

---

---

# Cryptographic Foundations and History

9 March 2010  
Lecture 2

Slide Credits: Steve Zdancewic (UPenn)

---

# Topics for Today

---

- What is computer security?
- Beginners Cryptography
  - Alphabetic Substitution Ciphers
  - Vigenere
  - Cracking (Cryptanalysis)
- Source: Anderson 1, Ford and Baum 4.2,4.3, HAC 1-2, SE 5

# What is Computer Security?

---

- What does security mean for a computer?
- What are we worried about?
- What should we be worried about?

# What is Computer Security?

---

You may think of hackers breaking into computer systems in the movies

- They do exist, but there is very little that we can do about them

Most (publicized) attacks are based on known vulnerabilities, misconfigured systems, or weak execution of security policies

Organized crime is alive and well

- Botnets and Spammets

Corporate (and national) espionage is alive and well

Most governments and intelligence agencies maintain active computer security divisions

- Israel has a decent electronic espionage unit, so does every major country (and terrorist group)

Insider attacks are the most common and difficult to prevent

# What is Computer Security?

---

The things on the previous slide are not our concern in this course

We want to understand the digital and mathematical tools of the trade

Technological and software security is another (more advanced) course

- You may learn this stuff on your own or through experience
- There are many good books on the subject (we will be studying from two of them)

# Security Atoms

---

- Authentication
  - Who are you talking to?
- Transaction Integrity and Accountability
  - Remember repudiation?
- Fault Tolerance
  - Know about failures
  - Recover gracefully
- Message Secrecy
  - Hide what you are saying
- Covertness
  - Hide that you are even communicating

# Security Atoms

---

- Security is built of many pieces
  - Electronic
  - Physical
  - Procedural
  - Mathematical
  - Psychological
- We focus on the **electronic** and **mathematical**
  - Not to minimize the others, but it's hard enough to get these two right

# What is a “System”?

---

1. a product or component, such as a cryptographic protocol, a smartcard or the hardware of a PC
2. a collection of the above plus an operating system, communications and other things that go to make up an organization’s infrastructure
3. the above plus one or more applications (media player, browser, word processor, accounts / payroll package, and so on)
4. any or all of the above plus IT staff
5. any or all of the above plus internal users and management
6. any or all of the above plus customers and other external users

# Definitions

---

- Subject – A physical person
- Principal - An entity which participates in a security system
- Role - A set of functions assumed by different people in succession
- Group - A set of principals
- Trusted - A system or component whose failure can break the security policy
- Trustworthy – A system or component is one that won't fail

# Definitions

---

---

- **Secrecy** סודיות is a technical term which refers to the effect of the mechanisms used to limit the number of principals who can access information, such as cryptography or computer access controls.
- **Confidentiality** involves an obligation to protect some other person's or organization's secrets if you know them.
- **Privacy** פרטיות is the ability and/or right to protect your personal information and extends to the ability and/or right to prevent invasions of your personal space (the exact definition of which varies quite sharply from one country to another).

# So far

---

- What is computer security?
- **Beginners Cryptography**
  - Alphabetic Substitution Ciphers
  - Vigenere
  - Cracking (Cryptanalysis)
- **Source: Anderson 1, Ford and Baum 4.2,4.3, HAC 1-2, SE 5**

# Κρυπτογραφία (Cryptography)

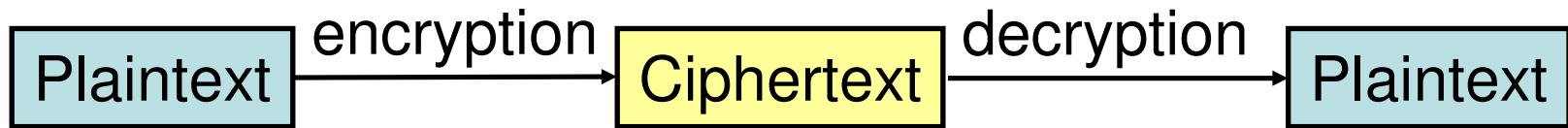
---

- Traditionally, most computer security courses start with an introduction to Cryptography
  - From the Greek "kryptos" and "graphia" for “secret writing”
- Confidentiality
  - Obscure a message from eaves-droppers
- Integrity
  - Assure recipient that the message was not altered
- Authentication
  - Verify the identity of the source of a message
- Non-repudiation
  - Convince a 3<sup>rd</sup> party that what was said is accurate

# Terminology

---

---



- Cryptographer
  - Invents cryptosystems
- Cryptanalyst
  - Breaks cryptosystems
- Cryptology
  - Study of crypto systems
- Cipher
  - Mechanical way of encrypting text or data
- Code
  - Semantic translation: “eat breakfast tomorrow” = “attack on Thursday”  
(or use Navajo!)

# Kinds of Cryptographic Analysis

---

Goal is to recover the key (& algorithm)

- **Ciphertext only attacks**
  - No information about content or algorithm
  - Very hard
- **Known Plaintext attacks**
  - Full or partial plaintext available in addition to ciphertext
- **Chosen Plaintext attacks**
  - Know which plaintext has been encrypted
- **Algorithm & Ciphertext attacks**
  - Known algorithm, known ciphertext, recover key

# The Caesar Cipher

---

- Purportedly used by Julius Caesar (c. 75 B.C.)
  - Add 3 mod 26

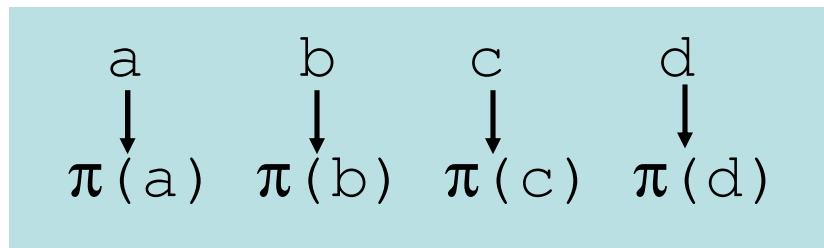
- Advantages
  - Simple
  - Intended to be performed in the field
  - Most people couldn't read anyway
- Disadvantages
  - Violates “no security through obscurity”
  - Easy to break (why?)

a	b	c	...	x	y	z
↓	↓	↓		↓	↓	↓
d	e	f	...	a	b	c

# Monoalphabetic Ciphers

---

- Also called *substitution* ciphers
- Separate *algorithm* from the *key*
  - Add N mod 26
  - rot13 = Add 13 mod 26
- General monoalphabetic cipher
  - Arbitrary permutation  $\pi$  of the alphabet
  - Key is the permutation



# Example Cipher

---

---

	a	b	c	d	e	f	g	h	i	j	k	l	...
$\pi$	z	d	a	n	c	e	w	i	b	f	g	h	...

Plaintext:    **he lied**

Ciphertext: **ic hbcn**

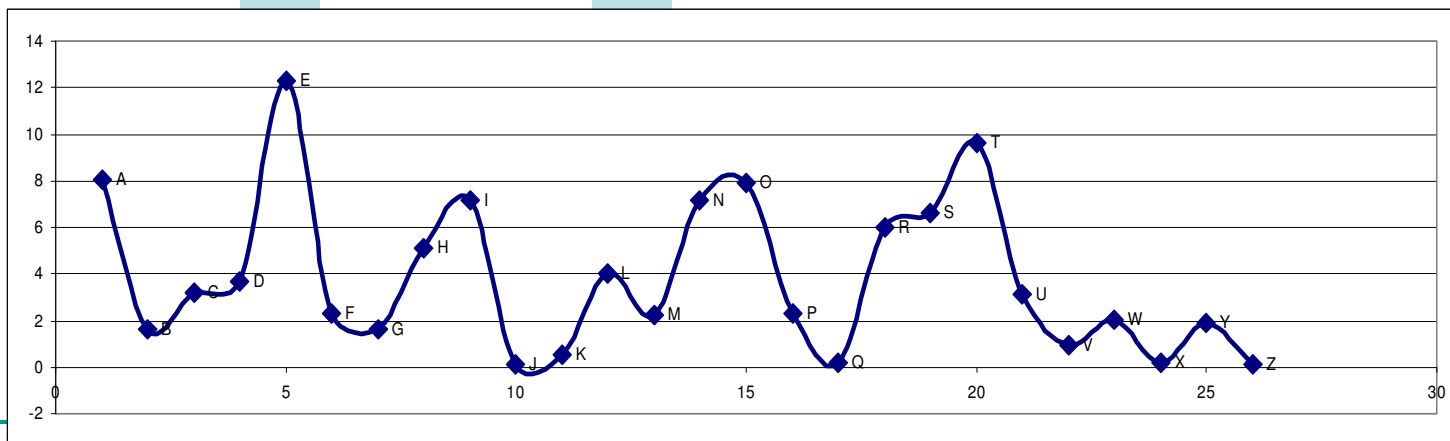
# Cryptanalysis of Monoalphabetic Ciphers

---

- Brute force attack: try every key
  - $N!$  Possible keys for N-letter alphabet
  - $26! \approx 4 \times 10^{26}$  possible keys
  - Try 1 key per  $\mu\text{sec}$  ... 10 trillion years
- ...but (!) monoalphabetic ciphers are *easy* to solve
  - One-to-one mapping of letters is bad
  - Frequency distributions of common letters

# Order & Frequency of Single Letters

<b>E</b>	<b>12.31%</b>	<b>L</b>	<b>4.03%</b>	<b>B</b>	<b>1.62%</b>
<b>T</b>	<b>9.59</b>	<b>D</b>	<b>3.65</b>	<b>G</b>	<b>1.61</b>
<b>A</b>	<b>8.05</b>	<b>C</b>	<b>3.20</b>	<b>V</b>	<b>0.93</b>
<b>O</b>	<b>7.94</b>	<b>U</b>	<b>3.10</b>	<b>K</b>	<b>0.52</b>
<b>N</b>	<b>7.19</b>	<b>P</b>	<b>2.29</b>	<b>Q</b>	<b>0.20</b>
<b>I</b>	<b>7.18</b>	<b>F</b>	<b>2.28</b>	<b>X</b>	<b>0.20</b>
<b>S</b>	<b>6.59</b>	<b>M</b>	<b>2.25</b>	<b>J</b>	<b>0.10</b>
<b>R</b>	<b>6.03</b>	<b>W</b>	<b>2.03</b>	<b>Z</b>	<b>0.09</b>
<b>H</b>	<b>5.14</b>	<b>Y</b>	<b>1.88</b>		



# Monoalphabetic Cryptanalysis

---

- Count the occurrences of each letter in the cipher text
- Match against the statistics of English
  
- Most frequent letter likely to be “e”
- 2<sup>nd</sup> most frequent likely to be “t”
- etc.
  
- Longer ciphertext makes statistical analysis more likely to work...

# Digrams and Trigrams

---

- Digrams in frequency order

TH HE AN IN ER RE

ES ON EA TI AT ST

EN NR OR

- Trigrams in frequency order

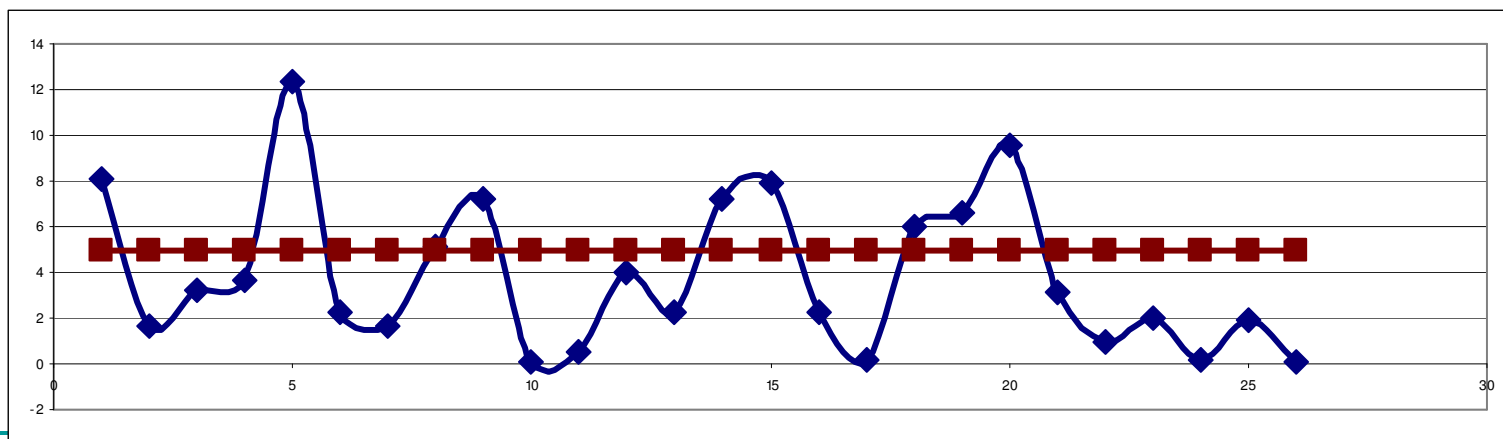
THE AND THA ENT ION TIO

FOR NDE HAS NCE EDT TIS

OFT STH MEN

# Desired Statistics

- Problems with monoalphabetic ciphers
  - Frequency of letters in ciphertext reflects frequency of plaintext
- Want a single plaintext letter to map to multiple ciphertext letters
  - “e” → “x”, “c”, “w”
- Ideally, ciphertext frequencies should be flat



# Polyalphabetic Substitutions

---

- Pick  $k$  substitution ciphers
  - $\pi_1 \pi_2 \pi_3 \dots \pi_k$
  - Encrypt the message by rotating through the  $k$  substitutions

<b>m</b>	<b>e</b>	<b>s</b>	<b>s</b>	<b>a</b>	<b>g</b>	<b>e</b>
$\pi_1(\mathbf{m})$	$\pi_2(\mathbf{e})$	$\pi_3(\mathbf{s})$	$\pi_4(\mathbf{s})$	$\pi_1(\mathbf{a})$	$\pi_2(\mathbf{g})$	$\pi_3(\mathbf{e})$
<b>q</b>	<b>a</b>	<b>x</b>	<b>o</b>	<b>a</b>	<b>u</b>	<b>v</b>

- Same letter can be mapped to multiple different ciphertexts
  - Helps smooth out the frequency distributions
  - *Diffusion*

# Vigenère Tableau

---

- Multiple substitutions
  - Can choose “complimentary” ciphers so that the frequency distribution flattens out
  - More generally: more substitutions means flatter distribution
- Vigenère Tableau
  - Invented by Blaise de Vigenère for the court of Henry III of France (c. 1500’s)
  - Collection of 26 permutations
  - Usually thought of as a 26 x 26 grid
  - Key is a word

# Vigenère Tableau

---

	a	b	c	d	e	f	g	.	.	.
A	a	b	c	d	e	f	g	.	.	.
B	b	c	d	e	f	g	h	.	.	.
C	c	d	e	f	g	h	i	.	.	.
D	d	e	f	g	h	i	j	.	.	.
E	e	f	g	h	i	j	k	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.

Plaintext:           a bad deed

Key "bed":           B EDB EDBE

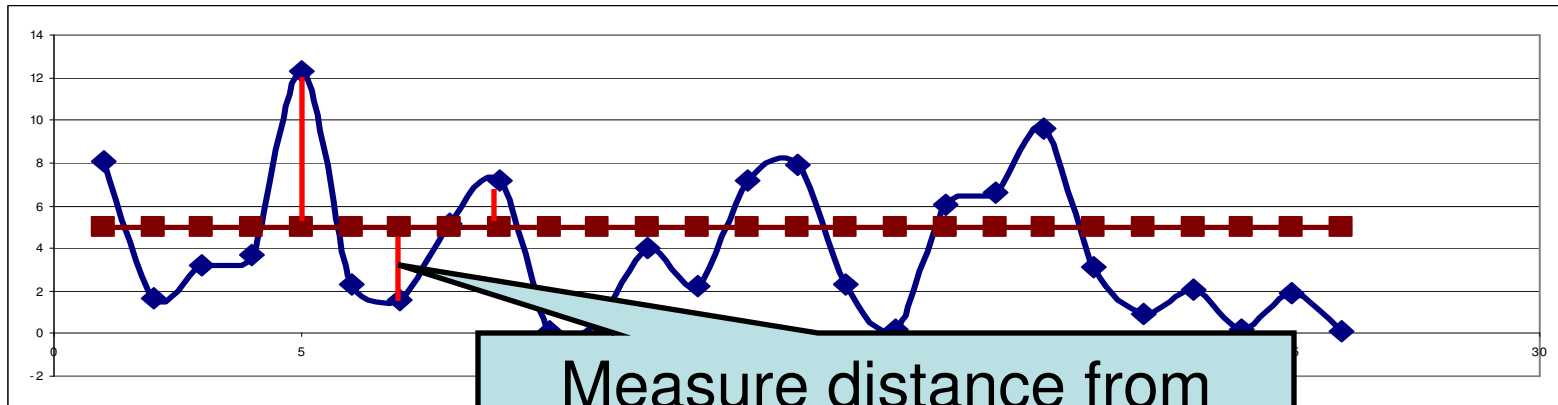
Ciphertext:         b fde hhfh

# Cryptanalysis on Polyalphabetics

---

- Once key length is guessed to be  $k$ ...
- Split ciphertext into  $k$  slices
  - Single letter frequency distribution for each slice should resemble English distribution
- How do we tell whether a particular distribution is a good match for another?
  - Let  $\text{prob}(\alpha)$  be the probability for letter  $\alpha$
  - In a perfectly flat distribution
    - $\text{prob}(\alpha) = 1/26 \approx 0.0384$

# Variance: Measure of “roughness”



Measure distance from  
“flat” dist.

$$\begin{aligned}\text{Var} &= \sum_{\alpha = a}^{\alpha = z} (\text{prob}(\alpha) - 1/26)^2 \\ &= \dots \\ &= \left( \sum_{\alpha = a}^{\alpha = z} \text{prob}(\alpha)^2 \right) - 1/26\end{aligned}$$

# Estimate Variance From Frequency

---

- $\text{prob}(\alpha)^2$  is probability that any two characters drawn from the text will be  $\alpha$
- Suppose there are  $n$  ciphertext letters total
- Suppose  $\text{freq}(\alpha)$  is the frequency of  $\alpha$
- What is likelihood of picking  $\alpha$  twice at random?
  - $\text{freq}(\alpha)$  ways of picking the first  $\alpha$
  - $(\text{freq}(\alpha) - 1)$  ways of picking the second  $\alpha$
  - But this counts twice because  $(\alpha, \beta) = (\beta, \alpha)$
  - So 
$$\frac{\text{freq}(\alpha) \times (\text{freq}(\alpha) - 1)}{2}$$

# Index of Coincidence

---

- But there are  $\frac{n \times (n-1)}{2}$  pairs of letters
- ...so  $\text{prob}(\alpha)$  is roughly  $\frac{\text{freq}(\alpha) \times (\text{freq}(\alpha) - 1)}{n \times (n - 1)}$
- Index of coincidence: approximates variance from frequencies

$$IC = \sum_{\alpha=a}^{\alpha=z} \frac{\text{freq}(\alpha) \times (\text{freq}(\alpha) - 1)}{n \times (n - 1)}$$

# What is it good for?

---

- If the distribution is flat, then  $IC = 0.0384$
- If the distribution is like English, then  $IC = 0.068$
- Can verify key length:

Keylen	1	2	3	4	5	many
IC	0.068	0.052	0.047	0.044	0.044	... 0.038

# Summary: Cracking Polyalphabetic

---

- Guess likely key length (if you are interested, look up “Kasiski” on Wikipedia)
- Compute the Index of Coincidence to verify key length  $k$
- $k$ -Slices should have similar IC to English
  
- Note: digram information harder to use for polyalphabetic ciphers...
  - May want to consider “split digrams”
  - Example: if tion is a common sequence  $k=2$  then “t?o” and “i?n” are likely “split digrams”

# Summary

---

- Beginners Cryptography
  - Alphabetic Substitution Ciphers
  - Cracking (Cryptanalysis)