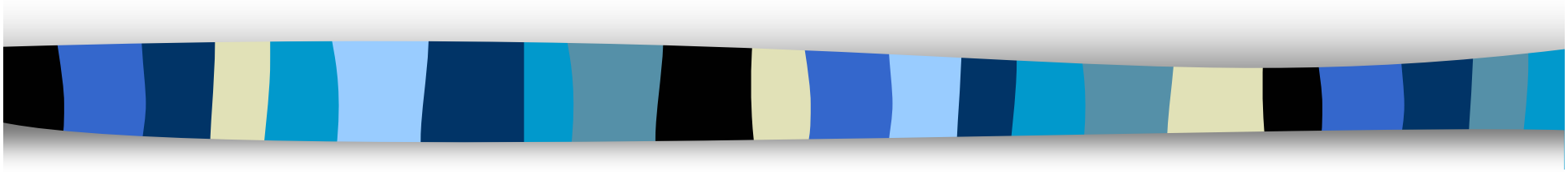
Cryptography 555

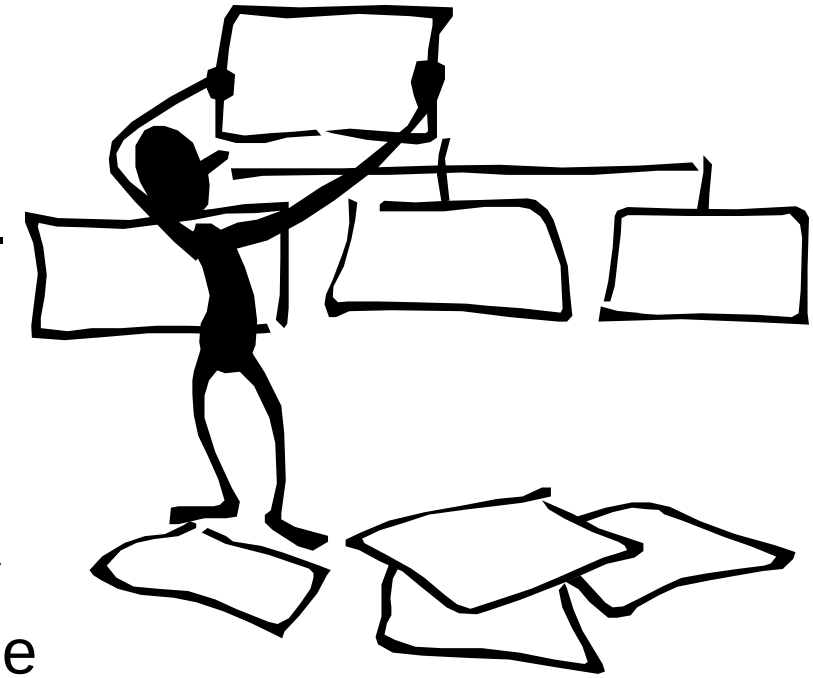
Lecture 2



Evolution of Classical Cryptography

Lecture Outline

- The Spartan scytale and transposition ciphers
- Shift and substitution ciphers.
- Frequency Analysis: attacks on substitution ciphers.
- Vigenere cipher.
- Attacks on Vigenere: Kasisky Test and Index of Coincidence
- Cipher machines: Rotor and Enigma machine.



History of Cryptography

- 2500+ years
- An ongoing battle between codemakers and codebreakers
- Driven by communication & computation technology
  - paper and ink
  - cryptographic engine & telegram, radio
  - modern cryptography: computers & digital communication

A Symmetric Cipher

- A Cipher ($\mathcal{K}$, $\mathcal{P}$, $\mathcal{C}$, **E**, **D**)
  - $\mathcal{K}$: the key space
  - $\mathcal{P}$: the plaintext space
  - $\mathcal{C}$: the ciphertext space
  - **E**: $\mathcal{K} \times \mathcal{P} \rightarrow \mathcal{C}$: the encryption function
  - **D**: $\mathcal{K} \times \mathcal{C} \rightarrow \mathcal{P}$: the decryption function
    - Given a key K and a plaintext P,
      D(K, **E**(K,P)) = P

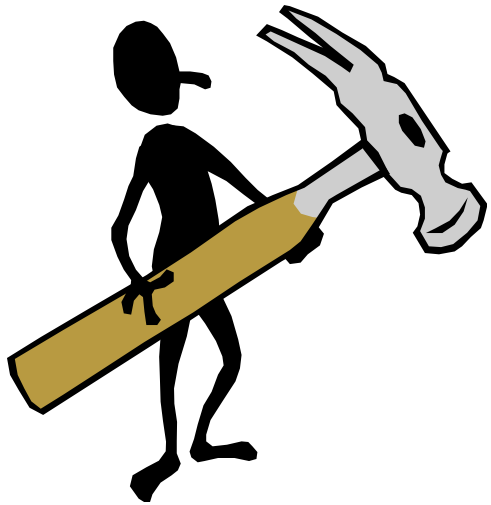
Cryptanalysis of Ciphers

- Two goals:
  - recover the encryption key
  - decrypt a given message
- There are different adversarial models, depending on:
  - the type of information available to the adversarial;
  - the interaction with the cipher machine.

Adversarial Models for Symmetric Ciphers

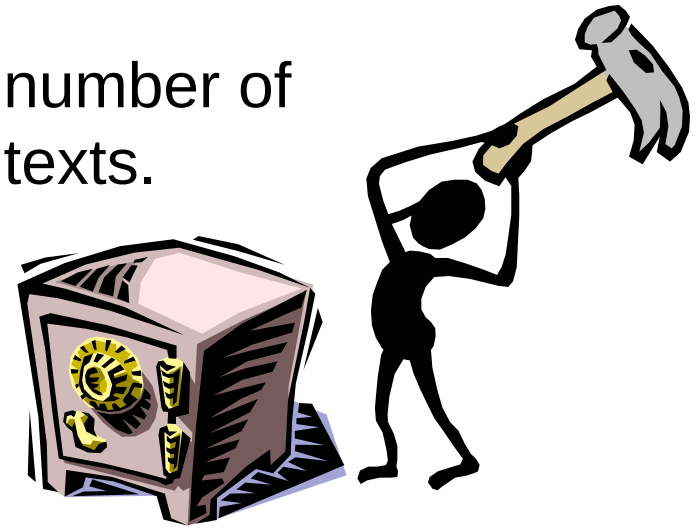
- **Ciphertext-only attack**: The cryptanalyst knows a number of ciphertexts. Sometimes the language of the plaintext and the cipher are also kown.

- **Known-plaintext attack**: The cryptanalyst knows some pairs of ciphertext and corresponding plaintext.



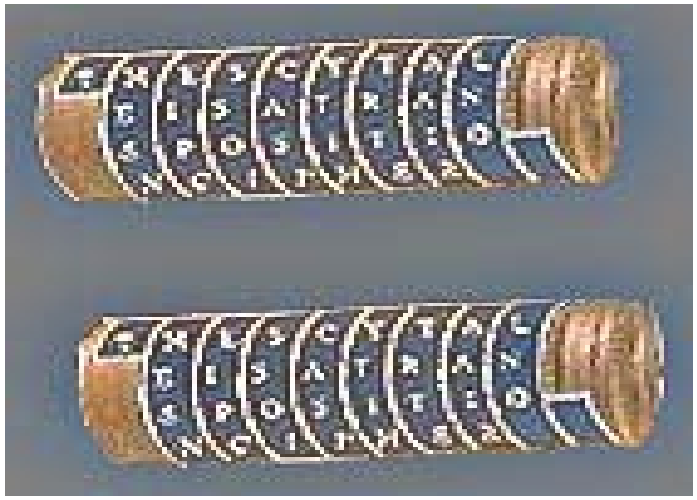
Adversarial Models for Symmetric Ciphers

- **Chosen-plaintext attack**
  The cryptanalyst can choose a number of messages and obtain the ciphertexts for them
- **Chosen-ciphertext attack**
  The cryptanalyst can choose a number of ciphertexts and obtain the plaintexts.



The Spartan Scytale Cipher

- Dating back to 5th centry B.C.
- A scytale is a wooden staff, around which a belt is wound; message is written along the length of the scytale
- It is a transposition cipher
  - the letters of a message are rearranged
- HW Problem: formally define the cipher
- Cryptanalysis?



Shift Cipher

- A substitution cipher
- The Key Space:
  - [1 .. 25]
- Encryption given a key K:
  - each letter in the plaintext P is replaced with the K'th letter following corresponding number (shift right)
- Decryption given K:
  - shift left

History: K = 3, Caesar's cipher

Shift Cipher Formally Defined

- Associate numbers with the alphabet

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
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

- The key space: Σ = {A, B, C, …, Z}
- The plaintext space: Σ^+
  - all strings consisting of letters in Σ
- The ciphertext space: Σ^+
- To encrypt a message P using key K,
  - each letter p_i in P is replaced by the letter $(p_i + K) \bmod 26$
- To decrypt a message C using key K,
  - each letter c_i in C is replaced by the letter $(c_i - K) \bmod 26$

Shift Cipher: An Example

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
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

P = CRYPTOGRAPHYISFUN

K = 11

C = NCJAVZRCLASJTDQFY

C $\rightarrow$ 2; 2+11 mod 26 = 13 $\rightarrow$ N

R $\rightarrow$ 17; 17+11 mod 26 = 2 $\rightarrow$ C

…

N $\rightarrow$ 13; 13+11 mod 26 = 24 $\rightarrow$ Y

Shift Cipher: Cryptanalysis

- Can an attacker find K?
  - YES: exhaustive search, key space is small (<= 26 possible keys).
- Once K is found, very easy to decrypt

General Monoalphabetic Substitution Cipher

- The key space: all permutations of Σ = {A, B, C, …, Z}
- Encryption given a key π:
  - each letter X in the plaintext P is replaced with $\pi(X)$
- Decryption given a key π:
  - each letter Y in the cipherext P is replaced with $\pi^{-1}(Y)$

Example:

	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
π=	B	A	D	C	Z	H	W	Y	G	O	Q	X	S	V	T	R	N	M	S	K	J	I	P	F	E	U

BECAUSE $\rightarrow$ AZDBJSZ

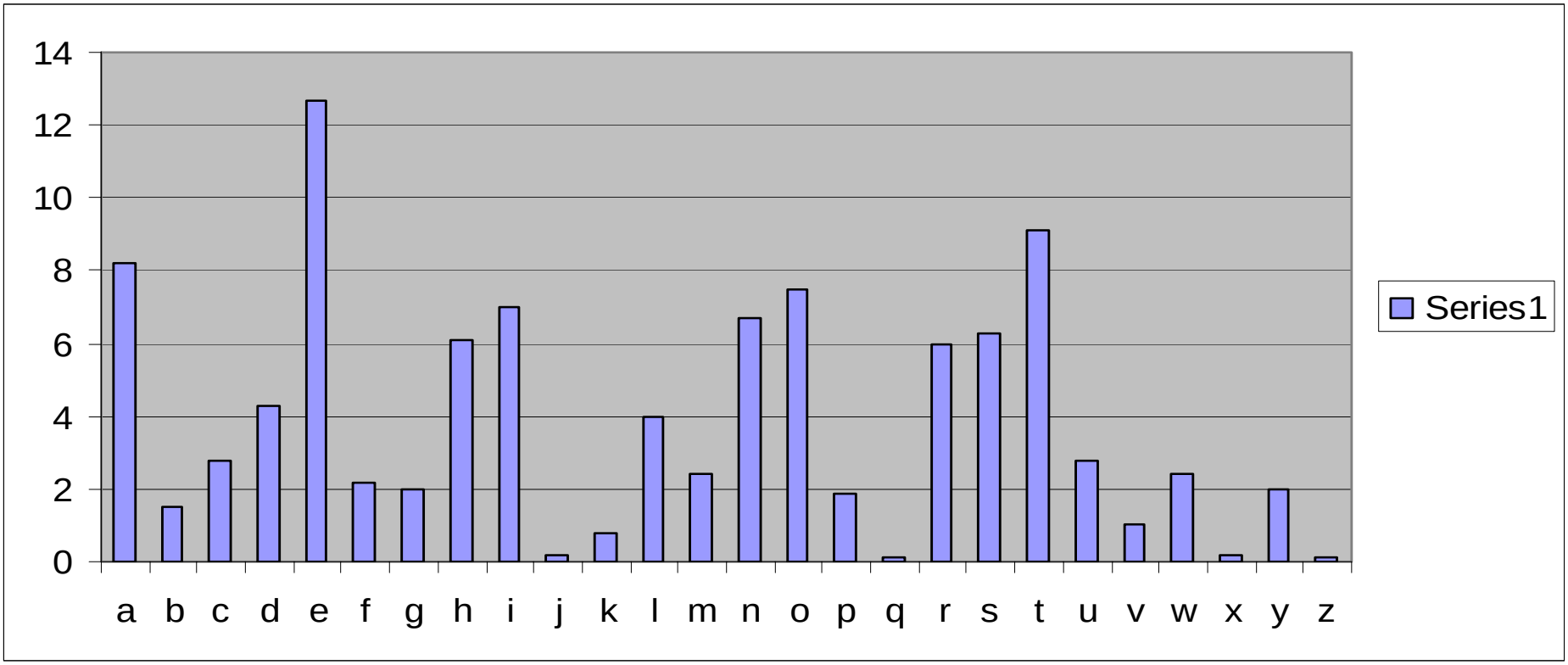
Strength of the General Substitution Cipher

- Exhaustive search is infeasible
  - key space size is $26! \approx 4 \times 10^{26}$
- Dominates the art of secret writing throughout the first millennium A.D.
- Thought to be unbreakable by many back then

Cryptanalysis of Substitution Ciphers: Frequency Analysis

- Basic ideas:
  - Each language has certain features: frequency of letters, or of groups of two or more letters.
  - Substitution ciphers preserve the language features.
  - Substitution ciphers are vulnerable to frequency analysis attacks.

Frequency of Letters in English

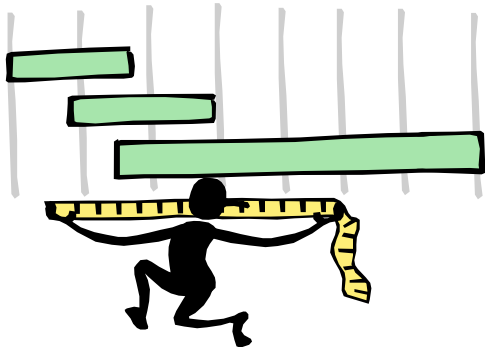


Other Frequency Features of English

- EN is the most common two-letter combination, followed by RE, ER, and NT.
- Vowels, which constitute 40 % of plaintext, are often separated by consonants.
- The letter A is often found in the beginning of a word or second from last. The letter I is often third from the end of a word.
- The letter Q is followed only by U
- And more …

Substitution Ciphers: Cryptanalysis

- The number of different ciphertext characters or combinations are counted to determine the frequency of usage.



- The cipher text is examined for patterns, repeated series, and common combinations.
- Replace ciphertext characters with possible plaintext equivalents using known language characteristics.

History

- Discovered by the Arabs
  - earliest known description of frequency analysis is in a book by the ninth-century scientist al-Kindi
- Rediscovered or introduced from the Arabs in the Europe during the Renaissance
- Frequency analysis made substitution cipher insecure

Ways to Improve the Security of Substitution Cipher

- Using nulls
  - e.g., using numbers from 1 to 99 as the ciphertext alphabet, some numbers representing nothing and are inserted randomly
- Deliberately misspell words
  - e.g., “Thys haz thi ifekkt off diztaughting thi ballans off frikwenseas”
- Homophonic substitution cipher
  - each letter is replaced by a variety of substitutes
- These make frequency analysis more difficult, but not impossible

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 Vigenere (published in 1586)

The Vigenere Cipher

Definition:

Given m, a positive integer, $P = C = (Z_{26})^n$, 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:

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

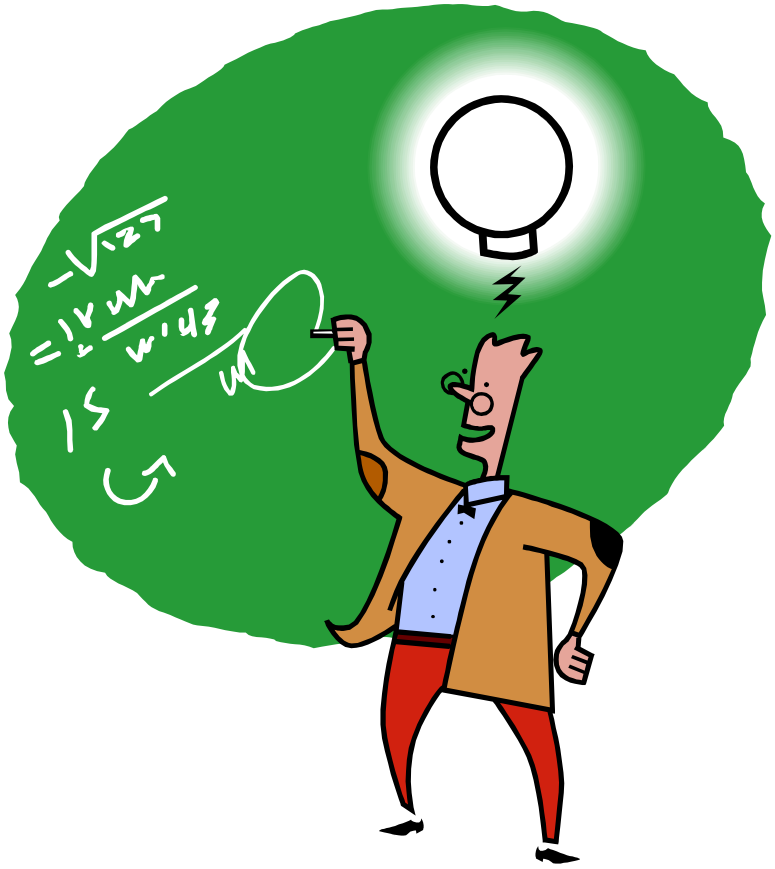
Ciphertext: N L A Z E I I B L J J I

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 simple substitution ciphers as there are letters in the key.

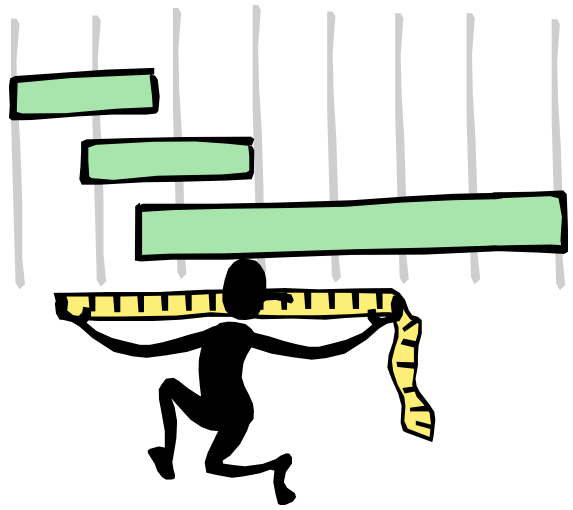
Vigenere Cipher: Cryptanalysis

- Find the length of the key.
- Divide the message into that many simple substitution encryptions.
- Use frequency analysis to solve the resulting simple substitutions.
  - how?



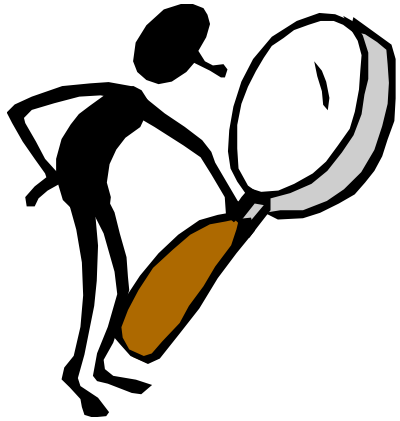
How to Find the Key Length?

- 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	B	U	K	W	I	A	O	X	B	U	K	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 $x = x_1x_2\ldots x_n$ is a string of n alphabetic characters. Then $I_c(x)$, the index of coincidence is:

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

Index of Coincidence (cont.)

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

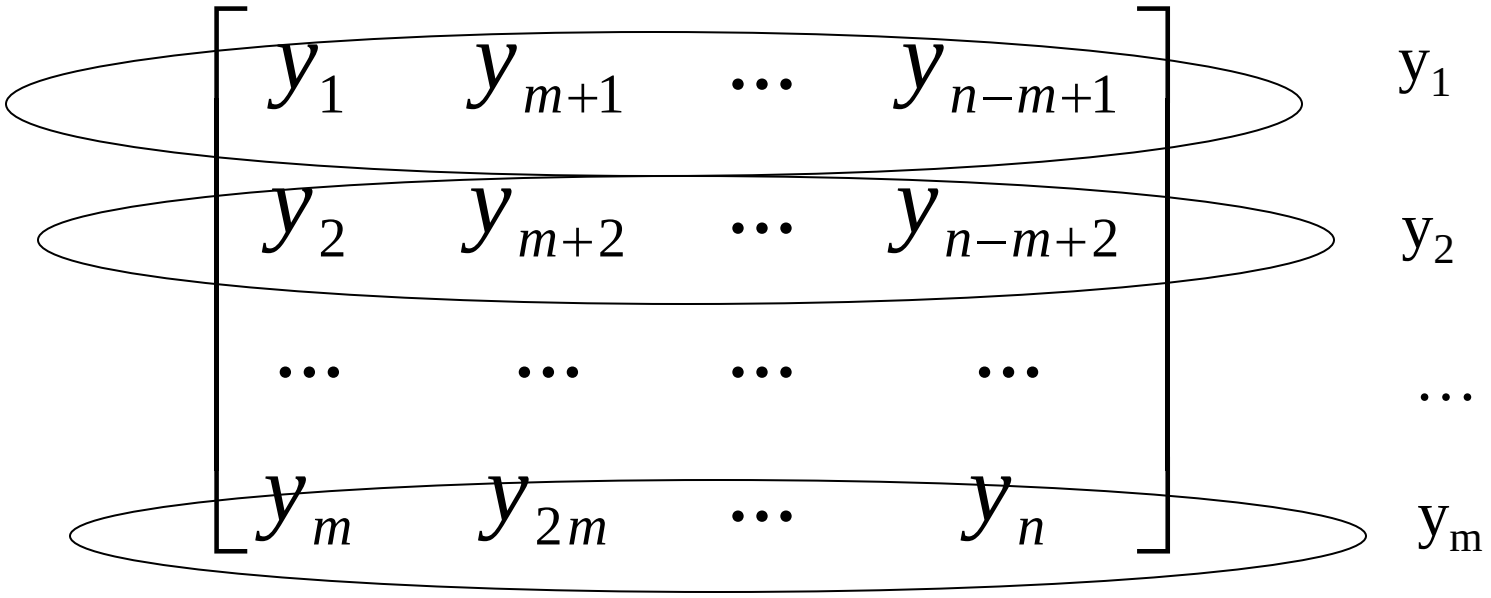
- For English, $S = 25$ and p_i can be estimated

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



Guessing the Key Length

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

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

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

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

Rotor Machines

- Basic idea: if the key in Vigenere cipher is very long, then the attacks won't work
- Implementation idea: multiple rounds of substitution
- A machine consists of multiple cylinders
  - each cylinder has 26 states, at each state it is a substitution cipher
  - each cylinder rotates to change states according to different schedule

Rotor Machines

- A m-cylinder rotor machine has
  - 26^m different substitution ciphers
    - $26^3 = 17576$
    - $26^4 = 456{,}976$
    - $25^5 = 11{,}881{,}376$

Enigma Machine

- Use 3 scramblers (motors): **17576 substitutions**
- 3 scramblers can be used in any order: 6 combinations
- Plug board: allowed 6 pairs of letters to be swapped before the encryption process started and after it ended.

$$\frac{26!}{14! \bullet 6! \bullet 64} = 100,391,791,500$$

- Total number of keys $\approx 10^{16}$

Using Enigma Machine

- A day key has the form
  - Plugboard setting: A/L–P/R–T/D–B/W–K/F–O/Y
  - Scrambler arrangement: 2-3-1
  - Scrambler starting position: Q-C-W
- Sender and receiver set up the machine the same way for each message
- Use of message key: a new scrambler starting position, e.g., PGH
  - first encrypt and send the message key, then set the machine to the new position and encrypt the message
  - initially the message key is encrypted twice

History of the Enigma Machine

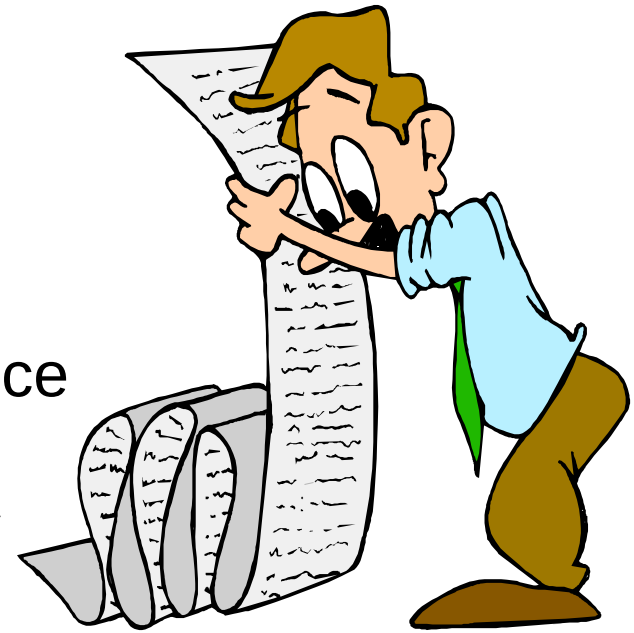
- Patented by Scherius in 1918
- Widely used by the Germans from 1926 to the end of second world war
- First successfully broken by the Polish's in the thirties by exploiting the repeating of the message key
- Then broken by the UK intelligence during the WW II

How the Polish Break Enigma Machine

- They have a copy of the machine
- Main idea: separating the effect of the plugboard setting from the starting position
- Exploiting the repetition of message keys

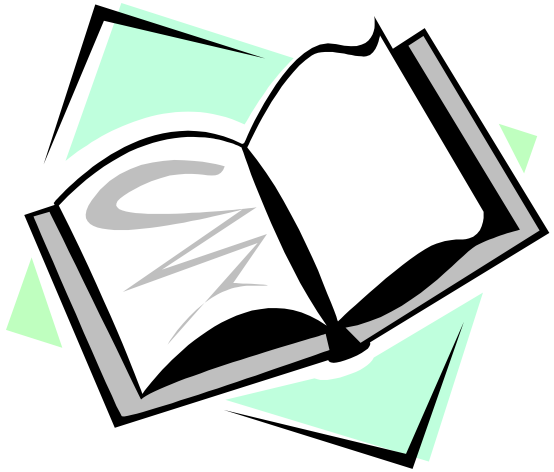
Summary

- Shift ciphers are easy to break using brute force attacks, they have small key space.
- Substitution ciphers vulnerable to frequency analysis attacks.
- Vigenere cipher is vulnerable: once the key length is found, a cryptanalyst can apply frequency analysis.



Recommended Reading for This Lecture

- Stinson
  - Section 1.1:
    - 0, 1, 2, 3, 4,
  - Section 1.2:
    - 0, 1, 2, 3,



- The Code Book by Simon Singh

Coming Attractions …

- Information-Theoretic secrecy (Perfect secrecy), One-Time Pad, Stream Ciphers

- Recommended reading for next lecture:

  Stinson 1.1.7, 1.2.5, 2.1, 2.2, 2.3