INTRODUCTION TO MODERN CRYPTOGRAPHY – 3RD EDITION SOLUTIONS MANUAL

HAO-MING PENG

ABSTRACT. This article presents my solutions to exercises in the 3rd edition of Katz and Lindell's textbook on cryptography "Introduction to Modern Cryptography".

Exercises in each chapter of the textbook are grouped under a section with the same title. Starred exercises are recommended to try, based on my own learning experience. Most of them are proofs of propositions, and there are also some that provide important insights for later chapters.

1. Introduction

Exercise 1.1. Decrypt the ciphertext provided at the end of the section on mono-alphabetic substitution ciphers.

Exercise 1.2. Provide a formal definition of the Gen, Enc, and Dec algorithms for the mono-alphabetic substitution cipher.

Solution. The message space $\mathcal{M}$ consists of all finite sequences of English letters, without distinction between upper and lower case. The key space $\mathcal{K}$ is the set of permutations over the set of English letters.

Gen: chooses a permutation $p \overset{u}{\leftarrow} \mathcal{K}$.

Enc: given a permutation p and a message $m \in \mathcal{M}$, where $m = m_1 m_2 \ldots m_\ell$, outputs the ciphertext:

$$c := c_1 c_2 \ldots c_\ell \quad \text{,where } c_i = p(m_i).$$

Dec: given a permutation p and a ciphertexct $c = c_1 c_2 \ldots c_\ell$, outputs the plaintext:

$$m := m_1 m_2 \ldots m_\ell \quad \text{,where } m_i = p^{-1}(c_i).$$

□

Exercise 1.3. Provide a formal definition of the Gen, Enc, and Dec algorithms for the Vigenère cipher. (Note: there are several plausible choices for Gen; choose one.)

Solution. Equate the English alphabet with the set $\mathbb{Z}_{26}$, that is $\mathsf{a} = 0$, $\mathsf{b} = 1$, etc. The message space $\mathcal{M}$ consists of all finite sequences of $\mathbb{Z}_{26}$. Index of sequence elements starts from 0.

Gen: given a length n, outputs a sequence:

$$k := k_0 k_1 k_2 \ldots k_n \quad \text{,where } k_i \overset{u}{\leftarrow} \mathbb{Z}_{26}.$$

Enc: given a key k and a message m, where $k = k_0 k_1 \ldots k_n$ and $m = m_0 m_1 \ldots m_\ell$, outputs the ciphertext:

$$c := c_0 c_1 \ldots c_\ell \quad \text{,where } c_i = m_i + k_{i \bmod n+1} \pmod{26}.$$

Dec: given a key k and a ciphertext c, where $k = k_0 k_1 \ldots k_n$ and $c = c_0 c_1 \ldots c_\ell$, outputs the plaintext:

$$m := m_0 m_1 \ldots m_\ell \quad \text{,where } m_i = c_i - k_{i \bmod n+1} \pmod{26}.$$

□

Exercise 1.4. Say you are given a ciphertext that corresponds to English-language text that was encrypted using either the shift cipher or the Vigenère cipher with period greater than 1. How could you tell which was the case?

Solution. Let q_i be the frequency of the ith letter in the ciphertext. If the ciphertext was encrypted by the shift cipher, then $\{q_i\}$ is a permutation of English-letter frequency. Therefore, $\sum_i q_i^2 \approx 0.65$. If it was encrypted by Vigenère cipher with period greater than 1, the frequency distribution might close to uniform. Hence, $\sum_i q_i^2 \approx \sum_i (\frac{1}{26})^2 \approx 0.38$. The encryption scheme then be determined by check which of 0.65 or 0.38 is closer to $\sum_i q_i^2$. □

Exercise 1.5. Implement the attacks described in this chapter for the shift cipher and the Vigenère cipher.

Exercise 1.6. The shift and Vigenère ciphers can also be defined on the 256-character alphabet consisting of all possible bytes (8-bit strings), and using XOR instead of modular addition.

(a) Provide a formal definition of both schemes in this case.
(b) Discuss how the attacks we have shown in this chapter can be modified to break these schemes.

Exercise 1.7. The index of coincidence method relies on a known value for the sum of the squares of plaintext-letter frequencies (cf. Equation (1.1)). Why would it not work using the $\sum_i p_i$ itself?

Solution. $\sum_i p_i$ always equal 1 for every distribution, so it doesn't provide statistical information of the underlying structure. □

***Exercise 1.8.** Show that the shift, substitution, and Vigenère ciphers are all trivial to break using a chosen-plaintext attack. How much chosen plaintext is needed to recover the key for each of the ciphers?

Solution. Suppose that random plaintext is encrypted, instead of English-language text, and the ciphertext is eavesrdropped by the adversary. Since that the plaintext is random, it can be decrypted only by recovering the key.

A single-letter plaintext is sufficient to mount the key-recovery attack on the shift cipher, since each plaintext letter is shifted forward the same value.

In the substitution cipher, the same plaintext letter always mapped to the same ciphertext letter. Choosing any plaintext that contains all letters in the alphabet is sufficient to build the full look-up table (key). Omiting one letter in the alphabet is fine, since the look-up table is one-to-one. Hence, if the alphabet contains n letters, then a $(n-1)$-length plaintext is sufficient.

For Vigenère cipher, there are three distinct threat models to consider, regarding the knowledge and adaptivity of the adversaries:

- The period of a Vigenère cipher is known: Any plaintext whose length is equal to the period is sufficient to recover the full key.
- The period is unknown and the adversary can choose plaintext adaptively, that is, *after* obtaining the ciphertext: The adversary can choose a plaintext that is of the same length as the ciphertext to be decrypted.
- The period is unknown and the adverary can choose plaintext only *before* obtaining the ciphertext: The adversary can choose a plaintext of arbitrary length, but this only reveals partial plaintext if the ciphertext to be decrypted is longer. (If shorter, it works the same as previous case.)

□

2. Perfectly Secrect Encryption

Exercise 2.1.

Exercise 2.2.

Exercise 2.3.

Exercise 2.4.

Exercise 2.5.

Exercise 2.6. Prove Lemma 2.7.

Solution. Prove perfect secrecy implies perfect indistinguishability by appeal to Lemma 2.5:

$$\begin{aligned}
&\Pr[\mathsf{PrivK}^{\mathsf{eav}}_{\mathcal{A},\Pi} = 1] \\
=&\sum_{b' \in \{0,1\}} \Pr[\mathrm{out}_{\mathcal{A}} = b' \wedge b = b'] \\
=&\sum_{c \in \mathcal{C}} \sum_{b' \in \{0,1\}} \Pr[\mathcal{A}(c) = b' \wedge \mathsf{Enc}_K(m_{b'}) = c \wedge b = b'] \\
=&\sum_{c \in \mathcal{C}} \sum_{b' \in \{0,1\}} \Pr[\mathcal{A}(c) = b'] \Pr[\mathsf{Enc}_K(m_{b'}) = c] \Pr[b = b'] \\
=&\frac{1}{2} \sum_{c \in \mathcal{C}} \Pr[\mathsf{Enc}_K(m) = c](\mathcal{A}(c) = 0 + \mathcal{A}(c) = 1) && \text{(Lem. 2.5)} \\
=&\frac{1}{2} \sum_{c \in \mathcal{C}} \Pr[\mathsf{Enc}_K(m) = c] = 1/2.
\end{aligned}$$

Prove perfect indistinguishability implies perfect secrecy by contradiction. Suppose that Π is perfectly indistinguishable. If Π is not perfectly secret, Lemma 2.5 tells that there exists a pair of messages m, m' and a cipherext c' such that

$$\Pr[\mathsf{Enc}_K(m) = c'] \neq \Pr[\mathsf{Enc}_K(m') = c'].$$

Consider an adverary $\mathcal{A}$ who (1) chooses $m_0 = m$ and $m_1 = m'$, and (2) outputs 0 if $c = c'$; otherwise, outputs 1. Show that $\Pr[\mathsf{PrivK}^{\mathsf{eav}}_{\mathcal{A},\Pi} = 1] \neq 1/2$:

$$\begin{aligned}
&\Pr[\mathsf{PrivK}^{\mathsf{eav}}_{\mathcal{A},\Pi} = 1] \\
=&\sum_{c \in \mathcal{C}} \sum_{b' \in \{0,1\}} \Pr[\mathcal{A}(c) = b'] \Pr[\mathsf{Enc}_K(m_{b'}) = c] \Pr[b = b'] \\
=&\frac{\Pr[\mathsf{Enc}_K(m_0) = c'] + \Pr[\mathsf{Enc}_K(m_1) \neq c']}{2} \\
=&\frac{\Pr[\mathsf{Enc}_K(m) = c'] + 1 - \Pr[\mathsf{Enc}_K(m') = c']}{2} \\
\neq&1/2.
\end{aligned}$$

The third line is due to $\Pr[\mathcal{A}(c') = 0] = 1$ and $\Pr[\mathcal{A}(c') = 1] = 0$. This contradits that Π is perfectly indistinguishable. Hence, it must be the case that Π is perfectly secret. □

Exercise 2.7.

Exercise 2.8.

Exercise 2.9.

Exercise 2.10.

Exercise 2.11.

Exercise 2.12.

Exercise 2.13.

Exercise 2.14.

Exercise 2.15.

Exercise 2.16.

Exercise 2.17.

Exercise 2.18.

***Exercise 2.19.** In this problem we consider definitions of perfect secrecy for the encryption of two messages (using the same key). Here we consider distributions on pairs of messages from the message space $\mathcal{M}$; we let M_1, M_2 be random variables denoting the first and second message, respectively. (These random variables are not assumed to be independent.) We generate a (single) key k, sample a pair of messages (m_1, m_2) according to the given distribution, and then compute ciphertexts $c_1 \leftarrow \mathsf{Enc}_k(m_1)$ and $c_2 \leftarrow \mathsf{Enc}_k(m_2)$; this induces a distribution on pairs of ciphertexts and we let C_1, C_2 be the corresponding random variables.

(a) Say encryption scheme $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ is *perfectly secret for two messages* if for all distributions on $\mathcal{M} \times \mathcal{M}$, all $m_1, m_2 \in M$, and all ciphertexts $c_1, c_2 \in C$ with $\Pr[C_1 = c1 \wedge C_2 = c_2] > 0$:

$$\begin{aligned}\Pr[M_1 = m_1 \wedge M_2 = m_2 \mid C_1 = c_1 \wedge C_2 = c_2]&\\ = \Pr[M_1 = m_1 \wedge M_2 = m_2].&\end{aligned}$$

Prove that *no* encryption scheme can satisfy this definition.

Hint: Take $c_1 = c_2$.

(b) Say encryption scheme $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ is *perfectly secret for two messages* if for all distributions on $\mathcal{M} \times \mathcal{M}$, all $m_1, m_2 \in M$ where he first and second messages are guaranteed to be different (i.e., distributions on pairs of distinct messages), and all ciphertexts $c_1, c_2 \in C$ with $\Pr[C_1 = c1 \wedge C_2 = c_2] > 0$:

$$\begin{aligned}\Pr[M_1 = m_1 \wedge M_2 = m_2 \mid C_1 = c_1 \wedge C_2 = c_2]&\\ = \Pr[M_1 = m_1 \wedge M_2 = m_2].&\end{aligned}$$

Show an encryption scheme that provably satisfies this definition.

Hint: The encryption scheme you propose need not be efficient, although an efficient solution is possible.

Solution.

(a) Let $m, m' \in \mathcal{M}$ with $m \neq m'$. Consider the distribution:

$$\Pr[M_1 = m \wedge M_2 = m] = 1/2$$
$$\Pr[M_1 = m \wedge M_2 = m'] = 1/2.$$

Then if $c_1 = c = c_2$:

$$\begin{aligned}&\Pr[M_1 = m \wedge M_2 = m' \mid C_1 = c \wedge C_2 = c]\\ =&\Pr[M_1 = m \wedge M_2 = m' \mid \mathsf{Enc}_K(M_1) = \mathsf{Enc}_K(M_2) = c]\\ =&\Pr[M_1 = m \wedge M_2 = m' \mid M_1 = M_2 = \mathsf{Dec}_K(c)]\\ =&\ 0 \neq \Pr[M_1 = m \wedge M_2 = m'].\end{aligned}$$

Hence, the definition can not be satisfied.

(b) The failure of one-time pad is due to that its encryption algorithm injects algebraic properties into its ciphertexts. The idea is to use a uniform permutaion to encrypt, so that each ciphertext is independent. [1]

[1] It is like a substitution cipher which enforces that all characters in a plaintext are distinct.

The message space is a single bit $\mathcal{M} = \{0, 1\}$ and $\Pi = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ works as:

Gen: chooses a uniform permutation of $\{0, 1\}$. The key space contains the two bijections $f(x) = x$ and $g(x) = \neg x$.

Enc: takes as inputs a permutation p and a message $m \in \{0, 1\}$, and outputs

$$c \coloneqq p(m).$$

Dec: takes as inputs a permutation p a ciphertext c, and outputs

$$m \coloneqq p^{-1}(m).$$

Consider without loss of generality that $m_1 = 0$ and $m_2 = 1$:

$$\begin{aligned}
&\Pr[M_1 = 0 \wedge M_2 = 1 \mid C_1 = c_1 \wedge C_2 = c_2] \\
=&\Pr[M_1 = 0 \wedge M_2 = 1 \mid \mathsf{Enc}_K(M_1) = c_1 \mathsf{Enc}_K(M_2) = c_2] \\
=&\frac{\Pr[\mathsf{Enc}_K(0) = c_1 \wedge \mathsf{Enc}_K(1) = c_2] \Pr[M_1 = 0 \wedge M_2 = 1]}{\Pr[\mathsf{Enc}_K(M_1) = c_1 \wedge \mathsf{Enc}_K(M_2) = c_2]}
\end{aligned}$$

The two statements $\mathsf{Enc}_K(0) = c_1$ and $\mathsf{Enc}_K(1) = c_2$ uniquely determines a permutation over $\{0, 1\}$. Therefore, $\Pr[\mathsf{Enc}_K(0) = c_1 \wedge \mathsf{Enc}_K(1) = c_2] = 1/2$. Similarly, $\Pr[\mathsf{Enc}_K(M_1) = c_1 \wedge \mathsf{Enc}_K(M_2) = c_2] = 1/2$. Hence,

$$\Pr[M_1 = 0 \wedge M_2 = 1 \mid C_1 = c_1 \wedge C_2 = c_2] = \Pr[M_1 = 0 \wedge M_2 = 1].$$

The scheme Π acheives the above definition. □