УДК 519.24; 53; 57.17
DOI: 10.22213/2410-9304-2019-2-30-36

ОЦЕНКА ЭНТРОПИИ ДЛИННЫХ КОДОВЫХ СЛОВ НА ВЫХОДЕ НЕЙРОСЕТЕВОГО ПРЕОБРАЗОВАТЕЛЯ БИОМЕТРИИ В ПРОСТРАНСТВАХ МНОЖЕСТВА СВЕРТОК ХЭММИНГА

А. И. Иванов, доктор технических наук, доцент, Пензенский научно-исследовательский электротехнический институт, Пенза, Россия
А. П. Юнин, Пензенский научно-исследовательский электротехнический институт, Пенза, Россия
М. А. Бояршинов, кандидат технических наук, ИжГТУ имени М. Т. Калашникова

Работа построена на учете дискретного характера спектра состояний нейросетевого преобразователя биометрических данных в код длиной 256 бит. Вычислить энтропию кодов такой длины по Шеннону технически невозможно. Предложено от обычных кодов перейти в пространстве расстояний Хэмминга. В этом случае число выходных состояний свертывается и вычисления оказываются выполнимы на обычной вычислительной машине. Даются примеры спектров расстояния Хэмминга для идеального «белого» шума, вычисленные в резных системах счисления (в системах с резным модулем). С ростом модуля, по которому вычисляется расстояние Хэмминга, быстро растет число учитываемых спектральных линий. Дана таблица значений амплитуд вероятности состояний простейшей хи-квадрат молекулы, настроенной на обработку сверток Хэмминга по модулю 2, 3, 4, …, 247. Приводится формула вычисления состояний двухуровневой молекулы Хэмминга, вычисляющей расстояния Хэмминга на двух уровнях. Проведением численных экспериментов доказывается, что предложенный алгоритм позволяет оценивать энтропию длинных кодов на малых выборках с привлечением обычных вычислительных машин. При этом преобразования для разных систем отсчета (для сверток Хэмминга, вычисленных по разному модулю) дополняют друг друга. Ошибки вычислений для сверток Хэмминга, вычисленных по разным модулям, оказываются не коррелированными. Это позволяет надеяться на то, что предложенный метод позволит достаточно точно оценивать близость кодов длиной 256 бит к идеальному «белому» шуму.

Принципиально важным является то, что все свертки Хэмминга, независимо от системы счисления, в которой они вычислены, всегда позволяют логарифмически уменьшать число рассматриваемых состояний. Это в конечном итоге и позволяет упростить задачу и оценивать энтропию длинных кодов на обычной вычислительной машине с привлечением малых выборок в несколько сотен тестовых примеров.

Ключевые слова: нейросетевой преобразователь, расстояния Хэмминга, дискретный спектр, вычисление энтропии длинных кодов, малые выборки.

Введение

В настоящее время активно создается цифровая экономика. Ее значительная часть всегда будет оказываться критической [1, 2], однако обходиться без интернет-«облаков» сегодня уже нельзя. Сегодня и завтра любой важный объект некоторой своей частью оказывается прицеплен к интернет-«облакам». Защита интернет-«облаков» может быть осуществлена только криптографией, и каждому из нас придется осваивать безопасное управление своими информационными ресурсами. Сегодня безопасность строится на коротких паролях (длинные пароли из случайных цифр человек запомнить не может).

Для того чтобы избавить обычных пользователей от запоминания длинных паролей, могут быть использованы нейросетевые преобразователи биометрического образа человека в код его личного криптографического ключа. Требования к нейросетевым преобразователям регламентирует пакет из 7 национальных стандартов с номерами ГОСТ Р 52633.xx–20xx. Автоматическое обучение больших нейронных сетей должно выполняться алгоритмом ГОСТ Р 52633.5–2011 («Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия-код доступа») на малой выборке в 20 примеров биометрического образа «Свой». Криптографическая защита таблиц связей и коэффициентов обученной нейронной сети должна выполняться по технической спецификации (ГОСТ Р xxxx-20xx «Криптографическая защита. Защита нейросетевых биометрических контейнеров с использованием криптографических алгоритмов»), которая будет введена в действие на территории России в ближайшее время после ее одобрения членами технического комитета по стандартизации № 026. Кроме того, в ближайшее время следует ожидать разработки в России ряда новых национальных биометрических стандартов [3].

После каждого обучения большой нейронной сети необходимо ее тестировать. Кажется, что задача тестирования качества (энтропии) выходных кодов нейронной сети длиной в 256 бит является вычислительно сложной. Это действительно так, если мы будем пытаться вычислять энтропию по Шеннону. В этом случае задача тестирования действительно имеет экспоненциальную вычислительную сложность.

Снизить вычислительную сложность задачи удается, если следовать рекомендациям ГОСТ Р 52633.3–2011 («Защита информации. Техника защиты информации. Тестирование стойкости средств высоконадежной биометрической защиты к атакам подбора») и перейти в пространство расстояний Хэмминга.

$$h_2(\bar{x},\bar{c}) = \sum_{i=1}^{256} \left| "x_i" - "c_i" \right| = \sum_{i=1}^{256} ("\neg x_i") \oplus ("c_i"), \quad (1)$$

где $"\neg x_i"$ – инверсия i-го разряда свертываемого кода с i-м разрядом; код «Свой» – $"c_i"$.

Следует отметить, что расстояния Хэмминга всегда являются целыми числами. Для кодов «Чужой» длиной в 256 бит математическое ожидание расстояний Хэмминга составит примерно 128 бит, что обусловлено свойствами алгоритма обучения ГОСТ Р 52633.5–2011. Пример сглаженного спектра распределения расстояний Хэмминга приведен на рис. 1.

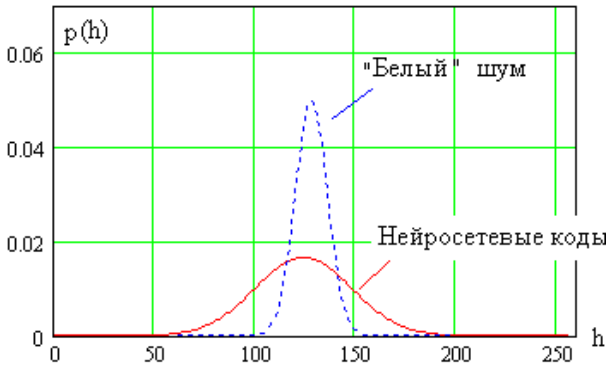

Рис. 1. Сглаженные распределения расстояний Хэмминга, вычисленных по модулю 2

Следует отметить, что с вычислением расстояния Хэмминга по модулю 2 можно пользоваться иным термином – свертка Хэмминга по модулю два пространства выходных кодов нейросети. Обоснованность такого термина обусловлена тем, что исходное поле выходных кодов нейросети имеет 2^{256} состояний, тогда как свертка Хэмминга этих кодов дает значения от 0 до 256 (всего 257 состояний). Мы имеем свертку, обеспечивающую сжатие пространства состояний пропорционально двоичному логарифму.

Наблюдение дискретного спектра расстояний Хэмминга нейросетевого преобразователя при неизвестном коде «Свой»

Для того чтобы оценить хэширующие свойства обученного нейросетевого преобразователя биометрия-код, достаточно воспользоваться несколькими тестовыми образами: «Чужой-1», «Чужой-2», …, «Чужой-32». При этом мы получим вектора кодов откликов: $\bar{x}_1, \bar{x}_2, ..., \bar{x}_{32}$. Этот численный эксперимент легко может быть воспроизведен с данными из среды моделирования «БиоНейроАвтограф» [4], которая имеет возможность обучать нейронную сеть и тестировать ее. При этом в файле mera.txt размещается последовательность расстояний Хэмминга, вычисленных по модулю два [5].

Если мы не знаем код «Свой» при вычислении энтропии, свертывать между собой следует кодовые отклики $\bar{x}_1, \bar{x}_2, ..., \bar{x}_{32}$.

При использовании N примеров в тестовой выборке мы получаем k вариантов расстояний Хэмминга (1):

$$k = \sum_{i=0}^{N-1} (N-1-i) \approx \frac{(N-1)^2}{2}. \quad (2)$$

Для выборки из 32 опытов мы получим 496 неповторяющихся взаимных расстояний Хэмминга между разными выходными кодами преобразователя. Этого вполне достаточно для тестирования.

Получение распределения расстояний Хэмминга для идеального «белого» шума

Для того чтобы получить коды практически идеального «белого» шума, следует воспользоваться функцией криптографического хэширования:

$$\bar{x}_i = hash("соль", round(rnd(256), 0)). \quad (3)$$

Желательно, чтобы эталонное хэширование (3) происходило с использованием вызова отечественной хэш-функции, выполненной по отечественному стандарту ГОСТ Р 34.11–2012 («Информационная технология. Криптографическая защита информации. Функция хэширования»). В этом случае длина выходного кода будет 256 бит. Если пользоваться хэш-функцией MD-5 с длиной кода 128 бит, потребуется дважды выполнять хэширование и объединять результаты конкатенацией.

Во всем остальном функция вычисления свертки Хэмминга (2) остается прежней. Следует также обратить внимание на то, что для идеального белого шума все 256 разрядов кода должны иметь равновероятные значения состояний «0» и «1». Более того, плотность вероятности сверток Хэмминга идеального белого шума должна описываться биномиальным законом распределения:

$$p(h) = \left[\frac{(256-h)!}{h!(256-h)!} \right] \cdot \left(\frac{1}{2} \right)^h \cdot \left(\frac{1}{2} \right)^{256-h}. \quad (4)$$

При этом математическое ожидание и стандартное отклонение распределений расстояний Хэмминга для идеального «белого» шума должно составлять:

$$E(h) = 128, \quad (5)$$

$$\sigma(h) = \sqrt{\frac{256}{4}} = 8. \quad (6)$$

Соотношения (5), (6) хорошо соответствуют действительности. Уже для выборок в несколько сотен расстояний Хэмминга:

$$\begin{cases} E(h) = 128 \pm 0,5, \\ \sigma(h) = 8 \pm 0,3. \end{cases} \quad (7)$$

Распределение расстояний Хэмминга для «белого» шума хорошо описывается нормальным законом распределения значений, отображенным на рис. 1 пунктиром.

Вычисления свертки расстояний Хэмминга по модулю четыре со сдвигом на один разряд

Следует отметить, что может существовать множество вариантов сверток Хэмминга [6–8]. Например, можно перейти от вычитания одинаковых разрядов двух свертываемых кодов к парам разрядов этих кодов со сдвигом на один разряд:

$$h_4(\overline{x}_k, \overline{x}_{k+j}, 1, 256) = \\ = \sum_{i=1}^{255} \left| "x_{i,k}, x_{i+1,k}" - "x_{i,k+j}, x_{i+1,k+j}" \right|. \quad (8)$$

Проведенный численный эксперимент показал, что такого вида свертки имеют число состояний примерно в три раза больше обычных расстояний Хэмминга:

$$\begin{cases} 0 \le h_4(\overline{x}_k, \overline{x}_{k+j}, 1, 256) \le 706, \\ E\left\{h_4(\overline{x}_k, \overline{x}_{k+j}, 1, 256)\right\} = 352, \\ \sigma\left\{h_4(\overline{x}_k, \overline{x}_{k+j}, 1, 256)\right\} = 24. \end{cases} \quad (9)$$

Этот тип сверток примерно в три раза меньше свертывает данные в сравнении со сверткой Хэмминга по модулю два (1).

Для нас принципиально важным является то, что отклики сравниваемых сверток являются некоррелированными. Кроме того, приведение к одному масштабу данных дает близкие значения стандартных отклонений:

$$\begin{cases} \operatorname{corr}(h_2(..,1,.), h_4(..,1,.)) \approx 0, \\ \dfrac{\sigma(h_2(..,1,.))}{256} = 0,032 \approx \dfrac{\sigma(h_4(..,1,.))}{706} = 0,034. \end{cases} \quad (10)$$

Вычисления свертки расстояний Хэмминга по модулю четыре со сдвигом на два разряда

Еще одним способом вычисления сверток Хэмминга по модулю четыре является сдвиг на два разряда при осуществлении сложений:

$$h_4(\overline{x}_k, \overline{x}_{k+j}, 2, 256) = \\ = \sum_{i=1}^{128} \left| "x_{2i-1,k}, x_{2i,k}" - "x_{2i-1,k+j}, x_{2i,k+j}" \right|. \quad (11)$$

Проведенный численный эксперимент показал, что такого вида свертки имеют число состояний примерно в три раза больше обычных расстояний Хэмминга:

$$\begin{cases} 0 \le h_4(\overline{x}_k, \overline{x}_{k+j}, 2, 256) \le 352, \\ E\left\{h_4(\overline{x}_k, \overline{x}_{k+j}, 2, 256)\right\} = 176, \\ \sigma\left\{h_4(\overline{x}_k, \overline{x}_{k+j}, 2, 256)\right\} = 12. \end{cases} \quad (12)$$

При этом отклики сравниваемых сверток Хэмминга по модулю 2 и по модулю 4 являются некоррелированными. Кроме того, приведение к одному масштабу данных дает близкие значения стандартных отклонений:

$$\begin{cases} \operatorname{corr}(h_2(..,1,.), h_4(..,2,.)) \approx 0, \\ \dfrac{\sigma(h_2(..,1,.))}{256} = 0,032 \approx \dfrac{\sigma(h_4(..,2,.))}{352} = 0,034. \end{cases} \quad (13)$$

Однако две свертки Хэмминга, вычисленные по одинаковому модулю, уже являются зависимыми:

$$\operatorname{corr}(h_4(..,1,.), h_4(..,2,.)) \approx 0,5. \quad (14)$$

Вычисления свертки расстояний Хэмминга по модулю восемь со сдвигом на один разряд

Очевидно, что мы можем при вычислении сверток Хэмминга повысить длину блока до трех сравниваемых разрядов:

$$h_8(\overline{x}_k, \overline{x}_{k+j}, 1, 256) = \\ = \sum_{i=1}^{254} \left| "x_{i,k}, x_{i+1,k}, x_{i+2,k}" - "x_{i,k+j}, x_{i+1,k+j}, x_{i+2,k+j}" \right|. \quad (15)$$

Проведенный численный эксперимент показал, что такого вида свертки имеют число состояний примерно в пять раза больше обычных расстояний Хэмминга, вычисленных по модулю два (12):

$$\begin{cases} 0 \le h_8(\overline{x}_k, \overline{x}_{k+j}, 1, 256) \le 1368, \\ E\left\{h_8(\overline{x}_k, \overline{x}_{k+j}, 1, 256)\right\} = 684, \\ \sigma\left\{h_8(\overline{x}_k, \overline{x}_{k+j}, 1, 256)\right\} = 32. \end{cases}$$

Корреляция между свертками разного модуля отсутствует, нормированные стандартные отклонения сопоставимы (13):

$$\begin{cases} \mathrm{corr}(h_2(..,1,.), h_4(..,1,.)) \approx 0 \approx \\ \approx \mathrm{corr}(h_4(..,1,.), h_8(..,1,.)) \approx 0, \\ \dfrac{\sigma(h_2(..,1,.))}{256} = 0,032 \approx \dfrac{\sigma(h_4(..,1,.))}{706} = \\ = 0,034 \approx \dfrac{\sigma(h_8(..,1,.))}{1368} = 0,023. \end{cases}$$

Спектральный анализ данных на близость к идеальному «белому» шуму

Анализ биометрических данных на их близость к «белому» шуму является важным направлением исследований биометрии. Так как мы знаем параметры, описывающие спектры значений сверток Хэмминга, мы можем построить простейшее правило, например, в виде хи-квадрат молекулы [9–11].

Как видно из рис. 1, для «белого» шума расстояния Хэмминга по модулю 2 могут принимать примерно 50 значений от 100 до 150. Появление других значений маловероятны на ограниченных выборках, состоящих из примерно 500 опытов. Формула преобразований хи-квадрат молекулы отличается от классической (в формуле нет суммирования) (14):

$$\chi^2(h_2) = \left\{\frac{h_2 - E(h_2)}{\sigma(h_2)}\right\}^2 = \left\{\frac{h_2 - 128}{8}\right\}^2.$$

Так как входных состояний хи-квадрат молекулы не более 50, выходных состояний оказывается в 2 раза меньше из-за возведения данных в квадрат, как это отображено на рис. 2.

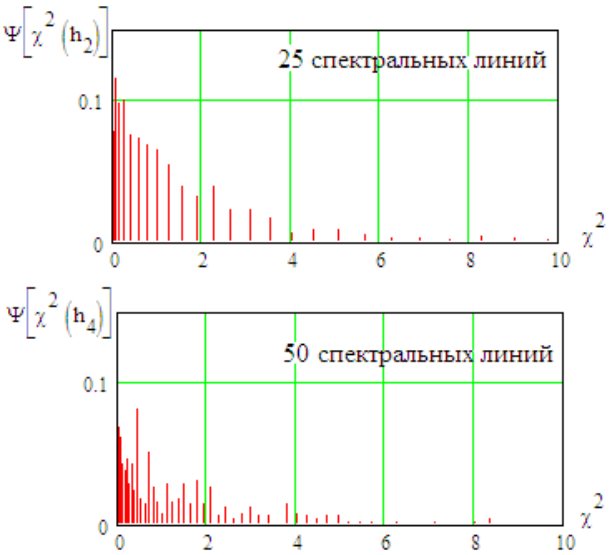

Рис. 2. Спектральные линии двух простейших хи-квадрат молекул

Как видно из рис. 2, наиболее вероятными являются малые значения хи-кадрат молекулы. Большие значения выходного состояния хи-квадрат молекулы (14) маловероятны (амплитуды вероятности их появления $\Psi(\chi^2(h_2))$ падают). Численные значения амплитуд вероятности значений хи-квадрат молекулы отражены в таблице.

Параметры разрешенных состояний простейшей хи-квадрат молекулы

n	0	1	2	4	4	5	6	7
$\Psi(\chi^2)$	0,078	0,117	0,099	0,105	0,074	0,072	0,076	0,064
χ^2	0	0,06	0,14	0,25	0,39	0,56	0,76	1
α	1,0	0,879	0,762	0,663	0,558	0,484	0,412	0,336
n	8	9	10	11	12	13	14	15
$\Psi(\chi^2)$	0,055	0,039	0,032	0,039	0,023	0,023	0,017	0,006
χ^2	1,26	1,56	1,89	2,25	2,64	3,06	3,51	4
α	0,272	0,217	0,178	0,146	0,107	0,084	0,061	0,044
n	16	17	18	19	20	21	22	23
$\Psi(\chi^2)$	0,009	0,009	0,005	0,003	0,003	0,001	0,004	0,001
χ^2	4,51	5,06	5,64	6,25	6,89	7,56	8,26	9
α	0,038	0,029	0,02	0,015	0,012	0,009	0,008	0,004

Таблица имеет 4 строки. В первой строке дан номер спектральной составляющей, в следующей строке дается амплитуда вероятности появления значений хи-квадарт молекулы χ^2. В следующей строке дается значение α доверительной вероятности того, что гипотеза «белого» шума верна.

Следует отметить, что переход к использованию сверток Хэмминга по модулю четыре, как источника данных хи-квадрат молекулы, дает удвоение ее состояний, как это показано в нижней части рис. 1. Спектр амплитуд вероятности появления содержит 50 линий. Для сверток Хэмминга по модулю 4 выходное состояние хи-квадрат молекулы вычисляются по формуле (15):

$$\chi^2(h_4) = \left\{ \frac{h_4 - E(h_4)}{\sigma(h_4)} \right\}^2 = \left\{ \frac{h_4 - 352}{24} \right\}^2 .$$

Переход к использованию двухуровневых хи-квадрат молекул

Следует отметить, что сегодня для оценки качества «белого» шума принято использовать несколько тестов. Так, национальный институт стандартов США (NIST) рекомендует применять 16 тестов на близость к белому шуму:

– тест на сопоставимое число «0» и «1» в коде (тест равной вероятности состояний);

– тест, анализирующий вероятность появления одинаковых соседних состояний;

– тест обезьяны, набирающей случайные знаки на пишущей машинке;

…

В [8] было показано, что рассматриваемые в данной работе процедуры обобщают рекомендации NIST по тестированию качества белого шума. В частности, их можно рассматривать как обобщение теста одной обезьяны, набирающей случайный текст на одной пишущей машинке, на тест «стаи обезьян», каждая из которых использует свою печатающую машинку. Чисто формально можно утверждать, что 16 не связанных между собой тестов NIST могут быть эквивалентны 16 обезьянам, случайно набирающим текст на 16 печатающих машинках с 16 разными кодировками букв 16 разных языков.

Для обобщения результатов этих тестов воспользуемся хи-квадрат критерием:

$$\chi^2(\overline{h}_{2^j}) = \sum_{j=1}^{16} \left\{ \frac{E(h_{2^j}) - h_{2^j}}{\sigma(h_{2^j})} \right\}^2 . \qquad (16)$$

Предположительно плотность распределения значений должна хорошо описываться формулой Пирсона для некоторого дробного числа степеней свободы m:

$$p(\chi^2) = \frac{1}{2^{\frac{m}{2}} \Gamma\left\{\frac{m}{2}\right\}} \cdot x^{\left(\frac{m}{2}-1\right)} \cdot \exp\left(\frac{-x}{2}\right), \qquad (17)$$

где $\Gamma(.)$ – гамма-функция Эйлера.

На данный момент неизвестно, как вычислять показатель степени числа степеней свободы, однако он, видимо, не может быть целым числом. Это связано с тем, что каждый из слагаемых формулы (16) имеет собственное дробное значение числа степеней свободы. Кроме того, компоненты формулы (16) могут быть коррелированны (14), что также приводит к дробности показателя числа степеней свободы [12].

По сути дела, преобразование (16) обобщает состояния нескольких частных хи-квадрат молекул, эту конструкцию можно рассматривать как некоторую обобщенную двухуровневую хи-квадрат молекулу. Хи-квадрат молекулы похожи на искусственные нейроны (нейросетевые молекулы [13]). Хи-квадрат молекулы можно объединять в сети, так же как и искусственные нейроны можно объединять в нейронные сети. При этом качество принимаемых решений существенно увеличивается.

Заключение

Изложенные в данной статье соотношения позволяют надеяться на то, что удастся создать новый стандарт оценки качества «белого» шума. Вместо 16 тестов NIST можно создать гораздо более эффективную вычислительную процедуру оценки качества кодов по их близости к идеальному «белому» шуму. Имеется возможность малое число тестов, стандартизованных в США (всего 16 тестов), дополнить 247 новыми тестами стаи обезьян [8]. При оценке качества кодов «белого шума» большой длины в 256 бит размерность решаемой задачи с 16-мерной может быть увеличена до 263-мерной. Рост увеличения размерности составляет примерно 16 раз, что является предпосылкой существенного роста достоверности принимаемых решений.

Библиографические ссылки

1. *Куделькин В. А., Янников И. М., Телегина М. В.* Принципы создания интегрированных систем безопасности критически важных и потенциально опасных объектов // Интеллектуальные системы в производстве. 2017. Т. 15. № 1. С. 105–109.

2. *Куделькин В. А., Янников И. М., Габричидзе Т. Г.* Особенности обработки данных в интеллектуальной интегрированной системе безопасности объектов и территорий // Интеллектуальные системы в производстве. 2017. Т. 15. № 4. С. 94–101.

3. *Иванов А. И.* Нейросетевая биометрия для облаков. Российские стандарты для защиты цифровых прав граждан // Системы безопасности. 2018. № 3. С. 134–143. URL: www.secuteck.ru/imeg/ss-3-2018.

4. *Иванов А. И., Захаров О. С.* Среда моделирования «БиоНейроАвтограф». Программный продукт создан лабораторией биометрических и нейросетевых технологий, размещен с 2009 г. на сайте АО «ПНИЭИ». URL: http://пниэи.рф/ activity/science/noc/ bioneuroautograph.zip.

5. *Иванов А. И.* Автоматическое обучение больших искусственных нейронных сетей в биометрических приложениях : учеб. пособие. Пенза, 2013. 30 с. URL: http://пниэи.рф/activity/ science/noc/tm_IvanovAI.pdf.

6. Многомерный портрет цифровых последовательностей идеального «белого шума» в свертках Хэмминга / В. И. Волчихин, А. И. Иванов, А. П.Юнин, Е. А. Малыгина // Известия высших учебных заведений. Поволжский регион. Технические науки. 2017. № 4. С. 4–13.

7. *Юнин А. П., Корнеев О. В.* Оценка энтропии легко запоминаемых, длинных паролей со смыслом в ASCII кодировке для русского и английского языков // Труды научно-технической конференции кластера пензенских предприятий, обеспечивающих безопасность информационных технологий. Т. 10. Пенза, 2016. С. 40–42. URL: http://пниэи.рф/activity/ science/BIT/T10-p40.pdf.

8. Оценка качества «белого» шума: реализация теста «стаи обезьян» через множество сверток Хэмминга, построенных на разных системах счисления / А. П. Юнин, А. И. Иванов, К. А. Ратников, Е. А. Кольчугина // Известия высших учебных заведений. Поволжский регион. Технические науки. 2018. № 4 (48). С. 32–36.

9. Дискретный характер закона распределения хи-квадрат критерия для малых тестовых выборок / Б. Б. Ахметов, А. И. Иванов, Н. И. Серикова, Ю. В. Фунтикова // Вестник Национальной академии наук Республики Казахстан. 2015. № 1. С. 17–25.

10. Циклические континуально-квантовые вычисления: усиление мощности хи-квадрат критерия на малых выборках / В. Кулагин, А. Иванов, А. Газин, Б. Ахметов // Аналитика. 2016. № 5 (30). С. 22–29.

11. Перспективы создания циклической континуально-квантовой хи-квадрат машины для проверки статистических гипотез на малых выборках биометрических данных и данных иной природы / В. И. Волчихин, А. И. Иванов, Д. В. Пащенко, Б. Б. Ахметов, С. Е. Вятчанин // Известия высших учебных заведений. Поволжский регион. Технические науки. 2017. № 1. С. 5–15.

12. *Volchikhin V. I., Ivanov A. I., Malygina E. A., Kupriyanov E. N., Serikova Yu. I.* Precision statistics: fractional number of degrees of freedom chi-square criterion for small samples of biometric data. /Journal of computational and engineering mathematics. Vol. 6, №1 (2019) p.p. 55-62.

13. *Волчихин В. И., Иванов А. И.* Нейросетевая молекула: решение обратной задачи биометрии через программную поддержку квантовой суперпозиции на выходах сети искусственных нейронов // Вестник Мордовского университета. 2017. Т. 27. № 4. С. 518–523.

References

1. Kudel’kin V.A., Yannikov I.M., Telegina M.V. [Principles of Developing the Integrated Security Systems of Critical and Potentially Dangerous Objects]. *Intellektual'nye sistemy v proizvodstve.* 2017. No. 1, pp. 105-109 (in Russ). DOI: 10.22213/2410-9304-2017-1-105-109.

2. Kudel'kin V.A., YAnnikov I.M., Gabrichidze T.G. [Features of data processing in an intelligent integrated security system of objects and territories]. *Intellektual'nye sistemy v proizvodstve.* 2017. No. 4, pp. 94-101 (in Russ.).

3. Ivanov A.I. [Neural network biometrics for clouds. Russian standards for the protection of digital rights of citizens]. *Sistemy bezopasnosti*. 2018. No. 3, pp. 134-143 (in Russ.). Available at: www.secuteck.ru/imeg/ss-3-2018 (in Russ.).

4. Ivanov A.I., Zaharov O.S. *Sreda modelirovaniya «BioNeiroAvtograf». Programmnyi produkt sozdan laboratoriei biometricheskikh i neirosetevykh tekhnologii, razmeshchen s 2009 g. na saite AO «PNIEI»* [Simulation environment «BioNejroAvtograf» The software product was created by the laboratory of biometric and neural network technologies; it has been posted on the website since 2009]. Available at http://pniehi.rf/activity/science/noc/bioneuroautograph.zip (in Russ.).

5. Ivanov A.I. *Avtomaticheskoe obuchenie bol'shikh iskusstvennykh neironnykh setei v biometricheskikh prilozheniyakh* [Automatic training of large artificial neural networks in biometric applications]. Tutorial. Penza. 2013. 30 p. (in Russ.).

6. Volchihin V.I., Ivanov A.I., YUnin A.P., Malygina E.A. [Multidimensional portrait of digital sequences of ideal “white noise” in Hamming convolutions]. *Izvestiya vysshih uchebnyh zavedenij. Povolzhskij region. Tekhnicheskie nauki.* 2017, no. 4, pp. 4-13. (in Russ.).

7. YUnin A.P., Korneev O.V. *Otsenka entropii legko zapominaemykh, dlinnykh parolei so smyslom v ASCII kodirovke dlya russkogo i angliiskogo yazykov* [Entropy assessment of easily remembered, long passwords with ASCII meaning for Russian and English]. Proceedings of the scientific and technical conference of the cluster of Penza enterprises, ensuring the security of information technology. Vol. 10, Penza-2016, pp. 40-42. Available at http://pniehi.rf/activity/science/BIT/T10-p40.pdf. (in Russ.).

8. Yunin A.P., Ivanov A.I., Ratnikov K.A., Kol'chugina E.A. [Quality assessment of "white" noise: the implementation of the test "flocks of monkeys"

through many Hamming convolutions, built on different number systems]. *Izvestiya vysshih uchebnyh zavedenij. Povolzhskij region. Tekhnicheskie nauki.* 2018, no. 4 (48), pp. 32-36 (in Russ.).

9. Ahmetov B.B., Ivanov A.I., Serikova N.I., *Funtikova YU.V. Diskretnyi kharakter zakona raspredeleniya khi-kvadrat kriteriya dlya malykh testovykh vyborok* [Discrete nature of the distribution law of the chi-square test for small test samples. Vestnik Nacional'noj akademii nauk Respubliki Kazahstan]. Almaty, 2015, no. 1, pp. 17-25 (in Russ.).

10. Kulagin V., Ivanov A., Gazin A., Ahmetov B. [Cyclic Continuous-Quantum Computations: Enhancing the Power of the Chi-Square Criterion on Small Samples]. *Analitika*, no. 5, 2016 (30), pp. 22-29 (in Russ.).

11. Volchihin V.I., Ivanov A.I., Pashchenko D.V., Ahmetov B.B., Vyatchanin S.E. [Prospects for creating a cyclic continual quantum chi-square machine for testing statistical hypotheses on small samples of biometric data and data of a different nature]. *Izvestiya vysshih uchebnyh zavedenij. Povolzhskij region. Tekhnicheskie nauki*. Penza: PGU, no.1, 2017, pp. 5-15 (in Russ.).

12. Volchikhin V.I., Ivanov A.I., Malygina E.A., Kupriyanov E.N., Serikova Yu.I. Precision statistics: fractional number of degrees of freedom chi-square criterion for small samples of biometric data. «Journal of computational and engineering mathematics» Vol. 6, №1 (2019) p.p. 55-62.

13. Volchihin V.I., Ivanov A.I. [Neural network molecule: solving the inverse problem of biometrics through software support for quantum superposition at the outputs of a network of artificial neurons]. *Vestnik Mordovskogo universiteta.* 2017, vol. 27, no. 4, pp. 518–523 (in Russ.).

* * *

Estimation of the Entropy of Long Code Words at the Output of a Neural Network Biometrics Converter in Spaces of Hamming Convolutions

A. I. Ivanov, DSc in Engineering, Associate Professor, Penza National Investigation Electrotechnical Institute
A. P. Yunin, Penza National Investigation Electrotechnical Institute
M. A. Boyarshinov, PhD in Engineering, Associate Professor, Kalashnikov ISTU

The work is based on accounting the discrete nature of the spectrum of states of a neural converter of biometric data into the code of 256 bits. It is technically impossible to compute the entropy of codes with such a length according to Shannon. It is proposed to move from usual codes to the space of Hemming distances. In this case, the number of output states of convolvong and calculations turn out to be workable at a usual computing machine. Examples of Hemming spectrum distances are given for an ideal "white" noise, computed in a divided system of numeration (in systems with the divided module). With growing of the module, according to which the Hemming distance is computed, the number taken into account spectral lines is quickly growing. The table is given for the values of amplitudes of state probabilities for the simplest chi-square molecule adjusted for processing the Hemming convolutions by modulo 2, 3, 4, , 247. The formula is given for calculation of states for the two-level Hemming molecule computing Hemming distances at two levels. The performed numerical experiments prove that the proposed algorithm allows to evaluate the entropy of long codes on small samples with attraction of usual computing machines. At that, the transformations for miscellaneous reference systems (for Hemming convolutions computed for miscellaneous modules) complement each other. The errors of calculations for Hemming convolutions computed for miscellaneous modules turn out to be not correlated. This allows to hope that the proposed method will allow to evaluate accurately enough the vicinity of codes with 256 bits length to the ideal "white" noise.

The principally important issue is that all Hemming convolutions regardless of numeration systems, in which they are computed, always allow to reduce logarithmically the number of the considered states. This finally allows to simplify the task and evaluate the entropy of long codes by a usual computing machine with application of small samples of several hundreds of test examples.

Keywords: neural converter, Hemming distances, discrete spectrum, calculation of entropy of long codes, small samples.

Получено: 24.04.19