Учёные записки
Комсомольского-на-Амуре государственного технического университета
№ IV-1 (12)
2012

Котляров В. П., Воробьев А. А.
V. P. Kotlyarov, A. A. Vorobiev

ИССЛЕДОВАНИЕ КРИПТОСТОЙКОСТИ МОДИФИКАЦИИ ШИФРА ГАММИРОВАНИЯ ПО ОПЕРАЦИИ XOR ПРИ ИСПОЛЬЗОВАНИИ КОНТИНУАЛЬНОГО МНОЖЕСТВА

ANALYSIS OF MODIFIED XOR GAMMA CIPHER USING A CONTINUUM SET

Котляров Валерий Петрович – кандидат технических наук, профессор кафедры информационных систем, декан факультета компьютерных технологий Комсомольского-на-Амуре государственного технического университета (Россия, Комсомольск-на-Амуре); 681013, г. Комсомольск-на-Амуре, ул. Ленина, д. 27; +7(4217)59-46-59. E-mail: fct@knastu.ru
Mr. Valeriy P.Kotlyarov – Ph.D., Professor of the Department of Information Systems, Dean of the Faculty of Computer Technologies, Komsomolsk-on-Amur State Technical University (Russia, Komsomolsk-on-Amur); 27, Lenina prospect, 681013 Komsomolsk-on-Amur, Khabarovsky region, Russian Federation; +7–(4217)–59–46–59. E–mail: fct@knastu.ru

Воробьев Антон Александрович – аспирант очной формы обучения (спец. 051318 «Математическое моделирование, численные методы и комплексы программ») Комсомольского-на-Амуре государственного технического университета (Россия, Комсомольск-на-Амуре); 681013, г. Комсомольск-на-Амуре, ул. Ленина, д. 27; 8-909-845-58-72. E-mail: zeromem@mail.ru
Mr. Anton A.Vorobiev – PhD Candidate, discipline No. 051318 «Mathematical modelling, numerical methods and software packages», Komsomolsk-on-Amur State Technical University (Russia, Komsomolsk-on-Amur); 27, Lenina prospect, 681013 Komsomolsk-on-Amur, Khabarovsky region, Russian Federation; 8-909-845-58-72. E–mail: zeromem@mail.ru

Аннотация. В **с**татье проводится оценка криптостойкости шифра гаммирования по модулю два к частотному криптоанализу с использованием решения о повышении криптостойкости шифра континуальным множеством. Показано изменение статистических характеристик графической и текстовой информации, а также повышение их энтропий при применении идеи со сферой Римана.

Summary. The paper deals with the assessment of the endurance of a gamma cipher algorithm at module two to frequency cryptanalysis using a continuum ser as a solution for cipher endurance improvement. The change of the statistical characteristics of graphical and textual information, as well as the increase of their entropy due to application of the idea of the Riemann sphere, is shown.

Ключевые слова: криптография, сфера Римана, варьирование запятой, защита информации, гаммирование, криптоанализ.

Key words: cryptography, Riemann sphere, comma varying, data security, gamma algorithm, cryptanalysis.

УДК 003.26, 004.056.5

В работах [4; 6] получены уравнения биективного отображения комплексной величины на сферу Римана при помощи следующих формул:

$$x = \frac{4x_i R^2}{4R^2 + y_i^2 + x_i^2}, y = \frac{4y_i R^2}{4R^2 + y_i^2 + x_i^2}, z = 2R - \frac{8R^3}{4R^2 + y_i^2 + x_i^2}. \quad (1)$$

Материал поступил 26.12.2011

Выдвинута гипотеза о варьировании запятой, которая в совокупности с числами с плавающей точкой из континуального множества повышает вычислительную сложность атаки методом «грубая сила» до k раз, где k – возможные различные положения запятой.

Любопытно провести исследование криптостойкости шифра гаммирования по операции исключающего ИЛИ (строгая дизъюнкция) при использовании отображения (1) на сферу Римана и варьирования запятой к частотному криптоанализу.

Дешифрованию различных криптограмм помогает частотный анализ появления отдельных символов и их сочетаний [1]. Как показал К. Шеннон [7], вследствие несовершенства языка, наличия огромного количества синтаксических и грамматических правил, отдельные частоты появления символов варьируются в некоторых определенных пределах для каждого языка. Так, в частности, буква «о» считается наиболее употребляемой в русском языке после знаков препинания (см. табл. 1 [3]), в то время как «е» считается самой употребляемой в английском. Очевидно, что шифры подстановки и замены совершенно не влияют на частотное распределение тех или иных символов используемого алфавита. Этим фактом пользуются для определения наиболее вероятного лингвистического языка открытого текста по шифротексту путем подсчета частот вхождения каждого символа в текст и сравнением с эталонными таблицами. Каждая эталонная таблица [3] представляет собою частоты появления того или иного элемента алфавита для конкретного языка.

Таблица 1

Эталонная таблица частот букв в русском тексте

Буква	Частота	Буква	Частота	Буква	Частота	Буква	Частота
"—"	0.145	р	0.041	я	0.018	х	0.009
о	0.095	в	0.038	з	0.016	ж	0.007
е, ё	0.072	л	0.035	ы	0.016	ш	0.006
а	0.062	к	0.028	б	0.014	ю	0.006
и	0.062	м	0.026	ь,ъ	0.014	ц	0.004
н	0.053	д	0.025	г	0.013	щ	0.003
т	0.053	п	0.023	ч	0.012	э	0.003
с	0.045	у	0.021	й	0.010	ф	0.002

Указанные выше умозаключения могут быть распространены на произвольный алфавит, составляющий суть некоторого сообщения, так как, исходя из [7], неравномерность распределения вероятностей появления того или иного символа алфавита повышает информацию о криптосистеме, что, в свою очередь, вызывает положительную корреляцию с возможностью её компрометации.

Это утверждение вытекает непосредственно из предложенной К. Шенноном в 1948 г. формулы количества информации:

$$I = -\sum_{i=1}^{K} p_i \log_2(p_i), \tag{2}$$

где K – общее количество возможных состояний системы, p_i – вероятность проявления i-го состояния системы. При $p_i = 0$ полагают, что $p_i \log_2(p_i) = 0$. Количество информации выражается в двоичных битах.

Так в свою очередь каждая буква английского языка несет в среднем 1.3 двоичных бит информации [8].

Формула (2) является обобщением формулы, полученной ранее в 1928 г. американским инженером Р. Хартли. Он впервые ввел количественную характеристику информации при условии равной вероятности состояний системы:

$$I = \log_2 K, \quad (3)$$

где K – общее количество возможных состояний системы.

Но, исходя из теории информации [3], количество информации, приобретаемое при полном выяснении состояния некоторой физической системы X, равно энтропии этой системы, то есть степени неопределенности системы:

$$I = H(X). \quad (4)$$

Так как злоумышленнику тем сложнее подвергать компрометации криптосистему, чем она неопределеннее, то повышение энтропии криптосистемы может являться одним из показателей повышения её криптостойкости.

Покажем изменения энтропии системы при использовании криптографического алгоритма метода гаммирования с ключом K в 4 байта после применения идеи со сферой Римана, воспользовавшись элементами частотного анализа и формулой (2) для графического изображения и текстовой информации.

В качестве графического изображения Γ для анализа выбран рисунок в RAW формате (см. рис. 1). К нему была применена идея со сферой Римана с помощью инструмента ECV [5] (см. рис. 2), а результат R сохранен в виде дампа памяти без соответствующих заголовков (RAW формат).

Полученные массивы данных были подвергнуты частотному анализу.

Рис. 1. Результат гаммирования с ключом в 4 байта

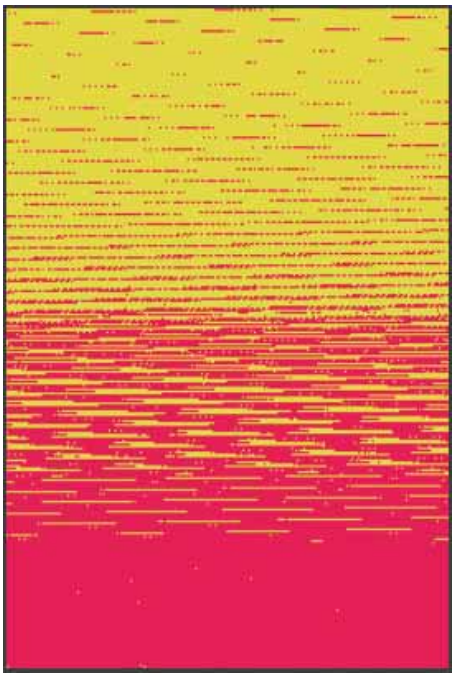

Рис. 2. Результат применения идеи со сферой Римана

Опишем алгоритм идеи со сферой Римана.

1) Пока входной поток данных X не пуст, выберем i-й элемент данных (байт) $a_i \in X$. Иначе – выход.

2) Выберем элемент ключа $k_j \in K \mid j = i(\mathrm{mod}\,4)$.

3) Составим комплексное число $q_i = (a_i, k_j)$.

4) Используя соотношения (1), вычислим тройку (x_i, y_i, z_i) по комплексной величине q_i.

5) Произведем последовательный вывод в поток P величин x_i, y_i, z_i;

6) Перейдем к пункту 1.

Подставляя вместо входного и выходного потоков соответствующие массивы данных, получаем требуемый результат. Заметим, что данный алгоритм можно легко распространить на общий случай длины ключа $|K| > 4$.

График распределения частот для исходного изображения имеет стохастический характер (см. рис. 3), что позволяет выделить класс наиболее вероятных проявлений того или иного элемента данных. Соответственно, после применения идеи со сферой Римана распределение частот для конечного изображения приняло практически сглаженный вид (см. рис. 4).

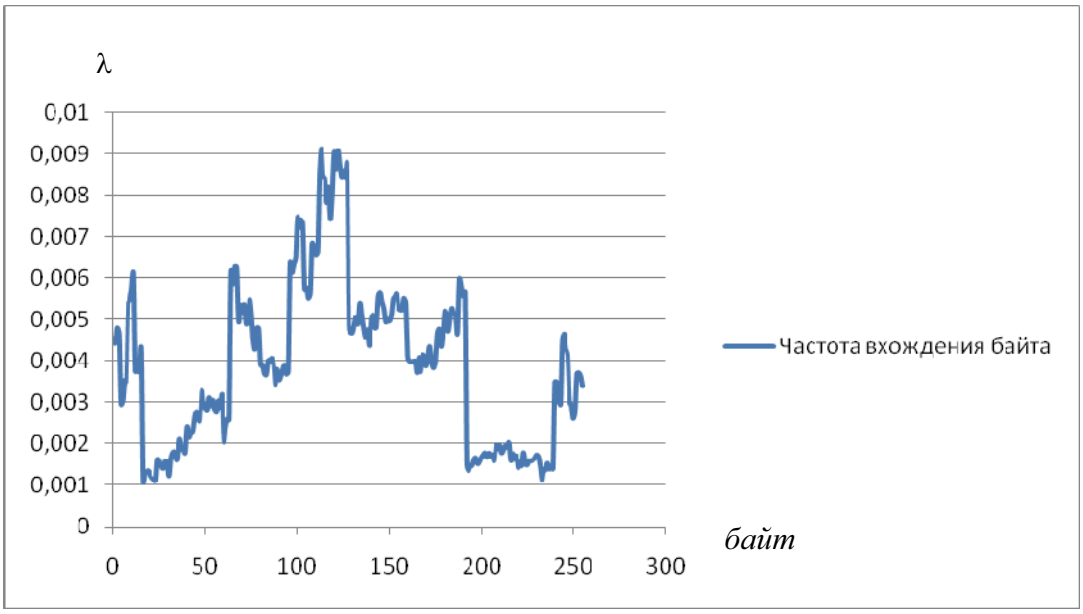

Рис. 3. Частота (λ) появления байта в потоке данных Γ при методе гаммирования по модулю два

Коэффициент корреляции k между частотным распределением исходных и конечных данных:

$$k = 0.104255. \tag{5}$$

Столь малое значение данного коэффициента позволяет говорить о достаточно сильном отличии начального набора данных от конечного набора, то есть об их малой корреляции. Наличие знака плюс объясняется прямо пропорциональной зависимостью между указанными массивами данных.

Энтропия $H(\Gamma)$ для исходных данных:

$$H(\Gamma) = 7.813, \tag{6}$$

Энтропия $H(R)$ для конечных данных:

$$H(R) = 6.848. \tag{7}$$

Соответственно имеем $H(\Gamma) > H(R)$, т.е. энтропия конечных данных (7) меньше, чем энтропия (6) исходных. Полученный результат является ожидаемым вследствие того, что любая естественная физическая система (фотография представима как мгновенное состояние некоторой естественной системы) имеет более высокую энтропию, чем энтропия её математической модели. Но из частотного распределения конечных данных (см. рис. 4)

очевидно, что сглаженность осложняет вероятностный анализ появления того или иного элемента данных.

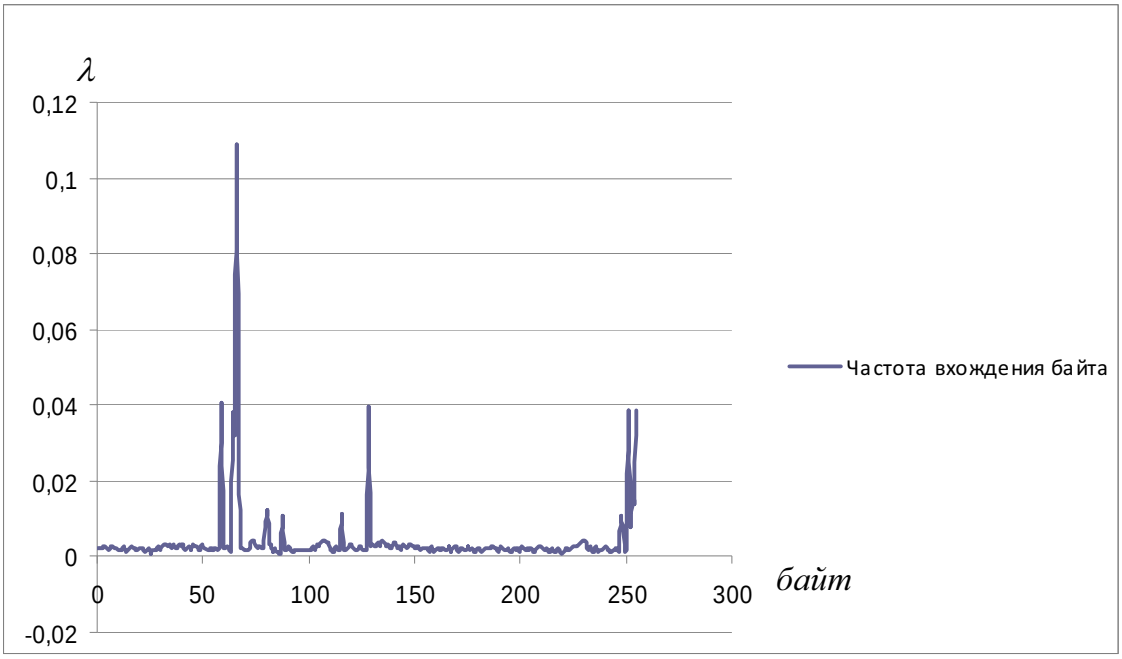

Рис. 4. Частота появления символа в потоке данных R при методе гаммирования по модулю два после применения отображения на сферу Римана

Опишем методику частотного анализа, используемую при получении значений (6) и (7), и применим ее для анализа некоторого английского текста до и после применения идеи со сферой Римана.

Для систем Γ и R были сформированы таблицы количества вхождения каждого байта методом, аналогичным при сортировке подсчетом.

Производя деление количества вхождения байта на общее количество байт для систем Γ и R, получим частоты появления каждого элемента данных в системе (см. рис. 3 и 4 соответственно).

Вычисление коэффициента корреляции (5) между массивами данных производилось при помощи встроенной в программный продукт Microsoft Excel 2007 функции «КОРЕЛ»[2].

Соответственно, вычисление энтропий (6) и (7) для каждой из систем выполнялось с использованием формулы (2) по полученным частотам при помощи встроенной математической функции «LOG».

В качестве текстовой информации T для анализа выберем английский текст:

«We send out a free English lesson by email most weeks. There is absolutely no charge and you can stop receiving them at any time, just by clicking on a link in the email. Just fill in this form with your name and email address. Pearson will immediately send you an email with a link to confirm you want to get our lessons. Click on that link and start receiving our free lessons».

Вычислим соответствующие показатели, аналогично формулам (5) – (7).

В результате имеем частотные распределения для исходных (см. рис. 5) и конечных (см. рис. 6) R' данных.

Соответствующие энтропии $H(T)$, $H(R')$ и коэффициент корреляции равны:

$$H(T) = 4.192, \tag{8}$$

$$H(R') = 6.713, \tag{9}$$

$$k = -0.00617. \tag{10}$$

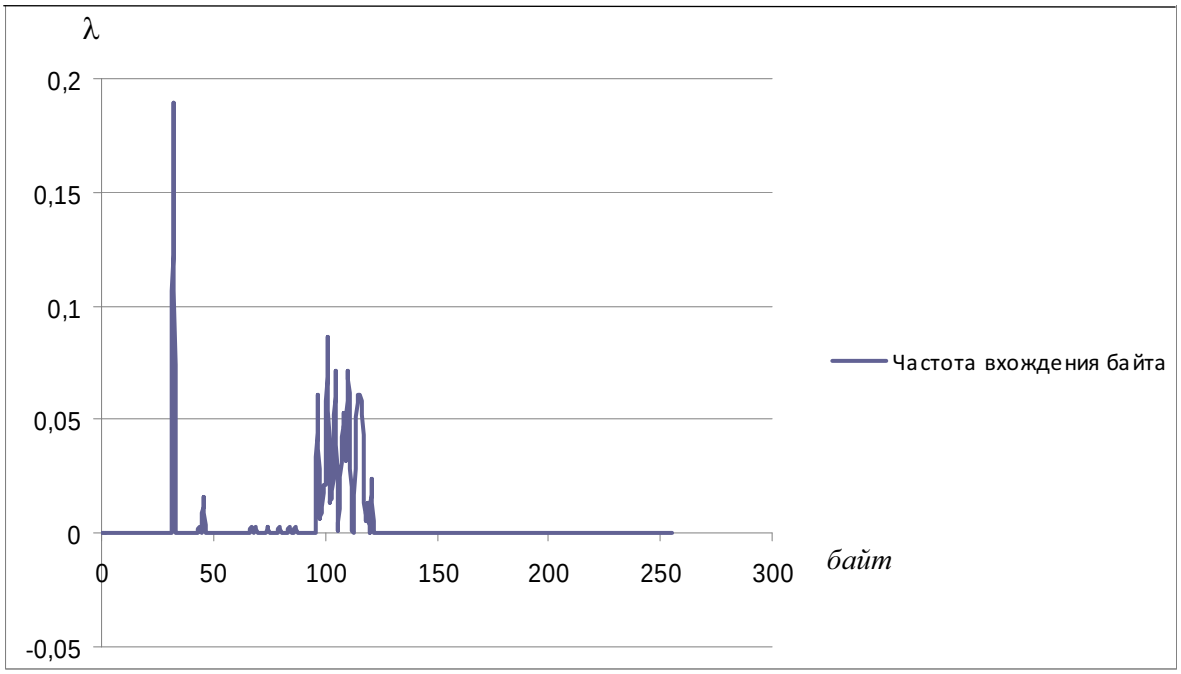

Рис. 5. Частота появления символа в потоке данных T при методе гаммирования по модулю два

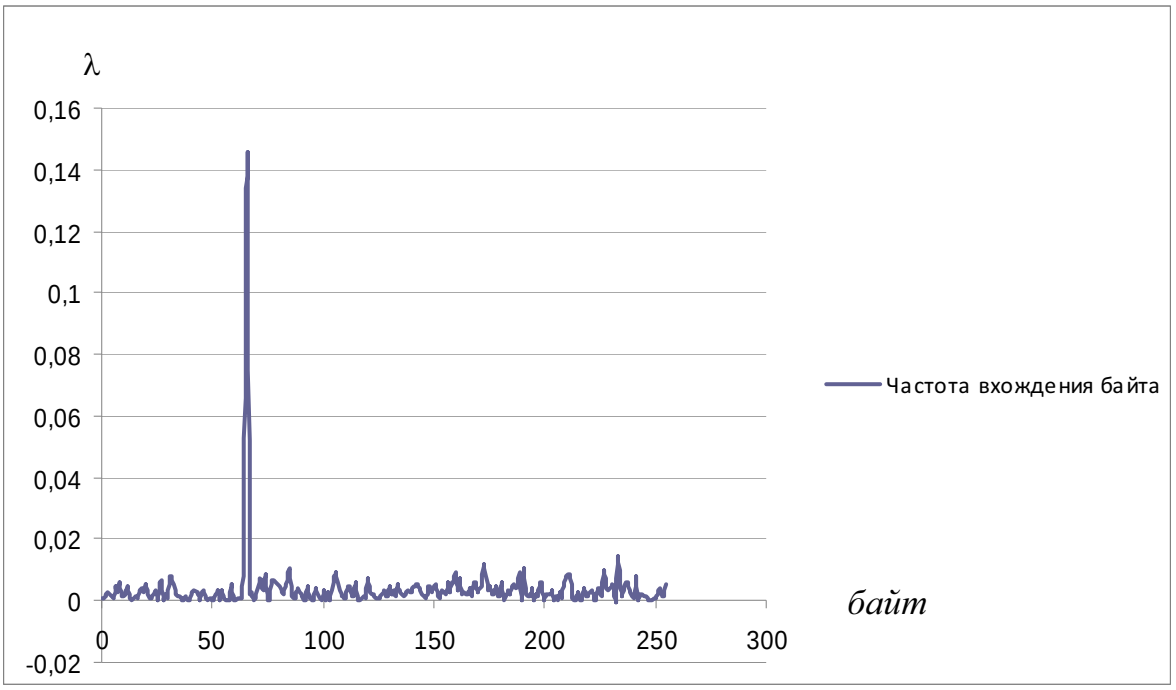

Рис. 6. Частота появления символа в потоке данных R' при методе гаммирования по модулю два после применения отображения на сферу Римана

Имеем $H(T) < H(R')$, т.е. энтропия конечных данных (9) больше, чем энтропия (8) исходных. Так как злоумышленник всегда стремится к полной компрометации криптосистемы, из соотношения (2) получаем, что злоумышленнику придется преодолеть большую энтропию, большую неопределенность в поведении системы после применения

идеи со сферой Римана, что и требовалось показать. Данный факт имеет огромное значение для текстовой информации, так как она содержит элементы неестественного, то есть неприродного характера, разработанные человеком, для которых методы частотного анализа являются более применимыми, нежели к потокам графических данных.

До этого момента, при использовании идеи со сферой Римана применялись числа с плавающей точкой с шестью знаками после запятой. Покажем, что повышение количества знаков также повышает энтропию исследуемых систем. Данный факт опытным путем подтверждает (косвенно) справедливость идеи о варьировании запятой предыдущей работы [6].

Повысив количество знаков до 15, а также подсчитав новые частотные характеристики, имеем следующие результаты.

Для потока R графической информации, исходя из нового частотного распределения (см. рис. 7), имеем энтропию

$$H_{15}(R) = 7.355\,. \tag{11}$$

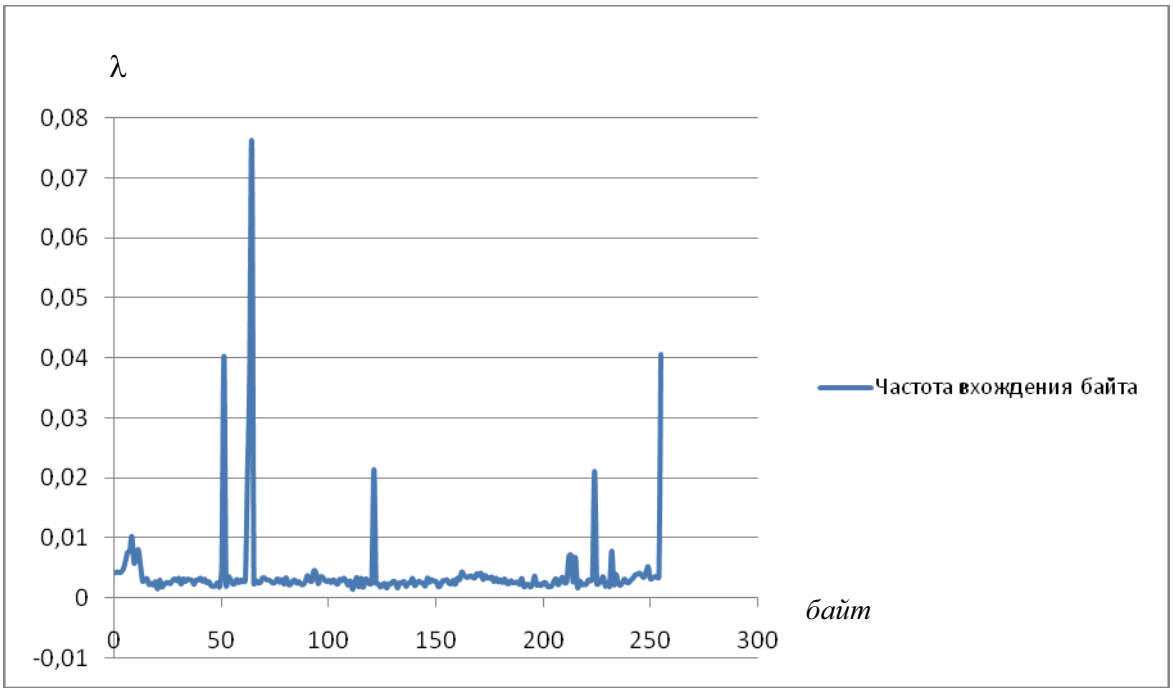

Рис. 7. Частота появления символа в потоке данных R при методе гаммирования по модулю два после применения отображения на сферу Римана (15 знаков)

Для потока R' текстовой информации, исходя из нового частотного распределения (см. рис. 8), имеем энтропию

$$H_{15}(R') = 7.281\,. \tag{12}$$

Коэффициент корреляции при использовании 15 знаков после запятой для графических данных и текстовой информации соответственно равны:

$$k = 0.020972\,, \tag{13}$$

$$k = -0.04054\,. \tag{14}$$

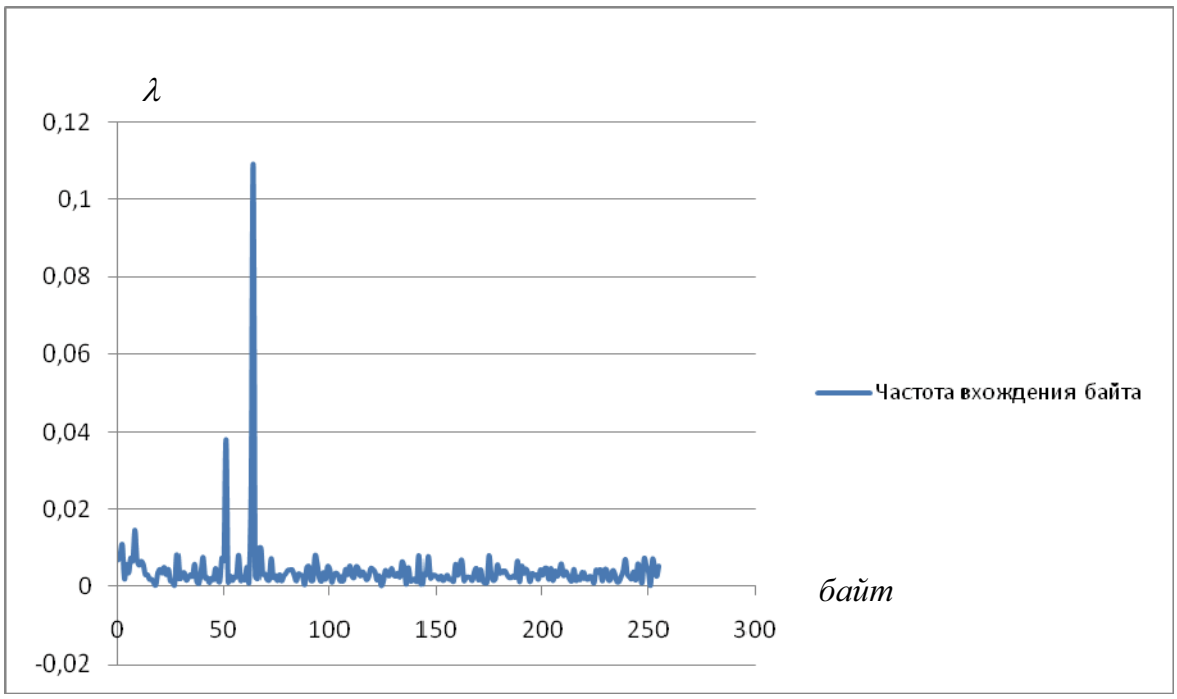

Рис. 8. Частота появления символа в потоке данных R' при методе гаммирования по модулю два после применения отображения на сферу Римана (15 знаков)

Используем полученные данные для построения кривой зависимости энтропии от числа используемых знаков (см. рис. 9), где в качестве значений при нуле знаков используются энтропии исходных данных.

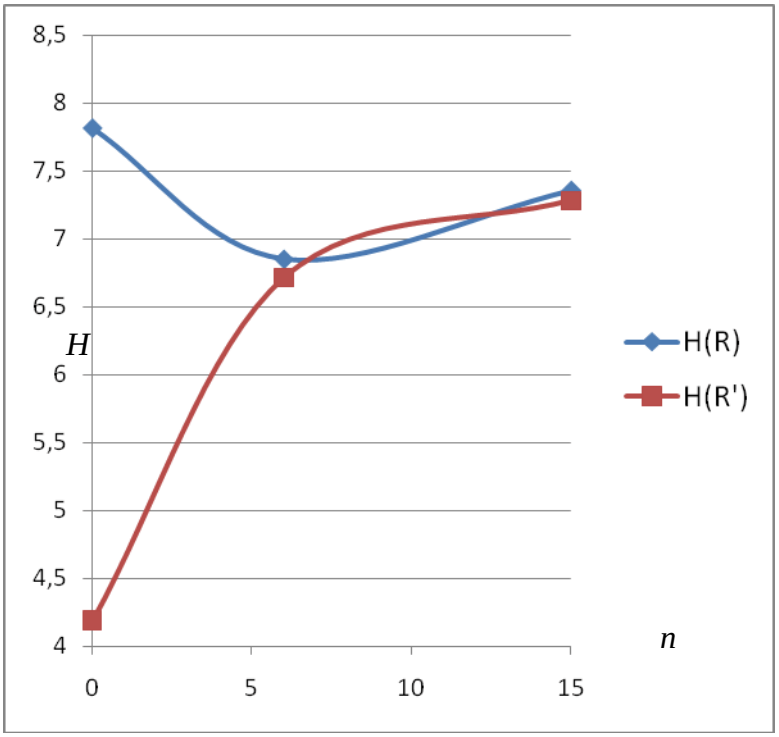

Рис. 9. Зависимость энтропии от числа используемых знаков (n)

Таким образом, можно сделать заключение о наличии прямо пропорциональной зависимости между количеством используемых знаков и энтропией, что опытным путем подтверждает справедливость идеи о варьировании запятой, – увеличение количества знаков после запятой приводит к повышению вычислительной сложности атаки методом «грубая сила».

ЛИТЕРАТУРА

1. Авдошин, С. М. Криптоанализ: вчера, сегодня, завтра / С. М. Авдошин, А. А. Савельева // Открытые системы. – 2009. – № 3. – С. 22-26.
2. Васильев, А. Н. Научные вычисления в Microsoft Excel / А. Н. Васильев. – М.: Вильямс, 2004. – 512 с.
3. Вентцель, Е. С. Теория вероятностей: учеб. для вузов / Е. С. Вентцель. – 8-е изд., стер. – М.: Высш. шк., 2002. – 575 с.
4. Воробьев, А. А. О проблеме взлома перебором и потенциальных решениях с помощью сферы Римана и варьирования запятой: доклад / А. А. Воробьев // Научная сессия ТУСУР–2010. – Томск: В–Спектр. – Ч. 3. – С. 230-235.
5. Воробьев, А. А. Визуализация процессов работы алгоритмов шифрования с дополнением преобразования сферой Римана / А. А. Воробьев // Открытый Дальневосточный конкурс программных средств студентов, аспирантов и специалистов «Программист-2010»: сб. докладов. – Владивосток: Дальневост. гос. ун-т, 2010. – С. 7-9.
6. Воробьев, А. А. О решениях повышения криптостойкости шифров с помощью континуального множества / А. А. Воробьев, В. П. Котляров // Ученые записки Комсомольского-на-Амуре государственного технического университета. – 2010. – № II–1(2). – С. 58-64
7. Shannon, C. E. A Mathematical Theory of Communication / C. E. Shannon // Bell System Technical Journal. – 1948. – Vol. 27, 28. – P. 379-423, 623-656.
8. Shannon, C. E. Prediction and Entropy of Printed English / C. E.Shannon // Bell Systems Technical Journal. – 1951. – Vol. 30. – P. 50-64.