CS915/435 Advanced Computer Security - Elementary Cryptography

Classical Cryptography

Quote of the day

"Any apparently contradictory set of requirements can be met using right mathematical approach."

— Ronald L. Rivest

At ACM Turing Award lecture 2007

What’s cryptography?

- The **art** and **science** of keeping information secure.

  - Bruce Schneier

- Computer security and crypto communities have drifted apart over the past 20 years
  - The former don’t always understand the available crypto tools
  - The latter don’t always understand the real-world problems
  - A security engineer must be able to be familiar with both

Basic terminology

- **Cipher**: a cryptographic algorithm to do encryption and decryption.
- **Key**: used for encryption and decryption
- **Keyspace**: the range of the key

- The secrecy must reside entirely on the key, not the cipher algorithm.

-- Kerckhoff’s principle

Cryptanalysis

- The art and science of analyzing weaknesses of cipher algorithms
- Also known as **attack**.
- **Cryptology** = **Cryptography** + **Cryptanalysis**

Four general types of attack

1. **Ciphertext-only** attack
2. **Known-plaintext** attack - In WWII, German ciphertext started with a date
3. **Chosen-plaintext** attack – Breaking secret code [AF at Midway]
4. **Chosen-ciphertext** attack – The job is to deduce the key (lunch time attack)

Dramatis Personae

• **Alice**	First participant in a crypto system
• **Bob**	Second participant
• **Carol**	Third participant
• **Eve**	Eavesdropper
• **Mallory**	Malicious active attacker

Large Numbers (storage)

No of atoms in the planet	2^{170}
No of atoms in the sun	2^{190}
No of atoms in the galaxy	2^{223}
No of atoms in the observable universe	2^{265}

- To store all 256-bit Keys **2^{264} bits**

Large Numbers (time)

- Time until the next ice age 2^{14} years
- Time until the sun dies 2^{30} years
- Age of the planet 2^{30} years
- Age of the observable universe 2^{34} years

- To brute-force a 256-bit key **2^{192} years**

(Assume guessing one billion keys in one ms)

Dyson, "Time Without End: Physics and Biology in an Open Universe," Reviews of Modern Physics, v. 52, n. 3, July 1979, pp. 447–460.

Roadmap

- Symmetric cryptography
  - Classical cryptography
  - Stream cipher
  - Block cipher I, II
  - Hash
  - MAC
- Asymmetric cryptography
  - Public key encryption
  - Digital signature
  - Key agreement

Some of the content is based on adapting and expanding Stanford Crypto Course
https://crypto.stanford.edu/~dabo/courses/OnlineCrypto/

Classical cryptography

- Classical cryptography
  - Based on characters (human)
- Modern cryptography
  - Based on binary inputs (computer)
- What has changed?
  - 26 elements to 2 elements.
  - But, the philosophy remains basically the same.

Confusion and Diffusion

- Two basic principles to obscure redundancies in a plaintext message (Shannon, 1949)
- **Confusion**:
  - Obscures the relationship between the plaintext and the ciphertext (e.g., by substituting letters)
- **Diffusion**:
  - Dissipates the redundancy of the plaintext by spreading it out over the ciphertext (e.g., by transposing plaintext).

Substitution ciphers

1. Monoalphabetic cipher
   – E.g. a->b, b->c
2. Polyalphabetic cipher
   – Made up of several monoalphabetic ciphers

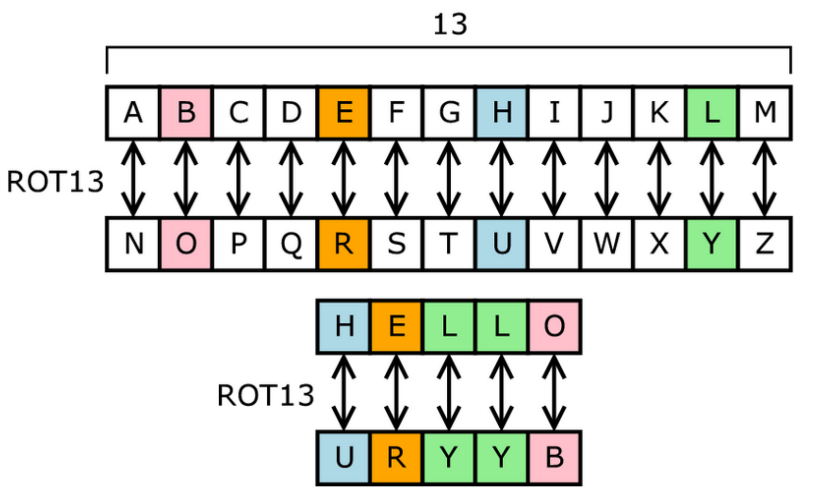
Example: Casesar cipher

- Named after Julius Caesar
- Used for hundreds of years
- Shifts every letter 3 positions to the right

- Example
  - attackatdawm -> dwwdfndwgdzp

ROT13

- Another example of monoalphabetic cipher
- Commonly found on UNIX systems
- Every letter is rotated by 13 positions
- Question
  - Why not ROT14?



(Source: Wikipedia.org)

Review of modular arithmetic

- Suppose a and b are integers, and m is a positive integer.
- We write a $\equiv b \ (mod\ m)$ if m divides b-a.
- The phrase a $\equiv b \ (mod\ m)$ is called a *congruence*.
- It is read as "a is congruent to b modulo m"
- The integer m is called the *modulus*.

Data range

- Some programming languages define $a \ mod \ m$ in the range of (-m, m). But for our purpose, we define it always to be non-negative.

- $101 \ mod \ 7 =$ $\boxed{}$

101 = 7x14 + 3
= 3 mod 7

- $-101 \ mod \ 7 =$ $\boxed{}$

-101 = -7x14 - 3
= -3 mod 7
= 4 mod 7

Shift cipher (formal definition)

- Arithmetic modulo Z_m is the set {0,…,m-1}

Let $P = C = K = Z_{26}$. For $0 \leq K \leq 25$, define

$$e_K(x) = (x + K)\ mod\ 26$$

And

$$d_K(y) = (y - K)\ mod\ 26$$

$(x, y \in Z_{26})$

Examples:

- K = 3 → Caesar cipher
- K = 13 → ROT13

Cryptanalysis of shift cipher

- Given JBCRCLQRWCRVNBJENBWRWN, can you find out the plaintext?

jbcrclqrwcrvnbjenbwrwn	(K=0)
iabqbkpqvbqumaidmavqvm	(K=1)
hzapajopuaptlzhclzupul	(K=2)
gyzozinotzoskygbkytotk	(K=3)
fxynyhmnsynrjxfajxsnsj	(K=4)
ewxmxglmrxmqiweziwrmri	(K=5)
dvwlwfklqwlphvdyhvqlqh	(K=6)
cuvkvejkpvkogucxgupkpg	(K=7)
btujudijoujnftbwftojof	(K=8)
astitchintimesavesnine	(K=9)

What went wrong?

- The shift cipher (modulo 26) is not secure, because it can be broken by ***exhaustive search***
- Only 26 possible keys
- On average, a plaintext can be computed after just 26/2=13 tries.
- ***Lesson***: for a cipher to be secure, the key space must be very large
- But, is the reverse true?

Substitution cipher (definition)

Let $P = C = Z_{26}$. K consists of all possible permutations of the 26 symbols. For each permutation $\pi \in K$, define

$$e_\pi(x) = \pi(x)\ mod\ 26$$

And

$$d_\pi(y) = \pi^{-1}(y)\ mod\ 26$$

($x, y \in Z_{26}$, and π^{-1} is the inverse permutation to π)

Example:

- $\pi = \{2,4,5,0, \dots, 7,16\}$
- π(0)=2, hence $a \rightarrow c$

Key space of substitution cipher

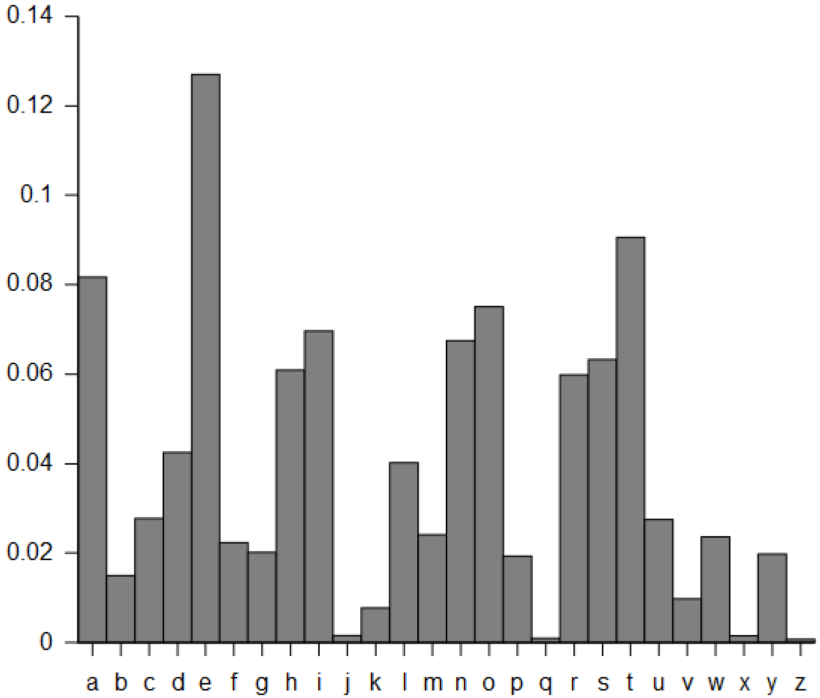
What's the key space?

a) $| K | = 26$
b) $| K | = 26!$ (26 factorial) $\approx 2^{88}$
c) $| K | = 2^{26}$
d) $| K | = 26^2$

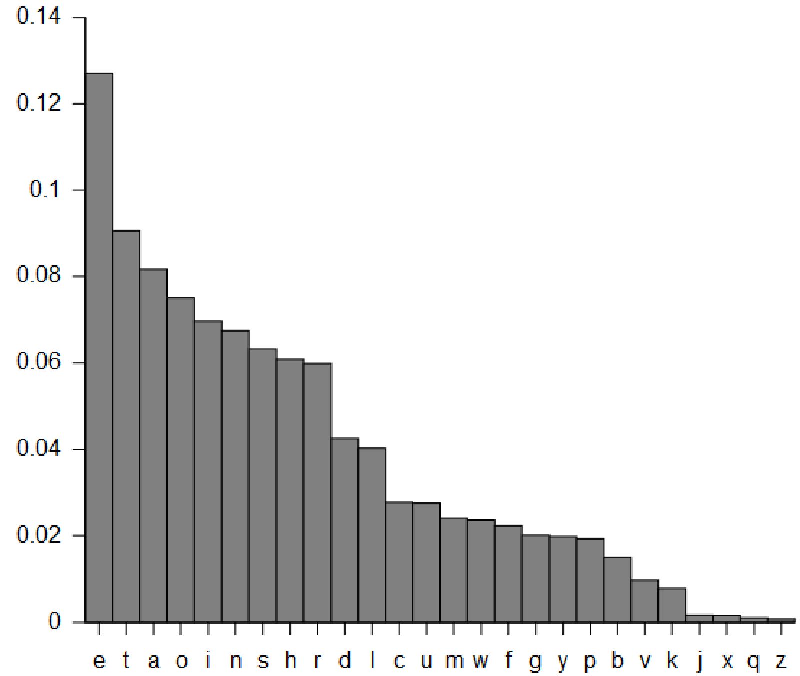
Cryptanalysis of substitution cipher

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

Cryptanalysis of substitution cipher



(a) Relative frequencies of English letters



(b) Relative frequencies sorted by frequency

(Source: wikipedia)

What went wrong?

- A large key space is not sufficient to ensure the cipher is secure.
- Substitution only provides confusion.
- ***Lesson***: a secure cipher should combine both confusion and diffusion.

Vigenère cipher

- A polyalphabetic cipher based on the idea of combining a few Caesar ciphers into one
- Named after Blaise De Vigenère, a French diplomat in 1586

k = A B C A B C A B C A B C A B C A

m = B E R E A D Y A C K A T D A W N (+ mod 26)

---

c = B F T E B F Y B E K B V D B Y N

Cryptanalysis of Vigenère cipher

- Two steps in the cryptanalysis
  1. Find out the key length m
  2. Find out each letter in the key

How to find out the key length?

- First method: Kasiski test
  - Described by Friedrich Kasiski in 1863
  - Search for identical segments and count how many positions they are apart

ABCDEABCDE … ABCDE … ABCDE

the……….the…….the…………………..

15 positions apart -> key length is either 3 or 5 or 15

Example: Vigenère cipher

0

CHREEVOAHMAERATBIAXXWTNXBEEOPHBSBQMQEQERBW
RVXUOAKXAOSXXWEAHBWGJMMQMNKGRFVGXWTRZXWIAK
LXFPSKAUTEMNDCMGTSXMXBTUIADNGMGPSRELXNJELX
VRVPPTULHDNQWTWDTYGBPHXTFALJHASVBFXNGLLCHR 165
ZBWELEKMSJIKNBHWRJGNMGJSGLXFEYPHAGNRBIEQJT
235
AMRVLCRREMNDGLXRRIMGNSNRWCHRQHAEYEVTAQEBBI
PEEWEVKAKOEWADREMXMTBHHCHRTKDNVRZCHRCLQOHP
275 285
WQAIIWXNRMGWOIIFKEE

How to find out the key length?

- Second method: index of coincidence
  - Described by William Friedman in 1920
  - Suppose $\boldsymbol{x} = x_1 x_2 \ldots x_n$ is a string of n alphabetic characters.
  - The ***Index of coincidence*** of $\boldsymbol{x}$ is defined to be the probability that two random elements of $\boldsymbol{x}$ are identical.

Index of coincidence

- Suppose a string of n English letters
- Occurrence of A = f_0 $(f_0/n) * (f_0-1/n-1)$
- Occurrence of B = f_1 $(f_1/n) * (f_1-1/n-1)$
- … …
- Occurrence of Z = f_{25}
- Hence, index of coincidence is calculated:

$$I_c(x) = \frac{\sum_{i=0}^{25} \binom{f_i}{2}}{\binom{n}{2}} = \frac{\sum_{i=0}^{25} f_i \, (f_i \cancel{- 1})}{n(n \cancel{- 1})} \approx \sum_{i=0}^{25} {p_i}^2 \quad (f_i/n)$$

Difference in index of coincidence

- Normal English text

  Index of coincidence: $\sum_{i=0}^{25} {p_i}^2 = 0.065$
- Completely random string of letters

  Index of coincidence: $\sum_{i=0}^{25} {p_i}^2 = 0.038$ 1/26
- Normal English text shifted by a fixed number

  Index of coincidence: $\sum_{i=0}^{25} {p_i}^2 = \boxed{0.065}$
- English text encrypted by Vigenère cipher

  Index of coincidence: $\sum_{i=0}^{25} {p_i}^2 = \boxed{}$

Same Example as before

CHREEVOAHMAERATBIAXXWTNXBEEOPHBSBQMQEQERBW
RVXUOAKXAOSXXWEAHBWGJMMQMNKGRFVGXWTRZXWIAK
LXFPSKAUTEMNDCMGTSXMXBTUIADNGMGPSRELXNJELX
VRVPPTULHDNQWTWDTYGBPHXTFALJHASVBFXNGLLCHR
ZBWELEKMSJIKNBHWRJGNMGJSGLXFEYPHAGNRBIEQJT
AMRVLCRREMNDGLXRRIMGNSNRWCHRQHAEYEVTAQEBBI
PEEWEVKAKOEWADREMXMTBHHCHRTKDNVRZCHRCLQOHP
WQAIIWXNRMGWOIIFKEE

Feed the text into a matrix [n, m] row by row where m is the guessed key length

index of coincidence

m	index of coincidence of each column
1	0.045
2	0.046, 0.041
3	0.043, 0.050, 0.047
4	0.042, 0.039, 0.045, 0.040
5	0.063, 0.068, 0.069, 0.061, 0.072

Next step: break each shift cipher

C H R E E
V O A H M
A E R A T
B I A X X
W T N X B
E E O P H
B S B Q M
…