
Classical Cryptography

CS461/ECE422

Fall 2007

Reading

- Chapter 2 from *Security in Computing*
- Chapter 9 from *Computer Science: Art and Science*
- *Applied Cryptography*, Bruce Schneier

Overview

- Classical Cryptography
 - Substitution Ciphers
 - Cæsar cipher
 - Keyed permutation
 - Vigènere cipher
 - One Time Pad
 - Book cipher
 - Enigma
 - Transposition Ciphers

Cryptosystem

- 5-tuple $(\mathcal{E}, \mathcal{D}, \mathcal{M}, \mathcal{K}, \mathcal{C})$
 - \mathcal{M} set of plaintexts
 - \mathcal{K} set of keys
 - \mathcal{C} set of ciphertexts
 - \mathcal{E} set of encryption functions $e: \mathcal{M} \times \mathcal{K} \rightarrow \mathcal{C}$
 - \mathcal{D} set of decryption functions $d: \mathcal{C} \times \mathcal{K} \rightarrow \mathcal{M}$
- Encrypting function: $E(p_i, k_i) = c_i$
- Decrypting function: $D(c_i, k_i) = p_i$

Example

- Example: Cæsar cipher (The most basic cipher)
 - $\mathcal{M} = \{ \text{sequences of letters} \}$
 - $\mathcal{K} = \{ i \mid i \text{ is an integer and } 0 \leq i \leq 25 \}$
 - $\mathcal{E} = \{ E \mid k \in \mathcal{K} \text{ and for all letters } m, \$
$$E(m, k) = (m + k) \bmod 26 \}$$
 - $\mathcal{D} = \{ D \mid k \in \mathcal{K} \text{ and for all letters } c, \$
$$D(c, k) = (26 + c - k) \bmod 26 \}$$
 - $\mathcal{C} = \mathcal{M}$

Attacks

- Opponent whose goal is to break cryptosystem is the *adversary*
 - Standard cryptographic practice: Assume adversary knows algorithm used, but not the key
- Three types of attacks:
 - *ciphertext only*: adversary has only ciphertext; goal is to find plaintext, possibly key
 - *known plaintext*: adversary has ciphertext, corresponding plaintext; goal is to find key
 - *chosen plaintext*: adversary may supply plaintexts and obtain corresponding ciphertext; goal is to find key

Basis for Attacks

- Mathematical attacks
 - Based on analysis of underlying mathematics
- Statistical attacks
 - Make assumptions about the distribution of letters, pairs of letters (digrams), triplets of letters (trigrams), *etc.*
 - Called *models of the language*
 - E.g. Caesar Cipher, letter E
 - Examine ciphertext, correlate properties with the assumptions.

Classical Cryptography

- Sender, receiver share common key
 - Keys may be the same, or trivial to derive from one another
 - Sometimes called *symmetric cryptography*
- Two basic types
 - Transposition ciphers
 - Substitution ciphers
 - Combinations are called *product ciphers*

Transposition Cipher

- Rearrange letters in plaintext to produce ciphertext
- Example (Rail-Fence Cipher or 2-columnar transposition)
 - Plaintext is HELLO WORLD
 - HE
LL
OW
OR
LD
 - Ciphertext is HLOOL ELWRD

Transposition Cipher

- Generalize to n-columnar transpositions
- Example 3-columnar
 - HEL
LOW
ORL
DXX
 - HLODEORXLWLX

Attacking the Cipher

- Anagramming
 - If 1-gram frequencies match English frequencies, but other n -gram frequencies do not, probably transposition
 - Rearrange letters to form n -grams with highest frequencies

Example

- Ciphertext: HLOOLELWRD
- Frequencies of 2-grams beginning with H
 - HE 0.0305
 - HO 0.0043
 - HL, HW, HR, HD < 0.0010
- Frequencies of 2-grams ending in H
 - WH 0.0026
 - EH, LH, OH, RH, DH ≤ 0.0002
- Implies E follows H

Example

- Arrange so the H and E are adjacent

HE

LL

OW

OR

LD

- Read off across, then down, to get original plaintext

Substitution Ciphers

- Change characters in plaintext to produce ciphertext
- Example (Cæsar cipher)
 - Plaintext is HELLO WORLD
 - Change each letter to the third letter following it (X goes to A, Y to B, Z to C)
 - Key is 3, usually written as letter 'D'
 - Ciphertext is KHOOR ZRUOG

Attacking the Cipher

- Exhaustive search
 - If the key space is small enough, try all possible keys until you find the right one
 - Cæsar cipher has 26 possible keys
- Statistical analysis
 - Compare to 1-gram model of English
 - CryptoQuote techniques

Statistical Attack

- Compute frequency of each letter in ciphertext:

G 0.1 H 0.1 K 0.1 O 0.3

R 0.2 U 0.1 Z 0.1

- Apply 1-gram model of English
 - Frequency of characters (1-grams) in English is on next slide
 - <http://math.ucsd.edu/~crypto/java/EARLYCIPH>

Character Frequencies

a	0.080	h	0.060	n	0.070	t	0.090
b	0.015	i	0.065	o	0.080	u	0.030
c	0.030	j	0.005	p	0.020	v	0.010
d	0.040	k	0.005	q	0.002	w	0.015
e	0.130	l	0.035	r	0.065	x	0.005
f	0.020	m	0.030	s	0.060	y	0.020
g	0.015					z	0.002

Statistical Analysis

- $f(c)$ frequency of character c in ciphertext
- $\varphi(i)$ correlation of frequency of letters in ciphertext with corresponding letters in English, assuming key is i

$\varphi(i) = \sum_{0 \leq c \leq 25} f(c)p(c - i)$ so here,

$$\varphi(i) = 0.1p(6 - i) + 0.1p(7 - i) + 0.1p(10 - i) + 0.3p(14 - i) + 0.2p(17 - i) + 0.1p(20 - i) + 0.1p(25 - i)$$

- $p(x)$ is frequency of character x in English

Correlation: $\varphi(i)$ for $0 \leq i \leq 25$

i	$\varphi(i)$	i	$\varphi(i)$	i	$\varphi(i)$	i	$\varphi(i)$
0	0.0482	7	0.0442	13	0.0520	19	0.0315
1	0.0364	8	0.0202	14	0.0535	20	0.0302
2	0.0410	9	0.0267	15	0.0226	21	0.0517
3	0.0575	10	0.0635	16	0.0322	22	0.0380
4	0.0252	11	0.0262	17	0.0392	23	0.0370
5	0.0190	12	0.0325	18	0.0299	24	0.0316
6	0.0660					25	0.0430

The Result

- Most probable keys, based on φ :
 - $i = 6$, $\varphi(i) = 0.0660$
 - plaintext EBIIIL TLOLA
 - $i = 10$, $\varphi(i) = 0.0635$
 - plaintext AXEEH PHKEW
 - $i = 3$, $\varphi(i) = 0.0575$
 - plaintext HELLO WORLD
 - $i = 14$, $\varphi(i) = 0.0535$
 - plaintext WTAAD LDGAS
- Only English phrase is for $i = 3$
 - That's the key (3 or 'D')

Cæsar's Problem

- Key is too short
 - Can be found by exhaustive search
 - Statistical frequencies not concealed well
 - They look too much like regular English letters
- Improve the substitution permutation
 - Increase number of mapping options from 26

Key the Mapping

- Caesar mapping (shift 3)
 - ABCEDFGHIJKLMNOPQRSTUVWXYZ
 - XYZABCDEFGHIJKLMNQPQRSTUVWXYZ
- Key mapping
 - ABCEDFGHIJKLMNOPQRSTUVWXYZ
 - **SECUR**ABDFGHIJKLMNOPQTVWXYZ
- Poor mapping at the end
- Still only one mapping of a character across whole message
 - Just a crypto quote

Vigènere Cipher

- Like Cæsar cipher, but use a phrase as key
- Example
 - Message THE BOY HAS THE BALL
 - Key VIG
 - Encipher using Cæsar cipher for each letter:

key	VIGVIGVIGVIGVIGV
plain	THEBOYHASTHEBALL
cipher	OPKWWECIYOPKWIRG

| 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

A | 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 | 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 a
C | c 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
D | 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
E | 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 d

F | f g h i j k l m n o p q r s t u v w x y z a b c d e
G | g h i j k l m n o p q r s t u v w x y z a b c d e f
H | h i j k l m n o p q r s t u v w x y z a b c d e f g
I | i j k l m n o p q r s t u v w x y z a b c d e f g h
J | j k l m n o p q r s t u v w x y z a b c d e f g h i
K | k l m n o p q r s t u v w x y z a b c d e f g h i j
L | l m n o p q r s t u v w x y z a b c d e f g h i j k
M | m n o p q r s t u v w x y z a b c d e f g h i j k l
N | n o p q r s t u v w x y z a b c d e f g h i j k l m
O | o p q r s t u v w x y z a b c d e f g h i j k l m n
P | p q r s t u v w x y z a b c d e f g h i j k l m n o
Q | q r s t u v w x y z a b c d e f g h i j k l m n o p
R | r s t u v w x y z a b c d e f g h i j k l m n o p q
S | s t u v w x y z a b c d e f g h i j k l m n o p q r
T | t u v w x y z a b c d e f g h i j k l m n o p q r s
U | u v w x y z a b c d e f g h i j k l m n o p q r s t
V | v w x y z a b c d e f g h i j k l m n o p q r s t u
W | w x y z a b c d e f g h i j k l m n o p q r s t u v
X | x y z a b c d e f g h i j k l m n o p q r s t u v w
Y | y z a b c d e f g h i j k l m n o p q r s t u v w x
Z | z 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

Relevant Parts of Tableau

	<i>G</i>	<i>I</i>	<i>V</i>
<i>A</i>	<i>G</i>	<i>I</i>	<i>V</i>
<i>B</i>	<i>H</i>	<i>J</i>	<i>W</i>
<i>E</i>	<i>L</i>	<i>M</i>	<i>Z</i>
<i>H</i>	<i>N</i>	<i>P</i>	<i>C</i>
<i>L</i>	<i>R</i>	<i>T</i>	<i>G</i>
<i>O</i>	<i>U</i>	<i>W</i>	<i>J</i>
<i>S</i>	<i>Y</i>	<i>A</i>	<i>N</i>
<i>T</i>	<i>Z</i>	<i>B</i>	<i>O</i>
<i>Y</i>	<i>E</i>	<i>H</i>	<i>T</i>

- Tableau shown has relevant rows, columns only
- Example encipherments(?):
 - key V, letter T: follow V column down to T row (giving “O”)
 - Key I, letter H: follow I column down to H row (giving “P”)

Useful Terms

- *period*: length of key
 - In earlier example, period is 3
- *tableau*: table used to encipher and decipher
 - Vigènere cipher has key letters on top, plaintext letters on the left
- *polyalphabetic*: the key has several different letters
 - Cæsar cipher is monoalphabetic

Attacking the Cipher

- Approach
 - Establish period; call it n
 - Break message into n parts, each part being enciphered using the same key letter
 - Solve each part
- We will show each step
- Automated in applet
 - <http://math.ucsd.edu/~crypto/java/EARLYCIPHERS/Vigenere.html>

The Target Cipher

- We want to break this cipher:

ADQYS MIUSB OXKKT MIBHK IZOOO
EQOOG IFBAG KAUMF VVTAA CIDTW
MOCIO EQOOG BMBFV ZGGWP CIEKQ
HSNEW VECNE DLAAV RWKXS VNSVP
HCEUT QOIOF MEGJS WTPCH AJMOC
HIUIX

Establish Period

- Kaskski: *repetitions in the ciphertext occur when characters of the key appear over the same characters in the plaintext*

- Example:

key VIGVIGVIGVIGVIGV

plain THEBOYHASTHEBALL

cipher OPKWECIYOPKWIRG

Note the key and plaintext line up over the repetitions (underlined). As distance between repetitions is 9, the period is a factor of 9 (that is, 1, 3, or 9)

Repetitions in Example

<i>Letters</i>	<i>Start</i>	<i>End</i>	<i>Distance</i>	<i>Factors</i>
MI	5	15	10	2, 5
OO	22	27	5	5
OEQOOG	24	54	30	2, 3, 5
FV	39	63	24	2, 2, 2, 3
AA	43	87	44	2, 2, 11
MOC	50	122	72	2, 2, 2, 3, 3
QO	56	105	49	7, 7
PC	69	117	48	2, 2, 2, 2, 3
NE	77	83	6	2, 3
SV	94	97	3	3
CH	118	124	6	2, 3

Estimate of Period

- OEQOOG is probably not a coincidence
 - It's too long for that
 - Period may be 1, 2, 3, 5, 6, 10, 15, or 30
- Most others (7/10) have 2 in their factors
- Almost as many (6/10) have 3 in their factors
- Begin with period of $2 \times 3 = 6$

Check on Period

- Index of coincidence is probability that two randomly chosen letters from ciphertext will be the same
- Tabulated for different periods:

1	0.066	3	0.047	5	0.044
2	0.052	4	0.045	10	0.041
Large	0.038				

Compute IC

- $IC = [n (n - 1)]^{-1} \sum_{0 \leq i \leq 25} [F_i (F_i - 1)]$
 - where n is length of ciphertext and F_i the number of times character i occurs in ciphertext
- Here, $IC = 0.043$
 - Indicates a key of slightly more than 5
 - This is a statistical measure, so it can be an error, but it agrees with the previous estimate (which was 6)

Splitting Into Alphabets

alphabet 1: AIKHOIATTOBGEEERNEOSAI

alphabet 2: DUKKEFUAWEMGKWDWSUFWJU

alphabet 3: QSTIQBMAMQBWQVLKVTMTMI

alphabet 4: YBMZOAFCCOFPHEAXPQEPOX

alphabet 5: SOIOOGVICOVCSVASHOGCC

alphabet 6: MXBOGKVDIGZINNVVCIJHH

- ICs (#1, 0.069; #2, 0.078; #3, 0.078; #4, 0.056; #5, 0.124; #6, 0.043) indicate all alphabets have period 1, except #4 and #6; consider them as the error of statistics

Frequency Examination

ABCDEFGHIJKLMNOPQRSTUVWXYZ

1 31004011301001300112000000

2 10022210013010000010404000

3 12000000201140004013021000

4 21102201000010431000000211

5 10500021200000500030020000

7 01110022311012100000030101

Letter frequencies are (H high, M medium, L low):

HMMMHHMMHHMMMMHHMLHHHMLLLLLL

Begin Decryption

- First matches characteristics of unshifted alphabet
- Third matches if I shifted to A
- Sixth matches if V shifted to A
- Substitute into ciphertext (bold are substitutions)

ADIYS RIUKB OCKKL MIGHK AZOTO
EIOOL IFTAG PAUEF VATAS CIITW
EOCNO EIOOL BMTFV EGGOP CNEKI
HSSEW NECSE DDAAA RWCXS ANSNP
HHEUL QONOF EEGOS WLPCM AJEOC
MIUAX

Look For Clues

- **AJE** in last line suggests “are”, meaning second alphabet maps A into S:

ALIYS RICKB OCKSL MIGH S AZOTO

MIOOL INTAG PACEF VATIS CIITE

EOCNO MIOOL BUTFV EGOOP CNESI

HSSEE NECSE LDAAA RECXS ANANP

HHECL QONON EEGOS ELPCM AREOC

MICAX

Next Alphabet

- **MICAX** in last line suggests “mical” (a common ending for an adjective), meaning fourth alphabet maps O into A:

ALIMS RICKP OCKSL AIGHS ANOTO
MICOL INTOG PACET VATIS QIITE
ECCNO MICOL BUTTV EGOOD CNESI
VSSEE NSCSE LDOAA RECLS ANAND
HHECL EONON ESGOS ELDCM ARECC
MICAL

Got It!

- QI means that U maps into I, as Q is always followed by U...So we get the key for the fifth alphabet:

**ALIME RICKP ACKSL AUGHS ANATO
MICAL INTOS PACET HATIS QUITE
ECONO MICAL BUTTH EGOOD ONESI
VESEE NSOSE LDOMA RECLE ANAND
THECL EANON ESSOS ELDOM ARECO
MICAL**

One-Time Pad

- A Vigenère cipher with a random key at least as long as the message
 - Provably unbreakable
 - Why? Look at ciphertext DXQR. Equally likely to correspond to plaintext DOIT (key AJIY) and to plaintext DONT (key AJDY) and any other 4 letters
 - Warning: keys *must* be random, or you can attack the cipher by trying to regenerate the key
 - Approximations, such as using pseudorandom number generators to generate keys, are *not* random

Book Cipher

- Approximate one-time pad with book text
 - Sender and receiver agree on text to pull key from
 - Bible, Koran, Phone Book
- Problem is that book text is not random
 - Combine English with English
 - Can still perform language based statistical analysis

Enigma - Rotor Machines

- Another approximation of one-time pad
- Substitution cipher
 - Each rotor is a substitution
 - Changes in rotor position change how substitutions are stacked
 - Key press passes through all rotors and back through a reflector rotor
 - Rotors advance after each key press changing the substitution.
- Key is initial position of the rotors
- More details
 - <http://www.codesandciphers.org.uk/enigma/>

Rotor Mappings

- **Rotor I**

- ABCDEF G HIJKLMNOPQRSTUVWXYZ
BDFHJL C PRTXVZNYEIWGAKMUSQO

- **Rotor II**

- AB C DEFGHIJKLMNOPQRSTUVWXYZ
AJ D KSIRUXBLHWTMCQGZNPYFVOE

- **Rotor III**

- ABC D EFGHIJKLMNOPQRSTUVWXYZ
EKM F LGDQVZNTOWYHXUSPAIBRCJ

- **Reflector**

- ABCDE F GHIJKLMNOPQRSTUVWXYZ
YRUHQ S LDPXNGOKMIEBFZCWVJAT

Lessons from Enigma

- The importance of known plaintext (cribs)
- Mechanical assisted key breaking
 - Leading to modern computers
- Information in the pattern of traffic
 - Traffic analysis
- Humans in the loop are important
 - Information from spies
 - Poor user procedures
 - Birthday messages – many cribs
 - Repeated patterns
 - Reluctance to believe cipher has been broken