Труды научно-технической конференции кластера пензенских предприятий,
обеспечивающих **БЕЗОПАСНОСТЬ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ.**
Том 8, Пенза-2012, с. 70-72 (http://пниэи.рф/activity/science/BIT/T8-p70.pdf)

ТРЕБОВАНИЯ К КАЧЕСТВУ РАБОТЫ МЕХАНИЗМА РАЗМНОЖЕНИЯ ОШИБОК НЕЙРОСЕТЕВОГО ПРЕОБРАЗОВАТЕЛЯ БИОМЕТРИЯ-КОД

Сомкин С.А., Китанин А.И., Чистякова Д.Д. (г. Пенза)

Шум является белым, если спектральные составляющие его равномерно распределены по всему диапазону задействованных частот. В природе и технике «чисто» белый шум (т.е. белый шум, имеющий одинаковую спектральную мощность на всех частотах) не встречается (ввиду того, что такой сигнал имел бы бесконечную мощность), однако под категорию белых шумов попадают любые шумы, спектральная плотность которых одинакова (или слабо отличается) в рассматриваемом диапазоне частот. В случае, если мы имеем дело с последовательностью цифровых отсчетов, то белый оцифрованный шум должен иметь отсчеты с независимыми состояниями к каждом разряде. То есть коэффициенты парной корреляции между любыми двумя случайно выбранными разрядами, вычисленные для цифровой последовательности белого шума должны давать нулевые значения.

Известно, что оцифрованные звуки в виде текста на русском языке не является белым шумом. Однако мы можем искусственно сделать коды текста на руссом языке, зашифровав их алгоритмом отечественного стандарта [1]. Зашифрованный текст может быть использован как эталон цифрового белого шума.

Предположим, что нас интересует выходные коды «Чужой» выдаваемые нейросетевым преобразователем биометрия-код. Эти коды так же как и коды текстов на естественных языках не являются белым шумом, однако они могут быть сделаны белым шумом, например, процедурой хэширования. Предположительно в преобразователь биометрия-код может быть встроен механизм хэширования данных (механизм размножения ошибок), который будет осуществлять самошифрование данных выходного кода. Применение термина самошифрование корректно в связи с тем, что часть выходного кода преобразователя может быть использована для шифрования параметров искусственных нейронов, еще не участвовавших в процессе связывания биометрических данных с выходным кодом [2]. То есть может быть сформирован так называемый механизм размножения ошибок биометрических данных, который хэширует данные при попытках предъявить образ «Чужой» и не работает (не искажает данные) при предъявлении образа «Свой».

Эффект от включения механизма размножения ошибок состоит в резком уменьшении дисперсии распределения расстояний Хэмминга между кодами «Свой» и «Чужой» [3]. В связи с этим возникает вопрос как оценивать качество работы механизма размножения ошибок.

При оценке качества работы механизма размножения биометрических ошибок будем исходить из того, что эталонным белым шумом является тексты на русском языке зашифрованные отечественным стандартов [1]. То есть встроенный механизм размножения биометрических ошибок должен давать такой же или чуть хуже белый шум.

Теоретическим обоснование такого подхода биномиальный закон для кодов длинной 256 бит:

$$P(n,m,p) = \frac{n!}{m!(n-m)!} \cdot p^m \cdot (1-p)^{n-m} \qquad (1)$$

где n – число разрядов биометрического кода;

m – число верно угаданных «Чужим» разрядов биометрического кода;

p – средняя вероятность угадывания состояний разрядов.

Для перехода от классической формы записи биномиального закона распределения значений к распределению расстояний Хэмминга необходимо осуществить замену:

$$m = n - h$$

В конечном итоге вероятность угадывания разрядов кода описывается биномиальным законом распределения значений можно выразить как:

$$P(n,h,p) = \frac{n!}{(n-h)!(h)!} \cdot p^{(n-h)} \cdot (1-p)^h$$

При p=0,5 математическое ожидание расстояний Хэмминга всегда оказывается равным половине длины кода:

$$E(h) = \frac{n}{2} = 256 \cdot p = 128 \text{ бит}$$

Среднеквадратическое отклонение для p= 0,5 составит:

$$\sigma(h) = \sqrt{n \cdot p \cdot (1-p)} = \frac{\sqrt{n}}{2} = 8 \text{ бит}$$

Эти теоретические положения хорошо согласуются с численным экспериментом.

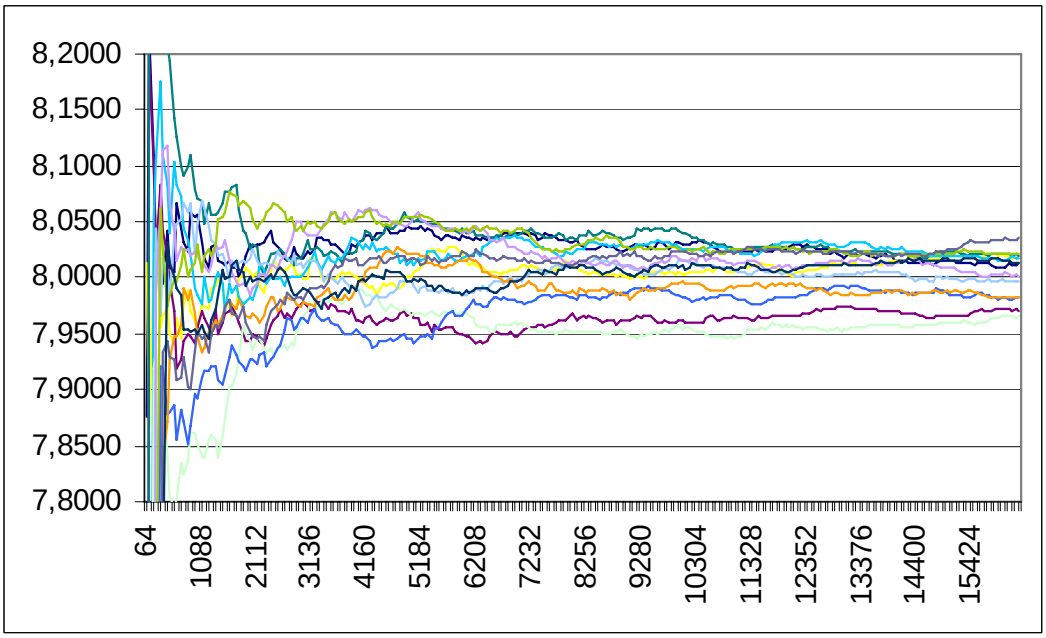

Рисунок 1 – Полученные значения дисперсии расстояний Хемминга в зависимости от увеличения длины последовательности зашифрованного текста

Как видно из полученного графика, для шифрованного текста распределения расстояний Хемминга близко к нормальному распределению, так как в шифрованных текстах отсутствуют «зависимости» между символами.

Если механизм размножения ошибок надежно работает (хэширование самошифрованием является криптографическим или близким к криптографическому), то увеличение выборки статистически исследуемых биометрических кодов «Чужой» должно приводить к монотонному росту точности выполнения последних условий. Ошибка оценки должна падать пропорционально $\sqrt{N}$, где N – размер тестовой выборки исследованных биометрических кодов[4].

Реализация алгоритма подсчета расстояний Хэмминга может быть использована для проверки оценки качества выходного нейросетевого преобразователя биометрия-код с включенным механизмом размножения ошибок. В качестве эталонных критериев оценки принимаются значения, полученные экспериментально при реализации подсчета распределения расстояний Хэмминга для шифрованного по алгоритму ГОСТ[1] русского осмысленного текста. Таким образом, если по результатам экспериментальной проверки выходной последовательности преобразователя биометрия-код с включенным механизмом размножения ошибок распределение дисперсии расстояний Хэмминга находится в диапазоне $8\pm\alpha\cdot0.05$ бит (для длины кода в 16384 байта), где α - некоторый коэффициент запаса. Предположительно значение коэффициент запаса должен выбираться в интервале от 2 до 5 в связи со сниженными требованиями к биометрии в сравнении с криптографией.

Литература:

1. ГОСТ 28147-89 «Система обработки информации. Защита криптографическая. Алгоритм криптографического преобразования».

2. Фунтиков В.А., Назаров И.Г., Бурушкин А.А. Национальные стандарты России: конфиденциальность персональных биометрических данных. «Стандарты и качество» № 7, 2010 г. с. 28-33.

3. ГОСТ Р 52633.3-2011 «Защита информации. Техника защиты информации. Тестирование стойкости средств высоконадежной биометрической защиты к атакам подбора».

4. Волчихин В.И., Иванов А.И., Назаров И.Г., Фунтиков В.А., Язов Ю.К. Нейросетевая защита персональных биометрических данных. 2012г. 160с.

Материалы поступили 30.11.2012, опубликовано в Интернет 12.12.2012 по положительной рецензии д.т.н., доцента Иванова А.И.