

Chapter 1

Introduction

Let us begin the study of cryptography with the classical problem of transmitting secret messages from a sender A to a receiver B . Both the sender and the receiver may be persons, organisations, various technical systems. Sometimes one speaks of A and B as of subscribers of some network, users of some computer system, or, more formally, abstract “parties” or “entities” participating in information exchange. But it often occurs more convenient to identify the participants of exchange with some humans and use the names Alice and Bob instead of A and B .

It is assumed that messages are transmitted via an open communications channel which can potentially be accessed by a third party different from the sender and receiver. Such a situation arises in radio transmission (say, from a mobile phone) and is possible even in such “trusted” systems as wire telephone, telegraph, as well as in ordinary mail. The Internet, as a means of communication gaining the leading positions all over the world, offers a special interest of being extremely vulnerable for unauthorised access of third parties. In this environment, not only copying of data is easily implemented but also deletion and substitution.

It is generally assumed in cryptography that the person who sends and/or receives messages has an adversary or enemy E , which can be a competitor in business, a member of a criminal group, a foreign intelligence agent, or even an excessively jealous spouse, and that the adversary can read and analyse the messages transmitted. The adversary is often thought of as a person called Eve who has powerful computing facilities and is able to use cryptanalytic methods. Of course, Alice and Bob want their messages to be completely unclear to Eve, and, to achieve this, they use appropriate ciphers.

Before transmitting a message from A to B over an open communica-

tions channel, A encrypts (or enciphers) the message. In his turn, B , after having received the encrypted message (ciphertext), decrypts (or decipher) it to recover the initial text (plaintext). It is important for the problem considered in this chapter that Alice and Bob can agree in advance about the cipher to be used (or rather about certain parameters thereof) *not via an open channel* but via a special “secure” channel which is inaccessible for Eva. Such a “secure channel” can be maintained with the aid of trusted messengers or couriers, or Alice and Bob can agree on the cipher during their private meeting, *etc.* It is necessary to take into account that, usually, maintaining the secure channel and transmitting messages over this channel is much more expensive compared with an open unsecured channel, and (or) the secure channel cannot be used at any time. For instance, courier post is far more expensive than the regular one, it transmits messages much slower than, say, electronic mail, and may be used not at any hour and not in any situation.

To be more concrete, consider an example of cipher. Since the problem of encryption has occurred long ago in centuries some ciphers are named after renowned historical persons. These ciphers are often used to introduce simple initial concepts and we shall follow that tradition. Let us start with a well-known cipher by Gajus Julius Caesar. In this cipher, each letter of a message is substituted by the other letter whose ordinal in the alphabet is increased by 3. For instance, the letter A is replaced by D , the letter B by E , and so on. The last 3 letters X , Y , Z are replaced by A , B , C , respectively. Thus the word $SEQUENCE$ transforms to $VHTXHQFH$ under the Caesar cipher.

Other Roman caesars have modified the cipher by using the shifts through 4, 5 and more letters in the alphabet. We can describe such ciphers in a general way if we enumerate (encode) the letters by their ordinal numbers (from 0 to 25). Then the rule of encryption will be

$$c = (m + k) \bmod 26, \quad (1.1)$$

where m and c are the ordinals of letters of plaintext and ciphertext, respectively, and k is an integer called the cipher key (in the Caesar cipher considered above, $k = 3$). (Here and after $a \bmod b$ denotes the remainder from division of integer a by integer b , the remainder being taken from the set $\{0, 1, \dots, b - 1\}$. For instance, $13 \bmod 5 = 3$.)

To decrypt the ciphertext one should apply an “inverse” algorithm

$$m = (c - k) \bmod 26. \quad (1.2)$$

Let's imagine that the sender and receiver have agreed to use the cipher (1.1) but, to make the adversary's job more difficult, decided to occasionally change the cipher key. For that purpose, Alice somehow generates the number k , sends it to Bob over a secure channel, after which they communicate messages encrypted with that k . The key may be changed prior to each communication session or after transmitting a specified number of letters (say, encipher every ten letters using a different key) and so forth. In this situation, the key is said to be generated by a key source. A schematic view of the cryptosystem considered is shown in Fig. 1.1.

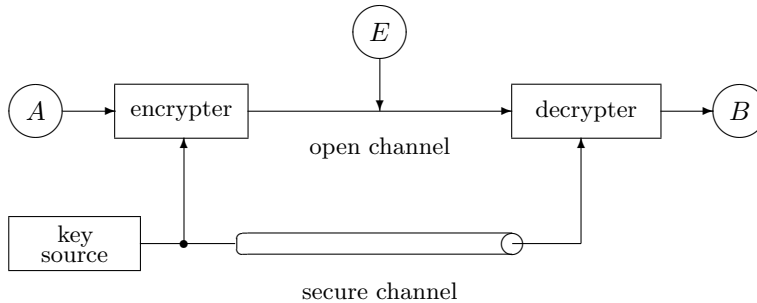


Fig. 1.1 Classical secret communications system.

Proceed now to the discussion about the actions of adversary who tries to recover the message and find the secret key, or, in other words, to break the cipher. Every attempt to break a cipher is called an *attack* to the cipher (or cryptosystem). It is generally assumed in cryptography that the adversary always has the ciphertext in her disposal and can learn everything about the encrypting algorithm used and the nature of data transmitted, but only does not know the secret key. These are the Kerckhoffs assumptions named after the scientist who was the first to formulate the main requirements to ciphers [Kerckhoffs (1883)]. Sometimes these assumptions may seem “overcautious” but such “overcautiousness” is by no means superfluous if, say, you send an order to transfer one million dollars from one account to another.

In our example, Eve knows that the plaintext was encrypted according to (1.1), the message was in English, and the ciphertext is VHTXHQFH. But the key is unknown to her.

The most obvious method to recover plaintext from ciphertext is to search through all possible keys (this is a so-called *brute-force attack*

or *exhaustive key search*). Thus Eva tries successively all possible keys $k = 1, 2, \dots$, applying them to decrypting algorithm and estimating the obtained results. Let us also try to use this method. The results of decrypting according to (1.2) under various keys and the ciphertext VHTXHQFH are shown in Table 1.1. In the majority of trials it was sufficient to decrypt only a few letters to reject the corresponding key due to the absence of the word in English that begins with these letters.

Table 1.1 Decrypting the word VHTXHQFH by exhaustive key search.

k	m	k	m	k	m	k	m
1	UGS	8	NZ	15	GS	22	ZL
2	TF	9	MYKOY	16	FRD	23	YK
3	SEQUENCE	10	LX	17	EQC	24	XJ
4	RD	11	KWI	18	DP	25	WIU
5	QC	12	JV	19	COAE	26	VHTXHQFH
6	PB	13	IU	20	BN		
7	OAM	14	HT	21	AMY		

We can see from Table 1.1 that the key $k = 3$ was used and hence the message SEQUENCE was enciphered. Moreover, when checking for the other values of key, it was not required to decrypt all 8 letters since the key might often be rejected after decrypting 2 or 3 initial letters. This example shows that the Caesar cipher is completely insecure: for breaking it, one needs to analyse several initial letters of the message after which the key is disclosed unambiguously and, consequently, the whole message may be deciphered.

What are the reasons of weakness of the considered cipher and how might its security be increased? Consider another example. Suppose that Alice hid some important documents in a safe with 5-digit combination lock. Now she would like to tell Bob the combination for opening the safe. She decided to use an analogue of the Caesar cipher adapted to the alphabet of decimal digits:

$$c = (m + k) \bmod 10. \quad (1.3)$$

Suppose she sent Bob the ciphertext 26047. Eve tries to decrypt it, as earlier, by searching through all possible keys. The results of her work are shown in Table 1.2.

We can see that all the variants are equivalent and Eve cannot decide on what combination is true. Based on the ciphertext only, she is unable to

Table 1.2 Decrypting the word 26047 by exhaustive key search.

k	m	k	m
1	15936	6	60481
2	04825	7	59370
3	93714	8	48269
4	82603	9	37158
5	71592	0	26047

find the secret key. Of course, before intercepting the encrypted message she had 10^5 possible lock combinations, and after that only 10. But it is important to note that in this particular example we have only 10 possible values of the key. Under such a key (one decimal digit) Alice and Bob could not count on better security.

The message in our first example is the text in natural language (English). So it obeys numerous rules, various letters and their combinations have different probabilities and, in particular, many combinations are forbidden (this property is referred to as redundancy of the text). And that is why the key was easily found and the message recovered, i.e. the redundancy had made it possible to break the cipher. But, in contrast, in our second example all combinations of digits are admissible. The “language” of combination lock does not possess any redundancy. Therefore even a simple cipher applied to messages in that language becomes unbreakable. In the classical work by C. Shannon [Shannon (1949)], a deep and elegant theory of secret key ciphers is constructed and, specifically, a “correct” measure of redundancy is suggested. We shall briefly touch upon these topics in Chap. 7, and in Chap. 8 some modern secret key ciphers will be described.

The attack to the cipher considered in the previous examples is said to be a *ciphertext-only attack*. But sometimes a so-called *known-plaintext attack* to the cipher may be maintained. This happens if Eve obtains in her disposal some plaintexts corresponding to previously transmitted ciphertexts. Eve tries to disclose the secret key by examining the pairs plaintext–ciphertext. If she succeeded, she would be able to decrypt all further messages from Alice to Bob.

One can imagine even a more “serious” attack, a so-called *chosen-plaintext attack*. In this attack, an adversary not only can access some plaintext–ciphertext pairs but is also able to create plaintexts on her own

and encrypt them under the key she wants to disclose. For instance, during World War II, Americans, having bribed the guards, stole the cipher-machine in Japan Embassy for two days at weekend and had an opportunity to input various plaintext and obtain corresponding ciphertexts. They could not open the machine to directly determine the installed key because the damage would be detected and all the keys immediately changed (this and many other stories can be found in [Kahn (1967)]).

It may seem that the known- and chosen-plaintext attacks are rather artificial and hard to maintain. It is so to some extent. But the designers of modern cryptosystems strive to make them invulnerable even to those kinds of attacks and there are great achievements in this direction. It is customary to think that it is more reliable to use a cipher secure against chosen-plaintext attacks rather than organisationally ensure the impossibility of such attacks. Extremely cautious people do both things.

So, we have acquainted with the main characters of cryptography — Alice, Bob, and Eve, and with important notions of that science — a cipher, a key, an attack, open and secure channels. Note that an intriguing fact is connected with the last item: secure cryptosystems are possible to construct without any secure channel! In such cryptosystems Alice and Bob compute the secret key so that Eve cannot do that. This discovery was made in the seminal works ([Diffie and Hellman (1976); Merkle (1979)]) and has constituted a new epoch in modern cryptography. The main part of this book will be devoted to this kind of cryptosystems, referred to as *public-key* or *asymmetric-key* schemes.

Problems and Exercises

- 1.1 Find the keys of the Caesar cipher if the following plaintext–ciphertext pairs are known:
 - (a) ORANGE – FIREXV
 - (b) APRICOT – XMOFZLQ
- 1.2 Decrypt the following messages encrypted with the Caesar cipher and an unknown key k , $0 < k < 26$:
 - (a) UNSJFUUQJ
 - (b) GUHAI