CIS551: Computer and Network Security

Jonathan M. Smith

jms@cis.upenn.edu

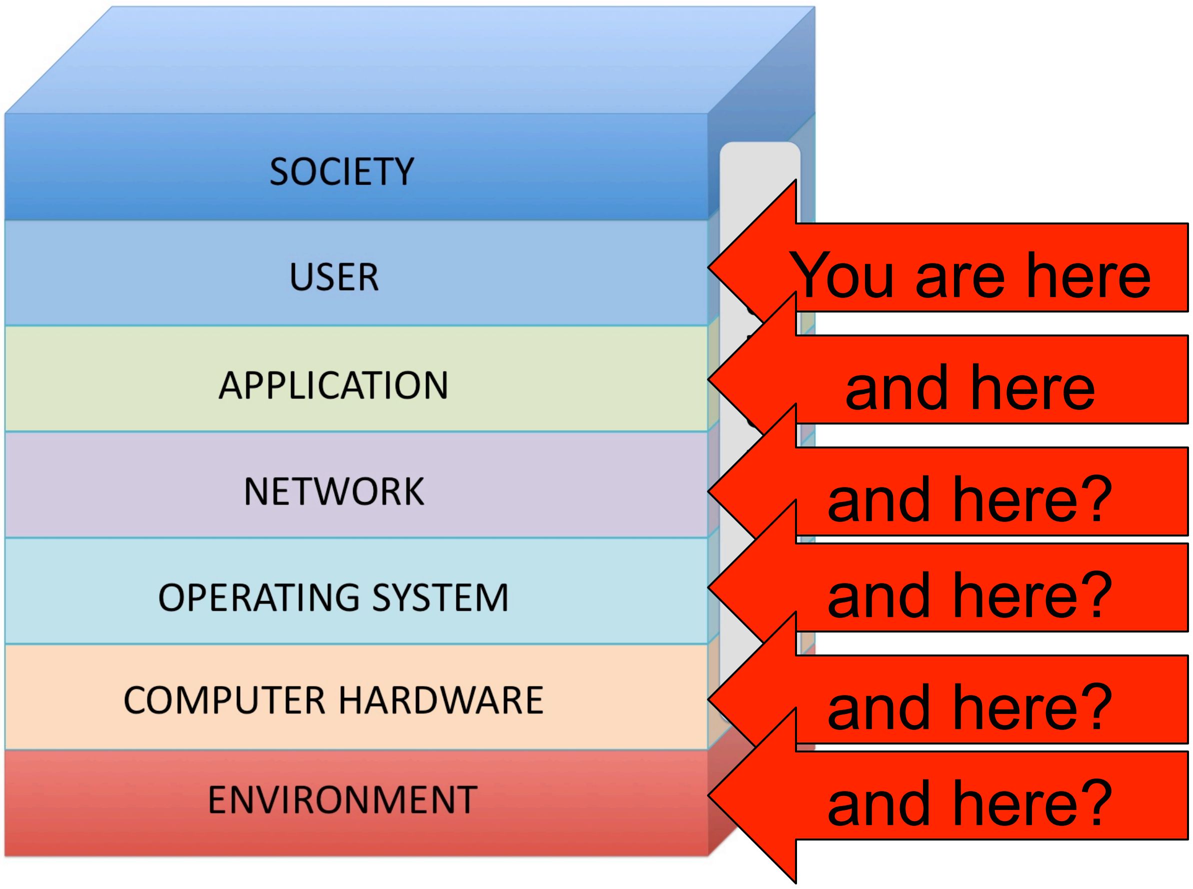
03/05/2014

Contains material from S. Zdancewic

CIS551 Topics

- Computer Security
  - Software/Languages, Computer Arch.
  - Access Control, Operating Systems
  - Threats: Vulnerabilities, Viruses
- Computer Networks
  - Physical layers, Internet, WWW, Applications
  - Cryptography in several forms
  - Threats: Confidentiality, Integrity, Availability
- Systems Viewpoint
  - Users, social engineering, insider threats

Sincoskie NIS model



W.D. Sincoskie, *et al.* “Layer Dissonance and Closure in Networked Information Security” (white paper)

Κρυϗτογραφία (Cryptography)

- From the Greek "kryptos" and "graphia" for "secret writing"
- Confidentiality
  - Obscure a message from eavesdroppers
- Integrity
  - Assure recipient that the message was not altered
- Authentication
  - Verify the identity of the source of a message
- Non-repudiation
  - Convince a 3rd 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!)
- Key
  - a parameter of the cipher algorithm

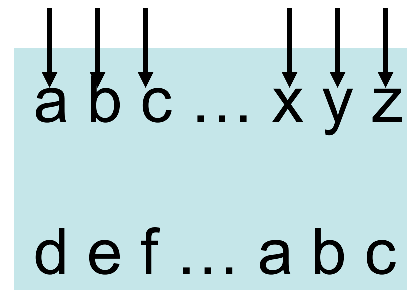
Contains material from S. Zdancewic

Kinds of Cryptographic Analysis

- Goal is to recover the key (& algorithm)
  - And hence recover the plaintext
- *Ciphertext Only* attacks
  - No information about content or algorithm
  - Very hard
- *Algorithm & Ciphertext* attacks
  - Known algorithm, known ciphertext, recover key
  - Common in practice
- *Known Plaintext* attacks
  - Full or partial plaintext available in addition to ciphertext
- *Chosen Plaintext* attacks
  - Attacker can choose which plaintext is encrypted, tries to reverse engineer the key. May be able to choose multiple plaintexts.

The Caesar Cipher

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

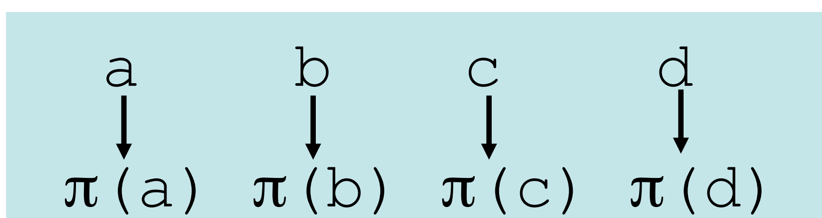


- Advantages
  – Simple
  – Intended to be performed in the field
  – Most people couldn't read anyway!
- Disadvantages
  – Violates "no security through obscurity"
  – Easy to break (why?)

Contains material from S. Zdancewic

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 π of the alphabet
  - “key” is the permutation



Contains material from S. Zdancewic

Example Cipher

	a	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**	**k**	**l**	**...**
π	**z**	**d**	**a**	**n**	**c**	**e**	**w**	**i**	**b**	**f**	**g**	**h**	**...**

Plaintext: **he lied**

Ciphertext: **ic hbcn**

Contains material from S. Zdancewic

Cryptanalysis of Monoalphabetic Ciphers

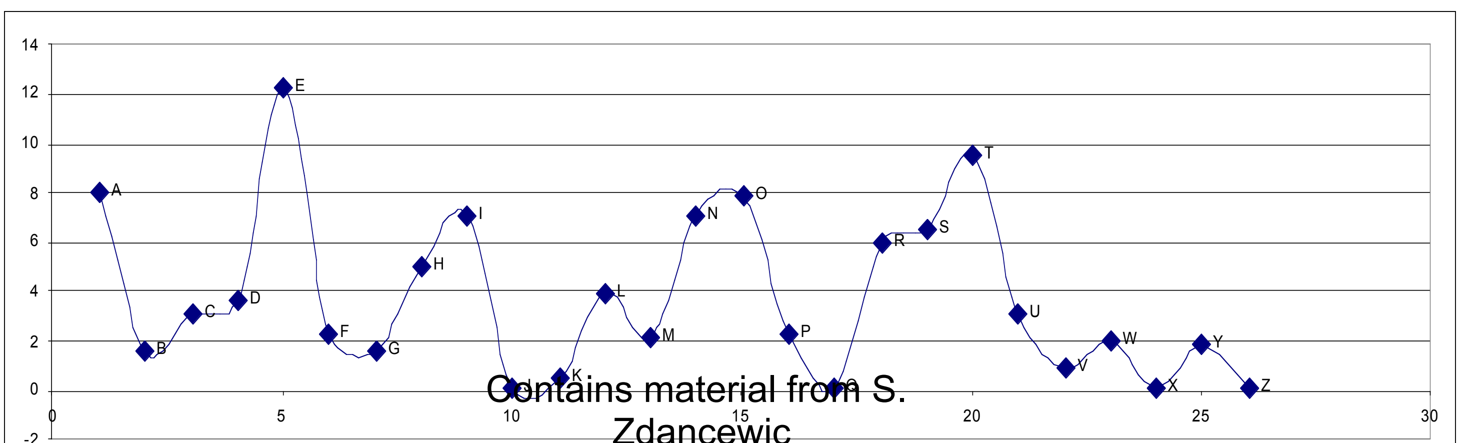
- Brute force attack: try every key
  - N! Possible keys for N-letter alphabet
  - 26! $\approx$ 4 x 10^{26} possible keys
  - Try 1 key per μsec … 10 trillion years

- …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

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



Contains material from S. Zdancewic

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"
- 2nd most frequent likely to be "t"
- etc.

- Longer ciphertext makes statistical analysis more likely to work…

Digrams and Trigrams

TH HE AN IN ER RE ES ON EA TI AT ST EN ND OR

- Digrams in frequency order (for English)

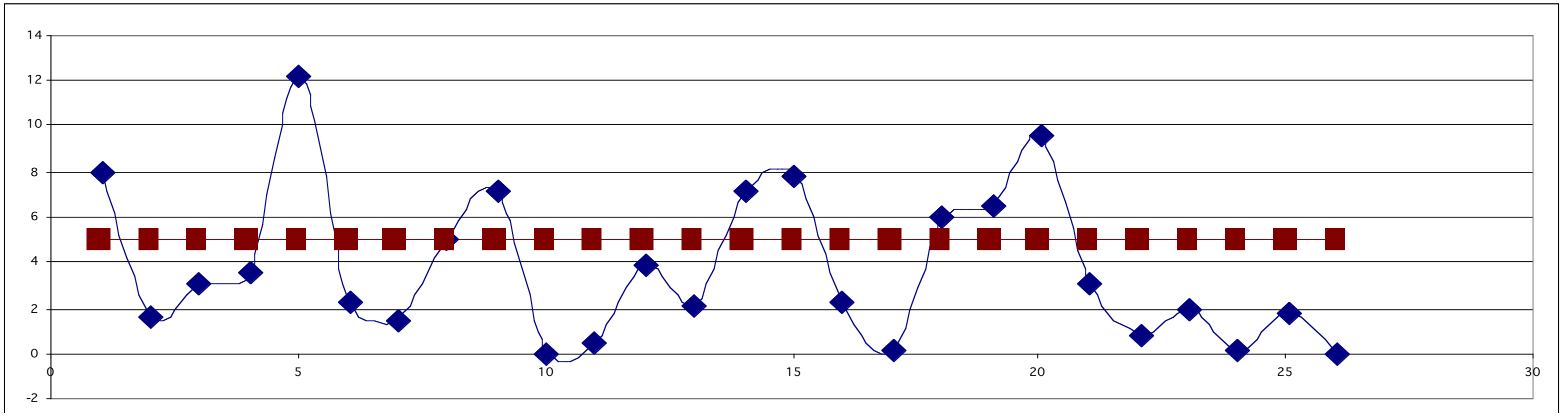
THE AND THA ENT ION TIO FOR NDE HAS NCE EDT TIS OFT STH MEN

- Trigrams in frequency order (for English)

Contains material from S. Zdancewic

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



Contains material from S. Zdancewic

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 hgfh**

Contains material from S. Zdancewic

Polyalphabetic Substitutions

m	**e**	**s**	**s**	**a**	**g**	**e**
π_1(**m**)	π_2(**e**)	π_3(**s**)	π_4(**s**)	π_1(**a**)	π_2(**g**)	π_3(**e**)
q	**a**	**x**	**o**	**a**	**u**	**v**

- Pick k substitution ciphers
  - $\pi_1\ \pi_2\ \pi_3\ \ldots\ \pi_k$
  - Encrypt the message by rotating through the k substitutions
- Same letter can be mapped to multiple different ciphertexts
  - Helps smooth out the frequency distributions
  - *Diffusion*

Contains material from S. Zdancewic

Cracking Polyalphabetic Substitutions

- Step 1:
  - Try to identify the number of substitutions used
  - For example, guess the length of the word used as a key in the Vigenère Tableau.

- Step 2:
  - Use frequency information to crack each of the substitutions as though it was a monoalphabetic cipher.

Kasiski Method

Goal: Identify key length of polyalphabetic ciphers

- Observation: If pattern appears k times and key length is n then it will be encoded k/n times by the same key

1. Identify repeated patterns of ≥ 3 chars.
2. For each pattern
   - Compute the differences between starting points of successive instances
   - Determine the factors of those differences
3. Key length is likely to be one of the frequently occurring factors

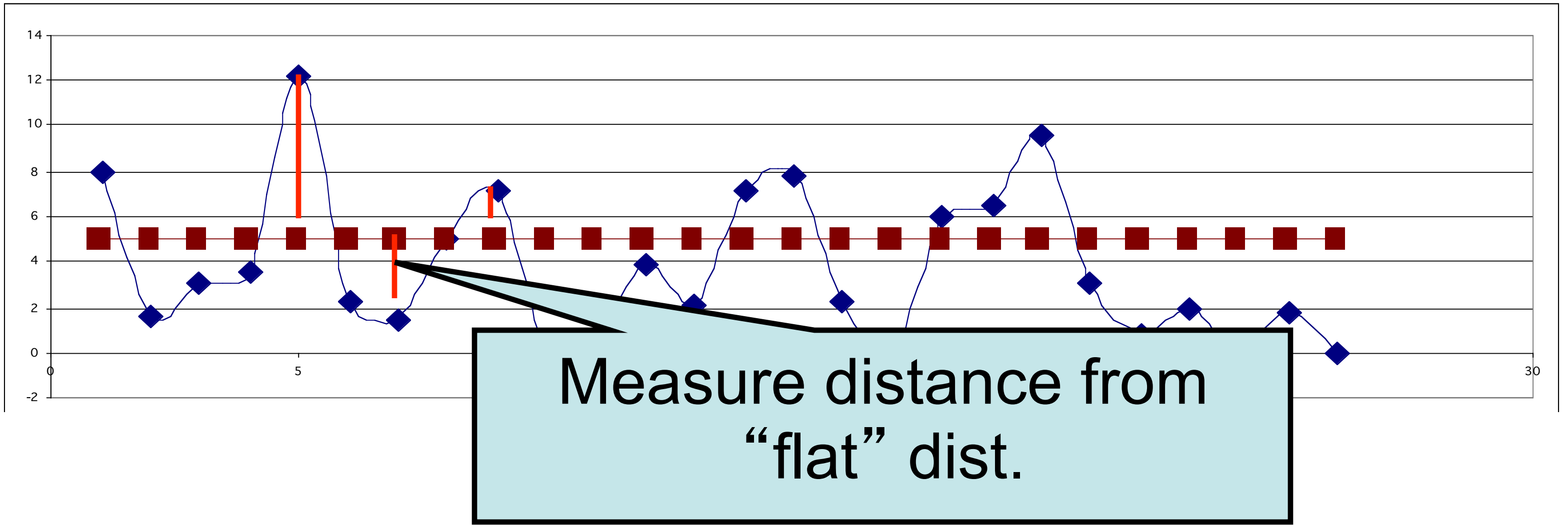
Cryptanalysis Continued

- 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 prob(α) be the probability for letter α
  - In a perfectly flat distribution

$$\text{prob}(\alpha) = 1/26 \approx 0.0384$$

Variance: Measure of "roughness"



$$\text{Var} = \sum_{\alpha = a}^{\alpha = z} (\text{prob}(\alpha) - 1/26)^2$$

$$= \ldots$$

$$= \left(\sum_{\alpha = a}^{\alpha = z} \text{prob}(\alpha)^2 \right) - 1/26$$

Contains material from S. Zdancewic

Estimate Variance From Frequency

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

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

- Index of coincidence: approximates variance from frequencies

What's 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 Polyalphabetics

- Use Kasiski method to guess likely key lengths
- Compute the Index of Coincidence to verify key length k
- k-Slices should have IC similar 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”