http://rosoperator.ru/rameev100/

Всероссийская научно-техническая конференция, посвященная 100-летию со дня рождения одного из основоположников советской вычислительной техники Б.И. Рамеева, на тему: «Информационно-управляющие и телекоммуникационные системы специального назначения». Секция «Биометрическая поддержка криптовалют, блокчейн реестров, облачных сервисов», проводится в конференц-зале Технопарка высоких технологий «Рамеев», доклад 16 мая 2018 года с 12^{00} до 12^{15} (ул. Центральная, д. 1), г. Пенза.

Юнин А.П., Иванов А.И.

Оценка качества «белого» шума: реализация теста «стаи обезьян» через множество сверток Хэмминга, построенных для разных систем счисления

Аннотация.

Актуальность и цели. Целью работы является повышение корректности вычисления энтропии длинных кодов со слабо зависимыми разрядами, порождаемыми нейросетевыми преобразователями биометрия-код или хэшированием биометрических данных средствами облегченной криптографии.

Материалы и методы. Классические процедуры Шеннона не могут быть использованы для вычислений, так как требуют использования огромного статистического материала. Для сокращения затрат вычислительных ресурсов используется отображение обычных кодов пространство сверток Хэмминга.

Результаты. Предложено рассматривать портрет белого шума в системе сверток Хэмминга, полученных в разных системах счисления. Это эквивалентно тесту множества обезьян, каждая из которых одновременно печатает на множестве печатающих машинок совершенно разных конструкций, созданных под разные языки и разные системы письменности. Чем больше используется печатающих машинок, тем быстрее множество обезьян совместными усилиями создадут осмысленную фразу на одном из возможных языков.

Выводы. В пространстве множества сверток Хэмминга оценка качества белого шума становится корректной, если нам заранее известны параметры распределений контролируемых сверток. При этом надежность оценок тем выше, чем больше контролируется различных сверток Хэмминга и чем выше размерность вычисляемых функционалов. Так как NIST США рекомендует порядка 16 тестов на случайность, предложено применять не менее 16 типов сверток Хэмминга, построенных для различных систем счисления.

Ключевые слова: энтропия длинных кодов со слабо зависимыми разрядами, регуляризация вычислений, многообразие сверток Хэмминга, свертки Хэмминга со случайным позиционированием разрядов

Проблема контроля уровня случайности данных, получаемых в биометрии

Активное развитие цифровой экономики в России и за рубежом делает необходимым повсеместное (массовое) использование облегченной криптографии и биометрии. Облегченная криптография применяется в том случае, когда стоимость операций не велика и они выполняются в доверенном процессоре SIM-карты, SD-карты, 4-х битном процессоре RFID-карты с микропитанием от созданного считывателем электромагнитного поля.

Возможность ослабления криптографии обусловлена так же тем, что она используется не самостоятельно, а в связке с биометрией пользователя [1, 2, 3]. Биометрия и ослабленная криптография в этом случае дополняют и усиливают друг друга.

Однако проблема локального получения действительно случайных чисел (действительно белого шума) существует. При этом, чем короче число, тем важнее становится контроль уровня его случайности. В этом отношении проблема контроля

случайности чисел для отечественной нейросетевой биометрии оказывается менее острой, чем та же проблема для зарубежных «нечетких экстракторов» [4, 5, 6, 7]. Зарубежные «нечеткие экстракторы» гораздо менее интеллектуальны нейросетевых преобразователей биометрия-код [8, 9]. «Нечеткие экстракторы» строятся на использовании кодов, обнаруживающих и исправляющих ошибки за счет высокой избыточности. Например, Даугман [7] использует самокорректирующийся код с 20-ти кратной избыточностью, то есть длина выходного кода «нечеткого экстрактора» в 20 раз короче, чем число контролируемых биометрических параметров.

В свою очередь число контролируемых биометрических параметров не может быть как угодно большим. Обычно производители средств биометрической защиты не указывают сколько параметров и какие параметры они контролируют. Единственным исключением является среда моделирования «БиоНейроАвтограф» [10], эта среда преобразует рукописный автограф (любой рукописный знак) в 416, контролируемых биометрических параметров (416 коэффициентов двухмерного преобразования Фурье рукописного образа). Положительным свойством среды моделирования «БиоНейроАвтограф» является то, что на данный момент времени это единственный достоверный источник биометрической информации для студентов, аспирантов, преподавателей университетов России, Беларуси, Казахстана, Армении, свободно читающих на русском языке. Все данные среды моделирования доступны для наблюдения [11], могут быть перегружены в иной математический редактор: MathCAD, MatLAB, Mathematica. Далее пользователь уже сам может написать собственную нейросетевую обработку биометрических данных в рамках курсового проекта, дипломной работы или диссертации.

Если вернуться к заявленной тематике, то для данных среды моделирования «БиоНейроАвтограф» «нечеткий экстрактор» будет иметь выходной личный ключ длинной 416/20=21 бит. Такая длина кода не сопоставимо меньше, чем 256 бит ключей, получаемых штатным нейросетевым преобразователем среды «БиоНейроАвтограф». При использовании ключа длинной в 21 бит требования к уровню его случайности на много выше, чем к уровню случайности ключа в 256 бит. Если уровень не случайности ключа в 21 бит составит 10% (реальная энтропия генератора может составить - 19 бит) – это на много страшнее для биометрических приложений, чем 10%-ная детерминированность генератора ключа в 256 бит, полученного от генератора, имеющего реальную энтропию - 230 бит. Все это следствие того, что собственная энтропия биометрического рукописного образа может находится в интервале от 11-ти до 97-ми бит. Для слабых биометрических образов энтропии в 19-ть бит генератора ключа «нечетких экстракторов» вполне достаточно. Атаковать нужно со стороны слабой биометрии с 11-ти битной собственной энтропией. Для сильной биометрии с собственной энтропией в 97 бит атаковать нужно, подбирая 21 выходных бит «нечеткого экстрактора».

Синтез личного криптографического ключа из неоднозначных компонент биометрических данных

Особенностью биометрии является то, что каждый пример реализации одного и того же биометрического образа отличается от других примеров. Для выделения из них нестабильной компоненты достаточно взять два примера биометрического образа, осуществить центрирование и нормирование их данных и получить их разность. Эта ситуация отображена на рисунке 1. Если теперь проквантовать эту разность, мы получим случайную последовательность.

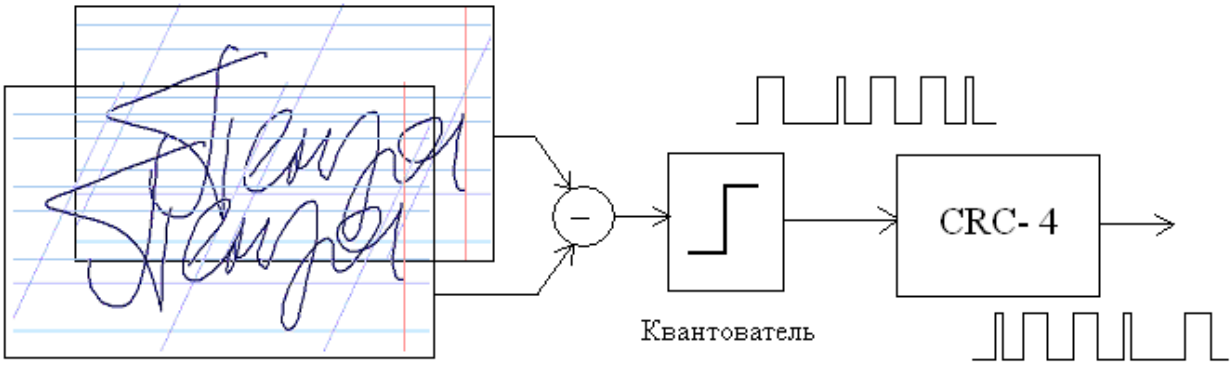

Рис. 1 Получение случайной последовательности из двух примеров одного биометрического образа

При вычитании стабильная часть примеров рукописного биометрического образа «*Пенза*» устраняется, подчеркивается его нестабильная часть. Появление нестабильной части биометрических примеров обусловлено влиянием множества факторов, точно повторить рукописный образ человек не может даже в случае попытки его обвода по шаблону.

Для того, что бы усилить, полученную случайную последовательность достаточно получить другую псевдослучайную последовательность от программного генератора и сложить их по модулю два. Можно поступить иначе и усилить случайную последовательность пропустив ее через нелинейную рекурсивную свертку, например, CRC-4 (подсчет контрольных сумм).

Если бы мы вычислили криптографическую хэш-функцию от полученной случайной последовательности [3], то можно было бы ее использовать для получения личного ключа без исследования ее качества. То, что мы, ориентируясь на малые вычислительные ресурсы, применили не криптографическое хэширование, в форме рекурренты CRC-4, приводит к необходимости оценки энтропии ключа. Заранее неизвестно достаточно ли стандартная рекуррента CRC-4 увеличивает энтропию естественной нестабильности примеров биометрического образа.

Контроль энтропии в пространстве однобитных сверток Хэмминга

Если нами был бы создан ключ в 256 бит (например, для последующего обучения нейросети среды «БиоНейроАвтограф»), то оценить его энтропию по Шеннону технически невозможно. В связи с этим приходится от обычных кодов переходить в пространство сверток Хэмминга [2]:

$$h = 256 - \sum_{i=1}^{256} ("c_i") \oplus ("x_i") \qquad (1),$$

где - "c_i" - i-тый разряд проверяемого разряда кода синтезированного ключа, "x_i" - эталонная последовательность «не белого» шума.

Если мы будем иметь ключ "$\overline{c}$" с почти независимыми разрядами, то множество сверток Хэмминга (1) с фрагментами «не белого» шума должно дать следующие статистические моменты:

$$\begin{cases} E(h) = 128 \ \ бит \\ \sigma(h) = 8 \ \ бит \end{cases} \qquad (2).$$

В случае, если математическое ожидание Хэмминга E(h) будет существенно отличаться от значения 128 бит мы получим отсутствие баланса равновероятных состояний «0» и «1» во всех разрядах кода. Если стандартное отклонение σ(h) будет

существенно больше 8 бит, то мы должны сделать вывод о сильной корреляционной связи разрядов синтезированного кода ключа.

Контроль энтропии в пространстве 8-ми битных кодировок

В качестве эталонных последовательностей «не белого» шума, например, могут быть взяты фрагменты текста на русском языке [12, 13, 14]. Каждый из фрагментов текста на русском языке при вычислении расстояний Хэмминга фактически накрывается гаммой синтезированного ключа. Выполнение условия (2) эквивалентно покрытию текста идеальной гаммой. После чего защищенный гаммой текст становится «белям» шумом.

Следует отметить, что снижение криптографических свойств синтезированного ключа "$\bar{c}$" возникают в очевидном случае замены его на код осмысленного пароля [13, 14]. Естественно, что улучшение криптографических качеств ключа "$\bar{c}$" будет монотонно приближаться к выполнению условий (2).

В [13, 14] показано, что при вычислениях энтропии русского языка приходится учитывать 8-ми кодировку русских и английских текстов. В этом случае расстояние Хэмминга будет вычисляться следующим образом:

$$h_8 = \sum_{i=1}^{32} \left| "c_i, c_{i+1}, ..., c_{i+7}" - "x_i, x_{i+1}, ..., x_{i+7}" \right| \qquad (3),$$

где "$c_i, c_{i+1}, ..., c_{i+7}$" - 8 бит кода в i-том фрагменте кода ключа, "$x_i, x_{i+1}, ..., x_{i+7}$" - 8 бит кода образцового текста, принадлежащего i-тому знаку эталонной последовательности.

Очевидно, что для свертки Хэмминга по модулю 256 (3) могут быть найдены ограничения на математическое ожидание и стандартное отклонение типа (2). Пользуясь близостью к этим ограничениям, мы всегда сможем оценить то, на сколько тестируемая последовательность "$\bar{c}$" , близка к «белому» шуму.

Для нас принципиально важно то, что результаты тестирования через однобитные свертки Хэмминга (1) и тестирование в 8-ми битных свертках Хэмминга по модулю 256 сильно различаются [14].

Множество сверток Хэмминга, вычисленных обычными процессорами в системах счисления кратных двум

Фактически мы имеем две свертки Хэмминга по модулю 2 (выражение (1)) и по модулю 256 (выражение (3)). Свертку по модулю два (1) можно рассматривать как эквивалент теста стаи обезьян набирающих текст на русском языке до его полного совпадения с контрольным фрагментом на однобитной печатающей машинке. Светку по модулю 256 32 знаков образцового русского текста (3) следует рассматривать как обычную печатающую машинку в руках стаи обезьян.

Исходя из этой интерпретации мы можем предложить промежуточные свертки, для других печатающих машинок:

$$h_2 = \sum_{i=1}^{128} \left| "c_i, c_{i+1}" - "x_i, x_{i+1}" \right| \qquad (4),$$

$$h_4 = \sum_{i=1}^{64} \left| "c_i, c_{i+1}, ..., c_{i+3}" - "x_i, x_{i+1}, ..., x_{i+3}" \right| \qquad (5),$$

$$h_{16} = \sum_{i=1}^{32} \left| "c_i, c_{i+1}, ..., c_{i+15}" - "x_i, x_{i+1}, ..., x_{i+15}" \right| \qquad (6),$$

$$h_{32} = \sum_{i=1}^{16} \left| "c_i, c_{i+1}, ..., c_{i+31}" - "x_i, x_{i+1}, ..., x_{i+31}" \right| \qquad (7),$$

$$h_{64} = \sum_{i=1}^{8} \left| "c_i, c_{i+1}, \ldots, c_{i+63}" - "x_i, x_{i+1}, \ldots, x_{i+63}" \right| \quad (8).$$

Множество процессоров с нечетным числом бит, на которых могут быть построены печатающие машинки для стаи обезьян

Все описанные выше конструкции легко согласуются с 2, 4, 8, 16, 32, 64 битными процессорами под создаваемые для стаи обезьян печатающие машинки с конвертируемой друг в друга арифметикой. Однако «белый» шум не знает о людях и о том, что для стаи обезьян людям удобно писать процессоры печатающих машинок с системами счисления кратными двум. Люди вполне могут пойти на отбрасывание малой части анализируемого кода и написать еще одну стаю процессоров под обезьян с 3-мя битами, 5-ю битами, 7 битами и так далее:

$$h_3 = \sum_{i=1}^{85} \left| "c_i, c_{i+1}, c_{i+2}" - "x_i, x_{i+1}, x_{i+2}" \right| \quad (9),$$

$$h_5 = \sum_{i=1}^{51} \left| "c_i, c_{i+1}, \ldots, c_{i+4}" - "x_i, x_{i+1}, \ldots, x_{i+4}" \right| \quad (10),$$

$$h_7 = \sum_{i=1}^{36} \left| "c_i, c_{i+1}, \ldots, c_{i+6}" - "x_i, x_{i+1}, \ldots, x_{i+6}" \right| \quad (11),$$

::

На ряду с 9 печатающими машинками, имеющими процессоры с четным числом бит, мы имеем дополнительно 247 печатающих машинок с процессорами, имеющими нечетное число разрядов, вполне пригодными для вычисления еще 247 сверток Хэмминга.

Мы имеем ситуацию, когда каждая обезьяна будет вооружена не одной, а 256 печатающими машинками, каждая из которых проверяет, не совпал ли набираемый обезьяной текст с осмысливаемыми эталонами на русском. Применяя подобную технологию, мы повышаем надежность контроля качества ключа (близость к эталонному «белому» шуму) в 256 раз надежнее, чем обычный тест стаи обезьян, рекомендуемый NIST. В место того, что бы рассматривать уровень случайности по этому тесту со стороны только одной пишущей машинки в 8-битной кодировке ASCII, мы одновременно рассматриваем ситуацию с точки зрения 256 пишущих машинок, каждая из которых работает в своей кодировке.

Естественно, что столь большое число наблюдателей избыточно. По рекомендациям NIST США обычно применяют 16 тестов на случайность. Видимо, при анализе качества ключей, полученных из нестабильной части биометрических данных будет достаточно применения первых 16-ти пишущих машинок под каждую из обезьян.

Следует так же отметить, что этот тип задач по созданию быстрых алгоритмов оценки энтропии длинных кодов применим достаточно широко. В частности такой подход может улучшить решение обратных задач нейросетевой биометрии [15].

Литература:

1. Волчихин В.И., Иванов А.И., Фунтиков В.А. Быстрые алгоритмы обучения нейросетевых механизмов биометрико-криптографической защиты информации. Монография. Пенза-2005 г. Издательство Пензенского государственного университета, 273 с.
2. Малыгин А.Ю., Волчихин В.И., Иванов А.И., Фунтиков В.А. Быстрые алгоритмы тестирования нейросетевых механизмов биометрико-криптографической защиты

информации /Пенза-2006 г., Издательство Пензенского государственного университета., 161 с.
3. Техническая спецификация (проект, публичное обсуждение начато с 01.02.2017 членами ТК 26 «Криптографическая защита информации») ЗАЩИТА НЕЙРОСЕТЕВЫХ БИОМЕТРИЧЕСКИХ КОНТЕЙНЕРОВ С ИСПОЛЬЗОВАНИЕМ КРИПТОГРАФИЧЕСКИХ АЛГОРИТМОВ
4. Dodis Y., Reyzin L., Smith A. Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy // Proc. EUROCRYPT, April 13, pages 523-540, 2004.
5. Monrose F., Reiter M., Li Q., Wetzel S. Cryptographic key generation from voice. //Proc. IEEE Symp. on Security and Privacy, pp. 202-213, 2001.
6. Ramírez-Ruiz J., Pfeiffer C., Nolazco-Flores J. Cryptographic Keys Generation Using FingerCodes. //Advances in Artificial Intelligence - IBERAMIA-SBIA 2006 (LNCS 4140), p. 178-187, 2006
7. Hao F., Anderson R., Daugman J. Crypto with Biometrics Effectively //IEEE TRANSACTIONS ON COMPUTERS, VOL. 55, NO. 9, SEPTEMBER, Page(s):1073 – 1074, 2006.
8. Иванов А.И. Нечеткие экстракторы: проблема использования в биометрии и криптографии. // Первая миля. № 1, 2015 г. с. 40-47.
9. Иванов А.И. Сопоставительный анализ показателей конкурирующих технологий биометрико-криптографической аутентификации личности. «Защита информации. ИНСАЙД» № 3 2014 г., с. 32-39.
10. 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» http://пниэи.рф/activity/science/noc./bioneuroautugraph.zi для свободного использования университетами России, Белоруссии, Казахстана, Армении.
11. Иванов А.И. АВТОМАТИЧЕСКОЕ ОБУЧЕНИЕ БОЛЬШИХ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ В БИОМЕТРИЧЕСКИХ ПРИЛОЖЕНИЯХ. Учебное пособие к пакету лабораторных работ, выполняемых в среде моделирования "БиоНейроАвтограф" http://пниэи.рф/activity/science/noc/tm_IvanovAI.pdf , электронная книга издательства ОАО «ПНИЭИ», Пенза-2013 г. 27 с.
12. Иванов А.И. Многомерная нейросетевая обработка биометрических данных с программным воспроизведением эффектов квантовой суперпозиции. Издательство АО «ПНИЭИ», Пенза-2016 г., 133 с. Свободный доступ http://пниэи.рф/activity/science/BOOK16.pdf
13. Волчихин В.И., Иванов А.И., Юнин А.П., Малыгина Е.А. Многомерный портрет цифровых последовательностей идеального «белого шума» в свертках Хэмминга // «Известия высших учебных заведений. Поволжский регион. Технические науки.» 2017 г. №4. (стр.??, принята к опубликованию, находится в печати)
14. Юнин А.П., Корнеев О.В. Оценка энтропии легко запоминаемых, длинных паролей со смыслом в ASCII кодировке для русского и английского языков //Тестирование стойкости средств высоконадежной биометрической защиты к атакам подбора». Труды научно-технической конференции кластера пензенских предприятий, обеспечивающих БЕЗОПАСНОСТЬ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ. Том 10, Пенза-2016, с. 40-42 (http://пниэи.рф/activity/science/BIT/T10-p40.pdf)
15. 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. Т27. №4, 2017, с 518-523.