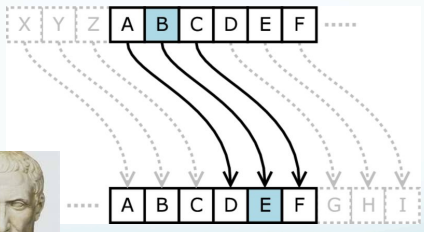
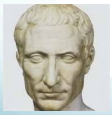
Substitution Ciphers

CMPUT 396

Caesar cipher





Julius Caesar

Caesar's memoirs

- Cicero receives Caesar's message (*De Bello Gallico* V:XLVIII:IX i.e. *The Gallic Wars* 5.48.9)
- Original Latin: *Ille perlectam in conventu militum recitat maximaque omnes laetitia adficit.*
- Google Translate: *He read it in the assembly of the greatest joy.*
- Human translation: *He, after perusing it, reads it out in an assembly of the soldiers, and fills all with the greatest joy.*

Caesar's cipher wheel

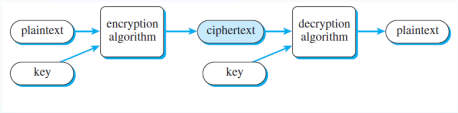
Each ciphertext letter is "the sum" of the plaintext letter and the shift value:

$C_i = M_i + K \bmod 26$

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

I	W	T		C	T	L		E	P	H	H	L	D	G	S		X	H		H	L	D	G	S	U	X	H	W
↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
T	H	E		N	E	W		P	A	S	S	W	O	R	D		I	S		S	W	O	R	D	F	I	S	H

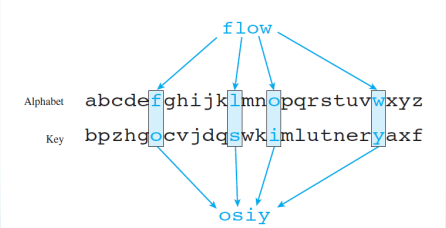
Encryption using a key



Secret Writing

- Steganography (hidden)
- Cryptography (encrypted)
  - Transposition (by reordering)
  - Substitution (by replacing)
    - Code (replace words)
    - Cipher (replace letters)

Substitution cipher



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

Security

- Kerckhoffs' Principle: The security of a cipher should depend only on keeping secret the key.
- The number of possible keys in a substitution cipher is:

  26! = 400,000,000,000,000,000,000,000,000

- If we could check 1,000,000 every second, it would take 12 trillion years.
- Later in this course, we will see how to break such ciphers within minutes.

In-class Exercise

- Decrypt this short message that young Alan Turing receives from his friend Christopher:

  AYCD2HVG

- Christopher used the following pangram as key:

"Quartz jock vends BMW glyph fix."

ABCDEFGHIJKLMNOPQRSTUVWXYZ

QUARTZJOCKVENDSBMWGLYPHFIX

- *See you in two long weeks, dearest friend.*

Creating a Key

- Random key is hard to remember
- Base key on a word or a short phrase
- For example, JULIUS CAESAR
- Remove repeated letters: JULISCAER
- Add remaining letters:

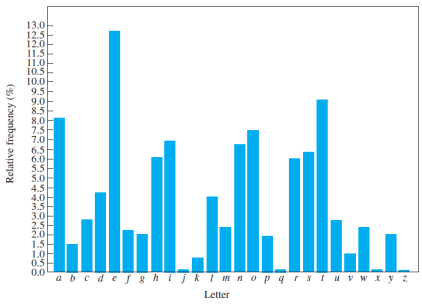
ABCDEFGHIJKLMNOPQRSTUVWXYZ

JULISCAERTVWXYZBDFGHKMNOPQ

The Arab Cryptanalysts

- The substitution cipher was unbreakable for centuries
- 610 AD Muhammmad's revelations
- 750 AD Abassid caliphate – golden age
- Flourishing of arts and science
- Papermaking acquired from the Chinese
- Translated and preserved Greek classics
- Administration: security of communications
- Scholars studied religious texts
- Ismail al-Kindi (801-873 AD) - cryptography

Relative letter frequencies



Cryptanalysing a ciphertext

- match relative letter frequencies in ciphertext to those in a large plain text
- letter doubles: *ss ee tt ff ll mm oo*
- 2-letter words: *of to in it is*
- 3-letter words: *the and*
- frequent bigrams: *th er he*
- guess words/phrases
- consonants vs. vowels

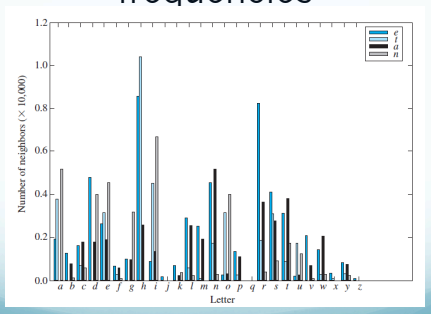
Pattern equivalence

- Word substrings are **pattern-equivalent** if there exists a monoalphabetic substitution that transforms one into the other:
  - 'will' is p-equivalent to 'jazz'
  - 'will' is not p-equivalent to 'said'
  - 'am I not' is p-equivalent to 'in a red'
- Formally, two substrings u and v are pattern-equivalent if and only if they satisfy the following three conditions:
  1. $|u| = |v|$
  2. $\forall i: u_i = \square \Leftrightarrow v_i = \square$
  3. $\forall i,j: u_i = u_j \Leftrightarrow v_i = v_j$

Substitution hacker program

- Find the word pattern for each cipherword in the ciphertext.
- Find the English word candidates that each cipherword could decrypt to.
- Create a dictionary showing potential decryption letters for each cipherletter to act as the cipherletter mapping for each cipherword.
- Combine the cipherletter mappings into a single mapping, which we'll call an intersected mapping.
- Remove any solved cipherletters from the combined mapping.
- Decrypt the ciphertext with the solved cipherletters.

Some letter pair frequencies



The Vigenere Cipher

- The idea of the Vigenere Cipher is to use a different key for each letter of the message.
- Unlike substitution cipher, the Vigenere cipher cannot be easily broken by frequency analysis.
- Invented in 1562, it was called "le chiffre indechiffrable" ("the indecipherable cipher").
- It was finally broken in 1854 by Charles Babbage, "the father of computers".

Vigenere Cipher

- The Vigenere cipher is like Caesar cipher, but with multiple keys/shifts.
- The keyword is aligned with the message:

```
Message: thesunandthemoon
    Key: KINGKINGKINGKING
 Cipher: DPRYEVNTXBUKWWBT
```

- Each ciphertext letter is "the sum" of the keyword letter and the plaintext letter:

$$C_i = (M_i + K_i) \bmod 26$$

The Vigenere square

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