ПРАКТИЧЕСКАЯ КРИПТОЛОГИЯ
ЛЕКЦИЯ 2
Специальность: 6.170101 – Бсіт
Лектор: Сушко С.А.

§6. ХАРАКТЕРИСТИКИ ОТКРЫТЫХ ТЕКСТОВ

Последовательность n соседних букв текста называется n-***граммой***. Общее количество n-грам, образованных из m-буквенного алфавита, равно m^n. Последовательность 2 соседних букв текста называют ***биграммой***, 3 букв – триграммой. Длина текста – это количество букв в тексте. Из текста длиной N можно образовать $N-(n-1)$ разных n-грам.

Например, слово СТАТТЯ (текст длиной 6) разбивается на 5 биграмм:

СТАТТЯ = СТ ТА АТ ТТ ТЯ

Частотные свойства открытых текстов:

1) частота появления букв, биграмм, триграмм в тексте приблизительно одинакова и не зависит от текста (*на частоты букв может слегка повлиять тематика текста и его длина. Например, редкая буква букву Ф часто встречается в техничных текстах со словами функция, дифференциал и др.*). Запомнить наиболее употребляемые буквы позволяют мнемонические правила:

- для русского языка – СЕНОВАЛИТР,
- для английского – TETRIS-HONDA,
- для украинского – НОВА-І-ТИРЕ.

Составлены специальные таблицы частот букв, биграм различных языков.

2) каждому языку свойственно определенное сочетание букв друг с другом;

3) в европейских языках наблюдается чередование гласных и согласных;

4) избыточность языка, благодаря которой можно с большой вероятностью «угадать» смысл сообщения даже при наличии неизвестных букв;

5) самым распространенным символом есть пробел.

Все n-граммы ($n=1,2,3\ldots$), которые можно образовать из букв алфавита любого языка, делятся на два типа:

- ***допустимые*** n***-граммы***, встречающиеся хотя бы в одном тексте
- ***запрещенные*** n***-граммы***, которых нет ни в одном тексте.

Суть ***детерминированной модели открытых текстов***: открытый текст – это последовательность букв алфавита, которая не должна содержать запрещенные n-граммы.

§7. ШИФРЫ ПРОСТОЙ ЗАМЕНЫ. ЧАСТОТНЫЙ КРИПТОАНАЛИЗ

Классические шифры с симметричным ключом делят на две категории:

1) ***шифры замены (подстановки)***, которые заменяют один символ открытого текста на другой символ в зашифрованном тексте.

2) ***шифры перестановки***, которые меняют местами позиции символов открытого текста.

В криптографии рассматриваются четыре типа подстановки: (замены): моноалфавитная, гомофоническая, полиалфавитная и полиграммная.

Гомофоническая замена одному символу открытого текста ставит в соответствие несколько символов шифротекста. Этот метод применяется для искажения статистических свойств шифротекста. При ***моноалфавитной замене*** каждой букве алфавита открытого текста ставится в соответствие одна буква одного алфавита шифротекста. ***Полиафавитная подстановка*** использует несколько алфавитов шифротекста. Полиграммная замена формируется из одного алфавита с помощью специальных правил для шифрования n-грамм.

Рассмотрим моноалфавитные подстановки.

1). ***Простая замена*** (***простая подстановка***) – каждой букве x алфавита открытого текста ставится в соответствие множество ***шифрообозначений буквы***. Если буквы $x \neq x'$, то множества их шифрообозначений не пересекаются.

Если множество шифрообозначений буквы состоит из нескольких символов, то это омофонная подстановка, а если шифрообозначение буквы состоит из одного символа – тоэто ***шифр простой однозначной замены***.

К шифрам простой замены относят:

- ***шифр моноалфавитной буквенной подстановки,*** который отображает алфавит открытого текста сам в себя $\pi : Z_m \to Z_m$, т.е. $x_i \to \pi(x_i)$, $i = 1,2,...,m$, $\pi(x_i)$ – шифрообозначение буквы x_i. Это сокращенно записывают:

$$\pi = \begin{pmatrix} x_1 & x_2 & \dots & x_m \\ \pi_1 & \pi_2 & \dots & \pi_m \end{pmatrix}.$$

Вторую строку матрицы называют ***ключом*** k ***буквенной подстановки.*** Количество разных ключей при алфавите из m букв,

равно количеству перестановок из m елементов $m!$. Для украинського алфавита из 33 букв это $33!$, или $\approx 10^{35}$.

$33!$=8683317618811886495518194401280000000

- ***Шифры сдвига (шифр Цезаря)*** – исторично первый шифр простой замени. Вторая строка матрицы – это последовательность букв, записаних по алфавиту, но с цикличным сдвигом.

Уравнение шифрования	Уравнение дешифрования
$y_i = (x_i + k) \bmod m$, где ключ $0 \le k < m$	$x_i = (y_i + k') \bmod m$, где ключ $k' = m - k$

m – количество букв алфавита. Общее количество ключей $m-1$ (ключ с нулевым сдвигом бесполезен). *По поручению римского императора Гая Юлия Цезаря этим шифром шифровал сообщения Марк Туллий Цицерон, используя ключ* $k = 3$.

Последовательное применение шифра сдвига, один раз з ключом k_1, а второй раз с ключом k_2, эквивалентно применению шифра сдвига с ключом $k_1 + k_2$.

- ***Лозунговые шифры***, в которых вторую строку матрицы заполняют по правилу: вначале вписывают слово-лозунг, а далее вписывают в алфавитном порядке буквы алфавита, которые не вошли в лозунг.

Криптограммы, полученные с помощью моноалфавитных подстановок, сохраняют, все частотные свойства открытого текста. На этом базируется ***частотный криптоанализ***.

I этап вскрытия шифра буквенной замены – подсчет в криптограмме частот повторения шифрообозначений букв, биграм, триграм. В длинном тексте эти частоты близки к среднестатистичным эталонным частотам букв (биграмм и триграмм) языка.

II этап вскрытия шифра буквенной замены –выдвижение гипотез относительно шифрообозначений.

Критерием неправильной дешифровки следует считать появление запрещенных биграм (триграмм). Дешифруемый текст должен быть читабельным и в нем должны появляться ***словоподобные структуры***.

- Сегодня моноалфавитная замена применяется в некоторых вирусах и **shell-кодах** (вредительских программах, которые проникают на серверы с помощью переполнения буфера).
- Брандмауэры, антивирусы и IDS (Intrusion Detection Software) находят такие программы, используя частотный анализ.

Кодировка «КОИ-8 в квадрате», дважды или трижды перекодированные e-mail — это тоже частный случай

моноалфавитной замены. Частотный анализ применительно к shell-кодам описан в 61-ом номере хакерского журнала Phrack.

С целью определении возможности использования морфологических свойств языка при контекстном моделировании для оценки вероятности появления очередного символа были проведены статистические исследования морфологического состава текстов различных стилей на русском и украинском языках. В результате этих исследований было установлено следующее [4].

1) Наиболее часто слова украинского языка начинаются с приставок «*в-*», «*з-*», «*с-*», после них в порядке убывания вероятностей располагаются приставки «*о-*», «*на-*», «*по-*». Для русского языка наиболее частыми приставками в предложениях являются «*по-*», «*не-*», «*на-*» и «*за-*», «*от-*», «*при*». Следует отметить, что вероятности первых трех приставок в два раза превышают значения вероятностей следующих за ними морфем.

2) Частота встречаемости приставок в текстах различных стилей заметно отличается. Так в научной и художественной украинской литературе вероятность появления приставки «*с-*» составляет 12—13 %, а для официально-делового стиля — 8 %.

3) Имеется существенное различие вероятности появления в текстах украинского языка приставки «*пре-*» для художественного, научного и официально-делового стилей; вероятность появления приставки в первом случае составляет 0,07 %, а для двух других — 0,5 %.

4) Наименее вероятными украинскими приставками являются «*понад-*» и «*анти-*», их суммарная вероятность составляет 0,02 %, русскими – «*сверх-*», «*анти-*».

5) Некоторые украинские приставки встречаются примерно с одинаковой вероятностью для всех стилей: «*з-*», «*за-*», «*під-*», «*ні-*», «*пере-*». Для текстов одного стиля, но различных тематик приставки используются практически с одинаковой частотой.

6) Наиболее часто для образования украинских слов используются суффиксы «*-в-*», «*-ість-*», «*-н-*», «*-ок-*», «*-ов-*», «*-ер-*», «*-ова-*», «*-ик-*», «*-ин-*», «*-ник-*», их суммарная вероятность составляет 87 %. В русском языке такой группой суффиксов являются «*-ость-*», «*-ность-*», «*-ик-*», «*-ова-*»,«*-ник-*», «*-тель-*», «*-ств-*», суммарная вероятность которых равна 85 %.

Вісник СевНТУ. Вип. 101: Інформатика, електроніка, зв'язок: зб. наук. пр. — Севастополь: Вид-во СевНТУ, 2010.

§8. НЕКОТОРЫЕ ДРУГИЕ КЛАССИЧЕСКИЕ ШИФРЫ

1. <u>Аффинные шифры</u> (число ключей указано для алфавита мощностью m, $\varphi(m)$ – функция Эйлера, единица отнимается из-за нешифрующего ключа).

Шифры	Уравнение шифрования	Уравнение дешифрования	Количество ключей
Сдвига	$y_i = (x_i + k) \bmod m$, где ключ $0 \le k < m$	$x_i = (y_i + k') \bmod m$, где ключ $k' = m - k$	$m - 1$
Линейные	$y_i = (kx_i) \bmod m$, где ключ $0 \le k < m$; НСД$(k, m) = 1$	$x_i = (k'y_i) \bmod m$, где ключ $k' = k^{-1} \bmod m$	$\varphi(m) - 1$ (не шифрует ключ k=1)
Аффинные	$y_i = (kx_i + t) \bmod m$ где ключи $0 \le t < m$; $0 \le k < m$; НСД$(k, m) = 1$	$x_i = (k'y_i + t') \bmod m$, где ключи $k' = k^{-1} \bmod m$; $t' = (-k't) \bmod m$	$m \cdot \varphi(m) - 1$ (не шифрует пара k=1, t=0).

§9. Шифр Виженера и его криптоанализ

Главный недостаток всех аффинных шифров и моноалфавитных подстановок – замена при шифровании одной буквы открытого текста на одно шифрообозначение и отсюда подверженность их частотному анализу. Метод борьбы с этим – использование полиалфавитных подстановок. Шифр называется ***полиалфавитной подстановкой*** (***шифром***), если алфавит шифротекста изменяется в процессе шифрования. Они маскируют естественную частоту появления символов в шифрованном тексте, поэтому надежнее моноалфавитных подстановок.

Самый известный полиалфавитный шифр – это шифр Виженера.

Первое точное документированное описание полиалфавитного шифра было сформулировано Альберти в 1467 году, для переключения между алфавитами использовался металлический шифровальный диск. Система Альберти переключает алфавиты после нескольких зашифрованных слов. То, что сейчас известно под шифром Виженера, впервые описал Джованни Беллазо. Блез Виженер представил своё описание простого, но стойкого шифра перед комиссией Генриха III во Франции в 1586 году, и позднее изобретение шифра было присвоено именно ему.

Напомним, суть полиалфавитного шифрования заключается в циклическом применении нескольких моноалфавитных шифров к определённому числу букв шифруемого текста. Например, пусть у нас имеется некоторое сообщение x_1 , x_2 , x_3 , ….. x_n , …… x_{2n} , ….., которое надо зашифровать. При использовании полиалфавитного шифра имеется несколько моноалфавитных шифров (например, n штук). И в нашем случае к первой букве применяется первый моноалфавитный шифр, ко второй букве — второй, к третьей — третий….. к n-ой букве — n-й, а к n+1 опять первый, ну и так далее. Таким образом, получаётся довольно-таки сложная последовательность, которую уже не так просто вскрыть, как один моноалфавитный шифр. Самым важным эффектом, достигаемым при использовании полиалфавитного шифра, является маскировка частот появления тех или иных букв в тексте, на основании которой обычно очень легко вскрываются моноалфавитные шифры.

Ключ шифра – последовательность букв алфавита $\gamma = (\gamma_1, \gamma_2, ..., \gamma_d)$ – ключевое слово (***лозунг***), называемый еще ***гаммой***. Его записывают под каждой буквой открытого текста (если текст длиннее ключа – несколько раз) и далее применяют так называемую таблицу Виженера – квадратную матрицу из букв

алфавита, строки которой представляют собой предыдущую строке с цикличным сдвигом на один символ влево.

При шифровании открытого текста буква x_i определяет номер столбца в таблице Виженера, а соответствующая буква γ_i гаммы показывает номер строки, на пересечении которых стоит шифрообозначение буквы x_i. При дешифровке под буквами криптограммы записывают буквы ключа и по строке таблицы Виженера, которая отвечает данной букве ключа, ищут букву криптограммы. Буква, стоящая над нею в первой строке таблицы, будет дешифрованной буквой.

Аппарат теории чисел позволяет отказаться от таблиц Виженера. Пусть $(x_1, x_2, ...x_l)$ – открытый текст длиной l; $\gamma = \gamma_1, \gamma_2, ..., \gamma_d$ –гамма, d – длина гаммы, которую еще называют ***периодом гаммы***. Уравнение шифрования i-ой буквы сообщения имеет вид

$$y_i = x_i + \gamma_i (\mathrm{mod}\, m),\ \ i = 1,2,...d,$$

а уравнение дешифрования

$$x_i = y_i - \gamma_i (\mathrm{mod}\, m),\ \ i = 1,2,...d.$$

где x_i, y_i, γ_i – номера букв в открытом тексте, криптограмме и ключе соответственно, m – мощность алфавита.

Количество разных ключей с периодом d в шифре Виженера равно m^d, где m – количество букв алфавита.

Приведем данные о количестве ключей для $m = 34$ (украинский алфавит+пропуск между словами) при разных d:

d	1	2	3	4	5
34^d	34	1156	$3{,}9 \cdot 10^4$	$1{,}3 \cdot 10^6$	$4{,}5 \cdot 10^7$
d	6	7	8	9	10
34^d	$1{,}5 \cdot 10^9$	$5{,}2 \cdot 10^{10}$	$1{,}8 \cdot 10^{12}$	$6 \cdot 10^{13}$	$2 \cdot 10^{15}$

Видно, что раскрыть шифр методом перебора практично невозможно.

Однако частотный анализ работает и против шифра Виженера.

Криптоанализ шифра Виженера состоит из двух этапов.

1) ***I этап*** – определение длины гаммы на основе ***теста Казиски***. Еще в XIX веке немецкий криптоаналитик Казиски заметил, что два одинаковых отрезка открытого текста будут одинаково зашифрованы, если начала отрезков расположены один от одного на расстоянии, кратном периоду гаммы d (в этом случае одинаковые

сочетания шифруются одинаковою частью ключа и в результате в криптограмме появляются одинаковые сочетания символов).

Например: КРИПТО КРАТКО ВІД КРИПТОГ РАФІ Ї
НАТОВПНАТОВПНАТОВ ПНАТОВПНАТОВ
ЮРЯГФГМКЖОФ Б ФЩЄОЮРЯГФГРРТЖЙЮ

Поэтому если $d_1, d_2, \ldots$ – расстояния в криптограмме между какими-то одинаковыми триграммами или n-граммами, то наибольший общий делитель НОД($d_1, d_2, \ldots$) чисел $d_1, d_2, \ldots$, должен делиться на число d.

Уточняют период гаммы с помощью индекса совпадений, (*впервые введен в практику У. Фридманом в 1920 г.*). Дадим определение.

Пусть буквы алфавита $A = \{a_1, a_2, \ldots, a_m\}$ пронумерованы и отождествлены с кольцом вычетов $Z_m = \{0, 1, 2, \ldots m-1\}$ за модулем m. Обозначим:

$x = (x_1, x_2, \ldots, x_n)$ строка, составленная из n букв алфавита. Например, в строке $x = (a_2, a_0, a_1, a_1, a_3)$ $n = 5$, $x_1 = a_2$, $x_2 = a_0$, $x_3 = a_1$, $x_4 = a_1$ и т.д.

f_i – частота появления буквы с номером i в строке.

Индексом совпадения $I_c(x)$ ***в строке*** называют вероятность того, что две наугад выбранные буквы этой строки совпадают. Он равен отношению количества благоприятствующих событий, когда на двух выбранных позициях в строке стоят две одинаковые буквы, к общему числу всех событий:

$$I_c(x) = \frac{\sum_{i=0}^{m-1} f_i(f_i - 1)}{n(n-1)}.$$

Для строк осмысленного текста индекс совпадения равен

$$I_c(x) = \sum_{i=0}^{m-1} p_i^2,$$

где p_i – вероятность появления букв алфавита в тексте на данном языке. Вместо них можно использовать эталонные частоты употребления букв в открытых текстах.

Подсчитано, что индекс совпадения строк осмысленного текста для разных языков такой:

$I_c(x) \approx 0{,}0575$ – украинский язык;

$I_c(x) \approx 0{,}0529$ – русский язык;

$I_c(x) \approx 0,0662$ – английский язык.

Для строк неосмысленного текста, написанного с использованием алфавита из m букв, т.е. для текста составленного из наугад выбираемых букв алфавита, этот индекс равен

$$I_c(x) = \sum_{i=0}^{m-1} p_i^2 = \sum_{i=0}^{m-1} \frac{1}{m^2} = m \cdot \frac{1}{m^2} = \frac{1}{m}.$$

В случае неосмысленного текста с использованием:

латиницы(+пробел) $I_c(x) = \frac{1}{26} \approx 0,037$,

украинского алфавита(+пробел) $I_c(x) = \frac{1}{34} \approx 0,029$.

Подобные расчеты будут правильными и для криптограмм, полученных при шифровании открытых текстов с помощью шифров простой замены, так как вероятности букв только переставляются местами, но их сумма $\sum_{i=0}^{m-1} p_i^2$ не меняется.

Используем индексы совпадений для уточнения периода гаммы. Пусть криптограмма $y_1, y_2, ..., y_n$, полученная с помощью шифра Виженера с гаммою периода d. Запишем ее буквы в d столбцов:

Y_1	Y_2	...	Y_d
y_1	y_2	...	y_d
y_{d+1}	y_{d+2}	...	y_{2d}
y_{2d+1}	y_{2d+2}	...	y_{3d}
...	...	...	...

Если период гаммы определен правильно, то каждый столбец Y_i – отрезок открытого текста, зашифрованного простой заменой. Тогда индексы совпадений каждого столбца совпадет с индексом осмысленных текстов на данном языке.

Если же период гаммы определен неверно, то столбцы Y_i будут «случайными», так как они – результат шифрования открытого текста с помощью полиалфавитной подстановки, а индексы совпадения столбцов таких столбцов ближе к числу $\frac{1}{m}$, где m – количество букв алфавита.

Этот метод определения длины d гаммы работает при $d \le 30$.

2) ***II этап криптоанализа*** – определение гаммы шифра. Пусть $x = (x_1, x_2, ..., x_n)$ и $y = (y_1, y_2, ..., y_N)$ – два строки, образованные из

букв алфавита $A=\{a_1,a_2,...,a_m\}$. ***Взаимным индексом совпадения*** $MI_c(x)$ ***в строках*** x ***и*** y называется вероятность того, что наугад выбранная буква из строки x совпадет с наугад выбранной буквою из строки y. Обозначим $f_0,f_1,...,f_{m-1}$ и $f'_0,f'_1,...,f'_{m-1}$ – частоты повторения букв в строках x и y соответственно. Взаимный индекс совпадения $MI_c(x,y)$ этих строк вычисляют по формуле:

$$MI_c(x,y)=\frac{\sum_{i=0}^{m-1} f_i\cdot f'_i}{n\cdot N}.$$

При правильно определенном периоде гаммы взаимный индекс совпадения $MI_c(Y_i,Y_j)$ двух столбцов криптограммы равен

$$MI_c(Y_i,Y_j)=\sum_{t=0}^{m-1} p_t\cdot p_{t+(s_i-s_j)\bmod m}.$$

Разность $(s_i-s_j)\bmod m$ называют ***относительным сдвигом столбцов*** Y_i и Y_j. Столбцы Y_i и Y_j с одинаковыми сдвигами s и $m-s$ имеют одинаковые взаимные индексы совпадения.

Так, для украинского алфавита ненулевым сдвигам соответствуют взаимные индексы совпадения от 0,021 до 0,037, а при нулевом сдвиге взаимный индекс совпадения наибольший 0,057.

Потому после вычисления периода гаммы шифра определяют значения относительных сдвигов s_i-s_j столбцов Y_i та Y_j.

Взаимные индексы совпадения столбцов $MI_c(Y_i,Y_j^s)$ вычисляют по формуле (следует из определения взаимного индекса):

$$MI_c(Y_i,Y_j^s)=\frac{\sum_{t=0}^{m-1} f'_t\cdot f''_{t-s(\bmod m)}}{n'\cdot n''},$$

где f'_t – частота повторений буквы в столбце Y_i (буква в алфавите имеет номер t); $f''_{t-s(\bmod m)}$ – частота повторений буквы, которая имеет в алфавите номер $t-s(\bmod m)$, в столбце Y_j; $0\le s\le m-1$; $1\le i<j\le d$; d – период гаммы; n', n'' – количество букв в столбцах Y_i и Y_j соответственно; m – количество букв алфавита. Всего можно посчитать $m\cdot C_d^2$ разных взаимных индексов совпадения.

Если $M I_c(Y_i, Y_j^s) \approx I_c(x)$ ($I_c(x)$-индекс совпадения осмысленных текстов), то $s = s_i - s_j$ и относительный сдвиг столбцов $(s_i - s_j) \bmod m = 0$. Если же $M I_c(Y_i, Y_j^s) \approx 1/m$ (индексу совпадения случайных буквенных последовательностей), то $s \neq s_i - s_j$ и $(s_i - s_j) \bmod m \neq 0$. Тогда для определения ключевого слова (гаммы) можно составить систему сравнений, которые свяжут сдвиги s разных столбцов, из решения которой и будет выбран ключ (при условии, что он осмысленный).

§10. Криптоанализ шифра Виженера методом вероятных слов

Метод вероятных слов использует известную тематику открытого текста. Например, если в криптограмме зашифрован финансовый отчет, то в тексте могут встретиться слова “дебет”, “кредит”, “баланс” и т.п. Чтобы проверить наличие вероятного слова в тексте, надо вычесть его из криптограммы по модулю m во всех возможных позициях, где m – мощность алфавита. Если такое слово присутствует в тексте и вычитается в правильной позиции, то в результате вычитания возникнет ключ шифрования или его часть. Если слово испытывается в неправильной позиции, то в результате вычитания появится бессмысленный набор букв. Понятно, что, если этого слова нет в тексте, все позиции будут неправильными.

Полученная в результате вычитания хотя бы часть осмысленного слова - показатель успеха. Далее надо попытаться расширить открытый текст или ключ в этом направлении.

Пример: криптограмма ВФЮЕОСЦШШЕРРПХХЧДЖРУЯФЯР есть результат применения шифра Виженера к украинскому тексту (33 буквы+пробел), тематика которого – криптография. Найти ключ.

Р е ш е н и е: 1) ключ КРИПТОГРАФИЯ в разных позициях:

Г Ю І О Р П І Д Х П П И Б Б К Г П Ї Л Є Ю Ж

К Р И П Т О Г Р А Ф И Я

У І Б Я Б ЖПХ Щ Ж Ї С …

Г Ю І О Р П І Д Х П П И Б Б К Г П Ї Л Є Ю Ж

К Р И П Т О Г Р А Ф И Я

Ґ Н Х Ж Б Ю Ч В Д П Щ Г В …

и т.д. сдвигаем ключ. Осмысленный текст не появляется ни в одной позиции.

2) ключ - АБЕТКА

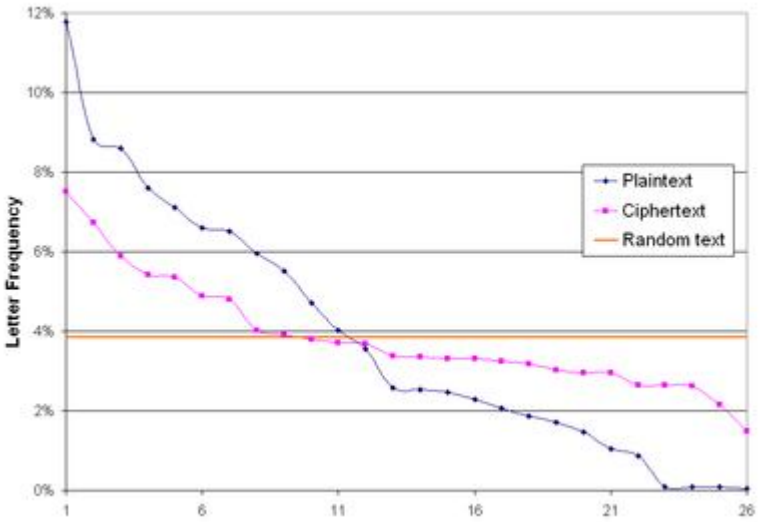
Г Ю І О Р П І Д Х П П И Б Б К Г П Ї Л Є Ю Ж

А Б Е Т К А

Г Ь Д Ь Е П І Ґ …

Г Ю І О Р П І Д Х П П И Б Б К Г П Ї Л Є Ю Ж
А Б Е Т К А
Ґ Я Ї П Р О Д Н І П Р І В В К В Й Ф Б Є Я З

Вероятно ключ ДНІПРО. Проверяем его и в результате возникает: ЯКА АБЕТКА ЦЬОГО ШИФРУ?



§11. ШИФРЫ ПЕРЕСТАНОВКИ

Простой перестановочный шифр с фиксированным периодом *n* подразумевает разбиение исходного текста на блоки по *n* символов и использование для каждого такого блока некоторой перестановки. Такие шифры так меняют правильный порядок расположения букв открытого текста, что он терцет смысл.

Ключ шифра – используемая при шифровании перестановочная матрица, указывающая правило перестановки. Общее число возможных ключей определяется длиной блока *n* и равно *n*!. Хотя число подходящих ключей $\leq n!$, так как не подходят перестановки, меняющие положение только двух букв, трех букв и т. д. При дешифровании используется матрица обратной перестановки.

Различают шифры горизонтальной, вертикальной, двойной перестановки, решетки, лабиринты, лозунговые и др.

Основные шифры перестановки

Шифр	Шифрование и дешифрование
Шифр скиталя	На цилиндр наматывали ленту пергамента, вдоль оси цилиндра в строку записывали открытый текст (V – VI ст. до н.е.). Ключом был диаметр цилиндра
Использование подстановки	Ключ шифра – перестановка σ, которая переставляет буквы открытого текста, а ключ дешифрования –

степени m	обратная перестановка σ^{-1}.
маршрутные перестановки Открытый текст записывают в некоторую геометрическую фигуру по определенной траектории, а выписывают в измененном порядке.	*Шифр вертикальной перестановки* Открытый текст вписывается в прямоугольник строками слева направо. Криптограмма выписывается по столбцам, которые выбирают в определенной последовательности по ключу
	Шифр перестановки строк Открытый текст вписывается в прямоугольник столбцами сверху вниз. Криптограмма выписывается по строкам, которые выбирают в определенной последовательности по ключу
	Лабиринты Маршрутные перестановки со сложными маршрутами записи открытого текста и криптограммы. Например, путь в лабиринте или как при обходе шахматной доски «ходом коня»
Поворотные решетки	Решетка Кардано Ключ – карточка с окошками, которая при наложении на лист оставляет открытыми только некоторые его части. Секретный текст вносят в окошки при разном положении решетки, зависящем от поворота карточки на 90^0 против часовой стрелки. При дешифровании криптограмму вписывают в диаграмму нужных размеров, а потом накладывают решетку и читают открытый текст

Пример. Используя ключ $(3,1,2,5,4)$ зашифровать открытый текст СТАНЦУЕМ ТАНГО (пробел исключить). Найти ключ дешифрования.

Р е ш е н и е. СТАНЦУЕМТАНГО = СТАНЦ УЕМТА НГО =
= АСТЦН МУЕАТ ОНГ= АСТЦНМУЕАТОНГ.

Ищем ключ дешифрования:

$$(3,1,2,5,4)=\begin{pmatrix}1,2,3,4,5\\3,1,2,5,4\end{pmatrix}=\begin{pmatrix}3,1,2,5,4\\1,2,3,4,5\end{pmatrix}=\begin{pmatrix}1,2,3,4,5\\2,3,1,5,4\end{pmatrix}=(2,3,1,5,4).$$

перестановка строк местами

упорядочение столбцов по возрастанию номеров в I строке

Криптоанализ шифров перестановки:

1) критерий неправильности перестановки при дешифровании – появление запрещенных n-грам;

3) первые столбцы таблицы, куда вписывался открытый текст, могут быть длиннее на одну букву по сравнению с остальными;
4) использование вероятностных слов, характерных для тематики открытого текста.

Замечание. При атаке с известным открытым текстом легко узнать, использовался ли шифр перестановки. Посчитать частоты повторения каждой буквы в открытом тексте и криптограмме. Совпадение частот свидетельствует о шифре перестановки.

ПРИЛОЖЕНИЕ

Повторение из дискретной математики. Упорядоченное размещение элементов множества $X=\{x_1,x_2,...,x_n\}$, *то есть размещения, в котором показано, какой элемент множества первый, какой – второй, и т.д., называется* ***перестановкой*** *множества. Элементы множества перенумеровывают числами* $1,2,...,n$, *и далее считают сами числа* $1,2,...,n$ *элементами перестановки. Перестановка обозначается* $(x_1,x_2,...,x_n)$, *где* x_1 *– первый,* x_2 *– второй, …,* x_n *– последний элемент перестановки или* $\sigma=\begin{pmatrix}1 & 2 & ... & n\\ \sigma(1) & \sigma(2) & ... & \sigma(n)\end{pmatrix}$, *где* $\sigma(i)$, $i=\overline{1,n}$ *– новый номер элемента, который до перестановки имел* i *в упорядоченной последовательности,* n *– период перестановки. Так, перестановки* (a,b,c,d,e), (c,b,a,e,d), (b,d,c,e,a) *– пример разных перестановок множества* $\{a;b;c;d;e\}$. *Если же перестановку* (c,b,a,e,d) *задать двустрочной матрицей, то она будет иметь вид* $\sigma=\begin{pmatrix}1 & 2 & 3 & 4 & 5\\ 3 & 2 & 1 & 5 & 4\end{pmatrix}$.

§12. Шифр Хилла

Пусть открытый текст $(x_0,x_1,...,x_{N-1})$ длиной N записан буквами алфавита, которые отождествлены с элементами кольца вычетов Z_m по модулю m (m – число букв алфавита). Разобьем открытый текст на n-граммы, где для удобства $N\vdots n$.

В ***шифре Хилла*** ключ шифру – некоторое линейное преобразование с невырожденной квадратной матрицей K порядка n. Уравнение шифрования

$$Y=KX,$$

где столбцами матрицы X служат n-граммы открытого текста. Криптограмма Y выписана по столбцам.

Уравнение дешифрования

$$X=K^{-1}Y,$$

где K^{-1} – обратная матрица к матрице K.

Шифр Хилла – это моноалфавитная подстановка на $n-$граммах.

Алгоритм использования шифра Хилла:

1. Пронумеровать буквы алфавита открытого текста от 0 до $m-1$ и записать открытый текст в цифровом виде.
2. Разбить открытый текст на $n-$граммы и записать их в столбцы матрицы X открытого текста.
3. Сформировать ключ шифрования, для чего выбрать квадратную матрицу K порядка n, определитель которой $|K| \neq 0$ и

   НОД $((|K|, m) = 1.$
4. Вычислить $Y = KX$. Криптограмму выписать из матрицы Y по столбцам.

Свойства шифра Хилла:

1^0 Шифр Хилла – линейный шифр n-го порядка.

2^0. Шифрование выравнивает частоты символов криптограммы.

3^0. Число ключей шифра Хилла равно числу $\phi_n(m)$ обратимых матриц порядка n с элементами кольца вычетов Z_m. При $m > 4$ число ключей оценивается неравенством

$$\phi_n(m) \geq \frac{m^{n^2}}{(6 \ln \ln m)^n} \quad .$$

Для алфавита из $m = 31$ буквы число ключей ≈ 65 млрд. при $n = 3$, $\approx 2 \cdot 10^{48}$ при $n = 6$.

Пример. Выбрать ключ шифрования шифра Хилла (матрица 3-го порядка) и зашифровать сообщение (алфавит украинский с условием Ґ=Г и І=Ї). Вычислить ключ дешифрования и дешифровать криптограмму.

Р е ш е н и е. После исключения букв Ґ и Ї из украинского алфавита пронумеруем буквы от 0 до 30 $(\mathrm{mod}\, 31)$

А	Б	В	Г	Д	Е	Є	Ж	З	И	І	Й	К	Л	М	Н	О	П	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ь	Ю	Я
↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
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	26	27	28	29	30

Ключ – матрица 3-го порядка $\Rightarrow$ текст разбиваем на триграммы:

ЗГОДЕН = ЗГО ДЕН

ЗГО=8 3 16; ДЕН= 4 5 15

Открытый текст $X = \begin{pmatrix} 8 & 4 \\ 3 & 5 \\ 16 & 15 \end{pmatrix}$.

Ключ $K = \begin{pmatrix} 3 & 2 & 21 \\ 4 & 2 & 1 \\ 9 & 13 & 4 \end{pmatrix}$. $|K| = 685 = 3 (\mathrm{mod}\, 31)$, НОД $((|K|, m) = 1$.

Шифрование:

$$Y = KX = \begin{pmatrix} 3 & 2 & 21 \\ 4 & 2 & 1 \\ 9 & 13 & 4 \end{pmatrix} \cdot \begin{pmatrix} 8 & 4 \\ 3 & 5 \\ 16 & 15 \end{pmatrix} = \begin{pmatrix} 366 & 337 \\ 54 & 41 \\ 175 & 161 \end{pmatrix} = \begin{pmatrix} 25 & 27 \\ 23 & 10 \\ 20 & 6 \end{pmatrix} \mathrm{mod}\, 31.$$

Криптограмма:

25 23 20 27 10 6 $\rightleftarrows$ Ч Х Т Щ І Є.

Ищем ключ дешифрования. $|K|^{-1} = 3^{-1} \equiv 21 (\mathrm{mod}\, 31)$;

$K_{11} = \begin{vmatrix} 2 & 1 \\ 13 & 4 \end{vmatrix} = -5 \equiv 26$; $K_{12} = -\begin{vmatrix} 4 & 1 \\ 9 & 4 \end{vmatrix} = -7 \equiv 24$;

$K_{13} = \begin{vmatrix} 4 & 2 \\ 9 & 13 \end{vmatrix} = 34$;

$K_{21} = 17;\ K_{22} = 9;\ K_{23} = 10;\ K_{31} = 22;\ K_{32} = 19;\ K_{33} = 29.$

$$K^{-1} = 21 \begin{pmatrix} 26 & 17 & 22 \\ 24 & 9 & 19 \\ 34 & 10 & 29 \end{pmatrix} \equiv \begin{pmatrix} 19 & 16 & 28 \\ 8 & 3 & 27 \\ 1 & 24 & 20 \end{pmatrix}.$$

Дешифрование: $X = K^{-1}Y = \begin{pmatrix} 19 & 16 & 28 \\ 8 & 3 & 27 \\ 1 & 24 & 20 \end{pmatrix} \cdot \begin{pmatrix} 25 & 27 \\ 23 & 10 \\ 20 & 6 \end{pmatrix} \mathrm{mod}\, 31 =$

$$= \begin{pmatrix} 1403 & 841 \\ 809 & 408 \\ 977 & 387 \end{pmatrix} \mathrm{mod}\, 31 = \begin{pmatrix} 8 & 4 \\ 3 & 5 \\ 16 & 15 \end{pmatrix} =$$ ЗГОДЕН.

§13. Криптоанализ шифра Хилла

Зная матрицы X и Y, умножим обе части уравнения $Y = KX$ на X^{-1}:

$$YX^{-1} = KXX^{-1} = K, \qquad \text{т.е. } K = YX^{-1}.$$

Поэтому криптоаналитик, который перехватит n^2 «пар символ сообщения/символ шифротекста» сможет составить систему линейных уравнений, которую обычно не сложно решить. Если окажется, что система не совместна, то необходимо всего лишь добавить еще несколько пар «символ сообщения/символ шифротекста».

Пример. Криптограмма

ФФЛ МДФ ІБН ДКХ РМВ ШІГ ЕІУ

получена с помощью шифра Хилла. Алфавит украинский из предыдущего примера (Ґ=Г, І=Ї, пропусков нет), текст разбит на триграммы. Последнее слово – подпись отправителя КОРАБЛЬОВ. Найти ключи зашифрования и расшифрования. Прочитать текст.

Р е ш е н и е. Подписи

$$\text{КОРАБЛЬОВ= 12 15 18 – 0 1 13 – 28 16 2} = \begin{pmatrix} 12 & 0 & 28 \\ 15 & 1 & 16 \\ 18 & 13 & 2 \end{pmatrix} = X$$

соответствует конец криптограммы

$$\text{РМВ ШІГ ЕІУ= 18 14 2 – 26 10 3 – 5 10 21} = \begin{pmatrix} 18 & 26 & 5 \\ 14 & 10 & 10 \\ 2 & 3 & 21 \end{pmatrix} = Y.$$

$$\Rightarrow \quad K = YX^{-1} = \begin{pmatrix} 18 & 26 & 5 \\ 14 & 10 & 10 \\ 2 & 3 & 21 \end{pmatrix} \cdot \begin{pmatrix} 12 & 0 & 28 \\ 16 & 1 & 16 \\ 18 & 13 & 2 \end{pmatrix}^{-1}.$$

$$|X| = 2848 \equiv 27 (\mathrm{mod}\, 31),\ 27^{-1} = 23 (\mathrm{mod}\, 31)$$

$$X^{-1} = 23 \begin{pmatrix} 11 & 23 & 3 \\ 8 & 16 & 8 \\ 4 & 30 & 12 \end{pmatrix} = \begin{pmatrix} 5 & 2 & 7 \\ 29 & 27 & 29 \\ 30 & 8 & 28 \end{pmatrix}.$$

$$K = YX^{-1} = \begin{pmatrix} 18 & 26 & 5 \\ 14 & 10 & 10 \\ 2 & 3 & 21 \end{pmatrix} \cdot \begin{pmatrix} 5 & 2 & 7 \\ 29 & 27 & 29 \\ 30 & 8 & 28 \end{pmatrix} = \begin{pmatrix} 994 & 778 & 1020 \\ 660 & 378 & 668 \\ 727 & 253 & 689 \end{pmatrix} \equiv$$

$$\equiv \begin{pmatrix} 2 & 3 & 28 \\ 9 & 6 & 17 \\ 14 & 5 & 7 \end{pmatrix} (\mathrm{mod}\, 31).$$

$$|K| = -653 \equiv 29 (\mathrm{mod}\, 31),\ 29^{-1} = 15 (\mathrm{mod}\, 31)$$

$$K^{-1} = 15\begin{pmatrix} 19 & 26 & 7 \\ 20 & 25 & 1 \\ 23 & 1 & 16 \end{pmatrix} \equiv \begin{pmatrix} 6 & 18 & 12 \\ 21 & 3 & 15 \\ 4 & 15 & 23 \end{pmatrix}.$$

ФФЛ МДФ ІБН ДКХ РМВ ШІГ ЕІУ = 22 22 13 – 14 4 22 – 10 1 15 – 4 12 23 – 18 14 2 – 26 10 3 – 5 10 21.

$$X = K^{-1}Y = \begin{pmatrix} 6 & 18 & 12 \\ 21 & 3 & 15 \\ 4 & 15 & 23 \end{pmatrix} \cdot \begin{pmatrix} 22 & 14 & 10 & 4 & 18 & 26 & 5 \\ 22 & 4 & 1 & 12 & 14 & 10 & 10 \\ 13 & 22 & 15 & 23 & 2 & 3 & 21 \end{pmatrix} \equiv$$

$$\equiv \begin{pmatrix} 2 & 17 & 10 & 20 & 12 & 0 & 28 \\ 10 & 16 & 4 & 0 & 16 & 1 & 16 \\ 4 & 2 & 28 & 12 & 18 & 13 & 2 \end{pmatrix} \Rightarrow$$ ВІДПОВІДЬ «ТАК». КОРАБЛЬОВ

По мере увеличения размерности ключа шифр быстро становится недоступным для расчетов на бумаге человеком. Поэтому Хилл с партнером собрал механическую шифровальную машину, которая упрощала расчеты – выполняла умножение матрицы 6×6 по модулю 26 (число букв в латинице) при помощи системы шестеренок и цепей. Авторы даже получили патент на устройство. К сожалению, машина могла использовать только ограниченное число ключей, и даже с машиной шифр Хилла применяли редко – только для шифрования некоторых государственных радиопередач.

Основной вклад Хилла – математический подход к конструированию надежных криптосистем.