

Chapter 1. Introduction

As Thomas Friedman described, the world becomes flat. Everything is changing to fit into a flattened world, so does the way to present the science and technologies. In the old days, each concept and terminology should be explained individually before introducing their relations. However, in today's world, people may have heard the terms in different contexts, and may have googled and checked them out from Wikipedia. What they really need is the idea, the logic, and the relationship of the individual terms.

This book will focus on the relationships of the basic concepts and functions. Based on this idea, we will first lay a real size blueprint on the ground, before we start to dig in at any specific point.

In this chapter, you may find that some of the concepts have not been defined as used. In that case, use your own intuitive to follow the logic flow. Do not worry about non-precision. The final purpose is to understand the big picture precisely.

This chapter introduces the basic aspects in establishing a secure communication system. Communication security includes information security and physical resource security. Logically, information is resources. Physical resources serve 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. This chapter tells what a secure communication system means in a general term.

1 Security Architecture

A communication system can be described as a set of nodes connected with links. Information can be processed or stored inside a node and transmitted from one node to another. Security architecture defines trust relationships among the nodes and protection mechanisms for the information processed, stored, and transmitted. Figure 1 is an abstract of 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. The transmission is conducted through a certain media, like radio, or physical materials, like copper wires to deliver the information from one node to another.

In most of the cases, security protection has to be applied together with an actual communication protocol, for example, Internet Protocol (IP). We employ layered architecture to describe security schemes. Figure 2 is an example of standard layered architecture, called Open System Interconnection (OSI) model, introduced by International Standards Organization (ISO). But in some

⁰Copyright ©2008 L.D. Chen and G. Gong. All rights reserved. May be freely reproduced for educational or personal use.

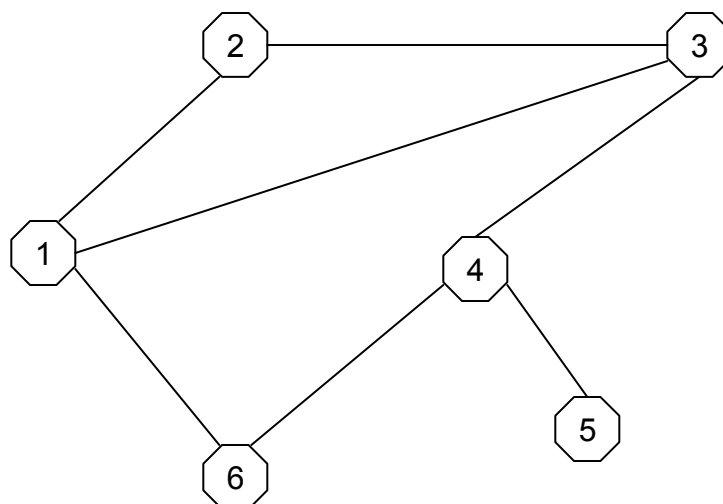


Figure 1: A Communication System

of our discussions, we may introduce some other layered architecture to describe certain security schemes.

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, from node 1 to node 5 may pass nodes 3 and 4 or 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-to-link fashion. That is, both nodes in a link will execute protection mechanisms. We will explain the term "link-to-link" in the next section when we introduce protection mechanisms.

For each of the links, the security protection may be applied at different layers. If we use OSI model, then it could be applied at data link layer, network layer, transport layer, or application layer. Therefore, the security protections will be described and introduced in a link-to-link and layer-by-layer manner.

2 Basic Information Security Concepts and Protection Mechanisms

The information security concepts described in this section includes confidentiality, integrity, and authenticity. To provide each of the basic security properties, it may employ one or more cryptographic functions. In this section, we will not introduce formal definitions and mathematical descriptions on each cryptographic functions used to achieve the security protections. Please refer to Appendix A for cryptography functions.

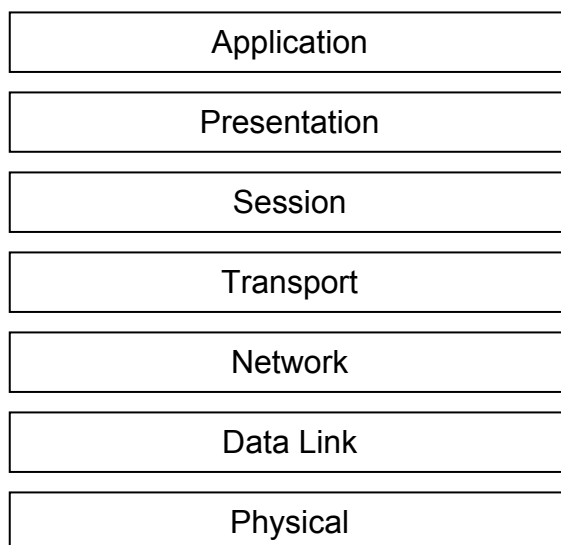


Figure 2: Layered Architecture in OSI Model

2.1 Confidentiality

Confidentiality is to protect information from accessing by non-eligible parties. The confidentiality is often provided by encryption. Encryption 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 need to provide confidentiality to the communication with node J , 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. Denote E as 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 public key based algorithm, for each communication parties, a pair of keys is involved. One key is called public key, denote as Pk and another is called private key, denote as Sk . The theory is called public key cryptography. It is a young science branch invented in 1976. Most of public key cryptography schemes are based on mathematically hard problems based on computing complexity theory, for example, 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. Here, being infeasible means by current computing capabilities when the problem is in certain sizes, it is not realistic to accomplish. Please notice that whether a given problem is hard depends on the assumption about the computing capacity. As aforementioned examples, integer factorization and discrete logarithm, the hardness is based on an assumption that quantum computers are still unavailable.

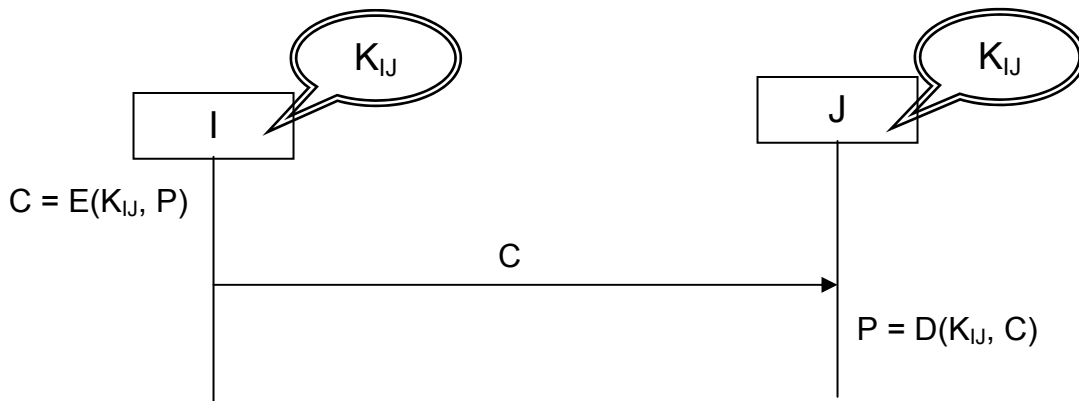


Figure 3: Symmetric Key Based Encryption

If it is hard to compute the private key from a public key, the public key is used for encryption and private key is used for decryption. With the above example, in order to communicate to node J , 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 $P = D(Sk_J, C)$. The Asymmetric (public) key based encryption algorithm is illustrated in Figure 4.

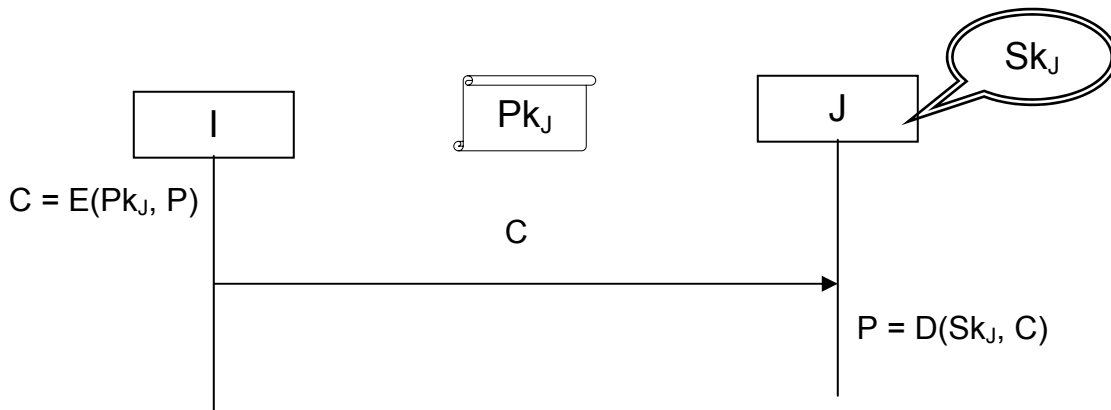


Figure 4: Asymmetric Key Based Encryption

Compared with symmetric key based encryption, public key based encryption will remove the requirement of distributing 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, either K_{IJ} in symmetric key case, or J 's private key Sk_J in public key case.

In this section, we have introduced confidentiality as a basic information security concept. It

can be achieved through encryption. In other words, encryption is considered as a protection mechanism. Now it is a good chance to further explain what we have mentioned in the previous section, that is, link-to-link or end-to-end, protection. We will use the path in Figure 1 from node 6 to node 5 as an example, where the path consists of link between node 6 and node 4, and then between node 4 to node 5.

For link-to-link encryption, information may be encrypted at node 6 and decrypted at node 4 with a shared key K_{64} in symmetric key situation. 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 can decrypt the information and encrypt again. We also assume that different keys are shared between node 6 and node 4 and between 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.

2.2 Integrity and Authenticity

Integrity and authenticity are two inseparable features in communication. Integrity is to guarantee that the information received is the same as it is sent. Authenticity is to guarantee that the originator appears 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.

These two are inseparable. It is often to use a single cryptographic function to provide both.

With symmetric key based cryptography method, integrity and authenticity can be provided 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 provided by digital signatures. Similar to message authentication code, a digital signature can be considered as a tag. 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 provide authenticity and integrity.

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 by intercepting and then forwarding, but cannot generate a valid message

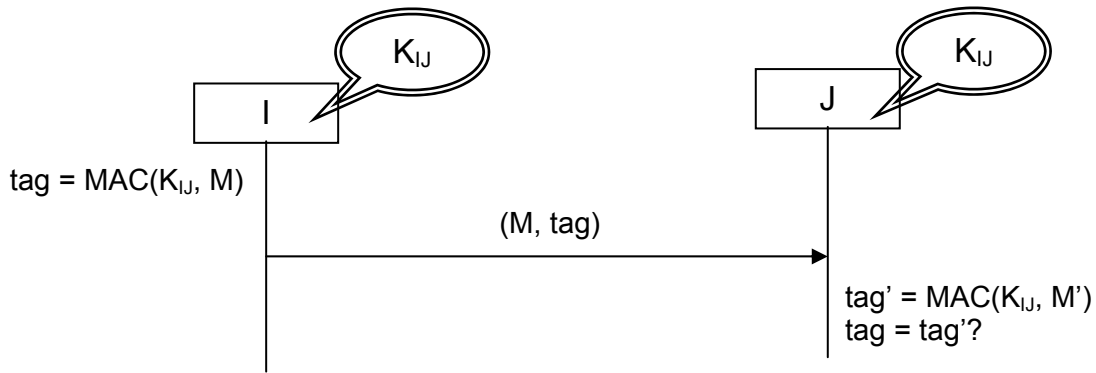


Figure 5: Message Authentication Code

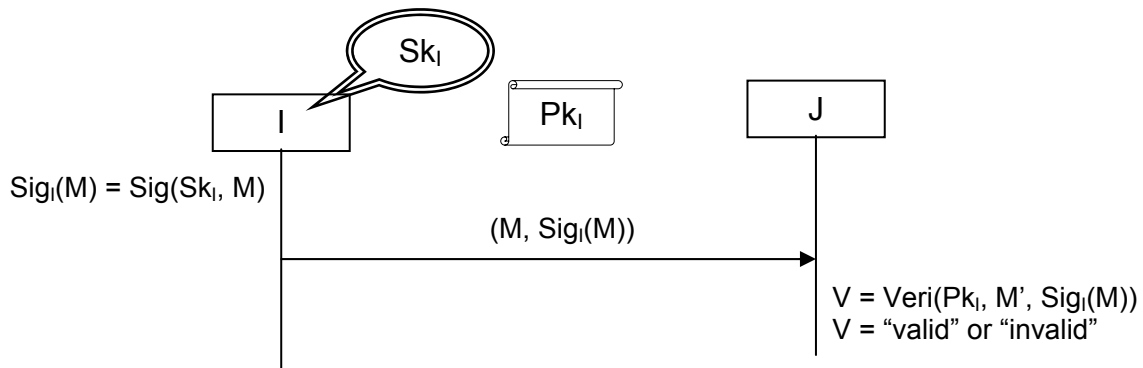


Figure 6: Digital Signature

authentication code or a valid digital signature unless it can access K_{IJ} in symmetric key schemes, assuming node J is the intended receiver, or the intended sender's private key Sk_I in public key situation.

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 beside protected key distribution requirement, symmetric key based message authentication code can only be verified by a single receiver. Furthermore, it cannot provide non-repudiation, since receiver can generate MAC tag for any message on the behalf of the sender. For public key based digital signature, it can be verified by multiple 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 are raised by distrust. For instance, we provide confidentiality because we cannot trust the non-eligible parties to access the information. However, we must implicitly assume 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 a asymmetric trust. It can also be defined on a set of parties with hierarchical trust. Some trust relations are assumptions for the security mechanisms, while the others are to be established through applying security mechanisms.

A trust model shall include security infrastructure support to the system. Informally, infrastructure support is to provide certain service to establish trust relation for other parties. For example, in Section 1.2, we introduced public key based cryptographic functions. For each such function, it is very important to trust the binding between a public key and its owner. 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.

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 cryptographic keys for authenticating subscribers.

A trust model may include assumptions on the physical environment. For example, we may trust a server located in a company's building more than a wireless access point installed in a rest area of highway. 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 and result in security holes. 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. Sometimes, the new relations are established partially on the old ones. A trust relation for one communication scenario may not still maintain the same for another scenario. As a result, a lot of assumptions have never been explicitly defined but mistakenly assumed by the system designers.

In this lecture, we will demonstrate how to define and establish trust model for each security mechanisms 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. We have to know which kind of attacks we are up to against. In this section, we will outline a threat model.

As we have discussed, the basic security properties can be achieved with cryptographic functions. The strength of each cryptographic function is defined in computing complexity measurement. That is, the strength is defined as the complexity to attack the function. Therefore, for a threat model, we will have to understand 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.

It also makes difference if an attacker can actively intercept communications between different nodes in real time. Therefore, it shall include the capability or possibility for an attacker to access

the communications to conduct a man-in-the middle attack. Figure 7 illustrate a man-in-the-middle attack between two nodes *I* and *J*.

In Figure 7, if the middle man *E* can actively intercept the message *X* and modify it to *X'* before forwarding to its destination. It can do the same on another direction, that is, intercept the message *Y* and modify it to *Y'* before forwarding, then the middle man *E* can manipulate the protocol between *I* and *J* and cheat either of them or both. This can lead to a more serious attack than simply observing the message flow between node *I* and node *J*, which is called passive interception.

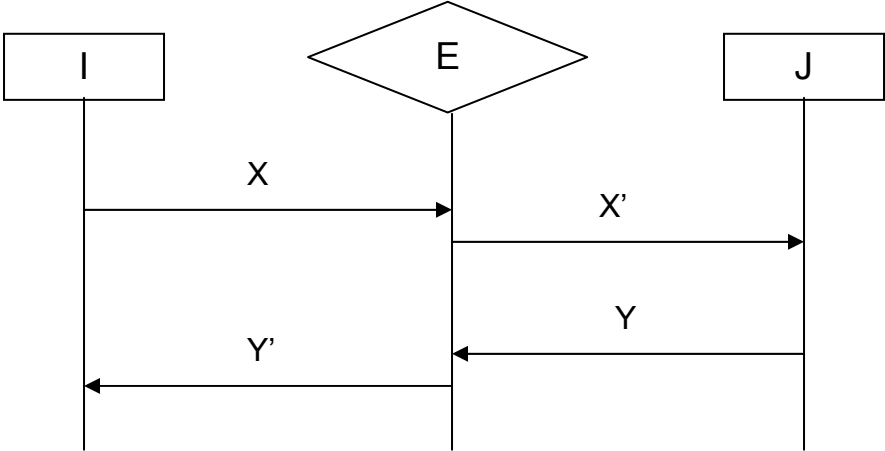


Figure 7: Active Man-in-the-middle Attack

However, whether the man-in-the-middle can conduct a successful attack depends on different factors and attackers' intentions. For example, when messages *X* and *Y* are authenticated, then the middle man, in order to conceive any of them, must have the capability to forge a message authentication codes or valid digital signatures in real time, which is unlikely without knowing the key. However, if the attacker *E* is only to disrupt the protocol, then it is easy to make it fail of the verification.

From the above example, we can see that a threat model should be defined for each specific communication protocol. In each of the specific case, the threat model shall include the capability to conduct active online attack like described in Figure 3 or passive offline attack by observing the communications between the nodes.

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 obtain the components.

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 will be 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 not issue an access to a file if the process is not 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 response tamper or intrusion. For example, it should be able to delete sensitive data, when a tamper is detected. 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 cryptanalysis based on the amount of consumed power[? Differential Power analysis].
- **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. For example, access control may be executed by system software. Validated system software can insure that the access control policy is executed in the same way as it is specified.
- **Authenticated applications** - Applications should be authenticated before 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 personal. 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 weird or even suspicious web sites. Furthermore, push type of service can easily distribute worms and virus to a large scale of devices to form an attack on network entities.

Therefore, our discussion on trusted platform will focus on terminals with wireless interface. We will get into details in Chapter 6.

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 requirement to establish and conduct protected communications between two nodes.

- **Mutual authentication** - Each of the nodes must be insured with whom to communicate. The entity authentication will be introduced in Chapter 2. Here we just understand entity authentication as an insurance that the entity is actually the same as it is claimed to be.
- **Key establishment** - The key establishment must be authenticated so that each node will know with whom the keys are established. For each key, both nodes must agree on its usage, that is, for which cryptographic function and with which algorithm. They may also agree on the key life time so that the key shall not be used after it is expired.
- **Protected negotiation** - The two nodes need to negotiate which mechanisms will be applied to the communications, for example, encryption, authentication, etc. It will also negotiate that for each mechanism, which algorithms are used. The negotiation should be authenticated so that none of the nodes or an attacker can degrade the security level.
- **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 is fail to apply the agreed protection, then the another node shall detect the failure and response 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 either established in a link-to-link or an end-to-end fashion. The protection can also be applied to different layers, depending on communication protocols. In Chapter 3, we will discuss the protections for different communication protocols.

For communication protection, wired and wireless links will be significantly different. We will discuss protections for wireless access network in Chapter 5.