

Лабораторная работа № 2

ИССЛЕДОВАНИЕ КРИПТОГРАФИЧЕСКИХ ШИФРОВ НА ОСНОВЕ ПОДСТАНОВКИ (ЗАМЕНЫ) СИМВОЛОВ

Цель: изучение и приобретение практических навыков разработки и использования приложений для реализации подстановочных шифров.

Задачи:

1. Закрепить теоретические знания по алгебраическому описанию, алгоритмам реализации операций зашифрования/расшифрования и оценке криптостойкости подстановочных шифров.
2. Ознакомиться с особенностями реализации и свойствами различных подстановочных шифров на основе готового программного средства (L_LUX).
3. Разработать приложение для реализации указанных преподавателем методов подстановочного зашифрования/расшифрования.
4. Выполнить исследование криптостойкости шифров на основе статистических данных о частотах появления символов в исходном и зашифрованном сообщениях.
5. Оценить скорость зашифрования/расшифрования реализованных способов шифров.
6. Результаты выполнения лабораторной работы оформить в виде описания разработанного приложения, методики выполнения экспериментов с использованием приложения и результатов эксперимента.

2.1. Теоретические сведения

Лабораторная работа является первой в цикле работ, относящихся к криптографическим шифрам. Основные теоретические сведения и определения из данной предметной области можно найти в [3].

В этом разделе материалов кратко рассмотрим лишь сведения, имеющие непосредственное отношение к цели и задачам лабораторной работы.

! Сущность подстановочного шифрования состоит в том, что исходный текст (из множества M) и зашифрованный текст (из множества C) основаны на использовании одного и того же или разных алфавитов, а тайной или ключевой информацией является алгоритм подстановки.

Если исходить из того, что используемые алфавиты являются конечными множествами, то в общем случае каждой букве a_x алфавита A_M ($a_x \in A_M$) для создания сообщения M_i ($M_i \in M$) соответствует буква a_y или множество букв $\{A_{xC}\}$ для создания шифртекста C_i ($C_i \in C$). Важно, чтобы во втором случае любые два множества (например, $\{A_{xC}\}_b$ и $\{A_{xC}\}_n$, $b \neq n$, $1 \leq b, n, x, y \leq N$, N – мощность алфавита), используемые для замены разных букв открытого текста, не пересекались:

$$\{A_{xC}\}_b \cap \{A_{xC}\}_n = \emptyset.$$

Если в сообщении M_i содержится несколько букв a_x , то каждая из них заменяется на символ a_y либо на любой из символов $\{A_{xC}\}$. За счет этого с помощью одного ключа можно сгенерировать различные C_i для одного и того же M_i . Так как множества $\{A_{xC}\}_b$ и $\{A_{xC}\}_n$ попарно не пересекаются, то по каждому символу C_i можно однозначно определить, какому множеству он принадлежит, и, следовательно, какую букву открытого сообщения M_i он заменяет. В силу этого открытое сообщение восстанавливается из зашифрованного однозначно.

Приведенные утверждения справедливы для следующих типов подстановочных шифров:

- *моноалфавитных* (шифры однозначной замены или простые подстановочные);
- *полиграммных*;
- *омофонических* (однозвучные шифры или шифры многозначной замены);
- *полиалфавитных*.

Кратко поясним особенности указанных шифров.

2.1.1. Моноалфавитные шифры подстановки

В данных шифрах операция замены производится отдельно над каждым одиночным символом сообщения M_i . Для наглядной демонстрации шифра простой замены достаточно выписать под заданным алфавитом тот же алфавит, но в другом порядке или,

например, со смещением. Записанный таким образом алфавит называют алфавитом замены.

Максимальное количество ключей для любого шифра этого вида не превышает $N!$, где N – количество символов в алфавите.

Для математического описания криптографического преобразования предполагаем, что зашифрованная буква a_y ($a_y \in C_i$), соответствующая символу a_x ($a_x \in M_i$), находится на позиции

$$y \equiv x + k \bmod N, \quad (2.1)$$

где x, y – индекс (порядковый номер, начиная с 0) символа в используемом алфавите; k – ключ.

Для расшифрования сообщения C_i необходимо произвести расчеты, обратные выражению (2.1), т. е.

$$x \equiv y - k \bmod N. \quad (2.2)$$

Соотношениям (2.1) и (2.2) соответствует классический шифр подстановки – **шифр Цезаря**. Согласно описаниям историка Светония в книге «Жизнь двенадцати цезарей», данный шифр использовался Гаем Юлием Цезарем для секретной переписки со своими генералами (I век до н. э.) [7] (в этой книге любознательный читатель найдет также много исторической информации по криптографии).

Пример 1. Имеем открытый текст $M_i = \langle cba \rangle$. На основе шифра Цезаря $C_i = \langle fed \rangle$.

Здесь $k = 3$, $N = 26$. Первый символ открытого текста (c) имеет индекс 2 (помним, что начальный символ алфавита (a) имеет нулевой индекс). Значит, первый символ шифртекста (c) будет иметь индекс $2 + k = 5$. А такой индекс в алфавите принадлежит символу f и т. д.

Известное послание Цезаря *VENI VIDI VICI* (в переводе на русский означает «пришел, увидел, победил»), направленное его другу Аминтию после победы над понтийским царем Фарнаком, выглядело бы в зашифрованном виде так: *YHQL YLGL YLFL*.

Применительно к русскому языку суть его состоит в следующем. Выписывается исходный алфавит (А, Б, ..., Я), затем под ним выписывается тот же алфавит, но с циклическим сдвигом на 3 позиции влево.

Существуют различные **модификации** шифра Цезаря, в частности, **Атбаш** и **лозунговый шифр**.

Атбаш. В Ветхом Завете существует несколько фрагментов из священных текстов, которые зашифрованы с помощью шифра замены, называемого Атбаш. Этот шифр состоит в замене каждой

буквы другой буквой, которая находится в алфавите на таком же расстоянии от конца алфавита, как оригинальная буква – от начала. Например, в русском алфавите буква А заменяется на Я, буква Б – на Ю и т. д. В оригинальном Ветхом Завете использовались буквы еврейского алфавита. Так, в книге пророка Иеремии слово «Бабель» (Вавилон) зашифровано как «Шешах» [7].

! Одним из существенных недостатков моноалфавитных шифров является их низкая криптостойкость. Зачастую метод криптоанализа базируется на частоте встречаемости букв исходного текста.

Если в открытом сообщении часто встречается какая-либо буква, то в зашифрованном сообщении также часто будет встречаться соответствующий ей символ. Еще в 1412 г. Шихаба ал-Калкашанди в своем труде «Субх ал-Ааша» привел таблицу частоты появления арабских букв в тексте на основе анализа текста Корана. Для разных языков мира существуют подобные таблицы. Так, например, для букв русского алфавита по данным «Национального корпуса русского языка» [8] (корпус – это информационно-справочная система, основанная на собрании текстов на некотором языке в электронной форме; национальный корпус представляет данный язык на определенном этапе (или этапах) его существования и во всем многообразии жанров, стилей, территориальных и социальных вариантов и т. п.). Частота их появления представлена в табл. 2.1.

Существуют подобные таблицы для пар букв (биграмм). Например, часто встречаемыми биграммами являются «то», «но», «ст», «по», «ен» и т. д. Другой прием взлома шифров основан на исключении возможных сочетаний букв. Например, в текстах (если они написаны без орфографических ошибок) нельзя встретить сочетания «чя», «щы», «ъь» и т. п. Таблицы с частотами (вероятностями) встречаемости пар и большего числа буквосочетаний существуют для разных алфавитов. Пример использования частотных свойств символов алфавита английского языка для криптоанализа можно найти на страницах 17–19 пособия [9].

Система шифрования Цезаря с ключевым словом (лозунгом). Также является одноалфавитной системой подстановки. Особенностью этой системы является использование ключевого слова (лозунга) для смещения и изменения порядка символов в алфавите подстановки (желательно, чтобы все буквы ключевого слова были различными). Ключевое слово пишется в начале алфавита подстановки.

Таблица 2.1

Частота появления букв русского языка в текстах

Номер п/п	Буква	Частота, %	Номер п/п	Буква	Частота, %
1	О	10,97	18	Ь	1,74
2	Е	8,45	19	Г	1,70
3	А	8,01	20	З	1,65
4	И	7,35	21	Б	1,59
5	Н	6,70	22	Ч	1,44
6	Т	6,26	23	Й	1,21
7	С	5,47	24	Х	0,97
8	Р	4,73	25	Ж	0,94
9	В	4,54	26	Ш	0,73
10	Л	4,40	27	Ю	0,64
11	К	3,49	28	Ц	0,48
12	М	3,21	29	Щ	0,36
13	Д	2,98	30	Э	0,32
14	П	2,81	31	Ф	0,26
15	У	2,62	32	Ъ	0,04
16	Я	2,01	33	Ё	0,04
17	Ы	1,90	-	-	-

Пример 2. Для шифра с использованием кодового слова «TABLE» исходный алфавит (первая строка) и алфавит подстановки (вторая строка) выглядят следующим образом:

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
T A B L E C D F G H I J K M N O P Q R S U V W X Y Z

Если $M_i = \text{«HELLO»}$, то $C_i = \text{«FEJJN»}$, если же $M_i = \text{«VENIVIDIVICI»}$, то $C_i = \text{«VEMGVGLGVGBG»}$.

Метод можно видоизменить, если ключевое слово записывать начиная не с первого символа (нулевой индекс) во второй строке, а в соответствии с некоторым числом a : $0 \leq a < N$. Рассмотрим систему на примере.

Пример 3. Выберем число $a = 10$ и слово *information* в качестве ключа.

Ключевое слово записывается под буквами алфавита, начиная с буквы, индекс которой совпадает с выбранным числом a , как это показано ниже:

0	1	2	3	4	5	10	15	20	25	
A	B	C	D	E	F	G	H	I	J	
K	L	M	N	O	P	Q	R	S	T	
U	V	W	X	Y	Z					
I N F O R M A T										

Как видим, повторяющиеся буквы (I, N и O в конце слова) во второй строке не дублируются. Оставшиеся буквы алфавита подстановки записываются после ключевого слова в алфавитном порядке:

5	10	15							
A	B	C	D	E	F	G	H	I	J
K	L	M	N	O	P	Q	R	S	T
U	V	W	X	Y	Z	I	N	F	O
R	Q	S	U	V	W	X	Y	Z	A
B	C	D	E	G	H	J	K		

Если $M_i = \langle \text{VENIVIDIVI} \rangle$, то $C_i = \langle \text{EUOYEYSYEQY} \rangle$.

Расшифрование сообщения производится по правилу, которое мы рассматривали на примерах, проанализированных выше.

Применяя одновременно операции сложения и умножения по модулю n над элементами множества (индексами букв алфавита), можно получить *систему подстановок*, которую называют **аффинной системой подстановок Цезаря**. Определим процедуру зашифрования в такой системе:

$$y \equiv ax + b \pmod{N}, \quad (2.3)$$

где a и b – целые числа.

При этом взаимно однозначные соответствия между открытым текстом и шифртекстом будут иметь место только при выполнении следующих условий: $0 \leq a, b < N$, наибольший общий делитель (НОД) чисел a, N равен 1, т. е. эти числа являются *взаимно простыми*.

Пример 4. Пусть $N = 26$, $a = 3$, $b = 5$. Тогда $\text{НОД}(3, 26) = 1$, и мы получаем следующее соответствие между индексами букв:

x	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
$3x+5$	5	8	11	14	17	20	23	0	3	6	9	12	15	18	21	24	1	4	7	10	13	16	19	22	25	2

Преобразуя числа в буквы английского алфавита, получаем следующее соответствие для букв открытого текста и шифртекста:

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
F I L O R U X A D G J M P S V Y B E H K N Q T W Z C

Для $M_i = \text{«VENIVIDIVI»}$ получение зашифрованного сообщения в деталях показывает табл. 2.2.

Таблица 2.2

Иллюстрация получения шифртекста на основе аффинной системы подстановок Цезаря

M_i	V	E	N	I	V	I	D	I	V	I	C	I
x	21	4	13	8	21	8	3	8	21	8	2	8
$y = 3x + 5$	16	17	18	3	16	3	14	3	16	3	11	3
C_i	Q	R	S	C	Q	C	O	C	Q	C	L	C

Таким образом, зашифрованное сообщение будет таким: $C_i = \text{«QRSCQCOCQCLC»}$.

Расшифрование основано на использовании соотношения

$$x \equiv a^{-1}(y + N - b) \pmod{N}, \quad (2.4)$$

где a^{-1} – обратное к a число по модулю N , т. е. оно удовлетворяет уравнению $aa^{-1} \equiv 1 \pmod{N}$.

2.1.2. Полиграммные шифры

В таких шифрах одна подстановка соответствует сразу нескольким символам исходного текста.

Первым известным шифром этого типа является **шифр Порты** [10]. Шифр представляется в виде таблицы. Наверху горизонтально и слева вертикально записывается стандартный алфавит. В ячейках таблицы записываются числа в определенном порядке. Одним

из возможных вариантов такой таблицы для алфавита русского языка будет показанный ниже фрагмент таблицы (табл. 2.3).

Таблица 2.3

**Фрагмент шифра Порты для алфавита русского языка,
состоящего из 33 букв**

	А	Б	В	...	Э	Ю	Я
А	001	002	003	...	031	032	033
Б	034	035	036	...	064	065	066
В	067	068	069	...	097	098	099
Г	100	101	102	...	130	131	132
...
Ю	1024	1025	1026	...	1054	1055	1056
Я	1057	1058	1059	...	1087	1088	1089

Шифрование выполняется парами букв исходного сообщения. Первая буква пары указывает на строку, вторая – на столбец. В случае нечетного количества букв в сообщении M_i к нему добавляется вспомогательный символ, например «А».

Пример 5. Исходное сообщение $M_i = \text{«АВВА»}$. Сообщение состоит из двух пар (биграмм): АВ и ВА – и будет зашифровано так: 003 067.

Другими известными полиграммными шифрами являются **шифр Плейфера** и **шифр Хилла** [10].

С точки зрения криптостойкости рассматриваемый тип шифров имеет преимущества перед моноалфавитными шифрами. Это связано с тем, что, во-первых, распределение частот групп букв значительно более равномерное, чем отдельных символов. Во-вторых, для эффективного частотного анализа требуется больший размер зашифрованного текста, так как число различных групп букв значительно больше, чем мощность алфавита.

2.1.3. Омофонические шифры

Омофонические шифры (омофоническая замена), или **однозвучные шифры подстановки,** создавались с целью увеличить сложность частотного анализа шифртекстов путем маскировки реальных частот появления символов текста с помощью **омофонии.**

В 1401 г. Симеоне де Крема стал использовать таблицы омофонов для сокрытия частоты появления гласных букв в тексте при помощи более чем одной подстановки. Такие шифры позже стали называться **шифрами многозначной замены**, или **омофонами** (от греч. *homos* – одинаковый и *phone* – звук; слова, которые звучат одинаково, но пишутся по-разному и имеют разное значение; очень много подобных слов содержит английский язык). Они получили развитие в XV в. В книге «Трактат о шифрах» Леона Баттисты Альберти (итальянский ученый, архитектор, теоретик искусства, секретарь папы Климентия XII), опубликованной в 1466 г. [11], приводится описание шифра замены, в котором каждой букве ставится в соответствие несколько эквивалентов, число которых пропорционально частоте встречаемости буквы в открытом тексте M_i . В этих шифрах буквы исходного алфавита соответствуют более чем одному символу из алфавита замены. Обычно символам исходного текста с наивысшей частотой дают большее количество эквивалентов, чем более редким символам. Таким образом, распределение частоты становится более равномерным, сильно затрудняя частотный анализ.

В табл. 2.4 представлен фрагмент таблицы подстановок для алфавита русского языка [12].

Таблица 2.4

Фрагмент таблицы подстановок для системы омофонов

Номер п/п	А	Б	В	...	М	...	О	...	Р	...	Я
1	311	128	175	...	037	...	248	...	064	...	266
2	357	950	194	...	149	...	267	...	189	...	333
...
16	495	990	199	...	349	...	303	...	374	...	749
...	...	–
20	519	–	427	...	760	...	306	...	469	...	845
...	...	–	–	–
32	637	–	524	...	777	...	432	...	554	–	–
...	...	–	...	–	–	–	–	–
45	678	–	644	–	–	–	824	...	721	–	–
...	...	–	–	–	–	–	–	–

Окончание табл. 2.4

Номер п/п	А	Б	В	...	М	...	О	...	Р	...	Я
47	776	–	–	–	–	–	828	...	954	–	–
...	...	–	–	–	–	–	...	–	–	–	–
80	901	–	–	–	–	–	886	–	–	–	–
...	–	–	–	–	–	–	...	–	–	–	–
110	–	–	–	–	–	–	903	–	–	–	–

При шифровании символ исходного сообщения заменяется на любую подстановку из «своего» столбца. Если символ встречается повторно, то, как правило, используют разные подстановки. Например, исходное сообщение «АБРАМОВ» после зашифрования может выглядеть так: «357 990 374 678 037 828 175» [12].

Заметным вкладом греческого ученого **Энея Тактика** в криптографию является предложенный им так называемый **книжный шифр** [11, 12]. После Первой мировой войны книжный шифр приобрел иной вид. Шифрозамена для каждой буквы определялась набором цифр, которые указывали на номер страницы, строки и позиции в строке (вспомните пример использования такого шифра известными героями фильма «Семнадцать мгновений весны»). Даже с формальной стороны отсутствие полной электронной базы изданных к настоящему времени книг делает процедуру взлома шифра практически невыполнимой.

2.1.4. Полиалфавитные шифры

Полиалфавитные (или **многоалфавитные**) шифры состоят из нескольких шифров однозначной замены. Выбор варианта алфавита для зашифрования одного символа зависит от особенностей метода шифрования.

Диск Альберти. В «Трактате о шифрах» [11] Альберти приводит первое точное описание **многоалфавитного шифра** на основе **шифровального диска** (см. рис. 2.1).

Он состоял из двух дисков – внешнего неподвижного и внутреннего подвижного, на которые были нанесены буквы алфавита. Процесс шифрования заключался в нахождении буквы открытого текста на внешнем диске и замене ее на букву с внутреннего диска, стоящую под ней. После этого внутренний диск сдвигался на одну

позицию, и шифрование второй буквы производилось уже по-новому шифралфавиту. Ключом данного шифра являлся порядок расположения букв на дисках и начальное положение внутреннего диска относительно внешнего.




Рис. 2.1. Реплика диска Альберти, используемого Конфедерацией во время Гражданской войны в Америке [13]

Таблица Трисемуса. В 1518 г. в развитии криптографии был сделан важный шаг благодаря появлению в Германии первой печатной книги по криптографии. Аббат Иоганнес Трисемус, настоятель монастыря в Вюрцбурге, написал книгу «Полиграфия», в которой он описал ряда шифров, один из которых развивает идею *многоалфавитной подстановки*. Зашифрование осуществляется так: заготавливается **таблица подстановки** (так называемая «**таблица Трисемуса**» – таблица со стороной, равной N , где N – мощность алфавита), где первая строка – это алфавит, вторая – алфавит, сдвинутый на один символ, и т. д. При зашифровании первая буква открытого текста заменяется на букву, стоящую в первой строке, вторая – на букву, стоящую во второй строке, и т. д. После использования последней строки вновь возвращаются к первой.

Пример 6. Рассмотрим процесс зашифрования сообщения $M_i = \langle \text{БГТУ} \rangle$, используя таблицу, фрагмент которой показан на рис. 2.2.

Стрелками на приведенном рисунке показан принцип зашифрования каждого символа открытого текста. Из этого следует, что шифртекст имеет вид: $C_i = \langle \text{БДФЦ} \rangle$.



А	Б	В	Г	Д	Е	Ё	Ж	З	И	Й	К	Л	М	Н	О	П	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ
Б	В	Г	Д	Е	Ё	Ж	З	И	Й	К	Л	М	Н	О	П	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы
В	Г	Д	Е	Ё	Ж	З	И	Й	К	Л	М	Н	О	П	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь
Г	Д	Е	Ё	Ж	З	И	Й	К	Л	М	Н	О	П	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э
Д	Е	Ё	Ж	З	И	Й	К	Л	М	Н	О	П	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю
Е	Ё	Ж	З	И	Й	К	Л	М	Н	О	П	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я

Рис. 2.2. Фрагмент таблицы Трисемуса для алфавита русского языка

В указанной книге Трисемус впервые систематически описал применение шифрующих таблиц, заполненных алфавитом в случайном порядке. Для получения такого шифра подстановки обычно использовались таблица для записи букв алфавита и *ключевое слово* (или фраза). Можно найти определенную аналогию с системой шифрования Цезаря с ключевым словом. В таблицу сначала вписывалось по стрелкам ключевое слово, причем повторяющиеся буквы также отбрасывались. Затем эта таблица дополнялась не вошедшими в нее буквами алфавита по порядку.

Таким образом, ключом в **таблицах Трисемуса** является **ключевое слово и размер таблицы**. При шифровании буква открытого текста заменяется буквой, расположенной ниже нее в том же столбце. Если буква текста оказывается в нижней строке таблицы, тогда для шифртекста берут самую верхнюю букву из того же столбца.

Указанный размер таблицы для алфавита русского языка может соответствовать 4×8 либо 8×4.

Пример 7. Пусть M_i = «ПРИШЕЛУВИДЕЛПОБЕДИЛ», а ключевое слово – «ЦЕЗАРЬ». Используем таблицу 8×4 (табл. 2.5).

Таблица 2.5

Пример заполнения таблицы с ключевым словом «ЦЕЗАРЬ»

Ц	Е (Ё)	З	А
Р	Ь	Б	В
Г	Д	Ж	И
Й	К	Л	М
Н	О	П	С
Т	У	Ф	Х
Ч	Ш	Щ	Ъ
Ы	Э	Ю	Я

Следуя вышеуказанного принципу подстановки, получим $C_i = \langle \Phi Г М Э Ъ П Ш И М К Ъ П Ф У Ж Ъ К М П \rangle$.

Шифр Виженера. В 1586 г. французский дипломат Блез Виженер представил перед комиссией Генриха III описание простого, но довольно стойкого шифра, в основе которого лежит таблица Трисемуса.

В этом шифре мы имеем дело с последовательностью сдвигов, циклически повторяющейся. Основная идея заключается в следующем. Создается таблица (таблица Виженера) размером $N \times N$ (N – число знаков в используемом алфавите). Эти знаки могут включать не только буквы, но и, например, пробел или иные знаки. В первой строке таблицы записывается весь используемый алфавит. Каждая последующая строка получается из предыдущего циклического сдвига последней на 1 символ влево. Таким образом, при мощности алфавита (английского языка), равной 26, необходимо выполнить последовательно 25 сдвигов для формирования всей таблицы. Более подробное описание шифра можно найти в [3, с. 41–43].

Листинг содержит часть кода, реализующего алгоритм шифрования Виженера.

```

: /*
:  *Главный цикл, который проходит по входной стороне
: */
: foreach (char symbol in input)
: {
:     /* characters – алфавит
:        keyword – ключевое слово
:        N – мощность алфавита
:        keyword_index – индекс текущей буквы ключевого слова
:        */
:     int c = (Array.IndexOf(characters, symbol) +
:     Array.IndexOf(characters, keyword[keyword_index])) % N;
:     /* result – результирующая строка */
:     result += characters[c];
:     keyword_index++;
:     /* циклический проход по ключевому слову */
:     if ((keyword_index + 1) == keyword.Length)
:         keyword_index=0;
: }

```

Листинг. Фрагмент кода, реализующего алгоритм шифра Виженера

Следует добавить, что в 1863 г. Фридрих Касиски опубликовал алгоритм атаки на этот шифр, хотя известны случаи взлома шифра некоторыми опытными криптоаналитиками и ранее. В частности, в 1854 г. шифр был взломан изобретателем первой аналитической вычислительной машины Чарльзом Бэббиджем. Этот факт стал известен только в XX в., когда группа ученых разбирала вычисления и личные заметки Бэббиджа [6]. Несмотря на это, шифр Виженера имел репутацию исключительно стойкого к «ручному» взлому еще долгое время. Так, известный писатель и математик Чарльз Доджсон (Льюис Кэрролл) в своей статье «Алфавитный шифр», опубликованной в детском журнале в 1868 г., назвал шифр Виженера невзламываемым. В 1917 г. научно-популярный журнал «Scientific American» также отозвался о шифре Виженера как о неподдающемся взлому [14].

Роторные машины. Идеи Альберти и Виженера использовались при создании электромеханических роторных машин первой половины XX в. Некоторые из них применялись в разных странах вплоть до 1980-х гг. В большинстве из них использовались роторы (механические колеса), взаимное расположение которых **определяло текущий алфавит шифрозамен**, используемый для выполнения подстановки. Наиболее известной из роторных машин является немецкая машина времен Второй мировой войны «**Энигма**». Более детальному изучению и практическому анализу «Энигмы» далее будет посвящена отдельная лабораторная работа.

К полиалфавитным относится также шифр на основе «одноразового блокнота».

Много полезной информации по рассмотренному классу шифров можно найти в [15].



Существует определенное сходство между подстановочными шифрами и шифрами на основе гаммирования. Последние рассматриваются как самостоятельный класс. Такие шифры схожи с подстановочными (и в определенном плане – с перестановочными) тем, что в обоих случаях можно использовать табличное представление выполняемых операций на основе ключей. В шифрах на основе гаммирования и в подстановочных шифрах при зашифровании происходит подмена одних символов другими.

Гаммирование будем рассматривать и изучать более подробно в лабораторной работе № 6.

2.2. Общие сведения о криптоанализе

Данная лабораторная работа посвящена анализу одного из разделов практической криптографии. В связи с этим здесь будет уместно охарактеризовать **противоположность криптографии – криптоанализ**. Данный термин введен американским криптографом Уильямом Ф. Фридманом в 1920 г.

! Еще раз вспомним, что криптоанализ – это раздел криптологии, занимающийся методами взлома шифров или методами организации криптографических атак на шифры.

Кратко проанализируем основные криптоатаки [3, 5, 12, 15].

Атака с известным шифртекстом (ciphertext only attack).

Предполагается, что противник знает алгоритм шифрования, у него имеется набор перехваченных шифрограмм, но он **не знает** секретный ключ.

Разновидности такой атаки:

- **полный перебор ключей** (лобовая атака, bruteforce attack);
- **атака по словарю**, перебор ключей по словарю (dictionary attack); применяется часто для взлома паролей;

- **частотный криптоанализ** – метод взлома шифра, основывающийся на предположении о существовании зависимости между частотой появления символов алфавита в открытых сообщениях и соответствующих шифрозамен в шифрограммах (этот вопрос с практической точки зрения мы анализировали при выполнении лабораторной работы № 2 из [2]).

Атака с выбором шифртекста (chosen cipher text attack).

Криптоаналитик имеет возможность выбрать необходимое количество шифрограмм и получить соответствующие им открытые тексты. Он также может воспользоваться устройством расшифрования один или несколько раз для получения шифртекста в расшифрованном виде. Используя полученные данные, он может попытаться восстановить секретный ключ.

Адаптивная атака с выбором шифртекста (adaptive chosen ciphertext attack). Криптоаналитик имеет возможность выбирать новые шифрограммы для расшифрования с учетом того, что ему известна некоторая информация из предыдущих сообщений. В некоторых криптографических протоколах при получении шифрограммы, несоответствующей стандарту (содержащей ошибки), отправитель получает ответное сообщение, иногда с детализированным

описанием этапа проверки и причины возникновения ошибки. Криптоаналитик может использовать эту информацию для последовательной посылки и уточнения параметров криптосистемы.

Атака с известным открытым текстом (known plaintext attack). То же, что и предыдущая, но противник для некоторых шифрограмм получает в свое распоряжение соответствующие им открытые тексты.

Атака с выбором открытого текста (chosen plaintext attack). Криптоаналитик обладает некоторыми открытыми текстами и соответствующими шифртекстами. Кроме того, он имеет возможность зашифровать несколько предварительно выбранных открытых текстов (до начала атаки).

Разновидности:

- **атака на основе получения временного неконфиденциального доступа к шифрующему устройству** для получения пар открытых и тайных текстов (известны случаи реализации таких атак спецслужбами);

- **атака на основе использования информации о структуре сообщений или стандартных фразах** – криптоаналитики из Блетчли-Парка (Bletchley Park) могли определить открытый текст сообщений «Энигмы» (см. далее лабораторную работу № 4) в зависимости от того, когда эти сообщения были посланы и как они подписывались;

- **перебор ключей по словарю** (dictionary attack) – криптоаналитик шифрует слова и фразы, наличие которых предполагается в шифрограмме, с использованием различных ключей; совпадение зашифрованных слов и фраз с частями шифрограммы может говорить о взломе ключа.

Адаптивная атака с выбором открытого текста (adaptive chosen plaintext attack). Криптоаналитик имеет возможность выбирать новые открытые тексты с учетом того, что ему известна некоторая информация из предыдущих сообщений – он может получить пары «открытое сообщение – шифрограмма», в том числе и после начала атаки.

Разновидности:

- **провоцирование противника на использование в сообщениях определенных слов или фраз**; придуман англичанами: Королевские военно-воздушные силы Великобритании минировали определенные участки Северного моря, этот процесс был назван «садоводством» (англ. *gardening*); практически сразу после этого немцами

посылались зашифрованные сообщения, включающие слово «мины» и названия мест, где они были сброшены;

- **дифференциальный криптоанализ** – метод вскрытия симметричных блочных шифров (и других криптографических примитивов, в частности хеш-функций), предложен в 1990 г. израильскими специалистами Эли Бихамом и Ади Шамиром и основан на изучении разностей между шифруемыми значениями на различных раундах для пары подобранных открытых сообщений при их зашифровании одним и тем же ключом;

- **интегральный криптоанализ** – аналогичен дифференциальному криптоанализу, но в отличие от него рассматривает воздействие алгоритма не на пару, а сразу на множество открытых текстов; предложен в 1997 г. Ларсом Кнудсенем;

- **линейный криптоанализ** – предложен японским криптологом Мицуру Мацуи в 1993 г.; основан на использовании некоторых **линейных приближений**, которые означают, например, следующее: если выполняется операция XOR над некоторыми битами открытого текста, затем – над некоторыми битами шифртекста, а затем над результатами, то получается бит (или биты), который представляет собой XOR некоторых битов ключа; это и есть линейное приближение, которое может быть верным с некоторой вероятностью;

- **использование открытых ключей в асимметричных системах** – криптоаналитик имеет возможность получить шифртекст, соответствующий выбранному сообщению, на основе открытого ключа.

Атака на основе связанных ключей (related key attack). Криптоаналитик знает не сами ключи, а некоторые различия (соотношения) между ними; реальные криптосистемы используют разные ключи, связанные известным соотношением, например, для каждого нового сообщения предыдущее значение ключа увеличивается на единицу или преобразуется на основе операции сдвига.

Атака с выбором ключа (chosen key attack). Криптоаналитик задает часть ключа, а на оставшуюся часть ключа выполняет атаку на основе связанных ключей.

2.3. Практическое задание

Рекомендация! Перед выполнением практического задания можно познакомиться с особенностями работы программного

средства *L_LUX*, реализующего подстановочные (и другие) методы зашифрования/расшифрования текстовой информации и являющегося приложением на компакт-диске к [6].

Программа понятна и проста в использовании с точки зрения интерфейса и функционала. Основная часть окна – текстовый редактор, в котором можно набирать текст либо размещать скопированный фрагмент из другого текстового документа (вставка – Ctrl + V). Здесь же отображается зашифрованный текст, а также сформированные программой распределения частот (в виде гистограмм) для исходного и зашифрованного текстов.

На рис. 2.3 представлено основное диалоговое окно программы после запуска исполнительного файла *L_LUX.EXE*.

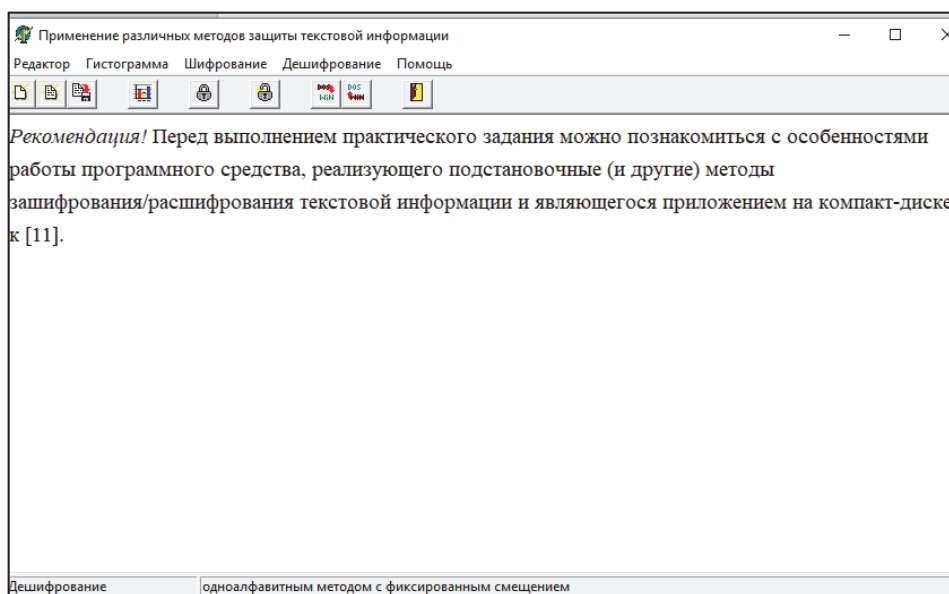


Рис. 2.3. Основное диалоговое окно программного средства *L_LUX*

На рис. 2.4 для примера и сравнения приведены гистограммы для исходного (в окне редактора на рис. 2.3) и зашифрованного документов (обратим внимание, что отдельные буквы – строчные и прописные – рассматриваются здесь как разные, что не соответствует традиционному подходу).

Основное задание

1. Разработать авторское приложение в соответствии с целью лабораторной работы. Приложение должно реализовывать следующие операции:

- выполнять зашифрование/расшифрование текстовых документов (объемом не менее 5 тысяч знаков), созданных на основе алфавита

языка в соответствии с нижеследующей таблицей вариантов задания; при этом следует использовать шифры подстановки из третьего столбца данной таблицы (варианты задания представлены в табл. 2.6);

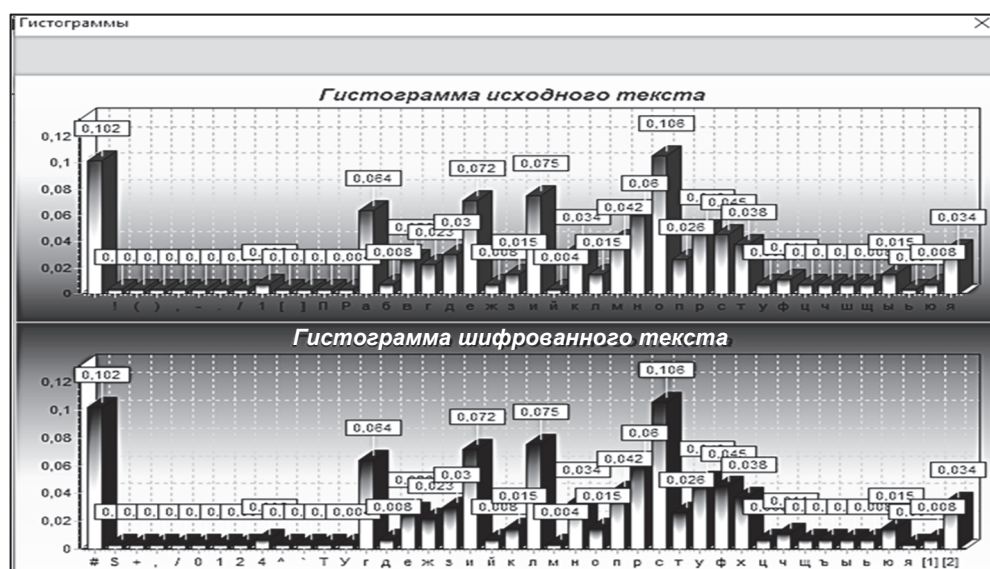


Рис. 2.4. Гистограммы для исходного и зашифрованного текстовых документов

Таблица 2.6

Варианты задания

Вариант	Алфавит	Шифр
1	Белорусский	1. На основе соотношений (2.1) и (2.2); $k = 5$ 2. Виженера, ключевое слово – собственная фамилия
2	Русский	1. На основе аффинной системы подстановок Цезаря; $a = 7, b = 10$ 2. Виженера, ключевое слово – собственная фамилия
3	Английский	1. Шифр Цезаря с ключевым словом, ключевое слово – собственная фамилия 2. Таблица Трисемуса, ключевое слово – собственное имя
4	Немецкий	1. На основе соотношений (2.1) и (2.2); $k = 7$ 2. Таблица Трисемуса, ключевое слово – enigma
5	Польский	1. На основе аффинной системы подстановок Цезаря; $a = 5, b = 7$ 2. Шифр Порты
6	Белорусский	1. На основе соотношений (2.1) и (2.2); $k = 21$ 2. Таблица Трисемуса, ключевое слово – собственное имя

Окончание табл. 2.6

Вариант	Алфавит	Шифр
7	Русский	1. Шифр Порты 2. Шифр Цезаря с ключевым словом, ключевое слово – собственная фамилия
8	Английский	1. Шифр Цезаря с ключевым словом, ключевое слово – собственная фамилия, $a = 24$ 2. Таблица Трисемуса, ключевое слово – собственное имя
9	Немецкий	1. На основе соотношений (2.1) и (2.2); $k = 7$ 2. Таблица Трисемуса, ключевое слово – enigma
10	Польский	1. На основе соотношений (2.1) и (2.2); $k = 28$ 2. Шифр Порты
11	Белорусский	1. Шифр Цезаря с ключевым словом, ключевое слово – інфарматыка, $a = 2$ 2. Таблица Трисемуса, ключевое слово – собственное имя
12	Русский	1. Шифр Цезаря с ключевым словом, ключевое слово – безопасность 2. Таблица Трисемуса, ключевое слово – безопасность
13	Английский	1. На основе аффинной системы подстановок Цезаря; $a = 6, b = 7$ 2. Таблица Трисемуса, ключевое слово – security
14	Немецкий	1. Виженера, ключевое слово – собственная фамилия 2. Шифр Порты
15	Польский	1. Виженера, ключевое слово – bezpieczeństwo 2. На основе соотношений (2.1) и (2.2); $k = 20$

- сформировать гистограммы частот появления символов для исходного и зашифрованного сообщений;

- оценить время выполнения операций зашифрования/расшифрования (напоминание: во многих языках программирования есть встроенные методы для замеров времени; при отсутствии такового в используемом языке можно воспользоваться разностью двух дат (например, в миллисекундах: время после выполнения программы – время до начала выполнения преобразования)).

При анализе полученных гистограмм можно сопоставить полученные данные с аналогичными результатами выполнения лабораторной работы № 2 из [2].

Если указанный в таблице язык исходного текста не известен разработчику программного средства, можно взять документ на требуемом языке и воспользоваться доступным электронным переводчиком (возникающие при этом отдельные семантические неточности не следует считать существенным недостатком выполняемого анализа).

2. Результаты оформить в виде отчета по установленным правилам.

ВОПРОСЫ ДЛЯ КОНТРОЛЯ И САМОКОНТРОЛЯ

1. В чем заключается основная идея криптографических преобразований на основе шифров замены?

2. Привести классификационные признаки и дать сравнительную характеристику разновидностям подстановочных шифров.

3. Сколько разновидностей шифров, подобных шифру Цезаря, можно составить для алфавитов русского и белорусского языков?

4. Найти ключ шифра, с помощью которого получен шифртекст: «*byajhvfwbjyyfzgjcktljdfntkmyjcnm*».

5. Расшифровать (с демонстрацией каждого шага алгоритма) текст $C_i = \text{«}qrscqcocqclc\text{»}$, зашифрованный аффинным шифром Цезаря при $N = 26$, $a = 3$, $b = 5$.

6. Зашифровать и расшифровать свою фамилию (на основе кириллицы), используя аффинный шифр Цезаря.

7. Можно ли использовать в качестве ключевого в шифре Виженера слово, равное по длине открытому тексту? Обосновать ответ.

8. По какому признаку можно определить, что текст зашифрован шифром Плейфера?

9. Имеются ли предпочтения в выборе размеров таблицы Трисемуса для виртуального алфавита мощностью 40: 4×10 ; 10×4 ; 5×8 ; 8×5 ; 2×20 ; 20×2 ?

10. Охарактеризовать основные виды атак на шифры.

11. Сравнить криптостойкость шифра Цезаря и шифра Виженера.

12. Охарактеризовать основные методы взлома подстановочных шифров.