## Лабораторная работа №2

## Исследование статистических свойств программного датчика ПСП

Цель: исследование статистических характеристик ПСП и их влияния на результат криптографического преобразования.

### Теоретические сведения

### Определение статистических характеристик ПСП

Двоичные случайные последовательности характеризуются следующими основными статистическими характеристиками:

- частота появления комбинаций бит заданной длины;
- закон распределения вероятностей появления комбинаций заданной длины, содержащих одинаковое количество единиц в пределах комбинации;
- функция автокорреляции.

1) Частота появления комбинаций бит заданной длины

Частота появления $p_k$ комбинации $k$ бит определяется отношением числа появлений этой комбинации $M_k$ к общему числу проанализированных комбинаций в последовательности $M$:

$$p_k = \frac{M_k}{M}.$$

Для вычисления частоты появления комбинаций длиной один бит необходимо подсчитать количество нулей и единиц в последовательности и разделить полученные числа на общее количество бит последовательности. Для определения частоты появления комбинаций длиной два и более бит, необходимо анализировать комбинации в скользящем режиме, как показано в примере:

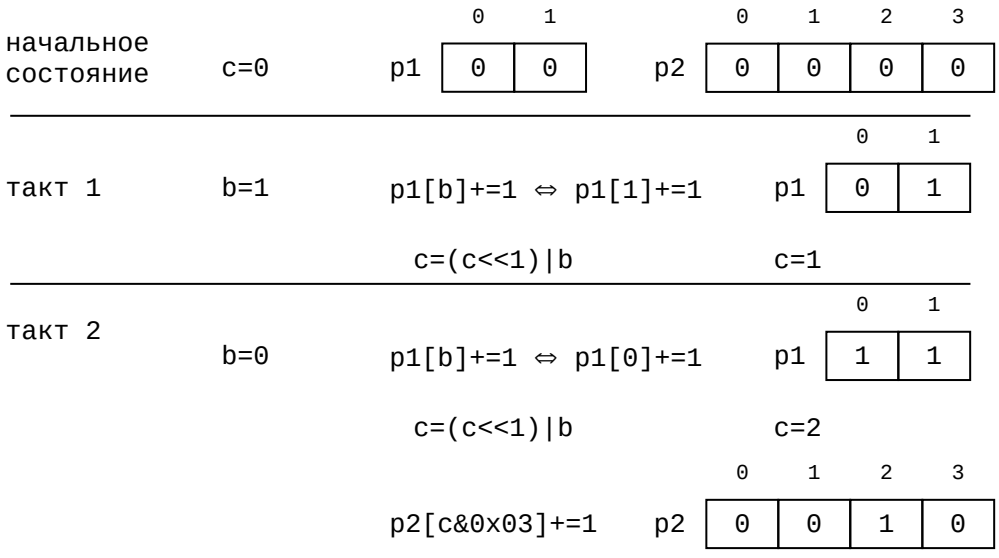
```
последовательность:   01101100101
        комбинации:   01
                       11
                        10
                         01
                          11
                           10
                            00
                             01
                              10
                               01
количество комбинаций: M=10
                       M00=1, p00=0,1
                       M01=4, p01=0,4
                       M10=3, p10=0,3
                       M11=2, p11=0,2
```

При подсчете частоты встречи комбинаций нет необходимости сначала формировать всю ПСП, а только затем определять значения частот встречи комбинаций. Подсчет частоты реализуется в скользящем режиме с использованием целочисленной переменной для хранения части ПСП и массивов, хранящих накопленное число появлений комбинаций определенной длины. Например, требуется определить частоту появления комбинаций длиной 1 и 2 бита в последовательности. Для этого надо определить целочисленную переменную, хранящую часть ПСП. Разрядность переменной должна быть больше, чем необходимая длина комбинации, поэтому будем использовать однобайтовую символьною переменную (обозначим `c`). Объявим два массива: первый массив (обозначим `p1`) из двух элементов будет накапливать число встречи комбинаций из одного бита (нулей и единиц), второй массив (обозначим `p2`) из четырех элементов будет накапливать число встречи комбинаций из двух бит. Нулевой элемент массива `p1` будет накапливать число встречи нуля в ПСП, первый элемент – число появления единицы. В массиве `p2` элементы будут соответствовать комбинациям следующим образом:

| массив p2 | | |
|---|---|---|
| элементы | индексы | индексы в двоичном виде |
| 0 | 0 | 00 |
| 0 | 1 | 01 |
| 0 | 2 | 10 |
| 0 | 3 | 11 |

Таким образом, значение комбинации можно использовать сразу для увеличения соответствующего элемента массива (счетчика числа встречи комбинации).

Рассмотрим процесс подсчета частот подробнее. Пусть переменная b хранит значение бита обратной связи на текущем такте работы генератора. Она может принимать значения 0 или 1. Используя переменную b в качестве индекса массива p1 можно сразу увеличивать число появления в ПСП единиц или нулей. Затем необходимо сдвинуть переменную c в сторону старших разрядов, и в младший разряд записать значение бита обратной связи. Два младших разряда переменной c будут определять индекс элемента массива p2, значение которого надо увеличить.




Из рисунка видно, что на первом такте можно увеличивать счетчики числа появления комбинаций длиной один бит, на втором также – число появлений комбинаций длиной один и два бита, и т.д. В итоге частота

встречи комбинаций будет определяться следующим образом. Пусть сформировано $10^3$ бит последовательности.

количество комбинаций: $M=10^3$

Массив p1:

| | 0 | 1 |
|---|---|---|
| p1 | 498 | 502 |

Частоты встречи:

$k=0,\ p_0=p1[k]/M=0,498$
$k=1,\ p_1=p1[k]/M=0,502$

Массив p2:

| | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| p2 | 251 | 247 | 249 | 252 |

Частоты встречи:

$k=00,\ p_{00}=p2[k]/(M-1)=0,251$
$k=01,\ p_{01}=p2[k]/(M-1)=0,247$
$k=10,\ p_{10}=p2[k]/(M-1)=0,249$
$k=11,\ p_{11}=p2[k]/(M-1)=0,253$

При вычислении частоты появления комбинации, длиной два бита, из общего количества следует вычитать единицу, поскольку на первом такте работы генератора ПСП комбинации из двух бит еще не было. Аналогично, при определении частоты появления комбинации, длиной три бита, из общего количества следует вычитать два, и т.д.

Распределение частот появления комбинаций различной длины в ПСП должно стремиться к равновероятному, что свидетельствует о правильной реализации генератора ПСП. При проведении статистических исследований на точность полученных результатов влияет объем выборки (в данном случае – количество сгенерированных бит ПСП). Например, теоретическая вероятность появления комбинаций 0 или 1 равна 0,5. Однако экспериментально полученные значения могут отличаться в зависимости от объема выборки.

| Вероятности | Теоретическая вероятность | Объем выборки | | |
|---|---|---|---|---|
| | | 10 | $10^3$ | $10^6$ |
| появления нуля $P_0$ | 0,5 | 0,3 | 0,502 | 0,500002 |
| появления единицы $P_1$ | 0,5 | 0,7 | 0,498 | 0,499998 |

Для правильной оценки статистических характеристик необходимо выбрать объем выборки, при котором увеличение объема выборки на порядок не приводит к существенному уменьшению отклонения вероятностей от теоретических.

Графически результат статистического анализа представляется в виде гистограмм вероятностей появления комбинаций определенной длины. Внешний вид гистограммы вероятностей появления комбинаций длиной 2 бита представлен на рисунке 1, где *P* – вероятности появления комбинаций, *к* – значения комбинаций, ряд «до» соответствует вероятностям появления комбинаций в открытом тексте до маскирования, ряд «после» соответствует вероятностям появления комбинаций в преобразованных данных после маскирования.

Если в задании требуется выполнить сравнение статистических характеристик, то на гистограмму надо наносить данные, соответствующие сравниваемым результатам (в примере сравниваются распределения до и после маскирования).

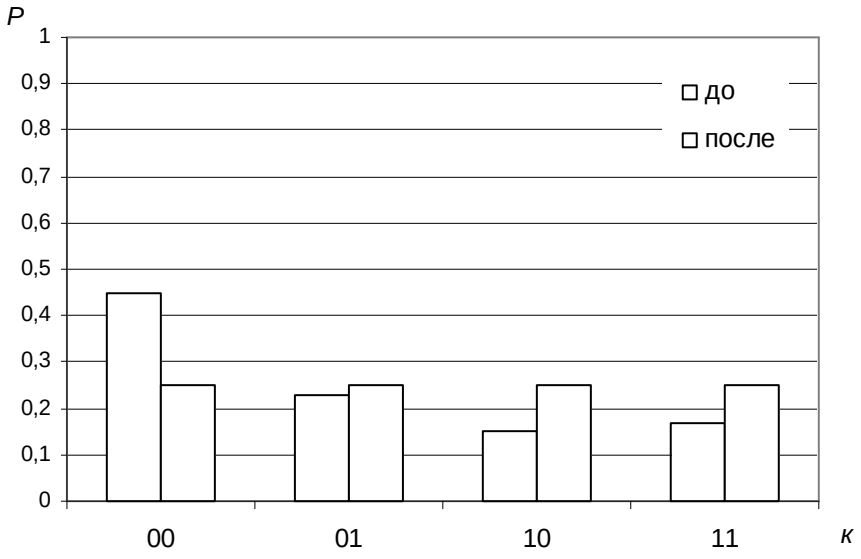



Рисунок 1 – Вероятности появления комбинаций длиной 2 бита

При построении гистограммы следует отключать автоматическое задание масштабов осей – шкала вероятности должна быть от 0 до 1. Допускается установить максимальное значение на шкале менее 1 для

лучшей наглядности. Однако *все гистограммы с одинаковыми данными в отчете должны иметь одинаковый масштаб*.

2) Закон распределения вероятностей появления комбинаций заданной длины, содержащих одинаковое количество единиц

Закон распределения вероятностей появления комбинаций заданной длины, содержащих одинаковое количество единиц в пределах комбинации, можно определить на основе распределения частоты встречи комбинаций. Например, для комбинаций длиной два бита, искомые вероятности появления комбинаций определяются следующим образом:

| Число единиц | Вероятность появления |
|---|---|
| 0 | $p_{00}$ |
| 1 | $p_{01}+p_{10}$ |
| 2 | $p_{11}$ |

Если рассчитать вероятности появления комбинаций заданной длины, содержащих одинаковое количество единиц на основе предыдущего примера, то получатся следующие значения:

| Число единиц | Вероятность появления |
|---|---|
| 0 | $p_{00}=0,251$ |
| 1 | $p_{01}+p_{10}=0,496$ |
| 2 | $p_{11}=0,253$ |

Распределение вероятностей появления комбинаций заданной длины, содержащих одинаковое количество единиц в ПСП должно стремиться к нормальному закону.

Результаты также представить в виде гистограмм.

3) Функция автокорреляции

Функция автокорреляции (АКФ) используется для оценки степени статистической зависимости бит, отстоящих друг от друга в последовательности на $\tau$ бит:

$$K(\tau)=\frac{A(\tau)-B(\tau)}{A(\tau)+B(\tau)},$$

где $A(\tau)$ – количество совпавших бит последовательности, отстоящих друг от друга на $\tau$ отсчетов;

$B(\tau)$ – количество различающихся бит последовательности, отстоящих друг от друга на $\tau$ отсчетов.

Пример расчета АКФ приведен ниже.

```
τ=0   01101100101
      01101100101
      ++++++++++        A=11   B=0    K(0)=1

      01101100101
τ=1    01101100101      A=3    B=7    K(1)=-0,4
       -+--+-+---

      01101100101
τ=1     01101100101     A=3    B=6    K(1)=-0,33
        --+----++
```

АКФ изменяется в пределах [-1;1]. Причем, если $K(\tau)=0$ для всех $\tau\neq0$, то биты последовательности статистически независимы друг от друга. В точке $K(0)$ АКФ ПСП всегда равна 1, что означает совпадение последовательности самой с собой. Значение $K(\tau)=-1$ говорит о том, что последовательность в этой точке инверсная сама себе.

Графически АКФ удобно представить в виде «точечного графика», где значению оси аргумента ($\tau$) соответствует определенное значение $K(\tau)$. Пример приведен на рисунке 2. При выборе масштаба следует учесть, что АКФ изменяется от -1 до 1 включительно. При оформлении не надо изображать линии графика между точками, т.к. АКФ является дискретной (точки $\tau=0,5$ быть не может, соответственно никакой линии, соединяющей точки быть не должно)

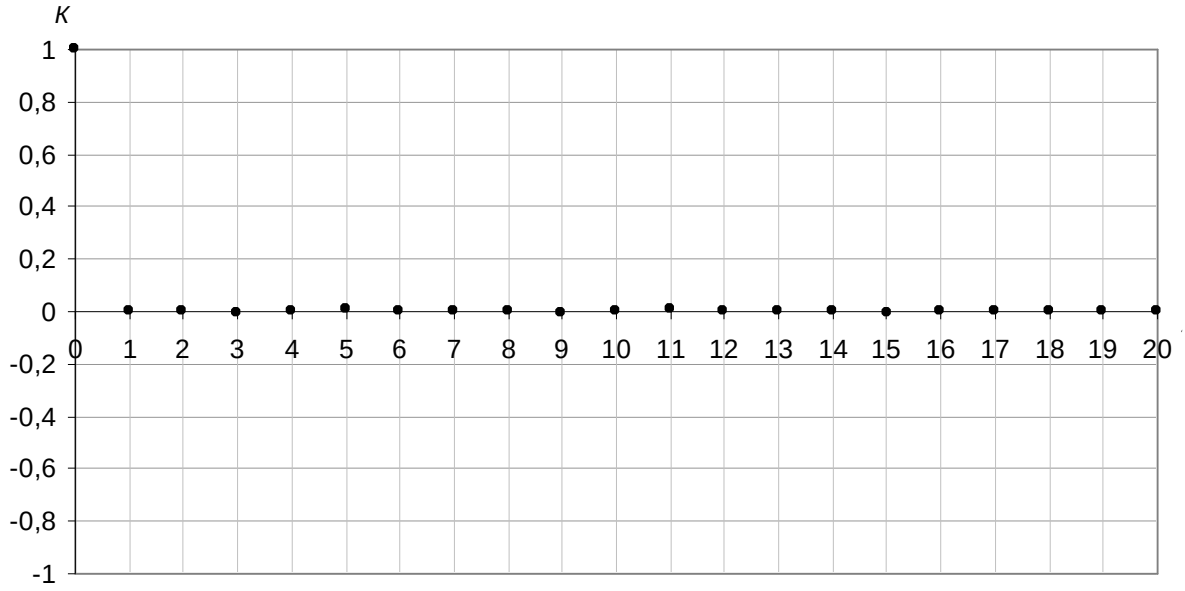



Рисунок 2 – Функция автокорреляции ПСП

## Задание на лабораторную работу

1 Дополнить программу, реализующую комбинирующий генератор ПСП, функционалом, позволяющим определять статистические характеристики вырабатываемой последовательности, маскируемых данных и замаскированных данных. В программе должна быть предусмотрена возможность указания длины комбинаций, для которых определяются вероятности появления, и возможность указания значения сдвига при вычислении АКФ.

2 Провести статистические исследования ПСП. Определить:

- объем выборки для проведения статистических исследований ПСП;
- вероятности появления комбинаций длиной от 1 до 4 бит;
- вероятности появления комбинаций длиной от 2 до 4 бит, содержащих одинаковое количество единиц;
- автокорреляционную функцию ПСП для $\tau$ от 0 до 32 бит;

Результаты исследования представить в графическом виде. Сделать выводы по результатам исследования.

2 Провести маскирование текстового файла (размер файла не менее 20 Кб, текст должен иметь семантическое содержание). Определить статистические характеристики файла до и после маскирования (распределение вероятностей появления символов кодовой таблицы). Результаты исследования представить в графическом виде. Сделать выводы по результатам исследования. Привести содержимое файла до и после маскирования.

3 Внести преобладания в маскирующую гамму (путем инверсии каждой *n*-й единицы в последовательности). Провести маскирование текстового файла. Определить статистические характеристики файла после маскирования. Сделать выводы по результатам исследования о влиянии качества ПСП на качество маскирования.

4 (дополнительное) Замаскировать файл формата bmp без маскирования заголовка. Для маскирования использовать полиномы из таблицы вариантов (гамма с малым периодом). Размер файла должен быть не менее 200 Кб. Сделать выводы по результатам исследования о влиянии периода ПСП на качество маскирования.

Таблица вариантов

| Номер варианта | Полином 1 | Полином 2 | шаг преобладания |
|---|---|---|---|
| 1 | $x^6+x^1+1$ | $x^5+x^4+x^2+x^1+1$ | 3 |
| 2 | $x^7+x^3+1$ | $x^6+x^5+x^2+x^1+1$ | 4 |
| 3 | $x^7+x^3+1$ | $x^7+x^3+x^2+x^1+1$ | 5 |
| 4 | $x^6+x^1+1$ | $x^5+x^4+x^2+x^1+1$ | 3 |
| 5 | $x^5+x^2+1$ | $x^6+x^5+x^2+x^1+1$ | 4 |
| 6 | $x^5+x^2+1$ | $x^7+x^3+x^2+x^1+1$ | 5 |
| 7 | $x^7+x^3+1$ | $x^5+x^4+x^2+x^1+1$ | 3 |
| 8 | $x^7+x^3+1$ | $x^6+x^5+x^2+x^1+1$ | 4 |
| 9 | $x^6+x^1+1$ | $x^7+x^3+x^2+x^1+1$ | 5 |
| 10 | $x^5+x^2+1$ | $x^5+x^4+x^2+x^1+1$ | 3 |
| 11 | $x^7+x^3+1$ | $x^6+x^5+x^2+x^1+1$ | 4 |
| 12 | $x^9+x^4+1$ | $x^7+x^3+x^2+x^1+1$ | 5 |
| 13 | $x^7+x^3+1$ | $x^5+x^4+x^2+x^1+1$ | 3 |
| 14 | $x^6+x^1+1$ | $x^6+x^5+x^2+x^1+1$ | 4 |
| 15 | $x^5+x^2+1$ | $x^7+x^3+x^2+x^1+1$ | 5 |
| 16 | $x^5+x^2+1$ | $x^5+x^4+x^2+x^1+1$ | 3 |
| 17 | $x^7+x^3+1$ | $x^6+x^5+x^2+x^1+1$ | 4 |
| 18 | $x^7+x^3+1$ | $x^7+x^3+x^2+x^1+1$ | 5 |
| 19 | $x^6+x^1+1$ | $x^5+x^4+x^2+x^1+1$ | 3 |
| 20 | $x^5+x^2+1$ | $x^6+x^5+x^2+x^1+1$ | 4 |
| 21 | $x^5+x^2+1$ | $x^7+x^3+x^2+x^1+1$ | 5 |
| 22 | $x^6+x^1+1$ | $x^5+x^4+x^2+x^1+1$ | 3 |
| 23 | $x^7+x^3+1$ | $x^6+x^5+x^2+x^1+1$ | 4 |
| 24 | $x^6+x^1+1$ | $x^7+x^3+x^2+x^1+1$ | 5 |
| 25 | $x^5+x^2+1$ | $x^5+x^4+x^2+x^1+1$ | 3 |