Many savages at the present day regard their names as vital parts of themselves, and therefore take great pains to conceal their real names, lest these should give to evil-disposed persons a handle by which to injure their owners.
—The Golden Bough, Sir James George Frazer
Main Problems

• The main problems for cryptography today
  ○ Privacy
  ○ Integrity

• Privacy
  ○ The right of an individual to be secure from unauthorized disclosure of information about oneself that is contained in documents

• Integrity
  ○ The ability to ensure that information is not modified except by people who are explicitly intended to modify it
**Scenario**

**Purpose**
Electronic deposit to employees’ accounts

- Make sure that
  - The message is in unreadable form except for the Bank

**Message**

- Make sure that
  - The message comes from the Organization
  - The message is original

**Organization** → **Message** → **Bank**
Solutions by Cryptography

- Cryptography is the way to protect aspect of message
- Solutions
  - Privacy: Encryption
  - Integrity: Authentication
- Encryption
  - The process of disguising a message or data in such a way as to hide its substance.
Solutions by Cryptography

• **Authentication**
  - Verification of the identity of the entities
    - Impersonate
  - Verification that the original contents of information have not been altered or corrupted
    - Substitute
Encryption

Signifies
Space: Communication line
Time: Storage

Alice  Message  Bob

Network Sniffer
Trapping the line

Eve
Encryption

Condition:

\[ C = E(m_1, K') \]
\[ m_2 = D(C, K) = m_1 \]

Must be kept secret

Easy

K' must be kept secret
“Symmetric Cryptography”

Hard

K' can be made public
“Asymmetric Cryptography”
Symmetric VS Asymmetric

<table>
<thead>
<tr>
<th></th>
<th>Symmetric</th>
<th>Asymmetric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of secret key per person</td>
<td>$n-1$</td>
<td>1</td>
</tr>
<tr>
<td>Total number of secret key</td>
<td>$n(n-1)/2$</td>
<td>$n$</td>
</tr>
<tr>
<td>Speed</td>
<td>Fast</td>
<td>Slow</td>
</tr>
<tr>
<td>Capability</td>
<td>Less powerful</td>
<td>More powerful</td>
</tr>
</tbody>
</table>
Symmetric Encryption

- or conventional / private-key / single-key
- sender and recipient share a common key
- all classical encryption algorithms are private-key
- was only type prior to invention of public-key in 1970’s
- and by far most widely used
Some Basic Terminology

- **plaintext** - original message
- **ciphertext** - coded message
- **cipher** - algorithm for transforming plaintext to ciphertext
- **key** - info used in cipher known only to sender/receiver
- **encipher (encrypt)** - converting plaintext to ciphertext
- **decipher (decrypt)** - recovering ciphertext from plaintext
- **cryptography** - study of encryption principles/methods
- **cryptanalysis (codebreaking)** - study of principles/methods of deciphering ciphertext *without* knowing key
- **cryptology** - field of both cryptography and cryptanalysis
Symmetric Cipher Model

Plaintext input

Encryption algorithm (e.g., DES)

Transmitted ciphertext

Decryption algorithm (reverse of encryption algorithm)

Plaintext output

Secret key shared by sender and recipient

Secret key shared by sender and recipient
Requirements

- two requirements for secure use of symmetric encryption:
  - a strong encryption algorithm
  - a secret key known only to sender / receiver
- mathematically have:
  \[ Y = E_K(X) \]
  \[ X = D_K(Y) \]
- assume encryption algorithm is known
- implies a secure channel to distribute key
Cryptography

- characterize cryptographic system by:
  - type of encryption operations used
    - substitution / transposition / product
  - number of keys used
    - single-key or private / two-key or public
  - way in which plaintext is processed
    - block / stream
Cryptanalysis

- objective to recover key not just message
- general approaches:
  - cryptanalytic attack
  - brute-force attack
Cryptanalytic Attacks

- **ciphertext only**
  - only know algorithm & ciphertext, is statistical, know or can identify plaintext

- **known plaintext**
  - know/suspect plaintext & ciphertext

- **chosen plaintext**
  - select plaintext and obtain ciphertext

- **chosen ciphertext**
  - select ciphertext and obtain plaintext

- **chosen text**
  - select plaintext or ciphertext to en/decrypt
Secure Algorithm

- Only relatively weak algorithms fail to withstand a ciphertext-only attack.
- Generally, an encryption algorithm is designed to withstand a known-plaintext attack.
More Definitions

- **unconditional security**
  - no matter how much computer power or time is available, the cipher cannot be broken since the ciphertext provides insufficient information to uniquely determine the corresponding plaintext
  - there is no encryption algorithm that is unconditionally secure, except “One-time Pad”

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

- always possible to simply try every key
- most basic attack, proportional to key size
- assume either know / recognize plaintext

<table>
<thead>
<tr>
<th>Key Size (bits)</th>
<th>Number of Alternative Keys</th>
<th>Time required at 1 decryption/µs</th>
<th>Time required at 10^6 decryptions/µs</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>(2^{32} = 4.3 \times 10^9)</td>
<td>(2^{31} \mu s) = 35.8 minutes</td>
<td>2.15 milliseconds</td>
</tr>
<tr>
<td>56</td>
<td>(2^{56} = 7.2 \times 10^{16})</td>
<td>(2^{55} \mu s) = 1142 years</td>
<td>10.01 hours</td>
</tr>
<tr>
<td>128</td>
<td>(2^{128} = 3.4 \times 10^{38})</td>
<td>(2^{127} \mu s) years</td>
<td>(5.4 \times 10^{24}) years</td>
</tr>
<tr>
<td>168</td>
<td>(2^{168} = 3.7 \times 10^{50})</td>
<td>(2^{167} \mu s) years</td>
<td>(5.9 \times 10^{36}) years</td>
</tr>
<tr>
<td>26 characters (permutation)</td>
<td>(26! = 4 \times 10^{26})</td>
<td>(2 \times 10^{26} \mu s) = (6.4 \times 10^{12}) years</td>
<td>(6.4 \times 10^6) years</td>
</tr>
</tbody>
</table>
Classical Substitution Ciphers

- where letters of plaintext are replaced by other letters or by numbers or symbols
- or if plaintext is viewed as a sequence of bits, then substitution involves replacing plaintext bit patterns with ciphertext bit patterns
Caesar Cipher

- earliest known substitution cipher
- by Julius Caesar
- first attested use in military affairs
- replaces each letter by 3rd letter on
- example:
  meet me after the toga party
  PHHW PH DIWHU WKH WRJD SDUWB
Caesar Cipher

- **can define transformation as:**
  
  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
  
  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

- **mathematically give each letter a number**
  
  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
  
  0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

- **then have Caesar cipher as:**
  
  \[ c = E(p) = (p + k) \mod (26) \]
  
  \[ p = D(c) = (c - k) \mod (26) \]
Cryptanalysis of Caesar Cipher

- only have 25 possible ciphers
  - A maps to A,B,..Z
- could simply try each in turn
- a **brute force search**
- given ciphertext, just try all shifts of letters
- do need to recognize when have plaintext
### Brute Force of Caesar Cipher

<table>
<thead>
<tr>
<th>KEY</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>oggv og chvgt vjg vqic rctva</td>
</tr>
<tr>
<td>2</td>
<td>nffu nf bgufs uif upbb qhsuz</td>
</tr>
<tr>
<td>3</td>
<td>meet me after the toga party</td>
</tr>
<tr>
<td>4</td>
<td>ldds ld zesdq sgd snfz ozqsx</td>
</tr>
<tr>
<td>5</td>
<td>kccr kc ydrcp rfc rmey nyprw</td>
</tr>
<tr>
<td>6</td>
<td>jbbq jb xcqbo qeb qldx mxoqv</td>
</tr>
<tr>
<td>7</td>
<td>iaap ia wbpan pda pkcw lwnpu</td>
</tr>
<tr>
<td>8</td>
<td>hzzo hz vaozm ocz ojbv kvmot</td>
</tr>
<tr>
<td>9</td>
<td>gyny ny uzynl nby niau julns</td>
</tr>
<tr>
<td>10</td>
<td>fxxm fx tymxk max mhzt itkmr</td>
</tr>
<tr>
<td>11</td>
<td>ewwl ew sxlwj lzw lgys hsjlq</td>
</tr>
<tr>
<td>12</td>
<td>dvvk dv rwkvi kyy kfxr grikp</td>
</tr>
<tr>
<td>13</td>
<td>cuuj cu qvjuh jux jewq fghjo</td>
</tr>
<tr>
<td>14</td>
<td>btti bt puitg iwt idvp epgin</td>
</tr>
<tr>
<td>15</td>
<td>assh as othsf hvs hcuo dofhm</td>
</tr>
<tr>
<td>16</td>
<td>zrrg zr nsgre gur gbtn cnegl</td>
</tr>
<tr>
<td>17</td>
<td>yqqf yq mrfqd ftq fasm bmdfk</td>
</tr>
<tr>
<td>18</td>
<td>xppe xp lqopo osp ozrl aloej</td>
</tr>
<tr>
<td>19</td>
<td>wood wo kpdob dro dykq zkbdi</td>
</tr>
<tr>
<td>20</td>
<td>vnnc vn jocna cqn cxpj yjach</td>
</tr>
<tr>
<td>21</td>
<td>ummb um inbmz bpm bwoi xizbg</td>
</tr>
<tr>
<td>22</td>
<td>tlll tl hmalu aol avnh whyaf</td>
</tr>
<tr>
<td>23</td>
<td>skkz sk glzkx znk zumg vgxze</td>
</tr>
<tr>
<td>24</td>
<td>rjvy rj fkyjw ymj ytlf ufwyd</td>
</tr>
<tr>
<td>25</td>
<td>qiix qi ejxiv xli xske tevxc</td>
</tr>
</tbody>
</table>
Monoalphabetic Cipher

- rather than just shifting the alphabet
- could shuffle (jumble) the letters arbitrarily
- each plaintext letter maps to a different random ciphertext letter
- hence key is 26 letters long

Plain:  abcdefghijklmnopqrstuvwxyz
Cipher:  DKVQFIBJWPESCXHTMYAUOLRGZN

Plaintext:  ifwewishtoreplaceletters
Ciphertext:  WIRFRWAJUHYFTSDVFSFUUFYA
Monoalphabetic Cipher Security

- now have a total of $26! = 4 \times 10^{26}$ keys
- with so many keys, might think is secure
- but would be !!!WRONG!!!
Language Redundancy and Cryptanalysis

- problem is language characteristics
- human languages are **redundant**
- eg "th lrd s m shphrd shll nt wnt"
- letters are not equally commonly used
- in English E is by far the most common letter
  - followed by T,R,N,I,O,A,S
- other letters like Z,J,K,Q,X are fairly rare
- have tables of single, double & triple letter frequencies for various languages
English Letter Frequencies

![Bar chart showing the relative frequency of English letters.]}
Use in Cryptanalysis

- key concept - monoalphabetic substitution ciphers do not change relative letter frequencies
- discovered by Arabian scientists in 9th century
- calculate letter frequencies for ciphertext
- compare counts/plots against known values
- if caesar cipher look for common peaks/troughs
  - peaks at: A-E-I triple, NO pair, RST triple
  - troughs at: JK, X-Z
- for monoalphabetic must identify each letter
  - tables of common double/triple letters help
Example Cryptanalysis

- given ciphertext:
  UZQSOVUOHXMOPVGPOZPEVSGZWSZOPFPESXUBMETSXAIZ
  VUEPHZHMDZHSHZOWSFPAPPDTSVPQUZWYMXUZUHSX
  EPYEPOPDZSZUFPMBZWPFPUPZHMDJUDTMOHQ

- 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 viet cong in moscow
Monoalphabetic Cipher

Let’s play!
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
- invented by Charles Wheatstone in 1854, but named after his friend Baron Playfair
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

<table>
<thead>
<tr>
<th>M</th>
<th>O</th>
<th>N</th>
<th>A</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>H</td>
<td>Y</td>
<td>B</td>
<td>D</td>
</tr>
<tr>
<td>E</td>
<td>F</td>
<td>G</td>
<td>I/J</td>
<td>K</td>
</tr>
<tr>
<td>L</td>
<td>P</td>
<td>Q</td>
<td>S</td>
<td>T</td>
</tr>
<tr>
<td>U</td>
<td>V</td>
<td>W</td>
<td>X</td>
<td>Z</td>
</tr>
</tbody>
</table>
Encrypting and Decrypting

- plaintext is encrypted **two** letters at a time
  1. Repeating plaintext letters that are in the same pair are separated with a filler letter, such as x
  2. Two plaintext letters that fall in the same row of the matrix are each replaced by the letter to the right
  3. Two plaintext letters that fall in the same column are each replaced by the letter beneath
  4. Otherwise, each plaintext letter in a pair is replaced by
     - the letter that lies in its own row
     - and the column occupied by the other plaintext letter.
Security of 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. by US & British military in WW1
- it can be broken, given a few hundred letters
- since still has much of plaintext structure
Frequency of Occurrence of Letters
Polyalphabetic Ciphers

- **polyalphabetic substitution ciphers**
- improve security using multiple cipher alphabets
- make 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
- 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
Vigenère Tableau

| Plaintext | 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 |
| a | 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 |
| b | 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 | A |
| c | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B |
| d | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C |
| e | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D |
| f | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E |
| g | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F |
| h | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G |
| i | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H |
| j | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I |
| k | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J |
| l | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K |
| m | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L |
| n | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M |
| o | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
| p | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O |
| q | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P |
| r | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q |
| s | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R |
| t | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S |
| u | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T |
| v | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U |
| w | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V |
| x | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W |
| y | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X |
| z | Z | 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 | A |

Key: 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

Vigenère Tableau is a polyalphabetic substitution cipher that uses a repeating key to encrypt plaintext. The Vigenère Tableau is a grid used to perform the encryption and decryption process. Each column of the grid represents a letter of the alphabet, and each row represents a letter of the key. To encrypt a message, the plaintext letter is located in the first column, and the corresponding key letter is located in the first row. The intersection of these two letters gives the cipher letter. Decryption is done in a similar manner, using the plaintext key.
Example of Vigenère Cipher

- 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: deceptivedeceptivedeceptivedeceptivedeceptive
plaintext: wearediscoveredsaveyourself
ciphertext: ZICVTWQNGRZGVTWAVZHCQYGLMGJ
Security of Vigenère Ciphers

- have multiple ciphertext letters for each plaintext letter
- hence letter frequencies are obscured
- but not totally lost
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
- 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:       deceptivewearediscoveredsave
plaintext: wearediscoveredsaveyourself
ciphertext: ZICVTWQNGKZEIIIGASXSTSLVVWLA
One-Time Pad

- "Perfect" encryption scheme
  - Unbreakable
    - Implemented properly
- Gilbert Vernam in 1917
  - Vernam Cipher
- Using a Pad
  - Made up of random Values
- Using exclusive-OR
  - Add Circuit
One-Time Pad

- To get the cipher stream, the message stream will be XORed with the Keystream
  - One-Time Pad
- The receiver receives the cipher stream
- He/She must have the same one-time pad to decrypt the message
- By reversing the process
  - Doing the (same old) exclusive-OR!
I love you.
-Jenny

Plaintext

01010001
01100110
10100101

XOR

01110011
01101001
01111011
10010101
10110010

One-time pad

11010101
00010010
11000110
10100101

Ciphertext

11010101
00010010

XOR

I love you.
-Jenny
One-Time Pad

- **Perfect encryption scheme if**
  - The pad must be used only one time
  - The pad must be at least as long as the message
  - The pad must be securely distributed
  - The pad must be kept/protected at the sites
  - The pad must be made up of truly random values

- **Impractical in most situation**
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:
  \[ \text{memat} \ \text{rhtgpry} \ \text{etefeteoaaat} \]
- giving ciphertext
  \[ \text{MEMATRHTGPRYETEFETEAOAAT} \]
Row Transposition Ciphers

- a more complex transposition
- 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: attackp
  o s t p o n e
  d u n t i l t
  w o a m x y z
Ciphertext: TTNAAPTMSTSUOAODWCOIXKNLYPETZ
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 complex ciphers in use
- 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
Hagelin Rotor Machine
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 cipher
  - polyalphabetic ciphers
  - transposition ciphers
  - product ciphers and rotor machines
  - stenography