Basic Ciphers
Ahmet Burak Can
Hacettepe University
abc@hacettepe.edu.tr

Books
- Textbook:
- Supplementary books:

Outline of the Course
- Basic ciphers
- Block ciphers, Encryption modes and Stream ciphers
- Hash functions, message digests, HMAC
- Number Theory, Public Key Cryptography, RSA
- Digital certificates and signatures, X509
- Authentication: Two-Three factor authentication, Biometrics, Smart Cards
- Security Handshake
- Real-time Communication Security, SSL/TLS, IPSEC
- Kerberos

Outline of the Course
- Threshold cryptography
- Operating System Security
- Malicious Software: Trojans, logic bombs, viruses, worms, botnets, rootkits, trapdoors and cover channels
- Firewalls, VPNs, Intrusion detection systems
- If time permits:
  - Program Security
  - HTTP and Web Application Security, XSS
  - Wireless Security: WEP and WPA

Which Security Concept?

Information Security
- Computer Security:
  - Ensure security of data kept on the computer
- Network Security:
  - Ensure security of communication over insecure medium
- Approaches to Secure Communication
  - Steganography
    - hides the existence of a message
  - Cryptography
    - hide the meaning of a message
Basic Security Goals

- Privacy (secrecy, confidentiality)
  - only the intended recipient can see the communication
- Authenticity (integrity)
  - the communication is generated by the alleged sender
- Authorization
  - limit the resources that a user can access
- Availability
  - make the services available 99.999...% of time
- Non-repudiation
  - no party can refuse the validity of its actions
- Auditing
  - Take a log of everything done in the system

Basic Terminology in Cryptography – 1

- Cryptography: the study of mathematical techniques related to aspects of providing information security services.
- Cryptanalysis: the study of mathematical techniques for attempting to defeat information security services.
- Cryptology: the study of cryptography and cryptanalysis.

Basic Terminology in Cryptography – 2

- Encryption (encipherment): the process of transforming information (plaintext) using an algorithm (cipher) to make it unreadable to anyone except those possessing special knowledge
- Decryption (decipherment): the process of making the encrypted information readable again
- Key: the special knowledge shared between communicating parties
- Plaintext: the data to be concealed.
- Ciphertext: the result of encryption on the plaintext

Encryption & Decryption

![Encryption and Decryption Diagram](image)

Breaking Ciphers

- There are different methods of breaking a cipher, depending on:
  - the type of information available to the attacker
  - the interaction with the cipher machine
  - the computational power available to the attacker

Breaking Ciphers – Attack Types

- Ciphertext-only attack: The cryptanalyst knows only the ciphertext. Sometimes the language of the plaintext is also known.
  - The goal is to find the plaintext and the key.
  - Any encryption scheme vulnerable to this type of attack is considered to be completely insecure.
- Known-plaintext attack: The cryptanalyst knows one or several pairs of ciphertext and the corresponding plaintext.
  - The goal is to find the key used to encrypt these messages or a way to decrypt any new messages that use that key.
Breaking Ciphers - Attack Types

- **Chosen-plaintext attack**: The cryptanalyst can choose a number of messages and obtain the ciphertexts for them.
  - The goal is to deduce the key used in the other encrypted messages or decrypt any new messages using that key.
- **Chosen-ciphertext attack**: Similar to the chosen-plaintext attack, but the cryptanalyst can choose a number of ciphertexts and obtain the plaintexts.

Today's Ciphers

- Shift Cipher
- Mono-alphabetical Substitution Cipher
- Polyalphabetical Substitution Ciphers
- Rotor Machine
- Enigma

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 key K:**
  - shift left
- **History:**
  - K = 3, Caesar's cipher

Shift Cipher: An Example

- **Ciphertext attack**
- **Decryption given a key ̟:**
  - key space is small (≤ 26 possible keys)
  - P = CRYPTOGRAPHY IS FUN
  - K = 11
  - C = NCJAVZRCLAS JTDQFY

  \[
  \begin{align*}
  C &\rightarrow 2 & 2+11 \mod 26 = 13 &\rightarrow N \\
  R &\rightarrow 17 & 17+11 \mod 26 = 2 &\rightarrow C \\
  \ldots \\
  N &\rightarrow 13 & 13+11 \mod 26 = 24 &\rightarrow Y \\
  \end{align*}
  \]

Shift Cipher: Cryptanalysis

- Can an attacker find K?
  - YES: exhaustive search,
  - key space is small (≤ 26 possible keys)
  - the attacker can search all the key space in very short time
- 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 π:**
  - each letter X in the plaintext P is replaced with π(X)
- **Decryption given a key π:**
  - each letter Y in the ciphertext P is replaced with π⁻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 Q X S V T R N M S K J I P F E U

BECAUSE → AZDJSZ
**General Substitution Cipher: Cryptanalysis**

- Exhaustive search is infeasible
  - for the letter A, there are 26 probabilities
  - for the letter B, there are 25 probabilities
  - for the letter C, there are 24 probabilities
  - ... and so on
- Key space size is $26! \approx 4 \times 10^{26}$

**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.
- History of frequency analysis:
  - 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 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.
- Some words are more frequent, such as the, and, at, is, on, in

**Cryptanalysis using Frequency Analysis**

- 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 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., "Thy hox thi ifekett off ditsaugthng 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.
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, \( P = \mathbb{Z}_m^* \), and \( K = (k_1, k_2, \ldots, k_n) \)
  - a key, we define:
- **Encryption:**
  - \( E_k(p_1, p_2, \ldots, p_n) = (p_1+k_1, p_2+k_2, \ldots, p_n+k_n) \pmod{26} \)
- **Decryption:**
  - \( D_k(c_1, c_2, \ldots, c_n) = (c_1-k_1, c_2-k_2, \ldots, c_n-k_n) \pmod{26} \)

Example:
- Plaintext: CRYPTOGRAPHY
- Key: LUCKLUCKLUCK
- Ciphertext: NLAZEIBLJII

Security of Vigenère Cipher

- Vigenère 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 Vigenère cipher is a collection of as many shift ciphers as there are letters in the key.

Vigenère 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.
- Vigenère cipher is vulnerable: once the key length is found, a cryptanalyst can apply frequency analysis.
- **How to Find the Key Length?**
  - For Vigenère, 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:
    - Kasiski test
    - Index of coincidence (Friedman)

Kasiski Test

- Two identical segments of plaintext will be encrypted to the same ciphertext, if they occur in the text at the distance \( \Delta \), \( \Delta \equiv 0 \pmod{m} \), \( m \) is the key length.
- Algorithm:
  - Search for pairs of identical segments of length at least 3
  - Record distances between the two segments: \( \Delta_1, \Delta_2, \ldots \)
  - \( m \) divides \( \gcd(\Delta_1, \Delta_2, \ldots) \)

PT: THE SUN AND THE MAN IN THE MOON
Key: KI NG K I NG K I NG K I NG K I NG K I NG
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

Rotor Machines-I

- Basic idea: if the key in Vigenere cipher is very long, then the attacks won’t work
- Implementation idea: multiple rounds of substitution
- A machine consists of multiple cylinders
  - each cylinder has 26 states, at each state it is a substitution cipher: the wiring between the contacts implements a fixed substitution of letters
  - each cylinder rotates to change states according to different schedule changing the substitution
Rotor Machines-2

- A m-cylinder rotor machine has 26th different substitution ciphers
  - $26^1 = 17576$
  - $26^2 = 456,976$
  - $26^3 = 11,881,376$

Enigma Machine

- Patented by Scherius in 1918
  - Came on the market in 1923, weighted 50 kg (about 110 lbs), later cut down to 12 kg (about 26 lbs)
  - It cost about $30,000 in today’s prices
  - $34 \times 28 \times 15$ cm
- Widely used by the Germans from 1926 to the end of second world war
  - First successfully broken by Polish in the thirties by exploiting the repeating of the message key and knowledge of the machine design
  - During the WWII, Enigma was broken by Alan Turing (1912 - 1954) in the UK intelligence. He was an English mathematician, logician and cryptographer, father of modern computer science.

Enigma

- Use 3 scramblers (rotors): 17576 substitutions
- 3 scramblers can be used in any order: 6 combinations
- Plug board: allowed 6 pairs of letters to be swapped before the scramblers process started and after it ended.
- Total number of keys $\approx 10^{16}$
- Later versions use 5 rotors and 10 pairs of letters

Key Mapping

- A reflector enables to map a character twice with each rotor
- First rotor rotates after each key press
- Second rotor rotates after first had a complete revolution,
- and so on

Encrypting with Enigma

- Machine was designed under the assumption that the adversary may get access to the machine
- Daily keys: The settings for the rotors and plug boards changed daily according to a codebook received by all operators
  - A day key has the form
    - Scrambler arrangement: 2-3-1
    - Scrambler starting position: Q–C–W
- Message keys: Each message was encrypted with a unique key defined by the position of the 3 rotors

How to Break the Enigma Machine?

- Recover 3 secrets
  - Internal connections for the 3 rotors
  - Daily keys
  - Message keys
- With 2 months of day keys and Enigma usage instructions, the Polish mathematician Rejewski succeeded to reconstruct the internal wiring
Lessons Learned From Breaking Enigma

- Keeping a machine (i.e., a cipher algorithm) secret does not help
  - The Kerckhoff’s principle
  - Security through obscurity doesn’t work
- Large number of keys are not sufficient
- Known plaintext attack was easy to mount
- Key management was the weakest link
- People were also the weakest link
- Even a strong cipher, when used incorrectly, can be broken

Kerckhoffs’s Principle

- Auguste Kerckhoff (1835 – 1903) was a Dutch linguist and cryptographer who was professor of languages at the School of Higher Commercial Studies in Paris in the late 19th century.

- The security of a protocol should rely only on the secrecy of the keys, protocol designs should be made public. (1883)
  - Secrecy of a protocol does not work