Information Security
CS 526
Topic 2

Cryptography: Terminology & Classic Ciphers
Readings for This Lecture

Required readings:
  – Cryptography on Wikipedia

Interesting reading
  – The Code Book by Simon Singh
Goals of Cryptography

• The most fundamental problem cryptography addresses: ensure security of communication over insecure medium

• What does secure communication mean?
  – confidentiality (privacy, secrecy)
    • only the intended recipient can see the communication
  – integrity (authenticity)
    • the communication is generated by the alleged sender

• What does insecure medium mean?
  – Two possibilities:
    • Passive attacker: the adversary can eavesdrop
    • Active attacker: the adversary has full control over the communication channel
Approaches to Secure Communication

• Steganography
  – “covered writing”
  – hides the existence of a message
  – depends on secrecy of method

• Cryptography
  – “hidden writing”
  – hide the meaning of a message
  – depends on secrecy of a short key, not method
Basic Terminology

- Plaintext: original message
- Ciphertext: transformed message
- Key: secret used in transformation
- Encryption
- Decryption
- Cipher: algorithm for encryption/decryption
Shift Cipher

- The Key Space:
  - $[0 \ldots 25]$

- 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: Cryptanalysis

• Can an attacker find K?
  – YES: by a bruteforce attack through exhaustive key search,
  – key space is small (≤ 26 possible keys).

• Lessons:
  – Cipher key space needs to be large enough.
  – Exhaustive key search can be effective.
Mono-alphabetic Substitution Cipher

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

Example:

$\text{BECAUSE} \rightarrow \text{AZDBJSZ}$
Strength of the Mono-alphabetic Substitution Cipher

- Exhaustive search is difficult
  - key space size is $26! \approx 4 \times 10^{26} \approx 2^{88}$

- Dominates the art of secret writing throughout the first millennium A.D.

- Thought to be unbreakable by many back then

- How to break it?
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

![Frequency of Letters in English](image-url)
How to Defeat Frequency Analysis?

• Use larger blocks as the basis of substitution. Rather than substituting one letter at a time, substitute 64 bits at a time, or 128 bits.
  – Leads to block ciphers such as DES & AES.

• Use different substitutions to get rid of frequency features.
  – Leads to polyalphabetical substitution ciphers
  – Stream ciphers
Towards the Polyalphabetic Substitution Ciphers

• Main weaknesses of monoalphabetic substitution ciphers
  – In ciphertext, different letters have different frequency
    • 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
    • As result, frequencies of letters in ciphertext are similar

• Developed into a practical cipher by Vigenère (published in 1586)
The Vigenère Cipher

Treat letters as numbers: [A=0, B=1, C=2, ..., Z=25]

Number Theory Notation: \( Z_n = \{0, 1, \ldots, n-1\} \)

**Definition:**
Given \( m \), a positive integer, \( P = C = (Z_{26})^n \), and \( K = (k_1, k_2, \ldots, k_m) \) a key, we define:

**Encryption:**
\[
e_k(p_1, p_2 \ldots p_m) = (p_1+k_1, p_2+k_2 \ldots p_m+k_m) \pmod{26}
\]

**Decryption:**
\[
d_k(c_1, c_2 \ldots c_m) = (c_1-k_1, c_2-k_2 \ldots c_m-k_m) \pmod{26}
\]

**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 there are letters in the key.
Vigenere Cipher: Cryptanalysis

• Find the length of the key.
  – Kasisky test
  – Index of coincidence

• Divide the message into that many shift cipher encryptions.

• Use frequency analysis to solve the resulting shift ciphers.
  – How?
Kasisky Test for Finding Key Length

- Observation: two identical segments of plaintext, will be encrypted to the same ciphertext, if the 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: $\Delta_1$, $\Delta_2$, …
  - $m$ divides $\text{gcd}(\Delta_1, \Delta_2, \ldots)$
Example of the Kasisky Test


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

Repeating patterns (strings of length 3 or more) in ciphertext are likely due to repeating plaintext strings encrypted under repeating key strings; thus the location difference should be multiples of key lengths.
Adversarial Models for Ciphers

- The language of the plaintext and the nature of the cipher are assumed to be known to the adversary.
- **Ciphertext-only attack**: The adversary knows only a number of ciphertexts.
- **Known-plaintext attack**: The adversary knows some pairs of ciphertext and corresponding plaintext.
- **Chosen-plaintext attack**: The adversary can choose a number of messages and obtain the ciphertexts.
- **Chosen-ciphertext attack**: The adversary can choose a number of ciphertexts and obtain the plaintexts.

What kinds of attacks have we considered so far? When would these attacks be relevant in wireless communications?
Security Principles

• Kerckhoffs's Principle:
  – A cryptosystem should be secure even if everything about the system, except the key, is public knowledge.

• Shannon's maxim:
  – "The enemy knows the system."

• Security by obscurity doesn’t work

• Should assume that the adversary knows the algorithm; the only secret the adversary is assumed to not know is the key

• What is the difference between the algorithm and the key?
Coming Attractions …

- Cryptography: One-time Pad, Informational Theoretical Security, Stream Ciphers