

## Principles of Information Security CS 5323 Lecture Nine

Prof. William Winsborough  
October 7, 2008

## Business

- Thursday October 9 will be devoted to review
  - Bring your questions
- Tuesday October 14 will be midterm 1
  - It will cover material from the first 7 chapters and lectures through Thursday 9/25
  - This change of date reflects my inability to provide you with a study guide last week
- Questions from previous lectures?

7 October 2008

Winsborough CS 5323 Lecture 9

2

## Cryptosystem

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

7 October 2008

Winsborough CS 5323 Lecture 9

3

## Example

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

7 October 2008

Winsborough CS 5323 Lecture 9

4

## 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

7 October 2008

Winsborough CS 5323 Lecture 9

5

## 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.

7 October 2008

Winsborough CS 5323 Lecture 9

6

## 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*

7 October 2008

Winsborough CS 5323 Lecture 9

7

## Transposition Cipher

- Rearrange letters in plaintext to produce ciphertext
- Example (Rail-Fence Cipher)
  - Plaintext is HELLO WORLD
  - Rearrange as  
HLOOL  
ELWRD
  - Ciphertext is HLOOL ELWRD

7 October 2008

Winsborough CS 5323 Lecture 9

8

## 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

7 October 2008

Winsborough CS 5323 Lecture 9

9

## 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  $\leq 0.0002$
- Implies E follows H

7 October 2008

Winsborough CS 5323 Lecture 9

10

## Example

- Arrange so the H and E are adjacent  
HE  
LL  
OW  
OR  
LD
- Read off across, then down, to get original plaintext

7 October 2008

Winsborough CS 5323 Lecture 9

11

## 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

7 October 2008

Winsborough CS 5323 Lecture 9

12

## 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

7 October 2008

Winsborough CS 5323 Lecture 9

13

## 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

7 October 2008

Winsborough CS 5323 Lecture 9

14

## 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

7 October 2008

Winsborough CS 5323 Lecture 9

15

## 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

7 October 2008

Winsborough CS 5323 Lecture 9

16

## 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

7 October 2008

Winsborough CS 5323 Lecture 9

17

## The Result

- Most probable keys, based on  $\varphi$ :
  - $i = 6$ ,  $\varphi(i) = 0.0660$ 
    - plaintext EBIL 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')

7 October 2008

Winsborough CS 5323 Lecture 9

18

## 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