**CLASSICAL ENCRYPTION TECHNIQUES**

**Plaintext**- original message

**Ciphertext** – coded message

**Enciphering, encryption** – process of converting from plaintext to ciphertext

**Deciphering, decryption** – restoring the plaintext from the ciphertext

**Cryptography** – area of study schemes for enciphering

**Cryptographic system, cipher** – scheme of enciphering

**Cryptanalysis** – techniques for deciphering a message without knowledge of the enciphering details

**Cryptology** – areas of cryptography and cryptanalysis

**Symmetric cryptography**



**From W. Stallings**



**Fig. 2.2. Key sharing and cryptanalysis (from W. Stallings)**

We can write:

Y=EK(X)

X= DK(Y)

Opponent knows Y, E, D. He may be interested to recover X or/and K. Knowledge of K gives him opportunity to read future messages.

Cryptographic systems are characterized by

1. The type of operations used for transforming plaintext to ciphertext (substitution, transposition). Fundamental requirement – no information be lost
2. The number of keys used (1 key – symmetric, single-key, secret-key; 2 keys – asymmetric, two-key, public-key)
3. The way in which the plaintext is processed (block cipher, stream cipher). Stream cipher may be viewed as a block cipher with block size equal to 1 element.

**Unconditionally secure encryption scheme** – ciphertext generated by the scheme does not contain enough information to determine uniquely the corresponding plaintext, no matter how much ciphertext is available. Excepting a scheme known as one-time pad, there is no encryption algorithm that is unconditionally secure. Therefore, encryption algorithm should meet one or both of the following criteria:

* + The cost of breaking the cipher exceeds the value of the encrypted information
	+ The time required to break the cipher exceeds the useful lifetime of the information

Such algorithm is called **computationally secure**. Table below shows how much time is involved for various key sizes. The 56-bit key size is used with the DES (Data Encryption Standard), 168-bit – for triple DES, 128-bit – for AES (Advanced Encryption Standard). Results are also shown for substitution codes that use 26-character key, in which all possible permutations of the 26 characters serve as keys. It is assumed that it take 1 μs to perform a single decryption or encryption (in last column – 106 decryptions per 1 μs)



Substitution ciphers

The Caesar cipher involves replacing each letter of the alphabet with the letter standing three places further down the alphabet

For example

*Plain: meet me after the toga party*

*Cipher: PHHW PH DIWHU WKH WRJD SDUWB*

Transformation is made using the following mapping:

*Plain: 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*

*Cipher: D 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*

With only 25 keys Caesar cipher is far from secure. A dramatic increase in the key space may be achieved by allowing an arbitrary substitution.Then there are 26! or greater than 4\*1026 possible keys. There is however another line of attack. If the cryptanalyst knows the nature of the plaintext (e.g., non-compressed English text), then the analyst can exploit the regularities of the language. Another line of attack is statistical analysis. Relative frequency of the letters can be determined and compared to a standard frequency distribution for English:



(from W. Stallings)

For the ciphertext

UZQSOVUOHXMOPVGPOZPEVSGZWSZOPFPESXUDBMETSXAIZ

VUEPHZHMDZSHZOWSFPAPPDTSVPQUZWYMXUZUHSX

EPYEPOPDZSZUFPOMBZWPFUPZHMDJUDTMOHMQ

the relative frequencies of the letters in the ciphertext (in percentages) are

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| P 13.33Z 11.67S 8.33U 8.33O 7.50M 6.67 | H 5.83D 5.00E 5.00V 4.17X 4.17 | F 3.33W 3.33Q 2.50T 2.50A 1.67 | B 1.67G 1.67Y 1.67I 0.83J 0.83 | C 0.00K 0.00L 0.00N 0.00R 0.00 |

Comparing this with Fig.2.5, it seems likely that cipher letters P and Z are the equivalents of plain letters e and t, but it is not certain which is which. The letters S,U,O,M, and H are all of the relatively high frequency and probably correspond to plain letters from the set {a,h,i,n,o,r,s}. The letters with the lowest frequencies (A,B,G,Y,I,J) are likely included in the set {b,j,k,q,v,x,z}.

Also, statistics on bigrams, trigrams may be used for statistical cryptanalysis.

**Hill Cipher**

It was developed by the mathematician Lester Hill in 1929. The encryption algorithm takes m successive plaintext letters and substitutes for them m ciphertext letters. The substitution is determined by m linear equations in which each character is assigned a numerical value:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 |
| 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 |

For m=3, the system can be described as follows:

C1=(k11p1+k12p2+k13p3) mod 26

C2=(k21p1+k22p2+k23p3) mod 26

C3=(k31p1+k32p2+k33p3) mod 26

This can be expressed in terms of column vectors and matrices:

C=KP mod 26,

where C and P are column vectors of length 3, representing the plaintext and ciphertext, and K is 3x3 matrix, representing the encryption key.

Operations are performed mod 26.

For example, consider the plaintext “paymoremoney”, and use the encryption key

K=

|  |  |  |
| --- | --- | --- |
| 17 | 17 | 5 |
| 21 | 18 | 21 |
| 2 | 2 | 19 |

The first 3 letters of the plaintext are represented by the vector (15 0 24). Then K(15 0 24) = (375 819 486) mod 26 = (11 13 18) = LNS. Continuing in this fashion, the ciphertext for the entire plaintext is LNSHDLEWMTRW.

Decryption requires using the inverse of the matrix K. The inverse K-1 of a matrix K is defined by K K-1 = K-1 K=I, where I is the unit matrix (1-s on the diagonal, other elements – zeroes). The inverse of the matrix does not always exist, but when it does, it satisfies the preceding equation. In this case, the inverse is

K-1=

|  |  |  |
| --- | --- | --- |
| 4 | 9 | 15 |
| 15 | 17 | 6 |
| 24 | 0 | 17 |

This is demonstrated as follows:

K K-1 =

|  |  |  |
| --- | --- | --- |
| 443 | 442 | 442 |
| 858 | 495 | 780 |
| 494 | 52 | 365 |

And after taking mod 26 of the elements above, unit matrix is obtained.

In general terms, the Hill system can be expressed as follows:

C=EK(P)=KP mod 26

P= DK(C)=K-1C mod 26 = K-1KP = P

As with Playfair, the strength of the Hill cipher is that it completely hides single-letter frequencies.

Although the Hill cipher is strong against a ciphertext-only attack (opponent has only ciphertext), it is easily broken with a known plaintext attack (opponent has pairs plaintext – ciphertext). For an m\*m Hill cipher, suppose we have m plaintext-ciphertext pairs, each of length m. We label the pairs Pj=(p1j, p2j,…, pmj) and Cj=(c1j, c2j,…, cmj) such that Cj=KPj for 1<=j<=m and for some unknown key matrix K. Now define two m\*m matrices X=( pij) and Y=( cij).

Then we can form matrix equation Y=KX. If X has an inverse, then we can determine K=YX-1. If X is not invertible, then a new version of X can be formed until an invertible X is obtained.

Suppose that the plaintext “friday” is encrypted using a 2\*2 Hill cipher to yield the ciphertext PQCFKU. Thus, we know that

K(5 17) = (15 16);

K(8 3) = (2 5);

K(0 24) = (10 20).

Using the first 2 plaintext-ciphertext pairs, we have



The inverse of X can be computed:





Let’s check now that this key matrix produces required transformation:







**One-time pad**

An US Army Signal Corps Captain, Joseph Mauborgne, in 1918, proposed an improvement (<http://en.wikipedia.org/wiki/One-time_pad> ) to Vernam cipher that yields the ultimate in security. He suggested using of a random key that was truly as long as the message, with no repetitions. Such a scheme, known as one-time pad, is unbreakable. It produces random output that bears no statistical relationship to the plaintext. Because the ciphertext contains no information whatsoever about the plaintext, there is no way to break the code.

Let’s consider example. Suppose that we are using a Vigenere scheme with 27 characters in which the 27th character is the space character, but with a one-time key that is as long as the message. Thus, Vigenere tableau must be expanded to 27x27. Consider the ciphertext:

ANKYODKYUREPFJBYOJDSPLREYIUNOFDOIUERFPLUYTS

We now show 2 different decryptions using 2 different keys:

Ciphertext: ANKYODKYUREPFJBYOJDSPLREYIUNOFDOIUERFPLUYTS

Key: px l mv ms yd o f t y rv zwc t nl e b n ecv g du p a hf zz l mn y ih

Plaintext: mr mu s t ar d with th e can d l est i ck i n th e h a ll

and

Ciphertext: ANKYODKYUREPFJBYOJDSPLREYIUNOFDOIUERFPLUYTS

Key: mf u g p mi yd g axg ou f hkl l l mhs qd qo g t e wbqf q yov u h wt

Plaintext: mi s s s c ar l e t wi t h t he kn if e i n t he l ib r a r y

If cryptanalyst would manage to find these keys, he will not be able to decide which key is correct, and which plaintext is correct. Randomness of the key leads to randomness of the ciphertext, and therefore cryptanalyst can’t use patterns or regularities to attack the ciphertext, the code is unbreakable.

But in practice, one-time pad has 2 fundamental difficulties:

* there is the practical problem of making large quantities of random keys. Any heavily used system might require millions of random characters on a regular basis. Supplying truly random characters in this volume is a significant task
* even more daunting is the problem of key distribution and protection. For every message to be sent, a key of equal length is needed by both sender and receiver. Thus, a mammoth key distribution problem exists.

**Transposition technique**

Another approach to enciphering is usage of transpositions, or permutations on the plaintext letters. The simplest such cipher is the rail fence technique, in which the plaintext is written down as a sequence of diagonals and then read off as the sequence of rows. For example, to encipher the message “meet me after the toga party” with a rail fence of depth 2, we write

m e m a t r h t g p r y

 e t e f e t e o a a t

The encrypted message is

MEMATRHTGPRYETEFETEOAAAT

A more complex scheme is to write the message in a rectangle, row by row, and read the message off, column by column, but to permute the order of columns. The order of columns then becomes the key to the algorithm. For example,

Key: 4 3 1 2 5 6 7

Plaintext: a t t a c k p

 o s t p o n e

 d u n t i l t

 w o a m x y z

Ciphertext: TTNAAPTMTSUOAODWCOIXKNLYPETZ

A pure transposition cipher is easily recognized because it has the same letter frequencies as the original plaintext.

The transposition cipher can be made more secure by performing more than 1 transposition

**Steganography**

The methods of steganography conceal the existence of the message, whereas the methods of cryptography render the message unintelligible to outsiders by various transformations of the text.

Use of:

The 1st letters of some text spells out the hidden message

Character marking (by pencil, pin punctures) of necessary letters in the text

Invisible inks

Typewriter correction ribbon: print between lines printed with a black ribbon – visible only under a strong light

Least significant bit of each 24-bit pixel

Steganography requires a lot overhead to hide a relatively few bits of information. Once the system is discovered, it becomes worthless. But a message can be at first encrypted, then hidden.

The advantage of steganography is that it can be employed by parties wanting to hide the fact of their secret communication

**Diffusion and confusion**

These are measures to thwart cryptanalysis based on statistical analysis. In diffusion, the statistical structure of the plaintext is dissipated into long range statistics of the ciphertext. This is achieved by having each plaintext letter affect the value of many ciphertext digits, which is equivalent to saying that each ciphertext digit is affected by many plaintext digits. An example of diffusion is to encrypt a message M=m1,m2,m3,.. of characters with an averaging operation:



adding k successive letters to get a ciphertext letter yn. The letter frequencies in the ciphertext will be more nearly equal than in the plaintext (structure dissipated).

Confusion seeks to make the relationship between the statistics of the ciphertext and the value of the encryption key as complex as possible. This is achieved by the use of a complex substitution algorithm. These operations became the cornerstone of modern block cipher design.