Шифр простой замены

Шифр простой замены это $(\Sigma^*, \Sigma^*, S_\Sigma, E, D)$

Где

- S_Σ-это группа перестановок на Σ

Открытый текст m $= m_1 m_2 ... m_n$
Криптограмма c $= {}_{12} ... c_n$

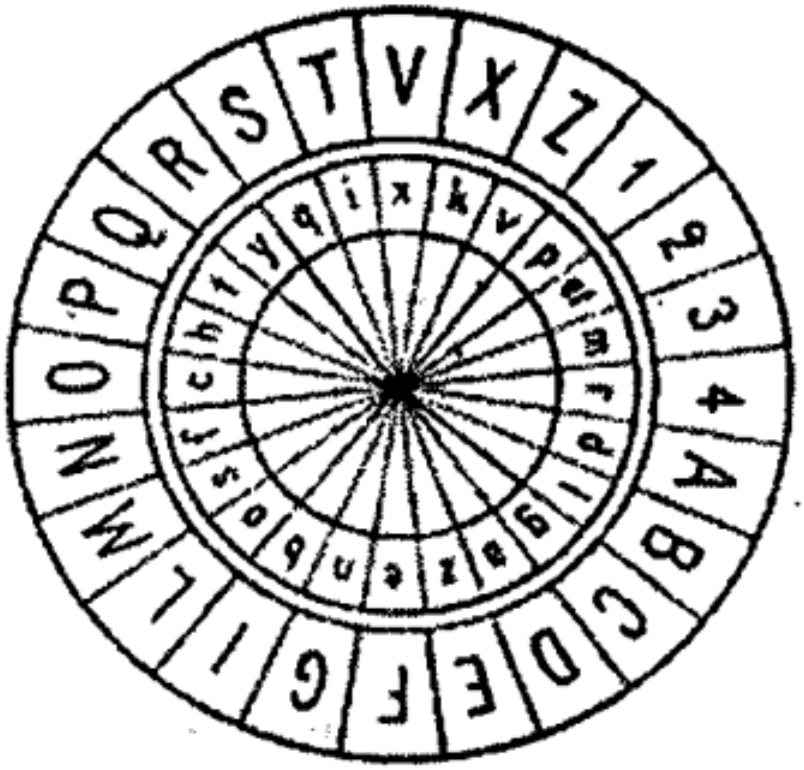
$k \in S_\Sigma$ такое что $k : \Sigma \to \Sigma$ (эта стрелка обозначает биекцию, в латехе нет таких стрелок блять)

$\forall i \in \{1, ...n\} : c_i = k(m_i)$

$\forall i \in \{1, ...n\} : m_i = k^{-1}(c_i)$

Исторические шифры простой замены

Диск Альберти



шифр недешифруем. Реализация шифра осуществлялась с помощью шифровального диска, положившего начало целой серии многоалфавитных шифров. Устройство представляло собой пару дисков – внешний, неподвижный (на нем были нанесены буквы в естественном порядке и цифры от 1 до 4) и внутренний – подвижный – на нем буквы были переставлены. Процесс шифрования заключался в нахождении буквы открытого текста на внешнем диске и замену ее на соответствующую (стоящей под ней) букву шифрованного текста. После шифрования нескольких слов внутренний диск сдвигался на один шаг. Ключом данного шифра являлся порядок расположения букв на внутреннем диске и его начальное положение относительно внешнего диска.

Шифр Цезаря

Если сопоставить каждому символу алфавита его порядковый номер (нумеруя с 0), то шифрование и дешифрование можно выразить формулами модульной арифметики[1][2]:

$y = (x + k) mod(n)$

$x = (y - k) mod(n)$

где

- x—символ открытого текста,
- y—символ шифрованного текста,
- n—мощность алфавита, а
- k—ключ.(Ключ это число, на которое сдвигаем букву)

Шифр Виженера

Шифр виженера это $(\Sigma^*, \Sigma^*, \Sigma^*, E, D)$

Открытый текст m $= m_1 m_2 ... m_n$

Криптограмма c $= {}_{12} ... c_n$

Ключ(какое-то слово) k $= k_1, ... k_p$, где $p \ll n$ (т.е длина ключа намного меньше длины открытого текста)

$c_i = (m_i + k_{[(i-1)\%p]+1}) mod(|\Sigma|)$

$m_i = (c_i - k_{[(i-1)\%p]+1}) mod(|\Sigma|)$

Про криптоанализ

Находимся в условиях, когда криптоаналитик знает какую криптосистему мы используем, но он не знаетключ

ОПР

Атака - Совокупность условий, в которых находится крипто аналитик называется атакой

Виды атак и методы атак

Виды атак:

1. **С известным шифротекстом**. Мы имеем доступ к зашифрованному сообщению и взлом осуществляется с помощью его исследования
2. **С известным открытым текстом**. Известны пары открытого текста и шифр текста
3. **С наиболее вероятным словом**, т.е знаем, что в криптограмме обязательно содержится определенное слово
4. **С избранным открытым текстом**, т.е Атакующий может расшифровывать выбранные шифртексты (кроме целевого). Либо атакующий может шифровать фрагменты открытого текста

Методы атак:

1. **Brute-force attack** (Полный перебор) Суть: Перебор всех возможных ключей. Эффективность: Зависит от длины ключа (например, 128-битный AES требует ~2^{12}⊠ попыток). Защита: Использование длинных ключей (256 бит и более).
2. **Частичная индукция**, т.е восстановление части открытого текста по криптограме
3. **Информационная индукция**, Получение некоторой инфы об открытом тексте по криптограме

Криптоанализ шифра простой замены

Атака №3(не устойчив к этой атаке)

Атакующий может зашифровать весь алфавит. После зашифровки мы получаем сразу весь ключ

Атака №2 (Взлом стал труднее, но ещё возможен)

По паре (m_i, c_i) находим часть ключевой перестановки, (смотрим как переходят буквы из открытго текста в криптограмму, таким образом получаем часть ключа).

Затем частично расшифровываем целевую криптограмму, а затем пытаемся дополнять рассшифрованный текст по осмысленности(по другим свойствам языка)

Атака №2'(Взлом стал труднее, но ещё возможен)

Метод протяжки слов, т.е подставляем в возможные места известные слова, получая фрагмент ключа. Если мы правильно подставили вероятные слова, то дальше подбираем ключ по осмысленности.

Пример от Ананичева: Если есть Террористы, которые переписываются с шифром простой замены, то скорее всего они говорят слова "Бомба" или "Алах ак бар". Эти слова можно протягивать сквозь текст, получая ключ

Атака №1

- **Частотный криптоанализ (подходит для длинных текстов $>$ 3000 символов)**.
    1. В зашифрованном тексте подсчитывают, какие символы встречаются чаще.
    2. Сравнивают с частотностью букв в языке (например, в русском чаще всего "О", "Е", "А").
    3. Подбирают замену, пока текст не станет осмысленным.

Для некоторых текстов, частоты могут меняться, например перед тем как зашифровать текст, можно удалить из него все знаки препинания, пробелы.

Также если текст - тематический, то наиболее вероятные слова могут сместить частоты.

Нотесравил - гласные и согласные идут по убыванию частот. Гласные стоят раньше согласных

Частоты букв в русском языке

буква

ранг

употреблений

частотность

графика

а

3

40 487 008

8,01%

б

21

8 051 767

1,59%

в

9

22 930 719

4,54%

г

19

8 564 640

1,70%

д

13

15 052 118

2,98%

е

2

42 691 213

8,45%

ё

33

184 928

0,04%

ж

25

4 746 916

0,94%

з

20

8 329 904

1,65%

и

4

37 153 142

7,35%

й

23

6 106 262

1,21%

к

11

17 653 469

3,49%

л

10

22 230 174

4,40%

м

12

16 203 060

3,21%

н

5

33 838 881

6,70%

о

1

55 414 481

10,97%

п

14

14 201 572

2,81%

р

8

23 916 825

4,73%

с

7

27 627 040

5,47%

т
6
31 620 970
6,26%

у
15
13 245 712
2,62%

ф
31
1 335 747
0,26%

х
24
4 904 176
0,97%

ц
28
2 438 807
0,48%

ч
22
7 300 193
1,44%

ш

26
3 678 738
0,73%

щ
29
1 822 476
0,36%

ъ
32
185 452
0,04%

ы
17
9 595 941
1,90%

ь
18
8 784 613
1,74%

э
30
1 610 107
0,32%

ю
27
3 220 715
0,64%

я

16

10 139 085

2,01%

Если текст короткий, то: * Делаем много разных гиппотез о возможных соотвествиях букв открытого текста буквам из криптограммы

- Чтобы сократить кол-во гиппотез, можно использовать частоты биграм, триграм и т.д сравнив их с их реальными частотами
- Также можно использовать информацию о паросочетаниях букв
  - ГГ - встречается с частотой 0.065
  - ГС - встречается с частотой 0.383
  - СГ - встречается с частотой 0.383
  - СС - встречается с частотой 0.168