Wireless IP Telephony

Manish Marwah and Shivakant Mishra, *University of Colorado, Boulder*

INTRODUCTION

Internet protocol (IP) telephony or *voice over IP* (VoIP) has seen tremendous growth in the past few years and is predicted to grow rapidly in the coming years. For instance, according to the Web site ZDNet, total VoIP equipment revenue will rise some threefold to $11.9 billion by 2010 from $3.95 billion in 2005. Both business and residential customers are switching to IP telephony. The main attractions of IP telephony are (1) cost savings, (2) better resource utilization, (3) numerous compelling services and applications arising from wireless IP telephony and its integration with Internet-based applications, and (4) reduced network administration and management.

Similarly, *wireless local area network* (WLAN) technologies, mainly *wireless fidelity* (WiFi), have grown rapidly and are becoming ubiquitous. Considering the popularity of these two important technologies (namely, IP telephony and wireless LAN), wireless IP telephony has emerged as an attractive combination that reduces costs, increases access, and provides greater control over enhanced enterprise communications tools. Current literature shows that significant savings are realized with VoIP using WiFi: Cellular phone roaming at $1.25 per minute, for instance, can be reduced to $0.02 per minute using wireless IP telephony. In fact, "dual-mode" phones such as Nokia N80 and Motorola A910 that support both cellular and wireless IP telephony are available now. This allows organizations to save on cellular phone expenses as employees use IP telephony over wireless LAN when located on company premises. Furthermore, cellular networks and WiFi complement each other in their data-transfer rates and coverage areas. Although WiFi can support high data rates (currently, as high as 54 Mbps), cellular networks support only a maximum data rate of a few Mbps. However, cellular networks have especially good geographical coverage unlike WiFi, which is only available in limited areas (sometimes referred to as *hot spots*). In addition, cellular networks typically do not have good coverage inside buildings, and in such cases wireless IP telephony can provide better connectivity and voice quality.

There are a number of technological challenges associated with transporting real-time data such as voice and video over wireless. In this chapter, we will discuss the state of the art of wireless IP telephony. In particular, we will discuss issues related to (1) mobility management; (2) *quality of service* (QoS), including mechanisms for call admission control; (3) security and privacy; and (4) *enhanced 911* (E911). Because 802.11 (IEEE 1999) (or WiFi) is the most popular WLAN technology, we discuss most of these issues in the context of 802.11-based standards, which are briefly described below.

802.11 Standards

The most commonly used *layer 2* (L2) WLAN technology is based on the 802.11 (or WiFi) standards (IEEE 1999). This umbrella encompasses various formalized standards that define different aspects of the WiFi technology. The WiFi physical layer is defined by standards such as 802.11b (11 Mbps) and 802.11g (54 Mbps). The forthcoming 802.11n standard, which is now in the draft stage, aims to increase data throughputs to as much as 600 Mbps.

In addition to letter standards that specify data rates, other 802.11 standards relevant to wireless IP telephony are:

- 802.11i, which defines security mechanism such as authentication, encryption, and access control;
- 802.11e, which provides QoS support for real-time applications such as voice and video; and
- 802.11r (not approved yet), which aims to reduce handoff latency during transitions between access points such that time-sensitive applications such as IP telephony are not disrupted.

For a more detailed description of 802.11 standards, see Chapter 116.

Devices using 802.11 have two modes of operation: (1) ad hoc, by which the devices directly communicate with each other, and (2) infrastructure, by which the 802.11 devices communicate through an access point.

In the infrastructure mode, an 802.11 system consists of *stations* (STAs) and *access points* (APs) as shown in Figure 1. One AP and multiple STAs that receive service from that AP constitute a *basic service set* (BSS). Multiple

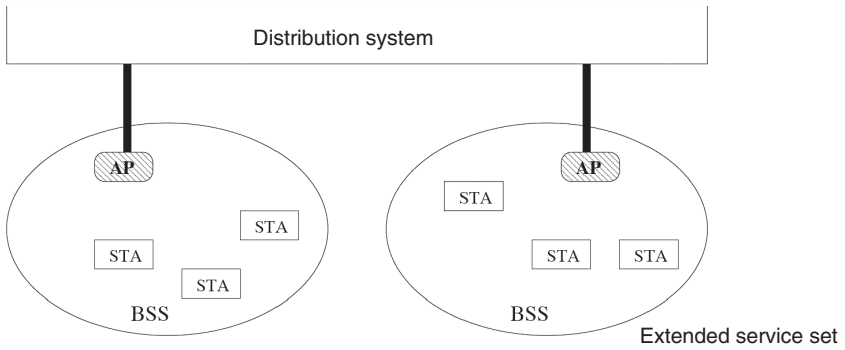

Figure 1: 802.11 infrastructure-mode architecture

BSSs connected through a wired network, such as a LAN, are referred to as an *extended service set*.

Because most 802.11 installations use infrastructure mode, in the rest of this chapter, we will assume that a 802.11 network operates in this mode (unless stated otherwise).

WiMAX and IP Telephony

WiMAX (802.16) is an upcoming standard for broadband wireless wide area network access. It has a range of approximately 6 miles radius and speeds of as high as 40 Mbps per channel. The initial version of WiMAX (also called *fixed WiMAX*) is aimed at stationary or nonmobile applications, which are not sensitive to longer handoff times; a later version (802.16e or *mobile WiMAX*) aims to provide support for mobility and real-time applications such as IP telephony. For further information on WiMAX, see Chapter 133.

Organization

This rest of this chapter is organized as follows. In the next section, we discuss mobility management at various levels of the network stack. This is followed by a section on quality of service issues pertinent to wireless IP telephony. In "Security and Privacy" below, we present those considerations. E911 is discussed in the section after that and before our conclusion.

MOBILITY MANAGEMENT

For IP telephony over wireless, mobility management has to be addressed at various levels. Standards set forth by the International Telecommunications Union (ITU) (see ITU-T 2003) recommend that for real-time voice communication, the one-way delay should be less than 150 ms. Latencies greater than this value result in poor voice quality and are discernible by a user. Handoffs between wireless APs must be done in a fast and seamless manner so that existing calls are not disrupted. In addition to L2 mobility, L3 (layer 3, the network layer) mobility is also important because the user could move to a different wireless subnet or network. Again, this must be done without any service disruption. Various mobility management techniques are discussed in Sharma, Zhu, and Chiueh 2004; Dutta et al. 2002; Ramani and Savage 2005; Mishra, Shin, and Arbaugh 2004; Park et al. 2004; Pack and Choi 2002.

Layer 2 Mobility

WiFi clients receive service through APs that typically provide coverage to distances up to 100 meters from an AP. Thus, a person using a WiFi-enabled device for a VoIP call could traverse the coverage area of several APs while walking from one area of a building to another. Furthermore, at any time, a WiFi device can only be associated with one AP (this limitation, which is part of the WiFi standards, enables packets to be efficiently routed to a WiFi device).

The process involved in changing the association from one AP to another is called a *handoff* or *handover*. A handoff usually occurs when a device moves out of the range of its current AP, causing its signal strength or signal-to-noise ratio to become unacceptable. Handoffs can be *hard* or *soft*. In a soft handoff, also referred to as a *make-before-break* handoff, the device first establishes connection with the new point of attachment before breaking the existing connection. This results in a more seamless handoff as compared to a hard handoff, also called a *break-before-make* handoff, in which the connection with the existing attachment point is broken before the new one is established.

Handoffs routinely happen during a call in cellular networks. There are two main reasons why handoffs in a cellular network are much smoother as compared to WiFi (Ramani and Savage 2005). First, in 802.11, the clients manage a handoff autonomously and independently without any knowledge of the network topology or other network characteristics. In cellular networks, by contrast, a *base station controller* or a *mobile switching center* initiates a handoff with full knowledge of the local topology and conditions such as the number of clients in the area. Furthermore, in cellular networks such as *code division multiple access* (CDMA) (Viterbi 1995), the phones perform a soft handoff in which the phone simultaneously communicates with multiple *base transceiver stations* (BTSs) before the handoff is executed to one of them. In 802.11, a handoff is hard.

Second, in cellular networks, a client continuously measures its connectivity with all the BTSs within its

range and assists the network in choosing the best alternative. However, in 802.11, a client only measures the quality of its current channel because it can only send and receive on a single channel at a time. When this quality goes below a threshold, the client switches to other channels to explore the possibility of a handoff to an AP with acceptable quality. (Note that while this switchover is in progress, frames on the current channel are dropped.) There is a trade-off involved in setting this voice quality threshold value. If it is too high, it is possible that the client does not find any other AP with better quality and thus unnecessarily causes a "gap" in the service because of the search for other APs. On the other hand, if the threshold is too low, the client may continue to be attached to an AP with poor signal quality even when a better quality AP is available. Unlike the cellular network, the time taken to perform handoffs in 802.11 networks is high, ranging from several hundred milliseconds to a second, which is unsuitable for real-time traffic such as voice.

The 802.11 standard provides excellent support for "portable" entities—that is, entities that migrate frequently but do not have strict connectivity requirements while they are in motion. However, because of the handoff mechanism used and the high handoff latency, 802.11 does not provide much support for "mobile" entities—that is, entities that have tight latency and response time requirements while they are in motion. We now provide a brief description of the L2 handoff process. This is followed by a discussion of the recent research aimed at improving the L2 handoff latency.

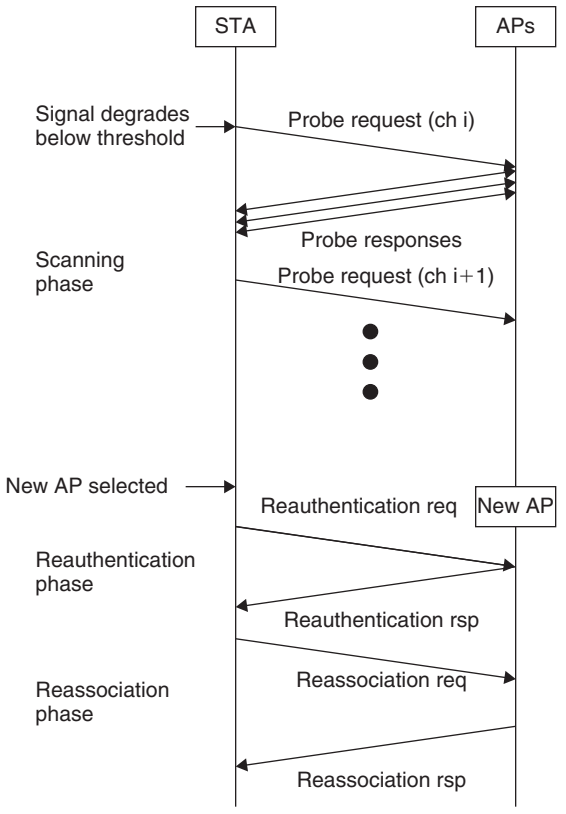

Figure 2: 802.11 handoff phases

L2 Handoff

A handoff is required as a station moves from one BSS to another. It is critical that the time taken for a handoff, or, the handoff latency, is minimal. As mentioned earlier, this latency should ideally be less than 150 ms. For seamless service, it is important that the handoff is done at the right time and to the right AP.

The steps involved in a handoff are described below and shown in Figure 2. It consists of three main phases: (1) discovery, (2) reauthentication, and, (3) reassociation.

Discovery. The discovery phase is entered when the signal quality on the current channel has degraded below a threshold level. The goal of this phase is to find another AP with acceptable signal quality. Depending on the version of 802.11 that the client is running, it switches to the corresponding channel frequencies one at a time. For each channel, the client either (1) waits for the periodic beacon (which is usually sent by an AP every 100 ms) to measure the quality of the corresponding AP's channel (*passive probing*) or (2) actively broadcasts a probe on the channel so that an AP operating in that channel replies back (*active probing*). Normally, active probing is much faster than passive probing.

The time taken by active probing is as follows (Ramani and Savage 2005):

$$\sum_{c=1}^{c=NumChannels} (1 - P(c)).MinChannelTime + P(c).MaxChannelTime$$

where $P(c)$ is the probability that an AP operates in channel c, MinChannelTime is the time the client waits if there is no AP operating in channel c (in other words, if it receives no response at all), and MaxChannelTime is the time the client waits if it receives at least one response (in order to allow other APs operating in that channel to respond). Suggested values of MinChannelTime vary from 1 ms (Velayos and Karlsson 2004) to 7 ms (Mishra, Shin, and Arbaugh 2003), whereas those of MaxChannelTime are approximately 11 ms (Velayos and Karlsson 2004; Mishra, Shin, and Arbaugh 2003). Using an average value of 4 ms for MinChannelTime and 11 ms for MaxChannelTime, and assuming eleven total channels (as is the case in 802.11b/g) gives the probing latency of from 40 ms (if no other AP is found) to 110 ms (if at least one AP is found on each channel). In addition, *channel switching delay* is incurred whenever a channel is changed. This can add a substantial amount of delay to the probing process. Ramani and Savage (2005) find that the probing delay using active probing for a popular 802.11b network interface card to be between 350 ms and 400 ms.

Reauthentication. After a new AP is selected, the client exchanges authentication messages with it. These establish the client's identity and its permission to use that particular AP. The reauthentication delay is approximately 20 ms (Mishra, Shin, and Arbaugh 2004).

Reassociation. After authentication, the client sends a reassociation request to the new AP, expecting a reassociation response in return to finalize the handoff. When an

AP receives a reassociation request, it exchanges station context information with the client's earlier AP. Interaccess point protocol or 802.11f is used for this exchange. The reassociation delay is on the order of approximately 20 ms.

The total time taken by the handoff process, which is the sum of the three phases, is approximately 500 ms to 1 sec (Ramani and Savage 2005). Furthermore, the probing delay is roughly 90 percent of the total taken (Mishra, Shin, and Arbaugh 2003).

Improving L2 Handoff Latency

An extensive body of research exists on enabling faster layer 2 handoffs in order to make WLANs more suitable for IP telephony (Ramani and Savage 2005; Brik, Mishra, and Banerjee 2005; Velayos and Karlsson 2004; Mishra, Shin, and Arbaugh 2004; Park et al. 2004; Shin et al. 2004; Pack and Choi 2002). Research aimed at reducing the time taken for each of the three phases of the handoff is presented in these papers. However, because the discovery phase takes almost 90 percent of the total handoff time, a number of researchers have suggested techniques to reduce the time taken for this phase.

In SyncScan (Ramani and Savage 2005), the authors propose synchronizing the clocks on the APs and the mobile devices so that the devices can predict the exact time when the periodic beacon from an AP will be sent. This would allow a device to continuously monitor the signals from all of the APs within its range. A concern with this technique is that the device will not be able to send or receive frames on its current channel while it reads beacons on the other channels. In other words, the frames sent or received on the current channel will be dropped during these times. This problem is addressed by using send and receive buffers. Whenever the device is tuned to other channels, the frames are buffered and later processed when the device returns to its current channel. Kim et al. (2004) construct a neighbor graph using network topology to narrow the list of channels, and only a few selected channels are probed. Furthermore, the authors propose algorithms that can be used to automatically generate a neighbor graph. Shin et al. (2004) use a combination of selective scanning and caching. In the selective scanning algorithm, the client remembers the channels on which replies were received on previous occasions. This reduces the handoff latency by 30 percent to 60 percent. The caching algorithm maintains a record of previous handoffs. Each entry in the cache table, keyed by an AP, stores as many as two other APs (and their corresponding channels) to which a handoff was performed from that AP. If a handoff needs to be performed, the cache table is checked first. If there is a cache hit and at least one of the two APs is still available, then the client does not need to perform scanning. However, if there is a cache miss, the selective scanning algorithm is used.

In Multiscan (Brik, Mishra, and Banerjee 2005), the authors propose to eliminate handoff latencies through use of additional client hardware. They suggest using two wireless interfaces, one of which is dedicated to looking for alternate APs. This continuous monitoring of the available APs allows the client to seamlessly handoff ongoing connections. Unlike SyncScan, Multiscan only requires changes to the clients; no changes are required to the AP infrastructure. Furthermore, unlike some techniques, Multiscan does not require any knowledge of the network topology.

In Mishra, Shin, and Arbaugh (2004), the authors aim to reduce the reassociation delay of the handoff process. Network topology information, stored as "neighbor graphs," is used to send the station context information ahead of time to APs adjacent to the client's current AP such that the context information is transferred to an AP before a client performs a handoff to it. Using this technique, the reassociation delay is reduced to 1.69 ms from 15.37 ms in the experiments performed by the authors. Although this is an order of magnitude improvement, it should be noted that the reassociation delay constitutes only a small part of the total handoff delay, and therefore the improvement made to the total handoff latency by this technique is not especially significant.

Pack and Choi (2002) examine ways to reduce the reauthentication delay. They propose to minimize this delay by having a client perform authentication (preauthentication) with multiple APs and not just its current AP.

Layer 3 Mobility

Layer 3 (or network layer) mobility is provided by the use of Mobile IP (Perkins 2002), which is an extension of IP to support mobile applications. It allows an Internet end point to retain its IP address (home address) irrespective of its point of attachment to the Internet. In other words, a host can roam across IP subnets or networks and still is contactable via its home address. This is accomplished by maintaining a home agent, whose location is fixed, and a foreign agent that registers on behalf of a mobile host. When the host is roaming, it additionally acquires a care-of address that is registered with the home agent. The host always uses its home address as the source address in any packet it sends out. The far end responds back to the host home address, which is received by the home agent. The home agent then tunnels it to the host's care-of address (located on the foreign agent). The foreign agent detunnels packets and delivers them to the mobile host. In some situations, to avoid issues with egress filters, the packets sent out by a mobile host may first be tunneled to the home agent and then sent to its intended destination. It should be noted that here the home agent acts much like a proxy for the mobile host.

For wireless IP telephony, Mobile IP allows greater mobility to end-point devices that can seamlessly move between different wireless IP subnets or networks without any disruption to ongoing calls.

Although support for network layer mobility as provided by Mobile IP is necessary for wireless IP telephony to work, it is not sufficient. Because of the strict latency requirements for voice, network layer handoff must be executed fast enough so that ongoing calls are not disrupted. In the context of Mobile IP, network layer handoff would be necessary whenever a mobile host moves to a different subnet or network. This would involve (1) tearing down the old tunnel, (2) acquiring a new care-of address, (3) registering this address with the home agent, and (4) establishing a new tunnel. Sharma, Zhu, and Chiueh (2004) discuss optimization of each of the phases involved in the

handoff and achieve a reduction in the network handoff latency by a factor of 5.

In some extensions of Mobile IP, several care-of addresses can be registered simultaneously. The home agent tunnels the packets to all these addresses. This is particularly useful in a wireless environment in which the host is rapidly changing its point of attachment to the Internet. Mobile IP is supported in IPv6 (Johnson, Perkins, and Arkko 2004) with some enhancements and better integration in TCP/IP stack as compared to IPv4. Among other advantages (Johnson, Perkins, and Arkko 2004), this obviates the need for a foreign agent. Even after a successful handoff, Mobile IP has drawbacks such as tunneling overhead and triangular routing, which is inefficient especially for delay-sensitive real-time traffic such as voice and video. To address this, Mobile IPv6 supports route optimization that allows a peer to send traffic directly to the care-of address, instead of the home address, of the mobile host (Johnson, Perkins, and Arkko 2004).

Application Layer Mobility

An alternative to supporting mobility management at the network layer is to handle it at the application layer, particularly in view of the inefficiencies incurred by network layer mechanisms such as Mobile IP. *Session initiation protocol* (SIP) can be used for performing a handoff at the application layer. SIP, described in Chapter 90, is the most popular and industry accepted protocol for call control in IP telephony.

Each mobile host belongs to a home region where a SIP proxy server provides registrar service. Whenever a host moves to another region, it communicates with the local SIP proxy. However, before providing service, the local (visited) proxy authenticates and registers the host with the host's home registrar. When a mobile host on an active call moves from one network to another, the following steps are involved in the handoff:

1. The host acquires a new IP address (e.g., by use of DHCP).
2. The host exchanges a set of messages with the visited SIP proxy for authentication and registration (note that the address of the SIP proxy server is included in the DHCP response).
3. After successful registration, the host sends a re-INVITE SIP message to the peer host using the call identifier of the original call to reestablish the call session. This is also referred to as *midcall mobility*, compared to *precall mobility* in which the host moves when there is no active call.

Because a number of messages need to be exchanged during handoff, the handoff delay can easily exceed 1 sec and thus is not suitable for real-time traffic (Nakajima et al. 2003). Several solutions based on SIP that minimize this delay have been proposed (Wang, Abu-Rgheff, and Akram 2004; Vali et al. 2003; Banerjee, Das, and Acharya 2006; Kwon, Gerla, and Das 2002). In Kwon, Gerla, and Das (2002), the authors propose using *shadow registration*. The key idea is that the host's current region a priori transfers authentication information related to that host to all of the neighboring regions in anticipation of the host moving to one of them. When the host does move to one of the neighboring regions, the registration step is much faster because the visited SIP registrar does not have to communicate to the host's home registrar. In Banerjee, Das, and Acharya (2006), the authors propose a soft handoff for SIP clients who have multiple network interfaces.

Vertical Handoff

A *vertical handoff* is one between heterogeneous networks—for example, between a cellular network and a WLAN. With the emergence of multimode phones (i.e., with multiple network interfaces), mobile users expect seamless operation and connectivity while moving from one network to another. For example, an employee using cellular service expects her call to seamlessly switch to wireless IP telephony once she is within the range of a WiFi service available on her company premises. A vertical handoff cannot be performed without application-level support. In Wu et al. (2005), a case study is performed of the delay incurred during vertical handoff in a WLAN–*universal mobile telephone service* (UMTS) internetwork using SIP as the mobility management protocol. The results show that a WLAN-to-UMTS handoff incurs a much larger delay (because of error-prone and bandwidth-limited wireless links) than a UMTS-to-WLAN handoff. It is suggested that soft handoff techniques be used to bring the handoff delay within acceptable limits.

QUALITY-OF-SERVICE ISSUES

Wireless IP telephony must provide same or better call quality as traditional telephony systems. Traditionally, IP (and wireless IP) has been a best effort service with each packet treated equally irrespective of its content. Although this works well for non–real-time applications such as file transfers and e-mail, real-time applications such as voice and video require strict performance guarantees from the network. If packet latency or delay, jitter, and loss exceed acceptable limits, then it has real user impact in terms of poor voice quality and dropped calls.

As real-time applications have become more prevalent on the TCP/IP networks, various QoS mechanisms have emerged. These mechanisms apply to various layers in the network stack. For the wired Ethernet MAC layer, the Institute of Electrical and Electronics Engineers' (IEEE's) 802.1p provides priority tagging. For the network layer, integrated services (Braden 1997) and differentiated services (Nichols 1998) provide QoS support. However, a QoS mechanism for 802.11 MAC had been lacking until 802.11e was approved in 2005.

The original 802.11 MAC defines two coordination functions for accessing the shared media: *distributed coordination function* (DCF) and *point coordination function* (PCF). DCF essentially uses carrier sense and collision avoidance (through use of the 802.11 RTS-CTS mechanism) for coordinating access. PCF provides channel access through polling by a centralized coordinator that usually resides at the AP. It is optional and only available in the infrastructure mode of operation (that is, when all communication to and from a device goes through an AP). Neither DCF nor PCF provides any QoS mechanisms, and all traffic has

the same priority. More detailed descriptions of DCF and PCF are presented elsewhere in this handbook.

Standard 802.11e

To overcome the lack of QoS support in 802.11, 802.11e was proposed. It provides QoS in two ways.

1. *Prioritized* QoS. 802.11e enhances the DCF channel access mechanism to support different priorities for different traffic categories (similar to DiffServ—see Nichols 1998). This enhanced MAC mechanism is called *enhanced DCF*. Before a frame is transmitted, a 802.11e station listens on a channel to determine if it is busy. This channel-sensing interval depends on the priority of the frame to be transmitted. In other words, the lower the priority of the frame, the larger the sensing time. This increases the probability of transmission of higher-priority frames compared to lower-priority frames.
2. *Parameterized* QoS. 802.11e also allows parameterized QoS where an 802.11e station can effectively request a certain bandwidth to be reserved (similar to RSVP—see Braden 1997). This is provided through *enhanced PCF*. It requires a coordinator (called *hybrid coordinator*, HC) that is usually located at the AP. A station may send a *reservation request* to a HC and be allocated a *transmission opportunity* that indicates the duration for which it can transmit (once it gets a chance to transmit).

A detailed description of 802.11e is presented in Mangold et al. (2002). For wireless IP telephony, 802.11e provides two important services. First, it allows voice frames to be treated at a higher priority as compared to non–real-time data frames. Second, it provides a mechanism for implementing call admission control—that is, it allows restrictions (depending on the capacity of the 802.11 network) to be imposed on the number of voice calls that can be initiated at the same time.

SECURITY AND PRIVACY

Security and privacy are important requirements of any communication network. These requirements become mandatory if the communication network is used for transferring personal or confidential information. Currently, such traditional telephone networks as the *public switched telephone network* (PSTN) provide an acceptable level of security and privacy. This security and privacy support has been achieved from a number of years of use and experience. IP telephony–based networks have introduced several new security and privacy concerns that have not been fully addressed yet (Ransome and Rittinghouse 2005). Combining wireless communications with IP telephony further introduces new caveats and security concerns that do not occur in PSTN or IP telephony-based networks. It is critical that these security and privacy issues are adequately addressed before implementing a widespread deployment of wireless IP telephony.

Unfortunately, as with IP telephony–based networks, the security and privacy concerns of wireless IP telephony have been overshadowed by the increased convenience, flexibility, cost effectiveness, and other advantages that it provides. Wireless IP telephony faces greater security and privacy risks because of a combination of several factors. These include the use of TCP/IP, the inherent characteristics of the wireless communications medium, and the relatively new protocols that have yet to be fully developed and tested.

Because wireless IP telephony uses IP as the underlying communication protocol, it inherits all of the known and unknown security weaknesses of IP. The issues of security and privacy were not considered when IP was first designed. As a result, many vulnerabilities are well known. For example, by spoofing source IP addresses, an attacker can impersonate a legitimate user and make phone calls disguised as that user. IP networks are common and easily accessible to the attackers. Furthermore, different parts of IP networks are controlled by different entities—for example, different service providers. Enforcing a common set of security protocols over these different parts of the networks is an extremely challenging task.

The wireless communications medium exacerbates the security and privacy challenges of IP telephony. We identify five sources of increased vulnerability resulting from the introduction of the wireless communications medium. First, the wireless communications medium increases the risk of eavesdropping by someone snooping packets over the air. It is relatively straightforward for an attacker to eavesdrop on a wireless communications medium using commonly available, off-the-shelf, and inexpensive equipment. In fact, it is significantly simpler for an attacker to eavesdrop over a wireless communications medium than over a wired communications medium. Thus, the privacy of telephone conversation is greatly jeopardized by incorporating wireless communication.

Second, it is relatively straightforward to launch denial-of-service attacks in the wireless communications medium. It is relatively easy to tap into a wireless communications medium; in the simplest case, an attacker can jam a wireless communications medium by injecting junk information. Of course, this simple attack requires an attacker to expend a large amount of energy. However, several intelligent denial-of-service attacks are well known that exploit specific communication patterns of network- and transport-level protocols. These attacks selectively jam specific control packets of TCP/IP protocols resulting in large number of retransmissions, or they jam for just long enough to interfere with the tail end of an IP packet. Such intelligent attacks require an attacker to expend little energy, and they cause significant damage in service availability.

Third, an attacker can take advantage of the vulnerabilities of the wireless communications medium to inject legitimate packets. This can enable the attacker to make unauthorized telephone calls, impersonate a legitimate telephone user, or even modify the contents of an ongoing telephone conversation. It is difficult to launch such an attack in traditional PSTN networks, but the wireless communications medium makes launching such attacks simpler.

Fourth, the existence of the wireless communications medium makes it easier for an attacker to compromise telephone user databases. Once again, the main reason

for this simplicity is the wide pervasiveness of the wireless communications medium and the ease with which one can tap into this network. This can allow an attacker not only to steal important private information of a telephone user (including name, address, and social security and credit card numbers) but also to modify this information in insidious ways.

Finally, some parts of the radio spectrum are licensed while others are unlicensed. Cellular telephone service uses licensed radio spectrum, whereas WiFi uses unlicensed radio spectrum. As a result of this and because of high Gaussian (white) noise generated by numerous RF-emitting systems (such as APs that may have been placed in close proximity), chances of interference in WiFi are significantly high, thus making the problem of providing security and privacy more difficult.

These security and privacy vulnerabilities associated with the wireless communications medium are quite well known, and in fact have been addressed to some extent in other applications that use the wireless communications medium. As a result, solutions to address the problems of eavesdropping, tampering, and denial-of service attacks have been developed. The key problem is that these solutions are typically not suitable for wireless IP telephony. This is because either the proposed solutions do not work well in resource-constrained mobile devices that users of wireless IP telephony tend to use or the quality of service issues such as latency, jitter, and packet loss are adversely affected.

Cryptographic algorithms, particularly asymmetric key algorithms, are highly computation-intensive. They require large amounts of memory and consume lots of power. Mobile wireless users typically tend to use relatively small, resource-constrained devices that run on low power. As a result, traditional cryptographic solutions do not work well in such devices. Furthermore, traditional solutions such as firewalls can delay or even block call setups, while computation-intensive encryption and decryption can cause unacceptable latency and jitter.

It is clear that wireless IP telephony brings in significantly higher security and privacy risks that are more than a mere nuisance. It is critical that these security and privacy concerns be addressed adequately before this promising technology is widely used.

ENHANCED 911

As people depend more on IP telephony (in many cases for all of their telephony service needs), it is vital that emergency services can accurately determine the location of a caller in an emergency in a timely manner. In PSTN systems, when a person makes a 911 call, the call is routed to the nearest *public safety answering point* (PSAP). The PSAP receives the caller's phone number and the exact location of the phone from which the call was made. The Federal Communications Commission (FCC) introduced enhanced 911 to allow mobile or cellular phone users to make 911 emergency calls and enable emergency services to locate the geographic position of the caller. This is typically implemented in current cellular telephone networks by using some form of radio location from the cellular network or a *global positioning system* (GPS) device built into the phone itself.

The packet-switched technology of IP telephony allows a caller to place a call from anywhere, unlike the PSTN in which a phone's physical location is fixed or a cellular network in which a phone's physical location is tied to the radio tower of the cell in which the phone is currently located. IP telephony severely complicates the provision of E911 service. In fact, earlier IP telephones were not integrated with the 911 system at all. However, the FCC has now mandated all IP telephony service providers to provide 911 service, including the E911 feature. There are several complicated technological problems with integrating E911 with IP telephony, and currently most IP telephony service providers only provide an ad hoc solution. For example, some service providers have encouraged their customers to register their locations from which their 911 calls should be routed to the local PSAP. Other service providers are attempting to connect customers to E911 services through the PSTN. The problem here is that the PSTN is controlled by telecom carriers who are their economic competitors.

The key problem associated with providing E911 service in IP telephony is that it is hard for an IP device to figure out its actual geographic location in a dynamic manner. This problem becomes more complicated in the case of wireless IP telephony because not only is a 911 call using a packet-switched network but also the user is mobile. It can be argued that a user is more likely to be within a short distance from a wireless access point with a fixed location. This provides a simple, though not highly accurate, solution for most situations. However, with the advent of wireless broadband services such as WiMAX or newer ad hoc wireless networks that allow an IP device that is several hops away from a wireless access point to be connected, a user can be significantly far away from a wireless access point. Another solution—although a bit expensive—is to install a GPS chip, which periodically transmits location information, in all handsets. This will enable accurate, near real-time user location information to be determined.

Determining the location of a WiFi client only from the RF signal (that is, without using other technologies such as GPS) is being actively researched. Several WLAN localization techniques have been proposed (e.g., Youssef and Agrawala 2005). Many of these use machine learning techniques to model the location of a WiFi client as a function of the properties of the signal (such as signal strength) that it receives. However, they are not always accurate and reliable. In fact, some commercial WLAN localization solutions (e.g., from Ekahau Inc.) have also emerged in the recent years.

CONCLUSIONS

In this chapter, we have provided a survey of the current state of the art in wireless IP telephony. We discussed key technical challenges that face wireless telephony related to mobility management, quality of service, security, privacy, and E911. Another challenge is seamless vertical handoff that would allow mobility between a WLAN and a cellular network. To be universally accepted, wireless

IP telephony must provide telephony services with the same degree of quality, high availability, and security as traditional circuit-switched telephony systems. To this end, substantial progress has been made in the past few years, and research solutions exist for most technical issues. However, it remains to be seen how fast these will be adopted and realized by industry.

GLOSSARY

802.11 Ad Hoc Mode: In this mode of operation, 802.11 wireless devices can directly communicate with each other without a need for an AP; also referred to as *peer-to-peer mode*.

802.11 Infrastructure Mode: In this mode of operation, 802.11 wireless devices communicate with each other through an AP, which also provides the connectivity of wireless devices to the wired network.

802.11 Standards: 802.11 (or WiFi) is a set of IEEE specifications that define various aspects of a wireless local area network. It consists of standards such as 802.11a and 802.11g (which define PHY and MAC layer of the WLAN), 802.11e (which defines QoS enhancements), and so on.

Access Point (AP): An AP is a layer 2 device that allows hosts in a wireless LAN to communicate with each other. It usually also provides the wireless devices with connectivity to the wired network.

Distributed Coordination Function (DCF): DCF is a MAC mechanism for providing channel access to 802.11 devices. Unlike PCF, it does not require a central coordinator.

Enhanced 911 (E911): A location technology mandated by FCC that would allow telephony service providers to locate callers in case of emergency calls. This is also required for emergency calls made using wireless IP telephony or IP telephony when it is harder to implement because of its mobile nature.

Handoff: In general, a handoff is passing an association from one entity to another. In the context of 802.11, a handoff is a transfer of association from one AP to another. A handoff is required when the current AP's signal quality becomes poor and another AP with a better signal quality is available.

IP Telephony: Use of the packet-based IP network for transporting voice calls instead of the traditional circuit-switched networks.

Jitter: In the context of IP telephony, jitter is the variation in the arrival rate of voice packets. A high jitter value leads to poor voice quality as perceived by a user.

Media Access Control (MAC): A mechanism for mediating access to each of the contending devices sharing a common resource for communication, such as a cable in case of Ethernet or the electromagnetic spectrum in case of 802.11.

Mobile IP: An IP extension for facilitating seamless operation of mobile applications. It allows hosts to roam—that is, change their point of attachment to the IP network while keeping the same IP address.

Mobility Management: In the context of wireless IP telephony, refers to mechanisms (at various layers of the network stack) to enable seamless communication (for ongoing as well as new calls) for a user who is mobile.

Packet Latency: In the context of IP telephony, refers to the delay in transporting a voice packet from a sender to a receiver. A high packet latency leads to poor voice quality.

Point Coordination Function (PCF): A MAC technique for providing channel access to 802.11 devices that uses a coordinator (unlike DCF) that is usually located at an AP.

Public Switched Telephone Network (PSTN): The publicly available, circuit-switched traditional telephony network.

Quality of Service (QoS): Refers to a mechanism for providing performance guarantees in the transportation of packets over a network. In the context of IP telephony, it refers to providing guarantees for transportation of voice packets on the IP network, which has traditionally only provided a best effort service in terms of packet loss, jitter, and latency.

Vertical Handoff: A handoff between heterogeneous access technologies—for example, between WLAN and 3G (cellular).

Voice over Internet Protocol (VoIP): Transmission of voice over an IP network as packets.

Wireless Fidelity (WiFi): WLAN based on 802.11 standards.

Wireless IP Telephony: Use of IP telephony over a wireless IP network, such as that based on the 802.11 standards (WiFi).

CROSS-REFERENCES

See *Network QoS*; *Voice over IP (VoIP)*; *Wireless LAN Standards*; *Wireless LANs (WLANs)*.

REFERENCES

Banerjee, N., S. Das, and A. Acharya. 2006. Seamless sip-based mobility for multimedia applications. *IEEE Network*, 20(2): 6–13.

Braden, R. 1997. *Resource reservation protocol (RSVP)*. RFC 2205, September.

Brik, V., A. Mishra, and S. Banerjee. 2005. Eliminating handoff latencies in 802.11 WLANs using multiple radios: Applications, experience, and evaluation. In *Proceedings of the Fifth Conference on Internet Measurement 2005* (USENIX IMC), Oct. 19–21, Berkeley, CA, USA. Pp. 299–304.

Dutta, A., O. Altintas, W. Chen, and H. Schulzrinne. 2002. Mobility approaches for all IP wireless networks. In *Sixth World Multiconference on Systemics, Cybernetics, and Informatics,* July 14–8, Orlando, FL, USA.

IEEE. 1999. Standard 802.11. Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications, July.

ITU-T. 2003. Recommendation G.114. *One way transmission time*, May.

Johnson, D., C. Perkins, and J. Arkko. 2004. *Mobility support in IPv6*. RFC 3775, June.

Kwon, T., M. Gerla, and S. Das. 2002. Mobility management for VoIP service: Mobile IP vs. SIP. *IEEE Wireless Communications*, 9(5): 66–75.

Mangold, S., S. Choi, P. May, O. Klein, G. Hiertz, and L. Stibor. 2002. IEEE 802.11e wireless LAN for quality of service. In *Proceedings of European Wireless '02*. Pp. 32–9.

Mishra, A., M. Shin, and W. A. Arbaugh. 2003. An empirical analysis of the IEEE 802.11 MAC layer handoff process. *Computer Communication Review*, 33(2): 93–102.

Mishra, A., M. Shin, and W. A. Arbaugh. 2004. Context caching using neighbor graphs for fast handoffs in a wireless network. In *Proceedings of the Twenty-Third Conference of the IEEE Communications Society* (INFOCOM), March 7–11, Hong Kong. Pp. 351–61.

Nakajima, N., A. Dutta, S. Das, and H. Schulzrinne. 2003. Handoff delay analysis and measurement for sip-based mobility in IPv6. In *IEEE International Conference on Communications* (ICC). Vol. 2: 1085–89.

Nichols, K. 1998. *Definition of the differentiated services field (DS field) in the IPv4 and IPv6 headers*. RFC 2474, December.

Pack, S., and Y. Choi. 2002. Fast inter-ap handoff using predictive authentication scheme in a public wireless LAN. *Networks: The Proceedings of the Joint International Conference on Wireless LANs and Home Networks (ICWLHN 2002) and Networking (ICN 2002)*, Aug. 26–9, Atlanta. Pp. 15–26.

Park, S.-H., H.-S. Kim, C.-S. Park, J.-W. Kim, and S.-J. Ko. 2004. Selective channel scanning for fast handoff in wireless LAN using neighbor graph. In *Proceedings of the Ninth International Conference on Personal Wireless Communications*, Sept. 21–23, Delft, the Netherlands. Pp. 194–203.

Perkins, C. 2002. *IP mobility support for IPv4*. RFC 3344, August.

Ramani, I., and S. Savage. 2005. Syncscan: Practical fast handoff for 802.11 infrastructure networks. In *Proceedings of the IEEE Conference on Computer Communications* (INFOCOM), March, Miami. Pp. 675–84.

Ransome, J. F., and J. W. Rittinghouse. 2005. *VoIP security*. New York: Elsevier.

Sharma, S., N. Zhu, and T. Chiueh. 2004. Low-latency mobile IP handoff for infrastructure-mode wireless LANS. *IEEE Journal of Selected Areas in Communications*, 22(4): 643–52.

Shin, S., A. G. Forte, A. S. Rawat, and H. Schulzrinne. 2004. Reducing MAC layer handoff latency in IEEE 802.11 wireless LANs. In *Proceedings of the Second International Workshop on Mobility Management and Wireless Access Protocols*, edited by A. Boukerche, K.M. Sivalingam, and S. E. Nikoletseas, 19–26. Philadelphia: ACM.

Szabó, I. On call admission control for IP telephony in best effort networks. *Computer Communications*, 26(4):304–313, 2003.

Vali, D., S. Paskalis, A. Kaloxylos, and L. Merakos. 2003. An efficient micromobility solution for sip networks. In *IEEE GLOBECOM*, 2003. Vol. 6: 3088–92.

Velayos, H., and G. Karlsson. 2004. Techniques to reduce IEEE 802.11b handoff time. In *IEEE International Conference on Communications*, June, Paris.

Viterbi, A. J. 1995. *CDMA: Principles of spread spectrum communication*. Redwood City, CA: Addison Wesley Longman.

Wang, Q., M. Abu-Rgheff, and A. Akram. 2004. Design and evaluation of an integrated mobile IP and SIP framework for advanced handoff management. In *IEEE International Conference on Communications* (ICC), June, Paris. Pp. 3921–5.

Wu, W., N. Banerjee, K. Basu, and S. K. Das. 2005. SIP-based vertical handoff between WWANs and WLANs. *IEEE Wireless Communications*, 12(3): 66–72.

Youssef, M., and A. Agrawala. 2005. The Horus WLAN location determination system. In ACM MobiSys '05, June (retrieved from www.usenix.org/publications/library/proceedings/mobisys05/tech/full_papers/youssef/youssef_html/).