конференция
РусКрипто'2013

http://www.ruscrypto.ru/conference/

Дешифрование шифра перестановки

Бабаш А.В., Романова Е.В.,
Александров А.А., Ларионов И.П.,
Тарелина А.В., Тащилин Е.С.

- Пусть I – некоторый алфавит, например, I=(а,б,в,г,…,я,_), I* - множество всех последовательностей конечной длины алфавита I, а U – подмножество всех содержательных (открытых) текстов из I*. Ключами шифра перестановки являются пары (d, P) – натуральное число d и подстановка степени d.

$$P = \begin{pmatrix} 1 & 2 & \cdots & d \\ P(1) & P(2) & \cdots & P(d) \end{pmatrix}$$

Если

$$i_1 i_2 \ldots i_n \in I^*$$

- некоторый текст подлежащий шифрованию, то он предварительно дополняется буквами до текста длины N=kd кратного d и записывается в таблицу с d столбцами:

$$\begin{array}{cccc} i_1 & i_2 & \ldots & i_d \\ i_{d+1} & i_{d+2} & \ldots & i_{2d} \\ . & . & . & . \\ i_{(k-1)d+1} & \ldots & . & i_N \end{array}$$

Затем столбцы переставляются по подстановке P: первый столбец встает на место P(1), второй – на место P(2), … столбец с номером d встает на место P(d). Обозначим через

$$\begin{array}{cccc} i^*_1 & i^*_2 & \dots & i^*_d \\ i^*_{d+!} & i^*_{d+2} & \dots & i^*_{2d} \\ . & . & . & . \\ i^*_{(k-1)d+1} & \dots & . & .i^*_N \end{array}$$

Результирующую таблицу.

Буквы результирующей таблицы записываются в одну строку .

- В работе
Проскурин Г.В. Принципы и методы защиты информации. Учебное пособие. Московский государственный институт электроники и математики. 1977.
был приведен подход к дешифрованию шифра перестановки. Он заключался в опробовании столбцов таблицы 2. Для каждой упорядоченной пары столбцов (*j*j’) просматривалось множество всех биграммы на их горизонталях. При нахождении в этом множестве «запретной биграммы», то есть биграммы, появляющейся в открытых текстах с вероятностью 0, пара столбцов получала пометку «не соседние». Далее строился граф из всех пар упорядоченных столбцов не имеющих данную пометку. В соответствии с этим графом проводилось упорядоченное опробование столбцов с учетом «читаемости текста». Недостатком этого подхода являлась большая неопределенность в выборе путей этого графа на основе «читаемости» текста.

Постановка задачи дешифрования шифра.

Известно, что на данном шифре перестановки шифруются тексты в алфавите I из множества U – всех содержательных (открытых) текстов длиной не превосходящих величины T. Известно начало

$$i^*_1 i^*_2 \ldots i^*_L$$

длины L шифрованного текста

$$i^*_1 i^*_2 \ldots i^*_T$$

длины T, причем L не меньше d. Найти начало

$$i_1 i_2 \ldots i_L$$

открытого содержательного текста.

Тотальный метод

- Число возможных ключей равно величине 2!+3!+…+L! , а среднее число опробований без возвращения этих ключей до получения искомого ключа равно

$$\frac{1}{2}\left(2!+3!+\ldots+L!\right)$$

Предыстория задачи

- Лет 10 назад Валентин Петрович Зязин спросил меня о возможности определения порядка d подстановки шифра перестановки по шифрованному тексту. Ответа сразу я не нашел. Через несколько лет по по электронной почте один из студентов Украины попросил посоветовать ему тему курсовой работы по историческим шифрам. Я сформулировал ему задачу Зязина в надежде, что он поделится идеями ее решения. Ответа пока не получил. Задача Зязина остается открытой.

Метод 1 решения задачи при известной части d ключа

- Записываем шифрованный текст в таблицу

$$
\begin{array}{cccc}
i*_1 & i*_2 & \dots & i*_d \\
i*_{d+!} & i*_{d+2} & \dots & i*_{2d} \\
. & . & . & . \\
i*_{(k-1)d+1} & \dots & .... & i_{kd}
\end{array}
$$

Для определения множества возможных вариантов следования друг за другом столбцов в истиной таблице открытого текста используются вероятности встречаемости биграм в открытом тексте

$$P(xy)$$

Для неправильных вариантов следования используются вероятности случайного, равновероятного и независимого выбора букв открытого текста.

$$p(x,y) = p(x)p(y)$$

- В случае правильного выбора следования в столбцах

$$\begin{pmatrix} i_{\nu} \\ i_{d+\nu} \\ \cdot \\ i_{(k-1)d+\nu} \end{pmatrix} \begin{pmatrix} i_{\nu+1} \\ i_{d+\nu+1} \\ \cdot \\ i_{kd+\nu+1} \end{pmatrix}$$

пары последующих букв алфавита - биграммы (по предположению) являются выборкой размера k из вероятностного распределения биграмм открытого текста – гипотеза H(0)

Множество соответствующих пар букв столбцов, не являющихся соседними в таблице 1, будем считать двумерной выборкой из вероятностного распределения букв открытых содержательных текстов – гипотеза H(1)

Суть метода 1 состоит

- Для каждого столбца *j из множества столбцов {1,2,…,d} с помощью статистического критерия определяется возможный последующий столбец *j’ .
- Обозначим множество полученных (прошедших критерий К) столбцов через П(*j) Будем говорить, что столбцы П(*j) согласованные столбцы со столбцом *j . Заметим, что для получения согласованных столбцов П(*j) мы провели d-1 опробований столбцов, а для получения всех множеств П(*j), потребовалось d(d-1) операций опробования столбцов.

- . Используя найденное бинарное отношение согласованности столбцов представим его в виде графа с d вершинами. Решение задачи находится теперь перебором путей длины d и составления для каждого из них таблицы столбцов и чтения открытого текста (возможно частично искаженного). Дополнительная информация, позволяющая частично уменьшить число промежуточных решений задачи, состоит в учете специальной связи столбца d и столбца 1 из таблицы (1)

Известно начало длины L шифрованного текста длины T и L не меньше d. Задача состоит в нахождении начала открытого содержательного текста

- Решение. Перебор d из множества {2,3,…L} и применения для каждого d решения изложенной ранее задачи.
- Трудоемкость в числе опробуемых пар столбцов $2\sum_{2}^{L} d(d-1)$

Надежность

$$\prod_{i=2}^{d-1}(1-\alpha)(1-\beta)^{d-i}$$

Результаты дешифрования шифра перестановки при известной части d ключа с использованием весового критерия.

- Положим L=kd+r, D=d. При малом значении k рекомендуем использовать в качестве критерия К весовой комбинированный критерий: для опробуемых столбца *j и на предмет следования за ним столбца *j’

$$\begin{matrix} i*_{j} & i*_{j'} \\ i*_{d+j} & i*_{d+j'} \\ . & . \\ i*_{kd+j} & i*_{kd+j'} \end{matrix}$$

вычисляется сумма (1)

$$Z(*j,*j') = P(i*_{j}\ i*_{j'}) + P(i*_{d+j}\ i*_{d+j'}) + \ldots + P(i*_{kd+j}\ i*_{kd+j'})$$

В качестве согласованного столбца *j’ со столбцом *j берется столбец с максимальным значением суммы.
С использованием этого критерия Для реализации метода было разработано два приложения на языке C++: одно для шифрования открытых текстов, другое для дешифрования шифрованных текстов. Тексты использовали алфавит из 34 символов - строчных букв русского алфавита и пробелов. Тексты не содержали цифр и знаков препинания. Шифровались тексты длин: 250, 1000 и 3000 знаков. При каждой из этих длин известная часть ключа D=d варьировалась. При каждой длине d ключа, как правило, выбиралось случайно и равновероятно двадцать пять ключей (подстановок степени d) из общего возможного числа ключей заданной длины. Использовались вероятности двухграмм открытого текста представленные в /3/ таблицей вероятностей биграмм русского текста. В результате эксперимента были выявлены ключи, при которых открытые тексты восстанавливались: однозначно правильно, частично правильно, не полностью, но частично читаемы.

Для описания результатов была введена следующая качественная шкала разборчивости частично дешифрованных сообщений:

- Отличная: большая часть слов и слогов (от 76%-100%) понятны, дешифрованный текст дает достаточно полное представление о смысле и содержании сообщения, незначительные детали сообщения непонятны.
- Хорошая: значительная часть слов и слогов (66-75%) понятны, дешифрованный текст дает представление о содержании и смысле сообщения, отдельные фрагменты непонятны.
- Средняя: около половины слов и слогов (46%-65%) понятны, дешифрованный текст дает представление об общем характере текста и основной мысли или теме сообщения.
- Слабая: около трети всех слов и слогов (26-45%) понятны, отдельные дешифрованные фрагменты дают представление о отдельных темах сообщения, содержание и характер всего сообщения непонятен.
- Плохая: только некоторые слова и слоги понятны (0-25%), сам дешифрованный текст не связан, нелогичен и не осмыслен, понятны части слов и фраз, остальное разобрать невозможно.
- Отдельно отмечались случаи правильного определения 1 и d-того столбца таблицы 2.

Таблица 1 «Оценка эффективности дешифрования при длине текста 250 символов, включая пробелы»

Длина ключа	Эффективность дешифрования сообщения	Разборчивость частично дешифрованного сообщения	Первый столбец перестановки найден	Последний столбец перестановки найден	Одновременно первый и последний столбцы перестановки найдены	Число испытанных ключей
2	100%	отличная	100%	100%	100%	2
3	100%	отличная	100%	100%	100%	6
4	100%	отличная	100%	100%	100%	24
5	96%	отличная	96%	100%	96%	25
6	88%	отличная	92%	96%	92%	25
7	76%	хорошая	88%	80%	80%	25
8	60%	хорошая	76%	72%	72%	25
9	52%	хорошая	68%	60%	60%	25
10	40%	средняя	56%	52%	52%	25
11	28%	средняя	48%	40%	40%	25
12	12%	средняя	32%	28%	24%	25
13	4%	слабая	24%	20%	20%	25
14	0%	слабая	16%	12%	12%	25
15	0%	слабая	8%	4%	4%	25
16	0%	плохая	4%	0%	0%	25

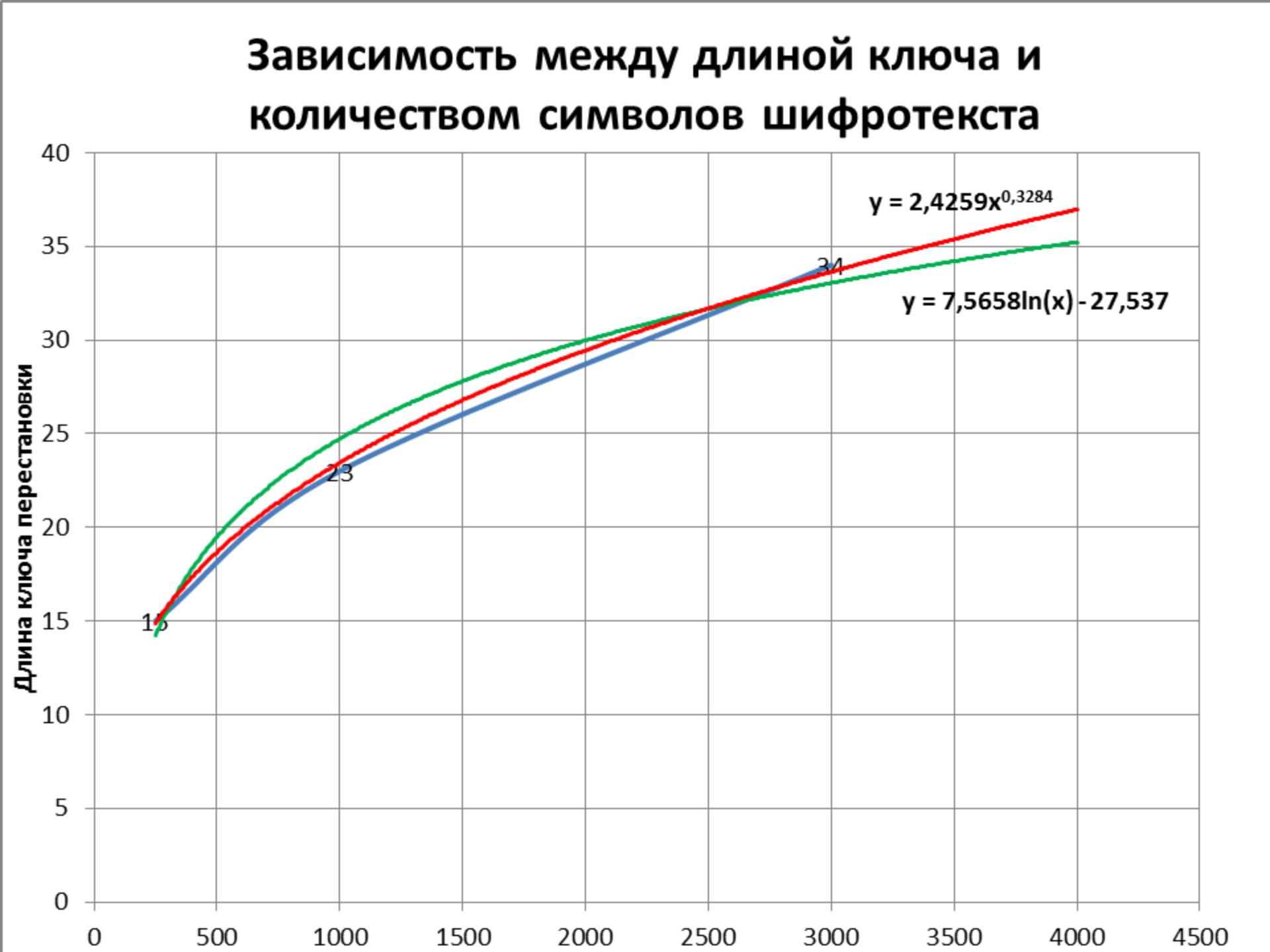
Таблица 2 «Оценка эффективности дешифрования при длине текста 1000 символов, включая пробелы»

Длина ключа	Эффективность дешифрования сообщения	Разборчивость частично дешифрованного сообщения	Первый столбец перестановки найден	Последний столбец перестановки найден	Одновременно первый и последний столбцы перестановки найдены	Число испытанных ключей
2	100%	отличная	100%	100%	100%	2
3	100%	отличная	100%	100%	100%	6
4	100%	отличная	100%	100%	100%	24
5	100%	отличная	100%	100%	100%	25
6	100%	отличная	100%	100%	100%	25
7	92%	отличная	100%	100%	100%	25
8	84%	отличная	100%	100%	100%	25
9	76%	отличная	96%	92%	92%	25
10	72%	хорошая	88%	84%	84%	25
11	64%	хорошая	80%	84%%	80%	25
12	52%	хорошая	76%	80%	76%	25

Таблица 3 «Оценка эффективности дешифрования при длине текста 3000 символов, включая пробелы

Длина ключа	Эффективность дешифрования сообщения	Разборчивость частично дешифрованного сообщения	Первый столбец перестановки найден	Последний столбец перестановки найден	Одновременно первый и последний столбцы перестановки найдены	Число испытанных ключей
2	100%	Отличная	100%	100%	100%	2
3	100%	Отличная	100%	100%	100%	6
4	100%	Отличная	100%	100%	100%	24
5	100%	Отличная	100%	100%	100%	25
6	100%	Отличная	100%	100%	100%	25
7	100%	Отличная	100%	100%	100%	25
8	100%	Отличная	100%	100%	100%	25
9	100%	Отличная	100%	100%	100%	25
10	100%	Отличная	100%	100%	100%	25
11	100%	Отличная	100%	100%	100%	25
12	100%	Отличная	100%	100%	100%	25
13	100%	Отличная	100%	100%	100%	25

С увеличением длины ключа увеличивается общее число возможных его вариантов, и тем сильнее «рассеивается» исходный текст (буквы переставляются на большее расстояние), вследствие чего сложнее дешифровать текст. Однако часть ключа восстанавливается, текст частично может читаться. Условимся употреблять термин «критическая разборчивость», если смысл дешифрованного текста на грани его установления. В обсуждаемом методе дешифрования для каждой длины ключа D шифртекста существует минимальная длина шифртекста L, при которой наступает критическая разборчивость дешифрованного текста. Представляет интерес исследование зависимости L от D (или D от L). На основе проведенных экспериментов выявлено, что «критическая разборчивость» равна: 15% для 250 символов текста; 23% для 1000 символов и 34% для 3000 символов. По этим трем точкам, с использованием средства MS Excel, был построен график (рис. 1) прогнозируемого тренда зависимости длины ключа и количества символов в шифртексте, при которых наступает «критическая разборчивость» дешифрованного текста.


Зависимость между длиной ключа и количеством символов шифротекста
Длина ключа перестановки
y = 2,4259x^0,3284
y = 7,5658ln(x) - 27,537
34
23
15
0
5
10
15
20
25
30
35
40
0
500
1000
1500
2000
2500
3000
3500
4000
4500

- Из рисунка видно, что зависимость носит логарифмический характер. Это означает, что при увеличении длины текста максимальная длина ключа, которую данным методом можно восстановить с критической разборчивостью растет незначительно.
- Доверяясь приведенному графику можно считать, что при длине ключа 40 и выше для дешифрования нашим методом потребуются очень длинные шифртексты.