МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное
учреждение высшего образования
«Юго-Западный государственный университет»
(ЮЗГУ)

Кафедра защиты информации и систем связи

УТВЕРЖДАЮ
Проректор по учебной работе
_____О.Г. Локтионова
«___» _____2015 г.

КРИПТОАНАЛИЗ ШИФРА ПОЛИАЛФАВИТНОЙ ПОДСТАНОВКИ

Методические указания по выполнению лабораторной работы
для студентов специальностей 10.05.03, 10.05.02, 10.03.01

Курск 2015

УДК 004.056.55 (076.5)

Составители: М.А. Ефремов

Рецензент
Кандидат технических наук, доцент *М.О. Таныгин*

Криптоанализ шифра полиалфавитной подстановки: методические указания по выполнению лабораторной работы / Юго-Зап. гос. ун-т; сост.: М.А. Ефремов. Курск, 2015. 17 с.: табл. 11, Библиогр.: с. 17.

Содержат основные сведения о сущности криптоанализа полиалфавитных подстановок, на основе использования статистических методов и частотного анализа. Указывается порядок выполнения лабораторной работы, правила оформления и содержание отчета.

Методические указания соответствуют требованиям программы, утвержденной учебно-методическим объединением по образованию в области информационной безопасности (УМО ИБ).

Предназначены для студентов специальностей 10.05.03, 10.05.02, 10.03.01 дневной формы обучения.

Текст печатается в авторской редакции

Подписано в печать . Формат 60x84 1/16.
Усл.печ. л. . Уч.-изд.л. . Тираж 30 экз. Заказ. Бесплатно.
Юго-Западный государственный университет.
305040, г. Курск, ул. 50 лет Октября, 94.

СОДЕРЖАНИЕ

1. ЦЕЛЬ РАБОТЫ

Цель лабораторной работы - определив индекс соответствия предлагаемой криптограммы, дешифровать криптограмму и найти ключ.

2. ЗАДАНИЕ

Ознакомьтесь с руководством пользователя и с теоретическим материалом. Запустите исполняемый файл и выберите необходимый номер варианта. Требуется дешифровать криптограмму, зашифрованную методом полиалфавитной подстановки, для этого необходимо определить индекс соответствия предлагаемой криптограммы и найти исходный ключ шифрования.

3. ПОРЯДОК ВЫПОЛНЕНИЯ РАБОТЫ

1. Получить задание.
2. Изучить теоретическую часть.
3. Определить индекс соответствия.
4. На основе частотного анализа отдельных групп криптограммы получить ключ.
5. Составить отчет.

4. СОДЕРЖАНИЕ ОТЧЕТА

1. Титульный лист.
2. Краткая теория.
3. Описание выбора индекса соответствия.
4. Краткий протокол криптоанализа.
5. Расшифрованный исходный текст и ключ.
6. Вывод.

5. ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

5.1 Введение

В случае моноалфавитных подстановок используется только один алфавит шифрования. Существуют шифры, где используется целый набор алфавитов шифрования. Такие шифры называются полиалфавитными и позволяют, в отличие от моноалфавитных подстановок, скрыть естественную частоту появления символов в тексте.

Простая полиалфавитная подстановка (или шифр Виженера) последовательно и циклически меняет используемые алфавиты шифрования. Число используемых алфавитов называется периодом шифра. Для шифрования используется ключ - слово или бессмысленный набор символов нормативного алфавита. Каждая буква ключа определяет свой алфавит шифрования, который получается из нормативного циклического сдвига на количество символов, равное числовому эквиваленту буквы ключа. Очевидно, что длина ключа равна периоду шифра.

Чтобы зашифровать сообщение шифром Виженера, поступают следующим образом. Под каждой буквой открытого текста помещается буква ключа. Ключ циклически повторяется необходимое число раз. Буквы ключа определяют величину смещения символов криптограммы относительно символов открытого текста.

Зашифруем, текст “полиалфавитная_подстановка” ключом “краб”. Будем использовать алфавит, приведенный в таблице 1.

Таблица 1 – Исходный алфавит

Нормативный алфавит	А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П
Числовые эквиваленты	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Таблица 1 (продолжение)

Нормативный алфавит	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ы	Ь	Э	Ю	Я	“_“
Числовые эквиваленты	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31

Процесс шифрования приведен в таблице 2.

Таблица 2 – Шифрование текста ключом «краб»

П	15	К	10	(15 + 10) mod 32	25	Щ
О	14	Р	16	(14 + 16) mod 32	30	Я
Л	11	А	0	(11 + 0) mod 32	11	Л
И	8	Б	1	(8 + 1) mod 32	9	Й
А	0	К	10	(0 + 10) mod 32	10	К
Л	11	Р	16	(11 + 16) mod 32	27	Ь
Ф	20	А	0	(20 + 0) mod 32	20	Ф
А	0	Б	1	(0 + 1) mod 32	1	Б
В	2	К	10	(2 + 10) mod 32	12	М
И	8	Р	16	(8 + 16) mod 32	24	Ш
Т	18	А	0	(18 + 0) mod 32	18	Т
Н	13	Б	1	(13 + 1) mod 32	14	О
А	0	К	10	(0 + 10) mod 32	10	К
Я	30	Р	16	(30 + 16) mod 32	14	О
_	31	А	0	(31 + 0) mod 32	31	_
П	15	Б	1	(15 + 1) mod 32	16	Р
О	14	К	10	(14 + 10) mod 32	24	Ш
Д	4	Р	16	(4 + 16) mod 32	20	Ф
С	17	А	0	(17 + 0) mod 32	17	С
Т	18	Б	1	(18 + 1) mod 32	19	У
А	0	К	10	(0 + 10) mod 32	10	К
Н	13	Р	16	(13 + 16) mod 32	29	Ю
О	14	А	0	(14 + 0) mod 32	14	О
В	2	Б	1	(2 + 1) mod 32	3	Г
К	10	К	10	(10 + 10) mod 32	20	Ф
А	0	Р	16	(0 + 16) mod 32	16	Р

В результате получилась криптограмма: “ЩЯЛЙКЬФБМШТОКО_Р ШФСУКЮОГФР”.

Чтобы вычислить числовой эквивалент буквы шифртекста, числовой эквивалент буквы ключа складывается по модулю L с числовым эквивалентом буквы открытого текста, где L - мощность нормативного алфавита. Т.е. шифр Виженера описывается следующим выражением:

$$E_i = (M_i + K_i \bmod U) \bmod L, \qquad (1)$$

E_i, M_i - числовые эквиваленты символов криптограммы и открытого текста соответственно;

$K_i \bmod U$ - числовой эквивалент буквы ключа;

L - мощность исходного алфавита;

U - длина ключа или период шифра.

Шифр Цезаря является частным случаем шифра Виженера с периодом, равным единице.

5.2 Криптоанализ полиалфавитных подстановок

Полиалфавитные подстановки маскируют естественную частоту появления символов в шифруемом тексте. Поэтому полиалфавитные подстановки значительно надежнее моноалфавитных. Однако метод частотного анализа применим и здесь. Разобьем криптограмму на блоки так, чтобы число символов в каждом блоке равнялось длине ключа. Символы криптограммы, занимающие одинаковое положение в блоках, имеют одинаковое смещение относительно символов открытого текста, т.е. при их шифровании используется один и тот же алфавит шифрования. В приведенном выше примере 1-ая, 5-ая, 9-ая, ... , (4*i + 1) - ая, буквы имеют смещение, равное десяти - числовому эквиваленту 1-ой буквы ключа “К”.

Описанное свойство дает возможность применить частотный анализ отдельно для каждой группы символов криптограммы, соответствующих определенной букве ключа. Такие группы символов криптограмм называют *группой периода*. Понятно, что число групп периода равно длине ключа.

Частотный анализ по группам ключа позволяет криптоаналитику узнать величину смещения для каждой группы, т.е. ключ шифрования.

Подобный метод криптоанализа применим, если число символов в криптограмме превышает число 20*U, где U - длина ключа.

5.3 Индекс соответствия

Чтобы иметь возможность применить частотный анализ к группам периода, криптоаналитик должен прежде высказать предположение о том, чему может быть равен период. Для этой цели используется *индекс соответствия* (ИС).

ИС представляет собой оценку суммы квадратов вероятностей каждого символа. Теоретически ожидаемые значения ИС вычисляются по формуле

$$ИС_{теор} = A \cdot \frac{(N-m)}{m \cdot (n-1)} + \frac{N \cdot (\mathrm{m}-1)}{L \cdot m \cdot (\mathrm{N}-1)}, \text{где} \qquad (2)$$

$$A = \sum_{i=1}^{L} (p[i]^2);$$

N - число символов в криптограмме;

m - длина ключа;

$p[i]$ - вероятность встречаемости i-ой буквы алфавита;

L - мощность алфавита.

Чтобы вычислить значение ИС для конкретной криптограммы, используют следующую формулу:

$$ИС_{прак} = \frac{\sum_{i=1}^{L} f[i] \cdot (f[i]-1)}{N \cdot (N-1)}, \text{где} \qquad (3)$$

N - число символов в криптограмме;

$f[i]$ - сколько раз i-ая буква встретилась в криптограмме.

Существуют таблицы, содержащие теоретически ожидаемые значения ИС для разных длин ключа. Здесь представлена таблица 3 для алфавита, состоящего из русских букв и пробела. Криптоаналитик, рассчитав ИС анализируемой им криптограммы, может определить ее период по такой таблице.

Однако из-за погрешностей в оценке ИС, его использование становится неэффективным при длине ключа большей, чем десять символов.

Таблица 3 – Теоретически ожидаемые значения ИС

Период	Min значение ИС	Max значение ИС	Среднее знач. ИС
1	0.0684		0.0684
2	0.0409	0.0498	0.0453
3	0.0364	0.0436	0.0400
4	0.0335	0.0405	0.0370
5	0.0327	0.0386	0.0356
10	0.0319	0.0350	0.0337

6. ВЫПОЛНЕНИЕ РАБОТЫ

6.1 Запуск программы

Запустить на выполнение файл rela12.exe.

Выбрать в меню пункт “Полиалфавитные подстановки”.

Нажать на клавишу Enter и выбрать в появившемся списке свой вариант.

6.2 Выбор длины ключа

Задайте длину ключа. Для этого вызовите таблицу индекса соответствия (по клавише F5). В этой таблице содержатся теоретически ожидаемые значения ИС для различных длин ключа и значение ИС, подсчитанное для конкретной криптограммы. На основании значения ИС выберете длину ключа. Для этого нажмите клавишу Esc и введите длину ключа в появившемся окне.

6.3 Подстановка букв ключа

На основе частотного анализа для каждой группы периода осуществить подстановку (одну). Таблица частот символов по группам периода и вероятностей букв русского языка вызывается по клавише F4. Номер анализируемой буквы ключа (т.е. номер текущей группы периода) указан в заголовке правого верхнего окна на экране. Номер текущей группы периода легко изменить, воспользовавшись клавишами Left и Right. Ключ изображается на экране в левом верхнем окне. Существует возможность изменить букву прямо в ключе. Для этого воспользуйтесь клавишей F8. В окне с ключом появится синий прямоугольник, который указывает на текущую букву ключа. Клавишами left и right подвести прямоугольник к нужной букве ключа и нажмите Enter. Теперь введите с клавиатуры букву и нажмите Enter еще раз.

Проанализируйте полученные текст и ключ. При неудовлетворительном результате измените подстановки в некоторых группах периода. При неудаче изменить длину ключа.

6.3 Пример дешифрации криптограммы, зашифрованной полиалфавитной подстановкой

Текст криптограммы:

```
ОНГПРЛЩГЮЧМУРМ-ЬТЯК-ОЗЫЬЫДЫРЫЛЗАСХСКЗВМИЮЬРТ-ОЗАПЧФЗХШМЬК-У-
----------------------------------------------------------------

ТИЭШБЩЦРТ--ЛБЗМЬЗЯКПСКСХ-ЮК-З-ЫНВПНЦЫРЛЫЧЗАПРОГЬЙЧМЩРЧОГРРМ-
----------------------------------------------------------------

РЦЮП-ОЩЦЙУТНРХЕПУЩ
------------------
```

Прежде всего необходимо определить период шифра. Это можно сделать с помощью ИС. Для данной криптограммы ИС=0,0384.

Таблица 4 – Таблица теоретически ожидаемых значений для ИС

Период	Min значение ИС	Max значение ИС	Среднее знач. ИС
1	0.0684		0.0684
2	0.0409	0.0498	0.0453
3	0.0364	0.0436	0.0400
4	0.0335	0.0405	0.0370
5	0.0327	0.0386	0.0356
10	0.0319	0.0350	0.0337

Исходя из таблицы, возможные значения длины ключа равны 4 или 5.

Предположим, длина ключа равна 4.

Теперь проведем анализ по группам периода.

Таблица 5 – Статистика относительно 1-ой буквы ключа

Криптограмма		Русский язык	
Символ	Частота	Буква	Вероятность
Р	0.200	Пробел	0.175
Т	0.086	о	0.090
Ы	0.086	е	0.072
З	0.086	а	0.062
-	0.086	и	0.062
О	0.057	н	0.053
и т.д.		и т.д.	

Сделаем в первой группе периода замену “Р”-“ ”.

Поскольку для шифра Виженера все символы, принадлежащие одной группе периода, имеют одинаковый сдвиг относительно исходного алфавита, программа вычисляет этот сдвиг, исходя из сделанной нами замены, как (“Р”- “ ”)mod32 = 17(“С”) и производит замены для остальных букв данной группы периода.

```
ОНГПРЛЩГЮЧМУРМ-ЬТЯК-ОЗЫЬЫДЫРЫЛЗАСХСКЗВМИЮЬРТ-ОЗАПЧФЗХШМЬК-У-
Ю---  ---М---  ---Б---Ю---К---К---А---Ц---М---О---Я---Д---Щ---

ТИЭШБЩЦРТ--ЛБЗМЬЗЯКПСКСХ-ЮК-З-ЫНВПНЦЫРЛЫЧЗАПРОГЬЙЧМЩРЧОГРРМ-
Б---Р---Б---Р---Ц---А---О---Ц---С---К---Ж---  ---Ш---  ---  ---

РЦЮП-ОЩЦЙУТНРХЕПУЩ
 ---О---Ш---  ---В-
```

Проанализируем вторую группу. Ее статистика имеет вид:

Таблица 6 – Статистика относительно 2-ой буквы ключа

Криптограмма		Русский язык	
Символ	Частота	Буква	Вероятность
Ч	0.114	Пробел	0.175
З	0.086	о	0.090
О	0.086	е	0.072
-	0.086	а	0.062
Л	0.057	и	0.062
Я	0.057	н	0.053
и т.д.		и т.д.	

Заменим “Ч” - “ ”. Но при таких заменах текст должен начинаться с “ЮХ”, что маловероятно. Попробуем другие варианты замен.

Как видно из статистики, для первой группы периода самые частые символы это “Р”, “Т”, “Ы”, “З” и “ ”. А для второй группы периода самые частые символы “Ч”, “З”, “О”, “ ” и “Л”.

Попытаемся подобрать такие замены для первой и второй групп периода, чтобы в тексте и в ключе не встречалось недопустимых диграмм для русского языка.

Замена в первой группе “Т” на “ ” не годится, т.к. тогда текст должен начинаться с “Ы”.

Замена в первой группе “З” на “ ”, а во второй группе “Ч” на “ ” не годится, т.к. тогда текст начинается с “ЖХ”.

Замена в первой группе “З” на “ ”, а во второй группе “З” на “ ” не годится, т.к. тогда ключ начинается с “ИИ”.

Проверив всевозможные замены в первой и во второй группах периода, и, делая замены в других группах для сомнительных случаев, приходим к выводу, что наша гипотеза о длине ключа в 4 символа неверна. Предположим, длина ключа равна 5.

Результаты статистического анализа относительно первой буквы ключа имеют вид:

Таблица 7 – Статистика относительно 1-ой буквы ключа

Криптограмма		Русский язык	
Символ	Частота	Буква	Вероятность
О	0.179	Пробел	0.175
Ь	0.107	о	0.090
М	0.071	е	0.072
Ю	0.071	а	0.062
Т	0.071	и	0.062
Л	0.036	н	0.053
и т.д.		и т.д.	

Сделаем в первой группе периода замену “О” - “ ”. Но тогда текст начинается с пробела.

Сделаем замену “Ь” - “ ”. Но тогда ключ начинается с “Ы”.

Сделаем замену “М”-“ ”.

Результаты статистического анализа относительно второй буквы ключа имеют вид:

Таблица 8 – Статистика относительно 2-ой буквы ключа

Криптограмма		Русский язык	
Символ	Частота	Буква	Вероятность
З	0.214	Пробел	0.175
Н	0.107	о	0.090
Щ	0.107	е	0.072
К	0.107	а	0.062
У	0.071	и	0.062
Ы	0.071	н	0.053
и т.д.		и т.д.	

Сделаем во второй группе периода замену “З”-“ ”.

```
ОНГПРЛЩГЮЧМУРМ-ЬТЯК-ОЗЫЬЫДЫРЫЛЗАСХСКЗВМИЮЬРТ-ОЗАПЧФЗХШМЬК-У-
БЕ---ЯС--- Л---НК---Б ---ЧУ---ЬШ---Ю ---РТ---Б ---З ---НВ---

ТИЭШБЩЦРТ--ЛБЗМЬЗЯКПСКСХ-ЮК-З-ЫНВПНЦЫРЛЫЧЗАПРОГЬЙЧМЩРЧОГРРМ-
ЕА---МО---ТГ---Н ---ДВ---РВ---ОЕ---ЙУ---К ---БЫ--- С---ЦИ---

РЦЮП-ОЩЦЙУТНРХЕПУЩ
ГО---БС---ЕЕ---ВЛ-
```

Результаты статистического анализа относительно третьей буквы ключа имеют вид:

Таблица 9 – Статистика относительно 3-ей буквы ключа

Криптограмма		Русский язык	
Символ	Частота	Буква	Вероятность
Р	0.286	Пробел	0.175
Г	0.071	о	0.090
Я	0.071	е	0.072
С	0.071	а	0.062
В	0.071	и	0.062
А	0.071	н	0.053
и т.д.		и т.д.	

Сделаем в третьей группе периода замену “р” - “ ”. Результаты статистического анализа относительно четвертой буквы ключа имеют вид:

Таблица 10 – Статистика относительно 4-ой буквы ключа

Криптограмма		Русский язык	
Символ	Частота	Буква	Вероятность
П	0.185	Пробел	0.175
М	0.111	о	0.090
Х	0.111	е	0.072
К	0.074	а	0.062
Т	0.074	и	0.062
Ш	0.074	н	0.053
и т.д.		и т.д.	

Сделаем в четвертой группе периода замену “П”-“ ”.

Результаты статистического анализа относительно пятой буквы ключа имеют вид:

Таблица 11 – Статистика относительно 5-ой буквы ключа

Криптограмма		Русский язык	
Символ	Частота	Буква	Вероятность
-	0.333	Пробел	0.175
Ч	0.111	о	0.090
Р	0.074	е	0.072
Ы	0.074	а	0.062
М	0.074	и	0.062
Л	0.037	н	0.053
и т.д.		и т.д.	

Сделаем в пятой группе периода замену “-” - “ ”.

Получили:

ОНГПРЛЩГЮЧМУРМ-ЬТЯК-ОЗЫЬЫДЫРЫЛЗАСХСКЗВМИЮЬРТ-ОЗАПЧФЗХШМЬК-У-
БЕТ РЯСТНЧ Л Э НКНЬ Б ККЫЧУ ЛЛЬШАЕСЮ СЭИРТ В Б П ЧЗ ДИМНВОГ

ТИЭШБЩЦРТ--ЛБЗМЬЗЯКПСКСХ-ЮК-З-ЫНВПНЦЫРЛЫЧЗАПРОГЬЙЧМЩРЧОГРРМ-
ЕАЛИБМО В ТГРЧМН НЬПДВАЕ РВОЧ ОЕС НЙУ ЫЫК П РБЫЙЩЧ С ЗОЦИ Э

РЦЮП-ОЩЦЙУТНРХЕПУЩ
ГОМ БСЕЩУЕЕ ЕЕВЛИ

Т.к. отрыв по частоте самых частых символов в первой и четвертой группах периода менее значителен, чем в остальных группах, попробуем изменить подстановки именно в них.

Заменим в первой группе “Ю” на “ ”. Безуспешно.

Заменим в первой группе “Т” на “ ”. Безуспешно.

Заменим в первой группе “Л” на “ ”. Обратим внимание на ключ “МИСРА”. Возможно, это искаженное “МИСКА”. Изменим соответствующим образом ключ:

```
ОНГПРЛЩГЮЧМУРМ-ЬТЯК-ОЗЫЬЫДЫРЫЛЗАСХСКЗВМИЮЬРТ-ОЗАПЧФЗХШМЬК-У-
ВЕТЕР СТУЧАЛ В ОКНА В КРЫШУ СЛЬШАЛСЯ СВИСТ И В ПЕЧИ ДОМОВОЙ

ТИЭШБЩЦРТ--ЛБЗМЬЗЯКПСКСХ-ЮК-З-ЫНВПНЦЫРЛЫЧЗАПРОГЬЙЧМЩРЧОГРРМ-
ЖАЛОБНО И УГРЮМО НАПЕВАЛ СВОЮ ПЕСЕНКУ БЫЛ ПЕРВЫЙ ЧАС НОЧИ В

РЦЮП-ОЩЦЙУТНРХЕПУЩ
ДОМЕ ВСЕ УЖЕ ЛЕГЛИ
```

Криптограмма расшифрована.

Исходный текст:

ВЕТЕР СТУЧАЛ В ОКНА В КРЫШУ СЛЫШАЛСЯ СВИСТ И В ПЕЧИ ДОМОВОЙ ЖАЛОБНО И УГРЮМО НАПЕВАЛ СВОЮ ПЕСЕНКУ БЫЛ ПЕРВЫЙ ЧАС НОЧИ В ДОМЕ ВСЕ УЖЕ ЛЕГЛИ

Ключ: МИСКА.

7. КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Какие подстановочные шифры вам известны, назовите их?
2. Что такое шифр Виженера?
3. Возможно ли применение статистических методов криптоанализа к полиалфавитным шифрам?
4. Что такое индекс соответствия криптограммы?

8. СПИСОК ИСПОЛЗОВАННЫХ ИСТОЧНИКОВ И ЛИТЕРАТУРЫ

1. Н. Смарт. Криптография [текст] Издательство: М.: Техносфера, 2005. – 528 с.

2. Джесси Рассел, Рональд Кон. Полиалфавитный шифр. Издательство: VSD. 2013. – 72с.

3. Кузьминов Т. В. Криптографические методы защиты информации. Новосибирск, 1998.

4. Нильс Фергюсон, Брюс Шнайер. Практическая криптография [текст] Издательство: Вильямс. 2005.- 416 с.

5. Нечаев В. И. Элементы криптографии (Основы теории защиты информации). М.: Высшая школа, 1999.

6. Брюс Шнайер. Прикладная криптография. Протоколы, алгоритмы, исходные тексты на языке Си. М.: Триумф, 2002