ЛАБОРАТОРНАЯ РАБОТА №1.
КРИПТОАНАЛИЗ МЕТОДОВ ПРОСТОЙ ПОДСТАНОВКИ

Простейшие шифры подстановки реализуют замену каждого символа исходного текста на один из символов алфавита шифротекста. В общем случае, подстановочный шифр описывается таблицей подстановки, состоящей из двух строк и *n* столбцов. Количество столбцов таблицы подстановки соответствует количеству различных символов в алфавите исходного текста. Верхняя строка таблицы подстановки содержит все возможные символы исходного текста, а нижняя – соответствующие им символы шифротекста.

Моноалфавитные шифры характеризуются однозначным соответствием символов исходного текста и символов шифротекста. В случае, когда алфавиты исходного текста и шифротекста состоят из одного и того же множества символов, алфавит шифротекста представляет собой простую перестановку лексикографического порядка символов в алфавите исходного текста. При выполнении шифрования каждый символ исходного текста заменяется соответствующим ему символом шифротекста.

Таблица подстановки, описывающая моноалфавитный шифр, преобразующий строчные буквы русского алфавита, будет состоять из 33 столбцов, что соответствует количеству букв в алфавите. Рассмотрим такую таблицу на следующем примере:

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**	**ы**	**ь**	**э**	**ю**	**я**
е	л	ц	н	й	и	в	а	ш	у	ь	т	с	я	ф	ы	э	ж	о	р	к	ч	м	ъ	х	п	м	б	г	ё	з	щ	д

В соответствии с приведённой таблицей шифрование строки «знание – сила» будет выполнено следующим образом:

«з» → «ш»;

«н» → «ф»;

…

«а» → «е».

В результате шифрования будет получен шифротекст «шфефуи – оусе».

Таблица подстановки, описывающая ключ моноалфавитного шифра, преобразующего ASCII-коды (однобайтовые значения), будет состоять из 256 столбцов. Таблица подстановки для шифра, осуществляющего подстановку 32-битных значений, будет состоять из 2^{32} столбцов, что уже вызовет сложности её хранения и передачи. По этой причине на практике вместо таблиц подстановок используются функции подстановки, аналитически описывающие соответствие между порядковыми номерами символов исходного текста в алфавите исходного текста и порядковыми номерами символов шифротекста в алфавите шифротекста.

Предположим, алфавит исходного текста ***M*** состоит из ***n*** символов ***M*={*a*$_0$, *a*$_1$, ..., *a*$_n$}**, тогда алфавит шифротекста ***C*** будет представлять собой ***n*-символьный** алфавит ***C*={*f*(*a*$_0$), *f*(*a*$_1$), ..., *f*(*a*$_n$)}**, где функция ***f*,** выполняющаяотображение $M \rightarrow C$, и будет являться функцией подстановки. В общем случае функция подстановки любого моноалфавитного шифра может быть задана в виде полинома степени *t*:

$$E_k(a) = (k_0 + k_1 \cdot a + k_2 \cdot a^2 + ... + k_{t-1} \cdot a^{t-1} + k_t \cdot a^t) \bmod n.$$

Классическим примером простейшего моноалфавитного шифра является ***метод Цезаря***. Свое название этот шифр получил по имени римского императора Гая Юлия Цезаря, который использовал этот шифр при переписке с Цицероном (около 50 лет до нашей эры).

Согласно методу Цезаря, при шифровании исходного текста каждая буква заменяется другой буквой того же алфавита путем смещения по алфавиту от исходной буквы на ***k*** букв. При достижении последнего символа алфавита осуществляется циклический переход к его началу. Согласно легенде, Цезарь использовал подобный шифр для ***k*=3**.

Математически метод Цезаря описывается выражением:

$$f(a)=(a+k) \bmod n,$$

где ***n*** - размерность алфавита (количество символов), ***k***-криптографический ключ, а величина ***a*** обозначает символ исходного текста его порядковым номером в алфавите ***M*.**

Для английского алфавита, используя последовательную нумерацию букв 0-A, 1-B, 2-C, 3-D, 4-E, 5-F, 6-G, 7-H, 8-I, 9-J, 10-K, 11-L, 12-M, 13-N, 14-O, 15-P, 16-Q, 17-R, 18-S, 19-T, 20-U, 21-V, 22-W, 23-X, 24-Y, 25-Z, процедура шифрования, предложенная Цезарем, будет описываться соотношением

$$f(a)=(a+3) \bmod 26.$$

Следует отметить, что как **a**, так и ***f*(*a*)** представляют собой номера букв в исходном алфавите. При шифровании буквы **G** исходного алфавита, имеющей номер **6**, получим ***f*(*a*)=6+3=9**, что соответствует букве **J**, используемой в качестве подстановочного элемента в шифротексте.

Поскольку для моноалфавитных шифров каждый символ исходного текста при шифровании может заменяться одним единственным символом шифротекста, для криптоанализа данных шифров возможно применение анализа частот встречаемости символов в шифротексте. Данная атака основана на том факте, что в естественных языках частоты встречаемости различных букв могут существенно отличаться. Так, в английском языке наиболее часто встречаемой является буква «e», а наименее встречаемой – буква «z». Зная типичную частоту встречаемости каждого из символов алфавита исходного текста, можно воссоздать использованную при шифровании таблицу подстановки, ставя каждому из символов исходного текста в соответствие символ шифротекста, частота встречаемости которого в зашифрованном тексте является наиболее близкой к типичной частоте встречаемости символа алфавита исходного текста.

Подстановочные шифры, называемые ***полиалфавитными***, используют более чем одну таблицу подстановки. Использование нескольких таблиц (алфавитов) обеспечивает возможность нескольких вариантов подстановки символов исходного сообщения, что повышает криптостойкость шифра.

Классическим полиалфавитным шифром является шифр ***Виженера*** (Vigenere Cipher). Так же как и в случае шифра Цезаря, данный шифр может быть задан таблицами подстановки, состоящими из ***n*** столбцов, где ***n*** – размер алфавита исходного текста. Количество таблиц подстановки для случая шифра Виженера будет равняться ***m***, где ***m*** – длина ключевого слова **(период ключа)**. Ключевое слово задаёт количество символов, на которое смещены относительно исходного алфавиты шифротекста в каждой из ***m*** таблиц подстановок. При шифровании исходного сообщения его записывают в строку, а под ним ключевое слово либо фразу. Если ключ оказался короче исходного текста, то его циклически повторяют необходимое число раз. На каждом шаге шифрования в верхней строке таблицы Виженера находят очередную букву исходного текста, а в левом столбце - очередное значение символа ключа. В результате очередная буква шифротекста находится на пересечении столбца, определенного символом исходного текста и строки, соответствующей строке символа ключа. Рассмотрим работу шифра Виженера на простом примере. Пусть дано ключевое слово «**MOUSE**». Тогда правила подстановки будут задаваться 5-ю таблицами, которые для компактности можно объединить в одну.

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

Пусть необходимо зашифровать текст «**CRYPTOGRAPHY AND DATA SECURITY**». Сперва запишем символы ключевого слова под символами исходного текста. Поскольку ключ в методе Виженера – периодический, повторим ключевое слово столько раз, сколько нам потребуется чтобы закрыть весь исходный текст.

C R Y P T O G R A P H Y A N D D A T A S E C U R I T Y
M O U S E M O U S E M O U S E M O U S E M O U S E M O

Каждый из символов ключа указывает, которую из таблиц подстановки нам необходимо использовать для подстановки рассматриваемого символа исходного текста. Символ ключа «M» указывает, что соответствующий символ шифротекста необходимо выбирать из таблицы подстановки, алфавит шифротекста в которой сдвинут относительно алфавита исходного текста на Pos(«M») – Pos(«A») = 12 позиций. Таким образом, первому символу исходного текста будет соответствовать символ шифротекста «O». Проведя аналогичную процедуру для всех символов исходного текста, мы получим искомый шифротекст:

«**OFSHXAULSTTM SRP XSXM MWGGFCLC**».

Как видно из примера, в результате шифрования по методу Виженера символы исходного текста «**P**» и «**H**» были преобразованы в один и тот же подстановочный элемент «**T**». Таким образом, отсутствует однозначное соответствие между символами исходного и зашифрованного текстов, что делает применение атаки на шифротекст путём частотного анализа «в лоб» невозможным.

Поскольку ключ шифратора Виженера является периодическим, зашифрованный текст можно представить как ***m*** текстов, зашифрованных по методу Цезаря. В рассмотренном примере для текста «CRYPTOGRAPHY AND DATA SECURITY» символы с позициями 1, 6, ..., 26 шифровались по методу Цезаря с ключом $k = 12$; символы с позициями 2, 7, ..., 27 – ключом $k = 14$; с позициями 3, 8, ..., 28 – ключом $k = 20$; с позициями 4, 9, ..., 29 – ключом $k = 18$; с позициями 5, 10, ..., 30 – ключом $k = 4$.

CRYPT**O**GRAP**H**Y AN**D** DAT**A** SEC**U**RITY
k = 12 («M»):C O H D A U
k = 14 («O»): R G Y - - R
k = 20 («U»): Y R - D S I
k = 18 («S»): P A A A E T
k = 4 («E»): T P N T C Y

Таким образом, зная длину ключевого слова шифра Виженера (период ключа), можно произвести взлом шифротекста, выполнив анализ частот встречаемости символов в отдельности для каждого из m компонентов шифротекста.

Различают три возможных варианта использования криптографического ключа: прямое использование (***Straight Keyword***); прогрессивный ключ (***Progressive Key***); самогенерирующийся ключ (***Auto Key***). Рассмотрим использование всех трех вариантов для исходного сообщения "*Wish you were here*".

Пример. Для шифрования сообщения "Wish you were here", используем ключ SIAMESE. Тогда для первого случая прямого использования ключа получим:

M=	W	I	S	H	Y	O	U	W	E	R	E	H	E	R	E
K=	S	I	A	M	E	S	E	S	I	A	M	E	S	E	S
C=	O	Q	S	T	C	G	Y	O	M	R	Q	L	W	V	W

Идея использования прогрессивного ключа заключается в циклическом сдвиге символов ключа на одну позицию в упорядоченном алфавите символов при повторном применении ключа. Тогда для ключа SIAMESE при повторном его использовании по прогрессивной схеме имеем TJBNFTF, а при третьемUKCOGUG, и так далее.

Пример . Для шифрования сообщения "Wish you were here", используем ключ SIAMESE и прогрессивную схему его применения получим.

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

В случае самогенерирующегося ключа в качестве его последующих символов используется исходный текст.

Пример . Для шифрования сообщения "Wish you were here", используем самогенерирующийся ключ SIAMESE. В результате имеем.

M=	W	I	S	H	Y	O	U	W	E	R	E	H	E	R	E
K=	S	I	A	M	E	S	E	W	I	S	H	Y	O	U	W
C=	O	Q	S	T	C	G	Y	S	M	J	L	F	S	L	W

Одним из методов определения длины ключевого слова, использованного при шифровании текста по методу Виженера, является **метод *Касиски*** (Kasiski).

Данный метод основан на предположении, что наличие повторяющихся ***l*-грамм** (*l*-символьных последовательностей) в зашифрованном тексте будет в большинстве случаев обусловлено наличием соответствующих повторяющихся *l*-грамм в исходном тексте. Предполагается, что случайное появление в шифротексте повторяющихся *l*-грамм маловероятно.

Одинаковым *l*-граммам, присутствующим в исходном тексте, будут соответствовать одинаковые *l*-граммы, расположенные на тех же позициях в шифротексте, только в том случае, если при шифровании они будут преобразованы с использованием тех же *l* символов ключа. Это условие будет выполняться для всех повторяющихся *l*-грамм, расположенных друг от друга на расстояниях, кратных длине ключевого слова шифра.

Тест Касиски состоит из следующих шагов:

1. Анализируется шифротекст на предмет присутствия в нём повторяющихся *l*-грамм.
2. Для каждой из встретившихся в шифротексте более одного раза *l*-граммы вычисляются расстояния между её соседними вхождениями.
3. Вычисляется наибольший общий делитель полученного на предыдущем шаге множества расстояний с учётом того, что среди найденных повторений *l*-грамм могут в незначительном количестве присутствовать случайные повторения. Полученное значение и будет являться длиной ключевого слова.

Эксперименты показывают, что данный метод является достаточно эффективным при анализе зашифрованных текстов на русском и английском языках в случае, если в тексте присутствуют повторяющиеся *l*-граммы длиной в 3 и более символов.