CS408
Cryptography & Internet Security

Lecture 3:

Classical cryptosystems

(Vigenère cipher)

Reza Curtmola

Department of Computer Science / NJIT

Towards Poly-alphabetic Substitution Ciphers

- Main weaknesses of mono-alphabetic substitution ciphers
  - each letter in the ciphertext corresponds to only one letter in the plaintext letter
- Idea for a stronger cipher (1460's by Alberti)
  - use more than one cipher alphabet, and switch between them when encrypting different letters
- Developed into a practical cipher by Blaise de Vigenère (published in 1586)
  - Was known at the time as the "indecipherable cipher"

The Vigenère Cipher

Definition:

Given m (a positive integer), $\mathcal{P} = \mathcal{C} = (Z_{26})^n$, and $K = (k_1, k_2, \ldots, k_m)$ a key, we define:

Encryption:

$e_k(p_1, p_2 \ldots, p_m) = (p_1+k_1, p_2+k_2 \ldots, p_m+k_m) \pmod{26}$

Decryption:

$d_k(c_1, c_2 \ldots, c_m) = (c_1-k_1, c_2-k_2 \ldots, c_m- k_m) \pmod{26}$

Example: key = LUCK (m = 4)

plaintext:	C	R	Y	P	T	O	G	R	A	P	H	Y
key:	L	U	C	K	L	U	C	K	L	U	C	K
ciphertext:	N	L	A	Z	E	I	I	B	L	J	J	I

Security of Vigenère Cipher

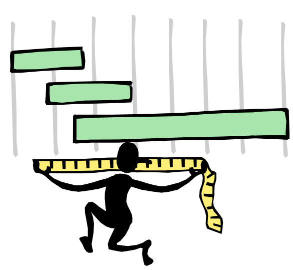
- Vigenère masks the frequency with which a character appears in a language: one letter in the ciphertext corresponds to multiple letters in the plaintext. Makes the use of frequency analysis more difficult.
- Any message encrypted by a Vigenère cipher is a collection of as many shift ciphers as there are letters in the key.

Vigenère Cipher: Cryptanalysis

- Find the length of the key.
- Divide the message into that many shift cipher encryptions.
- Use frequency analysis to solve the resulting shift ciphers.
  - how?

How to Find the Key Length?

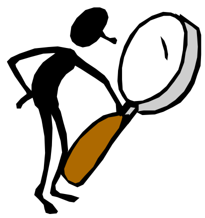
- For Vigenere, as the length of the key increases, the letter frequency shows less English-like characteristics and becomes more random.
- Two methods to find the key length:
  - Kasisky test
  - Index of coincidence (Friedman)



Kasisky Test

- Note: two identical segments of plaintext, will be encrypted to the same ciphertext, if they occur in the text at the distance Δ,

  Δ≡0 (mod m), m is the key length
- Algorithm:
  - Search for pairs of identical segments of length at least 3
  - Record distances between the two segments: Δ1, Δ2, …
  - m divides gcd(Δ1, Δ2, …)



Example of the Kasisky Test

P	T H E S U N A N D T H E M A N I N T H E M O O N
Key	K I N G K I N G K I N G K I N G K I N G K I N G
C	D P R Y E V N T N **B U K** W I A O X **B U K** W W B T

Kasisky Test: Another Example

- `Moonsunstarsmoonsunsmooth`
- Key: alfa
- **<u>MZT</u>**N***<u>SFS</u>***STLWS**<u>MZT</u>**N***<u>SFS</u>***S**<u>MZT</u>**THSWW
- 12,12,8
- Key length divides 12, 8, it's not 3, it's either 2 or 4

Index of Coincidence (Friedman)

Informally: Measures the probability that two random elements of an n-letter string x are identical.

Definition:

Suppose $x = x_1x_2\ldots x_n$ is a string of n alphabetic characters.

Then $I_c(x)$, the index of coincidence is:

$$I_c(x) = P(x_i = x_j)$$

where *i, j* are chosen at random from [1, 2, …, n]

Index of Coincidence (Friedman)

The I_C is specific to each language.

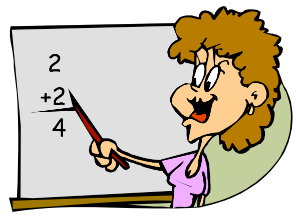
If we have a text (a string) in English and another text in Spanish, then the I_C for the two strings will be different.

Index of Coincidence (cont.)

- Reminder: binomial coefficient $\binom{n}{k} = \frac{n!}{k!(n-k)!}$
- Consider the plaintext x= $x_1 x_2 \ldots x_n$
- Let $f_0, f_1, \ldots f_{25}$ be the number of occurrences of characters A, B, … Z in x (frequencies of letters)
- Let $p_0, p_1, \ldots p_{25}$ be the probabilities with which A, B, … Z appear in x (i.e., $p_i = f_i / n$, for i=0..25)
- We want to compute .

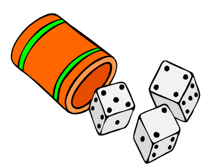
$$I_c(x) = P(x_i = x_j)$$

Begin Math



Elements of Probability Theory

A random experiment has
an unpredictable outcome.



Definition

The sample space (S) of a random phenomenon is the set of all outcomes for a given experiment.

Definition

The event (E) is a subset of a sample space (an event is any collection of outcomes).

Basic Axioms of Probability

If E is an event, we use *Pr(E)* to denote the probability that event E occurs.

The following hold true:

(a) $0 \le Pr(A) \le 1$ for any set ***A in S***.

(b) Pr(S) = 1 , where S is the sample space.

(c) If E_1, E_2, … E_n is a sequence of mutually exclusive events (that is $E_i \cap E_j = 0$, for all $i \neq j$), then:

$$\Pr(E_1 \cup E_2 \cup \ldots \cup E_n) = \sum_{i=1}^{n} \Pr(E_i)$$

Probability: More Properties

If E is an event and *Pr(E)* is the probability that the event E occurs. then

- *Pr(Ê)* = 1 - *Pr(E)* where Ê is the complimentary event of E
- If outcomes in S are equally likely, then

  Pr(E) = |E| / |S|

(where | | denotes the cardinality of the set)

So Pr(E) equals the ratio between the number of outcomes that result in the event occurring (positive outcomes) and the total number of possible outcomes.

Example

Random throw of a pair of dice.
What is the probability that the sum is 3?

Solution: Each die can take six different values {1,2,3,4,5,6}. The number of possible events (value of the pair of dice) is 36, therefore each event occurs with probability 1/36.

Examine the sum: 3 = 1+2 = 2+1
The probability that the sum is 3 is 2/36.

What is the probability that the sum is 11?
What is the probability that the sum is 12?

End Math

}

Index of Coincidence (cont.)

- We can choose two elements out of the string of size n in $\binom{n}{2}$ ways
- For each i, there are $\binom{f_i}{2}$ ways of choosing the two elements to be i:

  (S is the size of the alphabet)

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

THIS IS AN APPROXIMATION IF n is VERY LARGE

Example: IC of a String

- Consider the text: **THE INDEX OF COINCIDENCE**

$$I_C(x) = \frac{\sum_{i=0}^{S} f_i(f_i - 1)}{n(n-1)}$$

- There are 21 characters, so n = 21, S = 25

  I_c = (3*2+ 2*1+ 4*3+ 1*0+ 1*0+ 3*2+ 3*2+ 2*1+ 1*0+ 1*0) / 21*20 = **34/420 = 0.0809**

Example: IC of a Language

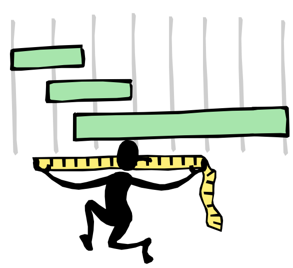
For English text, $S = 25$ and p_i can be estimated
(p_i is the probability with which character i appears in a large corpus of English text)

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

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

Find the Key Length

- For Vigenère, as the length of the keyword increases, the letter frequency shows less English-like characteristics and becomes more random.
- Two methods to find the key length:
  - Kasisky test
  - Index of coincidence (Friedman)



Finding the Key Length

Ciphertext x = $x_1 x_2 \ldots x_n$

m is the guessed key length (this is guessed, we start with m=3, then try 4,5,6,…)

$$\begin{bmatrix} x_1 & x_{m+1} & \cdots & x_{n-m+1} \\ x_2 & x_{m+2} & \cdots & x_{n-m+2} \\ \cdots & \cdots & \cdots & \cdots \\ x_m & x_{2m} & \cdots & x_n \end{bmatrix} \begin{matrix} y_1 \\ y_2 \\ \ldots \\ y_m \end{matrix}$$

Guessing the Key Length

- Try various values for m
- If m is the key length, then the texts y_i "look like" English text

$$I_c(y_i) \approx \sum_{i=0}^{25} p_i^2 = 0.065 \quad \forall 1 \le \mathrm{i} \le \mathrm{m}$$

- If m is not the key length, then the texts "look like" random text and:

$$I_c \approx \sum_{i=0}^{25} (\frac{1}{26})^2 = 26 \times \frac{1}{26^2} = \frac{1}{26} = 0.038$$

Finding the Key, if Key Length Known

Once the correct key length is found, apply frequency analysis method

- Consider vectors y_i, and look for the most frequent letter, etc.
- Look at the shift of the mapping, that represents the letter of the key
- Repeat for each vector. Each vector will yield a letter of the key.

The Vigenère Cipher

Definition

Given m (a positive integer), $\mathcal{P} = \mathcal{C} = (Z_{26})^n$, and $K = (k_1, k_2, \ldots, k_m)$ a key, we define:

Encryption:

$e_k(p_1, p_2 \ldots, p_m) = (p_1+k_1, p_2+k_2 \ldots, p_m+k_m) \pmod{26}$

Decryption:

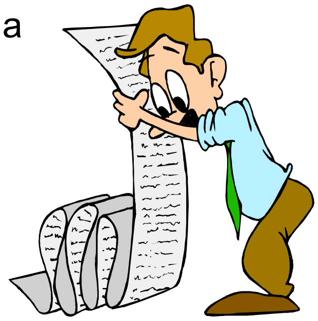
$d_k(c_1, c_2 \ldots, c_m) = (c_1-k_1, c_2-k_2 \ldots, c_m-k_m) \pmod{26}$

Example: key = LUCK (m = 4)

plaintext:	C	R	Y	P	T	O	G	R	A	P	H	Y
key:	L	U	C	K	L	U	C	K	L	U	C	K
ciphertext:	N	L	A	Z	E	I	I	B	L	J	J	I

Summary

- Vigenère cipher is vulnerable: once the key length is found, a cryptanalyst can apply frequency analysis.



Recommended Reading

- Chapter 2.1, 2.3, 2.4