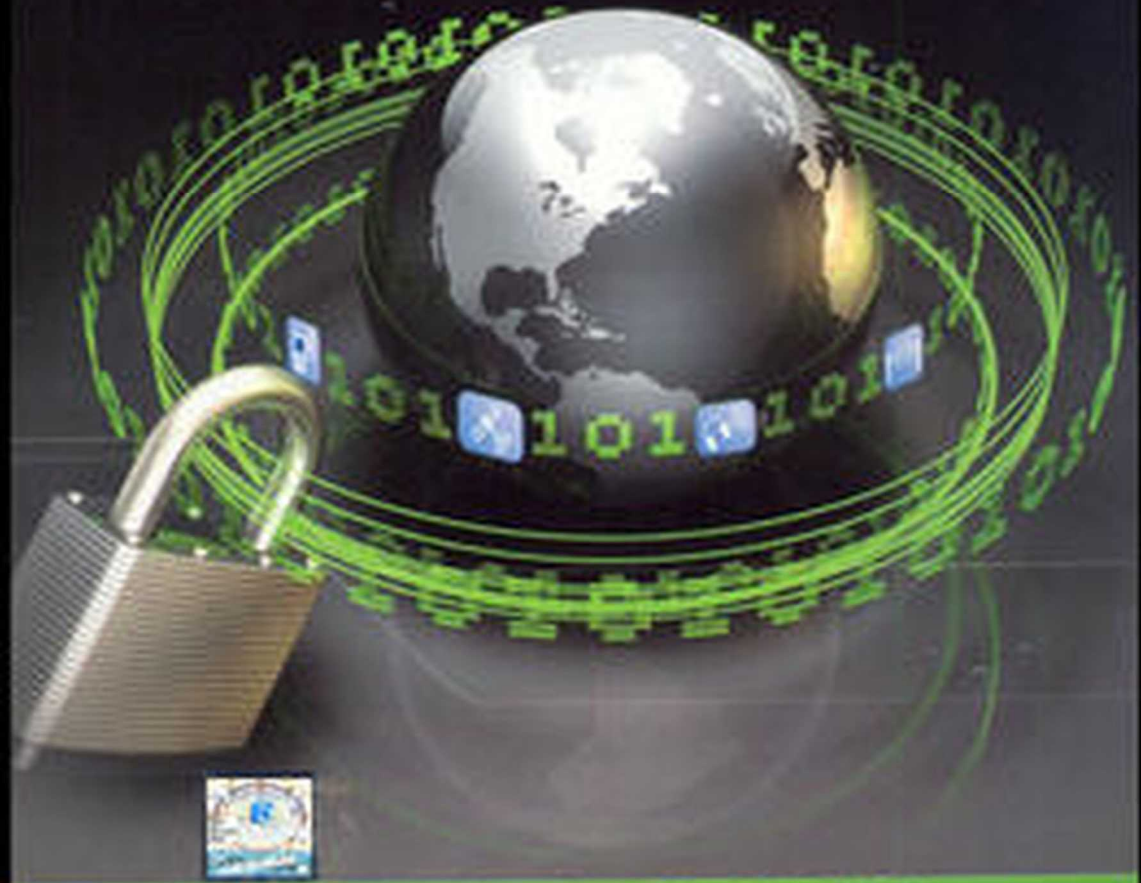
WIRELESS & MOBILE NETWORK SECURITY



PALLAPA VENKATARAM
SATHISH BABU B.

The McGraw-Hill Companies

Wireless and Mobile Network Security

About the Authors

Pallapa Venkataram received his PhD degree in Information Sciences from the University of Sheffield, UK, in 1986. He is currently a Professor of Electrical Communication Engineering with the Indian Institute of Science, Bangalore, India. He is also the Chief Program Executive of the Protocol Engineering Technology (PET) unit where he is involved in research on protocol engineering, wireless networks, network management, computational intelligence applications in communication, mobile computing security and multimedia systems. He is a Fellow of IEE (England), Fellow of IETE(India) and a Senior member of IEEE Computer Society. Dr Pallapa is the holder of a Distinguished Visitor Diploma from the Orrego University, Trujillo, Peru. He has published and presented over 150 papers in international/national journals/conferences, and authored/contributed chapters for 12 books in the area of networking, protocol engineering, and security.

B Sathish Babu received his PhD degree in Electrical Communication Engineering from the Indian Institute of Science, Bangalore, India, in 2010. He did his bachelors and masters degree in Computer Science and Engineering from Bangalore University. He has more than 15 years of teaching experience and is currently a Professor in the Department of Computer Science and Engineering at Siddaganga Institute of Technology, Tumkur. His research interests include intrusion and fraud-detection systems for mobile commerce environment, mobile communication security, and application of cognitive agents in proposing dynamic transaction-based authentication systems for mobile communications. He has published and presented ten papers in international journals/conferences on the subject of mobile security.

The McGraw-Hill Companies

Wireless and Mobile Network Security

Pallapa Venkataram
Professor
Department of Electrical Communication Engineering
Indian Institute of Science, Bangalore, Karnataka

B Sathish Babu
Professor
Department of Computer Science and Engineering
Siddaganga Institute of Technology
Tumkur, Karnataka

Tata McGraw Hill Education Private Limited
NEW DELHI

McGraw-Hill Offices
New Delhi New York St Louis San Francisco Auckland Bogotá Caracas
Kuala Lumpur Lisbon London Madrid Mexico City Milan Montreal
San Juan Santiago Singapore Sydney Tokyo Toronto

Tata McGraw-Hill

Published by the Tata McGraw Hill Education Private Limited,
7 West Patel Nagar, New Delhi 110 008.

Wireless and Mobile Network Security

ISBN (13 digits): 978-0-07-070024-6
ISBN (10 digits): 0-07-070024-9

Managing Director: *Ajay Shukla*

Head—Higher Education Publishing & Marketing: *Vibha Mahajan*
Manager—Sponsoring SEM & Tech Ed: *Shalini Jha*
Asst. Sponsoring Editor: *Surabhi Shukla*
Development Editor: *Surbhi Suman*
Executive—Editorial Services: *Sohini Mukherjee*
Jr Manager—Production: *Anjali Razdan*

Dy Marketing Manager—SEM & Tech Ed: *Biju Ganesan*

General Manager—Production: *Rajender P Ghansela*
Asst General Manager—Production: *B L Dogra*

Typeset at The Composers, 260, C.A. Apt., Paschim Vihar, New Delhi 110 063 and printed at Pushp Print Services, B-39/12 A, Gali No. 1, Arjun Mohalla, Moujpur, Delhi 110 053

Cover Design: K Anoop

Cover Printer: Rashtriya Printers

RZXYCRAZRQBDQ

In memory of

Dr Bharathi Pallapa

Contents

Preface

Overview

In today's world of high mobility, there is a growing need for people to communicate with each other and have timely access to information regardless of the location of individuals or information. This need is supported by the advances in the technologies of networking, wireless communications, and portable computing devices with reduction in the physical size of computers, leading to the rapid development in mobile communication infrastructure. Hence, mobile and wireless networks present many challenges to hardware, software, network and application designers and implementers. One of the biggest challenges is to provide a secure mobile environment. Security plays a more important role in mobile communication systems than in systems that use wired communication. This is mainly because of the ubiquitous nature of the wireless medium that makes it more susceptible to security attacks than wired communications.

There are various technologies for wireless networks, such as Cellular, Bluetooth, Ultra Wideband (UWB), Wi-Fi, Pervasive, and hybrid wireless networks. All of these wireless systems are vulnerable to security issues which are either general or specific to the networks. The required level of security depends on the design and type of the wireless network. For example, public Wi-Fi hotspots are typically unsecure from a user perspective because there is no data encryption, something that can allow a hacker to monitor your organization files, emails, and passwords flying unguarded over the airwaves. Public users may need to offer credentials to access Internet services, but the immediate wireless LAN is still wide open to hackers.

Objective

The available books on wireless network security mainly focus on security issues, challenges, attacks, and solutions pertaining to Wi-Fi networks, cellular networks, ad-hoc networks, home networking, ubiquitous networks, and so on. But they lack in providing a dimension on how security schemes could be linked with categories of applications, and the quantum of security required for their transaction running over these networks.

The proposed book comprehensively deals with various security issues, challenges, attacks, protocols, and available security solutions for wireless technologies. However its main focus is to bring together available research literature, and techniques about making the security dynamic, and adaptive to applications. The

security algorithms/schemes for encryption/decryption, authentication, integrity maintenance, non-repudiation, and so on, are not required to be statically applied for all applications. These algorithms can be dynamically selected depending on application characteristics as well as the environment of execution. The application characteristics are mainly concerned about deploying required type, and level of security, privacy, and trust for successful execution of an application. The environment can include profiles of communicating parties, network characteristics, device characteristics, content sensitivity, context of execution, and so on.

Target Audience

This book can be used as text/reference for the following subjects:

System and Network Security; Information Privacy and Computer Security; Telecommunications, Network and Internet Security; Special Topics in Information Security; Wireless Networks and Security; Secure Communications; Ubiquitous Computing; Network and Distributed Systems Security; Mobile Communication Networks; Information Security.

These subjects are offered as compulsory/elective papers in BTech final year and MTech 1st/2nd Semester in CSE/IT/Telecom/Information Security streams.

Also, graduate students preparing for PhD or Masters degree in network communication or information systems and researchers in information security or communication networks who require an introduction to the issues of mobile security will find this book to be very useful.

Salient Features

- Provides information on latest researches and developments in the field of security related to various networking technologies
- Hands-on approach for implementing different techniques linked with network security
- Well-organized presentation of security issues, attacks, solutions and challenges
- Covers conventional as well as updated attacks pertaining to various wireless networks
- Detailed coverage on Ubiquitous and Integrated Network Security

Chapter Organisation

The book is divided into 8 chapters. **Chapter 1** is on *Security Issues in Mobile Communications*. This chapter deals with mobile communication history, wired vs wireless security, security issues in wireless and mobile communications, security requirements, security for mobile applications, and so on.

Chapter 2 on *Security at Device, Server, and Network Levels* describes mobile device security requirements, mobile wireless network level security, and server level security issues.

Chapter 3 discusses *Application Level Security in Wireless Networks.* This chapter highlights security issues involved in wireless applications independent of networks and devices. It also details wireless security issues; wireless threats;

targeted attackers; vulnerabilities and attack methods; today's wireless security; first Generation WLANS; second generation WLANs; and 802.1X.

Chapter 4 is on *Application Level Security in Cellular Networks*. This chapter offers in-depth explanations on generation of cellular networks; 3G-UMTS architecture; security issues in cellular networks; limitations of cellular networks; attacks on cellular networks; security mechanisms in 3G-UMTS; Wireless Application Protocol (WAP); GSM Security; and GSM Security/Authentication/Access Control Features.

Chapter 5 on *Application Level Security in MANETs* is dedicated to security in ad-hoc networks; attacks; physical layer attacks; link layer attacks; network layer attacks; transport layer attacks; application layer attacks; security challenges; passive eavesdropping; active interference; internal threats; selfish nodes; denial of service attacks; attacking neighbor sensing protocols; exploiting route maintenance; attacks on protocol specific optimizations; trust based authentication; and threshold cryptography in mobile ad-hoc networks.

Chapter 6 is on *Application Level Security in Ubiquitous Networks*. This chapter discusses the Ubiquitous Computing (UC) vision; need of novel security schemes for UC; security issues; security challenges for UC; security characteristics of ubiquitous networks; privacy and trust characteristics; attacks; security provisions; trusted computing base; network anomaly detection; role-based access control system; and local proof of secret.

Chapter 7 is on *Application Level Security in Heterogeneous Wireless Networks*. This chapter deals with network architectures like iCAR, SOPRANO and Roofnet. It also explains in detail some security solutions for heterogeneous wireless networks.

Finally, **Chapter 8** which is on *Security for Mobile Commerce Application* provides coverage on M-commerce initiatives; M-commerce security; security challenges; types of attacks on mobile E-commerce; a secure M-commerce model based on wireless local area network; secure mobile commerce; an asymmetric authentication protocol for M-commerce applications; a personal authentication scheme using mobile technology; and transaction-based authentication schemes.

Over 200 *Review Questions* are present in the book to help students refresh their memory and apply the concepts to real-life examples.

Web Supplements

The book is well supported by an exhaustive website which can be accessed at http://www.mhhe.com/venkataram/wmns which includes the following:

For Instructors

- PowerPoint Slides (chapterwise)

For Students

- Additional practice questions
- Chapterwise links to important websites and important text materials

Acknowledgements

First of all, we would like to express our thanks to all those reviewers whose inputs proved valuable while shaping this text. Their names are given below:

Dinesh Kumar Tyagi
Birla Institute of Technology and Science (BITS), Pilani, Rajasthan
Karm Veer Arya
Atal Bihari Vajpayee Indian Institute of Information Technology and Management (ABV-IIITM), Gwalior, Madhya Pradesh
C Vijay Kumar
Dhirubhai Ambani Institute of Information and Communication (DAIICT), Gandhinagar, Gujarat
Jigisha Patel
Sardar Vallabbhai National Institute of Technology (SVNIT), Surat, Gujarat
M P Sebastian
Indian Institute of Management, Kozhikode, Kerala

The Tata McGraw-Hill team of Shalini Jha, Surabhi Shukla, Surbhi Suman, Sohini Mukherjee and Anjali Razdan deserves a special note of appreciation for the enthusiasm they showed while handling the text in all its stages—development, editing, and proofreading.

Finally, we are indebted to our families for their love, patience and wholehearted support in everything that we do. Without their contribution, this milestone would not be achieved.

We have taken care to present the concepts in a simple manner in this book and hope that the teaching and student community will wholeheartedly welcome and appreciate it. The readers should feel free to convey their criticism and suggestions for further enhancement at the below mentioned publisher's e-mail id.

Pallapa Venkatram
B Sathish Babu

Publisher's Note

Tata McGraw Hill Education looks forward to receiving your views, feedback and suggestions for improvement of the book. These may be emailed at tmh.csefeedback@gmail.com. Please mention the book title and author's name in the subject line. Please report any piracy spotted by you as well!!

List of Important Abbreviation

AAA	Authentication, Authorization, Accounting
API	Application Program Interface
BS	Base Station
DoS	Denial of Service
FH	Foreign Host
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communications
IPsec	Internet Protocol Security
IDS	Intrusion Detection System
MH	Mobile Host
MU	Mobile User
MitM	Man in the Middle
NIC	Network Interface Card
OTP	One Time Password
PDA	Personal Digital Assistant
PIN	Personal Identification Number
QoS	Quality of Service
SSID	Service Set Identifier
TBAS	Transaction Based Authentication Selection
TBSS	Transaction Based Security Selection
TSL	Transaction Sensitivity Level
VPN	Virtual Private Network
WEP	Wired Equivalent Privacy
WG	Wireless Gateway
WLAN	Wireless Local Area Network
WPA	Wi-Fi Protected Access

Security Issues in Mobile Communications

1

OBJECTIVES

- To understand briefly the history of mobile communications.
- To study salient features of various wireless and mobile technologies.
- To study the need of securing wireless networks.
- To understand various security issues in wireless and mobile networks.
- To know the limitations of some security technologies available at hardware level.
- To introduce the need of application-level security.
- To give glimpse of some of the security schemes proposed at the application level.

In a world of increasing mobility, there is a growing need for people to communicate with each other and have timely access to information regardless of the location of the individuals or the information. A phone call placed from a commuter train may close a business deal, remote access to medical records by a paramedic may save a life, or a request for reconnaissance updates by a soldier with a handheld device may affect the outcome of a battle. Each of these instances of mobile communications poses an engineering challenge that can be met only with an efficient, reliable, wireless communication network. The demand for wireless communication systems of increasing sophistication and ubiquity has led to the need for a better understanding of fundamental issues in communication theory and electromagnetics and their implications for the design of highly capable wireless systems.

Mobile Communications have been an essential part of communication in the last century. Early adopters of wireless technology primarily have been the military, emergency services, and law enforcement organisations. Scenes from World War II movies, for example, show soldiers equipped with wireless communication equipment being carried in backpacks and vehicles. Common commercial applications are product distribution and marketing, railway and shipping cargos, express package deliveries and trucking industry.

1.1 MOBILE COMMUNICATION HISTORY

There are three lines of evolution of wireless mobile network architectures that have been defined to support wireless technologies and the related services:

1. Those of interpersonal communication, based on the use of electromagnetic bands are generally subject to regulation and for this reason are highly estimated.
2. Those of communication among information technology and network devices, that substitutes the shared physical media (the wire) with technologies using part of the electromagnetic spectrum. The bandwidth involved can be both under or without the control of regulation authority, according to the specific country considered.
3. Those generally used for man-machine interaction, mostly for sending a command or an exchange of information.

1.1.1 Cellular Networks

Starting from 1928 when the first mobile system went into operation in Detroit with the Police Department, began the process of "communication mobility" from the commercial point of view. The so-called First Generation moved its first steps in early 1970s with AMPS (Advanced Mobile Phone System). At that time, FCC (Federal Communications Commission) allocated spectrum space for cellular systems as a result of a process of rationalising the deployment of several nonstandardised technologies, based on broadcasting techniques [such as IMTS (Intelligent Multi-mode Transit System)]. At the same time, on the opposite side of Atlantic Ocean, the same process was taking place for a quite limited environment of subscribers, using broadcasting telephone systems, not based on cellular technique. All those systems were "analog" while some encryption functionality was added as a feature only for privileged communications.

The first cellular mobile telephone services opened in 1981 by NMT (Nordic Mobile Telephone) was 450 system. It was an analog system, based on 450MHz

band, working in the North-European countries (Sweden, Norway, Denmark and Finland). Contemporary Siemens introduced in Germany the C-Netz 450 system (similar to the previous one). With a delay of a couple of years, due to administrative obstacles more than technical aspects, AMPS (Advanced Mobile Phone Service) went into service. In the second half of 80's, most European countries started the cellular environment adopting analog systems in the 800MHz band (probably the most exploitable for the purpose).

FDMA (Frequency Division Multiple Access) was the king of technology for the radio spectrum exploitation by systems. All the systems foreseen geographically distributed radio base stations [RBS (Robbed-bit Signalling)], with one or more sectors, typically submitted to an "intelligent equipment" which performs all the basic features to provide services:

- radio controller for the RBSs,
- switching functions for communications,
- mobility management (limited to the hard hand-over management),
- subscribers database management,
- signaling management, and
- interconnections functionalities.

While for the network architectures there were no significant differences among the systems, in the case of radio access side, some negligible differences occur, in spite of the radio spectrum allocation like the channel bandwidth.

GSM

GSM (Fig.1.1) is a classic digital mobile system built on a radio interface of TDM type. Its basis was within the CEPT (Conference of European Post and Telecommunications) with the constitution of the "Groupe Special Mobile". That was the origin of the name 'GSM', today changed to "Global System for Mobile communications".

Basic GSM was designed for voice communications at 9 Kbit/s, but data capability was with the CSD service (Circuit Switched Data at a maximum of 9.6 Kbit/s), the SMS (Short Message Service, the well-known packet-oriented messaging service), the ability to support FAX, and minimal broadcasting capabilities (the broadcast of cell information). But substantially remains linked to the circuit switched genesis of the system. ETSI (European Telecommunications Standard Institute) carried on the extension of GSM with the inclusion of the HCSD (High speed CSD) capable to reach 38.4 Kbit/s. In parallel ETSI put the basis for a further evolution of the GSM radio interface defining EDGE (Enhanced Data rates for GSM) Evolution, based on the introduction of a new modulation and consequent increasing of the GSM/CSD/GPRS data rates. ETSI also started the work on the W-CDMA based UMTS (Universal

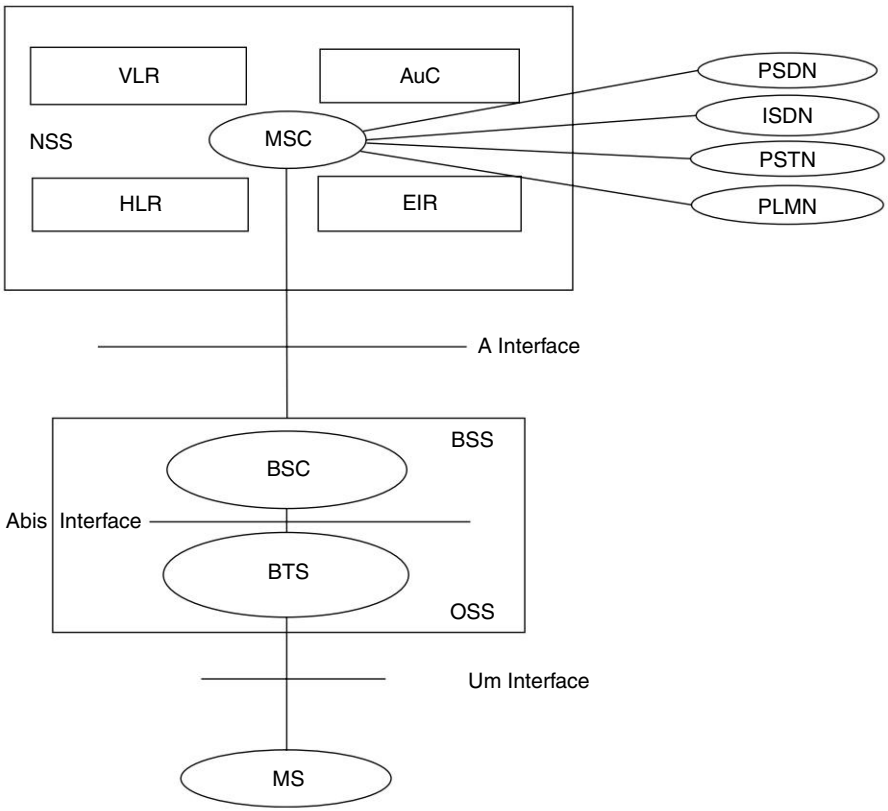


Fig. 1.1 *GSM Architecture*

Mobile Telecommunication System), being replaced in this task by 3GPP (3rd Generation Partnership Project), a world association of the major standardisation organisation related to the mobile telecommunication business. The 3GPP is responsible for all the GSM and UMTS specifications, and the result is a single system, that associates to the basement of a common core network, the GSM/GPRS radio access (and its EDGE enhancements) and the UMTS radio access. UMTS is, therefore, entitled to own the full inheritance of GSM, of which represents an evolution and integration.

GPRS

GPRS (General Packet Radio Service) shown in Fig. 1.2, is the first packet-oriented extension of GSM capable of managing a minimum of bit rate. The radio interface is the same TDMA scheme used for GSM, but the use of the slot is significantly different. Each slot is temporarily assigned to a terminal just for the time of the transmission, allowing the statistical multiplexing of

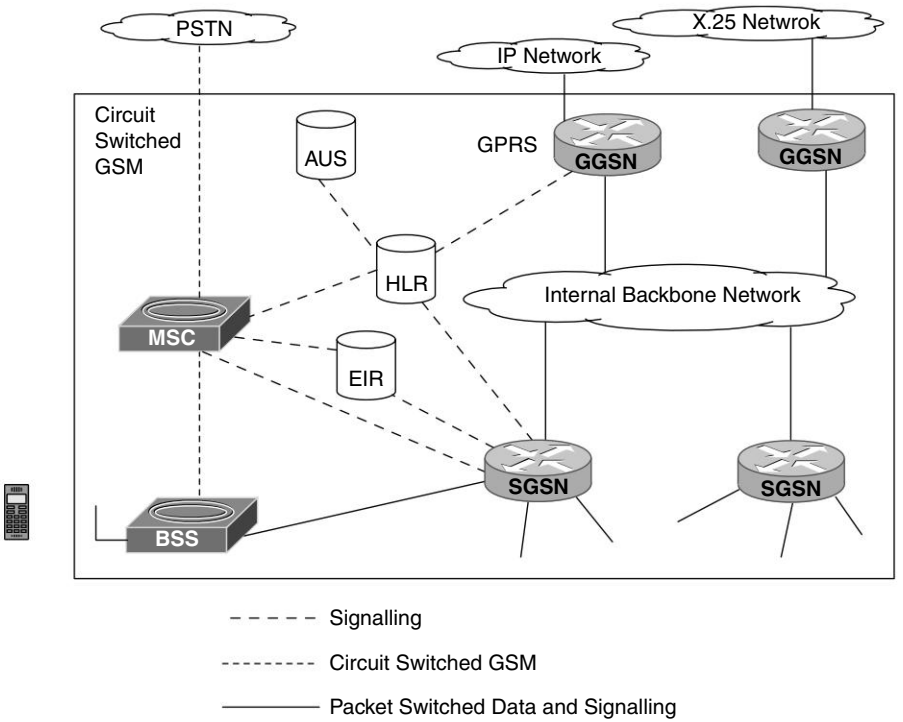


Fig. 1.2 *GPRS Architecture*

the slots among different terminals. GPRS foresees the allocation of multiple slots to the same terminal, increasing significantly the maximum bit rate. Typical numbers are 4+2 (4 for downlink and 2 for uplink) to privilege download and 2+2 for symmetric communications. GPRS also allows reducing the error protection, incrementing the net bit rate of each slot. This is related to different considerations—GSM was presumably designed to excessively protect against errors with respect to the radio condition obtained in present networks. A GPRS user is less mobile than a GSM one, so he/she has to face more stable radio conditions. Packet transmission implements a different and shorter interleaving, so the burst errors are less distributed on the transmission, being recovered by RLC re-transmission for wrongly received packets. This is allowed by the non-real time characteristics of GPRS. Depending on the radio condition, four levels of protection are possible, namely Coding Scheme 1, 2, 3, and 4.

EDGE and UMTS

EDGE (Enhanced Data Rate for GSM Evolution) is the solution designed for the support of higher bit rates on top of GPRS. It is based on the change of the

modulation from GSMK to 8PSK. The GSM 8-slot TDM frame is maintained but the number of bit transmissible in each slot is increased. EDGE is fully compatible with GPRS in the sense that as a fallback it could work as GMSK. UMTS is essentially based on the addition to the GSM/GPRS systems of a new type of access network, based on a completely different radio technique, the CDMA one. The UMTS, GPRS, EDGE and GSM can be a single system with a single CS/PS core network capable to serve all the type of accesses.

1.1.2 Wireless Data Networks

Wireless data networks exist in so many types and varieties that they are difficult to categorise and compare. However, we can classify wireless data networks in three categories—(1) Wireless Personal Area Network (WPAN), the average coverage is about 10 meters and typical applications are interconnection of personal devices, typically in point-to-point configuration. (2) Wireless Local Area Network (WLAN), the average coverage is about 100 meters and typical applications are PCs and servers interconnection, typically in point-to-multipoint configuration. (3) Wireless Metropolitan Area Network (WMAN), the average coverage is about 10 kilometers and typical applications are interconnection of buildings (houses, offices), typically in point-to-multipoint configuration.

Wireless LAN

Frequency hopping is a radio transmission technology where the signal is divided in multiple parts and then sent across the air in a random pattern of jumping, or "hopping" frequencies. When transmitting data, these "multiple parts" are data packets. The hopping pattern can be anywhere from several times per second to several thousand times per second. The Secret Communications System designed for World War II was the beginning of Wireless LAN. A Wireless LAN or WLAN can be considered as a cabled Local Area Network (LAN) shown in Fig. 1.3, where the physical medium is substituted by a radio frequency. Instead of towers, Wireless LANs use a base station to talk to their devices.

When the Wireless Ethernet Compatibility Alliance (WECA) launched the Wi-Fi certification, it was used to certify the interoperability among different vendors. WECA's mission is to certify interoperability of IEEE 802.11 products and to promote Wi-Fi as the global wireless LAN standard across all market segments. Even if a WLAN does not (and probably will never) offer the same performance as a cabled LAN, which can now almost reach the Gigabit, it can give a certain number of benefits—mobility, reduced installation costs, and high integration among different mobile devices. The fundamental

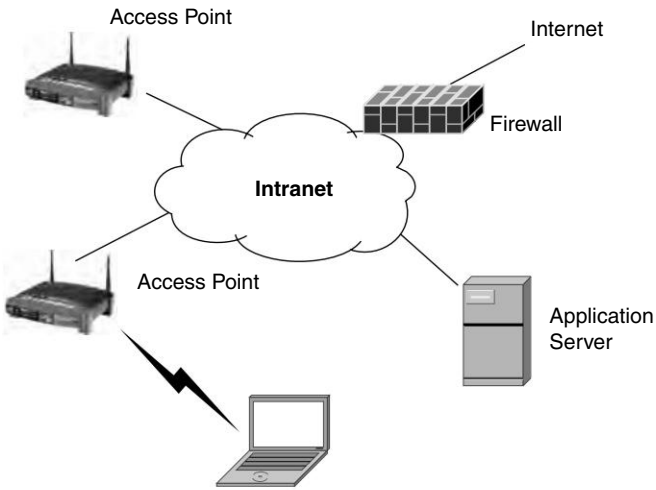


Fig. 1.3 *WLAN Architecture*

component of a wireless LAN is the Basic Service Set (or BSS). The BSS contains a single Access Point (AP) and a certain number of peripheral nodes. To establish the wireless link, each node and an AP uses a small RF radio, which includes an antenna, a transceiver, a modem and some signal processing electronics. All the nodes communicate with each other by using an AP as a bridge to relay the signals.

The first role of the AP is to get all the nodes in the BSS to talk to each other. The WLAN frequency bands are license exempt, which means that the spectrum is available to any operator wishing to deploy a WLAN infrastructure without any need to deal with the regulator. A wireless LAN can be set up without an AP in which all nodes communicate directly with each other. This peer-to-peer arrangement is sometimes referred to as an ad hoc network (See Fig. 1.4). As an example, a BSS may be a single floor in a big office building. It is possible to interconnect two BSSs through an Extended Service Set (ESS) by getting their individual APs to talk to each other. The 802.11 standard is really a series of standards updates reflecting improvement in technology and/or the employment of additional spectrum. There are three IEEE approved standards—802.11b, 802.11a, and 802.11g that exist. The 802.11b is the most popular and it works in the 2.4 GHz frequency spectrum, which is shared with other technologies, such as cordless phones and microwave ovens. The range is about 30–50 meters indoor. The maximum theoretical speed is 11 Mbps, while the real throughput is about 4–6 Mbps, due to the transmission protocol overheads. WLAN systems based on 802.11a use the 5 GHz frequency spectrum. The use of this frequency makes the IEEE 802.11a

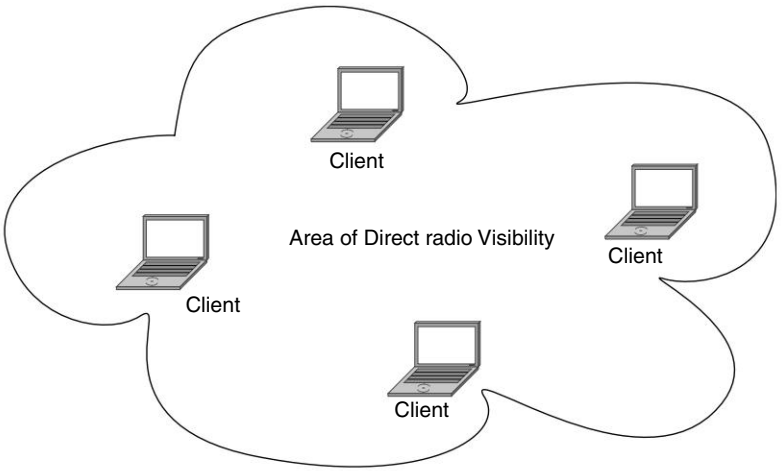


Fig. 1.4 *Ad hoc Network Architecture*

incompatible with the 802.11b and it is not widely used in Europe, due to specific restrictions in the use of this frequency by private and commercial users. The maximum theoretical speed is 54 Mbps, while the real throughput is about 22 Mbps, but the range is only no more than 25 meters. The 802.11g is the most recent standard, it operates in the same 2.5 GHz frequency spectrum as 802.11b, which makes it compatible with the previous generation products. The maximum theoretical speed is 54 Mbps, as for 802.11a, while the real throughput is about 15–20 Mbps. The range is about 30–50 meters indoor. The three standards "a", "b" and "g" can coexist in the same product.

Mobile Ad Hoc Networks

Mobile Ad Hoc Network (MANET) shown in Fig. 1.4, can be considered as a collection of autonomous wireless nodes or terminals that communicate with each other by forming a multihop radio network and maintaining connectivity in a decentralised manner. Ad hoc here is to be intended in the sense of "pragmatic", "dynamically constructed" or "transient". A mobile ad hoc network can also be considered a "mobile routing infrastructure". Ad hoc networks use no fixed network infrastructure or access point, like in the case of WLAN. Nodes form a network of highly mobile platforms that are not dependent on pre-existing or fixed communications infrastructure.

An ad hoc network node has routing functionalities and it can relay traffic to/from other nodes—this is why a single node is considered both as terminal node and part of the network. This is a very different approach compared to other mobile networks, where terminal and network functionalities are implemented in totally different network elements. Nodes can communicate

following a peer-to-peer paradigm, meaning that each node can communicate with all the other nodes.

The ad hoc networks can be established using GSM, ODMA, WLAN IEEE 802.11 (802.11a, 802.11b, 802.11g), IEEE 802.15 (Bluetooth), and WMAN IEEE 802.16a. The network topology is in general dynamic, because the connectivity among the nodes may vary with time due to node departures, new node arrivals, and the possibility of having mobile nodes. Hence, there is a need for efficient routing protocols to allow the nodes to communicate over multihop paths consisting of possibly several links in a way that does not use any more of the network "resources" than necessary. Significant examples of use of ad hoc networks include establishing survivable, efficient, and dynamic communication for emergency/rescue operations, disaster relief efforts, and military networks. Such network scenarios cannot rely on centralised and organised connectivity, and can be conceived as applications of ad hoc networks.

Wireless PAN

A WPAN (Wireless Personal Area Network) shown in Fig. 1.5, is a personal area network for interconnecting devices centered around an individual person's workspace in which the connections are wireless. Typically, a wireless personal area network uses some technology that permits communication within about 10 meters, in other words, a very short range. One such technology is Bluetooth, which was used as the basis for a new standard, IEEE 802.15, a WPAN could serve to interconnect all the ordinary computing and communicating devices that many people have on their desk or carry with them today, or it could serve a more specialised purpose such as allowing the surgeon and other team members to communicate during an operation.

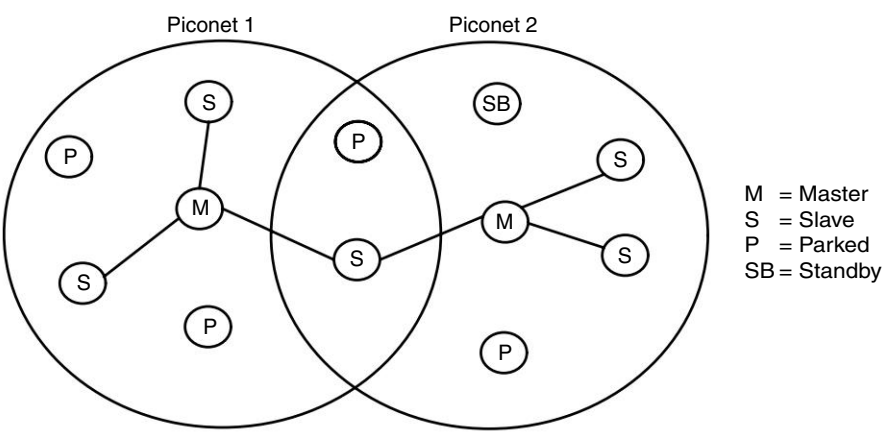


Fig. 1.5 *WPAN Arcthitecture*

A key concept in WPAN technology is known as plug and use. In the ideal scenario, when any two WPAN-equipped devices come into close proximity (within several meters of each other) or within a few kilometres of a central server, they can communicate as if connected by a cable. Another important feature is the ability of each device to lock out other devices selectively, preventing needless interference or unauthorised access to information. The technology for WPANs is in its infancy and is undergoing rapid development. Proposed operating frequencies are around 2.4 GHz in digital modes. The objective is to facilitate seamless operation among home or business devices and systems. Every device in a WPAN will be able to plug in to any other device in the same WPAN, provided they are within physical range of one another. In addition, WPANs worldwide will be interconnected.

Bluetooth is a specification for short-range wireless communication which has emerged as the leading candidate for an international standard for PAN. The bluetooth technology implements a simple low cost and low power system which allow devices to communicate in a "piconet". It supports many simultaneous and private connections, with hundreds of private piconets within range of each other. It supports both voice and data. It is very low power and compact to support the small portable devices into which the technology will be integrated, such as mobile computer, the mobile phone, and small personal digital assistants. The technology must be secure as a cable, which means to support application/link-layer authorisation, authentication, and encryption.

1.1.3 Integrated Networks

Figure 1.6 shows an IP-based integrated network architecture where access is through a variety of wireless technologies, with the intelligence residing in the access and the backbone primarily providing the packet transport. The various functions, features, and capabilities of the network and the services will be supported by specialised servers potentially attached anywhere in the IP network. The new architecture will need to be based on the paradigm that every user is potentially mobile, and the network has to be able to carry billions and billions of micro flows and mega flows concurrently, some requiring strict QoS and others requiring only best effort service. Heterogeneous networks with varying bandwidth capabilities give rise to heterogeneous traffic profiles. Whereas in the near-term, a vast number of applications will be transaction-oriented, in the long-term, multimedia traffic (e.g., MP3 audio, video streaming, gaming—to name just a few), will pose its own challenges.

The operating environment of networks has changed drastically. Many nations are continuously rationalising their communication capabilities, making them more flexible, and suited to respond to the need for fast deployments. In this con-

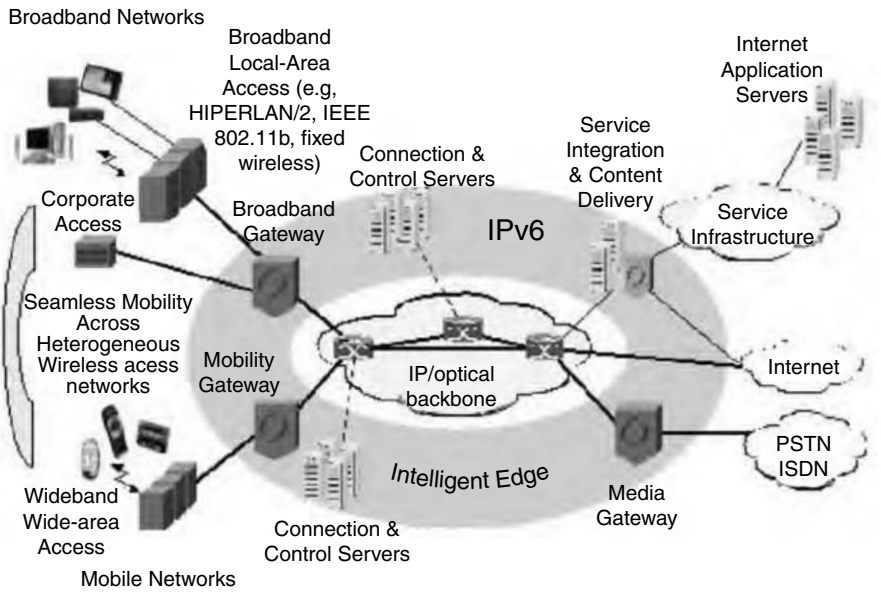


Fig. 1.6 *An example of Integrated Network*

text, several networks with different technical characteristics are employed. Satisfying the requirement for secure multi-services in such environments has become an even greater challenge than in the past. Many technical and industrial contributions favour an IP-based concept for an integrated network, where all services are delivered above the IP-layer and all data is encapsulated within IP-packets.

Users are prone to roaming across multiple geographical and organisational adjacent 'domains'. The term 'domain' refers to an unique local access network with characteristic security requirements and attacker capabilities, autonomous management and enforcement of different (security) policies. Additionally, in a wireless network environment, nodes are highly mobile under dynamic network conditions. Thus, in an integrated network environment, mobility management is needed to ensure that nodes can be located quickly and packet delivery operates properly in the presence of mobility of nodes, and networks without affecting the ongoing session. Thus, seamless hand-over is an important feature, protecting the user from taking notice during any roaming between adjacent domains. Particularly, the current user's security and multi-service environment are kept invariant and thus, enabling an automatic and high-mobility facility, e.g., for mobile military units or even whole headquarters transferred to the theater.

Any network application may communicate across multiple realms. This puts an additional burden on the security infrastructure of integrated networks

requiring it to be able to adapt to environments with scarce resources in terms of power, bandwidth and jamming-freeness and evolve once more resources become available. Other resources are inherent in the device, e.g., computing performance and size.

1.2 SECURITY—WIRED vs WIRELESS

Security plays an important role in wireless communication systems than in systems that use wired communication (like PSTN, LAN, etc). This is mainly because of the ubiquitous nature of the wireless medium that makes it more susceptible to security attacks than wired communications. In the wireless medium, anyone can listen to whatever is being sent over the network. Also, the presence of communication does not uniquely identify the originator (as it does in the case of a pair of coaxial cable or optical fibers). To make things worse, any tapping or eaves-dropping cannot even be detected in a medium as ubiquitous as the wireless medium. Thus, security plays a vital role for the successful operation of a mobile communication system.

The early first generation and the second generation systems did not have robust security features and thus suffered a large number of security attacks by hackers. Infact, eavesdropping of the 1G AMPS system can be accomplished using a very simple police-scanner which can be built without much effort. Even, in the relatively complicated 2G GSM systems, eavesdropping was not much of a trouble as cellular scanners that work on the 2G systems were built in spite of a ban imposed by the Electronic Communications Privacy Act passed by the FCC during that time. The only way to prevent such unauthorised use of communication resources and to protect user privacy, is the use of cryptographic techniques to provide security, authentication and access-control.

1.3 SECURITY ISSUES IN WIRELESS AND MOBILE COMMUNICATIONS

Communication security is often described in terms of confidentiality, integrity, authentication and nonrepudiation of transmitted data. These security services are in turn implemented by various mechanisms, that are usually cryptographic in nature. In addition, there is a confidentiality of traffic (i.e., whether or not communication is taking place), of location (where the communicating parties are located) and of the communicating parties addresses, all of which are important for privacy. A casual level of security is usually provided implicitly

even without taking any extra measures. For example, in order to eavesdrop on a particular person's mobile phone conversations, the eavesdropper has to be located in physical proximity to the person and carry special radio equipment which in itself represents a certain level of protection. Casual authentication, between mobile phone users, is indirectly provided by the calling and called party numbers. In case of voice telephony, authentication results from recognising the other person's voice.

Cryptography

Cryptography, on the other hand, gives the possibility of designing strong security services but often creates inconveniences while using the application. The use of cryptography, therefore, makes most sense in case of sensitive applications. When strong cryptographic security mechanisms are in place, the remaining vulnerabilities are usually due to poor management and operation and not by weaknesses in the cryptographic algorithms themselves.

Confidentiality

Confidentiality of transmitted data can be provided by encrypting the information flow between the communicating parties, and the encryption can take place end-to-end between the communicating parties or, alternatively, on separate legs in the communication path. In GSM networks for example, only the radio link between the mobile terminal and the base station is encrypted, whereas, the rest of the network transmits data in clear-text. Radio link confidentiality in GSM is totally transparent from the users point of view. Mechanisms for implementing confidentiality of traffic, location and addresses will depend on the technology used in a particular mobile network.

Authentication

Authentication of transmitted data is an asymmetric service, meaning for example that when A and B are communicating, the authentication of A data by B is independent from the authentication of B data by A. The types of authentication available will depend on the security protocol used. In the Internet for example, SSL (Secured Socket Layer) allows encryption with four different authentication options—1) server authentication, 2) client authentication, or 3) both server and client authentication or 4) no authentication, i.e. providing confidentiality only.

Non-repudiation

Non-repudiation is similar to authentication, but it is an asymmetric security service. A simple way to describe the difference between authentication and

non-repudiation is that with authentication the recipient himself is confident about the origin of a message but would not necessarily be able to convince anybody else about it, whereas for non-repudiation the recipient is also able to convince third parties. Digital signature is the mechanism used for non-repudiation. Cryptographically seen, a message authentication code and non-repudiation code can be identical, and the difference between the two services might only depend on the key distribution. In general, if a signature verification key has been certified by a trusted third party, the corresponding digital signature will provide nonrepudiation, whereas it can only provide authentication if the key has simply been exchanged between the two communicating parties. Different parties will have different interests regarding authentication and non-repudiation services. Network operators are interested in authenticating the users for billing purposes and to avoid fraud. Users and content service providers are interested in authenticating each other and might also be interested in authenticating the network service provider. How and where in the network, authentication services are implemented, will depend on the technology used and the business models involved.

A secure environment is a must for any network to operate successfully. The design and operating characteristics of the mobile networks make them susceptible to various types of attacks. Following are the various characteristics of mobile networks, which makes the security vulnerable in mobile communications:

1. **It is an open medium:** Wireless links renders the network susceptible to attacks, the attacks are ranging from passive eavesdropping to active interfering. The damages can include leaking secret information, message contamination and node impersonation.
2. **Decision-making is distributed:** Decision-making in mobile computing environment is sometimes decentralised, and some wireless network algorithms rely on the cooperative participation of all nodes and the infrastructure.
3. **Dynamically changing topology:** Disconnected operations and location-dependent operations results in unique communication patterns for wireless networks.
4. **Absence of central authority:** Tracking down a particular mobile node in a global scale network cannot be done easily.
5. **Lack of clear line of defence:** In terms of lack of standard security softwares and security protocols.

1.4 SECURITY REQUIREMENTS IN WIRELESS AND MOBILE COMMUNICATIONS

The various works in the field of wireless communication have an opinion that, protecting wireless communications and the applications that use this medium will be more difficult than securing desktop computer applications. The mobile device, by its very nature of operation, is vulnerable to unauthorised use and theft, the radio interface is open to attacks from all directions and the heterogeneous operation of mobile service network may lead to security gaps at any stage. The security approaches for the wireless environments are either proactive or reactive. The proactive approaches try to prevent the attacker from exploring the possible attacks, and reactive approaches perform posterior analysis of the attacks and come out with suitable remedies for future use. An effective security solution is needed to incorporate both the approaches. Following are some of the requirements while designing the security solutions for mobile networks:

Support of Roaming

Most mobile communication systems support "roaming" of users, wherein the user is provided service even if he/she moves into a region handled by a different service provider or a different network of the same service provider. Thus, there is requirement in the network for authenticating mobile users who roam into its area. The main problem is that the subscriber related information, that is useful for authentication, is present only in the home network of the user and is generally not accessible by the visited (or serving) network. Thus, there must be a method by which a subset of handset credentials are supplied to the serving network that is enough to authenticate the user. A complete disclosure of handset credentials may result in a security compromise.

Integrity Protection of Data

In addition to securing the data (signalling or user), there must be a provision in the network and the handset to "detect" or "verify" whether the data it receives has been altered or not. This property is called 'data integrity'. Signalling and user data that are considered to be sensitive must be protected using this method.

Requirements for Preventing Theft of Service or Equipment

Theft of service and equipment is a very serious problem in mobile personal communications. The network subsystem doesn't care whether a call has

originated from a legitimate or from a stolen terminal as long as it bills the call to the correct account. There are two kinds of thefts that could be possible here, the theft of personal equipment and theft of the services offered by the service provider. The Cryptographic systems must be designed to make the reuse of stolen terminals difficult (or even impossible). Further, it should block the theft of service by techniques such as "cloning" of mobile phones, which can be done both by the hackers using stolen equipment, as well as legitimate users.

Cloning and Clone Resistant Design

"Cloning" is a serious problem in contemporary mobile communication systems. Cloning refers to the ability of an imposter to determine information about a personal terminal and "clone" (or create a duplicate copy) of that personal terminal using the information collected. This kind of fraud can be easily accomplished by legitimate users of the network themselves, since they have all the information they need to clone their own personal terminal. In this way, multiple users can use one account by cloning personal equipments. This is where equipment cloning causes a lot of problems. The cryptographic system for the mobile network must incorporate some kind of "clone-resistant" design. The most obvious requirement for this design is the security of personal equipment information. This security must be provided for the air-interface, the network databases and the network interconnections so that personal equipment information is secure from imposters. Cloning also typically occurs in the installation and repair stages of the mobile equipment use and must be prevented.

User IDs and Provisioning

Since, the handset can be used by anyone, it is necessary to identify the correct person for billing purposes (i.e., the user must be identified to the network). This may take the form of a smart-card or a plug-in that plugs into a handset and is unique to a user. The process by which the network identifies the user is the authentication process.

Equipment Identifiers

In systems where the account information is separated (both logically and physically) from the handset (which is the case in all current mobile communication systems), stolen personal equipment and its resale could be an attractive and lucrative business. To avoid this, all personal equipment must have a unique identification information that reduces the potential of stolen

equipment to be re-used. This may take the form of "tamper-resistant" identifiers permanently plugged into the handsets.

Requirements on Power/Bandwidth/Computational Usage

Cryptographic systems and algorithms have varying computational complexities and some of them might produce encrypted outputs that occupy too many bits of information more than the original. Since the mobile communication channels are limited by power, bandwidth and battery life of handsets, the cryptographic system must take into consideration the following requirements.

- The algorithm used must be of limited computational complexity (this applies to handsets limited by battery-power).
- The encrypted outputs of the cryptographic algorithm must be of limited size, so that it does not add much overhead to the system.
- The number of transactions between the mobile user and the network (e.g., in the case of authentication) must be as minimal as possible to conserve bandwidth and power.

1.5 SECURITY FOR MOBILE APPLICATIONS

The network security's first aim is network protection—preventing the unauthorised reading or modification of data, or the unauthorised use of resources, including in the form of denial of service. Traditionally, protection has been based on the idea of protecting users from each other. The operating system and other programs authenticate a user, and the user is the basis of protection policies. Later at the stage of user's applications it is assumed they perform according to their specifications. It is clear that applications must be better managed to prevent unintended behaviour from having far-reaching consequences. Two concepts can aid in security design at the application level. They are the following:

The Application as a Unit of Protection

They rescind the conventional notion of users as the basis for security, and focus instead on applications. Applications provide the right level of detail for assigning privileges—they are usually written by a single entity, they exist to perform a cohesive task, and they are subject to a particular set of vulnerabilities. We can enforce security by independently verifying that the effect of the program falls within our bounds of correct behaviour.

The Application as a Unit of Reasoning

Much of the difficulty in enforcing per-application security is in deciding what an application is. In the era of shared libraries, components, and services, the traditional notion of an application as a single process is outdated. Users cannot reason about code executing on their system from an increasingly large and arcane set of processes. It is often difficult to detect spyware because it does not present itself as an installed application in the users view. What is needed is a consistent view of installed code with enough information to allow users to make informed choices.

1.5.1 Application-level Security Architecture

Figure 1.7 depicts a generic view of the application components which reside on several hosts. The architecture may support a number of different cryptographic security modules; for instance, a software module, a tamper-resistant hardware device or even (for an individual user) a smart card, as depicted in the Fig. 1.7. In general, the device drivers must handle multiple devices, transaction queueing and device failures, as well as variety of reporting functions. From the application perspective, all that is required is a high-level call to the API (e.g., sign this message). The details of the processing and the keys, or even which cryptographic module is being used, should be completely hidden

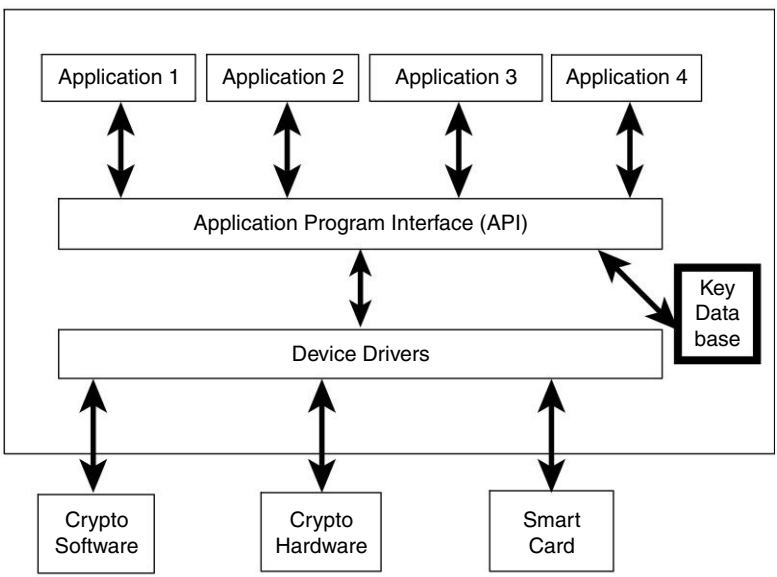


Fig. 1.7 *Application-level Security Architecture*

from the application. We provide a brief note of some of the application level security schemes by classifying them into five categories based on their basic functions in the following subsections.

1.5.2 Trust-based Context-aware Security

Trust naturally leads to a decentralised security management approach that can tolerate partial information, albeit one that has inherent risks for the trusting entity. By clarifying the trust relationship between parties, logical and computational trust analysis and evaluation can be deployed. As a result, it becomes much easier to take proper security measures, and make correct decisions on any security issues. Fundamentally, the ability to reason about trust and risk is what lets parties accept risk when interacting with other parties.

Context has been defined in various ways. Some definitions are broader than others. Some define a context as only location and the identity of nearby people and objects, and some other broaden the meaning to encompass location, identity, environment, and time. We can go further and define context as any information that can be used to characterise the situation of entities which is typically the location, identity and state of people, groups, and computational and physical objects. The context can encompass more than just the user's location, because other things of interest are also mobile and changing. Context includes lighting, noise level, network connectivity, communication costs, communication bandwidth and even the social situation, e.g., whether a user is with a manager or with a co-worker.

Traditionally, security requirements did not need to be context-sensitive as computing existed within a static environment. However, as computing technology becomes more and more integrated into everyday life, it is essential that security mechanisms become more flexible and less intrusive. Security in pervasive computing must be able to assimilate changes in context and situational information effortlessly. For example, access control decisions should factor in time or location and they should be able to change dynamically to limit permissions to times or situations when they are needed. However, viewing what the security policy might become in a particular time or under a particular situation may not be possible.

1.5.3 Role-based Security

At the core of role-based security is the concept of collecting permissions in Roles, which can be granted to Principals (a Principal corresponds to a Login). For example, imagine a Principal admin-Standard who is granted the Role Administrator which comes with all the permissions required for a seg-

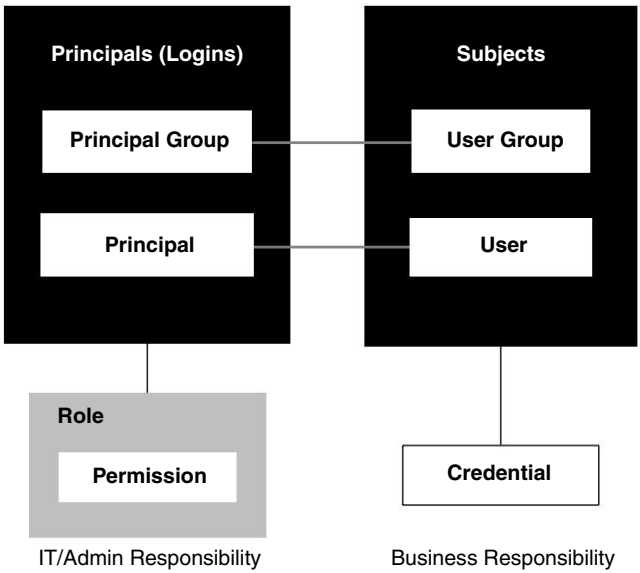


Fig. 1.8 *A Role-based Security Scheme*

ment administrator to do his work, or a Principal sales-representative who is granted the Role which comes with all the Permissions required for a sales representative to do his work. For better manageability several Principals can be collected in a Principal Group and roles can be directly attached to Principal Groups.

Users (e.g., human beings) can be assigned multiple Principals to reflect the fact that some users connect to the system in different roles depending on the tasks at hand. For example, User X might be assigned the Principal head-Sales (because X is head of sales) as well as the Principal admin-Standard (because X is also segment administrator). If X wants to work as segment administrator he/she logs in as Principal admin-Standard, if X wants to work as head of sales he/she logs in as Principal head-Sales. Credentials (like passwords, certificates, SecureIDs) are attached to Users. Like this, it is possible to let X connect to the system with the same password, regardless of whether he/she acts as segment administrator or head of accounting. For better manageability, several Users can be collected in a User Group. The term Subject is commonly used to refer to Users and User Groups.

The other role-based scheme called, Role-Based Access Control (RBAC) has received considerable attention as a promising alternative to traditional discretionary and mandatory access controls. In RBAC, permissions are associated with roles, and users are made members of appropriate roles, thereby acquiring the roles permissions. This greatly simplifies management of permissions.

Roles can be created for the various job functions in an organisation and users then assigned roles, based on their responsibilities and qualifications. Users can be easily reassigned from one role to another. Roles can be granted new permissions as new applications and systems are incorporated, and permissions can be revoked from roles as needed.

An important characteristic of RBAC is that by itself it is policy neutral. RBAC is a means for articulating policy rather than embodying a particular security policy. The policy enforced in a particular system, is the net result of the precise configuration and interactions of various RBAC components, as directed by the system owner. Moreover, the access control policy can evolve incrementally over the system life cycle, and in large systems it is almost certain to do so. The ability to modify policy, to meet the changing needs of an organisation, is an important benefit of RBAC.

1.5.4 Behaviours-based Security

Traditional network-based security examines traffic for code patterns, or signatures, that have been part of past intrusions or virus attacks. If known malicious code is found, security systems stop the suspect transmission. Although this approach can be effective, it also has limitations. For example, signature-based security frequently has trouble recognising new types of attacks or older kinds, in which known code strings have been altered somewhat—an approach many hackers use. Behaviour-based security, on the other hand, learns the normal behaviour of traffic and systems, and then continually examines them for potentially harmful anomalies and for behaviour, that frequently accompanies incidents. This approach recognises attacks based on what they do, rather than whether their code matches strings used in a specific past incident. Several vendors are thus beginning to make behavior-based security widely available

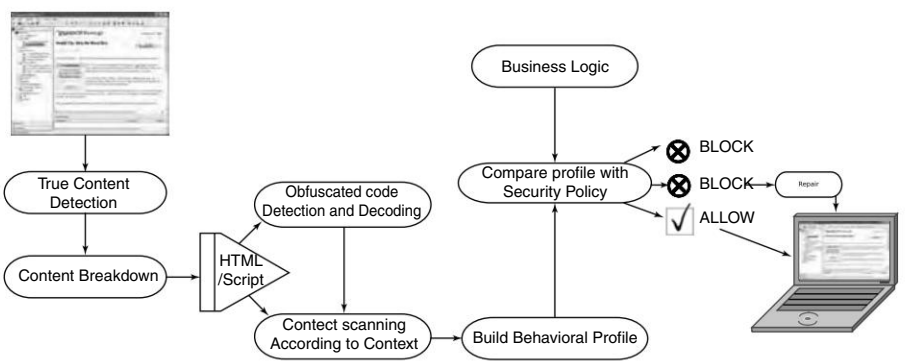


Fig. 1.9 *Behaviour-based Security Scheme*

to organisations via services, appliances, and software products. And some ISPs are protecting their entire networks via behavior-based services. The drawback is—behaviour-based security software can generate a much higher level of false alerts. The advantage is—it can spot and block new threats, based on their behaviours, before a menace is identified and assigned a threat signature and name.

Behaviour-based security products, start by studying behaviour and traffic patterns for a given system and determining which are normal. Some products adapt over time to learn new normal behaviors. Generally, the products are based on an approach which compares system behaviour and actual traffic to normal patterns, to detect anomalies. For example, they look for traffic flowing from one IP address to many others, indicating a worm might be sending itself from one e-mail client to people listed in a victim's address book. Upon identifying potential problems, the systems further analyse traffic to verify whether an attack is occurring, such as by comparing anomalies with entries in a dictionary of harmful behaviours.

Behaviour-oriented security can be based in different parts of a network system.

ISP-based Behavioural Security

This is a global approach in which security systems sit on a few large carriers of networks. When the equipment detects traffic exhibiting anomalous behaviour passing through the network circuits, they send it through alternate routers, scrub out the packet sets causing problems, and route the rest to the customer. Running security on ISPs systems frees up bandwidth and other resources on customer networks. In addition, because they monitor the flow of data packets across much of the Internet, ISPs contend their systems can spot attacks even before they hit corporate LANs. Because ISP-based systems work across an entire network, they are highly scalable. However, they do not offer highly granular control for individual users.

Security-provider-based Behavioural Security

Security vendors place these systems on the edge of a corporate LAN to examine traffic entering the network for questionable behaviour. When these systems find such behaviour, they send alerts and filter out malicious packets.

User-based Security

Organisations often implement these systems as part of ISP's installed on individual computers to provide security against attacks, using or targeting individual applications. These systems look for anomalies in application be-

haviour or for changes that applications cause in systems behaviour. Upon finding anomalies, the systems either block traffic or ask the user for permission to let it pass.

1.5.5 Agents-based Security

Intrusion Detection System (IDS) implemented using Mobile Agents (MAs) is one of the new paradigms for intrusion detection. MAs are a particular type of software agent, having the capability to move from one host to another. Intrusion detection is one of the key techniques behind protecting a network against intruders. An intrusion detection system tries to detect and alert on attempted intrusions into a system or network, where an intrusion is considered to be any unauthorised or unwanted activity on that system or network. Extensive research has been done in this field and efficient IDS systems have been designed for wired networks. These systems usually monitor user, system and network-level activities continuously, and normally have a centralised decision-making entity. While these architectures have proven to be effective, most of the techniques will not produce expected results when applied to wireless networks, due to some inherent properties of wireless networks.

1.5.6 Transactions-based Security

Client authentication is a common requirement for modern websites as more and more personalised and access-controlled services move online. Unfortunately, many sites use authentication schemes that are extremely weak and vulnerable to attack. These problems are most often due to careless use of authenticators stored with the client. For the illustration, consider mobile commerce transactions.

A tentative definition for M-commerce transactions can be stated as follows: "A mobile E-commerce transaction is any type of business transaction of an economic value that is conducted using a mobile terminal that communicates over a wireless telecommunications or personal area network with the E-commerce infrastructure."

Traditional transactions are used to encapsulate database operations so as to provide Atomicity, Consistency, Isolation, and Durability (ACID). They provide clean semantics to concurrent executions and a powerful abstraction for an application developer. There are three main issues which differentiates Mobile E-commerce transactions. The first issue is the *possible hostility of the open environment*. From transactional point of view it means there is a risk that the parties engaged in an E-commerce transaction might be more easily

disguised or forged ones. The second issue that is related with the previous one is the *vulnerability of the mobile hand-held devices*. They can easily be stolen and misused. Thus, from transactional point of view, the transactional mechanism should not rely on the device identity (such as phone number or IP number) and it should not deduce user's identity based on the device identity. The third issue that is closely related with the mobility concept is the *communication autonomy of the devices*. It means that the devices are not always reachable through the network and it is natural that they are rather often disconnected. From transactional point of view this means that transactional mechanisms should not assume continuous capability of the terminals to communicate, nor it should expect that there would be periods during which the terminal is able and willing to communicate with other components with (nearly) 100 per cent certainty.

1.6 ADVANTAGES AND DISADVANTAGES OF APPLICATION-LEVEL SECURITY

Advantages of Application-level Security

Some of the existing application-level security schemes for mobile communication have yielded following advantages.

- Provides true end-to-end security: Takes the security back to the user or the business application.
- ***Flexibility:*** Can use the intelligence of the application to provide a wide range of cryptographic services. For example, non-repudiation via digital signatures or selective field encryption.
- ***Protection against attack:*** In particular, protection against insider attacks, but also against external attacks to a certain extent (does not stop an outsider penetrating systems, but helps to limit the damage, thereafter).
- ***Device re-use:*** A single security device can provide services to a number of different applications (transparent to the applications).
- ***Multiple devices:*** The architecture can support multiple device types (possibly, for different applications) and also multiple devices for throughput and redundancy purposes.
- ***Secure audit trails:*** For example, a hardware-generated sequence number and a MAC or signature can be supplied for each audit trail entry, if appropriate.
- ***Mandated use:*** Application-level security using tamper-resistant hardware security modules is mandated for many banking applications (for example, inter-bank clearing, automatic teller systems (cash machines) and point-of-sale systems).

Disadvantages of Application-level Security

- ***Application dependence:*** Modifications to applications may have the knock-on effect of having to modify parts of the security solution.
- ***Maintenance difficulties:*** Security maintenance across a large organisation may be difficult (e.g., distribution of security upgrades).
- ***Application security weaknesses:*** The interdependence of the applications and the security solutions may lead to weaknesses via the applications themselves (e.g., always verify the PIN for this card number).
- ***Difficult to retro-fit:*** Because the security is tied so closely to the applications, adding security later may be difficult and expensive.

SUMMARY

In this chapter, we did brief discussions on various mobile communication technologies, including cellular networks, Wi-Fi networks, mobile ad hoc networks (MANETs) and Bluetooth networks. We have provided various security issues in these technologies, and discussed few limitations. We have set a stage for need for application-level security.

REVIEW QUESTIONS

1. Explain wireless system architecture.
2. What is radio resource management? Explain various aspects of RRM.
3. Explain the working of GSM with its architecture.
4. Explain the working of GPRS with its architecture.
5. Explain the working of EDGE and UMTS.
6. Explain the working of Wireless LANs.
7. Explain the working of Mobile ad hoc networks.
8. Explain the working of Bluetooth Networks.
9. How security in wireless differs from that of wired networks.
10. What are the various security issues in wireless networks ?
11. Explain the requirement of integrity protection of data in wireless networks.
12. Explain user-ids and provisioning in wireless networks.
13. Explain the role of equipment identifies in wireless networks.
14. Explain WAP gap security problem.
15. What are the main reasons for WEP vulnerability ?
16. What are the security weaknesses for VPN ?
17. What are the advantages of WPA over WEP ?
18. How could a WPA protocol be attacked ?

19. Why is VPN called a hackers paradise ?
20. Briefly explain trust-based security for applications.
21. Briefly explain role-based security for applications.
22. Briefly explain behaviours-based security for applications.
23. Briefly explain agents-based security for applications.
24. Briefly explain transactions-based security for mobile applications.
25. What are advantages and disadvantages of an application-level security ?

Security at Device, Network, and Server Levels

2

OBJECTIVES

- To understand the need of security at mobile device, network and server levels.
- To understand some of the security threats for mobile devices, mobile wireless networks and the mobile servers.
- To study the wireless technology challenges for designing security schemes at these three levels.
- To study various design requirements.
- To understand the reasons for security failures at these levels.

Modern technologies are becoming ever more integrated with each other. Mobile phones are becoming increasingly intelligent, and handsets are growing more like computers in functionality. The smart devices, such as PDAs, on-board computers, and new generation household appliances are equipped with communication functions. We are entering into a new era—the age of smart houses, global networks which encompass a wide range of devices, all of them exchanging data with each other. Such trends clearly open new horizons to malicious users, and the potential threats are self evident.

The security for mobile devices is already an issue; and the first indication of trouble came, not surprisingly, in the form of a virus. At first glance, a problem of malicious code appears only to be a part of the broader information security picture. However, malicious programs inevitably mutate from innocent amusements created by bored programmers and students who wish to show off their skills to professional solutions created for financial gain. There is always a need for virus research on the one hand, and network security,

program vulnerabilities, advertising software and criminal structures on the other, has effectively been eliminated. Therefore researching the techniques used by the authors of malicious code cannot be over estimated as an essential part of ensuring digital security.

Wireless and mobile networks have undergone a tremendous evolution since their start. This was mainly motivated by the need for connectivity everywhere, as exemplified by the philosophy of always on access. Mobile telecommunications networks were the first focus of mobile telephony, followed by a set of wireless technologies with or without embedded mobility. This large deployment of services over mobile and wireless networks demands enormous network security.

The servers in an organisation provide a wide variety of services to internal and external users, and many servers also store or process sensitive information for the organisation. Some of the most common types of servers are Web, email, database, infrastructure management, and file servers. Servers are frequently targeted by attackers because of the value of their data and services. Server security is as important as network security because servers often hold a good deal of an organisation's vital information. If a server is compromised, all of its contents may become available for the cracker to steal or manipulate at will.

2.1 MOBILE DEVICES' SECURITY REQUIREMENTS

According to a survey, many business users utilise their smart phones, PDAs, laptops, not only for company business, but for e-mail, instant messaging, browsing the Web, downloading and sharing files over the Internet, as well as for checking financial accounts. The survey also found that the majority of smart phone users (55.7 per cent) store confidential personal, business or client data on their devices. More than 54 per cent of smart phone owners use their devices to send and receive e-mails that include confidential personal data; 40 per cent access bank accounts using their smart phones; and nearly one-third of respondents' access credit card accounts.

Data communications and connectivity increase vulnerability of mobile devices. Mobile devices that communicate and exchange data with applications inside the enterprise can help increase worker productivity. Knowledge workers use mobile devices for various types of connectivity from construction supervisors entering change requests to retail workers using wireless connectivity for inventory management. However, as mobile devices move data and files to and from networks, and tap into existing applications, the risks increase for compromising sensitive data and transmitting infections.

Although PDAs are still selling at a fairly decent rate, many people are using the PDA functions in their mobile phones, as well. Clearly, as the technology matures, more people will be carrying sensitive data with them. So, how safe is this data stored on handheld device? There are several kinds of threats that afflict PDAs. First and foremost is the threat of simply losing the device. Whoever finds it not only has PDA, but also access to all the sensitive data stored on the device. Viruses also pose a threat to PDA. Several viruses that are PDA-specific have come up in recent years, although at the moment the cause for worry is still fairly low. Another threat to be aware of, is various kinds of Internet-related break-ins. Now that many PDAs come with built-in wireless and Bluetooth functionality, it is possible for someone over the Internet, or even across in the coffee shop, to access the data on PDA or even plant a virus.

Laptops have become a valuable part of the computing store. They allow users powerful mobile computers with the capacity and software of many desktops. They also allow connectivity, even outside the office, thus freeing people to take their workplace with them. This is extremely valuable for employees who must travel frequently while remaining in continual communication with their offices. Unfortunately, the mobility, technology and information that make laptops so useful to employees and organisations also makes them valuables for thieves. Informal surveys indicate that about 10–15% of those laptops are stolen by criminals intent on selling the data. Indeed, in case of laptops belonging to CEOs, the information cost carried upon was thought to be worth millions.

Device Loss

The challenge for users is that the inherently small form factors of PDAs and mobile phones make them more likely to be lost or stolen. Most users carry critical data on their devices such as e-mails, address books, meeting notes, and calendar appointments. Also, most platforms come with a simple software-based login scheme that allows configuring a password to protect access to the device. Such mechanisms can easily be bypassed by reading the device memory directly without starting the operating system. Despite the high numbers of mobile devices that go missing, companies are apparently not doing enough employee education to help secure their mobile assets. The problem is not unique, nearly two-thirds of business users do not use a password when they logon to their laptops, and of the users who do use passwords, 15% use their name and 10% give password details to colleagues. A third of the respondents have not changed their passwords for years.

In the hands of someone with malicious intent, a stolen device could be used to launch an attack against the device's owner, the enterprise to which

the device belongs and/or the mobile operator on whose network the device runs. As a result, there are several factors to consider when a device is lost or stolen:

- Replacement cost of the lost/stolen device.
- Cost of restoring data to the device, i.e., the time it takes for reconfiguring the device.
- Possible compromise of confidential data (personal and/or organisation), which could be almost anything depending on the device lost and the data kept on it, e.g., customer records (any industry), patient information (health care), etc.
- Possible breach in the security of the network to which the wireless device connects, e.g., if network passwords are stored on the device, then an unauthorised user could gain access to the network via an authorised device.
- Alternatively, if the device identity itself can be altered, then it can be re-purposed to possible affect the network and to bypass network security mechanisms based upon that identity. This is typical of the so-called cloning threat wherein a stolen device is reprogrammed with a new identity to bypass black-list enforcement.

Data Damage by Mobile Malwares

The term "mobile malware" refers to malicious software such as viruses, worms and Trojan Horses (Trojans) that have been specifically designed to infect wireless handsets. There are several mobile phone viruses "in the wild" right now, but the majority of that malware only affect smart phones and, mostly, only smart phones based on the Symbian Operating System (OS). The small target market for mobile malware is one of the reasons why it barely registers on the "threat meters" published by the leading anti-virus software firms.

The fear, of course, is that mobile malware may soon become far more dangerous to the data resident on those devices, the wireless networks over which those devices communicate and the corporate networks. Undetected malware on a smart phone could get transferred to a corporate network at which point it might be coded to activate and replicate itself.

Because of scenarios like this, more and more mobile enterprises are realising that security and administration policies must be extended to all mobile devices. They need remote interrogation systems to determine whether a device seeking a network connection is really an authorised device. They also need tools that interrogate a device to see if it is current in terms of firewall settings, antivirus updates, and software patches. These security measures are

a matter of policy to reduce risk, ensure business continuity, comply with regulations, etc. This should be a policy irrespective of actual threats because it is a matter of risk management to key business assets, processes and proprietary information.

2.1.1 Mobile Devices' Security Threats and Solutions

The threat model considers the Confidentiality, Integrity and Availability (CIA) of sensitive data being sent to or from the mobile device, being processed on the mobile device by an application, being stored (intentionally or unintentionally) on the mobile device. The generic threats which are applicable to mobile device usage are given below:

Environmental Threats and Solutions

These are threats to the sensitive data arising from the way the mobile device interacts with its environment. The confidentiality of the data may be subverted due to an architectural change required to the fixed IT network to support mobile device working. Out of all three generic threats mentioned here, this is the most complex to decompose, as this is the one with the greatest degrees of freedom. Therefore, this is the area which requires the greatest scrutiny. The degrees of freedom include (but are not limited to) the functionality of the hardware and the physical environments in which the mobile device is used, the threat arises due to the very nature of the mobile device and the way it is being used. A brief list of issues arising from these diverse threats is presented here.

- Threats arising due to the richness of interfaces furnished on current mobile devices. The threat is amplified by the fact that the interfaces tend to be controlled by software and therefore are very difficult to turn off with any meaningful degree of assurance. These additional communication capabilities represent additional vectors through which data may be compromised.
- Threats arising from unauthorised mobile devices in corporate environments. A PDA can be readily configured as a covert wireless LAN (WLAN) sniffer/ encryption cracker. It is therefore in the ITSOs (International Telecommunications Satellite Organisation) interest to control PDAs in the vicinity of vulnerable corporate IT networks.
- Threats arising from the use of the mobile device in busy environments in which little or no security control could be exercised. An example of this is shoulder surfing on a train. This type of threat can be mitigated

with good security operating procedures, supported by user training.

- Some mobile device vendors have fixed gateway products as part of their enterprise solutions, which are intended to manage the communication between the enterprise environment and the mobile fleet of mobile devices. Inclusion of these entities brings the possibility of additional vulnerabilities which will need to be countered if the risk is to be managed.
- Giving bonafide users the ability to access the enterprise network with PDAs may also make it easier for hackers to do the same. This may affect the threat profile of the fixed network. The simple connection of PDA to a fixed computer via Universal Serial Bus (USB) for the purposes of synchronisation introduces additional threat vectors.

It is difficult to propose specific countermeasures for this genre of threat, as the deployed architectures may be specific to the individual business requirements of the enterprise. However, a number of high level principles can be asserted which will aide in making the risk of operating mobile devices more manageable.

- Do not allow unauthorised devices into sensitive environments.
- Be sure that the remote entities accessing fixed infrastructure are who or what they declare themselves to be. This is an authentication issue. Designers should ensure that an authentication mechanism of some strength is used to filter out unauthorised remote access requests.
- For the functions and facilities which are present with the mobile device hardware, make sure that unused functions and facilities are turned off. If possible, turn the features off in hardware as opposed to software, as this provides higher assurance.
- Test vendor security claims by subjecting the end to end architectures to security testing in a controlled, benign environment.

Computer Security (COMPUSEC) Threats and Solutions

There are threats to the sensitive data (and sensitive applications) whilst processing occurs on the mobile device. The confidentiality of the data may be subverted when the mobile device is stolen, as an attacker may be able to access unprotected areas of permanent storage on the device.

If the assertion is made that the baseline security of the mobile device is no worse than an isotropic fixed equivalent, (i.e. good identification/authentication, good system accounting policy, etc.) the additional mobile device threats which arise are attributable to the large amounts of sensitive data being stored/processed on the device and perhaps the use of applications which themselves are sensitive. The use of these applications and data are in a security context

which has a heightened threat associated with it when compared to the fixed context. A facile example is the use of a mobile device on a busy train, where there is an increased risk of theft of the complete device compared to the fixed context in which there is a greater degree of physical protection from threat agents.

These arguments allow us to postulate that there is a need to look after the data and applications on the PDA, using counter-measures such as password controlled access to device processing and also password controlled access to device data. The nature of the countermeasure may need to be subtly different depending on the electronic technology being used, but in general will rely on some form of cryptography which will protect the data unless valid access credentials are presented. Therefore, the arguments presented in the COMSEC section apply, i.e. the strength and therefore assurance of the implementation can only be derived after an objective assessment of all of the aspects of the cryptographic implementation.

The following safety measures could reduce the risk that confidential information will be accessed from lost or stolen mobile devices:

- Provide training to personnel using mobile devices. People cannot be held accountable to secure their information if they have not been told how.
- Remove data from devices that are not in use. Several incidents have occurred with people obtaining "hand-me-down" mobile devices that still had confidential company data.
- Establish procedures to disable remote access for any mobile devices that are lost or stolen. Many devices store user names and passwords for Web site portals, which could allow a thief to access even more information than on the device itself.
- Centralise management of mobile devices. Maintain an inventory about who's using what kinds of devices.
- Patch management for software on mobile devices should not be overlooked. This can often be simplified by integrating patching with syncing, or patch management with the centralised inventory database.

Fortunately, security products that can detect malicious code exist for most mobile device operating systems. Security technologies that can protect both the organisation and the various types of mobile devices should also be implemented. Native mobile device security such as light encryption, basic passwords, and physical locks may deter some hackers, but rarely block a determined criminal.

Communications Security (COMSEC) Threats and Solutions

There are threats to the sensitive data in transit between the corporate network. The confidentiality of the data may be subverted by an attacker whilst the data is traversing an untrusted third party network.

The general counter-measure to COMSEC threats is the use of cryptography, sometimes supported by additional Transmission Security (TRANSEC) measures such as transmission padding. It is worth noting that whilst some vendors relate the efficacy of cryptography to specific aspects such as algorithm and key length, there are many other facets of cryptography which affect its security.

The main COMPUSEC countermeasure is the use of an appropriate grade of cryptography to protect the sensitive data whilst on the device. The effectiveness of this countermeasure is greatly dependent on the implementation of a mobile device, in particular the assertion that only known executable may run. Many OS vendors are expending effort in providing trusted execution functionality as part of their feature set usually, this is based on digital signatures of applications.

A multi-layered approach to security is important; securing the endpoint, gateway and network is key. Endpoint security must go with security at the edge and core of the enterprise network; they are complementary and address different threats and entry points. That said, mobile enterprises should seriously explore the following security solutions:

- Intrusion detection solutions act as a "security force" inside the perimeter to spot intruders that penetrate the outer defenses.
- Message security solutions filter spam and other undesired messages and content at the gateway and are essential to an overall e-mail security solution.
- Integrated firewall/VPN and virus protection/content filtering solutions offer protection from Internet-borne threats for the desktop and can protect data without slowing performance.
- Anti-spyware solutions can provide real-time scanning, automatic detection and removal, and integrated tools for re-mediating the side effects that spyware can have on a user's system.
- Policy compliance management solutions help define and enforce policies from a central location as well as probe for network vulnerabilities and suggest remedies.
- Administration solutions facilitate the management of hardware and software assets, and provide a way to plan, track, and apply system changes.

2.1.2 Mobile Device Security Policies

Improving device security involves the creation of security policies on how to use wireless devices, what types of corporate-related information can be stored on them, and the implementation of technology which performs the following:

1. Encrypts data transmissions to and from the device.
2. Encrypts data on the devices themselves.
3. Protects the identity International Mobile Equipment Identity (IMEI) of the device from being changed (i.e., cloning the device for illegal resale and use)
4. Measures the trustworthiness of the hardware, OS and applications to detect an unauthorised configuration.
5. Allows IT staffers to deactivate, lock and/or wipe devices which have been stolen or lost. This is a feature of many device management software products.
6. Provides strong user authentication both to activate the device and to access the network. In the case of loss/theft, user authentication can slow or halt an attacker entirely.
7. Management functions on the device and on the back-end which allow users to centrally create, rollout, change and enforce their security policies. These management functions are only feasible to the extent to which a device itself can be authenticated and its own integrity ensured.
8. Password protection on all devices at power-on. Most mobile devices ship with this feature; device management software products can allow users to enforce the use of this feature (which many users never activate).
9. As compared to laptop or desktop computers, achieving these objectives with smart phones, cell phones and/or PDAs is different for several reasons:

    - There are more operating systems to support: Symbian, Palm, Windows Mobile. The operating systems and product offerings different corporates have their own security which is not reliant on third-party solutions.
    - Mobile operators may control the WWANs, but they have no control over the public Internet. Data traffic that leaves a mobile operators network is in the clear unless the user (or enterprise) has some security in place (e.g., a VPN or application-level security).

      Mobile operators also control what applications are loaded onto smart phones and cell phones which function on their networks. All applications and content available for purchase through a mobile

operators Web portal, for example, goes through a stringent approval process. They have somewhat less control over the applications used on a PDA with a WWAN data card and the emergence of smart phones with expandable memory slots has also created the potential for unapproved content to be loaded onto those phones. That said, if the mobile operator has the proper tools in place, it can prevent certain applications from running over their network, e.g., VoIP software on a laptop or PDA. A mobile operator would prevent these types of applications from running if the operator were not involved in the revenue chain associated with that service.

- The smart phone devices themselves have comparatively short battery lives, slow CPUs and less memory than laptops and/or desktop computers although their performance improves with each new generation of devices. Lower performance as compared to laptop or desktop computers can limit the types of security applications that can effectively be run on mobile devices. On the positive side, however, the disparity in performance has encouraged innovative solutions to mobile device security.

2.1.3 Some of the Mobile OS Level Security Mechanisms

In this section, we discuss some of the security solutions provided in two popular mobile device operating systems—Symbian and Windows Mobile are discussed.

Symbian OS

Symbian OS is an open operating system, designed for mobile devices, with associated libraries, user interface frameworks and reference implementations of common tools. In this, a security enhancement for its operating system, called "platform security" is introduced. This enhancement is intended to increase the security awareness among software developers, and it provides tools for design and implementation.

- **Capability Model**

Application authentication and authorization is one of the improvements that Platform Security offers. When an application is installed, it is granted certain capabilities to do something. These capabilities are managed by the operating systems kernel, and they cannot change after the software is installed. Capabilities can be approached from two directions, they define the limits and

permissions that an application is bound to, but they also express the level of trust that the application has. Capabilities are used to verify the level of trust when linking static libraries or loading dynamic libraries. Capabilities are used to verify permissions when using inter process communication or requesting a service provided by a server.

- **Data Caging**

To protect both application executable and data files, Platform Security introduces a feature called data caging. Certain file types (for example, executables, resources) are stored in predefined directories, and only applications with sufficiently powerful capabilities can access these directories. Application-specific directories are automatically protected by the file system, so that only processes with the original application Secure Identifier (SID) can access the directory. This arrangement protects files from unauthorised access, thus providing more trust and integrity.

- Secure inter-process communication enhances the security features of client/server communication by adding capability checking to client/server interactions.
- Secure software installer allows authentication and authorisation of installed software.
- Secure backup and restore provides a way to retain the integrity of Platform Security features during backup and restore operations.
- Central repository is designed to store structured data securely. It is implemented as a Symbian OS server that manages the data storage. Server architecture also enables other security features of Platform Security, for example access control and authentication. The data is accessed asynchronously, and it supports transactions, logical tree structures and naming conventions, and types defined by the Open Mobile Alliance (OMA). Accessing is done with a Uniform Resource Identifier (URI) address, and applications have their own root identified by SID, which protects data from other applications. The old data-sharing mechanism (Shared Data) has been replaced with Central Repository.

Windows Mobile 6.x

Windows Mobile is a compact operating system combined with a suite of basic applications for mobile devices based on the Microsoft Win32 API. Devices that run Windows Mobile include Pocket PCs, Smartphones, Portable Media Centers, and on-board computers for certain automobiles. If the mobile devices on network run Windows Mobile 6.*x*, they can benefit from following security mechanisms which are built in.

- **Password Protection**

Windows software on mobile devices gives the option of using a simple 4 digit numerical PIN or an alphanumeric password up to 20 characters in length, which can be comprised of upper and lower case letters, numbers and symbols. The device should be set to lock after a reasonable period of time following power-down. It is possible to even configure the local device-wipe feature to do a hard reset and remove all the user data, if the wrong PIN or password is entered more than a specified number of times.

- **Support for Digital Certificates**

Windows on Mobile can use digital certificates to control which applications are allowed to run based on the digital signature.

- **Certificate Based Authentication**

For better security, Windows on mobile supports authentication using Transport Layer Security (TLS) with an encryption key up to 2048 bits. Desktop Enrollment is performed by connecting the Windows Mobile device to a PC in the domain where the certificate server resides. The certificate is installed on the mobile device through the PC.

- **Remote Wipe**

A remote wipe of the Windows based mobile device can be performed via exchange synchronisation or Outlook Web Access (OWA). All user data, keys and passwords, and configuration settings are overwritten.

- **Storage Card Protection**

With Windows Mobile 6 OS, the data can be encrypted on the storage card, so that it can read only on the device that encrypted it. This can be done via exchange server policies so that it can be controlled by the administrator and not left up to the user. Exchange server can also perform a remote wipe of the storage card.

- **Propagation of Policies**

Enterprise policies can be delivered to Windows based mobile devices when they synchronise with the Exchange server. Devices that do not comply with policies will not be allowed to synchronise with Exchange.

2.2 MOBILE WIRELESS NETWORK LEVEL SECURITY

In this section, some of the popularly used wireless technologies, and their security features and limitations at the hardware level are discussed. In particular, some of the issues pertaining to IEEE 802.11, GSM and GPRS networks.

2.2.1 Mobile Wireless Network Level Security

The wireless link is sensitive from the security point of view because, unlike wired connections, wireless signals can propagate to places well beyond the intended coverage area. This is true particularly of WLAN communication. If sufficient precautions are not taken, it is not difficult for eavesdroppers inside or outside the organisation to snoop on sensitive information with some basic sniffing equipment. To prevent this, mutual authentication, authorisation and encryption techniques need to be used. Location-enabling access control is another technique where selective access controls can be put in place, depending on the location from which access is made.

Security breaches in WLANs may potentially arise out of not changing default secure set identifiers (SSIDs), passwords and other access point settings. These can, however, be handled easily by having installation or configuration policies in place. Wired equivalent privacy (WEP) is the standard security protocol in WLAN (802.11*x*). Besides the many cases where security is compromised when the WEP security setting is left in its default state off, WEP in itself is not a sufficiently strong security mechanism, particularly when sensitive information is exchanged over the wireless network. WEP has been cracked several times in the past using techniques as simple as playing with driver settings. Therefore, it is recommended that upgrades (typically, firmware) to new IEEE 802.11 security standards be considered. Wi-Fi protected access (WPA) resolves most of WEPs serious problems and can be applied as a software (or firmware) upgrade to Wi-Fi certified devices. Periodic wireless sniffers and scanners can be used to check the airways for intrusion and security compliance, particularly when a security policy is placed into effect.

Mobile wireless networks, such as GSM, GPRS or the 3G networks, provide sufficient levels of access control and data encryption. However, for financial transactions, additional security is recommended using wireless access protocol (WAP) security (wireless transport layer security, or WTLS, end-to-end) by having WAP gateways located at the financial institutions or by employing security solutions based on mobile public key infrastructure (PKI). Transactions based on the use of the short messaging service (SMS) typically rely only on the security provided by the mobile network. For additional security,

cryptographic techniques based on the subscriber information module (SIM) toolkit may be employed.

2.2.2 Network Level Security Challenges

Transmission Security

The protection of wireless communication at physical, medium access and data link layers over wireless media. The services include counter measures against radio signal detection, jamming, control/user data acquisition, and eavesdropping.

Communication Security

The protection of data and voice communications between designated end-points. The services include message confidentiality, integrity, and end-point authentication. In addition, they may include optional nonrepudiation, anti-replay protection, and traffic analysis counter measures. Finally, military tactical networks often require rapidly supporting secure communications among dynamic groups of users or equipment, such as dynamically formed (or disbanded) coalitions.

Authorisation and Access Control

The support of multi-level security measures by implementing identity or role based access control on applications, application servers, and their proxies. Multi-level security requires segregation of levels, possibly via cryptography. Authorisation and access control require reliable authentication of human users and communication equipment.

Network Infrastructure Protection

The protection of routing and network management infrastructure against both passive and active attacks, such as rogue devices masquerading as switching elements, insertion, deletion, modification or replay of control messages, and introducing significant delays to message transport. This service may require strong authentication of switching equipment as well as confidentiality, integrity, non-repudiation, anti-replay protection and traffic analysis countermeasures for control traffic.

Robustness

The requirement to accommodate hardware and software failures, asymmetric and unidirectional links, or limited range of wireless communication. It in-

cludes the need for the networks to survive specific types of device overrun (physical seizure), network fragmentation and denial-of-service attacks.

Efficiency

Finally, even more than their commercial counterparts, military wireless networks are expected to be efficient in their use of electrical and computing power, silicon real estate, and communication bandwidth.

2.2.3 WLAN Security Issues

A wireless LAN (WLAN) is a wireless local area network, which is the linking of two or more computers or devices without using wires, also known as the basic service set. This gives users the mobility to move around within a broad coverage area and still be connected to the network. The IEEE 802.11 group of standards specify the technologies for WLANs.

Spread Spectrum Not Very Secure

Wireless LAN's uses spread-spectrum or Orthogonal Frequency-Division Multiplexing (OFDM) modulation technology based on radio waves to enable communication between devices in a limited area. The spread-spectrum is capable of changing the "spreading codes" in a secretive way, which makes it nearly impossible for someone to decipher the signal's intelligence unless they know the code. The problem, however, is that the 802.11 standard clearly describes the spreading codes publicly so that companies can design interoperable 802.11 components. As a result, a hacker only needs an 802.11-compliant radio NIC (Network Interface Card) as the basis for connectivity, which obliterates the security benefits of spread spectrum.

SSIDs

The 802.11 standard specifies the SSID (service set identifier) as a form of password for a user's radio NIC to join a particular wireless LAN. 802.11 requires that the user's radio NIC has the same SSID as the access point has to enable association and communications with other devices. In fact, the SSID is the only "security" mechanism that the access point requires to enable association in the absence of activating optional security features. The use of SSIDs is a fairly weak form of security, however, because most access points broadcast the SSID multiple times per second within the body of each beacon frame. A hacker can easily use an 802.11 analysis tool (e.g., AirMagnet, Netstumbler, or AiroPeek) to identify the SSID. Some network administrators turn off SSID broadcasting (which deletes the SSID from the

beacon frames), but a hacker can still sniff the SSID from frames that stations use when associating with an access point. It is shown that the most serious passive attack is the traffic analysis attack over the cipher texts because it does more than one can foresee. From traffic analysis, the attacker can passively simply figure out the frequency of transmission from a specific station, the size of packets being transmitted and the time taken to receive the response. They just have to wait until someone associates or re-associates (e.g., when roaming) with the network. Aside from sniffing the SSID, many wireless LAN administrators make it even easier by using the vendor's default SSIDs, which are pretty well known.

DHCP

Even if an intruder is capable of associating with an access point by using the correct SSID, they must often have an applicable IP address before they can directly access resources on the network. Many wireless LANs, though, use DHCP (dynamic host configuration protocol) to automatically assign IP addresses to users as they become active. With DHCP enabled, a hacker receives an applicable IP address just as other legitimate users do.

For example, a public wireless LAN may be at an Airport and someone associated to the same wireless LAN can easily use Windows to see other users connected to the network. If file sharing is turned on, the other person can click on the device and drill down to documents folder and open or copy files to their laptop. This is a serious problem that many end users overlook, especially when operating from home and public networks.

Man-in-the-middle Attacks

Through the use of an 802.11 analyser, a person can monitor 802.11 frames sent over the wireless LAN and easily fool the network through various "man-in-the-middle" attacks. One can view the frames sent back and forth between a user's radio NIC and access point during the association process. As a result, some one can learn information about the radio card and access point, such as IP address of both devices, association ID for the radio NIC, and SSID of the network. With this information, it is possible to set up a rogue access point (on a different radio channel) closer to a particular user to force the user's radio NIC to re-associate with the rogue access point, because 802.11 does not provide access point authentication.

WEP

On 802.11 networks, the WEP (wired equivalent privacy), which encrypts the body of each frame. This is supposed to keep away hackers from viewing sensitive

e-mails, user names and passwords, proprietary documents, etc. WEP has two generic limitations. First, use of WEP is optional, and as a result, many real installations never even turn on encryption. Second, by default, WEP uses a single shared key common to all users of a WLAN, and this common key is often stored in software-accessible storage on each device. If any device is lost, stolen, or compromised, the only recourse is to change the shared secret in all of the remaining devices. The Flurer-Mantin-Shamir (FMS) attack, describes the weaknesses in the key scheduling algorithm of RC4 (Rivest Cipher 4) which facilitates an attacker to work back to the key. RC4 is effective such that an attacker needs minimum 1 Gb of data to start inferring the randomness of its output. However, it has few issues as described by FMS attack, RC4 generates predictable key stream in the presence of weak keys and hence one can trace back to get the initial bytes of the secret key.

Wi-Fi Protected Access (WPA)

WPA is an encryption algorithm that takes care of a lot of the vulnerabilities inherent in WEP. A WPA key can be made good enough to make cracking it unfeasible. WPA basically comes in two flavors—Remote Authentication Dial In User Service (RADIUS) or Pre-Shared Key (PSK). PSK is crackable, RADIUS is not so much. PSK uses a user defined password to initialise the Temporal Key Integrity Protocol (TKIP). The TKIP is not really crackable as it is a per-packet key but upon the initialization of the TKIP, like during an authentication, the password could be obtained. A robust dictionary attack will take care of a lot of consumer passwords. RADIUS involves physical transferring of the key and encrypted channels. Look it up to learn more about it but 90% of commercial APs do not support it. It is more of an enterprise solution then a consumer one.

One of the widely-used authentication techniques is the address-based authentication which assumes that the identity of source could be inferred, based on the network address from which packets arrive. This network address could be either layer 3 address (IP address) or layer 2 address (MAC address). The operation of MAC-based authentication can be found in. There are various attack techniques developed over them:

1. Using the spoofed MAC address captured by passive listening the communication.
2. In case of two-factor authentication the DoS attack is launched against the authorised user to hijack authenticated-associated session.
3. Make the secret only one-time usable, so if somebody captures it, is not possible to use it for a second time. This is known as one-time password schemes. The one-time passwords may not defend consumers and e-businesses against real-time “man-in-the-middle” phishing attacks.

Biometrics, the application of statistical analysis to identify individuals through their biological or physiological characteristics, is emerging as a key aspect in new security systems. Biometric authentication systems may be very safe and secure and reliable but these systems are costly and need additional hardware and software support.

2.2.4 Cellular Networks Security Issues

A cellular network is a radio network made up of a number of radio cells (or just called as cells) each served by a fixed transmitter, known as a cell site or base station. These cells are used to cover different areas in order to provide radio coverage over a wider area than the area of one cell. Cellular networks are inherently asymmetric with a set of fixed main transceivers each serving a cell and a set of distributed (generally, but not always, mobile) transceivers which provide services to the network's users. Following are some of the security threats for cellular technology.

Phone Cloning

Let us look at some of the security issues w.r.t. most popular cellular networks of the world, GSM and GPRS. The most severe attack to the cellular systems through the air is phone cloning. A cellular phone is recognised by a pair of uniquely assigned numbers: ESN (Electronic Serial Number) and MIN (Mobile Identification Number). Such pairs of numbers are transmitted to a cell base station through the open air whenever the cellular phone is powered on. These numbers can be easily read by equipments and one can even possibly find the physical location of any powered-on cellular phone. In such cases, when placing a call, the PIN will need to be sent through the assigned voice channel after ESN and MIN are sent through a control channel. However, PINs are vulnerable to eavesdropping as well.

Hijacking

Another possible attack through the air is hijacking. Once a voice channel is established between a cellular phone and a cellular base station, a counterfeit cellular phone may seize the voice channel by increasing its power level above that of the legitimate cellular phone. An attacker could then make an illegal cellular call.

2.2.5 Security Issues in GSM

GSM (Global System for Mobile communications) is the most popular stan-

dard for mobile phones in the world. GSM differs from its predecessors in that both signalling and speech channels are digital, and thus is considered a second generation (2G) mobile phone system. One of the key features of GSM is the Subscriber Identity Module (SIM), commonly known as a SIM card. The International Mobile Subscriber Identity (IMSI) a unique number for every subscriber in the world.

This has also meant that data communication was easy to build into the system. GSM networks operate in four different frequency ranges. Most GSM networks operate in the 900 MHz or 1800 MHz bands. It includes information about the home network of the subscriber and the country of issue. GSM was designed with a moderate level of security. The system was designed to authenticate the subscriber using a pre-shared key and challenge-response. Communications between the subscriber and the base station can be encrypted.

SIM/MS Interface Tapping

It is possible for the SIM to be removed from one mobile station (MS) and put in another with which it has no previous association. As the SIM-ME interface is unprotected, it is possible for the SIM to be connected to terminal emulators instead of genuine MS and consequently for messages on the SIM-MS interface to be tapped. However, unless the algorithms on the SIM are substandard, there is no advantage in compromising the SIM-MS interface in this way.

Attacks on the Algorithm A3/8

Comp128 is new version of A3/8 algorithm used by many GSM operators. It has one weakness that if 160000 random challenge (RAND)-signed response (SRES) pairs are collected then authentication key (Ki) can be found out. Wagnner and Goldberg claimed in 1998 that they had cracked comp128 algorithm. A simple way of doing that is to steal a SIM card connecting it to PC emulator which sends 160000 chosen RANDS to the SIM card and receive the SRES. As the SIM card has very slow clock therefore it will take almost 10 hours to complete this process.

Flaws in A5/1 and A5/2 Algorithms

The attacks on A5/1 and A5/2 algorithms were made by Biryukov and Shamir, later on further improved by Wagner. In this method, a large database for algorithm states and related key streams is prepared. In the attack phase the database is searched for a match with a known key stream. After a match is found, the database gives the correct algorithm state, and then after simple computing, cipher key (Kc) can be calculated for decrypting purposes. Shamir and Birukov obtained this Kc after getting the key stream for 2 minutes

duration, but Wagner did it in only 02 seconds of plain text data (both up link and down link).

Attacks on the SIM Card

Subscriber identification module is implemented on a smart card and vulnerability in smart card will directly affect the security of the SIM as IMSI and Ki are stored on it. The attacks on SIM card are known as optical fault induction that the operation of the smart card processor can be interrupted if exposed to an electric camera flash bulb. Even by scratching the protective coating of the SIM micro processor circuit and focusing the flash light on individual transistors within the chip, by beaming the flash through a microscope, they were able to reverse engineer the memory address map and extract the secret data of IMSI and ki.

False Base Station

The GSM security system provides unilateral authentication, as the mobile station/mobile equipment (MS/ME) is authenticated to base station (BS), but the BS is not authenticated to mobile equipment. This drawback of unilateral authentication allows attacks on the GSM system as a false BS. At the time of GSM designing, it was assumed that the system should not be subject to such attacks due to high amount of expenses involved, as compared to other methods of attacking. But now the cost of GSM BS devices are too low and it is easy to use GSM BS emulators.

This method is based on the fact that ciphering of a call does not start automatically, rather the ciphering starts when BS instructs the ME to start encryption. The instruction from BS to ME to start encryption can be manipulated during transit not to start encryption by an intruder.

2.2.6 Security Issues in GPRS

General Packet Radio Service (GPRS) as a new data service uses a packet-mode technique to transfer high-speed and low speed data and signalling in an efficient manner. GPRS uses a new ciphering algorithm optimised for packet data transmission. GPRS offers the user an “always on” connection to Internet and Intranet. Some of the services may require high level of security. This can be financial transaction over the Internet, or exchange of confidential documents from a company Intranet to an employee. The threats to the GPRS are very different from the Circuit Switched GSM. The GPRS systems are much more exposed to intruders, because they have IP based backbone.

The signalling data or the control data are information that can be useful to conducting active attacks on the GPRS system and give the intruders access to secure management data. Manipulation of user traffic, signalling data or control data may occur in an accidental or a deliberate manner. The integrity is exposed if the traffic and the data in any way are modified, inserted, replayed or deleted. To jam users traffic is a physical intervention of denying someone the services. The user traffic, signalling data and control data are, by jamming, prevented from being transmitted over the air interface.

2.2.7 Mobile Ad hoc Networks Security Issues

The idea behind Mobile Ad hoc NETworks (MANETs) is to enable connectivity among any arbitrary group of mobile devices everywhere, at any time. People realising the need of security is of great importance in MANETs. However, the special properties of ad hoc networks, such as the lack of infrastructure, absence of trusted third parties (TTPs), as well as the constraints of the devices and the communication channel, make implementing security a very challenging task. Among the major challenges are: bootstrapping security, providing authentication and key exchange, and enabling key revocation and key renewing in public key infrastructures (PKIs).

2.3 SERVER LEVEL SECURITY

Security on the server level is one of the most important considerations for a network environment. Servers in an infrastructure not only handle critical network services, such as DNS, DHCP, directory lookups, and authentication, but they also serve as a central location for most, if not all, critical files in an organisation's network. In the current corporate environment, the nature of work is demanding employees to be mobile and this has been determining integration of mobile gadgets with enterprise applications. It is predicted that the global enterprise expenditure on mobile devices will grow year by year.

Most of the businesses are increasingly embracing new ways to communicate, whether its IP or unified communications. Email is an important application for organisations to disseminate information at all levels. Also, besides the email application, employees are using mobile devices for business applications like Customer Relationship Management (CRM), Enterprise Resource Planning (ERP), and Sales Force Automation Systems (SFA) to interact with customers, business partners, and other businesses, etc. Improved access to business information helps enhance employee productivity and collaboration.

Keeping these important mobility applications in view, organisations and institutions have to ensure that their confidential information is secure and protected at all levels. And hence, security concerns are the greatest barrier to implementation of enterprise mobility solutions. It is important for organisations to understand both front-end software, as well as backend infrastructure that play a critical role in addressing these concerns.

The messaging server is the main layer where data is stored and acts as a collaboration unit for users. There has to be a real-time synchronisation between the messaging server and mobile device to ensure better security for corporate information. Sometimes the data passes through the middleware servers which acts as a hindrance to corporate security. But these issues can be addressed by installing personal firewall settings before entering into the server.

2.3.1 Security Threats for Servers

To secure a server, it is essential to first define the threats that must be mitigated. Many threats against data and resources are possible because of mistakes, either bugs in operating system and server software that create exploitable vulnerabilities, or errors made by end users and administrators. Threats may involve intentional actors (e.g., attacker who wants to access information on a server) or unintentional actors (e.g., administrator who forgets to disable user accounts of a former employee.) Threats can be local, such as a disgruntled employee, or remote, such as an attacker in another geographical area. Organisations/Institutions should conduct risk assessments to identify the specific threats against their servers and determine the effectiveness of existing security controls in counteracting the threats; they then should perform risk mitigation to decide what additional measures (if any) should be implemented. The following are examples of common security threats to servers:

1. Malicious entities may exploit software bugs in the server or its underlying operating system to gain unauthorised access to the server.
2. Denial of Service (DoS) attacks may be directed to the server or its supporting network infrastructure, denying or hindering valid users from making use of its services.
3. Sensitive information on the server may be read by unauthorised individuals or changed in an unauthorised manner.
4. Sensitive information transmitted unencrypted or weakly encrypted between the server and the client may be intercepted.
5. Malicious entities may gain unauthorised access to resources elsewhere in the organisation‘s network via a successful attack on the server.

6. Malicious entities may attack other entities after compromising a server. These attacks can be launched directly (e.g., from the compromised host against an external server) or indirectly (e.g., placing malicious content on the compromised server that attempts to exploit vulnerabilities in the clients of users accessing the server).

2.3.2 Server Security Steps

A number of steps are required to ensure the security of any server. As a prerequisite for taking any step, however, it is essential that the organisation/institution has a security policy in place. Taking the following steps for server security within the context of the organisation's security policy should prove effective:

1. Planning

Plan the installation and deployment of the operating system (OS) and other components for the server. Developing such a plan enables organisations to make informed tradeoff decisions between usability and performance, and risk. In the planning stages of a server, the following items should be considered:

Identify the Purpose(s) of the Server

What information categories will be stored on the server? What information categories will be processed on or transmitted through the server?; What are the security requirements for this information? Will any information be retrieved from or stored on another host (e.g., database server, directory server, Web server, Network Attached Storage (NAS) server, Storage Area Network (SAN) server)? What are the security requirements for any other hosts involved?; and so on.

- Identify the network services that will be provided on the server, such as Hypertext Transfer Protocol (HTTP), File Transfer Protocol (FTP), Simple Mail Transfer Protocol (SMTP), Network File System (NFS), or database services (e.g., Open Database Connectivity [ODBC]). The network protocols to be used for each service (e.g., IPv4, IPv6) should also be identified.
- Identify any network service software, both client and server, to be installed on the server and any other support servers.
- Identify the users or categories of users of the server and any support hosts.
- Determine the privileges that each category of user will have on the server and support hosts.

- Determine how the server will be managed (e.g., locally, remotely from the internal network, remotely from external networks).

2. Install, Configure, and Secure the Underlying OS

Most commonly available servers operate on a general-purpose OS. Many security issues can be avoided if the OSs underlying the servers are configured appropriately. Because manufacturers are unaware of each organisations security needs, server administrators need to configure new servers to reflect their organisations security requirements and reconfigure them as those requirements change. The following basic steps are necessary to secure the OS: Patch and update the OS; Harden and configure the OS to address security adequately; Install and configure additional security controls, if needed; Test the security of the OS to ensure that the previous steps adequately addressed all security issues.

3. Install, Configure, and Secure the Server Software

The overarching principle, as before, is to install only the services required for the server and to eliminate any known vulnerabilities through patches or upgrades. Any unnecessary applications, services, or scripts that are installed should be removed immediately once the installation process is complete. During the installation of the server software, the following steps should be performed.

- Install the server software either on a dedicated host or on a dedicated guest OS if virtualisation is being employed.
- Apply any patches or upgrades to correct for known vulnerabilities in the server software.
- Create a dedicated physical disk or logical partition (separate from OS and server application) for server data, if applicable.
- Remove or disable all services installed by the server application but not required (e.g., gopher, FTP, HTTP, remote administration).
- Remove or disable all unneeded default user accounts created by the server installation.
- Remove all manufacturer's documentation from the server.
- Remove all example or test files from the server, including sample content, scripts, and executable code.

2.3.3 Security Solutions

Some servers check whether the security identifier (SID) or group membership SIDs of the user have been specifically denied access to the server. If so, the

user will not be granted access to the server. If the user is not specifically denied access, the server checks whether the user has been granted access directly or by virtue of a group membership. If access has been granted, the connection to server is maintained. The user then proceeds to the appropriate default database (where the user must also have been granted access). The access rights of the user are then checked for any objects the user is attempting to access. If access has not been granted for a particular set of logon credentials, the connection to the server is terminated. For non-trusted connections, such as when server is installed on the operating system, SIDs are not available. In this case, server generates a 16-byte globally unique identifier (GUID). The generated GUID is then used internally in the same way as SIDs are used for users and groups.

Firewalls are utilised as the main perimeter protection tool; they effectively determine which ports are closed or opened into the corporate network. The ports that are left open provide a conduit for the hacker to penetrate the firewall and break into a server machine. A good example is port 80 (HTTP protocol), which is used by Web servers and therefore is always left open. An attacker can pass a specifically crafted but legitimate HTTP message through the firewall to a Web server, exposing its vulnerabilities. The HTTP message can then exploit one or more of these vulnerabilities and cause a chain of events that ultimately allows the intruder to obtain privileged access to the Web server machine. This may seem to be a farfetched scenario. However, executable programs that do this are widely available for download off the Internet for those who are looking for them.

Intrusion detection systems are perceived as the next layer of defence in addition to the firewall. However, they only detect but do not provide real time prevention of attacks. Increasing evidence shows that Network IDS (NIDS) products have limited detection capabilities and inherent difficulties in properly identifying attack attempts. As a result, many attacks are left undetected, and false positives are generated as well.

The major drawbacks of NIDS are:

- NIDS cannot prevent attacks in real time. They listen to packets on the wire, but do not block their transfer. More often than not, the packet reaches its destination and is processed prior to interpretation by the NIDS. As a result, it is common for an attack to be successful before it is identified by the NIDS.
- NIDS cannot detect unknown attacks. Any signature-based system (like IDS) can handle only known attacks for which signatures exist in the product database.

2.3.4 Mobile Web Server

To give specific illustration, we consider a mobile web server.

The key target of Mobile Web Server is to facilitate straightforward content creation and service development for mobile websites (mobsites). To enable the effortless development of mobsites, Mobile Web Server contains a built-in mobsite application. The integrated features include, for example, a blog, a guestbook, and a calendar application.

One of the unique and innovative features of Mobile Web Server is interactivity with the mobsite owner. As the multimedia computer normally travels with the owner and is an active part of the Internet, multiple features can be used for ad-hoc interaction. Currently, features where a Web surfer requests a response from the mobsite owner include interactive photography and messaging. In practice, these features enable immediate content creation, which may also be linked to a certain location or moment in time.

WLAN may well be the most important factor influencing the adoption of mobile Web servers. In a WLAN, the transmission rate/price ratio is on a level of its own (although WLAN connectivity is far from being universally available). There are three alternatives for accessing Mobile Web Server via WLAN. First, the server can be connected to the gateway through a WLAN base station by using Internet connectivity. The other two methods only function within a strictly limited area. By using a WLAN base station as a hub, it is possible to access Mobile Web Server through a WLAN created by the base station. The last alternative is to create an ad-hoc WLAN network hosted by the multimedia computer.

Security Aspects of Mobile Web Servers

Mobile Web server security issues must be approached from several angles. It is necessary to protect the actual multimedia computer, mobsite content, and in some cases even the anonymity of the mobsite owner. The mobile Web server products support a traditional username/password-based access control system, with a drastically different approach in terms of user-friendliness.

If the content is not meant to be available to everyone, it must be protected using authentication. Especially in the Mobile Web Server scenario where personal content is expected to play a significant role, prohibiting unauthorised access is a key feature. In the Mobile Web Server case, it is easier to define folder access restrictions by using the folder access feature available.

The gateway is responsible for securing the entire data transport route between Mobile Web Server and the Web surfer. Two different techniques are used at different points along this route. The connection between the gateway and the Web browser used to access content on Mobile Web Server is secured

using HTTPS. The gateway is the endpoint of the HTTPS connection, and the last leg between the gateway and Mobile Web Server is secured using SSL. The security solutions are completely transparent from the end users point of view.

The entire Mobile Web Server system, including the gateway and the connector, is designed to take malware into consideration. The key issue here is the role of the gateway, which protects the relatively slow mobile connections and takes the pressure off the processor of the mobile device. As all traffic normally goes through the gateway (barring local connectivity scenarios), it can be filtered for malware in the process. This lowers the risk of Trojan horses or viruses making their way into Mobile Web Server, or moving the other way to the mobsite visitors computer.

SUMMARY

In this chapter we have made an attempt to categorically discussing security threats for mobile applications at three levels of mobile communication. The influence of device-level, network-level, and server-level vulnerabilities over the design and development of mobile-based applications is discussed. We have considered only cellular networks (GSM and GPRS) and WLAN's to keep the discussion simple.

REVIEW QUESTIONS

1. What are the security threats for mobile devices ?
2. What are the security challenges for mobile devices ?
3. What are the security impacts of loosing a mobile device ?
4. What are the design level security requirements for mobile devices ?
5. Mention various security solutions available at mobile device level.
6. Why is multi-layer security solution required for mobile device security ?
7. What are Gold's suggestions to mitigate the mobile device risks ?
8. Briefly explain the environmental threats for mobile devices and their counter measures.
9. Briefly explain the computer security threats for mobile devices and their counter measures.
10. Briefly explain the communications security threats for mobile devices and their counter measures.
11. What are the security issues of WLANs ?
12. Why is spread spectrum not secure ?
13. What are the security drawbacks of SSIDs ?
14. Why is DHCP a problem in Wi-Fi networks ?

15. Explain the weaknesses of WEP security.
16. What are the improvements of WPA over WEP ?
17. How are man-in-the-middle attacks performed in wireless networks ?
18. What is phone cloning ?
19. What is session-hijacking in cellular networks ?
20. What are the security issues in GSM ?
21. What are the flaws in A5/1 and A5/2 algorithms ?
22. What are the various attacks on SIM ?
23. What are the security issues in GPRS ?
24. Mention various security issues at server level.
25. What are the major drawbacks in using NID's ?

Application Level Security in Wireless Networks

3

OBJECTIVES

- To understand the need of wireless networks applications.
- To understand the security issues in wireless networks.
- To study the types and methods of attacks over wireless networks.
- To know the various security issues pertaining to development of applications over first and second generation Wi-Fi networks.
- To study some of the recent security schemes proposed for Wi-Fi applications.

In offices, airports and even at homes, Wi-Fi technology is rapidly gaining acceptance as the network connection of choice for many mobile applications. Easy and cost-effective to set up, maintain and change with the dynamic needs of an organisation, Wi-Fi empowers employees to move freely and work collaboratively anywhere within the range of a wireless LAN or public "hot spot".

The effect can be seen in jobs across the health care, government, retail, hospitality, entertainment, manufacturing, transportation and other industries. With wireless access to enterprise data and applications, waiters, loading dock employees, inventory-takers, health care professionals, hotel event managers and other workers no longer have to rely on tedious, error-prone paper-based systems.

As telecommunication companies install more hot spots and public wireless access points, mobile workers such as sales representatives, utility and maintenance crews, and government inspectors will have access to more timely information, especially in areas where wide-area wireless networks are not an

option because of cost, signal availability and performance issues. The idea of automating these workers is compelling—Wi-Fi applications offer the same access to information as typical wired applications, but without the mobility limitations of wires. So organisations can benefit from increased employee productivity and morale, the ability to provide better customer service, more timely and accurate data exchange, and reduced costs.

To take advantage of the opportunities created by Wi-Fi for point-of-activity computing, it's important to understand the unique needs of the users. Mobile field service workers, sales persons and professionals and on-premises users working in departments within larger organisations, remote branch offices to midsize businesses have common characteristics. Usually, they have minimal access to resources, including limited or zero on-site technical support as well as limited budget and hardware. However, once they have Wi-Fi access, their appetite for enterprise-class functionality and information-sharing capabilities will quickly rival that of wired users.

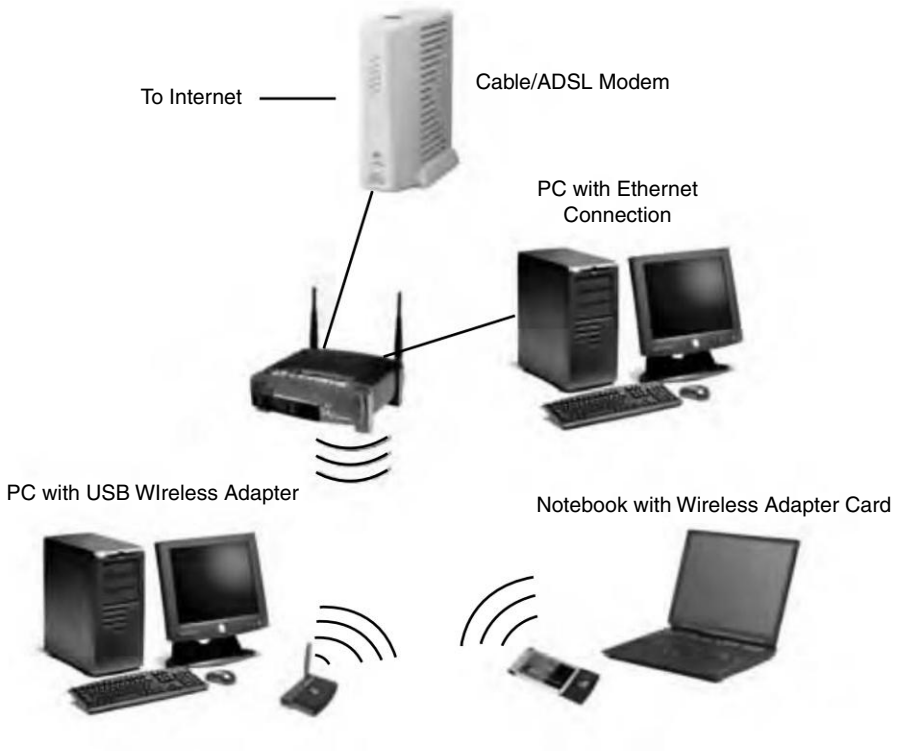


Fig. 3.1 *A typical wireless network with some wireless gadgets*

The security is the major area that developers must consider while developing application for Wi-Fi, along with areas such as roaming between hot spots; ease of use; and ease of administration. Wireless LANs and local hot spots can be accessed by unwelcome users unless proper security measures are taken. It is important to remember that application files may contain confidential business information and must be protected whether it is located on a PC in the office or a hand-held device in the field.

3.1 APPLICATION OF WLANs

Wireless LANs are useful for a wide range of applications. There are some applications, however, that are more effective and efficient than others. Let's examine some applications, which are convincing situations that prompt the use of WLANs.

3.1.1 Sharing Internet Access

The most compelling reason to install a WLAN is to share a single high-speed Internet connection. This ability can benefit almost anybody, from an enterprise businessman to a student to the person surfing the Web at home. The use of a WLAN to share a high-speed Internet connection allows the user to stay mobile and save money, because there are no wires to buy or install. One example of a WLAN for this purpose is within a small office or home setting. Every member of a family or small business can easily share a single high speed connection through the use of a cable/DSL modem, router, access point (or wireless router), and radio-equipped end user devices, see Fig. 3.2. This is very convenient and saves money because everybody can simultaneously have access to a single connection and roam anywhere in the house or office.

A WLAN increases the flexibility of the network because one can add new workstations at any time without having to run cable. The relocation of workstations, along with any printers or servers, is also very easy. WLANs also provide a high level of convenience in a larger enterprise environment because guests and corporate visitors with wireless devices can quickly connect to the network with very little configuration.

3.1.2 Transmitting Voice over WLANs

The use of a WLAN to transmit voice is a great solution when people need to constantly be in contact with each other (Fig. 3.3). WLAN phones, which work just like cell phones when they are in the coverage of the WLAN, are

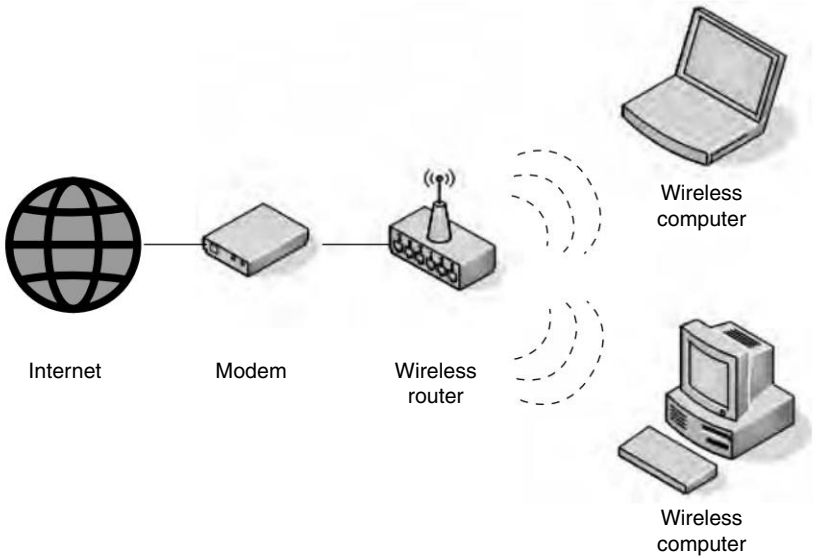


Fig. 3.2 *Internet access using wireless*

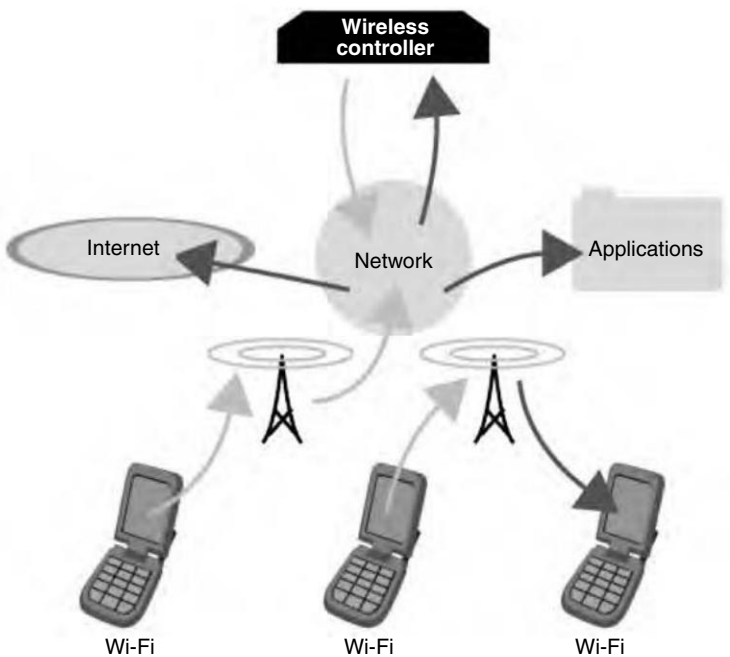


Fig. 3.3 *Voice over wireless*

very useful in places where workers are moving around. Some examples of WLAN phone solutions include the following.

- Retail stores where employees can communicate with each other to locate certain clothes for a customer and/or check inventory.
- Buildings where security guards can quickly call for onsite help, the police or an ambulance.
- Schools that need constant communication with teachers and administration.
- Hotels where staff members are spread out on different floors and need to respond quickly to requests.

3.1.3 Manufacturing Industry and Inventory Management

Many businesses benefit from using WLANs to manage their manufacturing processes. This lowers operating costs. Because the connections between the manufacturing equipment and main control systems are wireless, the company can reconfigure the assembly process at any time from anywhere, saving time and money. A WLAN can also track and update inventory in real-time, enabling efficiency and accuracy to increase dramatically. In a retail environment, as soon as a clerk purchases or stocks a product, a wireless management solution can update the inventory. In the manufacturing environment, WLANs can keep the raw materials and finished product statistics up-to-date. Employees equipped with wireless-enabled bar code scanners can check or change product prices and/or check the number in stock.

The improved accuracy provided by using a WLAN to manage inventory creates a chain reaction of benefits. Because the clerks enter the information directly into the main computer via handheld scanners, there is no paperwork to deal with. This significantly reduces human error when entering data, which leads to very accurate financial records. It is important to manufacturing companies because accurate financial records ensure correct taxes are paid and fines (and possible law suits) are kept to a minimum.

3.2 WIRELESS THREATS

To understand 802.11 and the advantages it offers to existing wireless security mechanisms, we must understand potential attackers and their threats. Knowledge of the real threats against wireless networks helps us place the complex landscape of security mechanism in context. Depending on an environment

and the assets that need to be protected, the risks posed by various attackers might vary.

3.2.1 Targetted Attackers

When IT professionals think about attackers (or hackers), they often envision a malicious individual dedicated to breaking into a trusted network. They think of an attacker with a grudge, sitting in a dark room, working late into the night doing stealthy scans, creating custom exploits, and quietly compromising their infrastructure. They think of someone who has nothing better to do than full-time attacking. A dedicated attacker who targets a specific enterprise is an IT pros worst nightmare, but an unlikely one.

Similarly, a targetted wireless attacker is also very unlikely. For an attacker to explicitly target a network, there must be a valuable enough asset for the attacker to pursue. For most home and small office networks, the payoff for breaking into a wireless network is simply too small for an attacker to expend the effort. However, if your enterprise contains valuable trade-secret, financial, or personal information, then the threat of a targetted attacker, even though it is remote, is likely to warrant a security solution.

3.2.2 Attackers of Opportunity

It is most likely that someone will attack a wireless network because it is a target of opportunity one that has no functional level of security and that an attacker can easily compromise. Attacker misuses vary wildly on targets of opportunity. For example, some attack pursue internal assets on the network, reveals that most attackers simply attempt to gain Internet access. While this type of misuse is technically an attack, the assets value is low. Using open wireless Internet access points to check e-mail and read news sites is a way of life for some people, so many could argue whether this really is an attack. However, if users access the open network to launch attacks against external resources using the wireless network to hide their identities, is a much more clear-cut case. Nonetheless, when a user accesses assets without permission, by most definitions, this type of use is an attack.

For attackers pursuing more valuable assets such as financial data and trade secrets, targets of opportunity are not a viable way of achieving their goals. Randomly attacking local targets on an open wireless network is not a high-yield activity for dedicated attackers. Typical home or small-office networks have few valuable assets (credit card numbers, personal information, product sales and so on) worth pursuing, assuming the hosts on the network are vulner-

able to attack in the first place. Spending an hour hunting for a single credit card number simply is not worth of most motivated attacker's time. Large enterprise environments are not likely to be targets of opportunity because they have rudimentary wireless security mechanisms in place.

3.2.3 Internal Attackers

There is another type of attacker that is called accidental. Employees within an enterprise potentially can subvert wireless network security better than targeted attackers. These employees' actions can punch through most defenses, and they underscore the need for wireless network internal auditing. For example, an employee might have difficulty with an existing enterprise wireless network: the network may be difficult to use or provide sub-par coverage in certain areas of a building. To overcome these difficulties, the employee could install a personal desktop wireless access point. While this solution is effective for the employee, it is not acceptable from an IT security standpoint. This rogue access point is outside the IT staff's control and might not adhere to enterprise security standards. Worse, the access point is an uncontrolled hole into a core network. Rogue access points are difficult to prevent with the current wireless security standards, in conjunction with a properly engineered wired network, minimises their vulnerabilities. Similarly, employees can leave their wireless interfaces connected to wireless networks when they dock their workstations using wired network docking stations, thereby providing uncontrolled dual-homed hosts. IEEE 802.11i and a properly engineered wired network could mitigate this threat.

3.3 SOME VULNERABILITIES AND ATTACK METHODS OVER WLANS

In this section we describe some situations in which WLANs and their applications are vulnerable to security attacks, and also, present some of the methods used for attacking these networks.

3.3.1 Human Error

It is understood that an individual with no understanding of networks can easily set up a flawed and vulnerable network. However, some executives need to be aware that even their system administrators could be lacking in their understanding of wireless network implementations. With a broad

number of floating, corporate hotspots being found everyday it has to be assumed that some of those hotspots were put into place by knowledgeable IT personnel. However, some of those techniques may have missed something. Maybe a manager gave some of his development staff permission to install a wireless router while providing no oversight to the installation. Though developers may know something about software architecture and design, they may or may not know anything about network security. Perhaps the local system administrator does not thoroughly understand networking principles. May be an individual lacks the tools necessary to carefully monitor network traffic and detect anomalies that could indicate the presence of a rogue access point. Worst case, someone might not even care. The companies should make sure that system administrators are well trained with a strong background in computer and network security.

In any system, the human components are the weakest link. Wireless networking is certainly no exception. An organisation should define strict policies and procedures related to wireless networking within a well-publicised company document. It is especially important with regard to wireless networking that employees are made aware of these rules. Sometimes, an unwitting employee with good intentions may compromise company data without even knowing the repercussions. Because wireless hardware is cheap and relatively easy to use, the risk of your network containing rogue access points is great. One should be sure to set standards for any wireless hardware configurations within the company network and perform routine network audits to ensure that there are no open doors.

3.3.2 Rogue Access Points

As discussed earlier, it is easy for even a novice to acquire equipment and set up a wireless network. If this is done within another network, it creates what is known as a subnet, which can create back doors to its parent. There are many easily overlooked mistakes that can be made in configuring a wireless network, many of which novice users will overlook. Individuals who wish to intrude upon a network can also plant rogue access points themselves. Network administrators must make sure to implement strict polices regarding the deployment of wireless hardware, and audit their networks often with reliable tools to ensure that these rogue access points do not exist.

Different Types of Rogue Devices

- **Employee Installed Rogue Access Points**

Driven by the convenience of Wireless home networking some employees

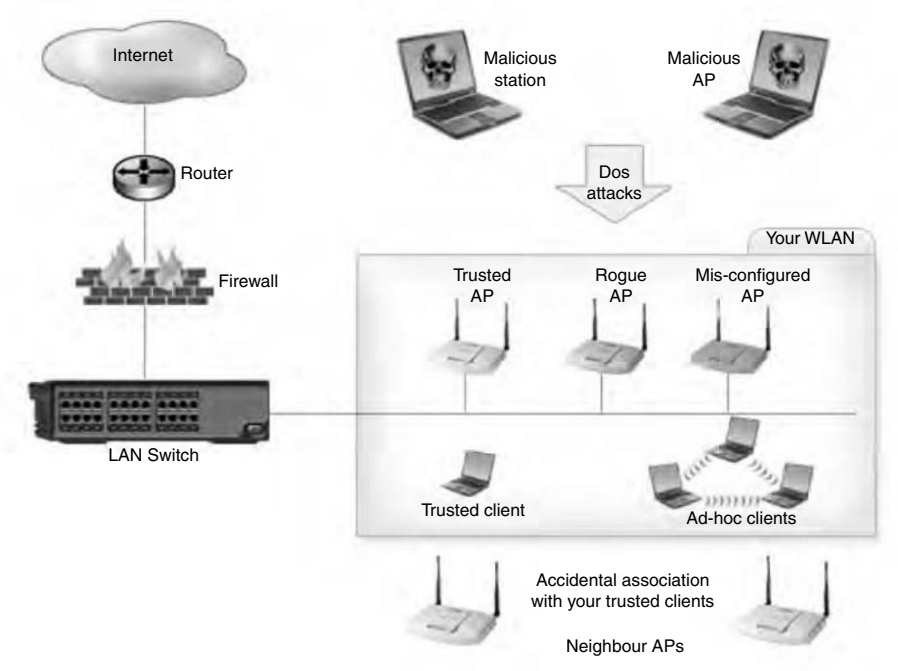


Fig. 3.4 *Different types of rouge devices*

plug cheap Small Office Home Office (SOHO) grade access points to corporate LAN. This unintentional act by the novice users punches a big hole on enterprise security exposing critical data to outsiders. The cheap AP may not follow enterprise standard deployment procedures thus compromising security on the wireless and wired network. Visitors inside your building and hackers outside your building can connect to such unauthorised APs to steal bandwidth, send objectionable content to others, retrieve confidential data, attack company assets, or use your network to attack others.

- **Mis-Configured Rogue Access Points**

Sometimes an authorised access point could suddenly turn into a rogue device due to a minor configuration flaw. Change in Service Set Identifier (SSID), authentication settings, encryption settings, etc., should be taken seriously as they could enable unauthorised associations if not configured properly. For example, in open mode authentication any wireless client device in state1 (unauthenticated & unassociated) can send authentication requests to an AP and on successful authentication would transit to state2 (authenticated but unassociated). If an AP doesn't validate the client properly due to a configuration flaw, an attacker can send lot of such authentication requests, overflow the

APs client-association-table, and make it reject access to other clients including the legitimate ones.

- **Rogue Access Points From Neighbour WLANs**

IEEE 802.11 clients automatically choose the best available AP nearby and connect with them. For example, Windows XP connects automatically to the best connection possible in the vicinity. Due to this behaviour, authorised clients of one organisation can connect to Access points from the neighboring organisation. Though the neighbour's APs have not intentionally lured the client, these associations can expose sensitive data.

- **Ad-hoc Devices**

Wireless clients can communicate among themselves without requiring a LAN bridging device such as Access Point. Though such devices can essentially share data among themselves, they pose significant threat to the application as they lack the necessary security measures such as IEEE 802.11x user authentication and the dynamic key encryption. As a result, ad-hoc networks risk exposing data in the air (as data is not encrypted). In addition, weak authentication may allow unauthorised devices to associate. If the ad-hoc mode clients are also connected to the wired network, the entire enterprise wired network is at risk.

- **Rogue Access Point that Do Not Adhere to Corporate Policies**

Applications or service providers can set policies on what constitutes an authorised AP. The basic one is MAC addressed based filtering. Applications can preconfigure the list of authorised devices MAC and identification of any other device outside the MAC list will signify the presence of a rogue device. Similarly applications can set various policies including SSID (Service Set IDentifier), Radio Media Type, and Channel. Whenever, a new access point is discovered in the network that falls outside the pre configured authorised list, it can be assumed to be a rogue AP.

- **Rogue Access Points Operated by Attackers**

Wireless LANs are prone to numerous attacks. Furthermore, freely available open-source attack tools ease the job of attackers. Attackers can install Access Points with the same ESSID (Extended Service Set IDentifier) as the authorised AP. Clients receiving stronger signal from the attacker operated AP would then attract legitimate clients to associate with it. The AP can then launch a man-in-the-middle attack. Attacker operated clients: Using a wireless enabled laptop and couple of tools an attacker can successfully disrupt wireless

service in networks few feet away. Most such denial-of-service attacks aim at exhausting AP resources such as the client-association-table.

3.3.3 Warchalking

Another point of vulnerability, which is possibly more of a compelling idea than a physical reality called warchalking. It is a modern version of the hobo sign language used to alert one another to places providing shelter, food, and potential trouble. Using a fairly universal hobo sign language, individuals mark structures that have hotspots associated with them. In many cases these symbols incorporate much information about each node and the type of security currently being implemented. Figure 3.5 gives some sample warchalking symbols and their interpretations.

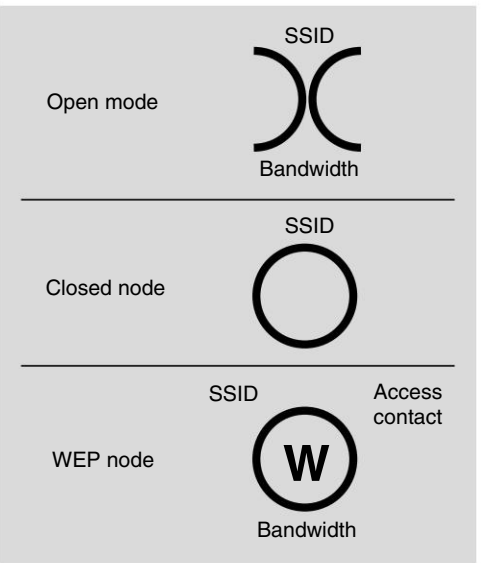


Fig. 3.5 *Warchalking symbols*

3.3.4 MAC Address Spoofing

Media Access Control (MAC) addresses act as personal identification numbers for verifying the identity of authorised clients on wireless networks. However, existing encryption standards are not foolproof. A hacker can pick off authorised MAC addresses and steal bandwidth, corrupt or download files, and wreak havoc on an entire network. While securing a wireless LAN by using

an authorised list of MAC addresses for authentication will provide some security, they were never intended to be used in this way.

There are a few legitimate reasons and examples of why MAC address spoofing is done:

- A firewall could be set up to only accept traffic from a certain MAC address at a certain time. An administrator could generate a list of MAC addresses that would change after certain number of days, hours, or even minutes. The user would have to set their MAC address within the time window to send packets to the firewall.
- Some ISPs (Internet Service Providers) keep track of the MAC address that a subscriber is using. They only allow registered addresses to connect to the Internet, and charge more money for additional IP clients. It might become inconvenient to be limited to a particular MAC address if a user needs to change the gateway or change cards in the gateway temporarily and would have to re-register a new MAC address just to move some equipment around for a few days.

Even if you are using encryption or virtual private networks (VPNs), MAC addresses are always in the air. With software such as Kismet or Ethereal, a hacker can capture the MAC address of an authorised user. They can then change their MAC address to the valid user's MAC address using any number of spoofing or cloning utilities, or even manually changing the Windows registry entry. Now the hacker can connect to the wireless LAN and bypass any MAC address filtering. Netstumbler can also be used with a MAC spoofing utility or MAC address modifying utility such as SMAC (Spoofer MAC) to achieve the same results.

3.3.5 Noisy Neighbours

The proximity of other wireless networks and equipment to that of your own is of utmost importance, for these can be a cause of noise within your network. Because we are dealing with radio waves passing through the air, unwanted radio signals can wander into our domain from outside sources such as cordless phones, microwave ovens, or the neighboring IEEE 802.11 router. This noise, or interference can have a drastic effect on network performance and reliability.

Aside from the noise related issues, network users within an earshot of your access point could be consuming the bandwidth. Windows XP's built-in Wireless Zero Configuration utility, for example, is set up by default to join the wireless network with the best signal. Once it has successfully connected,

it stores the network SSID as a "preferred network" and will connect to it each time it comes within range. Though this is convenient in most circumstances for the network client, it can lead to unwanted network users. Even with WEP enabled, which can keep unwanted clients from joining your network; would-be clients knocking on the door, requesting connections, can consume significant bandwidth.

3.3.6 Man-In-The-Middle Attacks

Hailing from the early days of cryptography, man-in-the-middle (MitM) attacks are an old strategy applied to a new technology. The key concept behind a MitM attack is exactly as it sounds, one entity with malicious intent intercepts a message between two communicating entities (See Fig. 3.6). The hijacker can then send the message onto the receiver as if it had never been delayed, and even alter the message's content. Used in war, this could be a valuable tool for intercepting and altering the enemy's message to suit the opposing side's purpose. In World War II, if the Axis forces needed to send information to deployed troops, they would send it with a decryption key. This key would be the primary tool for decoding the message and properly deciphering it. If the Allies could intercept this message and break the code, then they would be

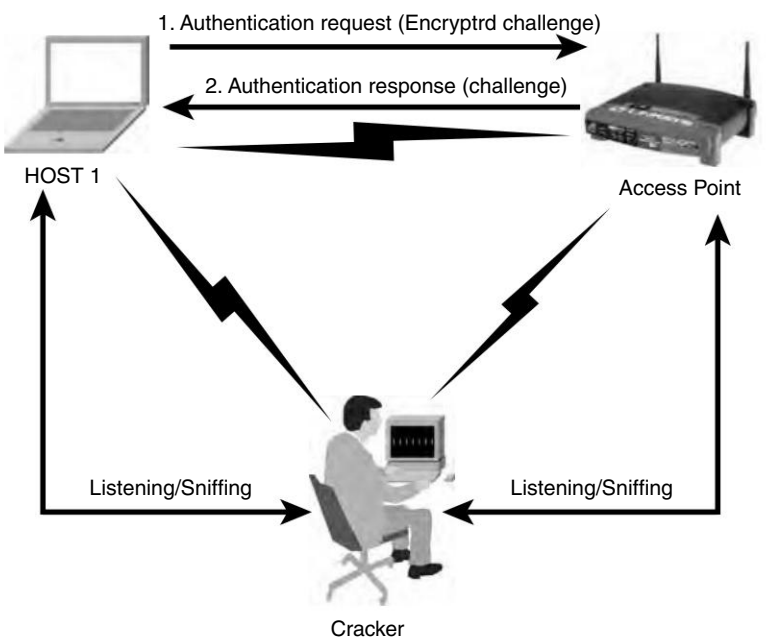


Fig. 3.6 *MitM attack*

executing a MitM attack. Upon successful completion of the attack, they would have three options as to how they would like to exploit the position—

- The message could be intercepted, altered and sent onto the recipient with fraudulent information.
- The message could be blocked and prevented from proceeding any further.
- The message could simply be read and sent on its way without the recipient's knowledge.

3.4 SECURITY FOR 1G WI-FI APPLICATIONS

3.4.1 Security Issues

In order to strengthen the security of wireless devices, it is necessary to understand the security concerns experienced with 1G (first generation) WLAN products. The three main security holes are

1. Equipment has security settings disabled by default,
2. Minimal security is easily broken, and
3. Rogue access points are easy to deploy and difficult to detect.

If users are to be productive no matter where they are, then WLANs have to be integrated seamlessly with the wired network in the organisation; such WLANs need to meet the twin requirements of security and mobility. Other issues also affect security in first generation WLANs. Human factors such as lack of awareness and lack of adherence to usage policies can cause loopholes in systems that have been secured in technical aspects. On the other hand, technical factors such as lack of encryption can cause loopholes in systems wherein the personnel are security-conscious and adhere strictly to security and usage policies.

3.4.2 Security Features

The minimal set of security features included in the 802.11b standard include the following:

1. Service Set Identifier (SSID)

Each access point has an SSID which identifies it to devices on the WLAN. The network can be configured so that clients are required to know the SSID of the access point before connecting to it.

2. MAC Address Filters

The access point can be configured to accept connections only from clients with MAC addresses registered with the access point.

3. WEP (Wired Equivalent Privacy) Encryption

The IEEE 802.11 standard included WEP as a mechanism for providing confidentiality that is subjectively equivalent to the confidentiality of a wired LAN medium that does not employ cryptographic techniques to enhance privacy.

3.4.3 Security Vulnerabilities

Several security vulnerabilities exist in 1G WLANs. Some of them include the following.

1. By default, the access point broadcasts its SSID in clear text. Even if the SSID broadcast is hidden, the client broadcasts the SSID to the access point while attempting to connect to it.
2. The MAC address of a valid client can be sniffed off the network and then spoofed by the rogue client.
3. WEP encryption is easily cracked. WEP only authenticates the client. This allows a rogue access point to capture data sent by an authenticated client.
4. A rogue access point (AP) can be installed that will intercept traffic from wireless clients.
5. Man-in-the-middle attacks can be launched by forcing an access point off its channel and then spoofing the SSID of the access point.
6. WLANs are easily crashed by denial of service (DoS) attacks with methods ranging from flooding the access point with spoofed MAC addresses to using devices like 2.4 GHz cordless phones to cause excessive radio interference.

3.4.4 Security Controls

Specific security methods that can be implemented to secure IEEE 802.11 wireless networks 1G applications include any or all of the following:

1. Turning off the broadcast SSIDs.
2. Introducing automated MAC-based access control mechanisms.
3. Enabling WEP encryption.

4. Lowering the power levels of the access points to limit the ability of hackers to connect from outside the specified boundary. This can also be accomplished by limiting connections to transmission rates of 11 Mbps and 5.5 Mbps.

3.5 SECURITY FOR 2G WI-FI APPLICATIONS

Most breaches in security of wireless devices are a result of a variety of layer two vulnerabilities. Protection against these vulnerabilities require defense in depth, i.e. multiple controls. The IEEE 802.1X task group addresses the problems of network security and access control. The IEEE 802.11i group has mandated the use of the 802.1X suite of protocols to improve and standardise wireless encryption. Such protocols include the Extensible Authentication Protocol (EAP), Protected EAP (PEAP) and Tunneled Transport Layer Security (TTLS), which supersede the weak WEP keys available for 1G WLANs. The 802.11i standard provides for using Temporal Key Integrity Protocol (TKIP) and Advanced Encryption Standard (AES) to encrypt data on the wireless network.

3.5.1 Encryption and Virtual Private Networks

Although WEP is flawed, it is still valuable as a first line of defence. WEP can prevent most novice hacker attacks and possibly delay intrusions by unauthorised users. The Wi-Fi Alliance come out with the industry-supported implementation of Wi-Fi Protected Access (WPA) as an interim protocol before AES becomes the standard. However, with WPA, enterprise managers discovered that the WLAN was secure but not truly mobile and reflective of user demands. Due to user roaming and the resulting changes in static IP addresses, WPA required re-authentication, which posed problems with 2G WLAN implementations. In addition, poorly chosen, short, human-readable passphrases used in WPA can be cracked with a robust dictionary attack offline and without access to the network. However, identity-based schemes are being adopted, which provide security without sacrificing mobility.

The draft release of the IEEE 802.11i standard includes features to address several of the vulnerabilities inherent to WEP. TKIP is one of several such protocols being offered by various vendors. Although TKIP still uses the RC4 encryption algorithm, it removes the weak key problem by forcing a new key to be generated every 10,000 packets or 10 kb. It also hashes the initialisation vector values that WEP sends as plaintext. TKIP includes a method for verifying the integrity of the data called the Message Integrity Check, which

mitigates the vulnerability that allows a hacker to inject data into a packet in order to deduce the encryption key. AES is the newest encryption standard and is under review for inclusion in 802.11i. AES is the strong encryption replacement for Data Encryption Standard. According to the 802.11i standard, AES will replace WEP and RC4 encryption. This will require a hardware optimisation to be able to handle the more robust algorithm.

Wireless networks are vulnerable by default. An additional safeguard that can be used to secure a wireless network is a Virutal Private Network (VPN). A VPN solution uses a combination of tunneling, encryption, authentication and access control. A VPN establishes a secure, encrypted network tunneled within a potentially hostile network like a wireless network.

IEEE 802.1x has a port-based access control method that provides a better way to control access to network ports. 802.1x does not specify an authentication method, although the most common approach for WLANs is EAP (Extensible Authentication Protocol), which is a framework for a variety of authentication methods. The specific method is determined by the client and access point during the authentication process. The EAP client (supplicant) contacts the access point (authenticator), which challenges the client for authentication information. The authenticator receives this information from the client and then passes it onto an authentication server for validation. No other communications from the client is permitted until the authentication server has validated the logon request. If the logon is accepted, the authentication server generates a WEP key specifically for the client and sends it through the access point to the client. The client is now permitted to access the network behind the access point.

There are several implementations of EAP, including the following:

1. Transport Layer Security (EAP-TLS): developed by Microsoft and used in IEEE 802.1x clients for Windows XP. EAPTLS provides strong security, but requires each WLAN user to run a client certificate.
2. Lightweight EAP (LEAP): developed by CISCO and used in their Aironet solution. LEAP supports dynamic WEP key generation and provides for fixed password user authentication.
3. Protected EAP (PEAP): PEAP does not require certificates for authentication. It supports dynamic WEP key generation and provides options for password, token or digital certificate based user authentication.
4. Tunneled Transport Layer Security (EAP-TTLS): It is developed as a competing standard for PEAP. EAP-TTLS supports password, token or certificate, side user authentication. Unlike EAP-TLS, EAP-TTLS requires only the server to be certified.

3.5.2 Wireless Gateways

Access points and WLAN clients are not designed to handle the large amount of overhead imposed by these additional layers of security. This becomes especially apparent as WEPs RC4 encryption is replaced by the more robust AES (Advanced Encryption Standard) encryption. One solution that is being implemented is the wireless gateway. Instead of access points connecting directly to the internal network, they are collectively connected to a device that contains the additional security levels where encryption and authentication are implemented. This configuration has the added advantage of simplifying roaming between access points without requiring additional authentication and the ability to implement Quality of Service at a single point.

3.5.3 Policies, Training and Awareness

No matter how much technology is employed at securing a wireless network, it will not be effective unless there are adequate policies in place along with security awareness training. Just as an enterprise has a security policy for a wired network, it should also have a strong policy on securing its wireless network. Important components of such a policy would include the following:

- **Physical Location of Access Points**

Suggestions range from concealing the access points to avoid vandalism, to shaping the radio waves by appropriate positioning of the antenna, to adjusting the power levels to prevent the signal from bleeding outside.

- **Logical Location of Access Points**

Access points should typically be placed in the DMZ (demilitarised zone) screened from the corporate network by a properly configured firewall.

- **Rogue Access Points**

A ban should be enforced on rogue access points. Consequences for violators should be stringent and strictly enforced.

- **Peer-to-Peer Mode**

The ad-hoc or peer-to-peer mode on clients should be disabled by default.

- **Configuration**

Properly configuring all devices, i.e., encryption, authentication and SSID is essential.

- **Interoperability**

Requiring that standard equipment be purchased from a single vendor wherever possible will increase interoperability and compatibility.

- **Site Surveys**

Frequent site surveys to locate any rogue access points and clients set up in ad-hoc mode.

- **Monitoring**

Frequent monitoring of the logs to ensure that intrusions have not occurred.

- **Updates and Patches**

Patch management policies are important to obtain timely updates.

- **Other**

References to other security policies to ensure consistency and integration with wired networks.

3.6 RECENT SECURITY SCHEMES FOR WI-FI APPLICATIONS

In this section we present some of security schemes developed for the benefit of Wi-Fi applications.

3.6.1 Software based Generic Authentication Schemes for Mobile Communication

Passwords

A password is a well-known example of a knowledge factor. Though this is frequently used, the use of only a knowledge factor for authentication is seen as a weak mechanism. One of the major problems is that users need to remember their password. Either a password is too easy and therefore simple to guess or complicated which often results in writing the password down. In combination with other authentication factors it is considered as a strong mechanism.

A generic password-level authentication system (usually hashes the password of the user with the help of hash function derived from a secret key cryptographic function. The hashed password is stored on the server in order

to preclude stealing the password by the adversary. The pseudocode of the password-based security is given in Table 3.1.

Table 3.1 *Generic Password Scheme*

SUMMARY: A identifies to B using static password.

1. *One-time setup.*
   (a) User A begins with a secret w. Let H be a one-way function.
   (b) A transfers (the initial shared secret) $w_1 = H(w)$, in a manner guaranteeing its authenticity, to the system B.
2. *Protocol messages*. $A \rightarrow B : w_1$
3. *Protocol actions*.
   (a) B computes $w_2 = H^{-1}(w_1)$
   (b) B accepts the password iff $w_1 = w_2$.

Image-based Authentication

Image-based authentication is developed over a user's successful authentication of image password set. After the username is sent to the authentication module, it responds by displaying an image set, which consists of images from the users password set mixed with other images. The user is authenticated by correctly identifying the password images. The experimental results clearly indicate that image-based authentication has an advantage over password or PIN-based authentication especially in regards to the human interface aspect. It is easier for a user to memorise an image than a text, and also easier to identify an image than recalling a text. Picture password authenticates a user through the selection of images displayed on a handheld device. Having the ability to tailor the display interface with personal images gives users a sense of freedom, and control.

An image recognition-based authentication scheme for wireless networks, which authenticates a user through his/her ability to recognise previously seen images. The working of Deja Vu scheme designed for wireless networks is given in Table 3.2.

Table 3.2 *Deja Vu Scheme*

Begin

1. The user A creates an image portfolio, by selecting a subset of p images out of a set of sample images.
2. The system presents a challenge set, consisting of n images.
3. To authenticate, the user must correctly identify the images which are part of portfolio.

End

Another algorithm based on image passpoints is given in Table 3.3, in which the image is coded by the server using JPEG2000, and subdivided into non-overlapping tiles of m x n pixels. The user access for the first time to the authentication system, and selects k points (passpoints) of the image.

Table 3.3 *Image Passpoints Algorithm*

Begin
1. The user authentication session consist of *p* input stages.
2. The user chooses *k* points in the image, these are called the passpoints.
3. Selected points are sent to the server.
4. The server evaluates the distance between selected points, and passpoints. If such distance is lower than a threshold then the user is authenticated.

End

Image-based authentication is not foolproof, brute force attacks can still be a problem for an image-based system. The user should not be overloaded with images, which results in limiting the number of combinations of images used for authentication. Another problem is shoulder surfing, where an intruder has a vantage point allowing them to see the password that is entered.

Tokens and Keys

Users present something they have in their possession for authentication. Examples of these are tokens or keys. Possession factors can be divided in two types: smart cards and tokens.

Many systems have built security proceedings on top of the challenge response system, as is the case with the idenitifier. In combination with the token, and its accompanying PIN code are required for access. Another type is tokens, which generate a different code every sixty seconds. Only the access control unit knows which number is valid at that moment of time for that user/authentication combination. Because the number changes every minute, a hacker cannot use a recorded code to login to the system later on. There are two methods for token authentication: Asynchronous (challenge/response) methods require the server to send the token device an encrypted message. The token device uses a preset algorithm and shared secret to decrypt the message and respond with the correct password encrypted with the shared secret. Synchronous methods require a server and token device to simultaneously calculate a challenge message using parameters from a time counter or login event counter, or both. If the calculated messages match, the authentication is successful.

For illustration we choose Hwang-Li's Scheme for smart card authentication (Table 3.4). There are three phases in the Hwang-Li's scheme: the registration phase, login phase, and the authentication phase. In the registration phase, the user, *U* sends a request to the *AS* (Authentication server) for the registration. The *AS* will issue a smart card, and a password to every user legal through a secure channel. In the login Phase, when the user, *U* wants to access the *AS*, it is required to insert the smart card to the smart card reader, and then keys the identity, and the password to access services. In the authentication phase, the *AS* checks the validity of the login request.

Table 3.4 *Hwang-Li's scheme for smart card authentication*

Begin

1. A user, *U* submits her/his *ID* to the *AS*. *AS* computes the password *PW* for the user *U*, as, $PW = ID^{x_s} \bmod p$, where X_s is secret key and *p* is a large prime number.
2. *AS* provides a password *PW*, and a smart card to the user *U* through a secure channel.
3. The smart card contains the public parameters (*f*, *p*), where *f* is a one-way function.
4. User *U* attaches her/his smart card to the smart card reader, and keys *ID*, and *PW*.
5. Smart card computes $C_1 = ID^r \bmod p$, and $t = f(T \oplus PW) \bmod p - 1$, where *T* is the current date, and time of the smart card reader and *r* is a random number.
6. Smart card computes $M = ID^t \bmod p$, and $C_2 = M(\mathrm{PW})^r \bmod p$.
7. Smart card $\rightarrow AS : C = (\mathrm{ID}, C_1, C_2, T)$.
8. *AS*: Check the format of *ID*. If the identity format is not correct, then *AS* will reject this login request.
9. *AS*: Check the legal time interval due to transmission delay, if not, then rejects the login request *C*.
10. IF $C_2(C_1)^{x_s - 1} = ID^{f(T \oplus PW)} \bmod p$, then the *AS* accepts the login request. Otherwise, the login request will be rejected.

End

Biometrics

The user is expected to present something of his/her physical attributes, e.g., eye, hand, face or voice. Better security is achieved when using a combination of these factors, such as a possession factor with a knowledge factor. Examples of being factors are various: hand geometry systems, a digital signature created with a pen, speaker recognition. These techniques are better known as biometric systems. Most important characteristic of it is the usage of a unique physical attribute of a person for authentication. Because the user is needed for authentication, it is more related to authentication of persons than the other two factors.

The securephone project's, primary aim is to realise a mobile phone prototype in which a biometrical authentication enables users to deal secure,

dependable transactions over a mobile network. The securephone is based on a commercial PDA-phone, supplemented with specific software modules, and a customised SIM card. It integrates in a single environment a number of advanced features: access to cryptographic keys through strong multimodal biometric authentication; appending, and verification of digital signatures; real-time exchange, and interactive modification of documents, and voice recordings. By mounting a camera on the mobile device, the face images of the user can be caught almost constantly. This is attractive, because the cost for a camera is low, and modern devices such as mobile phone, and PDA normally have a camera installed.

Mobile Authentication in PKI Infrastructures

In a mobile environment a certificate is assigned to a mobile device, hereby identifying the user. In case the device is stolen when it was switched on, there is no extra authentication mechanism required to block the device. To introduce an extra mechanism, for example, a password or a dual slot for SIM cards can solve this problem. In the latter, an extra card, carried by the user, is necessary in order to use the certificate installed on the mobile device. However, this puts the added value of a certificate under question.

Many mobile devices are not yet equipped with a dual slot. Next to this, the algorithms used in a PKI infrastructure usually require a level of processing power many mobile devices have difficulties to provide. To overcome this problem the Wireless Application Protocol Public Key Infrastructure (WPKI) is used as a protocol that offers a slimmed-down version of PKI optimised for wireless communications.

3.6.2 Generating Digital Signatures on Mobile Devices

Due to the rapid development of mobile communication technologies, people can now use cell-phones, palmtops and PDAs to access the Internet anywhere and anytime. However, many mobile communication applications are faced with some common problems: privacy and security. Such applications include: *mobile-payment system; remote walk-through system; electronic-wallet; e-ticket system; image authenticating and exchanging*.

Digital signature can be represented as a secure base in such applications because it provides authentication, data integrity and non-reputation cryptographic services. Traditional digital signature schemes were based on asymmetric cryptographic techniques which made the signature computation very expensive. In spite of handheld devices can come in many shapes and be used for different purposes, they have the common limitations: (1) limited computational capability, and

(2) short battery life. If the traditional asymmetric cryptographic computations are executed on mobile devices, those devices would be blocked for a period of time, and drain batteries quickly.

There are quite some work on digital signatures for mobile devices. A work based on server-supported signature scheme employs a one-way function and traditional digital signature scheme. Signature servers were responsible for generating signature tokens and certification authorities to verify these tokens. Therefore the schemes' robustness depends on the reliability of those servers. The modified version of this work is called Server Aided Signature Scheme. In this scheme, users are involved in the generation of the signature token, giving the on-line feature and defending the DOS attack. Unlike the traditional digital signatures which often use the pair public/secret keys to generate non-reputation signature token, one way hash function to generate senders secret key and use this key to produce nonreputation signature token. This is due to the fact that generation of robust pair public/secret keys is computation intensive for handheld devices. The asymmetric cryptographic computation used in these schemes is expensive in clients because during the period of verification of digital signature, clients should decrypt the encrypted signature token to verify the associated content added by signature servers.

A Server Based Signature (SBS) scheme was proposed for mobile devices. Besides achieving the same security level of the traditional digital signature protocols, the SBS scheme also: (1) reduces the computation complexity on the mobile devices; (2) reduces the communication consumption between signer and verifier. Application results show that the scheme is very useful for mobile communication systems.

3.6.3 Reputation Systems in Wi-Fi Networks

Repunet was implemented as a series of data mining scripts and a Web visualisation of the collected data. The data mining scripts were designed to periodically read the states of the set of access points. The access points had to contain a custom Linux distribution and only Z-Com XI-626 Wi-Fi cards. If the scripts were found to cooperate with any other unrecognised access point type, they would have to be rewritten. The read state is then stored in a database. No optimising measures as to the database were taken, which is the source of unmaintainable growth of the data-base size. Most of the code is bound to a specific database structure and even though it uses XSL templates and XML configuration, it has been proven inadaptable and unmaintainable for the future use (Refer Table 3.5).

Table 3.5 *Reputation system algorithm*

1. *Input:* Sensors: APs contain additional code keeping eye on active connections, and central server: computes reputation from sensor data 2. *Operation:* (a) Log data on sensors. (b) Collect data at central server. (c) Data evaluation at central server, computes the reputation. (d) Security countermeasures are selected and executed based on feedbacks.

3.6.4 Location Dependent Data Encryption/ Decryption

GPS receiver is popularly used in our daily life, such as car navigation, fleet management, and so on. GPS PDA is equipped with most of the wireless communication capabilities, including GSM/GPRS/EDGE, quad-band GSM phone capabilities, IEEE 802.11g, etc. The size and weight of GPS PDA is close to the mobile phone. However, its computing power and programming interface are better than mobile phones. Unlike the mobile phones in which data transmission is mostly based on SMS (Short Message Service), the types and quantities of data transmitted via GPS PDAs must be diverse and huge just like desktop PCs. That is, the data transmission via mobile devices will become more and more frequent according to the above trend.

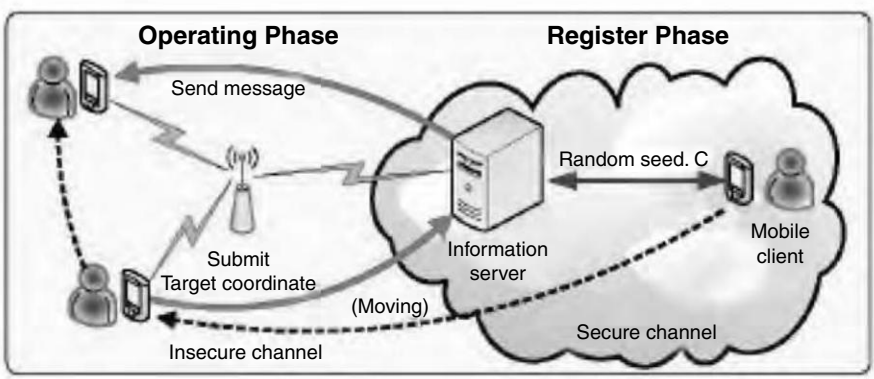


Fig. 3.7 *Location dependent security a scenario*

However, most of the data encryption techniques are location-independent. They cannot restrict the location of mobile clients for data decryption. A location-dependent approach is proposed for incorporating location information

into data transmission for mobile information system. The proposed approach is divided into two phases: in register phase and operation phase. A mobile client acquires a random seed of one-way hash function and MAC (message authentication code) function under a secure channel in the register phase. Then, data can be transmitted securely between information server and mobile clients in the operation phase. The mobile client transmits a target latitude/longitude coordinate for data encryption to information server. Then, the server encrypts the message and sends the ciphertext back to the mobile client. The client can only decrypt the ciphertext when the coordinate acquired from GPS receiver matches with the target coordinate, and the approach can meet the demands of mobile information system in the future. The pseudocode for working of the scheme is given in Table 3.6.

Table 3.6 *Location dependent data encryption*

1. *Registration:* User *A* acquires seed and MAC during the registration.
2. *Operation:*
   (a) $A \rightarrow$ *Information Server(IS)* : Target Latitude/Longitude coordinates.
   (b) $IS \rightarrow A$: *E(Message)*.
   (c) *A* aquires the coordinates from the GPS.
   (d) *A* performs *D(Message)* iff coordinates matches.

3.6.5 Personalised Firewalls

Traditionally, a firewall is considered as a set of components forming a gateway between two or more networks. Thus, a firewall has been a gateway which operates at the same time as a connector and a separator between the networks in a sense that the firewall keeps track of the traffic that passes through it from one network to another and restricts connections and packets that are defined as unwanted by the administrator of the system. Physically a firewall is a machine with appropriate software to do the tasks assigned to it. It can be a router, a personal computer (PC), or any other device that can be used for such purposes. Although firewalls are mostly used to connect Local Area Networks (LANs), i.e. internal networks, to the Internet and to protect against attackers or undesired traffic in general, they may also be used to separate and connect different segments of internal network for security purposes. The advantages of having a firewall are numerous. A firewall secures the network and can be used as a tool for monitoring the traffic especially from the outside to the inside of the network guarded by a firewall. Because all traffic intended for the internal network must pass through the firewall, most of the network security actions and policies can be concentrated in this particular point.

A computer device which can be connected to a home network and to a foreign network is provided with a local security mechanism, called a personal firewall, for protecting the computer device from attacks from the foreign network, in addition to or instead of a firewall in the internal network which protects the computer when connected to the internal network. The personal firewall is arranged to detect its current location, i.e., to determine the network to which it is connected at each particular moment, and to control its operation accordingly. The current location of the computer device is first determined on the basis of a currently used IP address of the computer device. Then this location determined on the basis of the current IP address of the computer device is verified by carrying out an additional location verification procedure with a predetermined network element.

One way to determine current location of the computer device is based on a currently used IP address of the computer device. This is based on the common practice that a computer device has a different IP address, either a fixed address or a dynamic address, in different networks. However, there are situations where the location determined on the basis of the current IP address is uncertain, i.e., the IP address fails to indicate the current location of the laptop. If the IP address does not match the current network, use of the Internet protocol (IP) to attack against the laptop is not likely, and one may reason that in that case a personal firewall does not need to be used. However, there is still a possibility that there is an attack using other protocols, such as NetBEUI or IPX. By detecting a situation where the IP address of the laptop is not an IP address of the current network, it is possible to block such protocols while in foreign networks. Further, NAT (network address translation) and private IP addresses are frequently used. This means that the same IP address can be in use in several networks. In that case it is not enough to trust IP address information only when determining the location of the network. It is even possible that while being connected to a hostile network, the DHCP (dynamic host configuration protocol) gives familiar IP address to make it easier to attack the laptop. The algorithm is given in Table 3.7.

Table 3.7 *Personalised firewalls*

1. Mobile device obtains IP address using DHCP.
2. *Operation:*
   (a) Determine the IP address of the network.
   (b) If the User changed the IP address of the device (which is not same as network sanctioned IP), then block that IP.

SUMMARY

In this chapter we have provided detailed discussions on various security level challenges and vulnerabilities for designing application for Wi-Fi networks. We have made an attempt to summarise all the possible attacks in first and second generation Wi-Fi applications. At the end we have provided various reserach initiatives towards securing the wireless-based applications.

REVIEW QUESTIONS

1. What are the Wi-Fi security issues?
2. What are the various types of attacks possible over Wi-Fi networks?
3. Mention various attack methods on Wi-Fi networks over applications.
4. Briefly explain the need of Wi-Fi networks in developing some of the applications.
5. Give two typical applications which use Wi-Fi networks.
6. What are the characteristics of first generation Wi-Fi networks applications?
7. What are the security features and issues of first generation Wi-Fi networks applications?
8. What are the various security controls of first generation Wi-Fi networks applications?
9. What are the characteristics of second generation Wi-Fi networks applications?
10. What are the security features and issues of second generation Wi-Fi networks applications?
11. What are the various security controls of second generation Wi-Fi networks applications?
12. Briefly explain the reputation systems in Wi-Fi networks.
13. Briefly explain the location dependent data encryption in Wi-Fi applications.
14. Briefly explain the role of personalised firewalls in protecting Wi-Fi applications.
15. Explain with algorithm the working of Hwang Li's scheme for smart card authentication.
16. Explain the working of Deja Vu scheme for image-based authentication.
17. What are the advantages of using multi-factor authentication over single-factor authentication?
18. What are the advantages and disadvantages of biometric authentication schemes?
19. Explain the mechanism of generating digital signatures on mobile devices.

Application Level Security in Cellular Networks

4

OBJECTIVES

- To understand security issues in cellular networks.
- To know about attacks on cellular networks.
- To discuss some of the available security attacks and solutions in various cellular technologies.
- To study a few security schemes proposed for various cellular technologies.
- To understand the working of public key infrastructure in mobile communications.

A cellular network provides cell phones or mobile stations (MSs), with wireless access to the public switched telephone network (PSTN). The service coverage area of a cellular network is divided into many smaller areas, referred to as cells, each of which is served by a base station (BS). The BS is fixed, and it is connected to the mobile telephone switching office (MTSO), also known as the mobile switching center. An MTSO is in charge of a cluster of BSs and it is, in turn, connected to the PSTN. With the wireless link between the BS and MS, MSs such as cell phones are able to communicate with wire-line phones in the PSTN. Both BSs and MSs are equipped with a transceiver. Figure 4.1 illustrates a typical cellular network, in which a cell is represented by a hexagon and a BS is represented by a triangle.

The main objective of security in cellular mobile communication systems is to secure the conversations and signalling data from interception as well as preventing the system from other frauds such as cloning and the like. With the older analog-based cellular telephone system such as AMPS (Advance Mobile

Phone System), it is relatively simple matter to intercept cellular telephone conversation with a police scanner. Cellular companies lose a substantial amount of money per year to cellular fraud. One cause of this fraud is cloning of cellular telephones. Federal Communications Commission (FCC) estimated that the cellular industry loses more than 650 million per year to fraud. As a result of this, the Wireless Telephone Protection Act, expanding the prior law to criminalise the use, possession, manufacture or sale of cloning hardware or software.

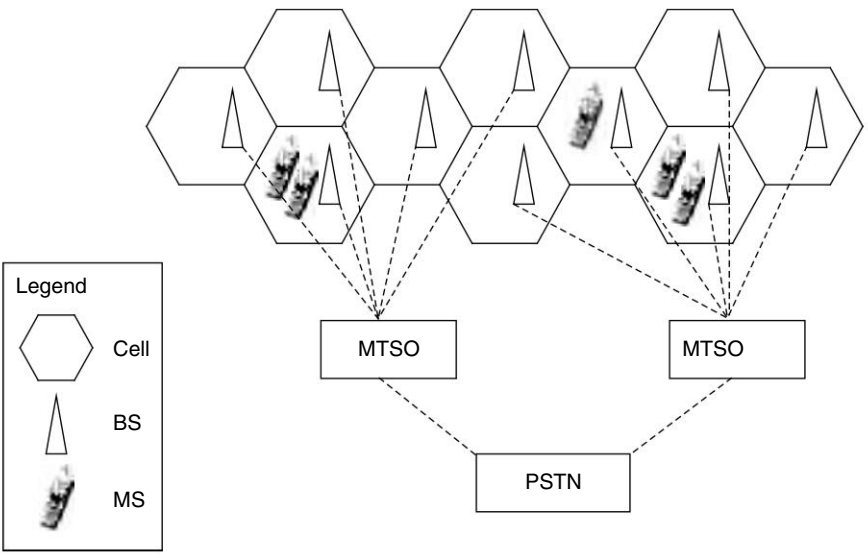


Fig. 4.1 *Typical cellular network*

4.1 GENERATIONS OF CELLULAR NETWORKS

Cellular Networks have been around since the beginning of 19th century. As the demand increased and technology improved they have gradually grown into versatile and sophisticated networks. We discuss the features of the cellular networks as per the generation.

4.1.1 1G

First generation (1G) networks were the first cellular networks introduced in the 1980s. They were only capable of transmitting voice at speeds of about 9.6 kbps max. In the US the system was known Advanced Mobile Phone

System (AMPS) and in Europe the Nordic Mobile Telephony (NMT). Both these technologies used analog modulation to transmit data as a continuously varying waveform.1G systems had some limitations such as no support for encryption, poor sound quality and inefficient use of the spectrum due to their analog nature.

4.1.2 2G and 2.5G

GSM is the most widely adopted 2G technology. Although it was initially employed in Europe, it has become a global technology with subscribers in many countries.

High-Speed Circuit-Switched Data (HSCSD)

This was the first attempt at providing data at high speeds data over GSM, with speeds of up to 115 kbps. This technique cannot support large bursts of data. HSCSD was not widely implemented and GPRS became a more popular technique.

General Packet Radio Service (GPRS)

This technique can support large burst of data transfers. In order to support this two new elements have to be added to existing networks. Service GPRS support node (SGSN) for security mobility and access control and Gateway GPRS support node (GGSN) in order to connect to external packet switched networks.

Enhanced Data Rates for GSM Evolution (EDGE)

The standard GSM uses GMSK modulation. Edge uses 8-PSK modulation. GPRS and EDGE combined provide data rates of up to 384 kbps.

Cellular Digital Packet Data (CDPD)

CDPD is a packet based data service. CDPD is able to detect idle voice channels and uses them to transfer data traffic without affecting voice communications. CDMA-One, also known as IS-95a was the initial technique. This technique allows users to use the entire spectrum and can support more users than TDMA and GSM. Speed between 4.8 and 14.4 kbps can be supported. The CDMA-Two extension can provide data rates of up to 115.2 kbps.

The 2.5G extension to this technology can be divided into two techniques—1xEV-DV uses one radio frequency channel for data and voice, whereas 1xEV-DO uses separate channels for data and voice. These are fully compatible with

both CDMA-One and its 3G replacement CDMA2000, to make the transition as easy as possible.

4.1.3 3G

3G is the next generation wireless cellular network whose aim is to provide a world wide standard and a common frequency band for mobile networking. The International Telecommunication Union (ITC) started the process in 1992, the result of this effort was a new network infrastructure called International mobile telecommunications 2000 (IMT-2000), with the 2000 signifying that this new technology has data rates of up to 2000 Kbps with 2000 MHz frequency range. The following are the list of objectives at IMT-2000.

1. To make a wide range of services, both voice and data available to users, irrespective of location.
2. To provide services over a wide coverage area.
3. To provide the best quality of service (QoS) possible.
4. To extend the number of services provided subject to constraints like radio transmission, spectrum efficiency and system economics.
5. To accommodate a great variety of mobile stations.
6. To admit the provision of service by more than one network in any area of coverage.
7. To provide an open architecture to permit the easy introduction of technology advancements as well as different applications.
8. To provide a modular structure which allows the system to start from small and simple configuration and grow as needed, both in size and complexity within practical limits.
9. The 3rd generation gives specifications for UMTS, a 3G technology based on Universal Terrestrial Radio Access (UTRA) radio interface and the extended GSM/GPRS network. A second radio interface also exists called IMT Multicarrier (IMT-MC) which is being promoted by the 3GPP2 organisation. This interface is backward compatible with IS-95 to make a seamless transition to 3G. This proposal is known as CDMA2000.

4.1.4 4G

4G is the fourth-generation wireless, which is going to provide the stage for broadband mobile communications which will supersede the 3G. 4G is expected to provide end-to-end IP and high-quality streaming video as its

distinguishing features. Fourth generation networks are likely to use a combination of WiMAX and WiFi.

Technologies employed by 4G includes SDR (Software-defined radio) receivers, OFDM (Orthogonal Frequency Division Multiplexing), OFDMA (Orthogonal Frequency Division Multiple Access), MIMO (multiple input/multiple output) technologies, UMTS and TD-SCDMA. All of these delivery methods provide high rates of data transmission and packet-switched transmission protocols. When fully implemented, 4G is expected to enable pervasive computing, in which simultaneous connections to multiple high-speed networks provide seamless handoffs throughout a geographical area.

Following are some of the services provided by 4G networks.

Push and Pull Services

These services rely on the networks ability to locate subscribers. In 4G, it is envisioned that networks will be able to pinpoint the exact location of subscribers, both indoors and outdoors. This ability will make it possible for value-added functionality to be offered by service providers. Push and pull services are further enhanced by user profiles. User profiles, established and updated by subscribers, assure that information to each user is truly customised. User profiles contain the subscribers' preferences (e.g., likes/dislikes, schedules, and formats) and permissions (i.e., who is allowed to know who and where they are).

The users profile would reside in a database maintained by the service provider. For example, if a user likes a particular type of food, the network will see the preference in the users profile and will push information regarding restaurants that serve that type of food in the general locale of the user. Similarly, the user will be able to request this same information from the network (pull) if he/she chooses not to have this information pushed to the wireless device.

Location-based Services

The challenge with location-based services is not in the applications but in the implementation. For location services to be of any real value, the network must be able to determine the location of subscribers to a high degree of accuracy perhaps to within a few feet.

4G involves Internet Protocol version 6 (IPv6) to route data packets to the handset. IPv6 has built in location tracking that will enhance the network's ability to pinpoint a subscriber's location. There are proposals for applying global positioning system (GPS) capabilities in handsets to locate subscribers.

Entertainment Services

Entertainment services are viewed by service providers as having the greatest potential for immediate return on investment. Entertainment services may include streaming audio, streaming video, chat, photo trading, and gaming.

4.2 SECURITY ISSUES AND ATTACKS IN CELLULAR NETWORKS

The infrastructure for cellular networks is massive, complex with multiple entities coordinating together, such as the Internet coordinating with the core network, and therefore throws a challenge for the network to provide security at every possible communication path. Compared to Wired Networks, Wireless Cellular Networks have the following limitations.

1. Open Wireless Access Medium

Since the communication is on the wireless channel, there is no physical barrier that can separate an attacker from the network.

2. Limited Bandwidth

Although wireless bandwidth is increasing continuously, because of channel contention everyone has to share the medium.

3. System Complexity

Wireless systems are more complex due to the need to support mobility and making use of the channel effectively. By adding more complexity to systems, potentially new security vulnerabilities can be introduced.

4. Limited Power

Wireless Systems consume high power and therefore with the existing technology they have a limited time battery life.

5. Limited Processing Power

The processors installed on the wireless devices are increasing in power, but still they are not powerful enough to carry out intensive processing.

6. Relatively Unreliable Network Connection

The wireless medium is an unreliable medium with a high rate of errors compared to a wired network.

There are several security issues that have to be taken into consideration when deploying a cellular infrastructure. They are the following:

1. Authentication

Cellular networks have a large number of subscribers, and each has to be authenticated to ensure the right people are using the network. Since the purpose of 3G is to enable people to communicate from anywhere in the world, the cross region and cross provider authentication becomes an issue.

2. Integrity

With services such as SMS, chat and file transfer it is important that the data arrives without any modifications.

3. Confidentiality

With the increased use of cellular phones in sensitive communication, there is a need for a secure channel in order to transmit information.

4. Access Control

The Cellular device may have files that need to have restricted access to them. The device might access a database where some sort of role based access control is necessary.

5. Operating Systems in Mobile Devices

Cellular Phones have evolved from low processing power, ad-hoc supervisors to high power processors and full fledged operating systems. Some phones may use a Java Based system, others use Microsoft Windows CE and have the same capabilities as a desktop computer. Issues may arise in the OS which might open security holes that can be exploited.

6. Location Detection

The actual location of a cellular device needs to be kept hidden for reasons of privacy of the user. With the move to wireless networks, the issue arises that a user may be associated with an access point and therefore their location might be compromised.

7. Viruses and Malware

With increased functionality provided in cellular systems, problems prevalent in larger systems such as viruses and malware arise. The first virus that appeared on cellular devices was Liberty. An affected device can also be used

to attack the cellular network infrastructure by becoming part of a large scale denial of service attack.

8. Downloaded Contents

Spyware or Adware might be downloaded causing security issues. Another problem is that of digital rights management. Users might download unauthorised copies of music, videos, wallpapers and games.

9. Device Security

If a device is lost or stolen, it needs to be protected from unauthorised use so that potential sensitive information such as emails, documents, phone numbers etc., cannot be accessed.

4.2.1 Attacks on Cellular Networks

Due to the massive architecture of a cellular network, there are a variety of attacks that the infrastructure is open to. We describe some of attacks in this subsection.

1. Denial of Service (DOS)

This is probably the most potent attack that can bring down the entire network infrastructure. This is caused by sending excessive data to the network, more than the network can handle, resulting in users being unable to access network resources.

2. Distributed Denial of Service (DDOS)

It might be difficult to launch a large scale DDOS attack from a single host. A number of hosts can be used to launch an attack.

3. Channel Jamming

Channel jamming is a technique used by attackers to jam the wireless channel and therefore deny access to any legitimate users in the network.

4. Unauthorised Access

If a proper method of authentication is not deployed then an attacker can gain free access to a network and then can use it for services that he might not be authorised for.

5. Eavesdropping

If the traffic on the wireless link is not encrypted then an attacker can eavesdrop and intercept sensitive communication such as confidential calls, sensitive documents, etc.

6. Message Forgery

If the communication channel is not secure, then an attacker can intercept messages in both directions and change the content without the users ever knowing.

7. Message Replay

Even if communication channel is secure, an attacker can intercept an encrypted message and then replay it back at a later time and the user might not know that the packet received is not the right one.

8. Man-in-the-Middle Attack

An attacker can sit in between a cell phone and an access station and intercept messages in between them and change them.

9. Session Hijacking

A malicious user can hijack an already established session, and can act as a legitimate base station.

The attacks on the cellular systems can take place through air (wireless) or through wirelines. To understand this, it is important to know that every connection from a cellular phone to a regular telephone involves the following types of communication: (1) air communication between the cellular phone to a nearest cell base station, (2) wirelined communication between the cell base station and a cellular switch station, and (3) wirelined communication between the cellular switch and the destination through the conventional Public Switched Telephone Network (PSTN).

The most severe attack to the cellular systems through the air is phone cloning. Unlike a regular telephone which can be recognised by a uniquely distinguishable wire, a cellular phone is only recognised by a pair of uniquely assigned numbers: ESN (Electronic Serial Number) and MIN (Mobile Identification Number). Such pairs of numbers are transmitted to a cell base station through the open air whenever the cellular phone is powered on. Some cellular phones are equipped with PINs (Personal Identification Numbers). In such cases, when placing a call, the PIN will need to be sent through the assigned voice channel after ESN and MIN are sent through a control channel. Such cellular phones are less likely to be cloned. However PINs are vulnerable to

eavesdropping as well. In fact, there are equipments which can be used to trace the transmitted ESN/MIN/PINs in real-time.

Another possible attack through the air is hijacking. Once a voice channel is established between a cellular phone and a cellular base station, a counterfeit cellular phone may seize the voice channel by increasing its power level above that of the legitimate cellular phone. A criminal could then make an illegal cellular call.

4.3 GSM SECURITY FOR APPLICATIONS

GSM (Global System for Mobile communications) is the most popular standard for mobile phones in the world. GSM differs from its predecessors in that both signaling and speech channels are digital, and thus is considered a second generation (2G) mobile phone system. This has also meant that data communication was easy to build into the system. GSM networks operate in four different frequency ranges. Most GSM networks operate in the 900 MHz or 1800 MHz bands. One of the key features of GSM is the Subscriber Identity Module (SIM), commonly known as a SIM card. GSM was designed with a moderate level of security. The system was designed to authenticate the subscriber using a pre-shared key and challenge-response. Communications between the subscriber and the base station can be encrypted.

4.3.1 GSM Architecture

A GSM network can be divided into three areas: Mobile Station, Base Station Subsystem and Network Subsystem. A Mobile Station (MS) consists of two main elements—The Mobile Equipment and the Subscriber Identity Module. The Base Station Subsystem controls the radio link with the Mobile Station. The Network Subsystem performs the switching of calls between mobile users and between mobile and fixed network (ISDN, PST, etc.) users. As shown in Fig. 4.2, security parameters of the GSM are distributed among the SIM card, the Mobile Equipment and the GSM network.

4.3.2 GSM Security Features

Anonymity or Subscriber Identity Confidentiality

To restrict a possible mobile traffic interceptor being able to identify which subscriber is using a given resource on the air interface is anonymity. The use of temporary identifiers safeguards the identity of the GSM user. Instead of

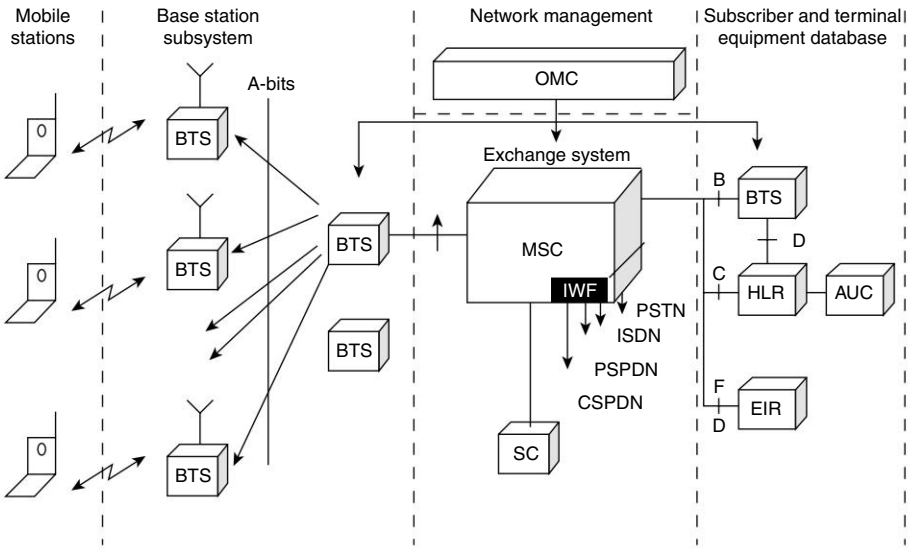


Fig. 4.2 *GSM Architecture*

using the IMSI, a new temporary mobile subscriber identity (TMSI) is allocated by the Public Land Mobile Network(PLMN) at least on every location update and used to identify a MS on the air interface. When a MS attempts access with a PLMN with which it is not presently registered, the MS uses its IMSI to identify itself. The IMSI is then authenticated by the PLMN, which results in the sharing of a cipher key (Kc). When the PLMN switch on encryption, the Visitor Location Register (VLR) generates a TMSI to the MS, storing the association of TMSI and IMSI in its database. The TMSI is then sent to the MS, encrypted with Kc. The next time the MS attempts access in that PLMN, it uses the TMSI previously allocated by the VLR instead of its IMSI. Then the PLMN looks up its table of TMSI to IMSI mapping to find the MS permanent identity. After a successful authentication and once an encrypted channel has been established, the PLMN assigns to the MS another TMSI. Figure 4.3 illustrates corresponding timing diagram.

Subscriber Identity Authentication

The authentication is required to identify the MS to the PLMN operator. The PLMN then knows who is using the system for billing purposes. This protects the PLMN from unauthorised use. GSM authentication is a one-way process, i.e., the visited PLMN is not authenticated. Authentication is performed by a challenge and response mechanism. Ki in the Home PLMN is held in the Authentication Center (AuC). A random challenge (RAND) is generated by the AuC and issued to the MS, via PLMN. The MS encrypts RAND using

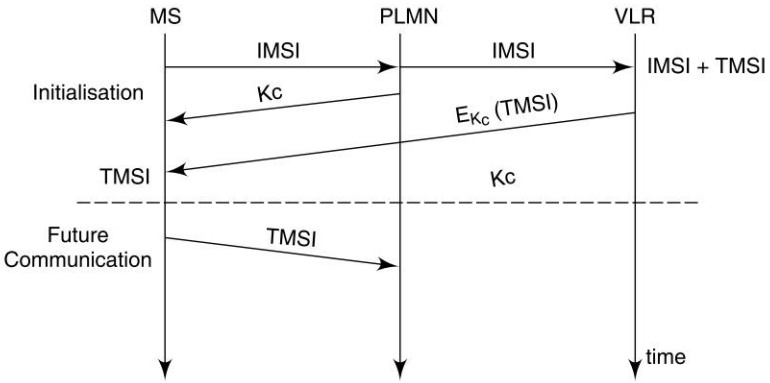


Fig. 4.3 *Timing diagram for Subscriber Identity Confidentiality*

Key (Ki) and the authentication algorithm A3 implemented within the SIM, and send a signed response (SRES) back to the PLMN. AuC performs the same process with RAND to compute the expected response (XRES), which is sent to the PLMN. The PLMN then compares the SRES and XRES and if equal then the user is authenticated. Figure 4.4 depicts timing diagram for subscriber ID Authentication.

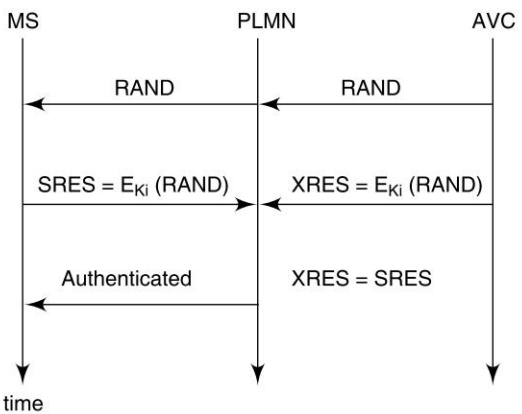


Fig. 4.4 *Timing diagram for Subscriber Identity Authentication*

Encryption of User Traffic and User Control Data

The GSM system uses symmetric cryptography and the data is encrypted using an algorithm which is seeded by the ciphering key the Kc. For decrypting the data same Kc is needed. The idea is that the Kc should only be known

by the phone and the network. Only in this case, the data is meaningless to anyone intercepting it. The Kc should also keep frequently changing, in case it is eventually compromised. The method of distributing the Kc to the phone is closely tied in with the authentication procedure discussed above.

Use of SIM as Security Module

The main security task of SIM is Key distribution, authentication and cipher key generation. SIM is implemented on a smart card tamper proof circuit, so that that it is impossible to extract the Ki. Technically, the SIM is required at the start of a call only. In GSM a call must close if the SIM is removed from the Mobile Equipment (ME) during a call to avoid parallel calls using a unique SIM (i.e., a stolen SIM). The ME passes the RAND received from the VLR to the SIM. Then SIM passes its Ki value and the received RAND through algorithm(s) A3/8. The resulting SRES produced by the SIM is passed back to the ME and then to the VLR, that verify if the SIM claimed identity can be authenticated. If the SIM is authenticated, the VLR passes Kc to serving BS. Then SIM passes Kc to the ME and as a result the BS and the ME can begin their operations. Figure 4.5 illustrates SIM Security aspects.

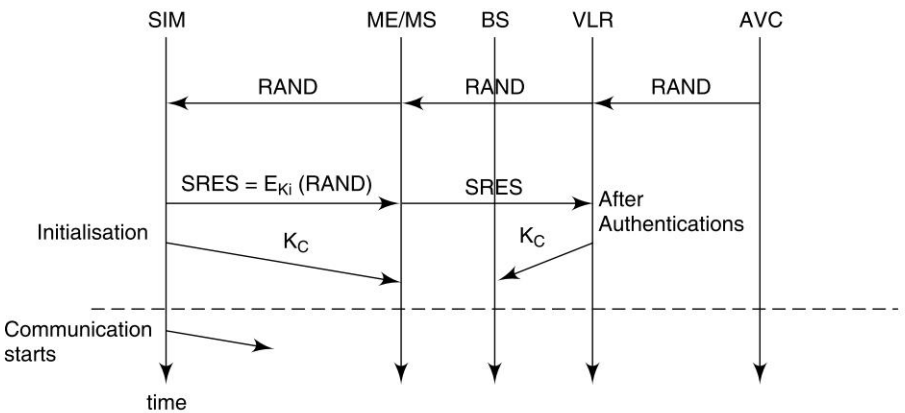


Fig. 4.5 *Use of SIM as Security module.*

4.3.3 GSM Security Attacks

SIM/MS Interface Tapping

It is possible for the SIM to be removed from one MS and put in another with which it has no previous association. As the SIM-ME interface is unprotected, it is possible for the SIM to be connected to terminal emulators instead of genu-

ine MS and consequently for messages on the SIM-MS interface to be tapped. However, unless the algorithms on the SIM are substandard, there is no advantage in compromising the SIM-MS interface in this way.

Attacks on the Algorithm A3/8

Comp128 is new version of A3/8 algorithm used by many GSM operators. It has one weakness, that if 160000 RAND-SRES pairs are collected then Ki can be found out some tried to crack comp128 algorithm using a PC emulator which sends 160000 chosen RANDS to the SIM card and receive the SRES. As the SIM card have very slow clock there fore it will take almost 10 hours to complete this process. Another method is to use false BS for sending the RAND over the air interface. However in this way a number of days will be required by the attacker to find the Ki. Thus the attacker can decrypt the data of the users.

Flaws in A5/1 and A5/2 Algorithms

In the attacks of these algorithms the database is searched for a match with a known key stream. After a match is found the database gives the correct Algorithm State, and then after simple computing, Kc can be calculated for decrypting purposes.

Attacks on the SIM Card

Subscriber identification module is implemented on a smart card and vulnerability in smart card will directly affect the security of the SIM as IMSI and Ki are stored on it. The attacks on SIM card are known as optical fault induction revealed by a study that showed that the operation of the smart card processor can be interrupted if exposed to an electric camera flash bulb. It has been done by using a camera flashgun and a microscope. The result suggests that the illumination of a target transistor causes it to conduct, thereby inducing a transient fault. Now by scratching the protective coating of the SIM micro processor circuit and focusing the flash light on individual transistors within the chip by beaming the flash through a micro scope. By this way they were able to reverse engineer the memory address map and extract the secret data of IMSI and Ki.

False Base Station

The GSM security system provides unilateral authentication. The MS/ME is authenticated to BS, but the BS is not authenticated to mobile equipment. This draw back of unilateral authentication allows attacks on the GSM system by a

false BS. At the time of GSM designing it was assumed that the system should not be subject to such attacks due to high amount of expenses as compared to other methods of attacking. But now the cost of GSM BS devices are too low and it is easy to use GSM BS emulators. This method is based on the fact that ciphering of a call does not start automatically, rather the ciphering starts when BS instructs the ME to start encryption. The instruction from BS to ME to start encryption can be manipulated during transit not to start encryption by an intruder.

4.3.4 GSM Security Solutions

The GSM specifications are not static; there have been many revisions to the standard, to add technologies such as GSM1800, HSCSD, GPRS and EDGE. In fact, the standard still evolves today into 3rd generation (3G) technologies such as UMTS. In recent versions of the standard, particularly UMTS, much has been done to improve upon the security flaws. Some of them are given below:

GSM—Newer A3/A8 Implementation

As discussed earlier, newer implementations of A3/A8 have been introduced, namely COMP128-2 and COMP128-3. So far these algorithms have held up reasonably well however they are developed in secret. COMP128-2 has the deliberate 10-bit weakening of the ciphering Key Kc. However. COMP128-3 is the same basic algorithm without this weakening (i.e. a truly 64-bit Kc). COMP128-2 and COMP128-3 have stopped SIM cloning and also make the serious attempts over the air Ki extraction unfeasible, even if they dont approach the ideal strength of 2^{128}.

GSM—A5/3 Ciphering

As mentioned previously, GSM supports up to 7 different algorithms for A5 (ciphering). Until recently, only the A5/1 and A5/2 algorithms were used. Later, GSM added a much stronger algorithm, A5/3 which is based on the Kasumi core (the core encryption algorithm for UMTS). Only few networks and handsets support this algorithm.

Public Key Infrastructure in Mobile Systems

Public key techniques are based on the use of asymmetric key pairs (see the Panel). Usually each user is in possession of just one key pair. One of the keys of the pair is made publicly available, while the other key of the pair is kept private. Because one of the keys is available publicly there is no need for a

secure out-of-band key exchange, however there is a need for an infrastructure to distribute the public key authentically.

Authentication is achieved by proving possession of the private key. One mechanism for doing this is a digital signature, which is generated with the private key and verified using the corresponding public key, i.e., by the public key bound to the entity generating the signature. Public key techniques make it possible to establish secret session keys dynamically. A simplified procedure is for one end-entity to calculate a secret session key and send it encrypted with the public key of the entity with which it wants to initiate a session. That entity then obtains the secret key by decrypting the received information with its private key.

As the public key of a key pair is usually published in a directory, the overhead associated with distributing key material to communicating parties is reduced significantly in comparison with solutions based solely on secret key techniques. For a group of n entities communicating with each other, only n key pairs are required. A drawback of public key techniques is that they are computationally very intensive, which makes them less suitable for devices of limited size and processing power, such as mobile phones. The advantages of the public key techniques described above do not come free. They must be paid for by additional organisational measures, and more sophisticated client logic. Furthermore, this additional overhead leads to extended user-interaction, for instance, in handling certificates.

Some of the examples of usage of PKI for mobile devices are given below:

- **Secure Browsing**

Network security protocols are probably the most common use of public key methodologies by wireless devices. The Open Mobile Alliance (OMA, formerly the WAP Forum) has specified a Wireless version of the IETF Transport Layer Security (TLS) protocol, known as WTLS, to secure mobile browsing. WTLS provided for a secure channel between the mobile phone and a WAP gateway, which however, did not satisfy the demand for end-to-end security in data networks. A later version of WAP (2.0) adopted the TLS protocol itself within WAP Transport Layer end-to-end Security specification.

- **Access to Enterprise Networks**

One of the greatest promises of 2.5G and 3G wireless networks is enabling mobile devices to access execute corporate applications, such as e-mail, file transfer, CRM and others. This raises the need for a Virtual Private Network (VPN) client application that will provide network layer security between the mobile device and the corporate gateway (or the end server). VPN clients

may be implemented at different layers, whereas the dominant implementation is within the Internet Protocol (IP) layer, using the IETF Internet Protocol Security (IPsec) protocol. VPN is already a powerful motive for enterprises to deploy public key infrastructure incorporating the set up of a Certificate Authority (CA) to deploy digital signatures. Once this infrastructure is in place for remote users, it can surely serve remote wireless users as well.

- **Mobile Payment Authentication**

Public key cryptography is considered as a preferred architecture for mobile commerce and banking. The most notable illustration for this is Visa Three-Domain Secure (3-D SecureTM) specification. Its architecture relies on the issuers ability to authenticate a remote cardholder by a pre-determined mechanism, where necessary data may be collected during the enrollment process. The 3-D Secure Wireless Authentication Scenarios specification presents several Using Public Key Cryptography in Mobile Phones authentication methods relevant for the wireless environment, including shared secret, signature and biometrics. The most secure scenario is that of a signature, that relies on public key cryptography. Local (proximity) transactions are also regarded as a future application of wireless phones. The Mobey forum has adopted the EMV (Europay, MasterCard, and VISA) protocol5 for these transactions, in Mobey Local Preferred Payment Architecture (Local PPA) specification.

- **Access Control**

The access control mechanism based on PKI is deployed for service access in cellular networks. The Mobile Electronic Transactions (MeT) group is working on a local authentication protocol called Personal Transaction Protocol (PTP) that will allow users to authenticate themselves at retail locations, ticket collection points, workstations, etc., using their cellular phones.

- **Digital Signatures on Mobile Transactions**

Digital signatures make public key cryptography a most practical tool in real-life applications, being the most reliable method for authentication and non-repudiation. As such, digital signatures are expected to become a fundamental element of mobile devices business applications, as they already are being used for signing transactions taking place in online banking and payment applications. A new concept for mobile transactions is called actionable alerts. These are constructed by a service provider sending a message to the mobile user, and the mobile user responding with an alert. A secure version of actionable alerts application, based on digital signatures and encryption, allows the banks to facilitate mobile platforms to secure banking transactions.

Similarly, other procurement transactions may be secured by engaging digital signatures, where the mobile user signs documents such as a contract, NDA, MOU, RFP, bids, etc.

- **Messaging**

Public key cryptography can also be used to secure other kinds of mobile messaging, such as SMS messages or wireless email applications using S/MIME (Secure/Multipurpose Internet Mail Extensions)—a specification for secure electronic mail messages in MIME format.

- **Content Authentication**

Code signing is an essential technology for mobile devices that enable application download over the air, such as Java applets. It is necessary, for such devices, to have the means to assure the safety of the downloaded code. The originator or the provider of the code may provide such assurance by digitally signing the code, via an XML digital signature, Java API or by other interfaces. The phone holds a trusted copy of the signers public key, for verifying the codes signature before using it. Code signing, does not in itself, certify the safety of the code, but it assures that the code was not originated or modified by illegal parties.

- **Digital ID**

A digital ID identifies its holder for multiple purposes, such as drivers' license, healthcare, insurance policy, etc. Digital IDs are implemented in the form of user credentials and associated certificates. The digital IDs are created and digitally signed by the relevant authority, according to their purpose. When used in wireless devices, digital IDs reside on the device, and can also be transferred (for example, in case of replacing the wireless device), either using a detachable card as intermediate medium, or over the air.

4.4 GPRS SECURITY FOR APPLICATIONS

GPRS as a new data service uses a packet-mode technique to transfer high-speed and low speed data and signalling in an efficient manner. GPRS optimises the use of network and radio resources. Strict separation between the radio subsystem and network subsystem is maintained, allowing the network subsystem to be reused with other radio access technologies. GPRS does not mandate changes to an installed MSC base. GPRS network elements (Refer Fig. 4.6) are SGSN (Serving GPRS Support Node), GGSN (Gateway GPRS Support Node), Bor-

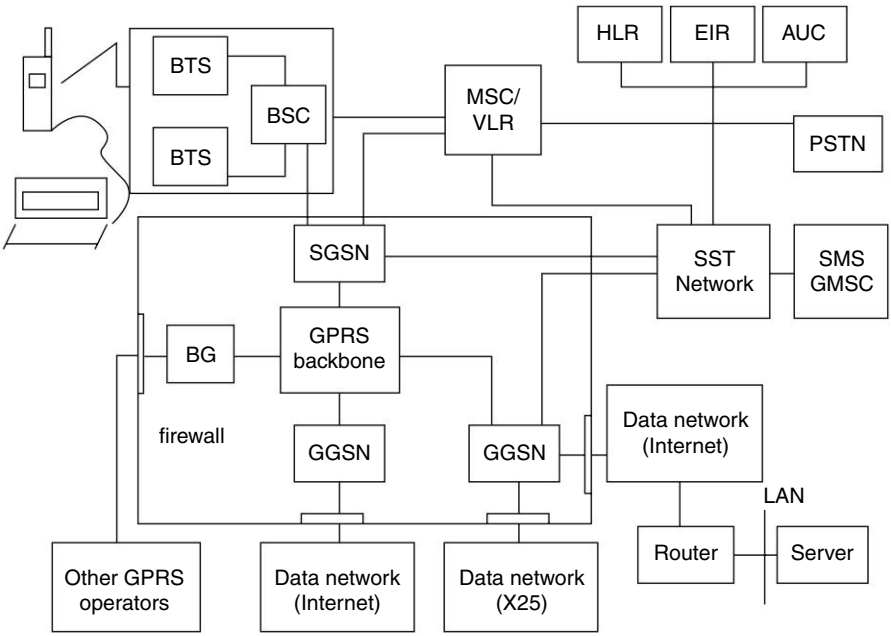


Fig. 4.6 *GPRS architecture*

der Gateway (BG), Backbone network (intra-PLMN and Inter-PLMN), HLR, MSC/VLR, SMS-GSMC.

GPRS introduces two new network nodes SGSN and GGSN in the GSM PLMN. The SGSN is at the same hierarchical level as the MSC. It is responsible for the delivery of packets to/from the MSs within its service area and communicates with the GGSN. The SGSN keeps track of the individual MSs location within its service area and performs security functions and access control. The SGSN is connected to the BSS with Frame Relay.

The GGSN provides interworking with external packet-switched networks, such as the Internet, X.25 networks or private networks, and is connected with SGSNs via an IP-based GPRS backbone network. It maintains routing information used to tunnel Protocol Data Units (PDU) to the SGSN that is currently serving the MS. The HLR is enhanced with GPRS subscriber information, and the SMS-GMSCs and SMS-IWMSCs are upgraded to support SMS transmission via the SGSN.

GPRS security functionality is equivalent to the existing GSM security. The SGSN performs authentication and cipher setting procedures based on the same algorithms, keys, and criteria as in existing GSM. GPRS uses a new ciphering algorithm optimised for packet data transmission.

GPRS offers the user an always on connection to Internet and Intranet. Some of the services may require high level of security. This can be financial

transaction over the Internet, or exchange of confidential documents from a company's Intranet to an employee. It is important to have strong focus on the security, so companies and persons that demand high level of security can take advantage of the services GPRS offers. What normally happens is that the services win against the security. This is in most cases adequate security for what a normal subscriber requires.

Security Issues in GPRS

Figure 4.7 illustrates five main areas where security in the GPRS system is exposed. These five areas are:

1. Security aspect related to the mobile phone and the SIM card.
2. Security between the MS and the SGSN. These include also the air interface from the MS to the BSS.
3. The PLMNs backbone network security that mainly referred to the traffic between the SGSN and the GGSN. But also handling the flow of subscriber information, like triplets between the HLR and SGSN.
4. Security among different operators.
5. Security between GGSN and the external connected networks, like Internet.

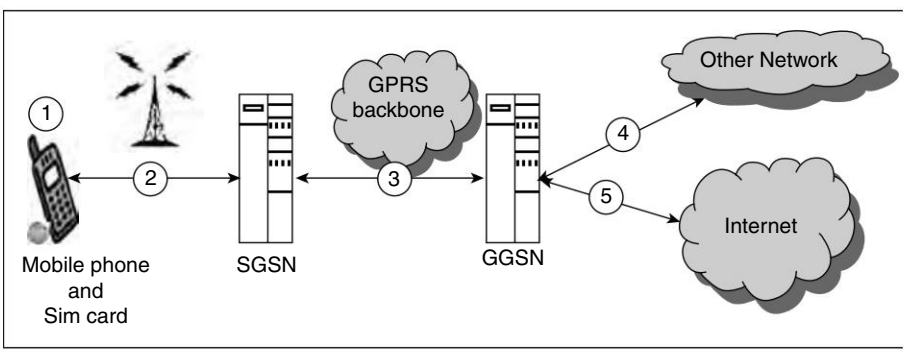


Fig. 4.7 *GPRS security issues*

4.4.1 Security Threats to the GPRS

The threats to the GPRS are very different from the Circuit Switched GSM. The security threats to the GSM are quite limited, there are not many hackers that can or will crack the obscurity of SS7 protocol. The GPRS systems are much more exposed to intruders, because of its IP-based backbone. There are many people who have thorough knowledge about the TCP/IP in proportion to the SS7. Intruders to the GPRS system can be people or organisation

that attempt to breach the confidentiality, integrity, availability or otherwise attempt to abuse the GPRS in order to compromise services, defraud users or any parts in the GPRS system. Generally following attempts are made to destabilise the systems:

Unauthorised Access to the Data

User traffic, signalling data or control data are information intruders may eavesdrop on the air interface. The signalling data or the control data are information that can be useful to conducting active attacks on the GPRS system and give the intruders access to secure management data.

It is possible for intruders to masquerade as a network element, e.g., a BTS and intercept user traffic, signalling data or control data over the air interface. Intruders may observe the time, rate, length, sources or destination of messages in order to get access to the information. This is a passive way to analyse the traffic. Intruders may actively initiate communication sessions and then observe in the same way as the passive traffic analysis to obtain access to information.

Threats to the Integrity

The integrity is exposed if the traffic and the data in any way are modified, inserted, replayed or deleted. For example, manipulation of user traffic, signalling data or control data may occur in an accidental or a deliberate manner.

Following are the various reasons for integrity protection:

- An active man-in-the-middle attacker could potentially compromise user traffic confidentiality by masquerading as a network to establish an unciphered connection towards the user. Since integrity protection can be made mandatory, this attack can be prevented as the user can always verify the instruction from the network to establish an unciphered connection. In GSM the instruction from the network to establish an unciphered connection is not integrity protected.
- The ability to integrity-protect ciphering algorithm negotiation messages provides protection against bidding-down attacks where an active attacker forces the use of an old ciphering algorithm which may, for instance, allow user traffic confidentiality to be compromised. This feature only becomes of interest when multiple algorithms are supported in the system, as is the case in GSM. In the first release of the 3GPP standards only one ciphering algorithm is available and all mobile stations must support this. However, it was considered desirable to design a future-proof system, which allowed new algorithms to be deployed in a way, which protects against bidding down attacks.

- Although ciphering of signalling traffic provides some integrity protection and the ciphering of user traffic severely limits the usefulness of any successful compromise of signalling message integrity, the application of a dedicated integrity protection mechanism with its own integrity key increases the security margin of the system. This is seen as an important enhancement, which will ensure that 3G offers adequate protection against increasingly sophisticated active attackers.

Denial of Service Attacks

To jam users traffic is a physical intervention of denying someone the services. The user traffic, signalling data and control data are by jamming, prevented from being transmitted over the air interface. Another way to prevent the information or data to be transmitted is by including specific protocol failures. It is possible for intruders to induce these failures by physical meanings. Another way to deny services is by masquerading as a network element and then prevent the user traffic, signaling data or control data from being transmitted.

Some of the example of Denial of service attacks are give below:

- **Attack from Valid Network**

An attacker using a valid network address could wreak havoc by making the attack appear to come from an organisation which did not, in fact, originate the attack and was completely innocent. In such cases, the administrator of a system under attack may be inclined to filter all traffic coming from the apparent attack source. Adding such a filter would then result in a denial of service to legitimate, non-hostile end-systems. In this case, the administrator of the system under attack, unwittingly becomes an accomplice of the attacker.

- **Randomly Changing Source Address**

The attacker launches the attack using randomly changing source addresses; the source addresses are depicted as from within some network, which are not generally present in the global Internet routing tables, and therefore, unreachable. However, any unreachable prefix could be used to perpetrate this attack method.

Unauthorised Access to Services

A possibility for an intruder to get unauthorised access to services can be by masquerading as a BST towards a user and then hijacking the users connecting after authentication has been established.

4.4.2 GPRS Security Solutions

Parwith A5/3, a new GPRS ciphering algorithm based on Kasumi has been added to the GPRS system, called GEA3. In GPRS the ciphering is performed at a higher layer in the protocol stack the LLC (Logical Link Control) layer. RLC/MAC messages are not ciphered. The FEC then applies at the physical layer. Similarly in UMTS ciphering occurs at the RLC/MAC layer, which sits just above the physical layer, at which FEC is performed.

4.5 UMTS SECURITY FOR APPLICATIONS

The aim of the 3G security architecture is to improve on the security of 2G systems. Any security holes present in the 2G systems are to be addressed and fixed. Also, since many new services have been added to 3G systems, the security architecture needs to provide security for these services.

There are five different sets of security features that are part of the UMTS architecture; shown in Fig. 4.8.

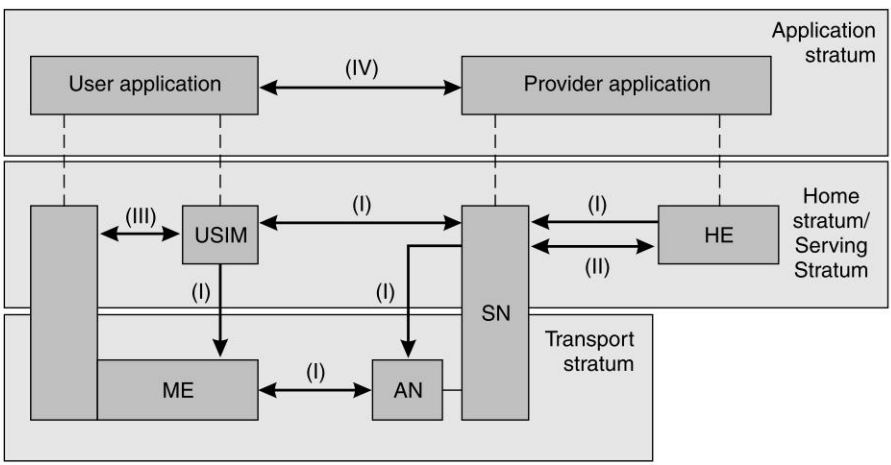


Fig. 4.8 *UMTS security architecture*

I. Network Access Security

This feature enables users to securely access services provided by the 3G network. It is responsible for providing identity confidentiality, authentication of users, confidentiality, integrity and mobile equipment authentication. User Identity confidentiality is obtained by using a temporary identity called the

International Mobile User Identity. Authentication is achieved using a challenge response method using a secret key. Confidentiality is obtained by means of a secret Cipher Key (CK) which is exchanged as part of the Authentication and Key Agreement (AKA) Process. Integrity is provided using an integrity algorithm and an integrity key (IK). Equipment identification is achieved using the International Mobile Equipment Identifier (IMEI).

II. Network Domain Security

This feature enables nodes in the provider domain to securely exchange signaling data, and prevent attacks on the wired network.

III. User Domain Security

This feature enables a user to securely connect to mobile stations.

IV. Application Security

This feature enables applications in the user domain and the provider domain to securely exchange messages.

V. Visibility and Configurability of Security

This feature allows users to enquire what security features are available.

4.5.1 UMTS AKA Security Mechanism

The UMTS Authentication and Key Agreement (UMTS AKA) mechanism is responsible for providing authentication and key agreement using the challenge/response mechanism. Challenge/Response is a mechanism where one entity in the network proves to another entity that it knows the password without revealing it. There are several instances when this protocol is invoked. When the user first registers with the network, when the network receives a service request, when a location update is sent, on an attach/detach request and on connection re-establishment. The current recommendation by 3GPP for AKA algorithms is MILENAGE. MILENAGE is based on the popular shared secret key algorithm called AES or Rijndael.

AKA provides mutual authentication for the user and the network. Also, the user and the network agree upon a cipher key (CK) and an integrity key (IK) which are used until their time expires. Control Signalling Communication between the mobile station and the network is sensitive and therefore its integrity must be protected. This is done using the UMTS Integrity Algorithm (UIA) which is implemented both in the mobile station and the RNC. This

is known as the f9 algorithm. First, the f9 algorithm in the user equipment calculates a 32 bit MAC-I for data integrity using the signaling message as an input parameter. This, along with the original signal message is sent to the RNC, where the XMAC-I is calculated and then compared to the MAC-I. If both are same, then the integrity of the message has not been compromised.

The UMTS-AKA protocol is an authentication and key agreement protocol. It is equipped by the 3GPP (3rd Generation Partnership Project). The objective is to meet the requirements of UMTS so that the mobile device can stay secure both, during the authentication process and during the telecommunication session. As shown in Fig. 4.9, the UMTS-AKA protocol has two phases. One is the phase of distribution of authentication vectors from HE (Home Environment) to SN (Service Network). The other is the phase of authentication and key establishment. Table 4.1 defines the relevant symbols.

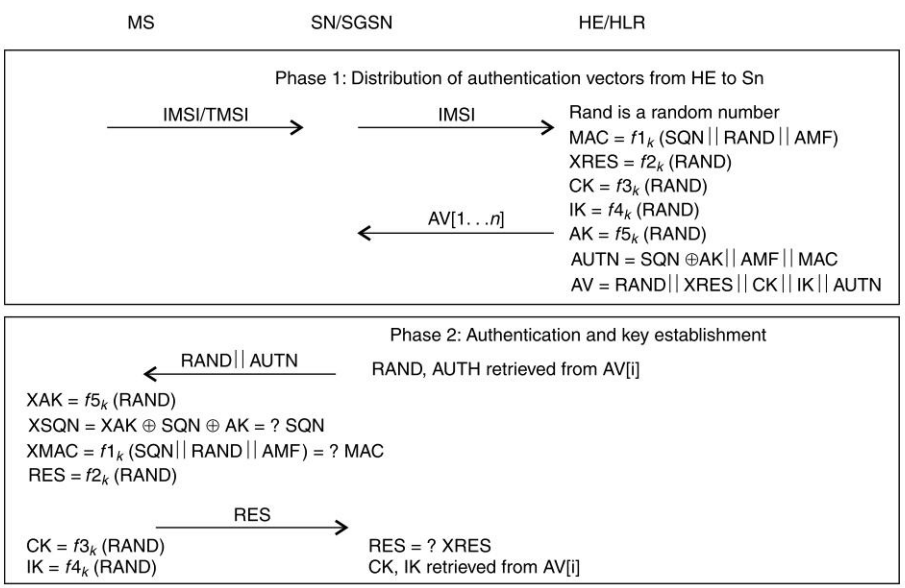


Fig. 4.9 *UMTS-AKA Protocol*

When an MS enters into the service domain of the SN, or VLR (Visitor Location Register), for the first time, it is executing the phase of distribution of authentication vectors from HE to SN and completing a registration procedure. This procedure, in addition to making MS's HE aware of the MS location, can let the SN obtain the AVs (Authentication Vectors) from MS's HE (for authentication with the MS in the future). AVs include the n set of authentication vectors and can provide n time authentication between MS and

Table 4.1 *Symbols used in UMTS-AKA-protocol / UMTS-IMS-AKA-protocol*

Symbol	*Description*
MS, SN, HE	Mobile Station, Service Network, Home Environment
VLR,HLR	Visitor Location Register, Home Location Register
AuC, SQN	Authentication Centre, Sequence Number
USIM	Universal Subscriber Identity Module
IMSI	International Mobile Subscriber Identity
TMSI	Temporary Mobile Subscriber Identity
AV, AUTN	Authentication Vector, Authentication Token
K	Secret Key which share between USIM and AuC
MAC	Messages Authentication Code
AMF	Authentication Management Field
Rand, RES	Random Number, User Response
XRES	Expected User Response
CK, IK, AK	Cipher Key, Integer Key, Authentication Key
f1 f5	Authentication and Key Generation Function
UE	User Equipment
IMPI, IMPU	IP Multimedia Private Identity , IP Multimedia Public Identity
P-CSCF	Proxy Call Service Control Function
I-CSCF	Integrating Call Service Control Function
S-CSCF	Service Call Service Control Function
P-CSCF	Proxy Call Service Control Function

SN. If MS is always registered and it wants to use a service in SN, it can execute the phase of authentication and key establishment leading to mutual authentication. Thus they confirm the legitimacy of each other. In this protocol, the basis of authentication is a secret key that is shared between MS's HE and MS. Through USIM (Universal Subscriber Identity Module) protection, this key can be recognised only by MS and its HE. SN and HE may be different system operators; this characteristic helps UMTS to expand its service range. Once authenticated, both parties can establish the cipher key and the integer key. These keys allow their messages to remain private.

UMTS Network Authentication to Phone

In UMTS the possibility of an attacker imitating the network has been removed by means of a 2-way authentication procedure. The procedure for which the mobile authenticates itself to the network is largely unchanged from GSM, however the network now sends an Authentication Token (AUTN) along with the RAND. The AUTN consists of a sequence number (SQN) encrypted using the RAND and the root key (K). It also consists of the MAC code, which works much like the GSM SRES but in the opposite direction. Thus if the XMAC does not match the MAC calculated by the SIM then phone

then send an authentication reject message to the network and the connection is over. Finally, to stop an attacker simply replaying the legitimate networks authentication request the SIM keeps track of the sequence numbers used (i.e., the same sequence number cannot be used twice, each sequence number must be newer).

4.6 3G SECURITY FOR APPLICATIONS

As the mobile operators move to 3G services, they are, for the most part, not deploying entirely new networks but instead leveraging their existing 2.5G network infrastructure GSM/GPRS/EDGE or CDMA/CDMA 1X equipment and backbone networks. For example, most UMTS cell sites can be co-located in GSM cell sites and much of the GSM/GPRS core network can be re-used. The Serving GPRS Support Node (SGSN) needs to be upgraded, but the mobile switching center (MSC) only requires a minor upgrade and the Gateway GPRS Support Node (GGSN) can remain the same. Because 3G networks were not all built from the ground up, they were not necessarily built with IP data security in mind. Moreover, the world of IP data is relatively new to mobile operators, they are used to dealing with comparatively more mundane voice-centric security threats.

4.6.1 3G Attacks

There are numerous attacks that can be perpetrated against a mobile network and they can originate from two primary vectors:

- **Outside the mobile network:** the public Internet, private networks, other operators networks.
- **Within the mobile network:** from devices such as data-capable handsets and smartphones, notebook computers or even desktop computers connected to the 3G network.

Denial of Service

One of the most prevalent security threats to wired ISPs is a distributed denial of service (DDoS) attack. Essentially, DDoS attacks use brute force methods to overwhelm the target system with data such that the response from the target system is either slowed or stopped. Creating enough traffic to inflict that kind of damage typically requires a network of compromised computers, which are often referred to as bots or zombies (sometimes collectively referred to as botnets). Essentially, botnets are computers that have been compromised

by attackers, generally through the use of Trojans (malware disguised as or embedded within legitimate software), which are then remotely controlled by the organisation orchestrating the DDoS attack. Laptops, smartphones, RIM BlackBerries and/or PDAs, connected to the Internet via a mobile broadband connection, could be similarly compromised and used as zombies in a DDoS attack.

Overbilling Attack

Another type of possible attack is called overbilling. Overbilling involves a malicious user hijacking a subscribers' IP address and then using that connection to initiate fee-based downloads or simply use that connection for their own purposes. In either case, the legitimate user is billed for activity which they did not authorise or actually conduct.

Spoofed PDP Context

These types of attacks exploit weaknesses in the GTP (GPRS Tunneling Protocol).

- Spoofed delete PDP context packets, which would cause service loss or interruption for end users
- Spoofed create PDP context packets, which would result in unauthorised or illegal access to the Internet or customer data networks.
- GTP packet floods, which is a type of Denial of Service attack. More on GTP and PDP follows in the Interfaces to other mobile networks section.

Signalling-level Attacks

The Session Initiation Protocol (SIP) is a signaling protocol used in IMS networks to provide voice over IP (VoIP) services. There are several well-known vulnerabilities with SIP-based VoIP systems. For example, there are vulnerabilities in the Call Manager function (which handles call routing and call signalling functions in VoIP systems) that might allow hackers to:

- Reconfigure VoIP settings and gain access to individual users' account information.
- Eavesdrop on VoIP communications.
- Hijack a user's VoIP subscription and subsequent communications.

4.6.2 Some Security Solutions for 3G

While there are several security mechanisms available in Wireless Cellular Networks, continued research is going on to provide new and even more secure mechanisms for cellular security, some of them are listed below.

A New Authentication Scheme with Anonymity for Wireless Networks

When a mobile user is roaming, it is necessary to provide anonymity to the users so that malicious parties are unable to associate the user with a particular session. The most basic method to provide anonymity is to have a temporary identity (TID) instead of the real-id of the user. There are several issues to consider when designing a security protocol for cellular networks. One, they have low computational power which means that algorithms that require high processing power are not suitable. Second, the error rate of messages increase on wireless networks as compared to cellular networks. Therefore, any mechanism that is designed should minimise message sizes and the number of messages in order to reduce the error rate.

Manual Authentication for Wireless Devices

This is a technique used by devices to authenticate one another by manually transferring data between the devices. This means that the users will enter some information using some form of input (e.g., keypad). Underneath they employ MAC algorithms for authentication. Although the scheme that is proposed is secure, it's usability depends upon how many numbers (or alphabets) the users have to input.

Elliptic Curve Cryptography for Wireless Security

Elliptic Curve Cryptography (ECC) is a mechanism which uses points on an elliptic curve to encrypt/decrypt data. It has an advantage over the popular RSA algorithm in that it is much faster. 163-bit ECC provides the same security as a 1024 bit RSA algorithm, and can be anywhere from 5 to 15 times faster depending on the platform. For example, in order to secure a 128-bit AES shared key and 521-bit ECC provides the same level of security as an 15,360 bit RSA while being about 400 times faster.

Channel Surfing and Spatial Retreats

Channel Surfing is a technique where the transmission frequency is changed to one where there is no interference. Spatial Retreats is a technique where the wireless users move to a location where there is no interference.

4.7 SOME OF SECURITY AND AUTHENTICATION SOLUTIONS

4.7.1 Protocol of Gong et al.

This solution contains a trusted third party which is continuously available online, as in Kerberos. The parties in the system authenticate each other by the help of the trusted server. In this protocol, unlike EKE, there is no need to generate fresh public/private key pairs per session, but there is a need for the trusted server's public key to be known by all parties in advance.

4.7.2 GSM User Authentication Protocol (GUAP)

The current GSM authentication scheme uses a cryptographic authentication key embedded in the SIM card of the device. By the new approach, the user can authenticate with a password instead of the embedded key, breaking the dependency on the SIM card during authentication. GSM authentication protocol resembles the approach of Gong et al. in that both schemes are based on three entities and that in both cases the third entity is a trusted server whose public key is known by all parties. In GSM authentication, VLR is an automated non-human entity, which is able to remember strong secrets. Another important difference is in regard to the use of timestamps. Lack of synchronised clocks in GSM does not allow use of timestamps for authentication which can be solved using random challenge numbers.

Mobile subscriber *MS* is to be authenticated to the HLR via the VLR, using password P. K_{hlr} is the public key of the HLR known to all parties, and K_{vlr} is the symmetric encryption key shared between the VLR and HLR. In the protocol, *RAND* is the challenge used to authenticate the mobile client. The client encrypts response $(RAND)_P$, and three other nonces $n1$, $n2$, and $n3$ with the HLR's public key and sends it to the VLR, who in turn passes it to the HLR along with the challenge *RAND*. The HLR verifies the correctness of the response $(RAND)_P$, generates a session key k, and passes it to the VLR and the MS, encrypted under K_{vlr} and P, respectively. In the last two messages, the VLR and the MS carry out a mutual challenge-response protocol and verify the new session key k. Table 4.2 illustrates the working of GUAP.

4.7.3 One-time Password Schemes

The purpose of a one-time password (OTP) is to make it more difficult to gain unauthorised access to restricted resources. There are basically three types of one-time passwords: the first type uses a mathematical algorithm to generate

Table 4.2 *GUAP Protocol*

- $MS \rightarrow VLR$: *IMSI*.
- $VLR \rightarrow MS$: *RAND*.
- $MS \rightarrow VLR : \{n_1, n_2, n_3, (RAND)_P\}\kappa_{hlr}, r_a$.
- $VLR \rightarrow HLR: \{ n_1, n_2, n_3, (RAND)_P \}\kappa_{hlr}, (RAND)\kappa_{vlr}$.
- $HLR \rightarrow VLR: \{k\}\kappa_{vlr}, \{n_1,n_2 \oplus k\}_P$.
- $VLR \rightarrow MS : \{n_1,n_2 \oplus k\}_P, \{r_a\}_k, r_b$.
- $MS \rightarrow VLR : \{r_b\}_k$.

a new password based on the previous, a second type that is based on time-synchronisation between the authentication server and the client providing the password, and a third type that is again using a mathematical algorithm, but the new password is based on a challenge (e.g., a random number chosen by the authentication server or transaction details) and a counter instead of being based on the previous password.

Mobile OTP

Mobile-OTP is a free "strong authentication" solution for java capable mobile devices like phones or PDAs. The solution is based on time synchronous one-time passwords. It consists of a client component (a J2ME MIDlet) and a server component (a Unix shell script). The MIDlet generates one time passwords by hashing the following data with MD5:

(a) the current epoch-time in a 10 second granularity.
(b) the 4-digit PIN that a user enters.
(c) a 16-hex-digit secret that has been created when the device was initialised.

When entering a PIN, the MIDlet displays the first 6 digits of the MD5-hash. This is the one time password. The password can be verified by the server, as the server also knows the current time, Init-Secret and PIN of the user. To compensate time differences, the server will accept passwords from 3 minutes in the past to 3 minutes in the future. In addition, different time offsets can be specified for each user on the token and/or the server. Each password will be accepted only once. After 8 successive failed authentication attempts a user gets locked out. Authentication is based on two factors: a PIN known by the user and the Init-Secret stored on the mobile device.

OTP using GSM

This scheme proposes a mobile OTP solution which combines simple and secure OTP principle with the ubiquitousness of a GSM mobile phone. A Java midlet which can be installed on any Java enabled mobile phone transforms the

phone into a secure OTP token which can be used to log in to any service on the Internet. The solution is based on a simple challenge-response protocol. When the user wants to log in, he/she presents his/her username, and a challenge is sent to the users mobile phone. The OTP midlet installed on the phone generates an OTP from the challenge and sends it back to the server. The server verifies that the OTP is correct and the user is authenticated, Table 4.3 protocol message used in OTP using GSM.

Table 4.3 *OTP using GSM*

SUMMARY: A identifies to B using OTP.
(a) *One-time setup.*
  i. The B generates a challenge and computes OTP = $H(challenge||$secret key).
  ii. The B computes the w_1=MAC(OTP).
(b) *Protocol messages.* $B \rightarrow A$: *challenge* & w_1
(c) *Protocol actions.*
  i. A computes $secretkey = D(w_1)$
  ii. A computes $OTP = H(challenge||secret\ key)$
  iii. A computes $w_2 = MAC(OTP)$
  iv. if $w_1 = w_2$ then authentication success.

OTP using GPRS

An authentication mechanism is presented here, which requires both a Web and a GPRS connection. The end user enters userid/password credentials via the web-based interface and receives a OTP via SMS on his mobile phone, which he must then type in to be granted access to the system. Table 4.4 describes a case study of server B authenticity user A.

Web/Mobile Authentication System with OTP

This authentication solution combines web/mobile authentication system. The basic authentication mechanism is integrated with a challenge/response process and an OTP. The challenge is issued from an authentication server and has to authenticate a mobile device, like a cell phone. This device can communicate with any other involved parts through a fixed terminal, typically a personal computer via a Bluetooth connection. The mobile device, once accepted, performs the authentication with the web site or application. Table 4.5 illustrates the working of the schemes.

Table 4.4 *OTP using GPRS*

SUMMARY: *A* identifies to *B* using basic authentication, challenge/response and temporary password.

(a) *One-time setup.*
   i. User *A* enters the Web site and submits his/her *username* and *password* via the browser.
   ii. The *Web-server* verifies that the user has a valid *username/password* combination.

(b) *Protocol messages*. *Web-server* → *Auth-server*: $Auth_{req}$, $t_{Web\text{-}server}$

(c) *Protocol actions.*
   i. *Auth-server* → *mobile-user*: *RAND*.®n
   ii. *mobile-user* computes $w_1 = E_k(RAND,F)$.
   iii. *mobile-user* → *Auth-server*: w_1.
   iv. *Auth-user* computes $w_2 = E_k^{-1}(RAND, w_1)$.
   v. *Auth-user* generates *OTP*.
   vi. *Auth-server* → *Web-server*: *OTP*, $t_{Auth\text{-}server}$
   vii. *Auth-server*} → *mobile-user*: $E_k(OTP)$.
   viii. *mobile-user* computes $d_1 = E_k^{-1}(RAND, E_k(OTP))$.
   ix. *mobile-user* → *Web-server*: d_1.
   x. if $d_1 = OTP$ then authentication success.

Table 4.5 *Web/Mobile authentication with OTP*

SUMMARY: *A* identifies to *B* using basic authentication, challenge/response and temporary password.

(a) *One-time setup*.
   i. User *A* enters the Web site and submits his/her *Username* and *Password* via the browser.
   ii. The *Web-server* verifies that the user has a valid Username/Password combination.

(b) Protocol messages. *Web-server* → *Support-server*: *SessionId*, *Username*

(c) *Protocol actions*.
   i. *Support-server* maps *Username* → $UserId_2$.
   ii. *Support-server* → *Auth-user*: $UserId_2$, *RAND*.
   iii. *mobile-user* → *Auth-server*: $UserId_1$.
   iv. *Auth-server* verifies $UserId_1 = UserId_2$.
   v. *Auth-server* → *mobile-user*: *RAND*.
   vi. *mobile-user* → *Auth-user*: $w_1 = E_k(RAND,Shared\text{-}data)$.
   vii. *Auth-user* computes $Shared\text{-}data = E_k^{-1}(RAND, w_1)$.
   viii. *Auth-user* generates OTP_1.
   ix. *Auth-user* → *Support-server*: OTP_1.
   x. *Auth-user* → *mobile-user*: $w_1 = E_k(RAND, OTP_1)$.
   xi. *mobile-user* → *Support-server* = OTP_1.
   xii. *mobile-user* computes $OTP_2 = E_k^{-1}(RAND, w_1)$.
   xiii. *mobile-user* → *Support-server*: OTP_2.
   xiv. *Support-server* verifies $OTP_1 = OTP_2$ then authentication success.

4.7.4 Location-based Encryption

Location based encryption scheme is proposed, which builds on established cryptographic algorithms and protocols in a way that provides an additional layer of security beyond that provided by conventional cryptography. It allows data to be encrypted for a specific place or broad geographic area, and supports constraints in time as well as space. It can be used with both fixed and mobile applications and supports a range of data sharing and distribution policies. The location coordinates are incorporated during the data encryption, the receiver is able to decrypt the data only when the location coordinates matches with its current location. A negotiating security technique across multiple terms of transaction, such as terminal types, service types, user's preference, and the level of sensitivity of information is proposed using a MAUT (Multi-Attribute Utility Theory). A MAUT is a systematic method that identifies and analyses multiple variables in order to provide a common basis for arriving at a decision. It is used as a decision-making tool to predict security levels depending on the security context (network state, the resource's and user's environments, etc.).

4.7.5 BioPasswords

BioPassword is a patented software-only authentication system based on the keystroke dynamics biometric. While a user enters password the system captures information about just how a user types, including any pauses between the pressings of different keys. Essentially the software observes the typing rhythm, pace and syncopation. This information is used to create a statistically reliable profile for an individual. In combination with a user password bio-password creates a so-called hardened password.

A behaviour-based passwords approach based on maps is proposed. A user is shown a map of some N cities with some routes selected and all other routes between all cities available but not activated. At the PassMap creation stage also known as the enrollment stage, a user is presented with a relatively large map of routes to which a user is asked to make any modifications. A possible list of atomic modifications includes: Selecting a direct route between any two cities, and Un-selecting a direct route between any two cities. In the PassMap system of authentication a user is not required to memorise any difficult character combinations, instead a user only needs to memorise the sequence of changes a user makes to the base map.

SUMMARY

In this chapter we have introduced various generations of cellular networks, followed by their security issues. We have presented the detailed discussions on security issues of GSM. GPRS, UMTS, and 3G networks for developing cellular-based services.

REVIEW QUESTIONS

1. Explain the working of 1G cellular network with architecture.
2. Explain the working of 2G cellular network with architecture.
3. Explain the working of 3G cellular network with architecture.
4. What are the various security issues in cellular networks ?
5. What are the various attacks possible over cellular networks ?
6. What security features does GSM offer for developing applications ?
7. Mention various security considerations while developing application for cellular networks.
8. Explain the impact of false base station attack on cellular applications.
9. Explain the impact of algorithm attacks on cellular applications.
10. Explain the impact of SIM interface tapping attack on cellular applications.
11. What are the security solutions available from GSM ?
12. Explain the working of public key infrastructure in mobile systems.
13. What security features GPRS offers for developing applications ?
14. What are the various attacks on GPRS networks ?
15. What are the security issues associated with deploying GPRS networks?
16. What are the security solutions available for GPRS applications ?
17. Explain five important features of UMTS security architecture.
18. Explain the working of the UMTS AKA security mechanism.
19. Explain the working of the UMTS network authentication to phone mechanism.
20. Explain the 3G security for applications development.
21. Explain some of the security solutions developed for 3G applications.
22. Explain few techniques used for application-level security in cellular networks.
23. Explain some of the techniques available for GSM user authentication.
24. Briefly discuss the working of the mobile OTP.
25. How OTP mechanism is implemented in GSM?
26. How OTP mechanism is implemented in GPRS?
27. How Web/mobile authentication system with OTP is implemented?
28. Explain the application of location-based encryption.
29. Explain the working of BioPassword system.

The McGraw-Hill Companies

Application Level Security in MANETs

5

OBJECTIVES

- To understand the need of mobile ad hoc networks.
- To discuss the salient features of MANETS.
- To illustrate some typical MANET applications.
- To know about what kind of applications best suits MANETS.
- To know about attacks at the network layer of MANETS.
- To understand various types of threats for MANET applications.
- To study some of the security schemes at application level proposed for MANETS.

Unlike a fixed wireless network, wireless ad-hoc or on-the-fly networks are characterised by the lack of infrastructure. Nodes in a mobile ad-hoc network are free to move and organise themselves in an arbitrary fashion. Each user is free to roam about while communicating with others. The path between each pair of the users may have multiple links, and the radio between them can be heterogeneous. This allows an association of various links to be a part of the same network. Mobile ad-hoc networks can operate in a stand-alone fashion or could possibly be connected to a larger network such as the Internet.

Ad-hoc networks are suited for use in situations where an infrastructure is unavailable or deploying is not cost effective. One of many possible uses of mobile ad-hoc networks is in some business environment, where the need for collaborative computing might be more important outside the office environment than inside, such as in a business meeting outside the office to brief clients on a given assignment. A mobile ad-hoc network can also be used to provide crisis management services applications, such as in disaster recovery, where the entire communication infrastructure is destroyed and resorting

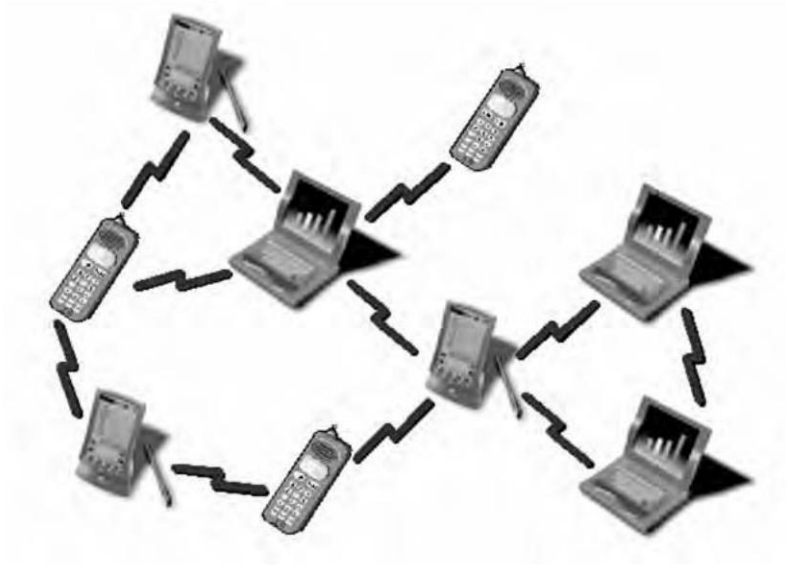


Fig. 5.1 *Mobile ad hoc networks*

communication quickly is crucial. By using a mobile ad-hoc network, an infrastructure could be set up in hours instead of weeks, as is required in the case of wired line communication. Another application example of a mobile ad-hoc network is Bluetooth, which is designed to support a personal area network by eliminating the need of wires between various devices, such as printers and personal digital assistants. The IEEE 802.11 (or Wi-Fi) protocol also supports an ad-hoc network system in the absence of a wireless access point.

5.1 MANETS

The mobile ad hoc networks (MANETs) have received tremendous attention because of their self-configuration and self-maintenance capabilities. Although security has long been an active research topic in wireline networks, the unique characteristics of MANETs present a new set of nontrivial challenges to security design. These challenges include open network architecture, shared wireless medium, stringent resource constraints, and highly dynamic network topology. Consequently, the existing security solutions for wired networks do not directly apply to the MANET domain.

There are basically two approaches to protecting MANETs: *proactive* and *reactive*. The proactive approach attempts to prevent an attacker from launching attacks in the first place, typically through various cryptographic

techniques. In contrast, the reactive approach seeks to detect security threats a posteriori and react accordingly. Due to the absence of a clear line of defence, a complete security solution for MANETs should integrate both approaches and encompass all three components: *prevention*, *detection*, and *reaction*. For example, the proactive approach can be used to ensure the correctness of routing states, while the reactive approach can be used to protect packet forwarding operations.

5.2 SOME APPLICATIONS OF MANETS

With the increase of portable devices as well as progress in wireless communication, ad hoc networking is gaining importance with the increasing number of widespread applications. Ad hoc networking can be applied anywhere, where there is little or no communication infrastructure or the existing infrastructure is expensive or inconvenient to use. Ad hoc networking allows the devices to maintain connections to the network as well as easily adding and removing devices to and from the network. The set of applications for MANETs is diverse, ranging from large-scale, mobile, highly dynamic networks, to small, static networks that are constrained by power sources. Besides the legacy applications that move from traditional infrastructured environment into the ad hoc context, a great deal of new services can and will be generated for the new environment. Some applications include the following:

Military Battlefield

Military equipment now routinely contains many computer based devices. Ad hoc networking would allow the military to take advantage of commonplace network technology to maintain an information network between the soldiers, vehicles, and military information head quarters. The basic techniques of ad hoc network came from this field (refer Fig. 5.2).

Commercial Sector

Ad hoc networks can be used in emergency/rescue operations for disaster relief efforts, e.g., in fire, flood, or earthquake. Emergency rescue operations must take place where non-existing or damaged communications infrastructure and rapid deployment of a communication network is needed. Information is relayed from one rescue team member to another over a small handheld device. Other commercial scenarios include, e.g., ship-to-ship ad hoc mobile communication, law enforcement, etc.

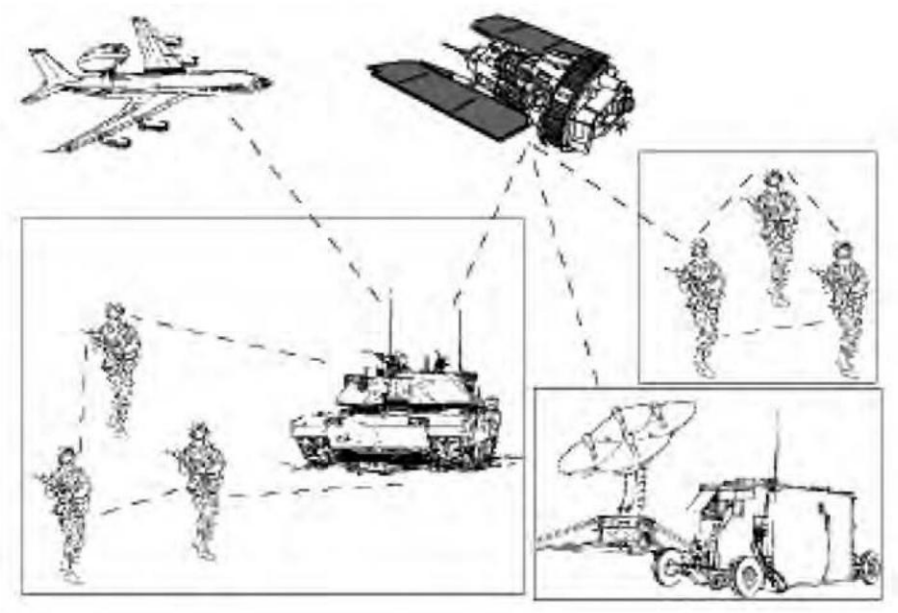


Fig. 5.2 *Possible scenario: war condition*

Local Level

Ad hoc networks can autonomously link an instant and temporary multimedia network using notebook computers or palmtop computers to spread and share information among participants at a conference or classroom. Another appropriate local level application might be in home networks where devices can communicate directly to exchange information. Similarly in other civilian environments like taxicab, sports stadium, boat and small aircraft, mobile ad hoc communications will have many applications.

Personal Area Network (PAN)

Short-range MANET can simplify the intercommunication between various mobile devices (such as a PDA, a laptop, and a cellular phone). Tedious wired cables are replaced with wireless connections. Such an ad hoc network can also extend the access to the Internet or other networks by mechanisms, e.g., Wireless LAN (WLAN), GPRS, and UMTS. The PAN is potentially a promising application field of MANET in the pervasive ubiquitous computing context.

5.3 MANET FEATURES

MANET has the following features:

Autonomous Terminal

In MANET, each mobile terminal is an autonomous node, which may function as both a host and a router. In other words, besides the basic processing ability as a host, the mobile nodes can also perform switching functions as a router. So usually endpoints and switches are indistinguishable in MANET.

Distributed Operation

Since there is no background network for the central control of the network operations, the control and management of the network is distributed among the terminals. The nodes involved in a MANET should collaborate amongst themselves and each node acts as a relay as needed, to implement functions, e.g., security and routing.

Multihop Routing

Basic types of ad hoc routing algorithms can be single-hop and multihop, based on different link layer attributes and routing protocols. Single-hop MANET is simpler than multihop in terms of structure and implementation, with the cost of lesser functionality and applicability. When delivering data packets from a source to its destination out of the direct wireless transmission range, the packets should be forwarded via one or more intermediate nodes.

Dynamic Network Topology

Since the nodes are mobile, the network topology may change rapidly and unpredictably and the connectivity among the terminals may vary with time. MANET should adapt to the traffic and propagation conditions as well as the mobility patterns of the mobile network nodes. The mobile nodes in the network dynamically establish routing among themselves as they move about, forming their own network on the fly. Moreover, a user in the MANET may not only operate within the ad hoc network, but may require access to a public fixed network (e.g., Internet).

Fluctuating Link Capacity

The nature of high bit-error rates of wireless connection might be more profound in a MANET. One end-to-end path can be shared by several sessions. The channel over which the terminals communicate is subject to noise, fad-

ing, and interference, and has less bandwidth than a wired network. In some scenarios, the path between any pair of users can traverse multiple wireless links and the link themselves can be heterogeneous.

Light-Weight Terminals

In most cases, the MANET nodes are mobile devices with less CPU processing capability, small memory size, and low power storage. Such devices need optimised algorithms and mechanisms that implement the computing and communicating functions.

5.4 SECURITY CHALLENGES IN MANETS

One fundamental vulnerability of MANETs comes from their open peer-to-peer architecture. Unlike wired networks that have dedicated routers, each mobile node in an ad hoc network may function as a router and forward packets for other nodes. The wireless channel is accessible to both legitimate network users and malicious attackers. As a result, there is no clear line of defence in MANETs from the security design perspective. The boundary that separates the inside network from the outside world becomes blurred. There is no well defined place/infrastructure where we may deploy a single security solution. Moreover, portable devices, as well as the system security information they store, are vulnerable to compromises or physical capture, especially low-end devices with weak protection. Attackers may sneak into the network through these subverted nodes, which pose the weakest link and incur a domino effect of security breaches in the system.

The stringent resource constraints in MANETs constitute another nontrivial challenge to security design. The wireless channel is bandwidth-constrained and shared among multiple networking entities. The computation capability of a mobile node is also constrained. For example, some low-end devices, such as PDAs, can hardly perform computation-intensive tasks like asymmetric cryptographic computation. Because mobile devices are typically powered by batteries, they may have very limited energy resources. The wireless medium and node mobility poses far more dynamics in MANETs compared to the wireline networks. The network topology is highly dynamic as nodes frequently join or leave the network, and roam in the network on their own will. The wireless channel is also subject to interferences and errors, exhibiting volatile characteristics in terms of bandwidth and delay. Despite such dynamics, mobile users may request for anytime, anywhere security services as they move from one place to another.

5.5 SECURITY ATTACKS ON MANETs

A MANET provides network connectivity between mobile nodes over potentially multihop wireless channels mainly through link-layer protocols that ensure one-hop connectivity, and network-layer protocols that extend the connectivity to multiple hops. These distributed protocols typically assume that all nodes are cooperative in the coordination process. This assumption is unfortunately not true in a hostile environment. Because cooperation is assumed but not enforced in MANETs, malicious attackers can easily disrupt network operations by violating protocol specifications.

The main network-layer operations in MANETs are ad hoc routing and data packet forwarding, which interact with each other and fulfill the functionality of delivering packets from the source to the destination. The ad hoc routing protocols exchange routing messages between nodes and maintain routing states at each node accordingly. Based on the routing states, data packets are forwarded by intermediate nodes along an established route to the destination. Nevertheless, both routing and packet forwarding operations are vulnerable to malicious attacks, leading to various types of malfunction in the network layer. Such network-layer vulnerabilities generally fall into one of two categories: routing attacks and packet forwarding attacks, based on the target operation of the attacks.

In addition to routing attacks, the adversary may launch attacks against packet forwarding operations as well. Such attacks do not disrupt the routing protocol and poison the routing states at each node. Instead, they cause the data packets to be delivered in a way that is intentionally inconsistent with the routing states. For example, the attacker along an established route may drop the packets, modify the content of the packets, or duplicate the packets it has already forwarded. Another type of packet forwarding attack is the denial-of-service (DoS) attack via network-layer packet blasting, in which the attacker injects a large amount of junk packets into the network. These packets waste a significant portion of the network resources, and introduce severe wireless channel contention and network congestion in the MANET.

Eavesdropping

Eavesdropping is the intercepting and reading of messages and conversations by unintended receivers. The mobile hosts in mobile ad hoc networks share a wireless medium. The majorities of wireless communications use the RF spectrum and broadcast by nature. Signals broadcast over airwaves can be easily intercepted with receivers tuned to the proper frequency. Thus, messages transmitted can be eavesdropped, and fake messages can be injected into network. Moreover, a radio signal can be jammed or interfered, which causes the

message to be corrupted or lost. If the attacker has a powerful transmitter, a signal can be generated that will be strong enough to overwhelm the targeted signals and disrupt communications. The most common types of this form of signal jamming are random noise and pulse. Jamming equipment is readily available. In addition, jamming attacks can be mounted from a location remote to the target networks.

Route Discovery Attacks

There are malicious routing attacks that target the route discovery or maintenance phase by not following the specifications of the routing protocols. Routing message flooding attacks, such as hello flooding, RREQ flooding, acknowledgment flooding, routing table overflow, routing cache poisoning, and routing loop are simple examples of routing attacks targeting the route discovery phase. Proactive routing algorithms, such as DSDV and OLSR, attempt to discover routing information before it is needed, while reactive algorithms, such as DSR and AODV, create routes only when they are needed. Thus, proactive algorithms are more vulnerable to routing table overflow attacks. Some of these attacks are listed below.

- **Routing Table Overflow Attack**

A malicious node advertises routes that go to non-existent nodes to the authorised nodes present in the network. It usually happens in proactive routing algorithms, which update routing information periodically. The attacker tries to create enough routes to prevent new routes from being created. The proactive routing algorithms are more vulnerable to table overflow attacks because proactive routing algorithms attempt to discover routing information before it is actually needed. An attacker can simply send excessive route advertisements to overflow the victims routing table.

- **Routing Cache Poisoning Attack**

In route cache poisoning attacks, attackers take advantage of the promiscuous mode of routing table updating, where a node overhearing any packet may add the routing information contained in that packet header to its own route cache, even if that node is not on the path. Suppose a malicious node M wants to poison routes to node X. M could broadcast spoofed packets with source route to X via M itself; thus, neighboring nodes that overhear the packet may add the route to their route caches.

- **Attacks at the Route Maintenance Phase**

There are attacks that target the route maintenance phase by broadcasting

false control messages, such as link-broken error messages, which cause the invocation of the costly route maintenance or repairing operation. For example, AODV and DSR implement path maintenance procedures to recover broken paths when nodes move. If the destination node or an intermediate node along an active path moves, the upstream node of the broken link broadcasts a route error message to all active upstream neighbours. The node also invalidates the route for this destination in its routing table. Attackers could take advantage of this mechanism to launch attacks by sending false route error messages.

- **Attacks at Data Forwarding Phase**

Some attacks also target data packet forwarding functionality in the network layer. In this scenario the malicious nodes participate cooperatively with the routing protocol for route discovery and maintenance phases, but in the data forwarding phase they do not forward data packets consistently according to the routing table. Malicious nodes simply drop data packets quietly, modify data content, replay, or flood data packets; they can also delay forwarding time-sensitive data packets selectively or inject junk packets.

End-to-End Attacks

The objectives of TCP-like Transport layer protocols in MANET include setting up of end-to-end connection, end-to-end reliable delivery of packets, flow control, congestion control, clearing of end-to-end connection. Similar to TCP protocols in the Internet, the mobile node is vulnerable to the classic "SYN" flooding attack or session hijacking attacks. However, a MANET has a higher channel error rate when compared with wired networks. Because TCP does not have any mechanism to distinguish between whether a loss was caused by congestion, random error, or malicious attacks, TCP multiplicatively decreases its congestion window upon experiencing losses, which degrades network performance significantly.

- **SYN Flooding Attack**

The SYN flooding attack is a denial-of-service attack. The attacker creates a large number of half-opened TCP connections with a victim node, but never completes the handshake to fully open the connection. For two nodes to communicate using TCP, they must first establish a TCP connection using a three-way handshake. The handshaking allows both nodes to learn that the other is ready to communicate and to agree on initial sequence numbers for the conversation.

During the attack, a malicious node sends a large amount of SYN packets to a victim node, spoofing the return addresses of the SYN packets. The

SYNACK packets are sent out from the victim right after it receives the SYN packets from the attacker and then the victim waits for the response of ACK packet. Without any response of ACK packets, the half-open data structure remains in the victim node. If the victim node stores these half-opened connections in a fixed-size table while it awaits the acknowledgment of the three-way handshake, all of these pending connections could overflow the buffer, and the victim node would not be able to accept any other legitimate attempts to open a connection. Normally there is a time-out associated with a pending connection, so the half-open connections will eventually expire and the victim node will recover. However, malicious nodes can simply continue sending packets that request new connections faster than the expiration of pending connections. The SYN-flooding attack scenario is given in the Fig. 5.3.

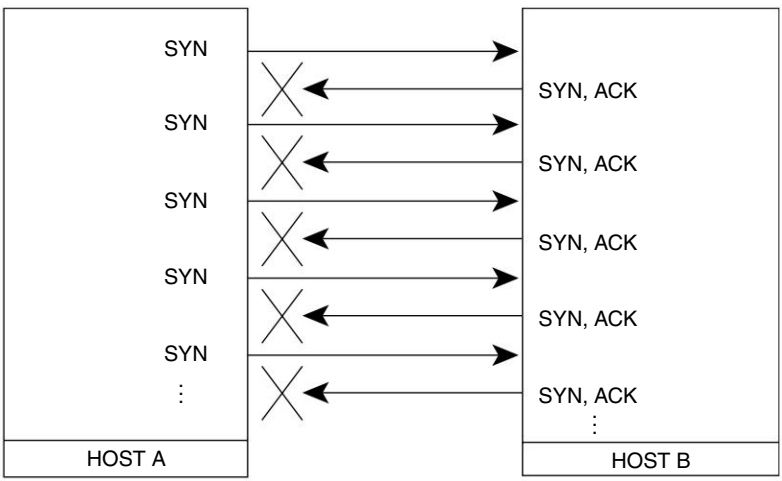


Fig. 5.3 *SYN flooding attack scenario*

- **Session Hijacking**

Session hijacking takes advantage of the fact that most communications are protected (by providing credentials) at session setup, but not thereafter. In the TCP session hijacking attack, the attacker spoofs the victims IP address, determines the correct sequence number that is expected by the target, and then performs a DoS attack on the victim. Thus the attacker impersonates the victim node and continues the session with the target. The TCP ACK storm problem, could be created when an attacker launches a TCP session hijacking attack. The attacker sends injected session data, and node A will acknowledge the receipt of the data by sending an ACK packet to node B. This packet will not contain

a sequence number that node B is expecting, so when node B receives this packet, it will try to resynchronise the TCP session with node A by sending it an ACK packet with the sequence number that it is expecting. The cycle goes on and on, and the ACK packets passing back and forth create an ACK storm. Hijacking a session over UDP is the same as over TCP, except that UDP attackers do not have to worry about the overhead of managing sequence numbers and other TCP mechanisms. Since UDP is connectionless, edging into a session without being detected is much easier than the TCP session attacks. A scenario of session hijacking attack is given in the Fig. 5.4.

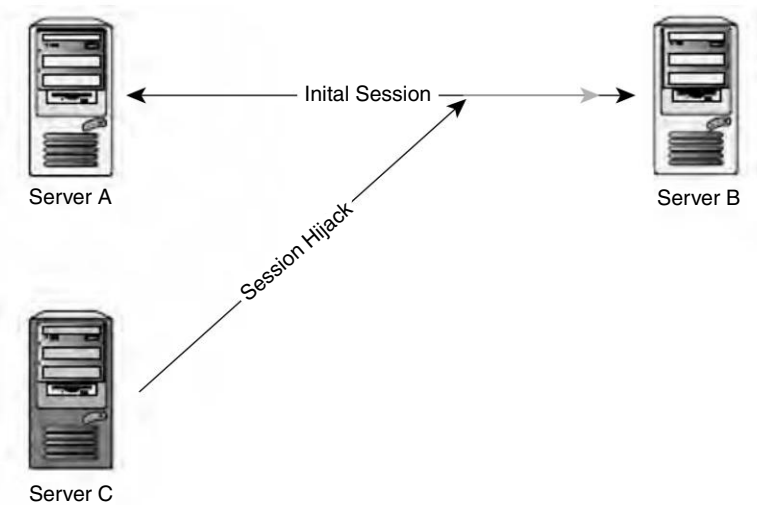


Fig. 5.4 *Session hijacking scenario*

Repudiation Attack

In the network layer, firewalls can be installed to keep packets in or keep packets out. In the transport layer, entire connections can be encrypted, end-to-end. But these solutions do not solve the authentication or non-repudiation problems in general. Repudiation refers to a denial of participation in all or part of the communications. For example, a selfish person could deny conducting an operation on a credit card purchase, or deny any on-line bank transaction, which is the prototypical repudiation attack on a commercial system.

5.5.1 Advanced Attacks

The blackhole (or sinkhole), Byzantine, and wormhole attacks are the typical examples, of advanced attacks which are described below.

- ***Wormhole Attack***

An attacker records packets at one location in the network and tunnels them to another location. Routing can be disrupted when routing control messages are tunneled. This tunnel between two colluding attackers is referred as a wormhole. Wormhole attacks are severe threats to MANET routing protocols. For example, when a wormhole attack is used against an on-demand routing protocol such as DSR or AODV, the attack could prevent the discovery of any routes other than through the wormhole. The tunneling procedure generates an illusion that the two nodes more than one hop away are in the neighbourhood of each other. The wormhole attack is possible even if the attacker has not compromised any hosts, and even if communication provides all authenticity and confidentiality. It is a severe attack and it is challenging to defend against.

- ***Blackhole Attack***

The blackhole attack has two properties. First, the node exploits the mobile ad hoc routing protocol, such as AODV, to advertise itself as having a valid route to a destination node, even though the route is spurious, with the intention of intercepting packets. Second, the attacker consumes the intercepted packets without any forwarding. However, the attacker runs the risk that neighbouring nodes will monitor and expose the ongoing attacks. There is a more subtle form of these attacks when an attacker selectively forwards packets. An attacker suppresses or modifies packets originating from some nodes, while leaving the data from the other nodes unaffected, which limits the suspicion of its wrong doing.

- ***Byzantine Attack***

A compromised intermediate node works alone, or a set of compromised intermediate nodes works in collusion and carry out attacks such as creating routing loops, forwarding packets through non-optimal paths, or selectively dropping packets, which results in disruption or degradation of the routing services. Byzantine attack disrupts the routing services by dropping, fabricating, modifying, or misrouting packets.

- ***Rushing Attack***

Two colluded attackers use the tunnel procedure to form a wormhole. If a fast transmission path (e.g., a dedicated channel shared by attackers) exists between the two ends of the wormhole, the tunneled packets can propagate faster than those through a normal multi-hop route. This forms the rushing attack. The rushing attack can act as an effective denial-of-service attack against all currently proposed on-demand MANET routing protocols, including protocols that were designed to be secure, such as ARAN and Ariadne. The attacker forwards the route request quickly (as fast as no legitimate node can do). When the neighbours of the target receiving requests from the attacker later receive packets from legitimate nodes, they discard them as duplicate.

- ***Resource Consumption Attack***

This is also known as the sleep deprivation attack. An attacker or a compromised node can attempt to consume battery life by requesting excessive route discovery, or by forwarding unnecessary packets to the victim node.

- ***Location Disclosure Attack***

An attacker reveals information regarding the location of nodes or the structure of the network. It gathers the node location information, such as a route map, and then plans further attack scenarios. Traffic analysis, one of the subtlest security attacks against MANET, is unsolved. Adversaries try to figure out the identities of communication parties and analyse traffic to learn the network traffic pattern and track changes in the traffic pattern. The leakage of such information is devastating in security sensitive scenarios.

5.6 EXTERNAL THREATS FOR MANET APPLICATIONS

In the presence of an authentication protocol to protect the upper layers, external threats are directed at the physical and data link layers. Physical layer security is intrinsically difficult to provide due to the possibly mobile nature of ad hoc nodes. External threats are divided into two major categories: passive eavesdropping, where the adversary simply listens to transmitted signals, and active interference, where the opponent sends signals or data designed to disrupt the network in some way.

5.6.1 Passive Eavesdropping

This can allow unauthorised principals to listen to and receive messages including routing updates. An unauthorised node will be able to gather data that can be used to infer the network topology, and other information such as the identities of the more heavily used nodes which forward or receive data. Hence, techniques may be needed to hide such information. Eavesdropping is also a threat to location privacy. Note that passive eavesdropping also allows unauthorised nodes to discover that a network actually exists within a geographical location, by just detecting that there is a signal present. Traffic engineering techniques have been developed to combat this.

5.6.2 Active Interference

The major threat from active interference is a denial of service attack caused by blocking the wireless communication channel, or distorting communications. The effects of such attacks depend on their duration, and the routing protocol in use. With regard to the routing of data packets, reactive routing protocols may see a denial of service attack as a link break. Route maintenance operations will cause most protocols to report the link as broken so that participating nodes can find an alternative route. Proactive routing protocols do not react immediately to non-delivery of data packets. If the route is believed to be broken, it will eventually be timed out and deleted.

There are also threats to integrity, e.g., where an external attacker can attempt to replay old messages, or change the order of messages. Old messages may be replayed to reintroduce out-of-date information. Out-of-date routing information could lead to further denial of service attacks as nodes try to use old but invalid routes, or delete current valid routes. If the routing protocol utilises neighbour sensing by monitoring received data packets, replaying old packets may falsely lead nodes into believing that an old link with a neighbour has become active and usable again.

5.7 INTERNAL THREATS FOR MANET APPLICATIONS

The threats posed by internal nodes are very serious, as internal nodes will have the necessary information to participate in distributed operations. Internal nodes can misbehave in a variety of different ways; there are four categories of misbehaviors: failed nodes, badly failed nodes, selfish nodes and malicious nodes. Note that two misbehaving nodes within the same category may exhibit

different degrees of incorrect node behavior. For example, some nodes will be more selfish than others. Also, a node may demonstrate behaviors from more than one category indeed, this may even be the typical case.

5.7.1 Failed Nodes

Failed nodes are simply those unable to perform an operation; this could be for many reasons, including power failure and environmental events. The main issues for ad hoc routing are failing to update data structures, or the failure to send or forward data packets, including routing messages. This is important as those data packets may contain important information pertaining to security, such as authentication data and routing information. A failure to forward route error messages will mean that originator nodes will not learn of broken links and continue to try to use them, creating bottlenecks. For example, the threat of having failed nodes is most serious, if failed nodes are needed as part of an emergency route, or form part of a secure route.

5.7.2 Badly Failed Nodes

Badly failed nodes exhibit features of failed nodes such as not sending or forwarding data packets or route messages. In addition they can also send false routing messages, which are still correctly formatted, but which contain false information and are a threat to the integrity of the network.

For example, false route requests for a node which does not exist may circulate in the ad hoc network using up valuable bandwidth, as no node can provide a suitable reply. Unnecessary route requests for routes which badly failed nodes already have, might also be sent. False route replies in response to a true route request may result in false routes being set up and being propagated through the network. False route error messages will cause working links to be marked as broken, potentially initiating a route maintenance procedure. Protocols which rely on neighbour sensing operations are also vulnerable, as false messages may cause nodes to sense extra neighbours.

Protocols such as AODV include within the route error messages a list of affected nodes to which the route errors should be unicast. If this list is large, then the threat not only affects network integrity, but is also a denial of service attack, as resources and bandwidth are being used up by the large volume of route error messages sent, and the unnecessary route requests and replies used to find alternative routes.

5.7.3 Selfish Nodes

Selfish nodes are typified by their unwillingness to cooperate as the protocol requires whenever there is a personal cost involved, and will exhibit the same behaviors as failed nodes, depending on what operations they decide not to perform.

Selfish nodes exploit the routing protocol to their own advantage, e.g., to enhance performance or save resources. Packet dropping is the main attack by selfish nodes, where most routing protocols have no mechanism to detect whether data packets have been forwarded, DSR being the only exception.

5.7.4 Malicious Nodes

Malicious nodes aim to deliberately disrupt the correct operation of the routing protocol, denying network services if possible. Hence, they may display any of the behaviors shown by the other types of failed nodes. The impact of a malicious nodes' actions is greatly increased if it is the only link between groups of neighbouring nodes.

5.7.5 Attacking Neighbour Sensing Protocols

Malicious nodes can either force nodes to incorrectly add neighbours when they do not exist, or cause nodes to ignore valid neighbour nodes. The method will depend on the neighbour sensing protocol but most require the receipt of some form of message. As with a badly failed node, a malicious node can send a neighbour sensing message with a false source address to cause the same effects.

For a malicious node to cause another node to ignore its neighbours, it could perform an active denial of service attack similar to external nodes. However, this could also be easily detected. Thus, this attack will be more successful for the malicious node if it could exploit some other operation such as a blacklist. If a bidirectional MAC protocol is in use, DSR uses a blacklist for neighbours a node believes it has asymmetrical links with. Thus, a malicious node could just try to block transmission in one direction to cause the node to be added to its blacklist. Blacklisted entries either expire or are deleted when bi-directional communication has been confirmed. So, conversely, a malicious node could try to force a node to delete neighbours from its blacklist by masquerading as a blacklisted node, and forward a route request, whose source header contains details of the blacklisted node (its IP address, etc.). A similar attack can be achieved with AODV.

5.7.6 Misdirecting Traffic

As previously mentioned, a malicious node can usually masquerade by just using a false source address in the data packets it sends, as described in FSR. In FSR, nodes examine the IP Header source address and use it as a neighbour address. If the malicious node uses a false address which belongs to another node, then it can affect network integrity by getting all nodes in the network to point their routes to the malicious node, instead of the true owner of the source address. A malicious node can do this in a reactive protocol by replying to route requests before the original owner can, and the same effect can be achieved in a proactive protocol where the malicious node just advertises false routes in the hope they get accepted before the true routes. The malicious node will then receive any information which was intended for the owner of the address. This attack has been named the black hole attack, akin to the celestial structure which sucks in all objects and matter.

Another reason for masquerading in this way is when the malicious node targets another node and cause excess traffic to be routed to it, causing a targeted sleep deprivation attack. A malicious node could send false route requests on behalf of this node, so that other nodes will then direct route replies to the node. Malicious nodes can advertise routes with attractive route metrics and high sequence number so that the likelihood of the false route being accepted is increased. However, the further away a malicious node is, the less successful this attack will be in getting the false routes accepted before the true routes.

5.7.7 Exploiting Route Maintenance

Malicious nodes can simply propagate false route error messages so that valid working links are marked as broken. Resources will be used in attempts to repair the links or find alternative routes. An alternative attack may be for a malicious node to coerce another node into sending route error messages by blocking an operational link (e.g., by blocking acknowledgments in DSR). This attack can also be performed by an external attacker.

5.7.8 Attacking Sequence Numbers and Duplicate Mechanism

Unique sequence numbers prevent replay attacks of old data packets. However, this mechanism can also be exploited to cause a denial of service. A malicious node could flood the network with as many messages with false source addresses containing as many high sequence numbers as possible. Thus any

true messages sent will be discarded as duplicated or out of sequence. This attack is possible because most protocols require nodes to maintain their own sequence number counter, and do not take into account the sequence numbers of received messages. Note that this discussion refers to message identifier sequence numbers and not the sequence numbers used to guarantee route freshness as in AODV and OLSR.

5.7.9 Attacks on Protocol Specific Optimisations

There are many protocol specific attacks. The following describes an attack on the DSR salvaging operation which is used to find alternative routes when a link break is detected. Using the attacks just described above, the malicious node injects into the network as many routes, with as many different next hops, as possible, all of which do not exist and all point to the same target. The malicious node then sends a data packet addressed for that non-existent target. This is a denial of service attack as an intermediate node will now attempt to send route error messages for each next-hop from which it tries to gain an acknowledgment. The intermediate node then tries to salvage the data packet by finding an alternative route. The alternative route is likely to be another false route and this carries on until the data packet has been the subject of MAX SALVAGE TIMES salvage attempts. Of course the malicious node can just send a data packet for the same target to repeat the denial of service.

5.8 SOME OF THE SECURITY SOLUTIONS

5.8.1 Threshold Cryptography in Mobile Ad Hoc Networks

For the construction of threshold cryptography schemes, the problem of distributed key generation, that is, generating a shared public key and individual shares as secret keys, is often crucial.

Threshold cryptography over MANET needs to be based on various assumptions: the MANET topology, the setup infrastructure that parties rely on, the adversary power and mode of operation, the intractability of computational problems, etc. The main problem to consider is that of designing threshold cryptography protocols in MANET that minimise any of the above assumptions, most notably the topology and setup assumptions, under which their security can be proved. In particular, improving topology assumptions is an important problem for secure MANET protocols, as restrictive MANET topologies may be hard to be maintained in practice. For instance, protocols

for wired networks, that implicitly assume (and often require) that the network graph is fully connected, cannot be directly deployed into MANET as partial connectivity and mobility features may make this assumption very often invalid. On the other hand, for certain security goals, such as security against Byzantine adversaries, certain topology assumptions are necessary to achieve security. Specifically, it directly follows from the results in the distributed computing area that threshold cryptography over MANETs cannot be securely implemented for very sparse ad hoc networks (regardless of mobility, which indeed only makes things worse). The other important problem is that of improving setup assumptions, or, in other words, relying as less as possible on the pre-execution of setup protocols realising, e.g., public-key infrastructures, group-based security associations, secure routing or even just physical identity-exchange. An implicit subproblem faced is that of building over threshold cryptography protocols for wired networks so to obtain protocols with similar security guarantees over the more challenging MANETs.

The proposed threshold password authentication in MANETS works as follows. The (t,n)-threshold password authentication scheme, is a collection of n nodes, sharing a system key, are deployed to act as server nodes S, and only if t (<n) of them cooperate, they fulfill mutual authentication with a registered user. The protocol steps are given in the Table 5.1.

Table 5.1 *Protocol steps in threshold password authentication*

1. *User* $\rightarrow$ *Login – Device*: $\{PW,ID\}$.
2. *Login-Device* $\rightarrow$ *S*: $\{ID,T,D = h(ID)^r \bmod p,\ C = h(T||E = B^r||B = \beta - h(PW) \bmod p) \bmod p\}$.
3. S_i: *Verifies the validity of ID*, and *T*.
4. $S_i \rightarrow$ *Dealer*: $\{E_i' \bmod p, B_i' = h(ID)^{x_i} \bmod p\}$.
5. *Dealer*: E', and B'.
6. *Dealer* checks $C = h(T \| E' \| B')$holds. If it does not hold, the request will be rejected. Otherwise, the login request will be accepted.

Key Definitions and Distribution Methodology

When a node joins the network, it is given a system public key and system private key. This pair of keys is shared by all the nodes of the network. Besides the system key, each node also needs a cluster key. This cluster key is unique to every cluster and a single cluster key is shared by all the nodes belonging to a cluster. This key is generated by the cluster head and distributed to all the cluster members. This key is encrypted with the system public key and broadcast by the head. Each cluster head also has a unique pair of public/private key called head key. This private key is known only to the head that

generates it. The corresponding public key is known to all the network nodes. This is done by means of a network wide broadcast that is initiated by each head immediately after it gets elected as the leader. Thus each member node needs to maintain a pair of system keys, a cluster key and a table consisting of cluster ids and the corresponding heads public key. The cluster head has an additional responsibility of storing securely its private key.

Authentication Scenarios

There are three different scenarios where authentication needs to be performed. They are:

1. When a Node Joins a Network for the First Time

This is a trivial case where a strong authentication is done by sending a challenge and receiving a response. The system key pair is used for mutual authentication between the joining node and a existing member of the network. When a new node joins the network and is detected by a cluster head (by means of hello messages), it gets the cluster key and also the table containing the cluster ids and head public keys.

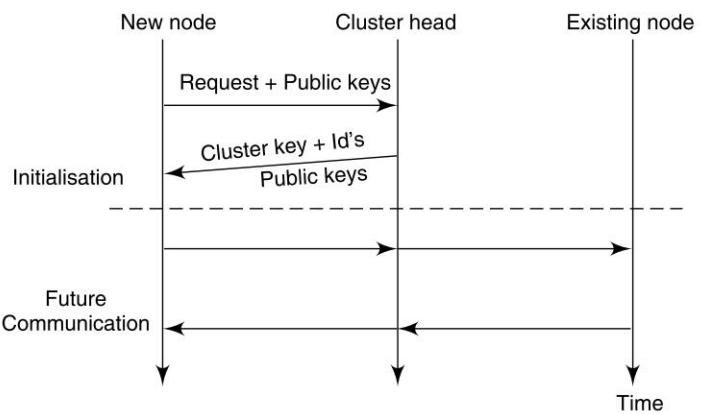


Fig. 5.5 *Timing diagram for node joining first time*

2. When a Node Leaves a Cluster and Joins Another Cluster

This situation arises due to the movement of nodes. When a node moves from a cluster to new one, the new cluster head treats it as any new node joining its cluster. A mutual authentication is performed between the moved node and its new cluster head using the system key pair. The cluster head then gives the node the cluster key for the new cluster. The old cluster purges the entry

for this node when it doesnt receive hello message for a certain predefined time interval.

- **When a Node From a Cluster Wishes to Communicate With a Node Belonging to Another Cluster**

This is a complex scenario and the scheme tries to minimise the overhead involved here. For complete confidentiality of the message, the entire packet has to be encrypted with a session key. The session key is shared solely by the two parties involved in the communication and therefore serves as authentication. But, in cases where the emphasis is on authentication alone and confidentiality is not very critical, it is unnecessary to encrypt the whole packet. A small encrypted tag appended to each packet, is sufficient to achieve authentication. In order to prevent the replay problems there is a need to perform strong authentication for each packet, i.e., a series of challenge and response back and forth. It is not feasible to do this for each packet as the delays and packet overhead would be too high.

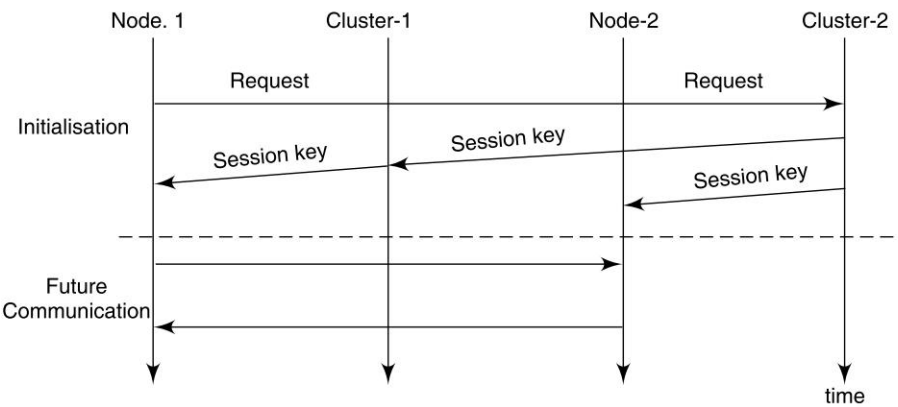


Fig. 5.6 *A node from one cluster communicating with a node in another cluster*

5.8.2 Group Member Authentication Protocol in Mobile Ad-hoc Networks

In a MANET architecture, there is no pre-existing fixed network infrastructure, and a mobile node in this network sends data packets to a destination node directly or through its neighbour nodes. This situation is of potential security concern since the neighbour nodes cannot be always trusted. A group member authentication protocol used in a MANET. It aims to allow a set of nodes to legitimately participate in group communication and then distribute a secret

group key to the approved nodes to establish secure communication with group members. The protocol provides knowledge-based group member authentication, which recognises a list of secret group keys held in a mobile node as the nodes group membership. It employs ZKP (Zero Knowledge Proof) and threshold cryptography.

Both of group member authentication and secret group key management are indispensable procedures for establishing the secure group communication in a MANET. While a secret group key is used only for encrypting and decrypting group communication in general, apply it for identifying the group that a mobile node belongs to. In other words, each secret group key can be defined as a unique identifier; hence the protocol examines secret group keys held by a mobile node and then recognises the criteria of group membership status of the node. In this protocol, a set of the secret group keys on a mobile node is called its knowledge, and recognise the nodes knowledge shows all groups the node previously joined.

When a mobile node wants to become a new group member, the node looks for legitimate group members in the same network and tries to communicate with them. These legitimate group members then investigate the nodes knowledge, compare the knowledge with pre-defined required group membership, and evaluate whether the node can join the group.

In this knowledge verification procedure, an adversary must not be able to succeed to steal any meaningful information even if he eavesdrops all the information exchanged between the new node and group members. Following this line of thought, the protocol employs ZKP algorithm, which gives a method to verify a secret key without disclosure of any secure information. In ZKP, a new node behaves as prover and legitimate group members behave as verifiers. In a ZKP session, verifier does not need to use a secret key for the key verification; while the nodes knowledge consists of secret group keys the node previously joined, the required group membership consists of publicly available verification keys corresponding to the secret group keys.

After the knowledge verification procedure is completed, the new node is ready to obtain the secret group key as the new group member. The threshold cryptography is a beneficial approach as in a secret group key management structure of this protocol; the secret group key is divided to n shares and later generated by a new group member by the response of t group members (among n nodes). This proposal makes the protocol be robust, because it does not require a key server, and hence it works even within a busy MANET.

When the node obtains a new secret group key, the protocol recognises that the node joins the new group and increases its current group membership. In other words, the nodes knowledge can be improved step by step when the node obtains different group keys in the proposed procedure. According to the knowledge-based group member authentication, there are several assumptions

in the protocol. The protocol does not protect the situation that an adversary invades (or cracks) a legitimate group member node and steals a secret group key from the disk or memory on that node. And since a mobile node that has all the required group membership can become the group member (i.e., can obtain the group key) in the protocol scheme.

5.8.3 Intrusion Detection Using Autonomous Agents

An autonomous agents for intrusion detection (AAFID) is intended for intrusion detection in wired networks with a fixed infrastructure. Its basic design, however, is decentralised, and proposed to modify to work in MANETs environment.

The AAFID architecture consists of three main components: *agents*, *transceivers*, and *monitors*. Each node in a network runs one or more agents, which continuously monitor the activity in the node for suspicious behavior, similar to the MLSI events. Each node also has a single transceiver running, which controls and communicates with the agents. The agents alert the transceiver each time they detect a suspicious event. The transceiver may start or stop other agents or issue control commands to running agents in order to reconfigure them to focus on certain aspects of system behavior. If enough suspicious events are discovered, the transceiver may raise an alarm. Hence, the transceiver functions as the data collection and analysis agency within a single host. Whereas agents cannot generate an alarm, transceivers can generate an alarm if they conclude that an intrusion has occurred. The third component monitors run on selected nodes in the network according to the original AAFID architecture. Monitors receive information from several transceivers and also from other monitors. They function as higher-level data collection and processing entities and collect network-wide data. All transceivers report their findings to at least one or more monitors. On the basis of these reports, a monitor may deduce the overall status of the network and take appropriate action. The pseudocode for working of the scheme is given in Table 5.2.

Table 5.2 *Intrusion detection using autonomous agents*

1. *Initialization:* Every node runs instance of an agent, which continuously monitors behaviors of the node.
2. *Operation:*
   (a) IF Behaviors are suspicious then pass the ALERT signal to transceiver.
   (b) Transceiver analyse the data sent from agents.
   (c) The analysis findings are passed onto monitors.
   (d) The monitors further process them by passing suitable messages to agents.

5.8.4 Transactions-Based Authentication Scheme: A Cognitive Agents Based Approach

This approach aims at sensitivity levels of mobile transactions, a user and environment circumstances behaviors, to allocate an authentication scheme dynamically. The scheme uses an authentication database which consists of huge collection of authentication protocols and certificates. Transaction-based authentication propose a scheme in which an authentication is performed at various sensitivity levels of transactions. Once a particular transaction-level is authenticated, the user is not allowed to perform any transactions above that level without further authentication. After initial authentication of a transaction, the scheme continues analysis on client transactions to identify the possibility of attacks.

The authentication scheme uses intelligent agents called Cognitive Agents (CAs): the *Mobile Cognitive Agent (MCA)*, and the *Static Cognitive Agent (SCA)*, which are secured with respect to their construction, and inter agent communications. The total authentication scheme is distributed into two logical parts: the service provider part, and the client part. The architecture of the scheme is given in the Fig. 5.7.

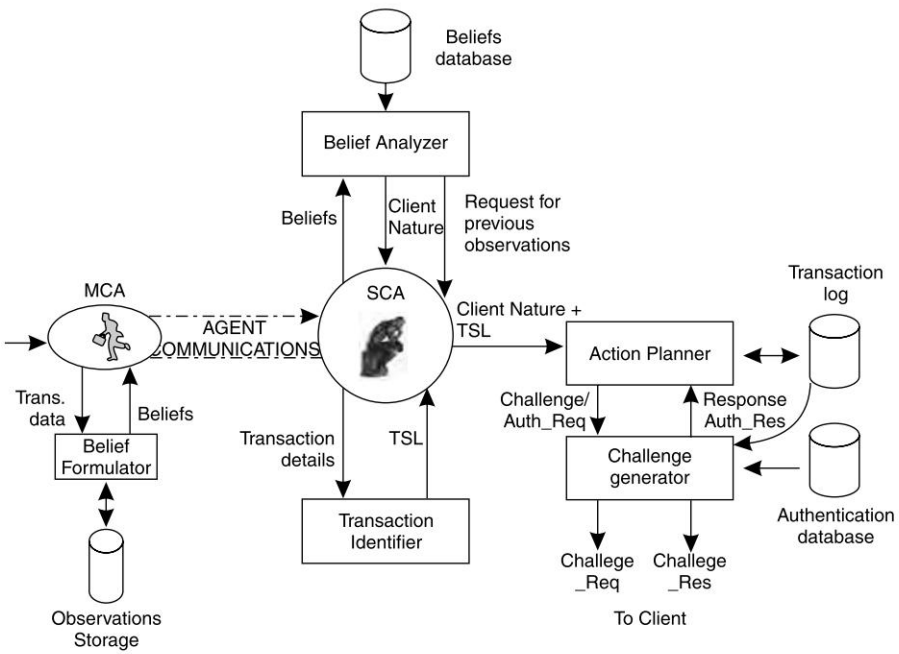


Fig. 5.7 *Transaction-based authentication system*

The MCA, migrates to a client mobile device along with *Belief Formulator* logic, during the service initiation request by mobile user (MU). The agent formulates beliefs using the *belief Formulator*, and communicates them to the SCA along with transaction details. The SCA co-ordinates functions of all the components at the authentication server. It is responsible for migrating the MCA to MU and carrying out communications with the MCA. Upon receiving the beliefs and transaction details from the MCA, the SCA submits them to *Belief Analyzer* and *Transaction Classifier* respectively. Based on the value of cumulative deviation factor, the SCA produces one of the following three types of opinions on the given MU: *NORMAL-USER, SUSPICIOUS-USER, ABNORMAL-USER*. The results obtained from these modules are passed onto *Action Planner* for suitable authentication actions. We explained the procedure of authentication by considering two transactions one at level-1 and another at level-3.

Example 1: Level-1 Transaction: Customer Requesting for Technical Information About a Product

- **Behaviors Observed**

Suspicious duration for data entry, More number of invalid data items entered, More number of mistakes/corrections, and Important data items are skipped.

- **Observation Generated**

Entering data in a hurry and careless data entry.

- **Beliefs Formulated**

Mischievous customer.

- **Belief Deviation Factor**

0.56.

- **Authentication Action**

Please enter your e-mail ID for verification? (since it is level-1 transaction certificates are not used).

Example 2: Level-3 transaction: Customer Placing an Order

- **Behaviors Observed**

Suspicious billing address, Suspicious shipping address, and Suspicious time-of-login.

- **Observation Generated**

Fraudulent behavior.

- **Beliefs Formulated**

Fraudster.

- **Belief Deviation Factor**

0.65.

- **Authentication Action**

Please enter the key for digital signature generation ?

5.8.5 Transactions-Based Security Scheme: A Cognitive Agents Based Approach

The objectives of the Transactions-based security scheme is to identify and select appropriate security technique to secure a given transaction of an on-going communication session between a user and the Application Service Provider (ASP)/Server.

The Transactions-Based Security Scheme (TBSS) architecture (Fig. 5.8) consists of a set of applications resources logically available with a mobile application service provider (ASP), along with the databases for user-profile, network-profile, and device-profile created by considering the history of transactions conducted by the users. The two cognitive agents (MCA and SCA) are used for belief generation, belief analysis and transaction analysis. The transactions security module (TSM) is responsible for identifying the suitable security technique out of repository of security techniques, and executing the selected technique. The TBSS analyses each of the mobile transactions and categorises them into one of the security levels before randomly selecting a security technique for preparation of the cipher.

Example: Consider a sequence of transactions in a customer-singleparty-indirect payment in a mobile commerce application using a mobile banking service. The transactions (their sensitivity level is given in parenthesis in this scenario and their corresponding sensitivity levels are as follows: (T_1: *Registration(1)*; T_2: *Client authentication(1)*; T_3: *Balance enquiry(1)*; T_4: *Transaction submission(2)*; T_5: *Vendor authentication(1)*; T_6: *Bank invoice(2);* T_7:*Client confirmation(2)*; T_8: *Vendor confirmation(2)*; T_9: *E-Money settlement(3)*; T_{10}: *Acknowledgment(1)*). The customer session which is believed to be *genuine*, has the transactions in the order $\{(T_1, T_2, T_3, T_4, T_5, T_6, T_7, T_8, T_9, T_{10})\}$.

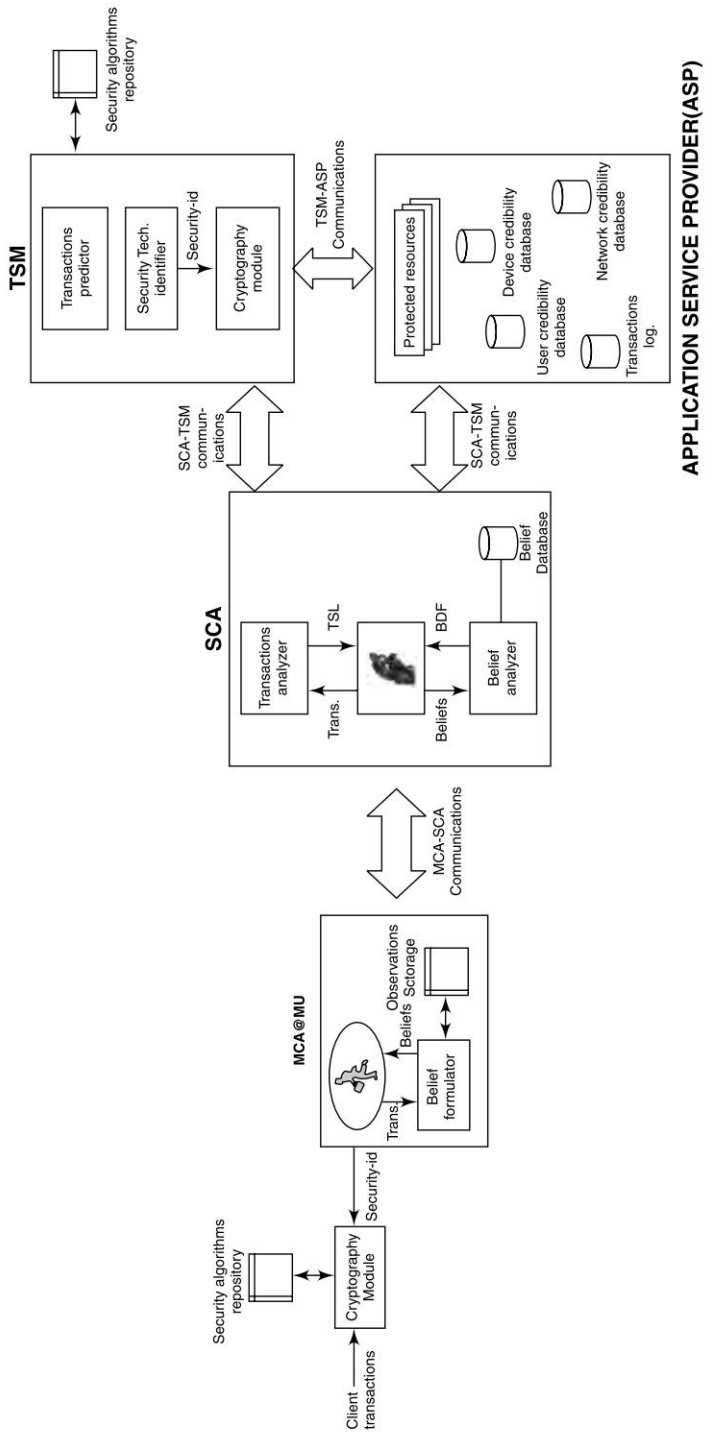


Fig. 5.8 *The TBSS architecture*

Following steps illustrates the dynamic selection of security technique, based on the transaction sensitivity levels and user behaviors.

1. Transaction: T_1, i.e, Registration

- *Sensitivity level:* 1.
- *Behaviors observed:* More number of invalid data items entered, and more number of mistakes/corrections.
- *Observation generated:* Entering data in a hurry and careless data entry.
- *Beliefs formulated:* Mischievous customer.
- *Belief deviation factor:* 0.56
- *Security technique used:* DES(64 bits) selected from the security repository.

2. Transaction: T_6, i.e, Bank Invoice

- *Sensitivity level:* 2.
- *Behaviors observed:* From regular timings, From known location, and From known device.
- *Observation generated:* Normal behaviors.
- *Beliefs formulated:* Regular visitor
- *Belief deviation factor:* 0.32
- *Security technique used:* DES(32 bits) selected from the security repository.

SUMMARY

In this chapter we have introduced the need of MANETs, by specifying some of the typical applications. Detailed discussions on attacks at various layers of MANET is provided. Most of the chapter focus is on summarising various internal and external threats for MANET-based applications. In the end we have provided few popular security schemes for MANET applications.

REVIEW QUESTIONS

1. How MANETs are different from infrastructure-based mobile networks.
2. What type of application are best suitable for MANETs ?
3. Give two typical applications for MANETs.
4. What are the important features of MANETs.

5. What are the security challenges faced while deploying MANET-based applications ?
6. What are the attacks over MANET applications at the physical layer?
7. What are the attacks over MANET applications at the link layer?
8. What are the attacks over MANET applications at the network layer?
9. What are the attacks over MANET applications at the transport layer?
10. What are the attacks over MANET applications at the application layer?
11. What are the advanced attacks over MANET applications?
12. Explain passive eavesdropping threat for MANET applications.
13. Explain active interference threat for MANET applications.
14. Explain threats caused by failed nodes in MANET applications.
15. Explain threats caused by selfish nodes in MANET applications.
16. Explain threats caused by malicious nodes in MANET applications.
17. Explain threats caused by neighbour nodes in MANET applications.
18. What is an impact of exploiting route maintenance attack on MANET applications ?
19. What is an impact of misdirecting traffic attack on MANET applications?
20. What is an impact of sequence numbers attack on MANET applications?
21. What is an impact of protocol specification attack on MANET applications?
22. Explain RSA and Diffie-Hellman based authentication for MANETs
23. Explain group member authentication for MANETs.
24. Explain trust based authentication for MANET applications.
25. Explain threshold cryptography in MANETs.
26. Explain intrusion detections in MANETs using agents.
27. Explain the working of transactions-based authentication scheme.
28. Explain the working of transactions-based security scheme.

6 Application Level Security in Ubiquitous Networks

OBJECTIVES

- To understand the need of ubiquitous computing (UC) and Networks.
- To familiarise with the salient features of UC.
- To discuss the need of the types of security schemes for UC applications.
- To know the various challenges for designing security schemes to UC applications.
- To describe the security characteristics and vulnerabilities of UC networks.
- To understand some of the available security solutions for UC applications.
- To aware of some of the ongoing research projects in UC.

The birth of a revolutionary computing paradigm promises to have an intensive effect on the way people interact with computers, devices, physical spaces and other people. This new technology envisions a world where processing power and digital communication are extremely inexpensive commodities that are widely available. This approach eliminates time and place barriers by making services available to users anytime and anywhere.

In the recent past, personal computers transformed computing into a one-to-one correspondence between people and computers. The natural sequel to this is a computing environment that advocates a one-user to many machines computing model. In such a scenario, users are surrounded with a comfortable

and convenient information environment that merges physical and computational infrastructures into an integrated habitat. Anytime and anywhere computing support for human activities offers novel opportunities to enhance human abilities and experience in business, science, education, medicine, day-to-day living, and entertainment.

The broadband, low-cost mobile communication infrastructure has grown in response to the spread of a wide range of electrical equipments. The list includes cellular phones, Personal Digital Assistants (PDAs), on-board car information equipment, intelligent electrical appliances, third-generation cellular phones, wireless hot spots, LANs, and so on. As a result, a ubiquitous environment that enables users to access services over the Internet anytime and anywhere has become a reality (Fig. 6.1). In the ubiquitous environment, users access the Internet in various places with various communication media and terminals. Therefore, services should provide contents that are personalised to suit the location, the capabilities of the communication terminal, and the users status. Moreover, because users cannot always be aware of changes occurring at the service side, services should be pushed autonomously according to the users status.

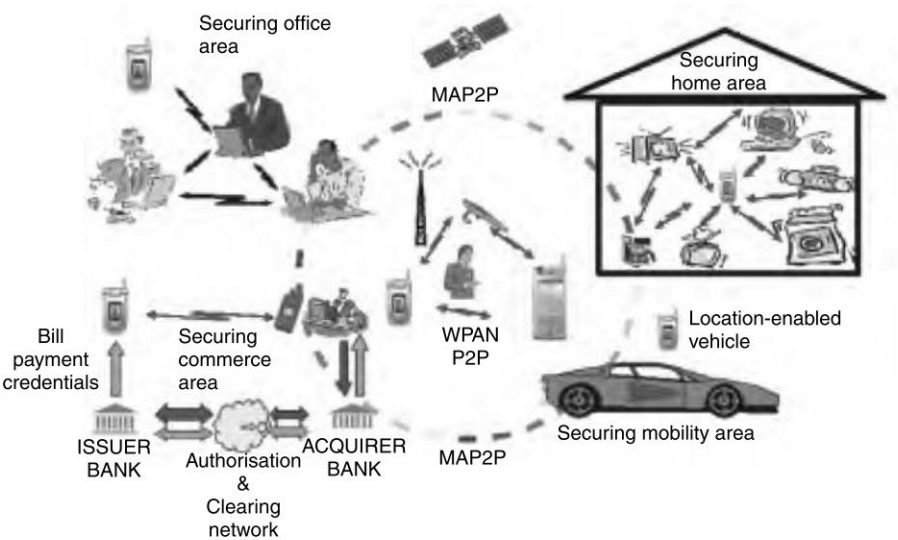


Fig. 6.1 *An example of a Ubiquitous Network Environment*

When services offer contents according to the users status, it is necessary to offer the most appropriate content for the communication media and terminals being used. For instance, consider the case in which restaurant information is provided through a car navigation system. The restaurant's location is dis-

played on the terminal's screen. However, because the driver cannot look at the screen until the car has stopped, the system automatically switches to voice guidance. In addition, when the driver stops the engine and leaves the car, the service continues to send information to the drivers cell phone. This requires a mechanism that makes it unnecessary for the service side to consider the communication media and terminal; otherwise, the extra processing that the service application program needs to perform will cause a cost increase.

6.1 UBIQUITOUS COMPUTING

In ubiquitous computing, user's availability will be maximised with the wired/wireless networks. Users get the proper service anytime, anywhere, without any restriction of time and location. Without user's concerns, user's contexts are sensed by the network to provide the proper service. However, with the increase of user's availability also will occur the increasing of security risks. Sensing users context cause privacy issues. Also, forgery of contexts is also a problem.

A ubiquitous computing environment can be defined as a wide variety of networks that offer attributes as broadband capabilities, mobile characteristics and continuous dedicated access via any mode or medium, such as stationary or mobile terminals, wired or wireless systems, and telecom or broadcasting, etc. The ubiquitous network is capable of supporting any information technology equipment, such as mobile phones, PDAs, car navigation terminals, and consumer information appliances, as well as desktop computers and mobile PCs that are connected or networked with each other via a access friendly interface. It creates an environment in which digital data can be exchanged in an interactive and seamless manner.

6.1.1 UC Vision

The term 'ubiquitous computing' is a very broad term that is often overloaded to mean diverse things to different applications. In many cases, researchers define ubiquitous computing by example, with respect to their own research. Therefore, it is important to define exactly what is the vision of ubiquitous computing. More precisely, ubiquitous computing refers to a proliferation of hundreds or thousands of computing devices, sensors and embedded processors that will provide new functionality, offer specialised services, boost productivity, and facilitate seamless interaction with the surrounding environment and available resources. Ubiquitous computing allows to realise additional abstractions that did not exist in traditional computing paradigms. The salient features of ubiquitous computing includes the following:

Extending Computing Boundaries

While traditional computing encompassed hardware and software entities, ubiquitous computing extends the boundaries of computing to include physical spaces, building infrastructures, and the devices contained within. This aims to transform dull, passive spaces into interactive, dynamic, and programmable spaces that are coordinated through a software infrastructure and populated with a large number of mobile users and devices.

Invisibility and Non-Intrusiveness

People have to change some of their behaviour and the way they perform tasks so that these tasks can be computerised. To boost productivity, it is important that computing machinery disappears from the spotlight. Computers should blend in the background allowing people to perform their duties without having machines at the center of their focus.

Creating Smart and Sentient Spaces

A dust of invisible embedded devices and sensors are incorporated to turn physical spaces into active, smart surroundings that can sense, see, and hear effectively, making the space sentient and personalised. Ultimately, the space should become intelligent enough to understand user's intentions and become an integral part of user's everyday life.

Context Awareness

An ubiquitous computing model should be able to capture the different contexts and situational information and integrate them with users and devices. This allows the active space to take on the responsibility of locating and serving users and automatically tailoring itself to meet their expectations and preferences.

Mobility and Adaptability

To be truly omnipresent, the ubiquitous computing environment should be as mobile as its users. It should be able to adapt itself to environments with scarce resources, while being able to evolve and extend once more resources become available.

6.1.2 UC Applications

This section lists some of the popular applications developed under UC.

Smart Tool Box

Tools are equipped with RFID tags, and the tool box contains a mobile RFID system (including a tag reader antenna integrated into the tool box). The tool box issues a warning for safety reasons if a worker attempts to leave the building site (or a sensitive maintenance area such as an airplane) while any tools are missing from the box. The box also monitors how often and for how long tools have been in use. Based on this information, tools can be replaced before they wear out. Additionally, the tool owner can charge for tool rental based on actual tool usage.

Smart Supply Chain

Smart identification technology can significantly improve the efficiency of supply chains and the internal logistics processes of companies. In such scenarios, the automatic identification and localisation of goods at instance level can help to prevent faulty deliveries and speed up the whole business process.

RFID Chef

In this application, grocery items are equipped with RFID tags (instead of the bar codes that are commonly used). When placed on a kitchen counter with an integrated RFID reader, a nearby display suggests dishes that could be prepared with the grocery items available, or shows missing ingredients. The suggested dishes not only depend on the available ingredients, but also on the preferences of the cook, who might for example prefer vegetarian or Asian dishes. To implement this functionality, the cook is identified by an RFID tag with the form factor of a credit card, carried in his or her wallet.

Context-Aware Mobile Phone

SenSay is a context-aware mobile phone that adapts to dynamically changing environmental and physiological states. In addition to manipulating ringer volume, vibration, and phone alerts, SenSay can provide remote callers with the ability to communicate the urgency of their calls, make call suggestions to users when they are idle, and provide the caller with feedback on the current status of the SenSay user. A number of sensors including accelerometers, light, and microphones are mounted at various points on the body to provide data about the users context. A decision module uses a set of rules to analyse the sensor data and manage a state machine composed of uninterruptible, idle, active and normal states.

Passenger Support System

Passengers can make their travel plans and purchase necessary tickets by accessing databases via the system. After starting the travel, a mobile terminal checks the travel schedule of its user by accessing several databases and gathering various kinds of information. In this application field, many kinds of data must be handled. Examples of such data are route information, fare information, area map, station map, planned operation schedule, real time operation schedule, vehicle facilities and so on. Depending on the user's situation, different information should be supplied and personalised. On the other hand, transport systems can gather information about situations and demands of users and modify their services offered for the users.

Smart Homes

The notion that we could eventually live in so-called "smart homes" domestic environments in which we are surrounded by interconnected technologies that are, more or less, responsive to our presence and actions seems increasingly plausible. The Aware Home project is noteworthy among the smart home researches. In the Aware Home project, they built a three-story, 5040 square-foot home that functions as a living laboratory for interdisciplinary design, development and evaluation.

Ubiquitous Healthcare

Ubiquitous healthcare is an emerging field of technology that uses a large number of environmental and patient sensors and actuators to monitor and improve patients physical and mental condition. Tiny sensors are being designed to gather information on bodily conditions such as temperature, heart rate, blood pressure, blood and urine chemical levels, breathing rate and volume, activity levels, and almost any other physiological characteristic that provides information that can be used to diagnose health problems. These sensors are worn on or implanted in the body, or installed in patients homes and workplaces. Actuators go further and trigger actions such as the release of small quantities of pharmaceuticals into the bloodstream or the electrical stimulation of brain areas. The main purpose of these sensors and actuators is to help patients and their carers monitor health status and design and implement interventions to improve that status.

Ubiquitous Guide for Museums

The ubiquitous museum is an environment which can provide museum-like visiting experience. It tightly combines the real objects and knowledge of

various fields, such as natural science, geological science, artificial scenery, historical relics, and so on. With high mobility of learning devices and interactive, attractive, and location-aware content, it is expected to produce a good visiting effect. Museums are organisations rich in content and their mission is to bring people closer to artifacts and the meanings they convey. The collections of artifacts are at the core of the museums. However, the visitors must be provided with information in order to be able to assign meaning and interpretations related to the artifacts.

6.2 NEED FOR NOVEL SECURITY SCHEMES FOR UC

Ubiquitous computing (UC) environment begins to receive increasing attention as a new paradigm after Internet. Ubiquitous computing is characterised by freedom of movement in both time and location, which means users expect to access resources anywhere and anytime. Therefore, the complexity of security is increased as the security model should consider the factor of location and time.

Due to the dynamism of ubiquitous communications, there exist numerous threats, for example, a hacker can gain control of users devices, eavesdropping of communication channels, modification of sensitive m-commerce transactions, Denial of Service (DoS), transaction of services or goods in other party's identities, etc. Therefore, one must not only provide the safeguards and countermeasures from these threats but also to develop ubiquitous security applications in an increasingly interconnected ubiquitous networks, where there is continuous, seamless use of wireless networking and broadband technologies. In addition, secure communications with Anyone, Any organisations, Anytime, Anywhere, using Any networks and Any devices (popularly called as A6) have to be accomplished.

In ubiquitous environment, most users do not need to know what networks they are currently accessing. The main focus is on personalisation and services and not so much on the infrastructure. The devices, applications and software must move across multiple heterogeneous wireless networks seamlessly without making these transitions apparent to users. The security in ubiquitous computing would be a major issue as individual, groups, and organisations are unlikely to put personal, important, and mission-critical information over an infrastructure that is either not secure or is not perceived to be secure. The security weaknesses of wireless and mobile infrastructure stems from both, the use of multiple incompatible security schemes and due to inherent weaknesses in certain wireless security algorithms (such as wireless LANs).

Depending on the type of data and the cost of possible loss, modification, and stolen data, a security strategy must be devised and implemented. In addition to security and privacy risks, new vulnerabilities arise due to the use of wireless devices. These could lead to possible change/deletion of information, and denial of service. In addition to these, many more security issues arise due to poor implementation, feature interactions, unplanned growth and new flaws that are created due to prior attacks (Fig. 6.2).

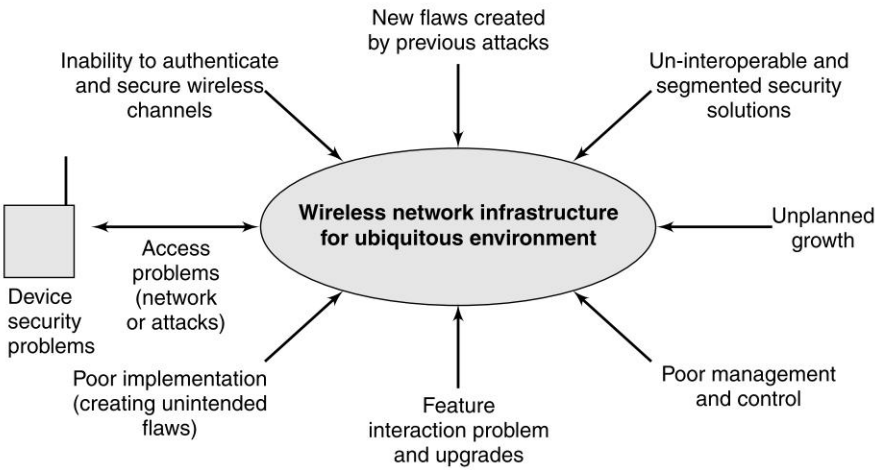


Fig. 6.2 *Security Issues in Ubiquitous Environments*

Traditional security mechanisms require much user interaction in the form of manual logins, logouts, and file permissions. These manual interactions violate the disappearing computer vision and imperil its ubiquitousness. The security requirements of an active space may vary according to the context of the space. Some situations (like during a confidential meeting or homeland security alerts) require greater security to be in place; while other situations may not require a very high level of security. Traditional security mechanisms are context-insensitive, i.e., they do not adapt their security policies to a changing context. Furthermore, traditional distributed systems rely on varying degrees of technical competence on the part of the user of the system to implement and enforce security policies and mechanisms.

As an example, the widespread use of PDAs and cellular phones, along with their value-added services, has already challenged these assumptions severely. An average cell-phone user may not have any knowledge that his or her voice-data is encrypted to prevent other users from listening in or modifying calls. However, the ability of the infrastructure to provide these guarantees at all times is the key to its usability. In a sense, the security mechanisms become invisible, but increasingly crucial, to the correct functioning of the technology.

Furthermore, the closer interaction between the real world and the virtual world presents its own set of challenges. The ubiquitous computing forces us to think about and evaluate security technology using a radically different approach. Issues such as dependability of the components and the infrastructure, as well as the impact of failures on security protocols, now become magnified and increasingly crucial to the survivability of the system. For the reasons stated above, new revolutionary security mechanisms need to be devised. These mechanisms need to be context aware, ubiquitous, and non-distracting.

6.3 SECURITY CHALLENGES FOR UC

Following are the some of security requirements and challenges for ubiquitous networks.

6.3.1 Security Requirements

Confidentiality and Integrity

This is a service to ensure authorised access of information. Ubiquitous network management information needs to be protected in storage and during transmission. One such protection is through password. Other protection could be done through the use of a cryptographic hash of a files contents as the key during the storing and retrieval of the file.

Authentication

This is the most important of all security services, as it allows one entity to verify the identity of another entity. Mutual authentication is required in the ubiquitous networks. Thus, we require mutual authentication protocols to prevent man-in-the-middle for User-to-Device (U2D), Device-to-Device (D2D), Device-to-Network (D2N), and User-to-Service-Provider (U2S) authentications.

Authorisation

This is the process of giving a ubiquitous network device the permission to execute tasks and assign users access rights on that device. For home devices, ubiquitous network environment authorisation corresponds to the users access rights on particular devices. For foreign devices, the owner of the device delegates certain access rights to foreign users who will need to pay for the use of these foreign devices in most cases.

Non-Repudiation

This is a service that prevents an entity from denying previous commitments or actions.

6.3.2 Specific Security Requirements

Following are some of the specific security requirements in UC.

Interoperability with Local Security Solutions

Ubiquitous networks comprise of devices in different security domains. Each domain has the local security solutions but it is doubtful that they will be well matched with security solutions in other domains and at the ubiquitous network level. Since these local security solutions are very difficult to be altered, the security for ubiquitous network architecture needs to be compatible with existing local security solutions.

Availability of Ubiquitous Network Management Functions

Ubiquitous networking is a very dynamic, self-adapting environment with devices joining and leaving the networks. If a device behaved as a gateway to a subnetwork, it will affect the entire subnetwork when it leaves. As the ubiquitous network environment requires to be in proper operation despite these dynamic changes, Ubiquitous Device Management (UDM) function to maintain such operation need to be globally available.

Protection, Revocation, and Renewal of Credentials

Ubiquitous network users credentials exist at different layers. For example, these credentials can exist at the link layer for wired and wireless communications, and IP (and IPSec) at the network layer. At the transport layer, Secure Socket Layer (SSL)/Transport Layer Security (TLS) security protocols could be embedded. The ubiquitous network user credentials also exist at the ubiquitous network overlays, above the transport layer, but below the application layer (middleware layer where the user services run). Of course, all these credentials need to be adequately protected, and protocols put in place for their revocation and renewal. In addition, depending on the technology, the end points of the security associations may differ. Different security protocols exist in the different subnetworks of the ubiquitous network infrastructure; uniform protocols are required at the ubiquitous network level. These protocols unify the existing solutions of a heterogeneous and dynamic ubiquitous environments.

Delegation

Ubiquitous networking has environments that engage numerous devices and services running on these devices on behalf of the ubiquitous network users. Because of the self-adapting characteristics of the ubiquitous networking, a service could change the device or the entire subnetwork where it is running, for example, a device moves from a car network environment into the home network environment. It is very much complicated for the ubiquitous network users to authorise all these changes and therefore it is necessary that the users delegate their rights to a management function acting on their behalf.

Platform Protection

A major motivation behind the development of the ubiquitous networking is the ability to download applications securely to the ubiquitous network devices and allowing the ubiquitous network devices to be reconfigured in a secured manner. Since the goal of the ubiquitous network devices is to give access to a vast variety of services, if restrictions are not placed on the source of downloaded applications, then there is a risk that malicious applications may reconfigure a device in an unauthorised manner. Therefore, it is important to provide some form of Secure Mobile Execution Environment (SMExE) to protect the platform from such attacks.

Single Sign-On

Ubiquitous networks inter-operate with other existing environments, each of which has a specific authentication infrastructure in place. Since the users need to authenticate different devices, networks, and services, all acting in different roles, it is necessary to implement a single sign-on solution. This will allow users to authenticate only once to initiate ubiquitous networks seamless operations in all network domains. This allows the ubiquitous network users to leave and join the ubiquitous networks without any interruptions.

Content Protection

Significant driving force behind the development of the ubiquitous networking is the capability to deliver new services to the ubiquitous network users. As the digital nature of the content allows perfect copies to be made, content providers are naturally concerned that their copyright is protected. For ubiquitous network environments to fully exploit the potential access to mobile content, some forms of Digital Rights Management (DRM) system will be required to be implemented in ubiquitous network devices.

6.3.3 The Extended Computing Boundary

Traditional computing is confined to the virtual computing world where data and programs reside. Current distributed computing research tends to abstract away physical locations of users and resources. Ubiquitous computing, however, extends its reach beyond the computational infrastructure and attempts to encompass the surrounding physical spaces as well. Ubiquitous computing applications often exploit physical location and other context information about users and resources to enhance the user experience. Under such scenarios, information and physical security become interdependent. As a result, such environments become prone to more severe security threats that can threaten people and equipment in the physical world as much as they can threaten their data and programs in the virtual world. Therefore, traditional mechanisms that focus merely on digital security become inadequate.

6.3.4 Privacy Issues

The physical outreach of ubiquitous computing makes preserving of user's privacy a more difficult task. Augmenting active spaces with active sensors and actuators enables the construction of more intelligent spaces and computing capabilities that are truly omnipresent. Through various sensors and embedded devices, active spaces can automatically be tailored to user's preferences and can capture and utilise context information fully. Unfortunately, this very feature could threaten the privacy of users severely. For instance, this capability can be exploited by intruders, malicious insiders, or even curious system administrators to track or stalk particular users electronically. As a result, the entire system becomes a distributed surveillance system that can capture too much information about users. In some environments, like homes and clinics, there is usually an abundance of sensitive and personal information that must be secured. Moreover, there are certain situations when people do not want to be tracked.

6.3.5 Trust Security

Ubiquitous computing means that computers are increasingly involved in all aspects of everyday life. The security gets interwoven with human security, i.e., individual and public security including privacy and trust, even safety and dependability. Security issues become ubiquitous since virtually any (non) action in daily life can potentially imply juridical or financial consequences, and even the most banal activities cause privacy concerns when observed by computers.

Trust is a relationship between two entities such that one entity believes, expects, and accepts that the other trusted entity will act or intend to act beneficially. Trust represents the degree to which a node would be trustworthy, secure, or reliable in any interaction with the node.

In ubiquitous computing, interaction between mutually unknown smart artifacts can take place only if there is an adequate level of trust between the parties. A trust security task will provide devices with the ability to operate and make security related decisions autonomously. While trust defies stringent definition, it is proposed that a model with explicit trust values can be realised in sufficient detail to be used either to augment other security mechanisms or as a basis for unencrypted interactions. With a range of explicit values representing trust, a finer granularity of representation is achieved, providing entities with enhanced information on which to base decisions. Thus the trust security task provides the security to the devices in the network as well as it generates the trust on the new devices depending on their behaviors. Trust security can be provided to devices that are present in the ubiquitous network by formulation of a set of rules in a trust security task.

6.3.6 Social Issues

Social issues include individual, group, and organisational behaviours that are affected by ubiquitous computing. Following questions are asked to find social consequences of technologies: *What if technology was literally untethered by any physical connection to a network, to a workspace, or to an organisation? What new ways to communicate, collaborate, coordinate, organise, and manage*? Ubiquitous computing technologies not only enable new ways of acting and interacting, but also stimulate fundamental reassessments of the meaning of human action and interaction.

6.3.7 User Interaction Issues

One of the main characteristics of ubiquitous applications is a rich, straightforward, unobtrusive user-interface for interactions between users and the surrounding spaces, as well as interactions among the different users. A variety of multimedia mechanisms are used for input and output, and to control the physical aspects of the space. The set of users in the space affects the security properties of the space. Because of the nature of group interactions between users, users in the space cannot easily be prevented from seeing and hearing things happening in it, so this has to be taken into account while designing some security mechanisms. Thus the physical and virtual aspects of access control for such spaces have to be considered together.

6.3.8 Security Policies

It is important in ubiquitous computing to have a flexible and convenient method for defining and managing security policies in a dynamic, context-aware fashion. This is because the security rules of an active space may vary according to the context of the space. As a result, the security subsystem has to support a security policy language that is descriptive, well-defined, and flexible. The language should be able to incorporate rich context information as well as physical security awareness.

6.3.9 Information Operations

There is a great deal of concern over new types of threats, namely, Information Operations (info ops) and cyber-terrorism, which are natural consequences of the increasing importance of electronic information and the heavy reliance on digital communication networks in most civilian and military activities. Info ops, which can be defined as "actions taken that affect adversary information and information systems while defending ones own information and information systems," is a serious concern in the networks. In such a scenario, cyberterrorists and other techno-villains can exploit computer networks, inject misleading information, steal electronic assets, or disrupt critical services. Ubiquitous computing provides additional leverage and adds many more capabilities to the arsenal of "info warriors," making info ops a much more severe threat.

6.4 SECURITY ATTACKS ON UC NETWORKS

In this section we briefly discuss some of the security attacks on UC networks.

Man-in-the-Middle Attacks

When appliances offer physical services such as playing music or delivering goods or money, the user has to verify that the appliance he/she is holding or touching will really deliver the service. In other words, he/she has to authenticate the appliance. Otherwise he/she could pay for a service provided to someone else. When a user has to provide a secret (e.g., password, PIN code) to an artifact or has to delegate it some rights, it is also mandatory to authenticate the artifact.

Ubiquitous Computing Man-in-the-middle attacks occur when actors, which can be artifacts or users, forward challenges and responses in order to simulate

the presence of other actors. In a regular scenario a client plugs his credit card and uses inputs and outputs of the terminal. Even with correct security protocols and tamper-resistant point of sale terminals, a masquerade attack is possible: a dummy terminal is proposed to the client and his/her inputs and outputs are modified before being redirected. A dummy credit card is plugged in the real terminal and acts as a proxy. Mutual authentication between the right terminal and the users credit card succeed but the user is not holding this right terminal. As a result, the attacker can modify the transaction without tampering with the terminal and without stealing the card. This attack, which cannot occur in virtual context, is possible because there is no way for the card to verify if it is plugged in the right terminal.

A similar attack can be mounted against a tamper-resistant appliance offering services to visitors. For example, suppose that a shop offers a discount to any customer coming frequently enough. In this shop, a short-range local transmitter broadcasts random challenges periodically. Each visitor can return his ID certificate and a challenge signed with his private key. The shop is then able to list the users that are present and that will receive the discount. Unfortunately, any visitor can forward challenges to other remote users and build a peer-to-peer location sharing system in which any member of the group can pretend to be present in order to get discounts.

As a result, man-in-the-middle attacks allow the impersonation of artifacts and users. It is already a relevant attack against point of sale terminals and will become more frequent when numerous micro-payments and rights delegations will occur daily within ubiquitous computing. It is necessary to defeat that kind of attack.

Access Network Attacks

The access network is the one that connects the home gateway to the outside service provider. The financial data, user ID and other information can be exposed when the attacker collects the network packet at the household network connection point.

Illegal Connection Attacks

The household appliances can be connected to the wired or wireless network through the home gateway. Typically, home gateway has the Web based management program installed. Its problem is that the attacker can attain the administrator privilege using Web server or CGI vulnerability. Since the home gateway is the point that connects the household with outside, attack against it can directly lead to the attack against the whole household network.

Furthermore, there is the possibility that the attacker can disguise itself as the internal user through the interactive Digital TV, IP set top box or home pad or access it illegally through other means to control the home appliances. Physical problem or malfunction of the home appliance can also leak the information, or problems of the device can cause inconvenience to the user when needed.

Capturing Sensitive Data

The most vulnerable part in the ubiquitous systems is the sensor implanted, because it is the computationally weakest device involved in the monitoring process. If an attacker is able to place a receiver very close to a sensor, it might be possible to acquire sensitive data directly from the implanted sensor. Most of the time the computing power of a sensor will be spent on sensing tasks instead of cryptographic functions.

Stealing Intermediary Device

An attacker could steal an intermediary device after it has collected sensor data. If the intermediary device is designed in a way that makes physical tampering evident, the damage could be contained, since only a small subset of sensor data might be compromised, and that device cannot be reused. On many occasions such a device most likely be equipped with a maintenance interface (in either wired or wireless networks) on many occasions this forms a potential vulnerability.

Data Manipulation

The authenticity of a sensor data cannot be certified directly by the sensor itself, due to computational restrictions. However, an intermediary device can keep record logs of incoming sensor data, and certify the findings on its own behalf. This makes it technically impossible to firmly attribute data to the sensor, but by use of certain means, the trust in the authenticity of data can be increased. The data manipulation can be handled by using the encryption/decryption techniques, but it always remains as a question how these techniques could be deployed on computationally poor infrastructure.

Impersonation and Insiders

An attacker might trick a monitoring equipment by pretending to be a technician or a physician. The attacker might exchange devices with fakes, or install additional equipment for surveillance. In otherways, an impersonated attacker can get free services from the service providers.

Denial of Service (DoS)

There are many possibilities of performing a DoS attack on a weakly protected monitoring system. Transmission links can be jammed, batteries within the sensor and the intermediary device can be drained, the communication interfaces of the medical center could be overloaded (e.g., by a conventional DoS attack in the Internet), and computationally intensive processes be injected in the centers processing plants. Therefore it is necessary to identify these situations in advance, and the security system should be pro-active to take necessary measures.

6.5 SOME OF THE SECURITY SOLUTIONS FOR UC

This section briefly discussess some of the solutions designed for providing security and authentication to UC applications and networks.

6.5.1 Real-Time Intrusion Detection in UC Networks

Intrusion Detection Systems (IDSs) are widely used to protect computer networks. They detect and make alarms when intrusions have taken place or are taking place in the networks. Existing IDSs have several weaknesses that hinder their direct application to ubiquitous networks. These shortcomings are caused by their lack of considerations about the heterogeneity, flexibility and resource constraints of ubiquitous networks. To overcome these issues, a service-oriented and user-centric Intrusion Detection System (SUIDS) for ubiquitous computing environments like a smart home/office is developed. Briefly, in SUIDS, service-oriented event records and user profiles are used to audit users activities and protect various networked appliances against intrusions.

A user-centric approach is proposed to spontaneously compose a defense wall against malicious users. In SUIDS, a users long-term behaviour is represented by probability distributions. They indicate the possible results and corresponding probabilities of a user's each kind of action. For example, statistical results may suggest that the typical time for Mike to open his home door during a day is between 8 to 9am and 5 to 6pm. This action rarely happens during other time. In SUIDS, a string-based method to determine the detection window is used. The 'string' is used to indicate the users short-term behaviour. For example, if the last 100 printing operations can effectively represent

Mikes short-term behaviour regarding his usage of the printer, a string with the length of 100 will be set to follow the printing probability distributions in his profile.

Table 6.1 *Steps in Real-Time Intrusion Detection in UC Networks*

1. Accumulate users long-term behaviours.
2. Develop probability distribution for service users long-term behaviours.
3. Obtain the current behaviour of a user.
4. Compute the statistical deviation between established behaviours and current behaviours.
5. Decide is it intrusion or not.

6.5.2 Role-Based Access Control System

RBAC is based on the principle that access control decisions, which is nothing but various roles an individual can take on as part of an organisation. The key concept in RBAC is a role, which is a placeholder for a set of users. Each role is associated with a set of permissions, which are its rights on objects. These roles may be organised into a hierarchy to reflect the organisational hierarchy among different users or entities in a system. RBAC maintains two mappings: the User Role Assignment (URA) and the Role Permission Assignment (RPA). These two mappings can be updated independently. Users can be added to the URA without changing the RPA, automatically providing new users a predefined role in the system. Similarly, the RPA associates all users in the system to a limited set of permissible behaviours, and can be updated independent of the URA. The key insight in RBAC is that the URA and PRA change less frequently than the permissions of individual users. RBAC has been adapted for use in ubiquitous computing environments, and the concept of roles is extended to deal with context information. While networked applications have traditionally attempted to hide physical location, by providing uniform interfaces for local and remote users to access services, in ubiquitous computing environments.

Table 6.2 *Steps in RBAC*

1. Obtain the user role.
2. Enumerate the privileges associated with that role.
3. Obtain the current action of a user.
4. Check is it allowed under that role with sanctioned privileges.
5. Authenticate a user if it is success.

6.5.3 Trust-Based Security Solution

Large, open systems do not scale well with centralised security solutions. Instead, a security solution based on trust management is proposed, which involves developing a security policy and assigning credentials to entities. It depends heavily on the delegation of trust to third parties. In this environment, users can access nearby smart devices via their handhelds which are connected using Bluetooth. The solution extends SPKI and RBAC.

Every domain in the distributed system contains security agents that are responsible for authenticating and verifying entities within their domain. The agents are arranged hierarchically and use X.509 certificates for identification. The security policies are based on the roles assigned to the user accessing the service. These roles can be delegated and revoked by authorised users. This notion of delegation makes the authorisation service effective, dynamic and manageable. Nevertheless, the system depends heavily on delegation of trust, which may be difficult to have in a ubiquitous environment with mobile users and devices.

Table 6.3 *Steps in Trust-based Security*

1. For a given entity, assign credentials.
2. Set security policies.
3. List the trusted entities, and assign them to given entity.
4. Obtain the trust over user/artifact from trusted entities.
5. Establish the trust on a new user/artifact based on the feedback given from trusted entities.
6. Perform trust-based authentication/access control.

6.5.4 Local Proof of Secret

It is a protocol which is able to verify that a secret is locally known in order to forbid man-in-the-middle attacks in ubiquitous computing.

Figure 6.3 shows how a user A can authenticate (label 1) a virtual entity B. For instance, he can use a terminal to send a challenge to B and verify that the response is right (signed by the private key KS_B of B). A trusted third party can certify (label 2) that B has some properties. For example, A can verify and display an attribute certificate describing B. In the second part of Figure, A sees or touches (label 3) a physical artifact B_P. It is yet necessary to verify that the authenticated virtual entity B is embodied in the artifact B_P that is held by A. In (label 4), A generates a new secret and encrypts it with the public key $K_{P\text{-}B}$ of the entity pretending to be in front of him. Only the owner of the private

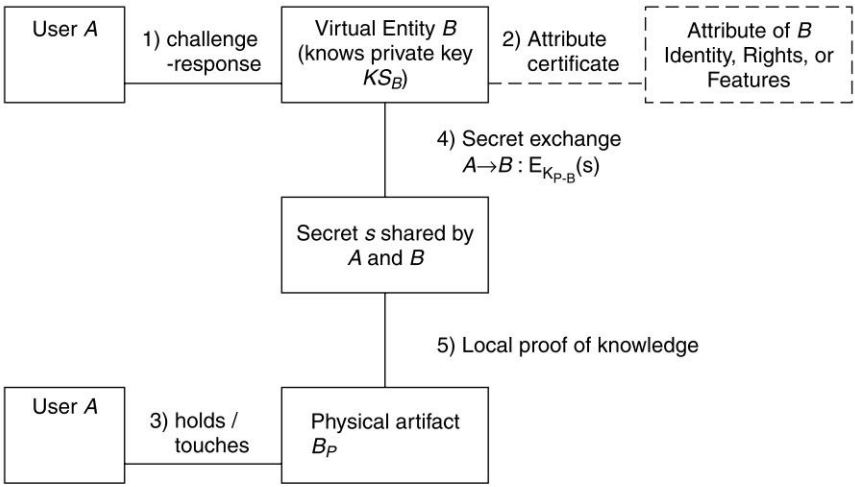


Fig. 6.3 *Local Proof of Secret for Authenticating Artifacts*

key can share this secret with A as long as it does not disclose it. Finally, in (label 5), The local proof of secret protocol is used to verify that the artifact knows the secret.

The local proof of secret is based on a message round trip time (RTT) measurement. If a user could check in one nanosecond that an artifact knows a secret, it could not be farther than fifteen centimeters (due to the physical limit imposed by the speed of light). To reach such a high performance, it is not possible to rely on application layer. The exchange has to occur at the physical layer and the messages have to be as short as possible. One-bit challenges and one-bit responses are exchanged by simple dedicated hardware (logical gates). As a first step, physical contact between artifacts has been chosen because it does not require any distance measurement.

6.5.5 RFID-Based Authentication Protocol

A Radio Frequency Identification (RFID) tag is a microchip that is capable of transmitting a unique serial number and other additional data through RF (radio frequency) signals. The goal of a RFID system is to identify objects remotely by embedding tags into the objects. For example, goods in shops can be tagged in order to provide automatic theft-detection, or to manage the goods inventory by using wireless scanning without any handwork. RFID tags are useful tools in manufacturing, supply chain management, inventory control, etc.

In ubiquitous computing environment, components of the RFID systems can exist anywhere. As schemes described, if a tag's ID should be dynamic value to

protect a user privacy, the tag only communicates with a fixed back-end database since the tag must synchronise the tag's dynamic ID value with the back-end database. However if a tag's ID is static value, then the tag can perform authentication protocol with any back-end database since the scheme does not need synchronisation of the tag's ID between a back-end database and the tag. Therefore, the tag holding static ID is able to communicate with any reader in ubiquitous computing environment.

In RFID systems, since an adversary can monitor all messages transmitted in wireless communication between a reader and a tag, the adversary can infringe upon a person's privacy using various methods. Therefore, RFID systems must be designed to be secure against attacks such as eavesdropping, traffic analysis, message interception and impersonation (e.g., spoofing and replay).

Information Leakage

A person is prone to carrying various tagged objects in every life. Some of objects such as expensive products and medicine are quite personal and provide information that the user does not want anyone to know. In RFID systems, the tag emits only distinguishable information in response to a query from a nearby reader. So, various personal information can be leaked without the acknowledgment of the user.

Traceability

When a target tag transmits a response to a nearby reader, an adversary can record the transmitted message and can establish a link between the response and the target tag. Once a link is established, the adversary is able to know the user's location history.

Table 6.4 *Steps in RFID-based Authentication Protocol*

1. Read RFID-tag from an entity.
2. Pass RFID-tag to database server.
3. Check the RFID-tag for validity.
4. Authenticate an entity is the RFID-tag matches.

6.5.6 Biometrics

Biometric authentication techniques sparked an enormous interest lately. Biometrics show good potential for providing seamless and automated methods for determining and confirming identity while being less obtrusive. Good fingerprint recognition or face recognition techniques, for example, are faster

than entering secure passwords and do not require users to carry special equipment (e.g., PDAs or badges). However, biometric authentication is plagued with several shortcomings. Many biometric authentication techniques have overt characteristics, i.e., the authentication data is often observable to everyone (handwriting signatures can be observed and forged, and fingerprints can be extracted relatively easily). Accuracy and seamlessness of biometric authentication techniques are very dependent on hardware. Finally, biometric authentication techniques still lack a good and secure method of storing the biometric features in a way that prevents compromise of sensitive data and preserves anonymity while providing enough flexibility to accommodate partial matches and deduce a suitable confidence level.

SUMMARY

In this chapter we have presented ubiquitous computing a new paradigm along with its vision and the need of novel security scheme. We have listed applications which demand the use of this computing paradigm. We have presented security challenges and attacks over the applications developed over UC. Some of security schemes proposed by research community is presented here.

REVIEW QUESTIONS

1. What is ubiquitous computing paradigm ? How it is different from earlier paradigms?
2. What is the vision of the UC?
3. Why are traditional security schemes not suitable for UC applications?
4. What are the security challenges for UC applications development?
5. What are the specific security challenges for UC applications development?
6. Explain user-interaction issues and privacy issues in implementation UC security.
7. What are various specific security requirements in UC security?
8. What are the user interaction issues for UC applications development?
9. Why the computing boundary is extended in UC applications development?
10. What are the security policies laid out for UC applications development?
11. What are the security characteristics of UC applications?
12. What are various security attacks on UC applications?
13. Explain the working of real-time intrusion detection in UC applications.

14. Explain the working of role-based access control in UC applications.
15. Explain the working of trust-based authentication in UC applications.
16. Explain the working of local proof of secret.
17. Explain the working of RFID authentication protocol in UC applications.
18. Explain the advantages and risks in Biometric authentication.
19. Explain the working of biometric authentication in UC applications.

7 Application Level Security in Heterogeneous Wireless Networks

OBJECTIVES

- To know about the importance of integrated heterogeneous wireless networks.
- To study some of the architectures of heterogeneous wireless networks.
- To discuss some applications of heterogeneous networks in disaster management.
- To understand the research issues, security problems, and security attacks on heterogeneous wireless networks.
- To illustrate a few security and authentication solutions available for heterogeneous wireless networks.

Wireless communication technologies cover a whole spectrum from Wireless Personal Area Networks (WPAN), such as Bluetooth, to third-generation cellular networks (3G), such as CDMA2000 and UMTS. Despite such variety, opinions differ on which technology is optimal for satisfying all communication needs because of differing coverage and bandwidth limitations. For example, 3G networks provide widespread coverage with limited bandwidth (up to 2 Mbps). However, Wireless Local Area Networks (WLAN, IEEE 802.11) provide high bandwidth (up to 54 Mbps) with relatively smaller coverage area. For ubiquitous and high-performance wireless networking services, the interworking between wireless networks is extremely important. Most inter-working studies have been dedicated to the integration of 3G and WLAN.

Cellular and WLAN systems face distinct security challenges, and each has addressed security in unique (although not necessarily perfect) ways. Although fraudulent access has been reduced in 3G systems compared to previous generations, the major role of 3G in packet-switched services introduces new challenges regarding security. And the weakness of WLANs original security architecture, WEP (Wired Equivalent Privacy), spurred the creation of the WPA (Wi-Fi Protected Access) security architecture by the Wi-Fi Alliance and the IEEE 802.11i task group. Security and performance are major challenges to the inter-working of 3G and WLAN, especially for access control and privacy of mobile stations. The composition of two secure architectures may produce an insecure result. This occurs because of differing, possibly contradictory, security assumptions, e.g., the compromise of a session in a WLAN network may endanger subsequent sessions in 3G systems. Furthermore, support for high bandwidth service with mobility demands a highly efficient authentication mechanism during handover. When a mobile station switches connectivity to a different network, the mobile station and the network have to authenticate each other. However, the authentication process required by each individual network tends to be complicated and costly. For example, the GSM technical specification on performance requirements assumes that the mobile station responds to an authentication request from the network in just under 1 second. In WLAN, EAP-TLS authentication takes about 800 ms. Long authentication delays during handover can cause a disruption of service that is perceivable by users.

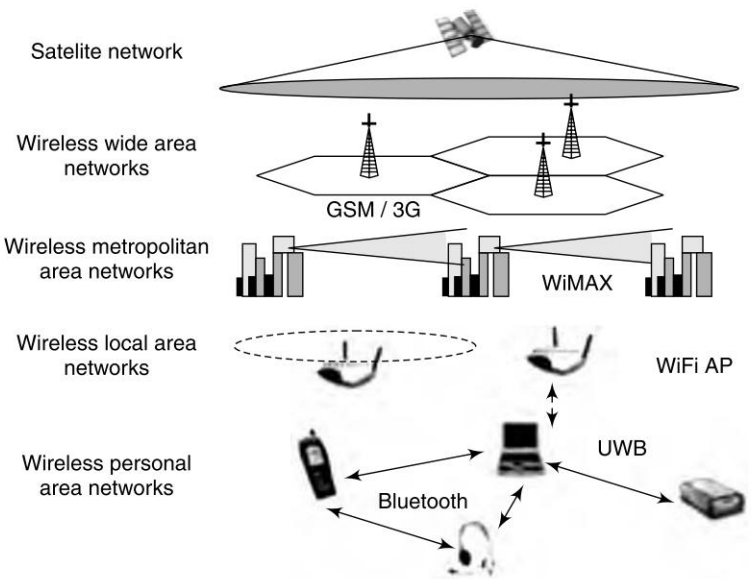


Fig. 7.1 *An example of heterogeneous wireless networks*

7.1 INTRODUCTION

The idea of combining heterogeneous access technologies is not a new idea initiated for wireless networks. In wired networks, multi-homing allows a user to access the Internet using multiple access technologies, e.g., dial up, ADSL, cable, Ethernet, etc. Integration of heterogeneous technologies is also found in wireless voice area. The DECT/GSM dual mode telephones allow the user to use the wide-area GSM cellular network whilst out of the office or away from home, while also using the same handset at home when in range of a cordless telephone BS.

However, the design of a network architecture to fulfill the objective of seamless and efficient integration of heterogeneous Radio Access Technologies (RATs) such as 3G and WLAN is still a challenging task. Wireless LANs, originally targeted at enterprise and home networks, lack many of the capabilities which are essential in public environments, such as unified and universally accepted AAA mechanisms, the integration of mobility mechanisms with QoS and application-level services, the support for roaming agreements, security issues. Conversely, although these characteristics are present in the design of 3G networks, their implementation depends on specific wireless access architectures such as CDMA2000 or UMTS, and their extension to other wireless technologies such as 802.11 presents several compatibility issues. In addition to tradition cell-based roaming (horizontal roaming), fast and transparent roaming across the constituent networks (vertical roaming) raises a challenge for mobility management, with the requirement of switching to a different network interface. Transmission protocol should also adapt to multiple interfaces accordingly.

There are certain open research issues in each of the converged network components and towards their emergence as a complete integrated framework, some of them are:

- QoS should be dealt with at different layers of the protocol stack such as Routing Layer, Application Layer, and Lower Layers.
- Scalability in multi-hop routing without drastically increasing the overhead.
- Credit charging and rewarding mechanisms in ad-hoc routing based converged networks.
- Efficient mobility and connection management approaches to reduce the delay and packet loss during inter-network handover.
- Interaction and messages exchange between Layer 2.5 and ad-hoc convergence layer.
- Service gateway discovery mechanisms and end-to-end security in heterogeneous wireless networks.

7.2 SOME OF THE HETEROGENEOUS WIRELESS NETWORK ARCHITECTURES

7.2.1 iCAR Architecture

An integrated cellular and ad hoc relaying (iCAR) system is proposed which enables a cellular network to achieve a throughput closer to its theoretical capacity. This approach is based on dynamically balancing the load among different cells. As illustrated in Figure 7.2 a number of Ad hoc Relaying Stations (ARS) are deployed at appropriate locations. The ARSs serve to relay excess traffic from a heavily loaded cell to a lightly loaded cell in the same vicinity. ARSs communicate with BS in infrastructure mode, while they communicate with MNs and other ARSs in ad hoc mode. The coordination between BSs, ARSs and MNs is handled by a control protocol. This work leverages the potential of ad hoc networks achieving a throughput closer to cells theoretical capacity. However, some limitations are noticed as signaling overhead, non-optimal routes and hardware complexity due to the number of radio interfaces that must be supported by the MNs. A similar approach, relaying excess traffic, is shown in MADF (Mobile Assisted Data Forwarding) architecture. A main difference is that MADF doesn't use dedicated nodes to relay excess traffic as iCAR does, but rather MADF uses MNs.

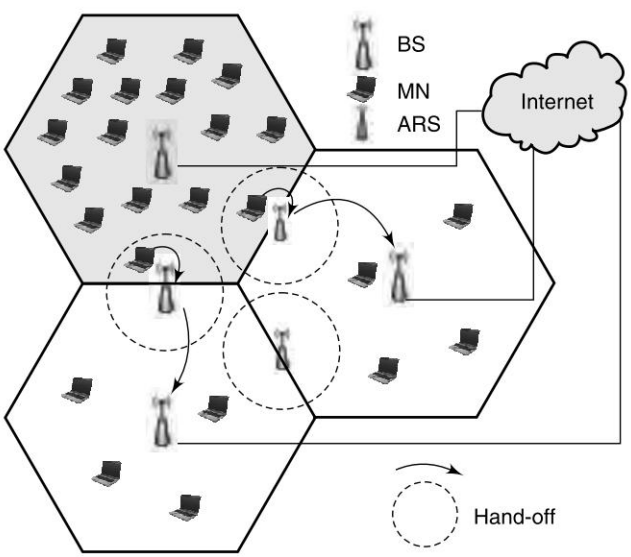


Fig. 7.2 *iCAR architecture*

7.2.2 SOPRANO Architecture

The SOPRANO (Self-Organizing Packet Radio Ad hoc with Overlay) is a wireless multi-hop network overlaid on a cellular structure as shown in the Figure 7.3. The goal is to provide high data rate Internet access by using inexpensive dedicated relay stations. More channels are available in the network and path losses are reduced, using the technique of cell splitting. Hence the network capacity can be maximized by choosing a suitable routing strategy which is a function of the techniques used in the physical layer. This approach can assure load balancing and potential throughput enhancement. Also, it succeeds in extending the BS zone without requiring multiple radio interfaces to be supported by MNs.

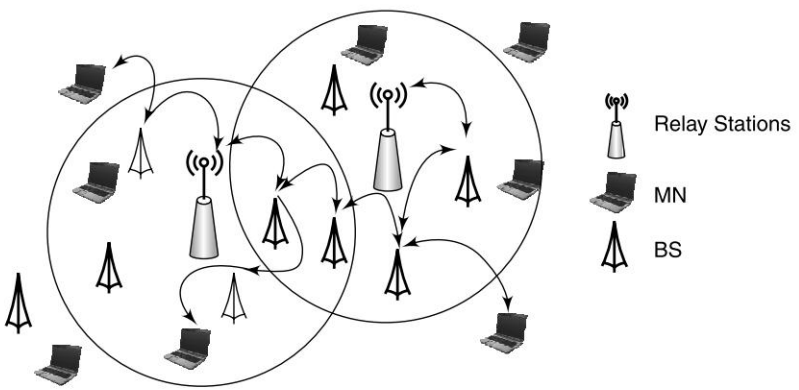


Fig. 7.3 *SOPRANO architecture*

7.2.3 Roofnet Architecture

A Roofnet 802.11 architecture consisting of unplanned node placement, omni-directional antennas and multi-hop routing. Figure 7.4 illustrates this architecture. The goal is to evaluate the performance of a wireless mesh architecture providing Internet access with little deployment planning and little operational management. A performance evaluation shows that Roofnet's multi-hop mesh increases both connectivity and throughput compared to a hypothetical single hop network. Since this architecture is quite static, it cannot be considered as a reference model in a real dynamic and ubiquitous environment.

Each Roofnet node consists of a PC, an 802.11b card, and a roof-mounted omni-directional antenna. The PCs Ethernet port provides Internet service to the user. Each PC has a hard drive for collecting traces and a CD reader in case an over-the-network upgrade fails. An entire Roofnet kit (PC, antenna,

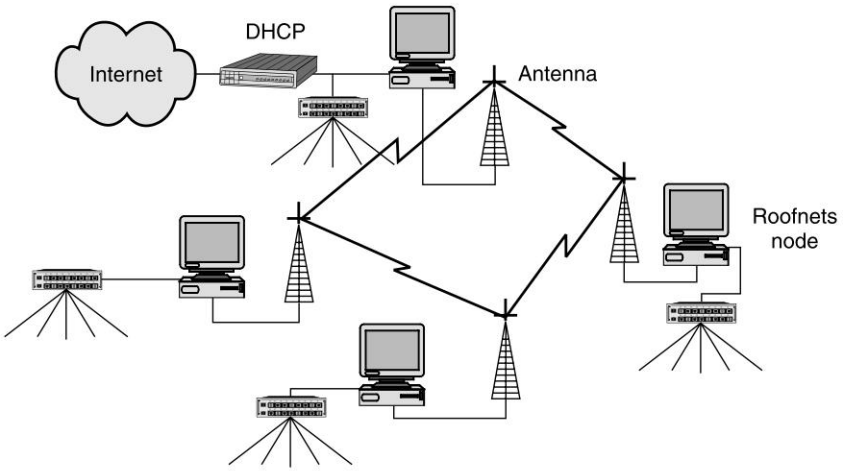


Fig. 7.4 *Roofnet architecture*

mounting hardware, and cable) can be carried by one person. Each Roofnet node runs identical turn-key software consisting of Linux, routing software implemented in Click, a DHCP server, and a web-server so users can monitor the network status.

From the users perspective, the node acts like a cable or DSL modem: the user connects a PC or laptop to the nodes Ethernet interface, and the node automatically configures the users computer via DHCP, listing the node itself as the default IP router. Some users choose to connect the node to their own wireless access point. In order that Roofnet nodes be completely self-configuring, the software must automatically solve a number of problems: allocating addresses, finding a gateway between Roofnet and the Internet, and choosing a good multi-hop route to that gateway.

7.3 HETEROGENEOUS NETWORK APPLICATION IN DISASTER MANAGEMENT

When a large-scale disaster strikes, first responders are sent to the site immediately. Once the most pressing needs of the disaster are addressed, the next step is to establish a command and control centre. To accommodate the need, a communication infrastructure is required to provide decision makers with data and information from the site to receive digital maps, data, and feedback from personnel in the field in a timely manner. Also, it should be able to provide a reliable connection with enough resources for a distributed command and control center.

Emergency planning and response/recovery approaches to a disaster vary from one incident to the next depending on the scale and the nature of each disaster. The degree of urbanization or the geographic spread may require different actions for a specific respond. The degree of urbanization is determined by the number of people affected by the disaster, but the handling of these incidents is different from those that are spread over a wider non-urban area. Wild fires are a good example of a disaster that affects a very wide area such as national parks at a first stage, however if it does not get under control in a timely manner, it may eventually lead to a larger scale disaster and impact more people.

A reliable robust communication technology is necessary to transmit information at all stages of an emergency situation to handle disasters more efficiently. This includes disaster mitigation, preparation, response, and recovery. Emergency response and recovery have a more specific need for quick deployment and easy reconfiguration of a communication infrastructure. These are more time-sensitive applications, while mitigation and preparation usually allow a longer planning time. At a disaster site, there may not be any communication infrastructure available. A mesh network infrastructure can be deployed quickly to provide a network for local communication. If there is any kind of Wide Area Network (WAN) or communication technology available, the local network at a disaster site can communicate to the outside world through this link. It is different from other mesh deployments in cities because of its application for emergency scenarios, portability, flexible infrastructure, and independence from power lines by being battery operated. A wireless mesh infrastructure is quickly deployable with minimal configuration and has multiple interface cards to communicate in a heterogeneous environment with different technologies. In this architecture, only gateways are connected through wireless long haul links, which is considered advantageous, as fewer nodes need to be configured/reconfigured.

The mesh architecture is resilient to the failure of nodes or links as there are alternate paths to take if any one link fails (Refer Figure 7.5). Similarly, a node can communicate through other nodes when a neighboring node fails. This characteristic improves reliability, as unavailability or failure of sub-components of the system does not affect the overall performance of the system and the service will be continuously available. This architecture is robust in the sense that it is able to operate in a heterogeneous environment with a variety of technologies. Additional wireless access nodes can join the network without causing a service interruption by finding the closest node with best signal strength and connecting to expand the existing network. Finally, at a disaster site, if we need to move the nodes at some point in time, reconfiguration is trivial since these wireless access nodes will automatically form a network as

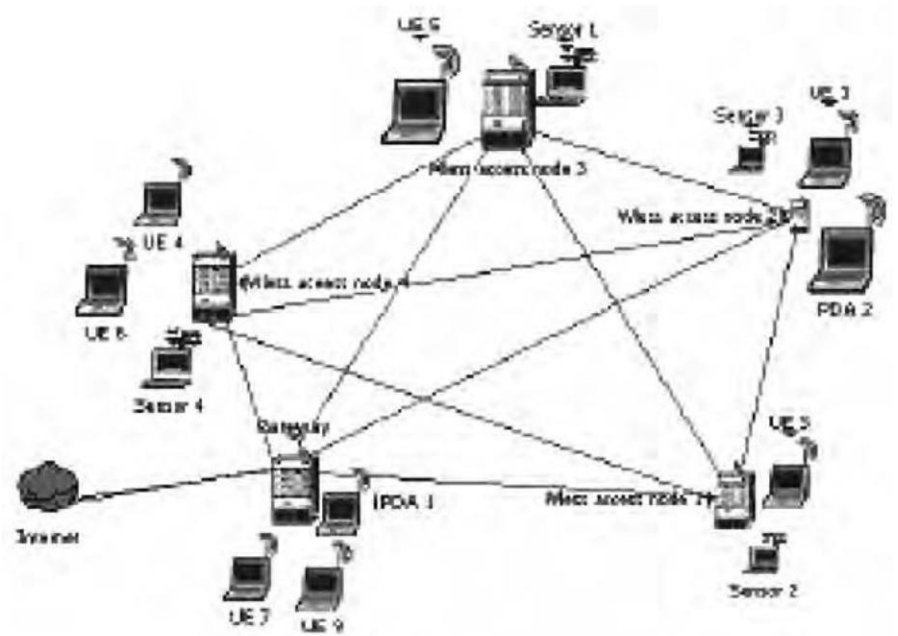


Fig. 7.5 *Mesh network infrastructure*

long as there is a line of sight between nodes. These wireless access nodes allow users to communicate with each other when there is no wired configuration. Figure 7.5 shows the infrastructure of the mesh network deployed at a disaster scene which provides connectivity to the command center and throughout the disaster site.

7.4 SECURITY PROBLEMS AND ATTACKS IN HETEROGENEOUS WIRELESS NETWORKS

Security is a core challenge in the emerging hybrid/heterogeneous wireless architectures. Many advances have already been done in the context of pure ad hoc networks such as those concerning key management, link layer security, secure routing and forwarding, secure end-to-end communications and application layer security. In the context of pure infrastructure networks, security solutions exist and are quite satisfying in 802.11i, GSM, GPRS, UMTS, etc. The main challenge for hybrid wireless networks is to leverage the infrastructure part of the network and adapt some approaches developed for pure ad hoc networks. The major trend when APs/BSs are not able to directly enforce

security in the whole network, is to distribute some of their security privileges to authenticated relay stations which in turn are responsible of enforcing security policies locally. This trend can be seen as a security grid between all the security enforcing entities. Some of open problems in secure integration of heterogeneous networks includes:

- Developing a generic security management protocol that can span the network clouds.
- Developing an efficient resource monitoring and planning mechanism.
- Creating techniques to defend against collusive attacks.
- Secure Integration of hybrid networks that operates at different rates, capacity and nature of transmission.
- DoS and collusive attacks by malicious nodes by using the nature of the network infrastructure.
- Need for a global authentication mechanism that can span over the network clouds.
- Reputation Mechanism to improve quality of transmission.

7.5 SOME SECURITY SOLUTIONS FOR HETEROGENEOUS WIRELESS NETWORKS

The present-generation mobile networks will no doubt consist of heterogeneous access networks, ranging from cellular 3G/4G, WiFi (IEEE 802.11), WiMax (IEEE 802.16) to other emerging access technologies such as mesh and ad-hoc networks. These access networks are deployed to support multimedia services for human-to-human, human-to-device and device-to-device communications in various operating environments where certain service areas may be covered by a multitude of access networks, while others are served by only one of these access technologies. The wide variety of access technologies require drastically different functionalities, capabilities and protocol standards. A key design objective for the future mobile networks is to enhance efficiency and ease of use. In turn, future mobile networks ought to be capable of enabling users communication devices to adapt automatically, dynamically, seamless and efficiently to various access networks for services available at a given time. The key requirement for such adaptation is to ensure end-to-end quality of service (QoS) in terms of data throughput, delay and error rate needed to support users applications despite the heterogeneous nature of the access technologies.

To maintain end-to-end security and privacy of homogeneous, wireless networks is already a big challenge, given the open nature of radio communications. As different access networks have their own security protocols and

controls, maintaining the overall security and privacy in the heterogeneous settings is challenging but necessary. The new security measures have to be distributed in nature so that they can be applicable to the multitude of access networks, possibly owned by many network operators.

7.5.1 End-to-End Security Solution for Wireless Mobile Adhoc NETworks (WMANET)

Wireless Mobile Ad hoc NETwork (WMANET) is a group of an independent wireless (Mobile/ Semi Mobile) nodes communicating on a peer-to-peer basis with no pre-established infrastructure, as shown in Figure 7.6. The unique characteristics of WMANET make such networks highly vulnerable to security attacks when compared with wired networks or infrastructure-based wireless networks.

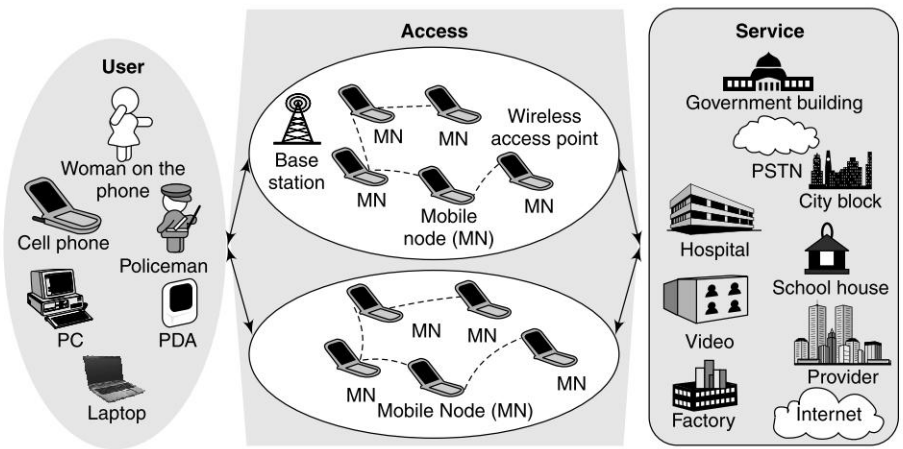


Fig. 7.6 *WMANET architecture*

Five different enterprise objects are used to make up the WMANET. These objects are the following:

User

A person or a device who requests a service. The user could be a mobile user, semi-mobile user, or fixed user.

Service

A set of meaningful capabilities or valuable functions offered to a user by the service provider. The service provider could be a retailer, broker, or third party service provider.

Access

The entity that the user contacts in order to get the service. It has two possible modes: *infrastructure-less wireless ad hoc mode and infrastructure-based wireless mode*. In the infrastructure-less wireless ad hoc mode, the access object consists of a group of an independent wireless mobile (semi-mobile) nodes communicating on a peer-to-peer basis with no pre-established infrastructure needed. The mobile stations could be laptops, Personal Digital Assistant (PDAs), cellphones, etc. In the infrastructure-based wireless mode, the Access Objects could be any infrastructure based access units available in wireless environment such as Base Stations (BSs) in cellular systems or access points in WLAN. Both infrastructure and infrastructure less modes can be integrated in a heterogeneous wireless mode in order to provide the user with a proper access to the service party under any conditions that could prevent the other two access modes from working individually in an effective way.

Core

Manages and controls the behaviour of three objects: *user*, *access*, and *service* in order to guarantee the delivery of the service. The core object is mainly concerned with the provisioning and support of the management functions: Operation, Administration, Maintenance and Provisioning (OAM&P) (ITU-T M.3020 2000, ITU-T M.3400 2000) and Fault, Configuration, Accounting, Performance, Security (FCAPS) (ITU-T M.3010 2000, ITU-T X.700 1992, ITU-T X.701 1997).

Application

Tools installed or used on the user side to facilitate effective communication and service provision. It could be a service-specific application or a generic control and management application. This object manipulates data, video, or voice that is received by the user to meet both the user and service requirements.

Based on the defined WMANET, a comprehensive, top-down, end-to-end security solution for WMANET is proposed. This security solution addresses the global security challenges of WMANET in order to detect, predict, and correct security vulnerabilities. Table 7.1 presents the security solution in a tabular form and illustrates a methodical approach to secure WMANET. Each

of the seven relationships represents a unique perspective for consideration of the seven security requirements. It should be noted that the security requirements applied to the different relationships have different objectives and consequently comprise different sets of security measures.

Table 7.1 *Security Techniques Selection*

Security Requirement	*Security Objective*	*Security Techniques used*
Authorization	Ensure that only authorized persons or devices are allowed to access user data that is transiting network access units (such as Mobile Stations (MSs), Base Stations (BSs) and Access Points (APs)) and their communications links.	Passwords,Access Control List (ACL), Firewall.
Authentication	Verify the identity of the person or device attempting to access user data that is transiting network access units (such as MSs, BSs APs) and their communications link.	Shared Secret, Public Key Infrastructure (PKI), Digital signatures, Digital certificates.
Privacy	Ensure that network access units (such as MSs, BSs APs) do not provide information belonging to the user's network activities (e.g., users geographic location, web sites visited, etc.) to unauthorized persons or devices.	(Partially by) Encryptions, Mechanisms to hide locations and Routing protocols used.
Data Confidentiality	Protect user data that is transiting network access units (such as MSs, BSs APs) and their communications links against unauthorized access or viewing.	Encryption.
Availability	Ensure that access to user data by authorized network access units (such as MSs, BSs APs) cannot be denied. This includes protection against active attacks such as Denial of Service (DoS) attacks as well as protection against the modification or deletion of authentication information (e.g., user identifications and passwords, administrator identifications and passwords).	Intrusion Detection Systems/Intrusion Prevention Systems (IDS/IPS)
Data Integrity	Protect user data that is transiting network access units (such as MSs, BSs APs) and their communications links against unauthorized modification, deletion, creation, and replication.	Hash functions, Digital certificates
Non-Repudiation	Provide a record identifying each individual or device that accessed user data while transiting network access units (such as MSs, BSs APs) and their communications links, and the action that was performed. This record is to be used as proof of access to the user data.	Digital signatures, System logs

7.5.2 Transparent End-User Authentication Across Heterogeneous Wireless Networks

Present wireless networks adopt an architecture that combines different networks in order to obtain wider area of coverage, higher data throughput, and better frequency reuse. While the wide coverage of technologies such as UMTS provides the end-user with always-on capability, technologies such as WLAN can provide hotspot data services as well as voice services at higher rates than UMTS. This will encourage end-users to connect to WLAN whenever possible to access their high bandwidth services.

The small coverage of WLAN makes handovers happen frequently for mobile users. These handovers can be either between WLANs or between WLAN and other access technologies, e.g., UMTS. In order for the end-users to achieve uninterrupted network access to services, fast-handover support is essential. One of the pre-requisites of fast-handover is transparent end-user authentication. The authentication mechanisms that are currently used in UMTS and WLAN are different. These mechanisms are developed separately, and were not designed to inter-operate.

Two authentication solutions for UMTS networks and WLANs are discussed here. One solution is based on the USIM and the other solution provides a way to integrate different authentication mechanisms in UMTS and WLAN. The end-users mobile terminal (MT) has dual or multiple access network interfaces that enable the terminal to access heterogeneous networks. There is a mechanism in the MT to decide which access network to attach to. In both solutions, the MT supports mobile IP (MIP). MIP is required to enable seamless roaming on IP level.

UMTS SIM (USIM) Based Solution

This solution is based on the authentication mechanism of the UMTS network. The solution described here is tailored to an architecture that consists of IEEE 802.11 based WLANs and UMTS networks. Some of the assumptions are as follows:

- The UMTS network and WLANs use the same authentication server (AS) located in the UMTS network. Each AS has digital certificate issued by a Certificate Authority (CA). The CAs can be different for the authentication servers in different UMTS networks.
- The MT is always connected to the UMTS network.
- The MT needs to obtain the public key of the serving AS if it wants to switch to the WLANs affiliated to it. The key can be passed to the MT by the UMTS network during a mutual authentication, either with the home network or with the visiting network.

- The WLANs are located within the range of UMTS networks they are affiliated with, which assures that a MT has the public key to validate the AS of a WLAN.
- The MT supports the PEAP, including USIM extension.

The mutual authentication procedure to a WLAN is as follows (see also Figure 7.7):

- The AS sends a PEAP/Start packet, which is an EAP-Request packet with EAP-Type=PEAP, the Start (S) bit set, and no data. The initial clear text identity exchange is omitted to protect the identification of the end-user from disclosure. The AP acts as a proxy and Network Access Server (NAS) between the MT and the AS.
- The MT sends an EAP response packet with EAP-Type=PEAP. The data filed of the packet contains the information needed to setup a TLS link.
- The AS sends its certificate to the MT while setting up the TLS secure link. The MT validates the certificate using the public key it received from the UMTS network during its association with current UMTS network. If the authentication succeeds, the procedure continues, as the secure link has been set up.
- The MT and AS negotiate and agree to use PEAP-USIM authentication method.

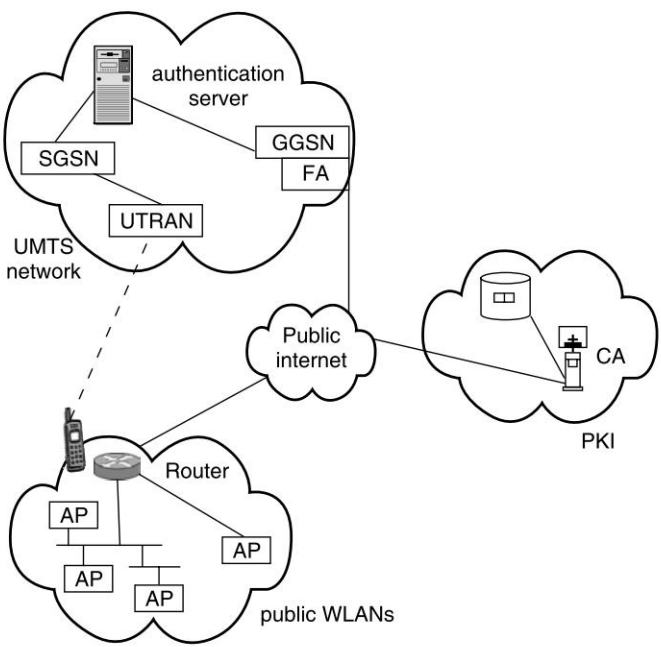


Fig. 7.7 *The PEAP-USIM based solution*

- The AS sends a RAND to the MT. The MT processes the challenge using the C_K in its USIM and sends the result to the AS.
- The AS verifies the MT by looking up the P-TMSI list and finds the C_K in the VLR of the end-user. It then carries out the same calculation as the MT did using the RAND and C_K and compares the results. (According to the assumptions, hotspots deployed by service provider A are within the range of UMTS networks of service provide A, the AS need only to check the data in its local VLR.)
- If the authentication succeeds, the AS sends an authentication success message to AP, which will enable its controlled port for the MTs MAC address and enables a WEP key.
- The MT obtains a new local IP address which will be used in the WLAN. The new local IP address will be registered at home agent.

Solution Using Multiple Authentication Mechanisms

In this solution the UTMS networks and WLANs use different authentication mechanisms. Authentication in UMTS network is based on USIM. Authentication in WLANs is based on digital certificates. The authentication mechanisms for WLAN such as 802.1x and EAP-TLS support mutual authentication between MT and AS. The AS has Internet access during the EAP authentication; it is capable of following a certificate chain to verify a peers identity. However, the MT does not have a connection to the Internet before the authentication succeeds. As the MT stays offline, in order to validate the certificate of an AS, the MT needs to hold public keys of different CAs, as well as the certificate revocation list (CRL) of each CA (Refer Fig. 7.8). This may be difficult for the MT, which has limited storage space. As the number of CA grows and when there is a certificate chain the situation becomes worse. The solution relies on the always-on ability enabled by the UMTS network. It means the UMTS link can be used to access to the Internet and validate the certificate of an AS when necessary. Some of the assumptions follows:

- The UMTS networks and WLANs are managed by different operators. They use different authentication servers.
- The UMTS and WLAN network operators have established roaming agreements. The MTs are always connected to the UMTS network because of its wide coverage.
- Every MT and each WLAN AS have a digital certificate issued by a CA. The CAs that issue the certificates can be different. The digital certificates are used during mutual authentication.
- The WLANs lie within the coverage of a UMTS network.
- MT supports PEAP.

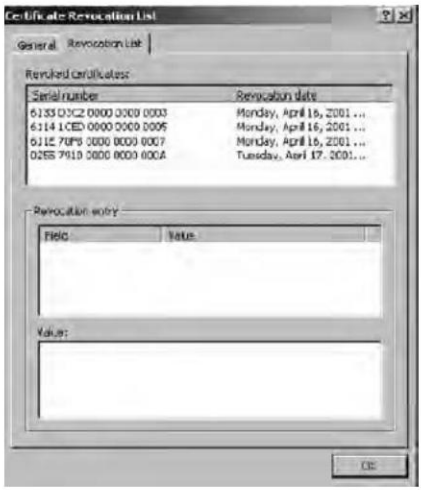


Fig. 7.8 *A sample CRL*

When a MT needs to switch from a UMTS network to a WLAN, the authentication begins and the procedure is described below (see also Figure 7.9):

- The AS sends a PEAP/Start packet, which is an EAP- Request packet with EAP-Type=PEAP, the Start (S) bit set, and no data. The initial clear text identity exchange is omitted to protect the identification of the end-user from disclosure. The AP acts as a proxy and NAS.
- The MT sends an EAP response packet with EAP-Type=PEAP.
- AS sends the certificate to the MT while setting up the TLS secure link. The mobile terminal validates the authentication server. Through the UMTS Internet connection, it is capable of following a certificate chain or verifying whether the certificate has been revoked.
- MT sends an EAP response to set up the TLS link. The MTs digital certificate is not sent in this message.
- After the TLS link has been set up, the AS and MT negotiate to use PEAP-Certificate method. The MT then sends its certificate to the AS. Because AS is on the Internet, it can always connect to the CA to validate the certificate of the MT.
- If the authentication succeeds, the AS sends a authentication success message to AP, which will enable its controlled port for the MTs MAC address.
- The MT obtains a new local IP address that will be used in the WLAN. The new local IP address will be registered with the home agent.

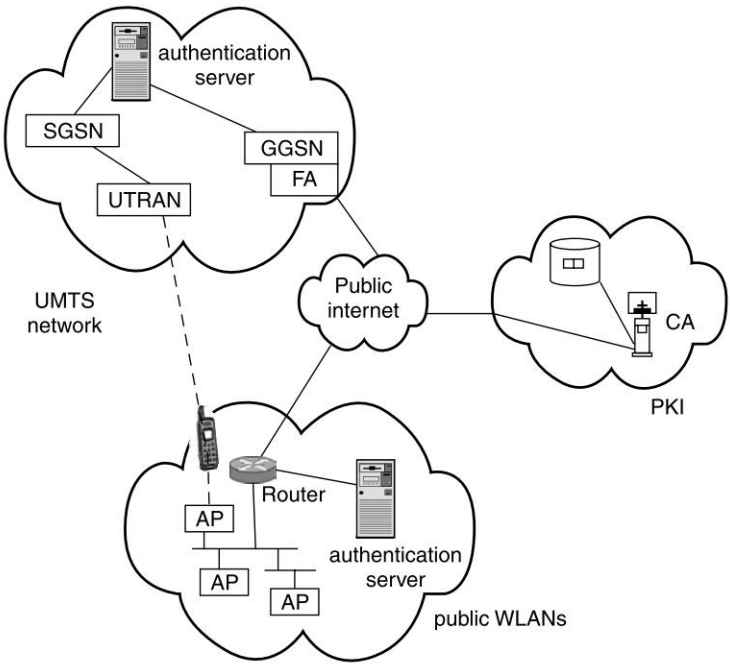


Fig. 7.9 *Solution using multiple authentication mechanisms*

SUMMARY

In this chapter we have discussed the importance of heterogeneous wireless networks in present scenario along with few typical applications. We have covered the security problems, issues, and attacks with respect to these networks. We have also presented some of security and authentication solutions available for heterogeneous wireless networks.

REVIEW QUESTIONS

1. Define heterogeneous wireless networks.
2. Why do we need heterogeneous wireless networks?
3. What are the potential applications of a heterogeneous wireless network?
4. Explain the working of iCAR architecture.
5. Explain the working of SOPRANO architecture.
6. Explain the working of Roofnet architecture.
7. Discuss the use of heterogeneous wireless networks in disaster management.

8. Discuss various security problems faced by heterogeneous wireless networks.
9. Explain the working of rushing attacks.
10. Explain the working of end-to-end security solution for Wireless Mobile Adhoc NETworks (WMANET).
11. Explain the choice of security techniques in WMANET.
12. Explain the need of transparent end-user authentication for heterogeneous wireless networks.
13. Explain the working of UMTS SIM (USIM) based solution for transparent end-user authentication.
14. Explain the authentication solution using multiple authentication mechanisms for transparent end-user authentication.

Security for Mobile Commerce Application

8

OBJECTIVES

- To know about the importance of commerce over mobile devices.
- To study some mobile commerce applications.
- To understand the key drivers and various initiatives taken towards promoting mobile commerce.
- To learn security challenges and attacks over mobile commerce services.
- To learn various security models available for mobile commerce.
- To study specifically mobile payment operations.
- To study few security and authentication schemes available for mobile commerce.

The term 'mobile commerce (m-commerce)' was coined in the late 1990s during the dot-com boom. The idea that highly profitable to m-commerce applications, would be the broadband mobile telephony provided by 2.5G and 3G cellphone services. The m-commerce is an e-commerce brought to mobile users via mobile devices such as palmtops, PDAs or most dominantly mobile phones. With an ever increasing number of devices in the market, mobile phones will undoubtedly play a crucial role in promoting the m-commerce. It allows users to conduct e-commerce on their mobile devices, obtain marketing and sales information, receive ordering information, make a purchase decision, pay for it, obtain the service or product and finally, receive customer support required.

M-commerce follows the user and is available anytime and anywhere. Although mobility is a valuable characteristic to the user in general, it is especially precious for m-commerce because it enables a key factor, which is

missing in other e-commerce forms, namely the ability to adapt to the user demands. In fact, the essence of commerce is to be able to satisfy the demands of the users. It is important not only to be able to offer whatever the user wants but also whenever he/she wants. M-commerce can also be customised such that it fits the preferences of the user in combination with time and location.

Another important aspect of m-commerce is the ability to mix electronic media with other media such as newspaper, TV, radio, natural communication in any of the commerce phases, i.e., presentation, selection, ordering, payment, delivery and customer care. For example, a mobile user can browse on his/her mobile phone and obtain the location of the closest shop. In this case, the presentation and selection are done electronically via the mobile phone while the rest is done in a traditional way via natural communication. In another situation, a user buys groceries and pays through a mobile phone. The presentation, selection, ordering, delivery and customer care phases are carried out in traditional way and only the payment phase is done electronically.

8.1 M-COMMERCE APPLICATIONS

The general m-commerce applications are categorised as transaction management, digital content delivery and telemetry services. The applications can be further subdivided into passive and active m-commerce applications. Active application relates with the applications in which the user has to take an initiative on a wireless device. In contrast, the passive applications themselves get activated towards accomplishing the assigned jobs or facilitate the users to carry forward.

8.1.1 Active Applications

M-commerce transactions point to online shopping customised to mobile phones and PDAs are equipped with the capabilities of browsing, selection, purchase, payment and delivery. An important m-commerce transaction is to initiate and pay for purchases and services in real time. The highest volume of m-commerce transactions using wireless devices for micro-transactions occur when individuals reach for their e-cash-equipped mobile phones or PDAs instead of coins to settle micro transactions, such as subway fees, widespread use of digital cash is becoming a reality.

The second important one is regarding digital content delivery. Wireless devices can retrieve status information, such as weather, transit schedules, flash news, sports scores, ticket availability and market prices, instantly from the providers of information and directory services. Digital products, such as MP3

music, software, high-resolution images and full-motion advertising messages, can be easily downloaded to and used in wireless devices. The better display screens and higher bandwidth will surely trigger the development of innovative video applications. This will help wireless users to access, retrieve, store and display high-resolution video content for a time of entertainment, product demonstration and e-learning.

Another major application of m-commerce is telemetry services, which include the monitoring of space flights, meteorological data transmission, videoconference, the Global Positioning System (GPS), wildlife tracking, camera control robotics, and oceanography. Thus in the near future, wireless phones and appliances can be used by people to contact and communicate with various devices from their homes, offices or any where at any time. For example, delivery drivers will ping intelligent dispensing machines or users can transmit messages to activate remote recording devices or service systems.

8.1.2 Passive Applications

This type of application seems manifold and exciting. Instead of using dedicated cash cards for automatic collection of toll charges, digital cash can be used by integrating cash cards with mobile devices. Mobile users can easily pay and record payment of toll, mass-transit, fast-food, and other transactions. As digital convergence becomes more common place, all kinds of mail, such as e-mail, fax documents and digitised voice mail, can be received passively. Thus it is felt that, in near future, there will be many novel services for mobile users for a fixed fee. Users may be offered some services free of cost for viewing audio or video advertisement delivered to their wireless devices. Any kind of security breach, illegal intrusion, unusual event or unacceptable condition will trigger automatic notification to users, irrespective of location. Airline companies are using this technology to alert frequent air passengers regarding seat availability and upgradation, to notify the changes made in the timings, etc., through wireless devices.

8.1.3 Banking

Mobile banking addresses the fundamental limitation of Internet banking, by reducing the customer requirement to just a mobile device. The main reason that Mobile banking scores over Internet banking is that it enables “Anywhere banking”. One of the ways to categorise the mobile banking services is by the nature of the service, which results in transaction-based and enquiry-based services. Example, a request for a bank statement is an enquiry-based service and a request for fund’s transfer to some other account is a transaction-based

service. Transaction-based services are considered more critical than enquiry based services in terms of additional security requirements across the channel from the mobile phone to the banks data servers. Following are some of the services offered by mobile banking industries, which could force the security risks.

Information Services

Bank advertisements, interest rates, exchange rates and news. Such information is usually available to the public. The information that is provided by the bank could be deliberately corrupted by the attacker, therefore it is important to protect against unauthorised modifications.

Bank Account Information

Bank account enquiry functions may be provided through mobile channels. Bank customers may use their mobile terminals to retrieve bank account balances, account summaries, transaction details and other details of a personal and confidential nature. Banks must take steps to ensure that only authorised parties have access to such information and that the information is conveyed in a manner where confidentiality and integrity is maintained.

Transfers Between Customer Linked Accounts

Customers may link multiple related accounts to their mobile banking facilities which enable them to perform transactions, including fund transfers on all linked accounts. Banks should allow fund transfers between a customer linked accounts only when stringent controls are instituted over the linking of accounts to an on-line mobile service. System checks and validations should be provided in the on-line banking systems to disallow and detect unauthorised or erroneous linking of unrelated customer accounts.

Transfers to Third-Party Accounts

Banks typically provide two types of third-party fund transfer services, one is the transfer to pre-approved third party accounts and second is transfers to third-party accounts without pre-approval. Transfers to pre-approved third-party accounts should be carried out only if these accounts have either been registered by the bank itself or if customers have given their banks specific written instructions and authorisations to do so. Examples of bank pre-approved accounts include government agencies and other reputable billing organisations. Transfers that do not require pre-approval enable bank customers to send their money to unrelated accounts within the same bank or at another bank. Such transfers pose a high level of security risk as they result in the immediate movement of funds from one customer account to another.

8.1.4 Locational Information and Marketing

As a consumer, the option to register for offers related to upcoming events, information about a location that they are visiting, or anywhere, anytime access to information and services will be available. For example, a service could tell a person how to find the nearest petrol station when they are concerned about their fuel, or to compare prices of an item elsewhere in town while they are in a shop. The provision of locational information will not be limited to pull information, but will also include push marketing, where consumers who are visiting or enter a particular location are provided with instant messages and information about specific services that are available in that area, or discounts. This will not necessarily be information that is requested or wanted, but could include information that is unsolicited, and is based purely on technology which will enable commercial companies to identify people who enter a vicinity.

8.1.5 Mobile Payment Systems

M-commerce involves m-payment, which is defined as the process of two parties exchanging financial value using a mobile device in return for goods or services. The various actors which are involved in mobile payment process are *Consumer*—who owns the mobile device and is willing to pay for a service or product through some payment provider; *Content provider or merchant* who sells product to the customer. Another actor in the payment procedure is the *payment service provider*, who is responsible for controlling the flow of transaction between mobile consumers, content providers and Trusted Third Party (TTP) for enabling and routing the payment message initiated from the mobile device.

Payment transaction process in a mobile environment is very similar to typical payment card transaction. The only difference is that the transport of payment detail involves wireless service provider. WAP/HTML based browser protocol might be used or payment details might be transported using technologies such as blue-tooth and infra-red.
Mobile payment life cycle has three main steps:

- **Registration**

Customer opens an account with payment service provider for payment service through a particular payment method.

- **Transaction**

Customer indicates the desire to purchase a content using a mobile device. Content provider forwards the request to the payment service provider.

Payment service provider then requests the trusted third party for authentication and authorisation. Payment service provider informs content provider about the status of the authentication and authorisation. If customer is successfully authenticated and authorised, content provider will deliver the purchased content.

- **Payment Settlement**

Payment settlement can take place during real-time, prepaid or postpaid mode.

A method of Implementing a m-Commerce Payment Procedure

The m-commerce procedure shown in Fig. 8.1 can be described by the following steps:

1. When the mobile user starts a m-commerce application, the MN (Mobile Node) will check for access to the wireless network.
2. If got access, go to step 4 directly; otherwise, requests for access by providing the certificate and signature.
3. After accepting the reply from the AP (Access Point), the MN sends the register request.
4. The MN determines whether it is at home network. If so, HA (Home Agent) returns the reply. Goto step 6. otherwise, the FA (Foreign Agent) forwards the request to the HA appending its certificate information.
5. The HA replies the request through the FA.
6. The mobile user can access the Internet successfully, and contact the merchant.
7. The mobile user and the merchant are certified by the CA (Certification Authority), and used SSL (Secure Socket Layer) or SET (Secure Electronic Transaction) to ensure the security at the higher layer.
8. The mobile user makes the payment through the payment gateway. The transaction is completed.

Some of the Existing Payment Solutions

- **PayPal System**

Electronic mails are used to send and receive money in PayPal system, the user of a PayPal system should provide his/her profile to the system including the credit card information. Once a payment takes place, the money is withdrawn from the credit card account to the automated clearinghouse. The receiver of the payment will be notified. PayPal offers instance notification and

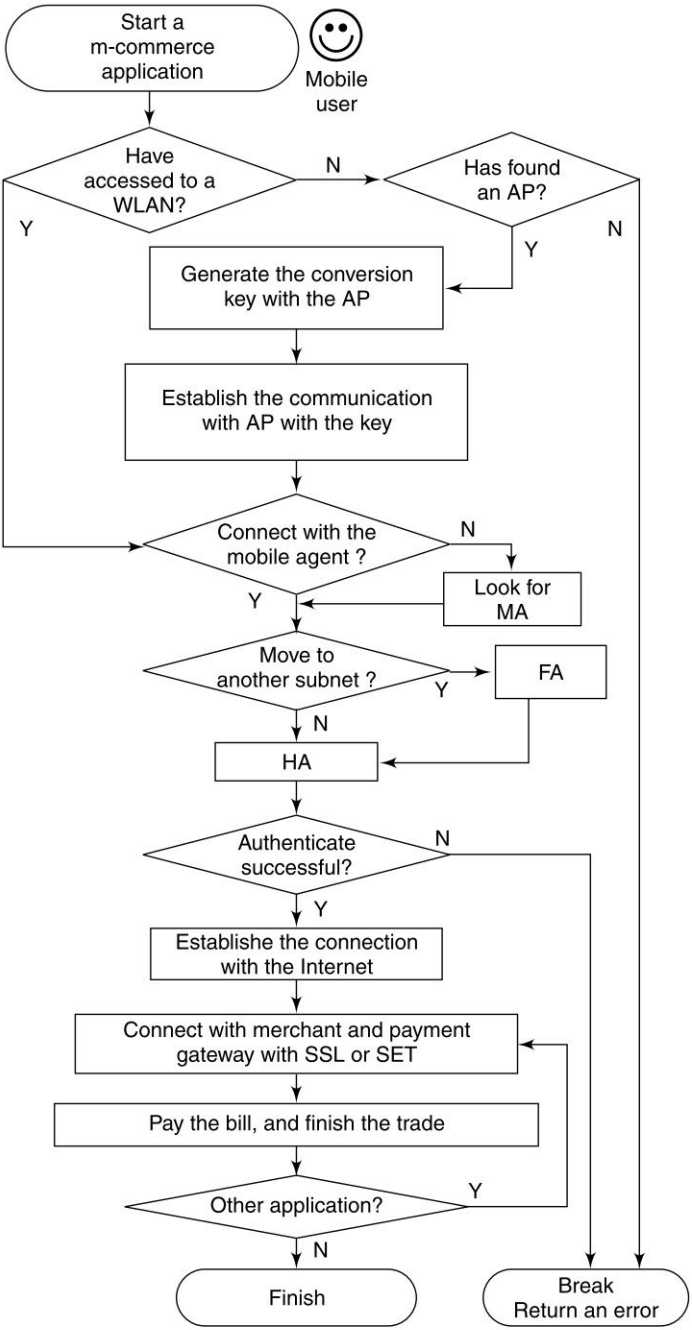


Fig. 8.1 *The m-commerce Procedure*

confirmation of the fund transfer. Thus, PayPal is a third party that organises the money transfer among users.

- **NetPay Model**

NetPay model allows customers to purchase items of micro and macro consumer payments without intervention of a third party (Broker). NetPay has three active parties: (1) broker who creates e-cash and handles micro-payment (2) customer (3) vendor. NetPay is an offline payment model that allows vendors to interact directly with customers via their e-wallet servers.

- **Judge Model**

This model allows the payer to remain anonymous during the transaction, however anonymity is controlled by trusted third party, which is called Judge. This model tries to make a balance between the two major characteristics of the e-payment system, which are anonymity and tractability. It allows a trusted third party to control the anonymity in suspicious payment transaction. The main objective is to prevent criminal use of funds such as money laundering and blackmailing.

- **PayWord Model**

PayWord model uses cryptographic properties of digital signature in e-cash generation with a simple scenario. The payment process has three major players: (1) broker (2) customer and (3) vendor. The customer creates an account at the broker website. The broker issues a digitally signed certificate, which allows the customer to make PayWord chain.

- **PayNow Model**

PayNow model supports micro payment in form of e-check. Cybercash has an e-wallet server that contains special checks for PayNow, which can be used in e-shops. The e-check works in a similar manner to stored-value chip card, where consumer can easily reload Cybercash wallet by using credit card or bank account.

- **Direct Cash**

In direct-cash-like payment systems, the customer withdraws money from the issuer, that is the third party interacting with the customer (for example, a bank or service provider), and hands payment tokens for the payment amount to the merchant. The merchant deposits the payment tokens with its acquirer, that is the third party interacting with the merchant. The issuer and acquirer then settle the payment. This payment scenario is sketched in Fig. 8.2. Since

digital cash is trivial to copy, direct-cash-like payment systems involve either tamper-proof hardware (i.e. smart cards) or online validation by the issuer (i.e. double spending test).

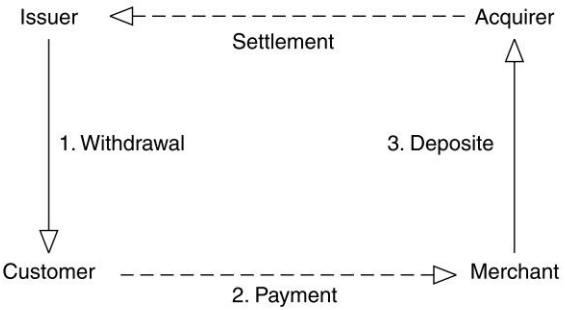


Fig. 8.2 *Direct Cashlike Payment System*

- **Cheque**

In this scenario, the customer hands a cheque (a payment authorisation) to the merchant. The cheque is presented to the acquirer who redeems it from the issuer. Cheque-like payments are sketched in Fig. 8.3 (a).

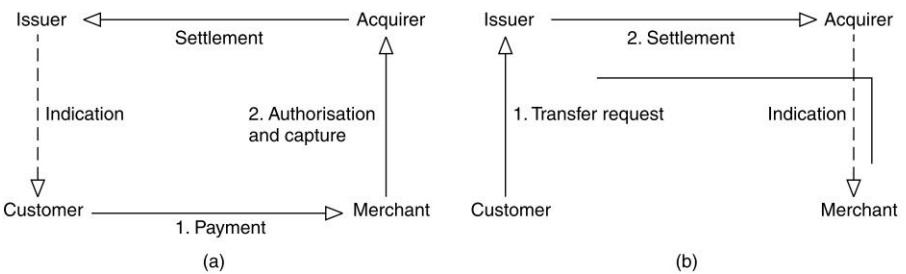


Fig. 8.3 *(a) Cheque like Payment System, (b) Bank Transfer Payment System*

- **Credit Card**

In terms of the information flow, credit card based payment systems are similar to cheque-like payment systems, with the difference that credit-card based payment systems use the existing credit card infrastructure for settling the payment.

- **Bank Transfer**

The bank transfer model is sketched in Fig. 8.3 (b). Here, the customer instructs the issuer to transfer money to the merchants account at the acquirer. The merchant is notified of the incoming payment.

- **Debit Advice**

This model describes the opposite case to the bank transfer model. The merchant instructs the acquirer to charge the account at the issuer. The customer is notified of the outgoing payment.

Categorisation of m-payment Systems

Most e-payment systems are not suitable for use in a mobile context, that is using a mobile device and communicating over a mobile telecommunication network. This is due to the special characteristics of mobile devices and mobile telecommunications. M-payment systems are categorised into following types according to the whereabouts of the customers money:

- **Software Electronic Coins**

In this case, monetary value is stored on the mobile device and the customer has full control of his/her money wherever the person goes and whatever he/she does. An electronic coin is represented as a file containing, among other information, a value, a serial number, a validity period, and the signature of the issuing bank. Since software electronic coins are easy to copy, the validity of an electronic coin depends on its uniqueness in terms of its serial number. The customer transfers electronic coins to the merchant, who forwards them to the issuing bank for the double spending test. In this test, it is checked whether the electronic coin has been spent beforehand. If yes, it is rejected. Otherwise, its serial number is entered into the double spending database and the money is credited to the merchant's account. The generation and storage of electronic coins is an orthogonal problem. Due to the limitations of mobile devices, electronic coins may have to be generated and stored externally, until they are downloaded onto the mobile device.

- **Hardware Electronic Coins**

In this case, monetary value is stored on a secure hardware token, typically a smart card, in the mobile device. The presentation of electronic money is not important, as long as it is stored securely on the smart card. Electronic money could be represented as a simple numeric counter. In order to get to the money, the customer's smart card and the merchant's payment server authenticate each

other and a secure channel is set up between them. Then, electronic money can be transferred from one to the other. This approach is quite attractive because smart cards provide an additional level of mobility. That means that the payment smart card can also be used in point-of-sale transactions.

- **Background Account**

Here, the money is stored remotely on an account at a trusted third party. Depending on the specific payment system, the account could be a credit card account, a bank account, or an account held at the network operator. Common to all scenarios is that, on receipt of an invoice, the customer sends an authentication and authorisation message to the merchant that allows the trusted third party (that holds the account) to identify the customer and to verify the payment authorisation. The accounts can then be settled.

There are numerous payment systems that fall into this category. The differences concern the nature of the trusted third party and the procedure to send authentication and authorisation data. For example, in some cases this data is sent in the clear (e.g., a credit card authorisation), not providing any security against eavesdropping, and in some cases this information is encrypted and digitally signed, providing anonymity to the customer (e.g., SET-Secure Electronic Transactions).

8.1.6 Content and Entertainment

In many countries, the wireless phone companies have announced their intention to launch a new service to allow customers to download games, entertainment and information on their phones. Using a new technology known as Binary Runtime Environment for wireless technology (BREW), consumers will be able to download applications and personalise their mobile devices through the Internet. Customers will also be able to send photos, track expenses and access directories using the service. The billing structure envisaged for the service is that charges would be made during the downloading process and for the product, and after that time users can access the applications at any time without additional charges.

8.2 M-COMMERCE INITIATIVES

A number of initiatives have been announced to encourage the use of mobile technology in financial services and to drive the adoption of open standards in this field. Some of them are given as follows:

The Trusted Mobile Platform Specification

The Trusted Mobile Platform has defined a set of hardware and software components that can be constructed to build devices offering different levels of security. Trusted Mobile Platform builds on well-established, strong security techniques and applies them to the hardware and software architectures to define a trusted execution environment that protects the device at boot time and during runtime.

3D Secure

Visa's 3D-Secure is Visa International's global specification that ensures the security of Internet payments made over mobile phones. Developed in conjunction with some 15 major industry players, the specification is part of Visa Authenticated Payment, a comprehensive e-commerce program designed to ensure safe and secure online payment transactions. The Mobile 3-D Secure specification extends payment authentication initiatives into mobile commerce, enabling Visa card issuers to validate the identity of their cardholders in real time. It ensures that payment data sent over open networks is not compromised and allows consumers to actively protect their Visa accounts from unauthorised use when shopping on-line over mobile devices.

Mobey

The Mobey Forum brings together the substantial expertise of the world's leading on-line financial institutions and the leading companies in mobile Internet technologies such as Wireless Application Protocol (WAP). The leading mobile phone manufacturers, Ericsson, Motorola and Nokia, acknowledge that the Mobey Forum will play a valuable part in the development of online wireless financial services.

Fundamo

There is also Fundamo, an initiative, which has developed the capability to deliver a working mobile payment solution to cellular networks using existing Phase 2+ compliant technology. This means the Fundamo solution can be used on both WAP handsets as well as Phase 2+ compliant handsets.

Radicchio

The Radicchio, a global initiative to define a standard security platform for mobile e-commerce using Wireless PKI (Public Key Infrastructure). Mobile commerce on WAP-phones is likely to be secured by Java and Wireless Identity Modules (WIMs) using PKI.

Sat Forum

The SAT (SIM Application Toolkit) initiative which is a part of the ETSI/SMG standard for Value Added Services and m-commerce using GSM phones to do the transactions. You will be able to check your bank account and pay bills using your SIM Toolkit-enabled phone with an appropriate SIM Toolkit-specific SIM card which will provide much of the intelligence to conclude a transaction over GSM. Wireless Internet Gateway (WIG) gives WAP and SIM Application Toolkit (SAT) terminals access to WML-based applications. It brings WAP to legacy terminals via SMS and supports end-to-end security, push and location-based services. A SIM-based WML browser not only lets a GSM operator deliver web-style content to the current large installed base of mobile phone subscribers, but it offers increased security inherent in the Smart Card technology.

eSIGN

eSign is an initiative between leading companies in the mobile marketplace to make mobile digital signatures. The consortium aims to develop a uniform application interface as the de-facto standard for the integration of the mobile phone into the Internet world and to use the mobile phone for implementing mobile digital signatures.

MEST

The MEST (Mobile Electronic Signature) Consortium is an association of companies active in the internet and mobile phone sectors. Its objective is to develop a secure and universal application infrastructure capable of employing mobile digital signatures.

8.3 SECURITY CHALLENGES IN MOBILE E-COMMERCE

As mentioned earlier, m-commerce is not possible without a secure environment, especially for those transactions involving monetary value. Depending on the point of views of the different participants in an m-commerce scenario, there are different security challenges. These security challenges relate to the following.

- The confidential user data on the mobile device as well as the device itself should be protected from unauthorised use. The security mechanisms employed here include user authentication (e.g., PIN or password

authentication), secure storage of confidential data (e.g., SIM card in mobile phones) and security of the operating system.

- The radio interface access to a telecommunication network requires the protection of transmitted data in terms of confidentiality, integrity, and authenticity. In particular, the users personal data should be protected from eavesdropping.
- The network operator infrastructure security mechanisms for the end user often terminate in the access network. This raises questions regarding the security of the users data within and beyond the access network. Moreover, the user receives certain services for which he/she has to pay. This often involves the network operator and he/she will want to be assured about correct charging and billing.
- The kind of m-commerce application, especially those involving payment, need to be secured to assure customers, merchants, and network operators. For example, in a payment scenario both sides will want to authenticate each other before committing to a payment. Also, the customer will want assurance about the delivery of goods or services. In addition to the authenticity, confidentiality and integrity of sent payment information, non-repudiation is important.

8.4 TYPES OF ATTACKS ON MOBILE E-COMMERCE

An attacker might want to gain access to an electronic message for numerous reasons. Gaining unauthorised access to information in order to violate someone's privacy, impersonating another user in order to shift the responsibility or originate a fraudulent activity are some of the reasons an attacker might want to access the information. In general, there is a flow of information from a source to a destination.

There are four general categories of attacks on a transmitted message.

- Interruption
- Interception
- Modification
- Fabrication

8.4.1 Interruption

Interruption (Refer Fig. 8.4) is the action of preventing a message from reaching its intended recipient. It can also occur when an asset of the system is destroyed or becomes unavailable or unusable. This is an attack on availability.

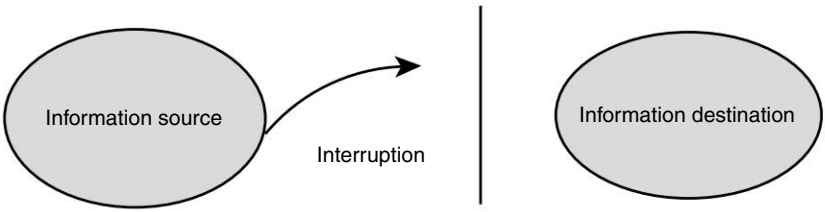


Fig. 8.4 *Interruption of Massage*

Example: Destruction of a piece of hardware: such as the intentional cutting of a communication line; Disabling of the file management system; Denial of service attack; and so on.

8.4.2 Interception

Interception (Refer Fig. 8.5) is where an unauthorised party gains access to information. This is an attack on confidentiality. The unauthorised party might be a person, program or a computing system. A loss due to this kind of attack might be noticed quickly, but a silent interceptor might leave no traces by which the interception can be detected. Examples of these kinds of attacks include: Wiretapping to capture data in a network; Illicit copying of files or programs; and so on.

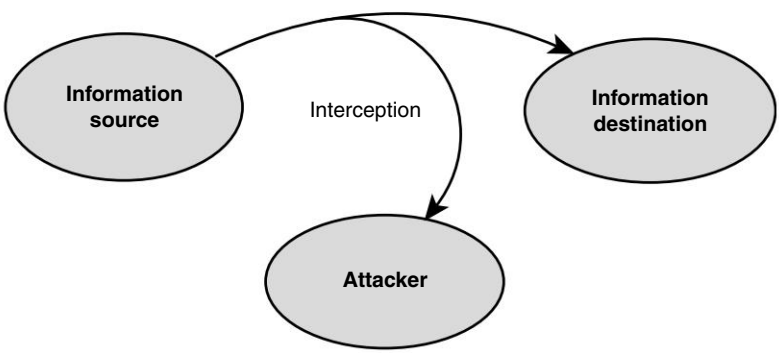


Fig. 8.5 *Interception of Message*

8.4.3 Modification

Modification is where an unauthorised party not only gains access to an asset, but tampers with it (Refer Fig. 8.6). This is an attack on the integrity of the message. Examples include: Changing of values in a database for personal

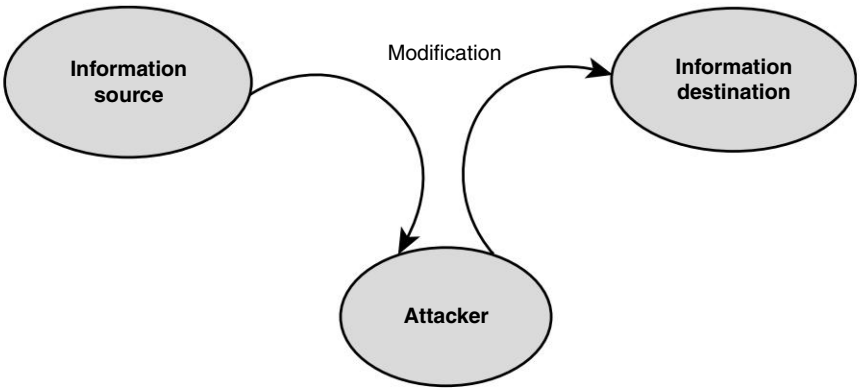


Fig. 8.6 *Modification of Message*

gain; Altering a program so that it performs an additional computation; Modifying the content of a message transmitted on a network; and so on.

8.4.4 Fabrication

Fabrication (Refer Fig. 8.7) occurs when an unauthorised party inserts counterfeit objects into the computing system. This is an attack on the authenticity of the message. These insertions can sometimes be detected as forgeries, but if skilfully done, they are virtually indistinguishable from the real thing. Examples include: Insertion of spurious information into the network communication system; Adding additional records to an existing file or database; and so on.

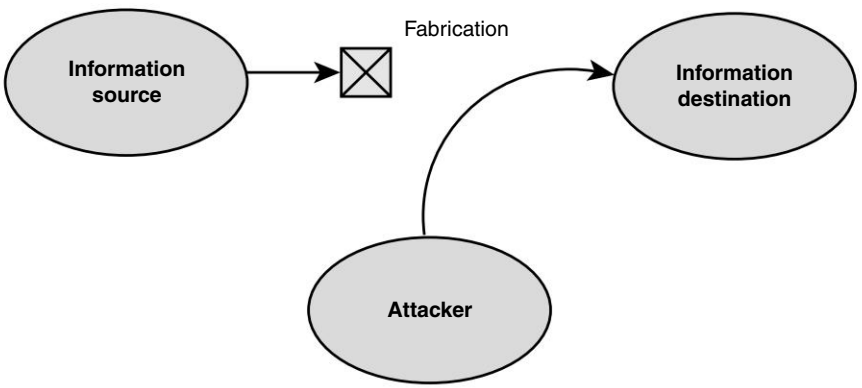


Fig. 8.7 *Fabrication of Message*

8.5 A SECURE M-COMMERCE MODEL BASED ON WIRELESS LOCAL AREA NETWORK

The model is investigated through the accessing procedure, the roaming management procedure, and the electronic trade procedure. To solve the security problem in the WLAN, a novel authentication method is used, in which the mobile node (MN) is validated twice by an access point (AP) and a mobile agent (MA), and all the devices are authenticated in a register procedure with the PKI/CA mechanism.

A model of m-commerce based on WLAN is shown in Fig. 8.8, in which a mobile user uses a MN to conduct e-commerce with a merchant connecting to the Internet. The user accesses the Internet through WLAN, and moves from subnet 1 to subnet 2, and is still able to finish the m-commerce transaction regardless of mobility. The most important device in WLAN is the access point (AP) through which the wireless stations can access into the Internet. The mobile devices (PDA, notepad, handheld, called as mobile node) with 802.11 adapter can contact other devices, for example, static hosts (SH) or other mobile hosts. If an MN moves into another WLAN, the MN can connect to the new AP to realise the re-association after the authentication.

A concept of mobile agent (MA) is introduced to mobile IP system, which can be divided into home agent (HA) and foreign agent (FA). When moving

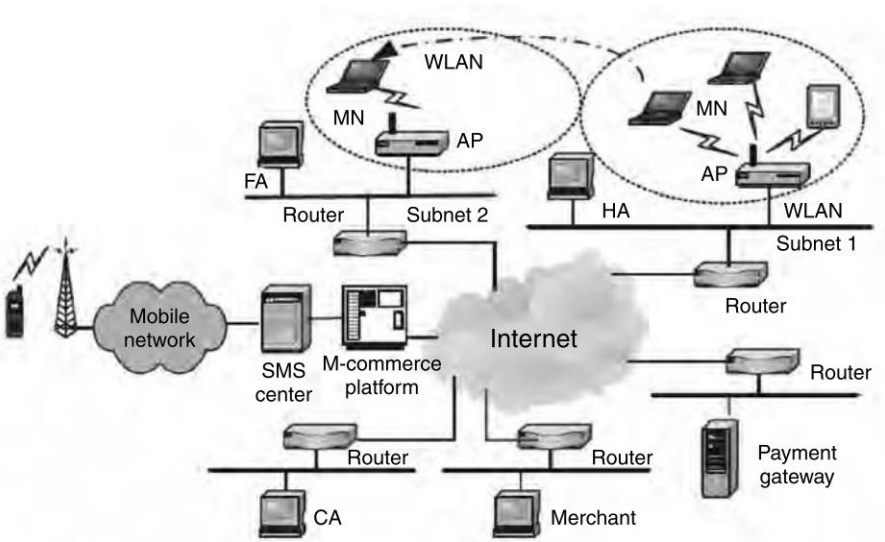


Fig. 8.8 *An m-commerce Model based on WLAN*

to another subnet, the MN selects the IP address of FA as its care-of address. The HA can capture the packet sending to the MN and forward the packet to the FA through a tunnel, who can transfer the packet to the MN ultimately.

The payment gateway accepts the bills, validates the accepted bills, contacts the bank, transfers the funds, manages and records the bills, applies for the certification, and manages the secret key. It is an important piece in the e-commerce system architecture. The m-commerce platform accepts the m-commerce information from all kind of mobile networks and connects the other mobile networks to the Internet. Figure 8.8 shows the m-commerce platform with the SMS (short message system) center. Although WLAN is widely used, there are many problems, such as the problems of quality of service and security. To solve the WLAN security problem, an authentication mechanism based on WLAN and mobile IP is proposed to authenticate MN, AP, FA, and HA using PKI/CA. The certification authority (CA) not only distributes and manages the certificates of the mobile users and the merchants, but also does it for the MN, the AP, and the MA.

8.6 SOME OF M-COMMERCE SECURITY SOLUTIONS

The security schemes for the m-commerce environment can be viewed from regular key-based schemes to application-level schemes. This section provides some of the security schemes available for m-commerce applications.

8.6.1 PKI/CA and ECC

The certificate authority (CA) is crucial in the PKI, which is the third authority that generates, issues, and manages the digital certificates. With the hierarchy structure, the superior CA assigns the certificates for the junior ones, and the bottom CA faced the end user directly. A digital certificate (or digital ID), is an electronic file granted and validated by the CA, including the public key of the certificate holder and the other related information.

Because the PKI may require extensive computations such as the PKI based on the RSA algorithm, the algorithms with fewer computations are needed for the wireless environment. An example of such an algorithm is Elliptic Curve Cryptography (ECC) system. ECC offers a higher encryption rate, a low energy consume, and bandwidth saving. Following steps illustrates the working of the security scheme.

The Procedure of the Access and the Registration

The e-commerce transactions based on the model shown in Fig. 8.9 are completed through three steps: (1) accessing to WLAN; (2) supporting mobile IP, and (3) applying m-commerce, which can be described as follows:

Security Mechanism for WLAN Accessing

When a MN wants to access the Internet, it first finds an AP to connect the wired network. MN learns which AP should be selected through AP periodical broadcasting. After receiving the AP broadcasting message, MN sends an access request packet (Req1) in which a certification and a signature with a time stamp are appended. The AP contacts with CA to check the validity of MN. After the successful authentication, a reply (Rep1) is returned to MN, with the success confirmation and the public key of the AP. The procedure is shown in Fig. 8.9.

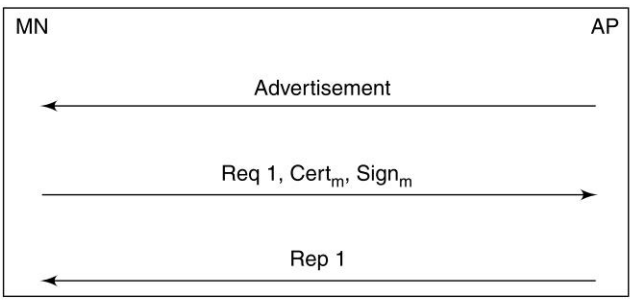


Fig. 8.9 *WLAN Access Control Procedure*

Then, the MN and the AP generate their conversation key. The ECDH (Diffie-Hellman key exchange algorithm of ECC) is used for generating the conversation key. The conversation key of MN(K_{MN}) is generated by the AP's public key and the MNs private key, while the conversation key of AP(K_{CAP}) is generated by the MN's public key and the AP's private key. This leads to $K = K_{AP} = K_{MN}$. The key is used in the entire session including the registration to MA and the m-commerce transactions.

8.6.2 Authentication in Mobile IP

The attacks to the mobile IP includes refuse service attack, resend attack, dialog filch attack, and initiative attack. Many works have been done to prevent the invader from attacking the network. The focus is on the prevention from the

refuse service attack generated by the forge AP or FA, where attacker may take its IP address as care-of address of a certain MN to intercept the message sent to the MN, while the genuine MN is refused. In addition, to ensure the MN is not an invader, an alternative authentication method with shared key is also used.

Before transferring the data, a register procedure must be initiated by MN through sending a register request. For example, the keyed-MD5 authentication algorithm "prefix+suffix" mode to compute a 128-bit "message digest" of the registration message is used. The message digest is combined with the PKI/CA to get a more reliable transmission shown in Fig. 8.10. At the same time, the authentication of AP is added.

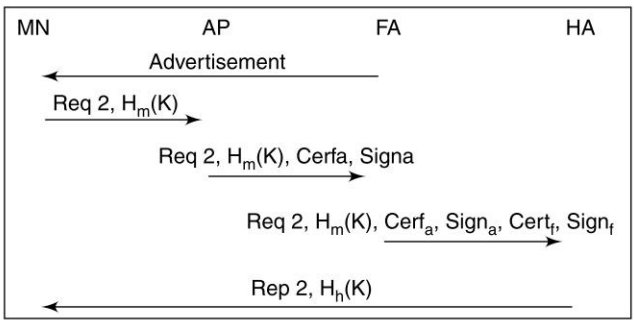


Fig. 8.10 *Security Mechanism of Mobile IP*

The notations in Fig. 8.10 are follows:

- Req2 is the register request from MN;
- K is the shared key between MN and HA, which is distributed by HA and saved in the database in HA;
- $H_m(K)$: the authentication extension to the register request of MN generated by 128-bit "message digest" with K;
- $H_h(K)$: the authentication extension to the register reply of HA generated by 128-bit "message digest" with K;
- $Cert_a$, $Sign_a$: the public key certification and digital signature of AP;
- $Cert_f$, $Sign_f$: the public key certification and digital signature of FA.

When a MN receives an advertisement from FA, which can indicate that the MN is located at the visited network, it sends a register request with the authentication extension using the shared key between the MN and the HA. The AP and FA forward the packet and append the digital certification with nonce and signature on it. The HA uses the digital signature signed in the APs and FAs certification to validate the AP and FA, and the shared key with

the MN to validate the MAC to prove the identification of the MN. The HA sends a register reply to the MN after the successful authentication. The register reply is forwarded by the FA and AP with authentication extension too, which can be used by the MN to authenticate it. To ensure the AP is valid when the MN moves out of the current AP covered area within a subnet, the MN initiates a new register procedure.

8.6.3 An Asymmetric Authentication Protocol for M-Commerce Applications

A asymmetric end-to-end authentication protocol based on the concept of using the wireless access Home Network of a Mobile Station to assist its authentication with a Service Provider is developed. The protocol uses an asymmetrical approach, i.e., computationally expensive cryptographic operations have been used, which are shifted away from mobile devices into their home networks and wired service providers while at the same time the level of security is not compromised. A verifiable authenticator is introduced in the protocol design so that the recipient of the authenticator can verify its correctness without revealing any sensitive datum outward. The following subsections provide a detailed description of the proposed authentication protocol.

Preliminaries

Ideally, the adequate authentication service for mobile users to access Internet m-commerce services should satisfy the following security requirements:

- **S1**

The authentication service should enable mutual authentication between end entities, i.e., the mobile user and the service provider.

- **S2**

The authentication service should enable secure session key establishment between the mobile user and the service provider.

- **S3**

The authentication service should ensure that the assisted party is accountable for all the messages it sends in the authentication process.

In addition to these security requirements, the service should also meet the following performance requirements:

- **I1**

The authentication service should mitigate the computational overhead on the mobile stations. Due to considerations about cost and computational delay, computationally expensive operations, such as public-key operations, should be kept minimal.

- **I2**

The overhead introduced by the authentication service, e.g. the number and size of authentication messages, should be minimal, especially for the wireless segment.

- **I3**

The authentication solution should impose minimum change to the protocols and procedures already in use in both wired and wireless networks so as to facilitate seamless service integration between two dissimilar networks.

Network Infrastructure

The wireless/wired integrated network infrastructure is illustrated in Fig. 8.11. The infrastructure consists of a number of functional entities: a Mobile Station (MS), a Home (wireless access) Network (HN), a Home Location Register (HLR) and a Service Provider (SP).

- An MS is a user equipment consisting of a mobile terminal and a User Service Identity Module (USIM). One of the most important parameters stored in USIM is the International Mobile User Identity (IMUI), which uniquely identifies the MS.
- An HN has overall responsibility for the provision of a set of services to users. It has a database that contains all the subscription data including all the security parameters such as the IMUI. The HLR is the operational entity of the HN.

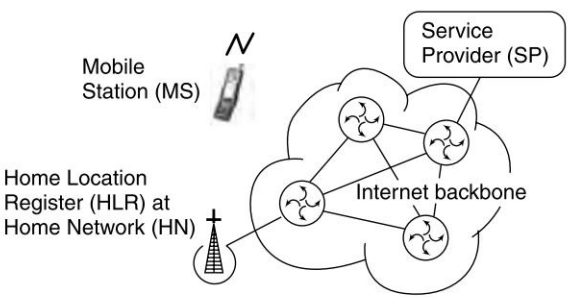


Fig. 8.11 *End-to-end Authentication Network Infrastructure*

- An SP provides a suite of services, tools and systems related to e-commerce/m-commerce. An example of an SP is an e-bank or an e-shop.

Notations

The notations used for the protocol presentation are summarised as follows:
A, B: Concatenation of data items A and B
$h(x)$: A one-way hash function with the following properties: (a) for any x, it is easy to compute $h(x)$; (b) given x, it is hard to find x ($\neq x$) such that $h(x) = h(x)$; and (c) given $h(x)$, it is hard to compute x. An example of such a one-way hash function is SHA-1.
$E_K(x)$: The cipher-text of a data item x encrypted with a key K. $E_K()$ is computed using a public-key cryptosystem if the corresponding decryption key is not K, and using a conventional cryptosystem otherwise.

Design Assumptions

The following assumptions have been used in the protocol design. The HN has a supplier-subscriber relationship with the MS, i.e., the MS is a subscriber of the HN. The HN provides the MS network access service. The responsibilities of the HN include the management of subscriber data and on-line interaction to ensure that users are properly authenticated to use the provided services. The HN keeps subscriber information confidential to itself. This supplier-subscriber relationship between the HN and MS is currently have with most of the wireless providers. It is thus sensible to let the HN to assist the MS in its authentication with the SP and to ensure that this assistance is provided with accountability assurance. In addition, as an MS and its HN appear together as one 'virtual' client in its mutual authentication with the SP, the SP needs to authenticate both the MS and the HN. It is also assumed that there is no collusion between any two of the three entities, i.e., an MS, the HN of the MS, and an SP, to cheat the third entity.

The Protocol Detail

The end-to-end authentication process between an MS and an SP can be fulfilled using three authentication processes—MS-HN authentication, HN-SP authentication and MS-SP authentication, which are depicted in the Fig. 8.12.

- **MS-HN Authentication Protocol**

Mutual authentication between an MS and its HN has already been defined in the wireless network standards, such as the UMTS standards, and this defined authentication standard as a primitive in this protocol design. The authentication process allows the HN to authenticate an MS and vice versa when the MS

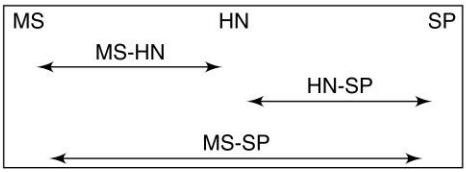


Fig. 8.12 *MS-HN, HN-SP and MS-SP Authentication Processes*

connects to the HN. After the MS-HN authentication process, the MS and HN will share a secret session key $K_{MS\text{-}HN}$ and a secret temporary identity TMUI which is used as the identity of the user during this communication session.

- **HN-SP Authentication Protocol**

As an HN and an SP are typically connected to wired networks, they usually have sufficient memory, computational power and bandwidth, and mutual authentication between them can be done by using many existing authentication protocols such as the TLS and Kerberos. These protocols enable the HN and the SP to mutually authenticate each other and establish a shared secret key $K_{HN\text{-}SP}$ that is then used as the session key for subsequent secure communication between them. This key will be used in the MS-SP authentication process, i.e. messages described in the section will be transmitted securely via the secure channel between the HN and the SP.

- **MS-SP Authentication Protocol**

This subsection presents the MS-SP protocol supporting mutual authentication between an MS and an SP with the assistance of the HN. It is assumed here that the two end entities are connected to a single wireless access domain. The authentication process is described as follows:

1. An MS initiates the authentication process by sending a request to the HLR at the HN. This request includes its identity id_{MS}, the SPs identity id_{SP} and an encrypted value E_{puSP} (x). Here, x is a random number selected by the MS and puSP is the public key of the SP. (id_{MS}/id_{SP} are public names of the MS and the SP, e.g. id_{SP} may be the URL address of the SP.) The message confidentiality is provided using the session key $K_{MS\text{-}HN}$ as only the HLR can decrypt this message to obtain the contents. The integrity and authenticity services are provided using a keyed hash function as only the key owners, the MS and the HLR, can produce this value correctly. These security services are also applied to transactions 4, 5 and 6 but different session keys are used. However, if confidentiality, integrity or authenticity verification fails, the HN requests the MS

to re-send the message. For repeated failures, the HN can terminate the protocol run. This is applied to all other transactions.

2. The HLR forwards $E_{puSP}(x)$ to the SP together with a hash value $h(TMUI)$ and its signature sig_{HN}. Here $h(TMUI)$ is generated by the HLR using the MSs TMUI stored in the HN and will be used to authenticate the MS to the SP. As TMUI is the identity of the MS, to protect its privacy, the HLR hashes it before sending it to the SP. The message confidentiality is provided using the session key $K_{HN\text{-}SP}$. The integrity and authenticity services are provided using the HLRs digital signature sig_{HN} (= E_{pvHN} ($h(h(TMUI))$, $E_{puSP}\}(x)$)), where pvHN is the private key of the HN. With this signature, the HN cannot later deny that it has not sent this message, and therefore is accountable for its action. These security techniques are also applied to transaction 3.
3. The SP replies to the MSs request with a random number y and its signature sig_{SP} (= $EpvSP(h(y))$), where pvSP is the private key of the SP. It then computes the secret session key $K_{MS\text{–}SP}$ (= $h(x, y)$).
4. The HLR forwards y to the MS with a hashed value $h(K_{HN\text{–}SP})$. $h(K_{HN\text{–}SP})$ will be used to authenticate the SP to the MS.
5. The MS computes the session key $K_{MS\text{–}SP}$ (= $h(x, y)$) using its secret x and the value y received, and sends a verifiable authenticator $h(h(TMUI), K_{MS\text{–}SP})$ to the SP. This verifiable authentication hash value is to authenticate the MS to the SP. As this authenticator is generated using a hash function, the HLR is unable to acquire the session key $K_{MS\text{–}SP}$.
6. Upon the receipt of $h(h(TMUI), K_{MS\text{–}SP})$ from the MS, the SP computes hash value $h(h(TMUI), K_{MS\text{–}SP})$, where the value $h(TMUI)$ is received in transaction 2 and $K_{MS\text{-}SP}$ is computed by the SP. The SP then compares this calculated value with the one received. If they are equal, the SP is assured that the MS is authenticated. Then the SP produces an authenticator h ($h(K_{HN\text{-}SP})$, $K_{MS\text{–}SP}$) and sends it to the MS. As in transaction 5, the HLR cannot access to the session key $K_{MS\text{-}SP}$ because it is hashed.
7. Once $h(h(K_{HN\text{–}SP}), K_{MS\text{–}SP})$ is received, the MS uses $h(K_{HN\text{-}SP})$ received in transaction 4 and the key $K_{MS\text{-}SP}$ to compute value $h(h(K_{HN\text{-}SP}), K_{MS\text{–}SP})$, and compares it with the one received. If they are equal, the MS is assured that the SP is authenticated and the authentication process is successfully completed.

8.6.4 A Personal Authentication Scheme Using Mobile Technology

A mobile architecture for strong personal authentication using mobile devices (cellular phones or PDAs) and wireless personal area networks (WPAN) like

Bluetooth is developed. The mobile device stores private information and performs cryptographic operations while WPAN connects mobile devices with access points. The risk of insecure wireless networks inclusion is studied and solved using suitable authentication protocols.

MASPA—Mobile Architecture for Strong Personal Authentication

MASPA is based on different entities, algorithms and protocols. The main entities are the user, the mobile device owned by the user and the access point.

The protocols used in MASPA architecture are basically the communication protocol and the authentication protocol. The former is the protocol used by the mobile device and the access point to communicate. The latter deals with the authentication process itself. Since security is a major concern in MASPA, the main algorithms used in the architecture are cryptographic primitives like hash functions, symmetric and asymmetric encryption including digital signatures.

The general authentication process is as follows. The user to be authenticated by an access point, must carry his personal mobile device. Once the physical distance between the mobile device and the access point is appropriate, the user initiates the authentication process. The authentication process starts with an authorisation procedure between the user and his mobile device when the mobile device checks the user identity. Then, the mobile device and the access point exchange different information using the authentication protocol. The information exchanged is validated using suitable cryptographic algorithms and the authentication process outcome is either the acceptance of the user identity for the access point or the termination without the acceptance. In fact, the authentication process also generates a symmetric key that can be used to obtain privacy in later communications. Information between the mobile device and the access point is exchanged through a WPAN. The specific technology used as a WPAN determines the physical distance within the user can initiate an authentication process. A particular example of MASPA implementation and application could be a user owning a PDA with a Bluetooth module who wants to get access to the intranet office through a desktop computer.

The Mobile Device

The mobile device is the entity that acts on behalf of the user. The role of the mobile device is the same of a smart card in a traditional authentication scheme. It stores and computes the cryptographic values needed in the authentication process. The basic properties of the mobile device are the following:

1. The mobile device must provide a limited amount of non-volatile memory.

2. The computational power of the mobile device must allow to execute cryptographic algorithms such as symmetric and public encryption in a reasonable time.
3. The mobile device must be able to transfer data through a WPAN such like Bluetooth.
4. The mobile device should be highly portable in the sense that it has to be carried by the user.

First and second properties allow to store secret information into the device and provide a secure and trusted environment where the user can execute his authentication protocol part. In this way, secret information needed in the authentication process, like shared or secret keys, never leaves the mobile device. The inclusion of a Bluetooth module in different electronic devices (mobile phones, PDAs, notebooks, etc.) starts to be in everyday use. The third property cited above ensures that such communication technology is available in the mobile device so data transfer between the mobile device and the access point is guaranteed.

Mobile device security is partly guaranteed by user physical control over his mobile device. Such physical control will become easier if the mobile device is portable as it is suggested in the fourth property. For instance, PDAs suit better this requirement than notebooks. Examples of mobile device could be a cellular phone or a PDA. Both devices already include a Bluetooth module. However, PDAs are more powerful than cellular phones since they have greater computer power, memory capacities together with the availability of different operating systems and programming languages.

The Access Point

The access point represents the entity that verifies the user identity. This is a basic scheme where the whole verification process is performed into and by the access point. Nevertheless, a more decentralised proposal can be obtained by performing the verification process outside the access point, for instance, in an authentication server.

The access point should provide the following features:

1. The access point must be able to store information exchanged during the authentication protocol.
2. Cryptographic computation power of the access point is also assumed in order to execute the authentication protocol.
3. The access point must be provided with the same WPAN technology used by the mobile device in order to transfer data between both entities.
4. For ad-hoc authentication applications, the access point must be able to store a user profile that fixes the relation between users and their privileges.

Notice that properties are similar to the mobile device ones since the interaction during the authentication process must be ensured. However, portability of the access point is not needed. The last property stated above anticipate that authentication process can be used to access different services and resources. Examples of access point can be very different. For instance, a desktop computer can be an access point to an intranet network. On the other hand, a electronic lock can also be regarded as an access point to control room physical access. As soon as WPAN technology will be extended to different devices, authentication using MASPA will increase its possibilities thanks to the wide range of possible access points (vending machines, phone boxes, etc.).

The Communication Model

Bluetooth is low cost, low-power, short-range wireless technology designed as a replacement for cables and other short-range technology like IrDA. MASPA can be implemented using BlueZ. BlueZ is the official Linux Bluetooth protocol stack.

The Authorisation Procedure

The authorisation procedure is performed by the user and his mobile device at the beginning of the authentication process. The authorisation procedure goal is to ensure that the user having the mobile device is its real owner. MASPA uses password-based procedure. The user enters a passphrase into his mobile device and the mobile device verifies the correction of such information. Although this procedure seems too weak for a strong authentication scheme, notice that smart cards use the same procedure with a four digit PIN.

The Authentication Protocol

The authentication protocol that has been chosen for architecture is a mutual entity authentication protocol. The protocol is a three-pass Diffie-Hellman variant that establishes a shared session key between the mobile device and the access point. Such authentication protocol is one of the candidates described in for mutually authenticate a mobile user and the network in upcoming third-generation mobile systems such as Universal Mobile Telecommunications Service.

The authentication protocol as described in is depicted in Fig. 8.13. This protocol assumes that both the mobile device (M) and the access point (A) have a public key pair (PK,SK) related to a public key cryptosystem and a certificate of the public key Cert. Subscripts are used to identify the owner of each element, so PK_m stands for the public key of the mobile device while r_a is a random number generated by the access point.

$$
\begin{aligned}
M \rightarrow A &: g^{r_m} && (1)\\
A &: K = (g^{r_m})^{r_a}\\
M \leftarrow A &: g^{r_a}, E_K\{Sig_a(g^{r_a}, g^{r_m})\}, Cert_a && (2)\\
M &: K = (g^{r_a})^{r_m}\\
M \rightarrow A &: E_K\{Sig_m(g^{r_a}, g^{r_m})\}, Cert_m && (3)
\end{aligned}
$$

Fig. 8.13 *The Authentication Protocol*

1. The mobile device takes g a generator of a multiplicative group in which discrete logarithms are hard to compute. Then it generates a random value r_m and then g^{r_m} is send to the access point. The access point then computes a symmetric key $K = (g^{r_m})^{r_a}$ using a random value r_a.
2. The access point computes g^{r_a} and signs g^{r_a}, g^{r_m} with his private key SK_a. The resulted signature is encrypted using a symmetric algorithm with key K generated in step (1). The value g^{r_a} together with the certificate of the access point $Cert_a$ and the encrypted value of the signature is sent to the mobile device. Then the mobile device can compute the symmetric key K, obtains the digital signature and validates it.
3. The mobile device signs g^{r_a}, g^{r_m} with his private key SK_m and encrypts the resulted signature and the certificate $Cert_m$ with the shared key K. The inclusion of the certificate into the cipher text offers anonymity for the mobile device against eventual eavesdropping of the authentication messages.

8.6.5 Transaction Based Authentication System for Mobile Commerce

The transaction-based authentication system can be performed in mobile commerce environment instead of session based authentication. The mobile commerce transactions have been assigned to various levels (Level 0-3), based on the sensitivity of operations they are performing and the degrees of severity of consequences that might arise from misappropriation of customers electronic identity/credentials. Upon receiving the transaction request, the scheme identifies under which level of authentication the current transaction should execute; if more sensitive the transaction, then more is the level of authentication required. After initial authentication of a transaction by accepting suitable identifiers, the scheme continues to perform analysis over transaction by collecting necessary temporal and symptomatic parameters from the transaction request and data. This analysis generates the deviation factor for each transaction, the accumulation of the deviation factors of various transactions of the

session results in certainty factor of possibility of attack. The continuation of customer authentication is performed depending on the value of certainty factor of attack, which uses challenge/response model. The proposed architecture for transaction based authentication scheme for mobile commerce is shown in the Fig. 8.14.

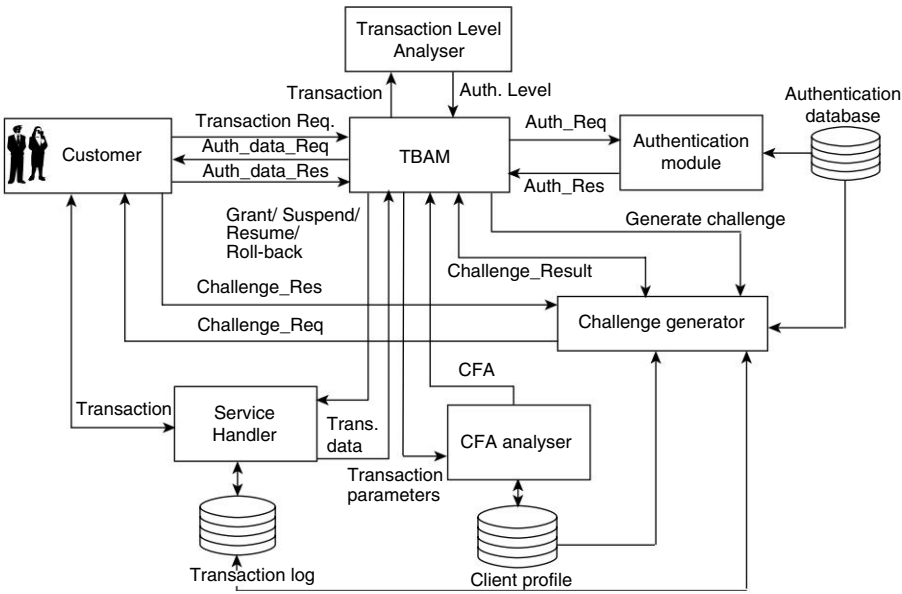


Fig. 8.14 *TBAM System Architecture*

It is the central co-ordinating module runs at mobile commerce service provider site, to perform transaction based authentication operation. Upon receiving the transaction request, the TBAM performs the following steps:

1. TBAM captures some initial set of parameters from the transaction request; to mention the few *(Time-of-Request, Location-of-Request, Device-address, Network-address)*.
2. The transaction request is passed to the *Transaction level analyser* to obtain the authentication level for the transaction.
3. After obtaining the authentication level, the TBAM requests for required authentication data from the customer.
4. The TBAM requests the *Authentication Module*, to validate the received authentication data from the customer, and after successful validation the TBAM issue grant transaction command to the *Service Handler*.
5. For every transaction the TBAM accepts transaction data submitted by the customer from the *Service Handler*.

6. The TBAM captures required symptomatic parameters like *(Transaction-amount, Account-numbers, Shipping-type, etc.)*, from the submitted transaction values and pass both initial set of parameters and symptomatic parameters to the *CFA Analyzer* for computation of the Certainty Factor of Attack (CFA).
7. If CFA is exceeding the threshold, then the TBAM issues suspend transaction command to *Service Handler* and requests the *Challenge Generator*, to generate challenge for the customer. If challenge is answered then the customer is allowed to continue the service by issuing resume transaction command to *Service Handler* or else the roll-back transaction command is issued to *Service Handler*.

8.6.6 Implementation of The TBSS and The TBAS Schemes for Mobile Commerce Application

The Fig. 8.15, gives the integrated system of the Transaction-based security selection (TBSS)-Scheme and the Transaction-based authentication selection (TBAS)-Scheme, and following entities in the m-commerce environment are considered, to implement integrated security system.

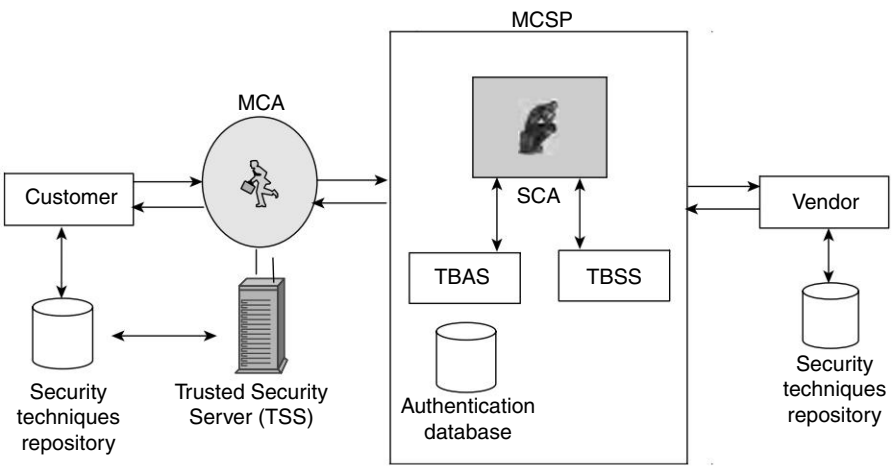


Fig. 8.15 *Transaction-based Security and Authentication System for Mobile Commerce*

Customer

Customer is an entity who buys goods or services using his/her mobile devices. The security techniques repository is either hosted on the customer device or

on the trusted security server. The Mobile Cognitive Agent (MCA) is migrated onto client device.

Vendor

An organisation, whether a profit business or a non-profit entity, conducting m-Commerce with customers. Vendor hosts security techniques repository.

Mobile Commerce Service Provider (MCSP)

The MCSP is a trusted service provider who provides the following services.

- Maintains index of vendors.
- Periodically updates the index.
- Provides vendor details to customers on request.
- Hosts the static cognitive agent (SCA), the TBSS, and the TBAS modules, and communicates with customer, and vendor. The TBSS of MCSP is responsible for deciding the security technique to be used for transactions security, the selected security technique identifier is communicated to both customer and the vendor. The TBAS of MCSP is responsible for forming authentication challenges and validating responses, using authentication database.

The sample list of transactions classification used in the mobile commerce application is given in the Table 8.1.

Table 8.1 *Sensitivity Levels and Transactions Categorisation*

TSL	*Example Transactions*
0	Product general information browsing, downloading free samples, browsing other user feedbacks, etc.
1	Request for technical information, requesting comparative statements, requesting after sales service options, placing low volume orders, requesting feedbacks, requesting account statement, etc.
2	Placing high volume orders, making macro payments, requesting purchase bills, requesting private information, making account transfers, etc.
3	Making large advance payments, Availing credits, etc.

The working of the proposed system is illustrated with the following example. A mobile client wish to purchase an electronic item, say, calculator from some mobile commerce vendor.

Step 1: The commerce service starts with the request from a client to the MCSP to know the available vendors who sells a calculator.

Step 2: The MCSP provides the list of vendors who sells the product calculator.

Step 3: After obtaining the list, the client selects a vendor from the list and forwards his/her selection to the MCSP.

Step 4: The MCSP migrates an instance of the MCA with initial security-ID to the client device, and indicates the selected vendor regarding this.

Step 5: The client communicates the transaction to the vendor, if the transaction is non-sensitive the vendor do not demand any authentication. Meantime, the MCA communicates the beliefs generated to the SCA at the MCSP.

Step 6: The SCA keep analysing the variations in belief and transaction sensitivity levels, if there is any increase in transaction sensitivity level, or the belief deviation factor is crossing the threshold. The SCA generates the next security technique to be used, and conveys the same to both the MCA and the vendor. The SCA also creates an authentication challenge and validates the client, based on the transaction sensitivity level and the belief deviation factor.

Step 7: This process continues for remaining transactions in the session.

The Table 8.2, discusses three test cases of purchasing scenario, which are distinguished based on the number of products the customer purchased. In each test case, we have listed the number of various levels of transactions took place during the purchase.

Table 8.2 *Breakup of Transactions into Various Levels*

Cases	*Number of products purchased*	*Number of transactions*	*TSL=0TS*	*L=1*	*TSl=2*	*TSL=3*
Case 1	1	20	7	8	3	2
Case 2	5	186	64	98	18	6
Case 3	10	965	284	436	100	145

Figure 8.16 shows the transactions arrival patterns, along with type of authentication challenges generated, and also demonstrates the instances where the deviation factor has gone above established threshold of normal behaviours. The distribution shows, in the first two cases sensitivity levels of transactions increases gradually with the progress of purchase, since the customer's buying behaviours is normally of type case-1 and case-2, we will not notice frequent authentication challenges, and changes in security levels. But, during the case-3, the deviation factor computed over behaviours of a customer, has

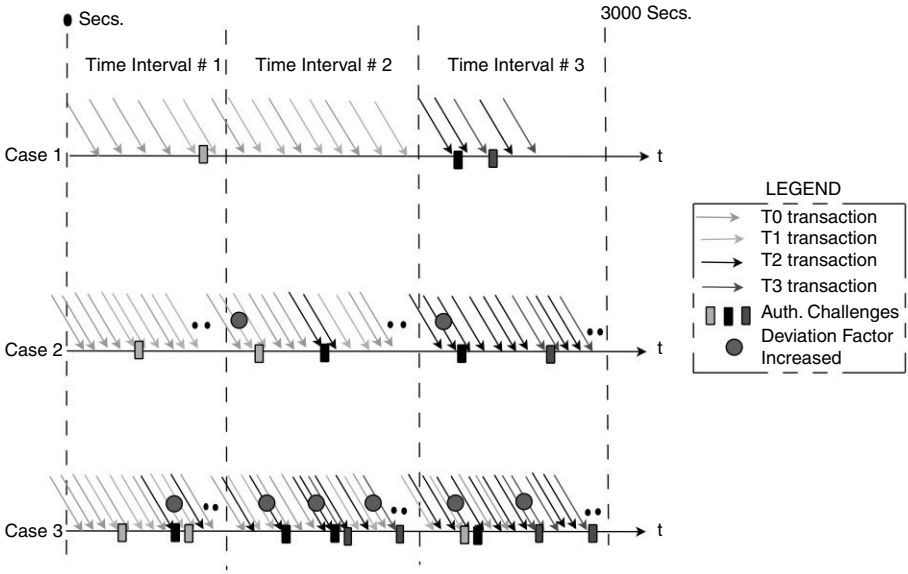


Fig. 8.16 *Transactions Arrival Pattern of a Mobile Commerce Customer*

crossed the threshold of suspiciousness many times, this is due to changes in buying behaviours of a customer. The customer is buying the products with increased frequency, results in increase in deviation factor, which in turn leads to increase in number of authentication challenges, and frequent changes in security levels used for transactions. The Table 8.3, provides the detailed list of observations made in every time slot.

Table 8.3 *Results Recorded in Every Time Slot, with Threshold as 0.45*

Cases	*Time slot*	T_0	T_1	T_2	T_3	*Deviation factor generated*	*Security levels used*	*Authentication levels used*
Case 1	#1	6	1	0	0	0.00	1	1
	#2	1	5	2	0	0.00	2	2
	#3	0	2	1	2	0.00	2	3
Case 2	#1	35	30	0	0	0.00	1	1
	#2	20	49	8	1	0.46	1,2	1,2
	#3	9	19	6	5	0.48	1,2,3	2,3
Case 3	#1	82	144	9	25	0.47	1,2	1,2
	#2	110	192	71	40	0.49, 0.54, 0.62	1,2,3	1,2,3
	#3	92	100	20	80	0.64, 0.66	1,2,3	1,2,3

The Table 8.4, shows the average time of cryptographic techniques executed by the system for the given three cases. At each level, there is a set of cryptographic operations, which perform similar operations, the methods are chosen randomly by the TBSS-Scheme. Since level-0 type of transactions are not coming under any security requirements, they were not listed in the table. The Table 8.5, shows the number of authentication challenges generated by the TBAS-Scheme, for various levels of transaction during the execution of test cases.

Table 8.4 *Avg. Time for Security Operations for Various Cases*

Cases	*Transaction type*	*Cryptography methods selected*	*Avg. time for encryption/ signature generation*	*Avg. time for decryption/ signature verification*
Case 1	T1	AES	53.375 ms.	54.75 ms.
	T2	DSA	2.6 ms	4.78 ms
	T3	Rabin	5.5 ms.	5.75 ms.
Case 2	T1	AES + RC4	57.74 ms.	58.95 ms.
	T2	ElGamal + DSA	7.19 ms.	9.82 ms.
	T3	RSA + ElGamal	5.25 ms.	5.98 ms.
Case 3	T1	AES + BlowFish + RC4	63.85 ms.	67.92 ms.
	T2	DSA + GQ + ElGamal	8.4 ms.	10.45 ms.
	T3	RSA + Knapsack + GM	6.34 ms.	7.3 ms.

Table 8.5 *Authentication challenges vs. transaction levels*

Cases	*Transaction Type*	*Authentication technique selected*	*Number of authentication challenges*
Case 1	T0	Nil	Nil
	T1	Device address	1
	T2	Login-name/Password	1
	T3	Digital certificate	1
Case 2	T0	Nil	Nil
	T1	Device address, IP, e-mail ID	3
	T2	Login-name/Password, PIN, TAN	3
	T3	Digital certificate	1
Case 3	T0	Nil	Nil
	T1	Device address, IP, e-mail ID, SSN, driving- license number, policy number …	27
	T2	Login-name/Password, PIN, TAN, favorite-product, favorite-shop, …	9
	T3	Digital certificate, Voice sample, favorite-day-of-purchase, favorite-vendor	4

SUMMARY

In this chapter, detailed discussions are made on the importance of mobile commerce in present business scenario along with few typical applications. We have mentioned what are the potential key drivers of mobile commerce. At the security side, we have presented various security challenges, attacks, and discussed some of the security schemes which are in practice for mobile commerce applications.

REVIEW QUESTIONS

1. Briefly explain the importance of mobile commerce in the current business scenario.
2. Differentiate active and passive applications in m-commerce.
3. What are the potential uses of mobile shopping?
4. How location-based services will help both customer and vendor in m-commerce environment?
5. Discuss various entertainment based applications in m-commerce environment.
6. What are the issues considered as the key drivers of mobile commerce?
7. What are various initiatives taken by industries and academia for promoting mobile commerce?
8. Explain various security challenges for mobile commerce applications.
9. Explain interruption attack over mobile commerce transactions, with an illustration.
10. Explain interception attack over mobile commerce transactions, with an illustration.
11. Explain modification attack over mobile commerce transactions, with an illustration.
12. Explain fabrication attack over mobile commerce transactions, with an illustration.
13. Describe a secure m-commerce model.
14. Explain the role of PKI/CA in mobile commerce.
15. Explain the process of authentication in mobile-IP.
16. How ECC type of cryptography is useful over mobile devices compared to RSA cryptography scheme?
17. Explain the role and operation of mobile payment systems in m-commerce environment.
18. Discuss some of the available payment solutions.
19. What are the various categories of e-payment systems?
20. What are the various categories of m-payment systems?
21. Explain the various requirements of asymmetric authentication protocol for mobile commerce applications.
22. Explain the working of MASPA protocol.

23. Explain the need of transaction-based authentication system for m-commerce applications.
24. Explain the working of transaction-based authentication system.
25. Explain the integrated working of TBSS and the TBAS schemes for mobile commerce.

Index