COMP7120/8120 Cryptography and Data Security

Introduction
Outline

• High-level Concepts:
  - security objectives, security services, threat, vulnerability, ...

• Introduction to cryptography
• Some earlier cryptographic methods.
Security Objectives

Confidentiality

Integrity

Availability
Security Objectives (CIA)

- **Confidentiality** — Prevent/detect improper disclosure of information
- **Integrity** — Prevent/detect improper modification of information
- **Availability** — Prevent/detect improper denial of access to services provided by the system
Commercial Example

• Confidentiality — An employee should not know the salary of his manager
• Integrity — An employee should not be able to modify the employee's own salary
• Availability — Paychecks should be printed on time as stipulated by law
Military Example

• Confidentiality — The target coordinates of a missile should not be improperly disclosed.

• Integrity — The target coordinates of a missile should not be improperly modified.

• Availability — When the proper command is issued the missile should fire.
Question:

- C, I, A
  - Which one is important than the other?
Security Services

- Security functions are typically made available to users as a set of **security services** through APIs or integrated interfaces.

- **Confidentiality**: protection of any information from being exposed to unintended entities.

- **Authentication**: assurance that an entity of concern or the origin of a communication is authentic - it’s what it claims to be or from.

- **Integrity**: assurance that the information has not been tampered with.
Security Services (Cont’d)

• **Non-repudiation**: offer of evidence that a party is indeed the sender or a receiver of certain information

• **Access control**: facilities to determine and enforce who is allowed access to what resources, hosts, software, network connections

• **Monitor & response**: facilities for monitoring security attacks, generating indications, surviving (tolerating) and recovering from attacks
Security Assurance

- **How well** your security mechanisms guarantee your security policy
  - Metrics to measure the level of security.

- Everyone wants high assurance
- High assurance implies high cost
  - May not be possible
- Trade-off is needed
Security by Cryptography

• Essential way to ensure the goals of integrity and confidentiality.

• Question: Can cryptography achieve the goal of availability?
Security by Obscurity

• Security by obscurity
  - If we hide the inner workings of a system it will be secure

• More and more applications open their standards (e.g., TCP/IP, 802.11)

• Widespread computer knowledge and expertise
Security by Legislation

• Security by legislation says that if we instruct our users on how to behave we can secure our systems

• For example
  - Users should not share passwords
  - Users should not write down passwords
  - Users should not type in their password when someone is looking over their shoulder

• User awareness and cooperation is important, but cannot be the principal focus for achieving security
Threat-Vulnerability

• Threats — Possible attacks on the system
  - The attacks targeting C, I, or A.

• Vulnerabilities — Weaknesses that may be exploited to cause loss or harm

• Requires assessment of threats and vulnerabilities
Threat Model and Attack Model

- Threat model and attack model need to be clarified before any security mechanism is developed

- Threat model
  - Assumptions about potential attackers
  - Describes the attacker’s capabilities

- Attack model
  - Assumptions about the attacks
  - Describe how attacks are launched
Introduction to Cryptography
Cryptography

- **Cryptography**: the art of secret writing

- Converts data into unintelligible (random-looking) form
  - Must be *reversible* (can recover original data without loss or modification)
Cryptography vs. Steganography

- **Steganography** concerns *existence*
  - Conceals the very existence of communication

Apparently neutral’s protest is thoroughly discounted and ignored. Isman hard hit. Blockade issue affects pretext for embargo on by-products, ejecting suets and vegetable oils.

Pershing sails from NY June I

- **Cryptography** conceals the contents of communication between two parties
Encryption/Decryption

- Plaintext: a message in its original form
- Ciphertext: a message in the transformed, unrecognized form
- Encryption: the process that transforms a plaintext into a ciphertext
- Decryption: the process that transforms a ciphertext to the corresponding plaintext
- Key: the value used to control encryption/decryption.
- Cipher: algorithm that performs encryption or decryption.
Cryptanalysis

- Cryptanalysis: the art of revealing the secret
  - Defeat cryptographic security systems
  - Gain access to the real contents of encrypted messages
  - Cryptographic keys can be unknown

- Difficulty depends on
  - Sophistication of the encryption/decryption
  - Amount of information available to the code breaker

- We call the party that performs cryptanalysis the attacker.
Ciphertext Only Attacks

• An attacker intercepts a set of ciphertexts

• Breaking the cipher: analyze patterns in the ciphertext
  - provides clues about the plaintext and key
Known Plaintext Attacks

• An attacker has samples of both the plaintext and its encrypted version, the ciphertext

• Makes some ciphers (e.g., mono-alphabetic ciphers) very easy to break
Chosen Plaintext Attacks

• An attacker has the capability to choose arbitrary plaintexts to be encrypted and obtain the corresponding ciphertexts
  – How could such attacks be possible?
  – Difference between known plaintext and chosen plaintext attacks
Perfectly Secure Ciphers

1. Ciphertext does not reveal any information about which plaintexts are more likely to have produced it
   - e.g., the cipher is robust against ciphertext only attacks

2. Plaintext does not reveal any information about which ciphertexts are more likely to be produced
   - e.g., the cipher is robust against known/chosen plaintext attacks
Computationally Secure Ciphers

1. The **cost** of breaking the cipher quickly exceeds the value of the encrypted information and/or

2. The **time** required to break the cipher exceeds the useful lifetime of the information
   - Under the **assumption** there is not a faster / cheaper way to break the cipher, waiting to be discovered

• Most ciphers today are computationally secure. Sometimes we also say:
  - computationally infeasible or computationally difficult to break a cipher.
Example

- If you are fast (2 tries/s), you need $1000/2=500s$ to try all combinations.

What is the lock has 10 digits?

$10^{10}/2=158.5$ years to try all combinations.
Computationally secure cipher

• Design goal:
  - Make ciphertext look completely random regardless of the content of plaintext.
  - There is no fast way to crack it, the only way is to try all combinations
  - Make sure there are a great number of possible combinations.
  - Ensure computational difficulty without key.
  - Computational efficiency with key to do encryption and decryption.
Keep what secret?

- We have
  - plaintext, key, cipher, and ciphertext

Definitely keep secret!
Question

• If cryptography is combined with compression
  - What is the right order?
Hide or Reveal Algorithms

- Keep algorithms secret
  - We can achieve better security if we keep the algorithms secret
  - Hard to keep secret if used widely

- Publish the algorithms
  - Security depends on the secrecy of the keys
  - Less unknown vulnerability if all the smart (good) people in the world are examine the algorithms

- Military
  - Both secret key and secret algorithm
Some Early Ciphers
Caesar Cipher

• Replace each letter with the one 3 letters later in the alphabet

plaintext alphabet

ciphertext alphabet

Trivial to break
A variant of Caesar Cipher

• Replace each letter by one that is \( \delta \) positions later, where \( \delta \) is selectable (i.e., \( \delta \) is the key)
  - example: IBM \( \rightarrow \) HAL (for \( \delta =25 \))

• Also trivial to break with modern computers (how many possibilities?)

plaintext alphabet

| A | B | C | D | E | F | G | H | I | J | K | ...
|---|---|---|---|---|---|---|---|---|---|---|---

ciphertext alphabet

| A | B | C | D | E | F | G | H | I | J | K | ...
|---|---|---|---|---|---|---|---|---|---|---|---|
Mono-Alphabetic Ciphers

- Generalized substitution cipher: randomly map one letter to another (How many possibilities?)
  - $26! \approx 2^{88}$
- The key must specify which permutation; how many bits does that take?
  - $\log_2(26!) \approx 88$ bits

<table>
<thead>
<tr>
<th>plaintext alphabet</th>
<th>ciphertext alphabet</th>
</tr>
</thead>
<tbody>
<tr>
<td>A B C D E F G H I J K ...</td>
<td>A B C D E F G H I J K ...</td>
</tr>
</tbody>
</table>
Attacking Mono-Alphabetic Ciphers

- Known plaintext attacks
- Frequency of single letters in English language, taken from a large corpus of text:

<table>
<thead>
<tr>
<th>Letter</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>8.2%</td>
</tr>
<tr>
<td>B</td>
<td>1.5%</td>
</tr>
<tr>
<td>C</td>
<td>2.8%</td>
</tr>
<tr>
<td>D</td>
<td>4.3%</td>
</tr>
<tr>
<td>E</td>
<td>12.7%</td>
</tr>
<tr>
<td>F</td>
<td>2.2%</td>
</tr>
<tr>
<td>G</td>
<td>2.0%</td>
</tr>
<tr>
<td>H</td>
<td>6.1%</td>
</tr>
<tr>
<td>I</td>
<td>7.0%</td>
</tr>
<tr>
<td>J</td>
<td>0.2%</td>
</tr>
<tr>
<td>K</td>
<td>0.8%</td>
</tr>
<tr>
<td>L</td>
<td>4.0%</td>
</tr>
<tr>
<td>M</td>
<td>2.4%</td>
</tr>
<tr>
<td>N</td>
<td>6.7%</td>
</tr>
<tr>
<td>O</td>
<td>7.5%</td>
</tr>
<tr>
<td>P</td>
<td>1.9%</td>
</tr>
<tr>
<td>Q</td>
<td>0.1%</td>
</tr>
<tr>
<td>R</td>
<td>6.0%</td>
</tr>
<tr>
<td>S</td>
<td>6.3%</td>
</tr>
<tr>
<td>T</td>
<td>9.1%</td>
</tr>
<tr>
<td>U</td>
<td>2.8%</td>
</tr>
<tr>
<td>V</td>
<td>1.0%</td>
</tr>
<tr>
<td>W</td>
<td>2.4%</td>
</tr>
<tr>
<td>X</td>
<td>0.2%</td>
</tr>
<tr>
<td>Y</td>
<td>2.0%</td>
</tr>
<tr>
<td>Z</td>
<td>0.1%</td>
</tr>
</tbody>
</table>
Attacking... (Cont’d)

• Suppose the attacker intercepts the following message

UXGPOGZCFJZJTFADADAJEJNJDZMZHBBGZGGKQGVVGVXCDIWGX

| 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 |
| 3 | 2 | 2 | 4 | 1 | 2 | 8 | 1 | 1 | 4 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 2 | 1 | 3 | 0 | 5 |

| A ≈ 8.2% | H ≈ 6.1% | O ≈ 7.5% | V ≈ 1.0% |
| B ≈ 1.5% | I ≈ 7.0% | P ≈ 1.9% | W ≈ 2.4% |
| C ≈ 2.8% | J ≈ 0.2% | Q ≈ 0.1% | X ≈ 0.2% |
| D ≈ 4.3% | K ≈ 0.8% | R ≈ 6.0% | Y ≈ 2.0% |
| E ≈ 12.7% | L ≈ 4.0% | S ≈ 6.3% | Z ≈ 0.1% |
| F ≈ 2.2% | M ≈ 2.4% | T ≈ 9.1% |
| G ≈ 2.0% | N ≈ 6.7% | U ≈ 2.8% |

FREQUENCY ANALYSIS IS AMAZING NOW WE NEED BETTER CIPHER
Letter Frequencies

![Letter Frequency Chart]

- Frequency (%)
- Letter
- Test Data
Vigenere Cipher

- A set of mono-alphabetic substitution rules (shift amounts) is used
  - the key determines what the sequence of rules is
  - also called a poly-alphabetic cipher

- Ex.: key = (3 1 5)
  - i.e., substitute first letter in plaintext by letter+3, second letter by letter+1, third letter by letter+5
  - then repeat this cycle for each 3 letters
Vigenere... (Cont’d)

- Ex.: plaintext = “BANDBAD”

plaintext message

B A N D B A D

shift amount

3 1 5 3 1 5 3

ciphertext message

E B S G C F G

What are the possible attacks?

- Known plaintext? Frequency analysis?
Hill Ciphers

• Encrypts $m$ letters of plaintext at each step
  - i.e., plaintext is processed in blocks of size $m$

• Encryption of plaintext $p$ to produce ciphertext $c$ is accomplished by: $c = Kp$
  - the $m \times m$ matrix $K$ is the key
  - decryption is multiplication by inverse: $p = K^{-1}c$
  - remember: all arithmetic mod 26
Hill Cipher Example

• For $m = 2$, let $K = \begin{pmatrix} 1 & 2 \\ 3 & 5 \end{pmatrix}$, $K^{-1} = \begin{pmatrix} 21 & 2 \\ 3 & 25 \end{pmatrix}$

Plaintext $p =$

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>X</td>
<td>Y</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>23</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>6</td>
</tr>
</tbody>
</table>

Ciphertext $c =$

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>F</td>
<td>T</td>
<td>H</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>N</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$(1 \times 0 + 2 \times 1) \mod 26 = 2 
$(3 \times 23 + 5 \times 24) \mod 26 = 19 
$(21 \times 15 + 2 \times 13) \mod 26 = 21$
Permutation Ciphers

• The previous codes are all based on substituting one symbol in the alphabet for another symbol in the alphabet

• Permutation cipher: permute (rearrange, transpose) the letters in the message
  - the permutation can be fixed, or can change over the length of the message
Permutation... (Cont’d)

• Permutation cipher ex. #1:
  - Permute each successive block of 5 letters in the message according to position offset <+1,+3,-2,0,-2>

plaintext message

W H Y O W | H Y C A N | T I F L Y

Y W W O H | C H N A Y | F T Y L I
ciphertext message
Permutation... (Cont’d)

• Permutation cipher ex. #2:
  • arrange plaintext in blocks of $n$ columns and $m$ rows
  • then permute columns in a block according to a key $K$

$$n = 4$$

Key (perm. of columns) $\rightarrow$ 4 3 1 2

Plaintext symbol positions

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
</tr>
</tbody>
</table>

$ciphertext$ sequence (by plaintext position) for one block

3 7 11 4 8 12 2 6 10 1 5 9
Permutation... (Cont’d)

- A longer example: plaintext = “ATTACK POSTPONED UNTIL TWO AM”

<table>
<thead>
<tr>
<th>Key:</th>
<th>4</th>
<th>3</th>
<th>1</th>
<th>2</th>
<th>5</th>
<th>7</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>plaintext</td>
<td>ATTACK</td>
<td>POSTPONE</td>
<td>DUNTIL</td>
<td>WOAMXYZ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ciphertext</td>
<td>TTNA</td>
<td>APTM</td>
<td>TSUO</td>
<td>AODW</td>
<td>COIX</td>
<td>PETZ</td>
<td>KNLY</td>
</tr>
</tbody>
</table>
A Perfectly Secure Cipher: One-Time Pads

- According to a theorem by Shannon, a perfectly secure cipher requires:
  - a key length at least as long as the message to be encrypted
  - the key can only be used once (i.e., for each message we need a new key)

- Very limited use due to need to negotiate and distribute long, random keys for every message
OTP... (Cont’d)

• Idea
  - generate a random bit string (the key) as long as the plaintext, and share with the other communicating party
  - encryption: XOR this key with plaintext to get ciphertext
  - decrypt: XOR same key with ciphertext to get plaintext
OTP... (Cont’d)

- Why can’t the key be reused?
Some “Key” Issues

Typical Ciphers Today
Types of Cryptography

• Number of keys
  - Secret key cryptography: one key
  - Public key cryptography: two keys
  - Hash functions: no key

• The way in which the plaintext is processed
  - Stream cipher: encrypt input message one symbol at a time
  - Block cipher: divide input message into blocks of symbols, and processes the blocks in sequence
    • May require padding
Secret Key Cryptography

- Same key is used for encryption and decryption
- Also known as
  - Symmetric cryptography
  - Conventional cryptography
Public Key Cryptography

- Invented/published in 1975
- A public/private key pair is used
  - Public key can be publicly known
  - Private key is kept secret by the owner of the key
- Much slower than secret key cryptography
- Also known as
  - Asymmetric cryptography
Hash Algorithms

- Also known as
  - Message digests
  - One-way transformations
  - One-way functions
  - Hash functions

- Length of $H(m)$ much shorter than length of $m$
- Usually fixed lengths: 128 or 160 bits
Summary

• Cryptography is a fundamental, and most carefully studied, component of security
  - not usually the “weak link”

• “Perfectly secure” ciphers are too expensive in practice

• Modern ciphers are based on computational difficulty.

• Early ciphers aren’t nearly strong enough