9 Biometrics Controls for Security

Learning Objectives

After completing this chapter you will be able to:

- understand what biometrics is all about.
- learn about various biometrics measurement techniques.
- appreciate how biometrics is used for user identification and authentication.
- learn about the biometrics matching and enrolment process.
- understand the key success factors for implementation of biometrics systems.
- appreciate the benefits offered by biometrics.
- get glimpses of issues and challenges involved in biometrics in order to get ready for the discussion in Chapter 10 focusing on this (issues and challenges in biometrics).

9.1 Introduction

Biometrics is the science for determining a person's identity (ID) by measuring his/her physiological characteristics. *Authentication* is a fundamental concept in security, especially with respect to human–computer interaction. In this chapter, biometrics methods are discussed. We start with the basics of biometrics to understand what it is, its place in user authentication for physical access control and various technologies and techniques used in biometrics. Issues and challenges in implementing a biometrics system will be discussed in the next chapter.

Biometrics has got interesting origin in the Chinese civilization. The earliest known use of biometrics dates back to the seventh century during China's Tang Dynasty. During this period fingerprints were used to sign and validate contracts. Over the last century, biometrics has grown enormously. Technologies are being developed to verify or identify individuals on the basis of measurements of the face, hand geometry, iris, retina, finger, ear, voice, speech, signature, lip motion, skin reflectance, deoxyribonucleic acid (DNA) and even body odor (readers are encouraged to review the IEEE paper by Jain & Prabhakar, 2004).

Biometrics techniques of today have been made possible by the advances in computing technology and the need that arises owing to universal presence and connectivity of computers all over the world. *Biometrics identification* is a much more sophisticated method of controlling access to computing facilities than badge readers (we discussed those in Chapter 8); however, the two methods operate in the same way. *Biometrics techniques* used for user identification typically include *fingerprint recognition*, *palm recognition*, *handprint recognition*, *voice pattern recognition*, *signature samples*, *retinal scans* and *iris scans*.

Biometrics provides a higher level of security than badges because it cannot be lost, stolen or shared. Thus, biometrics can provide a greater degree of security than traditional authentication methods; however, these methods are expensive as well as complex to deploy. Given this, as of now, biometrics identification techniques are suitable only for high-security, low-traffic entrance control for physical access. In this chapter, we explore the latest advances in biometrics.

9.2 Access Control, User Identification and User Authentication

These terms are important in the discussion of biometrics and therefore we first deal with them before proceeding with the rest of this chapter. *Access control* refers to the procedures and mechanisms used either to restrict entry into the premises where something confidential is stored – for example, the premises wherein information systems (IS)/computing facilities are hosed – or to restrict entry to the computing device, or to software and/or data within the computer and to those persons authorized to use such resources. In this context, *identification* and *authentication* of users are important for information systems security.

User identification refers to the action of the user claiming his/her ID when communicating with a device. *Authentication* is the process of proving that the claimed ID is genuine. Thus, the proof of ID is a critical process in access control. It may take one of the following three types either individually or in combination:

1. something that the user 'knows' [password, personal identification number (PIN), etc.];
2. something the user 'possesses' (badge, smart card, etc.);
3. something the user 'is' (user's biological characteristics).

Biometrics concerns itself with the third type. It is something so unique to a person and embedded with the person that it cannot be lost, stolen or copied. Given the unique nature of human biometrics ID, biometrics methods occupy an important place in user identification/authentication.

9.3 What is Biometrics?

The term biometrics comes from the Greek words *bios* meaning life and *metrikos* meaning measure. It is well known that humans intuitively use some body characteristics such as face, gait or voice to recognize each other. Since, today, a wide variety of applications require reliable verification schemes to confirm the ID of an individual, recognizing humans on the basis of their body characteristics has become more and more interesting in emerging technology applications. Biometrics is used as one of the methods for physical access control. It is, basically, a collection of methods for identification based on measuring the physiological characteristics that are unique to each and every individual. Some examples of such characteristics are:

1. voice;
2. fingerprints;
3. body contours;
4. retina and iris;
5. handwriting style/handwritten signature;
6. gait (not as commonly used as the characteristics mentioned above).

Readers would like to note that gait is the peculiar way one walks and is a complex spatiotemporal biometrics. Biometrics experts say that gait is not supposed to be very 'distinctive', but is sufficiently 'discriminatory' to allow verification in some low-security applications. It is important to be aware that 'gait' is a behavioral biometrics and may not remain invariant, especially over a long period of time, owing to the fluctuations in body weight, major injuries involving joints or brain or inebriety. However, because acquisition of gait (i.e., capturing the movement/walking style of an individual) is similar to acquiring a facial picture, it may be an acceptable biometrics. Since gait-based systems use the video-sequence footage of a walking person (see Figure 9.1) to measure several different movements of each articulate joint, it is input-intensive and computationally expensive (Box 9.1).

Biometrics methods, in general, involve performing some human action for configuring a system used to recognize the physiological parameters of the ID (human entity) to be authenticated, for example, most often, this could be:

Figure 9.1 Biometrics identification through gait recognition.
Photo Credit: Georgia Tech Research Institute.

Box 9.1 Walk the Walk – Gait Advances in Emerging Biometrics

A famous sentence in William Shakespeare's work (*The Tempest*) says 'Great Juno comes; I know her by her gait.'

Recognizing people by the way they walk (their *gait*), the shape of their ears, the rhythm they make when they tap and the involuntary response of ears to sounds all have the potential to raise the stock of biometrics techniques. The characteristics of your walk may not be as distinctive as the swaggering of some famous personalities, but your stride may still be unique enough to identify you at a distance – alone or among a group of people.

Researchers at the Georgia Institute of Technology and elsewhere are developing technologies to recognize a person's walk, or gait. Results indicate that these new identification methods hold promise as tools in the war on terrorism and in medical diagnosis. *Gait recognition technology* is considered a biometrics method – that is, a unique biological or behavioral identification characteristic, such as a fingerprint or a face.

Though still in its infancy, the technology is growing in significance owing to the continuous research being carried out by many agencies around the world. Because gait recognition technology is relatively new, researchers are assessing the uniqueness of gait and the methods by which it can be evaluated.

One study at the Georgia Tech Research Institute (GTRI) is addressing the issues of gait recognition by computer vision, and the other is exploring a novel approach – gait recognition with a radar system similar to those used by police officers to catch speeders (for details, visit http://www.gtri.gatech.edu/). The ultimate goal is to detect, classify and identify humans at distances up to 500 ft away under day or night, all-weather conditions (see Figure 9.1). Such capabilities will enhance the protection of defense forces and facilities from terrorist attacks.

Courtesy: http://gtresearchnews.gatech.edu/newsrelease/GAIT.htm.

1. drawing a few signatures so that the system can analyze and record their characteristics/pattern;
2. looking into a scanning apparatus in order to record retinal patterns;
3. intoning words for the analysis and recording of voice patterns;
4. collecting multiple video shots of a person walking (gait acquisition).

With some success, these methods have been adapted to computerized identification. However, the complexity of measuring physical attributes, their variation with time and the cost of identification apparatus are some of the reasons why biometrics identification/authentication techniques are not so rampant as compared with the traditional authentication methods (discussed in the Chapter 8). Also, biometrics methods, sometimes, give rise to social acceptance issues as often they are looked upon suspiciously (e.g., people may feel that exposing their eyes to the retina scanner may damage their eyes!). Box 9.2 illustrates the multi-disciplinary nature of biometrics.

Box 9.2 Biometrics: A Potpourri of Multi-Disciplines

Biometrics technology, while being an integral part of law enforcement, is also critical to the rapidly growing suite of civilian applications, such as citizen ID cards, electronic passports (e-passports) and driver licenses. These systems not only need advanced biometrics technology interfaces but also require the ability to deal with security and privacy issues (see Lewis and Statham, 2004).

The integration of biometrics with access control mechanisms and information security (InfoSec) is another area of growing interest. The challenge to the research community is to develop integrated solutions that address the entire gamut of research problems from sensors and data acquisition to biometrics data analysis and systems design.

One important aspect of biometrics that has not been adequately addressed by the research community thus far is that of *large-scale applications*. For example, visa authorities must handle populations of the order of several millions of individuals on a regular basis. Such applications require biometrics systems that scale efficiently, going beyond limited laboratory experiments to real-world installations and are yet 'reliable' and 'accurate'.

There is a need for scientists and practitioners from the diverse fields of computing, sensor technologies, law enforcement and social sciences to share a single forum to exchange ideas, research challenges and results. Some interesting aspects are as follows:

1. **Biometrics modalities:** fingerprints, face [two-dimensional (2D)/three-dimensional (3D)/thermal], iris, speech, signature, hand geometry, hand vein/palm print, gait, retinal patterns, ear (2D/3D), etc.
2. **Biometrics systems:** multi-modal biometrics, information fusion in biometrics, indexing and identification in large-scale databases, novel hardware architectures, security/privacy concerns, biometrics cryptography and storage and security of biometrics templates on smart cards.
3. **Novel biometrics:** soft biometrics, chemical biometrics, skin spectroscopy, etc.
4. **Systems evaluation:** performance metrics and statistical confidence measures.
5. **Applications:** law enforcement, security, travel safety and border control, physical and virtual access control and data storage standards.

Courtesy:

1. IEEE Announcement for a conference on the theme of Transactions on Systems, Man and Cybernetics – Part B" to discuss Special Issue on Recent Advances in Biometrics Systems [unable to trace more details of this document].
2. Paper of Speech Processing and Biometrics Group Signal from Processing Institute Ecole Polytechnique Fédérale de Lausanne (EPFL) that was accessed at http://scgwww.epfl.ch/courses.
3. Biometrics for Secure Identity Verification: Trends and Developments, thesis work by Joseph W. Lewis, at the University of Maryland Bowie State University (Yr. 2002).
4. Various biometrics related papers and other materials sited at http://biometrics.cse.msu.edu/publications.html, http://www.biometrics.org/bc2004/program.htm and http://www.itl.nist.gov/iad/tools.html#MEASUREMENT (Biometrics measurement products and tools information)

Basically, all biometrics techniques are based on similar concepts and employ common features and functions, the most important being the *procedures for enrolment* and physical access to premises housing the computing facilities/computer systems [it is said that the techniques are developed whereby laptop users can use their thumb impression as the way of authentication to start the laptop operating systems (OS) on power-on, e.g., the SONY notebook series]. Usually, the identification system is used in conjunction with other information, such as a PIN. In this type of inspection for identification, a reference pattern that is known to be secure is compared with freshly recorded identification data. Enrolment and template preparation are important procedures in biometrics; they are discussed in a later section.

9.4 Nature of Biometrics Identification/Authentication Techniques

It is very important to understand the inherent nature of biometrics before proceeding further in this chapter. In the world of security, identification and authentication techniques have 'accuracy' implications (more on this is discussed in Section 9.7) that are based on 'probabilistic' phenomena. This means that another issue surrounding the topic of biometrics is that of 'certainty as probability' and there is a good reason for this. When an individual's claims of ID and privilege are verified in a truly reliable way, the identification is 'authoritative'. The practical value of any identification/authentication scheme, however, generally exists in one of the following three states:

1. certain and unambiguous (*deterministic*);
2. certain, based on a low probability of error (*probabilistic*);
3. uncertain and ambiguous and therefore (for all practical purposes) false.

Unfortunately, a biometrics attribute is not necessarily unambiguously permanent; therefore, all biometrics schemes are probabilistic. Design and implementation steps that can reduce the likelihood of an error are essential to orderly deployment of the technology. Biometrics techniques are most reliable and effective when used as an authenticating technique as part of a *multi-factor scenario*. For example, if an individual makes a claim of ID at the bank with his/her name, and that claim is supported (authenticated) by a biometrics identifier, then the probability of error is very low. Errors are much more likely to occur where the system must figure out the ID of an individual on its own (identify). This point is a crucial one to remember. Now let us understand the nature of *biometrics identification* and *biometrics authentication* (see Box 9.3).

Box 9.3 Biometrics Identifications and Biometrics Authentication

Biometrics Identification

Biometrics identification is a sophisticated variation on a token-based, *single-factor security scheme*. In this case, the token is some physical attribute of the person – fingerprint, iris, retina, face, vein pattern etc. (details are discussed in Section 9.5). Biometrics identification systems typically follow three high-level processing steps (refer to Figure 9.2). First, the system must 'acquire' an image of the attribute through an appropriate scanning technique.

Once the scanned content is acquired, it must be 'localized' for processing purposes. During this step, extraneous informational content is discarded and 'minutiae' are isolated and turned into a 'template', a sort of internal canonical form for matching attributes stored in a database (these terms are discussed in Section 9.6).

Minutiae are the uniquely differentiating characteristics of the biometrics attribute. Whorls and loops and their relationship to one another on a fingerprint are examples of the minutiae that might be extracted (see Figure 9.4). Finally, templates stored in the database are searched for a match with

Box 9.3 *Continued...*

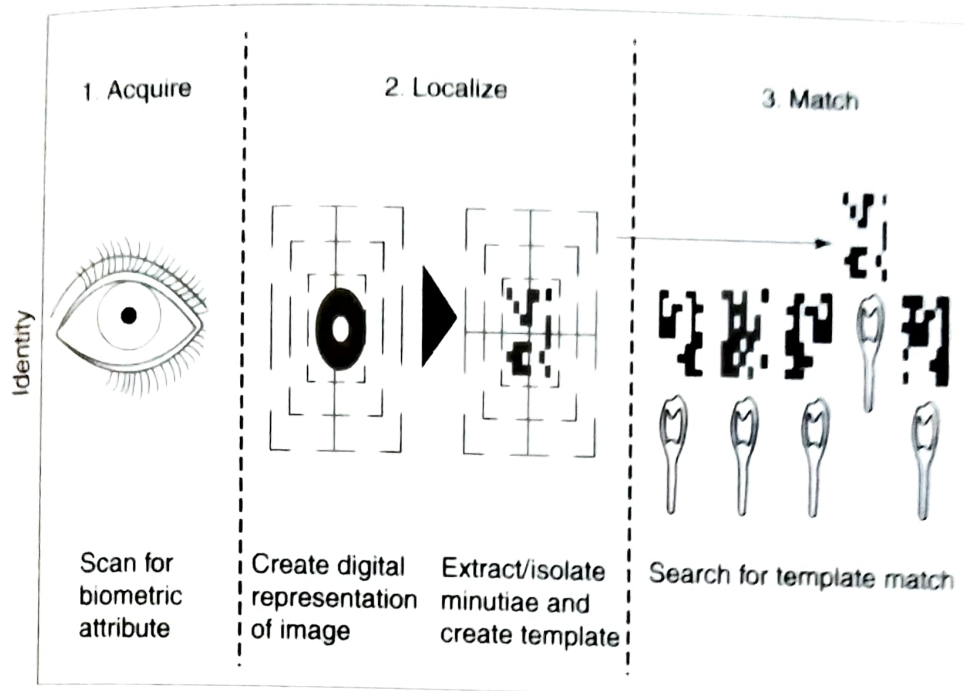


Figure 9.2 | Biometrics identification: acquisition, localization and matching.

the one just presented. If a match is found, the identification is a success and the succeeding steps of the security process can begin.

Biometrics Authentication

Biometrics authentication virtually eliminates the risk of anonymity in a 'two-factor security scenario by using a 'physical attribute' of the person to authenticate a token. *Two-factor authentication* is a security process in which the user provides *two means of identification*, one of which is typically a physical token, such as a card, and the other is typically something that can be memorized, such as a security code. In this context, the two factors involved are sometimes spoken of as *something you have* and *something you know*. Two-way authentication process is similar to biometrics identification. First, the requestor presents a token to assert the ID. For example, an automated teller machine (ATM) or credit card is inserted into a reader. A number encoded on the card is actually the token; the card is more like a container for the token, but treating the card as a token is appropriate. As with identification, the system must 'acquire' an image of the personal attribute. Second, the attribute must be 'localized', the minutiae extracted and a matching template created (see Figure 9.3).

Finally, the value of the token is used to look up the template previously stored for this individual. If it matches the template presented on this occasion, the requestor is authenticated.

Box 9.3 *Continued...*

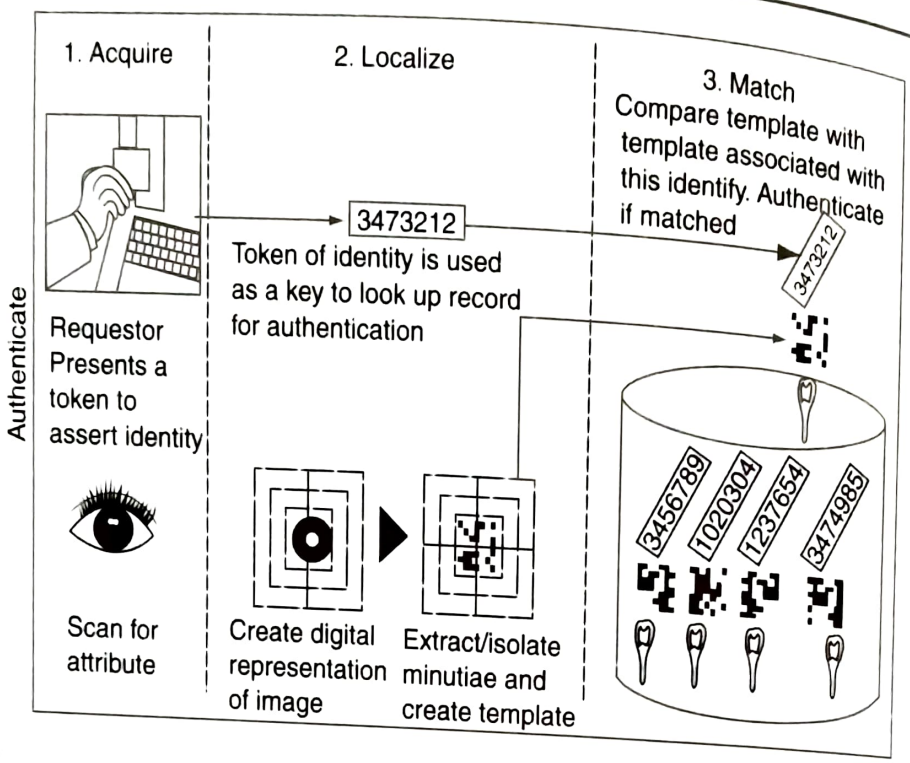


Figure 9.3 | Biometrics authentication: acquisition, localization and matching.

Courtesy:

1. http://www.tml.tkk.fi/Opinnot/Tik-110.501/1998/papers/12biometric/biometric.htm.
2. http://www.merkatum.com/docs/biometricsDell.pdf (accessed 1 June 2006).
3. http://www.idiap.ch/biometric-authentication.php and http://www.forensic-evidence.com/site/ID/ID_Biometric_jarvis.html.

9.5 Biometrics Techniques

Names of some biometrics techniques were mentioned in Section 9.3 and Box 9.3 (fingerprint scans, retinal and iris scans, etc.). In this section, major biometrics techniques are described in brief. A detailed treatment of each of the techniques is beyond the scope of this chapter. Readers interested in greater details may like to have a look at the reference materials solely devoted to the topic of biometrics as suggested in the *Further Reading* section. Some of the emerging technologies in biometrics are described here. They fall in major categories such as *hand-based techniques*, *eye-based techniques*, *face-based techniques*, *voice-based techniques* and *signature-based techniques*:

1. **Fingerprint:** Fingerprint identification techniques fall into two major categories – *Automated Fingerprint Identification Systems* (AFISs) and *Fingerprint Recognition Systems* (FRSs). AFIS is typically restricted to law-enforcement use. Fingerprint recognition derives a unique template from the attributes of the fingerprint without storing the image itself or even allowing for its reconstruction. Fingerprint recognition for identification acquires the initial image through a live scan of the finger (Figure 9.5) by direct contact with a reader device that can also check for validating attributes such as temperature and pulse. Since the finger actually touches the scanning device, the surface can become oily and cloudy after repeated use and therefore reduce the sensitivity and reliability of

optical scanners. Solid-state sensors overcome this and other technical hurdles because the coated silicon chip itself is the sensor. Solid-state devices use electrical capacitance to sense the ridges of the fingerprint and create a compact digital image, so they are less sensitive to dirt and oils. Fingerprint recognition is generally considered reliable enough for commercial use, and some vendors are already actively marketing readers as a part of local area network (LAN) login schemes. Figure 9.4 illustrates contours on a fingerprint of humans that is unique for each individual. Figure 9.5 illustrates a type of finger recognition equipment.

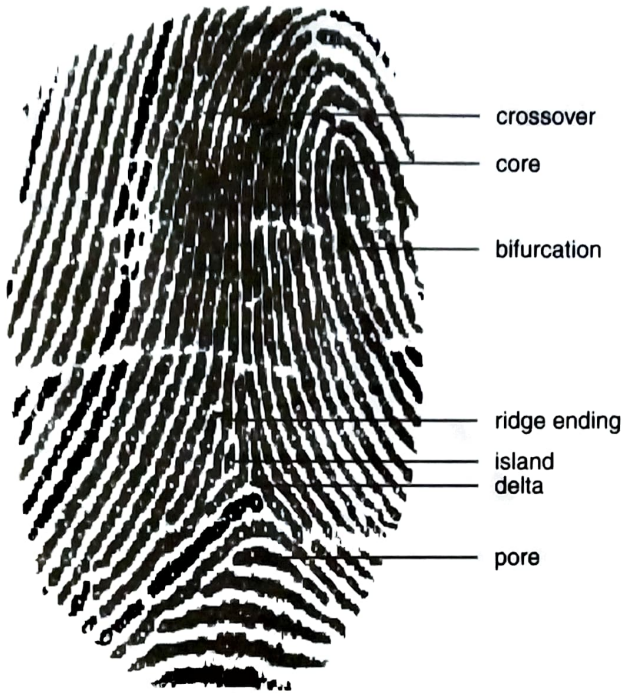


Figure 9.4 | Finger contour.

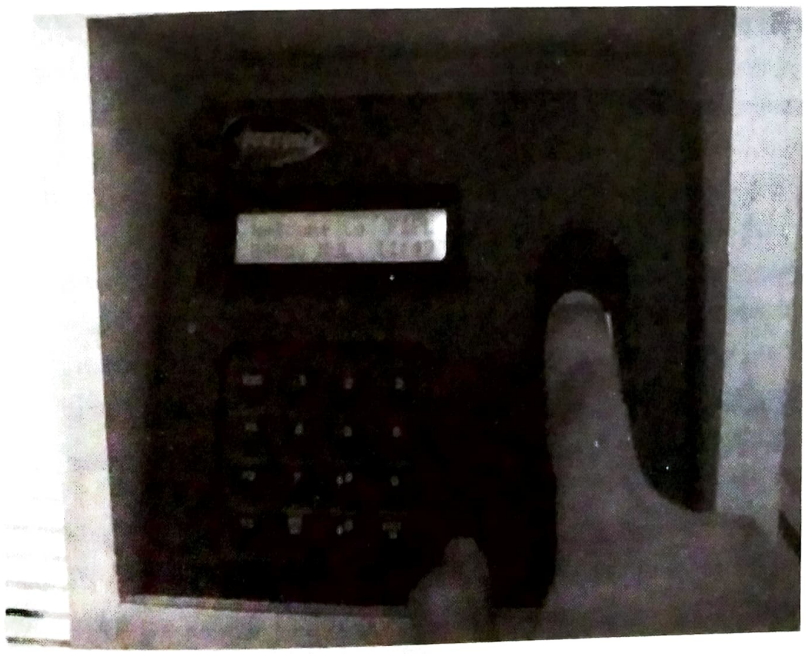


Figure 9.5 | Finger recognition system.

2. **Hand geometry**: The essence of hand geometry is the comparative dimensions of fingers and the locations of joints. Basically, the shape of a person's hand (the length and the width of the hand and fingers) measures hand geometry. This is a unique trait that differs significantly among people and hence is used in some biometrics systems to verify the ID of people. A person places his/her hand on a device that has grooves for each finger (see Figure 9.6). Reference marks on the plate allow calibration of the image to improve the precision of matching. The system compares the geometry of each finger, and the hand as a whole, to the information in a reference file (called the template) to verify that person's ID. Some systems perform simple, 2D measurements of the palm of the hand. Others attempt to construct a simple 3D image from which to extract template characteristics. Readers may find it interesting to note that one of the earliest automated biometrics systems, Identimat, was installed at the Shearson-Hamill investment bank on Wall Street (Manhattan, NY, USA) during the late 1960s. It used hand geometry and stayed in production for almost 20 years. In one of the most popular descendants of the Identimat, a small digital camera captures top and side images of the hand.

Figure 9.6 Hand geometry recognition system.
Courtesy: www.turnstile.us/index.

3. **Hand vein and palm vein biometrics**: Hand vein recognition attempts to distinguish individuals by measuring the differences in subcutaneous features of the hand using infrared (IR) imaging (see Figure 9.7).

Like face recognition system, vein recognition system, too, must deal with the extra issues of 3D space and the orientation of the hand. Like retinal scanning, it relies on the pattern of the veins in the hand to build a template with which to attempt matches against templates stored in a database. The use of IR imaging offers some of the same advantages as hand geometry over fingerprint recognition in manufacturing or shop-floor applications where hands may not be clean enough to scan properly using a conventional video or capacitance technique.

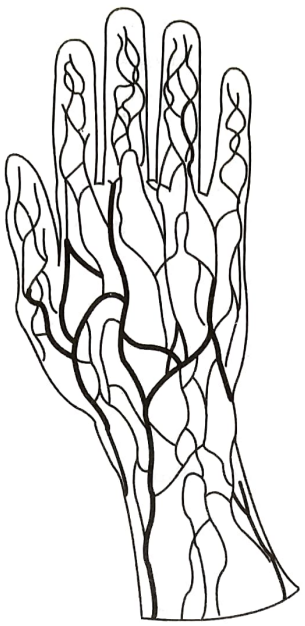


Figure 9.7 | Human palm vein geometry.

The pattern of blood veins (see Figure 9.7) is unique to every individual, even among identical twins. Palms have a broad and complicated vascular pattern and thus contain a wealth of differentiating features for personal identification. Furthermore, it will not vary during the person's lifetime. It is a very secure method of authentication because this blood vein pattern lies under the skin. This makes it almost impossible for others to read or copy.

Palm biometrics works by getting the vein pattern image captured (see Figure 9.7). An individual's vein pattern image is captured by radiating his/her hand with near-IR rays. The reflection method illuminates the palm using an IR ray and captures the light given off by the region after diffusion through the palm. The deoxidized hemoglobin in the vein vessels absorbs the IR ray, thereby reducing the reflection rate and causing the veins to appear as a black pattern. This vein pattern is then verified against a preregistered pattern to authenticate the individual. As mentioned before, given that veins are internal in the body and have a wealth of differentiating features, attempts to forge an ID are extremely difficult, thereby enabling a high level of security. In addition, the sensor of the palm vein device can only recognize the pattern if the deoxidized hemoglobin is actively flowing within the individual's veins. Palm vein recognition-based system is not dangerous; a near-IR ray is a component of sunlight and so there is no more exposure when scanning the hand than by walking outside in the sun.

4. **Signature:** Signature is the way a person signs his/her name and is known to be a characteristic of that individual. Although signatures require contact with the writing instrument and an effort on the part of the user, they have been accepted in government, legal and commercial transactions as a method of verification. Signatures are a *behavioral biometrics* that change over a period of time and are influenced by physical and emotional conditions of the signatories. Signatures of some people vary substantially: even successive impressions of their signature are significantly different. Further, professional forgers may be able to reproduce signatures that fool the system.

While a signature is not strictly biometrics (because it is not a part of human body), it is a simple, concrete expression of the unique variations in human hand geometry. Forensic experts have developed criteria over the years for verifying the authenticity of a signature. Automating this process allows computer automation to take the place of an expert in looking for unique identifying attributes. In

addition to the general shape of the signed name, a signature recognition system can also measure both the pressure and the velocity of the point of the stylus across the sensor pad. Signatures, however, are difficult to model for variation, and are reliable, especially when compared with other simpler alternatives.

Keystroke dynamics is a variation on signature recognition that measures the typing rates and intervals. Regarding keystroke dynamics, it is hypothesized that each person types on a keyboard in a characteristic way. This behavioral biometrics is not expected to be unique to each individual but it offers sufficient discriminatory information to permit ID verification. Owing to the fact that keystroke dynamics are a behavioral biometric, for some individuals, one may expect to observe large variations in typical typing patterns. Further, the keystrokes of a person using a system could be monitored unobtrusively as that person is keying in information. Box 9.4 explains the difference between biometric signature and digital signature.

Box 9.4 Biometrics Signature versus Digital Signature

The practice of signatures is probably as old as the human civilization starting from the time at which people learnt how to hand-sign. Thus, the act of signing a document has long been accepted by nearly every culture as one's recognition and agreement on the contents and implications of written words. However, now in the digital era, the increasing recognition of electronic signatures by lawmakers is bringing to the forefront concerns over electronic security (e-security) for privacy and protection of individuals.

For those conducting business transactions over private networks or the Internet, some form of official acknowledgement is now essential and legally binding. The security implications of producing or recognizing 'original' electronic documents (e-documents) will be more important than ever before. It is in this respect that an understanding to distinguish between the terms 'biometric' and 'digital' signatures becomes important.

What is a Digital Signature?

Digital signature is a term used to describe a long numeric code that is uniquely assigned to one person, hence the reference to 'signature'. Note that *it has nothing to do with a real signature*. The purpose of a digital signature is to be used in encryption systems. A digital signature is issued to an individual by a Certificate Authority (CA). This is a group or an organization responsible for maintenance and safekeeping of digital signatures. Because of their length, no one actually remembers or even knows their digital signatures.

An individual's digital signature will normally reside on his/her computer, or can be stored on a card (similar to banking cards). When someone wishes to encrypt an e-document, they will use a password or PIN that in turn allows the digital signature to be used.

Although secure once encrypted, digital signatures are only as safe as is the medium where they reside. Anyone obtaining access to your password, PIN or computer can potentially make unauthorized use of your digital signature. *The use of a digital signature*, however, *does not guarantee the ID of the originator*.

What is a Biometrics Signature?

Biometrics signature is a term used for referring to a signature that has been recorded/captured using a variety of input devices such as digitizing tablets, personal digital assistants (PDAs), computer displays or other contact-sensitive technologies. Typically, biometrics signatures can also be used to provide and control access security to buildings, networks, computers, documents and databases.

This method allows real handwritten signatures to be incorporated into e-documents during electronic transactions. Not every technology captures signature information in the same way. Some systems have a static approach and will only record an image of a signature and as such do not

Box 9.4 *Continued...*

record the unique behavioral elements associated with the execution of a signature. In a biometrics system, the geometric and dynamic characteristics of the signing process will be recorded and incorporated in an e-document. Most of the elements that make a signature unique and identifiable can be derived from the digital signature data. Furthermore, the data that are incorporated in an e-document can be used to lock and protect the contents from alteration.

Handwriting results from a highly complex series of dynamic neuromuscular tasks from the brain to the fingertips. A naturally developed signature represents the most often reproduced and habitual act of writing. Although we never sign exactly the same way twice, the signature adheres within certain boundaries unique to each individual. This natural variation is an essential component of handwriting. It also means that each signature is unique in that no two will be identical in all discrete features. Unlike fingerprints, retinal scans or DNA patterns that remain constant over time, the execution of a person's signature will be unique to each individual at that particular moment. Handwriting remains one of the most powerful human identifiers that exist today. Identical twins will have the same DNA pattern while their handwriting and signatures remain distinctively different.

Why Biometrics Signatures are Useful in the Security World ?

For a layman, the pictorial appearance of a conventional signature can be convincingly imitated. Forensically, when there is a question of whether or not the signature on a document is genuine, visual and microscopic examination by expert is required. This involves evaluating and comparing the general and discrete features of the contested signature with known signatures. With biometrics signatures, the authentication can be done in real time or after the fact. In the event that a biometrics signature is contested, the signature data can be extracted from the document and submitted to similar forensic investigation and analysis to verify the authenticity of the signature.

In fact, some of the biometrics data that are captured such as speed, acceleration, deceleration, and the amount of time the pen is on and off the paper are accurately measured. These data are either unavailable or qualitatively assessed at best in conventional forensic examinations of signatures. The additional behavioral features recorded from biometrics signatures make them even more difficult, if not impossible, to imitate.

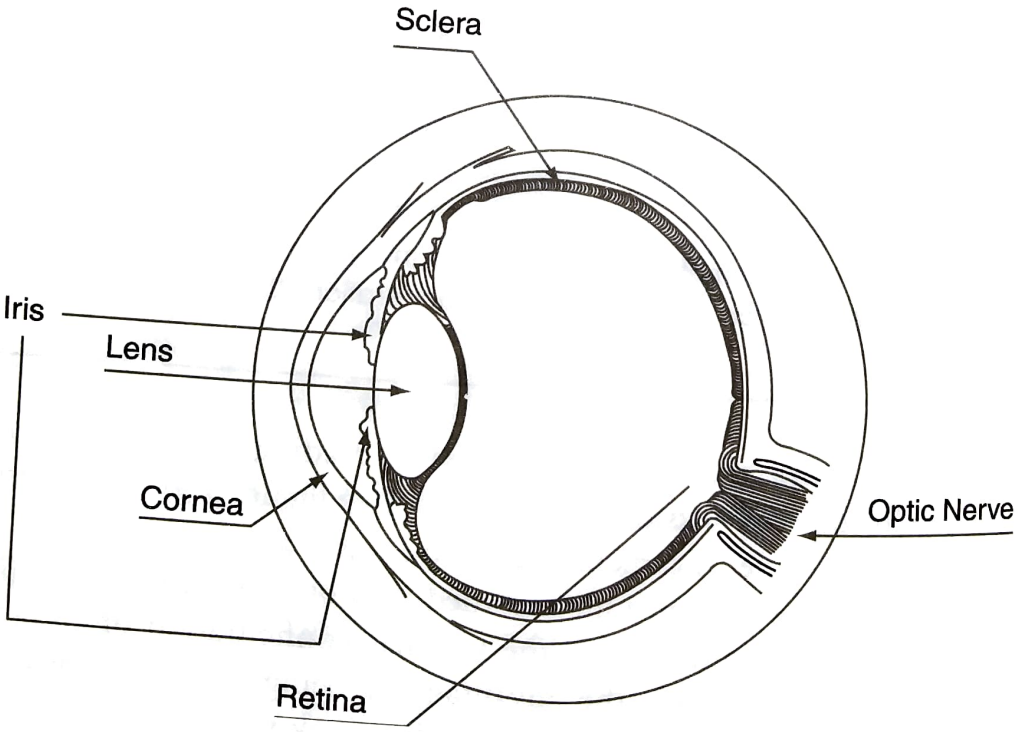
Biometrics signatures represent an ideal bridge between the long-recognized convention of signing a document and the need for e-documents to be uniquely recognized by individuals. This application provides individuals with security and control on documents originated, transacted and stored in the digital domain.

Courtesy:

1. http://www.bioenabletech.com/biometrics/signature_recognition.htm (accessed 11 June 2006).
2. http://www.cl.cam.ac.uk/~fh240/pdf/Private%20key%20generation%20from%20on-line%20handwritten%20signatures.pdf.

5. **Retinal scan:** For a *retinal scan*, there is a system used for reading a person's retina to scan the blood-vessel pattern of a retina on the backside of the eyeball (see Figure 9.9). This pattern is known to be extremely unique among people. A camera is used to project a beam inside the eye and capture the pattern and compare it to the reference file recorded previously (called the template). Thus, retinal recognition creates an 'eye signature' from the vascular configuration of the retina, an extremely consistent and reliable attribute with the advantage of being protected inside the eye itself. An image of the retina is captured by having the individual look through a lens at an alignment target (see Figure 9.8). Diseases or injuries that would interfere with the retina are comparatively rare in the general population, so the attribute normally remains both consistent and consistently available.
6. **Iris scan:** The 'iris' is the colored portion of the eye that surrounds the pupil. Refer Figure 9.9 to understand the anatomy of a human eye. The iris has unique patterns, rifts, colors, rings, coronas

Figure 9.8 | Retinal scanning equipment.



gure 9.9 | The anatomy of a human eye.

and furrows. It is the uniqueness of each of these characteristics that makes it amenable as a biometrics method for identification. These unique characteristics are captured by a camera and compared with the information gathered during the enrolment phase (enrolment is discussed in Section 9.6).

The issue of 'intrusiveness' of biometrics techniques is an important one and it is discussed in Section 9.7. At this point, it should be noted that *iris scanning* is less intrusive than retinal recognition because the iris is easily visible from several feet away. Responses of the iris to changes in light can provide secondary verification that the iris presented as a biometrics factor is genuine. Though empirical tests with the technology will improve its reliability, it appears quite promising and even practical for many applications, especially in two-factor scenarios. While some of the technical issues of iris scanning seem pedestrian, they present implementation challenges. A careful balance of light, focus, resolution and contrast is necessary to extract the attributes or minutiae from the localized image. While the iris seems to be consistent throughout adulthood, it does vary somewhat up to adolescence (Box 9.5).

Box 9.5 Biometrics Technology Holds a Promise – Iris Recognition That Works

According to a news item in the year 2001, the results of new studies in Britain showed that iris recognition is a viable, fast and easy way to authenticate a person's ID. The findings are so positive that the prospect of iris prints replacing PINs and other identification methods (e.g., passports) is much more viable.

A paper called 'Epigenetic Randomness, Complexity and Singularity of Human Iris Patterns' by John Daugman OBE and Cathryn Downing of the University of Cambridge contains the results from a study of 2,000 iris images from the United States, United Kingdom, and Japan collected over a three-year period. A total of 2,048-bit IrisCodes (a digital signature of the iris image) were taken from subjects using small video cameras, and these images represent a total of over 2.3 million possible pairs. Comparison of the IrisCodes found that the chances are 1 in 10,000,000 that two IrisCodes would be even two-thirds the same – even identical twins have different IrisCodes.

On the same token, if a live IrisCode matches an IrisCode in a database only up to 75%, the chances of a mismatch are one in a thousand, million, million. This finding allows for the normal variation in a person's iris under different conditions, when the digital signature might be off by an average of 11%.

The UK government's National Physics Laboratory also recently tested different biometrics identification systems. Out of fingerprint, face, voice, hand, vein and iris recognition systems, only Daugman and Downing's IrisCode system 'made no false matches among the 2.75 million comparisons, while working the fastest at 1.5 million matches per minute'. The IrisCode system's mathematical algorithms for iris identification are small enough and fast enough to make huge databases of millions of people easily searchable for identification purposes.

There is something furthermore interesting that makes a promise on iris recognition technology ticking well in the future as a method for biometrics identification of a person. Consider the following: according to another article reported on the website http://www.geek.com/news/geeknews/2001aug/gee20010809007215.htm, identical twins do not have the same IrisCodes, and identical twins are natural clones (i.e., they have the exact same DNA). So if identical twins do not have the same IrisCodes, neither will cloned persons! This means less chances of cheating using cloning!

Courtesy: http://www.geek.com/iris-recognition-can-work/.

7. **Face/facial thermogram:** Facial images are the most common biometrics characteristics used by humans to make a personal recognition, hence the idea to use this biometrics in technology. Face recognition technology is still in its early stages, and most tests and applications have been run against relatively small databases. The similarity score produced by each comparison determines the match – the highest score wins. Acquisition for biometrics identification purposes requires the individual's face to be presented to a video camera. An evident deficiency in some current schemes is the ability to fool or confuse some systems with makeup.
A facial thermogram works much like face recognition except that the image is captured by way of an IR camera, and the heat signature of the face is used to create the biometrics template used for matching. This is more reliable than simple imaging. Although a comparison of various technologies and algorithms shows that the results are promising and some approaches did yield impressive results, this technology is still considerably less reliable than some alternatives. As is the case with other technologies, practical usefulness increases dramatically in a two-factor scenario. Face recognition/facial thermogram method is a non-intrusive method and is suitable for covert recognition applications. The applications of facial recognition range from static ('mug shots') to dynamic, uncontrolled face identification in a cluttered background (subway and airport).
Face verification involves extracting a feature set from a 2D image of the user's face and matching it with the template stored in a database. The most popular approaches to face recognition are based on either (1) the location and shape of facial attributes such as eyes, eyebrows, nose, lips and chin and their spatial relationships or (2) the overall (global) analysis of the face image that represents a face as a weighted combination of a number of canonical faces. Although the performance of commercially available systems is reasonable, there is still a significant room for improvement since *false reject rate* (FRR) is about 10% and *false accept rate* (FAR) is 1%. These systems also have difficulties in recognizing a face from the images captured from two different angles and under different ambient illumination conditions. It is questionable if a face itself is a sufficient basis for recognizing a person from a large number of identities with an extremely high level of confidence. Facial recognition system should be able to automatically detect a face in an image, extract its features and then recognize it from a general viewpoint (i.e., from any pose) that is a rather difficult task. Another problem is the fact that the face is a changeable social organ displaying a variety of expressions.
8. **Voice:** Voice recognition techniques are generally categorized according to two approaches - automatic speaker verification (ASV) and automatic speaker identification (ASI). ASV uses voice as the authenticating attribute in a two-factor scenario. ASI attempts to use voice to identify who an individual actually is. Voice recognition distinguishes an individual by matching particular voice traits against templates stored in a database. Voice systems must be trained to the individual's voice at enrolment time, and more than one enrolment session is often necessary. Feature extraction typically measures formants or sound characteristics unique to each person's vocal tract. The pattern-matching algorithms used in voice recognition are similar to those used in face recognition. Readers are encouraged to refer the paper by Jain and Prabhakar (2004) for comparison of various biometrics techniques.

9.6 Matching and Enrolment Process in Biometrics

In the discussion so far, several times, there was a mention of the terms 'enrolment' and 'template'. In this section, we explain these extremely important terms associated with biometrics. As humans, we are more comfortable recognizing our friends and family members through their faces, voices, mannerisms and gaits (the way they walk). Also, most of us are more comfortable using PINs and passwords for proving who we are. However, teaching computers how we do this so easily is a challenge. For this purpose, enrolment

and template creation are the necessary steps in biometrics. Almost all biometrics systems share the same matching flow (illustrated in Figure 9.10).

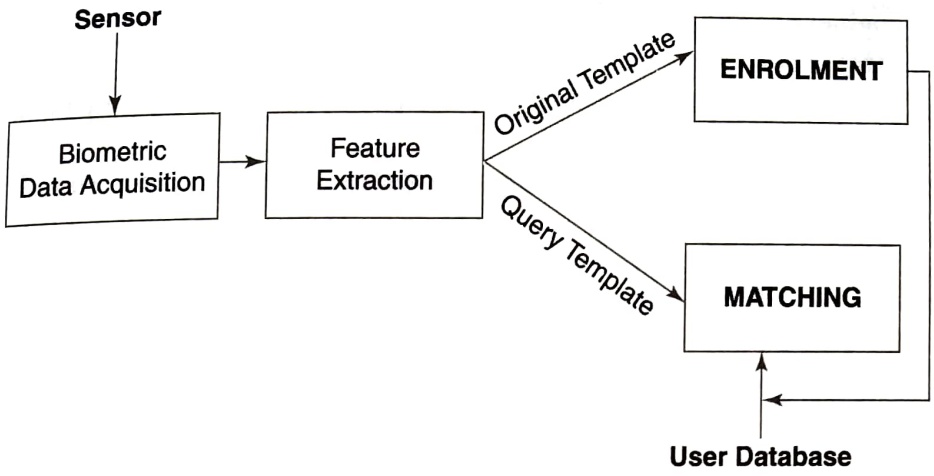


Figure 9.10 | Process flow in biometrics matching.

There are a number of key terms that appear in Figure 9.10 that are explained as follows:

1. **Biometrics:** A measurable physical characteristic or personal behavioral trait used to recognize the ID, or verify the claimed ID, of an enrollee.
2. **Behavioral biometrics:** This is a biometrics that is characterized by a behavioral trait that is learnt and acquired over time rather than a physiological characteristic. However, physiological elements may influence the monitored behavior.
3. **Biometrics data:** These are also known as biometrics sample. These data consist of biometrics characteristics of the entity under authentication and are physiological data in nature. They are the information extracted from the biometrics sample and used either to build a reference template (template data) or to compare against a previously created reference template (comparison data). Thus, biometrics sample/biometrics data are the raw data representing a biometrics characteristic of an end-user as captured by a biometrics system (e.g., the image of a fingerprint, retinal scan data, etc.).
4. **Enrolment:** It is the process by which a subject's (entity under authentication) biometrics data are initially acquired, accessed, processed and stored in the form of a template. Thus, it is the process of collecting biometrics samples from a person and the subsequent preparation and storage of biometrics reference templates representing that person's ID.
5. **Enrolment time:** It is the time period a person must spend to have his/her biometrics reference template successfully created.
6. **Template:** It is a crucial element in the working of biometrics systems as it is the deciding and defining element of biometrics technology. A template is nothing but a small file derived from the distinctive features of a users' biometrics data used for performing biometrics matches. It is important to note that the biometrics systems store and compare biometrics templates, and **not** biometrics data.
7. **Match/matching:** It is the process of comparing a biometrics sample against a previously stored template and scoring the level of similarity. Accept or reject decisions are based on whether this score exceeds the given threshold.

8. **Feature extraction:** This is the automated process of locating and encoding distinctive characteristics from biometrics data in order to generate a template. Feature extraction takes place during enrolment and verification process.
9. **Biometrics engine:** It is the software element of the biometrics system that processes biometrics data during the stages of enrolment and capture, extraction, comparison and matching.
10. **Biometrics device:** It is the part of a biometrics system containing the sensor that captures a biometrics sample from an individual.
11. **Comparison:** It is the process of comparing a biometrics sample with a previously stored reference template or templates.
12. **Minutiae:** It is the unique, measurable physical characteristic scanned as an input and stored for matching by biometrics systems. For fingerprints, minutiae include the starting and ending points of ridges, and ridge junctions among other features. Figure 9.11 shows, for example, how a fingerprint is processed to arrive at a *minutiae* starting with the original fingerprint and ending with the *minutiae graph*.

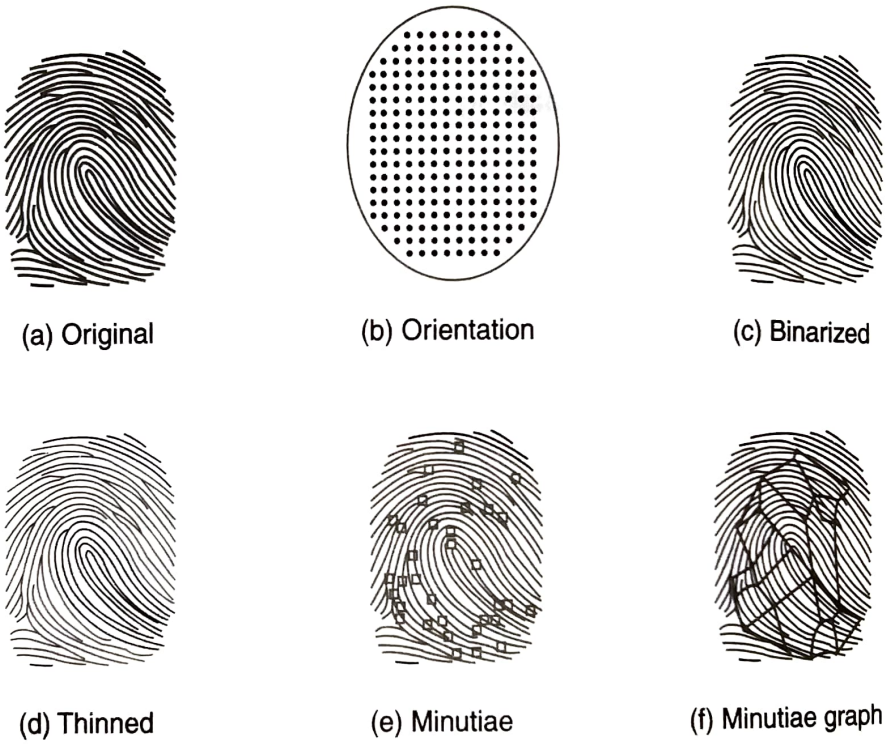


Figure 9.11 | Process flow in obtaining a minutiae.

9.7 Key Success Factors for Biometrics Systems

For any effective biometrics system, there are a few important factors associated with it: accuracy, speed and throughput rate, acceptance by users, uniqueness of biometrics organ and action (on which the techniques are based), reliability (i.e., resistance to counterfeiting), data storage requirements, enrolment time (this term was explained in the previous section), intrusiveness of data collection, etc. Effective functioning of biometrics systems would depend on these factors. In this section, these factors are explained in brief.

Accuracy

Accuracy is the most critical characteristic of a biometrics identification verification system. If the system cannot accurately separate an authentic person from an impostor, it should not even be termed a biometrics

identification system. There are two issues that arise – *false rejection rate (FRR)* and *false acceptance rate (FAR)*:

1. **FRR:** This rate is generally expressed as a percentage. It is the rate at which authentic, enrolled persons are rejected as unidentified or unverified persons by a biometrics system. The FRR is also known as *Type I error*. In access control, if the requirement is to keep the unauthorized persons out, false rejection is considered the least important error. However, in other biometrics applications (e.g., the visa application investigation discussed in Box 9.2), it may be the most important error. In the banking or retail business domain, when a biometrics system is used to authenticate the customer ID and account balance, false rejection means that the transaction or sale is lost, upsetting the customer. Most banks and retailers tend to be OK with a few false accepts as long as there are no false rejects. They do this keeping their business interests in mind.

   There is a reason why the matter of false rejection creates so much anxiety – it has a negative effect on throughput, frustrations and operations impediments by causing unnecessary delays in personnel movements. Another problem with FRR is that sometimes it is incorrectly attributed to the 'failure to acquire'. Failure to acquire occurs when the biometrics sensor is not presented with sufficient usable data to make an authentic or impostor decision. Examples of unusable data include smudged prints on a fingerprint system, improper hand positioning on a hand geometry system, improper alignment on a retina or iris system or mumbling on a voice recognition system. The issue is that subjects under identification (humans) cause failure-to-acquire problems, either accidentally (unintentionally) or on purpose to fudge the security system for their illicit purpose. When FRRs rise, it may be OK if it is a tight security area such as defense or medical institution but not OK if it is a retail business/commercial centre (where biometrics-based access controls are anyways not suitable). 'FAR' is a reverse situation. This is the rate (stated as percentage) at which unenrolled persons or impostors are accepted as authentic, enrolled persons by a biometrics system. *FAR* is also known as *Type II error*.
2. **Crossover error rate (CER):** It is the rate at which the FRR and FAR match. It is also known as 'equal error rate' (ERR) and is stated as a percentage. This is the most important measure of biometrics system accuracy. It is important to understand the meaning of false non-match rate (FNMR). A biometrics solution's FNMR is the probability that a user's template will be incorrectly judged to not match his/her enrolment template. Given this, we can say that ERR is the rate at which the FNMR is equal to the false match rate (FMR). Thus, ERR or CER presents the accuracy level at which the probability of a false match is the same as the probability of a false non-match. ERR is commonly used as a representation of overall system accuracy, because it is a general indicator of a system's resistance to break-ins and ability to match templates from authorized users.

Speed and Throughput Rate

For biometrics system characterization, *speed* and *throughput* are important. Data-processing capability of the biometrics system decides the speed; it is stated as how fast the accept or reject decision is enunciated. It relates to the authentication procedure; the system setup, card input or PIN (if a verification system); inputting the physical data by inserting the hand or finger, aligning the eye speaking access words or signing a name processing and matching of data files; enunciation of the accept or reject decision; in case of a portal system, movement through the door and closing of the door, etc. A system speed of 5 s from start-up through decision enunciation is good as per generally accepted standards. Another standard is a portal throughput rate of 6–10 per min, which equates to 6–10 s per person through the door. In spite of great strides in the biometrics research, it is not easy for most biometrics systems to meet these standards.

Acceptability by Users

User acceptability to-date is a big challenge for pervasive deployment of biometrics systems. This is mainly owing to the social stigma attached to the biometrics systems given their nature and lack of adequate awareness on biometrics identification systems. Biometrics system acceptance occurs when those who must use the system, that is, management and unions involved in the organizations, need to come to an agreement that biometrics should be deployed for the protection of organizational assets. Also consider the social stigma factor mentioned as well as the lack of awareness; fingerprinting is a particularly sensitive topic, given that it is associated with criminals. Eye retina scanning requires users to trust that the system will not damage their eyes, a feeling they carry possibly owing to rumors and inadequate information about how the retinal scanning technology works. It is clear that uncooperative users can overtly or covertly compromise, damage or sabotage the biometrics scanning equipment. Given this the management has to decide about the implementation based on the cost/benefit associated with a biometrics system. Biometrics has got privacy implications too (this aspect is discussed in Chapter 31).

Uniqueness of Biometrics Organ and Action

The purpose of a biometrics system is positive identification of the personnel – given this, it is important that the systems are based on unique characteristics of the employees. So, when the base is a unique characteristic, a file match is a positive identification rather than a statement of high probability that it is the right person. Out of the many physical characteristics that can be used, only three can really be considered unique enough for identification: the fingerprint, the retina of the eye (the blood-vessel pattern inside the black of the eyeball) and the iris of the eye (random pattern of features in the colored portion of the eye surrounding the pupil).

Reliability of Biometrics

When using biometrics verification systems, it is vital that they operate in an accurate fashion. The concept of a biometrics system's reliability is related to its 'selectivity'. *Reliability* is the probability that a matcher system will correctly identify the mate when the mate (i.e., entity whose biometrics unique character is being matched) is present in the system repository, whereas *selectivity* is the number of incorrect mates determined for a given search.

Only authorized persons must be allowed to access and it must preclude the others without breakdown or deterioration in performance accuracy or speed. In addition, these performance standards must be sustainable without high levels of maintenance or frequent diagnostics and system adjustments. The trade-off between reliability and selectivity offers the greatest system design challenge since these parameters are interdependent.

Data Storage Requirements in Biometrics Systems

Earlier computer systems had primary and secondary memory size constraints, that is, limited random access memory (RAM) and disk size. This is less of an issue today as computer technology has advanced in both hardware and software. Even then, the size of biometrics data files remains a factor of interest. Given the large size of biometrics match templates, even with the current ultra-high-speed processors, large data files take longer than small files to process. This is especially so in biometrics systems that perform 'full identification', that is, matching the input file against every file in the database. Typically, *biometrics file size* varies between 9 and 10,000 bytes, mostly falling in the 256–1,000 byte range.

Enrolment Time in Biometrics

We discussed 'matching and enrolment'; enrolment time is also not so much of an issue these days. In the early days, biometrics systems sometimes had enrolment procedures requiring many repetitions and several minutes to complete. Consider the following: a system requiring a 5-min enrolment instead of 2 min causes 50 h of expensive non-productive time if 1,000 users must be enrolled. In addition, line waiting time must also be considered. All this adds to the total cost. The accepted standard for enrolment time is 2 min per person. Most of the systems available in the market today meet this standard.

Data Collection Intrusiveness

Origins of this factor come from the users' concerns about the collection of biometrics data from inside the human body – especially the retina inside the eyeball. Early biometrics systems illuminated the retina with a red light beam that happened to coincide with increasing public awareness of lasers, sometimes demonstrated as red light beams cutting steel! Although there has never been any injury reported owing to retinal scans using the light beams, public fear and user sensitivity still remain. An advance form of such public concern is about intrusions into human body, a space that is considered private.

Requirements About Subject and System Contacts

Under biometrics methods for identification, the users are required to make a firm physical contact with the instrument for biometrics data collection. This factor is an extension of the concerns mentioned in the previous section, given that the users need to make mouth contacts (voice-based identification), eye contacts (retinal scans) and hand contact (hand geometry-based identification). This factor involves a lot of human psychology too – 'If I wish to make contact with the instrument as my own accord, then it is OK but if an organization and agency makes me do it then I am not OK with it', probably because the users imagine a possible misuse of what they think is 'private information' about their body. In a way, it is the attitude about the biometrics techniques that is having an impact on adoption rates of these techniques. Box 9.6 explains developments in 3D imaging techniques for use in biometrics.

Box 9.6 Special Camera for Biometrics Developers

Biometrics developers may soon benefit from new machine vision technology that utilizes real-time 3D imaging. A San Jose, CA, company is the inventor of a low-cost electronic perception technology, dubbed Equinox, that enables machines and ordinary electronic devices to perceive and react to nearby objects or individuals in real time. The single-chip Equinox 3D 'camera' has a universal serial bus (USB) interface that allows a connection to standard personal computers (PCs) and has a Windows-based software development environment.

The fundamental basics of the technology involve distance and timing, like a radar system using light reflection rather than radio waves. The light illuminating each individual pixel in an image sensor comes from a different feature in the scene being viewed. If the amount of time that light takes to reach each pixel can be determined, then the exact distance to that feature can be calculated with certainty.

This is developing a 3D relief map of the surfaces in the scene. 'In three dimensions, objects previously indistinguishable from the background, for example, metaphorically pop out. For a broad class of applications, this may prove extremely helpful in reducing the mathematical and physical complexity that has plagued computer vision applications from the start.' The chips are not fooled by ambient light conditions, either.

Box 9.6 *Continued...*

The technology measures the time duration it takes the pulse to reflect back to each pixel, using high-speed, on-chip timers in one method or simply counts the number of returning photons, an indirect measure of the distance, in another. In either case, the result is an array of 'distances' updated as often as 50 times/s that provides a mathematically accurate, dynamic 'relief' map of the surfaces being imaged. The image and distance information is then handed off to an on-chip processor running manufacturer's proprietary imaging software that further refines the 3D representation before sending it off chip to the original equipment manufacturer (OEM) application.

Courtesy: http://www.canesta.com/assets/pdf/news/2004/CanestaBiometricWatch09-04.pdf#search='biometric%20file%20sizes (accessed 1 June 2006).

9.8 Benefits of Biometrics over Traditional Authentication Methods

Traditionally, passwords and ID cards have been used to restrict access to secure systems but these methods can easily be breached and are unreliable. Biometrics cannot be borrowed, stolen or forgotten and forging one is practically impossible. From the preceding discussions, one can see that biometrics is an alternative to using passwords for authentication in logical or technical access control. Biometrics is based on the third type of authentication mechanism – *something you are* (recall Section 9.2). Biometrics is defined as *an automated means of identifying or authenticating the ID of a living person based on physiological or behavioral characteristics.* In biometrics, identification is a 'one-to-many' search of an individual's characteristics from a database of stored images. Authentication in biometrics is a 'one-to-one' search to verify a claim to an ID made by a person. Biometrics is used for identification in physical controls and for authentication in logical controls.

In the domain of physical security (discussed in Chapters 7 and 8), *passwords* and PINs are the most frequently used authentication techniques for controlling access. In higher security applications, handheld tokens are used instead of passwords. However, passwords, PINs and tokens have a number of problems that raise questions about their suitability for modern security access control applications, particularly high-security applications such as access to defense systems or medical data systems (Box 9.7). Biometrics provides a number of benefits compared to the traditional methods:

1. increased level of security;
2. greater convenience;
3. higher level of accountability;
4. fraud detection and fraud deterrence.

Box 9.7 Advances in Biometrics Set Off Debates

According to the Federal Bureau of Investigation (FBI) officials, the only way to trace a terrorist is through biometrics, as they feel traditional passports are getting meaningless given larger opportunities for forging with the help of modern document production technology.

British police will almost certainly be given access in the near future to US intelligence databases containing DNA samples, fingerprints and digital images of thousands of foreign nationals seized around the world by the United States as terror suspects. As the war on terror increasingly comes on biometrics technology – the use of physical characteristics unique to individuals such as iris pattern, DNA and fingerprints to verify the ID – western police and intelligence agencies are drawing up plans for sophisticated biometrics databases that would allow them to share sensitive information.

10.4 Architectural and Design Issues in Biometrics Systems

Refer to Figures 9.10 (Chapter 9), 10.1 and 10.2. A generic biometrics system goes through six basic steps as indicated in Figure 10.2. The last two steps are used only during the recognition phase.

Six Basic Steps in Biometrics Systems

1. **Sample acquisition:** In this first step, the biometrics data must be collected using an appropriate sensor, for example, an image capture in the case of iris recognition or a saliva sample in the case of DNA.
2. **Feature extraction:** This step performs the transformation from the sample into the template. In general, the template is numeric data. (This step can be omitted if full images are used.)
3. **Quality verification:** This step establishes a reference image or template by repeating the first two operations as many times as needed so as to ensure that the system has captured and recognized the data correctly.

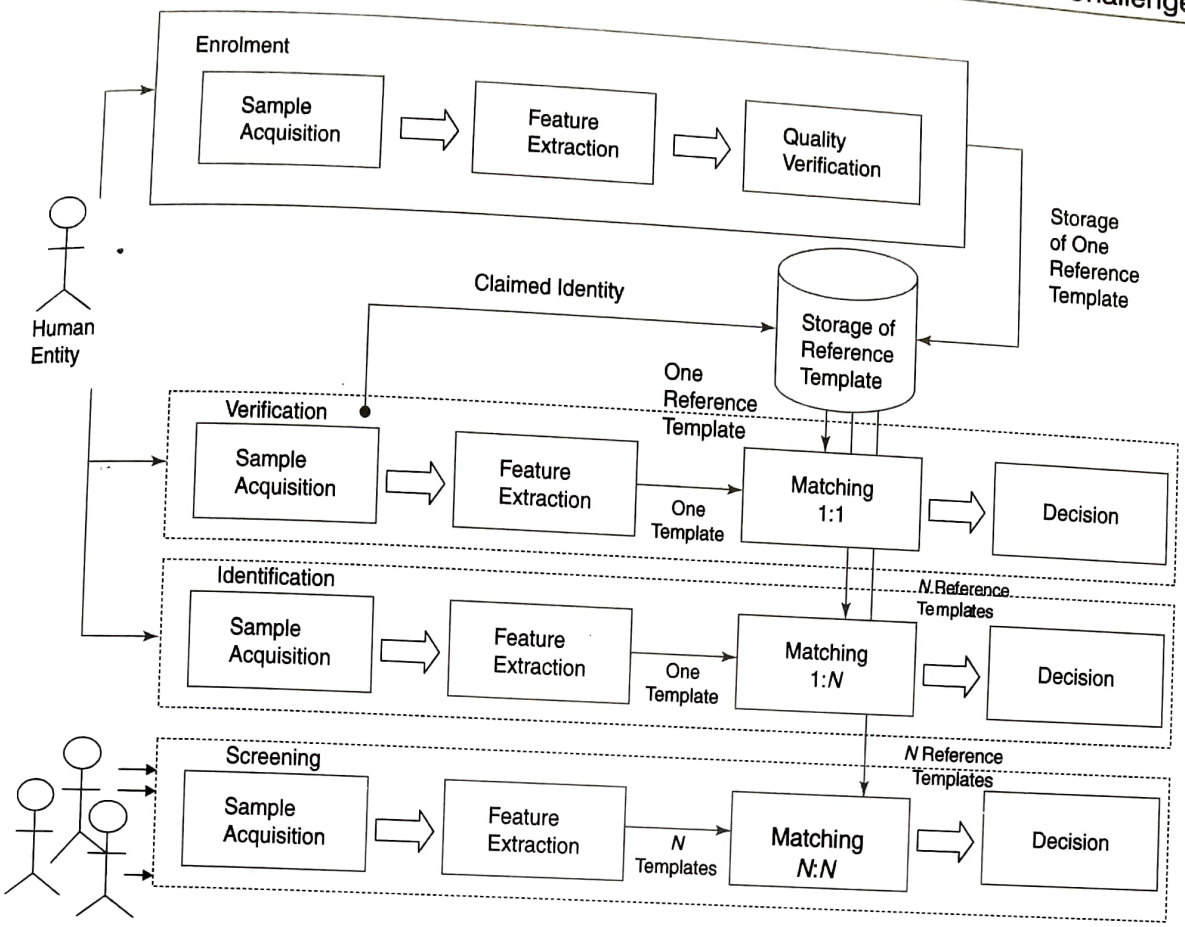


Figure 10.2 The six basic steps in a generic biometrics system. *Courtesy*: Adapted from Jain et al. (2001).

4. **Storage of reference template:** This step registers the reference template. Several storage media are possible and the choice depends on the requirements of the application (see Tables 10.1 and 10.2 and Section 10.2).
5. **Matching:** This step compares the real-time input data from an individual with the reference template(s) or image(s).
6. **Decision:** This step uses the result of the matching step to declare a result, in accordance with application-dependent criteria (e.g., decision threshold) – for example, for a verification task, the result would say whether the user claiming an ID should be authenticated.

Design Issues in Biometrics Systems

Biometrics systems, by their very nature, are complex systems with responsive decision-making involved in terms of physical access controls. The two most critical issues that designers of biometrics systems face are: storing and protecting the template and ensuring accuracy of the biometrics system steps that are mentioned above. These are briefly discussed as follows:

1. **Storage and protection of the template:** As shown in Figure 10.2, biometrics systems have to scan, store/retrieve a template and match. It is important to note that depending on the design of the system, the match can be performed in different locations: on the processor that is used to acquire the biometrics sample data, on a local PC or a remote server or on a portable medium such as a smart card (equipped with a strong enough processor). In addition, the reference template may be stored on the same three media leaving us with five different combinations and resulting in five different

levels of 'trust'. Also, there can be three different 'modes of protection' that may be used for the template: no protection, data encryption or digital signature. This means that there can be 15 possible configurations. Each use of combination has its own advantages and disadvantages; the choice of the combination is clearly application-dependent (based on risk and requirements analysis).

2. **Accuracy of biometrics system steps:** The evaluation of a biometrics system has to be based on the evaluation of all components: the recognition system performance, the communication interface, the matching and decision step and other key factors such as ease of use, acquisition speed and processing speed. We had discussed these factors in the previous chapter (Section 9.7). The performance of a biometrics system ultimately depends on the accuracy of the end decision only (refer Figure 10.2).

As mentioned in the previous chapter, in the case of a verification system, there are two possible types of error: *false non-match* (also known as *false negative* or *false rejection*, i.e., rejection of a legitimate user) and *false match* (also known as *false positive* or *false acceptance*, i.e., acceptance of an impostor). The corresponding error rates are the *false rejection rate* (FRR) that is equivalent to false non-match rate (FNMR) and the *false acceptance rate* (FAR) that is equivalent to false match rate (FMR). These error rates vary inversely, so for one technology under fixed operation conditions, lowering one error rate will necessarily raise the other.

Box 13.2 *Continued...*

certificates. It also includes a cryptographically secure pseudorandom number generator function CryptGenRandom.

For readers interested in seeking greater details on the topic of cryptovirology, some links are provided in the *Further Reading* section.

Courtesy:

1. http://en.wikipedia.org/wiki/Ransomware_(malware) (accessed 22 March 2008).
2. http://en.wikipedia.org/wiki/Cryptographic_Application_Programming_Interface (accessed 22 March 2008).

13.4 Role of Cryptography in Information Security

From electronic mail (e-mail) to cellular communications, from secure web access to digital cash, cryptography is an essential part of today's IS. In the light of the discussion so far, we can see that cryptography helps provide accountability, fairness, accuracy and confidentiality. It can prevent fraud in electronic commerce (e-commerce) and assure the validity of financial transactions. It can prove the ID of an entity or protect one's anonymity. It can keep vandals from altering a web page and prevent industrial competitors from reading confidential documents. And in the future too, as commerce and communications continue to move to computer networks, cryptography will become more and more vital. However, experts feel that the cryptography now available in the market does not provide the level of security it advertises. They find that most systems are not designed and implemented in concert with cryptographers.

One cannot make the systems secure by taking on cryptography as an afterthought. There has to be a vision of what is being done at every step of InfoSec design, from conception to installation. A point to lament is the following: a large amount of money is spent on computer security, and most of it is wasted on insecure products. After all, a weak cryptography looks the same on the shelf as a strong cryptography. Two e-mail encryption products may have almost the same user interface, yet one is secure while the other permits eavesdropping. A comparison chart may suggest that the two programs have similar features, although one has gaping security holes that the other does not have. An experienced cryptographer can tell the difference but so can a thief!

13.5 Digital Signature – A Method for Information Security

Recall that in Chapter 9 the term 'digital signature' was explained while comparing it with biometrics signature. In an earlier section, it was mentioned that the most recent and useful development in the use of cryptography is the digital signature. In this section, we explore the topic of *digital signature* to explore the underlying concepts using an illustrative scenario.

Suppose there is a student called Babu living in a college hostel who also works part-time. Let us say, he has been given two keys. One of Babu's keys is called a *public key*; the other is called a *private key*. Babu has some hostel mates: Pamila, D'Souza and Sushant who also work for some other organizations (see Figure 13.2).

Use of Keys for Data Encryption

Recall the key terminology in cryptography at the start of the chapter. Keys are used to encrypt information; only a person with the appropriate key can make it readable again. In this illustrative example, Babu's public

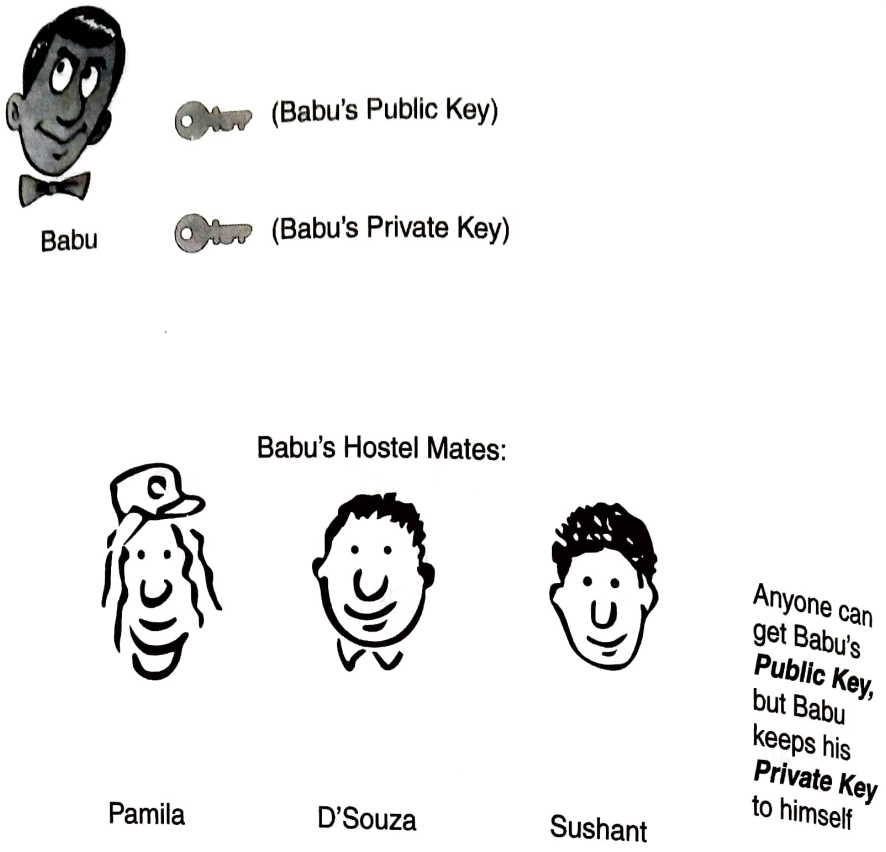


Figure 13.2 Public and private keys.
Courtesy: http://www.youdzone.com/signature.html – courtesy David Youd.

key is available to anyone who needs it, but he keeps his *private key* to himself. Any one of Babu's two keys can encrypt data, and the other key can decrypt those data. This point is important to note for understanding how encrypting and decrypting work. Sushant (see Figure 13.3) can encrypt a message using Babu's public key. Babu uses his private key to decrypt the message. Any of Babu's hostel mates might have access to the message Sushant encrypted, but without Babu's private key, the data are worthless.

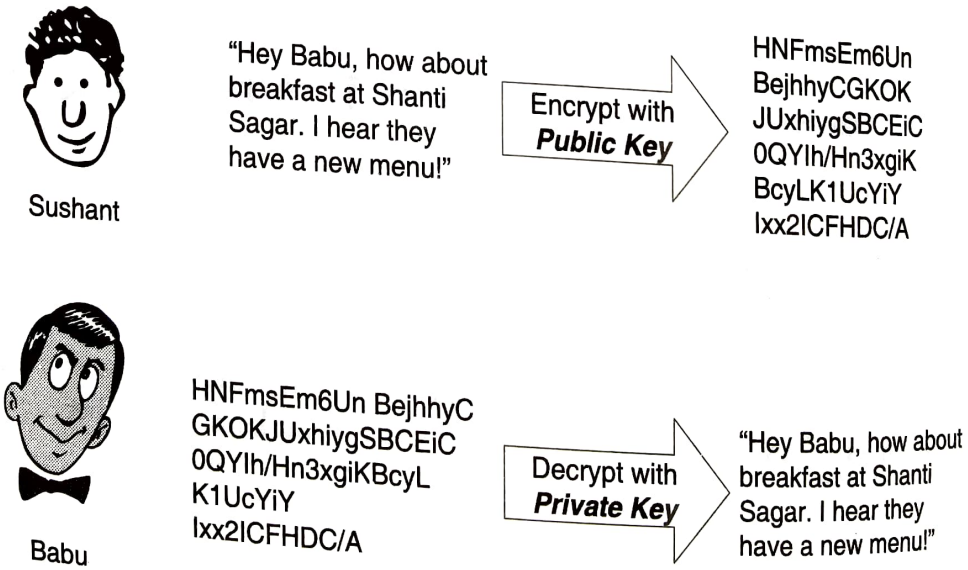


Figure 13.3 Message encryption and decryption using public and private keys.
Courtesy: http://www.youdzone.com/signature.html – courtesy David Youd.

Creating Digital Signature

With his private key and the right software, Babu can put digital signatures on documents and other data. A digital signature is a 'stamp' Babu places on the data that is unique to Babu, and is very difficult to forge. In addition, the signature assures that any changes made to the data that have been signed cannot go undetected (see Figure 13.3).

To sign a document, a person using the keys will use suitable software available to crunch down the data into just a few lines by a process called 'hashing'. A hash function is defined as the process that can take an arbitrary-length message and return a fixed-length value from that message. For practical use of hashing function, it is important that given a message, it should be easy to find the hash; given the hash, it should be hard to find the message and given the message, it should be hard to find another (specific or random) message that produces the same hash. Shown in Figure 13.4, these few lines as a result of hashing are called a *message digest.*

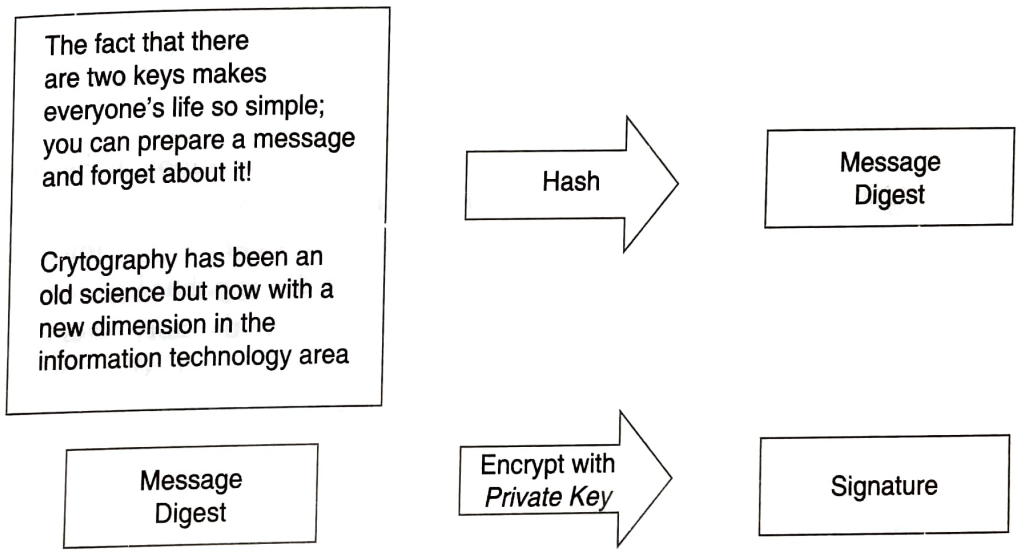


Figure 13.4 | Message digest creation.

Message Digest

A message digest is the product of a one-way hash function applied on a message; it is a fingerprint or a unique summary that can uniquely identify the message (see Figure 13.4). An important point to note is that it is not possible to change a message digest back into the original data from which it was created.

The software then encrypts the message digest with the private key. The result is the *digital signature* (see Figure 13.5). Finally, as shown in Figure 13.5, the software appends the digital signature to the document. All of the data that were hashed have been signed.

Illicit Message – Message Tampering

Now let us say that Babu passes the document on to Pamila. First, Pamila's software decrypts the signature (using Babu's public key), changing it back into a message digest. If this worked, then it proves that Babu signed the document, because only Babu has his private key. Pamila's software then hashes the document data into a message digest. If the message digest is the same as the message digest created when the signature was decrypted, then Pamila knows that the signed data have not been changed.

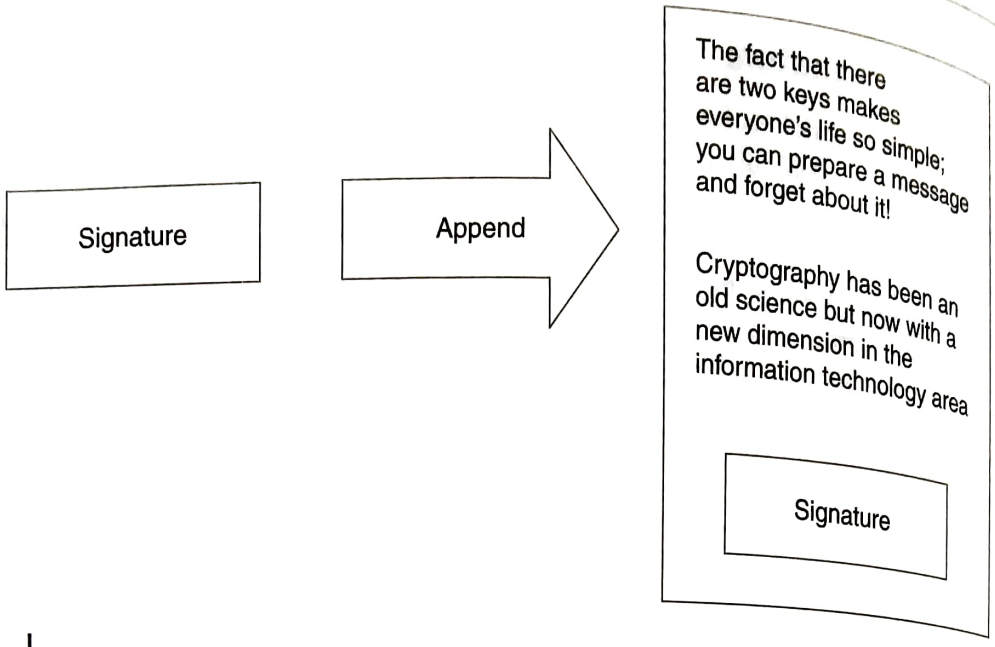


Figure 13.5 | Digital signature.

Now let us continue our original story and consider a possible complication in the situation. Suppose for some reason, D'Souza is disgruntled and decides to cheat Pamila. D'Souza makes sure that Pamila receives a signed message and a public key that appears to belong to Babu. Unknown to Pamila, D'Souza deceitfully sent a key pair he created using Babu's name. Short of receiving Babu's public key from him in person, how can Pamila be sure that Babu's public key is authentic? This is where the story takes an interesting turn! It just so happens that Sushant works at the company's certificate authority (CA) center. Sushant can create a digital certificate for Babu simply by signing Babu's public key as well as some information about Babu (see Figure 13.6).

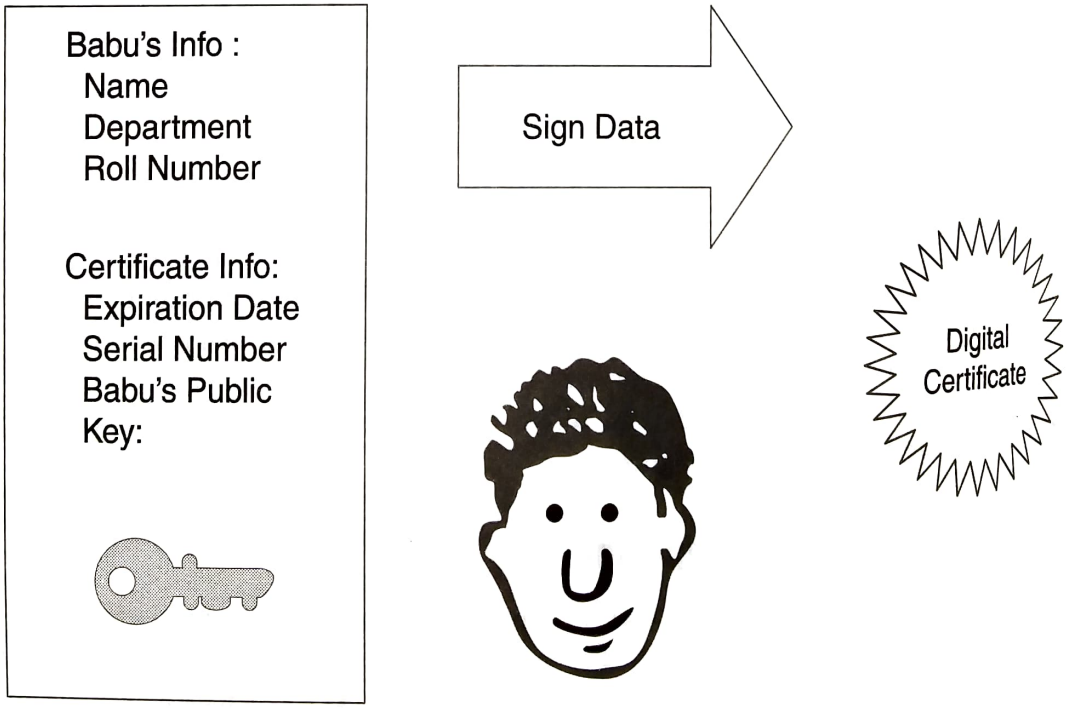


Figure 13.6 | Digital certificate.
Courtesy: http://www.youdzone.com/signature.htm courtesy David Youd.

Trusted Certificate

Now Babu's hostel mates can check Babu's trusted certificate to make sure that his public key truly belongs to him. In fact, no one at Babu's company accepts a signature for which there does not exist a certificate generated by Sushant. This gives Sushant the power to revoke the signatures if private keys are compromised, or no longer needed. There are even more widely accepted CAs that certify Sushant.

Now suppose Babu sends a signed document to Pamila. To verify the signature on the document, Pamila's software first uses Sushant's (the CA's) public key to check the signature on Babu's certificate. Successful de-encryption of the certificate proves that Sushant created it. After the certificate is de-encrypted, Pamila's software can check if Babu is in good standing with the CA and that all of the certificate information concerning Babu's ID has not been altered.

Use of Digital Certificate for Message Authentication

Pamila's software then takes Babu's public key from the certificate and uses it to check Babu's signature. If Babu's public key de-encrypts, that is, decrypts the signature successfully, then Pamila is assured that the signature was created using Babu's private key, because Sushant has certified the matching public key. And of course, if the signature is valid, then we know that D'Souza did not try to change the signed content. Although the scenario illustrated in this example may sound complicated, it is not. In reality, the specialized software for digital signature handles all this and the user (in this case Pamila) only has to click on it.

In summary of this illustrative section, we note that a digital signature is a message digest used to cryptographically sign a message. Digital signatures rely on asymmetric or public-key cryptography. We learn from the illustrative example above that to create a digital signature, you sign the message with your private key. The digital signature then becomes a part of the message. This has two effects:

1. Any changes to the message can be detected, owing to the message digest algorithm.
2. You cannot deny signing the message, because it was signed with your private key.

These two features, *message integrity* and *non-repudiation*, make digital signatures a very useful component for e-commerce applications. The example also brings forth a key point that a *message digest* is a number that is created algorithmically from a file and represents that file uniquely. If the file changes, the message digest also changes. In addition to allowing us to determine if a file has changed, message digests can also help to identify duplicate files or any other illicit action as seen in the scenario example. We shall discuss asymmetric and symmetric keys after addressing the topic of cryptographic algorithms.

13.6 Cryptographic Algorithms

The topic of cryptography and the algorithms involved in it is not only technically highly complex but also a very large one. It is not possible to cover everything in this chapter; only the basics are covered in this section to give an overall idea to the readers. Informative websites are mentioned in the *Further Reading* section for further exploration on the topic.

As the previous two chapters explained, modern networks are open and complex in nature. Of all the security mechanisms, cryptography is the one most suited to open and hostile environments where control is otherwise limited (recall Section 12.8 in Chapter 12 explaining the working of the Internet). Thus, cryptography is broadly applicable in today's modern, open broadcast, packet-switched heterogeneous networks. It helps us maintain the secrecy of the message, but the secrecy of the keys also needs to be kept in tact; this is done through what is known as *key management*.

Key Management

Readers have been introduced to the concept of 'key' through the illustrative scenario in Section 13.5. In that regard, 'key management' is an important concept to understand. It can be defined as the generation, recording, transcription, distribution, installation, storage, change, disposition and control of cryptographic key (such as the public and private keys whose use was illustrated in the Section 13.5). A brief overview of key management functions is provided in the next subsections.

Functions of Key Management

1. **Key generation:** This involves the selection of the number that is going to be used to tailor an encryption mechanism to a particular use. The use may be a sender and receiver pair, a domain, an application, a device or a data object. The key must be chosen in such a way that it is not predictable and that knowledge of it is not leaked in the process. Keys must be chosen randomly and in addition, they must not be disclosed at the time of the selection.
2. **Distribution:** Key distribution is the process of getting a key from the point of its generation to the point of its intended use. This problem is more difficult in symmetric key algorithms where it is necessary to protect the key from disclosure in the process (symmetric and asymmetric keys are discussed in the section 'Asymmetric and Symmetric Keys'). This step must be performed in a channel separate from the one that the traffic moves in (communication channels were discussed in the Chapter 12).
3. **Installation:** Key installation is the process of getting the key into the storage of the device or process that uses it. If this step involves manual operations, then such operations might result in leaking information about the key, key transcription errors or it might be so cumbersome as to discourage its use.
4. **Storage:** Keys need to be protected and the integrity of the storage mechanism itself is important. For example, the mechanism may be designed so that once the key is installed, it cannot be observed from outside the encryption machine itself. Some key-storage devices are designed to self-destruct when subjected to forces that might disclose the key or there are evidences that the key device is being tampered with. As another approach, the key may be stored in an encrypted form so that knowledge of the stored form does not disclose information about the behavior of the device under the key.
5. **Change:** This is about ending the use of one key and beginning that of another. This is determined by convention or protocol. Historical practices show that information about the key is prone to leakage during the key-change time. The longer the key is in use and the more traffic that is encrypted under it, the higher are the chances for its discovery and therefore more the traffic that will be compromised. This shows that there is value to key-changing practices.
6. **Control:** Controlling the key means the ability to exercise a directing or restraining influence over its content or use. For example, selecting which key from a set of keys is to be used for a particular application or party is a part of key control. Ensuring that a key which is intended for encrypting keys cannot be used for data is a part of key control. Key control is essential to the proper functioning of a key management system.

Asymmetric and Symmetric Keys

It is important that we emphasize the need for encryption in this section. Confidentiality of an electronic message is a necessity for e-commerce applications. As readers can appreciate by now, the primary method of achieving confidentiality is encryption. As discussed in Section 13.5, this involves creation of messages

by sender, initially in a form that is readable and understandable, and by other individuals as well if they have an access to the unencrypted message (cleartext also known as plaintext), etc., as already explained in Sections 13.1 and 13.2). Encryption is defined as the transformation of data, via a cryptographic mathematical process, into a form that is unreadable by anyone who does not possess the appropriate secret key (recall the story of Babu and his friends described in Section 13.5). It is the key that contains the binary code used to mathematically transform the message. True strength of the confidentiality service may depend on a number of variables associated with the encryption function listed as follows (a discussion on the encryption function is beyond the scope of this chapter):

1. the security protocol or application used to invoke the encryption function;
2. the trust in the platform executing the protocol or application;
3. the cryptographic algorithm;
4. the length of the key(s) used for encryption/decryption;
5. the protocol used to manage/generate those keys (recall the discussion in the section 'Key Management');
6. the storage of the secret keys (again recall the discussion mentioned above).

The strength of a cryptographic system usually increases as the key length increases. This is because a longer key length implies a larger number of possible keys. A 128-bit encryption has become the common practice; any key length less than 64 bits is no longer considered to be secure.

Secret Key Nuances

Most of the cryptographic algorithms use some form of secret keys to perform encryption functions. The differences in these keys are discussed as follows:

1. **Private/symmetric:** A private or symmetric key is a secret key that is shared between the sender and the receiver of the message. This key is usually the only key that can decipher the message (see Figure 13.7 and more details are provided in a later subsection 'Symmetric Key Cryptography').

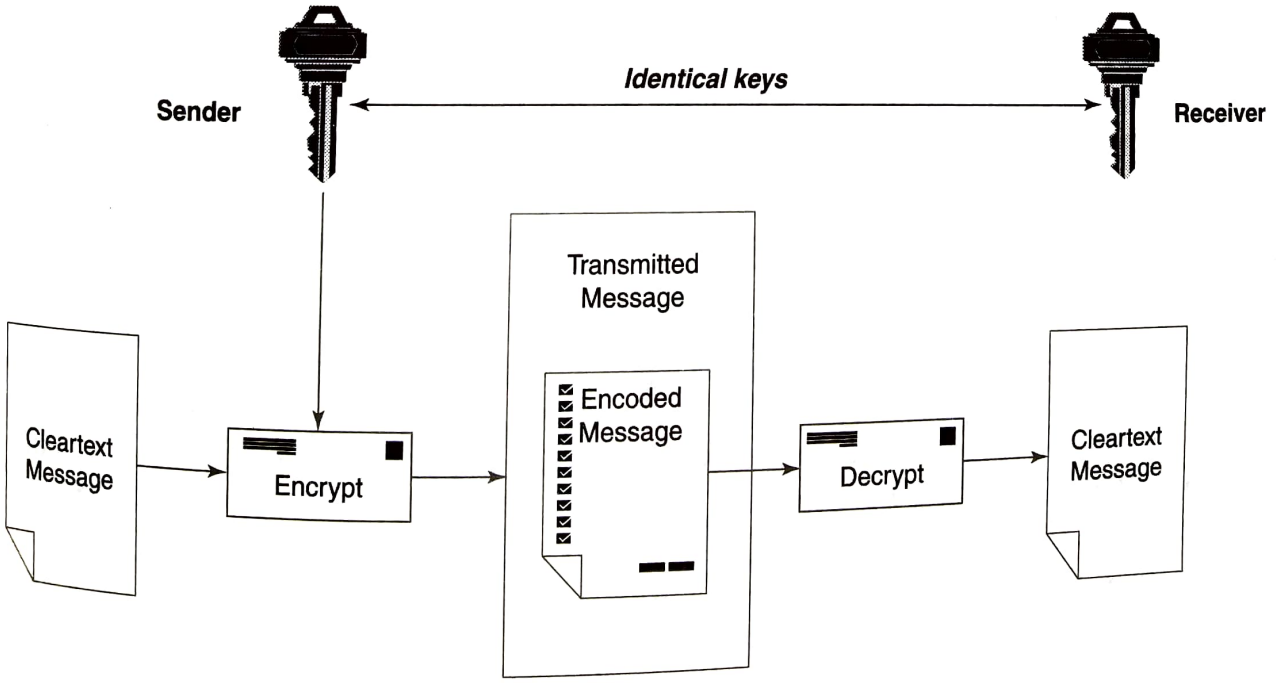


Figure 13.7 | Symmetric encryption method.

2. **Public/asymmetric:** A public or asymmetric key is the one that is made publicly available and can be used to encrypt data that only the holder of the uniquely and mathematically related private key can decrypt (asymmetric key cryptography details provided in the section 'Asymmetric Key Cryptography').
3. **Data/session:** This is a symmetric key, which may or may not be random or reused. It is used for encrypting data. Often, this key is negotiated using standard protocols or sent in a protected manner using a secret public or private key.
4. **Key encrypting:** These are the keys that are used to protect data encrypting keys. These keys are usually used only for key updates and not data encryption.
5. **Split keys:** To protect against intentional or unintentional key disclosure, it is possible to create and distribute parts of larger keys that only together can be used for encryption or decryption.

Symmetric Key Cryptography

As depicted in Figure 13.7, in a system that uses symmetric cryptography, both parties (sender and receiver) use the same key for encryption and decryption. This provides dual functionality to the key. Symmetric keys are also called 'secret keys' because this type of encryption relies on each user to keep the key a secret and properly protected. If this key falls into the wrong hands, that is, an intruder, the intruder would have the ability to decrypt any intercepted message encrypted with this key. Thus, the security of the symmetric encryption method is completely dependent on how well users protect the key (recall the discussion in the section 'Key Management').

Asymmetric Key Cryptography

In the traditional and conventional cryptography, symmetric cryptography has been the practice. However, the need for asymmetric cryptography arose from the imperative to provide a greater security to users. The development of asymmetric key cryptography has an interesting history; in 1976, Whitfield Diffie and Martin Hellman pointed out that although the relationship between the two keys (key for encrypting and key for decrypting) must be fixed, it need not be equality. Thus, the idea of asymmetric key cryptography was born (their work is famously known as *Diffie–Hellman algorithm*).

In this kind of cryptography, the key has two parts that are mathematically related to each other in such a way that what is encrypted with one part can be only decrypted with the other. In asymmetric cryptography, the idea is that the value of one of the keys does not necessarily imply the other, that is, it is not possible to calculate one from the other. Thus, one of the keys, plus a message encrypted under it, does not imply the other key. From a message and one part of the key, it is mathematically possible to calculate the other but it is not computationally feasible to do. This explains why asymmetric cryptography adds to the complexity. Now let us understand in detail the working of asymmetric cryptography.

Only one part, called the *private key*, needs to be kept secret. The other part, the *public key*, is published to the world. Anyone can use the public key to encrypt a message, but the message can only be decrypted and read by the owner of the private key. Conversely, anyone can read a message encrypted with the private key, but only the person with a beneficial use of that key could have encrypted it. Now let us consider the following scenario that is depicted in Figure 13.8:

1. The sender determines a secret value *a*.
2. A related value, *A*, is derived from *a*. *A* is made *public*.
3. The receiver determines a secret value *b*.
4. Another related value, *B*, is derived from *b*. *B* is made *public*.
5. The Diffie–Hellman algorithm is used to calculate a secret key corresponding to the key pairs (a, B) and (b, A).

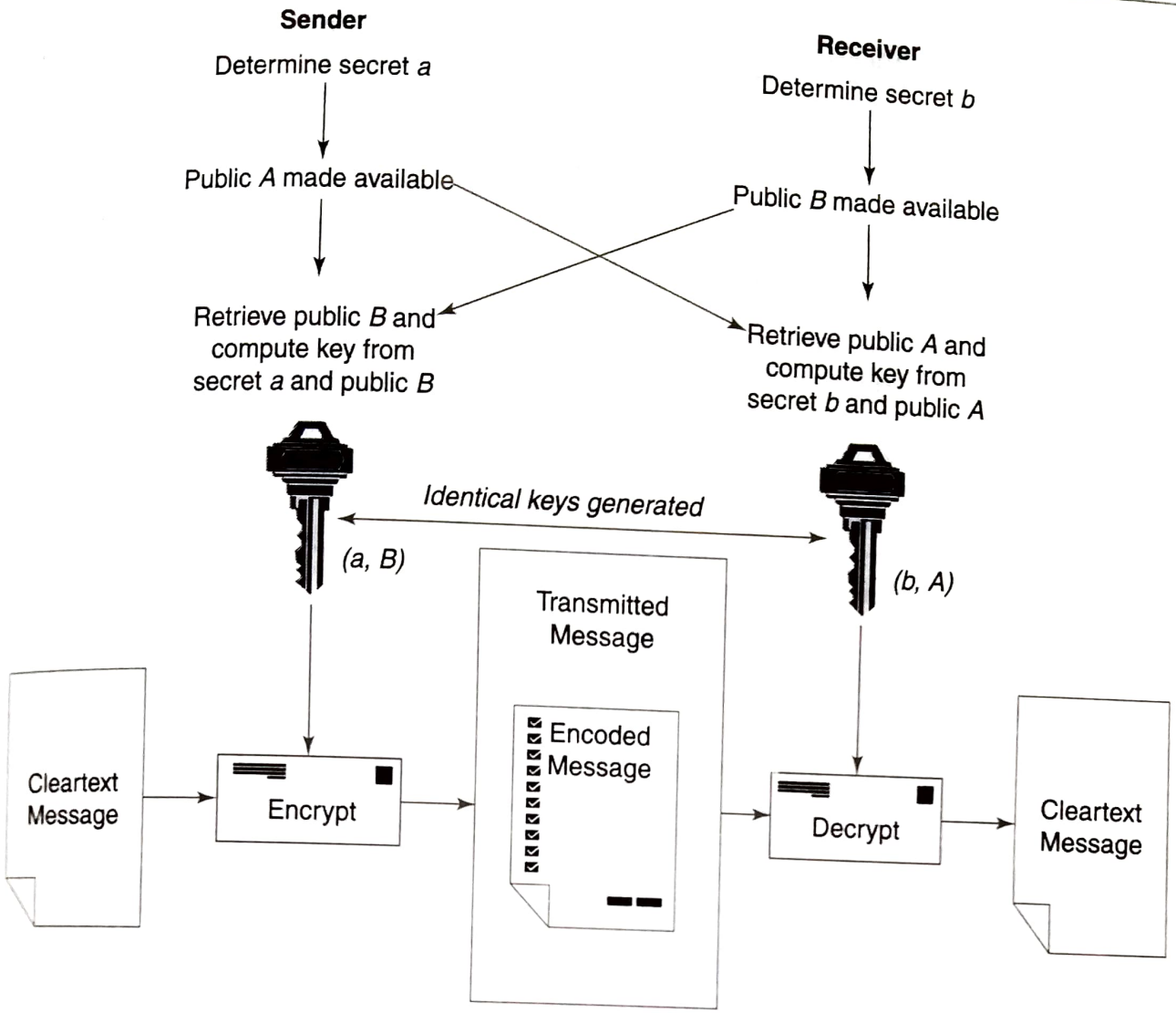


Figure 13.8 | Asymmetric encryption method.

Relative advantages and disadvantages of symmetric and asymmetric encryption methods are represented in Table 13.1. There are many other techniques for cryptography but these are beyond the scope of this chapter; most well-known among them is the technique of *steganography* (see Box 13.3).

Table 13.1 | Cryptographic methods: advantages and disadvantages

Symmetric cryptography	Asymmetric cryptography
Strengths:	Strengths:
1. Much faster than asymmetric systems	1. Better key distribution than symmetric systems
2. Hard to break with a large key size	2. Better scalability than symmetric systems
	3. Can provide authentication and non-repudiation
Weaknesses:	Weaknesses:
1. Requires secure delivery mechanism	1. Works more slowly than symmetric systems
2. Key management can become overwhelming	2. Involves mathematically intensive tasks
3. Does not provide authenticity or non-repudiation	

Box 13.3 Steganography – The Art of Hiding!

Given the sheer volume of data stored and transmitted electronically in the world today, it is no surprise that countless methods of protecting such data have evolved. One lesser known but rapidly growing method is *steganography*, the art and science of hiding information so that it does not even appear to exist!! Talking of lesser known methods, readers will recall the discussion about 'data obfuscation' in Chapter 5.

Steganography is a Greek word that means 'sheltered writing'. It is a method that attempts to hide the existence of a message or communication. This idea of data hiding is not a novelty; it has been used for centuries all across the world under different regimes. The practice dates back to ancient Rome and Greece, where the messages were etched into wooden tablets and then covered with wax or created by shaving a messenger's head and then tattooing a secret message on it, letting his hair grow back and then shaving it again after he arrived at the receiving party to reveal the message. Note that there is a slight difference in steganography and cryptography – steganography is the art and science of writing hidden messages in such a way that no one apart from the intended recipient knows of the existence of the message; this is in contrast to cryptography, where the existence of the message itself is not disguised, but the content is obscured. It is said that terrorists use steganography *techniques* to hide their communication in images on the Internet; most popular images are used such as those of film actresses or other celebrities. In its basic form, steganography is simple. For example, say every fourth letter of a memo could hide a message. This simple technique has an added advantage over encryption in that it does not arouse suspicion, that is, there is not much scope for getting started an investigation! Presence of an encryption could set off an investigation, but a message hidden in plain sight would get ignored.

The term 'cover' or 'cover medium' is used to describe the original, innocent message, data, audio, still, video and so on. It is the medium that hides the secret message. It must have parts that can be altered or used without damaging or noticeably changing the cover media. If the cover media are digital, these alterable parts are called 'redundant bits'. These bits or a subset can be replaced with the message that is intended to be hidden. Interestingly, steganography in digital media is very similar to 'digital watermarking'. In other words, when steganography is used to place a hidden 'trademark' in images, music and software, the result is a technique referred to as 'watermarking' (see the URLs quoted in the *Further Reading* section).

Courtesy: http://www.ansinet.org/fulltext/itj/itj33245-269.pdf#search='STEGANOGRPHY'.

Quantum Cryptography

A discussion on cryptography cannot be complete without introducing the topic of *quantum cryptography* (QC). Let us understand about this latest development. Ample links are provided in the *Further Reading* section to guide the readers for exploring greater details on this topic.

QC was proposed, in the early 1970s, first by Stephen Wiesner, and then at Columbia University in New York. QC uses quantum mechanics for secure communications. Unlike traditional cryptography, which employs various mathematical techniques to restrict eavesdroppers from learning the contents of the encrypted messages, QC is based on the 'physics of information'. *Eavesdropping* can be viewed as measurements on a physical object – in this case, the carrier of the information. Using quantum phenomena such as quantum superpositions or quantum entanglement, one can design and implement a communication system that can always detect eavesdropping. This is because measurements on the quantum carrier of information disturb it and therefore leave traces (Box 13.4).

QC is an effort to allow two users of a common communication channel to create a body of shared and secret information. This information, which generally takes the form of a random string of bits, can then be used as a conventional secret key for secure communication. It is useful to assume that the communica-

Box 13.4 IBM Contribution – Quantum Leap in Cryptography!

Early in the twentieth century, scientists realized that very small objects – atoms, electrons and other subatomic particles – do not obey the same laws of motion (mechanics) as everyday objects do. They devised a new set of laws, called quantum mechanics, to describe the behavior of these very small objects. Quantum mechanics is phrased in terms of randomness and probabilities of events occurring. In 1927, the German physicist Werner Heisenberg postulated the basic law of physics, now famous as the 'Principle of Uncertainty'. Scientists are using photons to develop data-security systems that may prove to be the ultimate defense against eavesdropping hackers. Whereas classic public-key cryptography relies on the computational difficulty of certain hard mathematical problems (such as integer factorization) for key distribution, quantum cryptography (QC) relies on the laws of quantum mechanics, a subset of quantum physics. Quantum cryptographic devices typically employ individual photons of light and take advantage of either the *Heisenberg Uncertainty Principle or quantum entanglement*. Charles H. Bennett, a fellow at IBM's *Thomas J. Watson Research Center*, and Gilles Brassard, a researcher at the University of Montreal in Canada, first devised QC in 1984 as a part of their study of the relationship between physics and information.

Principle of Uncertainty

The act of measurement is an integral part of quantum mechanics, not just a passive, external process as in classic physics. So it is possible to encode information into some quantum properties of a photon in such a way that any effort to monitor them necessarily disturbs them in some detectable way. The effect arises because in quantum theory, certain pairs of physical properties are complementary in the sense that measuring one property necessarily disturbs the other. This statement is known as the *Heisenberg Uncertainty Principle*. The two complementary properties that are often used in QC are the two types of a photon's polarization, for example, rectilinear (vertical and horizontal) and diagonal (at 45 and 135).

Principle of Entanglement

It is a state of two or more quantum particles, for example, photons, in which many of their physical properties are strongly correlated. The entangled particles cannot be described by specifying the states of individual particles and they may together share information in a form that cannot be accessed in any experiment performed on either of the particles alone. This happens no matter how far apart the particles may be at the time.

The Observed and the Altered

The Principle of Uncertainty (Heisenberg developed the theory around the Uncertainity Principle) holds that the mere act of observing or measuring a particle will ultimately change its behavior. At macroscopic levels, humans do not notice this law. Put your leg inside a magnetic resonance imaging (MRI) machine, for example, and it does not come out noticeably different. But at the atomic level, the MRI's application of strong magnetic forces alters the trajectory and spin of the electrons that are orbiting atoms inside your body.

Courtesy:

1. http://en.wikipedia.org/wiki/Quantum_cryptography (accessed 4 November 2006).
2. http://www.businessweek.com/technology/content/jul2003/tc20030715_5818_tc047.htm (accessed 4 November 2006).

parties initially share a small amount of secret information, which is used up and then renewed in the exchange process, but even without this assumption, exchanges are possible. The advantage of QC over the traditional key exchange methods is that the exchange of information can be shown to be secure in a very strong sense, without making assumptions about the intractability of certain mathematical problems. Even when assuming hypothetical eavesdroppers with an unlimited computing power, the laws of physics guarantee (probabilistically) that the secret key exchange will be secure, given a few other assumptions.

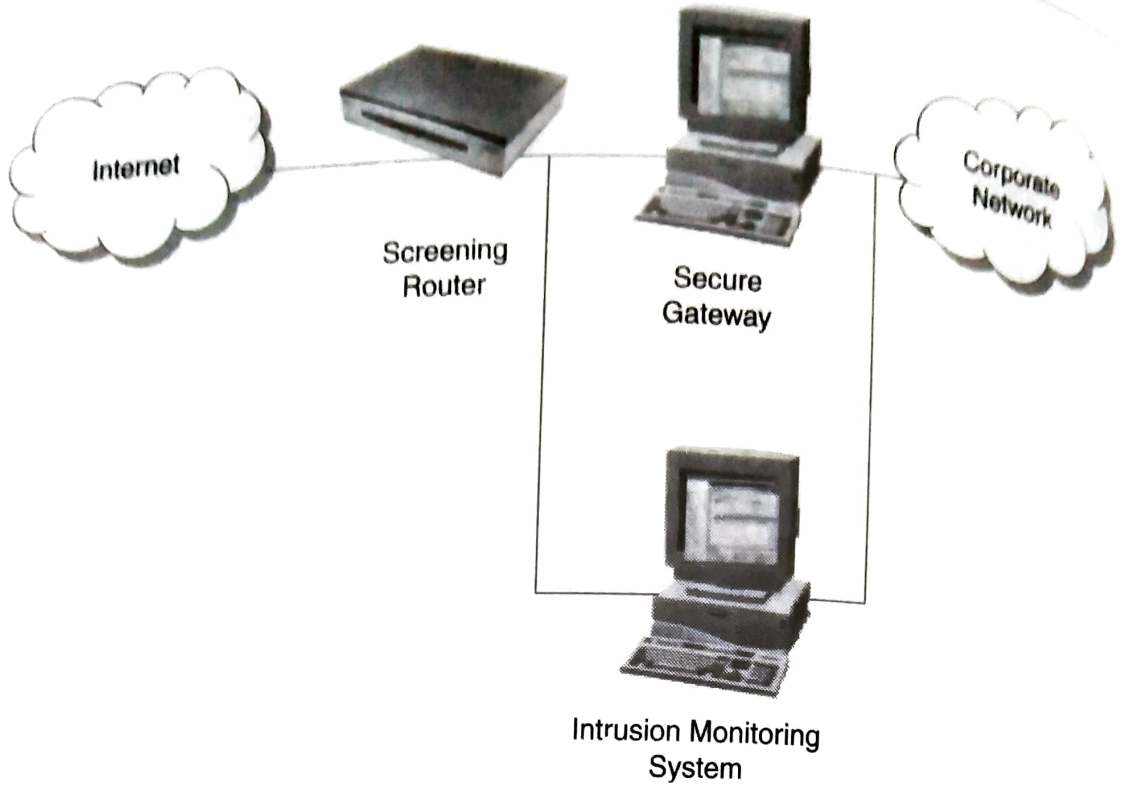


Figure 14.2 | Intrusion monitoring system.

14.4 Intrusion Detection for Information Systems Security

Having understood the need for IDSs, let us understand what IDSs are. As said in the introduction section, discussion about IDS in context of physical security was covered in Chapter 8. The purpose here is to discuss IDSs in a network security paradigm. An IDS inspects all inbound and outbound network activities. It can be set up to identify any suspicious network activity patterns that may indicate a network or system attack. Unusual patterns that are known to generally attack networks can signify someone attempting to break into the network system or trying to compromise the system.

The IDS can be hardware- or software-based security service that monitors and analyses system events for the purpose of finding and providing real-time or near real-time warning of events that are identified by the network configuration to be attempts to access system resources in an unauthorized manner. Typically the monitoring and warning is done by examining the network vulnerability scans (detailed discussion about vulnerability scanning is available in Chapter 35). There are a number of good network vulnerability tools available in the market; the *Further Reading* section points readers to some of those. Essentially network ports are scanned to assess if any potential vulnerabilities can be seen.

Conceptual Approaches to Intrusion Detection Methodologies

There are two popular approaches available currently to intrusion detection methodology: *knowledge-based IDSs* (also referred as *signature-based IDS*) and *behavior-based IDSs* (also referred as *anomaly-based IDS*).

Knowledge-based IDSs are more common than behavior-based IDSs. Knowledge-based IDS use a database of previous attacks and known system vulnerabilities to look for current attempts to exploit their vulnerabilities and trigger an alarm if a vulnerability is found. On the other hand, behavior-based IDSs take a dynamic approach in the sense that they detect deviations from the learned patterns of user behavior. An alarm is triggered when any activity that is considered outside of normal system use takes place. These types

...ms are more sophisticated but less common than the knowledge-based IDS. The relative advantages ...dvantages of both approaches are summarized in Table 14.1.

...ble 14.1 | Knowledge-based IDSs versus behavior-based IDSs: advantages and disadvantages

...owledge-based intrusion detection	Behavior-based intrusion detection
...vantages:	**Advantages:**
...False alarm rates are low, that is, intrusion ...ggers are more reliable	1. The systems can dynamically adapt to new or unique vulnerabilities
...Intrusion alarms are standardized and are clear ...w security personnel to understand	2. Systems are not dependent on specific operating systems (OSs) such as in the case of knowledge-based systems
...sadvantages:	**Disadvantages:**
...Systems place high demands on resources; knowledge databases need continuous maintenance and updates	1. False alarm rates are high; this can create high data noise, at times making the systems unusable
...New or unique attacks may go unnoticed if they happen not to be captured in the network attack history database	2. The activity and behavior of the users on the network may not be static enough to warrant an effective implementation of this type of system

Categories of Intrusion Detection System

Apart from the dichotomous division mentioned above, there are many ways in which an IDS can be categorized as follows depending on its use:

1. **Misuse detection:** Here, the IDS analyses the information it gathers and compares it to the databases of attack signatures. To be effective, this type of IDS depends on the attacks that have already been documented. Like virus detection systems, software for misuse detection is only as good as the databases of attack signatures that it can use to compare packets.
2. **Anomaly detection:** In this type of detection system, a baseline is established. It consists of things such as the network's traffic load state, breakdown, protocol and typical packet size. With anomaly detection, sensors monitor network segments to compare their present state against the baseline in order to identify anomalies.
3. **Network-based IDS (NIDS):** NIDSs monitor network traffic and uncover possible attacks or suspicious activities. In an NIDS, the IDS sensors evaluate the individual packets that are flowing through the network. The NIDS detects malicious packets that are designed by an attacker to be overlooked by the simplistic filtering rule of many firewalls. There is, however, a problem with an NIDS – it will not detect attacks against a host made by an intruder who is logged in at the host's terminal.
4. **Host-based IDS (HIDS):** HIDSs can be installed on individual workstations and/or servers to watch for an inappropriate or anomalous activity and insider attacks. They are usually used to make sure that the users do not accidentally delete system files, reconfigure important settings or put the system at risk in any other way. In an HIDS, the IDS examines the activity on each individual computer node or host. The kinds of items that are evaluated include modifications to important system files, abnormal or excessive central processing unit (CPU) activity and misuse of root or administrative rights.

5. **Passive IDS:** In a passive system, the IDS detects a potential security breach, logs the information and signals and alerts. Here, no direct action is taken by the system.
6. **Reactive IDS:** In a reactive system, the IDS can respond in several ways to the suspicious activity such as by logging a user off the system, closing down the connection or even reprogramming the firewall to block network traffic from the suspected malicious source.

Although vulnerability scanning and IDS are related, it is important to note that there is a subtle difference between the two concepts involved. It is best to explain this with a simple example; suppose we compare securing a network to securing a home; an IDS would be the burglar alarm. An IDS is triggered when someone attempts to enter your network. A vulnerability scanner, on the other hand, is like the home security consultant. Its role is to proactively examine the home, or a network, looking for vulnerabilities including various entry points, the integrity of the firewalls and IDS systems and so on. Network penetration testing is also important for securing the network systems (see Box 14.2).

Box 14.2 Network Penetration Testing

As a generic concept, a penetration test is the process of actively evaluating your information security measures. There are a number of ways by which this can be undertaken, but the most common procedure is that the security measures are actively analyzed for design weaknesses, technical flaws and vulnerabilities; the results are then delivered comprehensively in a report to executive management and technical audiences.

There are several reasons why organizations choose to perform a penetration test; they range from technical to commercial but the most common are the following:

1. Identify the threats facing your organization's information assets so that you can quantify your information risk and provide an adequate InfoSec expenditure.
2. Reduce your organization's information technology (IT) security costs and provide a better return on IT security investment by identifying and resolving vulnerabilities and weaknesses. These may be known vulnerabilities in the underlying technologies or weaknesses in the design or implementation.
3. Provide your organization with assurance – a thorough and comprehensive assessment of organizational security covering policy, procedure, design and implementation.
4. Gain and maintain certification to an industry regulation [BS7799, Health Insurance Portability and Accountability Act (HIPAA), etc. – these standards are discussed in Part V].
5. Adopt best practice by conforming to legal and industry regulations.

Different types of penetration tests are available – internal security assessment, application security assessment, wireless/remote access server (RAS) assessment, telephony security assessment and social engineering.

Readers may like to take a note that further discussion on penetration testing and vulnerability scanning is available in Chapter 35.

Characteristics of a Good Intrusion Detection System

In the light of the discussion so far, we can see that an IDS needs to address several issues in the interest of the security of the networks, a modern-day bastion of information systems (IS). There are many mechanisms for deploying the IDSs. However, regardless of the mechanisms on which they are based, the following are essential:

1. **Uptime:** Smooth and continuous running with minimal human intervention. It should run in the background. The internal working should get examined from outside, so it is not a black box.
2. **Fault tolerance:** This is needed for sustaining a system crash. Its knowledge base (recall the discussion on knowledge-based IDS in the section 'Conceptual Approaches to Intrusion Detection Methodologies') should not require a rebuild from restart.
3. **Robustness:** The IDS must be 'robust', that is, difficult to sabotage. The system should be self-healing in the sense that it should be able to monitor itself for suspicious network activities that might signify attempts to weaken the detection mechanism or shut it off.
4. **Performance:** This is always a critical concern. Without a good performance, an IDS will not get effectively used.
5. **Easy configurability:** Configuration of the IDS should be easy. This is important because every system has a different usage pattern, and the defense mechanism should adapt easily to these patterns.
6. **Easy adaptability:** Given the constantly changing network environment in today's business dynamics, the IDS should be like a chameleon in its ability to adapt to the changing environment. At the same time, it should stay current with the system as it changes, that is, new applications added, upgrades and any other modifications. In other words, the IDS must adapt to the changes of the system.
7. **Built-in defense mechanism:** An IDS must have built-in defense mechanisms, and the environment around it should be hardened to make it difficult to fool, that is, minimum opportunity for generating 'false alarm' or assurance on the positives, that is, a trigger is generated only in genuine situations.

Role of Router in Intrusion Detection System

This is a very intricate topic but we must have a basic understanding of it. Readers will recall that in Chapter 12, routers were discussed. Traditional network security technologies tend to focus on securing the perimeter with IDSs, antivirus software and firewalls. The security from within must also be considered: the network infrastructure – switches and routers – can play a crucial role in the network security. Security should exist between each of the seven layers [open system interconnections (OSI) layers 1–7] of the network (the OSI model is explained in Chapter 12). A layered approach to network security allows the creation of multiple layers of defense around key assets. Many switches and routers come with a rich set of security features. Knowing what they are, why they should be activated and how they should be deployed can result in security.

In the section 'Categories of Intrusion Detection System', we mentioned that an IDS can be a passive or a reactive system. In the case of a reactive type of IDS, it can respond reactively (communicate with firewalls/routers to block packets from a presumed attacker's IP net address). The classic IDS tools are NIDS and HIDS. They spot known attack indicators, that is, *attack signatures*. That works if attack network methods can be predicted in advance, that is, 'spotting known bad behavior'.

Apart from the NIDS and HIDS, there are other network components that also help, for example, routers and firewalls (dealt with in the Chapter 15) that block the data packets. The data thrown by routers also need to be considered. When using the IDS facility with other security features, it is important to note that a packet is subject to intrusion detection only if the router actually attempts to forward it. Packets dropped by an inbound access list, for example, would not be scanned. Also, if a packet is scanned and multiple, different signatures are detected, only the first one found is reported by the IDS.

Router Security Considerations

First, let us revisit the role of routers in networks. In Chapter 12, working of networks was explained. In larger, more complex computer networks, data need to be directed carefully. In almost all cases, large

networks are actually composed of a collection of LANs that are interconnected or 'internetworked'. This is where routers come in. Routers take network data messages from a LAN and convert them into packets suitable for transmission beyond the LAN on a wide area network (WAN). The goal is almost always to get these packets to another LAN and ultimately to the correct host on that LAN. Part of the 'conversion' process is to add a packet header. Other routers will generally only look at a packet's header information, not at the contents or data in the packet.

Routers also make decisions about where to send these packets, based on the addresses contained within the packet headers and a table of routes maintained within the router. Updating these routing tables and forwarding data packets between portions of a network are two of the primary tasks of a router. Building packets and unwrapping packets are additional router functions performed by the first and last routers, respectively, that a message passes through. In addition to directing packets, a router may be responsible for filtering traffic, allowing some packets to pass through and rejecting others. Filtering can be a very important function of routers; it allows them to help protect computers and other network components.

One of the tasks of a router is to maintain route tables that are used to decide where a packet is to go and thus which interface it should be sent out to. In the past, these tables were built and updated by hand and this is referred as static routing. In dynamic routing, the router learns about where various addresses are relative to itself and builds up route tables based on this information. There are a number of schemes or routing protocols for the routers to acquire and share route table information; however, a thorough treatment of the details is beyond the scope of this chapter.

As we understand, routers are an important part of a network, and their security is a vital part of the overall security for the networks they serve. What does it mean for a router to be secure? One simple way to define the security of a router is the following: do the operation, configuration and management of the router to satisfy your security policy. Figure 14.3 provides a layered view for router security.

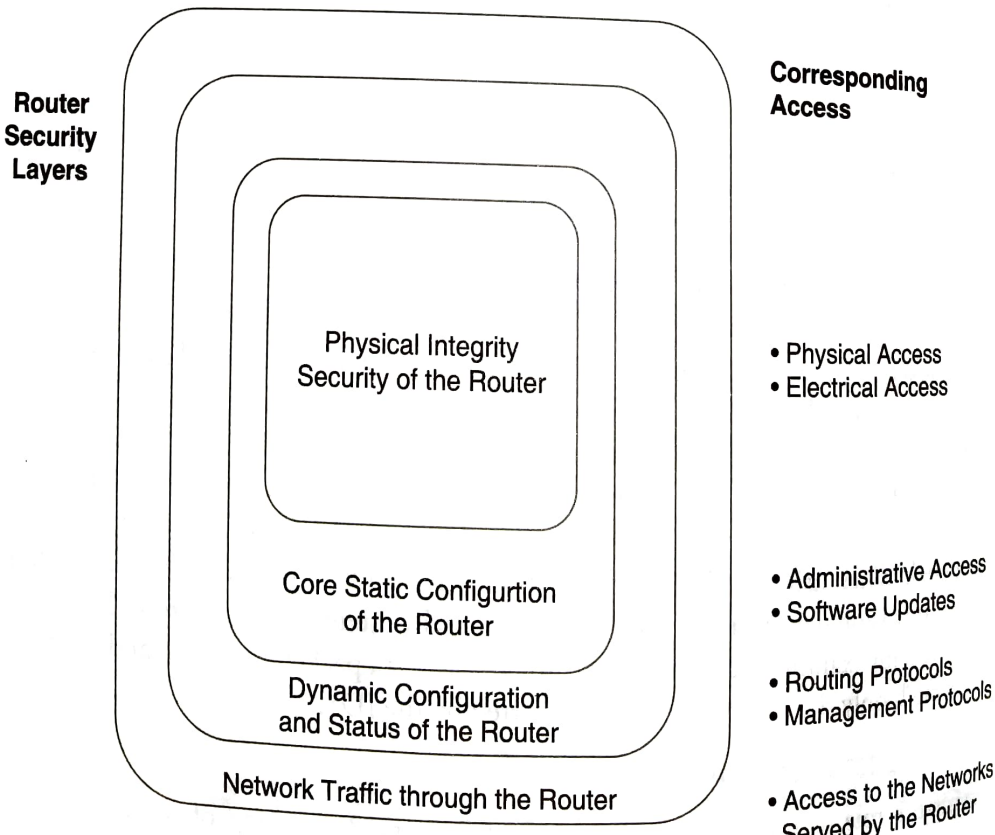


Figure 14.3 | Router security: Layered view.

Router security can be considered from two angles: one is its physical security and the other is its logical security, that is, its security through software by proper configuration of its features. In terms of the physical security of the router, the room that contains the router should be free of electrostatic or magnetic interference. It should have controls for temperature and humidity. If deemed necessary for availability or criticality reasons, an uninterrupted power supply (UPS) should be installed and spare components and parts kept on hand. Also, the router should be placed in a locked room with access by only a small number of authorized personnel. Finally, physical devices [e.g., Personal Computer Memory Card Industry Association (PCMCIA) cards and modems] used to connect to the router require storage protection. In terms of a router's logical security, to aid in protecting against some denial of service (DoS) attacks, and to allow it to support the widest range of security services, the router should be configured with the maximum amount of memory possible. For security purpose, each router should have a unique name to identify it, and each interface should have a unique network address associated with it. Also, basic security settings should be established on any router before it is connected to an operational network. Using a firewall and a router together can offer better security than either one alone.

Today's Challenges for Intrusion Detection Systems

Most of today's IDS products are focused on *signature detection* (see Box 14. 3). Typically, today's IDSs are designed for sub-100 Mbps shared media network environments, employing detection capabilities introduced a few years ago. IDS products have failed to keep up with the rapid advancement in switching and bandwidth growth and the increased sophistication of attacks – as well as the sheer volume of attacks that needs to be handled today. Current IDS products often operate in a monitoring-only mode – for example, 'sniffers', which can detect attacks but cannot effectively and reliably block malicious traffic before the damage is done. Network security managers deploying IDS products today face a number of challenges:

1. **Inaccurate detection:** IDS products' detection capabilities can be characterized in terms of accuracy and specificity. *Accuracy* is often measured in *true detection rate* – sometimes referred as the *false negative rate* – and the *false positive rate*. The true detection rate specifies how successful a system is in detecting attacks when they happen. The false positive rate tells us the likelihood that a system will misidentify benign activity as attacks. *Specificity* is a measure of how much detailed information about an attack is discovered when it is detected. It is said by security experts that IDS products today are lacking in both accuracy and specificity and generate too many 'false positives', alerting security engineers of attacks, when nothing malicious is taking place. In some cases, IDS products have delivered tens of thousands of 'false positive' alerts a day. There is nothing more corrosive to network vigilance than a jumpy security system that is continually issuing false alarms.
2. **Incomplete attack coverage:** IDS products typically focus on signature or anomaly or DoS detection. Network security managers have to purchase and integrate point solutions from separate vendors, or leave networks vulnerable to attack. See Box 14.3 and Figure 14.4.

Box 14.3 Understanding Signature Detection, Anomaly Detection and Denial of Service Detection

1. **Signature detection – to protect against 'known' threats:** In Section 14.2, we have mentioned about the 'stages of network attacks'; hackers often attack networks through tried and tested methods from previously successful assaults. These attacks have been analyzed by network security vendors and a detailed profile, or *attack signature*, has been created. Signature *detection techniques* identify network assaults by looking for the attack 'fingerprint' within network traffic and matching

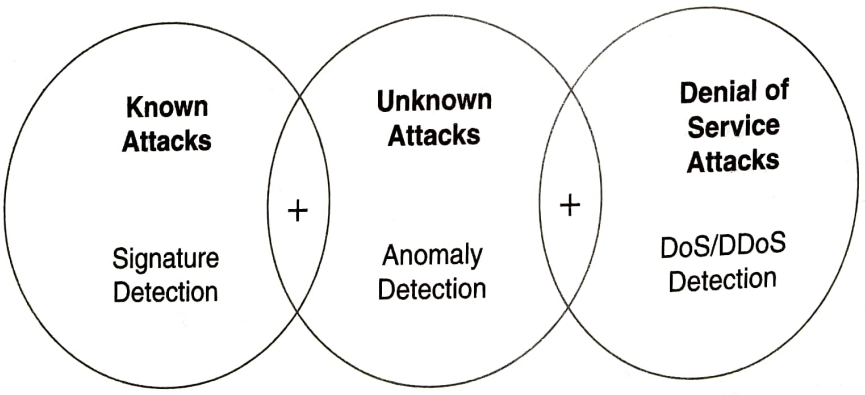
Box 14.3 *Continued...*

against an internal database of known threats. Once an attack signature is identified, the security system delivers an attack response, in most cases a simple alarm or alert. Success in preventing these attacks depends on an up-to-the-minute database of attack signatures, compiled from previous strikes. The drawback to the systems that rely mainly, or only, on signature detection is clear: they can only detect attacks for which there is a released signature. If signature detection techniques are employed in isolation to protect networks, infrastructure remains vulnerable to any variants of known signatures, first strike attacks and DoS attacks.

2. **Anomaly detection**: These techniques are required when hackers discover new security weaknesses and rush to exploit the new vulnerability. When this happens, there are no existing attack signatures. The 'Code Red' virus is an example of a new attack, or first strike, which could not be detected through an available signature. In order to identify these first strikes, IDS products can use anomaly detection techniques, where network traffic is compared against a baseline to identify abnormal and potentially harmful behavior. These anomaly techniques are looking for statistical abnormalities in the data traffic, as well as protocol ambiguities and atypical application activity. Today's IDS products do not generally provide enough specific anomaly information to prevent sophisticated attacks and if used in isolation, anomaly detection techniques can miss attacks that are only identifiable through signature detection.
3. **DoS detection**: The objective of DoS and distributed DoS attacks is to deny legitimate users access to critical network services. Hackers achieve this by launching attacks that consume excessive network bandwidth, host processing cycles or other network infrastructure resources. DoS attacks have caused some of the world's biggest brands to disappoint customers and investors as websites became inaccessible to customers, partners and users – sometimes for up to 24 h. IDS products often compare current traffic behavior with acceptable normal behavior to detect DoS attacks, where normal traffic is characterized by a set of preprogrammed thresholds. This can lead to *false alarms* or attacks being missed because the attack traffic is below the configured threshold.

Courtesy:

1. Papers on intrusion detection systems by Dr. Fengmin Gong, Intru Vert Networks, Inc.
2. IDS document by Rebecca and Peter Mell – published by NIST.
3. Paper by Andree Yee – NFR Security Inc (July 2003).



DDoS – Distributed denial of service attacks

Figure 14.4 | Types of attacks.

3. **More detection, less prevention:** Most of the systems nowadays concentrate on attack detection. Responding to attacks is a reactive activity, often too late to thwart the intrusion.
4. **Designed primarily for sub-100 Mbps networks:** Some security professionals are of the opinion that solutions have simply not kept up with the speed and sophistication of network infrastructure, and cannot accurately monitor higher speed or switched networks.
5. **Performance challenged:** Software applications running on general-purpose PC/server hardware do not have the processing power required to perform a thorough analysis. These underpowered products result in inaccurate detection and packet dropping, even on low-bandwidth networks.
6. **High-availability deployment not available:** Single port products are not able to monitor asymmetric traffic flows. Also, with networks becoming a primary mechanism to interact with customers and partners, forward-thinking organizations have developed backup systems should their current infrastructure fail in any way. The inability of current IDS products to cope with server failovers renders them virtually useless for any mission-critical network deployment.
7. **Scalability issues:** Today's IDS products are primarily designed for low-end deployments; as such, they do not scale for medium and large enterprise or government networks. Here, monitored bandwidth, the number of network segments monitored, the number of sensors needed, alarm rates and the geographical spread of the network exceed system limits.
8. **Security policy enforcement-related issues:** Current products generally support the selection of only one security policy for the entire system, even though the product may monitor traffic belonging to multiple administrative domains – in an enterprise, this could be the finance, marketing or human resource (HR) functions. This 'one size fits all' approach is no longer acceptable for organizations that require different security policies for each function, business unit or geography.
9. **Require significant IT resources:** IDS products today require substantial hands-on management – for example, the simple task of frequent signature updates can take up a lot of time and skilled engineering resources, delivering a very high total cost of ownership.

Implementing IDS

Given the challenges (described above) for today's IDSs, it is a good idea to discuss what it takes to implement them, the precautions, pros and cons, etc. An IDS is not just one component but a well-orchestrated system comprising the underlying technology, personnel and policy components along with the high-level components of incident handling scheme. This section offers some thoughts on planning and implementation. It may so happen that an IDS gets created in an organization because the top management said so; in that case, all that is required is 'convincing-looking-box-with-flashing-lights'. However, this is not the real purpose of implementing the IDS. Organizations want real data on 'incident' in terms of attempts by a perpetrator to break in and meaningful reports generated out of that data for decision-making use by the management.

The steps in implementing an IDS translate to planning and commissioning, event analysis and finally maintaining the IDS team. At the planning phase, organizations need to consider if they have an effective security policy that would support IDS deployment. For example, if an employee has acted inappropriately, would it be possible to prove that his/her action was not legitimate? The intrusion detection analysts and incident handling team may be wasting their time if this is not possible. When a major incident has taken place, does your policy guide you if it should or should not be informed to the media and stakeholders? It is only under policy guidance that the organizations can take a decision as to what circumstances call for waking up the owner, are there mechanisms to contact the Chief Information Officer (CIO) out of office hours when a major incident takes place, what escalation procedures are in place, where are the system and data backups and does the incident handler have the authority to shut down the organization's electronic

commerce (e-commerce) server if required? These are only a few examples that tell us why security policy is important. Readers should recollect that the topic of security policy was introduced in Chapter 4.

Event analysis is the next crucial step in IDS implementation. Once the IDS system is built and commissioned, the first task is to 'tune' it. A crucial responsibility is to recognize *false positive events* – correctly triggered but not indicative of network intrusion. Even after tuning is done, rapid isolation of these 'false positives' is crucial to identify those events that are genuinely significant, especially inside 'not known good' data. Another useful skill required for the IDS team is recognizing the 'traces' left by common attacks. The required skills need to be learnt thoroughly; training on penetration testing often helps here. All this typically requires generating effective test scenarios.

Finally comes the last step in IDS implementation: maintenance of the IDS team, that is, retaining the team members. Here, an organization's skill to encourage a creativity-supporting atmosphere comes into picture, especially given the lateral shift opportunities available to technical IS professionals, especially in India where most youngsters do not like to remain in the same job for a long time! Management needs to support their 'innovative thinkers'. It is no good penalizing individuals for 'rocking the boat'. In today's multi-skill technically challenging environment, it is no good for the organizations to expect that they can hold all the necessary skills and so, professional networking with other peer organizations is desirable within the provisions of intellectual property rights (IPRs).

The Future of Intrusion Detection Systems

IDSs, much like the security industry itself, have grown rapidly over the past few years. As seen in this chapter, IDS tools have become essential security components – as valuable to many organizations as a firewall. However, as in any environment, things change. Networks and crackers are evolving fast, demanding that security tools keep up. IDSs face several daunting, but exciting challenges in the future and are sure to remain one of our best weapons in the arena of network security. Although the original vision of IDSs has been a formal discipline for almost 50 years, the IDS research field is still young, with most research dating to the 1980s and 1990s. The large-scale commercial use of IDSs did not start until the mid-1990s. However, the intrusion detection and vulnerability assessment (VA) market has grown having a significant commercial presence. Technology market analysts predict continued growth in the demand for IDS and other network security products and services for the foreseeable future.

According to subject matter experts (SMEs) and market analysts, even while the IDS research field is maturing, commercial IDSs are still in their formative years. Some commercial IDSs have received a negative publicity owing to their large number of false alarms, awkward control and reporting interfaces, overwhelming numbers of attack reports, lack of scalability and lack of integration with enterprise network management systems (see the discussion in the section 'Implementing IDS'). However, the strong commercial demand for IDSs will increase the likelihood that these problems will be successfully addressed in the near future. It is anticipated that the improvement over time in quality of performance of IDS products will likely parallel that of antivirus software. Early antivirus software created false alarms on many normal user actions and did not detect all known viruses. However, over the past decade, antivirus software has progressed to its current state, in which it is transparent to users, yet so effective that few doubt its effectiveness.

There has been a rapid rise of the switched environment. Though, once upon a time, they were an expensive alternative to the ubiquitous hub, switches now dominate both the commercial and the home networking markets (switches, hubs and other network components were discussed in Chapter 12). Switches offer a better performance in most situations and protect against packet sniffers by sending traffic only to the intended ports. Hubs, on the other hand, simply copy all information to every available port. Since NIDSs are packet sniffers at their core, the switched network creates a definite problem. There are countermeasures, but they bring additional complexity and costs. High-end switches incorporate management and configuration

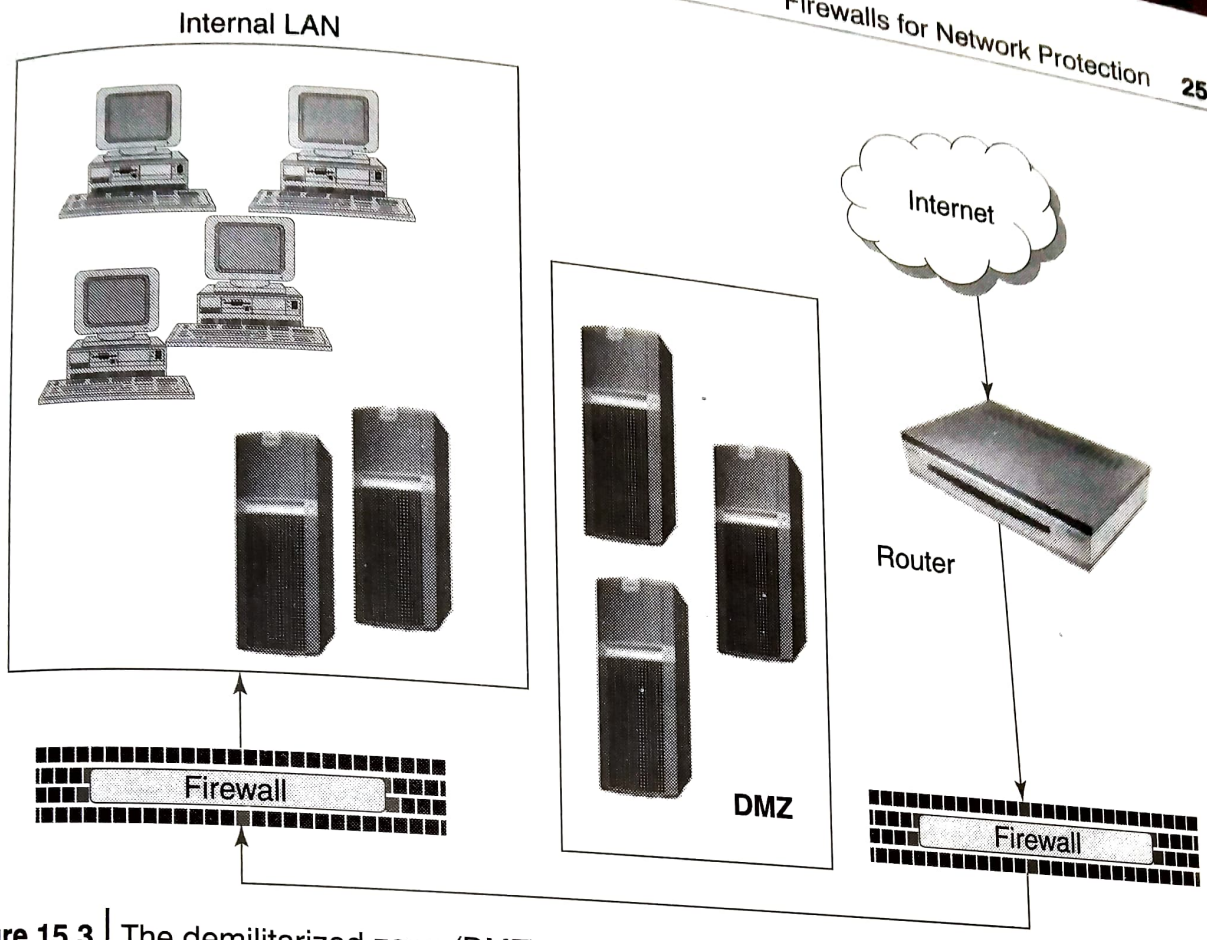


Figure 15.3 | The demilitarized zone (DMZ).

15.4 Why Firewalls are Needed – Protection Provided by Firewalls

Unscrupulous people, network intruders, hackers and perpetrators use many ways to access or abuse unprotected computers, such as remote log-ins, backdoor entries, session hijacking, exploiting the bugs in OSs, DoS attacks (DoS attacks are discussed in Chapter 2, session hijacking is explained in Chapter 11 and discussion on security patches to overcome operating system bugs is provided in Chapter 21), bombs implanted through electronic mails (e-mails), illicit design of macros, viruses and spam mails, routing exploitation, etc. All this is discussed in brief as follows:

1. **Remote log-in:** This is when someone is able to connect to a computer and control it in some form. This can range from being able to view or access your files to actually running programs on the connected computer.
2. **Application backdoors:** Some programs have special features that allow for remote access. Others contain bugs that provide a backdoor, or hidden access, that provides some level of control of the program. Readers will recall the discussion on Trojan horses and trap doors in some of the previous chapters.
3. **Simple mail transport protocol (SMTP) session hijacking:** SMTP is the most common method of sending an e-mail over the Internet. By gaining access to a list of e-mail addresses, a malicious attacker can send unsolicited junk e-mail (spam) to thousands of users. This is done quite often by redirecting the e-mail through the SMTP server of an unsuspecting host, making the actual sender of the spam difficult to trace.

4. **OS bugs:** Like applications, some OSs have backdoors. Others provide remote access with insufficient security controls or have bugs that an experienced hacker can take advantage of.
5. **DoS:** Readers will recall that this has been much discussed in previous chapters. We also hear this phrase used in news reports on the attacks on major websites. This type of attack is nearly impossible to counter. What happens is that the hacker sends a request to the server to connect to it. When the server responds with an acknowledgement and tries to establish a session, it cannot find the system that made the request. By inundating a server with these unanswerable session requests, a hacker causes the server to slow to a crawl or eventually crash.
6. **E-mail bombs:** An e-mail bomb is usually a personal attack. Someone sends you the same e-mail hundreds or thousands of times until your e-mail system cannot accept any more messages.
7. **Viruses:** We discussed about viruses in Chapter 2 (refer Box 2.1 and Section 2.5). Probably the most well-known threat is computer viruses. A virus is a small program that can copy itself to other computers. This way it can spread quickly from one system to the next. Viruses range from harmless messages to erasing all of your data.
8. **Macros:** Many applications allow you to create a script of commands that the application can run. This is done to simplify complicated procedures. The script written for this is known as a 'macro'. Hackers often take advantage of this to create their own macros that, depending on the application, can destroy your data or crash your computer.
9. **Spam:** Typically harmless but always annoying, a 'spam' is the electronic equivalent of junk mail. Spam can be dangerous though. Quite often, it contains links to websites. Be careful of clicking on these because you may accidentally accept a cookie that provides a backdoor to your computer.
10. **Source routing:** In most cases, the path taken by a packet (we discussed this term in Chapter 12) to travel over the Internet (or any other network) is determined by the routers along that path. But the source providing the packet can arbitrarily specify the route that the packet should travel. Hackers sometimes take advantage of this to make information appear to come from a trusted source or even from inside the network! Most firewall products disable source routing by default.
11. **Redirect bombs:** Hackers can use Internet control message protocol (ICMP) to change (redirect) the path information takes by sending it to a different router. This is one of the ways that a DoS attack is set up.

Refer Box 15.1 for useful and interesting information about cookies.

Box 15.1 Cookies – Monster or Your Diligent Assistant on the Internet!

Cookies started receiving tremendous media attention back in 2000 because of Internet privacy concerns, and the debate still rages. Are cookies bad? There are all kinds of discussion on this and the answer is not a plain yes or a plain no. The good part of cookies is that they provide capabilities that make the web much easier to navigate. When designers design an Internet site, they like to use them to provide a better user experience and make it much easier to gather accurate information about the site's visitors.

The following is a common misconception about cookies: Many people believe that 'Cookies are programs that websites put on your hard disk. They sit on your computer gathering information about you and everything you do on the Internet, and whenever the website wants to it can download all of the information the cookie has collected'.

Experts say that this is not right. Cookies are *not* programs, and they cannot run as programs do. Therefore, they cannot gather any information on their own. Nor can they collect any personal information (PI) about you from your computer. A valid definition of a cookie is the following – 'a

Box 15.1 *Continued...*

cookie is a piece of text that a web server can store on a user's hard disk. Cookies allow a website to store information on a user's machine and later retrieve it'.

For example, your common experience with cookies is when you happen to fill in a web-based form. When you do that once, next time, you would observe that the form somehow happens to 'remember' your name the moment you type in your name. The same may happen with your telephone number that you had once fed to the form. This is nothing but cookies at work! That website might have generated a unique identity (ID) number for each visitor and store the ID number on each user's machine using a cookie file.

15.11 Using Firewalls Effectively

For an effective usage of firewalls, organizations need to follow some best practices: choosing the right type of firewall based on a risk and vulnerability assessment, establishing meaningful firewall policies, physical security as well as maintenance of firewalls (in Section 15.10 we mentioned about regular upgrades of firewalls). To ensure that firewalls perform their intended function, it is important to choose an appropriate

firewall and to implement it correctly. Given this, it is necessary to ask some critical questions while making decisions about choosing a firewall:

1. Does the firewall support encryption?
2. Does the firewall support authentication?
3. Does the firewall allow you to manage it centrally and through a standard interface?
4. How easily can you establish rules for access to and from the firewall?
5. Does the firewall support filtering at the highest layers of the OSI model?
6. Does the firewall provide logging and auditing capabilities, or alert you to intrusions?
7. Does the firewall protect the ID of your internal LAN's addresses from the outside world?

We discussed various types of firewall configurations in this chapter; each type of firewall offers its own set of advantages and disadvantages. Some questions are presented above about making a right selection of firewalls.

Establishing a firewall policy is also a critical step in securing the firewall system. This was discussed in the previous section. Maintenance of firewalls is another critical issue as mentioned before. Often when organizations employ firewall, they feel a false sense of security. Firewall products have improved considerably over the past years and they will continue to improve. In today's 'extended enterprise' concept (this concept was discussed in Chapter 1), opportunities for third-party connections are ample; such opportunities also provide avenues for bypassing gate-based security mechanisms altogether. Therefore, an Internet security strategy that includes firewalls in addition to host-based security mechanism is invariably the most appropriate direction for achieving suitable levels of Internet security.