CS588:
Cryptology – Principles and Applications

Lecture 1: Introduction

With a magnetic card and his dog Buddy's name as a password, President Clinton e-signed a bill Friday that will make electronic signatures as real as those on paper.

FoxNews, 30 June 2000

CS588: Cryptology
University of Virginia
Computer Science

David Evans
http://www.cs.virginia.edu/~evans

Menu

- Course Introduction
  - Why you should or shouldn't take this course
  - Course Logistics: details on Syllabus
- Introduction to Cryptology
  - Terminology
  - A simple substitution cipher
  - Brief history of 4000 years of Cryptology

Send registration email by noon Friday.

Resources

- David Evans (call me "Dave"), devans@virginia.edu
  Office Hours (236A):
  Tuesdays, 10:30-11:30am; Weds after class
  Research: code safety, static analysis, programming and reasoning about swarms
- TAs:
  Danny Loffredo, dgl4b@virginia.edu
  CS Reading Room: Tuesdays, 3:30-4:30
  Anthony Wood, adw5p@virginia.edu
  TBA
- Web: http://www.cs.virginia.edu/cs588

Why you should take this course?

Reason #1: Fate of Humanity

Cryptology plays a central role in human history.

More than anything else, survival of humanity depends on computer security.

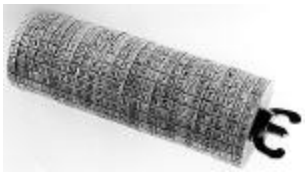
Why you should take this course? Reason #2: Intellectual

Cryptology is about making and solving puzzles.

Purest form of intellectual endeavor.

Why you should take this course? Reason #3: Be like Tom

Mr. Jefferson would have wanted you to.



Bad reasons to take this class

- You want to write the ultimate destructive virus.
- You want to break into (UVA's | the CIA's | your bank's) computer systems.

How to get an A in CS551

Problem Sets (**40**-50%)
  4 throughout term (1st is due 10 Sept)
Project (**30**-50%)
  Teams of 1 – 4
  Can involve design/implementation
  Can involve survey/analysis
Exams (**30**-50%)
  Midterm, Final
Class *Contribution* (**0**-10%)

"Easy ways" to get an A in CS551

- Break into my grades file (**on my home computer**) and change your grade to "Haha"
  - Physical attacks on my house, car or office are NOT eligible! (And NOT encouraged!)
  - Don't try to break into UVA's grade records:
    - Too easy (probably only worth a B, or C- for social engineering attack)
    - Honor code violation
- Discover a security flaw important enough to get reported in the New York Times
- Factor RSA-300 =

2769315567803442139028689061647233092237608363983953254005036722809375824714
9473946190060218756255124317186573105075074546238828817121274630072161346956
4396741836389979086904304472476001839015983033451909174663464663867829125664
4598955751571788169002287927112674719583575744167143664997220900115674047

Bonus Points / Demerits

(100 points = 1 problem set)

+100	Posting in RISKS
+(varies)	Solving a challenge problem
-100	Send me a virus
-200	Get arrested for computer attack
-1000	Get convicted for computer attack
-100000	I get arrested for something you do

Challenge Problems

- Open until solved or last day of class
- Usually only first satisfactory answer gets bonus
  - Better, later answer might still get bonus
- Solve in groups, each member gets
  $\sqrt{n} / n$ * value (e.g., 2 people = $\sqrt{2} / 2 = 0.7$)

First challenge problem starts tomorrow: Jefferson wheel cryptogram (see course web page)

Decrypting the Honor Code

- **Learn from your fellow students – they are your best resource!**
- Write down who you discussed assignments with, all external sources you used
- Don't use answers from last year's class
- **Be honest – you know what cheating is and isn't**
- Don't "pledge" your assignments, but let me know if you plan to cheat

Logistics Questions?

What is cryptology?

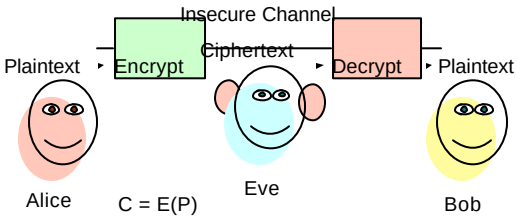
- Greek: “krypto” = hide
- Cryptology – science of hiding
  = cryptography + cryptanalysis + steganography
- Cryptography – secret writing
- Cryptanalysis – analyzing (breaking) secrets
  Cryptanalysis is what attacker does
  Decipher or *Decryption* is what legitimate receiver does
- Kryptonite – breaking ciphers all night?

Cryptology and Security

Cryptology is a branch of *mathematics*.

Security is about *people*.

Terminology



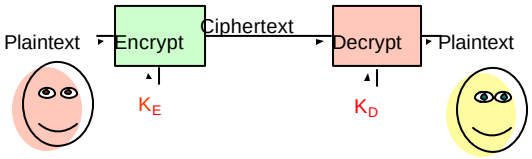
C = E(P)
P = D(C)
E must be invertible

Kerckhoff's Principle

- Cryptography **always** involves:
  - Transformation
  - Secret
- **Security should depend only on the key**
- Don't assume enemy won't know algorithm
  - Can capture machines, disassemble programs, etc.
  - Too expensive to invent new algorithm if it might have been compromised
- Security through obscurity isn't
  - Look at history of examples
  - Better to have scrutiny by open experts

"The enemy knows the system being used."
Claude Shannon

Alice and Bob



$C = E(K_E, P) = E_{K_E}(P)$
$P = D(K_D, C) = D_{K_D}(C)$
If $K_E = K_D$ it is *symmetric* encryption
If $K_E \neq K_D$ it is *asymmetric* encryption

Substitution Cipher

- $C = E_K(p)$
  $C_i = K[p_i]$
- Key is alphabet mapping:
  $a \rightarrow J$, $b \rightarrow L$, ...
- Suppose attacker knows algorithm but not key, how many keys to try? 26!
  If every person on earth tried one per second, it would take 5B years.

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."

W: 20
C: 11
F: 11
G: 11

"Normal" English:

e	12%
t	9%
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) are very redundant: about 1.3 bits of information per letter
  - Compress English with gzip – about 1:6

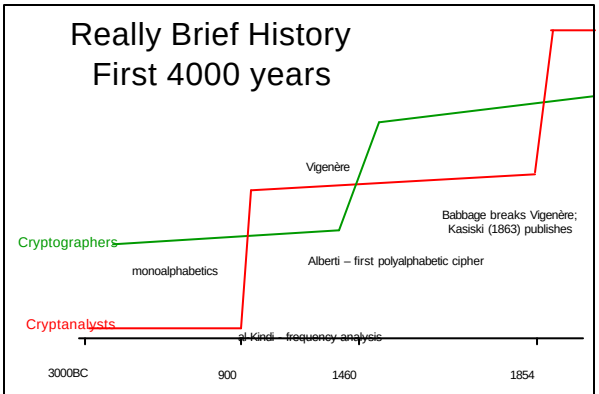
How to make it harder?

- 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

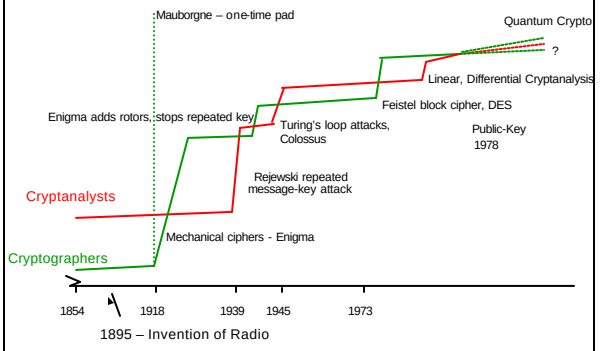
Types of Attacks

- 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 Not recommended in CS588
- Social Engineering
- "Rubber-hose cryptanalysis"
  - Cryptanalyst uses threats, blackmail, torture, bribery to get the key.

Really Brief History First 4000 years



Really Brief History - last 100 years



Themes

- Arms race between cryptographers and cryptanalysts
  - But, 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
- Multi-disciplinary field
  - Linguists, classicists, mathematicians, computer scientists, physicists
- Secrecy often means advances rediscovered and miscredited

Security vs. Pragmatics

- Trade-off between security and effort
  - one-time pad: perfect security, but requires distribution and secrecy of long key
  - DES: short key, fast algorithm, but breakable
  - quantum cryptography: perfect security, guaranteed secrecy of key, slow, requires expensive hardware
- Don't spend $10M to protect $1M.
- Don't protect $1B with encryption that can be broken for $1M.

Perfectly Secure Cipher: One-Time Pad

- Mauborgne/Vernam [1917]
- XOR ($\oplus$):
  $0 \oplus 0 = 0 \quad 1 \oplus 0 = 1$
  $0 \oplus 1 = 1 \quad 1 \oplus 1 = 0$
  $a \oplus a = 0$
  $a \oplus 0 = a$
  $a \oplus b \oplus b = a$
- $E(P, K) = P \oplus K$
  $D(C, K) = C \oplus K = (P \oplus K) \oplus K = P$

Why perfectly secure?

- For any given ciphertext, all plaintexts are equally possible.

  Ciphertext: **0100111110101**
  Key1: 1100000100110
  Plaintext1: 1000111010011 = “CS”
  Key2: 1100010100110
  Plaintext2: 1000101010011 = “BS”
- More formal proof next time

Go to the beach?

- Cannot reuse K
  - What if receiver has

    $C_1 = P_1 \oplus K$ 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

“One-Time” Pad’s in Practice

- Lorenz Machine – Nazi high command in WWII
  - Pad generated by 12 rotors
  - Receiver and sender set up rotors in same positions
  - One operator retransmitted a message (but abbreviated message header the second time!)
  - Enough for Bletchley Park to figure out key – and structure of machine that generated it!
  - But still had to try all configurations

Colossus – First Programmable Computer

- Bletchley Park, 1944
- Read ciphertext and Lorenz wheel patterns from tapes
- Tried each alignment, calculated correlation with German
- Decoded messages (63M letters by 10 Colossus machines) that enabled Allies to know German troop locations to plan D-Day
- Destroyed in 1960, kept secret until 1970s

Charge

- **Send me your registration survey by noon tomorrow**
- Start thinking about projects and teams
- Subscribe to comp.risks and Cryptogram (instructions on manifest)
- Next time:
  - Proving Ciphers are Perfect (in Theory)
  - Information Theory