3.16.21

ne the inverse of the matrix

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

3.16.22

ey for the affine linear cipher with alphabet {A,B,C,...,Z} k length three that encrypts "RED" as "ONE".

4 CHAPTER Probability and Perfect Secrecy

In the previous chapter, we have described a number of historical cryptosystems. It turned out that they were all affine linear and therefore insecure. Are there cryptosystems that are provably secure? In 1949, Claude Shannon [66] was able to describe such systems. Unfortunately, those systems are not very efficient. Also, they are not secure against active attacks if no further cryptographic techniques are used. In this chapter, we present Shannon's theory. At the same time, we will introduce a few notions and results of elementary probability theory.

4.1 Probability

et {A,B,C,. . . ,Z}
E".

4 CHAPTER Probability and Perfect Secrecy

In the previous chapter, we have described a number of historical cryptosystems. It turned out that they were all affine linear and therefore insecure. Are there cryptosystems that are provably secure? In 1949, Claude Shannon [66] was able to describe such systems. Unfortunately, those systems are not very efficient. Also, they are not secure against active attacks if no further cryptographic techniques are used. In this chapter, we present Shannon's theory. At the same time, we will introduce a few notions and results of elementary probability theory.

4.1 Probability

Let S be a finite nonempty set. We call it the *sample space*. Its elements are called *elementary events*. The elementary events model outcomes of experiments.

Example 4.1.1
If we flip a coin, we either obtain heads H or tails T. The sample space is $S = \{H,T\}$.

Exercise 3.16.21
Determine the inverse of the matrix

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

mod 2.

Exercise 3.16.22
Find a key for the affine linear cipher with alphabet {A,B,C,. . . ,Z} and block length three that encrypts "RED" as "ONE".

4 CHAPTER Probability and Perfect Secrecy

In the previous chapter, we have described a number of historical cryptosystems. It turned out that they were all affine linear and therefore insecure. Are there cryptosystems that are provably secure? In 1949, Claude Shannon [66] was able to describe such systems. Unfortunately, those systems are not very efficient. Also, they are not secure against active attacks if no further cryptographic techniques are used. In this chapter, we present Shannon's theory. At the same time, we will introduce a few notions and results of elementary probability theory.

4.1 Probability

Let S be a finite nonempty set. We call it the *sample space.* Its elements are called *elementary events.* The elementary events model outcomes of experiments.

Example 4.1.1
If we flip a coin, we either obtain heads H or tails T. The sample space is $S = \{H,T\}$.

If we throw a die, then we obtain a number in $\{1, 2, 3, 4, 5, 6\}$. Therefore, the sample space is $S = \{1, 2, 3, 4, 5, 6\}$.

An *event* (for S) is a subset of the sample space S. The *certain event* is the set S itself. The *null event* is the empty set $\emptyset$. We say that two events A and B are *mutually exclusive* if their intersection is empty. The set of all events is the *power set* $P(S)$ of S.

Example 4.1.2
An event is, for example, to obtain an even number when throwing a die. Formally, this event is $\{2, 4, 6\}$. It excludes the event $\{1, 3, 5\}$ to obtain an odd number.

A *probability distribution* on S is a map Pr that sends an event to a real number, namely

$$\Pr : P(S) \to \mathbb{R},$$

and has the following properties:

1. $\Pr(A) \geq 0$ for all events A,
2. $\Pr(S) = 1$,
3. $\Pr(A \cup B) = \Pr(A) + \Pr(B)$ for two events A and B, which are mutually exclusive.

If A is an event, then $\Pr(A)$ is the *probability* of this event. The probability of an elementary event $a \in S$ is $\Pr(a) = \Pr(\{a\})$.

It is easy to see that $\Pr(\emptyset) = 0$. Moreover, $A \subset B$ implies $\Pr(A) \leq \Pr(B)$. Therefore, $0 \leq \Pr(A) \leq 1$ for all $A \in P(S)$. Moreover, $\Pr(S \setminus A) = 1 - \Pr(A)$. If $A_1, \ldots, A_n$ are pairwise mutually exclusive events, then $\Pr(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \Pr(A_i)$.

Since S is a finite set, it suffices to define the probability distribution on elementary events. In fact, if A is an event, then $\Pr(A) = \sum_{a \in A} \Pr(a)$.

Example 4.1.3
The probability distribution on the set $\{1, 2, 3, 4, 5, 6\}$, which models throwing a die, maps each elementary event to $1/6$. The probability of the event "even result" is $\Pr(\{2, 4, 6\}) = \Pr(2) + \Pr(4) + \Pr(6) = 1/6 + 1/6 + 1/6 = 1/2$.

The probability distribution that maps each elementary event $a \in S$ to the probability $P(a) = 1/|S|$ is called the *uniform distribution*.

4.2 Conditional Probability

Let S be a sample space, and let Pr be a probability distribution on S. We explain conditional probability in an example.

Example 4.2.1
Again, we model throwing a die. The sample space is $\{1, 2, 3, 4, 5, 6\}$, and Pr sends any elementary event to $1/6$. Suppose Claus has thrown one of the numbers $4, 5, 6$, so we know that the event $B = \{4, 5, 6\}$ has happened. Under this assumption, we want to determine the probability that Claus has thrown an even number. Each elementary event in B is equally probable. Therefore, each elementary event in B has probability $1/3$. Since two numbers in B are even, the probability that Claus has thrown an even number is $2/3$.

Definition 4.2.2
Let A and B be events and $\Pr(B) > 0$. The conditional probability of "A given that B" occurs is defined to be

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}.$$

This definition can be understood as follows. We want to know the probability of A if B is certain to occur (i.e., the sum of the probabilities of all elementary events x in $A \cap B$). Such an elementary event has probability $\Pr(x)/\Pr(B)$ because $\Pr(B) = 1$. Therefore, the event $A \cap B$ has probability $\Pr(A \cap B)/\Pr(B)$.

Two events A and B are called *independent* if

$$\Pr(A \cap B) = \Pr(A)\Pr(B).$$

This condition is equivalent to

$$\Pr(A|B) = \Pr(A).$$

If the events are not independent, we call them *dependent*.

Example 4.2.3
If we flip two coins, then the probability of the event "the first coin comes up tails" is independent from the event "the second coin comes up tails". The probability that both events occur is 1/4. The probability of each individual event is 1/2.

If the coins are welded together such that they either both fall heads or both tails, then the probability of two tails is $1/2 \neq 1/2*1/2$. Hence, the events "the first coin comes up tails" and "the second coin comes up tails" are dependent.

We formulate and prove the theorem of Bayes.

Theorem 4.2.4
If A and B are events with $\Pr(A) > 0$ and $\Pr(B) > 0$, then

$$\Pr(B)\Pr(A|B) = \Pr(A)\Pr(B|A).$$

Proof. By definition, we have $\Pr(A|B)\Pr(B) = \Pr(A \cap B)$ and $\Pr(B|A)\Pr(A) = \Pr(A \cap B)$. This implies the assertion. □

4.3 Birthday Paradox

A good example for reasoning in probability theory is the birthday paradox. The problem is the following. Suppose a group of people are in a room. What is the probability that two of them have the same birthday? This probability is astonishingly large.

We will make a slightly more general analysis. Suppose that there are n birthdays and that there are k people in the room. An elementary event is a tuple $(b_1, \dots, b_k) \in \{1, 2, \dots, n\}^k$. If it occurs, then the birthday of the ith person is b_i, $1 \leq i \leq k$, so we have n^k elementary events. We assume that those elementary events are equally probable. Then the probability of an elementary event is $1/n^k$.

We want to compute the probability that two people in the room have the same birthday. Denote this probability by p. Then with probability $q = 1 - p$ any two people have different birthdays. We estimate this probability. The event in which we are interested is the set E of all vectors $(g_1, \dots, g_k) \in \{1, 2, \dots, n\}^k$ whose entries are pairwise different. Since the probability of an elementary event is $1/n^k$, the probability of E is the number of elements in E divided by n^k. The number of elements in E is the number of vectors in $\{1, \dots, n\}^k$ with pairwise different entries. This number is computed now. The first entry can be any of the n possibilities. If the first entry is fixed, then there are $n - 1$ possibilities for the second entry, and so on. Hence, we obtain

$$|E| = \prod_{i=0}^{k-1}(n-i)$$

and

$$q = \frac{1}{n^k}\prod_{i=0}^{k-1}(n-i) = \prod_{i=1}^{k-1}\left(1 - \frac{i}{n}\right). \tag{4.1}$$

Now $1 + x \leq e^x$ holds for all real numbers. Therefore, from (4.1) we obtain

$$q \leq \prod_{i=1}^{k-1} e^{-i/n} = e^{-\sum_{i=1}^{k-1} i/n} = e^{-k(k-1)/(2n)}. \tag{4.2}$$

If

$$k \geq (1 + \sqrt{1 + 8n\log 2})/2, \tag{4.3}$$

then (4.2) implies that $q \leq 1/2$. Then the probability $p = 1 - q$ that two people have the same birthday is at least 1/2. For $n = 365$, the choice $k = 23$ is sufficient for $q \leq 1/2$. In other words, if 23 people are in a room, then the probability that two of them have the same birthday is at least 1/2.

4.4 Perfect Secrecy

Following Shannon, we will now introduce perfect secrecy. We assume the following scenario. Alice uses a cryptosystem to send encrypted messages to Bob. If she sends such an encrypted message to Bob, the attacker, Oscar, can read the ciphertext. Oscar tries to obtain information concerning the plaintext from the ciphertext. A cryptosystem has perfect secrecy if Oscar learns nothing about the plaintext from the ciphertext. We want to formalize this property mathematically.

The cryptosystem has a finite plaintext space $\mathcal{P}$, a finite ciphertext space $\mathcal{C}$, and a finite key space $\mathcal{K}$. The encryption functions are E_k, $k \in \mathcal{K}$ and the decryption functions are D_k, $k \in \mathcal{K}$.

We assume that the probability of a plaintext p is $\Pr_{\mathcal{P}}(p)$. The function $\Pr_{\mathcal{P}}$ is a probability distribution on the plaintext space. It depends, for example, on the language that is used. The distribution Pr also depends on the application context. For example, if Alice is a university professor, then it is likely that she frequently uses the word "student". For the encryption of a new plaintext, Alice chooses a new key which is independent of the plaintext to be encrypted. The probability for a key k is $\Pr_{\mathcal{K}}(k)$. The function $\Pr_{\mathcal{K}}$ is a probability distribution on the key space. The probability that a plaintext p occurs and is encrypted with key k is

$$\Pr(p,k) = \underset{\mathcal{P}}{\Pr}(p)\underset{\mathcal{K}}{\Pr}(k). \tag{4.4}$$

This defines a probability distribution Pr on the sample space $\mathcal{P} \times \mathcal{K}$. We will now consider this sample space only. If p is a plaintext, then we also denote by p the event $\{(p,k) : k \in \mathcal{K}\}$ that p is encrypted. Clearly, we have

$$\Pr(p) = \underset{\mathcal{P}}{\Pr}(p).$$

Also, for a key $k \in \mathcal{K}$ we denote by k the event $\{(p,k) : p \in \mathcal{P}\}$ that the key k is chosen for encryption. Clearly, we have

$$\Pr(k) = \underset{\mathcal{K}}{\Pr}(k).$$

By (4.4), the events p and k are independent. For a ciphertext $c \in \mathcal{C}$, we denote by c the event $\{(p,k) : E_k(p) = c\}$ that the result of the encryption is c.

Oscar knows the probability distribution $\Pr_{\mathcal{P}}$ on the plaintexts because he knows, for example, the language that Alice and Bob use. Now Oscar sees a ciphertext. If the fact that this ciphertext has occurred makes some plaintexts more likely than they are according to the probability distribution $\Pr_{\mathcal{P}}$ and others less likely, then Oscar learns something from observing c. Otherwise, if the probability for each plaintext remains the same, then Oscar learns nothing. This motivates Shannon's definition of perfect secrecy, which we present now.

Definition 4.4.1

The cryptosystem of this section has *perfect secrecy* if the events that a particular ciphertext occurs and that a particular plaintext has been encrypted are independent (i.e., $\Pr(p|c) = \Pr(p)$ for all plaintexts p and all ciphertexts c).

Example 4.4.2

Let $\mathcal{P} = \{0,1\}$, $\Pr(0) = 1/4$, $\Pr(1) = 3/4$. Also, let $\mathcal{K} = \{A,B\}$, $\Pr(A) = 1/4$, $\Pr(B) = 3/4$. Finally, let $\mathcal{C} = \{a,b\}$. Then the probability that the plaintext 1 occurs and is encrypted with key B is $\Pr(1)\Pr(B) = 9/16$. The encryption function E_K works as follows:

$$E_A(0) = a, E_A(1) = b, E_B(0) = b, E_B(1) = a.$$

The probability of the ciphertext a is $\Pr(a) = \Pr(0,A) + \Pr(1,B) = 1/16 + 9/16 = 5/8$. The probability of the ciphertext b is $\Pr(b) = \Pr(1,A) + \Pr(0,B) = 3/16 + 3/16 = 3/8$.

We now compute the conditional probability $\Pr(p|c)$ for all plaintexts p and all ciphertexts c. It is $\Pr(0|a) = 1/10$, $\Pr(1|a) = 9/10$, $\Pr(0|b) = 1/2$, $\Pr(1|b) = 1/2$. Those results show that the cryptosystem described does not have perfect secrecy. If Oscar receives the ciphertext a he can be reasonably sure that the corresponding plaintext is 1.

We formulate and prove the famous theorem of Shannon.

Theorem 4.4.3

Let $|\mathcal{P}| = |\mathcal{K}| = |\mathcal{C}| < \infty$ and $\Pr(p) > 0$ for any plaintext p. Our cryptosystem has perfect secrecy if and only if the probability distribution on the key space is the uniform distribution and if for any plaintext p and any ciphertext c there is exactly one key k with $E_k(p) = c$.

Proof. Suppose that the cryptosystem has perfect secrecy. Let p be a plaintext. If there is a ciphertext c for which there is no key k with $E_k(p) = c$, then $\Pr(p) \neq \Pr(p|c) = 0$ since $\Pr(p) > 0$ by assumption. This contradicts the perfect secrecy. Hence, for any ciphertext c there is a key k with $E_k(p) = c$. But the number of keys is equal to the number of ciphertexts. Therefore, for each ciphertext c there is exactly one key k with $E_k(p) = c$. This proves the second assertion.

To prove the first assertion, we fix a ciphertext c. For a plaintext p, let $k(p)$ be the uniquely determined key with $E_{k(p)}(p) = c$. Then we have

$$\mathcal{K} = \{k(p) : p \in \mathcal{P}\} \tag{4.5}$$

since the number of plaintexts is equal to the number of keys. Below we show that for all $p \in \mathcal{P}$ the probability of $k(p)$ is equal to the probability of c. Then the probability of $k(p)$ does not depend on p. Hence the probability of all $k(p)$ is the same. Since by (4.5) every key $k \in \mathcal{K}$ is equal to $k(p)$ for some $p \in \mathcal{P}$, the probability distribution the key space is the uniform distribution.

Let $p \in \mathcal{P}$. As promised, we show that $\Pr(k(p)) = \Pr(c)$. It follows from Theorem 4.2.4 that

$$\Pr(p|c) = \frac{\Pr(c|p)\Pr(p)}{\Pr(c)} = \frac{\Pr(k(p))\Pr(p)}{\Pr(c)} \tag{4.6}$$

Since the cryptosystem has perfect secrecy, we have $\Pr(p|c) = \Pr(p)$. So (4.6) implies $\Pr(k(p)) = \Pr(c)$, as asserted.

Now we prove the converse. Assume that the probability distribution on the key space is the uniform distribution and that for any plaintext p and any ciphertext c there is exactly one key $k = k(p, c)$ with $E_k(p) = c$. Then

$$\Pr(p|c) = \frac{\Pr(p)\Pr(c|p)}{\Pr(c)} = \frac{\Pr(p)\Pr(k(p,c))}{\sum_{q \in \mathcal{P}} \Pr(q)\Pr(k(q,c))}. \tag{4.7}$$

Now $\Pr(k(q,c)) = 1/|\mathcal{K}|$ for all $q \in \mathcal{P}$, $c \in \mathcal{C}$. since all keys are equally probable. Hence,

$$\sum_{q \in \mathcal{P}} \Pr(q)\Pr(k(q,c)) = \frac{\sum_{q \in \mathcal{P}} \Pr(q)}{|\mathcal{K}|} = \frac{1}{|\mathcal{K}|}.$$

If we use this equation in (4.7), then we obtain $\Pr(p|c) = \Pr(p)$, as asserted. □

Example 4.4.4
Theorem 4.4.3 implies that the cryptosystem from example 4.4.2 has perfect secrecy if we set $\Pr(A) = \Pr(B) = 1/2$.

4.5 Vernam One-Time Pad

The most famous cryptosystem that has perfect secrecy is the *Vernam one-time pad*, which is explained in this section. Let n be a positive integer. The Vernam one-time pad encrypts bitstrings of length n. Plaintext space, ciphertext space, and key space are $\mathcal{P} = \mathcal{C} = \mathcal{K} = \{0,1\}^n$. The encryption function for key $k \in \{0,1\}^n$ is

$$E_k : \{0,1\}^n \to \{0,1\}^n, \quad p \mapsto p \oplus k.$$

The decryption function for key k is the same.

To encrypt a plaintext $p \in \{0,1\}^n$, Alice chooses a key k randomly with uniform distribution from the set $\{0,1\}^n$.. She computes the ciphertext $c = p \oplus k$. By Theorem 4.4.3, this cryptosystem is perfectly secure because the uniform distribution is used on the key space and for each plaintext p and each ciphertext c there is exactly one key k with $c = p \oplus k$, namely $k = p \oplus c$.

This cryptosystem was invented and patented in 1917 by Gilbert Vernam. However, it was not until 1949 that Shannon proved that the Vernam one-time pad has perfect secrecy.

Unfortunately, the one-time pad is not very efficient. To secretly communicate a plaintext of length n, Alice and Bob must randomly generate and exchange a key of length n. This is the reason for the name "one-time pad". Each key can be used only once.

If a key is used to encrypt several plaintexts, the one-time pad loses its perfect secrecy. Oscar can determine the key in a known plaintext attack. Suppose he knows a plaintext p and the corresponding ciphertext c. Then the key can be determined as $m \oplus c = m \oplus m \oplus k = k$.

Also, the One-Time-Pad is not secure against active attacks. This is demonstrated in the next example.

Example 4.5.1
Alice encrypts her electronic bank transactions using the one-time pad. If the attacker Oscar knows where in the ciphertext the amount is encrypted, then he can change that part of the ciphertext.

In Chapter 11 countermeasures against such active attacks are described.

4.6 Random Numbers

If Alice and Bob want to use the Vernam one-time pad, then they need a source for uniformly distributed random bits. It is a philosophical question whether such a source can exist or whether anything that happens is predetermined. In practice, random-bit generators are used which are software or hardware-based. Such generators are devices that use, for example, the randomness of radioactive decay or the time between two keyboard strokes. An overview can be found in [59].

If random-bit generators are used in cryptography, then it is important that an attacker have no way of predicting the bits that it outputs. Therefore, those generators are typically secure hardware devices.

In the following, we assume that we are given a random-bit generator that generates random bits according to the uniform distribution. We explain how such a device is used to generate random numbers.

We want to generate uniformly distributed random numbers in the set $\{0, 1, \ldots, m\}$, $m \in \mathbb{N}$. We set $n = \operatorname{size} m = \lfloor \log m \rfloor + 1$. Then we generate n random bits $b_1, \ldots, b_n$. If the number $a = \sum_{i=1}^{n} b_i 2^{n-i}$ is greater than m, then we forget it and generate a new one in the same way. Otherwise, a is the random number. It is easy to verify that the numbers a that are generated in this way are uniformly distributed random numbers in the set $\{0, 1, \ldots, m\}$.

If we want to generate uniformly distributed random n-bit numbers, $n \in \mathbb{N}$, then we generate $n - 1$ random bits $b_2, \ldots, b_n$ and set $b_1 = 1$ and output $a = \sum_{i=1}^{n} b_i 2^{n-i}$.

4.7 Pseudorandom Numbers

If it is too time-consuming to generate true random numbers, then pseudorandom number generators are used. A pseudorandom number generator is an algorithm that, given a short sequence of random bits, produces a long sequence of bits that "looks" random. This means that the output sequence cannot be distinguished in polynomial time from a true random sequence. A detailed description of the corresponding theory can be found in [31]. Pseudorandom number generators that are used in practice can be found in [49].

4.8 Exercises

Exercise 4.8.1

Let S be a finite set and Pr a probability distribution on S. Prove the following:

1. $\Pr(\emptyset) = 0$.
2. $A \subset B \subset S$ implies $\Pr(A) \leq \Pr(B)$.

Exercise 4.8.2

In an experiment, m is chosen with uniform distribution from $\{1, 2, \ldots, 1000\}$. Determine the following probabilities:

1. for choosing a square;
2. for choosing a number with i prime factors, $i \geq 1$.

Exercise 4.8.3

Find the sample space and probability distribution that model the experiment of flipping two coins. Describe the event "at least one coin comes up heads" formally and compute its probability.

Exercise 4.8.4

Determine the probability that a randomly chosen map $\{0, 1\}^* \to \{0, 1\}^*$ is affine linear.

Exercise 4.8.5

We throw two dice. Determine the probability that they both show different numbers under the condition that the sum of both numbers is even.

Exercise 4.8.6

Determine n such that the probability of two of n people having the same birthday is at least 9/10.

Exercise 4.8.7

Suppose that four-digit PINs are randomly distributed. How many people must be in a room such that the probability that two of them have the same PIN is at least 1/2?

Exercise 4.8.8

Prove that the Caesar cipher does not have perfect secrecy.

Exercise 4.8.9

Consider the linear block cipher with block length n and alphabet $\{0,1\}^n$. On the key space of matrices $A \in \{0,1\}^{(n,n)}$ with $\det(A) \equiv 1 \bmod 2$, choose the random distribution. Does this cryptosystem have perfect secrecy?

5 CHAPTER DES

In Chapter 3, we have defined cryptosystems and we have described some historical examples. All of the cryptosystems in Chapter 3 could all be broken because they are affine linear. A cryptosystem with perfect secrecy, the Vernam one-time pad, was presented in Chapter 4, but it turns out to be very inefficient. In this chapter, we describe the Data Encryption Standard (DES). For many years, this cryptosystem was the encryption standard in the U.S. and was used worldwide. Today, simple DES is no longer secure and the US National Institute of Standards (NIST) has chosen the Rijndael cryptosystem as the Advanced Encryption Standard (AES) [1] (see Chapter 6). AES is described in Chapter 6. Nevertheless, there are secure variants of DES (see Section 3.7) that are widely use. Also, DES remains an important model for the construction of secure block ciphers.

5.1 Feistel Ciphers

The DES algorithm is a so-called *Feistel cipher*. In this section, we explain Feistel ciphers.