Chapter 3

Traditional Symmetric-Key Ciphers

傳統對稱式金鑰加密法

又稱為Conventional /Secret-key / Single-key Encryption

Chapter 3

Objectives

❑ To define the terms and the concepts of symmetric key ciphers

❑ To emphasize the two categories of traditional ciphers: substitution and transposition ciphers

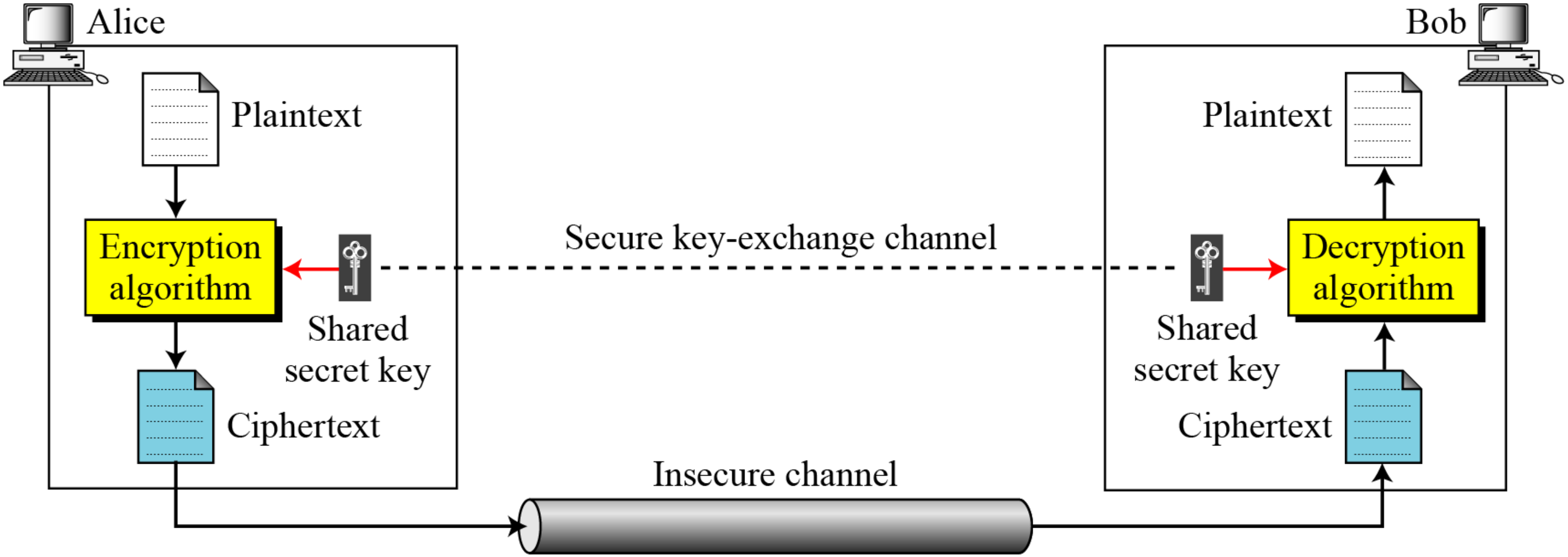
❑ To describe the categories of cryptanalysis used to break the symmetric ciphers

❑ To introduce the concepts of the stream ciphers and block ciphers

❑ To discuss some very dominant ciphers used in the past, such as the Enigma machine

3.1 Continued

Figure 3.1 ***General idea of symmetric-key cipher***



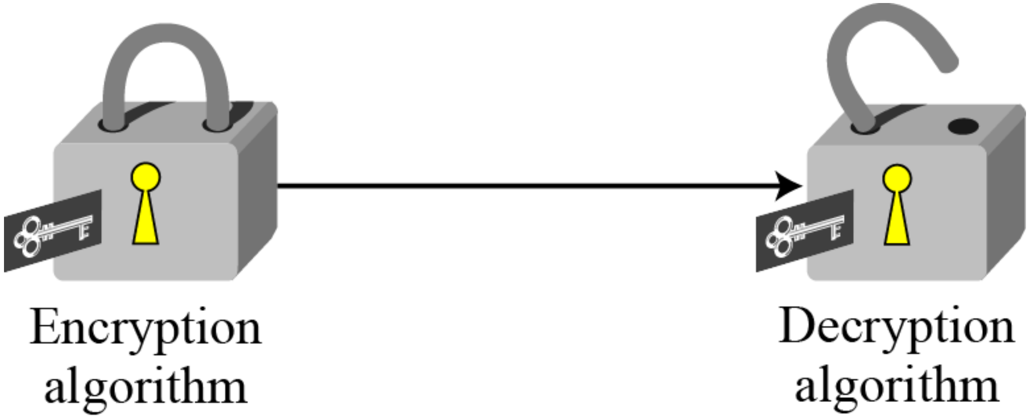
3.1 Continued

If P is the plaintext, C is the ciphertext, and K is the secret key,

Encryption: $C = E_k(P)$ Decryption: $P = D_k(C)$

In which, $D_k(E_k(x)) = E_k(D_k(x)) = x$

Figure 3.2 ***Locking and unlocking with the same key***



- *對稱式密碼系統有金鑰的管理問題，例如Alice若要與N個人做秘密通訊，那麼Alice就必須握有N把秘密金鑰*

3.1.1 Kerckhoff’s Principle 柯克霍夫斯原理

Based on Kerckhoff’s principle, one should always assume that the adversary(敵人), Eve, knows the encryption/decryption algorithm.

The resistance of the cipher to attack must be based only on the secrecy of the key.

Secret key cryptosystems vs Public key Cryptosystems

- *If k1=k2 ➔ symmetric key cryptosystem (對稱), one-key system, secret key system (私密密碼系統)*
  - *例子: DES, AES, webmail ID/ password, GSM pin*
  - *Authentication, privacy, integrity*
  - *缺點: 金鑰分配, 金鑰管理, 無法達成不可否認性*

- *If k1≠k2 ➔ asymmetric key cryptosystem (非對稱式), two-key system, public key system*
  - *例子: RSA, Elgamal, D-H key, ECC*
  - *Authentication, privacy, integrity, non-repudiation*
  - *缺點: 速度慢*

English

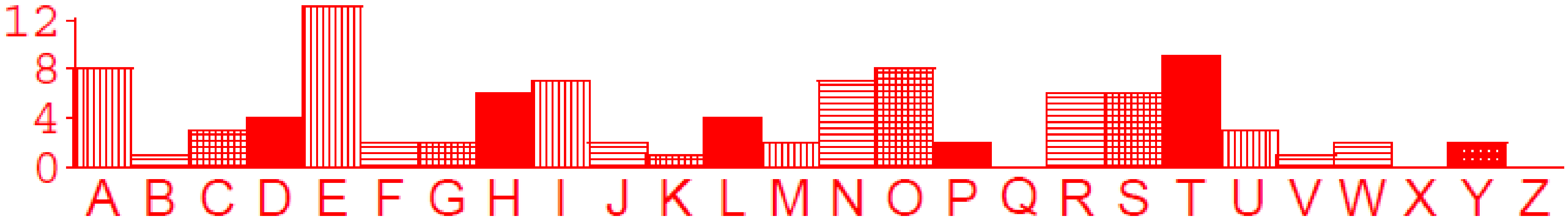
All of the examples in this section will have English plaintext

To break these early ciphers we make use of the statistics of English

In a later section on Information Theory we shall study this relationship between

- How easy the cipher is to break
- The underlying plaintext distribution

English Letter Frequencies



The most common bigrams are, in decreasing order,

- TH, HE, IN, ER, AN, RE, ED, ON, ES, ST, EN, AT, TO, NT, HA

The most common trigrams are, in decreasing order,

- THE, ING, AND, HER, ERE, ENT, THA, NTH, WAS, ETH, FOR

Shift Cipher

Each letter we identify with a number

- $A = 0$
- $B = 1$
- $C = 2$
- ...
- $Z = 25$

C= m +k mod 26

The key k is a number in the range $0 - 25$

- Encryption is add k onto each letter modulo 26.

Julius Caeser used the key $k = 3$.

- HELLO becomes KHOOR

We break a Shift cipher by using the statistics of the underlying language

Shift Cipher

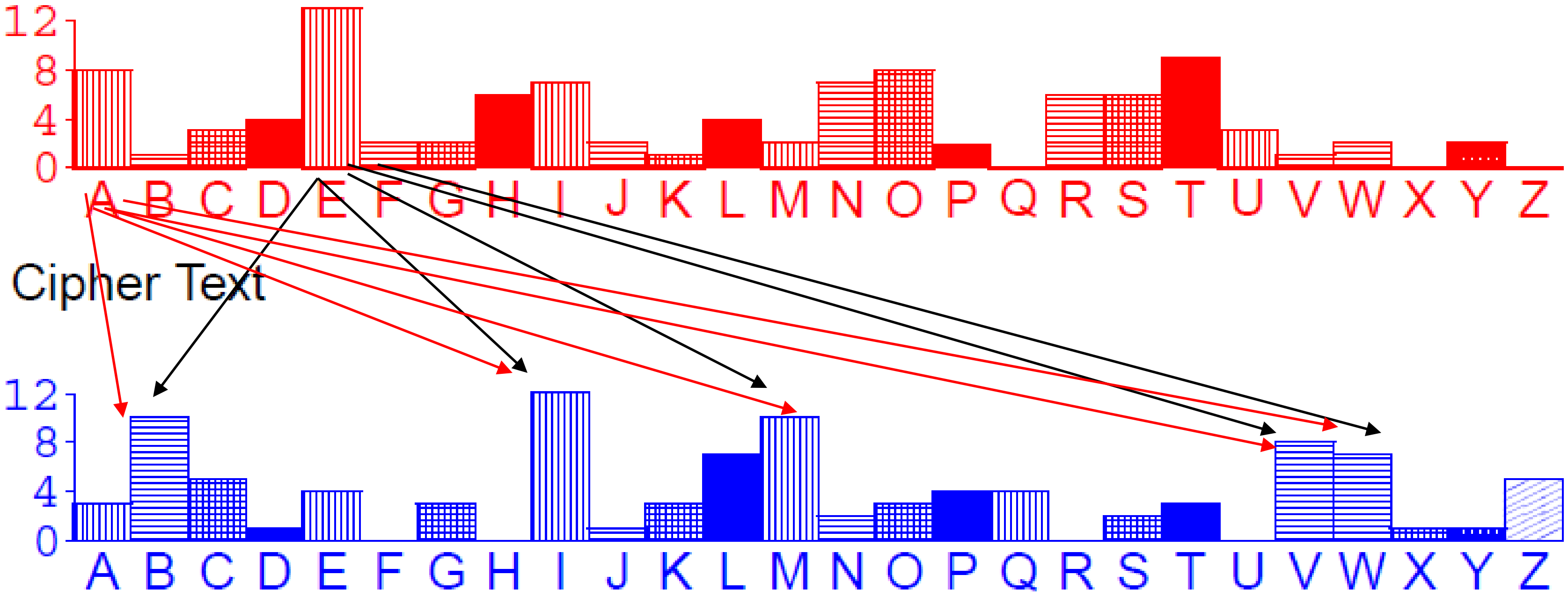
Take the following example cipher text

BPMZM WVKM EIA IV COTG LCKSTQVO
EQBP NMIBPMZA ITT ABCJJG IVL JZWEV
IVL BPM WBPMZ JQZLA AIQL QV AW UIVG EWZLA
OMB WCB WN BWEV
OMB WCB, OMB WCB, OMB WCB WN BWEV
IVL PM EMVB EQBP I YCIKS IVL I EILLTM IVL I YCIKS
QV I NTCZZG WN MQLMZLWEV
BPIB XWWZ TQBBTM COTG LCKSTQVO
EMVB EIVLMZQVO NIZ IVL VMIZ
JCB IB MDMZG XTIKM BPMG AIQL BW PQA NIKM
VWE OMB WCB, OMB WCB, OMB WCB WN PMZM
IVL PM EMVB EQBP I YCIKS IVL I EILLTM IVL I YCIKS
IVL I DMZG CVPIXXG BMIZ

We need to compare the frequency distribution of this text with standard English

Shift Cipher

Underlying Plain Text



The shift of E seems to be either 4, 8, 17, 18 or 23

The shift of A seems to be either 1, 8, 12, 21 or 22

Shift Cipher

Hence the key is probably equal to 8

We can now decrypt the cipher text to reveal

There once was an ugly duckling
With feathers all stubby and brown
And the other birds said in so many words
Get out of town
Get out, get out, get out of town
And he went with a quack and a waddle and a quack
In a flurry of eiderdown
That poor little ugly duckling
Went wandering far and near
But at every place they said to his face
Now get out, get out, get out of here
And he went with a quack and a waddle and a quack
And a very unhappy tear

Substitution Cipher

The problem with the Shift cipher is that the number of keys is too small.

- We only have 26 possible keys

To increase the number of keys a substituion cipher was invented.

Encryption involves replacing each letter by its permuted version.

Decryption involves use of the inverse permutation.

ABCDEFGHIJKLMNOPQRSTUVWXYZ
TMKGOYDSIPELUAVCRJWXZNHBQF

Hence HELLO encrypts to SOLLV

Substitution Cipher

Number of keys is $26! \approx 4.03 \cdot 10^{26} \approx 2^{88}$

- Feasible to only run a computer on a problem which takes under 2^{80} steps.

This is far too large a number to brute force search using modern computers.

Still we can break these ciphers using statistics of the underlying plaintext language.

Example

XSO MJIWXVL JODIVA STW VAO VY OZJVCO'W LTJDOWX KVAKOAXJTXIVAW VY SIDS XOKSAVLVDQ IAGZWXJQ.KVUCZXOJW, KVUUZAIKTXIVAW TAG UIKJVOLOKXJVAIKW TJO HOLL JOCJOWOAXOG, TLVADWIGO GIDIXTL UOGIT, KVUCZXOJ DTUOW TAG OLOKXJVAIK KVUUOJKO. TW HOLL TW SVWXIAD UTAQ JOWOTJKS TAG CJVGZKX GONOLVCUOAX KOAXJOW VY UTPVJ DLVMTL KVUCTAIOW, XSO JODIVA STW T JTCIGLQ DJVHIAD AZUMOJ VY IAAVNTXINO AOH KVUCTAIOW. XSO KVUCZXOJ WKIOAKO GOCTJXUOAX STW KLVWO JOLTXIVAWSICW HIXS UTAQ VY XSOWO VJDTAIWTXIVAW NIT KVLLTMVJTXINO CJVPOKXW, WXTYY WOKVAGUOAXW TAG NIWIXIAD IAGZWXJITL WXTYY. IX STW JOKOAXLQ IAXJVGZKOG WONOJTL UOKSTAIWUW YVJ GONOLVCIAD TAG WZCCVJXIAD OAXJOCJOAOZJITL WXZGOAXW TAG WXTYY, TAG TIUW XV CLTQ T WIDAIYIKTAX JVLO IA XSO GONOLVCUOAX VY SIDS-XOKSAVLVDQ IAGZWXJQ IA XSO JODIVA.

XSO GOCTJXUOAX STW T LTJDO CJVDJTUUO VY JOWOTJKS WZCCVJXOG MQ IAGZWXJQ, XSO OZJVCOTA ZAIVA, TAG ZE DVNOJAUOAX JOWOTJKS OWXTMLIWSUOAXW TAG CZMLIK KVJCVJTXIVAW. T EOQ OLOUOAX VY XSIW IW XSO WXJVAD LIAEW XSTX XSO GOCTJXUOAX STW HIXS XSO KVUCZXOJ, KVUUZAIKTXIVAW, UIKJVOLOKXJVAIKW TAG UOGIT IAGZWXJIOW IA XSO MJIWXVL JODIVA . XSO TKTGOUIK JOWOTJKS CJVDJTUUO IW VJDTAIWOG IAXV WONOA DJVZCW, LTADZTDOW TAG TJKSIXOKXZJO, GIDIXTL UOGIT, UVMILO TAG HOTJTMLO KVUCZXIAD, UTKSIAO LOTJAIAD, RZTAXZU KVUCZXIAD, WQWXOU NOJIYIKTXIVA, TAG KJQCXVDJTCSQ TAG IAYVJUTXIVA WOKZJIXQ.

Statistics

We have the following statistics for single letters

Letter	Freq	Letter	Freq	Letter	Freq
A	8.6995	B	0.0000	C	3.0493
D	3.1390	E	0.2690	F	0.0000
G	3.6771	H	0.6278	I	7.8923
J	7.0852	K	4.6636	L	3.5874
M	0.8968	N	1.0762	O	11.479
P	0.1793	Q	1.3452	R	0.0896
S	3.5874	T	8.0717	U	4.1255
V	7.2645	W	6.6367	X	8.0717
Y	1.6143	Z	2.7802		

Most common bigrams : TA, AX, IA, VA, WX, XS, AG, OA, JO, JV

Most common trigrams : OAX, TAG, IVA, XSO, KVU, TXI, UOA, AXS

Initial Analysis

Since O occures with frequency 11.479 we can guess O = E

Three common trigrams are then

- OAX = E * *
- XSO = * * E

Common similar trigrams in English are (from Stinson pp 25)

- ENT, ETH
- THE

Hence likely to have

- X = T
- S = H
- A = N

From now on we only look at the first two sentances

THE MJIWTVL JEDIVN HTW VNE VY EZJVCE’W LTJDEWT KVNKENTJTTIVNW VY HIDH TEKHNVLVDQ INGZWTJQ. KVUCZTEJW, KVUUZNIKTTIVNW TNG UIKJVELEKTJVNIKW TJE HELL JECJEWENTEG, TLVNDWIGE GIDITTL UEGIT,KVUCZTEJ DTUEW TNG ELEKTJVNIK KVUUEJKE.

This was after the changes...

- O = E, X = T, S = H, A = N

Since T occures as a single letter we must have

- T = I or T = A

T has probability of 8.07175, which is the highest probability left

Therefore, more likely to have

- T = A

The most frequent

- bigram is TA = AN
- trigram is TAG = AN*

Therefore highly likely that

- G = D

THE MJIWTVL JEDIVN HAW VNE VY EZJVCE’W LAJDEWT KVNKENTJATIVNW VY HIDH TEKHNVLVDQ INDZWTJQ. KVUCZTEJW, KVUUZNIKATIVNW AND UIKJVELEKTJVNIKW AJE HELL JECJEWENTED, ALVNDWIDE DIDITAL UEDIA,KVUCZTEJ DAUEW AND ELEKTJVNIK KVUUEJKE.

This was after the changes...

- O = E, X = T, S = H, A = N, T = A, G = D

We now look at two letter words...

- IX = *T
- Therefore I must be one of A,I due to English plaintext
- Already have A.

Hence

- I = I

- XV = T*
- Must have, due to English, V = O

More two letter words...

- VY = O*
- Hence Y must be one of F,N,R due to English
- Already have N
- Y has probability 1.6
- F has probability 2.2
- R has probability 6.0

Hence

- Y =F

- IW = I*
- Therefore W must be one of F,N,S,T
- Already have F,N,T

Hence

- W = S

THE MJISTOL JEDION HAS ONE OF EZJOCE'S LAJDEST KONKENTJATIONS OF HIDH TEKHNOLODQ INDZSTJQ. KOUCZTEJS, KOUUZNIKATIONS AND UIKJOELEKTJONIKS AJE HELL JECJESENTED, ALONDSIDE DIDITAL UEDIA,KOUCZTEJ DAUES AND ELEKTJONIK KOUUEJKE.

This was after the changes...

- O = E, X = T, S = H, A = N, T = A, G = D,
- I = I, V = O, Y = F, W = S

It is then easy to see what the underlying plaintext is

Vigenère Cipher

The problem with the Caeser and Substitution cipher was that each plaintext letter always encrypted to the same ciphertext letter.

Hence underlying statistics of the language could be used to break the cipher.

From the early 1800's onwards cipher designers tried to break this link between the plain and cipher texts.

The most famous cipher from the 1800's is the Vigenère Cipher

- Believed to be unbreakable for a number of years.

Vigenère Cipher

The Vigenère cipher again identifies letters with the numbers $0,\ldots,25$

The secret key is a short sequence of letters (e.g. a word).

Encryption involves adding the plaintext letter to a key letter.

- With the key letters used in rotation.

Thus if the key is SESAME, encryption works as follows,

S: 19th character, key=18; E:5th character, key=4

T	H	I	S	I	S	A	T	E	S	T	M	E	S	S	A	G	E	Message
S	E	S	A	M	E	S	E	S	A	M	E	S	E	S	A	M	E	Keystream
L	L	A	S	U	W	S	X	W	S	F	Q	W	W	K	A	S	I	Ciphertext

This is a polyalphabetic substituion cipher

- A will encrypt to a different letter depending on where it is in the message

Homework1. 一 vegenere cipher 金鑰為 abc, 密文為 hppetv, 請問明文為何

==>

Vigenère Cipher

But Vigenère is easy to break.

Once we have found the length of the keyword then breaking the message is the same as breaking the Shift Cipher a number of times.

Stinson gives a more automated way of breaking the system.

- We shall give a more down to earth way

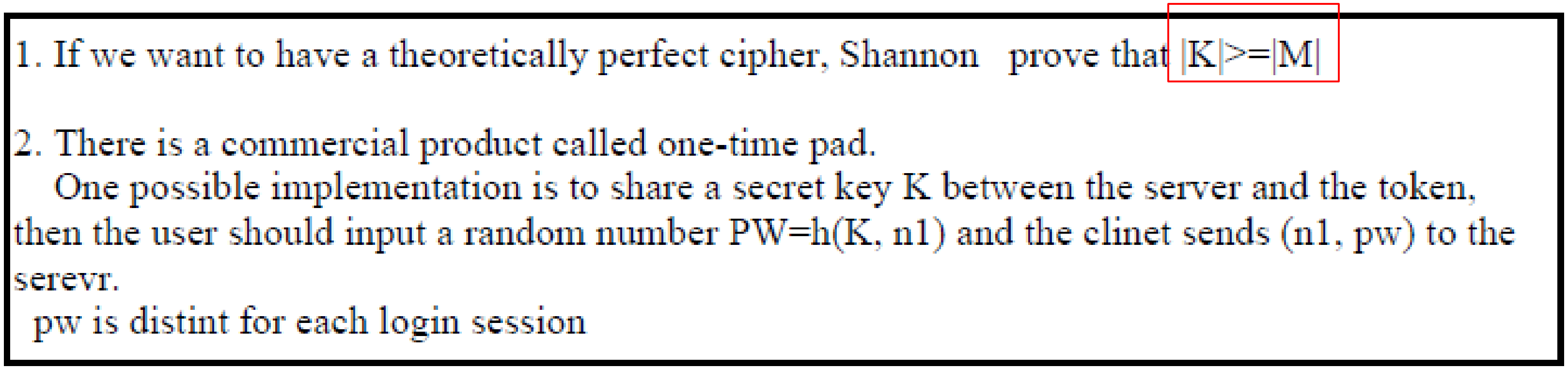
One Time Pad

During the first world war, extensive use was made of the one time pad.

This is often called the Vernam cipher.

We shall study this in more depth later

- It is totally secure if used correctly
- However it is very hard to use correctly.



1. If we want to have a theoretically perfect cipher, Shannon prove that |K|>=|M|

2. There is a commercial product called one-time pad.
One possible implementation is to share a secret key K between the server and the token, then the user should input a random number PW=h(K, n1) and the clinet sends (n1, pw) to the serevr.
pw is distint for each login session

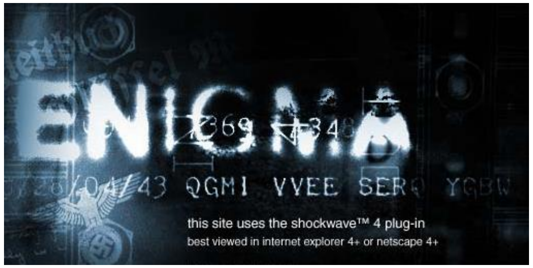
《獵風行動》（Windtalkers）

- 尼可拉斯凱吉的主要任務變成保護亞當比奇飾演的印第安人納瓦荷（Navajo）族密碼通訊兵。
- 原來美軍藉由他們的語言編構了一套日軍無法破解的密碼，因此印第安密碼通訊兵非常重要，但尼可拉斯凱吉真正的命令是「必要時不顧一切保住密碼」以保衛密碼不得外洩。

《獵殺U - 571》美德交戰，密碼決勝

- 這一天，盟軍偵測到在北大西洋上有一艘德軍的受創潛艦U-571，正在發出求救信號，美國海軍派出潛艇S-33偽裝成德國的救援潛艇截獲U-571，並奪取德國海軍Enigma密碼機

<<攔截密碼戰>>



this site uses the shockwave™ 4 plug-in
best viewed in internet explorer 4+ or netscape 4+

- **劇情簡介**
- 在海上有著橫越大西洋的同盟國護航艦隊正運送一萬名乘客及重要的物資，被俄國U型潛艇發現遭到埋伏攻擊。但是德軍修改了原本的密碼，使得原本破解德軍的密碼本失去效果。
- 以英國布雷奇利園區X工作站做背景英國政府網羅了無數優秀的譯碼員，天才數學家湯姆傑瑞柯協助在德軍換了密碼簿（code book） 之後從事破解鯊魚--德軍的「謎」密碼機（Enigma Machine）

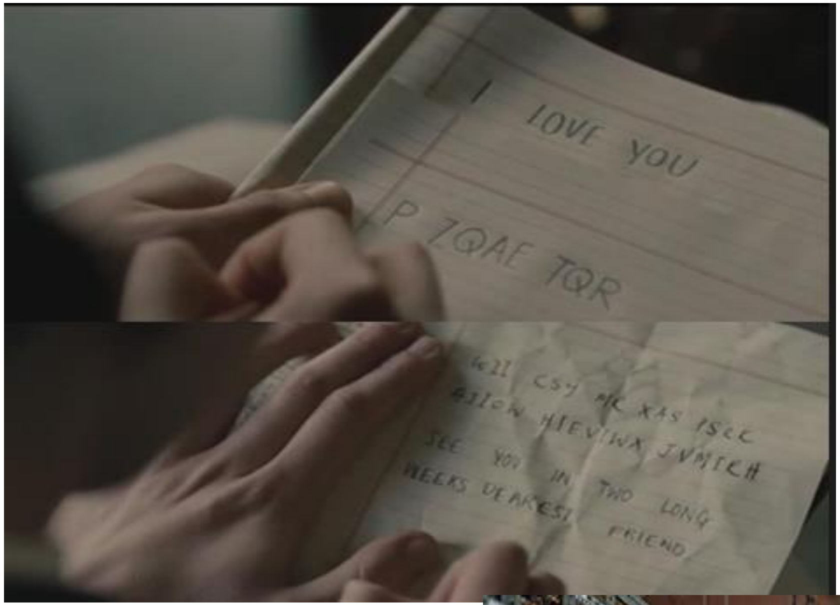
- 「謎」密碼機需要配合密碼簿、旋轉盤 （Rotors）以及接線板（Plugboard）的設定。每次按下的明文經過編譯後所得到的密文都會不同
- http://enigmaco.de/enigma/enigma.html

@www.kingnet.com.tw 照片提供/遠拓

模仿遊戲 (The Imitation Game, 2014)

- 《模仿遊戲》係改編「電腦之父」英國數學家與邏輯學家艾倫·圖靈(Alan Turing)傳奇一生.
- 描述二戰爆發後，德國納粹發明軍用密碼機Enigma「**英格瑪機**」，可將所有機密轉換成亂碼發送，被認為是最難破解的機器。英國參戰在布萊切利莊園（Bletchley Park）設置秘密基地，聚集科學家努力破解Enigma，Turing受政府延攬從事破解德軍密碼工作--幫助聯軍破解德軍潛艇和最高指揮部的通訊密碼。
- 大戰後，因為布萊切利的一切都是機密，圖靈和科學家們無法被當成英雄。但他堅持在機器研究提出許多科學理論，「圖靈機」（Turing Machine）就是人工智慧。只是他的秘密也在同時被發現：他是位同性戀者，後被化學閹割。 https://www.youtube.com/watch?v=DkqkTKqkJf8

模仿遊戲 (The Imitation Game, 2014)



- 《模仿遊戲》裡面的替換性密碼(Substitution cipher)
- P ZQAE TQR 翻譯後成為了I LOVE YOU
- 上圖：沒有固定偏移量
- 下圖：固定偏移量(key= - 4)
- 電影的另一組同樣重要的密碼

- WII CSY ME XAS PSRK AIIOW HIEVIWX JVMIRH
- 就有固定的偏移量了。
- 偏移量是4， 也就是把上面的字母回推4個，就能得到
- SEE YOU IN TWO LONG WEEKS DEAREST FRIEND

'Stay weird, stay different': Graham Moore's Oscars speech inspires

- 最佳改編劇本得主Graham Moore上台致詞時，表示把獎項獻給電影的主角圖靈（ Alan Turing ），他在現實中因自己的性取向被迫害，最終服毒自殺。Turing committed suicide at the age of 41
- http://mashable.com/2015/02/23/graham-moore-oscars-speech/

2001年7月，基金會在布萊切利園安放了一塊基石，上面刻著邱吉爾的名言「在人類歷史上，從未有如此多的人對如此少的人欠得如此多。」