



# Cryptanalysis

---

- Assumption: (Kerckhoffs' principle) the cryptosystem used is known to the opponent.
  - designer should not assume what system used can remain secret.
  - attack models: kind of information available to the adversary.



# Attack models

---

- $x$ =plaintext,  $y$ =ciphertext.
- 1. ciphertext only attack: only  $y$  is known.  
(weakest type of attack models)
- 2. known plaintext attack: some  $(x,y)$  where  $x \rightarrow y$  is known.
- 3. chosen plaintext attack: temporary access to  $e_k(x)$  [encryption machine]
- 4. chosen ciphertext attack: temporary access to  $d_k(y)$  [decryption machine]



# Cryptanalysis using statistical properties

---

- statistical analysis is useful for most of mono-alphabetic cryptosystems of English text.
- basic idea:
  - relative frequency of 26 letters are quite different. (e.g. "E" vs. "Z").
  - there are popular digrams (e.g. "TH") and trigrams (e.g. "ING").



# Table 1.1. Letter Frequency

A	0.082	N	0.067
B	0.015	O	0.075
C	0.028	P	0.019
D	0.043	Q	0.001
E	0.127	R	0.060
F	0.022	S	0.063
G	0.020	T	0.091
H	0.061	U	0.028
I	0.070	V	0.010
J	0.002	W	0.023
K	0.008	X	0.001
L	0.040	Y	0.020
M	0.024	Z	0.001

E	0.127	M	0.024
T	0.091	W	0.023
A	0.082	F	0.022
O	0.075	G	0.020
I	0.070	Y	0.020
N	0.067	P	0.019
S	0.063	B	0.015
H	0.061	V	0.010
R	0.060	K	0.008
D	0.043	J	0.002
L	0.040	Q	0.001
C	0.028	X	0.001
U	0.028	Z	0.001



# Common Digram and Trigram

---

- Common Digrams:
  - TH, HE, IN, ER, AN, RE, ED, ON, ES, ST, EN, AT, TO, NT, HA, ND, OU, EA, NG, AS, OR, TI, IS, ET, IT, AR, TE, SE, HI, OF
- Common Trigram:
  - THE, ING, AND, HER, ERE, ENT, THA, NTH, WAS, ETH, FOR, DTH



# Cryptanalysis: affine cipher

---

- Encryption  $e_k(x) = a x + b \pmod{26}$ .
  - "a" and "b" are unknown.
- Some known ciphertext as in e.g. 1.10:  
FMXVEDKAPHFERBNDKRXRSRE...
  - Frequency table in Table 1.2 (page 28).
- Matching most popular letters between Table 1.1 and Table 1.2 can be useful to solve "a" and "b". (two unknowns and two equations)



# Table 1.1 vs. Table 1.2

E	0.127	M	0.024
T	0.091	W	0.023
A	0.082	F	0.022
O	0.075	G	0.020
I	0.070	Y	0.020
N	0.067	P	0.019
S	0.063	B	0.015
H	0.061	V	0.010
R	0.060	K	0.008
D	0.043	J	0.002
L	0.040	Q	0.001
C	0.028	X	0.001
U	0.028	Z	0.001

R	8	X	2
D	7	B	1
E	5	N	1
H	5	O	1
K	5	Y	1
F	4	C	0
V	4	G	0
S	3	I	0
A	2	J	0
L	2	Q	0
M	2	T	0
P	2	W	0
U	2	Z	0



# Cryptanalysis: affine cipher

---

- Encryption  $e_k(x) = a x + b \pmod{26}$ .
  - "a" and "b" are unknown.
- Matching Table 1.1 and Table 1.2 can reduce the number of ways to solve "a" and "b". (read e.g. 1.10, page 28-29).
- NOTE: since the key space is small (how many?), we can easily solve by an exhaustive search program.



# Cryptanalysis: substitution cipher

---

- The key space is  $26!$ , too big for exhaustive search.
- We can use frequency table approach, along with digrams, and trigrams to solve.
  - For a more complicated example, read e.g. 1.11, page 29-32.



# Cryptanalysis: Vigenere cipher

---

- keyword of length  $m$ :  $\mathbf{k} = (k_1, k_2, \dots, k_m)$ 
  - $m$ =key word length.
- $\mathbf{y} = (y_1, y_2, \dots, y_n)$  is observed
  - $n$ =(large) ciphertext length. assume  $m|n$ .
- $e_{\mathbf{k}}(x_i) = x_i + k_i \pmod{26}$ ,  $i=1, \dots, m$  is the “position” in each block of size  $m$ .
- Cryptanalysis: need to find  $m$  and  $\mathbf{k}$ .



# Cryptanalysis: Vigenere cipher

---

- e.g. 1.12 (page 34) ciphertext:
  - **CHR**EEVOAHMAERATBIAXXWTNXBE...
- Q: how to find **m** and keyword **k** ?
- A: **Kasiski test**.
- NOTE: **CHR** appeared **five** times at position **1, 166, 236, 276, and 286**.
  - “distances” are multiple of 5. Hence **m=5**.
  - Other systematic method ?



# $I_c(\mathbf{x})$ : index of coincidence

---

- $\mathbf{x} = (x_1, x_2, \dots, x_n)$
- Let  $f_0, f_1, \dots, f_{25}$  be the frequency counts of letters 'A', 'B', ..., 'Z' in  $\mathbf{x}$ .
- Q: Randomly choosing two letters from  $\mathbf{x}$ , what is the probability of being identical letter ? [denoted as  $I_c(\mathbf{x})$ ]
- A:  $I_c(\mathbf{x}) = \sum [f_i (f_i - 1)] / [n(n-1)]$ . (why?)



## Using $I_c(\mathbf{x})$ to find $m$

---

- $\mathbf{y} = (y_1, y_2, \dots, y_n)$  be the ciphertext.
- Divide  $\mathbf{y}$  into  $m$  (guess) sub-strings as
  - $\mathbf{y}_1 = y_1, y_{m+1}, y_{2m+1}, \dots$
  - $\mathbf{y}_2 = y_2, y_{m+2}, y_{2m+2}, \dots$
  - ...
  - $\mathbf{y}_m = y_m, y_{2m}, y_{3m}, \dots$
- Compute  $I_c(\mathbf{y}_i)$ ,  $i=1, 2, \dots, m$ .
  - for correct  $m$ , the values of  $I_c(\mathbf{y}_i) \approx 0.065$  ?
  - for incorrect  $m$ , the values of  $I_c(\mathbf{y}_i) \approx 0.038$  ?



# Block length m determination

---

- Recall  $I_c(\mathbf{x}) = \sum [f_i (f_i - 1)] / [n(n-1)]$ 
  - If  $\mathbf{x}$  is a regular English text,  $I_c(\mathbf{x}) \approx \sum p_i^2 = 0.065$ .
    - $p_i$  = the relative frequency in Table 1.1. ( $i=0,1,\dots,25$ )
    - Note:  $I_c(\mathbf{x})$  remains unchanged with permutation.
  - If  $\mathbf{x}$  is a random text,  $I_c(\mathbf{x}) \approx \sum (1/26)^2 = 0.038$ .
- For  $m$  indices  $I_c(\mathbf{y}_i)$ ,  $i=1, 2, \dots, m$ .
  - if  $m$  is correct,  $\mathbf{y}_i$  is a sub-string of regular English text, and the values of  $I_c(\mathbf{y}_i) \approx 0.065$
  - if  $m$  is incorrect,  $\mathbf{y}_i$  is a sub-string of random test, and values of  $I_c(\mathbf{y}_i) \approx 0.038$



## Example 1.12. Find m.

- e.g. 1.12 (page 34) ciphertext:
  - CHREEVOAHMAERATBIAXXWTNXBE...
- if  $m=1$ , only one string,  $I_c(\mathbf{y}) = 0.045$
- if  $m=2$ ,  $I_c(\mathbf{y}_1) = 0.046$ ,  $I_c(\mathbf{y}_2) = 0.041$
- if  $m=3$ ,  $I_c(\mathbf{y}_i) = 0.043, 0.050, 0.047$
- if  $m=4$ ,  $I_c(\mathbf{y}_i) = 0.042, 0.039, 0.045, 0.040$
- if  $m=5$ ,  $I_c(\mathbf{y}_i) = 0.063, 0.068, 0.069, 0.0061, 0.072. (\approx 0.065)$



# Find key $k$

---

- Divide  $y$  into  $m$  sub-strings as
  - $y_1 = Y_1, Y_{m+1}, Y_{2m+1}, \dots$
  - $y_2 = Y_2, Y_{m+2}, Y_{2m+2}, \dots$
  - ...
  - $y_m = Y_m, Y_{2m}, Y_{3m}, \dots$
- Note: Each letter in  $y_i$  has been shifted by the same amount  $g = k_i$ . We search for  $g$  such that  $M_g = \sum p_j Q_{j+g} = \sum p_j^2 \approx 0.065$ . [why ?]
  - $Q_{j+g}$  is the relative letter frequency in  $y_i$ .



## Example 1.12. Find key $k$

---

- e.g. 1.12 (page 34) ciphertext:
  - CHREEVOAHMAERATBIAXXWTNXBE...
- $m=5$ . Divide the ciphertext into 5 substrings.  $y_1, y_2, \dots, y_5$ .
- For each  $g=0,1,2,\dots,25$  compute  $M_g(y_i)$  as shown in Table 1.4 (page 35)
- The correct key index  $g$  are boxed.
  - $k=(9,0,13,4,19)=\text{JANET}$ . (show!)

# Table 1.4 (page 35)

**TABLE 1.4**  
Values of  $M_{ij}$

$i$	value of $M_{ij}(y_i)$								
1	.035	.031	.036	.037	.035	.039	.028	.028	.048
	.061	.039	.032	.040	.038	.038	.045	.036	.030
	.042	.043	.046	.033	.049	.043	.042	.036	
2	.069	.044	.032	.035	.044	.034	.036	.033	.029
	.031	.042	.045	.040	.045	.046	.042	.037	.032
	.034	.037	.032	.034	.043	.032	.026	.047	
3	.048	.029	.042	.043	.044	.034	.038	.035	.032
	.049	.035	.031	.035	.066	.035	.038	.036	.045
	.027	.035	.034	.034	.036	.035	.046	.040	
4	.045	.032	.033	.038	.060	.034	.034	.034	.050
	.033	.033	.043	.040	.033	.029	.036	.040	.044
	.037	.050	.034	.034	.039	.044	.038	.035	
5	.034	.031	.035	.044	.047	.037	.043	.038	.042
	.037	.033	.032	.036	.037	.036	.045	.032	.029
	.044	.072	.037	.027	.031	.048	.036	.037	



# Hill cipher

---

- $P=C=(\mathbb{Z}_{26})^m$
- $K=(\mathbb{Z}_{26})^{m \times m}$ , key  $\mathbf{K}$  is a  $m \times m$  matrix.
- plaintext:  $\mathbf{x} = (x_1, x_2, \dots, x_m)$
- ciphertext :  $\mathbf{y} = (y_1, y_2, \dots, y_m)$
- $e_k(\mathbf{x}) = \mathbf{x} \mathbf{K} \pmod{26}$ ,
- $d_k(\mathbf{y}) = \mathbf{y} \mathbf{K}^{-1} \pmod{26}$ .



# Cryptanalysis: Hill cipher

---

- Can be hard to break with ciphertext only.
  - statistical frequency analysis is not useful. why not ?
- However, it is quite simple to break under known plaintext attack.
  - collect at least  $m$  pairs of  $(\mathbf{x}_i, \mathbf{y}_i)$  and solve a  $m \times m$  matrix equation. (how?)



# Break Hill cipher

---

- For  $i=1,2, \dots, m$ 
  - $i$ -th plaintext:  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{im})$
  - $i$ -th ciphertext :  $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{im})$
  - $\mathbf{y}_i = \mathbf{x}_i \mathbf{K} \pmod{26}$ ,  $\mathbf{K}$  unknown.
- Q: how find  $\mathbf{K}$  (and therefore  $\mathbf{K}^{-1}$ ) ?
- A: stack  $\mathbf{x}_i$  together as **matrix**  $\mathbf{X}$ , stack  $\mathbf{y}_i$  together as **matrix**  $\mathbf{Y}$ . We can solve
- $\mathbf{Y} = \mathbf{X} \mathbf{K}$  by  $\mathbf{Y} \mathbf{X}^{-1} = \mathbf{K} \pmod{26}$ .



## Example 1.13

---

- Suppose friday  $\rightarrow$  PQCFKU using Hill cipher with  $m=2$ . Find the key matrix.
  1. fr  $\rightarrow$  PQ:  $[5, 17] = [15, 16] K$
  2. id  $\rightarrow$  CF:  $[8, 3] = [2, 5] K$
  3. ay  $\rightarrow$  KU:  $[0, 24] = [10, 20] K$
- From first two equations, we can solve a  $2 \times 2$  matrix equation: (show!!)
$$Y = X K$$
- Q: what if we don't know  $m$  ?



# LFSR key stream cipher

---

- $Z_{i+m} = c_0 z_i + c_1 z_{i+1} + \dots + c_{m-1} z_{i+m} \text{ mod } 2$ 
  - $c_0 = 1$ , other  $c_i$  is 0 or 1.
  - Max period =  $2^m - 1$ .
    - How to choose "keys"  $c_i$  ?
- We can break the cryptosystem with a partial sequence (with length  $2m$ ) of  $z_i$ .
- Q: How ?



# Cryptanalysis: LFSR stream cipher

---

- $Z_{m+1} = C_0 Z_1 + C_1 Z_2 + \dots + C_{m-1} Z_m$
- $Z_{m+2} = C_0 Z_2 + C_1 Z_3 + \dots + C_{m-1} Z_{m+1}$
- $Z_{m+3} = C_0 Z_3 + C_1 Z_4 + \dots + C_{m-1} Z_{m+2}$
- .....
  
- $Z_{2m} = C_0 Z_m + C_1 Z_{m+1} + \dots + C_{m-1} Z_{2m-1}$
  
- re-written as (column)  $\mathbf{z} = \mathbf{M} \mathbf{c}$ ,  $\mathbf{c} = \mathbf{M}^{-1} \mathbf{z}$ .
  - m equations, m unknowns.



# LFSR Cryptoanalysis Example

- e.g. 1.14 (page 38). Assume  $m=5$  is known.
- Given a pair of  $(x,y)$  for  $x \rightarrow y = x+z \pmod{2}$ .
- key stream (LFSR) is  $z = x + y \pmod{2}$ . (why?)
- we can find the key stream generator.

x	1	0	1	1	0	1	0	1	1	1	1	0	0	1	0
y	0	1	1	0	0	1	1	1	1	1	1	1	0	0	0
$z = x+y \pmod{2}$	1	1	0	1	0	0	1	0	0	0	0	1	0	1	0

# Example 1.14 (page 38)

key z	1	1	0	1	0	0	1	0	0	0	0	1	0	1	0
left shift 1	1	0	1	0	0	1	0	0	0	0	1	0	1	0	
left shift 1	0	1	0	0	1	0	0	0	0	1	0	1	0		
left shift 1	1	0	0	1	0	0	0	0	1	0	1	0	0		
left shift 1	0	0	1	0	0	0	0	1	0	1	0	0			

$$\mathbf{c} = \mathbf{M}^{-1} \mathbf{z} = (1, 0, 0, 1, 0)'$$

$$z_{i+5} = (z_i + z_{i+3}) \bmod 2$$

c1	c2	c3	c4	c5	y
1	1	0	1	0	0
1	0	1	0	0	1
0	1	0	0	1	0
1	0	0	1	0	0
0	0	1	0	0	0



# Summary and HW2

---

- What we have learned so far ?
  - Crypto-analysis for various ciphers ?
  - Mathematical tools ?
- 
- HW2: Write a program to solve 1.21 or to verify Table 1.4.



# Chapter Review: Modulus and Matrix operations

---

- Modulus operations
  - $(-a) \bmod m, a^{-1} \bmod m.$
- Matrix Multiplication
- Matrix Inverse
  - 2x2 matrix
  - $m \times m$  matrix
  - Matrix Inverse mod  $m.$



# Chapter Review: Euler totient function $\phi(n)$

---

- $\phi(n)$ : number of integers between 1 and  $n$  that are relative prime to  $n$ .
- Computation of  $\phi(n)$ :
  1.  $\phi(p^e) = p^{e-1} (p-1)$
  2.  $\phi(P Q) = \phi(P) \phi(Q)$ , if  $\gcd(P, Q) = 1$ .
- E.g.
  - $\phi(20) = \#\{1, 3, 7, 9, 11, 13, 17, 19\} = 8$
  - $\phi(5) = \#\{1, 2, 3, 4\} = 4$
  - $\phi(4) = \#\{1, 3\} = 2$



# The Use of Encryption

---

- DES and AES (Ch 3)
- Cryptographic Hash Functions (Ch 4)
- Digital Signatures (Ch 7)
- Certificates (Ch 9)
- Key Exchange/Distribution (Ch 10)