COM 5335 Network Security
Lecture 1
Introduction

Scott CH Huang

Outline

- What is Cryptography?
- Steganography
- Cryptography: past, present and future
- Classical Cryptography
  - Shift cipher
  - Simple substitution cipher
  - Poly-alphabetic substitution cipher
- One-time pad

What is Cryptography?

- The American Heritage Dictionary:
  - *"The art or process of writing in or deciphering secret code."*
- Webster:
  - *"The science or study the techniques of secret writing."*
- More generally,
  - *"A science which studies how to provide communication channels with secrecy and/or authenticity properties."*
- Communication channels:
  - Network, hard drive, CD-ROM, etc.

Steganography

Dear George;

Greetings to all at Oxford. Many thanks for your letter and for the summer examination package. All Entry Forms and Fees Forms should be ready for final dispatch to the Syndicate by Friday 20th or at the very latest, I'm told by the 21st. Admin has improved here, thought there's room for improvement still, just give us all two or three more years and we'll really show you! Please don't let these wretched 16+ proposals destroy your basic O and A pattern. Certainly this sort of change, if implemented immediately, would bring chaos.

Sincerely yours;

Steganography (cont'd)

Susan eats truffles. Under pressure, that helps everything before owning Major Bullwinkle.

Steganography (cont'd)

Fishing freshwater bends and saltwater coasts rewards anyone feeling stressed. Resourceful anglers usually find masterful leapers fun and admit swordfish rank overwhelming anyday.

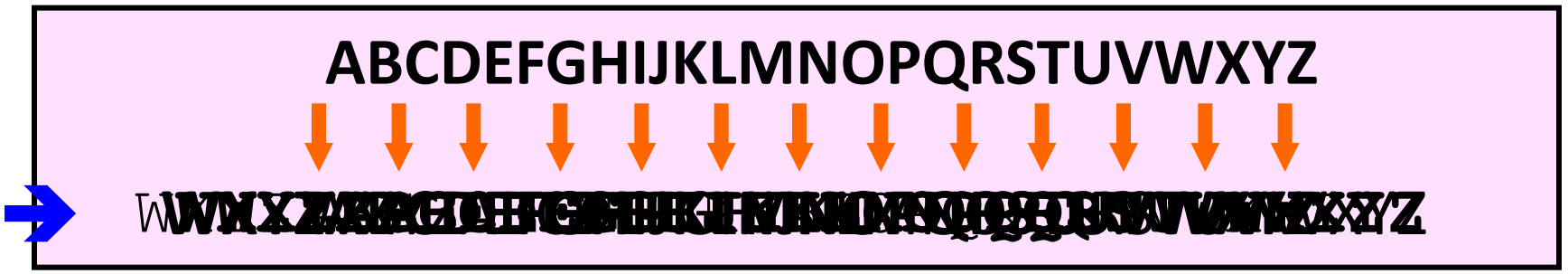
Cryptography: Past, Present, and Future

- Old days – Classical cryptography
  - Essentially secret writing (encryption)
- Today
  - Conventional Cryptography (symmetric key)
  - Public Key cryptography (asymmetric key)
  - Other cryptographic algorithms and schemes
    - proof of knowledge, digital signature, message authentication, secret sharing, etc.
- Future
  - Quantum Cryptography

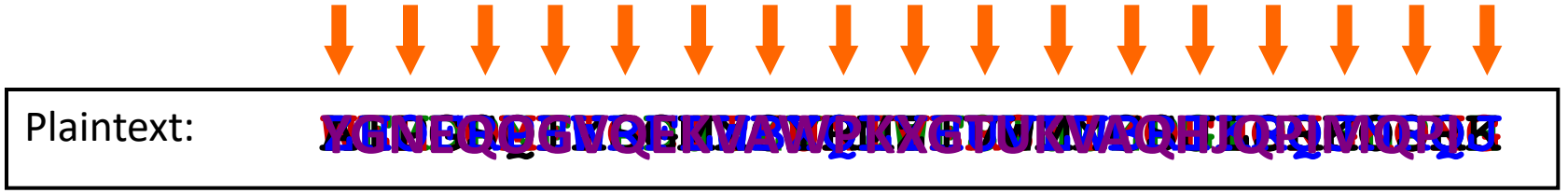
Classical Cryptography: Two Main Techniques

1. Shift
2. Substitution
   - Mono-alphabetic substitution
   - Poly-alphabetic substitution

Shift Cipher



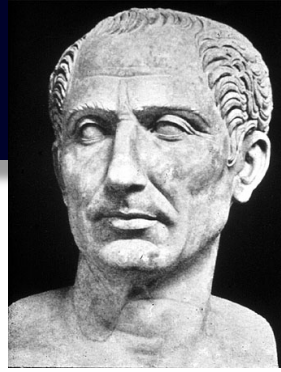
Ciphertext: **AIPGSQIXSGMXCYRMZIVWMXCSJLSRKOSRK**



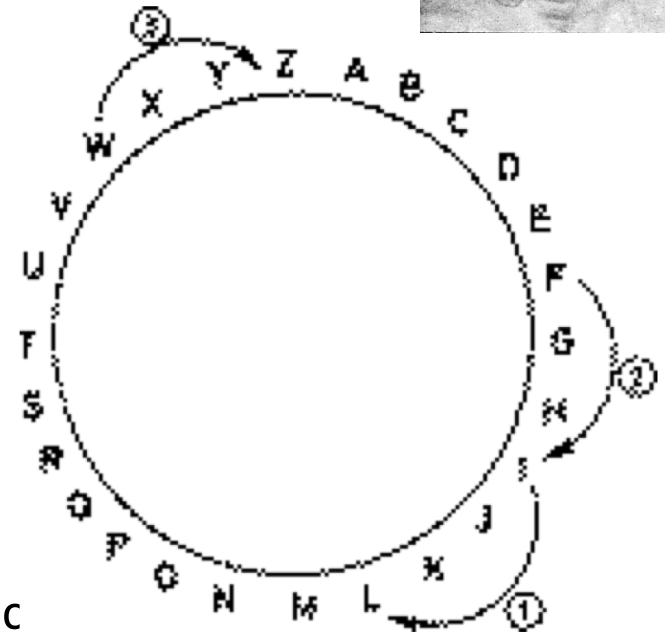
Total number of possible shifts = 26

…can you write a computer program to automate this cracking machine?…

Shift Cipher



- Famous shift cipher: **Caesar Cipher**
  - Shift by 3 letters
  - reputedly used by Julius Caesar (100 – 44 B.C.)
- Plaintext: **I CAME I SAW I CONQUERED**
  Ciphertext: **L FDPH L VDZ L FRQTXHUHG**



- A shift cipher can also be described as

  Encryption $E_K(x) = x + K \bmod 26$

  Decryption $D_K(x) = x - K \bmod 26$

  for English alphabet by setting up a correspondenc between alphabetic characters and residues modulo 26.
  - K is the Key
  - K=3 in Caesar Cipher

Security Strength of a Shift Cipher

- Only have 26 possible shifts
- Brute-force Attack : simply try each possible shift in turn
  - a.k.a. exhaustive key search
- Example:

  Ciphertext – **mjaiamwlxsvitpegipixxivw**

  Trial 1 lizhzlvkwruhsodfhohwwhuv (shift backward by 1)
  Trial 2 khygykujvotgrncegngvvgtu (shift backward by 2)
  Trial 3 jgxfxjtiupsfombdfmfuufst (shift backward by 3)

  Plaintext - **ifwewishtoreplaceletters** (shift backward by 4)

  Hence K=4.

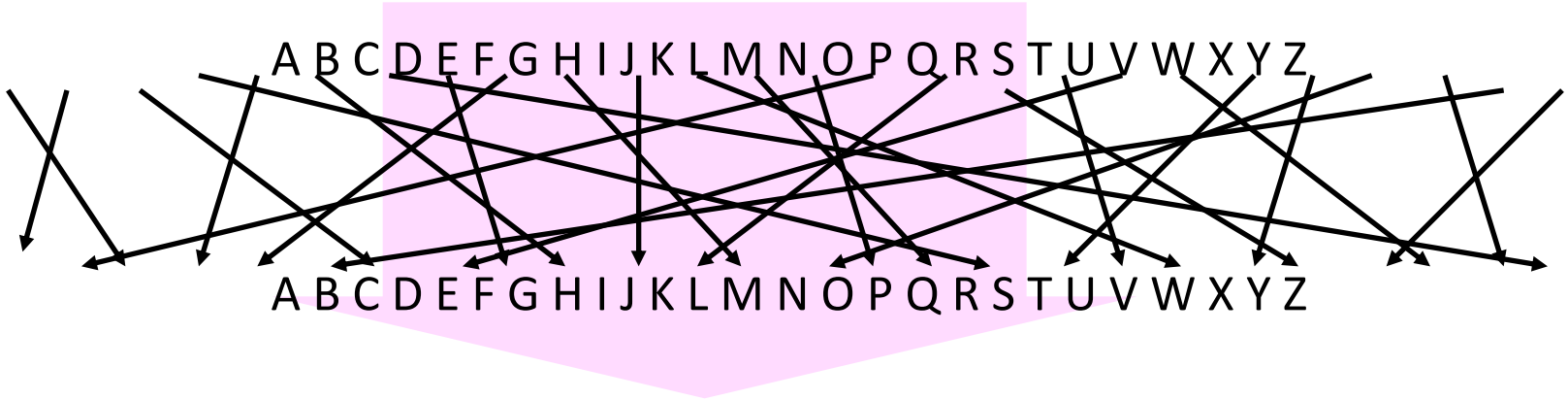
- The major problem of shift ciphers:

  "the key space is too small against brute-force attack"

- The complexity of brute-force attack is O(n).

Simple Substitution

EIMBULJIWLNYANJMVLIURAHIWAI



DEPARTMENTOFCOMPUTERSCIENCE

- A key is a *random permutation* of the alphabetic characters.
- E.g.

a	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*	*j*	*k*	*l*	*m*
X	*N*	*Y*	*A*	*H*	*P*	*O*	*G*	*Z*	*Q*	*W*	*B*	*T*

n	*o*	*p*	*q*	*r*	*s*	*t*	*u*	*v*	*w*	*x*	*y*	*z*
S	*F*	*L*	*R*	*C*	*V*	*M*	*U*	*E*	*K*	*J*	*D*	*I*

- What's the ciphertext of "**solutionsoffinalexam**"?

Simple Substitution

- Total number of possible permutations

$$26!$$

- 26! = 403,291,461,126,605,635,584,000,000 (i.e. 27 digits)

- Maybe… also write a computer program to try them all exhaustively… (so-called **Brute-force Attack**)

- **Calculation**: suppose we have one million 3GHz PCs can try 3 billion permutations per second, the machine will take **4263 years** to try all 26! permutations.

- Question: any better cracking algorithm?

Character Frequency Attack (Statistical Attack)

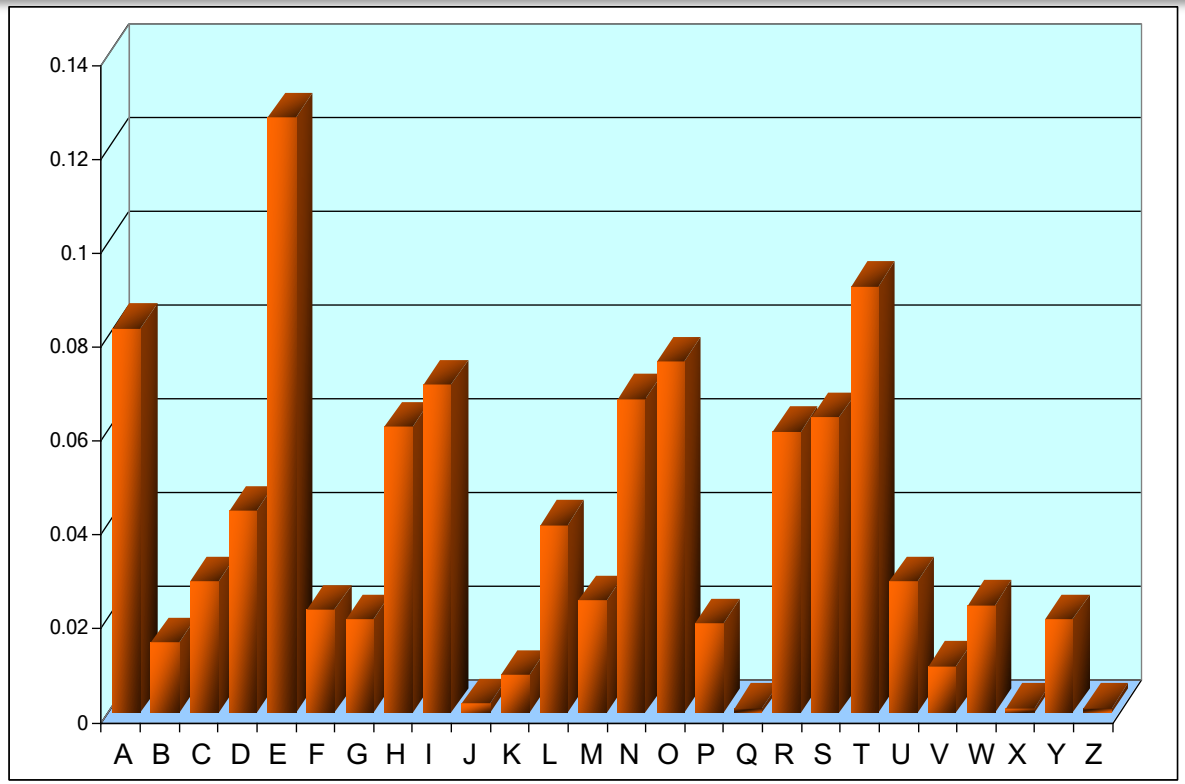
- Brute-force attack against simple substitution becomes less efficient for large alphabet size.
- However, simple substitution does not change relative letter frequencies.

Character Frequency Attack

- in most languages, letters are not equally common
- in English, **e** and **t** are by far the most common letters
- Probability of occurrences of the 26 English letters (obtained by Beker and Piper)

letter	probability	letter	probability
A	.082	N	.067
B	.015	O	.075
C	.028	P	.019
D	.043	Q	.001
E	**.127**	R	.060
F	.022	S	.063
G	.020	**T**	**.091**
H	.061	U	.028
I	.070	V	.010
J	.002	W	.023
K	.008	X	.001
L	.040	Y	.020
M	.024	Z	.001

Character Frequency Attack (Statistical Attack)



- May also be useful to consider sequences of two or three consecutive letters called **digrams** and **trigrams**, respectively.
- e.g. common diagrams (in decreasing order): TH, HE, IN, ER, AN, RE, ED, ON, ES, ST, EN, AT, TO, NT, HA, ND, OU, EA, NG, AS, OR, …
- e.g. common trigrams (in decreasing order): THE, ING, AND, HER, ERE, ENT, THA, NTH, WAS, …

Cryptanalysis of The Substitution Cipher

- May also be useful to consider sequences of two or three consecutive letters called **digrams** and **trigrams**, respectively.
- e.g. common diagrams (in decreasing order): TH, HE, IN, ER, AN, RE, ED, ON, ES, ST, EN, AT, TO, NT, HA, ND, OU, EA, NG, AS, OR, …
- e.g. common trigrams (in decreasing order): THE, ING, AND, HER, ERE, ENT, THA, NTH, WAS, …

Exercise: Ciphertext obtained from a substitution cipher

```
YIFQFMZRWQFYVECFMDZPCVMRZWNMDZVEJBTXCDDUMJ
NDIFEFMDZCDMQZKCEYFCJMYRNCWJCSZREXCHZUNMXZ
NZUCDRJXYYSMRTMEYIFZWDYVZVYFZUMRZCRWNZDZJJ
XZWGCHSMRNMDHNCMFQCHZJMXJZWIEJYUCFWDJNZDIR
```

What's the message?

Polyalphabetic Substitution Ciphers

- In both the Shift Cipher and the Substitution Cipher, once a key is chosen, each alphabetic character is mapped to a unique alphabetic character – monoalphabetic.
  - Small key space (O(n) for Shift Cipher)
  - Vulnerable to statistical attacks (Substitution Cipher).
- Alternative: **Polyalphabetic Substitution Ciphers**
  - We want large key space and less vulnerable to statistical attacks
- Approach: for different location or different time, same letter is to be substituted by different letter.

Vigenère Cipher

- Invented in the 16th century
- A kind of poly-alphabetic substitution cipher
- Using the correspondence A ↔ 0, B ↔ 1, …, Z ↔ 25.
- Keyword: **CIPHER** ↔ (2, 8, 15, 7, 4, 17)
- Plaintext: **thiscryptosystemisnotsecure**

19	7	8	18	2	17	24	15	19	14	18	24
2	8	15	7	4	17	2	8	15	7	4	17
21	15	23	25	6	8	0	23	8	21	22	15
18	19	4	12	8	18	13	14	19	18	4	2
2	8	15	7	4	17	2	8	15	7	4	17
20	1	19	19	12	9	15	22	8	25	8	19
				20	17	4					
				2	8	15					
				22	25	19					

- Ciphertext: **VPXZGIAXIVWPUBTTMJPWIZITWZT**

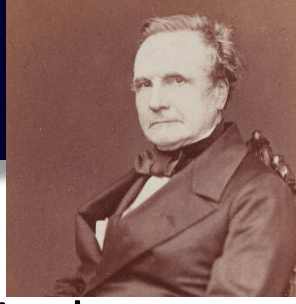
Vigenère Cipher and Enigma

- Number of possible keywords of length m = $\mathbf{26^m}$.
- Much larger than that of a simple substitution cipher.
- An alphabetic character of a plaintext can be mapped to one of m possible alphabetic characters (assuming that the keyword contains m distinct characters).
- In general, cryptanalysis is much more difficult for polyalphabetic than for monoalphabetic cryptosystems.

Rotor Machines

- In the 1920s, various mechanical encryption devcies were invented to automate the process of encryption.
- A rotor machine has a keyboard and a series of rotors, and implements a version of the Vigenère cipher.
- The best-known rotor device is the Enigma, used by the Germans during World War II.

Kasiski Test (1863)



- Idea: any two identical strings will be encrypted to the same ciphertext if they are km positions apart where m is the keyword length and k is a positive integer.
- Two Steps of Attack:
  - Find the keyword length
  - Conduct statistical attack
- An Example of Attack
  - Look for trigrams that are identical
  - Compute the distance between them, d_1, d_2, …
  - Let m' be a divisor of gcd(d_1, d_2, …)
  - Write the ciphertext in a rectangular array with m' columns, then statistical attack can be used on each column

One-Time Pad

- Polyalphabetic substitution with the keyword as long as the plaintext and the keyword has no statistical relationship to the plaintext.

Vernam Cipher (1918)

- Works on binary string rather than letters
- $C_i = P_i \oplus K_i$, where K_i is chosen randomly
- The only cryptographic system that can be proved to be unconditionally secure
- Can be shown to be information theoretically secure.

Weakness of One-time Pad

- **Malleable**: Provides secrecy but not authentication.
- **Keys must NOT be reused**:
  - cannot withstand Known Plaintext Attack
  - depending on known information about plaintexts, Eve can make use of

    $C_1 \oplus C_2 = (M_1 \oplus K) \oplus (M_2 \oplus K)$

    $= M_1 \oplus M_2$

    to figure out both messages
- Generating random bits
  - radioactive decay
  - noisy diode
  - flipping coins