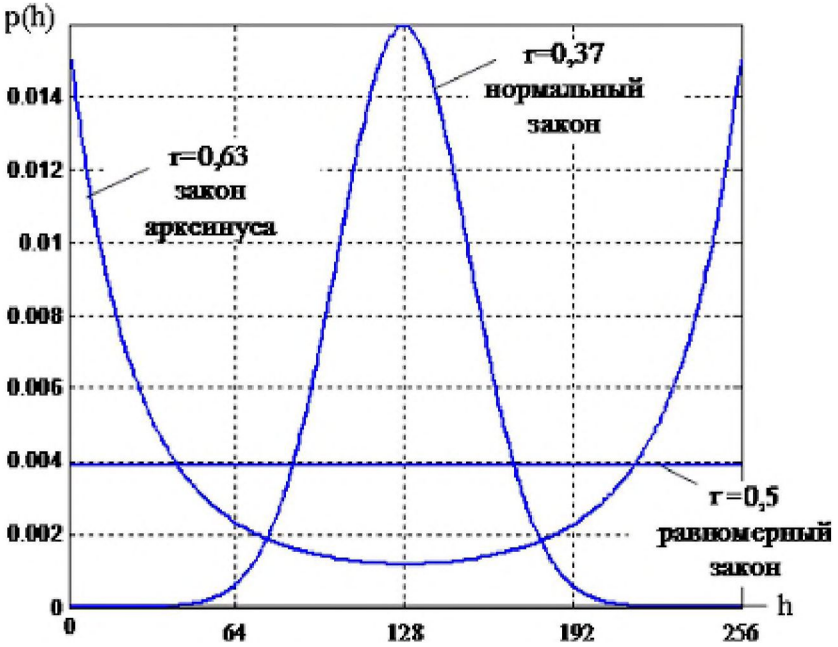

Рис. 2. Формы трех классических законов, являющихся этапами эволюции биномиального закона распределения значений зависимых данных

К сожалению, на сегодня отсутствует точное аналитическое описание эволюции плотности распределения значений меры Хэмминга соответствующей плотности распределения значений биномиального зависимого закона. На сегодняшний день имеются только таблицы аппроксимации этих плотностей, построенные для разных значений коэффициентов корреляции или среднеквадратического отклонения. Аппроксимировать плотно-

сти распределения значений биномиального зависимого закона удобно в нормированной метрике Хэмминга, изменяющейся в пределах от 0 до 1. Тогда приближающее распределение выражение записывается следующим образом для интервала модулей корреляции от 0,0 до 0,37:

$$p(h, \mathrm{E(h)}, \sigma(\mathrm{h})) = \frac{1}{\sqrt{2\pi}\cdot\sigma(h)} \exp\left\{\frac{-(E(h)-h)^2}{2\cdot\{\sigma(h)\}^2}\right\} \qquad (1)$$

где $p(h, E(h), \sigma(h))$ – псевдонепрерывная плотность распределения значений меры Хэмминга – h, которая на самом деле является огибающей дискретного распределения;

$\sigma(h)$ – среднеквадратическое отклонение нормированной меры Хэмминга;

$E(h)$ - математическое ожидание значений нормированной меры Хэмминга.

В интервале корреляции разрядов 256-битных кодов от 0.35 до 0.5 распределение расстояний Хэмминга хорошо описывается следующим приближением:

$$p(h) \approx \mathrm{a(r)}\cdot\frac{1}{257} + \frac{(1-a(r))\cdot\exp\left(\frac{-(128-h)^2}{2\sigma^2(h)\cdot(1-a^2(r))}\right)}{\sqrt{2\pi}\cdot\sigma(h)\cdot(1-a(r))} \qquad (2)$$

где . $a(r) = \frac{r-0.35}{0.15}$ (3)

Эволюция плотности распределения значений меры Хэмминга при монотонном росте модуля корреляционных связей в пределах от r = 0.5 до r = 0.92 приведена на рис. 3.

В интервале коррелированности разрядов 256-битных кодов от r = 0.5 до r = 0.87 распределение расстояний Хэмминга хорошо описывается следующим приближением:

$$p(h) \approx \mathrm{a(r)}\cdot\frac{1}{257} + \frac{1-a(r)}{\pi\sqrt{h(1-h)}} \qquad (4)$$

где $a(r) = \frac{r-0.87}{0.37}$ (5)

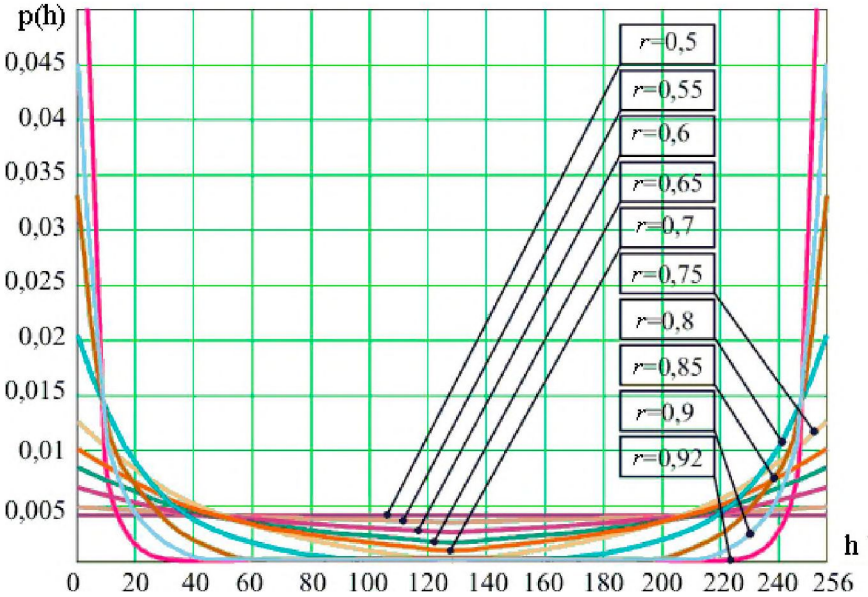

Рис. 3. Эволюция плотности распределения значений меры Хэмминга при монотонном росте модуля корреляционных связей в пределах от r=0.5 до r=0.92

При изменении модулей коэффициентов корреляции в интервале от 0,92 до 1,0 хорошая аппроксимация плотности распределения значений биномиального зависимого закона получается при приближении правой и левой ветви хи-квадрат распределениями, имеющими малые значения дробных показателей числа степеней свободы [8]:

$$p(h, F(h), \sigma(h)) = \frac{1}{2^{\frac{m}{2}+1}\Gamma\left(\frac{m}{2}\right)}\left\{h^{\frac{m}{2}-1}\cdot\exp\left(\frac{-h^2}{2}\right) + |1-h|^{\frac{m}{2}-1}\cdot\exp\left(\frac{-|1-h|}{2}\right)\right\} \quad (6)$$

где m - малое значение дробного показателя числа степеней свободы (число m может быть менее 1,0, но всегда больше 0,0), подбираемое при аппроксимации плотности распределения;

Г(.) - гамма функция.

По выражениям (1)-(6) может быть оценена вероятность ложной аутентификации "Чужого" обученного нейросетевого преобразователя биометрия-код с 256 выходами сбалансированного по вероятности появления "0" в разряде кода. Распределения, получаемые по этим выражениям, являются симметричными относительно математического ожидания нормированной меры Хэмминга. Эти распределения могут использоваться в методиках тестирования сбалансированных биометрико-нейросетевых преобразователей, изложенных в [4]. Ошибка оценки вероятности ложной аутентификации равна 0,03 % от 256 бит. При несбалансированности по вероятности появления "0" в разряде кода необходимо синтезировать таблицы биномиального зависимого закона распределения значений с учетом небаланса по этому параметру. Синтезировать таблицы вероятностей ложной аутентификации пользователей "Все Чужие" необходимо на основе результатов моделирования биномиального зависимого закона распределения значений, полученных в [7]. Она учитывает небаланс распределения нормированной меры Хэмминга относительно ее математического ожидания вероятностей появления "0" в бите кода и эффект больших хвостов при значениях корреляции стремящейся к единице. Исходя из требований [3], шаг по корреляции нужно выбирать равным 0,01. Исходя из наиболее часто встречающихся на практике ситуаций значений параметров тестируемых нейросетевого преобразователя биометрия-код, шаг по вероятности появления "0" в бите кода нужно выбирать равным 0,01.

Литература

1 *Dodis Y., Reyzin L., Smith A.* Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy. - 2004. April 13.

2 *Cavoukian A., Stoianov A.* Biometric Encryption: A Positive-Sum Technology that Achieves Strong Authentication, Security AND

Privacy, March 2007, http://www.ipc.on.ca.

3 ГОСТ Р 52633.0-2006. Защита информации. Техника защиты информации. Требования к средствам высоконадежной биометрической аутентификации.

4 *Малыгин А.Ю., Волчихин В.И., Иванов А.И., Фунтиков В.А.* Быстрые алгоритмы тестирования нейросетевых механизмов биометрико-криптографической защиты информации. - Пенза: Изд-во Пенз. гос. ун-а, 2006. - 161 с.

5 *Малыгин А.Ю., Надеев Д.Н., Иванов А.И.* Две причины не идеальности нейросетевых преобразователей биометрия-код по выходным случайным состояниям кодов "Чужие" // Вопросы защиты информации. - 2008. - № 2 (81). - С. 19-21.

6 *Надеев Д.Н.* Моделирование биномиального зависимого закона распределения значений вероятностей ошибок нейросетевых преобразователей для высоконадежной биометрической защиты // Вопросы защиты информации. - 2008. - № 3. - С. 31-35.

7 *Захаров О.С., Иванов А.И.* Учет корреляционных связей биометрических данных через дробный показатель степеней свободы закона распределения значений хи-квадрат // Инфокоммуникационные технологии. - 2008. - Т.6, № 1. - С. 12-15.