Cryptography
- Classical crypto

Avelino Zorzo

Review

- Two properties of a secure cipher
  - Confusion
    - » Substitution
    - » Intention: make the relationship between the key and ciphertext as complex as possible
  - Diffusion
    - » Transposition
    - » Intention: rearrange bits in the message so that an redundancy in the plaintext is spread out over the ciphertext

Four general types of attack

1. **Ciphertext-only** attack
   - Enigma: links and chains
2. **Known-plaintext** attack
   - Enigma: cribs
3. **Chosen-plaintext** attack
   - Breaking secret code at Midway
4. **Chosen-ciphertext** attack
   - The job is to deduce the key (lunch time attack)

Other types of attacks

- Meet-in-the-middle
  - 2^{2k} operations is reduced to 2k operations plus 2k storage
- Birthday paradox
  - Probability that 2 people in a room of 23 people have the same birthday is aprox. 0.507
  - $P_2(m,n) = 1 - e^{\wedge}(-n^2/2.m)$
  - $P_2(365, 23) = 0.507$
  - $P_2(365, 30) = 0.706$

Other types of attacks

- Brute force attack
- Man-in-the-middle attack
- Differential cryptanalysis
- Frequency analysis
- …

Modular Arithmetic

- Definition:
  - Suppose a and b are integers, and m is a positive integer. Then we write $a \equiv b$ if m divides a-b.
  - $a \equiv b \pmod m$ is called a *congruence*
  - m is called *modulus*
- Examples:

$2 \equiv 11 \pmod 3$ as 2 mod 3 = 11 mod 3 = 2

$12 \equiv 19 \pmod 7$ as 12 mod 7 = 19 mod 7 = 5

Modular Arithmetic

- Arithmetic modulus m is defined as follows:
  - $Z_n = \{0, \ldots, n-1\}$
  - Two operations: + and x
    - » Work as in real addition and multiplication
    - » Results are reduced to modulus n
    - » $a,b \in Z_n$, $a+b \in Z_n$
    - » $a,b \in Z_n$, $axb \in Z_n$
- Satisfy most of the familiar arithmetic rules, *e.g.* addition is closed, multiplication is commutative, etc.

Classical cryptography

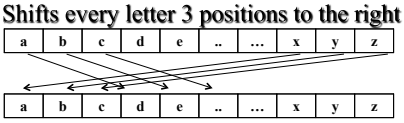
- Classical cryptography
  - Based on characters (human)
- Modern cryptography
  - Based on binary inputs (computer)
- What has changed?
  - 26 elements to 2 elements.
  - But, the philosophy remains basically the same.

Substitution ciphers

1. Monoalphabetic cipher
   - E.g. a → b, b → c
2. Polyalphabetic cipher
   - Made up of several monoalphabetic ciphers
3. Homophonic substitution cipher
   - E.g., a → 5 or 9 or 17
4. Polygram substitution cipher
   - Substitute groups of letters, AB → PQ

Caesar cipher – shift cipher

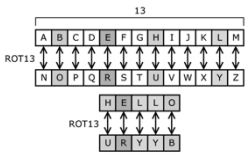
- Named after Julius Caesar
- Used for hundreds of years
- Shifts every letter 3 positions to the right



- Example
  - attackatdawm → dwwdfndwgdzp

ROT13 – shift cipher

- Another example of monoalphabetic cipher
- Commonly found on UNIX systems
- Every letter is rotated by 13 positions
- Question
  - Why not ROT14?



(Source: Wikipedia.org)

Cryptanalysis of shift cipher

- Brute force attack
  - Given JBCRCLQRWCRVNBJENBWRWN, can you find out the plaintext?

```
jbcrclqrwcrvnbjenbwrwn      (K=0)
iabqbkpqvbqumaidmavqvm      (K=1)
hzapajopuaptlzhclzupul      (K=2)
gyzozinotzoskygbkytotk      (K=3)
fxynyhmnsynrjxfajxsnsj      (K=4)
ewxmxglmrxmqiweziwrmri      (K=5)
dvwlwfklqwlphvdyhvqlqh      (K=6)
cuvkvejkpvkogucxgupkpg      (K=7)
btujudijoujnftbwftojof      (K=8)
astitchintimesavesnine      (K=9)
```

What went wrong?

- The shift cipher (modulo 26) is not secure, because it can be broken by ***exhaustive search***
- Only 26 possible keys
- On average, a plaintext can be computed after just 26/2=13 tries.
- ***Lesson***: for a cipher to be secure, the key space must be very large
- But, is the reverse true?

Substitution ciphers

- Number of symbols in the alphabet = q
  - q! distinct substitution ciphers
  - English alphabet = 26 letters
  - 26! aprox. 4×10^{26}

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

B D K Z Y U C A X W R L M E H F T Q N G I J O K S V

- Frequency analysis attack

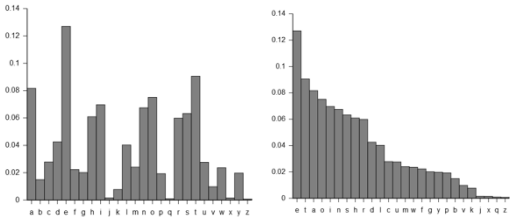
Substitution ciphers - cryptanalysis

- Letters frequency in English:

letter	prob	letter	prob	letter	prob
A	.082	J	.002	S	.063
B	.015	K	.008	T	.091
C	.028	L	.040	U	.028
D	.043	M	.024	V	.010
E	.127	N	.067	W	.023
F	.022	O	.075	X	.001
G	.020	P	.019	Y	.020
H	.061	Q	.001	Z	.001
I	.070	R	.060		

  - Most common diagrams: 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
  - Most common trigrams: THE, ING, AND, HER, ERE, ENT, THA, NTH, WAS, ETH, FOR, DTH

Substitution cipher - cryptanalysis



(a) Relative frequencies of English letters (b) Relative frequencies sorted by frequency

(Source: wikipedia)

Substitution ciphers

- Ciphertext-only attack
- Intercepted ciphertext:

BTLDXFETMDGLGMVMYFQEMQAPMVBZQMXZQEGZVXFTL
XGUWFVXBFWDYUXUQFQXUBGQZBMYMBBFHQXFPXGU
VHISUBXZVCMGQVXGUBFAUITUMCUTVXGZVIFFCXTMBUV
BTLDXFETMDGLPTFWZXVZQZXZMYMQAYZWZXUAHVUIL
XGUUELDXZMQVVFWUPFHTXGFHVMQALUMTVMEFXFXGU
XKUQXZUXGBUQXHTLKGUTUZXDYMLUAMBTHBZMYTFYU
ZQXGUFHXBFWUFPIFXGKFTYAKMTVBFWDYUXUAZQ
QZQUXUUQVZJXLXGTUUXGUIFFCBFOUTVXGFVUMVDUBXV
FPXGUGZVXFTLKGZBGKUTUWFVXVZEQZPZBMQXXFXGU
AUOUYFDWUQXFPXGUVHISUBX

Substitution ciphers - cryptanalysis

- Letters frequency in the ciphertext:

letter	prob	letter	prob	letter	prob
A	.023	J	.003	S	.005
B	.054	K	.015	T	.054
C	.010	L	.030	U	.120
D	.026	M	.061	V	.066
E	.018	N	.000	W	.023
F	.090	O	.005	X	.118
G	.066	P	.020	Y	.028
H	.023	Q	.059	Z	.064
I	.018	R	.000		

  - Most common diagrams:
    - » XG (16), GU (11), XF (8), QX (7), VX (7), BF (6), UX (6), ZQ (6)
  - Most common trigrams :
    - » XGU (10), BFW, FPX, FXG, GZV, LDX, LXG, MQA, PXG, UBX, UQX, UXU, VXG – (all 3 times)

Substitution ciphers - cryptanalysis

- Assumptions:
  - U and X appear the most often in the ciphertext.
  - Assume that they are E and T in the plaintext.
- Most common diagram in the ciphertext: XG.
- Most common trigram in the ciphertext: XGU
  - Assume X=T and G=H.
  - THE is the most common trigram in English: U=E.
- XF is a common diagrams.
  - Previously X = T. XF is TO or TI. O is a bit more frequent in English than I, so F = O.

Substitution ciphers - cryptanalysis

- Current ciphertext

BTLDtoETMDhLhMVMYoQEMQAPMVBZQMtZQEhZVtoTL
theWoVtBoWDYeteQoQteBhQZBMYMBBoHQtoPthe
VHISeBtZVCMhQVtheBoAeITeMCeTVthZVIoCtTMBeV
BTLDtoETMDhLPToWZtVZQZtZMYMQAYZWZteAHVeIL
theeELDtZMQVVoWePoHTthoHVMQALeMTVMEotothe
tKeQtZethBeQtHTLKheTeZtDYMLeAMBTHBZMYToYe
ZQtheoHtBoWeFPIothKoTYAKMTVBoWDYeteAZQ
QZQeteeQVZJtLthTeetheIooCBFOeTVthFVeMVDeBtV
oPthehZVtoTLKhZBhKeTeWoVtVZEQZPZBMQttothe
AeOeYoDWeQtoPtheVHISeBt

- X = T, G = H, U = E, F = O
  - Further analysis:
    - » QX (AT, IT, NT) and UQX → Q = N
    - » MQA = AND

Substitution ciphers - cryptanalysis

- Plaintext message (with spaces added):

cryptography has a long and fascinating history
the most complete nontechnical account of the
subject is kahns the codebreakers this book traces
cryptography from its initial and limited use by
the egyptians some four thousand years ago to the
twentieth century where it played a crucial role
in the outcome of both world wars completed in
nineteen sixty three the book covers those aspects
of the history which were most significant to the
development of the subject

(Taken from Handbook of Applied Cryptography, by A. Menezes, P. van Oorschot, and S. Vanstone)

What went wrong?

- A large key space is not sufficient to ensure the cipher is secure.
- Substitution only provides confusion.
- ***Lesson***: a secure cipher should combine both confusion and diffusion.

Vigenère cipher

- A polyalphabetic cipher based on the idea of combining a few Caesar ciphers into one
- Named after Blaise De Vigenère, a French diplomat in 1586

k = A L V A R O A L V A R O A L V A
m = B E R E A D Y A C K A T D A W N (+1 mod 26)

c = C G U F C G Z C F L C W E C Z O

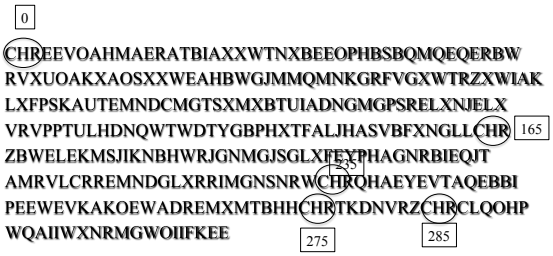
Vigenère cipher - cryptanalysis

- Two steps in the cryptanalysis
  1. Find out the key length m
  2. Find out each letter in the key

How to find out the key length?

- First method: Kasiski test
  - Described by Friedrich Kasiski in 1863
  - Search for identical segments and count how many positions they are apart

Example: Kasiski



How to find out the key length?

- Second method: index of coincidence
  - Described by William Friedman
  - Suppose $x = x_1x_2..x_n$ is a string of n alphabetic characters
  - The **index of coincidence** of x is defined to be the the probability that two random elements of x are identical

Index of coincidence

- Suppose a string of n English letters
  - Frequency of $a = f_0$
  - Frequency of $b = f_1$ ...
  - Frequency of $z = f_{25}$
  - Hence the index of coincidence is:

$$I_c(x) = \frac{\sum_{i=0}^{25}\binom{f_i}{2}}{\binom{n}{2}} = \frac{\sum_{i=0}^{25} f_i(f_i-1)}{n(n-1)} \approx \sum_{i=0}^{25} p_i^2$$

Index of coincidence

- Index of coincidence of sentence in English:

$$I_c(x) \approx \sum_{i=0}^{25} p_i^2 = 0.065$$

- The same reasoning applies if x is a ciphertext string obtained using a monoalphabetic cipher.
- Rewrite the ciphertext c as:

$$\begin{array}{lcll} c_1 & = & c_1c_{m+1}c_{2m+1} & \cdots \\ c_2 & = & c_2c_{m+2}c_{m+2} & \cdots \\ & & \cdots & \\ c_m & = & c_mc_{2m}c_{3m} & \cdots \end{array}$$

- If $c_1, c_2, \ldots, c_m$ are constructed in such a way that m is the keyword length, then each $I_c(c_i)$ should be approximately equal to 0.065

Index of coincidence

- If m is not the keyword length, the strings c_i would look random. Random strings have:

$$I_c \approx 26\left(\frac{1}{26}\right)^2 = \frac{1}{26} = 0.038$$

- Example: table with I_c for different values of m:

m	I_c
1	0.043
2	0.052; 0.051
3	0.05; 0.059; 0.045
4	0.049; 0.053; 0.052; 0.051
5	0.034; 0.05; 0.048; 0.038; 0.045
6	0.063; 0.07; 0.083; 0.062; 0.071; 0.048
7	0.033; 0.041; 0.038; 0.046; 0.041; 0.04; 0.047

- This method also shows that $m = 6$.

Next step: break each shift cipher

C H R E E
V O A H M
A E R A T
B I A X X
W T N X B
E E O P H
B S B Q M
...

Weakness of Vigenère

- Vigenère cipher involves some transposition
- However, the transposition doesn't randomly spread the information in the ciphertext.
- ***Lesson***: ciphertext should not contain any discernible patterns.
- In other words, a secure cipher should produce ciphertext that is indistinguishable from random.

First coursework

- Given one ciphertext find the plaintext
- Write a two pages report explaining how the cryptanalysis was done and part of the ciphertext and the plaintext

Transposition

- Cipher sentence: "this sentence is secret"

```
t h i s s
e n t e n
c e i s s
e c r e t
```

- Becomes: tecehnecitirsesesnst

```
3 1 2 5 4
t h i s s
e n t e n
c e i s s
e c r e t
```

- Becomes: hnecitirtecesnstsese