МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ НИЖЕГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ им. Н.И. ЛОБАЧЕВСКОГО»

РАДИОФИЗИЧЕСКИЙ ФАКУЛЬТЕТ

КАФЕДРА «БЕЗОПАСНОСТЬ ИНФОРМАЦИОННЫХ СИСТЕМ»

Научно-исследовательская работа

Работу выполнил: студент

Группы \_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_ Гусев Ю.С.

(подпись)

Номер зачётной книжки

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Проверил:

\_\_\_\_\_\_\_\_\_\_\_ Горбунов А.А.

(подпись)

Н. Новгород, 2018

**1. Введение.**

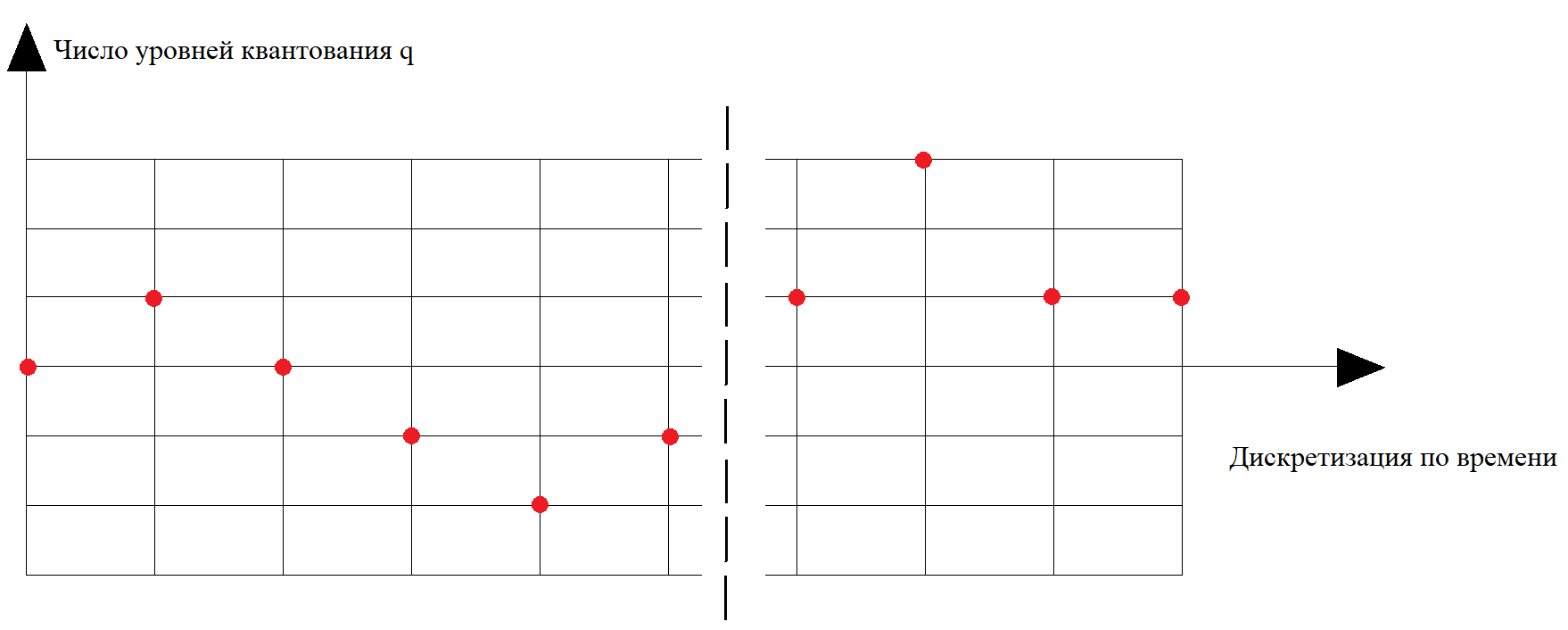
На современном этапе развития системы защиты информации получили довольно широкое распрстранение и задействованы в той или иной степени в жизни каждого человека. Поэтому задачи проверки устойчивости, оценки параметров и сравнения различных криптосистем не теряют своей актуальности.

В данной работе будет рассмотрена тема оценки параметров КС через методы структурной идентификации математической модели КС, базирующиеся на определение характеристик последовательностей, снятых в определенных точках автомата.

**2. Базовые параметры.**

Будем рассматривать текста на входе и выходе шифрующего автомата как сигналы, порождаемые источником экспериментальных данных. На каждом участке стационарности источник обладает набором базовых параметров

{q, n, Δt}. Δt – интервал квантования по времени, q – число уровней квантования, n – «сложность» источника на участке стационарности.

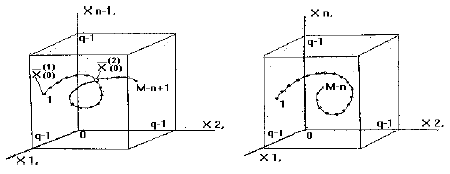


Для оценки численных значений БП требуются только сами «сигналы» источника данных и минимальная априорная информация (например, алфавит).

При определении базовых параметров математической модели источника данных решается задача структурной идентификации источника. После их определения можно перейти к нахождению «свободных» параметров математической модели. К ним относятся, например, коэффициенты матриц, вектора начального состояния.

**3. Поиск БП. Участки стационарности.**

Графически поиск БП можно изобразить в виде фазовой траектории, находящейся в гиперкубе со сторонами из q дискрет. Каждой точке этой ФТ будут соответствовать n последовательных выборок данных из сигнала. Так как число точек в дискретном фазовом пространстве ограничено, то траектория может замкнуться, тем самым выделяя участок стационарности. В таком случае при фиксированном q увеличивается n, то есть размерность гиперкуба, и процесс повторяется, до тех пор, вся ФТ не перестанет зацикливаться, а весь текст станет «q-n стационарным».

****

Если сигнал является не стационарным, то таким методом можно точно определить границы участков стационарности и БП для каждого из них. Помимо основных БП, сложный нестационарный сигнал обладает расширенным набором базовых параметров:

t(i) – координата конца i-го участка траектории,

s(i) – число точек до входа в цикл на i-м участке ФТ,

c(i) – число точек в цикле i-го участка ФТ,

cc(i) – полное число точек на цикле i-го участка ФТ.

После определения БП можно перейти к определению рабочих параметров, то есть к решению задачи параметрической идентификации ММ источника экспериментальных данных.

**4. Методы вычисления БП.**

Обозначим открытый текст ut, а шифрованный yt.

По своей сути q – размерность алфавита сигнала, а n – минимальная длина неповторяющейся подпоследовательности сигнала, состоящей из n последовательных выборок. Значит, вычислять q не нужно, так как он заранее известен.

Для вычисления параметра n есть несколько способов, например:

- Алгоритмы, основанные на последовательном переборе всех возможных последовательностей длины n. То есть, берется выборка последовательных n символов и сравнивается со всеми следующими выборками последовательных n символов. Если «двойник» не обнаружен, то берется следующая n-выборка и процесс повторяется. В итоге либо повторяющихся выборок не будет и данное n является базовым параметром текста, либо обнаружится «двойник», что приведет к увеличению n на 1 и повторению алгоритма. Алгоритм обладает небольшими затратами памяти, а время его работы подобного оценивается как О(М2), то есть он применим для относительно коротких текстовых последовательностей или в случаях, когда время не важно.

- Алгоритм бинарного поиска среди отсортированных по определенному правилу n-последовательностей. Общее время работы оценивается как

*.* Таким образом, данный алгоритм применим для более больших объемов текста, но при высокой скорости работы он затрачивает больше ресурсов памяти на хранение упорядоченного списка последовательностей.

- Алгоритмы, использующие деревья для поиска повторяющихся строк. Например, алгоритм Укконена с использованием суффиксального дерева или алгоритм Ахо-Корасик. Оба алгоритма выполняют свою работу за линейное относительно длины обрабатываемой текстовой последовательности. Ограничивают применение подобных алгоритмов большие затраты памяти при большой размерности алфавита q и большой длине самой последовательности.

В данной работе остановимся на первом методе, так как он является наиболее простым в реализации, а временные затраты на данный момент не играют особой роли.

**5. Ненадежность.**

Понятие энтропии сообщения введено Шеноном через вероятность появления различных вариантов текста.

– энтропия, или мера неопределенности, выходного сообщения.

Аналогично для входного сообщения u(i) :

Суммирование осуществляется по всем возможным вариантам сообщения.

Так же вводится понятие совместной энтропии входного и выходного сообщений:

Средняя условная энтропия исходного сообщения

получила название *функции ненадежности текста сообщения*.

Энтропия текста выражается через его базовые параметры как

Также Шеноном введено понятие «расстояние единственности», то есть длина криптограммы, при которой вероятность только одного варианта расшифровки близка к 1.

- энтропия ключа шифрования, где - средняя длина участка стационарности.

- оценка величины избыточности текста.

Допустим, что для шифрования сообщения на некотором реальном языке используется алгоритм шифрования Цезаря, в котором каждая буква исходного алфавита заменяется другой буквой этого же алфавита по определенному правилу. Если будет перехвачено сообщение большой длины (N), то велика вероятность существования единственного варианта расшифровки, который будет нести смысл. Если же будет перехвачено сообщение малой длины, то уже не получится так просто по смыслу определить правильный вариант расшифровки. Это N и есть расстояние единственности.

Таким образом, на основании выходного текста и БП можно сделать вывод о надежности шифрующего алгоритма или сравнить различные алгоритмы шифрования между собой. Также данные функции могут использоваться для описания канала связи с шумом, где шум является «шифрующей» составляющей. Отличия, во-первых, в том, что алгоритмы шифрующих преобразований обычно имеют более сложную природу, чем искажения, вызванные наложением шума, и, во-вторых, шум является непрерывным и выбирается из бесконечного множества, а ключ в КС – из конечного.

**6. Рассматриваемые алгоритмы шифрования.**

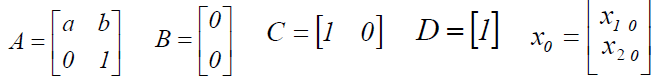
Рассмотрим несколько алгоритмов шифрования.

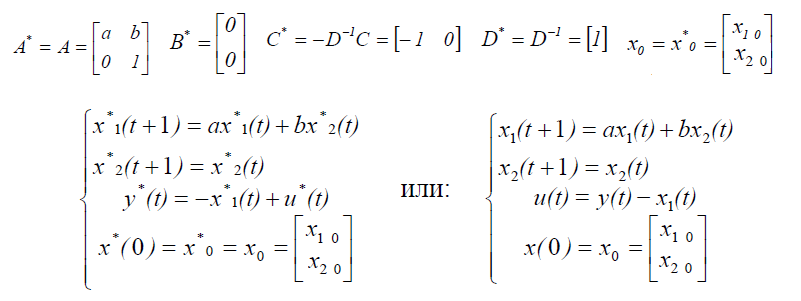
Первым будет один из простейших – метод шифрования гаммированием. В данном методе шифрование исходного сообщения происходит путем его сложения с гаммой по модулю, равному размерности алфавита исходного сообщения.

Гамма – это псевдослучайная последовательность чисел, порождаемая каким-либо алгоритмом.

Для получения гамма будет использован линейный конгруэнтный метод, с помощью которого будет генерироваться псевдослучайное 64-битное число, а затем находиться ближайшее к нему простое число .Система уравнений, описывающая шифратор, будет иметь следующий вид:

Матрицы будут иметь следующий вид:

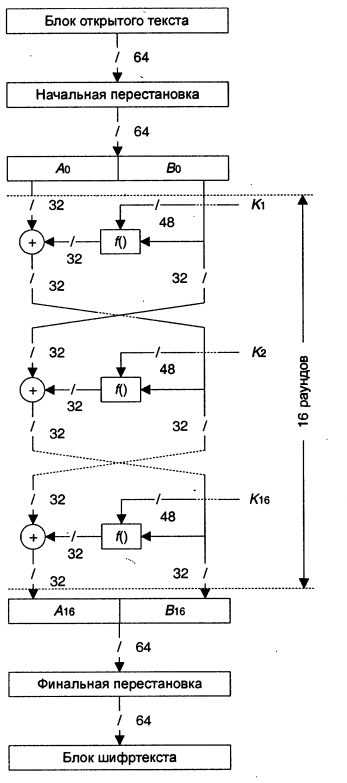


Соответственно, для дешифратора:  


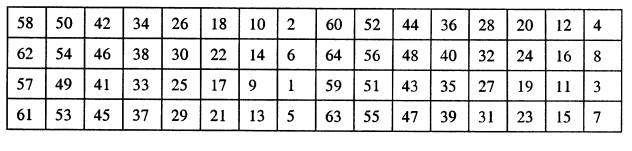
Дешифрование происходит путем применения алгоритма повторно с тем же ключем.

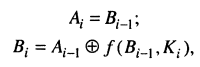
Второй рассматриваемый метод шифрования – RDES. Это обычный DES с добавленной перестановкой. Сам DES был принят в 1977 году как стандарт США на алгоритм шифрования данных.

Структура алгоритма DES представлена на рисунке ниже. Входной текст разбивается на 64-битные блоки, к которым применяется алгоритм.

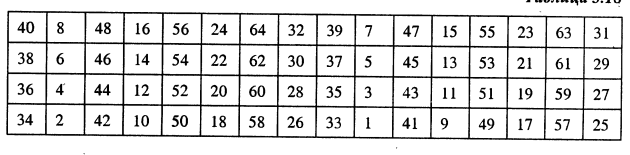


Начальная перестановка выполняется с использование таблицы перестановки.

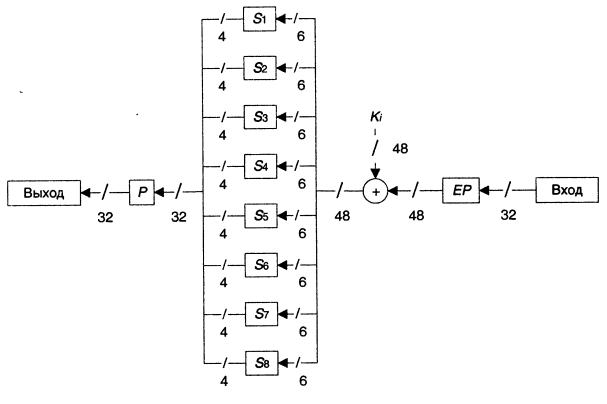
Далее выполняются 16 раундов преобразований, которые можно описать формулой



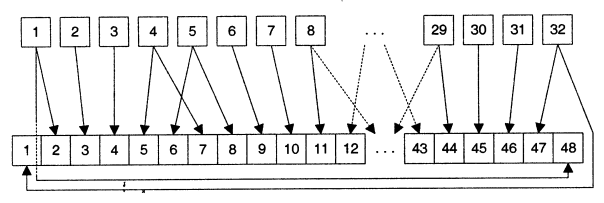
После применяется финальная перестановка

и на выходе получается зашифрованный текст.

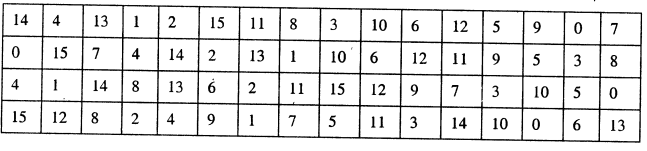
Структура функции f() представлена на рисунке ниже.

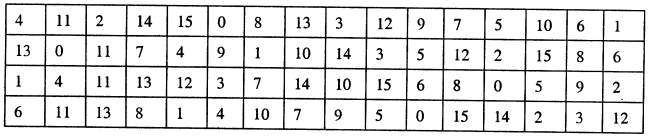
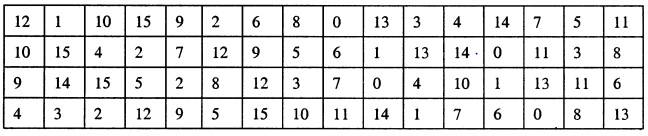
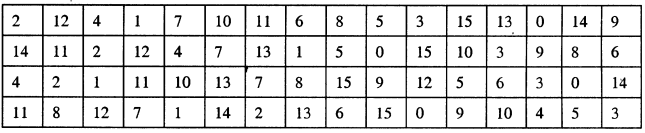
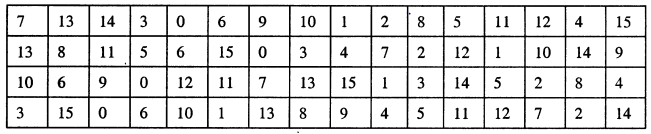
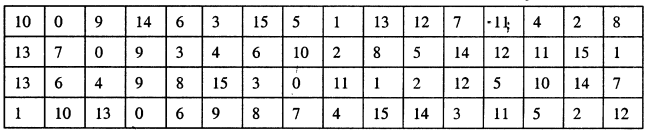
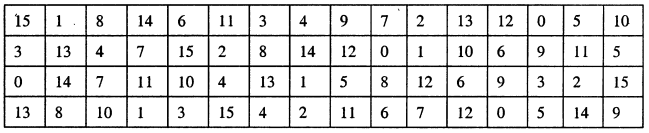


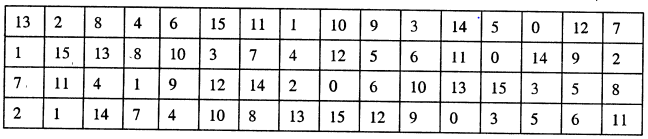
Расширяющая перестановка EP:



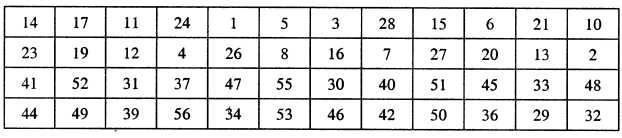
S-перестановки:



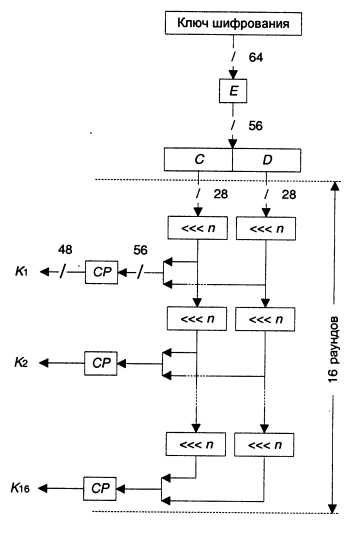




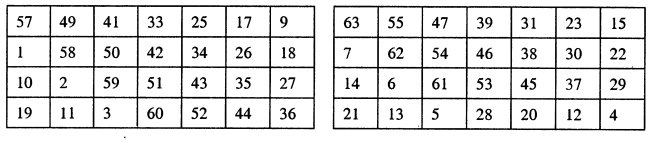
Сжимающая перестановка P:



Так как ключ изначально один, а раундов 16, то применяется процедура расширения ключа.



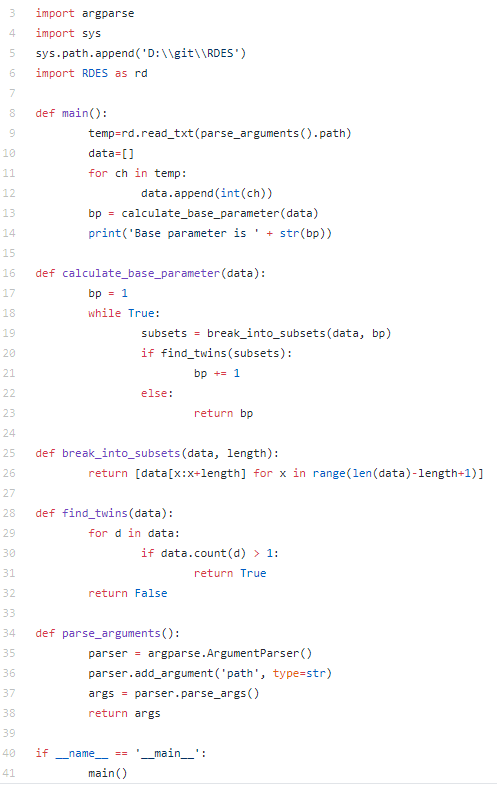
Перестановки С и D:



Дешифрование происходит путем применения алгоритма шифрования к шифрованному тексту, но с использованием ключей раундов в обратном порядке.

**7. Эксперимент по определению базовых параметров.**

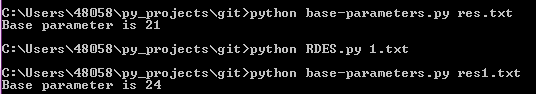
Как уже было описано выше, для нахождения базовых параметров остановимся на алгоритме перебора, так как он наиболее прост в реализации, а затрачиваемое время в данный момент не критично.

Алгоритм посимвольно считывает последовательность, подаваемую на вход, а затем последовательным перебором берет последовательность длины n и ищет ее «двойников». На выходе получаем n.

В параметрах запуска программы задается путь до анализируемого файла.

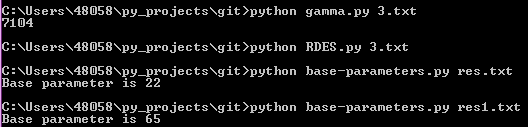
Ключи алгоритмов генерируются одинаковым образом и во всех случаях 64-битные.

Для начала, проверим последовательности небольшой длины, а именно 375 символов. Длина шифрованной последовательности после применения алгоритмов составляет 4224 двоичных символа. В данном случае базовый параметр в обоих случаях примерно совпадают.



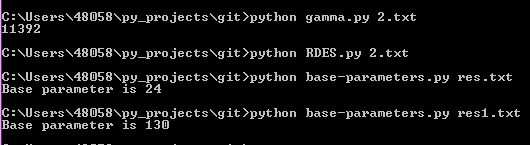
21 – n-параметр в случае алгоритма гаммирования, 24 – RDES.

Увеличим размер входного текста до 632 символов, при этом шифрованные последовательности будут иметь длину 7104 двоичных символа.



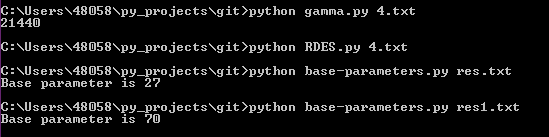
22 – гаммирование, 65 – RDES.

1020 символов входного текста и 11392 шифрованного.



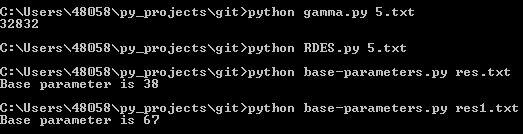
24 – гаммирование, 130 – RDES.

1937 символов входного текста и 21440 шифрованного.



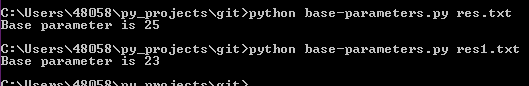
27 – гаммирование, 70 – RDES.

2955 символов входного текста и 32832 шифрованного.



38 – гаммирование, 67 – RDES.

290 символов входного текста и 3200 шифрованного.



Сведем полученные результаты в таблицу.

|  |  |  |  |
| --- | --- | --- | --- |
| Количество входных символов | Количество шифрованных двоичных символов | n-параметр / энтропия | |
| Гаммирование | RDES |
| 290 | 3200 | 25 | 23 |
| 375 | 4224 | 21 | 24 |
| 632 | 7104 | 22 | 65 |
| 1020 | 11392 | 24 | 130 |
| 1937 | 21440 | 27 | 70 |
| 2955 | 32832 | 38 | 67 |

Проверим зависимость на другом тексте:

|  |  |  |
| --- | --- | --- |
| Количество шифрованных двоичных символов | Гаммирование | RDES |
| 2304 | 24 | 20 |
| 10624 | 27 | 23 |
| 17472 | 28 | 27 |
| 23872 | 25 | 29 |
| 31360 | 29 | 33 |
| 35648 | 28 | 66 |

На примере второго текста также прослеживается закономерность роста базового параметра n с ростом количества символов.

Также можно заметить, что в обоих случаях на малых последовательностях n гаммирования больше ил равно n RDES, но на больших – n гаммирования меньше.

С учетом указанных ранее уравнений и их упрощения можно сделать вывод, что на одних и тех же входных последовательностях RDES позволяет получить более криптостойкую шифрограмму с большей энтропией.

**8. Вывод.**

Исследовав способы нахождения БП выходных последовательностей шифрующих автоматов и реализовав один из них практически, было выяснено, что БП параметры являются удобной мерой анализа сложности алгоритмов и их сравнения между собой.

**9.Список литературы.**

1) Ю.В. Романец, П.А. Тимофеев, В.Ф. Шаньгин – «Защита информации в компьютерных системах и сетях» издание второе, дополненное – Москва, «Радио и связь», 2001 г.

2) С.П. Панасенко – «Алгоритмы шифрования. Специальный справочник» – Санкт-Петербург, «БХВ-Петербург, 2009 г.

3) К.Г. Кирьянов – «Выбор оптимальных базовых параметров источников экспериментальных данных при их идентификации – Труды III Международной конференции «Идентификация систем и задач управления» SICPRO ’04 Москва 28-30 января 2004 г.

4) А.А. Горбунов, К.Г. Кирьянов – «Связь “Функции ненадежности” и “Расстояния единственности” криптосистем с базовыми параметрами шифратора в форме математической модели синхронного автомата Хаффмана»

5) А.А. Горбунов, К.Г. Кирьянов – «Динамические модели криптосистем с закрытым ключом. Синтез дешифраторов»

6) А.А. Горбунов – «Алгоритмы структурной идентификации математических моделей криптосистем на основе определения базовых параметров» – Доклады ТУСУРа, № 1 (19), часть 2, июнь 2009

7) К. Шенон – «Работы по теории информации и кибернетике»