Basic Terminology

- **plaintext** - the original message
- **ciphertext** - the coded message
- **key** - information used in encryption/decryption, and known only to sender/receiver
- **encipher (encrypt)** - converting plaintext to ciphertext using key
- **decipher (decrypt)** - recovering ciphertext from plaintext using key
- **cryptography** - study of encryption principles/methods/designs
- **cryptanalysis (code breaking)** - the study of principles/methods of deciphering ciphertext
Cryptographic Systems

Cryptographic Systems are categorized according to:

1. The operation used in transferring plaintext to ciphertext:
   - **Substitution**: each element in the plaintext is mapped into another element
   - **Transposition**: the elements in the plaintext are re-arranged.

2. The number of keys used:
   - **Symmetric (private-key)**: both the sender and receiver use the same key
   - **Asymmetric (public-key)**: sender and receiver use different key

3. The way the plaintext is processed:
   - **Block cipher**: inputs are processed one block at a time, producing a corresponding output block.
   - **Stream cipher**: inputs are processed continuously, producing one element at a time (bit,

---

**Symmetric Encryption Model**

- **Secret key shared by sender and recipient**
- **Plaintext input**
- **Encryption algorithm (e.g., DES)**
- **Transmitted ciphertext**
- **Decryption algorithm (reverse of encryption algorithm)**
- **Plaintext output**

---
Cryptographic Systems

Requirements

- two requirements for secure use of symmetric encryption:
  1. a strong encryption algorithm
  2. a secret key known only to sender / receiver

- $Y = E_k(X)$, where $X$: the plaintext, $Y$: the ciphertext
- $X = D_k(Y)$

- assume encryption algorithm is known
  - implies a secure channel to distribute key

Cryptographic Systems

Attacks:

1. Cryptanalytic Attacks: depends on the nature of the encryption algorithm used.
   - Uses information such as plaintext/ciphertext pairs to deduce the key

2. Brute-force Attack: try all the possible keys – depends on the key length.

<table>
<thead>
<tr>
<th>Key Size (bits)</th>
<th>Number of Alternative Keys</th>
<th>Time required at 1 encryption/ps</th>
<th>Time required at 10^6 encryptions/ps</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>$2^{32}$</td>
<td>35.6 minutes</td>
<td>2.15 milliseconds</td>
</tr>
<tr>
<td>56</td>
<td>$2^{56}$</td>
<td>1142 years</td>
<td>10.01 hours</td>
</tr>
<tr>
<td>128</td>
<td>$2^{128}$</td>
<td>5.4 x 10^4 years</td>
<td>5.4 x 10^8 years</td>
</tr>
<tr>
<td>168</td>
<td>$2^{168}$</td>
<td>5.9 x 10^6 years</td>
<td>5.9 x 10^9 years</td>
</tr>
<tr>
<td>26 characters</td>
<td>$26!$</td>
<td>$2 \times 10^5$ years</td>
<td>$6.4 \times 10^6$ years</td>
</tr>
<tr>
<td>(permutation)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Security Definitions

- unconditional security
  - no matter how much computer power is available, the cipher cannot be broken since the ciphertext provides insufficient information to uniquely determine the corresponding plaintext

- computational security
  - given limited computing resources (e.g., time needed for calculations is greater than age of universe), the cipher cannot be broken

Cryptographic Systems

Classical Ciphers-Substitution Ciphers

Shift Cipher

Letters of the alphabet are assigned a number as below

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th>O</th>
<th>P</th>
<th>Q</th>
<th>R</th>
<th>S</th>
<th>T</th>
<th>U</th>
<th>V</th>
<th>W</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
</tr>
</tbody>
</table>

Algorithm:
Let $P = C = K = \mathbb{Z}_{26}$ and $x \in P$, $y \in C$, $k \in K$

Encryption: $E_k(x) = x + k \mod 26$.
Decryption: $D_k(x) = x - k \mod 26$. 
Remark: When \( k = 3 \) the shift cipher is given a special name - *Caesar Cipher*.

**Example:** Let the key \( k = 17 \)

Plaintext: \( X = A \ T \ T \ A \ C \ K = (0, 19, 19, 0, 2, 10) \).

Ciphertext: \( Y = (0+17 \mod 26, 19+17 \mod 26, \ldots) \)

\( Y = (17, 10, 10, 17, 19, 1) = R \ K \ K \ R \ T \ B \)

**Attacks on Shift Cipher**

1. Exhaustive Search: Try all possible keys. \(|K|=26\).

2. Letter frequency analysis (Same plaintext maps to same ciphertext)

---

**Substitution Ciphers**

**Monoalphabetic Cipher**

- Jumble the letters arbitrarily
- each plaintext letter maps to a different random ciphertext letter.
- key is 26 letters long
- There are \( 26! = 4 \times 10^{26} \) Possible keys.
  - Plain: abcdefghijklmnopqrstuvwxyz
  - Cipher: DKVQFIBJWPESCXHTMYAUOLRGZN
  - Plaintext: ifwewishtoreplaceletters
  - Ciphertext: WIRFRWAJUHYFTSDVFSFUUFYA
Dr. Lo’ai Tawalbeh       Fall 2005

**Substitution Ciphers**

**Attacks: Frequency Analysis**

- Letter frequency analysis (Same plaintext maps to same ciphertext): language redundancy:
  - letters are not equally commonly used
  - in English, *e* is by far the most common letter, then *T,R,N,I,O,A,S*
  - other letters are fairly rare: cf. *Z,J,K,Q,X*

---

**English Letter Frequencies**

![Graph showing English letter frequencies](image-url)
Use in Cryptanalysis

- key concept - monoalphabetic substitution ciphers do not change relative letter frequencies
- calculate letter frequencies for ciphertext
- compare counts/plots against known values
- for monoalphabetic must identify each letter
  - tables of common double/triple letters help

- See the example in page 33

Example Cryptanalysis

- given ciphertext:
  UZQSOVU0KXMOPVSOSZ0PFPFESXUDMETSXMAI2
  U0FPHZHMDZSHZOWSFPAPPDTSVPQZWYMXUZUHSX
  EPYEPOPSZUFPQMBZWFUFSHDJUDM0HSQ

- count relative letter frequencies (see text)
- guess P & Z are e and t
- guess ZW is th and hence ZWP is the
- proceeding with trial and error finally get:
  it was disclosed yesterday that several informal but
direct contacts have been made with political
representatives of the vietcong in moscow
Playfair Key

- not even the large number of keys in a monoalphabetic cipher provides security
- one approach to improve security was to encrypt multiple letters
- a 5X5 matrix of letters based on a keyword
- fill in letters of keyword (sans duplicates)
- fill rest of matrix with other letters
- eg. using the keyword MONARCHY

<table>
<thead>
<tr>
<th>MONAR</th>
<th>CHYBD</th>
</tr>
</thead>
<tbody>
<tr>
<td>EFGIK</td>
<td>LPQST</td>
</tr>
<tr>
<td>UVWXZ</td>
<td></td>
</tr>
</tbody>
</table>

Encrypting and Decrypting

- plaintext encrypted two letters at a time:
  1. if a pair is a repeated letter, insert a filler like 'X', eg. "balloon" encrypts as "ba lx lo on"
  2. if both letters fall in the same row, replace each with letter to right (wrapping back to start from end), eg. "ar" encrypts as "RM"
  3. if both letters fall in the same column, replace each with the letter below it (again wrapping to top from bottom), eg. "mu" encrypts to "CM"
  4. otherwise each letter is replaced by the one in its row in the column of the other letter of the pair, eg. "hs" encrypts to "BP", and "ea" to "IM" or "JM" (as desired)
Security of the Playfair Cipher

- security much improved over monoalphabetic
- since have \(26 \times 26 = 676\) digrams
- would need a 676 entry frequency table to analyse (verses 26 for a monoalphabetic)
- and correspondingly more ciphertext
- was widely used for many years (eg. US & British military in WW1)
- it can be broken, given a few hundred letters
- since still has much of plaintext structure

Hill Cipher

**Example:** Let \(n=3\) and the key matrix be \(M\), \(C = PM\)

The key is an \(n \times n\) matrix whose entries are integers in \(Z_{26}\).

\[
M = \begin{pmatrix}
1 & 2 & 3 \\
4 & 5 & 6 \\
11 & 9 & 8 \\
\end{pmatrix}
\]

and the plaintext be \(ABC = (0, 1, 2)\) then the encryption operation is a vector-matrix multiplication

\[
(0,1,2) \times \begin{pmatrix}
1 & 2 & 3 \\
4 & 5 & 6 \\
11 & 9 & 8 \\
\end{pmatrix} = (0,23,22) \text{mod } 26 \Rightarrow AXW \text{ (ciphertext)}
\]

In order to decrypt we need the inverse of key matrix \(M\), which is

\[
N = \begin{pmatrix}
22 & 5 & 1 \\
6 & 17 & 24 \\
15 & 13 & 1 \\
\end{pmatrix}
\]
Polyalphabetic Ciphers

• another approach to improving security is to use multiple cipher alphabets
• called polyalphabetic substitution ciphers
• makes cryptanalysis harder with more alphabets to guess and flatter frequency distribution
• use a key to select which alphabet is used for each letter of the message
• use each alphabet in turn
• repeat from start after end of key is reached

Vigenère Cipher

• simplest polyalphabetic substitution cipher is the Vigenère Cipher
• effectively multiple caesar ciphers
• key is multiple letters long $K = k_1 \ k_2 \ ... \ k_d$
• $i^{th}$ letter specifies $i^{th}$ alphabet to use
• use each alphabet in turn
• repeat from start after $d$ letters in message
• decryption simply works in reverse
Example

- write the plaintext out
- write the keyword repeated above it
- use each key letter as a caesar cipher key
- encrypt the corresponding plaintext letter
- eg using keyword deceptive
  
  key: deceptive deceptivedeceptivedeceptive -- row
  plaintext: wearediscoveredsaveyourself -- column
  ciphertext: ZICVTWQNGRZGVTWAVZHCQYGLMJ

Security of Vigenère Ciphers

- have multiple ciphertext letters for each plaintext letter
- hence letter frequencies are obscured
- but not totally lost
- start with letter frequencies
  - see if look monoalphabetic or not
- if not, then need to determine number of alphabets, since then can attach each
Kasiski Method

- method developed by Babbage / Kasiski
- repetitions in ciphertext give clues to period
- so find same plaintext an exact period apart
- which results in the same ciphertext
- of course, could also be random
- eg repeated "VTW" in previous example
- suggests size of 3 or 9
- then attack each monoalphabetic cipher individually using same techniques as before

Autokey Cipher

- ideally want a key as long as the message
- Vigenère proposed the autokey cipher
- with keyword is prefixed to message as key
- knowing keyword can recover the first few letters
- use these in turn on the rest of the message
- but still have frequency characteristics to attack
- eg. given key deceptive
  
  key: decepтивeweadiscoveredsaveyourself
  plaintext: wearediscoveredsaveyourself
ciphertext: ZICWTVQNGKZEIIGASXSTSLEVWLA
One-Time Pad

- if a truly random key as long as the message is used, the cipher will be secure
- called a One-Time pad
- is unbreakable since ciphertext bears no statistical relationship to the plaintext
- since for any plaintext & any ciphertext there exists a key mapping one to other
- can only use the key once though
- have problem of safe distribution of key

Transposition Ciphers

- now consider classical transposition or permutation ciphers
- these hide the message by rearranging the letter order
- without altering the actual letters used
- can recognise these since have the same frequency distribution as the original text
Rail Fence cipher

- write message letters out diagonally over a number of rows
- then read off cipher row by row
- eg. write message out as:
  
  mematrhgpry
  etefeteoaat

- giving ciphertext
  
  MEMATRHTGPRYTEFETEAOAT

Row Transposition Ciphers

- a more complex scheme
- write letters of message out in rows over a specified number of columns
- then reorder the columns according to some key before reading off the rows

  Key: 4 3 1 2 5 6 7
  Plaintext: attack postponed untill womxyz
  Ciphertext: TTNAAPTMTSUOADW-OIXKNLYPETZ
Product Ciphers

- ciphers using substitutions or transpositions are not secure because of language characteristics
- hence consider using several ciphers in succession to make harder, but:
  - two substitutions make a more complex substitution
  - two transpositions make more complex transposition
  - but a substitution followed by a transposition makes a new much harder cipher
- this is bridge from classical to modern ciphers

Rotor Machines

- before modern ciphers, rotor machines were most common product cipher
- were widely used in WW2
  - German Enigma, Allied Hagelin, Japanese Purple
- implemented a very complex, varying substitution cipher
- used a series of cylinders, each giving one substitution, which rotated and changed after each letter was encrypted
- with 3 cylinders have $26^3 = 17576$ alphabets
Steganography

- an alternative to encryption
- hides existence of message
  - using only a subset of letters/words in a longer message marked in some way
  - using invisible ink
  - hiding in LSB in graphic image or sound file
- has drawbacks
  - high overhead to hide relatively few info bits

Summary

- have considered:
  - classical cipher techniques and terminology
  - monoalphabetic substitution ciphers
  - cryptanalysis using letter frequencies
  - Playfair ciphers
  - polyalphabetic ciphers
  - transposition ciphers
  - product ciphers and rotor machines
  - stenography