SECURITY & DISTRIBUTED SYSTEMS

JONATHAN D. MOFFETT
DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF YORK
ENGLAND

I. INTRODUCTION

This article is an introduction to the subject of security in distributed systems. It is assumed that the reader is generally familiar with the concepts of computer and communications security, by which we mean protection against risks which can compromise data integrity, allow unauthorised disclosure of information or lead to denial of service in systems. However, this article aims to be clear to the non-specialist, and additional background information can be found in the Further Reading.

The term **distributed systems** has not yet acquired a completely firm meaning. In this article we define is to be one in which several autonomous processors and data stores, supporting processes and/or databases interact in order to cooperate to achieve an overall goal. The processes coordinate their activities and exchange information by means of a communications network. See Further Reading for additional information.

As an illustration, consider the distributed system for a hypothetical business shown in figure 1. Here the financial department has its own network of clerical workstations with data storage and report printing facilities. This is linked to the production units, some of which may be on remote sites, which have specialist production control processing to perform. There are stores systems to maintain inventories, sales departments which generate invoices and so on. All these separate parts of the organisation are in communication with the corporate centre. Their information contributes to corporate planning and their actions respond to corporate policy. It is an illustration of a system which performs a vital part in the operation and management of a modern business. It therefore needs to be secured.

A. SECURITY RISKS OF DISTRIBUTED SYSTEMS

There are special factors of risk in distributed systems. Existing distributed systems offer significant opportunities for the introduction of insecure or malicious software. They also permit hacking and browsing. Even those distributed systems which are intended to support a low or medium risk area of business still have to be careful today not to leave themselves wholly unprotected. Vigilant management is required against attacks leading to denial of service. Even where these attacks do not compromise data integrity, they may be both inconvenient and expensive. The experience of people affected by the “Internet Worm” (Spafford, 1988) illustrates this. A deliberately created program propagated itself across several networks, especially in the USA. Although it did not itself cause any damage, it reproduced itself continuously until it had absorbed all the resources of computers it had invaded and brought them to a halt. The cost of recovery was estimated at millions of dollars. Other risks of this kind are described in (Denning, 1990).

Similar effects can be caused accidentally. In particular, the incorrect handling of error reports in electronic mail systems can cause "mail storms" which swamp the network. This can be caused if a message containing errors is broadcast to multiple sites If each receiver site reports the error back to the originator and separately triggers off a retry of the entire broadcast, the number of messages grows exponentially until the network halts.

Another risk is that unprotected systems may be used as an entry point into other inadequately protected but sensitive systems. The case of the German hackers who obtained access to many sensitive systems illustrated this risk (Stoll, 1989). They used unprotected systems as bases from which to probe systematically for security weaknesses in other more sensitive systems, with a surprisingly high degree of success. This resulted not only in the exposure of confidential information, but also in extra costs for several sites who only discovered that they had been penetrated when their bills for communication were unexpectedly high.

There is a direct risk of exposure of confidential information in the uncontrolled, unprotected use of public networks between nodes of the system for information transfer. There are many opportunities for network staff to gain access to transmitted information but, in addition, any satellite or point-to point radio link may be intercepted with the appropriate equipment. If a secure network is required, encryption and access controls are essential.

Distribution not only introduces additional risks to computer systems but also adds complications to dealing with the risks. For example:

* communication may introduce significant time-lag into the system in respect of security-related information; this may make it difficult for the security management system to correlate information which, taken together, would indicate a security breach;
* splitting the system into different geographical, political, technical or administrative domains complicates the setting and management of a coherent security policy; it also adds to the difficulty of tracing security breaches which are initiated from a different domain.

B. SECURITY BENEFITS OF DISTRIBUTED SYSTEMS

However, in addition to the down-side of introducing risks, there are compensating factors in distributed systems which can be used to enhance system security.

Unauthorised access to corporate data can provide the intruder with valuable strategic information. The advantage of distribution in this case is that it allows sensitive data to be distributed throughout the system. Thus only by knowing the way in which it is distributed and accessing it at all locations can an intruder obtain complete information.

The damage which can result when the processing capability of a system is disrupted can be very high, particularly when a high premium is placed on the ability to process information. Distribution can provide alternative locations from which to acquire processing resources. Accidental failures typically occur at one site at a time, and deliberate attempts to disrupt a service would require interference with a number of sites simultaneously.

There may be a variety of security requirements within a distributed system. One advantage of distribution is that it does not constrain all components of a system to accept the same security regime. If the environment is partitioned into separate security domains, each domain can reflect a different aspect of the organisation's policy concerning security. Overall control is obtained either by negotiated security interaction policies between the managers of domains or by a hierarchical structuring of domains with one manager taking responsibility for coordinating the interactions of all.

II. SECURITY FRAMEWORK

The objectives of security within distributed systems can be defined at a number of different levels, from a high-level objective such as “to safeguard the organisation's assets” to a low-level one such as “ensure that no dictionary words are used as passwords”, with a hierarchy of objectives in between. Each level helps to achieve the objectives of a higher level. These objectives may be achieved by mechanisms at several different architectural levels within a distributed system. An example of this, mentioned below in section II.D.1, is the protection of data in transmission. This can be achieved by link protection, by end-to-end protection, or at an intermediate level. The combination of security objectives and the architectural levels at which they may be supported together form a framework in which to describe security.

The International Standards Organisation (ISO) Open Systems Interconnection (OSI) Security Architecture (ISO, 7498-2) defines a set of security services based on generally agreed objectives and sets out the options for the architectural levels at which these may be provided. The objectives are described in more detail in the OSI Security Frameworks Overview (ISO, 10181-1). Section VI.C gives a summary of the OSI approach to security standards.

A. SECURITY OBJECTIVES

It is helpful to distinguish between the primary and secondary objectives of security. The primary objectives correspond to threats such as disclosure, corruption, loss, denial of service, impersonation, repudiation. The secondary objectives lead to the specification of services to support the primary ones.

There are three primary security objectives which apply to both stored data and messages in transit. They are:

* **Confidentiality** - maintaining confidentiality of information held within systems or communicated between them. This typically means the prevention of unauthorised access to stored data files and the prevention of eavesdropping on messages in transmission. However, in high-security applications there may also be a requirement for protection against revealing information which may be inferred solely from the fact that data is being transmitted and not from its contents. This information can be derived from **traffic analysis**, analysis of the source, destination and volume of communications. A classical case of traffic analysis is a military one in which preparation for troop movements could be revealed by the increased volume of communications between units.

* **Integrity** - maintaining the integrity of data held within systems or communicated between systems. This prevents loss or modification of the information due, for example, to unauthorised access, component failures or communication errors. In data communications, it may also be important to prevent the repetition of a message. For example, a message in an Electronic Funds Transfer system authorising the transfer of funds from one account to another must not be sent and acted on twice. Protection from this risk is known as prevention of replay. Integrity can be achieved in two different ways: either preventing the occurrence of failures at all, or detecting the occurrence and recovering from it. Prevention may be achieved by a number of means; by physical protection, by access control against unauthorised actions and by procedural measures to prevent mistakes. Detection and recovery require timely detection, combined with backup facilities which make it possible to start again from a situation of known integrity.
* **Availability** - maintaining the availability of information held within systems or communicated between systems, ensuring that the services which provide access to data are available and that data is not lost. Threats to availability may exist at a number of levels. A data file is unavailable to its user if the computer which provides the service is physically destroyed by fire, or if the file has been irretrievably deleted, or if the communication between user and computer has failed. As with integrity, two different modes of protection are available: prevention; and detection and recovery using backup facilities.

Two other primary security objectives apply specifically to communication between users and/or programs:

* **Authentication** - authenticating the identity of communicating partners and authenticating the origin and integrity of data which is communicated between them. It is important for several purposes. Authenticating the identity of the originator of a message gives confidence, in electronic mail systems, that messages are genuine. It also provides a basis for audit and accounting. It is a requirement for access control systems based on the identity of users of the system. Authentication of message contents enables the detection of integrity failures in messages.
* **Non-repudiation** - this is the prevention of a user wrongly denying having sent or having received a message. The first of these is known as proof of origin and the second as proof of delivery. Non-repudiation is important in any situation in which the interests of the sending and receiving parties may be in conflict. For example, in a stock transfer system it would be in the financial interest of the sender to repudiate a selling order if the value of the stock subsequently rises, and in the interest of the receiver to repudiate it if it falls. It is a key issue for contractual systems based on EDI (Electronic Data Interchange), for example, purchase and supply systems.

The secondary security objectives identified by the Security Architecture are as follows:

* **Access Control** - providing access control to services or their components to ensure that users can only access services, and perform data accesses, for which they are authorised. Access control is one means which is used to achieve Confidentiality, Integrity and Availability. It can be provided by physical and/or logical mechanisms. Unauthorised access to a personal computer may be prevented by a key lock disabling

the keyboard. Access to a shared system may be controlled by a logical access control system using access rules based on the authenticated identity of users.

* **Audit Trail** - providing an audit trail of activities in the system to enable user accountability. An audit trail provides evidence of who did what, and when. The important special case of audit of access control systems is discussed in section V.B.
* **Security Alarm** - the detection of occurrences indicating an actual or potential security failure should raise an alarm and cause the system to operate in a fail-safe mode. Some security failures are not detected at the time, and cannot be reported on, like the failure of the access control system to detect an unauthorised access because of its own weakness. Other activities may be indicative of possible security failures, and need investigation; for example, a changed pattern of access by a user. The objective in this situation is to minimise, simultaneously, the risk of loss if there really is a security failure and the inconvenience to the user if there is a false alarm.

The security objectives outlined above are interdependent, and should not be taken in isolation. Authentication is the basis for achieving many of the other objectives. Authenticated user identities are needed for identity-based Access Control, Non-Repudiation and Audit Trail, but password-based Authentication requires both Access Control to protect the password file and encryption-based Confidentiality for further protection if the Access Control fails. Access Control, besides requiring and supporting Authentication, is a basis for Confidentiality, Integrity and Availability. Audit Trails and Security Alarms both depend upon and support the other objectives.

B. ARCHITECTURAL LEVELS OF SECURITY SERVICES

The ISO Security Architecture identifies the possible communication protocol layers of the Open Systems Interconnection Basic Reference Model at which each security service could be provided. A security service, such as confidentiality, can be applied to communication at different layers in the model but it is not sensible to apply the service at all of the layers. For instance, a user who is obtaining end-to-end confidentiality through encryption (see III.E) at the Presentation Layer, has no need of Data-link encryption as well (see figure 2). Further standards work will identify appropriate profiles of security services for particular applications.

III. SECURITY MECHANISMS

A number of different mechanisms are used to achieve security objectives. They include:

* physical and electronic security of components of the system;
* authentication mechanisms;
* access control mechanisms;
* communication security mechanisms.

They are described briefly here. Interested reader are referred to Further Reading for more detail.

A. PHYSICAL SECURITY MECHANISMS

Physical security mechanisms are used for protection of equipment and for access control outside the scope of logical access control or encryption. They are necessary for protection against risks such as fire, tempest, terrorist attacks and accidental or malicious damage by users and technicians. Physical security requires a variety of mechanisms:

* Preventive Security - strong construction, locks on doors, fire resistance and waterproofing;
* Detection and Deterrence - movement detectors and door switches linked to alarms, security lighting and closed circuit television;
* Recovery - the provision of a backup site, with alternative computing and communication arrangements.

A basic level of physical security is always necessary even in the presence of logical access control and encryption. In some situations physical protection may be simpler and more secure than a logical solution; for example, by controlling physical access to terminals and personal computers and their data and by storing sensitive data on demountable media.

Figure 3 illustrates a situation in which encryption needs to be supplemented by physical line protection if complete end-to-end protection is to be achieved. It is necessary because the encryption unit is not an integral part of a secure terminal.

B. ELECTRONIC SECURITY MECHANISMS

Electronic security mechanisms may be needed for protection against interference from static electricity and RF (Radio Frequency) interference, both of which can cause computer and communication equipment to malfunction. They are also required for Radiation Security to avoid the passive eavesdropping of electromagnetic radiation from visual display units, printers and processors. The modulated signals can be detected by nearby radio receivers and analysed to reveal the data being displayed, printed or processed. Preventive devices are commercially available, and there are also military standards of protection (so-called "Tempest" proofing).

C. AUTHENTICATION

1. Personal Authentication

The aim of personal authentication in computer systems is to verify the claimed identity of a human user. There are a number of different mechanisms for it, all based on one or more of the following principles:

* a personal characteristic of the user, (fingerprint, hand geometry, signature, etc.) which is unique to the individual;
* a possession of the user, such as a magnetically or electronically coded card, which is unique to that person;
* information known only to the user, for example, a secret password or encryption key.

Secret personal passwords are the simplest and cheapest method to implement, and they provide an adequate level of protection for medium and low security applications. They need a number of supportive measures if they are not to be undermined. The measures include: regular change by the user, one-way encrypted storage, minimum length and controlled format (such as no dictionary words), limited number of permitted attempts, and logging and investigation of all failures. They can be reinforced by restricting users to logging in at specific physically protected terminals; for example, Payroll clerks may only log on in that capacity using one of the terminals sited within the Payroll Office.

The use of passwords across open communication channels in distributed systems is a particular problem because the password can be discovered by eavesdropping on the channel and then used to impersonate the user. One solution to this is the use of one-time passwords generated by smart cards (see below).

Magnetically coded cards have some advantages over passwords - they cannot be copied so easily and are less easy to forget. However, they also suffer from the potential exposure of their contents in open communication channels.

Smart cards offer increased security because they can be programmed to provide variable information. There are several modes in which they can be used for personal authentication. Two of these are:

* One-time password generators which generate a different password each time they are used. One commercial product changes the password every minute. In all cases the computing service must synchronise with the password generator.
* Challenge-response devices. The host sends a challenge number and the smart card has to calculate the correct response, including input from the user.

Smart cards are becoming cheaper and easier to use and they promise to provide a satisfactory way of overcoming the problems of personal authentication in distributed systems. However, any authentication system must cope with the problem of protecting the secure information upon which it is based. This is an aspect of security management (see section V.A).

2. Message Authentication

The aim of message authentication in computer and communication systems is to verify that the message comes from its claimed originator and that it has not been altered in transmission. It is particularly needed for EFT (Electronic Funds Transfer). The protection mechanism is generation of a Message Authentication Code (MAC), attached to the message, which can be recalculated by the receiver and will reveal any alteration in transit. See figure 4. One standard method is described in (ANSI, X9.9). Message authentication mechanisms can also be used to achieve non-repudiation of messages.

D. LOGICAL ACCESS CONTROL

Logical access control has to be used when physical access control is impossible, as is the case in multi-user systems. A model for logical access control is provided by a Reference Monitor, which intercepts all access attempts and allows them only if the access is

authorised. Otherwise the access is blocked, an error message is returned to the user and appropriate logging and alarm actions are taken.

There are two main forms of logical access control: mandatory access control, based upon fixed rules; and discretionary access control which permits users to share and control access (see section VI.A). The recommended discretionary access control approach is Identification/Authorisation. The system ensures that the identities of users are authenticated when they log on, and the Reference Monitor makes a decision based on access rules which relate the user, the entity being accessed, and the operation the user is attempting to carry out.

There are two main implementations of access rules:

* An Access Control List (ACL) is attached to the target entities, defining the users who are authorised to access them and the operations which they can perform;
* Users obtain authenticated Capabilities which act as tickets authorising them to access defined resources.

Many personal computing systems only provide access control based upon file passwords. These provide a minimal, easy-to-use level of protection which is adequate for low-security systems.

E. COMMUNICATION SECURITY MECHANISMS

There are two main mechanisms in ensuring communication security, in addition to physical protection of the lines and equipment: encryption; and traffic padding.

Encryption is one of the most important techniques in computer and communications security. Cryptography means literally "secret writing". Encryption transforms (enciphers) plain text into ciphertext, which is impossible to read, and decryption transforms it back again into readable plain text (deciphers it). Cryptography has been practised for thousands of years, but the advent of computer-based encipherment algorithms has changed it from a difficult and unreliable to a simple and powerful one. Algorithms such as the Data Encryption Standard, described below, are easily available, simple to use, and provide a high degree of protection against threats to the confidentiality and integrity of communications.

Traffic padding will only be dealt with briefly here. Its purpose of is to conceal the existence of messages on a communication line, by inserting dummy messages on the line to ensure that there is a uniform level of traffic at all times. It is mainly of interest in a military level of security.

Encryption can be used for several purposes: the prevention of eavesdropping; the detection of message alteration; and, in conjunction with the use of unique message identities, the detection of message deletion and replay.

1. Link or End-to-end Encryption

Encryption may be used on individual links or on an end-to-end basis. These options were illustrated in figure 2. Link encryption only covers the communication links, and the information is "in clear" at each communication processor. By contrast, end-to-end encryption is carried out directly between the initiating and target systems. An intermediate

level is network encryption, where encryption spans an entire network, but not the gateways between networks. In all cases where encryption is carried out by a separate hardware unit, the link between the terminal and the unit is not covered by encryption, and physical protection is required in addition. See figure 3.

2. Encryption Algorithms

There are two main types of encryption:

* Secret key encryption, which uses a single secret key shared between sender and receiver. See figure 5.
* Public key encryption, which uses a related pair of keys. One key is publicly available and may be used to encrypt messages, while the other key is secret, known only to the receiver, and may be used to decrypt messages. See figure 6.

Secret key encryption is available in a number of proprietary algorithms, and in the Data Encryption Standard (DES) which is an American, but not an international, standard. The DES is described in a number of text books (Davies & Price, 1989). The DES algorithm is available in software and also in hardware as a semi-conductor chip, providing higher performance. The chip is subject to export restrictions from the USA, and the hardware is therefore not suitable for multi-national applications. Any secret key algorithm suffers from key management problems, because of the need to transport secret encryption keys securely. There is a standard method of key management, described in (ANSI, X9.17), which covers: key generation and distribution; protection of the key management facility; and protocols for the cryptographic service.

The most widely accepted method of public-key encryption is the Rivest, Shamir, Adleman (RSA) algorithm (Rivest, Shamir, & Adleman, 1977). Its performance is much poorer than the DES algorithm, but key management is easier because there is no need for secrecy of the key used by the sender. Some mixed-mode systems use RSA for key distribution and DES for message security, thus gaining many of the advantages of each method.

IV. SECURITY POLICIES

Policies are the plans of an organisation to meet its objectives. Within the context of security, a security policy defines the overall objectives of an organisation with regard to security risks, and the plans for dealing with the risks in accordance with these objectives. Policies are usually hierarchical; the plans of a high-level policy are the objectives which a lower level policy must address.

All organisations should have a high-level security policy, defining the overall security goals of the organisation and setting out a framework of plans to meet the goals. These high-level objectives vary substantially from organisation to organisation. Military organisations place a high value on secrecy in contrast to academic institutions which value openness of information. Financial institutions are concerned above all with maintaining the integrity of data and messages which represent money. The default for simple social organisations is to have no security policy at all.

Security policies are not always precisely formulated or written down, but an effective computer security policy requires that the following questions are answered:

* What are the assets to be protected, and what is their value?
* What are the threats to these assets?
* Which threats should be eliminated and by what means?

The security policy for a distributed system should reflect the senior managers' expectations of the organisation's security objectives. Often, just as an organisation's security objectives may not be precisely formulated, those in a distributed processing environment are unstated and they must be extracted from other documents, or identified and agreed by persuasion and discussion with the staff of the organisation in question.

A high-level security policy can make general statement about the goals of the organisation, but in order to be effective it requires a risk analysis to be carried out so as to understand the vulnerability of the organisation and the consequences of security breaches. Risk management, discussed below in section V.C, is required because of the possible trade-offs between the anticipated cost of threats and the actual costs of security measures. The security measures taken to counter a threat should be commensurate with the threat itself. The results of a risk analysis may help to redefine or focus high-level policies, as well as define the lower-level policies for managing the system in a secure manner.

The choice of security services will have to reconcile a number of conflicting objectives which include the following:

* Security policy is often defined centrally but applied locally, or in each application, and at many intervening points in the communication between each user. There are therefore practical difficulties in ensuring that a security policy is met.
* Design of existing standard communication products and many operating systems has avoided or neglected considerations of security. Security requirements therefore have to be negotiated separately with system suppliers, requiring much greater effort in procurement than if they were defined as part of a standard or an intrinsic part of the operating system.

A. SECURITY INTERACTION POLICIES

It is possible to create a distributed system in which all aspects of security are centrally managed to a common standard. Because a distributed system is more likely to evolve by federation of a number of existing different (and heterogeneous) systems, they may have previously operated to a variety of security policies. It is always possible that these may be incompatible, either because the policies of the systems differ in the level of security they provide or tolerate, or because there is technical incompatibility, for example because different encryption algorithms have been selected.

ISO have recognised this problem and have introduced the concept of a Security Interaction Policy as part of the Security Frameworks. This is a policy which is acceptable to all parties in an interaction. It has to be negotiated between them before they can communicate. The issues which have to be resolved between them are both the level of security and the technical compatibility of their security mechanisms. So far as the level of security is

concerned, this is not limited to the security parameters relating to their communication. The security policy of one organisation may insist that compatible standards of security are in force at the other organisation's computing facilities before an interaction is permitted.

An inter-organisational Security Interaction Policy, agreed and committed to by all parties, may be difficult to negotiate because of the need for more widespread compatibility than simply of communication security standards. For example, there may be incompatibilities in the levels of security of their operating systems. If a common security policy cannot be agreed it may be decided to decline to communicate, either because the risk is unacceptable or to avoid the imposition of unacceptable or uncongenial security practices. For example, most organisations which are running their installations to government military security standards do not allow electronic mail to run on any of their networked computers, because of the known security exposures associated with it. They have to use a special free-standing electronic mail which is disconnected or buffered from the rest of their systems.

A Security Interaction Policy in a distributed system should lead to the generation of an agreed schedule of required security services and their supporting mechanisms.

B. THE PRACTICAL APPLICATION OF IT SECURITY POLICIES

To illustrate the practical application of information technology (IT) security polices in a distributed processing environment, some of the policies which could apply to a typical company are described. They are divided into the following areas:

* security administration policies;
* security levels;
* communication security;
* system access control;
* data access control;
* disaster planning;
* system auditability;
* legal and regulatory policies relating to security.

Note first of all the extent and limits of these policies. They include the areas needed to ensure the confidentiality, integrity and availability of information, with two notable exceptions. First, they do not cover the backup and recovery procedures which are part of normal day-today IT operations. It is assumed that, in addition to a IT Security Policy, the organisation has an IT Computer Operation Policy which covers day-to-day computer operating procedures, including recovery from incidents such as system failures and disc crashes; and also a IT Communication Management Policy covering procedures such as alternate routing in the event of line failures. The exclusion of these areas from security policies is to some extent arbitrary, but is common to many organisations, which prefer to regard them as aspects of system and communication operations. It is of course essential to ensure that these subjects are covered in one policy document or another.

Second, these security policies also exclude system change control. This too may be regarded as arbitrary, but the justification is again that it should be covered in other policies. Those aspects of change control relating to reconfiguration of the system by changing

hardware components such as processing systems and networks should be covered by IT Computer Operations and Communication Management Policies. Software change control should be covered by polices for IT Computer Operations and Development.

In a typical commercial organisation, security will make little or no direct contribution to each department's achievement of its immediate goals (until something goes wrong), while costing a substantial amount of effort and money. Therefore if the security policies are to be effective they need to be endorsed at the highest possible level of management, usually Board level, and included in the targets which are set for each department. Only then will business managers regard them as an integral part of their goals.

Further, they must be communicated effectively. Many organisations have informal security policies which can be deduced from other policy documents and from management decisions which may be buried in internal memoranda and difficult to find. Effective security policies need to be separately and clearly documented, preferably as a Security Policy document or as a section in an IT Policy document. It is then possible for IT and user personnel to find out easily what the policies are; there is no possibility of effective implementation of a policy which nobody knows about.

The first recommendation for a company is therefore that the director responsible for IT obtains Board approval for the creation and enforcement of IT security policies, as set out in an IT Security Policy document.

1. Security administration policies

The foundation of the security policies of the organisation will be an effective organisational structure. While information processing is centralised, it is sufficient to put someone in charge of enforcement of IT security throughout the organisation. However, as soon as it becomes distributed it is necessary to have a two-tier organisational structure: a central Security Coordinator; and Security Administrators covering every department of the organisation. To respect the autonomy of distributed systems, direct responsibility for security in each system should be held by its Security Administrator, but in order to ensure that the level of security is consistent throughout the organisation the Security Coordinator should have the goal of ensuring that each system is working to compatible standards and procedures. Typically the Security Coordinator has two tasks: ensuring that each Security Administrator is aware of the standards and procedures; and helping departmental business managers to ensure that they are adhered to.

The main concern of a company, apart from setting up its IT security organisation, must be to ensure that the net is spread widely enough; every independent system which has eluded the control of communication or systems management is a potential security risk. The personal computers which have “gone independent” may or may not be operating to company standards of security. Each must be brought in under the umbrella of the appropriate Security Administrator.

2. Security levels

Management policies should be made about groups of objects, rather than individuals. Most organisations will need to apply this concept to security measures also. If just a few levels of

security can be identified, and a package of security measures defined for each level, then detailed decisions about individual objects can be avoided.

Commercial organisations are likely to define two types of security level: for data and for users. Most aim to give the same level of protection to all their data, so that there will only be one security level defined for data. This is much easier to administer than multiple levels, and accords with the usual requirement, that sharing and mobility of data must be enabled but controlled. However, there may be a requirement to treat some data especially securely, either because its confidentiality is critical to the success of the business or, more commonly, because of the terms of a government contract.

On the other hand, it is quite common to wish to make distinctions between categories of users, especially when outsiders are given limited access to the system for special purposes, or when insecure dial-in access is permitted. So there may be a policy to treat three categories of user differently: normal users, who have the least restricted access; the same users when they dial in, who are to be allowed access to specifically pre-defined data; and outsiders, with a similar restriction. Note that this concept of security levels has similarities to the military-type security levels of mandatory security (Department of Defense (USA), 1985), but is much less formally defined.

3. Physical Security

Physical security is the basis of all other system security. There should be a policy which requires an appropriate level of physical protection for all physical assets within its control.

4. Communications security

Communications security policies typically define the required levels of confidentiality, integrity and availability. They are divided into two categories: the security of the corporate networks; and the criteria for applications to provide a higher level of security on an end-to-end basis. Based on today's technology, the policy is likely to require assured integrity and a defined percentage of availability from the network, but not to insist on assured network confidentiality. Within a few years, with increasing availability of cheap encryption products, the policy is likely to be upgraded to insist on network confidentiality also.

Assuming a relatively weak network security policy, an additional policy is required to ensure that applications requiring a higher level of security are provided with it by means of application-level security measures. Note that user authentication policies are dealt with below.

5. System access control

Perhaps the most important security measures to be taken in an organisation with open distributed systems are to do with controlling the access of users to the system. There needs therefore to be a policy about system access control, with two main parts. The first part states the requirement for unique user identifiers and the need for users to respect them. The second defines the strength of authentication which must be applied to users who attempt to log on to the system. Typically there will be two levels of authentication:

* Normal users logging in from terminals and workstations within ABC's premises. The policy will state the standards for passwords, like the minimum length and required format, and the frequency of change.
* Users, whether company staff or external, who log in from outside ABC's premises. There will be a much higher level of authentication, probably using smart cards or other one-time password generators.

6. Data access control

The policy for data access control contains the following elements:

* Ownership of all elements of the organisation's data should be defined, with the owner being responsible for decisions about the use of the data;
* All the data should be protected, and access should only be permitted when authorised by its owner;
* All systems should have access control systems which protect data to a defined standard. Typically the "Orange Book" C2 level of protection (see section VI.A) is suitable for a commercial organisation.

7. Disaster planning

Most organisations are now critically dependent upon the working of many of their communications and computer systems. If any one of them fails, the business will have difficulty in functioning at all. There needs therefore to be a disaster planning policy which requires each system to be considered for its criticality and defines how any critical system is to be recovered. It should define what is the maximum recovery time, and how all of data, communication and processors are to be backed up to enable rapid recovery in the event of any conceivable disaster. For distributed systems, the problems of back-up are eased by the existence of compatible systems at several sites.

8. System auditability

There are several reasons for having a policy requiring the auditability of all systems, that is the ability to trace any significant action which has taken place in the system. It gives a greater ability to control systems, it is usually a requirement of the auditors and it enables the company to demonstrate that it is complying with legal requirements. It is normally impractical to log all actions all of the time, so there should be a policy dividing systems and applications into three categories: events, such as security administrators' actions and financial transactions, which should be logged all the time; events which can be logged whenever necessary; and events which are so trivial that they never need logging.

9. Legal and regulatory policies relating to security

A policy is needed which makes it clear to all staff that all legal and regulatory requirements are to be complied with, for example data protection legislation.

10. Use of External Organisations

There are many applications in which external organisations are used for processing an organisation's data, for example EFT and EDI. The policy should require that an appropriate level of security should exist on the systems of any external organisation which processes this organisation's data.

11. Concluding Comment on the Security Policies

The security policies outlined above are a minimum set for a typical commercial organisation. Most of them "state the obvious", but it is notable how often obvious security needs are ignored. A comprehensive set of policies, on the lines outlined above, at least ensures that the appropriate questions are asked, and that no part of the organisation can claim to be immune to the need for security.

V. SECURITY AND MANAGEMENT

A. SECURITY MANAGEMENT

Security management is the activity of managing the functions and mechanisms which are used by the various services in a distributed system to implement security policies. The main security management functions include:

* The management of encryption keys (ANSI, X9.17) and other secret information such as passwords. This involves generation of keys as required and distribution of the keys to the relevant components in the system, storing keys and archiving keys. Keys should have a limited lifetime and so should be regenerated at regular intervals.
* Managing the registration of users and the information used to check their identity (public encryption key or password).
* Managing access control information relating to users and servers. This includes access control lists, capabilities, privileges and multilevel security labels.
* Providing security audit trails. These record all exceptional events (attempts at unauthorised access, etc.) and selected normal events such as log-ons and file accesses. Their purpose is to enable investigation of security breaches and audit of a security administrator's actions.

B. AUDIT OF ACCESS CONTROL SYSTEMS

Whatever the organisation for security and wherever the perceived threat, one vital aspect of security management is the ability to maintain an audit trail of significant actions. In this way the system can not only carry out security management, it can also be seen to be carrying out security management. Indeed, maintaining a security audit trail where a system requires access through public communication services can help to meet legal requirements for demonstrating data protection.

It is virtually impossible to guarantee that unauthorised access or viewing of sensitive material will never occur. At best it may only be possible to limit the potential size of an authorised access group and place its members under legal or contractual restraints over use

or disclosure of the information. If the use of due legal process is seriously contemplated as an option, it is essential to maintain an effective audit trail of accesses. That in itself implies the need for a high level of security for the access control mechanism and its associated audit trail. These must be sustained at the expense of performance and in the event of multiple failures, if any real value is to be gained from them. However, securing the record is insufficient; a prompt and effective audit analysis mechanism must also be available.

C. RISK ASSESSMENT AND MANAGEMENT

The term risk is frequently used in relation to information systems. Both the significance of risk, and the relevance of risk management, need to be clarified for distributed systems. Two quite different forms of risk need to be distinguished; *security* risk and *business* risk. Security risk is concerned with future events in which an occurrence leads only to loss. Examples are the risk of loss through fraud, breach of confidentiality or equipment failure. This is the type of risk dealt with here. Business risk can result in either loss or gain, and arises as a result of the normal management decisions of a business. It is not covered in this article

Risk assessment and management is an activity which takes a rational view of the assets of an organisation and the risks they face, and then makes decisions about the protection they are to be given. An essential element of the decisions is that the cost of protection should be commensurate with the expected costs arising from security losses. There are several computer-related security risk management methodologies available (Gilbert, 1989). They all have the following tasks in common:

* Identification and valuation of assets;
* Construction of security risk scenarios;
* Assessment of the probability of the scenarios and the losses which would ensue;
* Identification and costing of possible security measures for each scenario;
* Selection of a portfolio of security measures.

Risk analysis is at its most effective when carried out during the specification and design phases of a distributed system development. In these phases the risk implications of design decisions (identifying where the design shows its least robust characteristics) and the security requirements can be established and their consequences identified. However, this ideal policy is seldom practical for distributed systems since one of their characteristics is that they have often come into being by a process of evolution. For a system already in operation, risk analysis can help to identify where corrective action needs to be taken to remedy any vulnerabilities which had not previously been noted or had not been properly understood. And the analysis, once made, provides a base line from which subsequent analyses can be conducted as the distributed system continues to evolve over time.

VI. SECURITY STANDARDS

There are three main categories of security standards which concern distributed systems. The first are standards related to the security of individual computers, which have been in existence for some time, and are quite mature. The second are standards for the protection of communication transmission and remote authentication. These too are quite mature. The

third, still under development, are those which integrate computer and communication security standards to provide distributed systems security standards.

The standards do not generally prescribe physical or procedural mechanisms, nor do they prescribe risk management or risk assessment procedures or requirements for their use. These are all necessary elements to be considered and resolved for the resources that comprise the whole distributed system. Therefore, although security standards are an important support to distributed system security policies, they have to be viewed in the context of an overall security policy which uses other measures as well.

A. COMPUTER SECURITY STANDARDS

The USA Department of Defense Trusted Computer System Evaluation Criteria (TCSEC - the Orange Book) (Department of Defense (USA), 1985), deals with access control to individual systems. It defines a number of possible levels of trust which may be placed in a system, ranging from the certified high security of the A1 level down to a low level of informally defined security at the C1 level. It is commonly used as a means of indicating the level of security required or supplied in computer systems.

Access control is divided by this standard into two categories: mandatory and discretionary access control. The former enforces policies which are built into the design of the system and cannot be altered except by installing a new version of the system. An example is the policy that in multi-layer security systems data cannot be read by a user with a lower security classification than has been assigned to the data. Discretionary access control mechanisms are defined as those which allow users to specify and control sharing of resources with other users. For example the C2 level discretionary access control policy is defined as requiring mechanisms which ensure that information and resources are protected from unauthorised access and that access permission is only assigned by authorised users.

The Trusted Network Interpretation of the Trusted Computer System Evaluation Criteria (Department of Defense (USA), 1987) (the Red Book) extends the criteria of the Red Book to networks. It is chiefly concerned with the security criteria to be met when accessing remote hosts.

The Red Book is now quite old and it has always been more oriented to military type security than to commercial security. A standards effort which is now under way is the Information Technology Security Evaluation Criteria (ITSEC) (CEC, 1991). This is a joint undertaking by the UK, Dutch, French and German governments. Its aim is to take into account the needs of commercial users and improve on the Red Book by separating concerns about security levels from the way in which the security is evaluated. In the UK, the Department of Trade and Industry and the Communications-Electronics Security Group have established the UK IT Security Evaluation and Certification Scheme (CESG, 1991) which evaluates and certifies products using the criteria of ITSEC. Similar efforts are under way in other countries.

B. COMMUNICATION SECURITY STANDARDS AND ENCRYPTION

1. Transmission Security

Transmission security standards are mainly concerned with encryption methods. One algorithm in particular has been the subject of standards efforts: the DES algorithm for secret key encryption which is an American, but not an international, standard. On the other hand, the RSA algorithm for public key encryption is the subject of USA patents. It has become a de facto standard for public key cryptography but because of its patented status is not currently defined as a national or international standard. These two algorithms have been described briefly in section II.E.2, above.

Encryption depends for its strength upon the security of the hardware which is used, and (BSI, 86/67937) describes standards for the Physical Security of Cryptographic Equipment.

The basic standard for DES is (NBS, 46), supplemented by (ANSI, X3.92). There are supplementary standards describing its Modes of Operation (NBS, 81) and Guidelines for Installation and Use (NBS, 74),. The management of keys in banking applications is described in (ANSI, X9.17), but this standard is rather generally expressed and would apply to other applications also. A detailed discussion of DES standards is in (Davies & Price, 1989).

2. Authentication Standards

A number of standards have been developed for banking applications for peer-to-peer communication and message authentication. They too are quite general in format and could be used for other purposes. They include (ANSI, X9.19) for Message Authentication and (ANSI, X3.118) for personal authentication using a Personal Identification Number (PIN).

C. OSI SECURITY STANDARDS

Network security was not a primary concern when the Open Systems Interconnection (OSI) effort first got under way in the late 1970s. However there is now a series of ISO standards under development which aim to add security to OSI. The standards define the security services which the partners in a communication could agree upon, and the protocols to be used in setting up a secure interaction.

The security services which may be required for the communication facilities have been defined in the ISO 7498-2 Security Architecture (ISO, 7498-2). The protocols for their provision are still largely under development and are not yet available in OSI products.

The security services described in the Security Architecture are described in more detail in a series of Security Frameworks which are currently in production. They will eventually appear as International Standards 10181-1 to 10181-8. The planned framework parts are:

1 Overview - a general introduction (ISO, 10181-1);

2 Authentication;

3 Access Control;

4 Non-repudiation;

5 Integrity;

6 Data Confidentiality;

7 Audit framework;

8 Key management.

Many other standardisation efforts have security implications, and therefore have security-related standards, e.g. in the areas of OSI Directory, OSI Systems Management and Electronic Data Interchange (EDI). Several similar and related efforts are in progress, including profiles to describe the security characteristics of selected applications.

FURTHER READING

INFORMATION SECURITY

Caelli, W., & al, e. (1991). *Information Security Handbook*. MacMillan.

Denning, D. E. (1982). *Cryptography and Data Security*. Addison-Wesley.

DISTRIBUTED SYSTEMS

Sloman, M., & Kramer, J. (1987). *Distributed Systems and Computer Networks*. Prentice-Hall 1987.

Tanenbaum, A. S. (1988). *Computer Networks* (2nd ed.). Prentice-Hall International, Inc.

DISTRIBUTED SYSTEMS MANAGEMENT

Langsford, A., & Moffett, J. D. (1992). *Distributed Systems Management*. Addison-Wesley.

COMPUTER NETWORK SECURITY

Davies, D. W., & Price, W. L. (1989). *Security for Computer Networks*. John Wiley.

Muftic, S. (1989). *Security Mechanisms for Computer Networks*. John Wiley.

REFERENCES

ANSI X3.118 (1984). *Personal Identification Number - PIN Pad*. American National Standards Institution.

ANSI X3.92 (1981). *Data Encryption Algorithm*. American National Standards Institution.

ANSI X9.17 (1985). *Financial Institution Key Management, (Wholesale)*. American National Standards Institution.

ANSI X9.19 (1986). *Financial Institution Retail Message Authentication*. American National Standards Institution.

ANSI X9.9 (1986). *Financial Institution Message Authentication, (Wholesale)*. American National Standards Institution.

BSI 86/67937 (10 Dec 1986). *Physical Security of Cryptographic Equipment*. British Standards Institution.

CEC (1991). *Information Technology Security Evaluation Criteria (ITSEC): Provisional Harmonised Criteria; Version 1.2* (Report No. Office for Official Publications of the European Communities, Luxembourg.

CESG (1991). *UK IT Security Evaluation and Certification Scheme. Description of the Scheme* (Report No. UKSP 01, Issue 1.0). Communications-Electronics Security Group, Room 2/0804, Fiddlers Green Lane, Cheltenham GL52 5AJ, UK.

Davies, D. W., & Price, W. L. (1989). *Security for Computer Networks*. John Wiley.

Denning, P. J. (Ed.). (1990). *Computers Under Attack: Intruders, Worms and Viruses*. Addison-Wesley.

Department of Defense (USA) (1985). *Department of Defense Trusted Computer System Evaluation Criteria* (Report No. DOD 5200.78 - STD).

Department of Defense (USA) (1987). *Trusted Network Interpretation of the Trusted Computer System Evaluation Criteria* (Report No. NCSC-TG-005 version 1). Technical Guidelines Division, National Computer Security Center (USA).

Gilbert, I. E. (1989). *Guide for Selecting Automated Risk Analysis Tools* (NIST Special Publication 500-174).

ISO 10181-1 (1991). *Open Systems Interconnection - Security Frameworks - Part 1: Overview*. International Standards Organisation.

ISO 7498-2 (1988). *Open Systems Interconnection - Basic Reference Model - Part 2: Security Architecture*. International Standards Organisation.

NBS 46 (Jan 1977). *Data Encryption Standard*. National Bureau of Standards, US Department of Commerce.

NBS 74 (April 1981). *Guidelines for Installation and Use of the Data Encryption Standard*. National Bureau of Standards, US Department of Commerce.

NBS 81 (Dec 1980). *Data Encryption Standard Modes of Operation*. National Bureau of Standards, US Department of Commerce.

Rivest, R., Shamir, A., & Adleman, L. (1977). A Method for Obtaining Digital Signatures and, Public-Key Cryptosystems, MIT Laboratory for Computer Science, memo LCS/TM82.

Spafford, E. H. (1988). *The Internet Worm Program: An Analysis* (Report No. CSD-TR-823). Dept of Computer Sciences, Purdue University, West Lafayette, IND 47907-2004.

Stoll, C. (1989). *The cuckoo's egg: tracking a spy through a maze of computer espionage*. New York: Doubleday.