Introduction to Cryptography CS 355

Lecture 2

Classical Cryptography: Shift Cipher and Substitution Cipher

Announcements

- Join class mailing list
  - CS355_Fall2005@cs.purdue.edu
  - To join the list sent an email to mailer@cs.purdue.edu, with empty subject and the body containing the text 'add your_email to CS355_Fall2005'
- TA office hour:
  - Monday 9:30am to 10:30am
  - Wednesday 3:15pm to 4:15pm

Lecture Outline

- The Spartan scytale and transposition ciphers
- Shift and substitution ciphers.
- Frequency Analysis: attacks on substitution ciphers.

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

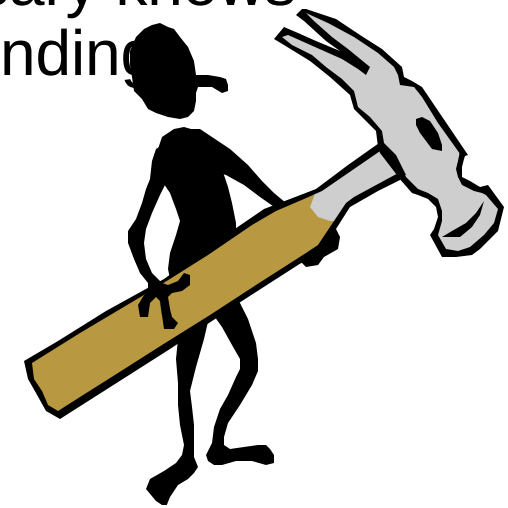
- A Cipher (K, P, C, **E**, **D**)
  - K : the key space
  - P : the plaintext space
  - C : the ciphertext space
  - **E**: $K \times P \rightarrow C$: the encryption function
  - **D**: $K \times C \rightarrow P$: the decryption function
    - Given a key K and a plaintext P,
      D(K, **E**(K,P))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

Cryptanalysis of Ciphers

- Goals:
  - recover the encryption key
  - decrypt a given message

Adversarial Models for Symmetric Ciphers

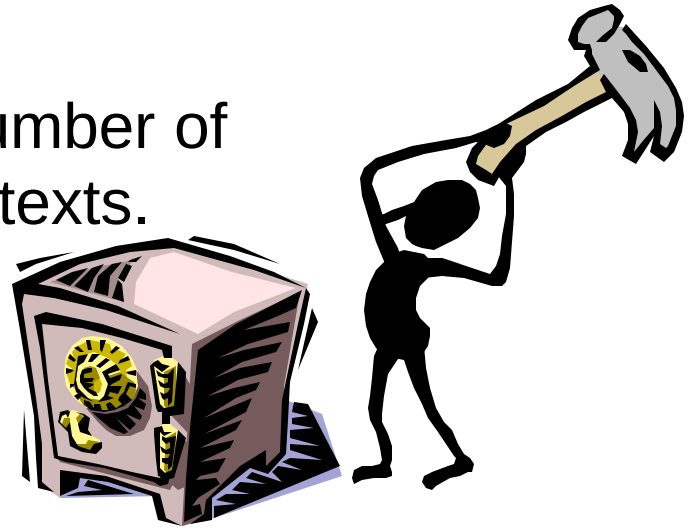
- The language of the plaintext and the nature of the cipher are assumed to be known to the adversary.
- **Ciphertext-only attack**: The adversary knows a number of ciphertexts.
- **Known-plaintext attack**: The adversary knows some pairs of ciphertext and corresponding plaintext.



Adversarial Models for Symmetric Ciphers

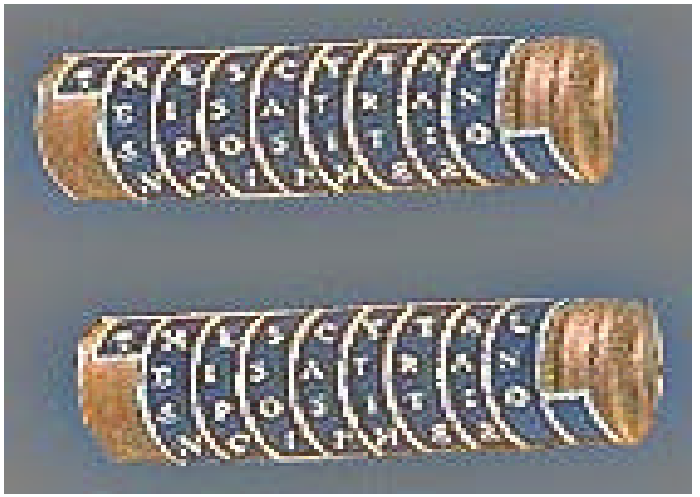
- **Chosen-plaintext attack**
  The adversary can choose a number of messages and obtain the ciphertexts for them

- **Chosen-ciphertext attack**
  The adversary 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
- 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: 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 $\Sigma = \{A, B, C, \ldots, Z\}$
- 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

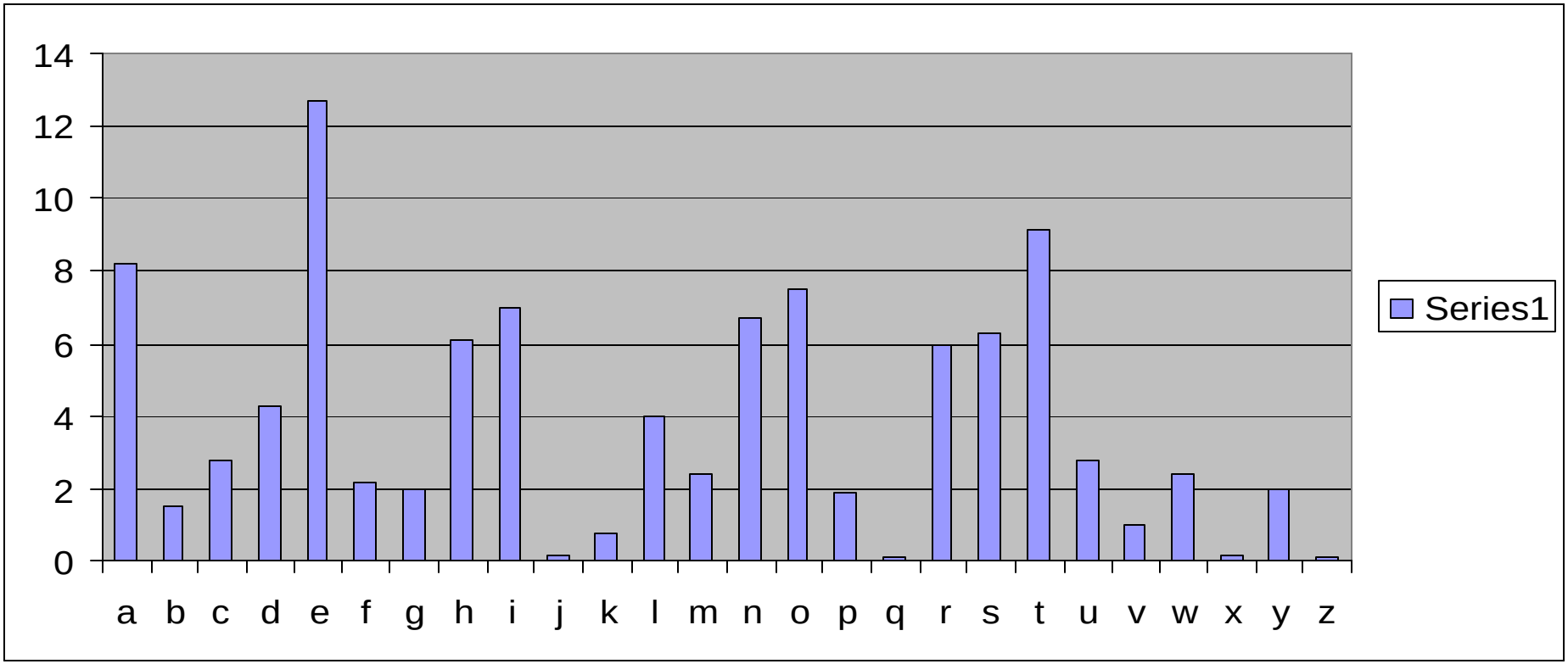
Strength of the General Substitution Cipher

- 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

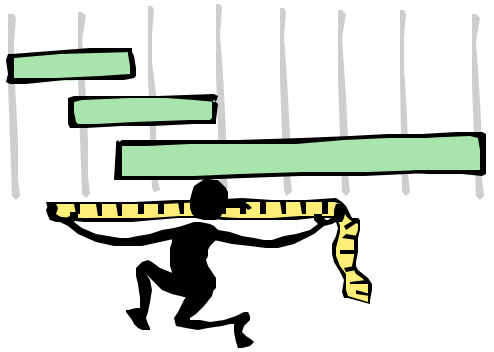


Other Frequency Features of English

- 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

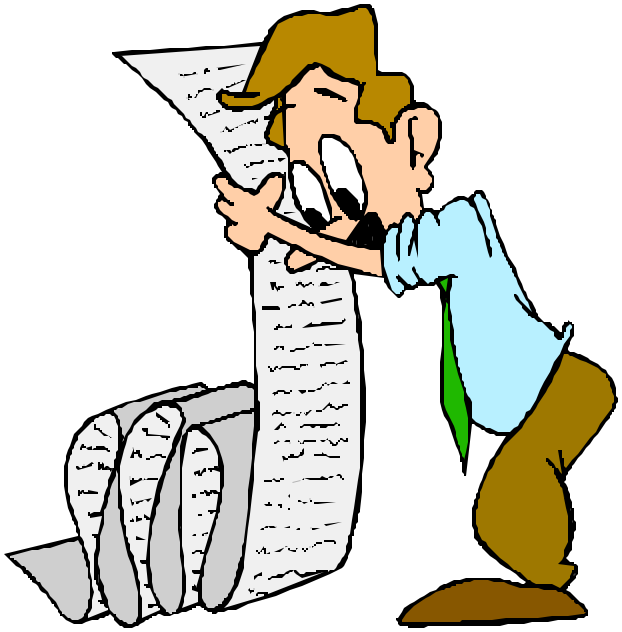
- Discovered by the Arabs
  - earliest known description of frequency analysis is in a book by the ninth-century scientist al-Kindi
- Rediscovered or introduced from the Arabs in the Europe during the Renaissance
- Frequency analysis made substitution cipher insecure

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

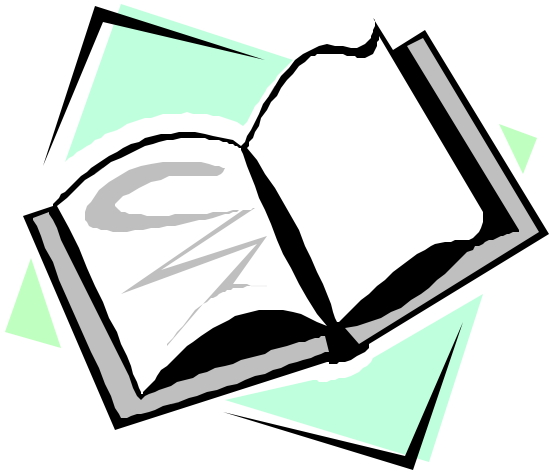
Summary

- Shift ciphers are easy to break using brute force attacks, they have small key space.
- Substitution ciphers vulnerable to frequency analysis attacks.



Recommended Reading for This Lecture

- The Code Book, Chapter 1



Coming Attractions …

- Basic modular arithmetic
- Affine cipher
- Recommended reading for next lecture:

  Trappe & Washington: 2.0, 2.1, 2.2