Cryptography and Network Security

Third Edition
by William Stallings

Lecture slides by Lawrie Brown
Basic Terminology

- **plaintext** - the original message
- **ciphertext** - the coded message
- **cipher** - algorithm for transforming plaintext/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)** - the study of principles/methods of deciphering ciphertext without knowing key
- **cryptology** - the field of both cryptography and cryptanalysis
Two kinds of Ciphers

State-of-the-art: two kinds of most popular encryption algorithms

• Symmetric ciphers
  – Sender and receiver share a common key

• Public key ciphers
  – Sender and receiver have asymmetric information of the key(s)
Symmetric Encryption

• or conventional / private-key / single-key
• was only type prior to invention of public-key in 1970’s
• remains very widely used
• sender and recipient share a common key
  – Both parties have full information of the key
• all classical encryption algorithms are common key (private-key)
  – Characteristic of conventional algorithms
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
Requirements

• two requirements for secure use of symmetric encryption:
  – a strong encryption algorithm (keeping key secret is sufficient for security)
  – a secret key known only to sender / receiver
    \[ Y = E_k(X) \]
    \[ X = D_k(Y) \]

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

• can characterize by:
  – type of encryption operations used
    • substitution / transposition / product systems
  – number of keys used
    • single-key or private / two-key or public
  – way in which plaintext is processed
    • Block: process one block of elements a time
    • Stream: continuous input, output one element a time
Types of Cryptanalytic Attacks

- **ciphertext only**
  - know a) algorithm b) ciphertext

- **known plaintext**
  - know some given plaintext/ciphertext pairs

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

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

- **chosen text**
  - select either plaintext or ciphertext to en/decrypt to attack cipher
Brute Force Search

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

<table>
<thead>
<tr>
<th>Key Size (bits)</th>
<th>Number of Alternative Keys</th>
<th>Time required at 1 encryption/µs</th>
<th>Time required at 10^6 encryptions/µs</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>$2^{32} = 4.3 \times 10^9$</td>
<td>$2^{31} \mu s = 35.8 \text{ minutes}$</td>
<td>2.15 milliseconds</td>
</tr>
<tr>
<td>56</td>
<td>$2^{56} = 7.2 \times 10^{16}$</td>
<td>$2^{55} \mu s = 1142 \text{ years}$</td>
<td>10.01 hours</td>
</tr>
<tr>
<td>128</td>
<td>$2^{128} = 3.4 \times 10^{38}$</td>
<td>$2^{127} \mu s = 5.4 \times 10^{24} \text{ years}$</td>
<td>5.4 x 10^{18} years</td>
</tr>
<tr>
<td>168</td>
<td>$2^{168} = 3.7 \times 10^{50}$</td>
<td>$2^{167} \mu s = 5.9 \times 10^{36} \text{ years}$</td>
<td>5.9 x 10^{30} years</td>
</tr>
<tr>
<td>26 characters</td>
<td>26! = $4 \times 10^{26}$</td>
<td>$2 \times 10^{26} \mu s = 6.4 \times 10^{12} \text{ years}$</td>
<td>6.4 x 10^{6} years</td>
</tr>
<tr>
<td>(permutation)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
More 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 (non-exist in real applications)

• computational security
  – given limited computing resources (eg time needed for calculations is greater than age of universe), the cipher cannot be broken
Classical Ciphers

• Examine a sampling of what might be called classical encryption techniques.

• Illustrate the basic approaches to symmetric encryption and the types of cryptanalytic attacks that must be anticipated.

• The two basic building blocks of all encryption techniques: substitution and transposition.
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 a letter \textit{three} places down the alphabet
- 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
  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

• mathematically give each letter a number
  a b c d e f g h i j k l m
  0 1 2 3 4 5 6 7 8 9 10 11 12
  n o p q r s t u v w x y z
  13 14 15 16 17 18 19 20 21 22 23 24 25

• then have Caesar cipher as:
  \[ C = E(p) = (p + k) \mod (26) \]
  \[ p = D(C) = (C - k) \mod (26) \]
  \[-\text{ modulo arithmetic: } 1 = 27 \mod 26, 3 = 29 \mod 26 \]
Cryptanalysis of Caesar Cipher

- only have 26 possible keys
  - Could shift $K = 0, 1, 2, \ldots, 25$ slots
- 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
- Test: break ciphertext
  - GCUA VQ DTGCM
Monoalphabetic Cipher

• rather than just shifting the alphabet
• could shuffle 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
  – The simplicity and strength of the monoalphabetic substitution cipher dominated for the first millennium AD.
• but would be !!!WRONG!!!
  – First broken by Arabic scientists in $9^{th}$ century
Frequency Analysis

• 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
• have tables of single, double & triple letter frequencies
English Letter Frequencies

![Bar graph showing the relative frequency of English letters]

- A: 8.167
- B: 1.492
- C: 2.282
- D: 4.253
- E: 12.702
- F: 2.228
- G: 2.015
- H: 6.094
- I: 6.996
- J: 0.153
- K: 0.772
- L: 4.025
- M: 2.406
- N: 6.749
- O: 7.807
- P: 1.294
- Q: 0.095
- R: 5.887
- S: 6.327
- T: 9.056
- U: 2.758
- V: 0.978
- W: 1.236
- X: 0.150
- Y: 1.974
- Z: 0.074
Use in Cryptanalysis

• key concept - monoalphabetic substitution ciphers do not change relative letter frequencies
• discovered by Arabian scientists in 9\textsuperscript{th} century
• 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:
  UZQSOVUOXMOPVGOZPEVSGZWSZOPFPESXUDBMETSXAI
  VUEPHZHMDZSHZOWSFPAAPPDTSPQQUZWYMXUZUHSX
  EPYEPOPDZSZUFPMOBZWPFPUPZHMDJUDTMOMQ

- 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
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

  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 "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
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
Example

key: deceptive deceptivedeceptivedeceptivedeceptivedeceptive
plaintext: wearediscovered saveyourself

• write the plaintext out
• write the keyword repeated above it
  – eg using keyword deceptive
• use each key letter as a caesar cipher key
• encrypt the corresponding plaintext letter

ciphertext: ZICVTWQNQGRZGVTWAVZHCQYGLMGJ
Vigenère Cipher

• simplest polyalphabetic substitution cipher is the **Vigenère Cipher**
• effectively multiple caesar ciphers
• key is d-letter 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
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

repetitions in ciphertext give clues to period
• so find same plaintext an exact period apart
• which results in the same ciphertext
• eg repeated “VTW” in previous example
  key: deceptive deceptive deceptive
  plaintext: wearediscovered saveyourself
  ciphertext: ZICVTWQNGRZGVTWAVZHCQYGLMGJ
  • suggests size of 3 or 9
• find a number of duplicated sequences, collect all their distances apart, look for common factors
• then attack each monoalphabetic cipher individually using same techniques as before
Autokey Cipher

- Use the plain text itself as part of the key
- eg. given key *deceptive*
  
  key: deceptive wearediscovered sav
  plaintext: wearediscovered sav eyourself
  ciphertext: ZICVTWQNGKZEIIGASXSTSLVVWLA

- but still have frequency characteristics to attack
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
  - No repetition of patterns
- 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:
  mematrhtgpry etefeteoaat
• giving ciphertext
  MEMATRHTGPRYETEFETEAOAT
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: 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
Ciphertext: TTNAAPTMTMSUOAODWCOIXKKNLYPETZ
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

- Multiple-stage substitution algorithms
- 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
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 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