

University of Rochester
CSC290B
Introduction to Computer Security

Basic Cryptography

March 19, 2009

Basic Cryptography

- Classical Cryptography
- Public Key Cryptography
- Cryptographic Checksums

Overview

- Classical Cryptography
 - Cæsar cipher
 - Vigènere cipher
 - DES
- Public Key Cryptography
 - Diffie-Hellman
 - RSA
- Cryptographic Checksums
 - HMAC

Cryptosystem

- Quintuple (E, D, M, K, C)
 - M set of plaintexts
 - K set of keys
 - C set of ciphertexts
 - E set of encryption functions $e: M \times K \rightarrow C$
 - D set of decryption functions $d: C \times K \rightarrow M$

Example

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

Attacks

- Opponent whose goal is to break cryptosystem is the *adversary*
 - Assume adversary knows algorithm used, but not 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*
 - 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)
 - Plaintext is HELLO WORLD
 - Rearrange as
H L O O L
E L W R D
 - Ciphertext is H L O O L E L W R D

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: H L O O L E L W R D
- 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

H E

L L

O W

O R

L D

- 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 ZRUG

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

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

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 EBIII 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
- So make it longer
 - Multiple letters in key
 - Idea is to smooth the statistical frequencies to make cryptanalysis harder

Vigènere Cipher

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

key	VIG VIG VIG VIG VIG V
plain	THEBOYHASTHEBALL
cipher	OPKW W ECIYOPKWIRG

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
 - Vigenere cipher has key letters on top, plaintext letters on the left
- *polyalphabetic*: the key has several different letters
 - Caesar 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
 - You can leverage one part from another
- We will show each step

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 DLA AV RWKXS VNSVP

HCEUT QOIOF MEGJS WTPCH AJMOC

HIUIX

Establish Period

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

```
key    VIGVIGVIGVIGVIGV  
plain  THEBOYHASTHEBALL  
cipher OPKWWECIYOPKWIRG
```

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
 - A statistical measure, so it can be in error, but it agrees with the previous estimate (which was 6)

Splitting Into Alphabets

alphabet 1: AIKH OIATTOBGEEERNEOSAI

alphabet 2: DUKKEFUAWEMGKWDWSUFWJU

alphabet 3: QSTIQBMAMQBWQVLKVTMTMI

alphabet 4: YBMZOAFCCOOPHEAXPQEPOX

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; assume statistics off

Frequency Examination

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

1 3 1 0 0 4 0 1 1 3 0 1 0 0 1 3 0 0 1 1 2 0 0 0 0 0 0

2 1 0 0 2 2 2 1 0 0 1 3 0 1 0 0 0 0 0 1 0 4 0 4 0 0 0

3 1 2 0 0 0 0 0 0 2 0 1 1 4 0 0 0 4 0 1 3 0 2 1 0 0 0

4 2 1 1 0 2 2 0 1 0 0 0 0 1 0 4 3 1 0 0 0 0 0 0 2 1 1

5 1 0 5 0 0 0 2 1 2 0 0 0 0 0 5 0 0 0 3 0 0 2 0 0 0 0

6 0 1 1 1 0 0 2 2 3 1 1 0 1 2 1 0 0 0 0 0 0 3 0 1 0 1

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

H M M M H M M H H M M M M H H M L H H H M L L L L L

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 **RI**UKB O**C**KKL M**I**GHK **A**ZOTO
E**I**OO**L** **I**FTAG **PA**U**E**F V**A**T**A**S C**I**TTW
EOCNO E**I**OO**L** **B**M**T**FV **E**GGOP C**N**E**K**I
 HS**S**E**W** **N**E**C**S**E** D**D**A**A**A R**W**C**X**S
ANSNP **H**H**E**U**L** Q**O**N**O**F **E**E**G**O**S** W**L**P**C**M
AJ**E**O**C** **M**I**U**A**X**

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 CITE
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
CNESIVSSEE 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:

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

Acknowledgements

- Substantial portions of these slides are ©2002-2004 Matt Bishop and used with permission