Computer Communications Security

Principles,
Standard Protocols and Techniques

Warwick Ford

Prentice Hall P T R
Englewood Cliffs, New Jersey 07632

Library of Congress Cataloging-in-Publication Data

Ford, Warwick
Computer communications security/Warwick Ford.
p. cm.
Includes bibliographical references and index.
ISBN 0-13-799453-2
1. Computer security. 2. Telecommunication systesm--Security measures. I. Title.
QA76.9.A25F65 1994
005.8--dc20 93-11666
CIP

Editorial/production supervision: *Harriet Tellem*
Cover design: *Lundgren Graphics*
Manufacturing buyer: *Alexis Heydt*

Published by Prentice Hall P T R
Prentice-Hall, Inc.
A Simon & Schuster Company
Englewood Cliffs, New Jersey 07632

The publisher offers discounts on this book when ordered in bulk quantities. For more information contact:

Corporate Sales Department
Prentice Hall P T R
113 Sylvan Avenue
Englewood Cliffs, New Jersey 07632

Phone: 201-592-2863
Fax: 201-592-2249

UNIX is a registered trademark of AT&T.
Kerberos and Project Athena are trademarks of the Massachusetts Institute of Technology (MIT).

Printed in the United States of America
10 9 8 7 6 5 4

ISBN 0-13-799453-2

Prentice-Hall International (UK) Limited, *London*
Prentice-Hall of Australia Pty. Limited, *Sydney*
Prentice-Hall Canada Inc., *Toronto*
Prentice-Hall Hispanoamericana, S.A.,
Prentice-Hall of India Private Limited, *New Delhi*
Prentice-Hall of Japan, Inc., *Tokyo*
Simon & Schuster Asia Pte. Ltd., *Singapore*
Editora Prentice-Hall do Brasil, Ltda., *Rio de Janeiro*

To NOLA,

TERRY,

and LOUISA

CONTENTS

Foreword

Security is becoming an essential requirement of information networks. Strong security technology is required to protect users' sensitive or valuable information, both within the communication network and within information processors connected to the network. Significantly, the network itself now consists of a collection of different types of processors used to route information rapidly and to provide network support services. Reliability and integrity of the complete system are of utmost importance.

The information presented in this book represents the broadest current thinking on network security from the perspectives of end users and system architects. The end user expects a certain level of reliability from a communication network and generally wants to protect information at a level commensurate with the perceived value of, and risk to, the information. Both ubiquitous security provisions to protect the network and user selectable security mechanisms transcending the network are needed.

Users of computer networks are largely unaware of the potential threats to their information, or they choose to ignore such threats. The number of incidents of disruption or loss, on a per-user basis, is still very low. Just as residents in most of the free world are not typically concerned about terrorists or sabotage, users of networks are not typically concerned about information loss or destruction ("It won't happen to me"). However, both could happen easily. The major cause for concern is the continual rise in the number of security incidents and the costs of these incidents throughout the world.

In order for security to become a pervasive network characteristic and for information to be provided a predetermined or selectable level of protection end-to-end, standards for security are of utmost importance. Standards are developed for various reasons. We need both baseline standards to assure a uniform level of information protection and interoperability standards to assure compatibility among network components.

In the U.S. federal government, computer security standards were identified as a high-priority initiative at the National Bureau of Standards in the early 1970s. Two technology milestones, time-sharing systems and remotely accessible computers, caused leading computer science experts to become concerned about potential risks to the security of the systems and the information they processed. Other organizations, such as the U.S. Department of Defense, had recognized the need for protecting classified information in their

systems in the mid-1960s but relied primarily on physical security of the stand-alone systems for the needed protection. Representatives of the two government organizations and several professional computing associations sponsored a workshop in the fall of 1972 that established a foundation for many of the computer security programs and standards that were subsequently developed.

In recent years, a large number of professional and standards-making organizations have contributed to the development of computer security programs and standards. The American Bankers Association, the American National Standards Institute, the International Organization for Standardization (ISO), and the International Telecomunication Union have led the development of voluntary industry standards in security. The U.S. Department of Defense established a series of standards and guidelines (including the *Orange Book* evaluation criteria) for its use. The National Institute of Standards and Technology (NIST; formerly the National Bureau of Standards) undertook development of Federal Information Processing Standards for protecting unclassified information (highlighted in the Computer Security Act of 1987). The NIST standards included the Data Encryption Standard (DES) and other standards supporting the application of DES. While each of these activities has contributed to a comprehensive security program, it has always been difficult for an end user or system architect to understand all aspects of the security problem, especially in a widely distributed network.

Warwick Ford has produced a thorough treatment of network security from the perspectives of end users and system architects. The content is broad enough to be used as a reference and deep enough to be used as a textbook. It takes a serious approach to a serious topic but in a readable style. The truly human nature of the computer security problem is made apparent. The fundamental desire for safety (i.e., security) and the typical unwillingness to meet the disciplinary requirements for achieving it are human characteristics that are often carried over into computer security. This book presents the big picture of network security, then elaborates on the details. It describes the elements needed for the network to exhibit and enforce the necessary discipline on a full time, automated basis.

I first met Warwick when he started participating in the NIST-sponsored Open Systems Interconnection (OSI) Implementors Workshop in the mid-1980s. I was the first chairman of the Workshop's security working group. In greeting this Australian-born Canadian I offered the Australian "G'Day." I remember the smile in response. Over the past seven years I have gotten to know Warwick well, and I responded with a similar smile when he asked me to write a foreword to his excellent book.

Warwick is an established expert in the world of commercial telecommunications security, contributing greatly to international network security standards. In addition to his individual contributions to many standards projects he has produced, in this book, a comprehensive compendium of

information on the field. Both formal and informal students of the field will find this a valuable document. While security books are not intended to produce laughter I believe that, for many readers, this book will produce a smile of appreciation.

Dennis K. Branstad, Fellow
National Institute of Standards
and Technology
Gaithersburg, Maryland, U.S.A.

Preface

Computer communications security involves protecting computer networks against penetration, eavesdropping, data alteration, or disruption by unauthorized persons. Concerns in this area are currently mounting steadily. These concerns interact with another important trend — the move away from a world of incompatible vendor-proprietary networking architectures to the *open-systems* world, in which networks can be constructed by straightforward interconnection of hardware and software components from a range of vendors. The key to the open-systems world is *standardization*.

To make a heterogeneous (i.e., mixed vendor) computer network secure, it is necessary to agree to common specifications for security-related protocols implemented in communicating systems and to specifications for the security techniques that underlie these protocols. These specifications are embodied in standards of various types including international and national standards, government standards, and community (e.g., Internet) standards. Many specifications for standard security protocols and security techniques have matured in the late 1980s and early 1990s. We can now assemble a picture of how such protocols and techniques can be used together to satisfy the full communications security needs of network users.

Objectives and Organization

My objectives in this book are to create an awareness and an understanding of standardized methods for securing computer networks and their applications, focusing on intersystem, as opposed to intrasystem, security functions.

The book is designed as a tutorial/reference for any technically oriented person concerned with the design, implementation, marketing, or procurement of computer communications networks. Readers should have a basic understanding of data networking principles. No prior knowledge of security technology is needed.

The book is organized into two parts. Part I provides a technical tutorial spanning the computer communications security field. It includes an introduction to the terminology, concepts, methods, and overall architectural approaches used throughout the field. This part is intended primarily for readers without a security background. However, material on newer topics,

such as non-repudiation and the placement of security in layered architectures, may be of value to the experienced security practitioner as well.

An understanding of the Part I material is a prerequisite to Part II, which describes specific security techniques and security protocols resulting from a wide range of standardization activities. Part II is targeted equally at readers with security experience and newcomers to the field.

The "roadmap" presented in this preface will assist in appreciating the organization of the material.

Stability of the Field

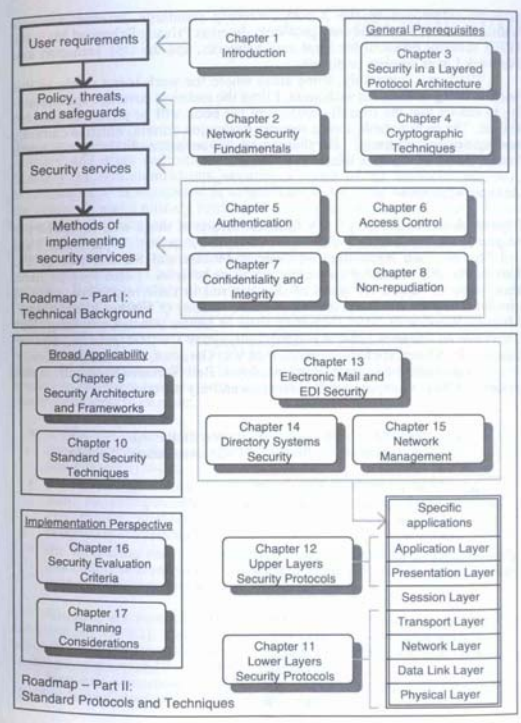
Is this a good time to publish a book on computer network security? Certainly, the subject is currently surrounded by its share of controversial issues, including:

(a) moves to promote encryption schemes that try to balance two conflicting interests — the interests of those seeking information privacy and the interests of law enforcement agencies which see uncontrolled information-hiding as an impediment to their functioning;
(b) the U.S. federal government move to standardize a new digital signature algorithm in competition with the worldwide *de facto* standard RSA algorithm; and
(c) the continuing competition between two "standard" network protocol development activities — the OSI and Internet protocol suites.

However, such issues do not necessarily imply instability in this field. Issues (a) and (b) simply underscore the need to adopt a "plug-in" philosophy when incorporating cryptographic algorithms into product or network designs. Recognizing that algorithm technology will continually advance, and that user needs, preferences, and biases will continually change, the wise designer will always use this philosophy.

Regarding issue (c), both the OSI and Internet suites have their strengths and weaknesses, and one would hope that the long term will bring a merging of the strong parts of both. From the security perspective, there is no significant difference in overall approach or in technological solutions for the two architectures. Consequently, the security aspects of both protocol suites can be covered in one coherent treatment of the field.

I believe this *is* a good time to publish a book on computer network security, because the field has just reached a solid ledge of stability, in terms of both technology and standards maturity. Technology-wise, the most significant factor is the bedding-in of public-key technology, which is now at the brink of large scale deployment. Standards-wise, the early 1990s have seen the completion of several important network security standards, such as the access



control extensions to the X.500 Directory standards, Internet network management (SNMP) security protocols, Internet Privacy Enhanced Mail, the IEEE security protocol for local area networks, and the OSI Transport and Network Layer security protocols.

There are, inevitably, some areas where the work is not stable as this book is being written. In such areas, I limit the technical coverage appropriately, to ensure that the overall accuracy of the book will be preserved into the future. The most notable area is security evaluation criteria, which is currently undergoing major change. For that reason, the coverage of this topic is much briefer than I would have liked.

Acknowledgments

This book has benefited greatly from the efforts of those who reviewed the original manuscript and provided such valuable comments. In particular, I am indebted to Vish Narayanan of General Motors and Stewart Lee of the University of Toronto for their comprehensive reviews. I also wish to thank those who reviewed the parts of the material in their respective areas of specialist expertise. These include Richard Ankney of Fischer International, Michael Ransom of NIST, Marshall Rose of Dover Beach Consulting, Bob Sharpton of Hughes, and my BNR colleagues Carlisle Adams, Sharon Boeyen, Bob Brett, Richard Thomas, Paul Van Oorschot, and Michael Wiener.

I appreciate the support of my employer, Bell-Northern Research, in this project. All views expressed, however, are entirely my own.

Warwick Ford
Ottawa, Ontario, Canada

1 Introduction

The importance of communications security has long been recognized in military and other environments where national security can be at stake. Mastery of communications security — and of its opposite number, cryptanalysis — is recognized as a significant factor in the winning of many of the century's major military conflicts, including the second world war. In this context, communications security is the means of hiding sensitive information and protecting it against tampering while in transit. Cryptanalysis is the means of compromising the communications security capabilities of one's enemies.

This book does not deal with national security grade communications security. Rather, it deals with the application of these same types of techniques to computer networks in commercial and *unclassified* government environments. Widespread application of such techniques in such environments has only recently become warranted. Sophisticated attacks, such as cryptanalysis, have long been considered so unlikely that the costs of sophisticated security measures were not justified. However, there are three major trends now leading to an urgent reassessment of this attitude and causing communications security concerns to escalate:

- the increasing interconnection of systems and of networks, making any system potentially accessible to a rapidly growing population of (known and unknown) users;
- the increasing use of computer networks for security-sensitive information; for example, electronic funds transfer, business data interchange, government *unclassified but sensitive* information, and corporate proprietary information; and
- the increasing ease of engineering a network attack, given the ready availability of increasingly sophisticated technology and the rapidly falling costs of such technology to a would-be attacker.

Hackers are now an ingrained element of the wide-area networking environment [STE1]. Networks of governments, financial institutions, telecommunications carriers, and private corporations have all been victims of hacker penetrations and they continue to be targets.

Network penetrations can have very wide-ranging impacts, as was apparent in the following well-documented cases:

- The sequence of hacker attacks into hundreds of U.S. military and government research facilities, described in detail by Cliff Stoll [STO1]. This was a case of successful (although not undetected) attacks over a period of many months. The objective was foreign espionage. The hacker's personal motivation was opportunistic financial gain. From Cliff Stoll's story, the most disconcerting message to network operators and users is the ease with which the penetrations were made, using very basic equipment and only a moderate level of technical expertise.
- The *Internet worm*, unleashed on the Internet in November 1988 by Cornell University student Robert Morris Jr. [SPA1]. The worm, a self-replicating penetration program, swept across the Internet, infecting and effectively disabling at least 1,200 (possibly up to 6,000) Internet computers running certain versions of the UNIX operating system.

Reliable statistics on hacker penetrations and other security incidents are difficult to obtain because of an unwillingness of many victims to publicly admit they are (or even were) vulnerable. An indication of trends can be obtained from security incident statistics maintained by the Internet Computer Emergency Response Team (CERT), which was formed in the wake of the Internet worm incident. The numbers of incidents reported in the 1989–1992 period are shown in Table 1-1. Note that an incident may involve one site or may involve up to thousands of sites, and incidents may be ongoing for long periods of time.

Year	*Incidents*
1989	132
1990	252
1991	406
1992	773

Table 1-1: Reported Internet Security Incidents 1989–1992

There are many potential motivations for attacks on commercial or unclassified government networks. They include financial fraud, theft of telecommunications resources, industrial espionage, illicit eavesdropping for financial or political gain, egotistical gratification, adventurism, and malicious attacks by disgruntled employees or wanton vandals. In addition to these *deliberate* types of attack, communications security needs to protect against

accidental exposures. Accidental connection of a sensitive communications session to the wrong address or accidental failure to properly protect sensitive information may prove as damaging as a successful deliberate attack.

1.1 Typical Security Requirements

The hacker threat is a concern in all networks which have public network access or which use public network facilities. However, this is not the only concern. In order to assemble a picture of other network security requirements, let us consider the main security concerns in some key network application environments.

Banking

Since the 1970s, electronic funds transfer (EFT) has been the focus for applying communications security in the financial industry [PAR1]. The primary concern is ensuring that nobody tampers with electronic transactions. There is a tremendous potential for massive fraudulent financial gain through quite simple modification of a transaction, e.g., of a dollar amount or of an account number.

This is a major issue with the financial institutions which generate and process transactions, as they face the prospect of bearing the cost of such a fraud. Even if the fraud is detected, prosecution may not be practical for many reasons, one being the public exposure damage to the institution. Because the financial system constitutes such a critical element of society, protection of this system also concerns governments. A sufficiently serious attack on a major financial system network could have the effect of destabilizing a nation's economy.

The seriousness of the financial industry's concerns and the support received from government have resulted in this industry leading the application of security technology in the commercial world.

In the 1980s, the introduction of automatic teller machine (ATM) networks and of EFT at point-of-sale (EFTPOS) raised a new set of communications security concerns in the retail banking arena [DAV1, MEY1]. The use of such facilities employs plastic cards and personal identification numbers (PINs). Because cards are frequently stolen and are easily forged, the security of the systems depends upon the secrecy of the PINs. Preservation of the secrecy of PINs is complicated by the fact that multiple separately administered networks are frequently involved in the processing of a transaction.

Furthermore, banks foresee continuing cost savings in replacing paper-based transaction services by electronic services. As new electronic services are progressively introduced, network security demands grow accordingly.

Increasingly, banks need to be sure that the party claiming to originate a transaction is genuine. An electronic equivalent of a customer's paper signature is needed. Banks are also concerned with protecting the privacy of their customers' transactions.

Electronic Trading

Electronic business data interchange (EDI) began to emerge as a major telecommunications application area in the 1980s [SOK1]. The goal of EDI is to replace the whole range of paper business transactions (e.g., purchase orders, invoices, and payments) with equivalent electronic transactions. EDI can potentially provide massive reductions in the costs of doing business.

For EDI to become widely accepted in commercial trading, security is an essential element. Users must be assured that the electronic system provides them with equivalent (if not better) protection against mistakes, misinterpretations, and fraudulent activities than the protection offered by the paper-and-signature scheme to which they are accustomed.

In EDI, there is a critical need for protection against deliberate or accidental modification of data and for assurance that the source of any transaction is legitimate. Confidentiality of transactions must also be preserved because of the corporate confidential information contained therein. In these respects EDI is similar to EFT. However, EDI introduces new security challenges because the community of users is much larger and the business associations are often much more tenuous.

EDI also introduces a major new requirement. EDI transactions constitute business contracts, which means they must have electronic signatures which have the same legal significance as paper signatures. For example, they should be acceptable as evidence in resolving disputes in courts of law. While the legal standing of electronic signatures was debated for several years, that legal standing is now being recognized. For example, in 1992 the American Bar Association approved a resolution[1] to:

> recognize that information in electronic form, where appropriate, may be considered to satisfy legal requirements regarding a writing or signature to the same extent as information on paper or in other conventional forms, when appropriate security techniques, practices, and procedures have been adopted.

Because of the cost savings to users and the extensive market opportunities for equipment vendors, EDI is viewed as a massive opportunity for users and vendors alike. Technical solutions and supporting standards for EDI security are becoming of major importance to virtually all industry.

[1] American Bar Association Resolution number 115, August 19, 1992.

Government

Governments are increasingly using computer communications networks for information transfer. Much of this information is not of a national security nature, hence is unclassified. However, security protection is required for other reasons, such as privacy legislation. This *unclassified but sensitive* information can be conveyed using commercially available networking equipment, provided adequate security provisions are employed.

For example, in the United States the Computer Security Act of 1987 introduced the concept of "sensitive information" defined as "any information, the loss, misuse or unauthorized access to or modification of which could adversely affect the national interest or the conduct of Federal programs, or the privacy to which individuals are entitled under section 552a of title 5, United States Code (the Privacy Act) , but which has not been specifically authorized under criteria established by an Executive Order or an Act of Congress to be kept secret in the interest of national defense or foreign policy."[2] The National Institute of Standards and Technology (NIST) was assigned "responsibility for developing standards and guidelines . . . to assure the cost-effective security and privacy of sensitive information." This assignment reinforced the distinction between sensitive and classified information, the latter being the province of the National Security Agency. Many other countries have comparable policies for recognizing and handling unclassified but sensitive data.

The most prominent security concern is ensuring that privacy is maintained, i.e., that information is not disclosed, either accidentally or deliberately, to anyone not authorized to have that information. Another concern is ensuring that information cannot be entered or modified by anyone not authorized to do so.

The cost savings of electronic transactions over paper transactions, e.g., the electronic filing of tax returns, are also being rapidly realized by governments. In addition to privacy guarantees, such systems introduce the need for legally enforceable electronic signatures. In 1991, with the removal of the main legal barrier, the way was paved for U.S. federal government use of electronic signatures. A Decision of the Comptroller General stated that contracts formed using electronic signatures constitute valid legal obligations of the government, provided properly secured systems are used (federal government message authentication code or digital signature standards must be followed).[3]

[2] U.S. Congress, The Computer Security Act of 1987, Public Law 100-235, January 8, 1988.

[3] Comptroller General of the United States, Decision B-245714, Decisions of the Comptroller General, vol. 71 Comp. Gen. 109 (1991), December 13, 1991.

Public Telecommunications Carriers

The management of public telecommunications networks covers a broad range of functions, collectively known as operations, administration, maintenance, and provisioning (OAM&P). These management functions themselves employ substantial data networking facilities, interconnecting a vast range of equipment and having a large population of human users (operations and maintenance personnel). While access to such networks was once tightly restricted, new access paths are continually opening up. Capabilities such as *customer network management* provide for customer personnel to access the management network to perform management operations on the public network resources used by that customer organization.

Telecommunications management networks and systems are susceptible to hacker penetrations [STE1]. A common motivation for such penetrations is theft of telecommunications services. Having penetrated network management, such theft can be engineered in various ways, such as invoking diagnostic functions, manipulating accounting records, and altering provisioning databases. Network management penetrations can also be directed at eavesdropping on subscriber calls.

A major concern of telecommunications carriers is the prospect of security compromises causing network down-time, which can be extremely costly in terms of customer relations, lost revenue, and recovery costs. Deliberate attacks on the availability of the national telecommunications infrastructure can even be viewed as a national security concern.

In addition to external penetrations, carriers are concerned about security compromises from internal sources, such as invalid changes to network management databases on the part of personnel who are not authorized to make such changes. Such occurrences may be accidental or deliberate, e.g., actions of a disgruntled employee. To protect against such occurrences, access to every management function needs to be restricted to only those individuals who legitimately require it. It is important to know correctly the identity of an individual attempting to access a management function.

Corporate/Private Networks

Virtually all private corporations have requirements to protect sensitive proprietary information. Disclosure of such information to competitors or other external entities can severely damage business, to the extent of winning or losing major contracts and possibly impacting corporate survival. Networks are being increasingly used for conveying proprietary information, e.g., between individuals, office locations, corporate subsidiaries, and/or business collaborators. Closed corporate networks are now a concept of the past — the growing trend of working at home (*telecommuting*) has ensured that.

Protection of proprietary information is not the only concern. Many organizations are entrusted with information about other organizations or individuals, for which they are obliged to provide privacy protection. Examples are the health-care and legal sectors.

Requirements for ensuring the authenticity of messages also arise in corporate networks. A sufficiently important electronic message always needs to be authenticated, in the same way that an important paper document requires a signature.

Until recently, corporations have operated under the assumption that comparatively simple protection mechanisms will satisfy their security requirements. They have not been concerned about technologically sophisticated penetrations, as might concern the government classified arena. However, there is increasing evidence that the resources of some foreign government intelligence agencies are now being used for industrial espionage purposes.[4] Commercial industry can no longer be complacent about the strength of security protective measures employed to protect sensitive proprietary information.

1.2 Security and Open Systems

While *network security* and *open systems* may appear to be contradictions in terms, this is not necessarily the case. The open-system concept represents the buyer's reaction to many years of lock-in to individual computer and communications hardware and software vendors. It is seen as the path to open choice of vendor for separate system components, with confidence that components from separate vendors will readily work together to satisfy a buyer's needs. The open-systems drive is tied to the establishment and widespread implementation of standards.

Computer networking and open systems go hand in hand. The flagship open-systems initiative — Open Systems Interconnection (OSI) — has been progressing since the 1970s, developing internationally agreed computer communications protocol standards. In addition to the formal OSI standards, open-system networking protocols have been established by other groups — notably the Internet community, with its TCP/IP protocol suite. Through these open-system networking activities, it is becoming possible to interconnect

[4] See, for example, the purported disclosures of the former head of the French Secret Service organization, DGSE [MAR1]. These disclosures indicate that the French Secret Service has a policy of supplying French industry with industrial espionage information gained through its international telecommunications interception facilities, and that such information has been successfully used by French corporations in procuring major international contracts.

equipment from many vendors, using virtually any communications technology and satisfying the needs of virtually any application.[5]

The incorporation of security protection into open-system networks is a comparatively recent endeavor. It has proven to be a complex task, largely because it represents a marriage of two technologies — security technology and communication protocol design. To provide open-system network security, it is necessary to employ *security techniques* in conjunction with *security protocols*, the latter being integrated with conventional network protocols.

Needed are compatible and complementary standards which span three broad fields:

- security techniques;
- general-purpose security protocols; and
- specific application protocols, e.g., banking, electronic mail.

Relevant standards for these fields are from four main sources[6]:

- international standards on information technology, developed under the International Organization for Standardization (ISO), the International Electrotechnical Commission (IEC), the International Telecommunication Union (ITU), and the Institute of Electrical and Electronics Engineers (IEEE);
- banking industry standards, developed either internationally under ISO or in the United States under the American National Standards Institute (ANSI);
- national government standards, especially those of the U.S. federal government; and
- Internet standards, developed by the Internet community.

Security-relevant standards from all of these sources are described in this book.

Summary

Network security is no longer a specialized requirement of military and national security environments. Network security requirements have emerged in

[5] The open systems concept has also spread beyond networking into areas such as database models and languages, computer operating system interfaces, and application software interfaces [NUT1]. However, the scope of this book is restricted to networking.

[6] Appendix A provides an overview of relevant standardization groups, including the organizations identified here, and describes the way they operate.

virtually all network application environments, including banking, electronic trading, government (unclassified), public telecommunications carriers, and corporate/private networks. A collection of typical requirements of these environments is summarized in Table 1-2.

Network security needs to be implemented in concert with the move to open-system (i.e., vendor-independent) networking. This means that the basic components of network security — security techniques and security protocols — need to be reflected in appropriate open-systems standards.

In Chapter 2, we use the collection of security requirements in Table 1-2 as an illustration of how such informally stated requirements translate to *threats*, and how *security services* can be used to counter these threats. Subsequent chapters describe methods of implementing these security services.

Application Environment	*Requirements*
All networks	Prevent outside penetrations (hackers)
Banking	Protect against fraudulent or accidental modification of transactions Identify retail transaction customers Protect PINs from disclosure Ensure customers' privacy
Electronic trading	Assure source and integrity of transactions Protect corporate privacy Provide legally binding electronic signatures on transactions
Government	Protect against unauthorized disclosure or manipulation of unclassified but sensitive information Provide electronic signatures on government documents
Public telecommunications carriers	Restrict access to administration functions to authorized individuals Protect against service interruptions Protect subscribers' privacy
Corporate/private networks	Protect corporate/individual privacy Ensure message authenticity

Table 1-2: Typical Network Security Requirements

REFERENCES

[DAV1] D.W. Davies and W.L. Price, *Security for Computer Networks*, Second Edition, John Wiley and Sons, New York, 1989.

[MAR1] P. Marion, *La Mission Impossible à la tête des Services Secrets*, Calmann-Lévy, France, 1991 (French language).

[MEY1] C.H. Meyer, S.M. Matyas, and R.E. Lennon, "Required Cryptographic Authentication Criteria for Electronic Funds Transfer Systems," *Proceedings of the 1981 Symposium on Security and Privacy*, Oakland Ca., IEEE Computer Society Press, 1981.

[NUT1] G.J. Nutt, *Open Systems*, Prentice Hall, Englewood Cliffs, NJ, 1992.

[PAR1] D.B. Parker, "Vulnerabilities of EFTs to Intentionally Caused Losses," *Communications of the ACM*, vol. 22, no. 12 (December 1979), pp. 654–660.

[SOK1] P.K. Sokol, *EDI: The Competitive Edge*, Intext Publications, McGraw-Hill Book Company, New York, 1988.

[SPA1] E.H. Spafford, "The Internet Worm: Crisis and Aftermath," *Communications of the ACM*, vol. 32, no. 6 (June 1989), pp. 678-687.

[STE1] B. Sterling, *The Hacker Crackdown: Law and Disorder on the Electronic Frontier*, Bantam Books, New York, 1992.

[STO1] C. Stoll, *The Cuckoo's Egg*, Doubleday, New York, 1989.

PART I

TECHNICAL BACKGROUND

2 Network Security Fundamentals

Traditionally, information security has been considered to have three fundamental objectives:

- *Confidentiality*: Ensuring that information is not disclosed or revealed to unauthorized persons.
- *Integrity*: Ensuring consistency of data; in particular, preventing unauthorized creation, alteration, or destruction of data.
- *Availability*: Ensuring that legitimate users are not unduly denied access to information and resources.

In contemporary computer network environments, there is one further fundamental objective which is not covered by the above:

- *Legitimate use*: Ensuring that resources are not used by unauthorized persons or in unauthorized ways.

To support these objectives, a network administration needs to have a *security policy* and needs to put in place a range of security measures to ensure that the goals of the security policy are met. Security measures fall into several categories. The two categories addressed in this book are *communications security* and *computer security*.[1] Communications security is the protection of information while it is being communicated from one system to another. Computer security is the protection of information within a computer system, and it embraces such subcategories as operating system security and database security. Communications security and computer security measures need to interwork with security measures in other categories, such as physical security and personnel security. The term *security service* is used to describe technology-based security functions provided in network systems and network products.

This chapter introduces the following fundamental concepts:

[1] This book deals mainly with communications security, but also addresses some aspects of computer security, e.g., the communication of management information between systems as needed to support computer security.

(1) security policy;
(2) threats and safeguards;
(3) the five generic security services: authentication, access control, confidentiality, data integrity, and non-repudiation; and
(4) intrusion detection and security audit.

Consideration of mechanisms for implementing network security is left to later chapters.

2.1 Security Policy

A *security policy* is a set of rules to apply to all security-relevant activities in a *security domain* (a security domain is typically the set of processing and communications resources belonging to one organization). The rules are established by an *authority* for that security domain.

Security policy is a very broad concept, and the term is used in many different ways in the literature and in standards. Some recent analyses of this issue [CHI1, STE1] have concluded that there are several different levels of security policy, such as:

- *Security policy objectives*: Statements of an organization's intent with respect to protecting identified resources.
- *Organizational security policy*: The set of laws, rules, and practices that regulate how an organization manages, protects, and distributes resources to achieve security policy objectives.[2]
- *System security policy*: Statements as to how a specific information technology system is engineered to support organizational security policy requirements.

In this book, use of the term *security policy* usually means the system security policy level, but readers should bear in mind that this forms part of the broader policy concept.

The following subsections identify some key aspects of security policy which impact network system and component design. Security policy models are discussed further in Chapter 6. For a more detailed coverage of the subject, see [NAT1].

[2] Organizational security policy may have implications upon the full range of safeguard categories, including communications security, computer security, physical security, and personnel security.

Authorization

Authorization is a fundamental part of a security policy. Authorization is the granting of rights — it amounts to establishing who may do what to what. Examples of authorization statements (at the organizational security policy level) are:

(a) File *Project-X-Status* can only be modified by person *G. Smith*, and can only be read by persons *G. Smith*, *P. Jones*, and members of the *Project-X* project team.
(b) A personnel record can only be created or modified by personnel division staff; it can be read by personnel division staff, by executive managers, and by the individual to whom it refers.
(c) Information classified within the hierarchy *confidential-secret-top secret* can only be disclosed to an individual with a clearance level equal to or greater than the classification level.

These policy statements place demands on various safeguard categories, e.g., personnel security for determining peoples' clearances. In the computer and communications categories, the main demands are reflected in a type of system security policy called *access control policy*.

Access Control Policies

Access control policies are system security policies which drive the automated enforcement of authorization in computer systems and networks. The example authorization statements (a), (b), and (c) above map to different types of access control policy, respectively:

(a) *Identity-based policy*: A policy that permits or denies access to explicitly identified individuals or groups of individuals;
(b) *Role-based policy*: A variation of an identity-based policy that assigns roles to individuals and applies authorization rules based on these roles; and
(c) *Multi-level policy*: A policy employing general rules, based on broad levels of the sensitivity of information and corresponding levels of clearances of people.

Access control policies are sometimes categorized as being *mandatory access control* or *discretionary access control* policies [DOD1]. Mandatory policies are imposed by the security domain authority and cannot be circumvented by individual users. Mandatory policies are most common in military

and other government classified environments; policy (c) above is an example. Discretionary policies provide particular users with access to resources (e.g., information), then leave it to these users to control further access to the resources. Policies (a) and (b) above are examples of discretionary policies. In classified environments, discretionary policies are used to enforce the *need-to-know* concept.

Access control policies are discussed further in Chapter 6.

Accountability

A fundamental principle underlying any security policy is *accountability*. Individuals taking actions which are governed by security policy need to be accountable for their actions. This provides an important link to personnel security. Some network security safeguards, involved with authenticating human identities and linking actions with such identities, contribute directly to support of this principle.

2.2 Threats and Safeguards

Basic Concepts

A *threat* is a person, thing, event, or idea which poses some danger to an asset, in terms of that asset's confidentiality, integrity, availability, or legitimate use. An *attack* is an actual realization of a threat. *Safeguards* are physical controls, mechanisms, policies, and procedures that protect assets from threats. *Vulnerabilities* are weaknesses in a safeguard, or the absence of a safeguard.

Risk is a measure of the cost of a realized vulnerability that incorporates the probability of a successful attack. Risk is high if the value of a vulnerable asset is high, and the probability of a successful attack is high. Conversely, risk is low if the value of the vulnerable asset is low and the probability of a successful attack is low. Risk analysis can provide a quantitative means of determining whether expenditure on safeguards is warranted.

Threats can sometimes be classified as being deliberate (e.g., hacker penetration) or accidental (e.g., message sent in error to the wrong address). Deliberate threats may be further classified as being *passive* or *active*. Passive threats involve monitoring but not alteration of information (e.g., wiretapping). Active threats involve deliberate alteration of information (e.g., changing the dollar amount of a financial transaction). In general, passive attacks are easier and less costly to engineer than active attacks.

There is no universally agreed way to identify, classify, or interrelate threats. The existence and significance of different threats varies from

environment to environment. However, in order to explain the role of network security services, we can assemble a picture of the threats typically encountered in contemporary computer networks. We do this in three stages — we identify *fundamental threats*, then *primary enabling threats*, then *underlying threats*.

Fundamental Threats

The four fundamental threats directly reflect the four security objectives identified at the start of the chapter:

- *Information leakage*: Information is disclosed or revealed to an unauthorized person or entity. We shall later see how this might involve direct attacks, such as eavesdropping or wiretapping, or more subtle types of information observation.
- *Integrity violation*: The consistency of data is compromised through unauthorized creation, alteration, or destruction of data.
- *Denial of service*: Legitimate access to information or other resources is deliberately impeded. This might involve, for example, making a resource unavailable to legitimate users through a heavy load of illegitimate, unsuccessful access attempts.
- *Illegitimate use*: A resource is used by an unauthorized person or in an unauthorized way. Examples are an intruder penetrating a computer system and using that system either as the basis of theft of telecommunications services or as a staging point for penetrating another system.

Primary Enabling Threats

The primary enabling threats are significant because a realization of any of these threats can lead directly to a realization of any of the fundamental threats. These threats therefore *enable* the fundamental threats. The primary enabling threats comprise *penetration* threats and *planting* threats. The main penetration threats are:

- *Masquerade*: An entity (person or system) pretends to be a different entity. This is the most common way of penetrating a security perimeter, e.g., a computer's *login perimeter*. An unauthorized entity convinces a perimeter guard that he is an authorized entity and thereafter assumes the rights and privileges of the authorized entity. Hackers succeed largely through the use of masquerade.
- *Bypassing controls*: An attacker exploits systems flaws or security weaknesses (e.g., system "features" whose existence was intended to be kept secret), in order to acquire unauthorized rights or privileges.

- *Authorization violation*: A person authorized to use a system or resource for one purpose uses it for another, unauthorized purpose. This is known as an *insider* threat.

The main planting threats are:

- *Trojan horse*: Software contains an invisible or apparently innocuous part which, when executed, compromises the security of its user. An example of a Trojan horse is a software application which has an outwardly legitimate purpose, e.g., text editing, but which also has a surreptitious purpose, e.g., copying user documents into a hidden private file which is read later by the attacker who planted the Trojan horse.
- *Trapdoor*: A feature is built into a system or system component such that the provision of specific input data allows security policy to be violated. An example is a login-processing subsystem which allows processing of a particular user-identifier to bypass the usual password checks.

Planting threats are usually realized by the planting party only after the planted capability has been left dormant for a period of time.

Underlying Threats

If we analyze any of the fundamental threats or primary enabling threats in a given environment, we can identify particular *underlying threats*, any of which may enable the more fundamental threats. For example, if we consider the fundamental threat of information leakage, we might find several underlying threats (apart from the primary enabling threats), such as:

- eavesdropping;
- traffic analysis;
- indiscretions by personnel; and
- media scavenging.

Figure 2-1 illustrates typical threats, and their interrelationships. Note that the path can become convoluted. For example, masquerade is a threat which can underlie all fundamental threats. However, masquerade can itself have information leakage as an underlying threat (because information leakage might reveal a password, which can enable masquerade). Table 2-1 lists the threats identified.

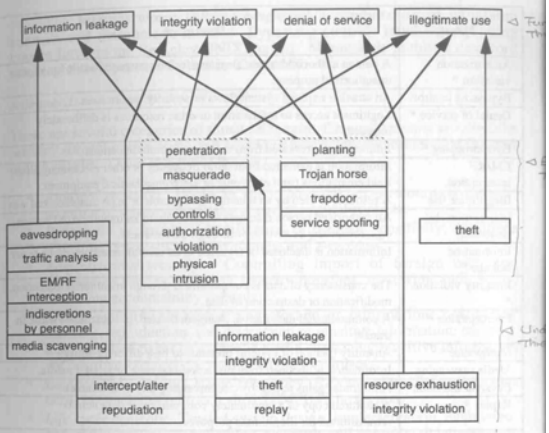


Figure 2-1: Threats Underlying Threats

An indication of the real-life significance of such threats can be obtained from [NEU1]. In a sampling of a collection of over 3,000 cases of computer system abuse, drawn from the media and personal reporting, the following threats were the most predominant (in order of decreasing frequency):

- authorization violation;
- masquerade;
- bypassing controls;
- Trojan horse or trapdoor; and
- media scavenging.

Threat	Description
Authorization violation *	A person authorized to use a system for one purpose uses it for another, unauthorized purpose.
Bypassing controls	An attacker exploits system flaws or security weaknesses.
Denial of service *	Legitimate access to information or other resources is deliberately impeded.
Eavesdropping *	Information is revealed from monitored communications.
EM/RF interception	Information is extracted from radio frequency or other electromagnetic field emanations from electronic or electromechanical equipment.
Illegitimate use	A resource is used by an unauthorized person or in an unauthorized way.
Indiscretions by personnel	An authorized person discloses information to an unauthorized person, e.g., for money or favors, or through carelessness.
Information leakage *	Information is disclosed or revealed to an unauthorized person or entity.
Integrity violation *	The consistency of data is compromised through unauthorized creation, modification or destruction of data.
Intercept/alter *	A communicated data item is changed, deleted, or substituted while in transit.
Masquerade *	An entity (person or system) pretends to be a different entity.
Media scavenging	Information is obtained from discarded magnetic or printed media.
Physical intrusion	An intruder gains access by circumventing physical controls.
Replay *	A captured copy of a legitimately communicated data item is retransmitted for illegitimate purposes.
Repudiation *	A party to a communication exchange later falsely denies that the exchange took place.
Resource exhaustion	A resource (e.g., access port) is deliberately used so heavily that service to other users is disrupted.
Service spoofing	A bogus system or system component aims to dupe legitimate users or systems into voluntarily giving up sensitive information.
Theft	A security-critical item, e.g., a token or identity card, is stolen.
Traffic analysis *	Information is leaked to unauthorized entities, through observation of communications traffic patterns.
Trapdoor	A feature is built into a system or system component such that the provision of specific input data allows security policy to be violated.
Trojan horse	Software that contains an invisible or apparently innocuous part which, when executed, compromises the security of its user.

* Threats that computer communications security can counter.

Table 2-1: Typical Network Threats

The Internet worm [SPA1] employed a combination of bypassing controls and masquerade attacks. Bypassing controls involved exploiting known flaws in the Berkeley UNIX system. Masquerade involved password cracking (this is pursued further in Chapter 5).

Safeguards

There are several categories of threat safeguard. *Communications security*, the primary subject of this book, and *computer security* were introduced at the start of this chapter. Other categories include:

- *Physical security*: Locks or other physical access controls; tamper-proofing of sensitive equipment; environmental controls.
- *Personnel security*: Identification of position sensitivity; employee screening processes; security training and awareness.
- *Administrative security*: Controlling import of foreign software; procedures for investigating security breaches, reviewing audit trails, reviewing accountability controls.
- *Media security*: Safeguarding storage of information; controlling marking, reproduction, and destruction of sensitive information; ensuring that discarded paper or magnetic media containing sensitive information are destroyed securely; scanning media for viruses.
- *Emanations security*: Radio frequency (RF) and other electromagnetic (EM) emanations controls (called TEMPEST protection).
- *Life cycle controls*: Trusted system design, implementation, evaluation, and endorsement; programming standards and controls; documentation controls.

**A security system is only as strong as its weakest link. For effective security, countermeasures from the different categories need to be used together. For example, a technically sound password system, designed to counter masquerade, will be ineffective if users leave written passwords in an unsecured place or can be duped into revealing a password to an unknown telephone caller.

Safeguards can be provided against most threats, but every safeguard has a cost. A network user or procurer needs to carefully consider whether the potential cost of a successful attack warrants expenditure on safeguards. For example, in commercial networks, countermeasures are not normally provided against EM/RF interception, because the risks are very small and the countermeasures are very costly (in a classified environment, the analysis may produce a different conclusion). The field of *risk management* assists in making such decisions. Various qualitative and quantitative risk management

tools have been developed; for further information see [COX1, GIL1, HOF1, MIG1, PFL1].

Viruses

A virus is a piece of executable program code that can "infect" other programs by modifying them to include a (possibly evolved) copy of itself. A virus generally contains two functions — a function which causes the "infection" of other programs, plus another function which is usually either a damage-causing function or a planted attack capability.

While there is little that can be done to totally eliminate viruses in "open" environments, their spread can be readily controlled through rigid software and media management practices in conjunction with some technological support. The technological support involves both proactive and reactive "antiviral" software.

Viruses are a computer security problem, not limited to network environments. Nevertheless, network environments can foster the spread of viruses, by facilitating transfer of executable program code between systems. Access controls to executable program resources are therefore particularly important. In countering viruses, the benefits of network environments should be exploited, e.g., by providing centrally initiated virus scans of distributed computers.

For an excellent study of the fundamental nature of viruses, see [COH1]. For a more recent insight into the practicalities of the virus problem, see [KEP1].

2.3 Security Services

In the computer communications context, the main security safeguards are known as *security services*. There are five generic security services[3]:

- *Authentication service*: Provides assurance of the identity of some entity (a person or a system).
- *Access control service*: Protects against unauthorized use or manipulation of resources.
- *Confidentiality service*: Protects against information being disclosed or revealed to unauthorized entities.

[3] The security service concept and the five generic security services are from the OSI Security Architecture standard ISO/IEC 7498-2.

- *Data integrity service*: Protects against data being changed, deleted, or substituted without authorization.
- *Non-repudiation service*: Protects against one party to a communication exchange later falsely denying that the exchange occurred.

Security policy for a security domain will govern whether any security service is used within that domain or in communications between that domain and other domains. It will also govern under what circumstances a security service is used, and what constraints are placed upon any variable parameters of such a service.

None of the generic security services was conceived specifically for data communications environments or even electronic environments. All of these services have non-electronic analogs, which employ supporting mechanisms familiar to most readers. Table 2-2 provides some examples.

Security Service	*Non-electronic Mechanism Examples*
Authentication	Photo-identification card Knowledge of mother's maiden name
Access Control	Locks and keys; master key system Checkpoint guard
Confidentiality	Sealed letter; opaque envelope Invisible ink
Integrity	Indelible ink Hologram on credit card
Non-repudiation	Notarized signature Certified or registered mail

Table 2-2: Non-electronic Security Mechanisms

In the following subsections, the purposes of the five security services are spelled out in more detail. The means of providing them are addressed in Chapters 5 through 8.

Authentication

Authentication services provide assurance of the identity of someone or something. This means that when someone (or something) claims to have a particular identity (e.g., a particular user name), an authentication service will provide a means of confirming that this claim is correct. Passwords are a well-known way of providing authentication.

Authentication is the most important of the security services, because all other security services depend upon it to some extent. Authentication is the means of countering the threat of masquerade which can directly lead to compromise of any of the fundamental security objectives.[4]

Authentication applies in a particular context, i.e., the context in which the identity is presented. Two important cases are:

- An identity is presented by a remote party participating in a communication connection or session. The authentication service in this case is known as *entity authentication*.
- An identity claiming to be that of the originator of a data item is presented along with that data item. The authentication service in this case is known as *data origin authentication*.

Note that data origin authentication can be used to authenticate the real source of a data item, regardless of whether that source is involved in current communications activities. For example, the data item may have been relayed through many systems whose identities may or may not have been authenticated.

Both types of authentication services have significant roles in meeting the fundamental security objectives. Data origin authentication is a direct means of ensuring part of the integrity objective, i.e., ensuring that the true source of a data item is known. Entity authentication contributes in various ways to meeting security objectives:

- As a necessary support for access control services, the operation of which depends upon assured knowledge of identities. (Access control services contribute directly to meeting confidentiality, integrity, availability, and legitimate use objectives.)
- As a possible means for providing data origin authentication (when used in conjunction with a data integrity mechanism).
- As a direct support for the accountability principle, e.g., providing assured identities associated with actions, for recording in audit trails.

An important special case of entity authentication is *personal authentication*, that is, authentication of a person at a network termination point. It requires special recognition for two reasons. The first is that different people can very easily replace each other at a termination point. The second is that special technologies may be applicable in identifying individual people.

The provision of authentication services is discussed in Chapter 5.

[4] Note also that, as masquerade is the number-one tool of the hacker, authentication is the primary countermeasure against hacker attacks.

Access Control

The goal of access control is to protect against unauthorized access to any resource (e.g., computing resource, communications resource, or information resource). The term *unauthorized access* includes unauthorized use, unauthorized disclosure, unauthorized modification, unauthorized destruction, and unauthorized issuing of commands. Access control contributes directly to achieving the security goals of confidentiality, integrity, availability, and legitimate use. The contribution to confidentiality, integrity, and legitimate use goals is obvious. The contribution to the availability goal lies in controlling:

- who can issue network management commands which may impact network availability;
- who can tie up resources in a way which does not productively use them; and
- who can learn information which may be used subsequently in a denial-of-service attack.

Access control is a means for enforcing authorization. It is as much a computer security (operating system) issue as a communications security issue. However, it places significant requirements on communications protocols, because of the need to communicate access control information between systems.

The general model for access control assumes a set of active entities, called *initiators*, or *subjects*, which attempt to access members of a set of passive resources called *targets*, or *objects*. While the subject/object terminology is used extensively in the literature, the term *object* can be confusing because of overloading of this term in modern computer and communications technologies. For this reason, the initiator/target terminology is being increasingly used, especially in security standards (and is favored in this book).

Authorization decisions govern which initiators may access which targets for which purposes and under which conditions. These decisions are reflected in an access control policy. Access requests are filtered through an access control mechanism which enforces the access control policy.

An access control mechanism can be modeled as comprising two conceptual components — an *enforcement function* and a *decision function*. The OSI access control model (in standard ISO/IEC 10181-3) uses these concepts; it is illustrated in Figure 2-2. In practice, the physical configurations of these components may vary widely. Typically, some are co-located. However, there is generally a need to communicate access control information between these components. The access control service provides for this communication.

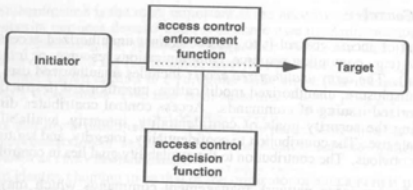


Figure 2-2: Access Control Model — Basic Conceptual Components

Another aspect of access control is preventing sensitive information from being transmitted through environments where it might be at risk. This involves the enforcement of *routing control* of network traffic or messages.

Further discussion of access control services depends on two factors: the type of access control policy and the configuration of the components. Such discussion is deferred to Chapter 6.

Confidentiality

Confidentiality services protect against information being disclosed or revealed to entities (e.g., people or organizations) not authorized to have that information.

At this point it is worthwhile emphasizing the difference between *information* and *data*. Information has semantics, or meaning. A data item is a string of bits, commonly used to store or communicate an encoded representation of a piece of information. Hence, a data item in storage or communication constitutes one form of *information channel* [DEN1]. However, this is not the only information channel in a computer communications environment. Other information channels include:

- observing the existence or non-existence of a data item (regardless of its contents);
- observing the size of a data item; and
- observing dynamic variations in data item characteristics (data item contents, existence, size, etc.).

For example, the information "the missile has been launched" may be communicated (intentionally or unintentionally) in any of the following ways:

- a one-bit data item, value 1 indicating "the missile has been launched" and value 0 indicating "the missile has not been launched";
- the existence or non-existence of a file called "missile launch report" in some file directory;
- the fact that the size of a data item containing a list of missile launch records is greater than it was after the preceding missile launch; or
- the fact that a counter of missile flight miles can be observed to be continually incrementing.

Achieving a confidentiality objective ideally requires protecting against disclosure of information through any such information channel.

In computer communications security, this leads to the recognition of two types of confidentiality service. A *data confidentiality service* makes it infeasible to deduce sensitive information from the content or size of a given data item (e.g., using encryption). A *traffic flow confidentiality service* makes it infeasible to deduce sensitive information by observing network traffic flows.

Considering data confidentiality services, there are several variations, depending on their granularity, i.e., to what data item(s) confidentiality is to apply. Three different cases are significant. The first, known as a *connection confidentiality* service, applies to all data transmitted on a connection.[5] The second, known as a *connectionless confidentiality* service, applies to all the data comprising one connectionless data unit. The third is a *selective field confidentiality* service, which applies only to nominated fields within a data unit (either a data unit sent on a connection or a connectionless data unit).

The means of providing confidentiality services are discussed in Chapter 7, building on the discussion of cryptographic techniques in Chapter 4.

Data Integrity

Data integrity services (or simply *integrity services*)[6] act as safeguards against the threat that the value or existence of data might be changed in a way inconsistent with the recognized security policy. Changing the value of a data item includes inserting additional data or deleting, modifying, or reordering parts of the data. Changing the existence of a data item means creating or deleting it.

[5] The terms *connection* and *connectionless* are explained in Section 3.1.

[6] The wider subject of information integrity has been expounded in the *Clark-Wilson Model* [CLA1]. This model defines integrity as those qualities which give data and systems both internal consistency and a good correspondence to real-world expectations for the systems and data. Controls are needed for both internal consistency and external consistency. Communications data integrity cannot help with external consistency. It helps with internal consistency of data by ensuring that a data item maintains its value while in communication.

Depending on the environment, any of the above forms of threat may be serious. Consider, for example, a (hypothetical) automatic teller machine (ATM) communicating with its controlling bank. If adequate consideration is not given to integrity, it is conceivable that any of the following attacks might be perpetrated by someone inserting his or her own equipment in the communications link:

- *Modifying* the dollar amount on a withdrawal-request message from ATM to bank (say from $500 to $50), then modifying the amount on the approval message from bank to ATM (back to the original $500), causing the ATM to pay out much more than was recorded for debit at the bank.
- *Repeating* the physical motions of an earlier withdrawal at the ATM, but passing nothing to the bank and replaying to the ATM all the responses recorded from the earlier transaction. The ATM pays out again, but nothing is debited at the bank.
- *Creating* a "cancel transaction — cash dispensing failed" message from ATM to bank in the dispense phase, even though the cash has been dispensed successfully.
- *Deleting* all black-list notifications from bank to ATM (used by those ATMs which sometimes operate in an off-line mode), then making a withdrawal with a voided (black-listed) card at an off-line time.

Similar to confidentiality, an important characteristic of a data integrity service is its granularity, i.e., to what data item(s) it is to apply. Three cases are significant. The first, known as a *connection integrity* service, applies to all data transmitted on a connection. The second, known as a *connectionless integrity* service, applies to all the data comprising one connectionless data item. The third is a *selective field integrity* service, which applies only to nominated fields within a data unit.

All data integrity services counter attempts to create or modify data. However, they do not all necessarily counter attempts to duplicate or delete data (which are harder to detect). Duplication can result from a replay attack. Connectionless and selective field integrity services, which are primarily concerned with detecting modification of individual pieces of data, may or may not detect replay. A connection integrity service is required to protect against replay of data within a connection. However, there may still be a vulnerability, as it might be possible for an intruder to *replay an entire connection*. Detecting deletion of pieces of data (e.g., one connectionless data unit, one selective field, or possibly one complete connection) is at least as difficult as detecting replay and may require special attention in specifying any data integrity service.

A connection integrity service may also offer the option of *recovery*. If so, when an integrity violation is detected within a connection, the service will

attempt to recover. For example, the communications will be rolled back to some checkpoint and restarted.

The means of providing data integrity services are discussed in Chapter 7, building on the discussion of cryptographic techniques in Chapter 4.

Non-repudiation

Non-repudiation is fundamentally different from the other security services. Its primary purpose is to protect communications users against threats from other legitimate users, rather than from unknown attackers. Repudiation was earlier defined to be a threat, whereby a party to a communication exchange later falsely denies that the exchange took place. Non-repudiation services counter this type of threat.

The word *non-repudiation* is not, in itself, very helpful. This service does not eliminate repudiation. It does not prevent any party from repudiating another party's claim that something occurred. What it does is ensure the availability of irrefutable evidence to support the speedy resolution of any such disagreement.

The motivation for non-repudiation services is not just the possibility that communicating parties may try to cheat each other. It is also a reflection of the reality that no system is perfect, and that circumstances can arise in which two parties end up with different views of something that happened.

Consider first some of the problems that can arise in the world of paper business transactions. Paper documents, such as contracts, quotations, bids, orders, invoices, and checks play a critical role in the conducting of business between organizations. However, many problems can occur in their handling, such as:

- document lost in the mail;
- document lost by recipient before processing;
- document generated by a person with insufficient authorization;
- document accidentally corrupted within an organization or while in transit between organizations;
- document fraudulently modified within an organization or while in transit between organizations;
- forged document; and
- disputed filing time of a document.

To aid in systematically dealing with such problems, various mechanisms are employed, such as signatures, countersignatures, notarized signatures, receipts, postmarks, and certified mail. If good business practices are followed, there will usually be an adequate paper trail to make dispute

resolution straightforward. If necessary, evidence will be available in the form of records held by the disputing parties plus third parties such as the post office, courier agents, and notaries. With the help of such evidence, parties may be able to resolve their differences themselves or, in some cases, may need to settle their dispute under arbitration, e.g., in a court of law.

With electronic business transactions, the problems that can arise are analogous to those for paper transactions. Non-repudiation services provide the protective mechanisms. In some respects the problems with electronic transactions are more difficult to resolve than those with paper transactions. There are fewer humans involved in the handling of documents (interrogating these humans assists greatly in resolving paper transaction problems). Also, original paper documents are not generally available for reference (another key basis for resolving paper transaction problems). However, in other respects, problems with electronic transactions are easier to solve. This results from the availability of sophisticated technologies such as digital signatures (which are described in Chapter 4).

In principle, non-repudiation can apply to any of a variety of events which affect two or more parties. In general, disagreements relate to whether a particular event occurred, when it occurred, what parties were involved with that event, and what information was associated with that event. If we restrict our concerns to data networking environments, repudiation scenarios can be separated into two distinct cases:

- *Repudiation of origin*: There is disagreement as to whether a particular party originated a particular data item (and/or disagreement as to the time this origination occurred).
- *Repudiation of delivery*: There is disagreement as to whether a particular data item was delivered to a particular party (and/or disagreement as to the time this delivery occurred).

These scenarios lead to two distinct variants of non-repudiation service.

The mechanisms for providing non-repudiation services are discussed in Chapter 8.

Application Examples

Chapter 1 (Table 1-2) identified typical security requirements of some important network application environments. That discussion can now be taken two steps further — mapping the requirements to threats and identifying security services which can constitute safeguards. Table 2-3 shows typical threats associated with the identified requirements (considering only those threats pertinent to computer communications security). Table 2-4 indicates the security services used to counter the threats identified in Table 2-3.

Requirements (Informal)	*Threats*
All Networks:	
Prevent outside penetrations (hackers)	Masquerade
Banking:	
Protect against fraudulent or accidental modification of transactions	Integrity violation
Identify retail transaction customers	Masquerade, repudiation
Protect PINs from disclosure	Eavesdropping
Ensure customers' privacy	Eavesdropping
Electronic Trading:	
Assure source and integrity of transactions	Masquerade, integrity violation
Protect corporate privacy	Eavesdropping
Provide legally-binding electronic signatures on transactions	Repudiation
Government:	
Protect against unauthorized disclosure or manipulation of unclassified but sensitive information	Masquerade, authorization violation, eavesdropping, integrity violation
Provide electronic signatures on government documents	Repudiation
Public Telecommunications Carriers:	
Restrict access to administration functions to authorized individuals	Masquerade, authorization violation
Protect against service interruptions	Denial of service
Protect subscribers' privacy	Eavesdropping
Corporate/Private Networks:	
Protect corporate/individual privacy	Eavesdropping
Ensure message authenticity	Masquerade, integrity violation

Table 2-3: Typical Threats in Particular Application Environments

2.4 Intrusion Detection and Security Audit

Intrusion detection is a general term for automated methods which, based on the analysis of real-time event sequences and/or accumulated records, can alert a security administrator to possible security violations. The main goal of these methods is to detect unusual activity, such as a large number of unsuccessful login attempts from one terminal or a large number of attempts to access a computer's password file. The methods used are typically based on statistical

Threat	*Security Service*
Masquerade	Authentication
Authorization violation	Access control
Eavesdropping	Confidentiality
Integrity violation	Integrity
Repudiation	Non-repudiation
Denial of service	Authentication, access control, integrity

Table 2-4: Security Services to Counter Typical Threats

analysis or, increasingly, on rule-based expert systems. Intrusion detection is a particularly powerful security tool because of its ability to counter attacks from insiders who misuse their privileges and attacks resulting from such events as lost or stolen passwords or cryptographic keys.

A *security audit* is an independent examination and review of system records and procedures. Its purposes include testing the adequacy of security policy, confirming compliance with security policy, recommending changes in security policy, assisting in the analysis of attacks, and gathering evidence for use in prosecuting an attacker.

A *security audit trail* is a journal of security-related events collected for potential use in intrusion detection and/or security audits.

A *security alarm* is a real-time indication of a situation which suggests a security violation. It involves sending a message to a system operator, administrator, or manager. Its purpose is to alert relevant staff so that damage from a security violation may be minimized, further violations may be prevented, and/or evidence may be gathered for use in prosecuting an attacker.

None of the above concepts is identified as a security service in the OSI Security Architecture standard. However, security audit trail and security alarm are recognized as *pervasive security mechanisms* which supplement the basic security services identified in Section 2.3. This distinction is moot. Security audit trail and alarm functions provide their own types of service to a system or network administrator — services to aid detection of weaknesses in security policy, procedures, and mechanisms, and to gather evidence. They also constitute a safeguard in their own right, by acting as a deterrent to would-be attackers.

Security audit trail and alarm functions are driven by *intrusion detection policy* and *audit policy*, which are elements of security policy. A vast range of security-related events occurs in a system, e.g., every request to establish a

connection or to access a sensitive information item. Intrusion detection policy and audit policy govern which of these events cause an audit trail record to be generated and which cause an alarm to be signaled. Such decisions may be based on complex intrusion detection procedures. The response to an event may depend on criteria such as the time of day, a threshold counter, the event type, and/or the event source. Audit policy further governs the type of post-event analysis to be systematically applied to an audit trail.

Protocols supporting the communications needs of security audit trail and alarm functions are addressed in Chapter 15. For further reading on intrusion detection and security audit see [LUN1, MCA1, SCH1, WIN1].

Summary

Information security has four fundamental objectives — confidentiality of information, integrity of data, availability of information and resources, and legitimate use of resources. To support these objectives, a network administration needs to have a security policy, and needs to put in place a range of security measures to ensure that the goals of the security policy are met. A security policy is a high-level set of rules typically applied across the set of communications resources belonging to one organization. Access control policy is the part of a security policy governing who may access information or other resources.

The determination of security requirements is based on assessing threats. Four fundamental threats follow directly from the above objectives: information leakage, integrity violation, denial of service, and illegitimate use. Primary enabling threats, which can directly enable any of the fundamental threats, involve either penetration (e.g., masquerade, bypassing controls, or authorization violation) or planting (e.g., Trojan horse or trapdoor). In a given environment, any threat can be considered to have underlying threats. Such threats include eavesdropping, intercept/alter, traffic analysis, and repudiation. Safeguards protect assets from threats. Safeguards fall into several categories, including communications security and computer security.

In the computer communications arena, safeguards are described in terms of security services. The five basic security services are authentication, access control, confidentiality, data integrity, and non-repudiation. Authentication services provide assurance of the identity of someone or something. Access control services protect against unauthorized access to resources, by enforcing an access control policy. Confidentiality services protect against information being disclosed or revealed to people who are not authorized to have it. Data integrity services counter the threat that the value or existence of data might be changed (in a way inconsistent with the recognized security policy). Non-

repudiation services protect against a communications user falsely denying that an earlier communication exchange took place.

Intrusion detection denotes a method which, based on the analysis of real-time event sequences and/or accumulated records, alerts a security administrator to a possible security violation. A security audit is an independent examination and review of system records and procedures. A security alarm is a real-time indication of a situation which suggests a security violation.

Exercises

1. List seven threats for which network security services can act as safeguards. Give a real-life example of each. For each threat identified, name the security service (or services) which may be used to counter it.

2. Describe:
   (a) the differences between entity authentication and data origin authentication;
   (b) the different variations of confidentiality services;
   (c) the different variations of data integrity services; and
   (d) the relationship between authorization and access control.

3. Suppose a password-based authentication system on a small network is managed by storing all user passwords on a password-server system. When a user wants to change his password, he does this through a short communications session with the password-server. When a host wants to authenticate a user, it does so by obtaining the user's password from the password-server and comparing it with that presented by the user. Identify the primary threats to such a system, and briefly describe how authentication, confidentiality, data integrity, and access control services might contribute to countering these threats.

REFERENCES

[CHI1] D.M. Chizmadia, "Some More Thoughts on the Buzzword 'Security Policy,'" *Proceedings of the 15th National Computer Security Conference*, October 1992, Baltimore, MD, pp. 651-660.

[CLA1] D.D. Clark and D.R. Wilson, "A Comparison of Commercial and Military Security Policies," in *Proceedings of the 1987 IEEE Symposium on Security and Privacy*, Oakland, CA, IEEE Computer Society Press, 1987, pp. 184-94.

[COH1] F. Cohen, "Computer Viruses: Theory and Experiments," *Computers and Security*, vol. 6, no. 1 (February 1987), pp. 22-35.

[COX1] R. Cox, M. O'Neill, and W. Price, "Risk Management of Complex Networks", *Proceedings of the 15th National Computer Security Conference*, October 1992, Baltimore, MD, pp. 544-553.

[DEN1] D.E. Denning, *Cryptography and Data Security*, Addison-Wesley, Reading, MA, 1982.

[DOD1] U.S. Department of Defense, *Department of Defense Trusted Computer System Evaluation Criteria*, DOD 5200.28-STD, National Computer Security Center, Fort Meade, MD, December 1985.

[GIL1] I.E. Gilbert, *Guide for Selecting Automated Risk Analysis Tools*, NIST Special Publication 500-174, U.S. Department of Commerce, National Institute of Standards and Technology, 1989.

[HOF1] L. Hoffman, "Risk Analysis and Computer Security: Bridging the Cultural Gap," *Proceedings of the 9th National Computer Security Conference*, October 1986, Baltimore, MD.

[KEP1] J.O. Kephart, S.R. White, and D.M. Chess, "Computers and Epidemiology," *IEEE Spectrum*, vol. 30, no. 5 (May 1993), pp. 20-26.

[LUN1] T.F. Lunt, "Automated Audit Trail Analysis and Intrusion Detection: A Survey", *Proceedings of the 11th National Computer Security Conference*, October 1988, Baltimore, MD, pp. 65-73.

[MCA1] N. McAuliffe, *et al.*, "Is Your Computer Being Misused? A Survey of Current Intrusion Detection System Technology," *Proceedings of the Sixth Annual Computer Security Applications Conference*, IEEE Computer Society Press, Los Alamitos, CA, December 1990, pp. 260-272.

[MIG1] J. Miguel, "A Composite Cost/Benefit/Risk Methodology," *Computer Security: A Global Challenge, Proc. IFIP Conference*, 1984, pp. 307-312.

[NAT1] National Research Council (U.S.), *Computers at Risk: Safe Computing in the Information Age*, National Academy Press, Washington, DC, 1990.

[NEU1] P.G. Neumann and D.B. Parker, "A Summary of Computer Misuse Techniques," *Proceedings of the 12th National Computer Security Conference*, October 1989, Baltimore, MD, pp. 396-407.

[PFL1] C.P. Pfleeger, *Security in Computing*, Prentice Hall International, Inc., Englewood Cliffs, NJ, 1989.

[SCH1] S.I. Schaen, "Network Auditing: Issues and Recommendations," *Proceedings of the Seventh Annual Computer Security Applications Conference*, IEEE Computer Society Press, Los Alamitos, CA, December 1991, pp. 66-79.

[SPA1] E.H. Spafford, "The Internet Worm: Crisis and Aftermath," *Communications of the ACM*, vol. 32, no. 6 (June 1989), pp. 678-687.

[STE1] D.F. Sterne, "On the Buzzword 'Security Policy,'" *Proceedings of the 1991 Symposium on Security and Privacy*, Oakland, CA, IEEE Computer Society Press, 1991, pp. 218-230.

[WIN1] J.R. Winkler and L.C. Landry, "Intrusion and Anomaly Detection: ISOA Update," *Proceedings of the 15th National Computer Security Conference*, October 1992, Baltimore, MD, pp. 272-281.

Standards

ISO/IEC 7498-2: *Information Technology — Open Systems Interconnection — Basic Reference Model — Part 2: Security Architecture* (Also ITU-T Recommendation X.800).
ISO/IEC 10181-3: *Information Technology — Security Frameworks in Open Systems — Access Control Framework* (Also ITU-T Recommendation X.812) (Draft).

3 Security in a Layered Protocol Architecture

Layered protocol architectures are fundamental to modern computer networking. They allow network designs to accommodate unlimited applications, unlimited underlying media technologies, and unlimited interconnection techniques. The primary purpose of layering is to modularize the protocol specification problem, such that separate pieces of the protocol puzzle can be developed independently and mixed and matched in different ways to give a "complete" protocol. To some extent, this modularization can also flow into implementation, such that different protocol components can be realized in different software modules or hardware products. This chapter addresses the important issue of how the provision of network security relates to architectural layering.

The Open Systems Interconnection (OSI) architecture is the recognized basis of protocol layering. The first OSI standard, the Basic Reference Model (ISO/IEC 7498-1), establishes this architectural model. Other OSI standards define specific protocols to fit into this model. Other protocol architectures, notably the Internet TCP/IP suite, define protocols which constitute alternatives to formal OSI protocols at some layers, but nevertheless fit into the same overall layering scheme.

This book requires a basic understanding of the OSI architecture and some knowledge of the internal structures and protocols of certain of the layers. To assist the reader in this respect, this chapter commences with an overview of the basic OSI concepts. Applicable international standards references are given. Relevant Internet protocol standards are also identified, and their relationships to the OSI architecture noted. For readers requiring tutorial coverage of these areas, the following texts are recommended. For a complete coverage of OSI, see [BLA1, DIC1]. For a more detailed coverage of the upper layers, see [HEN1]. For a more implementation-oriented perspective on OSI, see [ROS1]. For coverage of the Internet protocol suite, see [COM1].

This chapter also describes the issues underlying mapping of security services to architectural layers, and presents ground rules for making such mapping decisions. A four-level security architectural model is introduced as a simpler, more pragmatic, substitute for the OSI model when focusing on security placement issues. This four-level model is used throughout the book when discussing layer placement of security services.

The chapter is organized into sections addressing:

(1) general principles of protocol layering and associated terminology, as introduced in the OSI Basic Reference Model;
(2) the structures, services, and protocols of specific OSI layers;
(3) the Internet TCP/IP protocol suite and its relationship to the OSI architecture;
(4) the architectural placement of security services, including the four-level model; and
(5) the way management of security services relates to architectural layers.

3.1 Protocol Layering — Principles and Terminology

In the real world, communication occurs between *real systems*. For the purposes of defining protocols, the OSI standards introduce the concept of a *model* of a real system, known as an *open system*.[7] The model system is considered to be structured in layers. This does not necessarily require implementations of real systems to be structured in the same way — they can be constructed any way the implementor chooses, provided the resultant behavior of the complete implementation conforms to the behavior defined using the model. For example, an implementation may merge the functions of multiple adjacent layers into one piece of software, without interlayer boundaries needing to be apparent.

History

The OSI initiative was launched in 1977 by ISO Technical Committee TC97 (Information Processing Systems).[8] Subcommittee TC97/SC16 (Open Systems Interconnection) was established, with the goal of developing a model and identifying protocol standards to support the needs of an unlimited range of applications over multiple underlying communications media technologies. The project captured the attention of the International Telecommunication Union (ITU), which develops Recommendations used internationally by telecommunications carriers (until April 1993, these were called *CCITT* Recommendations). A collaborative arrangement was set up between ISO and the ITU to work

[7] Note that ISO/IEC 7498-1 applies this specific interpretation to the term *open system*, as distinct from the term's common usage.

[8] The organizations and committees mentioned here are described in more detail in Appendix A.

together to produce aligned ISO International Standards and ITU Recommendations on OSI.

The first significant output of this activity was the OSI Basic Reference Model. This was published in 1984 by ISO as international standard 7498 and by the ITU as Recommendation X.200. This document describes a seven-layer architecture, to be used as the basis for independently specifying individual layer protocols. Standards for the first protocols were published soon after the Basic Reference Model, and further standards have appeared in a steady stream since.

Layering Principles

The OSI model lays out certain principles for constructing communications protocols out of multiple layers. Some important concepts are illustrated in Figure 3-1.

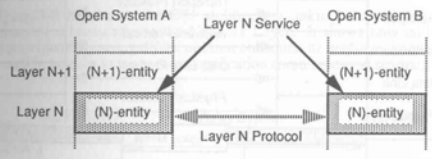


Figure 3-1: OSI Layering Concepts

Consider some middle layer, say layer *N*. Above it is layer *N*+1 and below it is layer *N*-1. In both open systems there is functionality supporting layer *N*. This is denoted the *(N)-entity* in each open system. The pair of communicating (*N*)-entities provide a *service* to (*N*+1)-entities in their respective systems. This service includes carrying data for the (*N*+1)-entities.

The pair of (*N*)-entities communicate with each other. They use an *(N)-protocol*, which embraces the syntax (format) and semantics (meaning) of data exchanged between them, plus rules for procedures to be followed. The (*N*)-protocol is conveyed by making use of a service provided by (*N*-1)-entities. Each message sent in an (*N*)-protocol is known as an *(N)-protocol-data-unit* (PDU).

An important principle underlying this layering concept is *layer independence*. The intention is that an (*N*)-layer service can be defined and can then be used in defining protocols for the (*N*+1)-layer, without any knowledge of the (*N*)-protocol used in providing that service.

The Seven OSI Layers

The OSI reference model defines seven layers, as shown in Figure 3-2. Protocols from each of the layers are grouped together into what is known as an OSI layer stack. An OSI layer stack fulfills the communication needs of an *application-process*, which is the part of a real system which performs information processing for a given application purpose.

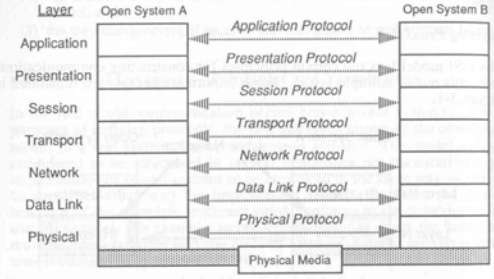


Figure 3-2: The Seven OSI Layers

The layers and their main functions are as follows:

- The *Application Layer* (Layer 7) provides a means for the application-process to access the OSI environment. Application Layer protocol standards deal with communications functions which apply to one particular application or a family of applications.
- The *Presentation Layer* (Layer 6) provides for the representation of information that application layer entities use or refer to in their communication.
- The *Session Layer* (Layer 5) provides the means for higher-layer entities to organize and synchronize their dialogue and to manage their data exchange.

- The *Transport Layer* (Layer 4) provides transparent transfer of data between higher-layer entities and relieves them from any concern with the detailed way in which reliable and cost-effective data transfer is achieved.
- The *Network Layer* (Layer 3) provides transmission between higher-layer entities, with independence from routing and relay considerations. This includes the case where multiple subnetworks are used in tandem or in parallel. It makes invisible to higher layers how underlying communications resources (i.e., data links) are used.
- The *Data Link Layer* (Layer 2) provides for transferring data on a point-to-point basis, and for establishing, maintaining, and releasing point-to-point connections. It detects and possibly corrects errors which may occur in the underlying Physical Layer.
- The *Physical Layer* (Layer 1) provides mechanical, electrical, functional, and procedural means to activate, maintain, and deactivate physical connections for bit transmission between data-link entities.

Figure 3-3 illustrates the OSI architecture, taking into account the significance of subnetworks in the Network Layer. It shows how multiple subnetworks (possibly using different interconnection or media technologies) may be used in tandem to support one application communication session.

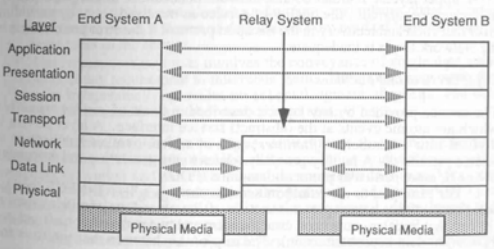


Figure 3-3: OSI Layering with Multiple Subnetworks

Upper Layers and Lower Layers

From a practical perspective, the OSI layers can be considered in terms of:

(a) application-dependent protocols;
(b) protocols associated with particular media; and
(c) a bridging function between (a) and (b).

The application-dependent protocols comprise the Application, Presentation, and Session Layers — the *upper layers*. Realizations of these layers are closely tied to the application being supported, and are entirely independent of the underlying communications technology or technologies.

The remaining layers, which span categories (b) and (c) above, are known as the *lower layers*. The media technology-dependent protocols are in the Physical and Data Link layers, and the lower (*subnetwork-dependent*) sublayers of the Network Layer.

The bridging function is provided by the Transport Layer and the upper sublayers of the Network Layer. The upper sublayers of the Network Layer enable a consistent network-service interface to be presented to the layer above, although the *quality-of-service* will be variable, dependent on the subnetwork(s) used. The Transport Layer makes the layers under it transparent to the upper layers. It either obtains network-connections of adequate quality-of-service, or upgrades the quality-of-service as necessary, e.g., by providing error detection and recovery in the transport protocol if the error performance of the Network Layer is inadequate.

Layer Services and Facilities

The service provided by any layer is described in terms of *service primitives*, which are atomic events at the (abstract) service interface. A layer service is divided into a number of *facilities*, each of which involves a collection of related primitives. A facility generally relates to the generation and processing of one or more particular protocol-data-units (PDUs).

For example, in the transport service, there is a T-CONNECT facility, which establishes a transport-connection. It involves four primitives (two at the end that initiates connection establishment and two at the other end) and two PDUs (one sent in each direction). The relationship between the primitives and PDUs is illustrated in Figure 3-4 as a time-sequence diagram.

The above type of sequence, involving two PDUs and four primitives, is very common, and is known as a *confirmed service*. Another common case, known as an *unconfirmed service*, involves only one PDU and two primitives. It is basically the same as the first half of the sequence shown in Figure 3-4.

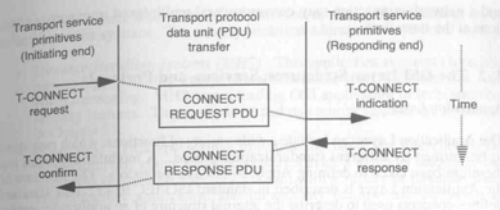


Figure 3-4: Layer Service Facility and Primitives

Connection-Oriented and Connectionless Services

There are two distinct modes of service at any layer:

- The *connection-oriented* mode is based on (*N*)-connections provided by the (*N*)-layer. A connection is an association between two (*N*)-entities, having an establishment phase, transfer phase, and release phase. During the transfer phase, a stream of data units is conveyed on behalf of higher-layer users of the service.
- The *connectionless* mode involves the conveyance of single data units, without any requirement to interrelate them. The service may route data units independently, provides no acknowledgment of receipt, and does not guarantee delivery in the same order as sent.

The main reason for the two types of service is that some underlying communications technologies are inherently connection-oriented (e.g., packet switching networks) and some are inherently connectionless (e.g., local area networks). The bridging function at the Network and Transport Layers includes supporting the operation of connection-oriented upper layers over connectionless communications technologies.

With connection-oriented upper layers, the connections at the individual layers map directly to each other. An application-association (which is the Application Layer's equivalent of a connection) maps directly to a presentation-connection, which maps directly to a session-connection. However, at layers below this, there is not necessarily this one-to-one mapping. For example, a transport-connection can be serially reused for multiple session-connections,

and a network-connection may convey several multiplexed transport-connections at the same time.

3.2 The OSI Layer Structures, Services, and Protocols

Application Layer

The Application Layer can include a wide variety of functions, which may need to be defined by different standardization groups. A modular approach has therefore been taken to defining Application Layer protocols. The structure of the Application Layer is described in standard ISO/IEC 9545. This standard defines concepts used to describe the internal structure of an application-entity, plus concepts used to describe active relationships between invocations of application-entities.

The most basic building block of an application-entity is called an application-service-element (ASE). (An ASE can be thought of as a piece of protocol specification which can be conveniently specified in one standard document.) The more general structuring component of an application-entity is an application-service-object (ASO), which is built up from ASEs and/or other ASOs. The structuring principles relating to application-entities, ASEs, and ASOs are discussed further in Chapter 12.

Two important concepts describing relationships between communicating application-entities are:

- *Application-association*: A cooperative relationship between two ASO-invocations which governs their bilateral use of the presentation service for communication purposes. This is the Application Layer's equivalent of a *connection*. It can also be considered the Application Layer view of a presentation-connection.
- *Application-context*: A set of rules shared by two ASO-invocations in order to support an application-association. This is effectively the complete Application Layer protocol in use on an application-association.

One ASE of special significance is the *Association Control Service Element* (ACSE). This ASE supports establishment and termination of application-associations, and it is required in all application-contexts. A practical view of ACSE is that it defines the Application Layer information conveyed in the protocol exchanges for establishing and terminating presentation-connnections and session-connections. The ACSE service is defined in standard ISO/IEC 8649, and the protocol in ISO/IEC 8650.

Several standard OSI-based applications have been defined. The standards include specifications of Application Layer protocols, plus supporting material such as definitions of information models and procedures to be followed within systems. The main applications addressed in this book are:

- *Message Handling Systems (MHS)*: This application supports electronic messaging, including interpersonal electronic mail, EDI transfer, and voice messaging. MHS was a leading OSI application in incorporating security features. This application and its security features are discussed in Chapter 13.
- *Directory*: This application provides the basis for interconnecting information processing systems so as to provide a logically integrated but physically distributed directory system, with a variety of potential uses. The Directory application and its security features are discussed in Chapter 14.
- *File Transfer, Access, and Management (FTAM)*: The FTAM application provides support for reading or writing of files in a remote computer system, accessing components of such files, and/or managing (e.g., creating or deleting) such files. FTAM is defined in ISO/IEC 8571.

The OSI network management facilities also constitute an OSI application. They are discussed in Chapter 15.

The OSI Application Layer standards include an important protocol modeling and construction tool called the *Remote Operations Service Element* (ROSE). ROSE is based on a general client-server model, in which one system (the client) invokes specified operations in the other system (the server). The protocol can be expressed in terms of arguments that accompany the invocation, and results or an error indication that can be returned from the operation. For an application which fits this model, use of ROSE can greatly facilitate protocol specification. ROSE is used in the MHS, Directory, and OSI network management protocols. The ROSE model, service, and protocol are defined in the ISO/IEC 9072 multi-part standard.

Presentation Layer

The Presentation Layer deals with how application information is represented (as a bit string) for purposes of transfer. An overview of the operation of this layer is provided in Chapter 12.

The standards for the presentation service and protocol are ISO/IEC 8822 and 8823 respectively.

Another pair of Presentation Layer standards, which are particularly important, are the standards ISO/IEC 8824 and ISO/IEC 8825 dealing with Abstract Syntax Notation One (ASN.1). ASN.1 is used heavily by OSI and non-OSI applications alike, for defining application layer information items and corresponding bit-string encodings for them. A brief introduction to ASN.1

is given in Appendix B. Readers unfamiliar with ASN.1 should read this appendix before starting Part II of this book. For a more detailed ASN.1 tutorial, see [STE1].

Session Layer

The Session Layer performs such functions as dialogue management and resynchronization, under the direct control of the Application Layer. Dialogue management supports full-duplex and half-duplex operational modes for applications. Resynchronization supports the insertion of synchronization marks in a data stream and the resynchronization back to an earlier mark in the event of an error condition. The standards for the session service and protocol are ISO/IEC 8326 and 8327, respectively.

Later discussion on security architecture will conclude that the Session Layer has virtually no role to play in providing security, hence readers unfamiliar with this layer can safely ignore it.

Transport Layer

The transport service is defined in ISO/IEC 8072. It supports end-system to end-system transparent transfer of data. It provides its (upper layers) users with independence from underlying communications technologies and with the ability to nominate a *quality-of-service* (in terms of such characteristics as throughput, residual error rate, and failure probability). If the quality-of-service of the available underlying network services is inadequate, the Transport Layer will upgrade the quality-of-service to the requisite level by adding value (e.g., error detection/recovery) in its own protocol. The transport service has both connection-oriented and connectionless variations.

The Transport Layer protocols to support the connection-oriented service are specified in ISO/IEC 8073. There are five different *classes* of protocol, as follows:

- *Class 0* adds no value to the network service.
- *Class 1* supports error recovery, upon detection of an error by the Network Layer.
- *Class 2* supports multiplexing of transport-connections on one network-connection.
- *Class 3* provides error recovery and multiplexing.
- *Class 4* provides error detection (checksum), error recovery, and multiplexing.

Using its error recovery features, the Class 4 protocol can operate over a connectionless network service, to provide a connection-oriented transport service.

The protocol to support the connectionless transport service is specified in ISO/IEC 8602.

Network Layer

The Network Layer is one of the more complex OSI layers, because it needs to accommodate a variety of subnetwork technologies and interconnection strategies. It has to deal with the problems of relaying between subnetworks of different technologies, and it has to deal with the problem of presenting a common service interface to the Transport Layer above. The existence of both connection-oriented and connectionless styles of operation contributes significantly to the complexities of the Network Layer standards.

The standards that best explain the working of the Network Layer are ISO/IEC 8880, ISO/IEC 8648, and ISO/IEC 8348. Figure 3-5 illustrates the relationships between these standards.

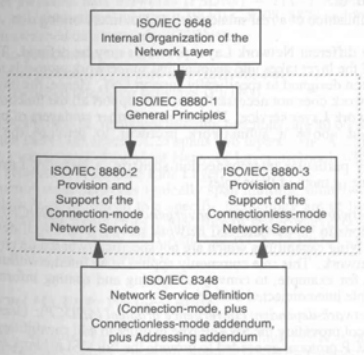


Figure 3-5: General Standards for the Network Layer

ISO/IEC 8648 introduces some important terms and concepts and describes how the OSI modeling concepts in this layer map to *real* network components. The concept of an *end system* (introduced in the OSI reference model) models a piece of equipment, or a collection of pieces of equipment, which implement a full seven-layer stack. The concept of an *intermediate system* is introduced in the Network Layer. An intermediate system performs only functions allocated to the lowest three OSI layers. An end system can communicate with another end system either directly or through one or a series of intermediate systems.

A *real subnetwork* is a collection of equipment and physical links used to interconnect other real systems, e.g., a public packet-switching network, a local area network (LAN), or an interconnected set of other real subnetworks. An interworking unit is a piece of equipment (or part of a piece of equipment) which realizes a network relay function. The term *intermediate system* can refer to the abstraction of any of:

(a) a real subnetwork;
(b) an interworking unit, connecting two or more real subnetworks (e.g., a router); or
(c) a combination of a real subnetwork and an interworking unit.

Many different Network Layer protocols may be defined. The internal structure of the layer takes into account that subnetwork protocols may or may not have been designed to specifically support OSI. Hence, the basic protocol of a subnetwork does not necessarily have to support all the functions required by the Network Layer service. If necessary, further sublayers of protocol can be provided above a subnetwork protocol to provide the necessary functionality.

In any particular interconnection scenario, a Network Layer protocol performs one or more of three *roles*:

- *Subnetwork-independent convergence protocol (SNICP)*: Provides functions to support the OSI network service over a well-defined set of underlying capabilities which are not specifically based on any particular subnetwork. This role commonly applies to an internetworking protocol used, for example, to convey addressing and routing information over multiple interconnected networks.
- *Subnetwork-dependent convergence protocol (SNDCP)*: Operates over a protocol providing the SNAcP role in order to add capabilities needed by a SNICP protocol or needed to provide the full OSI network service.
- *Subnetwork access protocol (SNAcP)*: A protocol which is inherently part of a particular type of subnetwork. It provides a subnetwork service

at its end points, which may or may not be equivalent to the OSI network service.

One of the more important Network Layer protocols is the Connectionless Network Protocol (CLNP) defined in ISO/IEC 8473. This protocol is generally used in a SNICP role, for providing the connectionless-mode network service. The ISO/IEC 8473 standard also defines how this protocol can operate over X.25 packet switched subnetworks or LAN subnetworks.

Subnetwork Technology Functions

OSI is designed to operate over a virtually unlimited range of underlying subnetwork technologies. These technologies have their own subnetwork-dependent (SNAcP and SNDCP role) Network Layer protocols, and Data Link Layer and Physical Layer protocols. Many standards have been developed for specific subnetwork technologies, including:

- local area networks (LANs) — the ISO/IEC 8802 series;
- packet switched data networks (PSDNs) — ITU-T Recommendation X.25, and international standards ISO/IEC 8208, 8878 and 8881;
- circuit switched data networks (CSDNs);
- point-to-point subnetworks;
- integrated services digital networks (ISDNs); and
- public switched telephone networks (PSTNs).

The LAN protocols, including bridging functions, are all considered to be located in the Data Link layer. X.25 spans two layers. The X.25 packet-level protocol is a subnetwork-dependent Network Layer protocol, while the X.25 link access protocol is in the Data Link Layer.

As open-system networks typically span multiple subnetwork technologies, security features linked to a specific technology are of limited value. This part of the OSI architecture is therefore of lesser relevance to this book than higher layers. Security for LANs and for X.25 packet switched data networks is addressed in Chapter 11.

3.3 Internet (TCP/IP) Protocol Suite

The Internet protocols have evolved since the mid-1970s, when the U.S. Defense Advanced Projects Research Agency (DARPA) started funding the development of packet switched network facilities for interconnecting universities and government research institutions across the United States. A

complete suite of protocols has been specified, spanning virtually the same range of functionality as the OSI reference model. The protocol suite is often known as the TCP/IP suite, named after two of the most important constituent protocols. These protocols are being rapidly deployed in many networks worldwide, especially the collection of interconnected networks known as the DARPA Internet.

The Internet protocol suite is sometimes seen as a head-to-head competitor with the OSI protocol suite. However, it is becoming increasingly apparent that both sets of protocols have their strengths and weaknesses, and a great deal can be gained by mixing member protocols of both families to provide complete networking solutions. Protocol layering makes this possible.

The Internet protocol suite can be modeled using the same layering approach as the OSI architecture and, although not all seven layers are apparent in the Internet suite, the protocols map readily to the OSI model. There are effectively four Internet layers. For the purposes of this book, we shall call them:

- *Application Layer*: This layer combines the functionalities of the OSI Application, Presentation, and Session Layers, i.e., the OSI *upper layers* as defined in Section 3.2.
- *Transport Layer*: This layer is functionally equivalent to the OSI Transport Layer.
- *Internet Layer*: This layer is functionally equivalent to the subnetwork-independent part of the OSI Network Layer. (Unless further qualified, the term *Network Layer* used in the remainder of this book should be taken to include the Internet Layer.)
- *Interface Layer*: This layer is functionally equivalent to the OSI subnetwork technology functions discussed in Section 3.2.

Assuming this mapping, it is possible to consider security architecture in terms equally applicable to the OSI and Internet suites. The upper layers architectural differences prove inconsequential because, from the security perspective, there is no need to separate out the OSI upper layers into Application, Presentation, and Session Layers. Similarly, in the lower layers, there is no need to separate out the subnetwork technology functions into constituent layers.

Application Layer Protocols

There are numerous Internet Application Layer protocols, some important ones being:

- *File Transfer Protocol (FTP)*: A protocol which allows users to log into a remote system, identify themselves, list remote directories, and copy files to or from the remote machine.
- *Simple Mail Transfer Protocol (SMTP)*: An electronic mail protocol based on [POS1, CRO1]. Internet electronic mail and related security features are discussed in Chapter 13.
- *Simple Network Management Protocol (SNMP)*: A protocol supporting network management. SNMP and related security features are discussed in Chapter 15.
- *TELNET*: A simple remote terminal protocol which allows a user at one site to establish a connection to a login server at another site, passing keystrokes and responses between them.

Transport and Network Layer Protocols

There are two primary Internet Transport Layer protocols:

- *Transmission Control Protocol (TCP)*: A connection-oriented transport protocol designed for operation over a connectionless network service [POS2]. This protocol is comparable to the Class 4 OSI transport protocol.
- *User Datagram Protocol (UDP)*: A connectionless transport protocol [POS3]. This protocol is comparable to the OSI connectionless transport protocol.

The primary Internet Network Layer protocol is the *Internet Protocol* (IP), which is a connectionless network protocol [POS4]. This protocol is comparable to the OSI connectionless network protocol (CLNP).

3.4 Architectural Placement of Security Services

The provision of security services in a layered communications architecture raises some significant issues. Protocol layering results in data items being embedded within data items and connections being carried within connections, potentially with multiple layers of nesting. Hence, there are major decisions to be made as to the layer(s) at which data-item or connection-based protection should be provided.

The first formal standard addressing layer assignment of security services was the OSI Security Architecture (ISO/IEC 7498-2), published in 1988. This standard (which is discussed in Chapter 9) provides guidance as to which OSI layers are appropriate for providing the various security services. However, it

does not provide all the answers, leaving many options open. Some services may need to be provided in different layers in different application scenarios; some may even need to be provided in multiple layers in the same scenario. One reason for the apparent inclusiveness of ISO/IEC 7498-2 is the approach taken of trying to assign fourteen security services to seven architectural layers. This can be crystallized into a much simpler and more pragmatic four-level model, based on the *real* security implications in *real* networks.

Figure 3-6 illustrates how a pair of end systems communicate with each other via a cascaded series of subnetworks. An end system is usually one piece of equipment, anywhere in the range from personal computer to workstation to

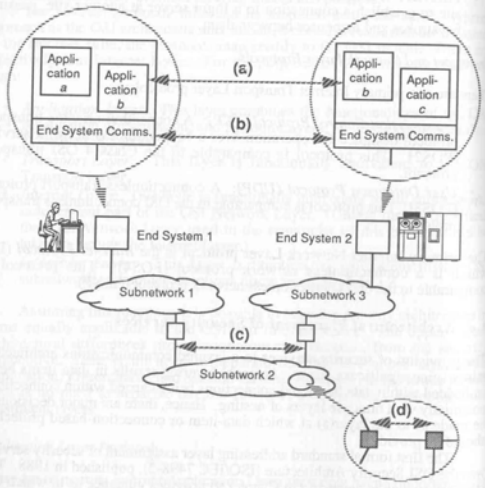


Figure 3-6: Basic Communications Architecture

minicomputer to mainframe computer. One characteristic that can be reasonably assumed for an end system is that, for security purposes, it has one policy authority.

A subnetwork is a collection of communications facilities employing the same communications technology, e.g., a particular local area network (LAN) or wide area network (WAN). It is also reasonable to assume that, for security purposes, any subnetwork has one policy authority. However, distinct subnetworks will often have different security environments and/or different policy authorities. An end system and the subnetwork to which it connects may or may not have the same policy authority. A typical scenario is an end system connecting to a LAN on one corporation's premises, with the LAN having a gateway to a public WAN. After possibly traversing multiple separately administered WANs, communications pass via another LAN to the other end system.

Another aspect introduced in Figure 3-6 is that of an end system simultaneously supporting multiple applications, such as electronic mail, directory access, and file transfer, for one or more users simultaneously. Another simultaneous application may be network management services for the system administrator. The security requirements of these applications often differ considerably.

We also need to recognize that security requirements may vary *within* a subnetwork. Subnetworks generally comprise multiple links connecting multiple subnetwork components and different links may pass through different security environments. Therefore, protection of individual links may need to be accommodated.

Figure 3-6 indicates four levels at which distinct requirements for security protocol elements arise:

(a) *Application level:* Security protocol elements that are application-dependent.
(b) *End-system level:* Security protocol elements providing protection on an end-system to end-system basis.
(c) *Subnetwork level:* Security protocol elements providing protection over a subnetwork which is considered less trusted than other parts of the network environment.
(d) *Direct-link level:* Security protocol elements providing protection internal to a subnetwork, over a link which is considered less trusted than other parts of the subnetwork environment.

From the communications protocol perspective, these four levels are all that need to be distinguished. An approximate mapping of these levels to the OSI architectural layers is illustrated in Figure 3-7.

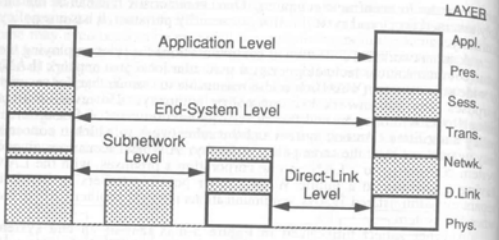


Figure 3-7: The Four Basic Architectural Levels for Security

What are the ramifications of locating security services at higher as opposed to lower levels? Before discussing each level individually, we can identify some general properties that vary between higher and lower levels:

- *Traffic mixing*: As a result of multiplexing, there is a greater tendency at lower levels to have data items from different source/destination users and/or applications mixed in a data stream than at higher levels. The significance of this factor varies with the type of security policy. If the security policy tends to leave it to individual users and/or applications to specify the protection required for their data, placing security services at a higher level tends to be better. With security at lower levels, the individual users/applications have inadequate control, and there is likely to be a cost in unnecessarily protecting some data because of the security requirements of other data sharing the data stream. On the other hand, if the security policy is such that an organization wants to ensure that all organizational traffic is protected to a certain degree, regardless of user or application, this is more easily achieved with security services at lower levels.
- *Route knowledge*: At lower levels, there tends to be more knowledge of the security characteristics of different routes and links. In an environment where such characteristics vary significantly, placing security at lower levels can have effectiveness and efficiency benefits. Appropriate security services can be selected on a subnetwork or direct-link basis, while eliminating security costs entirely on subnetworks or links where protection is unnecessary.

- *Number of protection points*: Placing security at a very high level (i.e., application level) requires security to be implemented in every sensitive application in every end system. Placing security at a very low level (i.e., direct-link level) requires security to be implemented at the ends of every network link. Placing security closer to the middle of the architecture (i.e., end-system or subnetwork level) will tend to require security features to be installed at significantly fewer points, which may significantly reduce costs.
- *Protocol header protection*: Security protection at higher levels cannot protect protocol headers of lower levels which, in at least some environments, may be sensitive. This tends to encourage placing security services at a low level.
- *Source/sink binding*: Some security services, such as data origin authentication and non-repudiation, depend upon associating data with its source or sink. This is most effectively achieved at higher levels, especially the application level. However, it can sometimes be achieved at lower levels, subject to special constraints, e.g., binding of a message originator to a particular end system through the use of trusted hardware and/or software.

Taking all the above considerations together, it becomes apparent why there is no simple answer to the question of the "best" architectural placement for security services. In the following subsections, the characteristics of each of the four levels are discussed further, in a service-independent way. Subsequent chapters discuss architectural placement of specific security services with reference to these four levels.

Application Level Security

In terms of the OSI architecture, application level security relates to the upper layers. (In OSI protocol terms, this means the Application Layer, possibly assisted by Presentation Layer facilities; the Session Layer does not contribute to the provision of security. The division of functions between Application and Presentation Layers is addressed in Chapter 12.)

For most security services, it is *possible* to locate the service at the application level. In many cases, lower-level alternatives are also available and frequently have advantages (e.g., lower equipment or operational costs). However, there are two situations in which the application level is the *only* viable level for locating a security service:

(a) Where security services are application-specific, either semantically or by virtue of being built into a particular application protocol.

(b) Where security services traverse application relays.

Some security requirements are inextricably linked to application semantics. For example, a file transfer application may need to deal with file access control, e.g., reading or updating access control lists attached to files. In other cases, the granularity of security protection is reflected in application protocol fields. This is very common with selective field confidentiality, selective field integrity, and non-repudiation services. Examples are giving confidentiality protection to a PIN field in a financial transaction, or digitally signing individual retrieval requests in a directory protocol. In all of these cases, the security services must be located at the application level, as layer independence prevents lower layers from having knowledge of the requisite semantics or the protocol boundaries.

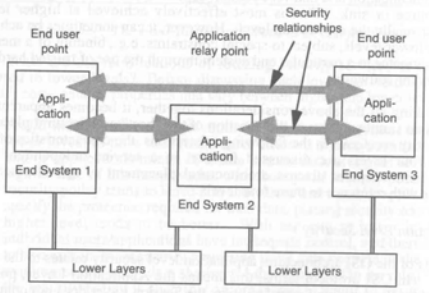


Figure 3-8: Application Relay Scenario

The other situation demanding an application level solution is an application relay scenario. Some applications inherently involve more than two end systems, as illustrated in Figure 3-8. Electronic mail systems are an example. A message originating at one end system may pass through multiple relay systems before arriving at a recipient in another end system. There may be a requirement to protect the *content* part of the message on an end-user to end-user basis, i.e., the keying relationship rests with the end-user systems, and the relay systems do not know the key(s) used. However, other parts of

the message, e.g., address fields and trace fields, are not protected this way, as the relay systems need to use, and possibly update, such fields. In such a scenario, all security services in the end-user to end-user security relationship need to be at the application level.

When deciding if a security requirement should be dealt with at the application level, or at a lower level, the above factors should be considered first. If none apply, consideration should be given to using lower levels.

End-System Level Security

The following types of security requirements point to an end-system level solution:

- requirements based on the assumption that the end systems are trusted, but that all underlying communications network(s) are untrusted;
- requirements dictated by the end-system authority, that are to be enforced upon all communications regardless of the application; and
- requirements relating to network connections (or all traffic thereon), that are not linked to any particular application, e.g., confidentiality and/or integrity protection of all traffic on a connection.

Some services, such as confidentiality and/or integrity protection of all user information on an end-system to end-system basis, could potentially be provided at either the application level or the end-system level. In deciding which level, there are several factors to be taken into account. Factors tending to favor an end-system level solution over a comparable application level solution include:

- the ability to make protection services transparent to the applications;
- superior performance of bulk data protection services, owing to the ability to operate on larger data units and to process the data of multiple applications in a common way;
- placing management of the security facilities in the hands of one end-system administrator, rather than in separate applications (supports a consistent security policy); and
- ensuring that the protocol headers of mid-layer protocols (i.e., Transport, Session, and Presentation Layer protocols) receive protection.

In OSI terms, end-system level security relates to either Transport Layer or subnetwork-independent Network Layer protocols. The decision between these two options has been the subject of ongoing debate in the standards forums for several years. There has been no real resolution to this debate, in

that standards have been developed supporting both options (standards ISO/IEC 10736 and ISO/IEC 11577 respectively).

Factors favoring the Transport Layer approach include:

- extension of the protection right to the end system, thereby protecting against vulnerabilities in local-access or front-end communications facilities; and
- the possibility of providing different grades of protection to different transport-connections multiplexed on a network-connection.

Factors favoring the Network Layer approach include:

- the ability to use the same solution at the end-system and subnetwork levels;
- the ease of transparently inserting security devices at standardized physical interface points, e.g., X.25 or LAN interfaces; and
- the ability to support *any* upper-layer architecture, including OSI, Internet, and proprietary architectures.

The irreconcilability of such factors explains why there is no simple resolution of this issue. Profiling groups and user communities need to make their own decisions, based on their own requirements.

Subnetwork Level Security

The difference between end-system level security and subnetwork level security is that the latter provides protection across one or more specific subnetworks only. There are two very important reasons for distinguishing this level from the end-system level:

- It is very common that subnetworks close to end-systems are trusted to the same extent as the end-systems themselves, since they are on the same premises and administered under the same authorities.
- In any network, the number of end systems usually far exceeds the number of subnetwork gateways. Therefore, equipment costs and operational costs for subnetwork level security solutions may be much lower than those for end-system level solutions.

Subnetwork level security should therefore always be considered as a possible alternative to end-system level security.

In OSI, subnetwork level security maps to the Network Layer or, in the case of LANs, to the Data Link Layer (where LAN protocols are located).

Direct-Link Level Security

The appropriate situations for using direct-link level security are those with comparatively few untrusted links in an otherwise-trusted environment. On a given link, a high level of protection can be provided at a low equipment cost. Provision of security at this level can be transparent to all higher communications layers, including network-protocols, hence it is not tied to any particular network architecture (e.g., OSI, TCP/IP, or proprietary). Security devices can be easily inserted at common standardized physical interface points. However, operational costs may be high, because of the need to independently manage devices on a link-by-link basis. It is important to realize that direct-link level security cannot protect against vulnerabilities *inside* subnetwork nodes, e.g., hubs, bridges, or packet switches.

In terms of OSI layers, direct-link level security usually relates to the Physical Layer. Protection is provided at the bit-stream level and is transparent to all higher-level protocols. For example, encryption processes can be applied to the bit-streams passing at any interface point. Protection transmission technologies, such as spread-spectrum or frequency hopping techniques, can also be employed [TOR1]. Direct-link level security can potentially relate to the Data Link Layer, e.g., if protection is provided at the frame level.

Human User Interactions

Some network security services involve direct interaction with a human user. Such interactions do not fit cleanly into any of the architectural options presented above. The most important case is *personal authentication*. The user is external to the communication facilities, i.e., beyond one of the end systems. Communications supporting personal authentication are either local (i.e., between the human user and his local end system) or they involve protocol elements at the application level, or they combine both of the above. Examples of these three cases are:

- The human user authenticates to his or her local end system. That end system then authenticates itself to the remote end system, and advises the user identity, which the remote end system accepts as authentic.
- The human user passes authentication information (e.g., a password) to his or her local end system, which conveys it on to the remote end system which performs the user authentication.
- The human user enters a password to his or her local end system, which this system uses to obtain an authentication certificate from an on-line authentication or key server. The certificate is passed to the remote end system which uses it as the basis of authenticating the user.

Personal authentication is addressed further in Chapter 5.

Figure 3-9 illustrates the relationship between human user interaction protocols and the application level architectural option.

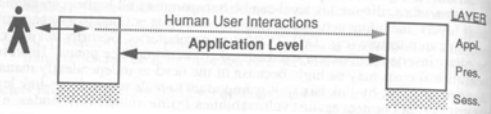


Figure 3-9: Human User Interactions

3.5 Management of Security Services

Security services need the support of management functions, such as:

- key management for cryptosytems used in providing a security service (this is discussed in Chapters 4 and 7);
- distribution of required information to decision-making points, e.g., for use in making authentication or access control decisions — this includes information enabling a positive decision to be made and notifications of the withdrawal of previously distributed information;
- central accumulation of information for purposes such as archival (for subsequent non-repudiation purposes) or security audit trail or alarm generation;
- operational functions, such as service activation and deactivation; and
- special security management functions, such as remotely invoking virus scans on network workstations or monitoring systems for illegitimate software.

Such management functions usually employ communications capabilities of the same network they are helping to protect. In this case it is essential that these management communications be protected to the maximal extent available. Any vulnerability in security management communications will translate to an equivalent or magnified vulnerability in protected communications generally.

In architectural terms, security management functions are generally provided through network applications. These can include applications

dedicated to network management (some examples are discussed in Chapter 15) or applications with other primary purposes. Exceptions to this level of placement can occur, e.g., when key management exchanges are closely linked to cryptographic processing at lower layers. The latter issue is taken up in Chapters 4 and 7.

Summary

Layered protocol architectures allow network designs to accommodate unlimited applications, unlimited underlying media technologies, and unlimited interconnection techniques. The OSI architecture provides a general model on which layering can be based. There are seven layers, which can be separated into the upper layers (Application, Presentation, and Session Layers) and lower layers (Transport, Network, Data Link, and Physical Layers). Other OSI standards define layer services and specific protocols for the seven layers. The Internet TCP/IP protocol suite defines alternative protocols, which can be mapped straightforwardly to the OSI model.

In providing security services, there are significant issues in deciding the layer(s) at which protection should be located. To assist in making these decisions, four security architectural levels can be identified — the application level, the end-system level, the subnetwork level, and the direct-link level. The application level involves security protocol elements which are application-dependent, and require support in upper layers protocols. Certain security requirements demand a solution at this level. The end-system level involves security protocol elements providing protection on an end-system to end-system basis. This can employ security protocols at either the Transport or Network Layer; both options are available, and there are various factors favoring each. The subnetwork level provides protection over particular subnetworks within the Network Layer or (in the case of LANs) the Data Link Layer. The direct-link level provides protection on a link-by-link basis over parts of a subnetwork environment; this level relates to the Physical or Data Link layer. Interactions with human users (especially for authentication purposes) do not fit fully into the above four levels, and require special consideration.

Management of security services involves various functions, most of which are provided through network management applications.

Exercises

1. When unprotected data exists within network switching equipment (such as bridges, routers, or packet switches), this equipment may need to be

physically secured to maintain adequate protection. Such physical security can be costly. To minimize this cost, at what security architectural level(s) should protection be located?

2. In an on-line financial transaction message, a PIN is to be conveyed encrypted while the other transaction details are transferred unprotected. Which of the four security architectural level(s) is (are) involved and why?

3. If sensitive information can be gleaned by monitoring the addressing information in a connection establishment exchange or in a connectionless data unit, which security architectural level(s) could be employed to ensure adequate protection?

4. A large corporation has a network which spans several sites. At the discretion of individual users, the traffic includes a substantial amount of corporate proprietary information. The corporation wants to apply blanket protection to vulnerable parts of the network to protect against disclosure of proprietary information to outsiders. In each of the following configuration scenarios, what appears the most appropriate architectural level for applying a confidentiality service, and why?

   (a) The network involves local area networks within corporate locations, with a public wide area network interconnecting these locations.

   (b) The network involves local area networks within corporate locations, with a small number of leased lines interconnecting LAN gateways at these locations.

   (c) The network involves a wide variety of communication links, trusted to varying extents, with the end user having no control over the security of the route used for any communicated item.

REFERENCES

[BLA1] U. Black, *OSI: A Model for Computer Communications*, Prentice Hall, Englewood Cliffs, NJ, 1991.

[COM1] D.E. Comer, *Internetworking with TCP/IP: Principles, Protocols, and Architecture*, Prentice Hall, Englewood Cliffs, NJ, 1988.

[CRO1] D.H. Crocker, *Standard for the Format of ARPA Internet Text Messages*, Request for Comments (RFC) 822, Internet Activities Board, 1982.

[DIC1] G. Dickson and A. Lloyd, *Open Systems Interconnection*, Prentice Hall, Englewood Cliffs, NJ, 1991.

[HEN1] J. Henshall and S. Shaw, *OSI Explained: End-to-End Computer Communication Standard*, Second Edition, Prentice Hall, Englewood Cliffs, NJ, 1990.

[POS1] J.B. Postel, *Simple Mail Transfer Protocol*, Request for Comments (RFC) 821, Internet Activities Board, 1982.

[POS2] J.B. Postel, *Transmission Control Protocol*, Request for Comments (RFC) 793, Internet Activities Board, 1981.

[POS3] J.B. Postel, *User Datagram Protocol*, Request for Comments (RFC) 768, Internet Activities Board, 1981.

[POS4] J.B. Postel, *Internet Protocol*, Request for Comments (RFC) 791, Internet Activities Board, 1981.

[ROS1] M.T. Rose, *The Open Book: A Practical Perspective on OSI*, Prentice Hall, Englewood Cliffs, NJ, 1990.

[STE1] D. Steedman, *Abstract Syntax Notation One (ASN.1): The Tutorial and Reference*, Technical Appraisals Ltd., Isleworth, England, 1990.

[TOR1] D.J. Torrieri, *Principles of Secure Communication Systems*, Second Edition, Artech House, Inc., Norwood, MA, 1992.

Standards

ISO/IEC 7498-1: *Information Technology — Open Systems Interconnection — Basic Reference Model* (Also ITU-T Recommendation X.200).

ISO/IEC 7498-2: *Information Technology — Open Systems Interconnection — Basic Reference Model — Part 2: Security Architecture* (Also ITU-T Recommendation X.800).

ISO/IEC 8072: *Information Technology — Open Systems Interconnection — Connection Oriented Transport Service Definition* (Also ITU-T Recommendation X.214).

ISO/IEC 8073: *Information Technology — Open Systems Interconnection — Connection Oriented Transport Protocol Specification* (Also ITU-T Recommendation X.224).

ISO/IEC 8208: *Information Technology — Data Communications — X.25 Packet Level Protocol for Data Terminal Equipment*.

ISO/IEC 8326: *Information Technology — Open Systems Interconnection — Basic Connection Oriented Session Service Definition* (Also ITU-T Recommendation X.215).

ISO/IEC 8327: *Information Technology — Open Systems Interconnection — Basic Connection Oriented Session Protocol Specification* (Also ITU-T Recommendation X.225).

ISO/IEC 8348: *Information Technology — Data Communications — Network Service Definition* (Also ITU-T Recommendation X.213).

ISO/IEC 8473: *Information Technology — Data Communications — Protocol for Providing the Connectionless-Mode Network Service*.

ISO/IEC 8571: *Information Technology — Open Systems Interconnection — File Transfer, Access and Management (FTAM)*.

ISO/IEC 8602: *Information Technology — Open Systems Interconnection — Protocol for Providing the Connectionless-Mode Transport Service*.

ISO/IEC 8648: *Information Technology — Data Communications — Internal Organization of the Network Layer*.

ISO/IEC 8649: *Information Technology — Open Systems Interconnection — Service Definition for Association Control* (Also ITU-T Recommendation X.217).

ISO/IEC 8650: *Information Technology — Open Systems Interconnection — Protocol Specification for the Association Control Service Element* (Also ITU-T Recommendation X.227).

ISO/IEC 8802: *Information Technology — Local and Metropolitan Area Networks*.

ISO/IEC 8822: *Information Technology — Open Systems Interconnection — Connection Oriented Presentation Service Definition* (Also ITU-T Recommendation X.216).

ISO/IEC 8823: *Information Technology — Open Systems Interconnection — Connection Oriented Presantation Protocol Specification* (Also ITU-T Recommendation X.226).

ISO/IEC 8824: *Information Technology — Open Systems Interconnection — Specification of Abstract Syntax Notation One (ASN.1)* (Also ITU-T X.680 series Recommendations).

ISO/IEC 8825: *Information Technology — Open Systems Interconnection — Specification of ASN.1 Encoding Rules* (Also ITU-T X.690 series Recommendations).

ISO/IEC 8878: *Information Technology — Data Communications — Use of X.25 to Provide the Connectionless-mode Network Service*.

ISO/IEC 8880: *Information Technology — Data Communications — Protocol Combinations to Provide and Support the OSI Network Service*.

ISO/IEC 8881: *Information Technology — Data Communications — Use of the X.25 Packet Layer Protocol in Local Area Networks*.

ISO/IEC 9072: *Information Technology — Open Systems Interconnection — Remote Operations*.

ISO/IEC 9545: *Information Technology — Open Systems Interconnection — Application Layer Structure* (Also ITU-T Recommendation X.207).

ISO/IEC 10736: *Information Technology — Telecommunications and Information Exchange Between Systems — Transport Layer Security Protocol* (Also ITU-T Recommendation X.824).

ISO/IEC 11577: *Information Technology — Telecommunications and Information Exchange Between Systems — Network Layer Security Protocol* (Also ITU-T Recommendation X.823) (Draft).

ITU-T Recommendation X.25: *Interface between data terminal equipment (DTE) and data circuit-terminating equipment (DCE) for terminals operating in the packet mode and connected to public networks by dedicated circuit* (ISO/IEC adaptation in ISO/IEC 8208).

4 Cryptographic Techniques

Cryptographic techniques, such as encipherment and digital signatures, are important building blocks in the implemention of all security services. This chapter introduces the major cryptographic techniques used in securing contemporary computer networks.

The most basic building block is called a cryptographic system (or *cryptosystem*). A cryptosystem defines a pair of data transformations. The first transformation is applied to an ordinary data item, known as *plaintext*, and generates a corresponding (unintelligible) data item called *ciphertext*. The second transformation, applied to ciphertext, results in the regeneration of the original plaintext. The two transformations are most commonly called *encryption* and *decryption*, respectively. The alternative terms *encipherment* and *decipherment* are also used, and are generally preferred in international standards.[1]

An encryption transformation uses as input both the plaintext data and an independent data value known as an *encryption key*. Similarly, a decryption transformation uses a *decryption key*. These keys are seemingly random bit-vectors.

An obvious use of a cryptosystem is for providing confidentiality. The plaintext is unprotected data. The corresponding ciphertext may be transmitted in untrusted environments because, if the cryptosystem is a good one, it will be infeasible for anyone to deduce the plaintext from the ciphertext without knowing the decryption key. Cryptosystems also have uses other than confidentiality, as will become apparent later in the chapter.

There are two basic types of cryptosystems — *symmetric* systems (sometimes called private-key or secret-key systems) and *public-key* (or asymmetric) systems. These have distinct characteristics and are used in different ways to provide security services.

This chapter is organized into sections addressing:

(1) symmetric cryptosystems;
(2) public-key cryptosystems;
(3) integrity check-values or seals (also known as message authentication codes);

[1] This is because "encryption" and "decryption" become confused with more traditional interpretations of "burying" and "digging up" in some languages.