Overview

1. Introduction, Pre-WWII cryptology
   - Lorenz Cipher (Fish)
     - Used by Nazis for high command messages
     - First programmable electronic computer built to break it
2. Enigma Cipher
   - Used by German Navy, Army, Air Force
   - Broken by team including Alan Turing
3. Post-WWII
   - Modern symmetric ciphers
   - Public-Key Cryptography

Menu

- Introduction to Cryptology
  - Terminology
  - Principles
  - Brief history of 4000 years of Cryptology
- Cryptology before World War II
  - A simple substitution cipher
  - [Break]
  - Breaking substitution cipher
  - Vigenère Cipher

What is cryptology?

- Greek: “krypto” = hide
- Cryptology = science of hiding
  - Cryptography, Cryptanalysis = hide meaning of a message
  - Steganography, Steganalysis = hide existence of a message
- Cryptography = secret writing
- Cryptanalysis = analyzing (breaking) secrets
  - Cryptanalysis is what attacker does
  - Decipher or Decryption is what legitimate receiver does

Cryptology and Security

Cryptology is a branch of mathematics.

Security is about people.

Attackers try find the weakest link. In most cases, this is not the mathematics.
Introductions

Encrypt

Decrypt

Plaintext → Ciphertext → Plaintext

Alice

Eve (passive attacker)

Bob

Insecure Channel

Introductions

Encrypt

Decrypt

Plaintext → Ciphertext → Plaintext

Alice

Malice (active attacker)

Bob

Insecure Channel

Cryptosystem

Ciphertext = \( E(\text{Plaintext}) \)

Required property: \( E \) must be invertible

Plaintext = \( D(\text{Ciphertext}) \)

Desired properties:

Without knowing \( D \) must be “hard” to invert

\( E \) and \( D \) should be easy to compute

Possible to have lots of different \( E \) and \( D \)

“The enemy knows the system being used.”

Claude Shannon

Kerckhoff’s Principle

- French handbook of military cryptography, 1883
- Cryptography **always** involves:
  - Transformation
  - Secret
- **Security should depend only on the key**
- Don’t assume enemy won’t know algorithm
  - Can capture machines, find patents, etc.
  - Too expensive to invent new algorithm if it might have been compromised

**Axis powers often forgot this**
Symmetric Cryptosystem

Ciphertext = E (K, Message)
Message = D (K, Ciphertext)

Desired properties:
1. Kerckhoff’s: secrecy depends only on K
2. Without knowing K must be “hard” to invert
3. Easy to compute E and D

All cryptosystems until 1970s were like this. Asymmetric cryptosystems allow encryption and decryption keys to be different.

Really Brief History
First 4000 years

Desired properties:
- Kerckhoff’s: secrecy depends only on K
- Without knowing K must be “hard” to invert
- Easy to compute E and D

Really Brief History - last 100+ years

Themes
- Armes race: cryptographers vs. cryptanalysts
  - Often disconnect between two (e.g., Mary Queen of Scots uses monoalphabetic cipher long after known breakable)
- Motivated by war (more recently: commerce)
- Driven by advances in technology, mathematics
  - Linguists, classicists, mathematicians, computer scientists, physicists
- Secrecy often means advances rediscovered and mis-credited

Types of Attacks
- Ciphertext-only - How much Ciphertext?
- Known Plaintext - often “Guessed Plaintext”
- Chosen Plaintext (get ciphertext)
  - Not as uncommon as it sounds!
- Chosen Ciphertext (get plaintext)
- Dumpster Diving
- Social Engineering
- “Rubber-hose cryptanalysis”
  - Cryptanalyst uses threats, blackmail, torture, bribery to get the key

Security vs. Pragmatics
- Trade-off between security and effort
  - Time to encrypt, cost and size of equipment, key sizes, change frequency
  - One-time pad (1918) offers theoretically “perfect” security, but unacceptable cost
    - Compromises lead to insecurity (class 2)
- Commerce
  - Don’t spend $10M to protect $1M
  - Don’t protect $1B with encryption that can be broken for $1M
- Military
  - Values (and attacker resources) much harder to measure
Simple Substitution Cipher

- Substitute each letter based on mapping
- Key is alphabet mapping:
  a → J, b → L, c → B, d → R, ..., z → F
- How secure is this cipher?

Key Space

- Number of possible keys
  26 (ways to choose what a maps to)
  * 25 (b can map to anything else)
  * 24 (c can map to anything else)
  ... * 1 (only one choice left for z)
  = 26! = 403291461126605635584000000

If every person on earth tried one per second, it would take 5B years to try them all.

Really Secure?

- Key space gives the upper bound
  - Worst possible approach for the cryptanalyst is to try all possible keys
- Clever attacker may find better approach:
  - Eliminate lots of possible keys quickly
  - Find patterns in ciphertext
  - Find way to test keys incrementally

Monoalphabetic Cipher

"XBW HGQW XS ACFPSUWG FWPGWXF CF AWWKZV CDQGJCDWA CD BHYJD DJXHGJ; WUWD XBW ZWJFX PHGCSHF YCDA CF GSHFWA LV XBW KGSYCFW SI FBJGCDQ RDSOZWAQW OCXBBWZA IGSY SXBWGF."

Frequency Analysis

"XBW HGQW XS ACFPSUWG FWPGWXF CF AWWKZV CDQGJCDWA CD BHYJD DJXHGJ; WUWD XBW ZWJFX PHGCSHF YCDA CF GSHFWA LV XBW KGSYCFW SI FBJGCDQ RDSOZWAQW OCXBBWZA IGSY SXBWGF."

W: 20 "Normal" English:
  e 12%
C: 11 t 9%
F: 11 a 8%
G: 11

Pattern Analysis

"XBe HGQe XS ACFPSUEG FePGeXF CF AeeKZV CDQGJCDeA CD BHYJD DJXHeG; eUeD XBe ZeJFX PHGCSHF YCDA CF GSHFeA LV XBe KGSYCFe SI FBJGCDQ RDSOZeAqE OCXBBeZA IGSY SXBeGF."

XBe = "the"
Most common trigrams in English:
  the = 6.4%
  and = 3.4%
Guessing
"the HGQe tS ACFPSUeG FePGetF CF AeeKZV CDQGJCDeA CD HHYJD DJtHGe; eueD the ZeJFt PHGCSHF YCDA CF GSHFeA LV the KGYSyFe SI FhJGCDQ RDOZeAQe OChthzeZA IGSY StheGF."

S = "0"

Guessing
"the HGQe to ACFPOUeG FePGetF CF AeeKZV CDQGJCDeA CD HHYJD DJtHGe; eueD the ZeJFt PHGCoHF YCDA CF GohFeA LV the KGoYCFe oI FhJGCDQ RDOOZeAQe OChthzeZA IGoY otheGF."

otheGF = "others"

Guessing
"the HRQe to ACSPoUer sePresCs AeeKZV CDqrJCeA CD HHYJD DJtHre; eueD the ZeJst PHrCHOs YCDA Cs roHseA LV the KroYCe oI shJrCDQ RDOOZeAQe OChthzeZA IroY others."

"sePresCs" = "secrets"

Guessing
"the HRQe to ACsCOUer secretsCs AeeKZV CDQjJCDeA CD HHYJD DJtHre; eueD the ZeJst CHRCHs YCDA CS roHseA LV the KroYCSe oI shJrCDQ RDOOZeAQe OChthzeZA IroY others."

"ACsCOUer" = "discover"

Guessing
"the HRQe to discover secrets is deekZv iDQRjiDeD iD HHYJD DJtHre; eved the ZeJst chirOs YiD ds roHSed LV the KroYise oI shJriDQ RDOOZedQe OithhezD IroY others."

Monoalphabetic Cipher
"The urge to discover secrets is deeply ingrained in human nature; even the least curious mind is roused by the promise of sharing knowledge withheld from others."
- John Chadwick,
The Decipherment of Linear B
Why was it so easy?
• Doesn’t hide statistical properties of plaintext
• Doesn’t hide relationships in plaintext (EE cannot match dg)
• English (and all natural languages) is very redundant: about 1.5 bits of information per letter (~68% of letters redundant)
  – Compress English with gzip ~ about 1:6

How to make it harder?
• Cosmetic
• Hide statistical properties:
  – Encrypt “e” with 12 different symbols, “t” with 9 different symbols, etc.
  – Add nulls, remove spaces
• Polyalphabetic cipher
  – Use different substitutions
• Transposition
  – Scramble order of letters

Ways to Convince
• “I tried really hard to break my cipher, but couldn’t. I’m a genius, so I’m sure no one else can break it either.”
• “Lots of really smart people tried to break it, and couldn’t.”
• Mathematical arguments – key size (dangerous!), statistical properties of ciphertext, depends on some provably (or believed) hard problem
• Invulnerability to known cryptanalysis techniques (but what about undiscovered techniques?)

Vigenère
• Invented by Blaise de Vigenère, ~1550
• Considered unbreakable for 300 years
• Broken by Charles Babbage but kept secret to help British in Crimean War
• Attack discovered independently by Friedrich Kasiski, 1863.

Vigenère Encryption
Key is a \( N \)-letter string.
\[ E_K (P) = C \]
\[ C_i = (P_i + K_i \mod N) \mod Z \]  
(size of alphabet)

\[ E_{KEY} ("test") = DIQD \]
\[ C_0 = (t + K) \mod 26 = D \]
\[ C_1 = (e + E) \mod 26 = I \]
\[ C_2 = (s + Y) \mod 26 = Q \]
\[ C_3 = (t + K) \mod 26 = D \]
Babbage’s Attack

- Use repetition to guess key length:
  Sequence XFO appears at 65, 71, 122, 176.
  Spacings = (71 – 65) = 6 = 3 * 2
  (122 – 65) = 57 = 3 * 19
  (176 – 122) = 54 = 3 * 18
  Key is probably 3 letters long.

Key length - Frequency

- Once you know key length, can slice ciphertext and use frequencies:
  \[ L_0: \text{DLQ}L\text{CN}Q\text{L}S\text{Q}R\text{KGBSEV}Y\text{NDOIOXYR}S\text{OS}GYKY} \]
  VZXVOCDNOSOOCOWDKOOGYROEVS\text{RBXENI}
  Frequencies: O: 12, S: 7, Guess O = e
  \[ C_i = (P_i + K_{i \mod N}) \mod Z \]
  ‘O’ = (‘e’ + K_0) \mod 26
  14 = 5 + 9 \Rightarrow K_0 = ‘K’

Sometimes, not so lucky...

L_1: "LMISQTVYJSSSHJYECYSGWGVJMRXEGWRPEJSI"
S: 9, X: 7, I: 6 guess S = ‘e’
  ‘S’ = (‘e’ + K_1) \mod 26
  19 = 5 + 14 \Rightarrow K_1 = ‘N’
  ‘X’ = (‘e’ + K_1) \mod 26
  24 = 5 + 19 \Rightarrow K_1 = ‘M’
  ‘I’ = (‘e’ + K_1) \mod 26
  10 = 5 + 5 \Rightarrow K_1 = ‘E’

Vigenère Simplification

- Use binary alphabet \{0, 1\}:
  \[ C_i = (P_i + K_{i \mod N}) \mod 2 \]
  \[ C_i = P_i \oplus K_{i \mod N} \]
- Use a key as long as P:
  \[ C_i = P_i \oplus K_i \]
- One-time pad – perfect cipher!

Perfectly Secure Cipher: One-Time Pad

- Mauborge/tern [1917]
- XOR (\(\oplus\)):
  \[
  \begin{align*}
  0 \oplus 0 &= 0 \\
  1 \oplus 0 &= 1 \\
  0 \oplus 1 &= 1 \\
  1 \oplus 1 &= 0
  \end{align*}
  \]
  a \oplus a = 0
  a \oplus 0 = a
  a \oplus b = a
- E(P, K) = P \oplus K
  D(C, K) = C \oplus K = (P \oplus K) \oplus K = P

Why perfectly secure?

For any given ciphertext, all plaintexts are equally possible.

Ciphertext: 01001
Key1: 01001
Plaintext1: 00000
Key2: 10110
Plaintext2: 11111
Perfect Security Solved?

- Cannot reuse $K$
  - What if receiver has
    \[ C_1 = P_1 \oplus K \text{ and } C_2 = P_2 \oplus K \]
    \[ C_1 \oplus C_2 = P_1 \oplus K \oplus P_2 \oplus K \]
    \[ = P_1 \oplus P_2 \]
- Need to generate truly random bit sequence as long as all messages
- Need to securely distribute key

Next week: “One-Time” Pads in Practice

- Lorenz Machine
- Nazi high command in WWII
  - Operator errors: reused key
- Pad generated by 12 rotors
  - Not really random

Public Lecture Tonight

- David Goldschmidt, “Communications Security: A Case History”
  - Director of Center for Communications Research, Princeton
  - Enigma and how it was broken
- 7:30pm Tonight
- UVa Physics Building, Room 203

We will cover some of the same material in the third class.