Cryptography

Cunsheng DING, HKUST

COMP3631

Lecture 04: One-Key Block Ciphers

Outline of this Lecture

- One-key block ciphers and their security
- Simple substitution ciphers and their security
- The one-time pad

Ciphers and Their Classifications

Definition: A cipher is an encryption-and-decryption system for hiding information. The Julius Caesar cipher covered in Lecture 3 is an example.

Classifications:

- Ciphers are classified into two types:
  "one-key ciphers" (also called **symmetric** ciphers) and
  "two-key ciphers" (also called **asymmetric** ciphers).
- One-key ciphers are further classified into another two subtypes:
  "one-key block ciphers" and **"one-key stream ciphers"**.

We will introduce and study them one by one.

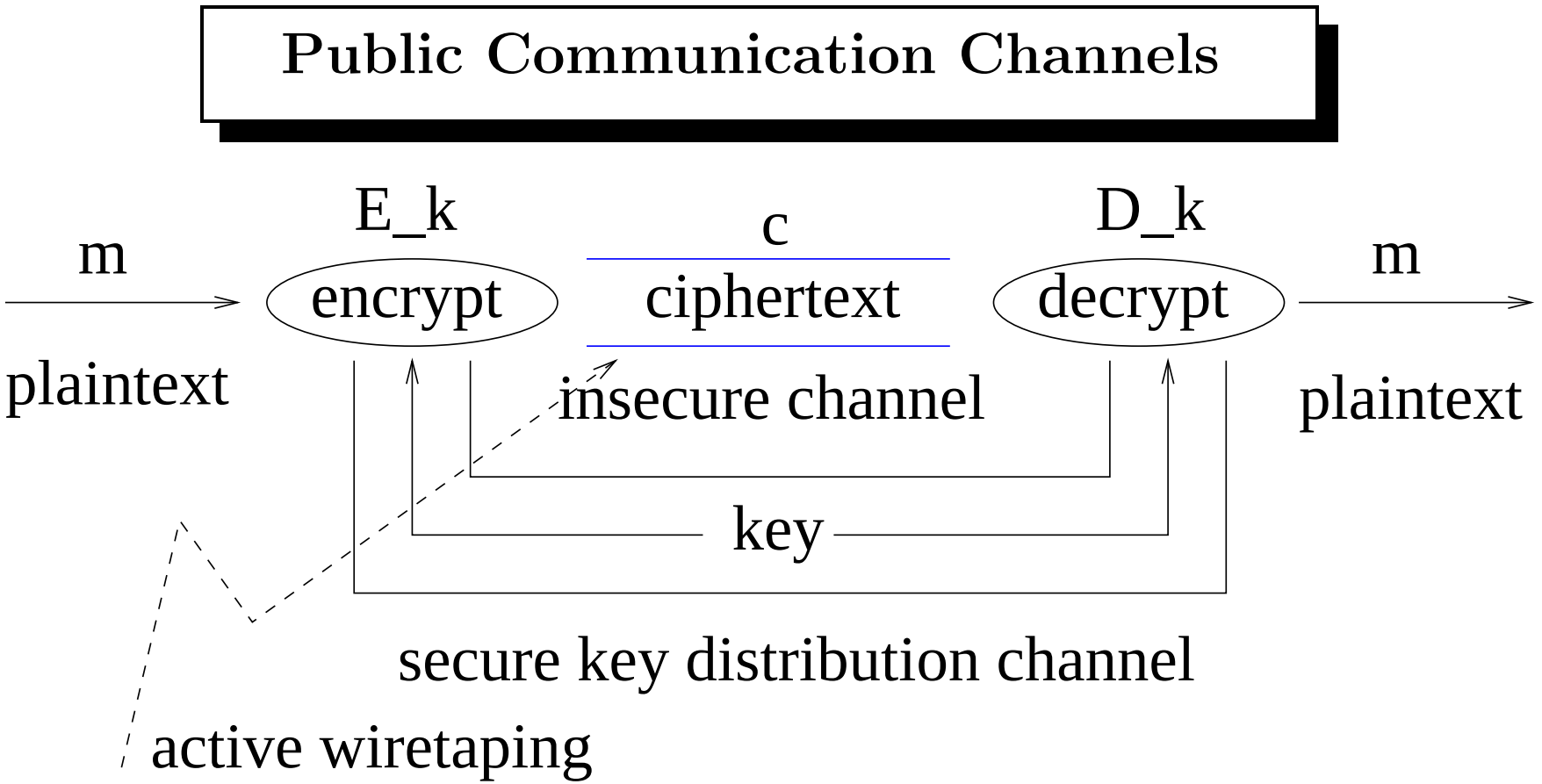
One-Key Block Ciphers

One-key Block Ciphers

A 5-tuple $(\mathcal{M}, \mathcal{C}, \mathcal{K}, E_k, D_k)$, where

- $\mathcal{M}$, $\mathcal{C}$, $\mathcal{K}$ are respectively the plaintext space, ciphertext space, and key space;
- Any $k \in \mathcal{K}$ could be the encryption and decryption key;
- E_k and D_k are encryption and decryption transformations with $D_k(E_k(m)) = m$ for each $m \in \mathcal{M}$.
- Encryption: $c = E_k(m)$, where E_k is usually applied to blocks or characters of the plaintext m.
- Decryption: $m = D_k(c)$, where D_k is usually applied to blocks or characters of the ciphertext c.

The ciphertext $c = E_k(m)$ depends **only on** k and m.


Public Communication Channels
E_k
c
D_k
m
encrypt
ciphertext
decrypt
m
plaintext
insecure channel
plaintext
key
secure key distribution channel
active wiretaping

Attacks on One-Key Block Ciphers

Ciphertext-only attack: An attacker determines the decryption transformation D_k or key k, or the plaintexts from a number of intercepted ciphertexts c.

Known-plaintext attack: An attacker determines the decryption transformation D_k or key k, from a number of ciphertext-plaintext pairs (c, m).

Security Requirements for One-key Block Ciphers

- The security should depend on the confidentiality of the key, so it is usually assumed that the algorithms E_k and D_k are known to a cryptanalyst.
- It should be **computationally infeasible** for a cryptanalyst to determine the plaintext m, given a ciphertext c.
- It should be **computationally infeasible** for a cryptanalyst to determine the decryption transformation D_k or the key k, given some ciphertext-plaintext pairs (c, m).

Question: How do you design a one-key block cipher meeting these requirements?

Answer: We will address this question in the next four lectures.

Attention: The algorithm E_k and the transformation E_k are different.

Simple Substitution Ciphers

A Special Type of One-Key Block Ciphers

Description of Simple Substitution Ciphers

Let A and B be two alphabets with $|A| \leq |B|$. We construct a simple substitution cipher $(\mathcal{M}, \mathcal{C}, \mathcal{K}, E_k, D_k)$, where

- $\mathcal{M} = A^*$, $\mathcal{C} = B^*$, i.e., all finite strings of characters from A (resp. B).
- $\mathcal{K}$ is the set of all one-to-one functions f from A to B (recall that $|\mathcal{K}| = |A|!\binom{|B|}{|A|}$, see Lecture 1).
- $k := f \in \mathcal{K}$ is the encryption and decryption key;
- For a message $m = m_0 m_1 m_2 \cdots$,

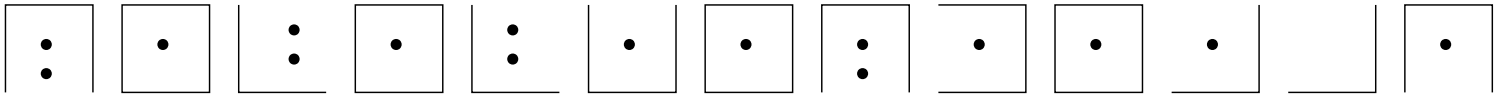
$$E_k(m) = f(m_0) f(m_1) f(m_2) \cdots$$

- For a ciphertext $c = c_0 c_1 c_2 \cdots$,

$$D_k(c) = f^{-1}(c_0) f^{-1}(c_1) f^{-1}(c_2) \cdots$$

First Example of Simple Substitution Ciphers

Example: Let A be the English alphabet and B the set of the 26 characters given in the following figure. The A and B define a simple substitution cipher, i.e., the churchyard cipher.



a	b	c
d	e	f
g	h	i

k	l	m
n	o	p
q	r	s

t	u	v
w	x	y
z	j	

Second Example of Simple Substitution Ciphers

Let $A = B$ be the English alphabet. Then A and B define a simple substitution cipher. We identify letters with digits:

a	b	c	$\cdots$	y	z
0	1	2	$\cdots$	24	25

Take any (k_0, k_1) with $\gcd(k_0, 26) = 1$ and $0 \le k_0 \le 25$, define the 1-to-1 function f by

$$f(\mu) = (\mu k_0 + k_1) \bmod 26.$$

It is called the **affine cipher**, where the key $k = (k_0, k_1)$ or $k = f$.

If $(k_0, k_1) = (1, 3)$, it is the **Caesar cipher**. RENAISSANCE is encrypted as UHQDLVVDQFH.

Question: Why should $\gcd(k_0, 26) = 1$?

The Security of Simple Substitution Ciphers

Claim 1: A simple substitution cipher for encrypting human languages is **not** secure with respect to known-plaintext attacks.

Claim 2: A simple substitution cipher for encrypting human languages is **insecure** with respect to ciphertext-only attacks!

Question: Why is a simple substitution cipher for encrypting human languages **insecure** with respect to ciphertext-only attacks?

Answer: Human languages have certain statistical properties.

Remark: We now consider the case that A is the English alphabet.

Frequency Distribution of Single English Letters

A	B	C	D	E	F	G	H	I	J	K	L	M
8.0	1.5	3.0	4.0	13.0	2.0	1.5	6.0	6.5	0.5	0.5	3.5	3.0

N	O	P	Q	R	S	T	U	V	W	X	Y	Z
7.0	8.0	2.0	0.2	6.5	6.0	9.0	3.0	1.0	1.5	0.5	2.0	0.2

Remark: In the table, 8.0 means 8.0%. E appears the most, and Z the least.

Frequency Distribution of Digraphs & Trigraphs

Definition: A digraph (also called bigram) is a sequence of two English letter, e.g., **th**
A trigraph is a sequence of three English letters, e.g., **the**

The most frequent digraphs: th, he, in, er, an, re, on, at, en, nd, ed, or, es, ti, te, it, is, st, to, ar, of, ng, ha, al

The most frequent trigraphs: the, and, tha, hat, ent, ion, for, tio, has, edt, tis, ers, res, ter, con, ing, men, tho

Remark: Some digraphs and trigraphs do not appear at all.

Question: What do the uneven frequency distributions (of single letters, digraphs and trigraphs) mean to the security of simple substitution ciphers?

Security of Simple Substitution Ciphers

Conclusion: Simple substitution ciphers for encrypting human languages are not secure with respect to ciphertext-only attacks, due to the uneven frequency distributions of single letters and digraphs.

Claim: For English, about 28 letters in a piece of ciphertext are needed for breaking a simple substitution cipher.

See, Denning, Cryptography and Data Security, 1982. You are not required to read this reference book about the claim above.

Breaking Simple Substitution Ciphers

Ciphertext-only attack: Given a piece of ciphertext c encrypted with a simple substitution cipher for encrypting the English language, we want to determine the key $k = f$ that is a 1-to-1 function from the English alphabet A to an alphabet B.

Cryptanalysis: For the given piece of ciphertext c, we compute the frequency distributions of letters and digraphs in B, and then compare them with those of the English letters, and try to match them. If the number of characters in c is long enough (in theory, 28 characters should work), the key is uniquely determined.

Exercise: On the course web page there are ten pieces of ciphertext. This will be an assignment problem.

The One-Time Pad

A secure but impractical cipher

The One-Time Pad

The One-Time Pad Cipher: $\mathcal{M} = \mathcal{C} = \mathcal{K} := \{0,1\}^*$.

- To encrypt a plaintext m, the sender generates a secret key k, which is a **"random binary string"** with the same length as the plaintext m.
- The sender distributes the key k to the receiver with another channel. (How? Key distribution will be treated later.)
- The ciphertext is the bitwise exclusive-or of the message with the secret key (the encryption process), i.e., $c = E_k(m) = m \oplus k$.
- The message is the bitwise exclusive-or of the ciphertext with the secret key (the decryption process), i.e., $m = D_k(c) = c \oplus k$.
- A secret key is used only for one message and is then discarded (a special key usage policy), and is called a **one-time pad**.

Security of the One-Time Pad

Conclusion 1: It is secure with respect to ciphertext-only attacks?

Proof: Let $m = m_\ell m_{\ell-1} \cdots m_2 m_1$ be the plaintext, and let $k = k_\ell k_{\ell-1} \cdots k_2 k_1$ be the secret key. Then the ciphertext

$$\begin{aligned} c = m \oplus k &= (m_\ell \oplus k_\ell)(m_{\ell-1} \oplus k_{\ell-1}) \cdots (m_2 \oplus k_2)(m_1 \oplus k_1) \\ &= c_\ell c_{\ell-1} \cdots c_2 c_1. \end{aligned}$$

By definition, $c_i = m_i \oplus k_i$. Knowing c_i does not give you any information about m_i, as $k_i = 0$ and $k_i = 1$ are equally likely (due to the fact k_i is a random bit).

Conclusion 2: It is secure with respect to known-plaintext attacks.

Proof: Given a pair (c, m), one can recover the key k as $k = c \oplus m$. However, each key k is used to encrypt only one message m.

The One-Time Pad Is Impractical

It is not practical due to the following:

- It is time-consuming to generate a **long random binary string**.
- It is not easy to distribute the secret key to the other party securely.
- Each key is used only once.

In summary, key generation and key distribution are not practical.

Summary

- We defined one-key block ciphers and talked about their security issues in general.
- We discussed simple substitution ciphers, and realized that a cipher may be insecure if it is not well designed.
- We saw that the one-time pad is secure in the information-theoretic sense but impractical.

Question: Is there any **secure** and **practical** cipher?

This is a very hard question!
We will spend three more lectures on one-key ciphers.

Appendix A: Transposition Ciphers

Permutations of $\mathbf{Z}_d$ for Transposition Ciphers

Let $\mathbf{Z}_d$ denote the set of integers 0 through $d-1$. A **permutation** f of $\mathbf{Z}_d$ is a one-to-one function from $\mathbf{Z}_d$ to itself.

Question: What is the total number of permutations on $\mathbf{Z}_d$?

Example: Let $d = 4$ and define f by

i:	0	1	2	3
$f(i)$:	2	0	3	1

Then f is a permutation of $\mathbf{Z}_4$.

Question: What is the inverse permutation f^{-1}?

Description of Transposition Ciphers

Let f be a permutation of $\mathbf{Z}_d$. It is a 5-tuple $(\mathcal{M}, \mathcal{C}, \mathcal{K}, E_k, D_k)$, where

- $\mathcal{M} = \mathcal{C} =$ set of all finite strings of English letters.
- $\mathcal{K}$ is the set of all possible pairs (d, f).
- $k = (d, f) \in \mathcal{K}$ is the secret key; and
- A message is divided into blocks of length d. For each message block $m = m_0 \cdots m_{d-1}$,

$$E_k(m) = m_{f(0)} \cdots m_{f(d-1)}$$

- For each ciphertext block $c = c_0 \cdots c_{d-1}$,

$$D_k(c) = c_{f^{-1}(0)} \cdots c_{f^{-1}(d-1)}$$

An Example of Transposition Ciphers

Example: Let $d = 4$ and define f by

i:	0	1	2	3
$f(i)$:	2	0	3	1

The message RENAISSANCES is broken into groups of 4 letters and encrypted into

position		0123	0123	0123
m	=	RENA	ISSA	NCES
$E_k(m)$	=	NRAE	SIAS	ENSC.

Exercise: Decrypt the ciphertext NRAESIASENSC.

The Security of Transposition Ciphers

Question: How do you detect a cipher as a transposition cipher?

Question: Is a transposition cipher secure with respect to known-plaintext attacks?

Question: Is a transposition cipher secure with respect to ciphertext-only attacks? If yes, justify your conclusion. If no, demonstrate how to **break** it.

Remark: The answers to these questions can be found in the Internet and some textbooks on cryptography.