In this lecture we mainly discuss perfect secrecy. We begin by briefly recalling some of the basic probability. Next, we define the problem of secure communication, identify the relevant agents and offer some assumptions about the model. Then we define the notion of perfect secrecy and study the one-time pad encryption which achieves this level of security. We remark that in most settings the limitation regarding the key length makes the one-time pad or any other perfectly secret scheme unusable.

1 Probability Theory

1.1 Discrete/finite probability

Definition 1. An experiment is a procedure that yields one of a given set of outcomes. The individual possible outcomes are called simple events. The set of all possible outcomes is called the sample space.

In this lecture, we only consider discrete sample spaces; that is, sample spaces with only finitely many possible outcomes. For example, the simple events of a sample space $S$ are labeled as $s_1, s_2, \ldots, s_n$.

Definition 2. A probability distribution $P$ on $S$ is a sequence of numbers $p_1, p_2, \ldots, p_n$ that are all non-negative and sum to 1. The number $p_i$ is interpreted as the probability of $s_i$ being the outcome of the experiment.

Definition 3. An event $E$ is a subset of the sample space $S$. The probability that event $E$ occurs, denoted $Pr[E]$, is the sum of the probabilities $p_i$ of all simple events $s_i$ which belong to $E$. If $s_i \in S$, $Pr[\{s_i\}]$ is simply denoted by $Pr[s_i]$.

Definition 4. Definition If $E$ is an event, the complementary event is the set of simple events not belonging to $E$, denoted $\overline{E}$.

Let $E \subseteq S$ be an event. It obvious that
(1) $0 < Pr[E] < 1$. Furthermore, $Pr[S] = 1$ and $Pr[\emptyset] = 0$.
(2) $Pr[\overline{E}] = 1 - Pr[E]$.

Definition 5. A random variable $X$ is a function from the sample space $S$ to the set of real numbers. For each simple event $s_i \in S$, $X$ assigns a real number $X(s_i)$.

Since $S$ is assumed to be finite, $X$ can only take on a finite number of values.
1.2 Conditional Probability

**Definition 6.** Let $E_1$ and $E_2$ be two events with $Pr[E_2] > 0$. The *conditional probability* of $E_1$ given $E_2$, denoted $Pr(E_1|E_2)$, is

$$Pr[E_1|E_2] = \frac{Pr[E_1 \cap E_2]}{Pr[E_2]}$$

$Pr[E_1|E_2]$ measures the probability of event $E_1$ occurring, given that $E_2$ has occurred.

**Definition 7.** Events $E_1$ and $E_2$ are said to be *independent* if $Pr[E_1 \cap E_2] = Pr[E_1]Pr[E_2]$.

Observe that if $E_1$ and $E_2$ are independent, then $Pr[E_1|E_2] = Pr[E_1]$ and $Pr[E_2|E_1] = Pr[E_2]$. That is, the occurrence of one event does not influence the likelihood of occurrence of the other.

**Theorem 1** (Bayes’ Theorem). If $E_1$ and $E_2$ are events with $Pr[E_2] > 0$, then

$$Pr[E_1|E_2] = \frac{Pr[E_1]Pr[E_2|E_1]}{Pr[E_2]}$$

**Theorem 2** (Law of Total Probability). If $E_i, \ i = 1, 2, 3, \ldots$, is a finite or countably infinite partition ($Pr[E_i \cap E_j] \neq \emptyset, i \neq j$) of a probability space and each set $E_i$ is measurable, then for any event $F$ we have

$$Pr(F) = \sum_i Pr[F \cap E_i]$$

or, alternatively,

$$Pr(F) = \sum_i Pr[F | E_i]Pr[E_i]$$

2 The Setting of Private-Key Encryption

2.1 Shannon’s Model of Secret Communication

![Shannon’s Model](image)

Figure 1: Shannon’s Model

Remarks on the model:

1. Shannon’s key setup only applied to symmetric (private key) system.
2. In this model, plain text source, encryption algorithm, decryption algorithm, key source and key transmission channel are all assumed to be secure.


4. Only consider the attack called eavesdrop.

2.2 Syntax of Encryption Scheme

A private-key encryption scheme comprises three algorithms: the first is a procedure for generating keys, the second a procedure for encrypting, and the third a procedure for decrypting.

1. Gen: A key-generation algorithm is a probabilistic algorithm that outputs a key $k$ chosen according to some distribution that is determined by the scheme.

   The set of all possible keys output by the key-generation algorithm is called the key space and is denoted by $\mathcal{K}$. WLOG, Gen simply chooses a key uniformly at random from the key space.

2. Enc: The encryption algorithm is a probabilistic algorithm that takes as input a key $k$ and a plaintext message $m$ and outputs a ciphertext $c$. We denote by $\text{Enc}_k(m)$ the encryption of the plaintext $m$ using the key $k$.

   The set of all “legal” messages is denoted $\mathcal{M}$ and is called the plaintext (or message) space. Since any ciphertext is obtained by encrypting some plaintext under some key, the sets $\mathcal{K}$ and $\mathcal{M}$ together define a set of all possible ciphertexts denoted by $\mathcal{C}$.

3. Dec: The decryption algorithm is a deterministic algorithm that takes as input a key $k$ and a ciphertext $c$ and outputs a plaintext $m$. We denote the decryption of the ciphertext $c$ using the key $k$ by $\text{Dec}_k(c)$.

An encryption scheme is fully defined by specifying the three algorithms (Gen, Enc, Dec) and the plaintext space $\mathcal{M}$.

The basic correctness requirement of any encryption scheme is that for every key $k$ output by Gen and every plaintext message $m \in \mathcal{M}$, it holds that

$$\text{Dec}_k(\text{Enc}_k(m)) = m$$

So decrypting a ciphertext (using the appropriate key) yields the original message that was encrypted.

3 Perfect Secrecy

Let’s first consider the probability distributions over $\mathcal{K}$, $\mathcal{M}$ and $\mathcal{C}$. The distribution over $\mathcal{K}$ is simply the one that is defined by running Gen and taking the output. For $k \in \mathcal{K}$, we let $Pr[K = k]$ denote the probability that the key output by Gen is equal to $k$. Similarly, for $m \in \mathcal{M}$, we let $Pr[M = m]$ denote the probability that the message is equal to $m$.

The distribution over $\mathcal{K}$ and $\mathcal{M}$ are independent. This is the case because the key is chosen and fixed before the message is known. Furthermore, the distribution over $\mathcal{K}$ is fixed by the encryption scheme itself while the distribution over $\mathcal{M}$ may vary depending on the parties who are using the encryption scheme.
For \( c \in \mathcal{C} \), we let \( P_r[C = c] \) denote the probability that the ciphertext is \( c \). Given the encryption algorithm \( \text{Enc} \), the distribution over \( \mathcal{C} \) is fully determined by the distributions over \( \mathcal{K} \) and \( \mathcal{M} \).

\[
P_r[C = c] \overset{\text{def}}{=} \sum_{m \in \mathcal{M}, \; k \in \mathcal{K} \; \text{s.t.} \; \text{Enc}_k(m) = c} P_r[M = m, K = k]
\]

We are now ready to define the notion of perfect secrecy. Intuitively, we imagine an adversary who knows the probability distribution over \( \mathcal{M} \), that is, the adversary knows the likelihood that different messages will be sent. The adversary then observes some ciphertext being sent by one party to the other. Ideally, obtaining the ciphertext gives no advantage to the adversary. In other words, the posteriori likelihood that some message \( m \) was sent should be the same as priori probability that \( m \) would be sent. This should hold for any \( m \in \mathcal{M} \). This means that a ciphertext reveals nothing about the underlying plaintext, and thus an adversary who intercepts a ciphertext learns absolutely nothing about the plaintext that was encrypted.

**Definition 8.** An encryption scheme \((\text{Gen}, \text{Enc}, \text{Dec})\) over a message space \( \mathcal{M} \) is perfectly secret if for every probability distribution over \( \mathcal{M} \), every message \( m \in \mathcal{M} \), and every ciphertext \( c \in \mathcal{C} \) for which \( P_r[C = c] > 0 \):

\[
P_r[M = m | C = c] = P_r[M = m]
\]

Notice here \( P_r[C = c] > 0 \) is a technical requirement needed to avoid conditioning on a zero-probability event.

### 4 One-Time Pad (Vernam’s Cipher)

The one-time pad was invented in 1917 and patented a couple of years later. It is derived from the Vernam cipher, named after Gilbert Vernam, one of its inventors. The "pad" part of the name comes from early implementations where the key material was distributed as a pad of paper, so the top sheet could be easily torn off and destroyed after use. One-time pad has been proven achieves perfect secrecy by Shannon.

#### 4.1 The One-Time Pad Encryption Scheme

Let \( a \) and \( b \) be two binary strings of length \( l \), denoted as \( a, b \in \{0, 1\}^l \). Define \( a \oplus b \) the bitwise exclusive-or (XOR) of \( a \) and \( b \). The one-time pad encryption is defined as follows:

1. Fix an integer \( l > 0 \), Then the message space \( \mathcal{M} \), key space \( \mathcal{K} \), and ciphertext space \( \mathcal{C} \) are all equal to \( \{0, 1\}^l \).
2. The key-generation algorithm \( \text{Gen} \) works by tossing \( l \) coins and generate a string \( k \) from \( \mathcal{K} \) according to the uniform distribution.
3. Given a key \( k \in \mathcal{K} \) and a message \( m \in \mathcal{M} \), the encryption algorithm \( \text{Enc} \) outputs \( c = m \oplus k \).
4. Given a key \( k \in \mathcal{K} \) and a ciphertext \( c \in \mathcal{C} \), the decryption algorithm \( \text{Dec} \) outputs \( m = c \oplus k \).

Intuitively, the one-time pad is perfectly secret because given a ciphertext \( c \), there is no way an adversary can know which plaintext \( m \) it originated from. We now prove the security formally.

**Theorem 3.** Vernam’ cipher (a.k.a one-time pad encryption) is perfectly-secret.
Proof. Let’s fix some distribution over $\mathcal{M}$ and fix an arbitrary $m \in \mathcal{M}$ and $c \in \mathcal{C}$. In order to prove the perfect secrecy, we need to show that $Pr[M = m|C = c] = Pr[M = m]$. From Bayes’ Theorem and the independence of $K$ from $M$, we know that

$$Pr[M = m|C = c] = \frac{Pr[C = c|M = m] \cdot Pr[M = m]}{Pr[C = c]} = \frac{Pr[M \oplus K = c|M = m] \cdot Pr[M = m]}{\sum_{m,k \in \{0,1\}^l \atop m \oplus k = c} Pr[M = m, K = k]}$$

$$= \frac{Pr[M \oplus K = c|M = m] \cdot Pr[M = m]}{\sum_{m,k \in \{0,1\}^l \atop m \oplus k = c} Pr[M = m] \cdot Pr[K = k]}$$

$$= \frac{Pr[K = c \oplus m|M = m] \cdot Pr[M = m]}{\frac{1}{2^l} \cdot \sum_{m \in \{0,1\}^l} Pr[M = m]$$

$$= \frac{1}{2^l} \cdot Pr[M = m]$$

$$= Pr[M = m]$$

Q.E.D.

4.2 Limitations of One-Time Pad

Shannon also provided a characterization of perfectly-secret encryption schemes

**Theorem 4** (Shannon’s Theorem). Let $(Gen, Enc, Dec)$ be an encryption scheme over a message space $\mathcal{M}$ for which $|\mathcal{M}| = |\mathcal{K}| = |\mathcal{C}|$. The scheme is perfectly secret if and only if:

1. Every key $k \in \mathcal{K}$ is chosen with equal probability $1/|\mathcal{K}|$ by algorithm $Gen$.

2. For every $m \in \mathcal{M}$ and every $c \in \mathcal{C}$, there exists a unique key $k \in \mathcal{K}$ s.t. $Enc_k(m)$ outputs $c$.

We can conclude that perfect secrecy is attainable. However, the one-time pad encryption scheme has a number of drawbacks.

- The key is required to be as long as the message.

- As the name indicates, one-time pad scheme is only secure if used once with the same key.

- If the messages correspond to English-language text, then given the XOR of two sufficiently-long message, it has been shown to be possible to perform frequency analysis.

5 Limitations of Perfect Secrecy

In fact, the aforementioned limitations of the one-time pad encryption scheme are inherent. Specifically, we can prove that any perfectly secret encryption scheme must have a key space that is at least as large as the message space. If the key space consists of fixed-length keys, and the message space consists of all messages of some fixed length, this implies that the key must be at least as long as the message. Thus, the problem of a large key length is not specific to the one-time pad, but is inherent to any scheme achieving perfect secrecy.
Theorem 5. Let \((\text{Gen}, \text{Enc}, \text{Dec})\) be a perfectly-secret encryption scheme over a message space \(\mathcal{M}\), and let \(\mathcal{K}\) be the key space as determined by \(\text{Gen}\). Then \(|\mathcal{K}| \geq |\mathcal{M}|\).

Proof. We can prove this theorem by contradiction. Assume \(|\mathcal{K}| < |\mathcal{M}|\). Consider the uniform distribution over \(\mathcal{M}\) and let \(c \in \mathcal{C}\) be a ciphertext that occurs with non-zero probability. Let \(\mathcal{M}(c)\) be the set of all possible messages which are possible decryptions of \(c\).

\[
\mathcal{M}(c) \overset{\text{def}}{=} \{ \hat{m} | \hat{m} = \text{Dec}_k(c) \text{ for some } \hat{k} \in \mathcal{K} \}
\]

Since \(\text{Dec}\) is deterministic, for each message \(\hat{m} = \mathcal{M}(c)\) we can identify at least one key \(\hat{k} \in \mathcal{K}\) s.t. \(\hat{m} = \text{Dec}_\hat{k}(c)\). It’s clear that \(|\mathcal{M}(c)| < |\mathcal{K}|\). Under the assumption that \(|\mathcal{K}| < |\mathcal{M}|\), there exist some \(m' \in \mathcal{M}\) s.t. \(m' \notin \mathcal{M}(c)\). Then

\[
Pr[M = m'|C = c] = 0 \neq Pr[M = m']
\]

and so the scheme is not perfectly secret. \(\square\)

References

