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
**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:** DKVQFIBJWPESCMYAUOLRGZN
- **Plaintext:** ifwewishtoreplaceletters
- **Ciphertext:** WIRFRWAUHYFTSDVFSFUUFYA

**Attacks:**

- Exhaustive Search: Try all possible keys. For Caesar \(|K|=26\).
- 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
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
Example Cryptanalysis

- given ciphertext:
  UZQSOVDHMMOPVGOZPEVSUGWZOPFPESXUSDHBTSXAIIZ
  VUERPHZHBMBZSHOZSFAPPDTSVPQZWYMUXZUHSX
  EPYEPPOPOZSUFPQMBZWPFFUPZHMDJUDMTQMNQ

- 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 Cipher

- not even the large number of keys in a monoalphabetic cipher provides security
- one approach to improving security was to encrypt multiple letters
- the Playfair Cipher is an example
Playfair Key Matrix

- 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

```
MONAR
CHYBD
EFGIK
LPQST
UVWXZ
```

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 "bal x 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 x 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
The key is an $n \times n$ matrix whose entries are integers in $\mathbb{Z}_{26}$.

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

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

$\begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} \times \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 11 & 9 & 8 \end{pmatrix} = \begin{pmatrix} 22 \\ 5 \\ 1 \end{pmatrix} \equiv (0, 23, 22) \mod 26 \Rightarrow AXW$ (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 \( K = k_1 \ k_2 \ldots \ k_d \)
- \( i^{th} \) 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 deceptive deceptive
  plaintext: wearediscovered saveyourself
  ciphertext: ZICVTW QNGRZGVTWAVZHCQYGLMJ

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 fluke
- 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: deceptivewearediscoveredsav
plaintext: wearediscoveredsaveyourself
ciphertext:ZICVTWQNGKZEIIGASXSTSVVWLA
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

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

<table>
<thead>
<tr>
<th>Key:</th>
<th>4 3 1 2 5 6 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plaintext:</td>
<td>a t t a c k p o s t p o n e d u n t i l t w o a m x y z</td>
</tr>
<tr>
<td>Ciphertext:</td>
<td>TTNAAPTMSONADWCOIXNLYPETZ</td>
</tr>
</tbody>
</table>
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