Chapter

4

The *M* in FMC

Give me a place to stand, and I will move the earth.

—Archimedes (287–212 B.C.)

Without a good understanding of fixed networks, which you already should have acquired by reading the preceding chapter, it would be quite difficult to fully appreciate the aspects of mobility we are going to delve into in this chapter. Indeed, much of the fundamental functionality and many architectural aspects of fixed networks have been borrowed and extended to create the core of modern mobile networks.

For instance, termination of a voice call in a mobile environment requires an interaction with a database storing subscriber profiles, the Home Location Register (HLR), much like the interaction that occurs in fixed networks with an SCP in order to route a toll-free phone call to the correct call center's call distributors. This example well illustrates how lessons learned and concepts invented for fixed environments are reapplied in a mobile setting, albeit with the necessary adaptations.

In this chapter we will identify these similarities and differences, and explore the most common cellular and noncellular radio access systems that may commonly be part of FMC solutions. The knowledge acquired in this and the preceding chapter will then make it possible for you to fully appreciate the ensuing discussion on convergence.

In the end, *fixed* and *mobile* communications can be seen (and they are) as two different facets of the same technology, which indeed have never been totally independent, so convergence for them is a consequential natural course of evolution. Thus, sooner or later, fixed and mobile networks will come together—the process

already evident in the industry with the launch of services allowing use of a mobile phone at fixed-line rates, and using fixed-line access, when in the home—with FMC being one of the technologies hastening this "homecoming." To paraphrase Archimedes' famous statement: "Give me a fixed network and I can build mobile networks around that."

Mobility

While mobility is not only about wireless or radio access systems (as we have established in earlier chapters, there is quite a good distinction between *wireless* and *mobile,* wireless referring to the type of access and mobile to the type of service), wireless systems are usually at the core of mobile communications, so in this chapter most of the discussion in fact revolves around such solutions. It should be noted that many wireless access systems such as Wi-Fi do not natively support mobility. In fact, most of today's applications have relied on Wi-Fi mainly for "cord cutting" while accessing fixed networks; that is, this has held true until the recently introduced uses of Wi-Fi for metro data coverage and wireless VoIP broadened its scope, albeit without the support for macromobility allowed by cellular systems.

In the future it is also likely that Wi-Fi and other technologies such as WiMAX will be used to connect to mobile packet cores (à la 3GPP TS 23.234 [82]), which is not surprising given their broadening support on cellular phones. Before turning to Wi-Fi, however, in the sections that follow we will first address cellular systems and then introduce WiMAX access, paying special attention to its mobile version.

Mobile Communications Systems Fundamentals

One of the fundamental characteristics of mobile systems is that, like fixed systems, they include a core network, the elements of which act as an anchor for mobility. The core network also contains points of interworking with other (fixed and mobile) networks and connects to an access network needed to support the radio interface technology that characterizes a specific mobile system.

Core and Access

Note that different mobile systems, defined by various radio interfaces and access networks, may still share the same core network. For instance, it is possible to use the same core network to support the Global System for Mobile Communications (GSM—originally an acronym for *Groupe Spécial Mobile* and now the most widespread cellular technology) and the Universal Mobile Telecommunications System (UMTS).

Core Network Functions The core network normally includes the "intelligence" necessary to make the system operate. The core provides the admission control to wireless

services via access authentication and stores subscriber profiles, letting the core determine the specific service set and treatment to be applied to a user. These are examples of the questions the core can decide on:

- Is the user allowed to roam to another network?
- Are there areas where the user cannot receive service?
- Is the user in and out of a "home zone"?
- What Quality of Service (QoS) for packet data services is the user is entitled to?
- What calls is the user entitled to receive or place, and which are barred?

The core also provides *mobility management* of the mobile terminal in idle mode, that is, at times when it is not engaged in a voice call or in the active transmission and reception of data. The mobility management function allows tracking the user's whereabouts, by means of location updates the terminal issues periodically or when it crosses boundaries of "location areas" as the user moves.

Further, the core participates in call control for speech services and session control for multimedia services as well as providing infrastructure that is ancillary to the delivery of services (such as location information). The core network is also the main source of charging information.[1] Finally, mandated services like lawful interception rely on the core network to provide support for tracking the location and specific behavior of a user, log its activity types, and deliver the content of communications to the authorized agencies.

All this intelligent functionality (typically supported by different servers and network elements) can be quite computationally intensive, and concentrating these in centralized locations provides efficiency gains in large part for statistical reasons (not all users will need to use the same resources at the same time, so centralizing some functions will increase efficiency). Another reason for centralization is the need to provide a "point of access" for external networks to the mobile system services. Examples of this include scalable call termination and roaming interfaces (on roaming interfaces, it is in fact necessary to minimize the number of peering relationships to simplify roaming operations).

Access Network Functions Typical functions of the radio access network are Physical-layer termination and Link-layer termination of radio interface protocols, ciphering and header compression,[2] mobility management within the radio access, paging, broadcast information distribution, and scheduling.

---

[1] In some systems, like in UMTS, the RAN cooperates with the core by delivering unsent data volume reports, but this feature is most likely to be discontinued in future systems, as it either was not implemented in most cases by vendors or was not used by providers, and of course the increase of data rates will make the value per bit unsent smaller, so it will be less important to count unsent data.

[2] In some technologies like GSM, ciphering and header compression for data sessions are done in the core network.

It should also be understood that for a mobile system to provide a satisfactory user experience, it must offer extensive coverage over large geographical areas. There is also the need to keep the transmission power of terminals and the complexity of receivers low to allow for longer battery life and lower terminal cost. To assist in that goal, the reach of the base transceiver stations (BTSs, which include the antennas in the network, radiating signals to mobiles and receiving signals from them) has to be limited (also to allow for frequency reuse, as explained later in the chapter), so the number of elements in a typical radio access network is much higher than in the core, and this is where most of the capital investment is needed.

Since BTSs represent the largest number of elements in a mobile radio access network, the functions in the BTS tend to be kept to the bare minimum necessary to provide radio access. However, with the introduction of high-speed wireless data services, there is a trend to offload some of the functions that were suitably kept in the access nodes higher up in the radio access network hierarchy down to the BTS. There is a host of technical reasons behind this, mostly linked to the delay introduced by the traversal of a backhaul network to a controller, which would make, for instance, running some retransmission algorithms less effective.

This trend of assigning additional functions in the BTS to handle data access has been taken to the extreme by the introduction of the concept of "Home BTS" or *femtocell,* which is promising the delivery of "personalized" wireless coverage in residential or business premises not reachable by the existing WAN infrastructure of BTSs. This technology is driving the use of IP connectivity to the BTS in order to reuse the infrastructure already available in those premises, and the simplification of the protocol layers supported on the interface between the core and the *femtocells,* to minimize the cost of backhaul.

Circuit and Packet

Mobile networks, like fixed networks, can provide a variety of services, including circuit-switched (CS) and packet-switched (PS) services.

Circuit services include speech, video, and circuit-switched data (akin to landline modem dial-up—fast becoming a thing of the past with the rise of broadband—based on dialing a circuit connection to a remote access server (RAS) and running PPP over it).

Packet services include high-speed data used for access to Internet and private networks, portal access, Web browsing, instant messaging and presence, and other innovative services, including IMS, multicast, and broadcast.

Circuit-switched services are supported by all traditional *cellular* systems, whether digital or analog (the latter actually have been discontinued globally, with possibly a few rare exceptions). However, not every *mobile* communication system supports circuit services. For instance, the evolution of the 3GPP system foresees that all services will be packet based and, more specifically, IP based. Other mobile systems such as WiMAX and CDMA EV-DO revA were built as packet only from the ground up.

On the other hand, packet-switched services were not supported by analog systems or by initial releases of digital systems, as at the times these systems were introduced, the demand for data services was not sufficient to justify the additional effort to define packet data support in cellular systems. This functionality was added in subsequent system releases as an evolution of both the radio and the core network of these systems.

The need to support both packet- and circuit-switched services requires the core and access of a mobile-system to support features that are often very different in nature. The mobile-system features dedicated to CS services are often identified as the "CS domain," whereas the set of features devoted to the support of PS services are defined as the "PS domain."

Having thus explained the fundamental differences between PS and CS, we are now ready to proceed with an overview of the specific mobile systems, starting with the discussion of the traditional cellular infrastructure.

Cellular Systems: The Sky Is the Limit

The concept of *cellular* has been invented to cope with one fundamental issue in wireless transmission technologies operating in a licensed spectrum.[3] The issue is the scarcity (and cost!) of the radio frequency (RF) bandwidth. When an operator is allowed to provide wireless service, a regulatory authority assigns it portions of the RF spectrum (known as "bands") to use for radiating and receiving signals from its subscribers.

Sometimes operators engage in fierce bidding to buy some of this spectrum, which results in staggering amounts of money being transferred to the government (or, in some cases, to previous owners of the spectrum bands). Therefore, the need to maximize the efficiency when using this precious resource, as well as the need to overcome basic physical limitations of wireless transmission, has led to the invention of what is called *frequency reuse.* With frequency reuse, multiple transmitters can use *the same* radio frequency to send and receive data as long as they are "sufficiently remote" so as to avoid interference. That is, a given (RF) channel can be used over and over again and thus cover an extremely wide area and support a very high number of simultaneous transmissions.

The concept of "sufficiently remote" requires some degree of clarification here. It is a common experience for everyone that the coverage of a radio station can be quite broad even without repeaters. The reason for this is that the radio station transmits at high power levels. In cellular systems, by contrast, the goal is to control the power of transmission so that interference is avoided between areas where the same frequency is reused. So, the concept of "sufficiently remote" is relative to the possible reach of a signal transmitted at a given power level. One of the most important characteristics of frequency reuse is the *reuse distance.* Intercell interference can be limited by changing

[3] Licensed spectrum is spectrum for which an operator needs to obtain authorization from a regulatory authority for the use of the radio frequency (RF) it needs for the operation of the wireless service.

the reuse distance and the power levels of BTS transmission. The resulting combination of power control and frequency planning is used to fine-tune the cellular network in a given area.

The elementary area of coverage in a cellular system roughly centered around the BTS is called a *cell*. Each BTS can use a subset of the RF channels the operator is allowed to operate. These channels cannot be reused in any neighboring and potentially interfering cells (that is, any cell within the "reuse distance"). With the introduction of cellular systems based on Code Division Multiple Access (CDMA) technology, there have been ways to advance the use of digital technology to reduce interference while allowing for greater reuse levels and thus increased capacity. Also, other techniques such as cell "sectorization" (using directional antennas, thus adding space division to frequency division, in order to maximize frequency reuse) and dynamic channel assignment to cells (thus introducing the concept of time division in addition to frequency division) have made it possible to further increase capacity by allowing higher reuse factors and availability of more channels per BTS.

In summary, the cellular systems have made it commercially viable to support mobility of a large number of subscribers in a wide area at a reasonable cost practically without any limitations as long as a sufficient number of base stations is installed. It is no surprise that, given the relative ease and speed of deployment allowed by cellular systems, and the reach of customers in even the remotest areas at reasonable costs, cellular communication is now a dominant and growing sector of the telecommunications industry. This sector is, however, relatively nascent in comparison with fixed networking (it is only a few decades old, compared to the more than one-hundred-year history of public wireline telephony), but it has dramatically changed the way we communicate. Its evolution has also been quite rapid from the early days of analog cellular telephony to the recent days of high-speed wireless data. The next section summarizes this evolution.

Cellular Generations

The evolution of cellular communication systems has been commonly identified via a series of generations (1G, 2G, 3G, and, by implication, 4G and yet further generations). This evolution has been somewhat linked to the evolution of computing. When digital computing was expensive, it was considered unfeasible to digitally process speech and use digital transmission of digitally encoded voice over the radio interface, in large part due to the high handset power requirements and limited memory capabilities. Therefore, the natural choice was to keep wireless telephony entirely in the analog domain.

1G Analog cellular systems are commonly known as 1G (or first-generation) cellular systems. 1G was introduced in the late 1970s and early 1980s. These systems delivered almost exclusively speech services; albeit some low-bit-rate data services were available as dial-up connections via wireless modems.

Analog systems are almost everywhere being phased out, and those that are not will probably be retired very soon. From a practical standpoint, the investment in these

systems has stopped on both vendors' and operators' sides, and for this reason they are not considered in the scope of convergence, which mostly lies in an area where the industry is going to develop.

Typical 1G systems include:

- **AMPS** Advanced Mobile Phone Service, adopted mainly in the America, using FDMA transmission in the 800 MHz band.
- **TACS** Total Access Communication System, adopted mostly in Europe. TACS was similar to the AMPS system. There were various flavors of it; for instance, in the UK, ETACS (Extended TACS) operated in the 871–904/916–949 MHz band, and Narrowband TACS (NTACS) operated in the 860–870/915–925 MHz band (by using narrower channel spacing, it supported more channels for the same amount of spectrum).
- **NMT** Nordic Mobile Telephone, deployed in many European countries, was first launched in the Scandinavian region (as its name may suggest) in 1979, and it was the first analog cellular system operating in both the 450 MHz and 900 MHz bands. The operation in the 450 MHz band affords particularly good coverage due to the propagation properties of radio waves, so it has been kept alive for quite a long time in Scandinavia, often used by boat owners in the Baltic region, or rural area dwellers, where it was not economical to deploy systems in different bands. The "450" frequencies are therefore particularly appealing for these reasons, and digital systems offering the capability to cover these subscribers (GSM, CDMA, and also 3G) have been offered to refarm this system.

1G systems were deployed on a country-by-country basis and did not offer international roaming. This became an apparent issue, especially in Europe where there were two different analog systems operating in different frequencies and according to different rules. This clearly was creating a problem for the development of the cellular industry, and therefore even at the political level it was clearly understood that the development of competing digital systems in that region would have been a critical mistake.

Beyond forcing compatibility issues and hampering roaming capabilities for European citizens, this would have fragmented the market and made it more difficult for European vendors to compete globally. European operators would also be unable to benefit from effects of scale for both equipment and, most important, handsets. In summary, this was at the core of the reason why European regulatory bodies encouraged operators and vendors to come together and agree on the common digital system, which was later named GSM.

2G Cellular digital systems in the first wave to appear in the mobile communications industry are known as 2G systems. These were rolled out worldwide in the course of the 1990s. These systems are characterized by the digital transmission of all services, including voice, which is digitally encoded for transmission over the radio interface.

Initially they were conceived to deliver circuit services only, but then their evolution to deliver packet data services was devised (these systems' evolution to support packet data later became known as 2.5G).

Typical examples of 2G systems include:

- **GSM** The Global System for Mobile Communications initially is the European 2G system, specified by ETSI and currently maintained by 3GPP. It is now adopted in practically every country or region of the world, except Japan, Korea, and a few others. GSM was initially operating in the 900 MHz band only, but then operation at 1800 MHz, 1900 MHz, 450 MHz, and other frequencies was also defined. GSM uses a time division multiple access (TDMA) multiplexing over-the-air interface. Its 2.5G system evolution is known as General Packet Radio Service (GPRS), and its 2.5G air interface evolution is known as Enhanced Data Rates for GSM Evolution (EDGE). EDGE has also been accepted as an IMT-2000 technology, so technically it is also part of the "3G" mobile systems; albeit it is deeply linked to its 2G heritage.
- **TDMA** Time division multiple access, defined in TIA IS-136 [96], or D-AMPS (for Digital-AMPS), has been adopted in North and South America. It uses a time division multiple access technology (like GSM), over-the-radio interface, but unlike in GSM, this technology was not globally adopted, so it has been largely phased out in favor of GSM in every region where it was deployed (except some South American countries, where it may still be available for a while).
- **PDC** Personal Digital Cellular (PDC) is the Japanese standard for digital cellular telephony. In Japan, this is quickly being replaced by 3G systems, and it has received quite phenomenal competition from a Japanese cordless telephony standard (PHS—Personal Handy-phone System), which has been made available in major urban areas to the large population of pedestrians and users of public transportation systems. PDC is also a TDMA system, replacing the analog NTT and JTACS Japanese systems. It operates in the 800 and 1400 MHz frequency bands.
- **CDMA** Code Division Multiple Access, specified in standard TIA IS-95A [98], uses a technique to assign to all transmitters their respective, specific, bit vectors, which are mutually orthogonal so that a receiver can recover the signal transmitted even if multiple transmitters use the same carrier at the same time to modulate the RF signal. This standard is now maintained by 3GPP2 and has been adopted in many countries in Asia and South America, after its initial deployment in North America. The evolution of IS-95 to support packet services is known as IS-95B, and it supports data rates up to 64 Kbps.

3G Similarly to what happened in Europe during the migration from 1G to 2G, during the migration from 2G to 3G it appeared that having many flavors of 3G standards across the globe was not advantageous. So various standards organizations from different countries came together to form the Third-Generation Partnership Project (3GPP)

to define a standard based on the evolution of the GSM core network and W-CDMA radio transmission technology (the most promising and efficient at the time).

In North America, this evolutionary path to 3G was not received too well, though, as major CDMA operators (and their main suppliers) perceived it not to be protecting their investment sufficiently. So a competing partnership was formed, called 3GPP2, developing a 3G standard based on the evolution of IS-95 CDMA radio transmission technology and the ANSI-41-based core network.

With the creation of 3GPP and 3GPP2, the intent of achieving a global cellular standard for 3G has failed. However, there was a considerable momentum toward harmonization for both competing standards as the number of different systems went down from four in 2G to two in 3G.

The 3G standardization activity took place under the supervision of the IMT-2000 in ITU-T, which was assigning frequencies and acknowledging technologies suitable for the 3G standards (also known hence as IMT-2000 technologies). 3G Systems, launched after the year 2000, promised to offer faster access to the Internet and other data services with typical speeds ranging in the hundreds of Kbps. They also offered circuit services like speech and real-time video. The systems resulting from the efforts of the 3GPPs were:

- **UMTS** The Universal Mobile Telecommunications System, which has been specified by 3GPP, is a multiradio interface system, which includes satellite communications, but its most widespread use for now is based on a W-CDMA radio interface, and it operates in different frequency bands, depending on the region of deployment. W-CDMA comes in FDD (frequency-division duplexing) and TDD (time-division duplexing) flavors. In FDD, uplink and downlink signals are allocated to different frequencies, while in TDD they share the same frequency on a time-division basis. UMTS offers both packet data and circuit services. In China, the TD-SCDMA (time division–synchronous code division multiple access) radio interface has been selected to constitute the homegrown UMTS radio interface. Japan launched 3G UMTS before any other country, due to the leadership of DoCoMo in 3G UMTS-based standards with its Freedom of Mobile Access (FOMA) system.
- **CDMA 2000** CDMA 2000 evolves IS-95 to include additional service capabilities based on packet data. It is a direct competitor of the other major 3G standard, UMTS, and operates at 450 MHz (used to refarm NMT, as mentioned earlier, along with GSM operating in the same band, and in developing countries to allow better coverage), 850 MHz, 900 MHz, 1800 MHz, 1900 MHz, and 2100 MHz. CDMA2000 1XRTT has been the first step in the evolution to 3G, which has been recognized as an IMT-2000 technology, and also the first 3G technology to be deployed worldwide.

Since the first releases of the 3G systems, they have evolved quite significantly. CDMA2000 EV-DO (Evolution–Data Only or –Data Optimized, also known as the High-Rate Packet Data air interface) and EV-DO revision A (DOrA), HSDPA (High-Speed Downlink Packet Data Access) are examples of the 3G high-speed data services–capable systems that grew out of the original UMTS and CDMA2000 standards.

If the industry continues to follow the trend established during the transition from 1G to 3G in the transition from 3G to 4G, we should expect to see only one global cellular system based on a common core and radio transmission technology. While this is not yet a certain fact, all evidence suggests that this is going to be the case, as major operators of 3GPP and 3GPP2 networks seem to be converging on a single evolutionary path to 4G.

We must mention, however, another trend developing in parallel with cellular standardization. It appears that systems such as WiMAX (and, to a certain extent and with some additional limitation, Wi-Fi) are being positioned as yet other candidates for deployment in many operators' networks as a complement to or even a complete replacement for 3GPP and 3GPP2 cellular systems.

So, at a time when major cellular families seem to be on an eventual path of convergence, there is a potential for competing technologies (if only on the radio interface side) to be adopted by both the cellular and wireline operators. As a consequence, this may result once again in a plurality of systems. On the other hand, due also to the potential existence of such a plurality of access technologies that need to provide access to the same set of services, it is also recognized that using a single access-independent core is indeed valuable. As such, the industry is experiencing more and more a drive toward using a common core for all mobile (and nonmobile) systems, providing access to the same set of services, maybe based on IMS, thus validating the overall direction toward convergence.

In the following sections we provide a more detailed review of the mobile systems, which are best positioned to participate in FMC solutions. In this discussion we are going to focus on the relevant aspects of these systems such as core network details and operation, and characteristics of the access network that need to be considered when being converged with the fixed network in the context of practical FMC solutions.

3GPP Systems

Both the GSM 2G mobile system and the UMTS 3G system are specified and maintained by 3GPP. The GSM system was originally specified by ETSI under the drive of the European countries[4] to share a common digital system to enable easier intraregion subscriber roaming, but when 3GPP started its work to define a 3G system evolved from the GSM core, it became clear that it would have made perfect sense to also maintain GSM specifications and define their evolution within 3GPP.

Today it is possible to roam using GSM in more than 210 countries. GSM operates in the 900 MHz and 1.8 GHz bands in Europe and the 1.9 GHz and 850 MHz bands in the U.S. The 850 MHz band is also used for GSM and 3GSM in Australia, Canada, and many South American countries. UMTS was originally defined to operate in the

[4] The European Conference of Postal and Telecommunications Administrations (CEPT) created the *Groupe Spécial Mobile (GSM)* in 1982 aiming at developing a standard for a mobile telephone system that could be used across Europe.

1885–2025 MHz band for uplink and 2110–2200 MHz for downlink, though, in addition to these spectrum ranges, today it is commonly run on 850 MHz and 1900 MHz in some countries, notably in the U.S. by AT&T.

3GPP subscribers are identified by means of an International Mobile Subscriber Identity (IMSI), and they are assigned one (or more) MSISDN (Mobile Station ISDN) number, that is, an E.164-compliant telephone number. These are stored on a Smart Card known as the Subscriber Identity Module (SIM) card for GSM, or Universal SIM (USIM) card for UMTS. The smart cards are given to subscribers when they sign up for service, and they can be installed on any 3GPP-compatible handset for its activation with the service. The SIMs are not commonly used in 3GPP2 systems, where subscriber identity is linked to a specific terminal; albeit in recent years the capability to use a smart card has been added to CDMA standards and is now being deployed (especially in China, among other countries) with the smart card–based Removable User Identity (RUID).

The GSM and the UMTS systems share a common core network for both the CS domain (also known as circuit core) and the PS domain (also known as GPRS core, or packet core).

The 3GPP CS core The 3GPP circuit core (shown in Figure 4.1) is based around the Mobile Switching Center (MSC), which acts as a switch for voice and circuit data calls and handles mobility management of "CS-attached" users. CS-attached users are authenticated and accepted by the circuit-switched mobile core and can therefore be handled by the 3GPP system. Authentication of 3GPP users is based on data and algorithms stored in the SIM. On the network side, the authentication function is based on data downloaded from the Home Location Register (HLR), specifically from the

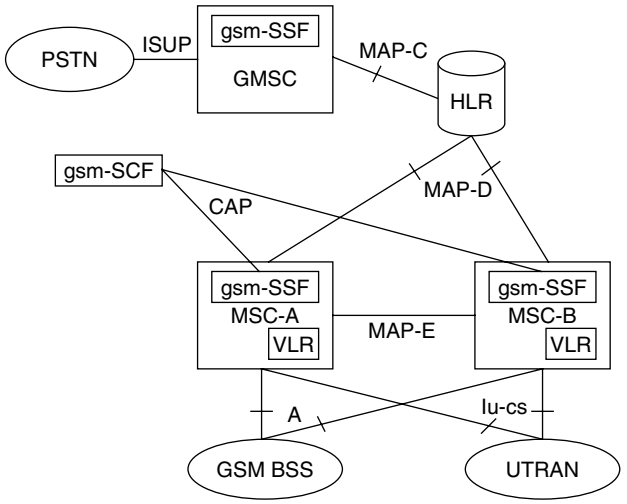

Figure 4.1 The 3GPP circuit core

authentication center (AUC) component of the HLR, into the serving MSC, and more precisely in the Visitor Location Register (VLR), also commonly a component of the MSC. The HLR and VLR are subscriber databases where the subscriptions to services, their parameters, and their activation status information are stored, along with the authentication information.

The HLR is always located in the home network, the network with which the subscriber has a customer-provider relationship. The VLR is located in the visited network. Since both the roamers' and home subscribers' data are stored in the VLR when a subscriber is registered with an MSC, the VLR effectively assumes a role as a cache of subscriber data, used to avoid continuous interrogations of the HLR, which would cause significant signaling load. Note that the VLR was defined to be potentially stand-alone, but in all practical cases today it is implemented within the MSC platform.

MSCs also host Customized Applications for Mobile Network Enhanced Logic (CAMEL) Intelligent Network triggers via the GSM Service Switching Function (gsm-SSF), which are armed to provide Intelligent Network–supported capabilities such as toll-free calling, caller ID, location determination, call forwarding, and so on, by interacting with the GSM Service Control Function (gsm-SCF).

In addition to these functions, the MSC often acts as a gateway to the PSTN, and as such it is identified as gateway MSC (GMSC). In this role the GMSC, as shown in Figure 4.2, would receive incoming ISUP call establishment signaling messages from the PSTN and query the HLR for the whereabouts of the user (more precisely, find out the MSC where the user is known to have last registered with the HLR).

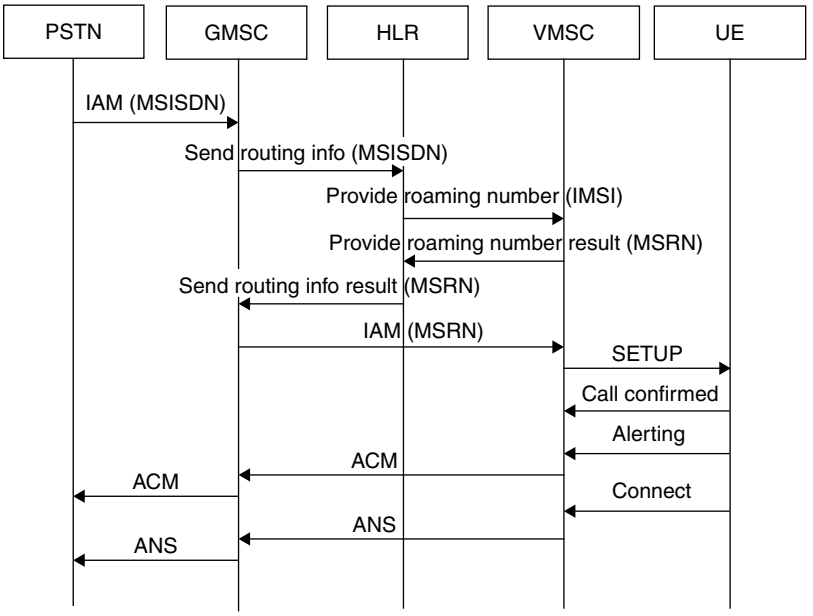

Figure 4.2 Mobile-terminated call setup

The HLR, as shown in Figure 4.2, then queries the visited MSC's VLR to obtain a Mobile Station Roaming Number (MSRN) for the subscriber; the VLR allocates one MSRN and returns it to the HLR, which passes it back to the GMSC. The GMSC would then use ISUP to establish a circuit using the MSRN as the destination number. When the call is set up, the MSRN can be released and used for another subscriber, thus limiting the number of E.164 addresses needed for mobile-terminated call routing at an MSC.

The approaches similar to the one used for such termination of calls have been used for a mechanism to route calls to the IMS or the CS domain in Voice Call Continuity (VCC—described in detail in the next chapter) for dual-mode terminals capable of receiving a voice call in the IMS-based PS or in the CS domain, depending on which network they are attached to. VCC enables the termination of calls in either domain and also enables continuing the call from one domain to another as the terminal changes the access technology it is camped on. VCC is therefore a potentially fundamental component in many FMC applications described in this book.

An MSC can act as a point of interconnection toward the PSTN for outgoing calls, so that they are optimally routed to the PSTN destination if necessary. MSCs interact with the HLR using the IS-41 [101] interface in CDMA and the Mobile Application Part (MAP [102]) interface in GSM. MAP interface variations include:

- MAP-D between VLR and HLR
- MAP-C between the GMSC and the HLR
- MAP-E between MSCs to prepare handover

The MSC also implements circuit-switched supplementary services based on settings and an activation status provided by the HLR (or retrieved from the VLR in the last visited MSC) when a subscriber is accepted by an MSC. This is in contrast to relying on a centralized execution of service logic as in the case of CAMEL-based services, which follow an Intelligent Network model of decoupling service execution from switching.

When integrating a CS network with a PS (for example, an IMS-based PS) network in a converged environment, the operator needs to make sure that services invoked and triggered in the CS network have their state updated or kept in sync in the PS network. Ideally, the services themselves should be evolved to be executed in a single centralized location (logical and/or physical). The example of such centralization is presented by IN and the use of IMS application servers, when dual-mode CS/PS terminals are being used.

Mobility Management MSCs are directly involved in *mobility management (MM)* of users in the idle and active states of both GSM and UMTS, by handling location update procedures and by being involved in handovers between MSCs or between different nodes in the radio access directly attached to the MSC. When the inter-MSC mobility events occur, the new MSC updates the location of the subscriber with the HLR, so that the HLR knows how to route mobile-terminated calls for the subscriber, or how to enforce some asynchronous actions (e.g., purge the subscriber identity and profile from the last known MSC).

Handling mobility management implies interaction with the terminal, which is also necessary for call control. The GSM and UMTS systems use an ISDN-access-signaling-like interaction with the terminal for call control, specified in 3G TS 24.008 [180].

The interfaces to the radio access network of GSM and UMTS, respectively, as shown in Figure 4.1 earlier in the chapter, are called the A interface and the $I_{u\text{-}cs}$ interface. The $I_{u\text{-}cs}$ interface is the interface between the UMTS access and the 3GPP CS core, and it also has an $I_{u\text{-}ps}$ component for the PS domain core. The A interface assumes an E1 (or T1) transport to be available, while the $I_{u\text{-}cs}$ interface assumes an ATM or IP-based transport.

Transcoding The transcoding between the PCM encoding used in the PSTN and the voice codecs used in GSM (AMR, or adaptive multirate voice coding) happens in the base station controller (BSC). Specifically, the *transcoding unit (TRAU)* is a function of the BSC dealing with transcoding. The *A* interface therefore carries the PCM-encoded voice-over-TDM multiplexed channel.

In UMTS, this model changes, with transcoding taking place in the core or preferably avoided altogether with the Transcoder-Free Operation (TrFO) option, where the same encoding of voice is used between terminals in mobile-to-mobile calls. In the TrFO model, the voice quality is therefore quite substantially improved because it avoids several transcodings necessary in the end-to-end data path. For communication with the PSTN, the transcoding between the AMR [63] (or wideband AMR [64]) codec used in UMTS and the PCM codec used in the PSTN takes place as close as possible to the point of interworking with the PSTN.

Bearer-Independent Circuit-Switched Network (BICSN) The UMTS TrFO feature has been defined in a server-based architecture introduced in 3GPP Release 4. By enabling packet-based transport of voice, it is possible to eliminate the need of transit switches in the core network and as such the need to convert AMR- or WB-AMR-encoded voice into PCM for switching in a classic TDM framing. This server-based architecture for the CS core, also known as a bearer-independent circuit-switched network (BICSN), is represented in Figure 4.3.

The attentive reader has already noticed that the difference between the classic architecture and the BICSN is that the MSC is split into a media gateway and MSC server components. MSC servers control the media gateway via the H.248-based M_c interface. The N_b interface supports the transport of the media components between gateways, and the N_c interface, based on BICC signaling, is used to let MSC servers interact with one another to establish calls or set up inter-MSC legs during handover. In addition, starting with Release 4 of the 3GPP specifications, the HLR has been replaced by the new functional element called the Home Subscriber Server (HSS) as a part of the IMS framework. The HSS extends the HLR functionality to support additional interfaces for the interaction with IMS entities, via DIAMETER-based interfaces.

The 3GPP PS Core The PS core of GSM and UMTS, shown in Figure 4.4, is in many aspects similar to the CS core, in that there is a gateway entity, called the gateway

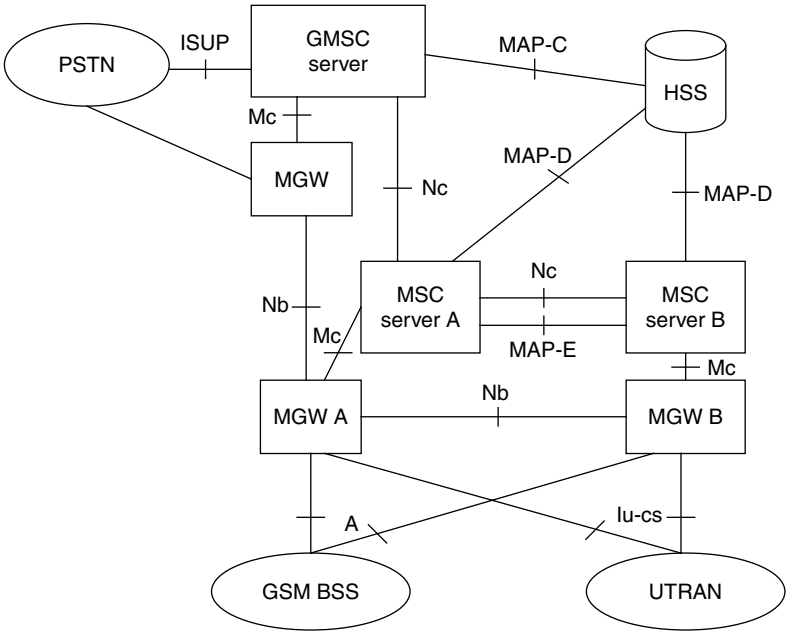

Figure 4.3 Bearer-independent CS network

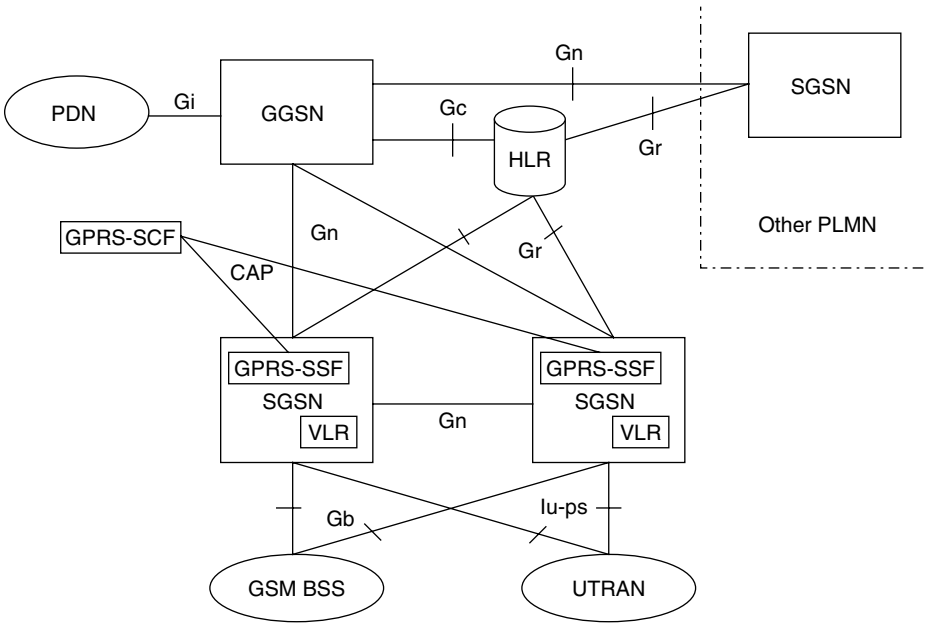

Figure 4.4 The GPRS core network

GPRS support node (GGSN), and a serving entity called the Serving GPRS support node (SGSN) supporting packet data mobility. GPRS itself stands for General Packet Radio Service, which is an extension of the GSM system designed to support packet data services. GPRS is optional to GSM but is an integral component of UMTS from the outset, so technically speaking, the support of GPRS core and services is "not an option," if we can say so, when deploying a UMTS system.

Similarly to the CS core, it is also possible to use CAMEL Intelligent Network services in GPRS, mainly to support prepaid charging models. This is based on the interaction of the SGSN-hosted gprs-SSF, where the CAMEL triggers are armed, with the gprs-SCF. Since the Intelligent Networking subsystem had its roots in the legacy circuit domain, it often does not provide optimal solutions for packet data applications. For example, since deep packet inspection and per-flow charging and QoS policies are better supported at the GGSN, the prepaid and other charging models and services traditionally associated with Intelligent Networking are now transitioning to being handled by the DIAMETER-based interfaces between the GGSN and servers specialized for these functions. This trend is also in alignment with other sectors of the industry (e.g., the RACS/RACF subsystems use DIAMETER-based interfaces for QoS policy control).

The interconnection between the elements of the GPRS core and the HLR (or HSS, in 3GPP Rel-4–based systems) happens via interfaces called G_r (between SGSN and HLR) and G_c (between GGSN and HLR). The G_r interface is used for user data download in the SGSN and location update, while the role of the G_c interface is linked to the support of the network-initiated data sessions feature. This feature works only for statically assigned IP addresses, and it has not proven to be very popular, as IPv4 addresses are a scarce resource and therefore are rarely statically assigned to mobiles. On the other hand, although IPv6 is supported by GPRS standards, and it does not suffer from the issue of scarcity of IP addresses like IPv4, this IP version has not been widely adopted in commercial deployments yet.

The interface between the SGSN and the UMTS access is called $I_{u\text{-}ps}$, and the interface toward the GSM access is called G_b. The $I_{u\text{-}ps}$ interface assumes that either IP transport or Asynchronous Transfer Mode (ATM) transports are available. The G_b interface assumes a Frame Relay transport. Other than these distinct Data Link–layer interfaces and some difference in QoS capabilities defined for GSM and UMTS, there is virtually no difference between the GPRS core for GSM and that for UMTS. The most notable functional allocation difference between GSM and UMTS is that the functions of header compression and ciphering of user data are in the core (and more precisely in the SGSN) for GPRS operating in G_b mode, and in the UTRAN for the I_u mode of operation. "G_b mode" is a way to identify the 3GPP PS core for GSM access only, and "I_u mode" is a way to identify the 3GPP PS core for UMTS access.

The GPRS Tunneling Protocol (GTP) is the protocol for the transfer of user-plane packets between the UMTS access and the UMTS core. GTP is also used over the G_n interfaces between SGSN and GGSN, and between SGSNs during handover. The G_n interface, unlike the I_u interface, uses GTP for both control and user planes (that is, both the GTP-C and GTP-U versions of GTP, specified in 3GPP TS 29.060 [181]). In the $I_{u\text{-}ps}$ interface, only GTP-U is used and the control is based on the RAN Application Part (RANAP) protocol.

In the course of 3GPP Release 7 development, an option to bypass the user plane of the SGSN under certain conditions (i.e., nonroaming user, no lawful interception activated, no CAMEL services enabled) has been introduced for the I_u mode of operation, thus enabling a direct tunnel between the access network and the GGSN. This feature seems to find its justification in higher data rates introduced in the UTRAN in 3GPP Rel-6 and Rel-7.

In the roaming case, the GGSN can be located in the home network or in the visited network, depending on the roaming agreements. This is different from the CS core, where the GMSC is always in the home network, regardless of the user location. When the GGSN is in the home network and the SGSN is in the visited network, the G_n interface connecting them becomes the G_p interface (which is only a difference in names, rather than a protocol-level difference).

Even though the capability to use a GGSN in the visited network is foreseen by the GPRS specifications, this option has not yet been used by operators as part of their roaming agreements. This is perhaps because of the practices linked to a legacy of operation of CS networks, but it is also because the operators want to have tight control of the user traffic at all times, so they wish to enforce policies or perform deep packet inspection in the home network when users are roaming.

The GGSN connects to external networks (also identified with the acronym PDN, or Packet Data Network) via the G_i interface, and the selection of an external network by the UE is done by submitting an access point name (APN) when the first Packet Data Protocol (PDP) context is created for a PDP session. A PDP context identifies a bearer used for a GPRS data session. For a single data session there can be one or multiple PDP contexts (up to 11 in UMTS), depending on the different levels of QoS used. Operator policies may determine the value of a default APN for the user when one is not specified at PDP context activation, and also the level of QoS allowed for a PDP context.

From 3GPP Release 7 onward, a framework to enforce QoS and charging policies on a per–IP flow basis has been introduced, so that the allowed QoS and the charging for each IP flow can be controlled based on the information derived from the applications associated with these flows. This framework is known as Policy Control and Charging (PCC), specified in 3GPP TS 23.203 [52]. As illustrated in Figure 4.5, PCC enables the

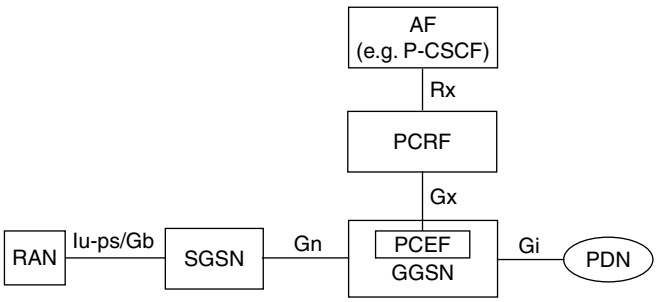

Figure 4.5 PCC framework

GGSN to act as the Policy and Charging Enforcement Function (PCEF) and interact with a Policy and Charging Rules Function (PCRF) via a DIAMETER-based G_x interface to enforce PCC decisions taken by the PCRF. The PCRF bases its decisions on the interactions with an Application Function (AF) via the DIAMETER-based R_x interface.

As VoIP becomes popular within wireline networks, it is likely that, given its acceptance, it will gain prominence in the cellular environment too. The IP Multimedia Subsystem has been defined to support VoIP service via the cellular PS domain in both 3GPP and 3GPP2. When VoIP is deployed in cellular, the PCC infrastructure is also introduced to support prioritization of voice flows (especially those related to emergency calls) and to provide fine-granularity QoS as well as gating of undesirable data that is not related to ongoing and accepted voice sessions. In the case of IMS, the Application Function is the Proxy CSCF (P-CSCF).

Radio Access Network and Terminal Aspects Since aside from minor differences, GSM and UMTS share a common core, the substantial distinction between GSM and UMTS mobile systems is confined to the radio access network. Besides Physical-layer differences, where multiple access to the radio resources is based on TDMA for GSM and W-CDMA for UMTS (at least in its terrestrial aspects), the radio access networks for GSM and UMTS differ in both the respective network elements and their capabilities.

The GSM Radio Access Network The GSM base station subsystem, depicted in Figure 4.6, includes a base station controller (BSC) connected via A-bis interfaces to a number of a base transceiver stations (BTSs). BTS supports radio frequency transmission to and reception from the mobile station (MS) via the U_m interface. The BSC, in addition to supporting inter-BTS mobility and encryption of voice communications, also performs transcoding between the voice-encoding format used over the radio interface and that used in the PSTN.

When the optional GPRS feature is added, the BSC supports the PCU (Packet Control Unit) used to add the packet data transmission capabilities without impact on the rest of the access network. When EDGE is deployed, theoretically raising the data rates to approximately 300 Kbps, then some modification to the BTS is needed.

Recently 3GPP has defined the I_u mode of operation for GSM, where the GSM BSS (depicted in Figure 4.6) is made compatible with the I_u interface used to access 3GPP core. However, since this would not change the overall GSM capabilities for the end user, which are limited by the BSS itself, and, especially, GSM terminal capabilities,

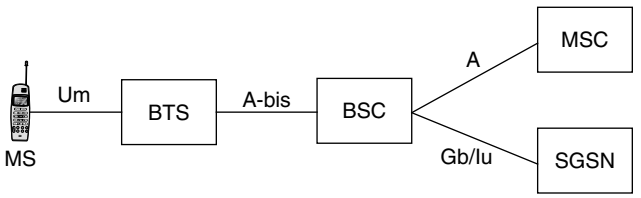

Figure 4.6 The GSM BSS (base station subsystem)

this mode of operation for the GSM BSS has not been implemented in real-life systems. Another recent set of enhancements is related to the potential need to support real-time services in GSM (to allow the reuse of the GSM BSS for VoIP applications), which also includes a packet handover capability (to allow shorter interruption time for real-time services when SGSN and BSS changes occur as the user moves).

GPRS Terminal Classes and Dual Transfer Mode (DTM) As specified in 3GPP TS 23.060 [182], there are different classes of GPRS terminals, with the class indicating the mobile phone capabilities:

- **Class A** These mobile phones can be attached to both GPRS and GSM services simultaneously.
- **Class B** These mobile phones can be attached to both GPRS and GSM, using one of them at a time. Class B enables making or receiving a voice call, or sending/receiving an SMS during a GPRS connection. During voice calls or while sending SMS, GPRS services are suspended and then resumed automatically after the call or SMS session has ended.
- **Class C** Mobile phones are attached to either a GPRS or GSM voice network; they cannot support two services simultaneously.

Since UMTS supports simultaneous PS and CS access, operators of GSM networks or mixed GSM/UMTS networks need to emulate the UMTS behavior in scenarios with patchy coverage (where data and voice sessions need to switch frequently between the two systems) and to provide subscribers with a consistent user experience whether they are camped on a 3G-capable network or not.

The definition of the GPRS class A mode of operation assumes a total independence between the CS and PS domains. This complicates the internal architecture of the terminal, so the optional capability known as Dual Transfer Mode (DTM) has been introduced, permitting the emulation of a class A terminal using DTM-capable terminals using *suspension of data transmission*. This is both a terminal and a network feature, so support on the terminal side alone is not sufficient for it to work.

The UMTS Terrestrial Radio Access Network (UTRAN) The UMTS terrestrial radio access network (depicted in Figure 4.7) by design supports concurrent PS and CS access, and it has interfaces (named $I_{u\text{-}cs}$ and $I_{u\text{-}ps}$, respectively) to both the CS and PS core networks. These interfaces are supported by a radio network controller (RNC), which acts as a concentrator of traffic to and from many Node-Bs. (Node-B is the name of the network entity equivalent to a BTS in GSM.) The Node-Bs are connected to the RNC via an I_{ub} interface.

Since in CDMA technology a mobile terminal, or user equipment (UE) in UMTS terminology, can send to and receive from multiple base stations, the RNC combines and splits signals over the "active set" of base stations, in what is called the "Soft Handover" capability.

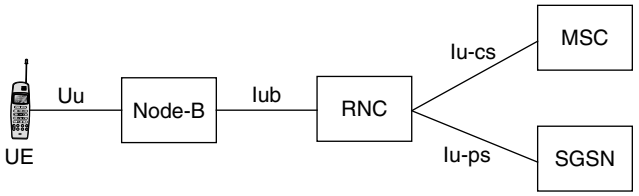

Figure 4.7 The UMTS terrestrial radio access network—UTRAN

3GPP2 Systems

In the family of 3GPP2 systems, we shall include the 2G and 3G CDMA systems, defined by ANSI/TIA and 3GPP2, respectively, despite the fact that the 2G system had been defined by TIA and not by 3GPP2 (like GSM was not defined by 3GPP, but by ETSI). The 3GPP2 system, also known as CDMA2000, is in fact an evolution of the cdmaOne system, known as the IS-95 family of CDMA technologies. These are defined in the TIA/EIA IS-95 specification, including the IS-95A [98] and IS-95B [99] revisions, which describe a complete mobile system. The world's first 3GPP2 3G commercial system was launched by SK Telecom in South Korea in October 2000 using CDMA2000 1X.

The IS-95A revision of IS-95 was published in 1995, and it was the one used for commercial 2G CDMA system deployments. It describes a wireless system based on 1.25 MHz CDMA channels, and it can provide voice services as well as circuit-switched data connections up to 14.4 Kbps. The IS-95B revision defines the support for *packet data services,* so it is sometimes categorized as a 2.5G system. IS-95B systems in fact offer 64 Kbps packet-switched data, in addition to voice services.

CDMA2000 technologies were defined by 3GPP2 and belong to a family of cellular wireless systems accepted as IMT-2000 technologies (i.e., 3G) by the ITU-T. This family includes

- CDMA2000 1x
- CDMA2000 1xEV-DO (Rev 0, Rev A, Rev B)
- Ultra Mobile Broadband, or UMB (previously known as CDMA2000 1xEV-DO Rev C), which represents an evolution toward 4G

CDMA2000 1x supports both circuit-switched voice communications and packet data speeds of up to 307 Kbps (as peak data rate). CDMA2000 1xEV-DO (Evolution-Data Optimized, or -Data Only) introduces a high-speed data broadband wireless network that can offer theoretical peak data rates beyond 2 Mbps in a mobile environment (in practice, the data rates consistently reach into 500 Kbps but rarely beyond that). CDMA2000 1xEV-DO is specified in IS-856, as the CDMA2000 High-Rate Packet Data Air Interface (also known as HRPD).

The 3GPP2 Circuit Core The CDMA CS core network (used for IS-95 and CDMA 1x, as EV-DO is a PS-only system and relies on CDMA 1x to provide CS services) follows

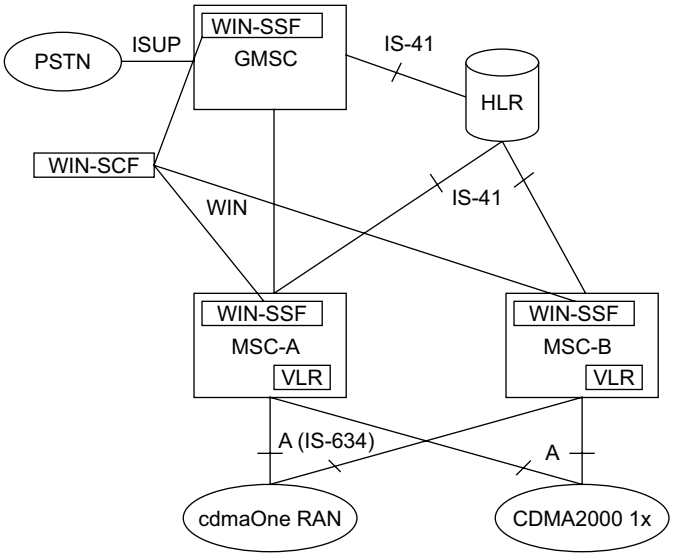

Figure 4.8 The CDMA CS core

similar principles to those of the GSM core, although there are protocol and mainstream implementation differences. The CDMA CS core in fact includes an MSC and a GMSC, as it is shown in Figure 4.8, and its interface to the RAN is also called the A interface; it follows the IS-634 [100] specifications.

Unlike in 3GPP systems, in CDMA, base station controller (BSC) and MSC are often implemented in a single physical node. In other cases, they communicate over a proprietary interface. The MAP protocol used in 3GPP systems to interface with the HLR is replaced by the TIA/EIA IS-41 [101]–specified protocol. Since this protocol needs to be used in roaming cases to access subscriber information for service authorization purposes, roaming between a CDMA network and a GSM/UMTS network is not possible without an interworking function between MAP and IS-41 (and of course the user needs a CDMA- and GSM/UMTS-capable phone). In addition, the flavor of the IN protocol (WIN) used in 3GPP2 is different from the CAMEL Application Part used in 3GPP.

The 3GPP2 Packet Core Unlike the 3GPP and 3GPP2 circuit cores, which are essentially similar, the 3GPP2 packet core, shown in Figure 4.9, is quite different from the 3GPP core not only in its protocols, but also in its overall design. CDMA operators have chosen to adopt a packet core that is fully optimized for data, so they have used design principles more aligned with practices used by the Internet service providers. Therefore, they based their standardization efforts on reusing existing IETF protocols as they were, or promoted the enhancements of existing IETF protocols, rather than inventing their own.

Specifically, there was no need to define a special IP routing or tunneling setup (and its special and limited address ranges) for a walled roaming network like in GPRS,

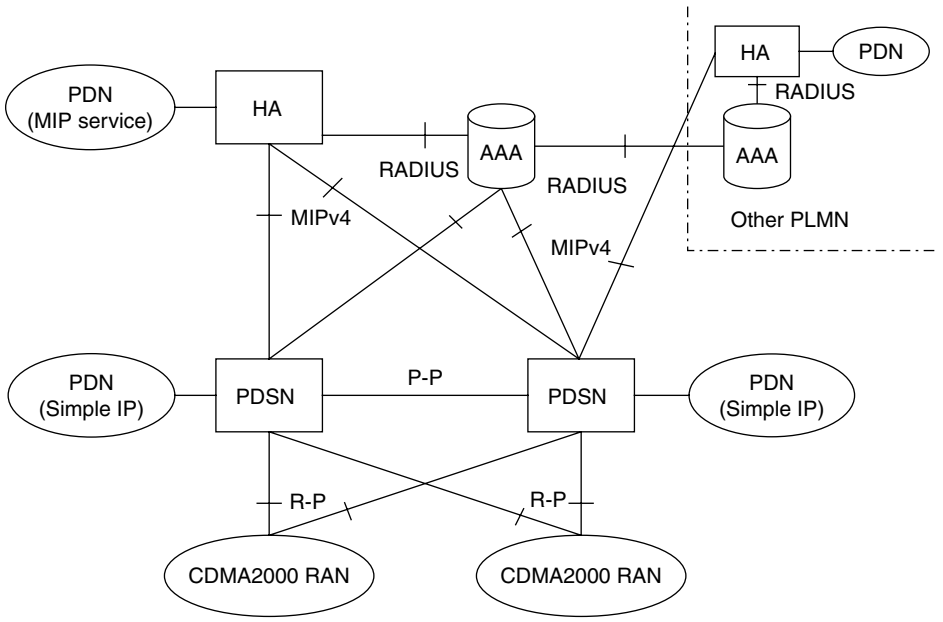

Figure 4.9 The CDMA2000 packet core

called GRX (GPRS Roaming Exchange), as roaming can work across the Internet using Internet Mobility augmented by the AAA support protocols that can withstand the security challenges encountered in the Internet.

Also, the HLR in CDMA networks is now used only for voice applications, as data subscribers' profiles are held in the AAA server, much like in all fixed IP networks. This approach is not aligned with the decision taken by 3GPP operators to keep storing subscribers' profiles for data in the HLR, or to evolve MAP to also support data applications. Similarly, there are no Intelligent Network–based services for the CDMA packet core, and all services and charging models are based on the use of AAA and non-IN-based frameworks.

The primary element in a CDMA packet core is the packet data serving node (PDSN). It is defined in the TSG-X X.S0011-001-D "cdma2000 Wireless IP Network Standard: Introduction" [108]. This node terminates PPP connections from CDMA terminals and performs user authentication for network access. It can also provide network access services and limited mobility within the scope of the BSCs directly attached to it via an R-P interface. The R-P interface is based on GRE encapsulation for the user plane and a protocol derived from Mobile IPv4 for the control plane. There is a homing relationship between PDSNs and CDMA2000 access network BSCs, so the PDSN's anchored mobility is limited geographically. The service provided to subscribers based on using a PDSN's anchored mobility only is called "Simple IP."

To achieve wider area mobility, that is, mobility between PDSNs, it is necessary for the terminal to support Mobile IP (MIP).[5] The network operator must deploy a MIP Home Agent (HA) and support a MIPv4 Foreign Agent (FA) for MIP4 or a MIPv6 access router for IPv6. Note that in real-life implementations the FA and access router functionality are typically supported in the PDSN platform. A MIP HA may be allocated to subscribers statically or dynamically. The latter is performed via AAA interaction during the Mobile IP registration phase. Simple IP and Mobile IP modes of operation are defined in X.S0011-002-D "cdma2000 Wireless IP Network Standard: Simple IP and Mobile IP Access Services" [109].

With the support of Mobile IP and a relatively technology-neutral approach to data services support, CDMA networks could be poised to be transformed into converged multiaccess technology networks in a more straightforward way than their GSM/UMTS counterparts.

Radio Access Network and Terminal Aspects Let's start the discussion of the CDMA systems RAN with the analysis of the data rates and services capabilities supported by various flavors of CDMA 3G systems.

CDMA2000 1xEV-DO Rev 0 supports bidirectional peak data rates of up to 153 Kbps and an average of 60–100 Kbps in commercial networks in a 1.25 MHz channel. Release 1 can deliver peak data rates of up to 307 Kbps. CDMA2000 1x handsets are backward compatible with cdmaOne systems, so dual-mode capability is not needed for a commercial terminal to use both 2G and 3G, unlike in 3GPP technologies.

CDMA2000 1xEV-DO Rev 0 offers peak data rates of up to 2.4 Mbps in the "forward link" (also known as downlink in 3GPP) and 153 Kbps in the reverse link (or uplink in 3GPP parlance), in a single 1.25 MHz FDD carrier. As usual, this is to be checked with reality, and in most cases in commercial networks, CDMA2000 1xEV-DO Rev 0 may deliver average (or sustained) throughput in the range of 300–700 Kbps in the forward link and 70–90 Kbps in the reverse link.[6]

To cope with initially spotty coverage of CDMA2000 1xEV-DO, CDMA2000 1xEV-DO devices include a CDMA2000 1x modem to be compatible with CDMA2000 1x and cdmaOne systems (it should not be forgotten that a CDMA 1x device is also compatible with cdmaOne).

CDMA2000 Evolution CDMA2000 1xEV-DO Revision A (also known as DOrA) is an evolution of CDMA2000 1xEV-DO Rev 0 providing higher peak data rates, offering more symmetric performance between forward and reverse links, and, most important, supporting QoS levels compatible with delay-sensitive multimedia applications, including VoIP and streaming video. Rev A supports 3.1 Mbps in the forward link and 1.8 Mbps in the reverse link in a 1.25 MHz FDD carrier.

---

[5] We invite those willing to probe further on Mobile IP details to check out a fundamental book on the subject by James Solomon, *Mobile IP, The Internet Unplugged* [165].

[6] Source: CDMA Development Group (CDG).

The first commercial deployments of DOrA systems took place in the course of the year 2006, at the same time as High-Speed Downlink Packet Access (HSDPA). Note that HSDPA enhanced with improved uplink performance, known as HSUPA, was not yet deployed at the time. In commercial networks, Rev A average (or sustained) throughput is in the range of 450–800 Kbps in the forward link and 300–400 Kbps in the reverse link[7] with typical latency as low as 50 ms. Multicast capabilities have also been added to DOrA by means of OFDM technology to enable multicast content delivery.

The CDMA2000 1xEV-DO Rev A capability to support VoIP makes it an ideal technology for converged networks where voice and other multimedia services such as video telephony and push-to-talk are supported over a single all-IP infrastructure. Improved reverse link speeds allow users to send and receive large amounts of data in ways previously possible only in fixed networks or over Wi-Fi.

DOrA networks support existing CDMA EV-DO Rev 0 applications and devices. This backward compatibility preserves an operator's network investments. Multimode devices, however, are needed to access CDMA 1x and cdmaOne. CDMA2000 1xEV-DO Revision B [110], defined in 3GPP2 C.S0024-B, evolves Rev A by aggregating multiple DOrA 1.25 MHz channels to provide higher performance for multimedia services, bidirectional data transmissions, and VoIP-based services.

Rev B Peak data rates are proportional to the number of carriers aggregated (Rev B is a "multicarrier" version of Rev A). So, in principle, if a 20 MHz spectrum bandwidth is available, fifteen 1.25 MHz Rev A channels can be combined, and peak data rates of 46.5 Mbps in the forward link and 27 Mbps in the reverse link could be achieved. By using a 64-QAM (Quadrature Amplitude Modulation) modulation scheme, forward link peak data rates of up to 4.9 Mbps per 1.25 MHz carrier are achievable, which means fifteen carriers can deliver up to 73.5 Mbps in the forward link direction. It should be noted that in these multicarrier setups, the 1.25 MHz carriers do not have to be adjacent to one another, giving operators significant flexibility in their deployments. Device compatibility considerations are similar to those illustrated for Rev A.

UMB UMB is yet another step on the way of CDMA evolution. UMB specs are being completed by 3GPP2 at the time of this writing. UMB promises to deliver peak data rates of up to 280 Mbps on the forward link and 68 Mbps on the reverse link; a very low latency below 20 ms; and scalable, noncontiguous, and dynamic channel (bandwidth) allocations of 1.25 MHz, 5 MHz, 10 MHz, and 20 MHz.

The CDMA radio access network architecture differs depending on whether a plain IS-95 and CDMA2000 1x access (see Figure 4.10) or a CDMA2000 1x EV-DO access is used. It should be remembered from the discussion so far that the only CDMA access networks supporting CS service are CDMA2000 1x and IS-95, and EV-DO is an enhancement that leads to data-only operation and to the possibility to support real-time communications over a PS domain (in Rev A).

[7] Source: CDMA Development Group.

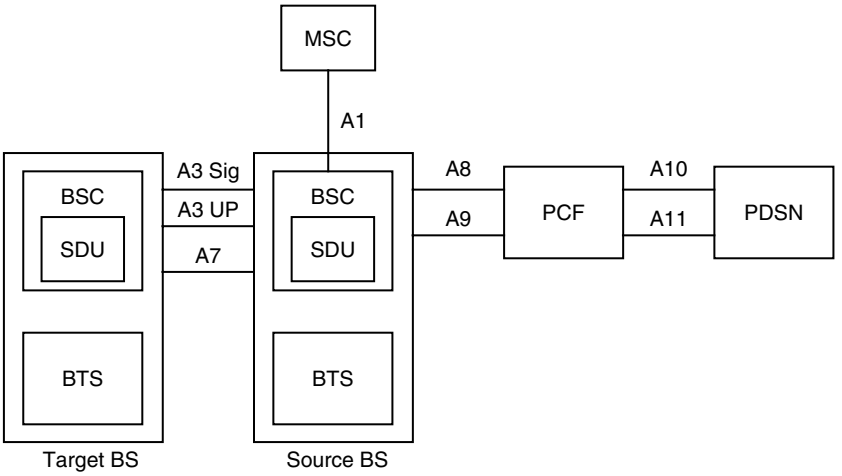

Figure 4.10 The CDMA access network

CDMA RAN The CDMA RAN for CS services only is made of a base station, in most commercial implementations comprising a base station controller (BSC) and a base transceiver station (BTS). The standards do not define an open interface between the BTS and the BSC, although they may be based on different physical nodes. The BSC supports functions such as the SDU (selection distribution unit) necessary to support CDMA soft handover (requiring the BSC to perform frame selection and distribution to an active set of base stations).

To support packet data in CDMA, the RAN is augmented with a packet control function (PCF). The PCF is normally implemented as a part of the same physical platform as the BSC. The session management and mobility management functions are normally located in the core network, but the option to place these functions in the PCF has also been introduced in the course of the HRPD specification. In the latter case, the interworking with an MSC supporting this function in the CDMA 1x network requires an IWS (interworking solution) module in the PCF. This option, depicted in Figure 4.11, is specified in the 3GPP2 A.S0009-A "Interoperability Specification (IOS) for High-Rate Packet Data (HRPD) Radio Access Network Interfaces with Session Control in the Packet Control Function" [95].

The RAN architecture also evolves to support the concept of IP-based RAN and more efficient distribution of functions between the BSC and the BS. One example is the placement of the scheduling function in HRPD. Up to CDMA 1x, the CDMA RAN is based on a BSC connecting to a base station using some form of TDM transport (T1s). The introduction of scheduling in the base station makes it possible for IP-based RAN to be deployed for CDMA 1x EV-DO networks, and thus benefit from lower backhaul costs and improved distribution of network elements in the network. This also reduces OPEX for operators, as they can reduce the number of BSC hosting locations.

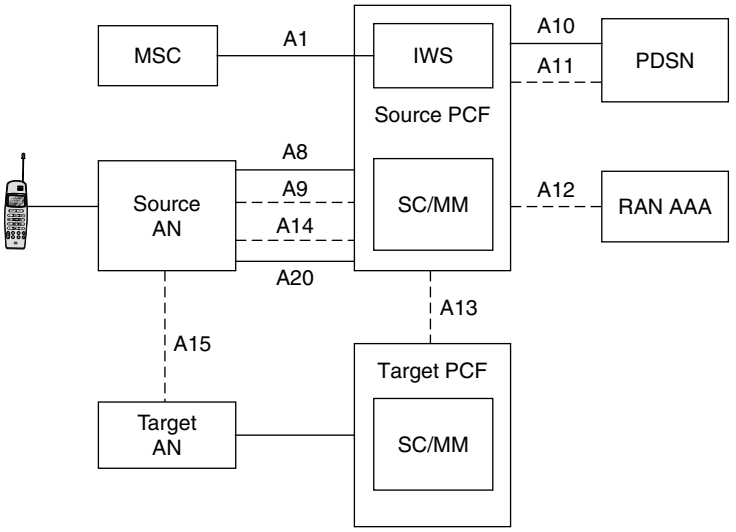

Figure 4.11 CDMA access network for packet data with Session Control in the packet control function

The Need for VCC Devices supporting 1x and Rev A cannot use both systems at the same time. When VoIP is rolled out in areas of Rev A coverage, the deployment of Rev A is unlikely to be ubiquitous, so when a voice call is started in an area of the RAN where Rev A is supported, it may be the case that due to varying radio conditions or limited Rev A coverage, the call will need to be "handed down" to 1x circuit-based voice and vice versa.

This requirement has lead to the definition of a Voice Call Continuity specification for the handover of voice calls between HRPD and CDMA 1x (that is, packet- and circuit-based systems), based on the assumption that the two technologies cannot be accessed at the same time. The system requirements for the support of this mechanism are defined in 3GPP2 S.R0108-0 "HRPD-cdma2000 1x Interoperability for Voice and Data System Requirements" [112], and the specification of the actual handover procedures is in 3GPP2 C.S0075-0, "Interworking Specification for cdma2000 1x and High-Rate Packet Data Systems" [183].

A Look into the Future

As 3G cellular systems have been defined, deployed, and enhanced, the 3GPP and 3GPP2 communities are looking ahead to shape the next generation of mobile systems. In doing so, they are targeting convergence of the standards by evolving 3GPP and 3GPP2 systems in the same direction. In large part, this direction is driven by some prominent 3GPP2 operators interested in getting access to the larger embedded base of 3GPP subscribers and realizing better economies of scale.

The evolution of the 3GPP system is being investigated as part of the System Architecture Evolution (SAE) work for the architectural aspects, and in the Long Term

Evolution (LTE) project for the radio access network aspects. The resulting system will be a PS-only system (no circuit services will be supported) delivering data rates in the order of 100 Mbps in the downlink and 50 Mbps in the uplink. The new radio transmission technology, known commonly as LTE after the project name or E-UTRA, is supported by the evolved UTRAN (E-UTRAN).

The 3GPP Evolved Packet System As defined in the standards, the E-UTRAN will be substantially simplified and will consist of a single element called *Evolved NodeB (E-NodeB).* The E-NodeB element, roughly speaking, groups all the functions once supported by the UMTS Node B and RNC. Like the UMTS RNC, the E-NodeB supports the encryption and header compression functions for the user plane. The interface between the E-NodeB and the core network, equivalent to the $I_{u\text{-}ps}$ interface in UMTS, is based on the GTP protocol for the user plane and the *S1* application protocol for the control part.

The 3GPP Evolved Packet Core network (EPC) is made by a control-plane entity called the *mobility management entity (MME),* whose functions include paging and mobility management. As the MME is responsible for the signaling interface with the UE, it also supports the relevant security functions, including authentication and signaling traffic encryption. The user plane is handled by two entities: the serving gateway (SGW) and the PDN gateway (PGW). The overall architecture, documented in 3G TS 23.401 [120] and 3G TS 23.402 [121], is depicted at a high level in Figure 4.12 (the references can provide more detail).

The SGW concentrates the S1 interface user plane to E-NodeBs, buffers data in the downlink, and triggers paging when the UE is in idle mode. The SGW may also anchor mobility toward legacy UMTS and GSM SGSNs. The SGW interacts with the MME via an open interface. There is a single SGW per UE at any given time, and there is

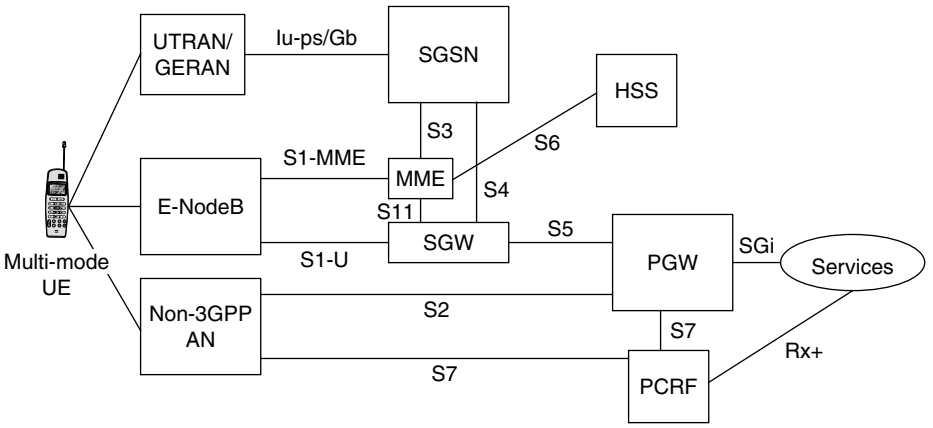

Figure 4.12 High-level SAE architecture

no geographical homing relationship between an E-NodeB and an SGW, improving reliability by avoiding a single point of failure in the system.

The PGW provides access to packet data networks. Potentially a UE may access multiple PDNs concurrently via one or multiple PGWs through the SGW. The PGW also supports a Mobile IP HA functionality to provide for network-based (Proxy-MIP [117]) or client-based (Mobile IP [118]) IP mobility support with non-3GPP access networks over the S2 interface.

Support of mobility across a variety of access technologies, documented in 3G TS 23.402 [121], is one of the most important features of the evolved system, and it is also among the most controversial ones. In fact, the deployment of IETF-based IP mobility in the 3GPP system to support non-3GPP access may lead to the eventual introduction of MIP-based roaming interfaces instead of or in addition to the GTP-based mobility support and roaming interfaces that have been so far typical for 3GPP systems. Changing roaming interfaces will of course imply operations changes as a consequence, a fact that, for obvious reasons, is not being accepted well by many operators, so there has been some friction on this topic.

The 3GPP2 System Evolution As far as 3GPP2 is concerned, its system evolution has identified a UMB radio interface (Ultra Mobile Broadband), and there is a great deal of cooperation with 3GPP in order to share the core network architecture and to define interworking between LTE and CDMA systems, thus placing 3GPP2 into an evolutionary path toward the same system architecture (and potentially also the same radio interface) as 3GPP.

WiMAX: A Migration Away from Traditional Cellular Systems

Unlike the majority of the cellular systems reviewed in previous sections, WiMAX is a mobile system designed to deliver only packet data services. WiMAX is based on the IEEE 802.16 [114] radio interface, which can operate in both licensed and unlicensed spectrum bands. As with cellular systems, it is possible to use WiMAX terminals equipped with RUID, USIM, or SIM cards. However, in contrast to cellular, which requires a purchase of a smart card or a telephone as part of the subscription, the WiMAX model also allows a casual setup of network access, similar to today's Wi-Fi hotspots. Average bit rates advertised of WiMAX systems are up to 70 Mbps, but of course, this figure needs to be regarded with some caution, as in practical deployments these may be significantly lower (in the order of 10 Mbps average data rate).

WiMAX was not born as a mobile system per se; it was thought of as more of a wireless alternative to cable and DSL for fixed broadband applications and the evolution of Wi-Fi. But when a wide area wireless system is available, the temptation to also make it usable for mobile applications is irresistible. For this very reason, the IEEE was prompted to define various flavors of IEEE 802.16.

The original IEEE 802.16 (also known as IEEE 802.16a) specifies a radio interface operating in the 10 to 66 GHz range. The IEEE 802.16d evolution of the standard added the operation in the 2 to 11 GHz bands. Its subsequent update, IEEE 802.16e, defines a radio interface suitable for mobile applications.

Let us also clarify the distinction between different modes of WiMAX operation. The terms "fixed WiMAX," "mobile WiMAX," "802.16d," and "802.16e" are frequently used. This is what they mean:

- **802.16d** This standard is more precisely identified as IEEE 802.16-2004 [115]. However, in the industry this is mostly known as 802.16d.
- **Fixed WiMAX** WiMAX systems use IEEE 802.16d as radio interface.
- **802.16e** Technically speaking, what is commonly known as 802.16e is an amendment to IEEE 802.16-2004. Its most accurate name would be IEEE 802.16e-2005 [116]. This standard supports full mobility, unlike 802.16d.
- **Mobile WiMAX** WiMAX systems use IEEE 802.16e as a radio interface; hence these systems enable mobility (although a mobile WiMAX system can also be used for Fixed Wireless Broadband Access applications).

Standardization and Deployment

The WiMAX Forum is the standards forum that has specified the WiMAX system. Release 1 enables, among other things, the support of interworking with 3GPP and 3GPP2 networks, at the level of providing access to the Internet based on sharing credentials (loosely coupled interworking), but no real handover capability. This and other advanced capabilities are, however, in the scope of the work of the WiMAX Forum.

One of the biggest challenges for WiMAX is the ability to be deployed in a regulated spectrum in a way that does not create a global lack of device-network interoperability (or better, compatibility) for roamers. In fact, regulators seem to give permission to use WiMAX in quite different bands across the globe. For instance, in the U.S., companies like Sprint, Nextel, and Clearwire own the spectrum around 2.5 GHz. In other regions deployments are foreseen in the 2.3/2.5 GHz (the dominant frequency expected in Asia), 3.5 GHz, or 5 GHz band. In the European Union it has also been proposed to allocate some spectrum (yet undefined at the time of writing) for noncellular wireless communications technologies such as WiMAX. Clearly, this situation will force device manufacturers to produce terminals compatible with a large number of operating frequencies, potentially even more than current quad-band GSM handsets support. This may drive complexity and cost in terminals, which is undesirable, as this tends to impact mass market acceptance.

Like cellular systems, the WiMAX system (depicted in Figure 4.13) defines, broadly speaking, an Access part also known as Access Service Network (ASN) and a Core part referred to as Connectivity Service Network (CSN), although some of the classic cellular core functions, such as accounting and user access control, are supported in the WiMAX Access part.

The WiMAX Connectivity Service Network The Connectivity Service Network depicted in Figure 4.14 serves as an interconnection point between the ASN and services networks (e.g., application service providers, the Internet, the operator's own services, etc.). Also, the CSN makes it possible for a subscriber to move between ASNs and

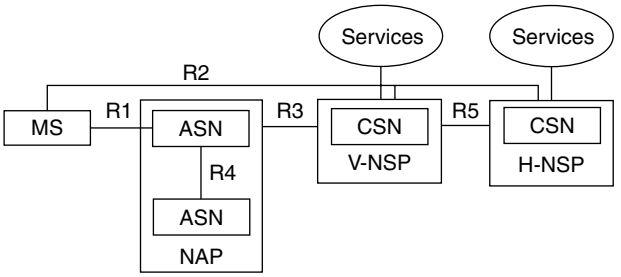

Figure 4.13 The WiMAX high-level architecture

to roam, by providing the necessary anchor point for mobility and the AAA mechanisms necessary to support roaming.

A CSN may in fact act as a AAA proxy toward the home network CSN, with the ASN playing the role of the AAA client, as shown in Figure 4.14. The CSN includes the MIP Home Agent terminating Layer-3 mobility support tunnels, whether they are based on PMIP [117] or client MIP defined in [118] and [119]. The AAA server in the CSN is used to determine whether a user has access rights to the wireless service, and it also may act as a AAA proxy toward the home network or toward customer/services networks connected to the CSN.

The AAA server stores WiMAX subscription profiles (including static QoS profiles) and performs accounting data collection. When dynamic QoS policies are supported, the CSN may include a policy decision point (e.g., a PCRF), but this is not defined in the WiMAX Release 1.0. The enforcement of QoS policies may take place in the ASN.

The visited network is connected to the home network via the R5 interface, and the visited ASN connects to the visited CSN via the R3 interface. It should be noted that the HA may be located in the V-CSN for local breakout of traffic or in the H-CSN for the home-routed

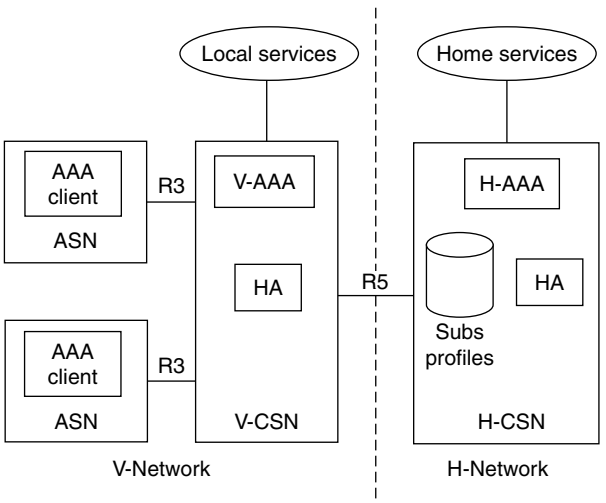

Figure 4.14 The CSN in the WiMAX mobile system

traffic roaming mode. The CSN, in its role of interfacing with the services network, also performs IP address allocation functions and network access authentication.

The WiMAX Access Service Network The WiMAX Access Service Network is composed of two network elements: a base station (BS) and an ASN gateway (also abbreviated as ASN GW henceforth), as shown in Figure 4.15. The functions in the BS and the ASN differ depending on which of the three WiMAX profiles a vendor complies with. In fact, the WiMAX Forum has defined the ASN as a set of functions that can be grouped differently depending on the profile being used. The compliance to a profile ensures that there is a known way to interoperate with nodes that are built according to it.

The ASN gateway may act as the anchor for inter-BS mobility, and support control-plane functions to coordinate with other entities in the ASN itself, the MS, and the CSN. The MS is associated with a single ASN GW; but the BS to which MS is attached can use multiple ASN GWs (for redundancy, to avoid single points of failure, and for load-sharing reasons). Figure 4.15 shows a typical ASN architecture. Different ASNs or different ASN GWs may interact via the R4 interface. The R3 interface connects the ASN GW to one or more CSNs.

An ASN GW may also include a decision point for Dynamic QoS policies and an enforcement point. This decision point may be separate from the enforcement point on the R7 open interface. The BS connects to one or more ASN GWs via the R6 interface; the R8 interface between BSs exists for handover purposes.

WiMAX Profiles The WiMAX standard framework includes three different usage and deployment profiles for a WiMAX system: A, B, and C. These profiles are characterized by different allocations of functions to BS and ASN GW.

The main characteristics of profile A are these:

- Handover control is in the ASN GW.
- Radio Resources Management (RRM) is in the ASN GW, so that it is a centralized function.
- Intra-ASN mobility is based on open interfaces (the open interfaces R6 and R4 are both used).

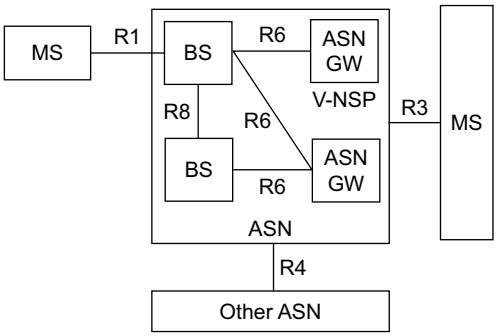

Figure 4.15 The ASN in the WiMAX mobile system

For profile B, *Intra-ASN* mobility is based on proprietary mechanisms, but *Inter-ASN* mobility uses the R4 interfaces. No allocation of functions to ASN GW or BS is specified, due to proprietary nature of the ASN.

For profile C, these are the essential features:

- Handover control is in the base station.
- RRC (Radio Resource Control) is in the BS (more specifically, the RRM in the BS). An "RRC Relay" is in the ASN GW, to relay the RRM messages sent from BS to BS via R6.
- Intra-ASN mobility is based on open interfaces (the open interfaces R6 and R4 are both used).

Table 4.1 summarizes the main characteristics of each profile.

QoS From a QoS perspective, the Release 1.0 specification of the WiMAX system defines the following potential capabilities:

- Preprovisioned service flow creation, modification, and deletion
- Initial service flow creation, modification, and deletion

Table 4.1 WiMAX Profiles and Their Function Allocations to ASN GW and BS

Profile	ASN GW Functions	BS Functions
Profile A	Authenticator Key Distributor Data Path function (user-plane handling) HO function Context server/client MIP Foreign Agent (MIPv4) MIP Access Router (MIPv6) Paging controller RRM Service flow authorization (QoS policy decisions)	Auth Relay Key Receiver Data Path function (user-plane handling) HO function Context server/client Paging agent Radio Resource Agent Service flow management (QoS policy enforcement)
Profile B	Not specified	Not specified
Profile C	Authenticator Key Distributor Data Path function (user-plane handling) HO function Context server/client MIP Foreign Agent (MIPv4) MIP Access Router (MIPv6) Paging controller RR relay Service flow authorization (QoS policy decisions)	Auth Relay Key Receiver Data Path function (user-plane handling) HO function Context server/client Paging agent Radio Resource Agent Radio Resource Control Service flow management (QoS policy enforcement)

- QoS policy provisioning between AAA and SFA
- Service flow ID management

The scope of Release 1.0 is limited to preprovisioned service flows and to IEEE 802.16 radio aspects only of QoS, with no definition of end-to-end behavior. The characteristics of the QoS service over the WiMAX radio are the following:

- Connection-oriented service
- Five QoS services at the air interface, namely:
  - UGS (Unsolicited Grant Service)—for Real Time Constant bit rate service
  - RT-VR (Real Time—Variable Rate)
  - ERT-VR (Extended Real Time—Variable Rate)
  - NRT-VR (Non-Real Time—Variable Rate)
  - BE (Best Effort)
- Provisioned QoS parameters for each subscriber based on a subscription profile
- Policy-based admission of service flow requests

A subscription profile for a user is defined as a number of service flows (which are quintuples {Source IP Address, Destination IP Address, Source Port, Destination Port, Protocol}), each of which is associated with the QoS parameters. This information is provisioned in a policy server (e.g., in an AAA subsystem). With the static service model, a WiMAX terminal is not allowed to change the parameters of provisioned service flows or create new service flows (which instead are both possible within the scope of the dynamic service model). The dynamic service flow creation is triggered by the terminal or the applications accessible via the WiMAX system.

Cellular Friend or Foe: Voice over Wi-Fi

So here we have it. Cellular is synonymous with wireless access and mobile voice communications. Indeed, up until recently that was the only technology supporting truly mobile communications. (Mobile WiMAX technology provides standard support for voice communications; however, at the time of this book's writing, no broad-scale WiMAX voice commercial deployments have taken place.)

Nowadays, however, new technologies such as Wi-Fi, originally designed to be mere extensions of wired Ethernet to provide wireless data network access, are being quickly adopted for mobile voice. With the proliferations of public hotspots and the rise of convergence technologies such as FMC, Voice over Wi-Fi is quickly maturing into a credible complement, if not an alternative, to traditional cellular.

In this section we provide a brief overview of this novel approach to the support of wireless voice service, and analyze the impact of Voice over Wi-Fi (VoWi-Fi) on both the end user and the service provider.

Technology Fundamentals

Wireless local area networking (WLANs) or Wi-Fi[8] (for Wireless Fidelity) has truly taken the world by storm in the recent years. This phenomenon coincided (albeit independently) with the rise of Voice over IP as one of the primary alternatives to traditional fixed PSTN voice communications. So it did not take long for the users and the industry to realize that the two are even more useful when combined into a service offering potentially capable of ending consumer dependency on both traditional PSTN and cellular telephony.

While in theory combining VoIP with Wi-Fi to make it mobile looked like a natural extension of fixed VoIP solutions, early real-world implementations proved to be difficult. This happened mainly because wireless networks based on IEEE 802.11 [126] (a group of standards underlying Wi-Fi) were originally prone to interference, did not support voice prioritization and guaranteed QoS, were not optimized to carry real-time traffic, and so forth (as seen in Figure 4.16). So, to address these deficiencies, the industry had to turn to short-term proprietary implementations while at the same time mounting the concerted standardization efforts needed to make Wi-Fi more voice friendly.

Wi-Fi Standardization Wi-Fi is currently in widespread use all around the world as both replacement and transport media complementary to wired LANs. Indeed, 802.11-based WLANs are essentially a wireless extension of the wired Ethernet LANs standardized

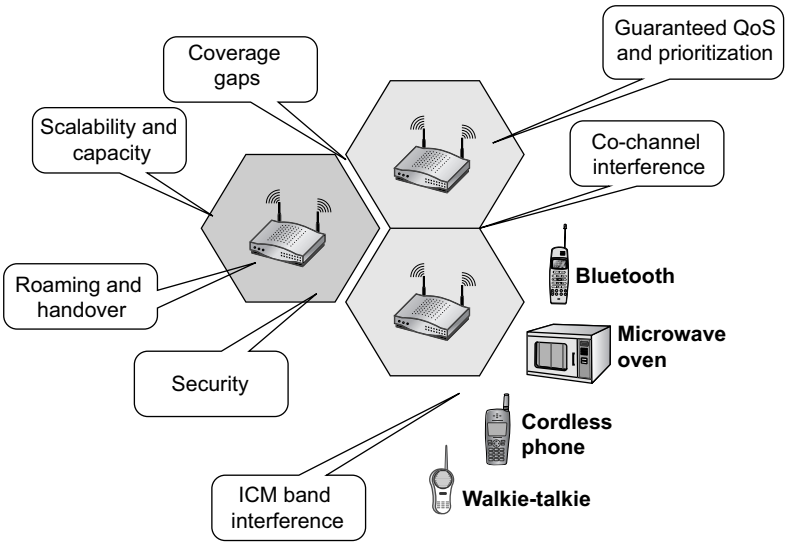

Figure 4.16 VoWi-Fi challenges

[8] We are going to use both terms interchangeably throughout the text for the purposes of discussion.

in IEEE Standard 802.3 [127]. Wi-Fi is another example of wireless technology originally designed to only carry data (similarly to WiMAX). Unlike WiMAX, however, WLANs were not intended to be deployed in wide areas but rather to serve the needs of local, not highly, mobile users. Therefore, original Wi-Fi specifications did not include roaming and handover support.

IEEE adapted standard 802.11[9]—which provided the foundation for Wireless LAN technology—in 1997. The standard was consequently revised in 1999 and continues to be enhanced with amendments and supplements by the same standards body. Following the introduction of 802.11, IEEE later coined a more popular term: WLAN. Independently of IEEE, the Wi-Fi Alliance, a trade organization specifying and testing commercial equipment for interoperability and compliance with IEEE standards, came up with the term Wi-Fi, which soon became a favorite with popular media (presumably for its "catchiness").

The latest version of the standard was approved by IEEE in March 2007 under the name IEEE 802.11-2007, which now combines eight previous amendments: a, b, d, e, g, h, i, and j.

Wi-Fi Components Unlike cellular systems, Wi-Fi technology and networks[10] are not very complex, and their components are fairly simple and inexpensive, which was one of the decisive factors in the world-wide proliferation of Wi-Fi just a few short years after its introduction. A Wi-Fi network requires a client device with a wireless *network interface card (NIC)* and an *access point (AP)* terminating a radio link to multiple clients and connecting the wireless LAN with the wired infrastructure.

While a variety of NICs continue to be available in all shapes and sizes from laptop PC cards to flash and USB dungles, the majority of commercial implementations nowadays incorporate them into mobile devices themselves in the form of a chipset combined with an RF subsystem with a dedicated antenna.

The APs are following a similar trend. While dedicated access points are still widely available, one can more often find them supporting additional functionality such as switching or routing and sold as multifunction devices or even bundled with NICs and other equipment such as desktop PCs and laptops. Strictly speaking, two distinct types of APs are available today: regular APs, also called "fat," or intelligent and simplified ones called "thin." The thin APs are most often deployed in combination with aggregating Wi-Fi switches also called WLAN controllers in large-scale enterprises, or in metro or campus installations.

Further, when used for VoIP traffic, Wi-Fi customer-premises equipment typically must perform even more duties. It must support routing and Wi-Fi networking and also provide support for both VoIP and PSTN telephony via terminal adapter capability for customer legacy equipment. Therefore it is not uncommon to find all three functions embedded into one device. Such an integrated device might be designed to support

[9] After the number of engineers assigned to design it.

[10] Note the important distinction.

two or more service set identifiers (SSIDs), one for VoIP traffic and one for regular data communications. The purpose of this approach is to provide service separation and enable QoS and prioritization solutions as described in the following sections. Figure 4.17 lists the examples of commercial AP implementations and their properties.

While we are at it, let's define SSID. It is a text string of up to 32 characters identifying a common access point domain in a Wi-Fi network. All clients intending to communicate with a particular AP must be programmed with its unique SSID. The original thinking behind the SSID was to use it as an additional security measure. This approach, however, does not provide real security, since even if the SSID is not broadcast by the AP, it can be easily obtained, e.g., by snooping using any of the widely available hardware or software protocol analyzers.

Radio Interface WLAN ranges specified in IEEE standards can reach as far as 300 meters outdoors under perfect conditions with no obstacles or interference present. With the use of amplification devices such as directional antenna arrays, the signals can reach up to 1 kilometer, often even in the presence of obstacles. The distance and obstacles, however, significantly affect the signal strength, which in most cases decreases exponentially with the distance. Once a client device has connected to an 802.11 access point, it makes the data rate determination based upon the available signal strength and sometimes other parameters such as QoS profiles and power-saving policies.

The spectrum used by Wi-Fi falls into the unlicensed Industrial, Scientific, and Medical (ISM) band category (2.4 to 2.5 and 5 GHz in the U.S. and Europe), and therefore does not require federal licenses in most countries to be used by individuals or offered commercially by service providers. Many other devices such as cordless phones, microwave ovens, Bluetooth devices, and headsets are permitted to emit radiation in this spectrum, which creates the potential for interference (see the accompanying sidebar).

Each Wi-Fi AP is assigned to a channel. That channel consists of frequencies in the 2.4 GHz, 2.5 GHz, or 5 GHz range of the radio spectrum, depending on the specific flavor of the 802.11 standard being used. For example, in the U.S. there are 11 different

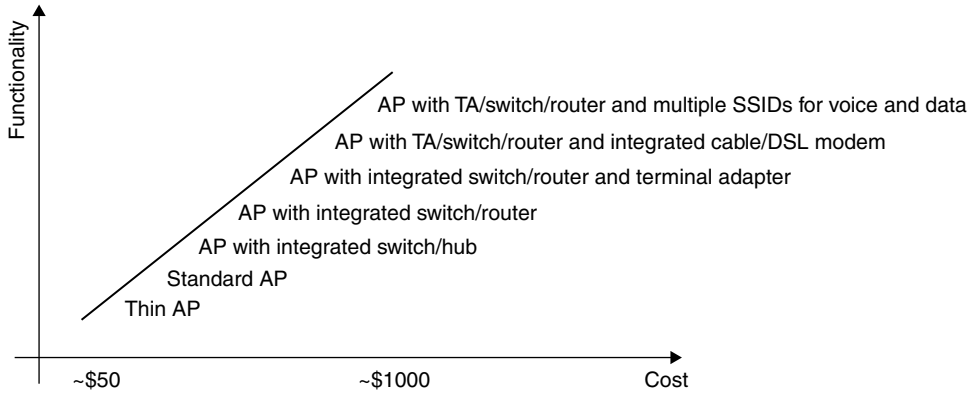

Figure 4.17 Wi-Fi AP types

but overlapping[11] channels available in 802.11b/g wireless networks (see the letter nomenclature description in the next section). In other regions of the world such as Europe and Japan, the standard supports 13 and 14 channels, respectively. Similar to its wired predecessor, Ethernet, the Wi-Fi channel access is based on the Carrier Sense Multiple Access–Collision *Avoidance* (CSMA-CA)[12] technique.

Interference

The 2.4 GHz ISM band reserved for free public use spans the frequency range 2400–2483.5 MHz. The devices operating in this band include two types: *unlicensed* such as microwave ovens, Bluetooth devices, and certain cordless phones, and *licensed* such as amateur radio and RF remote controls. Note that according to regulation (such as those by the FCC, the regulatory body in the U.S.), unlicensed devices are not allowed to interfere with the licensed ones.

A number of cordless telephones operate in the 2.4 GHz ISM bands using a randomly determined set of frequencies, or "frequency hopping." Potentially they can interfere with Wi-Fi traffic, but this does not happen often. For one thing, the user would have to be in an active data session and engaged in a phone conversation over a cordless set at the same time. Using VoWi-Fi phones simultaneously with analog PSTN ones can potentially exacerbate the problem, but the chance that a household or business would keep two types of phones, PSTN and Wi-Fi cordless, is slim.

Bluetooth devices are another potential source of WLAN interference. However, the transmission power of the Bluetooth devices is an order of magnitude lower than that of the cordless phones, so their potential for interference only arises in close proximity to WLAN devices (at ranges of 1–3 meters). Also, the Bluetooth specification starting from version 1.2 specifies the adaptive frequency hopping (AFH) method to avoid interference. The designers of dual-mode Wi-Fi/cellular phones, which typically support Bluetooth, and other Wi-Fi devices likely to experience interference have also created a number of proprietary solutions providing effective workarounds.

Finally, home microwave ovens radiate a narrowband signal between 2450 and 2460 MHz, creating interference potential. Most home microwave ovens, however, are well shielded, and experiments have shown that they would only interfere with an 802.11 network if the AP were within a few feet of one of the 802.11 endpoints.

[11] Channels 1, 6, and 11 were designated as nonoverlapping channels with additional spectrum provided for better separation.

[12] This protocol has its roots in the other standard created at the University of Hawaii in 1970 called ALOHA-NET, which was one of the most important milestones in data networking as we know it today. Note that the protocol underlying Ethernet is really called CSMA-CD for collision detection. The schema had been modified for the wireless environment to provide collision avoidance.

The networks defined by 802.11 standards originally relied on two types of radio frequency modulation techniques: frequency-hopping spread spectrum (FHSS) and direct-sequence spread spectrum (DSSS). While one of the 802.11 implementations called 802.11a was based on FHSS, most of the commercial solutions that followed (starting with 802.11b) used DSSS due to its higher tolerance to interference and potential to support higher bit rates.

Architecture The 802.11 standard defines two modes of operation for WLANs:

- *Ad hoc* mode, also known as independent basic service set (IBSS), where the clients communicate with each other
- *Infrastructure* mode, where an AP provides client access to a network

A WLAN in ad hoc mode essentially functions as a peer-to-peer network that does not require servers. It allows two or more clients with NICs to communicate directly with each other. An example of an ad hoc use case might be a group of visitors at a meeting with their laptops connected to each other, creating a mesh separate from a corporate network of their host.

The infrastructure mode is a more widespread architecture. Its topology includes one or more clients connected through APs to the IP core infrastructure. The architecture of Wi-Fi networks operating in an infrastructure mode consisting of one AP is referred to in standards as a basic service set (BSS). Multiple BSSs compose an extended service set, or ESS.

From the high-level view, functionality of APs is similar to that of a cellular network base station. In fact, multiple APs forming an ESS in enterprise, campus, or metro mesh setups are treated similarly to cells in a typical distributed wireless cellular system, as Figure 4.18 illustrates. One of the important distinctions between cellular *systems* and Wi-Fi *networks* is that the original 802.11a, b, and g standards did not support IP-layer mobility or the ability to maintain an active data session while changing access points and subnets[13] and many other functions necessary to provide wide area service.

802.11 Variations The initial success of Wi-Fi quickly resulted in considerable standards activity, which produced both numerous variations of the original 802.11 standard specifying different bands or data rates and extensions dealing with functionality omitted from the original specifications, such as QoS, inter-AP handoff, security, and others. What follows is a quick list of some of these standards and extensions to the original standards identified by lowercase letters following the 802.11 identification in alphabetical order. As mentioned in the earlier section "Wi-Fi Standardization," extensions a through j have recently been merged under the 802.11-2007 version of the original 802.11 standard.

[13] This capability is specified by the 802.11r extension discussed later in the chapter.

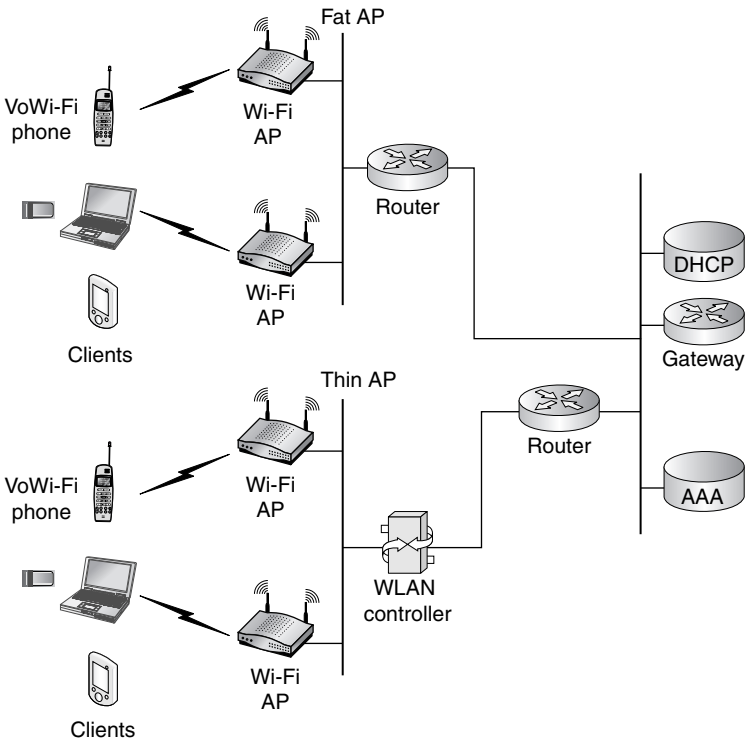

Figure 4.18 WLAN ESS example

802.11a The 802.11a [131] standard, along with 802.11b, was the first amendment to the original 802.11 specification. This standard defines operation in the 5 GHz band with 300 MHz of bandwidth. The theoretical maximum bit rate of IEEE 802.11a is 54 Mbps. The 802.11a devices operate in the 5 GHz frequency band using OFDM technology. The reason for selecting the 5 GHz band was the belief that the 2.4 spectrum may become congested.

By design, 802.11a supports 24 nonoverlapping channels (though only 8 can be used at any given time), which is a significant improvement over the b and g standards described next. The more channels are offered, the more options are available for the user, making it easier to avoid interference. The trade-off for the increased bandwidth with 802.11a and the move up the frequency ladder is a range typically limited to about 50 meters, which is roughly half of that provided by 802.11b. The limited range of 802.11a was partially responsible for its relatively cold reception by the enterprise market it was originally targeting.

802.11b The 802.11b [132] standard was the second of the two initial amendments to the original 802.11 specification. The 802.11b standard specifies data rates of 1, 2, 5.5,

and 11 Mbps in the 2.4 GHz spectrum and is based on DSSS. This standard is one of the most widely adopted in today's commercial products both in residential and enterprise environments.

802.11d The 802.11d [138] standard was developed by IEEE to extend the original 802.11 specification to countries where the original specification is not applicable or allowed due to unique spectrum allocations and other regulatory reasons or, in the standard words, "additional regulatory domains."

802.11e The 802.11e [139] standard specifies a set of QoS enhancements of the original 802.11 standards prompted mostly by low applicability of the standard Wi-Fi for VoIP and other types of real-time traffic. The 802.11e standard supports differentiation for various types of data on the network such as voice, video, and other multimedia and more delay-tolerant types of communications.

On the footsteps of 802.11e, the Wi-Fi alliance introduced the Wi-Fi Multimedia (WMM) certification used to ensure interoperability of 802.11e-compliant equipment. WMM defines four traffic categories (or access categories in WMM terminology): voice, video, best effort, and background.

802.11g The 802.11g [133] standard defines a higher bit rate alternative to 802.11b. The higher bit rates are achieved by transmission on the same 2.4 GHz radio frequency band used by 802.11b. Like 802.11a, 802.11g is also based on the OFDM technology and supports a maximum theoretical bit rate of 54 Mbps. 802.11g is backward compatible with 802.11b.

For backward compatibility with 802.11b devices, the standard specifies support of complementary code keying (CCK) modulation based on the RC4 cryptographic algorithm and providing both access authentication and encryption of traffic. The standards allow mixed b and g systems; however, g speeds will degrade significantly as b clients are associated with combined b/g access points.

802.11h The use of the 5 GHz frequency band in Europe caused some problems with regulatory requirements and interference with other services. To overcome these problems, an amended version of IEEE 802.11a, IEEE Standard 802.11h [134], was developed. The 802.11h standard also introduced some advanced techniques such as Transmit Power Control (TPC), potentialy reducing transition power by 50 percent, and Dynamic Frequency Selection (DFS), providing automatic channel hopping to avoid interference or overload. In Norway and a number of other European countries, the 802.11a/h systems can only be used indoors due to local government regulations.

802.11i Originally wireless LAN security was based on SSID and—fast becoming legacy—the Wired Equivalent Privacy (WEP) security scheme. WEP security is weak and vulnerable to eavesdropping because of its basic keying scheme and poor vector initialization. The IEEE 802.11i [136] security specification has addressed WEP shortcomings by introducing a new security framework superseding the original WEP and enabling robust authentication, encryption, and key rotation.

The Wi-Fi Alliance commercialized selected portions of the 802.11i standard as WPA, which stands for Wi-Fi Protected Access. The 802.11i standard is also referred to as WPA2 by the Wi-Fi Alliance. Instead of the weak RC4 cipher, the 802.11i standard relies on the stronger Advanced Encryption Standard (AES) cipher.

802.11n In response to the demand for even greater throughput, IEEE has began working on the 802.11n standard, specifying data rates of up to 248 MBps, and backward compatible with both 802.11b and 802.11g. 802.11n defined the elaborate MIMO-based antenna schema supporting separate antenna arrays for sending and receiving signals (similar to that specified by 802.16 standards) to significantly improve data rates through *spatial multiplexing* and ranges through *spatial diversity.* The 802.11n standard supports both 5 GHz and 2.4 GHz frequencies. Work on the 802.11n definition is still ongoing at the time of writing, with the current approval target set for the end of 2008.

802.11r The 802.11r standard—which is currently under development by IEEE—also known as *fast BSS transitions* or *fast roaming,* is an 802.11 extension specifying fast data handover between access points. The standard is especially applicable in enterprise or metro environments where there is a need to preserve the data session continuity while moving between multiple APs. It is expected that VoIP will eventually become one of the standard's main applications; however, for this to happen, IEEE must address real-time traffic delay concerns.

The WLAN supporting the 802.11r standard will in many respects behave similarly to today's cellular network (at least in terms of data handover). Interestingly, 802.11r is mostly focused on the handover setup and not the actual handover handling, which is currently left for vendor interpretation.

Addressing VoWi-Fi Challenges VoWi-Fi deployment presents a number of unique challenges to all players in its ecosystem. Some of the challenges (as shown earlier in Figure 4.16) include:

- Support for QoS and prioritization of voice in the environment originally created to carry data traffic
- Support for strong security mechanisms (especially in enterprise applications)
- Support for fast roaming and seamless in-call handover in wide area and metro deployments
- Slow introduction and limited selection of mass-market VoWi-Fi-capable mobile devices
- High VoWi-Fi device power consumption
- Low access point capacity and limited coverage

New 802.11 standards and extensions as well as numerous proprietary approaches are being successfully applied to address these challenges, enabling VoWi-Fi solutions,

originally relegated to vertical markets, to go mainstream. The examples include residential and small office systems complementing broadband VoIP services and replacing cordless telephony, enterprise solutions extending VoIP PBXs, and FMC solutions where VoWi-Fi is seamlessly combined with cellular—the focus of this book.

Next we explore how these challenges are being addressed.

QoS In general, running Voice over IP, especially in the enterprise and hotspot "WAN" environment (in other words in places where there are multiple APs and multiple streams of data), presents many challenges for a Wi-Fi network. The most critical among these is achieving and uniformly maintaining acceptable audio quality, for example, by minimizing network delay in a mixed voice and data environment.

Wired or wireless, 802.11 networks were originally not designed for real-time streaming media or support for guaranteed packet delivery rates. Therefore, congestion on the wireless network or drop in throughput, without traffic differentiation, can quickly make any kind of voice transmission unintelligible or at least severely degrade the user experience. When voice service is provided over the IP protocol running over a wireless interface, prioritization and quality of service become hard requirements, mainly because of the variable and finite throughput of the air link, which is further affected by interference, constantly changing medium characteristics, and varying distance to the AP.

Conducting real-time communication, such as voice or multimedia streaming, dependably is especially difficult in *public* Wi-Fi networks using ordinary "best effort" operation. Excessive latency in such environments can be caused by either one-way packet delay or extensive buffering needed to address jitter (variance of the packet arrival time). Latency exceeding 200–300 milliseconds (depending on the codec, device, and specific type of voice communication used) during a two-way conversation starts negatively affecting the user experience and quickly degrades the conversation. The original Wi-Fi standards were designed to operate using statistical multiplexing of user traffic contending for access to the air interface based on a "best effort" approach. That means that the quality of a particular type of service cannot be guaranteed, especially when the channel utilization is increasing, which in turn causes an incremental rise in packet collision and retransmission rates.

Guaranteed QoS[14] mechanisms combined with prioritization schemas were introduced to Wi-Fi in the 802.11e standard, primarily addressing latency, jitter, and error rate.

The 802.11e standard provides four QoS traffic categories:

- Voice
- Video
- Best-effort
- Background

[14] QoS is based on the idea that transmission rates, error rates, and other characteristics can be measured, controlled, and to some extent, guaranteed in advance.

802.11e relies on the Hybrid Coordination Function (HCF). HCF is a single-channel access protocol enhancing the original 802.11 Distributed Coordination Function (DCF) used by CSMA-CA for medium allocation and channel access coordination. HCF identifies and prioritizes different types of traffic by introducing a concept of Traffic Classes (TC) and applies the Controlled Channel Access (CCA) protocol to CBR traffic to ensure a constant bit rate above the minimum quality threshold set for a particular traffic type.

802.11e HCF introduces two main options for controlling channel access:[15]

- Enhanced Distributed Channel Access (EDCA)
- Hybrid-Coordinated Controlled Channel Access (HCCA)

EDCA The Enhanced Distributed Channel Access schema prioritizes traffic classes (higher-priority traffic sent first) by assigning each traffic class a Transmit Opportunity tag (TXOP), which identifies the period of time during which the client is allowed to transmit an unlimited amount of data.

HCCA The HCF-Coordinated Controlled Channel Access (HCCA) schema is used when a more precise QoS definition (than that provided by EDCA) is needed. HCCA also allows for the *reservation* of TXOPs with the AP and defines traffic streams (TSs) in addition to traffic classes. HCCA introduces a concept of Controlled Access Phase (CAP)—a period initiated by an AP for contentionless communication with a mobile station.

Along with ensuring consistent quality of service, the HCCA also provides a robust call admission control (CAC) mechanism to make sure the allocated channel bandwidth is sufficient to carry a given number of simultaneous voice conversations. CAC enables the APs to calculate the available bandwidth, make handover decisions (if other APs in the range have unused capacity), and throttle the traffic accordingly.

QoS Summary Both the HCCA and EDCA options have their pros and cons. The HCCA mechanism is more appropriate for setting up fine-grained QoS policies—allowing for precise definition of acceptable latency and jitter limitations—and therefore may be better suited for commercial VoIP and multimedia services. However, HCCA is more complex to implement than EDCA and to date has not seen much support in commercial Wi-Fi products, which almost uniformly support EDCA.

Needless to say, in addition to the approaches defined in 802.11e, all the common IP-layer QoS standards such as MPLS, DiffServ,[16] and RSVP (described in Chapter 2) can also be used in conjunction with techniques offered by 802.11e to improve overall QoS of VoWi-Fi system traffic.

---

[15] Interested readers are welcome to explore further in the original IEEE standard specification and the book by Frank Ohrtman, *Voice over 802.11* (Artech, 2004).

[16] DiffServ is an IP-layer QoS framework defined by IETF that takes the IP type of service (TOS) field, renames it in the Differentiated Services Code Point (DSCP) field, and uses it to carry information about IP packet service requirements.

Security The security support is as important for voice as for data traffic to ensure conversation privacy, enforce authentication and authorization, and preserve overall network integrity. However, the demands of strong security on a wireless network, which for example may inadvertently increase latency or jitter, can have a direct impact on voice quality. Security support becomes even more difficult in a Wi-Fi environment supporting fast roaming, as the properly secured network must still be able to unobtrusively support mobility (both roaming and handover).

All 802.11 specifications, such as a, b, and g, support a simple WEP security mechanism. By design, WEP may support a 40-bit key and a 104-bit key combined with a 24-bit initialization vector, resulting in 64- and 128-bit encryption, respectively. WEP is based on RC4[17] cryptography and provides both access authentication and encryption of traffic. WEP, however, allows for transmission of clear text (original data) and cipher text (encrypted data) in the open during session establishment between client and AP. This makes it easy to extract a shared secret key from the traffic between the AP and the client for anyone with a sufficiently sophisticated protocol analyzer.

Further, WEP is usually implemented with manual distribution of static keys shared between all clients associated with an AP. This deficiency in particular makes it clear why WEP-based security is not suitable for enterprise installations with multiple clients because of the need to change keys on all associated clients if one of them is lost.

The 802.11i standard, referred to as WPA2 by Wi-Fi Alliance, was developed to address most if not all of the original WEP shortcomings. The main security improvements introduced by 802.11i include the support for 802.1x/EAP (Extensible Authentication Protocol, providing robust authentication and authorization), along with advanced encryption algorithms such as the Temporal Key Integrity Protocol (TKIP) and the Advanced Encryption Standard–Counter Mode/CBC-MAC Protocol (AES-CCMP), designed to improve confidentially protection. 802.11i also provides mechanisms designed specifically to aid fast roaming, such as preauthentication.

The 802.11i standard proposes two authentication methods:

- *Personal,* which is a basic option based on preshared keys intended for residential use
- *Enterprise,* which is a more advanced option developed for enterprises and, potentially, certain metro deployments

While the 802.11i personal option essentially enhances WEP by eliminating key transmission and requiring instead distribution of master shared secret keys to users through other means, the "Enterprise" option provides a completely new security approach based on centralized authentication and dynamic key distribution.

[17] Named incidentally for Route Coloniale 4, a road in Vietnam and a famous battle of the first Vietnam war, which took place in 1950.

The 802.11i Enterprise option is based on 802.1x and EAP IEEE standards, which provide RADIUS-based mutual client/server authentication, support centralized policy definition, and offer dynamic distribution of encryption keys. The session time-out in the Enterprise option triggers reauthentication and new encryption key generation, which further enhance privacy protection.

Authentication The EAP/802.1x model is based on three main objects, depicted in Figure 4.19, involved in the authentication process:

- An *Authentication server* such as RADIUS or DIAMETER
- A *Supplicant,* a client supported by a device that needs to be authenticated
- An *Authenticator,* an element acting as a proxy of an authentication server, typically an AP

In this model the specific authentication method is defined by EAP between the Supplicant and the Authentication server. To gain access to a network, the Supplicant (client) must go through a two-step process of initial association with the server followed by mutual authentication between the client and the server. After the authentication is performed by the Authentication server, the Authenticator (AP) grants or disallows client network access based on the authentication process results. If the authentication is successful, the client and server originate the same encryption key, which is never transmitted in the open. When the encryption key is generated, the Authentication server distributes a session key to the Authenticator, which it uses to encrypt the broadcast key used to encrypt the session.

The solutions implementing EAP typically support login credentials based on *two-factor* authentication, requiring the user to supply an ID and a password for initial association.

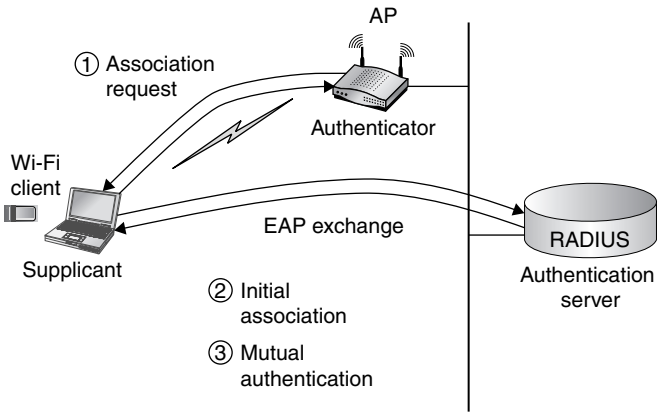

Figure 4.19 802.11i authentication model

Since EAP leaves many aspects of actual protocol execution up to the vendor, the industry came up with several commercial implementations called EAP types. Popular examples of EAP types include:

- EAP–Transport Layer Security (EAP-TLS), defined in IETF RFC 2716 [140]
- EAP–Subscriber Identity Module (EAP-SIM), described in IETF RFC 4186 [84] and used in 3GPP-defined solutions relying on SIM cards
- EAP-LEAP (for Lightweight EAP), a proprietary mechanism developed by Cisco Systems supporting dynamic WEP keys and mutual authentication between client and Authentication server
- Protected EAP (PEAP), a proprietary mechanism developed by a consortium of vendors, which provides client authentication but does not support encryption

EAP-TLS enhances the original EAP model with the support for digital certificates. When the client attempts association with the server, the Authentication server first sends the client a certificate for validation. Following a successful validation, the client sends its own certificate back to the server, which is validated in the same way. Upon successful mutual authentication the *EAP-Success* message is sent to the client, the Authenticator (AP) is notified, and the normal EAP procedures resulting in generation of dynamic WEP key take place.

The EAP-SIM option, based on the challenge-response mechanism, was developed in part with converged Wi-Fi/3GPP cellular solutions in mind. Unlike EAP-TLS, EAP-SIM requires the mobile device to support a SIM card à la GSM and also requires the Authentication server to be connected with the GSM core (specifically AuC and MSC), because this EAP implementation uses the SIM card and International Mobile Subscriber Identity (IMSI—a number uniquely associated with a particular SIM) to supply authentication credentials to identify the user. The AuC function in GSM allows the MSC to authenticate SIMs upon initial connection to the network. The AuC also generates an encryption key to protect session privacy and a random number called a triplet (a 64-bit random number, a 32-bit response [SRES], and a 64-bit K_c key), which is used as a shared secret between the SIM and the AuC.

The authentication process starts with the Authentication server requesting GSM triplets from the AuC and returning the random numbers with a checksum (derived using SRES and K_c components of the triplet) to the client with SIM as a challenge. Using the random number provided by the server, the client calculates the checksum using the device's SIM card to generate its own SRES and K_c numbers. The resulting checksum is then compared to the one received from the server; if they match, the authentication is successful and the client initiates a challenge to the server using the same procedure. Upon successful mutual authentication, the EAP-Success message is sent to the Authenticator and a sufficiently robust key is generated.

Despite its complexity and reliance on the access to AuC in the GSM core infrastructure, EAP-SIM is particularly well suited for use in dual-mode FMC handsets. On the other hand, it is not extendable to CDMA FMC solutions or any other systems with non-SIM-based handsets.

Confidentiality Protection Along with addressing the problem of weak authentication, one of the goals of 802.11i was to improve the original WEP encryption. This is achieved via the support of two encryption options: the Temporal Key Integrity Protocol (TKIP) enhancing the WEP RC4 cipher, and the Advanced Encryption Standard (AES) combined with the Counter Mode with Cipher Block Chaining Message Authentication Code Protocol (CCMP), providing more-advanced encryption and ciphering.

The TKIP enhancements include per-packet keying (PPK), a message integrity check (MIC), and extension of the original WEP initialization vector from 24 bits to 48 bits. The TKIP PPK mechanism supports generation of different unicast per-packet keys to allow multiple initialization vectors to use different keys (as opposed to WEP). The PPK is also used to provide the broadcast key rotation to protect broadcast and multicast WLAN traffic against the same threats.

Mobility To be considered for carrier-scale commercial and enterprise-wide deployments, VoWi-Fi solutions must support mobility to perform on a par with cellular systems. The mobility support requirements include:

- **Macromobility** Roaming and active call handover between Wi-Fi and other wireless access technologies such as cellular
- **Micromobility** Roaming and active call handover between Wi-Fi APs within the same subnet, in different subnets of the same network, or in different Wi-Fi networks

Clearly, both types of mobility must be supported in the environments with both multiple APs, where the VoWi-Fi mobile station is expected to change its physical location and roam between them, and multiple access networks such as Wi-Fi, WiMAX, cellular, and others. Examples of such environments include wide areas with mixed networks and VoWi-Fi-only locales such as corporate offices, college campuses, and metro installations, potentially capable of covering whole cities or conceivably even regions. VoWi-Fi micromobility, however, is less relevant in residential or SoHo installations, where VoWi-Fi technology is used for a cordless telephony or in combination with cellular for residential FMC services.

The majority of today's general-purpose access points are not equipped to support inter-AP roaming and handoffs. Wi-Fi technology was originally designed with the data users in mind and did not account for typical telephony use cases (walk or drive and talk vs. walk or drive and type messages or browse the Web).

To support micromobility, the mobile station must preserve its IP address to avoid reattachment to the network, and it must always perform fast reauthentication to protect overall network security and integrity. These requirements present additional challenges to handover timing, which must not exceed approximately 250 milliseconds (a threshold also set for cellular systems) to avoid choppy voice transmission negatively affecting the user experience.

In the section that follows we are going to analyze the technology allowing the support of IP-layer micromobility, while deferring discussion of macromobility to Chapter 5.

Inter-AP Handover The standardization of Wi-Fi roaming is still ongoing in IEEE "task group r," with the 802.11r standard expected to be ratified in 2008. The 802.11r specification defines fast BSS (basic service set) transitions for seamless handover for the a, b, g, and upcoming n standards. The issues currently under consideration by the working group include both VoIP and data mobility, minimization of handover delay, and compatibility with 802.11n and 802.11i (robust authentication). The 802.11r standard is based on the make-before-break scheme, in which the security association, QoS profile, and other connection properties are established with the neighboring *target* "to" AP before leaving the *serving* "from" AP.

With the 802.11r standard still in the works, in recent years the industry has come up with multiple proprietary roaming solutions designed to satisfy the requirement of the systems with multiple APs and provide a user experience on a par with that of cellular system users. The majority of such architectures are based on two types of access points:

- A thin AP/WLAN switching network architecture with the control consolidated in the core infrastructure components
- An intelligent AP network architecture in which most of control is concentrated in "smart" or "fat" APs themselves

Inter-AP roaming must be supported so that the handovers can take place on the OSI Data Link layer when the client roams between APs within the same subnet or on the Network (IP) layer using Mobile IP when the client roams between APs in different subnets or even networks (see Figure 4.20).

Typically, inter-AP handoff is a multistep process beginning when the mobile station monitors the RF link quality of the serving AP and evaluates it against other APs within range. The RF link condition deteriorating below a certain threshold will trigger handover based on a conditional algorithm. Along with RF monitoring, the mobile station also periodically scans the network for foreign APs to determine potential availability for roaming.

The process of handover itself starts with reauthenticating with the target AP. This procedure depends on a particular security protocol in use, and if not implemented properly, may have a significant impact on overall handover latency. After the mobile station is reauthenticated and granted permission to connect to a new AP, the transfer of QoS context occurs and the mobile device is associated with a new AP. In the process of handover involving multiple subnets or networks, the IP address of a mobile station must be kept constant, which can be accomplished utilizing protocols like Mobile IP.

Dealing with Handover Latency Keeping handover latency sufficiently low to make it imperceptible to the user is a significant technical challenge with the majority of the secure Wi-Fi authentication methods. Two of the standardized approaches to speeding up the authentication process defined in the 802.11i standard are based on the reuse of the previous authentication results for generating new encryption keys or modifying the results of the earlier authentication to derive the new keys to avoid full reauthentication.

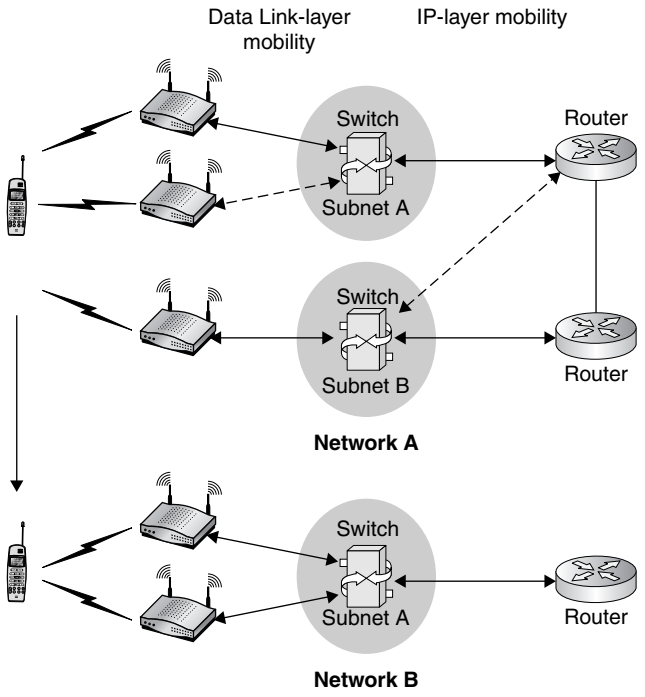

Figure 4.20 Wi-Fi micromobility

The first method is called *preauthentication.* This method requires the serving AP to be connected to potential target APs over the wired network. With such a connection in place, a target AP can "preauthenticate" the client prior to the actual move, so during the handover only minimal 802.1x/EAP authentication procedures will be required.

The second method is called pair-wise master key (PMK) caching. When a client associates with a particular AP supporting PMK using one of the EAP procedures, it caches the EAP encryption key. If the mobile station roams away from the AP and then "comes back," it will be able to reuse the cached key to avoid full reauthentication. The PMK flavor called opportunistic PMK, developed for use in an 802.11i enterprise security framework, is applied in thin AP architectures, allowing caching of an encryption key in a switch or AP controller.

Along with reducing reauthentication times, other proprietary implementations and IEEE draft standards are focusing on reducing the time needed for neighboring AP scans. One popular scheme enables a broadcast of the list of neighboring APs in the beacon. This reduces the scan time for the clients and saves battery life. Another scheme proposes deploying APs in preset channel numbers only, thereby limiting the scan channel list for the clients.

Handover Triggers In Wi-Fi networking, the client decision to initiate handover is usually triggered by reaching certain threshold conditions. These are examples of such conditions:

- The radio link quality falls below an acceptable threshold.
- The maximum retry count is exceeded.
- The data rate falls below an acceptable threshold.
- The serving AP becomes overloaded.

When one of these triggers fires, the mobile station initiates the scan of available 802.11 channels. On each discovered channel, the client station sends a probe and waits for probe responses or beacons from APs on that channel. The probe responses and beacons received from APs are discarded unless they have matching SSID and encryption settings. Thus qualified, remaining APs are then prioritized based on a set of predetermined parameters such as availability, throughput, or proximity. After the target AP is determined, the handover procedure is initiated.

Reducing Client Power Consumption The original Wi-Fi specification was written for data-centric mobile stations such as laptops with a relatively long battery life. When Wi-Fi is used for voice communication, the mobile devices are usually handheld with a form factor similar to that of a cordless phone, or even a cellular phone in the case of dual-mode FMC solutions. Smaller devices mean smaller batteries and consequently degraded active and standby time. In fact, early technology trials yielded Wi-Fi standby times roughly half those of cellular, a fact that was pretty much guaranteed to disappoint users.

To address these shortcomings along with proprietary approaches, two power-saving solutions were introduced in the 802.11e standard:

- Unsolicited Automatic Power Save Delivery (U-APSD)
- Scheduled Automatic Power Save Delivery (S-APSD)

In U-APSD, the client sends a "standby ready" notification to the AP before going on standby, which causes the AP to start backing up the data destined for the client and switch to "beacon" mode. The AP in beacon mode sends periodic beacons to the dormant client indicating if there is any data available. The client in a dormant state would periodically "wake up" only to read the beacons. Depending on the beacon type, the client will either go back to a dormant state or switch to an active state.

The S-APSD scheme provides a similar power-saving mechanism for clients operating in the HCCA mode. This scheme was designed specifically for real-time traffic such as VoIP. It goes a step farther than U-APSD by allowing the clients to go on standby for preprovisioned periods, without having to listen for the beacons.

Access Point Capacity The individual AP capability to support multiple simultaneous VoIP streams is an important measure of overall VoWi-Fi system capacity, which is

especially significant in enterprise and hotspot environments. Hotspots installed in public spaces like libraries, cafés, and hotel lobbies are expected to support significant numbers of simultaneous conversations. In enterprises, the same requirements are applicable to APs serving factory floors, conference rooms, reception areas, and the like.

The capability of 802.11b systems to support VoIP was first analyzed back in 2004 in a paper by Hole and Tobagi [175]. The paper provided theoretical calculations of the maximum possible number of VoIP sessions that could be supported by a single AP. Figure 4.21 shows the dependence of the number of sessions on data rate under three different sets of conditions:

- First (leftmost in each set) is a theoretical analysis assuming perfect RF conditions and no delay.
- Second (middle of each set) is a theoretical analysis assuming perfect RF conditions with a one-way Wi-Fi link delay of 10 ms or less.
- Third (rightmost in each set) is a theoretical analysis assuming the bit error rate (BER) of the radio channel to be 1.0×10^{-4} with a delay of less than 10 ms.

Not surprisingly, the number of supported sessions was shown to increase with the data rate. Supporting paper observations, later practical experience showed that, for example, a single 802.11b AP can support between 8 and 12 simultaneous VoIP sessions in an environment without interference. This number, however, can quickly drop to 2–3 with the introduction of interference or increasing distance between the AP and clients. The 802.11a or g solutions can increase this rate to 15–20 (albeit, at the expense of coverage radius). It was also observed that the AP throughput itself may not be as much of a limiting factor as latency, prioritization, and jitter.

Regulatory: E911, CALEA, and Other Mandates A commercial VoWi-Fi service must comply with the emergency calling framework whenever possible, behaving on a par with today's cellular and fixed VoIP systems. Currently, emergency calls in a cellular network, governed by E911 in the U.S., E112 in the EU, and E110/119 in Japan, are routed

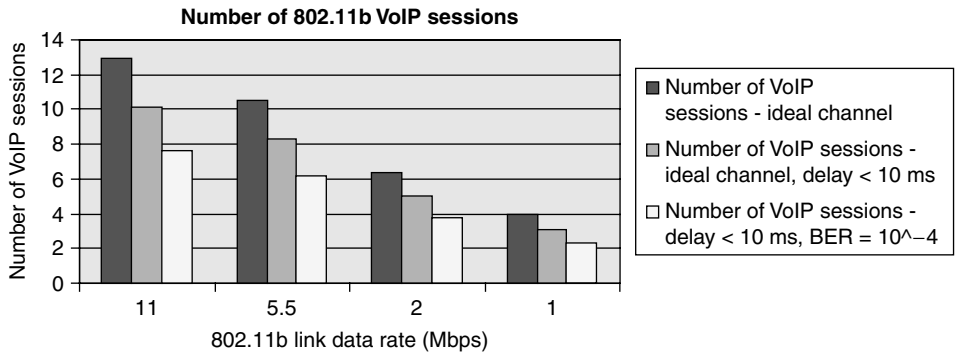

Figure 4.21 Access point capacity (source: Hole and Tobagy)

to the appropriate Public Safety Access Point (PSAP) using the routing digits received from the location services platform.

In a VoIP and hence in a commercial VoWi-Fi solutions, in most countries, emergency calling support is required by the government to deliver the caller's location or address of the subscriber to the PSAP along with the callback number. This can be accomplished by appropriately routing VoIP calls from media gateways in service provider networks or establishing IP connections directly to appropriately equipped PSAPs.

For example, in the U.S. the National Emergency Number Association (NENA) is currently evaluating a number of methods for bringing VoIP 911 calling into E911 systems that provide Selective Routing and Automatic Location Information (ALI). NENA is likely to recommend that Voice over IP providers use newly available interfaces in the E911 systems for customers that have fixed locations and are using telephone numbers from their local area code.

It must be noted that standard PSTN emergency calling approaches are often not applicable in commercial VoIP and, hence, VoWi-Fi implementations. Since VoIP service is access independent, the subscribers may change their address or even use several temporary points of attachment to the network during the course of the contract. The VoWi-Fi service users that can travel between hotspots within the same city or country are even harder to track; albeit during admission to Wi-Fi service, the identity and location of the Wi-Fi hotspot serving the user can be easily passed on to the PSAP via the Authentication infrastructure.

Convergence of Mobile Systems

At the end of this chapter we are going to take a quick look at convergence of mobile systems, which should provide a good segue to the discussion of convergence of fixed and mobile systems.

One of the most important aspects of the next-generation wireless communication systems is the capability to support simultaneous access to CS and PS services. This allows for more flexibility in circuit voice and packet data services creation (parallelism in access to service), and also to the transition from one technology, supporting circuit voice, to another, supporting packet-based transmission, using the methods like VCC (first mentioned in the section "3GPP Systems" of this chapter and described in detail in Chapter 5) defined in 3GPP Release 7. In addition, Mobile IP can support the *make before break* handover on the IP layer when it is possible to access both the source technology and the target technology simultaneously, when both support PS services.

Also, the ability to concurrently support access to CS and PS services via different access technologies minimizes the interruption time during an active voice call even when the network is not cooperating with the terminal in the handover process. This property is particularly desirable in systems with a significant installed base of both CS and PS technologies (such as Wi-Fi and circuit cellular systems), which need to be converged while minimizing the impact on the infrastructure. It is also possible to use DTM to support simultaneous access to GSM CS services and GPRS services, or to use UMTS to provide simultaneous access to its PS domain and its CS domain.

Table 4.2 Most Likely Terminal Operation (SR = single radio, DR = dual radio)

	GSM	UMTS	LTE	WiFi	WiMAX	CDMA
GSM	N/A	SR	SR	DR	DR	SR[1]
UMTS		N/A	SR	DR	DR	SR[2]
LTE			N/A	DR	SR	SR
Wi-Fi				N/A	SR	DR
WiMAX					N/A	DR
CDMA						N/A

[1] Applicable to provide GSM or CDMA customers with a terminal capable of global roaming support.

[2] Applicable to provide UMTS or CDMA customers with a terminal capable of global roaming support, but also in countries like Korea, where there are operators with both these access networks.

Often, to minimize the cost of multimode terminals, they are built to support only one cellular technology in an active transmission state at a time (that is, other supported technologies cannot be active at the same time—in the approach known as a "single radio terminal" (SR)—when one is in an active transmission state). In this case, the support of seamless mobility or change of domain requires cooperation by the network in preparing the handover. While it is always possible to implement "dual-radio" (DR) terminals, it is expected that single radio will be prevalent in many commercial implementations. Table 4.2 presents what the authors consider the most likely scenarios for dual-mode terminals. The case of multimode (e.g., tri-mode UMTS, E-UTRAN, Wi-Fi) terminals could be inferred by a logical combination of the relevant dual-mode options.

It turns out that in most cases the single-radio strategy applies when multiple cellular access technologies are used, while Wi-Fi and WiMAX matched with a cellular access would normally require a dual-radio operation (for instance to allow a user to place a circuit voice call while accessing data networks via the Wi-Fi or WiMAX network).

In summary, the type of terminal used to support the convergence solution will also impact the necessary support from the network side. If the convergence application implies using a dual-radio terminal, then seamless mobility and a seamless user experience can be provided without the need for radio access networks to cooperate in preparing handovers or domain changes.

Summary

This chapter brought us one step further in our quest to understand the FMC technology components. We have analyzed the relevant aspects of cellular systems, compared circuit and packet technologies, and looked at radio access and mobility.

We then provided an in-depth discussion on the modern WiMAX and Wi-Fi technology landscape, paying special attention to voice service support. We looked at both the main VoWi-Fi enablers and the biggest technical roadblocks to its widespread deployment. Armed with this knowledge, we are now ready to proceed to the final destination of our journey through FMC technology, the *C* in FMC.