Cryptography in World War II
Jefferson Institute for Lifelong Learning at UVa
Spring 2006 **David Evans**

Class 1: Introduction and Cryptography before World War II

Bletchley Park, Summer 2004

http://www.cs.virginia.edu/jillcrypto

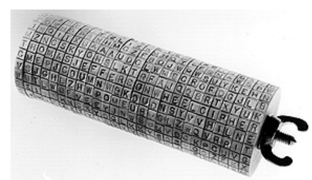
Overview

1. Introduction, Pre-WWII cryptology
2. Lorenz Cipher (Fish)
   – Used by Nazis for high command messages
   – First programmable electronic computer built to break it
3. Enigma Cipher
   – Used by German Navy, Army, Air Force
   – Broken by team including Alan Turing
4. 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

Jefferson Wheel 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

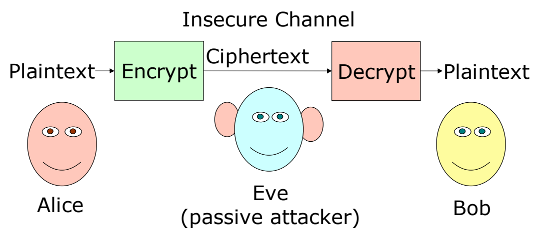
Cryptology and Security

Cryptology is a branch of *mathematics*.

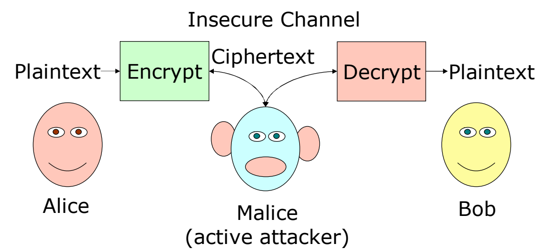
Security is about *people*.

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

Introductions



Introductions



Cryptosystem

Ciphertext = *E*(Plaintext)

Required property: *E* must be invertible

Plaintext = *D*(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."

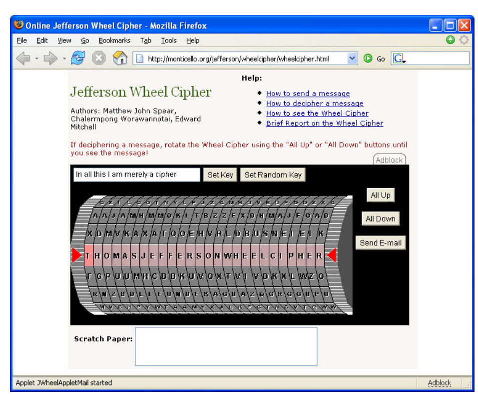
Claude Shannon

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

http://monticello.org/jefferson/wheelcipher/



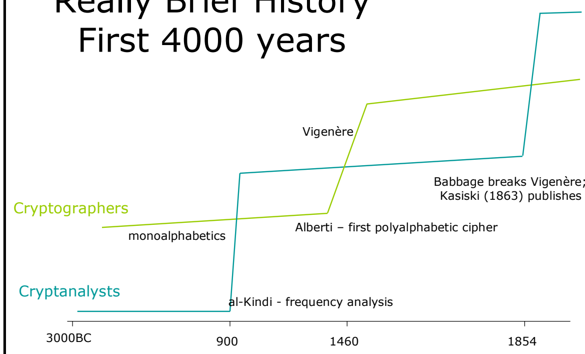
Symmetric Cryptosystem

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

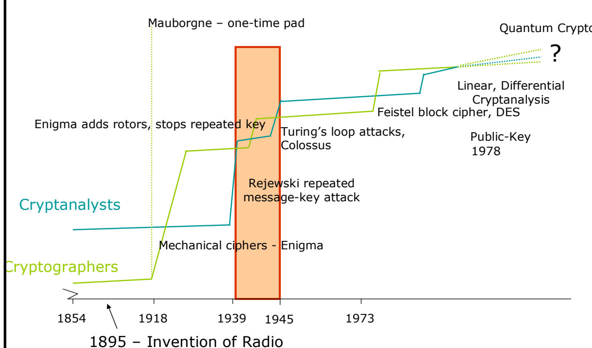
Desired properties:
1. Kerckhoff's: secrecy depends only on K
2. Without knowing K must be "hard" to invert
3. Easy to compute E and D

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

Really Brief History
First 4000 years



Really Brief History - last 100+ years



Themes

- Arms 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 patters in ciphertext
  - Find way to test keys incrementally

Monoalphabetic Cipher

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

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

Pattern Analysis

"XBe HGQe XS ACFPSUeG FePGeXF CF
AeeKZV CDQGJCDeA CD BHYJD DJXHGe;
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 KGSYCFe SI FhJGCDQ
RDSOZeAQe OCthheZA IGSY StheGF."

S = "o"

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 OCthheZA IGoY otheGF."

otheGF = "others"

Guessing

"the HrQe to ACsPoUer sePrets Cs
AeeKZV CDQrJCDeA CD hHYJD DJtHre;
eUeD the ZeJst PHrCoHs YCDA Cs
roHseA LV the KroYCse oI shJrCDQ
RDoOZeAQe OCthheZA IroY others."

"sePrets" = "secrets"

Guessing

"the HrQe to ACscoUer secrets Cs
AeeKZV CDQrJCDeA CD hHYJD DJtHre;
eUeD the ZeJst cHrCoHs YCDA Cs
roHseA LV the KroYCse oI shJrCDQ
RDoOZeAQe OCthheZA IroY others."

"ACscoUer" = "discover"

Guessing

"the HrQe to discover secrets is
deeKZV iDQrJiDed iD hHYJD DJtHre;
eveD the ZeJst cHrioHs YiDd is
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% f ltrs r redndnt)
  - 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
- Polyalphbetic 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.

$E_K(P) = C$ where

$C_i = (P_i + K_{i \bmod N}) \bmod Z$

(size of alphabet)

$E_{\text{"KEY"}}(\text{"test"}) = \text{DIQD}$

$C_0 = (\text{'t'} + \text{'K'}) \bmod 26 = \text{'D'}$

$C_1 = (\text{'e'} + \text{'E'}) \bmod 26 = \text{'I'}$

$C_2 = (\text{'s'} + \text{'Y'}) \bmod 26 = \text{'Q'}$

$C_3 = (\text{'t'} + \text{'K'}) \bmod 26 = \text{'D'}$


```
        12345678901234567890123456789012345678901234567890123456
Plain:  ThebiggerboysmadeciphersbutifIgotholdofafewwordsIusually
Key:    KEYKEYKEYKEYKEYKEYKEYKEYKEYKEYKEYKEYKEYKEYKEYKEYKEYKEYKE
Cipher: DLCLMEQIPLSWCQYNIASTFOVQLYRSJGQSRRSJNSDKJCGAMBHQSYQEEJVC

        7890123456789012345678901234567890123456789012345678901234
Plain:  foundoutthekeyTheconsequenceofthisingenuitywasoccasionally
Key:    YKEYKEYKEYKEYKEYKEYKEYKEYKEYKEYKEYKEYKEYKEYKEYKEYKEYKEYKEY
Cipher: DYYLNSSDXFOOCIXFOGMXWCAYCXGCYJRRMQSREORSSXWGEQYGAKWGYRYVPW

        56789012345678901234567890123456789012345678901234567 89
Plain:  painfultheownersofthedetectedcipherssometimesthrashedme
Key:    KEYKEYKEYKEYKEYKEYKEYKEYKEYKEYKEYKEYKEYKEYKEYKEYKEYKEYK
Cipher: ZEGXJSVXFOSUXIPCSDDLCNIROGROHASTFOVQCSKOXGWIQDLPKWFOHKO

        012345678901234567890123456789012345
Plain:  thoughthefaultlayintheirownstupidity
Key:    YKEYKEYKEYKEYKEYKEYKEYKEYKEYKEYKEYKE
Cipher: XFYYERXFOJYEPRVEWSRRRIGBSUXWRETGNMRI
```

Charles Babbage (quoted in Simon Singh, *The Code Book*)

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: DLQLCNSOLSQRNKGBSEVYNDOIOXAXYRSOSGYKY VZXVOXCDNOOSOCOWDKOOYROEVSRBXENI

Frequencies: O: 12, S: 7, Guess O = e

$C_i = (P_i + K_{i \bmod N}) \bmod Z$

$\text{'O'} = (\text{'e'} + K_0) \bmod 26$

$14 = 5 + 9 \Rightarrow K_0 = \text{'K'}$

Sometimes, not so lucky...

L_1: LMISQITVYJSSSJAHYECYSXOXGWYGJMRRXEGWRPEJXSI SLIGHTVSXILWHXYXJPERISWTM

S: 9, X: 7; I: 6 guess S = 'e'

$\text{'S'} = (\text{'e'} + K_1) \bmod 26$

$19 = 5 + 14 \Rightarrow K_1 = \text{'N'}$

$\text{'X'} = (\text{'e'} + K_1) \bmod 26$

$24 = 5 + 19 \Rightarrow K_1 = \text{'M'}$

$\text{'I'} = (\text{'e'} + K_1) \bmod 26$

$10 = 5 + 5 \Rightarrow K_1 = \text{'E'}$

Vigenère Simplification

- Use binary alphabet {0, 1}:

  $C_i = (P_i + K_{i \bmod N}) \bmod 2$

  $C_i = P_i \oplus K_{i \bmod N}$
- 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

- Mauborgne/Vernam [1917]
- XOR (⊕):

  0 ⊕ 0 = 0   1 ⊕ 0 = 1
  0 ⊕ 1 = 1   1 ⊕ 1 = 0

  a ⊕ a = 0
  a ⊕ 0 = a
  a ⊕ b ⊕ b = a
- E(P, K) = P ⊕ K
  D(C, K) = C ⊕ K = (P ⊕ K) ⊕ 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.