Mobile device is configured with security mechanisms and parameters to conform to organization security policy Traffic is encrypted; uses SSL or IPsec VPN tunnel Authentication/ access control server Mobile device configuration server Application/ database server M 230 chapter 7 / Wireless Network Security ■■ Either sensitive data should be prohibited from storage on the mobile device or it should be encrypted. ■■ IT staff should also have the ability to remotely access devices, wipe the device of all data, and then disable the device in the event of loss or theft. ■■ The organization may prohibit all installation of third-party applications, implement whitelisting to prohibit installation of all unapproved applications, or implement a secure sandbox that isolates the organization's data and applications from all other data and applications on the mobile device. Any application that is on an approved list should be accompanied by a digital signature and a public-key certificate from an approved authority. ■■ The organization can implement and enforce restrictions on what devices can synchronize and on the use of cloud-based storage. ■■ To deal with the threat of untrusted content, security responses can include training of personnel on the risks inherent in untrusted content and disabling camera use on corporate mobile devices. ■■ To counter the threat of malicious use of location services, the security policy can dictate that such service is disabled on all mobile devices. Traffic Security Traffic security is based on the usual mechanisms for encryption and authentication. All traffic should be encrypted and travel by secure means, such as SSL or IPv6. Virtual private networks (VPNs) can be configured so that all traffic between the mobile device and the organization's network is via a VPN. A strong authentication protocol should be used to limit the access from the device to the resources of the organization. Often, a mobile device has a single device-specific authenticator, because it is assumed that the device has only one user. A preferable strategy is to have a two-layer authentication mechanism, which involves authenticating the device and then authenticating the user of the device. Barrier Security The organization should have security mechanisms to protect the network from unauthorized access. The security strategy can also include firewall policies specific to mobile device traffic. Firewall policies can limit the scope of data and application access for all mobile devices. Similarly, intrusion detection and intrusion prevention systems can be configured to have tighter rules for mobile device traffic. 7.3 IEEE 802.11 Wireless Lan Overview IEEE 802 is a committee that has developed standards for a wide range of local area networks (LANs). In 1990, the IEEE 802 Committee formed a new working group, IEEE 802.11, with a charter to develop a protocol and transmission specifications for wireless LANs (WLANs). Since that time, the demand for WLANs at different frequencies and data rates has exploded. Keeping pace with this demand, the IEEE 802.11 working group has issued an ever-expanding list of standards. Table 7.1 briefly defines key terms used in the IEEE 802.11 standard. 7.3 / IEEE 802.11 Wireless Lan Overview 231 The Wi-Fi Alliance The first 802.11 standard to gain broad industry acceptance was 802.11b. Although 802.11b products are all based on the same standard, there is always a concern whether products from different vendors will successfully interoperate. To meet this concern, the Wireless Ethernet Compatibility Alliance (WECA), an industry consortium, was formed in 1999. This organization, subsequently renamed the Wi-Fi (Wireless Fidelity) Alliance, created a test suite to certify interoperability for 802.11b products. The term used for certified 802.11b products is Wi-Fi. Wi-Fi certification has been extended to 802.11g products. The Wi-Fi Alliance has also developed a certification process for 802.11a products, called Wi-Fi5. The Wi-Fi Alliance is concerned with a range of market areas for WLANs, including enterprise, home, and hot spots. More recently, the Wi-Fi Alliance has developed

certification procedures for IEEE 802.11 security standards, referred to as Wi-Fi Protected Access (WPA). The most recent version of WPA, known as WPA2, incorporates all of the features of the IEEE 802.11i WLAN security specification. IEEE 802 Protocol Architecture Before proceeding, we need to briefly preview the IEEE 802 protocol architecture. IEEE 802.11 standards are defined within the structure of a layered set of protocols. This structure, used for all IEEE 802 standards, is illustrated in Figure 7.3. Physical Layer The lowest layer of the IEEE 802 reference model is the physical layer, which includes such functions as encoding/decoding of signals and bit transmission/reception. In addition, the physical layer includes a specification of the transmission medium. In the case of IEEE 802.11, the physical layer also defines frequency bands and antenna characteristics. Access point (AP) Any entity that has station functionality and provides access to the distribution system via the wireless medium for associated stations. Basic service set (BSS) A set of stations controlled by a single coordination function. Coordination function The logical function that determines when a station operating within a BSS is permitted to transmit and may be able to receive PDUs. Distribution system (DS) A system used to interconnect a set of BSSs and integrated LANs to create an ESS. Extended service set (ESS) A set of one or more interconnected BSSs and integrated LANs that appear as a single BSS to the LLC layer at any station associated with one of these BSSs. MAC protocol data unit (MPDU) The unit of data exchanged between two peer MAC entities using the services of the physical layer. MAC service data unit (MSDU) Information that is delivered as a unit between MAC users. Station Any device that contains an IEEE 802.11 conformant MAC and physical layer. Table 7.1 IEEE 802.11 Terminology 232 chapter 7 / Wireless Network Security Media Access Control All LANs consist of collections of devices that share the network's transmission capacity. Some means of controlling access to the transmission medium is needed to provide an orderly and efficient use of that capacity. This is the function of a media access control (MAC) layer. The MAC layer receives data from a higher-layer protocol, typically the Logical Link Control (LLC) layer, in the form of a block of data known as the MAC service data unit (MSDU). In general, the MAC layer performs the following functions: ■■ On transmission, assemble data into a frame, known as a MAC protocol data unit (MPDU) with address and error-detection fields. ■■ On reception, disassemble frame, and perform address recognition and error detection. ■■ Govern access to the LAN transmission medium. The exact format of the MPDU differs somewhat for the various MAC protocols in use. In general, all of the MPDUs have a format similar to that of Figure 7.4. The fields of this frame are as follows. ■■ MAC Control: This field contains any protocol control information needed for the functioning of the MAC protocol. For example, a priority level could be indicated here. ■■ Destination MAC Address: The destination physical address on the LAN for this MPDU. ■■ Source MAC Address: The source physical address on the LAN for this MPDU. Figure 7.3 IEEE 802.11 Protocol Stack Logical Link Control Medium Access Control Physical Encoding/decoding of signals Bit transmission/reception Transmission medium Assemble data into frame Addressing Error detection Medium access Flow control Error control General IEEE 802 functions Specic IEEE 802.11 functions Frequency band denition Wireless signal encoding Reliable data delivery Wireless access control protocols 7.3 / IEEE 802.11 Wireless Lan Overview 233 ■■ MAC Service Data Unit: The data from the next higher layer. ■■ CRC: The cyclic redundancy check field; also known as the Frame Check Sequence (FCS) field. This is an error-detecting code, such as that which is used in other data-link control protocols. The CRC is calculated based on the bits in the entire MPDU. The sender calculates the CRC and adds it to the frame. The receiver performs the same calculation on the incoming

MPDU and compares that calculation to the CRC field in that incoming MPDU. If the two values don’t match, then one or more bits have been altered in transit. The fields preceding the MSDU field are referred to as the MAC header, and the field following the MSDU field is referred to as the MAC trailer. The header and trailer contain control information that accompany the data field and that are used by the MAC protocol. Logical Link Control In most data-link control protocols, the data-link protocol entity is responsible not only for detecting errors using the CRC, but for recovering from those errors by retransmitting damaged frames. In the LAN protocol architecture, these two functions are split between the MAC and LLC layers. The MAC layer is responsible for detecting errors and discarding any frames that contain errors. The LLC layer optionally keeps track of which frames have been successfully received and retransmits unsuccessful frames. IEEE 802.11 Network Components and Architectural Model Figure 7.5 illustrates the model developed by the 802.11 working group. The smallest building block of a wireless LAN is a basic service set (BSS), which consists of wireless stations executing the same MAC protocol and competing for access to the same shared wireless medium. A BSS may be isolated, or it may connect to a backbone distribution system (DS) through an access point (AP). The AP functions as a bridge and a relay point. In a BSS, client stations do not communicate directly with one another. Rather, if one station in the BSS wants to communicate with another station in the same BSS, the MAC frame is first sent from the originating station to the AP and then from the AP to the destination station. Similarly, a MAC frame from a station in the BSS to a remote station is sent from the local station to the AP and then relayed by the AP over the DS on its way to the destination station. The BSS generally corresponds to what is referred to as a cell in the literature. The DS can be a switch, a wired network, or a wireless network. When all the stations in the BSS are mobile stations that communicate directly with one another (not using an AP), the BSS is called an independent BSS (IBSS). An IBSS is typically an ad hoc network. In an IBSS, the stations all communicate directly, and no AP is involved. Figure 7.4 General IEEE 802 MPDU Format MAC Control redaeh CAM reliart CAM Destination MAC Address Source MAC Address MAC Service Data Unit (MSDU) CRC 234 chapter 7 / Wireless Network Security A simple configuration is shown in Figure 7.5, in which each station belongs to a single BSS; that is, each station is within wireless range only of other stations within the same BSS. It is also possible for two BSSs to overlap geographically, so that a single station could participate in more than one BSS. Furthermore, the association between a station and a BSS is dynamic. Stations may turn off, come within range, and go out of range. An extended service set (ESS) consists of two or more basic service sets interconnected by a distribution system. The extended service set appears as a single logical LAN to the logical link control (LLC) level. IEEE 802.11 Services IEEE 802.11 defines nine services that need to be provided by the wireless LAN to achieve functionality equivalent to that which is inherent to wired LANs. Table 7.2 lists the services and indicates two ways of categorizing them. 1. The service provider can be either the station or the DS. Station services are implemented in every 802.11 station, including AP stations. Distribution services are provided between BSSs; these services may be implemented in an AP or in another special-purpose device attached to the distribution system. 2. Three of the services are used to control IEEE 802.11 LAN access and confidentiality. Six of the services are used to support delivery of MSDUs between stations. If the MSDU is too large to be transmitted in a single MPDU, it may be fragmented and transmitted in a series of MPDUs. Figure 7.5 IEEE 802.11 Extended Service Set STA 2 STA 3 STA4 STA 1 STA 6 STA 7 STA 8 AP 2 AP 1 Basic Service Set (BSS) Basic Service Set (BSS) Distribution System 7.3 / IEEE 802.11 Wireless

Following the IEEE 802.11 document, we next discuss the services in an order designed to clarify the operation of an IEEE 802.11 ESS network. MSDU delivery, which is the basic service, already has been mentioned. Services related to security are introduced in Section 7.4. Distribution of Messages Within a DS The two services involved with the distribution of messages within a DS are distribution and integration. Distribution is the primary service used by stations to exchange MPDUs when the MPDUs must traverse the DS to get from a station in one BSS to a station in another BSS. For example, suppose a frame is to be sent from station 2 (STA 2) to station 7 (STA 7) in Figure 7.5. The frame is sent from STA 2 to AP 1, which is the AP for this BSS. The AP gives the frame to the DS, which has the job of directing the frame to the AP associated with STA 7 in the target BSS. AP 2 receives the frame and forwards it to STA 7. How the message is transported through the DS is beyond the scope of the IEEE 802.11 standard. If the two stations that are communicating are within the same BSS, then the distribution service logically goes through the single AP of that BSS. The integration service enables transfer of data between a station on an IEEE 802.11 LAN and a station on an integrated IEEE 802.x LAN. The term integrated refers to a wired LAN that is physically connected to the DS and whose stations may be logically connected to an IEEE 802.11 LAN via the integration service. The integration service takes care of any address translation and media conversion logic required for the exchange of data. Association-Related Services The primary purpose of the MAC layer is to transfer MSDUs between MAC entities; this purpose is fulfilled by the distribution service. For that service to function, it requires information about stations within the ESS that is provided by the association-related services. Before the distribution service can deliver data to or accept data from a station, that station must be associated. Before looking at the concept of association, we need Service Provider Used to support Association Distribution system MSDU delivery Authentication Station LAN access and security Deauthentication Station LAN access and security Disassociation Distribution system MSDU delivery Distribution Distribution system MSDU delivery Integration Distribution system MSDU delivery MSDU delivery Station MSDU delivery Privacy Station LAN access and security Reassociation Distribution system MSDU delivery Table 7.2 IEEE 802.11 Services 236 chapter 7 / Wireless Network Security to describe the concept of mobility. The standard defines three transition types, based on mobility: ■■ No transition: A station of this type is either stationary or moves only within the direct communication range of the communicating stations of a single BSS. ■■ BSS transition: This is defined as a station movement from one BSS to another BSS within the same ESS. In this case, delivery of data to the station requires that the addressing capability be able to recognize the new location of the station. ■■ ESS transition: This is defined as a station movement from a BSS in one ESS to a BSS within another ESS. This case is supported only in the sense that the station can move. Maintenance of upper-layer connections supported by 802.11 cannot be guaranteed. In fact, disruption of service is likely to occur. To deliver a message within a DS, the distribution service needs to know where the destination station is located. Specifically, the DS needs to know the identity of the AP to which the message should be delivered in order for that message to reach the destination station. To meet this requirement, a station must maintain an association with the AP within its current BSS. Three services relate to this requirement: ■■ Association: Establishes an initial association between a station and an AP. Before a station can transmit or receive frames on a wireless LAN, its identity and address must be known. For this purpose, a station must establish an association with an AP within a particular BSS. The AP can then communicate this information to other APs within the ESS to facilitate routing and delivery of addressed frames. ■■ Reassociation: Enables an established association to be transferred from one AP to another, allowing a mobile station to move from one BSS to another. ■■ Disassociation: A notification from either a station or an AP that an existing association is terminated. A station should give this notification before leaving an ESS or shutting down. However, the MAC management facility protects itself against stations that disappear without notification. 7.4 IEEE 802.11i Wireless Lan Security There are two characteristics of a wired LAN that are not inherent in a wireless LAN. 1. In order to transmit over a wired LAN, a station must be physically connected to the LAN. On the other hand, with a wireless LAN, any station within radio range of the other devices on the LAN can transmit. In a sense, there is a form of authentication with a wired LAN in that it requires some positive and presumably observable action to connect a station to a wired LAN. 2. Similarly, in order to receive a transmission from a station that is part of a wired LAN, the receiving station also must be attached to the wired LAN. On the other hand, with a wireless LAN, any station within radio range can receive. Thus, a wired LAN provides a degree of privacy, limiting reception of data to stations connected to the LAN. 7.4 / IEEE 802.11i Wireless Lan Security 237 These differences between wired and wireless LANs suggest the increased need for robust security services and mechanisms for wireless LANs. The original 802.11 specification included a set of security features for privacy and authentication that were quite weak. For privacy, 802.11 defined the Wired Equivalent Privacy (WEP) algorithm. The privacy portion of the 802.11 standard contained major weaknesses. Subsequent to the development of WEP, the 802.11i task group has developed a set of capabilities to address the WLAN security issues. In order to accelerate the introduction of strong security into WLANs, the Wi-Fi Alliance promulgated Wi-Fi Protected Access (WPA) as a Wi-Fi standard. WPA is a set of security mechanisms that eliminates most 802.11 security issues and was based on the current state of the 802.11i standard. The final form of the 802.11i standard is referred to as Robust Security Network (RSN). The Wi-Fi Alliance certifies vendors in compliance with the full 802.11i specification under the WPA2 program. The RSN specification is quite complex, and occupies 145 pages of the 2012 IEEE 802.11 standard. In this section, we provide an overview. IEEE 802.11i Services The 802.11i RSN security specification defines the following services. ■■ Authentication: A protocol is used to define an exchange between a user and an AS that provides mutual authentication and generates temporary keys to be used between the client and the AP over the wireless link. ■■ Access control:1 This function enforces the use of the authentication function, routes the messages properly, and facilitates key exchange. It can work with a variety of authentication protocols. ■■ Privacy with message integrity: MAC-level data (e.g., an LLC PDU) are encrypted along with a message integrity code that ensures that the data have not been altered. Figure 7.6a indicates the security protocols used to support these services, while Figure 7.6b lists the cryptographic algorithms used for these services. IEEE 802.11i Phases of Operation The operation of an IEEE 802.11i RSN can be broken down into five distinct phases of operation. The exact nature of the phases will depend on the configuration and the end points of the communication. Possibilities include (see Figure 7.5): 1. Two wireless stations in the same BSS communicating via the access point (AP) for that BSS. 2. Two wireless stations (STAs) in the same ad hoc IBSS communicating directly with each other. 1 In this context, we are discussing access control as a security function. This is a different function than media access control (MAC) as described in Section 7.3. Unfortunately, the literature and the standards use the term access control in both contexts. 238 chapter 7 / Wireless Network Security 3. Two wireless stations in different BSSs communicating via their respective APs

across a distribution system. 4. A wireless station communicating with an end station on a wired network via its AP and the distribution system. IEEE 802.11i security is concerned only with secure communication between the STA and its AP. In case 1 in the preceding list, secure communication is assured if each STA establishes secure communications with the AP. Case 2 is similar, with the AP functionality residing in the STA. For case 3, security is not provided across the distribution system at the level of IEEE 802.11, but only within each BSS. Endto-end security (if required) must be provided at a higher layer. Similarly, in case 4, security is only provided between the STA and its AP. Figure 7.6 Elements of IEEE 802.11i Access Control Services Protocols Services Algorithms IEEE 802.1 Port-based Access Control Extensible Authentication Protocol (EAP) Authentication and Key Generation (a) Services and protocols Condentiality, Data Origin Authentication and Integrity and Replay Protection TKIP CCMP Robust Security Network (RSN) Con-dentiality TKIP (Michael MIC) CCM (AESCBCMAC) CCM (AESCTR) NIST Key Wrap HMACMD5 HMACSHA-1 Integrity and Data Origin Authentication (b) Cryptographic algorithms Key Generation TKIP (RC4) Robust Security Network (RSN) HMACSHA-1 RFC 1750 CBC-MAC = Cipher Block Chaining Message Authentication Code (MAC) CCM = Counter Mode with Cipher Block Chaining Message Authentication Code CCMP = Counter Mode with Cipher Block Chaining MAC Protocol TKIP = Temporal Key Integrity Protocol 7.4 / IEEE 802.11i Wireless Lan Security 239 With these considerations in mind, Figure 7.7 depicts the five phases of operation for an RSN and maps them to the network components involved. One new component is the authentication server (AS). The rectangles indicate the exchange of sequences of MPDUs. The five phases are defined as follows. ■■ Discovery: An AP uses messages called Beacons and Probe Responses to advertise its IEEE 802.11i security policy. The STA uses these to identify an AP for a WLAN with which it wishes to communicate. The STA associates with the AP, which it uses to select the cipher suite and authentication mechanism when the Beacons and Probe Responses present a choice. ■■ Authentication: During this phase, the STA and AS prove their identities to each other. The AP blocks non-authentication traffic between the STA and AS until the authentication transaction is successful. The AP does not participate in the authentication transaction other than forwarding traffic between the STA and AS. ■■ Key generation and distribution: The AP and the STA perform several operations that cause cryptographic keys to be generated and placed on the AP and the STA. Frames are exchanged between the AP and STA only. ■■ Protected data transfer: Frames are exchanged between the STA and the end station through the AP. As denoted by the shading and the encryption module icon, secure data transfer occurs between the STA and the AP only; security is not provided end-to-end. Figure 7.7 IEEE 802.11i Phases of Operation Phase 1 - Discovery STA AP AS End Station Phase 5 - Connection Termination Phase 3 - Key Management Phase 4 - Protected Data Transfer Phase 2 - Authentication 240 chapter 7 / Wireless Network Security ■■ Connection termination: The AP and STA exchange frames. During this phase, the secure connection is torn down and the connection is restored to the original state. Discovery Phase We now look in more detail at the RSN phases of operation, beginning with the discovery phase, which is illustrated in the upper portion of Figure 7.8. The purpose of this phase is for an STA and an AP to recognize each other, agree on a set of security capabilities, and establish an association for future communication using those security capabilities. Figure 7.8 IEEE 802.11i Phases of Operation: Capability Discovery, Authentication, and Association STA AP AS Probe request Station sends a request to join network AP sends possible security parameter (security capabilities set per the security policy) AP performs null authentication AP sends the associated security parameters

Station sends a request to perform null authentication Station sends a request to associate with AP with security parameters Station sets selected security parameters Open system authentication request Probe response 802.1X EAP request Access request (EAP request) 802.1X EAP response Accept/EAP-success key material 802.1X EAP success Association request Association response Open system authentication response 802.1X-controlled port blocked 802.1X-controlled port blocked Extensible Authentication Protocol Exchange 7.4 / IEEE 802.11i Wireless Lan Security 241 Security Capabilities During this phase, the STA and AP decide on specific techniques in the following areas: ■■ Confidentiality and MPDU integrity protocols for protecting unicast traffic (traffic only between this STA and AP) ■■ Authentication method ■■ Cryptography key management approach Confidentiality and integrity protocols for protecting multicast/broadcast traffic are dictated by the AP, since all STAs in a multicast group must use the same protocols and ciphers. The specification of a protocol, along with the chosen key length (if variable) is known as a cipher suite. The options for the confidentiality and integrity cipher suite are ■■ WEP, with either a 40-bit or 104-bit key, which allows backward compatibility with older IEEE 802.11 implementations ■■ TKIP ■■ CCMP ■■ Vendor-specific methods The other negotiable suite is the authentication and key management (AKM) suite, which defines (1) the means by which the AP and STA perform mutual authentication and (2) the means for deriving a root key from which other keys may be generated. The possible AKM suites are ■■ IEEE 802.1X ■■ Pre-shared key (no explicit authentication takes place and mutual authentication is implied if the STA and AP share a unique secret key) ■■ Vendor-specific methods MPDU Exchange The discovery phase consists of three exchanges. ■■ Network and security capability discovery: During this exchange, STAs discover the existence of a network with which to communicate. The AP either periodically broadcasts its security capabilities (not shown in figure), indicated by RSN IE (Robust Security Network Information Element), in a specific channel through the Beacon frame; or responds to a station’s Probe Request through a Probe Response frame. A wireless station may discover available access points and corresponding security capabilities by either passively monitoring the Beacon frames or actively probing every channel. ■■ Open system authentication: The purpose of this frame sequence, which provides no security, is simply to maintain backward compatibility with the IEEE 802.11 state machine, as implemented in existing IEEE 802.11 hardware. In essence, the two devices (STA and AP) simply exchange identifiers. ■■ Association: The purpose of this stage is to agree on a set of security capabilities to be used. The STA then sends an Association Request frame to the AP. In this frame, the STA specifies one set of matching capabilities 242 chapter 7 / Wireless Network Security (one authentication and key management suite, one pairwise cipher suite, and one group-key cipher suite) from among those advertised by the AP. If there is no match in capabilities between the AP and the STA, the AP refuses the Association Request. The STA blocks it too, in case it has associated with a rogue AP or someone is inserting frames illicitly on its channel. As shown in Figure 7.8, the IEEE 802.1X controlled ports are blocked, and no user traffic goes beyond the AP. The concept of blocked ports is explained subsequently. Authentication Phase As was mentioned, the authentication phase enables mutual authentication between an STA and an authentication server (AS) located in the DS. Authentication is designed to allow only authorized stations to use the network and to provide the STA with assurance that it is communicating with a legitimate network. IEEE 802.1X Access Control Approach IEEE 802.11i makes use of another standard that was designed to provide access control functions for LANs. The standard is IEEE 802.1X, Port-Based Network Access Control. The authentication protocol that is used, the Extensible

Authentication Protocol (EAP), is defined in the IEEE 802.1X standard. IEEE 802.1X uses the terms supplicant, authenticator, and authentication server (AS). In the context of an 802.11 WLAN, the first two terms correspond to the wireless station and the AP. The AS is typically a separate device on the wired side of the network (i.e., accessible over the DS) but could also reside directly on the authenticator. Before a supplicant is authenticated by the AS using an authentication protocol, the authenticator only passes control or authentication messages between the supplicant and the AS; the 802.1X control channel is unblocked, but the 802.11 data channel is blocked. Once a supplicant is authenticated and keys are provided, the authenticator can forward data from the supplicant, subject to predefined access control limitations for the supplicant to the network. Under these circumstances, the data channel is unblocked. As indicated in Figure 5.5, 802.1X uses the concepts of controlled and uncontrolled ports. Ports are logical entities defined within the authenticator and refer to physical network connections. For a WLAN, the authenticator (the AP) may have only two physical ports: one connecting to the DS and one for wireless communication within its BSS. Each logical port is mapped to one of these two physical ports. An uncontrolled port allows the exchange of PDUs between the supplicant and the other AS, regardless of the authentication state of the supplicant. A controlled port allows the exchange of PDUs between a supplicant and other systems on the LAN only if the current state of the supplicant authorizes such an exchange. IEEE 802.1X is covered in more detail in Chapter 5. The 802.1X framework, with an upper-layer authentication protocol, fits nicely with a BSS architecture that includes a number of wireless stations and an AP. However, for an IBSS, there is no AP. For an IBSS, 802.11i provides a more complex solution that, in essence, involves pairwise authentication between stations on the IBSS. 7.4 / IEEE 802.11i Wireless Lan Security 243 MPDU Exchange The lower part of Figure 7.8 shows the MPDU exchange dictated by IEEE 802.11 for the authentication phase. We can think of authentication phase as consisting of the following three phases. ■■ Connect to AS: The STA sends a request to its AP (the one with which it has an association) for connection to the AS. The AP acknowledges this request and sends an access request to the AS. ■■ EAP exchange: This exchange authenticates the STA and AS to each other. A number of alternative exchanges are possible, as explained subsequently. ■■ Secure key delivery: Once authentication is established, the AS generates a master session key (MSK), also known as the Authentication, Authorization, and Accounting (AAA) key and sends it to the STA. As explained subsequently, all the cryptographic keys needed by the STA for secure communication with its AP are generated from this MSK. IEEE 802.11i does not prescribe a method for secure delivery of the MSK but relies on EAP for this. Whatever method is used, it involves the transmission of an MPDU containing an encrypted MSK from the AS, via the AP, to the AS. EAP Exchange As mentioned, there are a number of possible EAP exchanges that can be used during the authentication phase. Typically, the message flow between STA and AP employs the EAP over LAN (EAPOL) protocol, and the message flow between the AP and AS uses the Remote Authentication Dial In User Service (RADIUS) protocol, although other options are available for both STA-to-AP and AP-to-AS exchanges. [FRAN07] provides the following summary of the authentication exchange using EAPOL and RADIUS. 1. The EAP exchange begins with the AP issuing an EAP-Request/Identity frame to the STA. 2. The STA replies with an EAP-Response/Identity frame, which the AP receives over the uncontrolled port. The packet is then encapsulated in RADIUS over EAP and passed on to the RADIUS server as a RADIUS-Access-Request packet. 3. The AAA server replies with a RADIUS-Access-Challenge packet, which is passed on to the STA as an EAP-Request. This request is of the appropriate authentication type and contains relevant challenge information. 4. The STA formulates an EAP-Response message and sends it to the AS. The response is translated by the AP into a Radius-Access-Request with the response to the challenge as a data field. Steps 3 and 4 may be repeated multiple times, depending on the EAP method in use. For TLS tunneling methods, it is common for authentication to require 10 to 20 round trips. 5. The AAA server grants access with a Radius-Access-Accept packet. The AP issues an EAP-Success frame. (Some protocols require confirmation of the EAP success inside the TLS tunnel for authenticity validation.) The controlled port is authorized, and the user may begin to access the network. Note from Figure 7.8 that the AP controlled port is still blocked to general user traffic. Although the authentication is successful, the ports remain blocked 244 chapter 7 / Wireless Network Security until the temporal keys are installed in the STA and AP, which occurs during the 4-Way Handshake. Key Management Phase During the key management phase, a variety of cryptographic keys are generated and distributed to STAs. There are two types of keys: pairwise keys used for communication between an STA and an AP and group keys used for multicast communication. Figure 7.9, based on [FRAN07], shows the two key hierarchies, and Table 7.3 defines the individual keys. Figure 7.9 IEEE 802.11i Key Hierarchies Out-of-band path EAP method path Pre-shared key EAPOL key conrmation key EAPOL key encryption key Temporal key PSK 256 bits 384 bits (CCMP) 512 bits (TKIP) 128 bits (CCMP) 256 bits (TKIP) 40 bits, 104 bits (WEP) 128 bits (CCMP) 256 bits (TKIP) 256 bits 128 bits No modication Legend Possible truncation PRF (pseudo random function) using HMAC-SHA-1 128 bits User-dened cryptoid EAP authentication Following EAP authentication or PSK During 4-way handshake These keys are components of the PTK ≥ 256 bits PMK KCK PTK KEK KT AAAK or MSK Pairwise master key (b) Group key hierarchy (a) Pairwise key hierarchy AAA key Pairwise transient key 256 bits Changes periodically or if compromised Changes based on policy (dissociation, deauthentication) GMK (generated by AS) GTK Group master key Group temporal key 7.4 / IEEE 802.11i Wireless Lan Security 245 Abbreviation Name Description / Purpose Size (bits) Type AAA Key Authentication, Accounting, and Authorization Key Used to derive the PMK. Used with the IEEE 802.1X authentication and key management approach. Same as MMSK. Ú 256 Key generation key, root key PSK Pre-shared Key Becomes the PMK in pre-shared key environments. 256 Key generation key, root key PMK Pairwise Master Key Used with other inputs to derive the PTK. 256 Key generation key GMK Group Master Key Used with other inputs to derive the GTK. 128 Key generation key PTK Pair-wise Transient Key Derived from the PMK. Comprises the EAPOLKCK, EAPOL-KEK, and TK and (for TKIP) the MIC key. 512 (TKIP) 384 (CCMP) Composite key TK Temporal Key Used with TKIP or CCMP to provide confidentiality and integrity protection for unicast user traffic. 256 (TKIP) 128 (CCMP) Traffic key GTK Group Temporal Key Derived from the GMK. Used to provide confidentiality and integrity protection for multicast/broadcast user traffic. 256 (TKIP) 128 (CCMP) 40,104 (WEP) Traffic key MIC Key Message Integrity Code Key Used by TKIP's Michael MIC to provide integrity protection of messages. 64 Message integrity key EAPOL-KCK EAPOL-Key Confirmation Key Used to provide integrity protection for key material distributed during the 4-Way Handshake. 128 Message integrity key EAPOL-KEK EAPOL-Key Encryption Key Used to ensure the confidentiality of the GTK and other key material in the 4-Way Handshake. 128 Traffic key / key encryption key WEP Key Wired Equivalent Privacy Key Used with WEP. 40,104 Traffic key Table 7.3 IEEE 802.11i Keys for Data Confidentiality and Integrity Protocols 246 chapter 7 / Wireless Network Security Pairwise Keys Pairwise keys are used for communication between a pair of devices, typically between an STA and an AP. These keys form a hierarchy beginning with a master key from which other keys are derived dynamically and used for a limited period of time. At the top level of the hierarchy are two possibilities. A pre-shared key (PSK) is a secret key shared by the AP and a STA and installed in some fashion outside the scope of IEEE 802.11i. The other alternative is the master session key (MSK), also known as the AAAK, which is generated using the IEEE 802.1X protocol during the authentication phase, as described previously. The actual method of key generation depends on the details of the authentication protocol used. In either case (PSK or MSK), there is a unique key shared by the AP with each STA with which it communicates. All the other keys derived from this master key are also unique between an AP and an STA. Thus, each STA, at any time, has one set of keys, as depicted in the hierarchy of Figure 7.9a, while the AP has one set of such keys for each of its STAs. The pairwise master key (PMK) is derived from the master key. If a PSK is used, then the PSK is used as the PMK; if a MSK is used, then the PMK is derived from the MSK by truncation (if necessary). By the end of the authentication phase, marked by the 802.1X EAP Success message (Figure 7.8), both the AP and the STA have a copy of their shared PMK. The PMK is used to generate the pairwise transient key (PTK), which in fact consists of three keys to be used for communication between an STA and AP after they have been mutually authenticated. To derive the PTK, the HMAC-SHA-1 function is applied to the PMK, the MAC addresses of the STA and AP, and nonces generated when needed. Using the STA and AP addresses in the generation of the PTK provides protection against session hijacking and impersonation; using nonces provides additional random keying material. The three parts of the PTK are as follows. ■■ EAP Over LAN (EAPOL) Key Confirmation Key (EAPOL-KCK): Supports the integrity and data origin authenticity of STA-to-AP control frames during operational setup of an RSN. It also performs an access control function: proof-of-possession of the PMK. An entity that possesses the PMK is authorized to use the link. ■■ EAPOL Key Encryption Key (EAPOL-KEK): Protects the confidentiality of keys and other data during some RSN association procedures. ■■ Temporal Key (TK): Provides the actual protection for user traffic. Group Keys Group keys are used for multicast communication in which one STA sends MPDU's to multiple STAs. At the top level of the group key hierarchy is the group master key (GMK). The GMK is a key-generating key used with other inputs to derive the group temporal key (GTK). Unlike the PTK, which is generated using material from both AP and STA, the GTK is generated by the AP and transmitted to its associated STAs. Exactly how this GTK is generated is undefined. IEEE 802.11i, however, requires that its value is computationally indistinguishable from random. The GTK is distributed securely using the pairwise keys 7.4 / IEEE 802.11i Wireless Lan Security 247 that are already established. The GTK is changed every time a device leaves the network. Pairwise Key Distribution The upper part of Figure 7.10 shows the MPDU exchange for distributing pairwise keys. This exchange is known as the 4-way handshake. The STA and AP use this handshake to confirm the existence of the Figure 7.10 IEEE 802.11i Phases of Operation: 4-Way Handshake and Group Key Handshake STA AP Message 1 delivers a nonce to the STA so that it can generate the PTK. Message 1 delivers a new GTK to the STA. The GTK is encrypted before it is sent and the entire message is integrity protected. The AP installs the GTK. Message 3 demonstrates to the STA that the authenticator is alive, ensures that the PTK is fresh (new) and that there is no man-in-the-middle. Message 2 delivers another nonce to the AP so that it can also generate the PTK. It demonstrates to the AP that the STA is alive, ensures that the PTK is fresh (new) and that there is no man-in-the-middle. The STA decrypts the GTK and installs it for use. Message 2 is delivered to the AP. This frame serves only as an acknowledgment to the AP. Message 4 serves as an acknowledgment to Message 3. It serves no cryptographic function. This message also ensures the reliable start of the group key handshake. Message 2 EAPOL-key (Snonce, Unicast, MIC) Message 1 EAPOL-key (Anonce, Unicast) Message 1 EAPOL-key (GTK, MIC) Message 4 EAPOL-key (Unicast, MIC) Message 2 EAPOL-key (MIC) Message 3 EAPOL-key (Install PTK, Unicast, MIC) AP's 802.1X-controlled port blocked AP's 802.1X-controlled port unblocked for unicast trac 248 chapter 7 / Wireless Network Security PMK, verify the selection of the cipher suite, and derive a fresh PTK for the following data session. The four parts of the exchange are as follows. ■■ AP S STA: Message includes the MAC address of the AP and a nonce (Anonce). ■■ STA S AP: The STA generates its own nonce (Snonce) and uses both nonces and both MAC addresses, plus the PMK, to generate a PTK. The STA then sends a message containing its MAC address and Snonce, enabling the AP to generate the same PTK. This message includes a message integrity code (MIC)[2] using HMAC-MD5 or HMAC-SHA-1-128. The key used with the MIC is KCK. ■■ AP S STA: The AP is now able to generate the PTK. The AP then sends a message to the STA, containing the same information as in the first message, but this time including a MIC. ■■ STA S AP: This is merely an acknowledgment message, again protected by a MIC. Group Key Distribution For group key distribution, the AP generates a GTK and distributes it to each STA in a multicast group. The two-message exchange with each STA consists of the following: ■■ AP S STA: This message includes the GTK, encrypted either with RC4 or with AES. The key used for encryption is KEK. A MIC value is appended. ■■ STA S AP: The STA acknowledges receipt of the GTK. This message includes a MIC value. Protected Data Transfer Phase IEEE 802.11i defines two schemes for protecting data transmitted in 802.11 MPDUs: the Temporal Key Integrity Protocol (TKIP), and the Counter Mode-CBC MAC Protocol (CCMP). TKIP TKIP is designed to require only software changes to devices that are implemented with the older wireless LAN security approach called Wired Equivalent Privacy (WEP). TKIP provides two services: ■■ Message integrity: TKIP adds a message integrity code (MIC) to the 802.11 MAC frame after the data field. The MIC is generated by an algorithm, called Michael, that computes a 64-bit value using as input the source and destination MAC address values and the Data field, plus key material. ■■ Data confidentiality: Data confidentiality is provided by encrypting the MPDU plus MIC value using RC4. 2 While MAC is commonly used in cryptography to refer to a Message Authentication Code, the term MIC is used instead in connection with 802.11i because MAC has another standard meaning, Media Access Control, in networking. 7.4 / IEEE 802.11i Wireless Lan Security 249 The 256-bit TK (Figure 7.9) is employed as follows. Two 64-bit keys are used with the Michael message digest algorithm to produce a message integrity code. One key is used to protect STA-to-AP messages, and the other key is used to protect AP-to-STA messages. The remaining 128 bits are truncated to generate the RC4 key used to encrypt the transmitted data. For additional protection, a monotonically increasing TKIP sequence counter (TSC) is assigned to each frame. The TSC serves two purposes. First, the TSC is included with each MPDU and is protected by the MIC to protect against replay attacks. Second, the TSC is combined with the session TK to produce a dynamic encryption key that changes with each transmitted MPDU, thus making cryptanalysis more difficult. CCMP CCMP is intended for newer IEEE 802.11 devices that are equipped with the hardware to support this scheme. As with TKIP, CCMP provides two services: ■■ Message integrity: CCMP uses the cipher block chaining message authentication code (CBC-MAC), described in Chapter 3. ■■ Data confidentiality: CCMP uses the CTR block cipher mode of operation with AES for encryption. CTR is described in Chapter 2. The same 128-bit AES key is used for both integrity and confidentiality. The scheme uses a 48-bit packet number to construct a nonce to prevent replay attacks. The IEEE 802.11i Pseudorandom Function At a number of places in the IEEE 802.11i scheme, a pseudorandom function (PRF) is used. For example, it is used to generate nonces, to expand pairwise keys, and to generate the GTK. Best security practice dictates that different pseudorandom number streams be used for these different purposes. However, for implementation efficiency, we would like to rely on a single pseudorandom number generator function. The PRF is built on the use of HMAC-SHA-1 to generate a pseudorandom bit stream. Recall that HMAC-SHA-1 takes a message (block of data) and a key of length at least 160 bits and produces a 160-bit hash value. SHA-1 has the property that the change of a single bit of the input produces a new hash value with no apparent connection to the preceding hash value. This property is the basis for pseudorandom number generation. The IEEE 802.11i PRF takes four parameters as input and produces the desired number of random bits. The function is of the form PRF(K, A, B, Len), where K = a secret key A = a text string specific to the application (e.g., nonce generation or pairwise key expansion) B = some data specific to each case Len = desired number of pseudorandom bits For example, for the pairwise transient key for CCMP: PTK = PRF (PMK, “Pairwise key expansion”, min (APAddr, STA-Addr) || max (AP-Addr, STA-Addr) || min (Anonce, Snonce) || max (Anonce, Snonce), 384) 250 chapter 7 / Wireless Network Security So, in this case, the parameters are K = PMK A = the text string “Pairwise key expansion” B = a sequence of bytes formed by concatenating the two MAC addresses and the two nonces Len = 384 bits Similarly, a nonce is generated by Nonce = PRF (Random Number, “InitCounter”, MAC || Time, 256) where Time is a measure of the network time known to the nonce generator. The group temporal key is generated by GTK = PRF (GMK, “Group key expansion”, MAC || Gnonce, 256) Figure 7.11 illustrates the function PRF(K, A, B, Len). The parameter K serves as the key input to HMAC. The message input consists of four items concatenated together: the parameter A, a byte with value 0, the parameter B, and a counter i. The counter is initialized to 0. The HMAC algorithm is run once, producing a 160-bit hash value. If more bits are required, HMAC is run again with the same inputs, except that i is incremented each time until the necessary number of bits is generated. We can express the logic as PRF (K, A, B, Len) R S null string for i S 0 to ((Len + 159)/160 − 1) do R SR || HMAC-SHA-1 (K, A || 0 || B || i) Return Truncate-to-Len (R, Len) Figure 7.11 IEEE 802.11i Pseudorandom Function HMAC-SHA-1 | | K A 0 B i R = HMAC-SHA-1(K, A || 0 || B || i) + 1 7.5 / Key Terms, Review Questions, And Problems 251 7.5 Key Terms, Review Questions, And Problems Key Terms 4-way handshake access point (AP) basic service set (BSS) Counter Mode-CBC MAC Protocol (CCMP) distribution system (DS) extended service set (ESS) group keys IEEE 802.1X IEEE 802.11 IEEE 802.11i independent BSS (IBSS) logical link control (LLC) media access control (MAC) MAC protocol data unit (MPDU) MAC service data unit (MSDU) message integrity code (MIC) Michael pairwise keys pseudorandom function Robust Security Network (RSN) Temporal Key Integrity Protocol (TKIP) Wi-Fi Wi-Fi Protected Access (WPA) Wired Equivalent Privacy (WEP) Wireless LAN (WLAN) Review Questions 7.1 What is the basic building block of an 802.11 WLAN? 7.2 List and briefly define threats to a wireless network. 7.3 List and briefly define IEEE 802.11 services. 7.4 List some security threats related to mobile devices. 7.5 How is the concept of an association related to that of mobility? 7.6 What security areas are addressed by IEEE 802.11i? 7.7 Briefly describe the five IEEE 802.11i phases of operation. 7.8 What is the difference between TKIP and CCMP? Problems 7.1 In IEEE 802.11, open system authentication simply consists of two communications. An authentication is requested by the client, which contains the station ID (typically the MAC address). This is followed by an authentication response from the AP/router containing a success or failure message. An example of when a failure may occur is if the client’s MAC address is explicitly excluded in

the AP/router configuration. a. What are the benefits of this authentication scheme? b. What are the security vulnerabilities of this authentication scheme? 7.2 Prior to the introduction of IEEE 802.11i, the security scheme for IEEE 802.11 was Wired Equivalent Privacy (WEP). WEP assumed all devices in the network share a secret key. The purpose of the authentication scenario is for the STA to prove that it possesses the secret key. Authentication proceeds as shown in Figure 7.12. The STA sends a message to the AP requesting authentication. The AP issues a challenge, which is a sequence of 128 random bytes sent as plaintext. The STA encrypts the challenge with the shared key and returns it to the AP. The AP decrypts the incoming value and compares it to the challenge that it sent. If there is a match, the AP confirms that authentication has succeeded. a. What are the benefits of this authentication scheme? b. This authentication scheme is incomplete. What is missing and why is this important? Hint: The addition of one or two messages would fix the problem. c. What is a cryptographic weakness of this scheme? 252 chapter 7 / Wireless Network Security 7.3 For WEP, data integrity and data confidentiality are achieved using the RC4 stream encryption algorithm. The transmitter of an MPDU performs the following steps, referred to as encapsulation: 1. The transmitter selects an initial vector (IV) value. 2. The IV value is concatenated with the WEP key shared by transmitter and receiver to form the seed, or key input, to RC4. 3. A 32-bit cyclic redundancy check (CRC) is computed over all the bits of the MAC data field and appended to the data field. The CRC is a common error-detection code used in data link control protocols. In this case, the CRC serves as a integrity check value (ICV). 4. The result of step 3 is encrypted using RC4 to form the ciphertext block. 5. The plaintext IV is prepended to the ciphertext block to form the encapsulated MPDU for transmission. a. Draw a block diagram that illustrates the encapsulation process. b. Describe the steps at the receiver end to recover the plaintext and perform the integrity check. c. Draw a block diagram that illustrates part b. 7.4 A potential weakness of the CRC as an integrity check is that it is a linear function. This means that you can predict which bits of the CRC are changed if a single bit of the message is changed. Furthermore, it is possible to determine which combination of bits could be flipped in the message so that the net result is no change in the CRC. Thus, there are a number of combinations of bit flippings of the plaintext message that leave the CRC unchanged, so message integrity is defeated. However, in WEP, if an attacker does not know the encryption key, the attacker does not have access to the plaintext, only to the ciphertext block. Does this mean that the ICV is protected from the bit flipping attack? Explain. Figure 7.12 WEP Authentication; refer to Problem 7.2 STA AP Request Station sends a request for authentication AP sends challenge message containing 128-bit random number AP decrypts challenge response. If match, send authentication success message Station responds with encrypted version of challenge number Response Challenge Success 253 8.1 Internet Mail Architecture 8.2 E-mail Formats 8.3 E-mail Threats and Comprehensive E-mail Security 8.4 S/MIME 8.5 Pretty Good Privacy 8.6 DNSSEC 8.7 Dns-Based Authentication of Named Entities 8.8 Sender Policy Framework 8.9 DomainKeys Identified Mail 8.10 Domain-Based Message Authentication, Reporting, and Conformance 8.11 Key Terms, Review Questions, and Problems Chapter Electronic Mail Security 254 chapter 8 / Electronic Mail Security 8.1 Internet Mail Architecture For an understanding of the topics in this chapter, it is useful to have a basic grasp of the Internet mail architecture, which is currently defined in RFC 5598 (Internet Mail Architecture, July 2009). This section provides an overview of the basic concepts. E-mail Components At its most fundamental level, the Internet mail architecture consists of a user world in the form of Message User Agents (MUA), and the transfer world, in the form

of the Message Handling Service (MHS), which is composed of Message Transfer Agents (MTA). The MHS accepts a message from one user and delivers it to one or more other users, creating a virtual MUA-to-MUA exchange environment. This architecture involves three types of interoperability. One is directly between users: messages must be formatted by the MUA on behalf of the message author so that Learning Objectives After studying this chapter, you should be able to: ◆◆ Summarize the key functional components of the Internet mail architecture. ◆◆ Explain the basic functionality of SMTP, POP3, and IMAP. ◆◆ Explain the need for MIME as an enhancement to ordinary e-mail. ◆◆ Describe the key elements of MIME. ◆◆ Understand the functionality of S/MIME and the security threats it addresses. ◆◆ Understand the basic mechanisms of STARTTLS and its role in e-mail security. ◆◆ Understand the basic mechanisms of DANE and its role in e-mail security. ◆◆ Understand the basic mechanisms of SPF and its role in e-mail security. ◆◆ Understand the basic mechanisms of DKIM and its role in e-mail security. ◆◆ Understand the basic mechanisms of DMARC and its role in e-mail security. In virtually all distributed environments, electronic mail is the most heavily used network-based application. Users expect to be able to, and do, send e-mail to others who are connected directly or indirectly to the Internet, regardless of host operating system or communications suite. With the explosively growing reliance on e-mail, there grows a demand for authentication and confidentiality services. Two schemes stand out as approaches that enjoy widespread use: Pretty Good Privacy (PGP) and S/MIME. Both are examined in this chapter. This chapter concludes with a discussion of DomainKeys Identified Mail. 8.1 / Internet Mail Architecture 255 the message can be displayed to the message recipient by the destination MUA. There are also interoperability requirements between the MUA and the MHS— first when a message is posted from an MUA to the MHS and later when it is delivered from the MHS to the destination MUA. Interoperability is required among the MTA components along the transfer path through the MHS. Figure 8.1 illustrates the key components of the Internet mail architecture, which include the following. ■ Message User Agent (MUA): Operates on behalf of user actors and user applications. It is their representative within the e-mail service. Typically, this function is housed in the user’s computer and is referred to as a client e-mail program or a local network e-mail server. The author MUA formats a message and performs initial submission into the MHS via a MSA. The recipient MUA processes received mail for storage and/or display to the recipient user. ■ Mail Submission Agent (MSA): Accepts the message submitted by an MUA and enforces the policies of the hosting domain and the requirements of Internet standards. This function may be located together with the MUA or Figure 8.1 Function Modules and Standardized Protocols Used between them in the Internet Mail Architecture Message user agent (MUA) Message author Message recipient ESMTP (Submission) SMTP SMTP SMTP ESMTP (Submission) (SMTP, local) (IMAP, POP, local) Mail submission agent (MSA) Message transfer agent (MTA) Message transfer agent (MTA) MESSAGE HANDLING SYSTEM (MHS) Message transfer agent (MTA) Mail delivery agent (MDA) Message store (MS) Message user agent (MUA) 256 chapter 8 / Electronic Mail Security as a separate functional model. In the latter case, the Simple Mail Transfer Protocol (SMTP) is used between the MUA and the MSA. ■ Message Transfer Agent (MTA): Relays mail for one application-level hop. It is like a packet switch or IP router in that its job is to make routing assessments and to move the message closer to the recipients. Relaying is performed by a sequence of MTAs until the message reaches a destination MDA. An MTA also adds trace information to the message header. SMTP is used between MTAs and between an MTA and an MSA or MDA. ■ Mail Delivery Agent (MDA): Responsible for transferring the message from the MHS to the MS. ■ Message Store (MS): An MUA can employ a long-term MS. An MS can be located on a remote server or on the same machine as the MUA. Typically, an MUA retrieves messages from a remote server using POP (Post Office Protocol) or IMAP (Internet Message Access Protocol). Two other concepts need to be defined. An administrative management domain (ADMD) is an Internet e-mail provider. Examples include a department that operates a local mail relay (MTA), an IT department that operates an enterprise mail relay, and an ISP that operates a public shared e-mail service. Each ADMD can have different operating policies and trust-based decision making. One obvious example is the distinction between mail that is exchanged within an organization and mail that is exchanged between independent organizations. The rules for handling the two types of traffic tend to be quite different. The Domain Name System (DNS) is a directory lookup service that provides a mapping between the name of a host on the Internet and its numerical address. DNS is discussed subsequently in this chapter. E-mail Protocols Two types of protocols are used for transferring e-mail. The first type is used to move messages through the Internet from source to destination. The protocol used for this purpose is SMTP, with various extensions and in some cases restrictions. The second type consists of protocols used to transfer messages between mail servers, of which IMAP and POP are the most commonly used. Simple Mail Transfer Protocol SMTP encapsulates an e-mail message in an envelope and is used to relay the encapsulated messages from source to destination through multiple MTAs. SMTP was originally specified in 1982 as RFC 821 and has undergone several revisions, the most current being RFC 5321 (October 2008). These revisions have added additional commands and introduced extensions. The term Extended SMTP (ESMTP) is often used to refer to these later versions of SMTP. SMTP is a text-based client-server protocol where the client (e-mail sender) contacts the server (next-hop recipient) and issues a set of commands to tell the server about the message to be sent, then sending the message itself. The majority of these commands are ASCII text messages sent by the client and a resulting return code (and additional ASCII text) returned by the server. 8.1 / Internet Mail Architecture 257 The transfer of a message from a source to its ultimate destination can occur over a single SMTP client/server conversation over a single TCP connection. Alternatively, an SMTP server may be an intermediate relay that assumes the role of an SMTP client after receiving a message and then forwards that message to an SMTP server along a route to the ultimate destination. The operation of SMTP consists of a series of commands and responses exchanged between the SMTP sender and receiver. The initiative is with the SMTP sender, who establishes the TCP connection. Once the connection is established, the SMTP sender sends commands over the connection to the receiver. Each command consists of a single line of text, beginning with a four-letter command code followed in some cases by an argument field. Each command generates exactly one reply from the SMTP receiver. Most replies are a single-line, although multiple-line replies are possible. Each reply begins with a three-digit code and may be followed by additional information. Figure 8.2 illustrates the SMTP exchange between a client (C) and server (S). The interchange begins with the client establishing a TCP connection to TCP port 25 on the server (not shown in figure). This causes the server to activate SMTP and send a 220 reply to the client. The HELO command identifies the sending domain, which the server acknowledges and accepts with a 250 reply. The SMTP sender is transmitting mail that originates with the user Smith@bar.com. The MAIL command identifies the originator of the message. The message is addressed to three users on machine foo.com, namely, Jones, Green, and Brown. The client S: 220 foo.com Simple Mail Transfer Service Ready C: HELO bar.com S:

250 OK C: MAIL FROM: S: 250 OK C: RCPT TO: S: 250 OK C: RCPT TO: S: 550 No such user here C: RCPT TO: S: 250 OK C: DATA S: 354 Start mail input; end with . C: Blah blah blah . . . C: . . . etc. etc. etc. C: S: 250 OK C: QUIT S: 221 foo.com Service closing transmission channel Figure 8.2 Example SMTP Transaction Scenario 258 chapter 8 / Electronic Mail Security identifies each of these in a separate RCPT command. The SMTP receiver indicates that it has mailboxes for Jones and Brown but does not have information on Green. Because at least one of the intended recipients has been verified, the client proceeds to send the text message, by first sending a DATA command to ensure the server is ready for the data. After the server acknowledges receipt of all the data, it issues a 250 OK message. Then the client issues a QUIT command and the server closes the connection. A significant security-related extension for SMTP, called STARTTLS, is defined in RFC 3207 (SMTP Service Extension for Secure SMTP over Transport Layer Security, February 2002). STARTTLS enables the addition of confidentiality and authentication in the exchange between SMTP agents. This gives SMTP agents the ability to protect some or all of their communications from eavesdroppers and attackers. If the client does initiate the connection over a TLS-enabled port (e.g., port 465 was previously used for SMTP over SSL), the server may prompt with a message indicating that the STARTTLS option is available. The client can then issue the STARTTLS command in the SMTP command stream, and the two parties proceed to establish a secure TLS connection. An advantage of using STARTTLS is that the server can offer SMTP service on a single port, rather than requiring separate port numbers for secure and cleartext operations. Similar mechanisms are available for running TLS over IMAP and POP protocols. Historically, MUA/MSA message transfers have used SMTP. The standard currently preferred is SUBMISSION, defined in RFC 6409 (Message Submission for Mail, November 2011). Although SUBMISSION derives from SMTP, it uses a separate TCP port and imposes distinct requirements, such as access authorization. Mail Access Protocols (POP3, IMAP) Post Office Protocol (POP3) allows an e-mail client (user agent) to download an e-mail from an e-mail server (MTA). POP3 user agents connect via TCP to the server (typically port 110). The user agent enters a username and password (either stored internally for convenience or entered each time by the user for stronger security). After authorization, the UA can issue POP3 commands to retrieve and delete mail. As with POP3, Internet Mail Access Protocol (IMAP) also enables an e-mail client to access mail on an e-mail server. IMAP also uses TCP, with server TCP port 143. IMAP is more complex than POP3. IMAP provides stronger authentication than POP3 and provides other functions not supported by POP3. 8.2 E-mail Formats To understand S/MIME, we need first to have a general understanding of the underlying e-mail format that it uses, namely, MIME. But to understand the significance of MIME, we need to go back to the traditional e-mail format standard, RFC 822, which is still in common use. The most recent version of this format specification is RFC 5322 (Internet Message Format, October 2008). Accordingly, this section first provides an introduction to these two earlier standards and then moves on to a discussion of S/MIME. 8.2 / E-mail Formats 259 RFC 5322 RFC 5322 defines a format for text messages that are sent using electronic mail. It has been the standard for Internet-based text mail messages and remains in common use. In the RFC 5322 context, messages are viewed as having an envelope and contents. The envelope contains whatever information is needed to accomplish transmission and delivery. The contents compose the object to be delivered to the recipient. The RFC 5322 standard applies only to the contents. However, the content standard includes a set of header fields that may be used by the mail system to create the envelope, and the standard is intended to facilitate the acquisition of such information by programs. The

overall structure of a message that conforms to RFC 5322 is very simple. A message consists of some number of header lines (the header) followed by unrestricted text (the body). The header is separated from the body by a blank line. Put differently, a message is ASCII text, and all lines up to the first blank line are assumed to be header lines used by the user agent part of the mail system. A header line usually consists of a keyword, followed by a colon, followed by the keyword's arguments; the format allows a long line to be broken up into several lines. The most frequently used keywords are From, To, Subject, and Date. Here is an example message: Date: October 8, 2009 2:15:49 PM EDT From: "William Stallings" Subject: The Syntax in RFC 5322 To: Smith@Other-host.com Cc: Jones@Yet-Another-Host.com Hello. This section begins the actual message body, which is delimited from the message heading by a blank line. Another field that is commonly found in RFC 5322 headers is Message-ID. This field contains a unique identifier associated with this message. Multipurpose Internet Mail Extensions Multipurpose Internet Mail Extension (MIME) is an extension to the RFC 5322 framework that is intended to address some of the problems and limitations of the use of Simple Mail Transfer Protocol (SMTP) or some other mail transfer protocol and RFC 5322 for electronic mail. RFCs 2045 through 2049 define MIME, and there have been a number of updating documents since then. As justification for the use of MIME, [PARZ06] lists the following limitations of the SMTP/5322 scheme. 260 chapter 8 / Electronic Mail Security 1. SMTP cannot transmit executable files or other binary objects. A number of schemes are in use for converting binary files into a text form that can be used by SMTP mail systems, including the popular UNIX UUencode/ UUdecode scheme. However, none of these is a standard or even a de facto standard. 2. SMTP cannot transmit text data that includes national language characters, because these are represented by 8-bit codes with values of 128 decimal or higher, and SMTP is limited to 7-bit ASCII. 3. SMTP servers may reject mail message over a certain size. 4. SMTP gateways that translate between ASCII and the character code EBCDIC do not use a consistent set of mappings, resulting in translation problems. 5. SMTP gateways to X.400 electronic mail networks cannot handle nontextual data included in X.400 messages. 6. Some SMTP implementations do not adhere completely to the SMTP standards defined in RFC 821. Common problems include: — Deletion, addition, or reordering of carriage return and linefeed —Truncating or wrapping lines longer than 76 characters —Removal of trailing white space (tab and space characters) —Padding of lines in a message to the same length —Conversion of tab characters into multiple space characters MIME is intended to resolve these problems in a manner that is compatible with existing RFC 5322 implementations. Overview The MIME specification includes the following elements. 1. Five new message header fields are defined, which may be included in an RFC 5322 header. These fields provide information about the body of the message. 2. A number of content formats are defined, thus standardizing representations that support multimedia electronic mail. 3. Transfer encodings are defined that enable the conversion of any content format into a form that is protected from alteration by the mail system. In this subsection, we introduce the five message header fields. The next two subsections deal with content formats and transfer encodings. The five header fields defined in MIME are as follows: ■ MIME-Version: Must have the parameter value 1.0. This field indicates that the message conforms to RFCs 2045 and 2046. ■ Content-Type: Describes the data contained in the body with sufficient detail that the receiving user agent can pick an appropriate agent or mechanism to represent the data to the user or otherwise deal with the data in an appropriate manner. 8.2 / E-mail Formats 261 ■ Content-Transfer-Encoding: Indicates the type of transformation that has been used to represent the body of the message in a way that is acceptable for mail transport. ■ Content-ID: Used to identify MIME entities uniquely in multiple contexts. ■ Content-Description: A text description of the object with the body; this is useful when the object is not readable (e.g., audio data). Any or all of these fields may appear in a normal RFC 5322 header. A compliant implementation must support the MIME-Version, Content-Type, and ContentTransfer-Encoding fields; the Content-ID and Content-Description fields are optional and may be ignored by the recipient implementation. MIME Content Types The bulk of the MIME specification is concerned with the definition of a variety of content types. This reflects the need to provide standardized ways of dealing with a wide variety of information representations in a multimedia environment. Table 8.1 lists the content types specified in RFC 2046. There are seven different major types of content and a total of 15 subtypes. In general, a content type declares the general type of data, and the subtype specifies a particular format for that type of data. Type Subtype Description Text Plain Unformatted text; may be ASCII or ISO 8859. Enriched Provides greater format flexibility. Multipart Mixed The different parts are independent but are to be transmitted together. They should be presented to the receiver in the order that they appear in the mail message. Parallel Differs from Mixed only in that no order is defined for delivering the parts to the receiver. Alternative The different parts are alternative versions of the same information. They are ordered in increasing faithfulness to the original, and the recipient's mail system should display the "best" version to the user. Digest Similar to Mixed, but the default type/subtype of each part is message/rfc822. Message rfc822 The body is itself an encapsulated message that conforms to RFC 822. Partial Used to allow fragmentation of large mail items, in a way that is transparent to the recipient. External-body Contains a pointer to an object that exists elsewhere. Image jpeg The image is in JPEG format, JFIF encoding. gif The image is in GIF format. Video mpeg MPEG format. Audio Basic Single-channel 8-bit ISDN m-law encoding at a sample rate of 8 kHz. Application PostScript Adobe Postscript format. octet-stream General binary data consisting of 8-bit bytes. Table 8.1 MIME Content Types 262 chapter 8 / Electronic Mail Security For the text type of body, no special software is required to get the full meaning of the text aside from support of the indicated character set. The primary subtype is plain text, which is simply a string of ASCII characters or ISO 8859 characters. The enriched subtype allows greater formatting flexibility. The multipart type indicates that the body contains multiple, independent parts. The Content-Type header field includes a parameter (called boundary) that defines the delimiter between body parts. This boundary should not appear in any parts of the message. Each boundary starts on a new line and consists of two hyphens followed by the boundary value. The final boundary, which indicates the end of the last part, also has a suffix of two hyphens. Within each part, there may be an optional ordinary MIME header. Here is a simple example of a multipart message containing two parts—both consisting of simple text (taken from RFC 2046): From: Nathaniel Borenstein To: Ned Freed Subject: Sample message MIME-Version: 1.0 Content-type: multipart/mixed; boundary="simple boundary" This is the preamble. It is to be ignored, though it is a handy place for mail composers to include an explanatory note to non-MIME conformant readers. —simple boundary This is implicitly typed plain ASCII text. It does NOT end with a linebreak. —simple boundary Content-type: text/plain; charset=us-ascii This is explicitly typed plain ASCII text. It DOES end with a linebreak. —simple boundary— This is the epilogue. It is also to be ignored. There are four subtypes of the multipart type, all of which have the same overall syntax. The multipart/mixed subtype is used when there are multiple independent body parts that need to be bundled in a particular order. For the multipart/ parallel subtype, the order of the parts is not significant. If the recipient's system is appropriate, the multiple parts can be presented in parallel. For example, a picture or text part could be accompanied by a voice commentary that is played while the picture or text is displayed. For the multipart/alternative subtype, the various parts are different representations of the same information. The following is an example: From: Nathaniel Borenstein To: Ned Freed Subject: Formatted text mail 8.2 / E-mail Formats 263 MIME-Version: 1.0 Content-Type: multipart/alternative; boundary=boundary42 —boundary42 Content-Type: text/plain; charset=us-ascii . . . plain text version of message goes here. . . . —boundary42 Content-Type: text/enriched . . . RFC 1896 text/enriched version of same message goes here . . . —boundary42— In this subtype, the body parts are ordered in terms of increasing preference. For this example, if the recipient system is capable of displaying the message in the text/enriched format, this is done; otherwise, the plain text format is used. The multipart/digest subtype is used when each of the body parts is interpreted as an RFC 5322 message with headers. This subtype enables the construction of a message whose parts are individual messages. For example, the moderator of a group might collect e-mail messages from participants, bundle these messages, and send them out in one encapsulating MIME message. The message type provides a number of important capabilities in MIME. The message/rfc822 subtype indicates that the body is an entire message, including header and body. Despite the name of this subtype, the encapsulated message may be not only a simple RFC 5322 message, but also any MIME message. The message/partial subtype enables fragmentation of a large message into a number of parts, which must be reassembled at the destination. For this subtype, three parameters are specified in the Content-Type: Message/Partial field: an id common to all fragments of the same message, a sequence number unique to each fragment, and the total number of fragments. The message/external-body subtype indicates that the actual data to be conveyed in this message are not contained in the body. Instead, the body contains the information needed to access the data. As with the other message types, the message/external-body subtype has an outer header and an encapsulated message with its own header. The only necessary field in the outer header is the Content-Type field, which identifies this as a message/external-body subtype. The inner header is the message header for the encapsulated message. The Content-Type field in the outer header must include an access-type parameter, which indicates the method of access, such as FTP (file transfer protocol). The application type refers to other kinds of data, typically either uninterpreted binary data or information to be processed by a mail-based application. MIME Transfer Encodings The other major component of the MIME specification, in addition to content type specification, is a definition of transfer encodings for message bodies. The objective is to provide reliable delivery across the largest range of environments. 264 chapter 8 / Electronic Mail Security The MIME standard defines two methods of encoding data. The ContentTransfer-Encoding field can actually take on six values, as listed in Table 8.2. However, three of these values (7-bit, 8-bit, and binary) indicate that no encoding has been done but provide some information about the nature of the data. For SMTP transfer, it is safe to use the 7-bit form. The 8-bit and binary forms may be usable in other mail transport contexts. Another Content-Transfer-Encoding value is x-token, which indicates that some other encoding scheme is used for which a name is to be supplied. This could be a vendor-specific or application-specific scheme. The two actual encoding schemes defined are quoted-printable and base64. Two schemes are defined to provide a choice between a transfer technique that is essentially human readable and one that is safe for all types of data in a way that is reasonably compact. The quoted-printable transfer encoding is useful when the data consists largely of octets that correspond to printable ASCII characters. In essence, it represents nonsafe characters by the hexadecimal representation of their code and introduces reversible (soft) line breaks to limit message lines to 76 characters. The base64 transfer encoding, also known as radix-64 encoding, is a common one for encoding arbitrary binary data in such a way as to be invulnerable to the processing by mail-transport programs. It is also used in PGP and is described in Appendix H. A Multipart Example Figure 8.3, taken from RFC 2045, is the outline of a complex multipart message. The message has five parts to be displayed serially: two introductory plain text parts, an embedded multipart message, a richtext part, and a closing encapsulated text message in a non-ASCII character set. The embedded multipart message has two parts to be displayed in parallel: a picture and an audio fragment. Canonical Form An important concept in MIME and S/MIME is that of canonical form. Canonical form is a format, appropriate to the content type, that is standardized for use between systems. This is in contrast to native form, which is a format that may be peculiar to a particular system. RFC 2049 defines these two forms as follows: ■ Native form: The body to be transmitted is created in the system's native format. The native character set is used and, where appropriate, local end-of-line conventions are used as well. The body may be any format that corresponds to 7 bit The data are all represented by short lines of ASCII characters. 8 bit The lines are short, but there may be non-ASCII characters (octets with the high-order bit set). binary Not only may non-ASCII characters be present but the lines are not necessarily short enough for SMTP transport. quoted-printable Encodes the data in such a way that if the data being encoded are mostly ASCII text, the encoded form of the data remains largely recognizable by humans. base64 Encodes data by mapping 6-bit blocks of input to 8-bit blocks of output, all of which are printable ASCII characters. x-token A named nonstandard encoding. Table 8.2 MIME Transfer Encodings 8.2 / E-mail Formats 265 MIME-Version: 1.0 From: Nathaniel Borenstein To: Ned Freed Subject: A multipart example Content-Type: multipart/mixed; boundary=unique-boundary-1 This is the preamble area of a multipart message. Mail readers that understand multipart format should ignore this preamble. If you are reading this text, you might want to consider changing to a mail reader that understands how to properly display multipart messages. —unique-boundary-1 . . . Some text appears here . . . [Note that the preceding blank line means no header fields were given and this is text, with charset US ASCII. It could have been done with explicit typing as in the next part.] —unique-boundary-1 Content-type: text/plain; charset=US-ASCII This could have been part of the previous part, but illustrates explicit versus implicit typing of body parts. —unique-boundary-1 Content-Type: multipart/parallel; boundary=unique-boundary-2 —unique-boundary-2 Content-Type: audio/basic Content-Transfer-Encoding: base64 . . . base64-encoded 8000 Hz single-channel mu-law-format audio data goes here . . . . —unique-boundary-2 Content-Type: image/jpeg Content-Transfer-Encoding: base64 . . . base64-encoded image data goes here . . . . —unique-boundary-2— —unique-boundary-1 Content-type: text/enriched This is richtext. as defined in RFC 1896 Isn't it cool? —unique-boundary-1 Content-Type: message/rfc822 From: (mailbox in US-ASCII) To: (address in US-ASCII) Subject: (subject in US-ASCII) Content-Type: Text/plain; charset=ISO-8859-1 Content-Transfer-Encoding: Quoted-printable . . . Additional text in ISO-8859-1 goes here . . . —unique-boundary-1— Figure 8.3 Example MIME Message Structure 266 chapter 8 / Electronic Mail Security the local model for the representation of some form of information. Examples include a UNIX-style text file, or a Sun raster image, or a VMS indexed file, and audio data in a system-dependent format stored only in memory. In