



Cryptography and Security

Cunsheng DING
HKUST, Hong Kong

Version 3



Lecture 02: 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 1 is an example.

Classifications:

- Ciphers are classified into two types:
“one-key ciphers” and “two-key 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 later.



One-Key Block Ciphers



One-key Block Ciphers

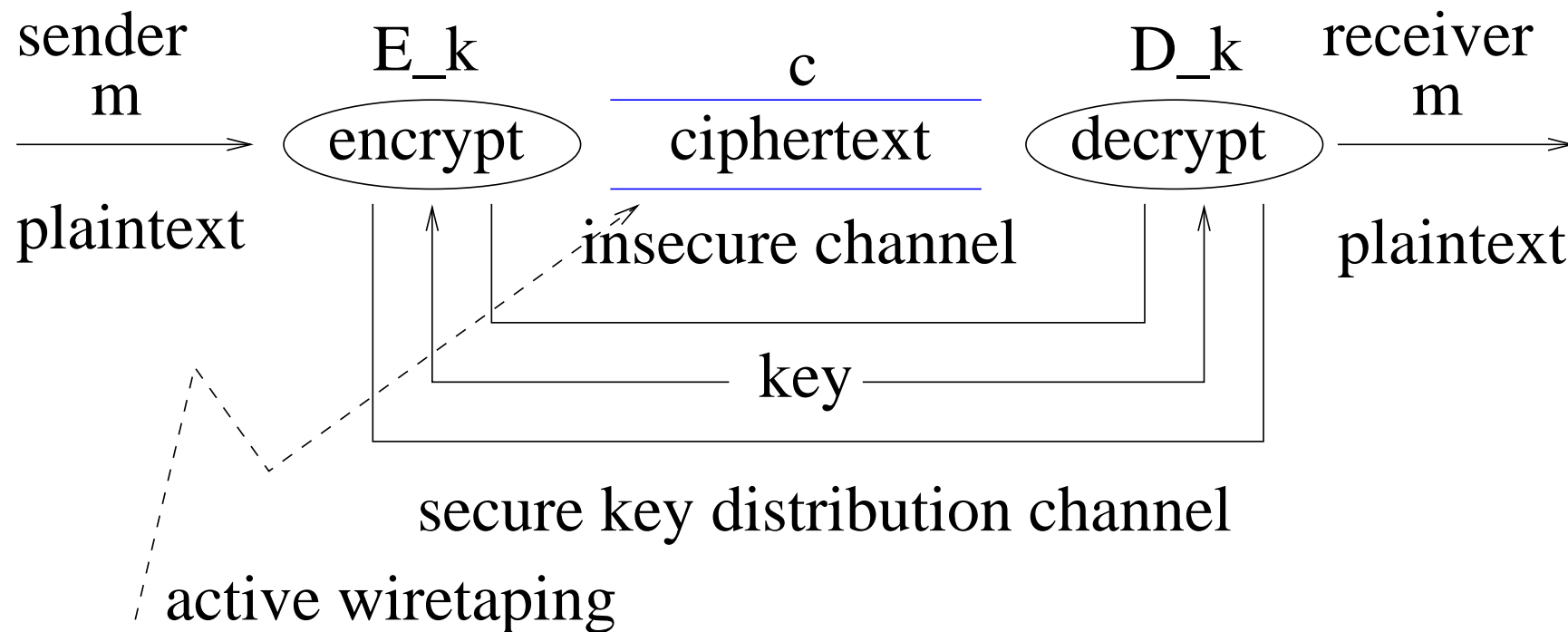
A 5-tuple $(\mathcal{M}, \mathcal{C}, \mathcal{K}, E_k, D_k)$, where

- $\mathcal{M}, \mathcal{C}, \mathcal{K}$ 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 , is time-independent.



Classical Information Channel





Attacks on One-Key Block Ciphers

Ciphertext-only attack: A cryptanalyst determines the decryption transformation D_k or key k , or the plaintext from intercepted ciphertext c .

Known-plaintext attack: A cryptanalyst determines the decryption transformation D_k or key k , from a ciphertext-plaintext pair (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 systematically determine the decryption transformation D_k or key k from intercepted ciphertext c , even if the corresponding plaintext m is known.

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



Simple Substitution Ciphers

A Special Type of One-Key Block Ciphers



Description of Simple Substitution Ciphers

Let f be a 1-to-1 mapping from alphabet A to alphabet B . It is a 5-tuple $(\mathcal{M}, \mathcal{C}, \mathcal{K}, E_k, D_k)$, where

- $\mathcal{M} = A^*$ and $\mathcal{C} = B^*$, i.e., all finite strings of characters.
- \mathcal{K} is the set of all possible f .
- $k = f \in \mathcal{K}$ is the encryption and decryption key;
- For a message $m = m_0m_1m_2 \cdots$,

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

- For a ciphertext $c = c_0c_1c_2 \cdots$,

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



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 following mapping f defines 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. We identify letters with digits:

$$\begin{array}{cccccc} a & b & c & \cdots & y & z \\ 0 & 1 & 2 & \cdots & 24 & 25 \end{array}$$

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

$$f(a) = (ak_0 + k_1) \bmod 26.$$

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 is **not** secure with respect to known-plaintext attacks.

Claim 2: A simple substitution cipher is insecure with respect to ciphertext-only attacks!

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



Frequency Distribution of Single English Letters

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>	<i>I</i>	<i>J</i>	<i>K</i>	<i>L</i>	<i>M</i>
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
<i>N</i>	<i>O</i>	<i>P</i>	<i>Q</i>	<i>R</i>	<i>S</i>	<i>T</i>	<i>U</i>	<i>V</i>	<i>W</i>	<i>X</i>	<i>Y</i>	<i>Z</i>
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. The uneven distribution of letters makes it easy to break simple substitution ciphers.



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 distributions (of single letters, digraphs and trigraphs) mean to the security of simple substitution ciphers?



Redundancy in Human Languages

Language redundancy: E.g., in “hwever”, “hoever” and “howeer”, you can easily determine the missing letters.

Comment: Shannon information theory can be used to give a rigorous measure of redundancy in a human language.

See, Denning, Cryptography and Data Security, 1982.

Why redundancy in human languages?

Comment: The uneven distributions of single English letters and digraphs are due to the redundancy in a human language.

Comment: The amount of redundancy in a human language affects the security of a one-key block cipher.

Remark: Chinese has less redundancy than English!



Security of Simple Substitution Ciphers

Claim: Simple substitution ciphers are not secure with respect to ciphertext-only attacks. **Why?**

Claim: For English, about 28 letters in a piece of ciphertext are needed to “break” a simple substitution cipher.

See, Denning, Cryptography and Data Security, 1982.



Breaking Simple Substitution Ciphers

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

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.



The One-Time Pad: A Special Cipher



The One-Time Pad

Question: Is there any unbreakable cipher?

One-time pad:

- Each message is encoded into a binary string (e.g., using the ASCII code).
- Generate a secret key, which is a “**random binary string**” with the same length as the message.
- The ciphertext is the bitwise exclusive-or of the message with the secret key (the encryption process).
- The message is the bitwise exclusive-or of the ciphertext with the secret key (the decryption process).
- A secret key is used only for one message and is then discarded (a special key usage policy).



Questions Regarding the One-Time Pad

Question: How do you prove that it is unbreakable?

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

$$c = m \oplus k = (m_\ell \oplus k_\ell)(m_{\ell-1} \oplus k_{\ell-1}) \cdots (m_2 \oplus k_2)(m_1 \oplus k_1).$$

Intuitive Proof: Let x be a random bit and let y be a message bit. Define $z = x \oplus y$. Knowing z does not give you any information about y , as $y = 0$ and $y = 1$ are equally likely (due to the fact x is a random bit).

Question: Is this a practical cipher?

- How to generate a long random binary string?
- How to distribute the secret key to the other party?



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 learnt a secure cipher (the one-time pad), which is not practical.

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: 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

$$\begin{array}{rcccc} i : & 0 & 1 & 2 & 3 \\ f(i) : & 2 & 0 & 3 & 1 \end{array}$$

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

$$\begin{array}{cccc} i : & 0 & 1 & 2 & 3 \\ f(i) : & 2 & 0 & 3 & 1 \end{array}$$

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

$$\begin{array}{rcccc} \text{position} & & 0123 & & 0123 & & 0123 \\ m & = & \text{RENA} & & \text{ISSA} & & \text{NCES} \\ E_k(m) & = & \text{NRAE} & & \text{SIAS} & & \text{ENSC.} \end{array}$$

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: These are left to students as exercises.