CS-3002 Information Security

Lecture # 2: Introduction to Cryptography and Classical Cryptography

Prof. Dr. Sufian Hameed
Department of Computer Science
FAST-NUCES

Cryptography

- **Cryptography** (*kryptos*: secret; *graphein*: writing)
  = art and science of keeping information secure
  ↦ protection of confidentiality and integrity
- **Cryptanalysis** = study of attacks against cryptography
- **Cryptology** = cryptography and cryptanalysis
- **Steganography** (*steganos*: covered; *graphein*: writing)
  = art and science of hiding information
  ↦ deniability and unobservability of communication

Examples

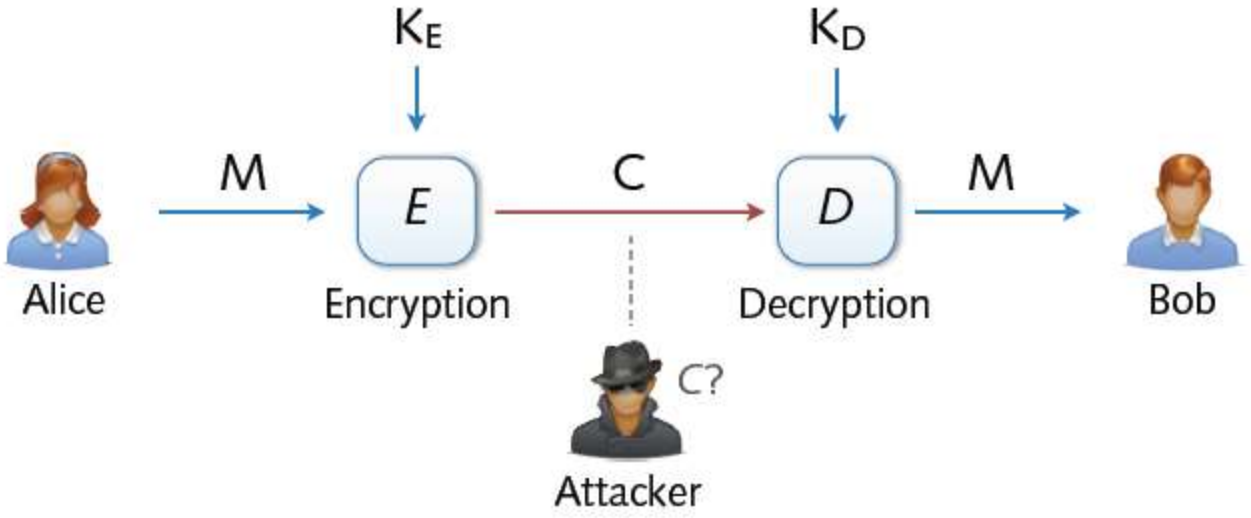
Cryptography

Message encrypted using the Enigma during WW2

Steganography

Message of 500 bytes hidden within image

Cryptosystem



- **Cryptographic system for en/decrypting messages**
  - M = plaintext message C = ciphertext message
  - K_E = encryption key K_D = decryption key

Attacks against Cryptosystems

1.) Cipher text-only: Attacker possesses a string y of the cipher text

2.) Known plaintext: Attacker possesses a string x of the plaintext and the corresponding cipher text y. The problem now is to find out the key which produces y from x

3.) Chosen plaintext: Attacker has access to the encryption machinery. Hence he can chose a plaintext string x and construct the corresponding cipher text string y.

4.) Chosen cipher text: Attacker has access to the decryption machinery. Hence, he can chose a cipher text string y and construct the corresponding plaintext string x.

Security of Keys

- **Kerckhoffs's Principle**
  - Cryptosystem is known, security depends on key only
  - Contrasting concept: "security by obscurity"
- **Keyspace defined over bits of key**
  - n-bit key $\mapsto$ size of keyspace 2^n
  - Time of brute-force attacks grows exponentially in n

Bits of key	10^9 checks per second	Cluster of 100,000 nodes
16	0.07 milliseconds	0.000 milliseconds
32	4.29 seconds	0.004 milliseconds
64	585 years	5 hours
128 considered secure	10^{22} years	10^{16} years

Cryptography is everywhere

Secure communication:

- web traffic: HTTPS
- wireless traffic: 802.11i WPA2 (and WEP), GSM, Bluetooth

Encrypting files on disk:

- EFS (Encrypting File System)
- TrueCrypt (open-source disk encryption software)

Content protection

- DVD --- Content Scramble System (**CSS**) is a Digital Rights Management (DRM) and encryption system employed on almost all commercially produced **DVD**-Video
  - Easy to break
- Blu-Ray --- Advance Access Content System (AACS)

User authentication

… and much much more

Things to remember

Cryptography is:

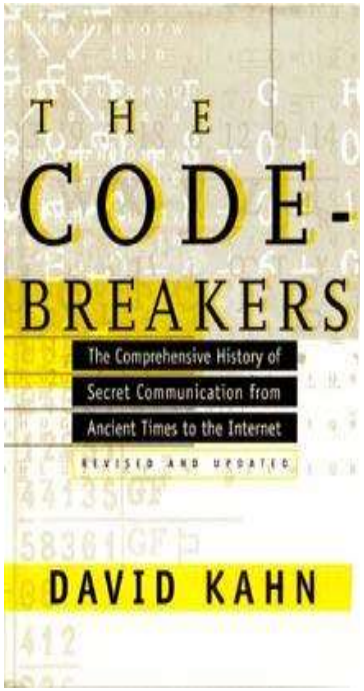
- A tremendous tool
- The basis for many security mechanisms

Cryptography is not:

- The solution to all security problems
  - Software bugs
  - Social engineering attacks
- Reliable unless implemented and used properly
  - Wired Equivalent Privacy (WEP -- good example on how not to use cryptography)
- ***Something you should try to invent yourself***
  - many examples of broken ad-hoc designs
  - Propriety ciphers, once re-engineered are easily broken

History

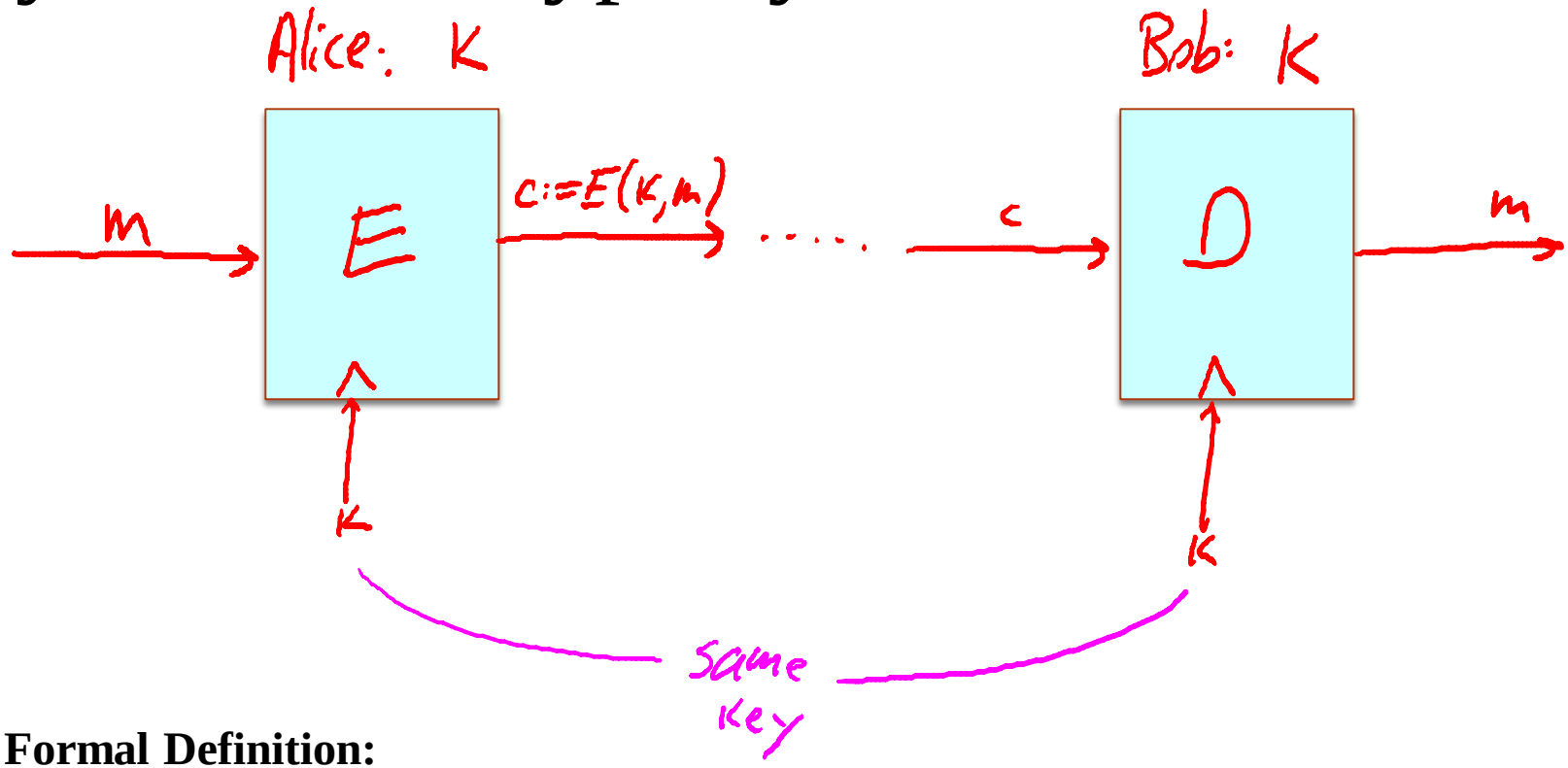
David Kahn, "The code breakers" (1996)



Historical Cryptosystems

- **Monoalphabetic cipher:** Each alphabetic character is mapped onto a unique alphabetic character
  - Examples: Shift Cipher, Substitution Cipher, Affine Cipher
- **Polyalphabetic cipher:** Each alphabetic character is mapped onto various alphabetic characters
  - Examples: Vigenere Cipher, Hill Cipher, Permutation Cipher

Symmetric Cryptosystems



Formal Definition:

Cryptosystem is defined over (K,M,C) and a pair of "efficient" algorithms (***E***, ***D***) s.t.

$$\forall\ m \in M,\ k \in K \text{ and } c \in C : \boldsymbol{E}(k, m) = c,\ \boldsymbol{D}(k, \boldsymbol{E}(k, m)) = m$$

Efficient means run in polynomial time

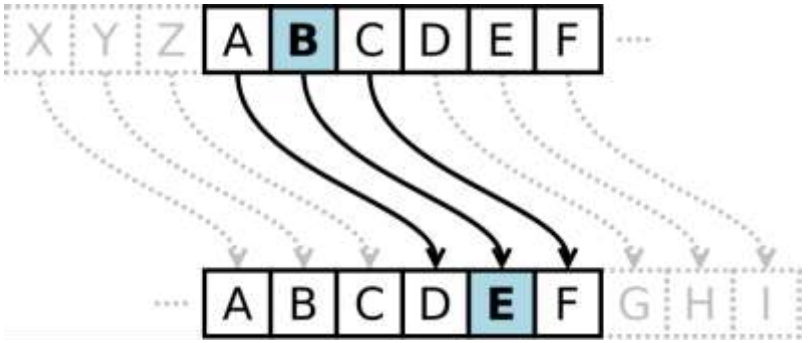
Shift Cipher

- Cipher in which each letter in the plaintext is replaced by a letter some fixed number of positions down the alphabet. Example includes Ceasar cipher, ROT13
- **Ceasar Cipher**
  - Each letter is replaced with a fixed shift of 3 letters

Example of Ceasar cipher using left rotation of 3 places

Plain: ABCDEFGHIJKLMNOPQRSTUVWXYZ

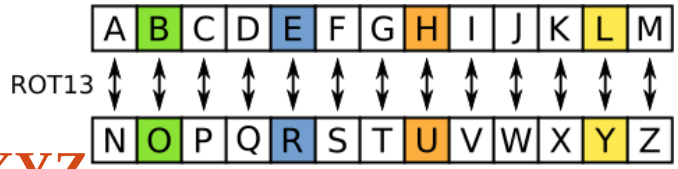
Cipher: DEFGHIJKLMNOPQRSTUVWXYZABC



Source: wikipedia

Shift Cipher (ROT13)

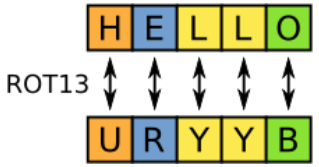
- **ROT13**
  - Each letter is replaced with a fixed shift of 13 letters



The transformation can be done as follows

Plain: ABCDEFGHIJKLMNOPQRSTUVWXYZ

Cipher: NOPQRSTUVWXYZABCDEFGHIJKLM



Modular arithmetic representation:

- Encryption of a letter **x** by a shift ***n*** can be described mathematically as

  $E_n(x) = (x+n) \bmod 26$

- Decryption is performed in a similar manner

  $D_n(x) = (x-n) \bmod 26$

Key space is ridiculously small, very easy to break

Source: wikipedia

Substitution Cipher

Idea: *use a permutation over the set of characters as key to get a more flexible scheme as in the shift cipher*

- Keyspace significantly larger
- Character frequencies are preserved

• a	• b	• c	• d	• e	• f	• g	• h	• i	• j	• k	• l	• m
• F	• G	• N	• E	• A	• T	• X	• Z	• O	• I	• Q	• B	• H
• n	• o	• p	• q	• r	• s	• t	• u	• v	• w	• x	• y	• z
• K	• Y	• W	• V	• C	• P	• J	• L	• S	• D	• M	• U	• R

computerscience → NYHWLJACPNOAKNA

What is the size of key space in the substitution cipher assuming 26 letters?

$|\mathcal{K}| = 26$

$|\mathcal{K}| = 26!$ (26 factorial)

$|\mathcal{K}| = 2^{26}$

$|\mathcal{K}| = 26^2$

$26! \simeq 2^{88}$

Breaking Monoalphabetic Ciphers

Monoalphabetic ciphers preserve the frequency of alphabetic characters, pairs, etc.

→ *Identify alphabetic characters due to their frequency*

Method to decipher natural languages:

1. Determine frequency of alphabetic characters of the cipher text
2. Identify alphabetic characters according to their frequency: *e*, *n*, *i*, *s*, *r*, *a*, *t* *(in Germany: e, n, r, i, s, t, u, d, a, g, l, o, ...)*
3. Determine frequency of pairs
4. Identify e.g. *th he*
5. Look at identified text, re-substitute, guess, ...

Breaking Monoalphabetic Ciphers

letter	probability
a	.082
b	.015
c	.028
d	.043
e	.127
f	.022
g	.020
h	.061
i	.070
j	.002
k	.008
l	.040
m	.024

letter	probability
n	.067
o	.075
p	.019
q	.001
r	.060
s	.063
t	.091
u	.028
v	.010
w	.023
x	.001
y	.020
z	.001

Partition into five groups:

1. **E**, having probability about 0.12
2. **T,A,O,I,N,S,H,R**, each having probabilities between 0.06 and 0.09
3. **D,L**, each having probabilities around 0.04
4. **C,U,M,W,F,G,Y,P,B**, each having probabilities between 0.015 and 0.028
5. **V,K,J,X,Q,Z**, each having probabilities less than 0.01

Digram frequencies

th	.0315	in	.0169
he	.0251	er	.0154
an	.0172	re	.0148

Vigenere Cipher

Popular polyalphabetic substitution cipher

- Known as “le chiffre indéchiffrable” (‘the indecipherable cipher’);-)
- Combination of simple substitution ciphers
- Rotations determined by a word (key)

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

Message	T	H	I	S	A	T	E	S	T
Running key	K	E	Y	K	E	Y	K	E	Y
	+10	+4	+24	+10	+4	+24	+10	+4	+24
Ciphertext	D	L	G	C	E	R	O	W	R

Polyalphabetic

Period

Breaking Vigenere Cipher

- Frequency analysis trivial if period can be guessed
- **Kasiski test**
  - Repeated words may, by chance, sometimes be encrypted using the same key letters, leading to repeated groups in the ciphertext
  - Consider the following encryption using the keyword ABCD

    Key: ABCD ABCD ABCD ABCD ABCD ABCD ABCD

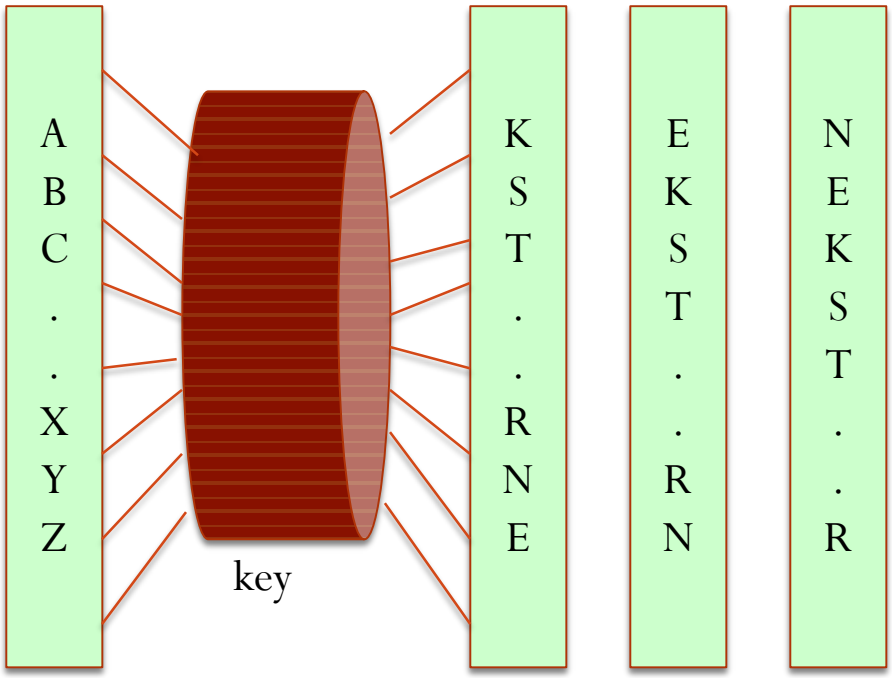
    Plaintext: **CRYP TO**IS SHOR TFOR **CRYP TO**GR APHY

    Ciphertext: ***CSAS TP***KV SIQU TGQU ***CSAS TP***IU AQJB
  - Repetitions of CSASTP is at a distance16
  - Assuming that the repeated segments represent the same plaintext segments, this implies that the key is 16, 8, 4, 2, or 1 characters long

Source: wikipedia

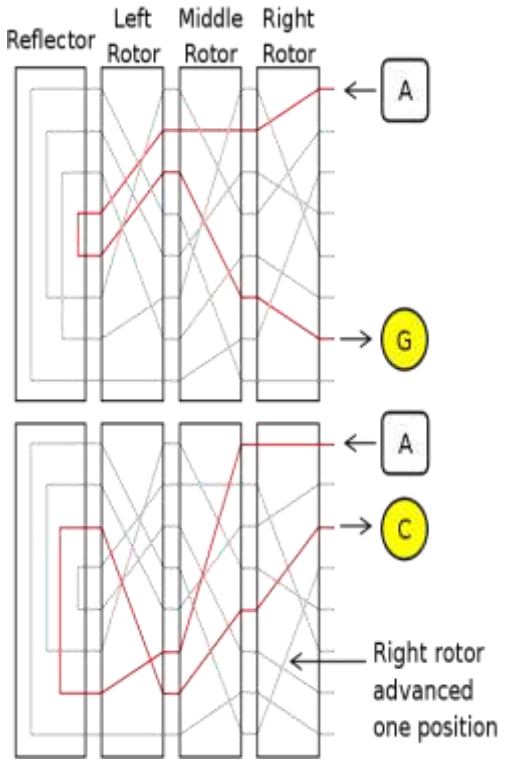
Rotor Machines (1870-1943)

- The Hebern Machine (single rotor)
  - Easily broken (CT only) using letter frequency, diagram frequency, trigram frequency



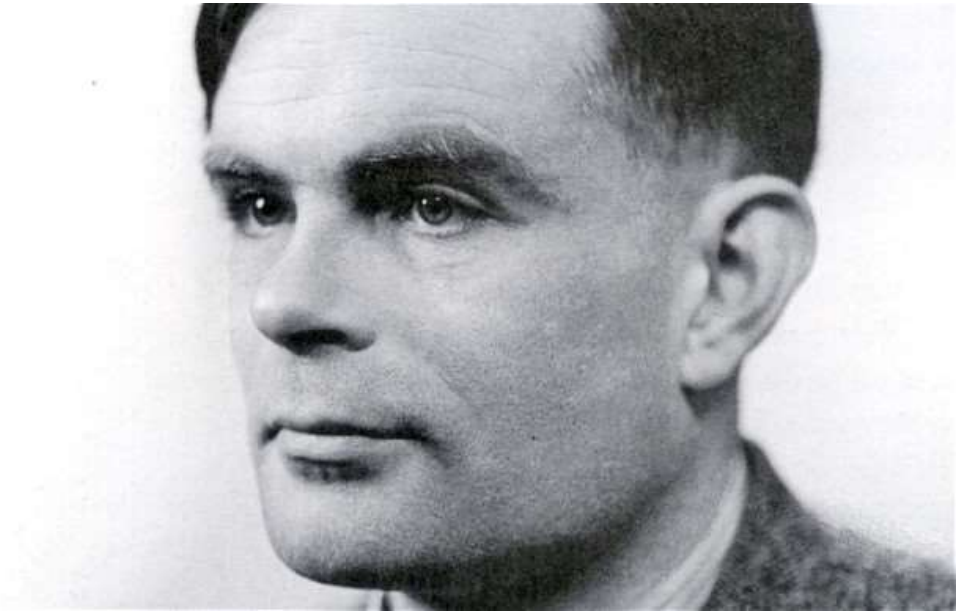
Rotor Machines (cont.)

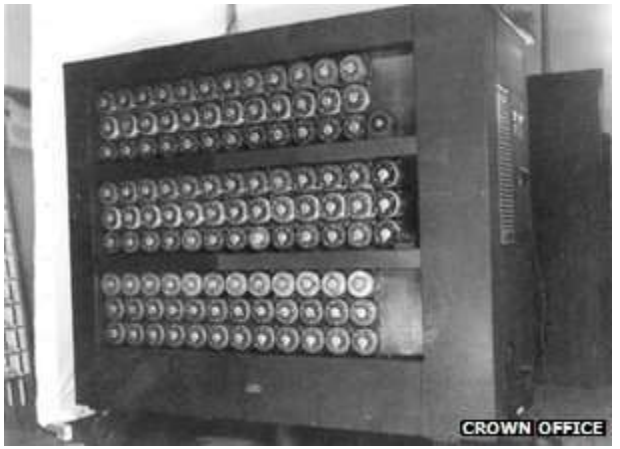
Most famous: the Enigma (3-5 rotors)



With 4 rotors keys = $26^4 = 2^{18}$ (actually 2^{36} due to optional plugboard)

Turing Bombe





Must watch

“The Man Who Cracked Enigma”

Acknowledgements

Material in this lecture are taken from the slides prepared by:

- Prof. Dr. Konrad Rieck (Uni-Göttingen)
- Prof. Dan Boneh (Stanford)