Information Security
CS 526
Topic 2

Cryptography: Terminology & Classic Ciphers

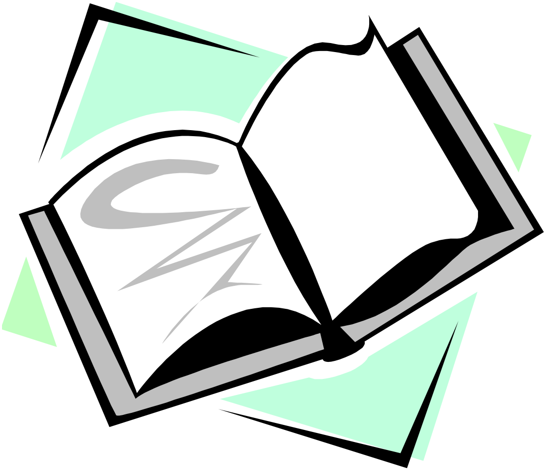
Readings for This Lecture

Required readings:

- [Cryptography on Wikipedia]

Interesting reading

- The Code Book by Simon Singh



Goals of Cryptography

- The most fundamental problem cryptography addresses: ensure security of communication over insecure medium
- What does secure communication mean?
  - confidentiality (secrecy)
    - only the intended recipient can see the communication
  - integrity (authenticity)
    - the communication is generated by the alleged sender
- What does insecure medium mean?
  - Two basic possibilities:
    - Passive attacker: the adversary can eavesdrop
    - Active attacker: the adversary has full control over the communication channel

Approaches to Secure Communication

- Steganography
  - "covered writing"
  - hides the existence of a message
  - depends on secrecy of method
- Cryptography
  - "hidden writing"
  - hide the meaning of a message
  - depends on secrecy of a short key, not method

Basic Terminology for Encryption

• Plaintext	original message
• Ciphertext	transformed message
• Key	secret used in transformation
• Encryption	
• Decryption	
• Cipher	algorithm for encryption/decryption

Shift Cipher

- The Key Space:
  - [0 .. 25]
- Encryption given a key K:
  - each letter in the plaintext P is replaced with the K'th letter following corresponding number (shift right)
- Decryption given K:
  - shift left

History: K = 3, Caesar's cipher

Shift Cipher: Cryptanalysis

- Can an attacker find K?
  - YES: by a bruteforce attack through exhaustive key search.
    - key space is small (<= 26 possible keys).
  - How much ciphertext is needed?

- Lessons:
  - Key space needs to be large enough.
  - Exhaustive key search can be effective.

Mono-alphabetic Substitution Cipher

- The key space: all permutations of $\Sigma = \{A, B, C, \ldots, Z\}$
- Encryption given a key π:
  - each letter X in the plaintext P is replaced with $\pi(X)$
- Decryption given a key π:
  - each letter Y in the cipherext P is replaced with $\pi^{-1}(Y)$

Example:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
$\pi=$	B	A	D	C	Z	H	W	Y	G	O	Q	X	S	V	T	R	N	M	L	K	J	I	P	F	E	U

BECAUSE $\rightarrow$ AZDBJSZ

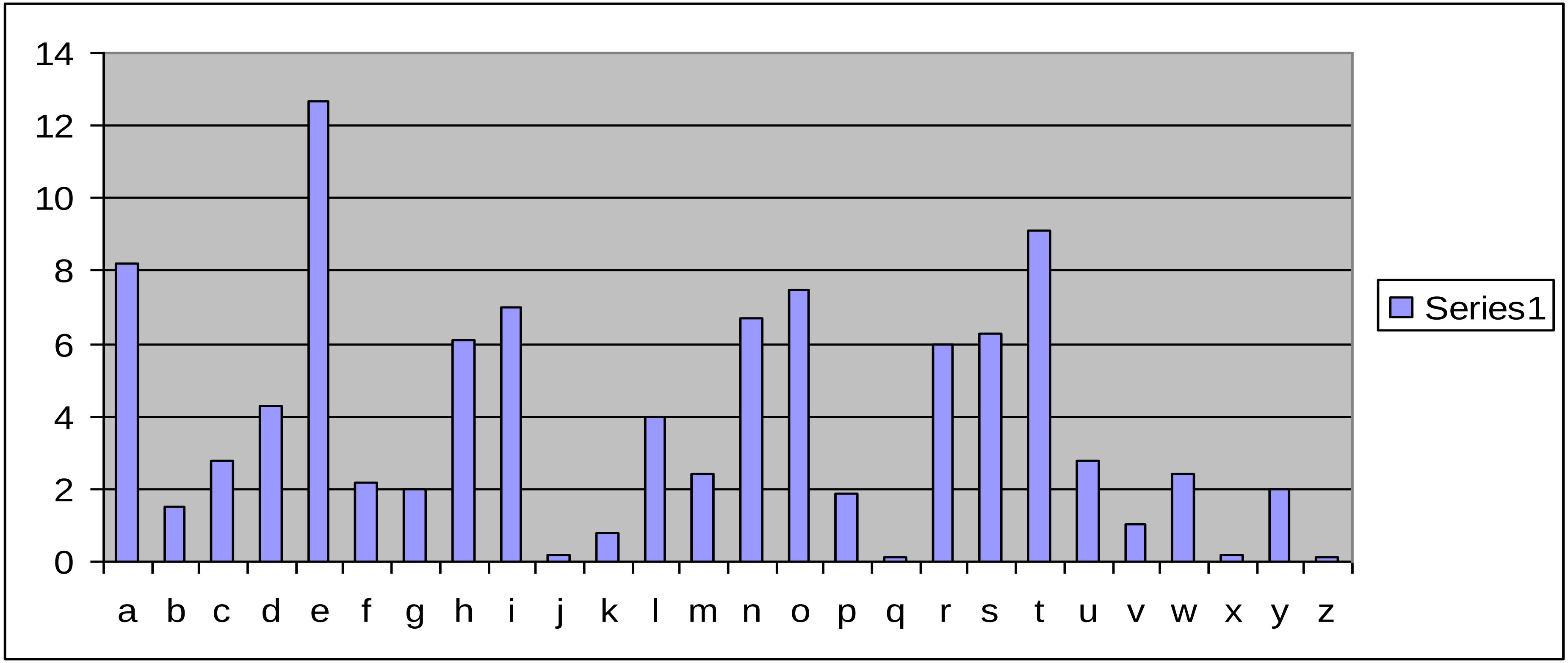
Strength of the Mono-alphabetic Substitution Cipher

- Exhaustive search is difficult
  - key space size is $26! \approx 4 \times 10^{26} \approx 2^{88}$
- Dominates the art of secret writing throughout the first millennium A.D.
- Thought to be unbreakable by many back then
- How to break it?

Cryptanalysis of Substitution Ciphers: Frequency Analysis

- Basic ideas:
  - Each language has certain features: frequency of letters, or of groups of two or more letters.
  - Substitution ciphers preserve the language features.
  - Substitution ciphers are vulnerable to frequency analysis attacks.
  - How much ciphertext is required?

Frequency of Letters in English



How to Defeat Frequency Analysis?

- Use larger blocks as the basis of substitution. Rather than substituting one letter at a time, substitute 64 bits at a time, or 128 bits.
  – Leads to block ciphers such as DES & AES.

- Use different substitutions to get rid of frequency features.
  – Leads to polyalphabetical substituion ciphers, and to stream ciphers such as RC4

Towards the Polyalphabetic Substitution Ciphers

- Main weaknesses of monoalphabetic substitution ciphers
  - In ciphertext, different letters have different frequency
    - each letter in the ciphertext corresponds to only one letter in the plaintext letter
- Idea for a stronger cipher (1460's by Alberti)
  - Use more than one substitutions, and switch between them when encrypting different letters
    - As result, frequencies of letters in ciphertext are similar
- Developed into an easy-to-use cipher by Vigenère (published in 1586)

The Vigenère Cipher

Treat letters as numbers: [A=0, B=1, C=2, …, Z=25]

Number Theory Notation: $Z_n = \{0, 1, \ldots, n-1\}$

Definition:

Given m, a positive integer, $P = C = (Z_{26})^n$, and $K = (k_1, k_2, \ldots, k_m)$ a key, we define:

Encryption:

$e_k(p_1, p_2 \ldots p_m) = (p_1+k_1, p_2+k_2 \ldots p_m+k_m) \pmod{26}$

Decryption:

$d_k(c_1, c_2 \ldots c_m) = (c_1-k_1, c_2-k_2 \ldots c_m-k_m) \pmod{26}$

Example:

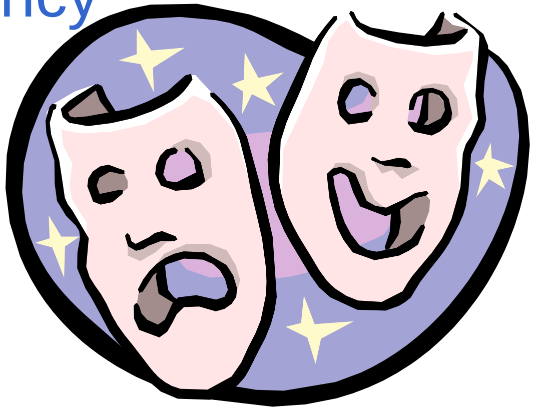
Plaintext: C R Y P T O G R A P H Y

Key: L U C K L U C K L U C K

Ciphertext: N L A Z E I I B L J J I

Security of Vigenere Cipher

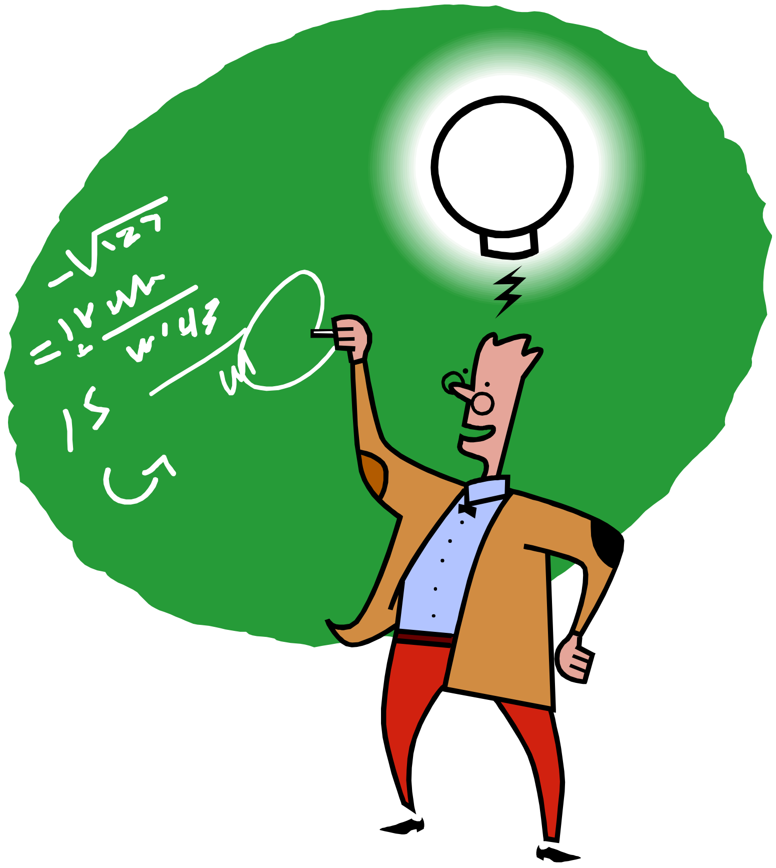
- Vigenere masks the frequency with which a character appears in a language: one letter in the ciphertext corresponds to multiple letters in the plaintext. Makes the use of frequency analysis more difficult.



- Any message encrypted by a Vigenere cipher is a collection of as many shift ciphers as there are letters in the key.

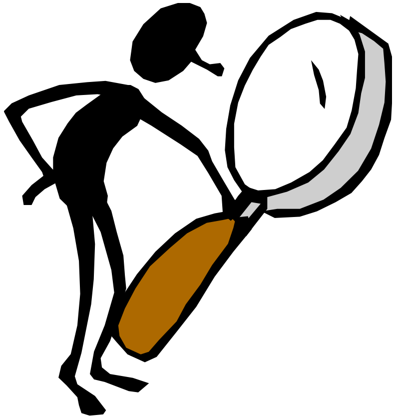
Vigenere Cipher: Cryptanalysis

- Find the length of the key.
  - Kasisky test
  - Index of coincidence (we won't cover here)
- Divide the message into that many shift cipher encryptions.
- Use frequency analysis to solve the resulting shift ciphers.
  - How?



Kasisky Test for Finding Key Length

- Observation: two identical segments of plaintext, will be encrypted to the same ciphertext, if they occur in the text at a distance Δ such that Δ is a multiple of m, the key length.
- Algorithm:
  - Search for pairs of identical segments of length at least 3
  - Record distances between the two segments: $\Delta 1$, $\Delta 2$, …
  - m divides gcd($\Delta 1$, $\Delta 2$, …)



Example of the Kasisky Test

Key	K	I	N	G	K	I	N	G	K	I	N	G	K	I	N	G	K	I	N	G	K	I	N	G
PT	t	h	e	s	u	n	a	n	d	t	h	e	m	a	n	i	n	t	h	e	m	o	o	n
CT	D	P	R	Y	E	V	N	T	N	<u>B</u>	<u>U</u>	<u>K</u>	W	I	A	O	X	<u>B</u>	<u>U</u>	<u>K</u>	W	W	B	T

Repeating patterns (strings of length 3 or more) in ciphertext are likely due to repeating plaintext strings encrypted under repeating key strings; thus the location difference should be multiples of key lengths.

One-Time Pad

- Fix the vulnerability of the Vigenere cipher by using very long keys

- Key is a random string that is at least as long as the plaintext
- Encryption is similar to shift cipher
- Invented by Vernam in the 1920s

One-Time Pad

Let Z_m ={0,1,…,m-1} be the alphabet.

Plaintext space = Ciphtertext space = Key space = $(Z_m)^n$

The key is chosen uniformly randomly

Plaintext $X = (x_1\ x_2\ \ldots\ x_n)$

Key $K = (k_1\ k_2\ \ldots\ k_n)$

Ciphertext $Y = (y_1\ y_2\ \ldots\ y_n)$

$e_k(X) = (x_1+k_1\ \ x_2+k_2\ \ldots\ x_n+k_n) \bmod m$

$d_k(Y) = (y_1-k_1\ \ y_2-k_2\ \ldots\ \ y_n-k_n) \bmod m$

The Binary Version of One-Time Pad

Plaintext space = Ciphtertext space =

Keyspace = $\{0,1\}^n$

Key is chosen randomly

For example:

- Plaintext is 11011011
- Key is 01101001
- Then ciphertext is 10110010

Bit Operators

- Bit AND

  $0 \wedge 0 = 0$ $0 \wedge 1 = 0$ $1 \wedge 0 = 0$ $1 \wedge 1 = 1$

- Bit OR

  $0 \vee 0 = 0$ $0 \vee 1 = 1$ $1 \vee 0 = 1$ $1 \vee 1 = 1$

- Addition mod 2 (also known as Bit XOR)

  $0 \oplus 0 = 0$ $0 \oplus 1 = 1$ $1 \oplus 0 = 1$ $1 \oplus 1 = 0$

- Can we use operators other than Bit XOR for binary version of One-Time Pad?

Key Randomness in One-Time Pad

- One-Time Pad uses a very long key, what if the key is not chosen randomly, instead, texts from, e.g., a book are used as keys.
  - this is not One-Time Pad anymore
  - this can be broken
  - How?
- Corrolary: The key in One-Time Pad should never be reused.
  - If it is reused, it is Two-Time Pad, and is insecure!
  - Why?

Usage of One-Time Pad

- To use one-time pad, one must have keys as long as the messages.
- To send messages totaling certain size, sender and receiver must agree on a shared secret key of that size.
  - typically by sending the key over a secure channel
- Key agreement is difficult to do in practice.
- Can't one use the channel for sending the key to send the messages instead?
- Why is OTP still useful, even though difficult to use?

Usage of One-Time Pad

- The channel for distributing keys may exist at a different time from when one has messages to send.

- The channel for distributing keys may have the property that keys can be leaked, but such leakage will be detected
  - Such as in Quantum cryptography

Adversarial Models for Ciphers

- The language of the plaintext and the nature of the cipher are assumed to be known to the adversary.
- **Ciphertext-only attack**: The adversary knows only a number of ciphertexts.
- **Known-plaintext attack**: The adversary knows some pairs of ciphertext and corresponding plaintext.
- **Chosen-plaintext attack:** The adversary can choose a number of messages and obtain the ciphertexts
- **Chosen-ciphertext attack:** The adversary can choose a number of ciphertexts and obtain the plaintexts.

What kinds of attacks have we considered so far?

When would these attacks be relevant in wireless communications?

The Open Design Security Principle

- **Kerckhoffs's Principle:**
  – A cryptosystem should be secure even if everything about the system, except the key, is public knowledge.
- **Shannon's maxim**:
  – "The enemy knows the system."
- Security by obscurity doesn’t work
- Should assume that the adversary knows the algorithm; the only secret the adversary is assumed to not know is the key
- What is the difference between the algorithm and the key?

Coming Attractions …

- Cryptography: Informational Theoretical Security, Stream Ciphers