One-Time Pad

The sender uses each key letter to encrypt exactly one plaintext character.

Encryption is the addition modulo 26 of the plaintext character and the

one-time pad key character.

The sender encrypts the message and then destroys the used pages of the pad or used section of the tape. The receiver has an identical pad and uses each key on the pad, in turn, to decrypt each letter of the ciphertext. The receiver destroys the same pad pages or tape section after decrypting the message.

ONETIMEPAD

and the key sequence from the pad is TBFRGFARFM

then the ciphertext is IPKLPSFHGQ

because

 $O + T \mod 26 = I$ $N + B \mod 26 = P$ $E + F \mod 26 = K$ etc.

Binary case:

 $\begin{array}{rcl} x = & 01001101 \ 01011101 \ \dots \\ k = & 11010000 \ 11101011 \ \dots \\ y = & 10011101 \ 10110110 \ \dots \end{array}$

y + k = (x + k) + k = x + (k + k) = x

+ : mod 2 addition (XOR) **Probabilistic model (C. Shannon)**

X, Y, K random variable K has uniform distribution (coin flipping sequence) X and K are independent $Y = E_K(X)$

Perfect encryption: I(X,Y)=0.

Theorem 1: Perfect encryption exists.

Proof: One time pad

 $Y=X+K \quad (+=\oplus)$

 $P(Y=y|X=x) = P(X+K=y|X=x) = P(K=y-x|X=x) = P(K=y-x) = 2^{-N}$

(X and K are independent)

$$P(Y=y) = Σx P(X+K=y|X=x)P(X=x) = 2-N = P(Y=y|X=x)$$

◆

Theorem 2: $H(K) \ge H(X)$

$$\begin{split} H(X) &= H(X|Y) + I(Y;Y) = H(X|Y) \\ H(X|Y) &\leq H(X,K|Y) = H(K|Y) + H(X|Y,K) \\ &= H(K|Y) \leq H(K) \\ (H(U,V) = H(U) + H(V|U), H(X|Y,K)=0) \blacklozenge \end{split}$$

Corrollary: $H(X) \le H(K) = H(K_1, K_2, ..., K_N) \le |K|$

Use data compession before applying one time pad encryption!

Consider the case when |P|=|C|=|K|, i.e. the size of message space, key space and ciphertext space is the same.

Theorem 3: Assume |P| = |C| = |K|. The cryptosystem is perfect

if and only if

- A1.) every key is used with the same probability (1/|K|), and
- A2.) for every message **x** and cipertext **y** there exists a unique key **k**, such that $E_k(x)=y$.

Proof:

- 1. A1, A2 \rightarrow perfect (similar proof as for Th.1.)
- 2. perfect \rightarrow A1, A2:

perfect \rightarrow for each x,y there exists at least one k such that $E_k(x)=y$

|C| = |K|

 \rightarrow A2.

fix y

Bayes th. $\rightarrow P(x_i|y) = P(y|x_i)P(x_i)/P(y) = P(k_i)P(x_i)/P(y)$ (E_{ki} (x_i)=y_i)

 $perfect \rightarrow P(x_i|y) = P(x_i)$

 $\rightarrow 1 = P(k_i)/P(y)$

 \rightarrow P(k_i) is constant \rightarrow P(k_i)= 1/|K|.

Classical simple cryptosystems

- 1. Shift Cipher
- 2. Affine Cipher
- 3. Substitution Cipher
- 4. Vigenere Cipher
- 5. Hill Cipher
- 6. Permutation Cipher
- 7. Stream Cipher (with LFSR)

1. Shift Cipher

 $P = C = K = Z_{26}$

|K|=26

 $E_k(x)=x+k \mod 26$

 $D_k(y)=y-k \mod 26$

Character conversion (preprocessing):

					<u>`I</u> .			<u> </u>
Α	В	C	•••	W	Х	Y	Ζ	
0	1	2		22	23	24	25	

Example encryption: x=wewillmeet, k=11 y=HPHTWWXPPE attack: exhaustive key search (meaningful plaintext)

2. Affine Cipher

 $P=C=Z_{26}$ $K=\{(a,b)\in Z_{26}xZ_{26}: gcd(a,26)=1),$ $|K|=26\cdot\phi(26)=26\cdot12=312$ k=(a,b) $E_{k}(x)=ax+b \mod 26$ $D_{k}(y)=a^{-1}(y-b) \mod 26$

 $\mathbf{x}, \mathbf{y} \in \! \mathbf{Z}_{\!26}$

Example:

 $k=(7,3), 7^{-1}=15 \pmod{26} (7.15=105=4.26+1)$

 $E_k(x) = 7x + 3 \mod 26$

 $D_k(y)=15(y-3)=15y -19 \mod 26$

cryptanalysis:

lette	prob.	letter	prob.
r			
Α	.082	N	.067
В	.015	Ο	.075
С	.028	Р	.019
D	.043	Q	.001
Е	.127	R	.060
F	.022	S	.063
G	.020	Т	.091
Н	.061	U	.028
Ι	.070	V	.010
J	.002	W	.023
K	.008	X	.001
L	.040	Y	.020
Μ	.024	Ζ	.001

Letter probability distribution in English texts

y=FMXVEDKAPHFERBNDKRXRSREFMORUDSDKDVSHV UFEDKAPRKDLYEVLRHHRH

Letter frequency in ciphertext y

letter	freq.	letter	freq.
А	2	Ν	1
В	1	0	1
С	0	Р	2
D	7	Q	0
Е	5	R	8
F	4	S	3
G	0	Т	0
Η	5	U	2
Ι	0	V	4
J	0	W	0
K	5	X	2
L	2	Y	1
Μ	2	Ζ	0

R(8),

D(7),

E,H,K(5),

F,V(4)

R(8), D(7), E,H,K(5), F,V(4)

guess:

 $R \to e, \, D \to t$

```
E_k(4)=17

E_k(19)=3

1. 4a+b=17 mod 26

2. 19a+b=3 mod 26

a=6, b=19

(2.-1.: 15a=-14=12, 15<sup>-1</sup>=7, a=7.12=6 mod 26)

gcd(a,26)=2>1 incorrect guess
```

next guess: $R \rightarrow e, E \rightarrow t \rightarrow a=13$ incorrect

next guess: $R \rightarrow e, H \rightarrow t \rightarrow a=8$ incorrect

next guess: $R \rightarrow e, K \rightarrow t \rightarrow a=3, b=5$ legal key

decryption trial with (check if we get *meaningful decrypted text*) $D_k(y)=3^{-1}(y-5)=9y-19 \mod 26$

algorithms are quite general definitions of arithmetic processes

algorithms are quite general definitions of arithmetic processes

result of the analysis: k=(3,5) is the correct key.

3. Substitution Cipher

 $P=C=Z_{26}$ K={all possible permutations of the ABC}, |K|=26!

Example permutation π : (selected randomly from K)

a	b	c	•••	W	X	у	Z
Х	Ν	Y		Κ	J	D	Ι

 $E_{\pi}(x)=\pi(x)$

 $D_{\pi}(y) = \pi^{-1}(y)$

attack (cryptanalysis): letter frequency digram frequency trigram frequency

4. Vigenere Cipher

Extension of shift cipher (substitution cipher) to a m parallel shifts (substitutions)

keyword: key of **m** characters (letters): **k**=(**k**₁,**k**₂,...,**k**_m)

periodical substitution of characters with period m:

the 1st char. of x is ciphered with the 1st char. of k the 2nd char. with the 2nd char. of k ... the m-th char. with the m-th char. of k the m+1-th char. with the 1-st char. of k ...

 $|K|=26^{m}$

 \rightarrow exhaustive key search is not feasible for not too small m

cryptanalysis:

Step 1.: determination of keyword length m (Kasiski test, Friedman test)

Step 2: finding the shifts (key chars) (Friedman test)

Step 1: 1.1. The Kasiski test:

The idea:

i.) two identical segments of plaintext are encrypted to the same ciphertext, if their distance within the text is a multiple of m ii.) conversely, if we see two identical segments of ciphertext each with length at least 3 chars., there is a good chance that they correspond to identical segments of plaintext

The test: search the ciphertext for pairs of identical segments of length at least 3, and record the distances. Calculate the gcd of the distances.

1.2. The Friedman test (index of coincidences):

Assume we have a string z of letters of length n, where the frequency of letters (A,B,C...Z) is $f_0,f_1,...,f_{25}$. The probability $I_c(x)$ of the event that selecting two randomly selected elements from x are identical is the following:

$$I_{c}(z) = \frac{\sum_{i=0}^{25} f_{i}(f_{i}-1)}{n(n-1)} \approx \sum_{i=0}^{25} p_{i}(=0.0.65 \text{ for English})$$

where \mathbf{p}_{i} is the letter frequency of the actual language for the *i*-th letter in the alphabet. For random text

$$\mathbf{p_i}=26 \cdot (1/26)^2 = 1/26 \rightarrow \mathbf{I_c}(\mathbf{x}) \approx 1/26 = 0.038.$$

The test: Take the letters from the ciphertext at distances m=1,2,3,... and calculate the index of coincidences for the obtained

substring(m). For the right guess of m index value close to 0.065 is obtained.

Step 2: Finding the key

Mutual index of coincidences:

Let **z** and **z'** two strings of letters with length **n** and **n'** respectively. The corresponding letter frequencies are f_0, f_1, \dots, f_{25} and $f'_0, f'_1, \dots, f'_{25}$.

The probability of the event, that a randomly selected letter from z is identical to a randomly selected letter from z' is the following:

$$MI_{c}(z,z') = \frac{\sum_{i=0}^{25} f_{i} \cdot f_{i}'}{n \cdot n'} \approx \sum_{i=0}^{25} p_{i} p_{i}'$$

We have a guess on $\mathbf{k}=(\mathbf{k}_1,\mathbf{k}_2,\ldots,\mathbf{k}_m)$. Let this guess be $\mathbf{K}=(\mathbf{K}_1,\mathbf{K}_2,\ldots,\mathbf{K}_m)$.

Consider the pair K_1 , K_2 . Take two substrings of the ciphertext: ${\boldsymbol{z}}$ and ${\boldsymbol{z}}$, where

z consist of the letters at positions 1,m+1,2m+1,...,im+1,...

z' consist of the letters at positions 2,m+2,2m+2,...,im+2,....

Calculate the mutual index of coincidences for the pair z, z':

$$MI_{c}(z, z') \approx \sum_{l=0}^{25} p_{l-K_{1}} p_{l-K_{2}} = \sum_{l=0}^{25} p_{l} p_{l+K_{1}-K_{2}}$$

Note: Only the relative shift is important!

Calculate the relative shifts for all pairs of characters of the guessed key (i.e. for m(m-1)/2 pairs). 26 possible keys remain, resolve with exhaustive search.

Homework: Write a program for analysis of Vigenere encryption.

5. Hill Cipher

 $P=C=(Z_{26})^{m}$

Linear block cipher transforming m character long plaintext into m character long ciphertext by matrix multiplication:

 $E_k(x)=xK \mod 26$

 $D_k(y)=yK^{-1} \mod 26$

where the key k is an mxm invertible matrix over Z_{26}

attack (cryptanalysis):

known plaintext attack

x=friday, m=2, y=PQCFKU

$fr \rightarrow PQ$ id $\rightarrow CF$ ay $\rightarrow KU$:	$E_k(5,17) = E_k(8,3) = 0$ $E_k(0,24) = 0$	=(15,16) (2,5) =(10,20)
$\begin{pmatrix} 15 & 16 \\ 2 & 5 \end{pmatrix} =$	$\begin{pmatrix} 5\\ 8 \end{pmatrix}$	$ \begin{array}{c} 17 \\ 3 \begin{pmatrix} k_{11} \\ k_{21} \end{pmatrix} $	$egin{array}{c} k_{12} \ k_{22} \end{array}$

$$\begin{pmatrix} 5 & 17 \\ 8 & 3 \end{pmatrix}^{-1} = \begin{pmatrix} 9 & 1 \\ 2 & 15 \end{pmatrix} \rightarrow K = \begin{pmatrix} 9 & 1 \\ 2 & 15 \end{pmatrix} \begin{pmatrix} 15 & 16 \\ 2 & 5 \end{pmatrix} = \begin{pmatrix} 7 & 19 \\ 8 & 3 \end{pmatrix}$$

6. Stream Cipher (with LFSR)

 $y_i = x_i + k_i$ i=1,2,..., where y_i, x_i, k_i are binary variables

LFSR: Linear Feedback Shift Register

Defined by the following linear recursion:

$$k_{m+i} = \prod_{j=0}^{m-1} f_j k_{i+j} \mod 2, \quad i=0,1,2,\dots$$

 k_0, k_1, \dots, k_{m-1} initial state (random seed)

 f_0, f_1, \dots, f_{m-1} feedback coefficients

e.g.

$$k_m = f_0 k_0 + f_1 k_1 + \dots + f_{m-1} k_{m-1}$$

 $k_{m+1} = f_0 k_1 + f_1 k_2 + \dots + f_{m-1} k_m$
...

attack (cryptanalysis):

known plaintext attack System of m linear equations can be obtained in unknown feedback coefficients.

$$(k_m, k_{m+1}, \dots, k_{2m-1}) = (f_0, f_1, \dots, f_{m-1}) \begin{pmatrix} k_0 & k_1 & \dots & k_{m-1} \\ k_1 & k_2 & \dots & k_m \\ \dots & \dots & \dots & \dots \\ k_{m-1} & k_m & k_{2m-2} \end{pmatrix}$$

Unicity distance

$$K(y) = \{K : \exists x, E_K(x) = y\}$$
$$S_n = p(y)(|K(y)| - 1)$$
$$y \in C^n$$

Theorem 3: If |P| = |C|

$$S_n \ge \frac{\mid K \mid}{\mid P \mid^{nR_L}} - 1$$

where R_L is the redundancy per letter of the language.

Unicity distance: n_0 , such that $S_{n0}=0$

$$n_0 \approx \frac{\log_2 |K|}{R_L \log_2 |P|}$$

Example: substitution cipher

 $|\mathsf{P}|=26$, $|\mathsf{K}|=26!$, $\mathsf{R}_{\mathrm{L}}=0.75$ (English) \rightarrow $n_0 \approx 88.4/(0.75x4.7) \approx 25$. Given a cipheetext string of length at least 25, (usually) unique decryption is possible.