Global edition Network Security Essentials Applications and Standards sixth edition William Stallings Network Security Essentials: Applications and Standards Sixth Edition Global Edition William Stallings Harlow, England • London • New York • Boston • San Francisco • Toronto • Sydney • Dubai • Singapore • Hong Kong Tokyo • Seoul • Taipei • New Delhi • Cape Town • Sao Paulo • Mexico City • Madrid • Amsterdam • Munich • Paris • Milan Vice President and Editorial Director, ECS: Marcia J. Horton Executive Editor: Tracy Johnson (Dunkelberger) Editorial Assistant: Kristy Alaura Program Manager: Carole Snyder Project Manager: Robert Engelhardt Media Team Lead: Steve Wright Acquisitions Editor, Global Edition: Sourabh Maheshwari Assistant Project Editor, Global Edition: Shaoni Mukherjee Manager, Media Production, Global Edition: Vikram Kumar Senior Manufacturing Controller, Production, Global Edition: Trudy Kimber R&P Manager: Rachel Youdelman R&P Senior Project Manager: William Opaluch Senior Operations Specialist: Maura Zaldivar-Garcia Inventory Manager: Meredith Maresca Marketing Manager: Demetrius Hall Product Marketing Manager: Bram Van Kempen Marketing Assistant: Jon Bryant Cover Designer: Marta Samsel Cover Art: Africa Studio Full-Service Project Management: Chandrasekar Subramanian, SPi Global Credits and acknowledgments borrowed from other sources and reproduced, with permission, in this textbook appears on page 448. Pearson Education Limited Edinburgh Gate Harlow Essex CM20 2JE England and Associated Companies throughout the world Visit us on the World Wide Web at: www.pearsonglobaleditions.com ISBN 10: 1-292-15485-3 ISBN 13: 978-1-292-15485-5 Typeset by SPi Global Printed and bound in Malaysia. British Library Cataloguing-in-Publication Data A catalogue record for this book is available from the British Library Authorized adaptation from the United States edition, entitled Network Security Essentials: Applications and Standards, 6th Edition, ISBN 978-0-134-52733-8, by William Stallings published by Pearson Education © 2017. 10 9 8 7 6 5 4 3 2 1 For Tricia never dull never boring the smartest and bravest person I know This page intentionally left blank Contents

 Preface In this age of universal electronic connectivity, of viruses and hackers, of electronic eavesdropping and electronic fraud, there is indeed no time at which security does not matter. Two trends have come together to make the topic of this book of vital interest. First, the explosive growth in computer systems and their interconnections via networks has increased the dependence of both organizations and individuals on the information stored and communicated using these systems. This, in turn, has led to a heightened awareness of the need to protect data and resources from disclosure, to guarantee the authenticity of data and messages, and to protect systems from network-based attacks. Second, the disciplines of cryptography and network security have matured, leading to the development of practical, readily available applications to enforce network security. What's New In The SIXTH Edition In the four years since the fifth edition of this book was published, the field has seen continued innovations and improvements. In this new edition, I try to capture these changes while maintaining a broad and comprehensive coverage of the entire field. To begin this process of revision, the fifth edition of this book was extensively reviewed by a number of professors who teach the subject and by professionals working in the field. The result is that, in many places, the narrative has been clarified and tightened, and illustrations have been improved. Beyond these refinements to improve pedagogy and user-friendliness, there have been substantive changes throughout the book. Roughly the same chapter organization has been retained, but much of the material has been revised and new material has been added. The most noteworthy changes are as follows: ■■ Fundamental security design principles: Chapter 1 includes a new section discussing the security design principles listed as fundamental by the National Centers of Academic Excellence in Information Assurance/Cyber Defense, which is jointly sponsored by the U.S. National Security Agency and the U.S. Department of Homeland Security. ■■ Attack surfaces and attack trees: Chapter 1 includes a new section describing these two concepts, which are useful in evaluating and classifying security threats. ■■ Practical use of RSA: Chapter 3 expands the discussion of RSA encryption and RSA digital signatures to show how padding and other techniques are used to provide practical security using RSA. ■■ User authentication model: Chapter 4 includes a new description of a general model for user authentication, which helps to unify the discussion of the various approaches to user authentication. ■■ Cloud security: The material on cloud security in Chapter 5 has been updated and expanded to reflect its importance and recent developments. ■■ Transport Layer Security (TLS): The treatment of TLS in Chapter 6 has been updated, reorganized to improve clarity, and now includes a discussion of the new TLS version 1.3. 10 Preface 11 ■■ E-mail Security: Chapter 8 has been completely rewritten to provide a comprehensive and up-to-date discussion of e-mail security. It includes: — New: discussion of e-mail threats and a comprehensive approach to e-mail security. — New: discussion of STARTTLS, which provides confidentiality and authentication for SMTP. — Revised: treatment of S/MIME has been substantially expanded and updated to reflect the latest version 3.2. — New: discussion of DNSSEC and its role in supporting e-mail security. — New: discussion of DNS-based Authentication of Named Entities (DANE) and the use of this approach to enhance security for certificate use in SMTP and S/MIME. — New: discussion of Sender Policy Framework (SPF), which is the standardized way for a sending domain to identify and assert the mail senders for a given domain. — Revised: discussion of DomainKeys Identified Mail (DKIM) has been revised. — New: discussion of Domain-based Message Authentication, Reporting, and Conformance (DMARC), allows e-mail senders to specify policy on how their mail should be handled, the types of reports that receivers can send back, and the frequency those reports should be sent. Objectives It is the purpose of this book to provide a practical survey of network security applications and standards. The emphasis is on applications that are widely used on the Internet and for corporate networks, and on standards (especially Internet standards) that have been widely deployed. Support Of ACM/IEEE Computer Science Curricula 2013 The book is intended for both academic and professional audiences. As a textbook, it is intended as a one-semester undergraduate course in cryptography and network security for computer science, computer engineering, and electrical engineering majors. The changes to this edition are intended to provide support of the current draft version of the ACM/ IEEE Computer Science Curricula 2013 (CS2013). CS2013 adds Information Assurance and Security (IAS) to the curriculum recommendation as one of the Knowledge Areas in the Computer Science Body of Knowledge. The document states that IAS is now part of the curriculum recommendation because of the critical role of IAS in computer science education. CS2013 divides all course work into three categories: Core-Tier 1 (all topics should be included in the curriculum), Core-Tier-2 (all or almost all topics should be included), and elective (desirable to provide breadth and depth). In the IAS area, CS2013 recommends topics in Fundamental Concepts and Network Security in Tier 1 and Tier 2, and Cryptography topics as elective. This text covers virtually all of the topics listed by CS2013 in these three categories. The book also serves as a basic reference volume and is suitable for self-study. 12 Preface Plan Of The Text The book is organized in three parts: ■■ Part One. Cryptography: A concise survey of the cryptographic algorithms and protocols underlying network security applications, including encryption, hash functions, message authentication, and digital signatures. ■■ Part Two. Network Security Applications: Covers important network security tools and applications, including key distribution, Kerberos, X.509v3 certificates, Extensible Authentication Protocol, S/MIME, IP Security, SSL/TLS, IEEE 802.11i WiFi security, and cloud security. ■■ Part Three. System Security: Looks at system-level security issues, including the threat of and countermeasures for malicious software and intruders, and the use of firewalls. The book includes a number of pedagogic features, including the use of numerous figures and tables to clarify the discussions. Each chapter includes a list of key words, review questions, homework problems, and suggestions for further reading. The book also includes an extensive glossary, a list of frequently used acronyms, and a list of references. In addition, a test bank is available to instructors. Instructor Support Materials The major goal of this text is to make it as effective a teaching tool for this exciting and fastmoving subject as possible. This goal is reflected both in the structure of the book and in the supporting material. The following supplementary materials that will aid the instructor accompany the text: ■■ Solutions manual: Solutions to all end-of-chapter Review Questions and Problems. ■■ Projects manual:

Suggested project assignments for all of the project categories listed below. ■■ PowerPoint slides: A set of slides covering all chapters, suitable for use in lecturing. ■■ PDF files: Reproductions of all figures and tables from the book. ■■ Test bank: A chapter-by-chapter set of questions with a separate file of answers. ■■ Sample syllabi: The text contains more material than can be conveniently covered in one semester. Accordingly, instructors are provided with several sample syllabi that guide the use of the text within limited time. These samples are based on real-world experience by professors who used the fourth edition. All of these support materials are available at the Instructor Resource Center (IRC) for this textbook, which can be reached through the Publisher's Website www.pearsonglobaleditions.com/stallings. To gain access to the IRC, please contact your local Pearson sales representative. Projects And Other Student Exercises For many instructors, an important component of a network security course is a project or set of projects by which the student gets hands-on experience to reinforce concepts from the text. This book provides an unparalleled degree of support, including a projects component in the course. The IRC includes not only guidance on how to assign and structure the projects, but also a set of project assignments that covers a broad range of topics from the text: ■■ Hacking project: This exercise is designed to illuminate the key issues in intrusion detection and prevention. ■■ Lab exercises: A series of projects that involve programming and experimenting with concepts from the book. ■■ Research projects: A series of research assignments that instruct the student to research a particular topic on the Internet and write a report. ■■ Programming projects: A series of programming projects that cover a broad range of topics and that can be implemented in any suitable language on any platform. ■■ Practical security assessments: A set of exercises to examine current infrastructure and practices of an existing organization. ■■ Firewall projects: A portable network firewall visualization simulator is provided, together with exercises for teaching the fundamentals of firewalls. ■■ Case studies: A set of real-world case studies, including learning objectives, case description, and a series of case discussion questions. ■■ Writing assignments: A set of suggested writing assignments, organized by chapter. ■■ Reading/report assignments: A list of papers in the literature—one for each chapter— that can be assigned for the student to read and then write a short report. This diverse set of projects and other student exercises enables the instructor to use the book as one component in a rich and varied learning experience and to tailor a course plan to meet the specific needs of the instructor and students. See Appendix B in this book for details. Online CONTENT For Students For this new edition, a tremendous amount of original supporting material for students has been made available online. Preface 13 Purchasing this textbook new also grants the reader one year of access to the Companion Website, which includes the following materials: ■■ Online chapters: To limit the size and cost of the book, three chapters of the book are provided in PDF format. This includes a chapter on SHA-3, a chapter on SNMP security, and one on legal and ethical issues. The chapters are listed in this book's table of contents. ■■ Online appendices: There are numerous interesting topics that support material found in the text but whose inclusion is not warranted in the printed text. A number of online appendices cover these topics for the interested student. The appendices are listed in this book's table of contents. ■■ Homework problems and solutions: To aid the student in understanding the material, a separate set of homework problems with solutions are available. These enable the students to test their understanding of the text. ■■ Key papers: A number of papers from the professional literature, many hard to find, are provided for further reading. ■■ Supporting documents: A variety of other useful documents are referenced in the text and provided online. To access the Companion Website, click on the Premium Content link at the Companion Website or at pearsonglobaleditions.com/stallings and enter the student access code found on the card in the front of the book. Relationship To Cryptography And

Network Security This book is adapted from Cryptography and Network Security, Seventh Edition, Global Edition (CNS7eGE). CNS7eGE provides a substantial treatment of cryptography, key management, and user authentication, including detailed analysis of algorithms and a significant mathematical component, all of which covers nearly 500 pages. Network Security Essentials: Applications and Standards, Sixth Edition, Global Edition (NSE6eGE), provides instead a concise overview of these topics in Chapters 2 through 4. NSE6eGE includes all of the remaining material of CNS7eGE. NSE6eGE also covers SNMP security, which is not covered in CNS7eGE. Thus, NSE6eGE is intended for college courses and professional readers whose interest is primarily in the application of network security and who do not need or desire to delve deeply into cryptographic theory and principles. Acknowledgments This new edition has benefited from review by a number of people who gave generously of their time and expertise. The following professors reviewed the manuscript: Jim Helm (Arizona State University, Ira A. Fulton College of Engineering, Information Technology), Ali Saman Tosun (University of Texas at San Antonio, Computer Science Department), Haibo Wang (DIBTS, Texas A&M International University), Xunhua Wang (James Madison University, Department of Computer Science), Robert Kayl (University of Maryland University College), Scott Anderson (Southern Adventist University, School of Computing), and Jonathan Katz (University of Maryland, Department of Computer Science). 14 Preface Thanks also to the people who provided detailed technical reviews of one or more chapters: Kashif Aftab, Alan Cantrell, Rajiv Dasmohapatra, Edip Demirbilek, Dan Dieterle, Gerardo Iglesias Galvan, Michel Garcia, David Gueguen, Anasuya Threse Innocent, Dennis Kavanagh, Duncan Keir, Robert Knox, Bo Lin, Kousik Nandy, Nickolay Olshevsky, Massimiliano Sembiante, Oscar So, and Varun Tewari. Nikhil Bhargava (IIT Delhi) developed the set of online homework problems and solutions. Professor Sreekanth Malladi of Dakota State University developed the hacking exercises. Sanjay Rao and Ruben Torres of Purdue developed the laboratory exercises that appear in the IRC. The following people contributed project assignments that appear in the instructor's supplement: Henning Schulzrinne (Columbia University), Cetin Kaya Koc (Oregon State University), and David Balenson (Trusted Information Systems and George Washington University). Kim McLaughlin developed the test bank. Finally, I thank the many people responsible for the publication of this text, all of whom did their usual excellent job. This includes the staff at Pearson, particularly my editor Tracy Johnson, program manager Carole Snyder, and production manager Bob Engelhardt. Thanks also to the marketing and sales staffs at Pearson, without whose efforts this text would not be in front of you. ACKNOWLEDGEMENTS FOR THE GLOBAL EDITION The publishers would like to thank the following for contributing to and reviewing the Global Edition: A. Kannammal (Coimbatore Institute of Technology), Somitra Sanadhya (IIIT Delhi), Atul Kahate (Symbiosis University and Pune University), Anwitaman Datta (NTU Singapore), and Khyat Sharma. Preface 15 About the Author Dr. William Stallings has authored 18 titles, and counting revised editions, over 40 books on computer security, computer networking, and computer architecture. His writings have appeared in numerous publications, including the Proceedings of the IEEE, ACM Computing Reviews, and Cryptologia. He has 13 times received the award for the best Computer Science textbook of the year from the Text and Academic Authors Association. In over 30 years in the field, he has been a technical contributor, technical manager, and an executive with several high-technology firms. He has designed and implemented both TCP/IP-based and OSI-based protocol suites on a variety of computers and operating systems, ranging from microcomputers to mainframes. As a consultant, he has advised government agencies, computer and software vendors, and major users on the design, selection, and use of networking software and products. He created and maintains the Computer Science Student Resource Site at ComputerScienceStudent.com. This site provides documents and links on a variety of subjects of general interest to computer science students (and professionals). He is a member of the editorial board of Cryptologia, a scholarly journal devoted to all aspects of cryptology. Dr. Stallings holds a Ph.D. from MIT in Computer Science and a B.S. from Notre Dame in electrical engineering. 16 17 Introduction 1.1 Computer Security Concepts A Definition of Computer Security Examples The Challenges of Computer Security 1.2 The OSI Security Architecture 1.3 Security Attacks Passive Attacks Active Attacks 1.4 Security Services Authentication Access Control Data Confidentiality Data Integrity Nonrepudiation Availability Service 1.5 Security Mechanisms 1.6 Fundamental Security Design Principles 1.7 Attack Surfaces and Attack Trees Attack Surfaces Attack Trees 1.8 A Model for Network Security 1.9 Standards 1.10 Key Terms, Review Questions, and Problems Chapter 18 chapter 1 / Introduction The requirements of information security within an organization have undergone two major changes in the last several decades. Before the widespread use of data processing equipment, the security of information felt to be valuable to an organization was provided primarily by physical and administrative means. An example of the former is the use of rugged filing cabinets with a combination lock for storing sensitive documents. An example of the latter is personnel screening procedures used during the hiring process. With the introduction of the computer, the need for automated tools for protecting files and other information stored on the computer became evident. This is especially the case for a shared system, such as a time-sharing system, and the need is even more acute for systems that can be accessed over a public telephone network, data network, or the Internet. The generic name for the collection of tools designed to protect data and to thwart hackers is computer security. The second major change that affected security is the introduction of distributed systems and the use of networks and communications facilities for carrying data between terminal user and computer and between computer and computer. Network security measures are needed to protect data during their transmission. In fact, the term network security is somewhat misleading, because virtually all business, government, and academic organizations interconnect their data processing equipment with a collection of interconnected networks. Such a collection is often referred to as an internet,1 and the term internet security is used. 1 We use the term internet with a lowercase "i" to refer to any interconnected collection of network. A corporate intranet is an example of an internet. The Internet with a capital "I" may be one of the facilities used by an organization to construct its internet. Learning Objectives After studying this chapter, you should be able to: ◆◆ Describe the key security requirements of confidentiality, integrity, and availability. ◆◆ Describe the X.800 security architecture for OSI. ◆◆ Discuss the types of security threats and attacks that must be dealt with and give examples of the types of threats and attacks that apply to different categories of computer and network assets. ◆◆ Explain the fundamental security design principles. ◆◆ Discuss the use of attack surfaces and attack trees. ◆◆ List and briefly describe key organizations involved in cryptography standards. chapter 1 / Introduction 19 There are no clear boundaries between these two forms of security. For example, a computer virus may be introduced into a system physically when it arrives on a flash drive or an optical disk and is subsequently loaded onto a computer. Viruses may also arrive over an internet. In either case, once the virus is resident on a computer system, internal computer security tools are needed to detect and recover from the virus. This book focuses on internet security, which consists of measures to deter, prevent, detect, and correct security violations that involve the transmission of information. That is a broad statement that covers a host of possibilities. To give you a feel for the areas covered in this book, consider the following examples of security violations: 1. User A transmits a file to user B. The file contains sensitive information (e.g., payroll records) that is to be protected from disclosure. User C, who is not authorized to read the file, is able to monitor the transmission and capture a copy of the file during its transmission. 2. A network manager, D, transmits a message to a computer, E, under its management. The message instructs computer E to update an authorization file to include the identities of a number of new users who are to be given access to that computer. User F intercepts the message, alters its contents to add or delete entries, and then forwards the message to E, which accepts the message as coming from manager D and updates its authorization file accordingly. 3. Rather than intercept a message, user F constructs its own message with the desired entries and transmits that message to E as if it had come from manager D. Computer E accepts the message as coming from manager D and updates its authorization file accordingly. 4. An employee is fired without warning. The personnel manager sends a message to a server system to invalidate the employee's account. When the invalidation is accomplished, the server is to post a notice to the employee's file as confirmation of the action. The employee is able to intercept the message and delay it long enough to make a final access to the server to retrieve sensitive information. The message is then forwarded, the action taken, and the confirmation posted. The employee's action may go unnoticed for some considerable time. 5. A message is sent from a customer to a stockbroker with instructions for various transactions. Subsequently, the investments lose value and the customer denies sending the message. Although this list by no means exhausts the possible types of security violations, it illustrates the range of concerns of network security. This chapter provides a general overview of the subject matter that structures the material in the remainder of the book. We begin with a general discussion of network security services and mechanisms and of the types of attacks they are designed for. Then we develop a general overall model within which the security services and mechanisms can be viewed. 20 chapter 1 / Introduction 1.1 Computer Security Concepts A Definition of Computer Security The NIST Computer Security Handbook [NIST95] defines the term computer security as Computer Security: The protection afforded to an automated information system in order to attain the applicable objectives of preserving the integrity, availability, and confidentiality of information system resources (includes hardware, software, firmware, information/data, and telecommunications). This definition introduces three key objectives that are at the heart of computer security. ■ Confidentiality: This term covers two related concepts: Data2 confidentiality: Assures that private or confidential information is not made available or disclosed to unauthorized individuals. Privacy: Assures that individuals control or influence what information related to them may be collected and stored and by whom and to whom that information may be disclosed. ■ Integrity: This term covers two related concepts: Data integrity: Assures that data (both stored and in transmitted packets) and programs are changed only in a specified and authorized manner. System integrity: Assures that a system performs its intended function in an unimpaired manner, free from deliberate or inadvertent unauthorized manipulation of the system. ■ Availability: Assures that systems work promptly and service is not denied to authorized users. These three concepts form what is often referred to as the CIA triad. The three concepts embody the fundamental security objectives for both data and for information and computing services. For example, the NIST Standards for Security Categorization of Federal Information and Information Systems (FIPS 199) lists confidentiality, integrity, and availability as the three security objectives for information and for information systems. FIPS 199 provides a useful characterization of these three objectives in terms of requirements and the definition of a loss of security in each category. ■ Confidentiality: Preserving authorized restrictions on information access and disclosure, including means for protecting personal privacy and proprietary information. A loss of confidentiality is the unauthorized disclosure of information. 2 RFC 4949 defines information as "facts and ideas, which can be represented (encoded) as various forms of data," and data as "information in a specific physical representation, usually a sequence of symbols that have meaning; especially a representation of information that can be processed or produced by a computer." Security literature typically does not make much of a distinction, nor does this book. 1.1 / Computer Security Concepts 21 ■ Integrity: Guarding against improper information modification or destruction, including ensuring information nonrepudiation and authenticity. A loss of integrity is the unauthorized modification or destruction of information. ■ Availability: Ensuring timely and reliable access to and use of information. A loss of availability is the disruption of access to or use of information or an information system. Although the use of the CIA triad to define security objectives is well established, some in the security field feel that additional concepts are needed to present a complete picture (Figure 1.1). Two of the most commonly mentioned are ■ Authenticity: The property of being genuine and being able to be verified and trusted; confidence in the validity of a transmission, a message, or message originator. This means verifying that users are who they say they are and that each input arriving at the system came from a trusted source. ■ Accountability: The security goal that generates the requirement for actions of an entity to be traced uniquely to that entity. This supports nonrepudiation, deterrence, fault isolation, intrusion detection and prevention, and after-action recovery and legal action. Because truly secure systems are not yet an achievable goal, we must be able to trace a security breach to a responsible party. Systems must keep records of their activities to permit later forensic analysis to trace security breaches or to aid in transaction disputes. Examples We now provide some examples of applications that illustrate the requirements just enumerated.3 For these examples, we use three levels of impact on organizations or individuals should there be a breach of security (i.e., a loss of confidentiality, integrity, or availability). These levels are defined in FIPS 199: ■ Low: The loss could be expected to have a limited adverse effect on organizational operations, organizational assets, or individuals. A limited adverse effect means that, for example, the loss of confidentiality, integrity, or availability might 3 These examples are taken from a security policy document published by the Information Technology Security and Privacy Office at Purdue University. Figure 1.1 Essential Network and Computer Security Requirements Data and services Availability Integrity Accountability Authenticity Condentiality 22 chapter 1 / Introduction (i) cause a degradation in mission capability to an extent and duration that the organization is able to perform its primary functions, but the effectiveness of the functions is noticeably reduced; (ii) result in minor damage to organizational assets; (iii) result in minor financial loss; or (iv) result in minor harm to individuals. ■ Moderate: The loss could be expected to have a serious adverse effect on organizational operations, organizational assets, or individuals. A serious adverse effect means that, for example, the loss might (i) cause a significant degradation in mission capability to an extent and duration that the organization is able to perform its primary functions, but the effectiveness of the functions is significantly reduced; (ii) result in significant damage to organizational assets; (iii) result in significant financial loss; or (iv) result in significant harm to individuals that does not involve loss of life or serious, life-threatening injuries. ■ High: The loss could be expected to have a severe or catastrophic adverse effect on organizational operations, organizational assets, or individuals. A severe or catastrophic adverse effect means that, for example, the loss might (i) cause a severe degradation in or loss of mission capability to an extent and duration that the organization is not able to perform one or more of its primary functions; (ii) result in major damage to organizational assets; (iii) result in major financial loss; or (iv) result in severe or catastrophic harm to individuals involving loss of life or serious, life-threatening injuries. Confidentiality Student grade information is an asset whose confidentiality is considered to be highly important by students. In the United States, the release of such information is regulated by the Family Educational Rights and Privacy Act (FERPA). Grade information should only be available to students, their parents, and employees that require the information to do their job. Student enrollment information may have a moderate confidentiality rating. While still covered by FERPA, this information is seen by more people on a daily basis, is less likely to be targeted than grade information, and results in less damage if disclosed. Directory information (such as lists of students, faculty, or departmental lists) may be assigned a low confidentiality rating or indeed no rating. This information is typically freely available to the public and published on a school's Web site. Integrity Several aspects of integrity are illustrated by the example of a hospital patient's allergy information stored in a database. The doctor should be able to trust that the information is correct and current. Now suppose that an employee (e.g., a nurse) who is authorized to view and update this information deliberately falsifies the data to cause harm to the hospital. The database needs to be restored to a trusted basis quickly, and it should be possible to trace the error back to the person responsible. Patient allergy information is an example of an asset with a high requirement for integrity. Inaccurate information could result in serious harm or death to a patient and expose the hospital to massive liability. An example of an asset that may be assigned a moderate level of integrity requirement is a Web site that offers a forum to registered users to discuss some specific topic. Either a registered user or a hacker could falsify some entries or deface the Web site. If the forum exists only for the enjoyment of the users, brings in little or no advertising revenue, and is not used for something important such 1.1 / Computer Security Concepts 23 as research, then potential damage is not severe. The Web master may experience some data, financial, and time loss. An example of a low-integrity requirement is an anonymous online poll. Many Web sites, such as news organizations, offer these polls to their users with very few safeguards. However, the inaccuracy and unscientific nature of such polls are well understood. Availability The more critical a component or service, the higher is the level of availability required. Consider a system that provides authentication services for critical systems, applications, and devices. An interruption of service results in the inability for customers to access computing resources and for the staff to access the resources they need to perform critical tasks. The loss of the service translates into a large financial loss due to lost employee productivity and potential customer loss. An example of an asset that typically would be rated as having a moderate availability requirement is a public Web site for a university; the Web site provides information for current and prospective students and donors. Such a site is not a critical component of the university's information system, but its unavailability will cause some embarrassment. An online telephone directory lookup application would be classified as a lowavailability requirement. Although the temporary loss of the application may be an annoyance, there are other ways to access the information, such as a hardcopy directory or the operator. The Challenges of Computer Security Computer and network security is both fascinating and complex. Some of the reasons include: 1. Security is not as simple as it might first appear to the novice. The requirements seem to be straightforward; indeed, most of the major requirements for security services can be given self-explanatory, one-word labels: confidentiality, authentication, nonrepudiation, and integrity. But the mechanisms used to meet those requirements can be quite complex, and understanding them may involve rather subtle reasoning. 2. In developing a particular security mechanism or algorithm, one must always consider potential attacks on those security features. In many cases, successful attacks are designed by looking at the problem in a completely different way, therefore exploiting an unexpected weakness in the mechanism. 3. Because of point 2, the procedures used to provide particular services are often counterintuitive. Typically, a security mechanism is complex, and it is not obvious from the statement of a particular requirement that such elaborate measures are needed. It is only when the various aspects of the threat are considered that elaborate security mechanisms make sense. 4. Having designed various security mechanisms, it is necessary to decide where to use them. This is true both in terms of physical placement (e.g., at what points in a network are certain security mechanisms needed) and in a logical sense [e.g., at what layer or layers of an architecture such as TCP/IP (Transmission Control Protocol/Internet Protocol) should mechanisms be placed]. 24 chapter 1 / Introduction 5. Security mechanisms typically involve more than a particular algorithm or protocol. They also require that participants be in possession of some secret information (e.g., an encryption key), which raises questions about the creation, distribution, and protection of that secret information. There also may be a reliance on communications protocols whose behavior may complicate the task of developing the security mechanism. For example, if the proper functioning of the security mechanism requires setting time limits on the transit time of a message from sender to receiver, then any protocol or network that introduces variable, unpredictable delays may render such time limits meaningless. 6. Computer and network security is essentially a battle of wits between a perpetrator who tries to find holes and the designer or administrator who tries to close them. The great advantage that the attacker has is that he or she need only find a single weakness, while the designer must find and eliminate all weaknesses to achieve perfect security. 7. There is a natural tendency on the part of users and system managers to perceive little benefit from security investment until a security failure occurs. 8. Security requires regular, even constant, monitoring, and this is difficult in today's short-term, overloaded environment. 9. Security is still too often an afterthought to be incorporated into a system after the design is complete rather than being an integral part of the design process. 10. Many users (and even security administrators) view strong security as an impediment to efficient and user-friendly operation of an information system or use of information. The difficulties just enumerated will be encountered in numerous ways as we examine the various security threats and mechanisms throughout this book. 1.2 The Osi Security Architecture To assess effectively the security needs of an organization and to evaluate and choose various security products and policies, the manager responsible for computer and network security needs some systematic way of defining the requirements for security and characterizing the approaches to satisfying those requirements. This is difficult enough in a centralized data processing environment; with the use of local and wide area networks, the problems are compounded. ITU-T4 Recommendation X.800, Security Architecture for OSI, defines such a systematic approach.5 The OSI security architecture is useful to managers as a way of organizing the task of providing security. Furthermore, because this architecture 4 The International Telecommunication Union (ITU) Telecommunication Standardization Sector (ITU-T) is a United Nations-sponsored agency that develops standards, called Recommendations, relating to telecommunications and to open systems interconnection (OSI). 5 The OSI security architecture was developed in the context of the OSI protocol architecture, which is described in Appendix D. However, for our purposes in this chapter, an understanding of the OSI protocol architecture is not required. 1.3 / Security Attacks 25 was developed as an international standard, computer and communications vendors have developed security features for their products and services that relate to this structured definition of services and mechanisms. For our purposes, the OSI security architecture provides a useful, if abstract, overview of many of the concepts that this book deals with. The OSI security architecture focuses on security attacks, mechanisms, and services. These can be defined briefly as ■ Security attack: Any action that compromises the security of information owned by an organization. ■ Security mechanism: A process (or a device incorporating such a process) that is designed to detect, prevent, or recover from a security attack. ■ Security service: A processing or communication service that enhances the security of the data processing systems and the information transfers of an organization. The services are intended to counter security attacks, and they make use of one or more security mechanisms to provide the service. In the literature, the terms threat and attack are commonly used to mean more or less the same thing. Table 1.1 provides definitions taken from RFC 4949, Internet Security Glossary. 1.3 Security Attacks A useful means of classifying security attacks, used both in X.800 and RFC 4949, is in terms of passive attacks and active attacks. A passive attack attempts to learn or make use of information from the system but does not affect system resources. An active attack attempts to alter system resources or affect their operation. Passive Attacks Passive attacks (Figure 1.2a) are in the nature of eavesdropping on, or monitoring of, transmissions. The goal of the opponent is to obtain information that is being transmitted. Two types of passive attacks are the release of message contents and traffic analysis. The release of message contents is easily understood. A telephone conversation, an electronic mail message, and a transferred file may contain sensitive or confidential information. We would like to prevent an opponent from learning the contents of these transmissions. Threat A potential for violation of security, which exists when there is a circumstance, capability, action, or event that could breach security and cause harm. That is, a threat is a possible danger that might exploit a vulnerability. Attack An assault on system security that derives from an intelligent threat. That is, an intelligent act that is a deliberate attempt (especially in the sense of a method or technique) to evade security services and violate the security policy of a system. Table 1.1 Threats and Attacks (RFC 4949) 26 chapter 1 / Introduction A second type of passive attack, traffic analysis, is subtler. Suppose that we had a way of masking the contents of messages or other information traffic so that opponents, even if they captured the message, could not extract the information from the message. The common technique for masking contents is encryption. If we had encryption protection in place, an opponent still might be able to observe the pattern of these messages. The opponent could determine the location and identity of communicating hosts and could observe the frequency and length of messages being exchanged. This information might be useful in guessing the nature of the communication that was taking place. Passive attacks are very difficult to detect, because they do not involve any alteration of the data. Typically, the message traffic is sent and received in an Figure 1.2 Security Attacks (a) Passive attacks Alice (b) Active attacks Bob Darth Bob Darth Alice Internet or other comms facility Internet or other comms facility 1 2 3 1.4 / Security Services 27 apparently normal fashion, and neither the sender nor the receiver is aware that a third party has read the messages or observed the traffic pattern. However, it is feasible to prevent the success of these attacks, usually by means of encryption. Thus, the emphasis in dealing with passive attacks is on prevention rather than detection. Active Attacks Active attacks (Figure 1.2b) involve some modification of the data stream or the creation of a false stream and can be subdivided into four categories: masquerade, replay, modification of messages, and denial of service. A masquerade takes place when one entity pretends to be a different entity (path 2 of Figure 1.2b is active). A masquerade attack usually includes one of the other forms of active attack. For example, authentication sequences can be captured and replayed after a valid authentication sequence has taken place, thus enabling an authorized entity with few privileges to obtain extra privileges by impersonating an entity that has those privileges. Replay involves the passive capture of a data unit and its subsequent retransmission to produce an unauthorized effect (paths 1, 2, and 3 active). Modification of messages simply means that some portion of a legitimate message is altered, or that messages are delayed or reordered, to produce an unauthorized effect (paths 1 and 2 active). For example, a message meaning “Allow John Smith to read confidential file accounts” is modified to mean “Allow Fred Brown to read confidential file accounts.” The denial of service prevents or inhibits the normal use or management of communications facilities (path 3 active). This attack may have a specific target; for example, an entity may suppress all messages directed to a particular destination (e.g., the security audit service). Another form of service denial is the disruption of an entire network—either by disabling the network or by overloading it with messages so as to degrade performance. Active attacks present the opposite characteristics of passive attacks. Whereas passive attacks are difficult to detect, measures are available to prevent their success. On the other hand, it is quite difficult to prevent active attacks absolutely because of the wide variety of potential physical, software, and network vulnerabilities. Instead, the goal is to detect active attacks and to recover from any disruption or delays caused by them. If the detection has a deterrent effect, it also may contribute to prevention. 1.4 Security Services X.800 defines a security service as a service that is provided by a protocol layer of communicating open systems and that ensures adequate security of the systems or of data transfers. Perhaps a clearer definition is found in RFC 4949, which provides the following definition: A processing or communication service that is provided by 28 chapter 1 / Introduction a system to give a specific kind of protection to system resources; security services implement security policies and are implemented by security mechanisms. X.800 divides these services into five categories and fourteen specific services (Table 1.2). We look at each category in turn.6 6 There is no universal agreement about many of the terms used in the security literature. For example, the term integrity is sometimes used to refer to all aspects of information security. The term authentication is sometimes used to refer both to verification of identity and to the various functions listed under integrity in this chapter. Our usage here agrees with both X.800 and RFC 4949. Authentication The assurance that the communicating entity is the one that it claims to be. Peer Entity Authentication Used in association with a logical connection to provide confidence in the identity of the entities connected. Data-Origin Authentication In a connectionless transfer, provides assurance that the source of received data is as claimed. Access Control The prevention of unauthorized use of a resource (i.e., this service controls who can have access to a resource, under what conditions access can occur, and what those accessing the resource are allowed to do). Data Confidentiality The protection of data from unauthorized disclosure. Connection Confidentiality The protection of all user data on a connection. Connectionless Confidentiality The protection of all user data in a single data block. Selective-Field Confidentiality The confidentiality of selected fields within the user data on a connection or in a single data block. Traffic-Flow Confidentiality The protection of the information that might be derived from observation of traffic flows. Data Integrity The assurance that data received are exactly as sent by an authorized entity (i.e., contain no modification, insertion, deletion, or replay). Connection Integrity with Recovery Provides for the integrity of all user data on a connection and detects any modification, insertion, deletion, or replay of any data within an entire data sequence, with recovery attempted. Connection Integrity without Recovery As above, but provides only detection without recovery. Selective-Field Connection Integrity Provides for the integrity of selected fields within the user data of a data block transferred over a connection and takes the form of determination of whether the selected fields have been modified, inserted, deleted, or replayed. Connectionless Integrity Provides for the integrity of a single connectionless data block and may take the form of detection of data modification. Additionally, a limited form of replay detection may be provided. Selective-Field Connectionless Integrity Provides for the integrity of selected fields within a single connectionless data block; takes the form of determination of whether the selected fields have been modified. Nonrepudiation Provides protection against denial by one of the entities involved in a communication of having participated in all or part of the communication. Nonrepudiation, Origin Proof that the message was sent by the specified party. Nonrepudiation, Destination Proof that the message was received by the specified party. Table 1.2 Security

Services (X.800) 1.4 / Security Services 29 Authentication The authentication service is concerned with assuring that a communication is authentic. In the case of a single message, such as a warning or alarm signal, the function of the authentication service is to assure the recipient that the message is from the source that it claims to be from. In the case of an ongoing interaction, such as the connection of a terminal to a host, two aspects are involved. First, at the time of connection initiation, the service assures that the two entities are authentic (i.e., that each is the entity that it claims to be). Second, the service must assure that the connection is not interfered with in such a way that a third party can masquerade as one of the two legitimate parties for the purposes of unauthorized transmission or reception. Two specific authentication services are defined in X.800: ■ Peer entity authentication: Provides for the corroboration of the identity of a peer entity in an association. Two entities are considered peers if they implement the same protocol in different systems (e.g., two TCP modules in two communicating systems). Peer entity authentication is provided for use at the establishment of or during the data transfer phase of a connection. It attempts to provide confidence that an entity is not performing either a masquerade or an unauthorized replay of a previous connection. ■ Data origin authentication: Provides for the corroboration of the source of a data unit. It does not provide protection against the duplication or modification of data units. This type of service supports applications like electronic mail, where there are no prior interactions between the communicating entities. Access Control In the context of network security, access control is the ability to limit and control the access to host systems and applications via communications links. To achieve this, each entity trying to gain access must first be identified, or authenticated, so that access rights can be tailored to the individual. Data Confidentiality Confidentiality is the protection of transmitted data from passive attacks. With respect to the content of a data transmission, several levels of protection can be identified. The broadest service protects all user data transmitted between two users over a period of time. For example, when a TCP connection is set up between two systems, this broad protection prevents the release of any user data transmitted over the TCP connection. Narrower forms of this service can also be defined, including the protection of a single message or even specific fields within a message. These refinements are less useful than the broad approach and may even be more complex and expensive to implement. The other aspect of confidentiality is the protection of traffic flow from analysis. This requires that an attacker not be able to observe the source and destination, frequency, length, or other characteristics of the traffic on a communications facility. 30 chapter 1 / Introduction Data Integrity As with confidentiality, integrity can apply to a stream of messages, a single message, or selected fields within a message. Again, the most useful and straightforward approach is total stream protection. A connection-oriented integrity service deals with a stream of messages and assures that messages are received as sent with no duplication, insertion, modification, reordering, or replays. The destruction of data is also covered under this service. Thus, the connection-oriented integrity service addresses both message stream modification and denial of service. On the other hand, a connectionless integrity service deals with individual messages without regard to any larger context and generally provides protection against message modification only. We can make a distinction between service with and without recovery. Because the integrity service relates to active attacks, we are concerned with detection rather than prevention. If a violation of integrity is detected, then the service may simply report this violation, and some other portion of software or human intervention is required to recover from the violation. Alternatively, there are mechanisms available to recover from the loss of integrity of data, as we will review subsequently. The incorporation of automated recovery mechanisms is typically the more attractive alternative. Nonrepudiation Nonrepudiation prevents either sender or receiver from denying a transmitted message. Thus, when a message is sent, the receiver can prove that the alleged sender in fact sent the message. Similarly, when a message is received, the sender can prove that the alleged receiver in fact received the message. Availability Service Both X.800 and RFC 4949 define availability to be the property of a system or a system resource being accessible and usable upon demand by an authorized system entity, according to performance specifications for the system (i.e., a system is available if it provides services according to the system design whenever users request them). A variety of attacks can result in the loss of or reduction in availability. Some of these attacks are amenable to automated countermeasures, such as authentication and encryption, whereas others require some sort of physical action to prevent or recover from loss of availability of elements of a distributed system. X.800 treats availability as a property to be associated with various security services. However, it makes sense to call out specifically an availability service. An availability service is one that protects a system to ensure its availability. This service addresses the security concerns raised by denial-of-service attacks. It depends on proper management and control of system resources and thus depends on access control service and other security services. 1.5 / Security Mechanisms 31 Specific Security Mechanisms May be incorporated into the appropriate protocol layer in order to provide some of the OSI security services. Encipherment The use of mathematical algorithms to transform data into a form that is not readily intelligible. The transformation and subsequent recovery of the data depend on an algorithm and zero or more encryption keys. Digital Signature Data appended to, or a cryptographic transformation of, a data unit that allows a recipient of the data unit to prove the source and integrity of the data unit and protect against forgery (e.g., by the recipient). Access Control A variety of mechanisms that enforce access rights to resources. Data Integrity A variety of mechanisms used to assure the integrity of a data unit or stream of data units. Authentication Exchange A mechanism intended to ensure the identity of an entity by means of information exchange. Traffic Padding The insertion of bits into gaps in a data stream to frustrate traffic analysis attempts. Routing Control Enables selection of particular physically secure routes for certain data and allows routing changes, especially when a breach of security is suspected. Notarization The use of a trusted third party to assure certain properties of a data exchange. Pervasive Security Mechanisms Mechanisms that are not specific to any particular OSI security service or protocol layer. Trusted Functionality That which is perceived to be correct with respect to some criteria (e.g., as established by a security policy). Security Label The marking bound to a resource (which may be a data unit) that names or designates the security attributes of that resource. Event Detection Detection of security-relevant events. Security Audit Trail Data collected and potentially used to facilitate a security audit, which is an independent review and examination of system records and activities. Security Recovery Deals with requests from mechanisms, such as event handling and management functions, and takes recovery actions. Table 1.3 Security Mechanisms (X.800) 1.5 Security Mechanisms Table 1.3 lists the security mechanisms defined in X.800. The mechanisms are divided into those that are implemented in a specific protocol layer, such as TCP or an application-layer protocol, and those that are not specific to any particular protocol layer or security service. These mechanisms will be covered in the appropriate places in the book, so we do not elaborate now except to comment on the 32 chapter 1 / Introduction definition of encipherment. X.800 distinguishes between reversible encipherment mechanisms and irreversible encipherment mechanisms. A reversible encipherment mechanism is simply an encryption algorithm that allows data to be encrypted and subsequently decrypted. Irreversible encipherment mechanisms include hash algorithms and message authentication codes, which are used in digital signature and message authentication applications. Table 1.4, based on one in X.800, indicates the relationship between security services and security mechanisms. 1.6 Fundamental Security Design Principles Despite years of research and development, it has not been possible to develop security design and implementation techniques that systematically exclude security flaws and prevent all unauthorized actions. In the absence of such foolproof techniques, it is useful to have a set of widely agreed design principles that can guide the development of protection mechanisms. The National Centers of Academic Excellence in Information Assurance/Cyber Defense, which is jointly sponsored by the U.S. National Security Agency and the U.S. Department of Homeland Security, list the following as fundamental security design principles [NCAE13]: ■ Economy of mechanism ■ Fail-safe defaults ■ Complete mediation Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Peer entity authentication Service Mechanism Encipherment Digital signature Access control Data integrity Authentication exchange Trac padding Routing control Notarization Data origin authentication Access control Con-dentiality Trac ow con-dentiality Data integrity Nonrepudiation Availability Table 1.4 Relationship between Security Services and Mechanisms 1.6 / Fundamental Security Design Principles 33 ■ Open design ■ Separation of privilege ■ Least privilege ■ Least common mechanism ■ Psychological acceptability ■ Isolation ■ Encapsulation ■ Modularity ■ Layering ■ Least astonishment The first eight listed principles were initially proposed in [SALT75] and have withstood the test of time. In this section, we briefly discuss each principle. Economy of mechanism means that the design of security measures embodied in both hardware and software should be as simple and small as possible. The motivation for this principle is that relatively simple, small design is easier to test and verify thoroughly. With a complex design, there are many more opportunities for an adversary to discover subtle weaknesses to exploit that may be difficult to spot ahead of time. The more complex the mechanism, the more likely it is to possess exploitable flaws. Simple mechanisms tend to have fewer exploitable flaws and require less maintenance. Further, because configuration management issues are simplified, updating or replacing a simple mechanism becomes a less intensive process. In practice, this is perhaps the most difficult principle to honor. There is a constant demand for new features in both hardware and software, complicating the security design task. The best that can be done is to keep this principle in mind during system design to try to eliminate unnecessary complexity. Fail-safe default means that access decisions should be based on permission rather than exclusion. That is, the default situation is lack of access, and the protection scheme identifies conditions under which access is permitted. This approach exhibits a better failure mode than the alternative approach, where the default is to permit access. A design or implementation mistake in a mechanism that gives explicit permission tends to fail by refusing permission, a safe situation that can be quickly detected. On the other hand, a design or implementation mistake in a mechanism that explicitly excludes access tends to fail by allowing access, a failure that may long go unnoticed in normal use. For example, most file access systems work on this principle and virtually all protected services on client/server systems work this way. Complete mediation means that every access must be checked against the access control mechanism. Systems should not rely on access decisions retrieved from a cache. In a system designed to operate continuously, this principle requires that, if access decisions are remembered for future use, careful consideration should be given to how changes in authority are propagated into such local memories. File access systems appear to provide an example of a system that complies with this principle. However, typically, once a user has opened a file, no check is made to see if permissions change. To fully implement complete mediation, every time a user 34 chapter 1 / Introduction reads a field or record in a file, or a data item in a database, the system must exercise access control. This resource-intensive approach is rarely used. Open design means that the design of a security mechanism should be open rather than secret. For example, although encryption keys must be secret, encryption algorithms should be open to public scrutiny. The algorithms can then be reviewed by many experts, and users can therefore have high confidence in them. This is the philosophy behind the National Institute of Standards and Technology (NIST) program of standardizing encryption and hash algorithms and has led to the widespread adoption of NIST-approved algorithms. Separation of privilege is defined in [SALT75] as a practice in which multiple privilege attributes are required to achieve access to a restricted resource. A good example of this is multifactor user authentication, which requires the use of multiple techniques, such as a password and a smart card, to authorize a user. The term is also now applied to any technique in which a program is divided into parts that are limited to the specific privileges they require in order to perform a specific task. This is used to mitigate the potential damage of a computer security attack. One example of this latter interpretation of the principle is removing high privilege operations to another process and running that process with the higher privileges required to perform its tasks. Day-to-day interfaces are executed in a lower privileged process. Least privilege means that every process and every user of the system should operate using the least set of privileges necessary to perform the task. A good example of the use of this principle is role-based access control, described in Chapter 4. The system security policy can identify and define the various roles of users or processes. Each role is assigned only those permissions needed to perform its functions. Each permission specifies a permitted access to a particular resource (such as read and write access to a specified file or directory, connect access to a given host and port, etc.). Unless a permission is granted explicitly, the user or process should not be able to access the protected resource. More generally, any access control system should allow each user only the privileges that are authorized for that user. There is also a temporal aspect to the least privilege principle. For example, system programs or administrators who have special privileges should have those privileges only when necessary; when they are doing ordinary activities, the privileges should be withdrawn. Leaving them in place just opens the door to accidents. Least common mechanism means that the design should minimize the functions shared by different users, providing mutual security. This principle helps reduce the number of unintended communication paths and reduces the amount of hardware and software on which all users depend, thus making it easier to verify if there are any undesirable security implications. Psychological acceptability implies that the security mechanisms should not interfere unduly with the work of users, while at the same time meeting the needs of those who authorize access. If security mechanisms hinder the usability or accessibility of resources, then users may opt to turn off those mechanisms. Where possible, security mechanisms should be transparent to the users of the system or at most introduce minimal obstruction. In addition to not being intrusive or burdensome, security procedures must reflect the user's mental model of protection. If the protection procedures do not make sense to the user or if the user must translate his image of protection into a substantially different protocol, the user is likely to make errors. 1.6 / Fundamental Security Design Principles 35 Isolation is a principle that applies in three contexts. First, public access systems should be isolated from critical resources (data, processes, etc.) to prevent disclosure or tampering. In cases where the sensitivity or criticality of the information is high, organizations may want to limit the number of systems on which that data is stored and isolate them, either physically or logically. Physical isolation may include ensuring that no physical connection exists between an organization's public access information resources and an organization's critical information. When implementing logical isolation solutions, layers of security services and mechanisms should be established between public systems and secure systems responsible for protecting critical resources. Second, the processes and files of individual users should be isolated from one another except where it is explicitly desired. All modern operating systems provide facilities for such isolation, so that individual users have separate, isolated process space, memory space, and file space, with protections for preventing unauthorized access. Finally, security mechanisms should be isolated in the sense of preventing access to those mechanisms. For example, logical access control may provide a means of isolating cryptographic software from other parts of the host system, and for protecting cryptographic software from tampering and the keys from replacement or disclosure. Encapsulation can be viewed as a specific form of isolation based on objectoriented functionality. Protection is provided by encapsulating a collection of procedures and data objects in a domain of its own so that the internal structure of a data object is accessible only to the procedures of the protected subsystem, and the procedures may be called only at designated domain entry points. Modularity in the context of security refers both to the development of security functions as separate, protected modules and to the use of a modular architecture for mechanism design and implementation. With respect to the use of separate security modules, the design goal here is to provide common security functions and services, such as cryptographic functions, as common modules. For example, numerous protocols and applications make use of cryptographic functions. Rather than implementing such functions in each protocol or application, a more secure design is provided by developing a common cryptographic module that can be invoked by numerous protocols and applications. The design and implementation effort can then focus on the secure design and implementation of a single cryptographic module and including mechanisms to protect the module from tampering. With respect to the use of a modular architecture, each security mechanism should be able to support migration to new technology or upgrade of new features without requiring an entire system redesign. The security design should be modular so that individual parts of the security design can be upgraded without the requirement to modify the entire system. Layering refers to the use of multiple, overlapping protection approaches addressing the people, technology, and operational aspects of information systems. By using multiple, overlapping protection approaches, the failure or circumvention of any individual protection approach will not leave the system unprotected. We will see throughout this text that a layering approach is often used to provide multiple barriers between an adversary and protected information or services. This technique is often referred to as defense in depth. 36 chapter 1 / Introduction Least astonishment means that a program or user interface should always respond in the way that is least likely to astonish the user. For example, the mechanism for authorization should be transparent enough to a user that the user has a good intuitive understanding of how the security goals map to the provided security mechanism. 1.7 Attack Surfaces and Attack Trees In Section 1.3, we provided an overview of the spectrum of security threats and attacks facing computer and network systems. Section 11.1 goes into more detail about the nature of attacks and the types of adversaries that present security threats. This section elaborates on two concepts that are useful in evaluating and classifying threats: attack surfaces and attack trees. Attack Surfaces An attack surface consists of the reachable and exploitable vulnerabilities in a system [MANA11, HOWA03]. Examples of attack surfaces are the following: ■ Open ports on outward facing Web and other servers, and code listening on those ports ■ Services available on the inside of a firewall ■ Code that processes incoming data, e-mail, XML, office documents, and industry-specific custom data exchange formats ■ Interfaces, SQL, and Web forms ■ An employee with access to sensitive information vulnerable to a social engineering attack Attack surfaces can be categorized in the following way: ■ Network attack surface: This category refers to vulnerabilities over an enterprise network, wide-area network, or the Internet. Included in this category are network protocol vulnerabilities, such as those used for a denial-of-service attack, disruption of communications links, and various forms of intruder attacks. ■ Software attack surface: This refers to vulnerabilities in application, utility, or operating system code. A particular focus in this category is Web server software. ■ Human attack surface: This category refers to vulnerabilities created by personnel or outsiders, such as social engineering, human error, and trusted insiders. An attack surface analysis is useful for assessing the scale and severity of threats to a system. A systematic analysis of points of vulnerability makes developers and security analysts aware of where security mechanisms are required. Once an attack surface is defined, designers may be able to find ways to make the surface 1.7 / Attack Surfaces and Attack Trees 37 smaller, thus making the task of the adversary more difficult. The attack surface also provides guidance on setting priorities for testing, strengthening security measures, or modifying the service or application. As illustrated in Figure 1.3, the use of layering, or defense in depth, and attack surface reduction complement each other in mitigating security risk. Attack Trees An attack tree is a branching, hierarchical data structure that represents a set of potential techniques for exploiting security vulnerabilities [MAUW05, MOOR01, SCHN99]. The security incident that is the goal of the attack is represented as the root node of the tree, and the ways that an attacker could reach that goal are iteratively and incrementally represented as branches and subnodes of the tree. Each subnode defines a subgoal, and each subgoal may have its own set of further subgoals, etc. The final nodes on the paths outward from the root, that is, the leaf nodes, represent different ways to initiate an attack. Each node other than a leaf is either an AND-node or an OR-node. To achieve the goal represented by an AND-node, the subgoals represented by all of that node's subnodes must be achieved; and for an OR-node, at least one of the subgoals must be achieved. Branches can be labeled with values representing difficulty, cost, or other attack attributes, so that alternative attacks can be compared. The motivation for the use of attack trees is to effectively exploit the information available on attack patterns. Organizations such as CERT publish security advisories that have enabled the development of a body of knowledge about both general attack strategies and specific attack patterns. Security analysts can use the attack tree to document security attacks in a structured form that reveals key vulnerabilities. The attack tree can guide both the design of systems and applications, and the choice and strength of countermeasures. Figure 1.4, based on a figure in [DIMI07], is an example of an attack tree analysis for an Internet banking authentication application. The root of the tree is Figure 1.3 Defense in Depth and Attack Surface Attack surface Medium security risk High security risk Low Deep security risk Layering Shallow Small Large Medium security risk 38 chapter 1 / Introduction the objective of the attacker, which is to compromise a user's account. The shaded boxes on the tree are the leaf nodes, which represent events the comprise the attacks. Note that in this tree in this example, all the nodes other than leaf nodes are OR-nodes. The analysis to generate this tree considered the three components involved in authentication: ■ User terminal and user (UT/U): These attacks target the user equipment, including the tokens that may be involved, such as smartcards or other password generators, as well as the actions of the user. ■ Communications channel (CC): This type of attack focuses on communication links. ■ Internet banking server (IBS): These types of attacks are offline attacks against the servers that host the Internet banking application. Figure 1.4 An Attack Tree for Internet Banking Authentication Bank account compromise User credential compromise User credential guessing UT/U1a User surveillance UT/U1b Theft of token and handwritten notes Malicious software installation Vulnerability exploit UT/U2a Hidden code UT/U2b Worms UT/U3a Smartcard analyzers UT/U2c E-mails with malicious code UT/U3b Smartcard reader manipulator UT/U3c Brute force attacks with PIN calculators CC2 Sning UT/U4a Social engineering IBS3 Web site manipulation UT/U4b Web page obfuscation CC1 Pharming

Redirection of communication toward fraudulent site CC3 Active man-in-the middle attacks IBS1 Brute force attacks User communication with attacker Injection of commands Use of known authenticated session by attacker Normal user authentication with speci-ed session ID CC4 Pre-de-ned session IDs (session hijacking) IBS2 Security policy violation 1.8 / A Model for Network Security 39 Five overall attack strategies can be identified, each of which exploits one or more of the three components. The five strategies are as follows: ■ User credential compromise: This strategy can be used against many elements of the attack surface. There are procedural attacks, such as monitoring a user's action to observe a PIN or other credential, or theft of the user's token or handwritten notes. An adversary may also compromise token information using a variety of token attack tools, such as hacking the smartcard or using a brute force approach to guess the PIN. Another possible strategy is to embed malicious software to compromise the user's login and password. An adversary may also attempt to obtain credential information via the communication channel (sniffing). Finally, an adversary may use various means to engage in communication with the target user, as shown in Figure 1.4. ■ Injection of commands: In this type of attack, the attacker is able to intercept communication between the UT and the IBS. Various schemes can be used to be able to impersonate the valid user and so gain access to the banking system. ■ User credential guessing: It is reported in [HILT06] that brute force attacks against some banking authentication schemes are feasible by sending random usernames and passwords. The attack mechanism is based on distributed zombie personal computers, hosting automated programs for username- or password-based calculation. ■ Security policy violation: For example, violating the bank's security policy in combination with weak access control and logging mechanisms, an employee may cause an internal security incident and expose a customer's account. ■ Use of known authenticated session: This type of attack persuades or forces the user to connect to the IBS with a preset session ID. Once the user authenticates to the server, the attacker may utilize the known session ID to send packets to the IBS, spoofing the user's identity. Figure 1.4 provides a thorough view of the different types of attacks on an Internet banking authentication application. Using this tree as a starting point, security analysts can assess the risk of each attack and, using the design principles outlined in the preceding section, design a comprehensive security facility. [DIMI07] provides a good account of the results of this design effort. 1.8 A Model for Network Security A model for much of what we will be discussing is captured, in very general terms, in Figure 1.5. A message is to be transferred from one party to another across some sort of Internet service. The two parties, who are the principals in this transaction, must cooperate for the exchange to take place. A logical information channel is established by defining a route through the Internet from source to destination and by the cooperative use of communication protocols (e.g., TCP/IP) by the two principals. 40 chapter 1 / Introduction Security aspects come into play when it is necessary or desirable to protect the information transmission from an opponent who may present a threat to confidentiality, authenticity, and so on. All of the techniques for providing security have two components: 1. A security-related transformation on the information to be sent. Examples include the encryption of the message, which scrambles the message so that it is unreadable by the opponent, and the addition of a code based on the contents of the message, which can be used to verify the identity of the sender. 2. Some secret information shared by the two principals and, it is hoped, unknown to the opponent. An example is an encryption key used in conjunction with the transformation to scramble the message before transmission and unscramble it on reception.7 A trusted third party may be needed to achieve secure transmission. For example, a third party may be responsible for distributing the secret information to the two principals while keeping it from any opponent. Or a third party may be needed to arbitrate disputes between the two