УДК 004.7

Н.Л. Корепанова, доцент, канд. техн. наук
Севастопольский национальный технический университет
ул. Университетская 33, г. Севастополь, Украина, 99053
E-mail: kvt@sevgtu.sebastopol.ua

АНАЛИЗ СТОЙКИХ КРИПТОСИСТЕМ НА ОСНОВЕ ИНФОРМАЦИОННО-ТЕОРЕТИЧЕСКОГО И ПРАКТИЧЕСКОГО ПОДХОДА

Анализируются криптосистемы на основе критерия стойкости при использовании информационно-теоретического и практического подходов.

***Ключевые слова**: криптосистема, шифрование, стойкость.*

Основным назначением криптосистем является обеспечение передачи секретных сообщений через несекретные, незащищенные каналы связи. Поэтому важнейшей характеристикой любой криптосистемы является ее стойкость, то есть способность противостоять попыткам дешифрования перехваченных шифротекстов или раскрытию ключей шифра. Цель исследований – оценка стойкости криптосистем при использовании информационно-теоретического и практического подходов.

Информационно-теоретический подход определения стойкости криптосистем основан на понятии информации, ее количественной оценке и анализе количества информации об открытом тексте, необходимой для обеспечения стойкости криптосистемы. Информация, несмотря на разнородность ее качества, допускает разумное количественное измерение. Кроме того, оказывается возможным оценить количество информации, необходимой криптоаналитику для успешного проведения криптоатаки - фактически оценить стойкость шифра. Введем основные положения указанных теорий, необходимые для изложения информационно-теоретического подхода к оценке стойкости криптосистем.

Способ измерения количества информации при помощи ее сравнения с информацией представленной в виде некоторого числа двоичных знаков (битов 0 или 1) стал сейчас стандартом в теории информации [1]. Пусть задан конечный алфавит $A = \{a_1 \ldots a_k\}$. Тогда формально можно составить $N = k^n$ «текстов» представленных из n букв. При этом за количество информации, содержащейся в указании одного из текстов, принять величину

$$k \log I = n \log_2 k. \quad (1)$$

Однако на практике реальные языковые тексты можно передавать значительно более кратким способом, так как число N' «осмысленных» текстов из n букв значительно меньше N.

Поэтому в принципе возможна идеальная система записи, которая требовала бы для кодирования этих N' осмысленных текстов лишь $I' \approx \log_2 N'$ двоичных знаков.

Особенностью языковых систем является то, что буквы алфавита встречаются в них с различной частотой. Если обозначить через $n_1,\ldots, n_k$ - количество вхождений букв $a_1,\ldots,a_k$ в текст длины n, причем $n1 + n2 + \ldots + nk = n$, то число текстов длины n сократится до

$$N^* = \frac{n!}{n1!\,n2!\ldots nk!}.$$

Используя слабую форму формулы Стирлинга легко подсчитать, что при достаточно больших ni

$$\mathrm{Log}_2(n!) \quad n \log_2 n,$$

$$I^* = \log_2 N^* - n \sum_{i=1}^{k} p_i \log_2 p_i,$$

где p_i - частоты появления отдельных букв в тексте, т.е.

$$p_i = \frac{n_i}{n}$$

Таким образом, количество информации, приходящейся на одну букву текста, равно

$$H = -\sum_{i=1}^{k} p_i \log_2 p_i. \quad (2)$$

В случае же равных частот, вхождения букв алфавита в текст получим прежнюю оценку количества информации в виде

$$H = \log_2 k, \quad (3)$$

а при любых других распределениях частот

$$H < \log_2 k.$$

Обобщим понятие количества информации, конкретизируя его применительно к криптоанализу. Формула (2) являясь статистической (в ней p_i являются частотами) допускает также вероятностную интерпретацию. Пусть X - некоторое сообщение, которое может быть выбрано случайным образом из n возможных текстов x_1, …, x_n. И пусть дано распределение вероятностей случайной величины X:) $P\{X = x_i\} = P(x_i)$.

Если дано только распределение $P(x_i)$, то остается неопределенность в ответе на вопрос, каким именно из текстов x_1, …, x_n является сообщение X. Числовое значение неопределенности распределения $P(x_i)$ или его энтропию можно вычислить по формуле:

$$H(X) = -\sum_{i=1}^{k} P(x_i) \log_2 P(x_i). \quad (4)$$

Формула (4) выражает количество информации, необходимое для устранения неопределенности в задании X с помощью распределения $P(x_i)$, т.е. количество информации, необходимой для точного указания сообщения X.

Аналогично, как в формуле (3), в случае, когда все сообщения равновероятны, величина $H(X)$ принимает максимальное значение, равное

$$H_{\max(x)} = \log_2 n .$$

Энтропия уменьшается (с увеличением количества информации о сообщении X), когда распределение вероятностей сообщений начинает отличаться от равновероятного и может достичь минимального значения при $P(x_i) = 1$, для некоторого сообщения x_i:

$$H_{\min(x)} = 0 .$$

Таким образом, применительно к криптоанализу, энтропия сообщения X измеряет его неопределенность числом битов информации, которые должны быть восстановлены, когда сообщение скрыто от криптоаналитика в шифротексте.

Сформулируем определение стойкости криптосистемы в рамках информационно теоретического подхода. По Шеннону, криптосистема называется теоретически стойкой, если криптоаналитик не может уточнять распределение вероятностей возможных открытых текстов по имеющемуся у него шифротексту, сформированному в этой системе, даже если он обладает всеми необходимыми для этого средствами [2].

Совершенная секретность криптосистемы достигается, если открытый текст M и шифротекст C статистически независимы, т.е.

$$P(\mathrm{M}|\mathrm{C}) = P(\mathrm{M}),$$

для всех возможных M и C, а значит, получение шифротекста не дает криптоаналитику никакой информации о посланном открытом тексте, все они равновероятны.

Также можно сделать вывод о том, что неопределенность секретного ключа должна быть не меньше, чем неопределенность шифротекста, полученного с его помощью. А значит, размер секретного ключа k должен быть не меньше размера открытого текста M. Такое условие является неудобным с практической точки зрения, так как передаваемые сообщения могут быть достаточно длинными.

Криптосистема называется практически стойкой, если в течение разумного количества времени шифротексты этой системы не могут быть вскрыты всеми известными методами криптоанализа.

Из этого определения можно сделать вывод о том, что проблема создания практически стойких криптосистем может быть сведена к проблеме нахождения решений наиболее сложных задач - факторизации и потенцирования больших чисел (более 1024 бит).

Этот подход к определению стойкости криптосистем основан на понятии вычислительной сложности криптоаналитических алгоритмов, в отличие от информационно-теоретического подхода, рассмотренного выше. В нем рассматривается вопрос не о том, возможно ли извлечь информацию об открытом тексте из анализа шифротекста, а осуществимо ли это в приемлемое время.

Можно создать шифр таким образом, чтобы криптоатака на него включала в себя некоторую задачу, про которую известно, что для ее решения требуется достаточно большой объем вычислительной работы. Этот подход позволяет достичь свойства совершенной секретности криптосистемы даже для случаев, когда используемые секретные ключи значительно меньше по размерам, чем длина открытого текста.

Перспективой дальнейших исследований является выбор наиболее сложных задач, для которых известно решение, и использование их для построения криптосистем с практически стойкими шифрами.

Библиографический список использованной литературы

1. Колмогоров А.Н. Теория информации и теория алгоритмов / А.Н. Колмогоров. — М.: Наука, 1987. — 345 с.

2. Галуев Г.А. Математические основы криптологии: учебно-методическое пособие / Г.А. Галуев. — Таганрог: Изд-во ТРТУ, 2003. — 40 с.

Поступила в редакцию 07.11.2011 г.

Корепанова Н.Л. Аналіз стійкості криптосистем на основі інформаційно-теоретичних і практичних підходів

Аналізуються криптосистеми на основі критерію стійкості під час використання інформаційно-теоретичного та практичного підходу.

Ключові слова: криптосистема, шифрування, стійкість.

Korepanova N.L. Analysis of stable cryptosystems based on the information-theoretical and practical approaches

Analyzed are the cryptosystem based on the criteria of stability in the use of information and theoretical and practical approaches.

Keywords: cryptography, encryption, strength.