Introduction to Cryptography CS 355

Lecture 4

The Vigenère Cipher

Lecture Outline

- Vigenère cipher.
- Attacks on Vigenere:
  - Kasisky Test
  - Index of Coincidence
  - Frequency analysis

Towards the Polyalphabetic Substitution Ciphers

- Main weaknesses of monoalphabetic 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 Vigenère (published in 1586)

The Vigenère Cipher

Definition:

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

Encryption:

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

Decryption:

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

Example:

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

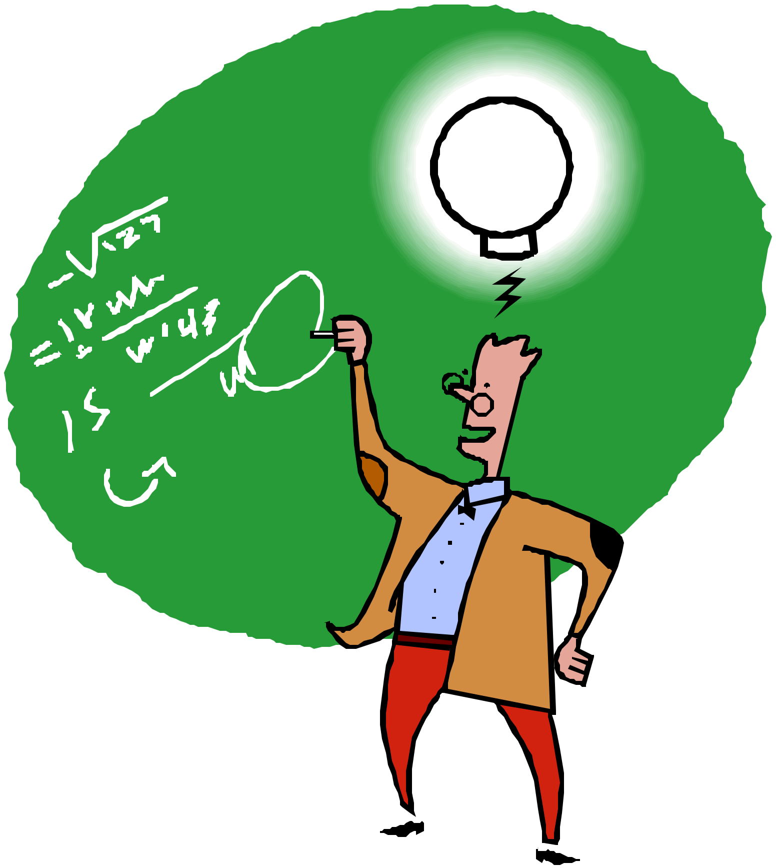
Ciphertext: N L A Z E I I B L J J I

Security of Vigenere Cipher

- Vigenere 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 Vigenere cipher is a collection of as many shift ciphe are letters in the key.

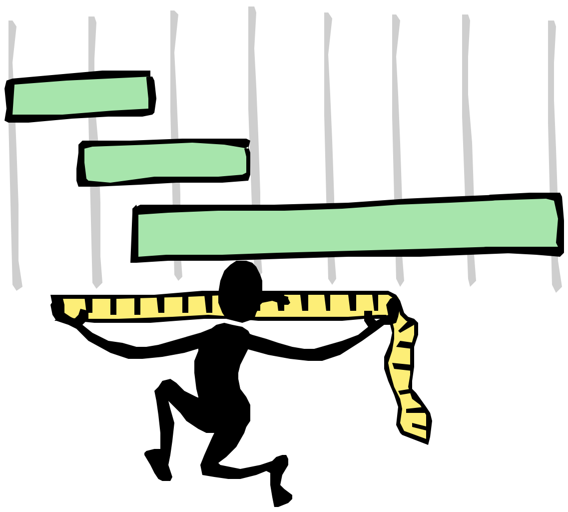
Vigenere Cipher: Cryptanalysis

- 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?



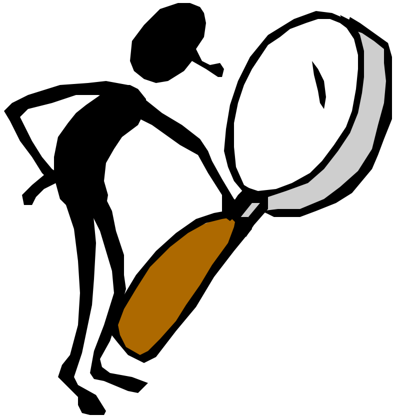
How to Find the Key Length?

- For Vigenere, as the length of the 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)



Kasisky Test

- Note: two identical segments of plaintext, will be encrypted to the same ciphertext, if the they occur in the text at the distance Δ, ($\Delta \equiv 0 \pmod{m}$, m is the key length).
- Algorithm:
  – Search for pairs of identical segments of length at least 3
  – Record distances between the two segments: $\Delta 1$, $\Delta 2$, …
  – m divides gcd($\Delta 1$, $\Delta 2$, …)



Example of the Kasisky Test

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
PT	t h e s u n a n d t h e m a n i n t h e m o o n
CT	D P R Y E V N T N <u>B U K</u> W I A O X <u>B U K</u> W W B T

Index of Coincidence (Friedman)

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

Definition:

Suppose x = $x_1 x_2 \ldots x_n$ is a string of n alphabetic characters. Then $I_c(x)$, the index of coincidence is:

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

Index of Coincidence (cont.)

- Reminder: binomial coefficient $\binom{n}{k} = \frac{n!}{k!(n-k)!}$
- Consider the plaintext x, and $f_0, f_1, \ldots f_{25}$ are the frequencies with which A, B, … Z appear in x and $p_0, p_1, \ldots p_{25}$ are the probabilities with which A, B, … Z appear in x.
- We want to compute $I_c(x)$.

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 elements to be i

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

Index of Coincidence of English

- For English, S = 25 and p_i can be estimated

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		

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

Finding the Key Length

$y = y_1 y_2 \ldots y_n$, m is the key length

$$\begin{bmatrix} y_1 & y_{m+1} & \cdots & y_{n-m+1} \\ y_2 & y_{m+2} & \cdots & y_{n-m+2} \\ \cdots & \cdots & \cdots & \cdots \\ y_m & y_{2m} & \cdots & y_n \end{bmatrix}$$

$\mathbf{y}_1$

$\mathbf{y}_2$

…

$\mathbf{y}_m$

Guessing the Key Length

- If m is the key length, then the text ``looks like'' English text

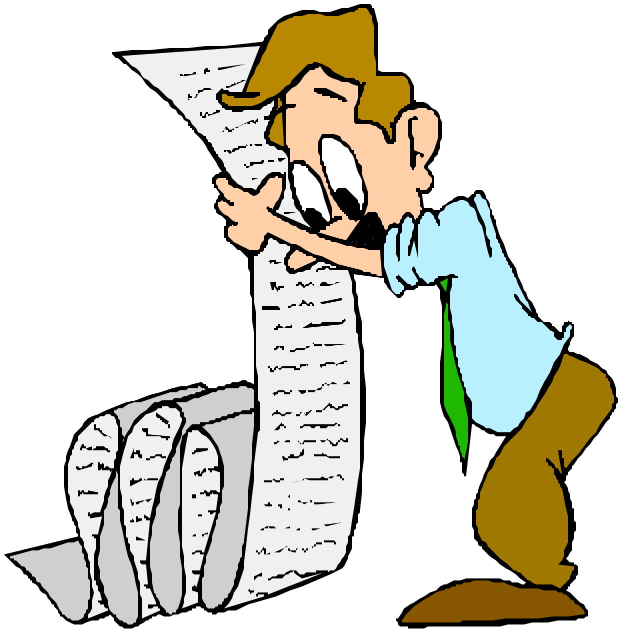
$$I_c(y_i) \approx \sum_{i=0}^{i=25} p_i^{\,2} = 0.065 \quad \forall 1 \leq \mathrm{i} \leq \mathrm{m}$$

- If m is not the key length, the text ``looks like'' random text and:

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

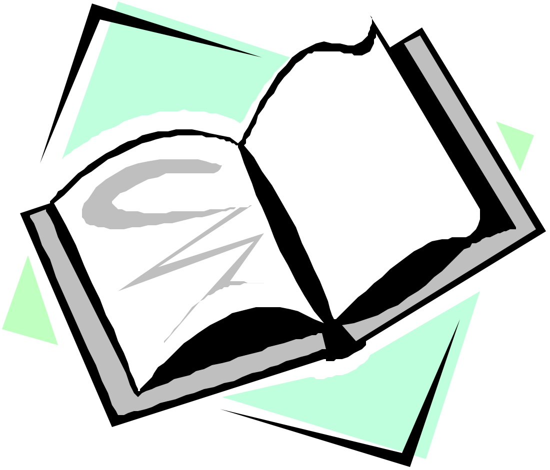
Summary

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



Recommended Reading for This Lecture

- Trappe & Washington
  - Section 2.3
- The Code Book
  - Chapter 2



Coming Attractions …

- Enigma Machine

- Recommended Reading
  - Trappe & Washington: 2.12
  - The Code Book: Chapters 3 & 4