

CS355: Cryptography

Lecture 3: Shift cipher,
substitution cipher,
Vigenere cipher.

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 \square 2; $2+11 \bmod 26 = 13 \square$ N

R \square 17; $17+11 \bmod 26 = 2 \square$ C

...

N \square 13; $13+11 \bmod 26 = 24 \square$ 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 Mono-alphabetical Substitution Cipher

- The key space: all permutations of $\Sigma = \{A, B, C, \dots, Z\}$
- Encryption given a key σ :
 - each letter X in the plaintext P is replaced with $\sigma(X)$
- Decryption given a key σ :
 - each letter Y in the ciphertext P is replaced with $\sigma^{-1}(Y)$

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
$\sigma =$	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 σ AZDBJSZ

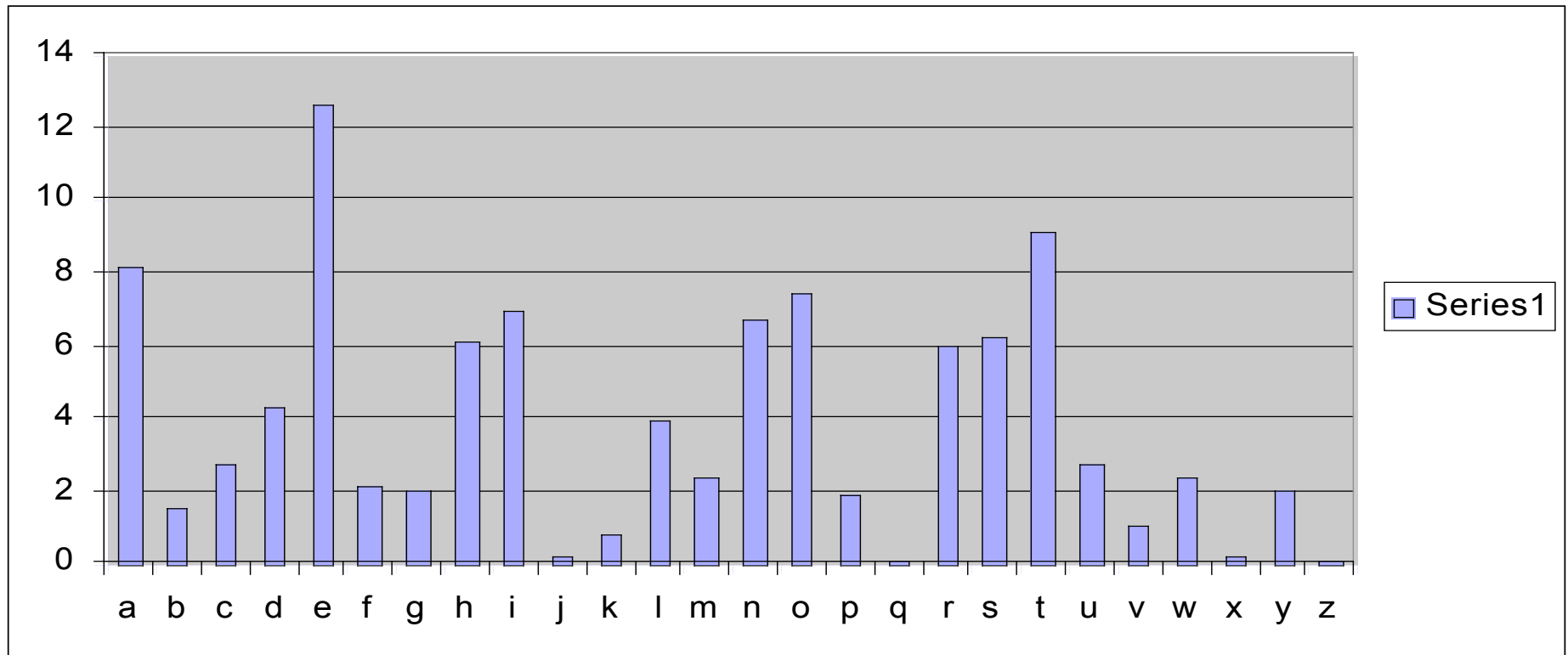
Strength of the General Substitution Cipher

- Exhaustive search is infeasible
 - key space size is $26! \approx 4 \cdot 10^{26}$
- 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

- Vowels, which constitute 40 % of plaintext, are often separated by consonants.
- Letter A is often found in the beginning of a word or second from last.
- Letter I is often third from the end of a word.
- 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.



Frequency Analysis History

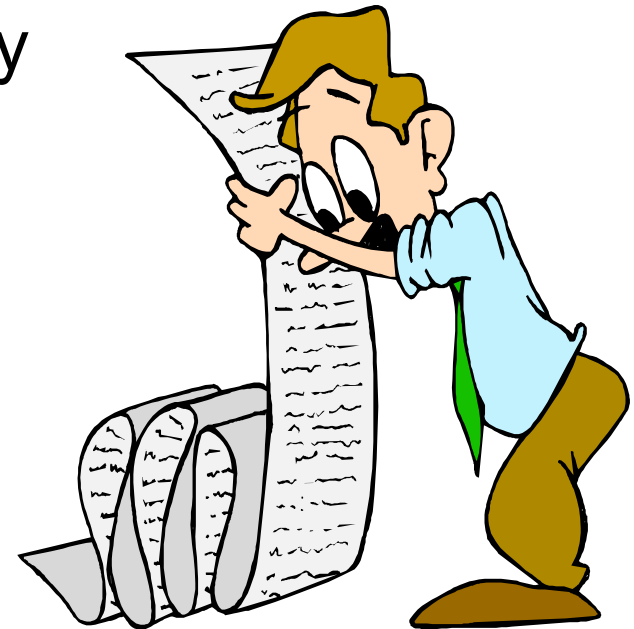
- Discovered by the Arabs
 - earliest known description of frequency analysis is in a book by the ninth-century scientist al-Kindi
- Rediscovered or introduced from the Arabs in the Europe during the Renaissance
- Frequency analysis made substitution cipher insecure

Improve the Security of Substitution Cipher

- Using nulls
 - e.g., using numbers from 1 to 99 as the ciphertext alphabet, some numbers representing nothing 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 preserve language features and are vulnerable to frequency analysis attacks.



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, $P = C = (\mathbb{Z}_{26})^n$, 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

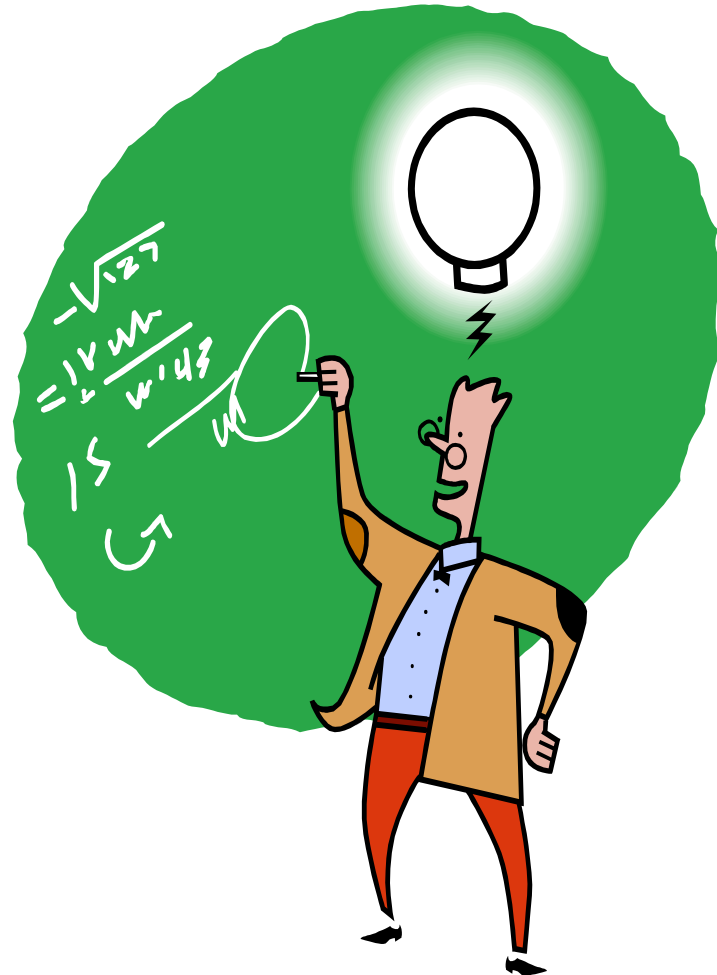
Security of Vigenere Cipher

- 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 ciphers** as letters in the key.



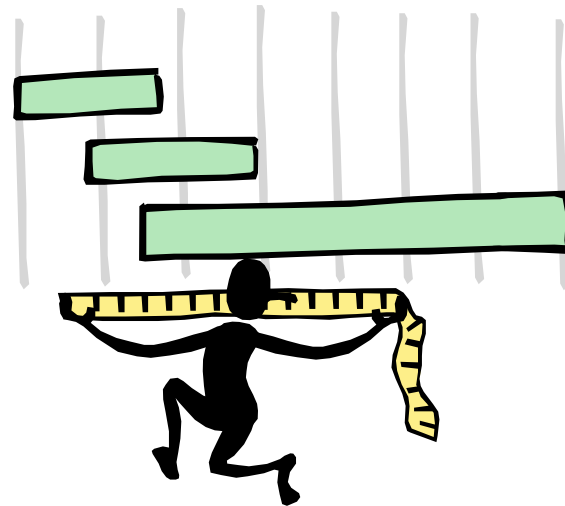
Vigenere Cipher: Cryptanalysis

- 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?



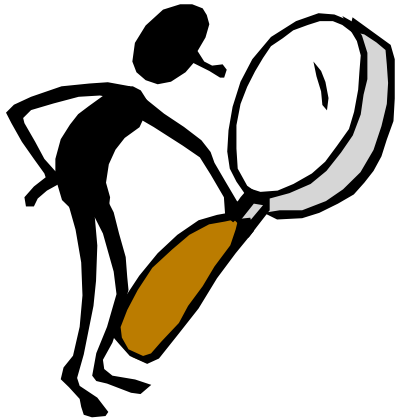
How to Find the Key Length?

- 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:
 - 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 , ($\Delta \equiv 0 \pmod{m}$), m is the key length).
- Algorithm:
 - Search for pairs of identical segments of length at least 3
 - Record distances between the two segments: $\Delta_1, \Delta_2, \dots$
 - m divides $\gcd(\Delta_1, \Delta_2, \dots)$



Recommended Reading for This Lecture

- Chapter 2.1, 2.3, 2.4

