Cryptography

Acknowledgment

- Based on slides by prof. Cristina Nita-Rotaru

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 Mono-alphabetical Substitution Cipher

- The key space: all permutations of Σ = {A, B, C, ..., Z}
- Encryption given a key (permutation) 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

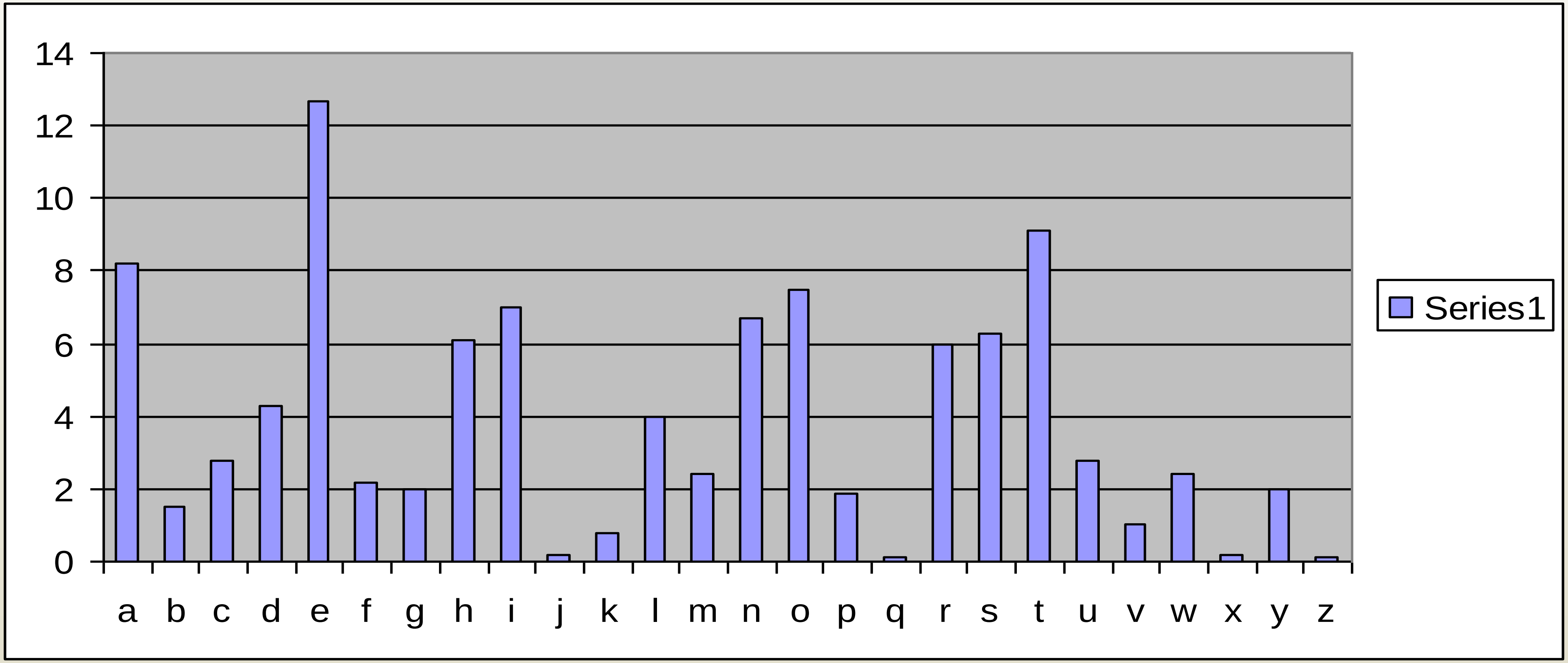
Strength of the General Substitution Cipher

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

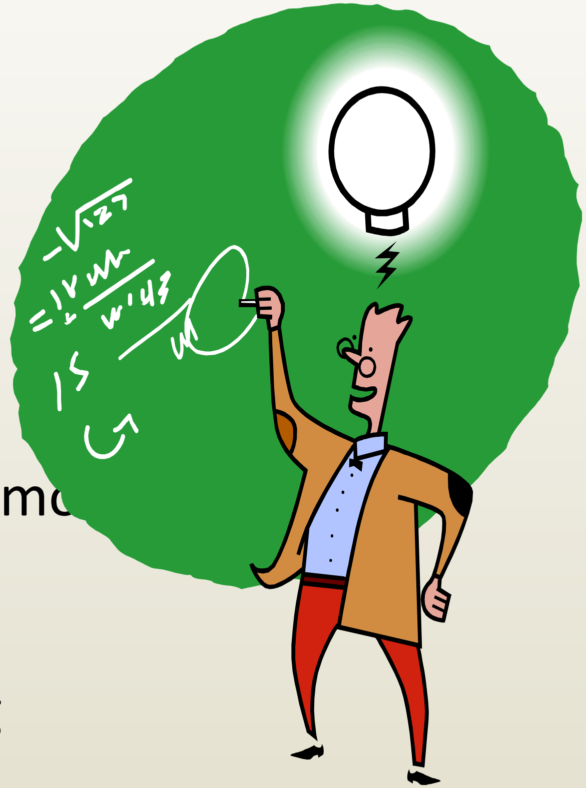


Other Frequency Features of English

- 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 commo combinations.
- Replace ciphertext characters with possible plaintext equivalents using known language characteristics.



Frequency Analysis History

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

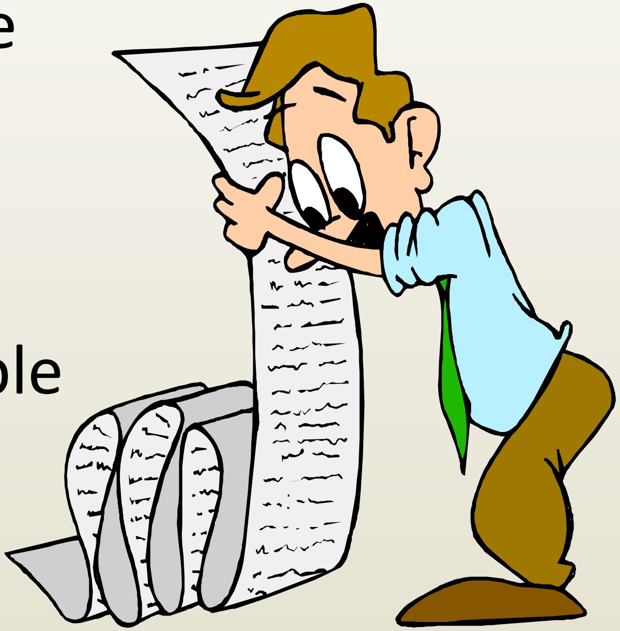
Improve the Security of Substitution Cipher

- 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 a key $K = (k_1, k_2, \dots, k_m)$ and a plain text message $P = (p_1, p_2, \dots, p_n)$

Encryption:

$c_i = (k_{i(\text{mod } m)} + p_i) \pmod{26}$, for $i = 1$ to n

Decryption:

$p_i = (e_i - k_{i(\text{mod } m)}) \pmod{26}$, for $i = 1$ to n

Example:

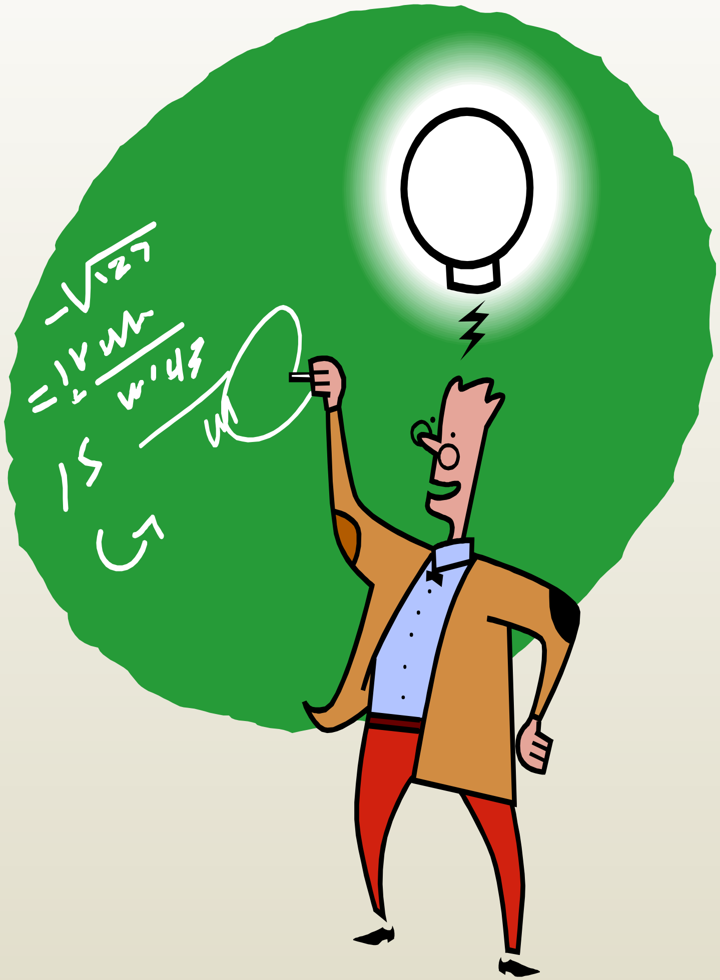
Plaintext P (n = 12):	C R Y P T O G R A P H Y
Key K (m = 4):	L U C K L U C K L U C K
Ciphertext C:	N L A Z E I I B L J J I

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 there are letters in the key.

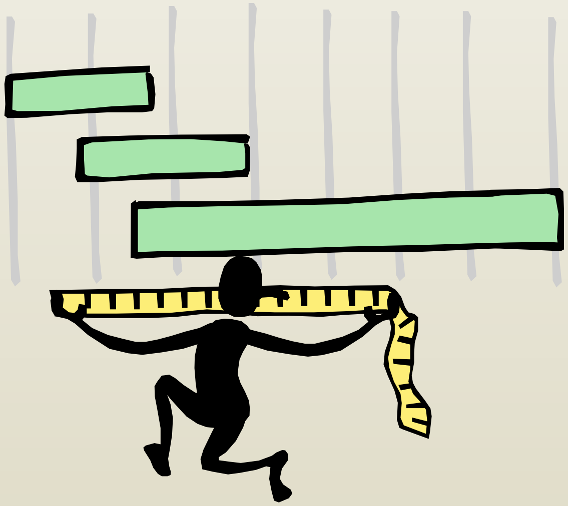
Vigenere Cipher: Cryptanalysis

- 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.
- One method to find the key length:
  - Kasisky test

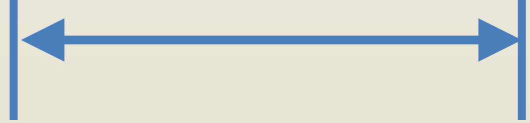


Kasisky Test

- Two identical segments of plaintext, will be encrypted to the same ciphertext, if the they occur at a distance that is a multiple of m
  - m is the key length
- Algorithm:
  - Search for pairs of identical segments of length at least 3
  - Record distances between the two segments: $\Delta 1, \Delta 2, \ldots$
  - m divides greatest common divisor of $\Delta 1, \Delta 2, \ldots$

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



8 characters =
4 x 2

Summary

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

Cryptography today—RSA algorithm

- Invented in **1978** by Ron **R**ivest, Adi **S**hamir and Leonard **A**dleman
- Security relies on the difficulty of factoring large composite numbers
  - Numbers with 1,024 bits
- Essentially the same algorithm was discovered in 1973 by Clifford Cocks, who works for the British intelligence