MH4311 Cryptography

Lecture 3

One–Time Pad & Information Theory

Wu Hongjun

Lecture Outline

- Classical ciphers
- Symmetric key encryption
  - One-time pad & information theory
  - Block cipher
  - Stream cipher
- Hash function and Message Authentication Code
- Public key encryption
- Digital signature
- Key establishment and management
- Introduction to other cryptographic topics

Recommended Reading

- **CTP Section 2.1, 2.2, 2.3, 2.4, 2.6**
- **HAC Section 2.1.1, 2.1.2, 2.1.3, 2.2**

- **Wikipedia:**
  - **One-time pad**
    - **http://en.wikipedia.org/wiki/One-time_pad**
  - **Information theory**
    - **http://en.wikipedia.org/wiki/Information_theory**

Weakness of Vigenere cipher

- **Key is expanded (repeated) so as to encrypt a long message using shift ciphers**
  - **Similar scheme being used in Microsoft Word 95**
    - **Key word is XORed with the plaintext**
- **Attack Vigenere cipher**
  - **Find key length:**
    - **Kasiski test or**
    - **Index of Coincidence**
  - **Then use frequency analysis to break each shift cipher**

One-Time Pad (OTP)

- **To strengthen Vigenere cipher, we can use the secure One-Time Pad:**
  - **Key generation**
    - **1) truly random key**
    - **2) key is as long as the message**
  - **Encryption**
    - **3) each key is used to encrypt only one message (using shift ciphers)**

⇒The resulting cipher is unconditionally secure (it achieves perfect secrecy), i.e., unbreakable to attackers with unlimited computing resource

One-Time Pad

- **Example**

```
Plaintext:  nanyang
Key:        XRTRPLK
Ciphertext: K
```

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
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

```
n->13  X->23 (13+23)mod 26 = 10->K
```

One-Time Pad

- Why do we call it One-Time Pad ?
  - One-time: each key is used to encrypt only one message
  - Pad: in early implementations, key material distributed as a pad of paper
    - top sheet can be easily torn off and destroyed after use

- OTP is also called Vernam cipher
  - Invented by Gilbert Vernam (an AT&T engineer) in 1917

One-Time Pad

- **Modern one-time pad deals with bit sequence**
- **Bit**
  - **A bit has value either 1 or 0**
  - **It is the most basic information unit**
    - **All the information on computer (such as operating system, application software, music, video, pdf files, word document files …) are stored/processed as a sequence of bits**
  - **One byte consists of 8 bits**
    - **Byte is the commonly used information unit on computer**

One-Time Pad

- **Modern one-time pad deals with bit sequence**
  - **Instead of using "addition mod 26"**
  - **"addition mod 2" is used for bit sequence**
    - **"addition mod 2", also called XOR (exclusive OR)**
      - **"(a + b) mod 2" is denoted as "a XOR b", "$a \oplus b$"**
      - **In C programming language, "a XOR b" is encoded as "a ^ b"**
- **Example:**

```
Plaintext:    1010011000
Key:        ⊕ 0110101110
Ciphertext: =
```

One-time Pad

- **The key must be randomly generated**
  - **In 1944–1945, the U.S. Army's Signals Intelligence Service was able to solve a one-time pad system used by the German Foreign Office for its high-level traffic, since the pads were not completely random — the machine used to generate the pads produced predictable output**

One-time Pad

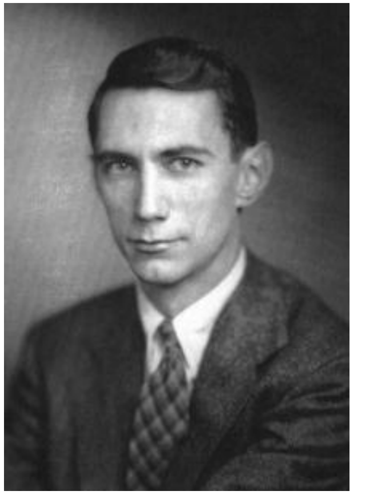
- **Mainly limited to diplomacy and intelligence applications in history**
  - **Used by British Special Operations Executive in World War II**
  - **Used by spies in the Cold War**
  - **Used to protect the hotline between Moscow and Washington D.C. after the Cuban missile crisis**

One-time Pad

- **Advantage**
  - easy to encrypt/decrypt
  - perfect security
- **Disadvantage**
  - The key should not be reused
    - Due to key distribution mistake, the embassies of Soviet Union used the key of one-time pad more than once in WWII
  - Large key size for long message

One-Time Pad

- **How to prove that one-time pad achieves perfect security?**
  - **One-time pad was believed to be secure**
  - **Its perfect secrecy was proven by Shannon (1948)**



Claude Shannon

(1912-2001)

"the father of information theory and cryptography"

Elementary Probability Theory

- **Random variable**
  - **A discrete random variable X takes certain values with certain probabilities**
  - **The probability that the discrete random variable X takes on a particular value *x* is denoted as Pr[X = *x*]**
  - **Let *X* denote the set of all the possible values of *x*, it must be true that**

$$\sum_{x \in X} \mathbf{Pr}(\mathbf{X} = x) = 1$$

- **Example: Coin Toss**
  - **The random variable X is the result of coin toss: head or tail**
  - **The set of all the possible values of X: *X* = {tail, head}**
  - **Pr[X = tail] = Pr[X = head] = ½ (for a fair coin toss)**

Elementary Probability Theory

- **Random variable example 2: English text**
  - **Let X be the random variable representing letters in English text**
  - **The set of all the possible values of X:**

    $$X = \{a, b, c, d, \ldots, z\}$$

  - **Pr[X = a] = 0.082, Pr[X = b] = 0.015, …**

    Pr[X = z] = 0.01

Elementary Probability Theory

- Join Probability
  - **X** and **Y** are two random variables
  - The join probability $\mathbf{Pr}[\mathbf{X}=x, \mathbf{Y}=y]$ is the probability that **X** takes the value x and **Y** takes the value y
- **X** and **Y** are independent if

$$\mathbf{Pr}[\mathbf{X}=x, \mathbf{Y}=y] = \mathbf{Pr}[\mathbf{X}=x]\cdot\mathbf{Pr}[\mathbf{Y}=y]$$

for all values of x and y

Elementary Probability Theory

- Conditional Probability
  - **X** and **Y** are two random variables
  - The conditional probability $\mathbf{Pr}[\mathbf{X}=x \mid \mathbf{Y}=y]$ is the probability that **X** takes the value *x* given that **Y** takes the value *y*
- Joint Probability and conditional probability are related:

$$\mathbf{Pr}[\mathbf{X}=x, \mathbf{Y}=y] = \mathbf{Pr}[\mathbf{X}=x \mid \mathbf{Y}=y]\cdot\mathbf{Pr}[\mathbf{Y}=y]$$

Elementary Probability Theory

$$\mathbf{Pr}[\mathbf{X}=x, \mathbf{Y}=y] = \mathbf{Pr}[\mathbf{X}=x \mid \mathbf{Y}=y]\cdot\mathbf{Pr}[\mathbf{Y}=y] \quad (1)$$

$$\mathbf{Pr}[\mathbf{Y}=y, \mathbf{X}=x] = \mathbf{Pr}[\mathbf{Y}=y \mid \mathbf{X}=x]\cdot\mathbf{Pr}[\mathbf{X}=x] \quad (2)$$

$\mathbf{Pr}[\mathbf{Y}=y, \mathbf{X}=x]$ is the same as $\mathbf{Pr}[\mathbf{X}=x, \mathbf{Y}=y]$ (3)

From (1), (2), (3),

$$\mathbf{Pr}[\mathbf{X}=x \mid \mathbf{Y}=y]\cdot\mathbf{Pr}[\mathbf{Y}=y] = \mathbf{Pr}[\mathbf{Y}=y \mid \mathbf{X}=x]\cdot\mathbf{Pr}[\mathbf{X}=x]$$

$$\mathbf{Pr}[\mathbf{X}=x \mid \mathbf{Y}=y] = \frac{\mathbf{Pr}[\mathbf{Y}=y \mid \mathbf{X}=x]\cdot\mathbf{Pr}[\mathbf{X}=x]}{\mathbf{Pr}[\mathbf{Y}=y]}$$

Bayes' Theorem

Elementary Probability Theory

Bayes' theorem example: Dice Throwing

- A pair of dice are randomly thrown
- **X** is a random variable defined as the sum of two dice
  - The set of all the possible values of **X** is $X = \{2,3,4,\ldots,12\}$
- **Y** is a random variable
  - **Y** = *d* if the two dice are the same (throw "doubles")
  - **Y** = *n* if the two dice are not the same

Elementary Probability Theory

Bayes' theorem example: Dice Throwing (cont.)

- Now we perform the following computation to test Bayes' theorem

$$\mathbf{Pr}[\mathbf{X} = 4] = \mathbf{Pr}[\text{1st dice} = 1] \cdot \mathbf{Pr}[\text{2nd dice} = 3] + \mathbf{Pr}[\text{1st dice} = 2] \cdot \mathbf{Pr}[\text{2nd dice} = 2] + \mathbf{Pr}[\text{1st dice} = 3] \cdot \mathbf{Pr}[\text{2nd dice} = 1] = 1/6 \times 1/6 + 1/6 \times 1/6 + 1/6 \times 1/6 = 1/12$$

$$\mathbf{Pr}[\mathbf{Y} = \mathrm{d} \mid \mathbf{X} = 4] = \frac{\mathbf{Pr}[\text{1stdice}=2] \bullet \mathbf{Pr}[\text{2nddice}=2]}{\mathbf{Pr}[\text{1stdice}=1] \bullet \mathbf{Pr}[\text{2nddice}=3] + \mathbf{Pr}[\text{1stdice}=2] \bullet \mathbf{Pr}[\text{2nddice}=2] + \mathbf{Pr}[\text{1stdice}=3] \bullet \mathbf{Pr}[\text{2nddice}=1]}$$

$$= \frac{1/6 \times 1/6}{1/6 \times 1/6 + 1/6 \times 1/6 + 1/6 \times 1/6} = 1/3$$

$$\mathbf{Pr}[\mathbf{Y} = d] = \mathbf{Pr}[\text{1st dice} = 1] \cdot \mathbf{Pr}[\text{2nd dice} = 1] + \ldots\ldots + \mathbf{Pr}[\text{1st dice} = 6] \cdot \mathbf{Pr}[\text{2nd dice} = 6] = (1/6 \times 1/6) \times 6 = 1/6$$

$$\mathbf{Pr}[\mathbf{X} = 4 \mid \mathbf{Y} = d] = \frac{\mathbf{Pr}[\text{1stdice}=2] \bullet \mathbf{Pr}[\text{2nddice}=2]}{\mathbf{Pr}[\text{1stdice}=1] \bullet \mathbf{Pr}[\text{2nddice}=1] + \mathbf{Pr}[\text{1stdice}=2] \bullet \mathbf{Pr}[\text{2nddice}=2] + \cdots + \mathbf{Pr}[\text{1stdice}=6] \bullet \mathbf{Pr}[\text{2nddice}=6]}$$

$$= \frac{1/6 \times 1/6}{(1/6 \times 1/6) \times 6} = 1/6$$

$$\text{-->}\ \mathbf{Pr}[\mathbf{Y}=\mathrm{d} \mid \mathbf{X}=4] \cdot \mathbf{Pr}[\mathbf{X}=4] = \mathbf{Pr}[\mathbf{X}=4 \mid \mathbf{Y}=\mathrm{d}] \cdot \mathbf{Pr}[\mathbf{Y}=\mathrm{d}] = 1/36$$

Perfect Secrecy of OTP

- **A cryptosystem has perfect secrecy if knowing ciphertext reveals no information about the plaintext**
- **Definition:**
  - **A cryptosystem has perfect secrecy if for every plaintext *p* and every ciphertext *c*,**

$$\Pr[P = p \mid C = c] = \Pr[P = p]$$

    - **$\Pr[P = p \mid C = c]$ is *a posteriori* probability that the plaintext is *p*, given that the ciphertext *c* is observed.**
    - **$\Pr[P = p]$ is *a priori* probability that the plaintext is *p***
  - **<u>an attacker cannot correctly guess the plaintext with higher probability after knowing the ciphertext</u>**

Perfect Secrecy of OTP

- One-time pad
  - $\mathbf{P} = \mathbf{C} = \mathbf{K} = \{0,1\}^n$ (n-bit sequence)
  - Key is chosen randomly
    - $\mathbf{Pr}(\mathbf{K} = k) = 1/2^n$
  - Show that $\mathbf{Pr}[\mathbf{P} = p \mid \mathbf{C} = c] = \mathbf{Pr}[\mathbf{P} = p]$ (perfect secrecy)
- Proof.
  - $\mathbf{Pr}[\mathbf{C} = c \mid \mathbf{P} = p] = \mathbf{Pr}[\mathbf{K} = p \oplus c] = 1/2^n$
  - $\mathbf{Pr}[\mathbf{C} = c] = \sum_{p \in P} (\mathbf{Pr}[\mathbf{P} = p] \cdot \mathbf{Pr}[\mathbf{C} = c \mid \mathbf{P} = p])$
    $= \sum_{p \in P} (\mathbf{Pr}[\mathbf{P} = p]) \times 1/2^n = 1/2^n$
  - Using Bayes' theorem:

$$\mathbf{Pr}[\mathbf{P} = p \mid \mathbf{C} = c] = \mathbf{Pr}[\mathbf{P} = p] \cdot \mathbf{Pr}[\mathbf{C} = c \mid \mathbf{P} = p] \,/\, \mathbf{Pr}[\mathbf{C} = c]$$
$$= \mathbf{Pr}[\mathbf{P} = p] \cdot (1/2^n) \,/\, (1/2^n)$$
$$= \mathbf{Pr}[\mathbf{P} = p]$$

Entropy

- **One-time pad**
  - Key length = message length
  - Perfect secrecy
- **How about the security of the following cipher?**
  - key length < message length,
  - the attacker has unlimited computing resource
- **The concept "Entropy" is needed to answer the above question**

Entropy

- Entropy in information theory
  - Claude Shannon's information theory
    - "A Mathematical Theory of Communication", 1948
  - a measure of the uncertainty associated with a random variable
- Definition

Suppose that **X** is a discrete random variable which takes on values from a finite set *X*. The entropy of the random variable **X** is defined as (in bits):

$$H(\mathbf{X}) = -\sum_{x \in X} \mathbf{Pr}[x] \cdot \log_2 \mathbf{Pr}[x]$$

Entropy

- Example: Let **X** denote the outcome of coin toss
  - For coin toss, the head and tail appear with prob. 0.5

$$H(\mathbf{X}) = -0.5\log_2 0.5 - 0.5\log_2 0.5 = 1$$

  - If the coin is not perfect, the head appears with prob. 0.7

$$H(\mathbf{X}) = -0.7\log_2 0.7 - (1-0.7)\log_2(1-0.7) = 0.881$$

  - If coin toss is wrongly performed, and the head appears with prob. 1 (note that $\lim_{y->0} y\log_2 y = 0$)

$$H(\mathbf{X}) = -1\log_2 1 - (1-1)\log_2(1-1) = 0$$

$\Rightarrow$ Entropy is used to measure uncertainty

(as uncertainty decreases, entropy drops)

Entropy

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

- Entropy (per letter) of a natural language L
  - For random message: $H(\mathbf{P}) = (-\frac{1}{26}\log_2\frac{1}{26}) \times 26 = \log_2 26 \approx 4.70$
  - "First order approximation": single letters

$$H(\mathbf{P}) = -0.082\log_2 0.082 - 0.015\log_2 0.015 - .... - 0.001\log_2 0.001 = 4.19$$

  - "Second order approximation": digrams

$$H(\mathbf{P}^2)/2 \approx 3.9$$

  - we consider large segment of letters:

$$1.0 \le H_L = \frac{\lim_{n\to\infty} H(\mathbf{P}^n)}{n} \le 1.5$$

In average, each English letter carries about 1.5-bit information !

Entropy

- Redundancy of a natural language L

$$R_L = 1 - \frac{H_L}{\log_2 |P|}$$

information in a letter

information in a random letter

  - $|P|$ is the number of letters in a language (26 for English)
  - $\log_2 |P|$ denotes the entropy (per letter) of a random message: $(-\frac{1}{|P|}\log_2\frac{1}{|P|}) \times |P| = \log_2 |P|$
  - For the English language, if using H_L=1.25,

$$R_L = 1 - \frac{1.25}{\log_2 26} \approx 1 - \frac{1.25}{4.7} \approx 0.75$$

The English language is about 75% redundant!

Entropy

- Unicity distance
  - The unicity distance of a cryptosystem is defined as the average amount of ciphertext required to determine the key, **given unlimited computing resource.**

$$n_0 \approx \frac{\log_2 |K|}{R_L \log_2 |P|}$$

The uncertainty in the key

The information leaked from each ciphertext letter

- Examples
  - For a substitution cipher,
    - $|K| = 26!$, $n_0 \approx \log_2 26!/(0.75\times4.7) \approx 25$
  - For Vigenere cipher with key length 100,
    - $|K| = 26^{100}$, $n_0 \approx \log_2 26^{100}/(0.75\times4.7) \approx 133$

Summary

- One-Time Pad
  - Perfect secrecy
- Information theory
  - Entropy
  - Entropy & redundancy of a language
  - Unicity distance