

Chapter 1. Introduction

This chapter introduces the basic aspects in establishing a secure communication system. Communication security includes information security and physical resource security. Logically, information is a kind of resources. However, in this book, we consider physical resources as information carriers. For example, a computer hard disk is the resources to store information. A cable is also the resources to transmit information. Radio frequency is resources too, over which, information is transmitted wirelessly. Therefore, communication security is the theory and technology to protect information and resources in an integrated way.

In this chapter, we will start from a general model of a communication system. Then we will introduce the basic information security objectives. Each of the mechanisms to achieve the objectives is designed based on two essential aspects, that is, trust assumptions and threat assessments. These two aspects are described as trust model and threat model in Section 1.3 and Section 1.4 respectively. At the end of this chapter, we will tell what a secure communication system means through a group of basic components.

1 A Communication System

A communication system can be described as a set of nodes connected with links. Information can be processed and stored inside a node and transmitted from one node to another through the links. Figure 1 is an abstract of a communication system.

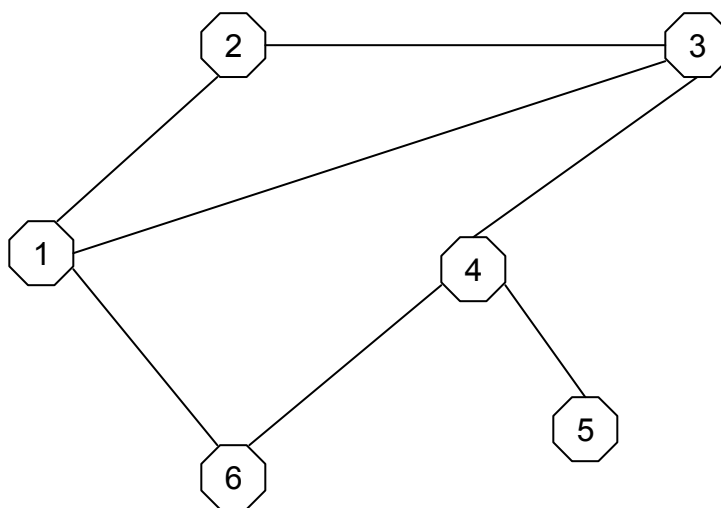


Figure 1: A Communication System

A node may consist of hardware, system software, and application software. It has capability of information storage, procession, and transmission. Practically, a node can either be a terminal, like a personal computer, or a network entity, like an internet router. The information processing may convert the information representation from one format to another. For example, it may transform IP packet to other data format. It may also apply protections to the information to be transmitted or stored. For example, encryption operation provides confidentiality.

A link connects two nodes where information is transmitted from one node to another. On each link, the transmission is conducted through a certain media, e.g. radio wave, or physical materials, like copper wires. The information is represented and interpreted through different formats based on different rules when transmitted. These formats and rules are different transport protocols. The different protocols can be described through a layered architecture. Figure 2 is an example of standard layered architecture, called Open System Interconnection (OSI) model, introduced by International Standards Organization (ISO).

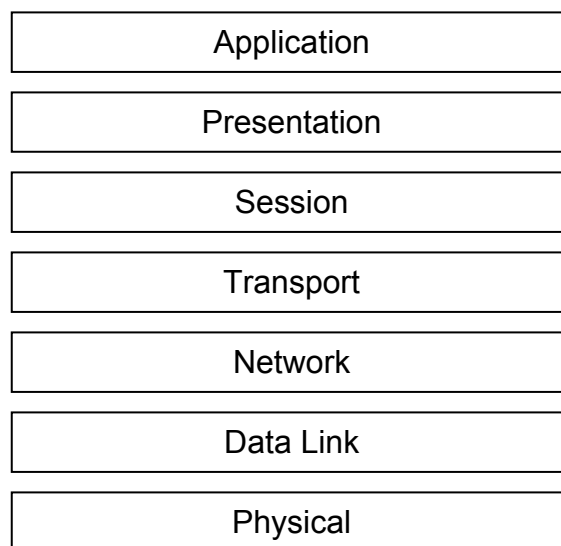


Figure 2: Layered Architecture in OSI Model

In most of the cases, security protection has to be applied together with an actual communication protocol, for example, Internet Protocol (IP). In this book, we employ layered architecture to describe security schemes. However, in some of our discussions, we may introduce a layer not included in what presented in Figure 2 to describe a specific protocol or a security scheme.

In this chapter, the security architecture is defined in terms of nodes and links. For information transmission from one node to another, it may pass several nodes. For example, in Figure 1, information transmitted from node 1 to node 5 may pass nodes 3 and 4 or nodes 6 and 4. Each combination of the links between two nodes can be called a *path*. A link could be bidirectional or unidirectional, so is a path.

The information protection when transmitted through a path may be applied in a link-by-link or an end-to-end fashion. We will explain the term “link-by-link” and “end-to-end” in the next section when we introduce protection mechanisms.

In each of the links, the security protection may be applied together with a communication protocol. If we use the OSI model, then the protection could be applied at data link layer, network layer, transport layer, or application layer. In this book, we will introduce protection mechanisms applied to different layers.

2 Information Security Objectives and Protection Mechanisms

The information security objectives described in this section includes confidentiality, integrity, and authenticity. To achieve each of the basic security objectives, it may employ one or more cryptographic functions. In this section, we will not introduce formal definitions and mathematical descriptions on each cryptographic algorithms used to achieve the security protections. Please refer to Appendix A for cryptography algorithms. In this section, each security protection is provided through what called mechanism without specifying which cryptographic algorithms are used. However, we will get into general descriptions on different classes of cryptographic algorithms.

2.1 Confidentiality

Confidentiality is to protect information from accessing by non-eligible parties. The confidentiality is achieved through encryption mechanisms. An encryption mechanism is a cryptographic transformation from plaintext P to ciphertext C such that it is computationally infeasible to inverse the transformation, called decryption, unless using the cryptographic key. An encryption algorithm can be symmetric key based or public key based.

For symmetric key based algorithm, encryption and decryption will use the same key. For example, if node I is to transmit information to node J with confidentiality, they must share a key K_{IJ} . The information will be encrypted with K_{IJ} when node I prepares the message. The node J will decrypt with K_{IJ} when receiving the ciphered message to recover the information. Let's use E to denote an encryption function, and D the corresponding decryption function. Then $C = E(K_{IJ}, P)$ and $P = D(K_{IJ}, C)$. The symmetric key based encryption algorithm is illustrated in Figure 3.

For a public key based algorithm, a pair of keys are involved. One key is called a *public key*, denote as Pk and another is called a *private key*, denote as Sk . The theory is called *public key cryptography*. It is a young science branch and invented in 1976. Most of public key cryptography schemes are based on mathematically hard problems. The hardness is measured by computing complexity theory. The examples of the hard problems, based on which cryptography algorithms are designed, include integer factorization and discrete logarithm. The hardness of the problem will make it computationally infeasible to obtain the private key Sk from the public key Pk . Here, being infeasible means by current computing capabilities when the problem is in certain sizes, it is

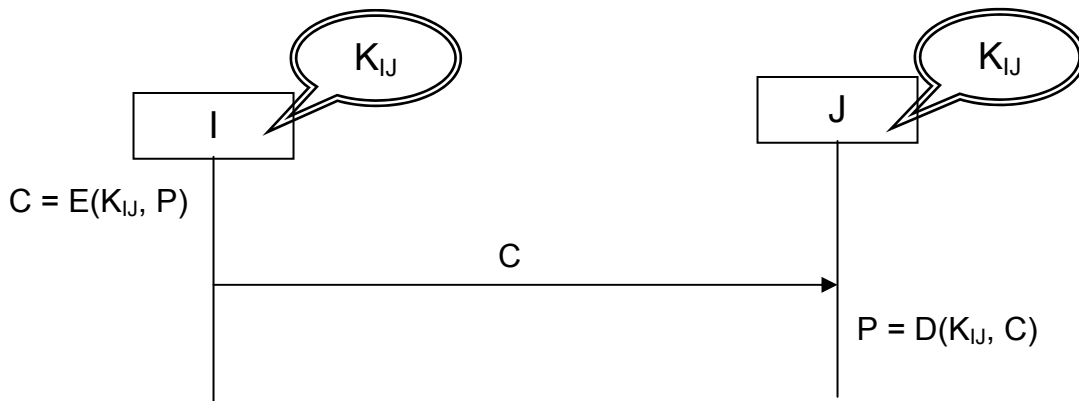


Figure 3: Symmetric Key Based Encryption

not realistic to find a solution. The current size to make integer factorization and discrete logarithm hard is based on an assumption that quantum computers are still unavailable.

In a public key based encryption algorithm, the public key is used for encryption and private key is used for decryption. In order to communicate to node J with confidentiality, node I will use node J 's public key Pk_J to encrypt the information, that is $C = E(Pk_J, P)$. Once node J receives the encrypted message, it will use its own private key Sk_J to decrypt the information such that $P = D(Sk_J, C)$. A public key based algorithm is also called asymmetric key based algorithm. The asymmetric (public) key based encryption algorithm is illustrated in Figure 4.

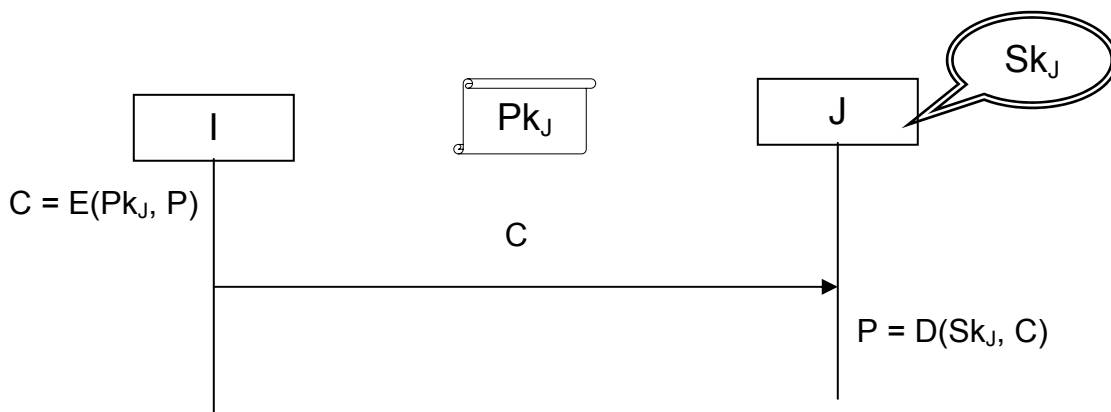


Figure 4: Asymmetric Key Based Encryption

Compared with symmetric key based encryption, public key based encryption does not require to distribute the key K_{IJ} in a protected manner. We will get into key distribution in Chapter 2.

Encryption can provide confidentiality independent to physical security in the sense that even the message is intercepted on the link from node I to node J , the interceptor will not be able to

get the information unless it can access the key, K_{IJ} in symmetric key case, and J 's private key Sk_J in public key case.

In this section, we have introduced confidentiality as one of the basic information security objectives. It can be achieved through encryption. In other words, encryption is considered as a protection mechanism. Now it is a good time to explain the protection mode we have mentioned in the previous section. That is, protection can be applied link-by-link or end-to-end. We will use the path in Figure 1 from node 6 to node 5 as an example, where the path consists of the link between node 6 and node 4 and the link between node 4 to node 5. In the following we will use symmetric key based encryption to illustrate.

For link-by-link encryption, information is encrypted at node 6 and decrypted at node 4 with a shared key K_{64} . Then it will be encrypted again at node 4 with key K_{45} and decrypted at node 5 with the same key. In this case, we assume that node 4 decrypts the information and encrypts again. We also assume that different keys are shared for node 6 and node 4 and for node 4 and node 5.

For end-to-end encryption, node 6 will encrypt the information with a key shared with node 5, say, K_{65} . The encrypted message will not be decrypted at node 4, but at its destination node 5, assuming that node 6 knows the final destination and also shares the key with node 5.

Here we will not go deeper on the security and practical implications of the different protection modes. However, we like to point out that which mode is adopted depends not only on its security implications but also on the practical communication protocols. In some of the protection protocols we will introduce later, the different protection modes will be further discussed.

2.2 Integrity and Authenticity

Integrity and authenticity are two inseparable information security objectives. Integrity is to guarantee that the information received is the same as it is sent. Authenticity is to guarantee that the originator appearing to the receiver is its actual originator. In other words, integrity is to prevent from altering the message content, while authenticity is to prevent from altering the message source. Therefore, it is often to use a single mechanism to achieve both integrity and authenticity.

With symmetric key based cryptography method, integrity and authenticity can be achieved by message authentication code (MAC). For a message M , message authentication code is a data tag, calculated with M and a key K . For example, if node I will send a message M to node J , then it will input M and key K_{IJ} to a MAC function. The output is a tag. That is, $tag = MAC(K_{IJ}, M)$. The tag will be sent together with the message M . When node J receives a message M' and tag , it uses M' and K_{IJ} to compute $tag' = MAC(K_{IJ}, M')$ and compare it with the tag attached to the message M' . If they are identical, then J can conclude that M' is not altered as transmitted and M' is indeed from I . That is, it has verified both integrity and authenticity of the received message M' . Figure 5 illustrates using symmetric key based message authentication code to provide authenticity and integrity.

With public key based cryptographic method, integrity and authenticity can be achieved by

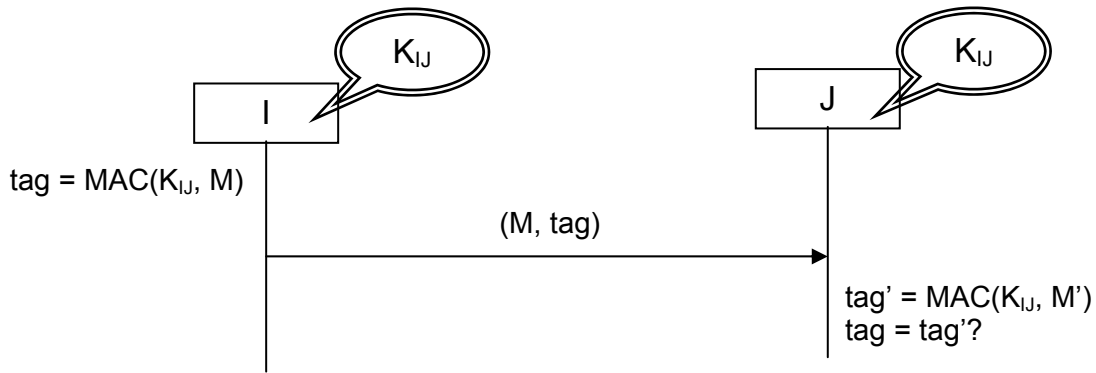


Figure 5: Message Authentication Code

digital signatures. Similar to message authentication code, a digital signature can be considered as a tag as well. Differently from the message authentication code, the node I will generate a digital signature with its private key Sk_I as $Sig_I(M) = Sig(Sk_I, M)$, which can be verified with I 's public key Pk_I at the receiving end, with input $Sig_I(M)$ and node I 's public key Pk_I . A verification function will output either valid or invalid. Notice that any one can verify the signature $Sig_I(M)$ since I 's public key Pk_I is public. Figure 6 illustrates using asymmetric key based digital signature to achieve authenticity and integrity.

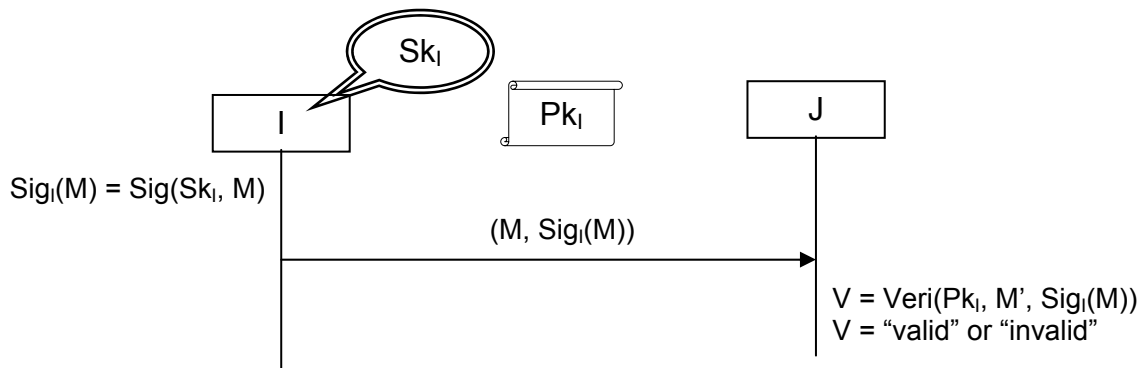


Figure 6: Digital Signature

With either message authentication code or digital signature, integrity and authenticity can be guaranteed without physical security in the sense that an interceptor may modify the message and message source but cannot generate a valid message authentication code or a valid digital signature on the behalf of node I unless it can access K_{IJ} in symmetric key case, assuming node J is the intended receiver, or the sender's private key Sk_I in public key case.

We will not further elaborate the difference between symmetric key based message authentication code and public key based digital signature. However, it is worth to notice that symmetric key based message authentication code can only be verified by an assigned receiver who shares the key.

Furthermore, it cannot provide non-repudiation, since receiver, once sharing the key, can generate MAC tag for any message on the behalf of the sender. A public key based digital signature can be verified by any receivers. It can also provide non-repudiation since the private key owner is the only one who can generate the valid digital signatures.

3 Trust Model

It may sound ironical to discuss trust here since all the needs for security protections are raised by distrust. For instance, we apply encryption mechanisms, because we cannot trust the non-eligible parties to access the information. However, we must assume, often implicitly though, that we trust the two nodes to keep the cryptographic key secret. Any security mechanism is based on some, explicitly or implicitly, assumed trustiness. This section will introduce main aspects in a trust model.

Generally speaking, a trust model is to define trust relations among different parties. A trust relation can be defined for any two parties with a mutually equal trust or an asymmetric trust. It can also be defined on a set of parties with hierarchical trust. For each security mechanism, the existed trust relationships can provide support to establish the new trust relationships. This service is called infrastructure support to the system. For example, when a symmetric key based encryption algorithm is used, it may demand a trusted key distributor to distribute the symmetric key K_{IJ} to node I and J in a protected manner. For public key based cryptographic functions, it is very important to trust the binding between a public key and its owner identity. Otherwise, if an interceptor E generates a pair of keys Pk_E, Sk_E , then E can claim that Pk_E is node I 's public key and therefore forge signatures on behalf of I . In many practices, a party called *certificate authority* will be trusted by the other parties to bind a public key with its owner. We will discuss certificate authority and other infrastructure support in Chapter 2.

In cryptography and security research, for a protection mechanism, the trust model is a set of well defined assumptions, based on which the security and robustness of the mechanism can be proved theoretically. However, for security practice in a communication system, a trust model is also a practical concept. It is often, if not always, determined by business relations. For example, a cellular service provider may play a role of trusted party for its subscribers in the sense that the service provider can hold the subscribers' long term keys. The cryptographic keys used to protect the user traffic are derived from the long term keys and therefore, accessible by the service provider.

A trust model may include assumptions on the physical environment. For example, we may trust a server located in a company's building rather than a wireless access point installed in a rest area of freeway. The physical environment can be a part of threat model, which will be discussed in the next section.

A trust model is crucial in establishing a secure communication system. It can go wrong in many ways. As a result, a lot of security holes are caused by ambiguous assumptions. It is, in general, not an easy task to explicitly and exclusively define a trust model. One reason, among a lot of others, is that many relations may co-exist in a large system. A trust relation for one

communication scenario may not maintain for another scenario. As a result, a lot of assumptions have never been explicitly defined but mistakenly assumed by the system designers.

In this book, we will demonstrate how to define and establish trust model for security mechanisms to be introduced when it is possible. We will also point out some possible pitfalls on the trust with each mechanism.

4 Threat Model

In the famous “Art of War” by Sun Tzu, it was said that “the enemy, military is invincible.” Securing a communication system is like a war. For each protection mechanism, we have to know which kind of attacks the mechanism is up to against. In this section, we will discuss threat model for security mechanisms.

As we have discussed, the information security objectives can be achieved with application of cryptographic algorithms. The strength of each cryptographic algorithm is defined as the computing complexity to break the algorithm. Therefore, for a threat model, we must have an estimation on the processing and communicating capacity of an attacker.

The physical threat must be a part of threat model. For example, how easy it can break in a node without destroying it. On the other hand, whether a link is wired or wireless also makes difference in a threat model.

A threat model often includes a set of possible attacks to a communication system. For example, *man-in-the-middle attack* is often an attack that a security communication protocol must be able to against. However, when assessing the threat of a man-in-the-middle attack, whether an attacker can actively intercept communications in real time is an important factor. An attacker E can launch an *active man-in-the-middle attack* if E can intercept the message and modify it before forwarding to its destination. This attack is more serious than simply observing the message flow between two nodes, which is called *passive man-in-the-middle attack*. In some protocols, an attacker can launch active man-in-the-middle attack on the message in both directions. However, in some other protocols, active man-in-the-middle attack can only be launched on the messages in one direction. In Figure 7 illustrate an active man-in-the-middle attack in two directions.

A threat model should be defined for each specific communication protocol so that proper protection mechanisms can be applied. For example, for a communication protocol, if an active man-in-middle attack is possible, then messages should be authenticated through message authentication code or digital signatures.

5 Security Components

This section will informally introduce the security components for a communication system. The purpose of this section is to understand what are on our “wish list” before going into a formal discussion on how to establish the components.

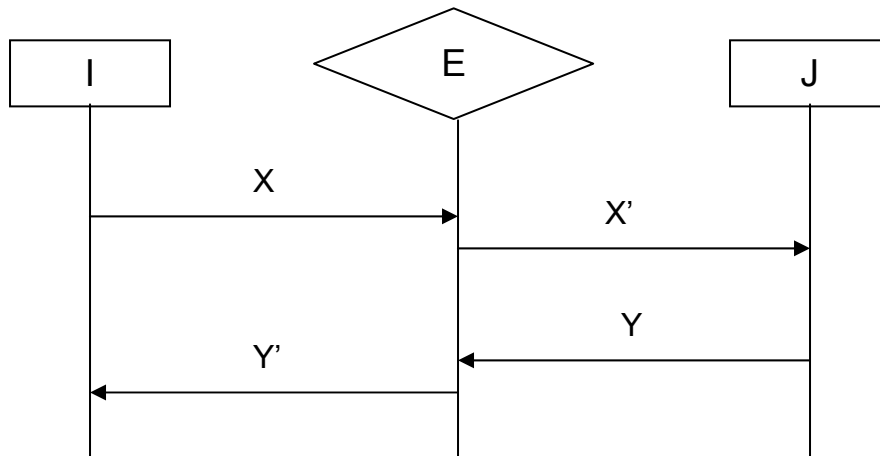


Figure 7: Active Man-in-the-middle Attack

As we described in Section 1.1, a communication system is considered as a set of nodes and links which connect nodes. The security components for a communication system will include security components to make each node a trusted platform and also to protect communications on each link.

5.1 Trusted Platform

In this section, platform should be understood as a generic term for a node. Logically, platform is a relative concept. For example, for a personal computer, relative to operating system (OS) and system software, hardware processors and memories will form a platform. However, for applications, the hardware and system software are integrated to a platform.

Generally, a trusted platform is a platform to operate as it is supposed to. For example, a trusted platform shall not bypass an execution of encryption function before transmitting a message, if it is supposed to encrypt it. Another example is that a trusted platform should allow a process accessing a file unless the process is entitled to. A trusted platform should include but not limited to the following components.

- **Robust hardware:** Hardware is the basis for providing a trusted platform. It should be able to detect and respond to tamper or intrusion. One of the effective responses to tamper and intrusion is to delete the sensitive data in use on the platform. Robust hardware should also be able to shelter physical characters of execution from observing. For example, it should hinder observation on power consumption variables for cryptographic operations to prevent from obtaining useful information for the cryptanalysis.
- **Validated system software:** System software, for example, operating system, includes

mechanisms to execute security enhancements. It must be able to validate the status of system software. In other words, all the security enhancements will function in the same way as they are supposed to. When access control is executed by system software, validated system software can assure that the access control policy is executed in the same way as it is specified.

- **Authenticated applications:** Applications should be authenticated before installed and executed to prevent harmful applications, which are either maliciously installed or poorly designed, from weakening the security.

For a given node, the effort to satisfy these security requirements varies. If the node is a high end network device, then its physical environment may be well controlled so that it is not easy for an attacker to get close to the hardware. Its software installation and execution may also be restricted to the assigned system management. However, if a node is a user portable device, like a personal computer or a mobile phone, then it is likely for attackers to get the hands on it. The open platform will allow downloading applications from all kinds of suspicious web sites. Push type of service can easily distribute worms and virus to a large scale of devices to form an attack on network entities. We will get into the theory and technologies for trusted platform in Chapter 7.

5.2 Protected Communications

The protected communications are the communications with one or all the security properties, confidentiality, integrity, and authenticity. As we discussed in Section 1.2, these security properties can be achieved through cryptographic functions. In order to apply cryptographic functions, the cryptographic keys need to be established between two nodes. This section will introduce some basic procedures to establish and conduct protected communications between two nodes.

- **Mutual authentication:** Each of the nodes must be assured with whom it is to communicate. The entity authentication will be introduced in Chapter 2. Here entity authentication is understood as a procedure to make sure that the entity is actually the same as it is claimed to be.
- **Key establishment:** The keys used for protection mechanisms, for example, encryption and message authentication, must be established between the two nodes. If the keys are established between two mutually authenticated nodes, then each node will be assured with whom the keys are established. For each key, both nodes must agree on its usage, that is, for which cryptographic function and used in which algorithm. They may also agree on the key life time for each key so that the key shall not be used after it is expired.
- **Protected negotiation:** The two nodes may need to negotiate which mechanisms will be applied to the communications, for example, encryption, message authentication, etc.. It will also negotiate that for each mechanism, which algorithms are used. Usually, the negotiation

will select from a set of algorithms, among which some algorithms may weaker than others. In most of the cases, the negotiation is to make sure that both nodes have implemented the selected algorithms. The negotiation should be authenticated so that an attacker can degrade the security level by forcing them to choose weaker algorithms.

- **Failure detection:** Once the protected communication starts, each node shall be able to detect failures for the protection. That is, if one of the nodes fails to apply the agreed protection, then the another node shall detect the failure and respond properly.

For the protected communications, the property of the communication protocols will affect the security. For example, some of the protocols are session based. That is, once the session is established, the two nodes will be dedicated to the protected communications. However, some of the communication protocols are not session based. That is, unprotected communications may happen alternatively between the same two nodes as the protected communications.

As we discussed in the beginning of this chapter, in a communication system, two nodes might be connected by more than one link. The information flow will go through some other nodes before arriving to the final destination. In this case, the protection can be established in a link-by-link or an end-to-end fashion. The protection can also be applied to different layers, depending on communication protocols. For communication protection, wired and wireless links will be significantly different. We will discuss protections for wireless access network in Chapter 5.

At the end of this section, we like to point out that trusted communications are protected communications. But protected communication may not be trusted communications. In Chapter 3, we will discuss how to establish trusted communications between two nodes.