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)

УДК: 519.2, 612.087, 621.319.7

Юнин А.П., Корнеев О.В.

Оценка энтропии легко запоминаемых, длинных паролей со смыслом в ASCII кодировке для русского и английского языков

В настоящее время безопасность информационных ресурсов, как правило, построена на использовании запоминаемых паролей доступа. При этом пользователи саботируют использование длинных паролей, сформированных из случайных символов. Пользователи не могут и не хотят запоминать длинные случайные пароли, однако легко могут запоминать длинные осмысленные тексты. В связи с этим возникает задача оценки стойкости легко запоминаемых, осмысленных паролей доступа.

Очевидно, что с этой задачей легко можно справиться, определив энтропию длинного осмысленного пароля. При этом энтропию длинного, осмысленного пароля следует рассматривать в кодировке букв (знаков) носителя языка и в некоторой альтернативной кодировке. По IP адресу всех пользователей можно разделить на несколько групп, предпочитающих работать в той или иной кодировке. В России пользователи, скорее всего, будут набирать длинный осмысленный пароль в кириллической кодировке или в латинской кодировке. Переключение кодировок перед вводом парольной фразы может являться одним из секретов парольного доступа.

Следует подчеркнуть, что определение энтропии пароля по Шеннону является задачей высокой вычислительной сложности. Это связано с тем, что ориентируясь на Шеннона, придется использовать большие объемы текстов и ждать появление очень редких событий. Определение энтропии по Шеннону является задачей экспоненциальной вычислительной сложности $(2^8)^{a_r \cdot n}$ для ASCII кодировки. Степенной показатель вычислительной сложности - $(a_r \cdot n)$ линейно связан с длиной пароля – n. Так же этот показатель линейно связан с коэффициентом понижения размерности - a_r, возникающим из-за корреляционных связей между буквами в языке осмысленного пароля.

Легко показать, что оценка энтропии длинных паролей по Шеннону на обычной вычислительной машине оказывается технически не выполнима даже для паролей, состоящих из 8 букв. В связи с этим, для ускорения вычислений следует использовать переход в пространство расстояний Хэмминга [1, 2] между кодом пароля и скользящим кодом текста на языке осмысленного пароля. При этом оценка энтропии пароля строится на предсказании редких событий (ожидание редких событий по Шеннону замещается на предсказание вероятности редких событий в пространстве расстояний Хэмминга). В свою очередь, возможность предсказаний опирается на факт нормального распределения расстояний Хэмминга для кодов с длиной более 32 бит (паролей из 4 букв в 8 битной ASCII кодировке). Это позволяет в частности оценивать энтропию сильно коррелированных откликов нейронной сети на биометрические образы [3, 4, 5].

Для парольной фразы «Мишка косолапый» (два пробела между словами, 16 символов в парольной фразе) распределение расстояний Хэмминга между

ASCII кодом и тестовым текстом на русском длиной в 316 символов приведено на рисунке 1.

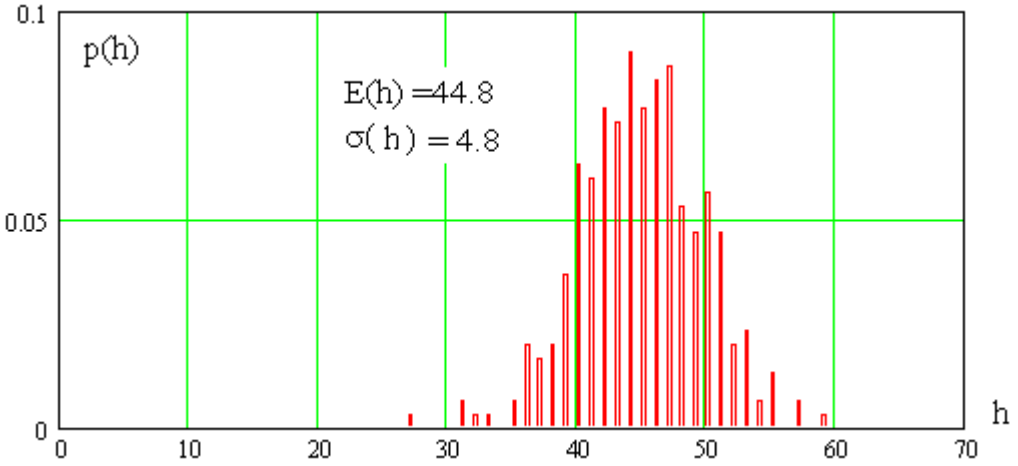

Рис. 1. Распределение расстояний Хэмминга для пароля на русском языке «Мишка косолапый», длина проверочного текста на русском языке 316 знаков

Если считать это распределение нормальным, вероятность ошибок второго рода оценивается следующим образом:

$$P_2 = \frac{1}{\sigma(h)\sqrt{2\pi}} \int_{-\infty}^{1} \exp\left\{\frac{-(E(h)-u)^2}{2\cdot(\sigma(h))^2}\right\} du = 3.6\cdot 10^{-20} \qquad (1).$$

Если же мы для парольной фразы на русском языке будем вычислять расстояния Хэмминга для тестового текста на английском языке, мы получим распределение расстояний Хэмминга, приведенное на рисунке 2.

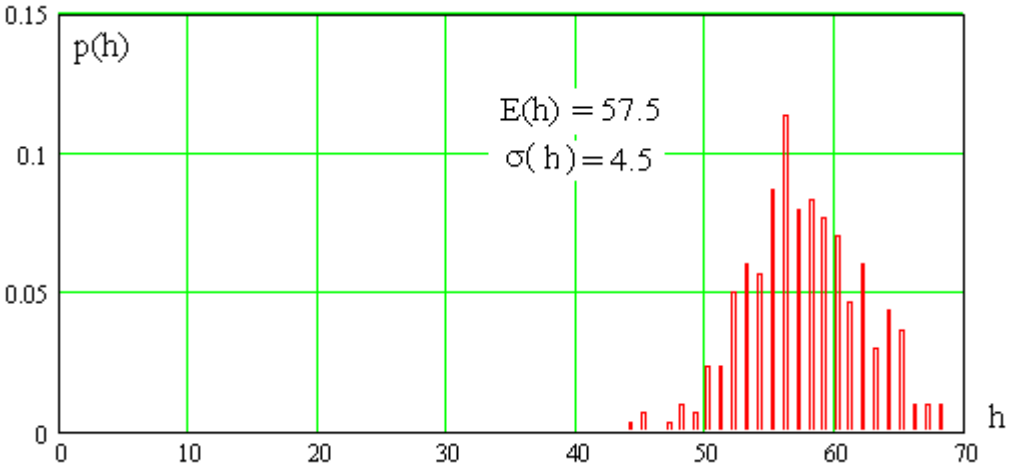

Рис. 2. Распределение расстояний Хэмминга для пароля на русском языке «Мишка косолапый», длина проверочного текста на английском языке 316 знаков

Сравнивая рисунок 1 и рисунок 2 легко заметить увеличение математического ожидания - E(h) и стандартного отклонения - σ(h) расстояний Хэмминга при тестировании русской парольной фразы на английском тексте. Это приводит к ощутимому снижению вероятности ошибок второго рода до величины $P_2 = 6.6\cdot 10^{-46}$.

К сожалению, подобные оценки вероятности ошибок второго рода, приведенные ранее в работах [1, 2], оказались слишком оптимистичными. Более достоверная оценка получаются, если вычислять расстояния Хэмминга по модулю

256 (2^8) (модуль соответствует 8 битной ASCII кодировке). В этом случае, наблюдается значительное снижение относительного значения математического ожидания и значительное увеличение стандартного отклонения (рисунок 3).

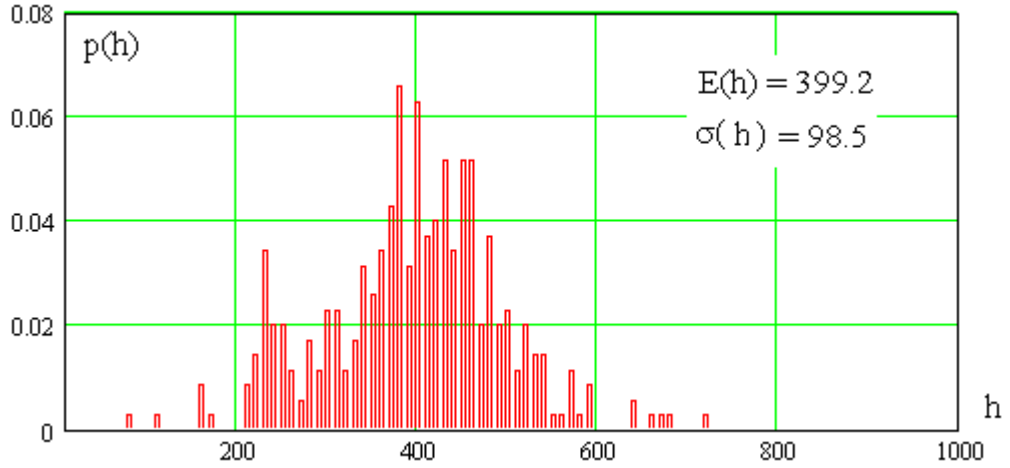

Рис. 3. Распределение расстояний Хэмминга по модулю 256 для пароля на русском языке «Мишка косолапый», длина проверочного текста на русском 316 знаков

Распределение расстояний Хэмминга по модулю 256 дает повышение значения вероятности ошибок второго рода до величины $P_2 = 2.6 \cdot 10^{-4}$. Вероятность ошибки второго рода выросла на 16 десятичных порядков. Это означает, что операция переноса двоичных разрядов при бинарном суммировании 8 битных кодов играет роль хеширующей (перемешивающей) функции в пространстве расстояний Хэмминга. Энтропия паролей на русском языке составляет $-\log_2(P_2)/16 = 0.95$ бита, на один символ.

ЛИТЕРАТУРА:

1. Иванов А.И., Фунтиков В.А., Майоров А.В., Надеев Д.Н. Моделирование кодовых последовательностей с энтропией естественных и искусственных биометрических языков. Инфокоммуникационные технологии Том 8, № 4, 2010 г., с. 75-79, http://ikt.psuti.ru.
2. Малыгина Е.А., Иванов А.И., Язов Ю.К., Надеев Д.Н. Прогнозирование значений энтропии длинных кодовых последовательностей, порождаемых естественными и искусственными языками «Инфокоммуникационные технологии» том 12, № 2 2014, с.12-15 http://ikt.psuti.ru
3. ГОСТ Р 52633.3-2011 «Защита информации. Техника защиты информации. Тестирование стойкости средств высоконадежной биометрической защиты к атакам подбора».
4. Иванов А.И. Нечеткие экстракторы: проблема использования в биометрии и криптографии. // Первая миля. № 1, 2015 г. с. 40-47.
5. 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

Статья поступила 29.09.2016, опубликована 13.11.2016 по положительной рецензии к.т.н. Зефирова С.Л.