

MAS 433: 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 length < message length
 - 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
 - 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 (perfect secrecy) ⇒ unbreakable even to attackers with unlimited computing resource
- The above encryption scheme is called *One-Time Pad*

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

$n \rightarrow 13$ $X \rightarrow 23$ $(13 + 23) \bmod 26 = 10 \rightarrow K$

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

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)
 - “ $(a + b) \bmod 2$ ” is denoted as “ $a \text{ XOR } b$ ”, “ $a \oplus b$ ”
 - In C programming language, “ $a \text{ XOR } b$ ” is “ $a \wedge b$ ”
- Example:

Plaintext: 1010011000

Key: \oplus 0110101110

Ciphertext: =

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
- 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

- 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 \mathbf{X} takes certain values with certain probabilities
 - The probability that the discrete random variable \mathbf{X} takes on a particular value x is denoted as $\mathbf{Pr}[\mathbf{X} = x]$
 - 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 \mathbf{X} is the result of coin toss: head or tail
 - The set of all the possible values of \mathbf{X} : $X = \{\text{tail}, \text{head}\}$
 - $\mathbf{Pr}[\mathbf{X} = \text{tail}] = \mathbf{Pr}[\mathbf{X} = \text{head}] = \frac{1}{2}$ (for a fair coin toss)

Elementary Probability Theory

- Random variable example 2: English text
 - Let \mathbf{X} be the random variable representing letters in English text
 - The set of all the possible values of \mathbf{X} :
$$X = \{a, b, c, d, \dots, z\}$$
 - $\Pr[\mathbf{X} = a] = 0.082, \Pr[\mathbf{X} = b] = 0.127, \dots$
 $\Pr[\mathbf{X} = z] = 0.01$

Elementary Probability Theory

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

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

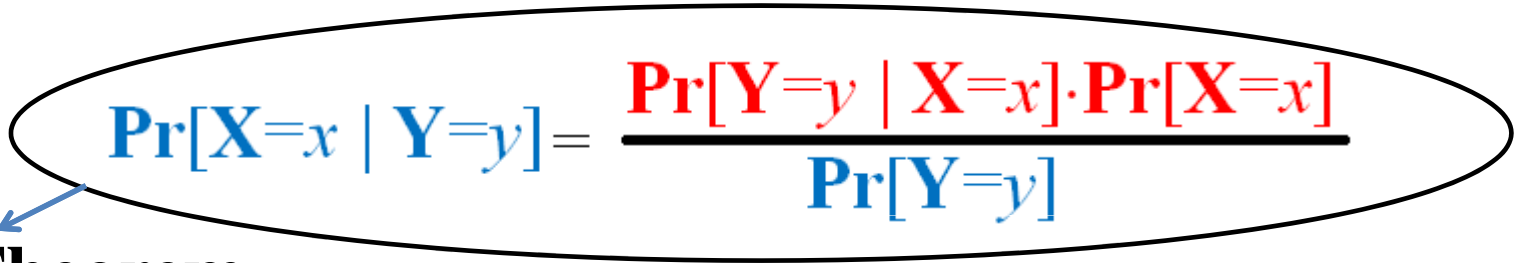
$$\Pr[X=x, Y=y] = \Pr[X=x \mid Y=y] \cdot \Pr[Y=y] \quad (1)$$

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

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

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

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


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

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 $X = \{2, 3, 4, \dots, 12\}$
- **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
- Now we perform the following computation to test Bayes' theorem:

$$\begin{aligned}\Pr[X = 4] &= \Pr[1\text{st dice} = 1] \cdot \Pr[2\text{nd dice} = 3] + \Pr[1\text{st dice} = 2] \cdot \Pr[2\text{nd dice} = 2] \\ &\quad + \Pr[1\text{st dice} = 3] \cdot \Pr[2\text{nd dice} = 1] \\ &= 1/6 \times 1/6 + 1/6 \times 1/6 + 1/6 \times 1/6 = 1/12\end{aligned}$$

$$\begin{aligned}\Pr[Y = d | X = 4] &= \frac{\Pr[1\text{st dice} = 2] \cdot \Pr[2\text{nd dice} = 2]}{\Pr[1\text{st dice} = 1] \cdot \Pr[2\text{nd dice} = 3] + \Pr[1\text{st dice} = 2] \cdot \Pr[2\text{nd dice} = 2] + \Pr[1\text{st dice} = 3] \cdot \Pr[2\text{nd dice} = 1]} \\ &= \frac{1/6 \times 1/6}{1/6 \times 1/6 + 1/6 \times 1/6 + 1/6 \times 1/6} = 1/3\end{aligned}$$

$$\begin{aligned}\Pr[Y = d] &= \Pr[1\text{st dice} = 1] \cdot \Pr[2\text{nd dice} = 1] + \dots + \Pr[1\text{st dice} = 6] \cdot \Pr[2\text{nd dice} = 6] \\ &= (1/6 \times 1/6) \times 6 = 1/6\end{aligned}$$

$$\begin{aligned}\Pr[X = 4 | Y = d] &= \frac{\Pr[1\text{st dice} = 2] \cdot \Pr[2\text{nd dice} = 2]}{\Pr[1\text{st dice} = 1] \cdot \Pr[2\text{nd dice} = 1] + \Pr[1\text{st dice} = 2] \cdot \Pr[2\text{nd dice} = 2] + \dots + \Pr[1\text{st dice} = 6] \cdot \Pr[2\text{nd dice} = 6]} \\ &= \frac{1/6 \times 1/6}{(1/6 \times 1/6) \times 6} = 1/6\end{aligned}$$

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

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 ,
$$\Pr[\mathbf{P} = p \mid \mathbf{C} = c] = \Pr[\mathbf{P} = p].$$
 - $\Pr[\mathbf{P} = p \mid \mathbf{C} = c]$ is *a posteriori* probability that the plaintext is p , given that the ciphertext c is observed.
 - $\Pr[\mathbf{P} = p]$ is *a priori* probability that the plaintext is p
- i.e., an attacker cannot guess the plaintext with higher probability after knowing the ciphertext

Perfect Secrecy of OTP

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

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
- “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 \mathbf{X} is a discrete random variable which takes on values from a finite set X . The entropy of the random variable \mathbf{X} is defined as (in bits):

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

Entropy

- Example: \mathbf{X} denotes 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 $\lim_{y \rightarrow 0} y \log_2 y = 0$)
$$H(\mathbf{X}) = -1 \log_2 1 - (1 - 1) \log_2 (1 - 1) = 0$$
- \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

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

- “Second order approximation”: digrams

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

- we consider large segment of letters:

$$1.0 \leq \underline{H_L} = \lim_{n \rightarrow \infty} \frac{H(\mathbf{P}^n)}{n} \leq 1.5$$

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

Entropy

- Redundancy of a natural language L

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

$\xrightarrow{\text{information in a letter}}$
 $\xrightarrow{\text{information in a random letter}}$

- $|P|$ 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,
 - $|K| = 26!$, $n_0 \approx \log_2 26! / (0.75 \times 4.7) \approx 25$
 - For Vigenere cipher with key length 100,
 - $|K| = 26^{100}$, $n_0 \approx \log_2 26^{100} / (0.75 \times 4.7) \approx 133$

Summary

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