Наименьшее натуральное число γ, такое что $f|(x^\gamma - 1)$, называется *порядком многочлена* f, и обозначать символом $ord\ f(x)$. (Если $f(0) = 0$, то полагаем $f(x) = x^\alpha g(x)$, $g(0) \neq 0$, и $ord\ f(x) = ord\ g(x)$.)

Многочлен $f \in F_p[x]$ степени n над полем $\mathbb{F}_p$ называется *примитивным*, если $ord\,(f(x)) = p^n - 1$. То есть примитивным будет многочлен, имеющий максимальный возможный порядок [5]. Таким образом, если характеристический многочлен $f(\lambda)$, порождающий линейное рекуррентное уравнеие, неприводим и примитивен, то генерируемые соответствующим линейным рекуррентным уравнением последовательности будут иметь максимально возможный период.

Учитывая что для любой пары (p, n) формула

$$b_p(n) = \tfrac{\varphi(p^n-1)}{n}[5].$$

вычисляет число $b_p(n)$ примитивных многочленов степени n над полем $\mathbb{F}_p$ и эти числа неизменно положительны, то получаем, что для любого простого числа p и любого натурального числа n существует примитивный над полем $\mathbb{F}_p$ многочлен степени n.

Учитывая же свойства порядка многочленов над конечным полем, а именно что

$$\operatorname{ord}(f(x))^n = p^t \operatorname{ord} f(x),$$

где t - наименьшее целое число, такое что $p^t \geq n$, и $f(x) \in F_p[x]$ - неприводимый многочлен, получаем получаем что всегда можно сгенерировать последовательность с периодом $(p^n-1)p^t$ для любого простого p и натуральных n и t.

Также рассматриваются более сложные конструкции n непочредственно для поля $\mathbb{F}_2$, так как именно такие последовательности нужны для решенияч практических задач.

СПИСОК ЦИТИРОВАННОЙ ЛИТЕРАТУРЫ

1. Деза Е.И. Специальные числа натурального ряда. - М.: URSS, 2010.
2. Деза Е.И., Котова Л.В. Введение в криптографию. - М.: URSS, 2018.
3. Деза Е.И., Котова Л.В. Рекуррентные числовые последовательности: теория и приложения // Чебышевский сборник. 2022. Том 23, № 3. С. 77-101.
4. Конечные поля : В 2 т. / Р. Лидл, Г. Нидеррайтер ; перевод с англ. А. Е. Жукова, В. И. Петрова ; под ред. В. И. Нечаева. - Москва : Мир, 1988.
5. Нечаев В.И. Основы защиты информации. - М.: МГУ, 1999.

УДК 511.42

О влиянии неравновероятности выходной последовательности на качество криптографических преобразований

А. Б. Лось (Россия, г. Москва)
Московский институт электроники и математики им. А. Н. Тихонова Национального исследовательского университета «Высшая школа экономики»
e-mail: alos@hse.ru
А. Ю. Нестеренко (Россия, г. Москва)
Московский институт электроники и математики им. А. Н. Тихонова Национального исследовательского университета «Высшая школа экономики»
e-mail: nesterenko_a_y@mail.ru

О. А. Рогачёва (Россия, г. Москва)
Московский институт электроники и математики им. А. Н. Тихонова Национального исследовательского университета «Высшая школа экономики»
e-mail: oarogacheva@hse.ru

On the Influence of Unequal Probability of the Output Sequence on the Quality of Cryptographic Transformations

A. B. Loss (Russia, Moscow)
HSE Tikhonov Moscow Institute of Electronics and Mathematics (MIEM HSE)
e-mail: alos@hse.ru
A. Yu. Nesterenko (Russia, Moscow)
HSE Tikhonov Moscow Institute of Electronics and Mathematics (MIEM HSE)
e-mail: nesterenko_a_y@mail.ru
O. A. Rogacheva (Russia, Moscow)
HSE Tikhonov Moscow Institute of Electronics and Mathematics (MIEM HSE)
e-mail: oarogacheva@hse.ru

1. Постановка задачи

Пусть знаки $a_i, a_i \in A_n, i = 1, 2, \ldots, N$, входного сообщения $\bar{a}^N = a_1, a_2, \ldots, a_N$ некоторого источника сообщений $M^N = \{\bar{a}^N\}$, выбираются из алфавита $A_n = \{1, 2, \ldots, n\}$, а знаки $b_i, i = 1, 2, \ldots, N$ выходного (шифрованного) сообщения $\bar{b}^N = (b_1, b_2, \ldots, b_N)$, $b_i \in A_n$, образуются из знаков входного сообщения в соответствии с уравнением:

$$\bar{b}^N = F_k(\bar{a}^N),$$

где $F_k(\cdot)$ - некоторая функция, зависящая от ключа k, $k \in K$ и определяемая применяемым алгоритмом шифрования ([1-2]), K – множество всех ключей криптографического алгоритма.

Для дальнейшего изложения введем следующую теоретико-вероятностную модель.

Обозначим $E^N = \{\bar{b}^N = (b_1, b_2, \ldots, b_N) | b_i = F_k(a_i), i = 1, 2, \ldots, N, k \in K\}$ - множество всех шифрованных текстов, которые могут быть получены в результате применения уравнения шифрования (1) при различных ключах k из множества K.

Зададим на множестве входных сообщений M^N и множестве ключей K некоторые независящие друг от друга вероятностные распределения $P(\cdot)$ и $\theta(\cdot)$:

$$P(M^N) = \{P(\bar{a}^N), \bar{a}^N \in M^N\}, \sum_{\bar{a}^N \in M^N} P(\bar{a}^N) = 1,$$

$$\theta(K) = \{\theta_k, k \in K\}, \sum_{k \in K} \theta(k) = 1.$$

Очевидно, что распределения $P(\cdot)$ и $\theta(\cdot)$ индуцируют на множестве шифрованных сообщений некоторое вероятностное распределение $Q(\cdot)$:

$$Q(E^N) = \{Q(\bar{b}^N) | \bar{b}^N \in E^N\}, \sum_{\bar{b}^N \in E^N} Q(\bar{b}^N) = 1.$$

Обозначим через $H(X)$ - энтропию некоторого ансамбля X, через $H(X/Y)$ - условную энтропию ансамбля X при заданном ансамбле Y, а через $I(X, Y) = H(X) - H(X/Y)$ - взаимную информацию ансамблей X и Y ([3-4]).

Далее в работе нас будет интересовать значение взаимной информации ансамбля входных M^N и шифрованных E^N сообщений $I(M^N, E^N) = H(M^N) - H(M^N/E^N)$.

2. Оценка взаимной информации входных и выходных сообщений

Для дальнейшего изложения введем дополнительные обозначения, полагая:

A_n^i - множество векторов длины i в алфавите A_n,

$M^i = \{\alpha^i = (\alpha_1, \ldots, \alpha_i) \in A_n^i \mid$ существует $\bar{a}^N = (a_1, \ldots, a_N) \in M^N$, такое что $a_j = \alpha_j, j = \{1, \ldots, i.\}\}$,

$E^i = \{\beta^i = (\beta_1, \ldots, \beta_i) \in A_n^i \mid$ существует $b^N = (b_1, \ldots, b_N) \in E^N$ такое, что $b_j = \beta_j, j = \{1, \ldots, i\}\}$,

$E_i = \{\beta_i \in A_n \mid$ существует $b^N = (b_1, \ldots, b_N) \in E^N$ такое что $b_i = \beta_i\}$.

Будем также обозначать через A^i, B^i и B_i случайные величины, принимающие значения соответственно из множеств M^i, E^i, E_i и имеющие распределение вероятностей:

$P\{A^i = a^i\} = \sum_{\alpha^N \in M^N, \alpha_j = a_j, j=1,\ldots,i} P(\alpha^N)$

$P\{B^i = b^i\} = \sum_{\bar{\beta}^N \in E^N, \beta_j = b_j, j=1,\ldots,i} Q(\beta^N)$

$P\{B_i = b_i\} = \sum_{\beta^N \in E^N, \beta_i = b_i} Q(\beta^N)$.

Теорема 1. *Пусть распределение $Q(\cdot)$ таково, что вероятность появления знака шифрованного сообщения b_i не зависит от последующих знаков входного и шифрованного сообщения $(a_{i+1}, \ldots, a_N)$ и $(b_{i+1}, \ldots, b_N)$, а для любого значения величины $i = 1, 2, \ldots, N$ имеет место соотношение:*

$$p\{B_i = b_i / A^i = a^i; B^{i-1} = b^{i-1}\} = \frac{1}{n} + \frac{1}{n}\Delta_{b_i}(a^i, b^{i-1}), \tag{1}$$

где $|\Delta_{b_i}(a^i, b^{i-1})| \leq \delta_i, 0 \leq \delta_i < 1, a^i \in M^i, b^{i-1} \in E^{i-1}, b_i \in E_i$.

Тогда имеет место равенство:

$$H(M^N/E^N) = H(M^N) - \sum_{i=1}^{N} \varepsilon_i, \tag{2}$$

где

$$0 \leqslant \varepsilon_i \leqslant \frac{1}{2 \cdot \ln 2}\left[\delta_i^2 + \frac{\delta_i^3}{3\,(1-\delta_i)^3}\right].$$

Условие невозможности восстановления входного сообщения

Результат теоремы 1 позволяет при известных величинах δ_i оценить условную энтропию $H(M^N/E^N)$ ансамбля входных сообщений M^N при заданном ансамбле криптограмм E^N.

В соответствии с [4], при достаточно больших значениях величины N, можно полагать:

$$H(M^N) = N \cdot H_0, \tag{3}$$

где H_0 — средняя энтропия на знак открытого текста.

Какие либо содержательные результаты прикладных и научных исследований по оценке энтропии H_0 русского языка авторам неизвестны. Среди электронных ресурсов можно указать работы [5] и [6], в которых указаны значения средней энтропии на знак текстов на русском языке H_0, соответственно, $0,6$ и $0,7$ дв. ед.

Из результатов теоремы 1 далее получаем:

$$H(M^N/E^N) \geqslant N \cdot H_0 - \varepsilon \cdot N, \tag{4}$$

где

$$\varepsilon = \frac{1}{2\ln 2}\left[\delta^2 + \frac{\delta^3}{3\,(1-\delta)^3}\right], \quad \delta = \max \delta_i.$$

Отсюда получаем неравенство:

$$\frac{H(M^N)}{H(M^N/E^N)} \leqslant \frac{H_0}{H_0 - \varepsilon} = 1 + \frac{\varepsilon}{H_0 - \varepsilon}. \tag{5}$$

Для иллюстрации полученных результатов приведем (таблицы 1 и 2) данные численных расчетов величины

$$\Lambda = \frac{\varepsilon}{H_0 - \varepsilon}$$

для значений энтропии $H_0 = 0,5$ и $H_0 = 1$, значений параметра $\delta = n \cdot 10^{-5}$ и $\delta = n \cdot 10^{-3}$, а также ряда значений величины n.

Таблица 1: Значение параметра Λ для величины $\delta = n \cdot 10^{-5}$.

H_0/n	2	10	26	32
0,5	$5,77 \cdot 10^{-10}$	$1,14 \cdot 10^{-8}$	$9,75 \cdot 10^{-8}$	$1,47 \cdot 10^{-7}$
1	$2,89 \cdot 10^{-10}$	$7,22 \cdot 10^{-9}$	$4,88 \cdot 10^{-8}$	$7,39 \cdot 10^{-8}$

Таблица 2: Значение параметра Λ для величины $\delta = n \cdot 10^{-3}$.

H_0/n	2	10	26	32
0,5	$1,2 \cdot 10^{-5}$	$2,8 \cdot 10^{-4}$	$2,0 \cdot 10^{-3}$	$3,0 \cdot 10^{-3}$
1	$5,8 \cdot 10^{-6}$	$1,44 \cdot 10^{-4}$	$9,7 \cdot 10^{-4}$	$1,48 \cdot 10^{-3}$

Заметим далее, что в соответствии с подходом, предложенным К. Шенноном ([3]), величину $H(M^N/E^N)$ можно рассматривать как характеристику числа возможных исходных сообщений длины N, соответствующих известному выходному сообщению. В частности, можно полагать оценку числа таких сообщений равной $2^{H(M^N/E^N)}$.

Тогда, если ввести некоторый параметр β, характеризующий допустимую многозначность восстановления исходного сообщения при известном шифрованном сообщении, например, учитывающий различное написание окончания слов, то число допустимых вариантов входного сообщения оценивается величиной $2^{\beta \cdot N}$. В этом случае условие невозможности восстановления передаваемого сообщения принимает вид:

$$2^{H(M^N/E^N)} \geqslant 2^{\beta \cdot N}$$

или

$$\frac{1}{N} \cdot H(M^N/E^N) \geqslant \beta. \tag{6}$$

С учетом неравенства (4), условие невозможности восстановления передаваемого сообщения принимает вид:

$$H_0 - \varepsilon \geq \beta. \tag{7}$$

Учитывая выражение для величины ε, условие (7) принимает вид:

$$\frac{1}{2\ln 2}\left[\delta^2 + \frac{\delta^3}{3\,(1-\delta)^3}\right] \leqslant H_0 - \beta,$$

либо при $\delta \leqslant 0,46$:

$$\frac{1}{\ln 2} \cdot \delta^2 \leqslant H_0 - \beta,$$

откуда следует условие на величину неравновероятности знаков преобразованного сообщения δ, при котором невозможно восстановление исходного сообщения:

$$\delta \leqslant (\ln 2 \cdot (H_0 - \beta))^{1/2}. \tag{8}$$

Для иллюстрации полученных результатов приведем численные расчеты параметра $\lambda = \frac{1}{n}\delta$ (таблица 3) для значения величины $\beta = 0,1$ и ряда значений величин H_0 и n.

Таблица 3: Значение величины $\lambda = \frac{1}{n}\delta$.

H_0/n	2	10	26	32
0,5	0,26	0,052	0,02	0,016
0,7	0,32	0,064	0,024	0,02
0,85	0,35	0,072	0,028	0,025

3. О допустимой неравновероятности знаков ключа при применении преобразования гаммирования

Применим полученные выше результаты к исследованию преобразования гаммирования ([1,2]). В этом случае знаки выходного сообщения $b^N = (b_1, b_2, \ldots, b_N)$, $b_i \in A_n$, образуются из знаков входного сообщения $a^N = (a_1, a_2, \ldots, a_N)$ в соответствии с уравнением:

$$b_i = a_i + \gamma_i \pmod n, \quad a_i, \gamma_i \in A_n, \quad i = 1, 2, \ldots, N, \tag{9}$$

где $\{\gamma_i | \gamma_i \in A_n, i = 1, 2, \ldots, N\}$ - последовательность знаков гаммы, не зависящих от входного сообщения.

Обозначим через Γ случайную величину, принимающую значение из множества A_n и имеющую некоторое вероятностное распределение $\theta(\cdot)$, заданное на этом множестве.

Теорема 2. *Теорема 2. Пусть распределение $\theta(\cdot)$, заданное на множестве ключей K таково, что для любых $i = 1, \ldots, N$ выполнено равенство:*

$$P\{\Gamma_i = \gamma_i / \Gamma_j = \gamma_j, j = 1, \ldots, i-1\} = \frac{1}{n} + \frac{1}{n} \cdot \Delta_{\gamma_i}(\gamma_j, j = 1, \ldots, i-1), \tag{10}$$

где $|\Delta_{\gamma_i}(\gamma_j,\ j = 1, \ldots, i-1)| \leqslant \delta_i,\ 0 \leqslant \delta_i < 1$.

Тогда

$$H\left(M^N/E^N\right) = H\left(M^N\right) - \sum_{i=1}^{N} \varepsilon_i, \tag{11}$$

где $0 \leqslant \varepsilon_i \leqslant \frac{1}{2\ln 2}\left[\delta_i^2 + \frac{\delta_i^3}{3(1-\delta_i)^3}\right]$.

Аналогично результатам предыдущего раздела можно найти допустимое отклонение знаков ключевой последовательности, выбираемой из алфавита n, из уравнения:

$$\delta = (\ln 2 \cdot (H_0 - \beta))^{1/2}. \tag{12}$$

Дальнейшие расчеты будем проводить для алфавита мощности $n = 32 = 2^5$, соответствующему русскому языку.

Полагая для примера $H_0 = 0,5$ и $\beta = 0,1$, из выражения (12) получаем граничное значение величины $\delta : \delta = 0,51$. При этом условие, налагаемое на вероятность появления знаков ключевой последовательности выбираемой из множества мощности $n = 2^5$, имеет вид:

$$p\{\Gamma_i = \gamma_i / \Gamma^{i-1} = \gamma^{i-1}\} = \frac{1}{2^5}[1 + \Delta_i],$$

где $|\Delta_i| \leqslant 0,51$. Отсюда следует, что в рассматриваемом случае относительные частоты знаков ключевой информации должны отклоняться от величины $2^{-5} = 0,031$ не более чем на величину $0,51 \cdot 2^{-5} = 1,6 \cdot 10^{-2}$.

Далее заметим, что, как правило, в криптографических алгоритмах ключевая информация представляется в виде двоичной последовательности. С целью получения аналогичных результатов для случая, когда знаки входного сообщения и знаки гаммы представляются в виде двоичных векторов, воспользуемся следующими соображениями.

Пусть знак ключа $\gamma_i \in A_n$ представляется в виде двоичного вектора

$$\gamma = \left(\gamma^{(1)}, \gamma^{(2)}, \ldots, \gamma^{(m)}\right), \gamma^{(j)} \in \{0; 1\}, n = 2^m.$$

Предполагая независимость случайных величин $\gamma^{(i)}$, получаем:

$$P\{\gamma = \rho\} = P\{\gamma^{(1)} = \rho^{(1)}, \gamma^{(2)} = \rho^{(2)}, ..., \gamma^{(m)} = \rho^{(m)}\} = \prod_{i=1}^{m} P\{\gamma^{(i)} = \rho^{(i)}\}, \tag{13}$$

где $\rho = \left(\rho^{(1)}, \rho^{(2)}, \ldots, \rho^{(m)}\right)$, $\rho^{(j)} \in \{0, 1\}$, $n = 2^m$.

При этом условие наличия необходимых статистических качеств ключевой последовательности, будет иметь вид:

$$P\{\gamma_i = \rho\} = \frac{1}{2^m}\left[1 + \Delta_i\right] = \prod_{i=1}^{m} \frac{1}{2} \cdot \left(1 + \Delta_i'\right), \tag{14}$$

где $|\Delta_i| \leq \delta_i, |\Delta_i'| \leq \delta_i', 0 \leq \delta_i < 1, 0 \leq \delta_i' < 1$.

Очевидно, что отклонение Δ_i соответствует алфавиту из $n = 2^m$ знаков, а отклонение Δ' — двоичному алфавиту.

Пусть, как и ранее,

$$\delta = \max_i \delta_i, \delta' = \max_i \delta_i'$$

С учетом соотношения (18) значение величины δ' может быть найдено из условия:

$$\frac{1}{2^m} \cdot (1 + \delta')^m = \frac{1}{n} \cdot (1 + \delta). \tag{15}$$

Принимая во внимание, тот факт, что величина δ_i' близка к 0, и разлагая в ряд функцию $(1 + \delta')^m$, получаем, оценку для величины δ':

$$\delta' = \frac{1}{m} \cdot \delta. \tag{16}$$

Учитывая, что при значении $m = 5$, величина δ равна

$$\delta = 0,51,$$

получаем значение для допустимого отклонения распределения знаков бинарных последовательностей ключевой информации от равновероятного распределения:

$$\delta' = 0,016.$$

Таким образом, при выработке ключевой двоичной последовательности, вероятность появления двоичного знака не должна отклоняться от значения $\frac{1}{2}$ на величину более $\delta' = 0,016$.

4. Заключение

В работе рассмотрены вопросы влияния возможного неравновероятного распределения знаков преобразованного сообщения при применении криптографического алгоритма и неравновероятного распределения знаков ключевой информации при применении схемы простого гаммирования, на криптографические качества алгоритмов с позиции теоретико-информационного подхода.

Очевидно, что полученные результаты неприменимы в отношении криптографических алгоритмов, обеспечивающих практическую стойкость, поскольку в случае наличия расстояния единственности всегда есть возможность определить ключ, применяя переборные методы и однозначно дешифровать входное сообщение.

В работе в рамках конкретной теоретико-вероятностной модели получены оценки взаимной информации ансамблей входных и выходных сообщений. На основании данных оценок получены выражения для определения границ возможного отклонения от равновероятного распределения вероятности знаков выходного (шифрованного) сообщения и знаков ключевой информации, при которых обеспечивается невозможность восстановления входного сообщения при известном выходном (шифрованном) сообщении. Для наиболее интересных с практической точки зрения случаев проведены численные расчеты исследуемых параметров.

СПИСОК ЦИТИРОВАННОЙ ЛИТЕРАТУРЫ

1. А. Б. Лось. Криптографические методы защиты информации. Учебник для вузов. Серия: для изучающих компьютерную безопасность / А. Б. Лось, А. Ю. Нестеренко, М. И. Рожков 2-е изд., испр. М.: Изд. Юрайт.2021. 423 с.

2. А. П. Алферов Основы криптографии/А.П. Алферов, А.Ю. Зубов, А.С. Кузьмин, А.В. Черемушкин. — М.: Гелиос АРВ, 2001. 479 с.

3. К. Шеннон Работы по теории информации и кибернетике. — М.: ИЛ. 1963.829 с.

4. Колесник В. Д. Курс теории информации / В.Д. Колесник, Г.Ш. Полтырев. — М.: Наука, 1982.416 с.

5. Энтропийный анализ текстов / Электронный ресурс (режим доступа — свободный): https://studme.org/245864/prochie/entropiynyy_analiz_tekstov (дата обращения 05.12.22).

6. Энтропия сложных сообщений / Электронный ресурс (режим доступа — свободный): https://studopedia.su/9_31691_entropiya-slozhnih-//soobshcheniy-izbitochnost-istochnika.html (дата обращения 05.12.22).

7. М. Я. Выгодский. Справочник по высшей математике/ М.Я. Выгодский.-М: Наука, 1964.-870 с.