2

Next-Generation Wireless Standards and Their Integration with the Internet

Hang Liu

Abstract

Standards provide the foundation for developing innovative technologies and enabling them to be widely adopted in market. Several major international standard bodies are developing next-generation wireless standards, including the Institute of Electrical and Electronics Engineers (IEEE), the Internet Engineering Task Force (IETF), the International Telecommunication Union Radiocommunication Sector (ITU-R), the European Telecommunications Standards Institute (ETSI), and the Third Generation Partnership Project (3GPP). The standardization activities of IEEE 802 committee mainly focus on physical (PHY) and media access control (MAC) layers, that is, layers 1 and 2 of the network protocol stack, including WLAN, WMAN, and WPAN network interfaces. IETF standards deal with layer 3 and above, in particular with standards of the TCP/IP and Internet protocol suite, including mobile IP and mobile ad hoc networks (MANET) related protocols. ITU-R is one of the three sectors of the ITU and is responsible for radio communications. It plays a vital role in the global management of the radio-frequency spectrum and satellite orbits, and developing standards for radio communications systems to assure the necessary performance and quality and the effective use of the spectrum. ETSI is a European standards organization for producing globally applicable standards for information and communications technologies (ICT), including fixed, mobile, broadcast, and Internet technologies. ETSI inspired the creation of, and is a partner of, 3GPP – a collaboration project between groups of telecommunications associations worldwide. 3GPP's original scope was to produce technical specifications and technical reports for a globally applicable 3G cellular mobile system based on evolved Global System for Mobile communications (GSM) core networks and radio access technologies, as well as maintain and develop GSM technical specifications and reports. It is currently developing 4G mobile network system. 3GPP standardization

https://doi.org/10.1017/CBO9780511921117.003 Published online by Cambridge University Press

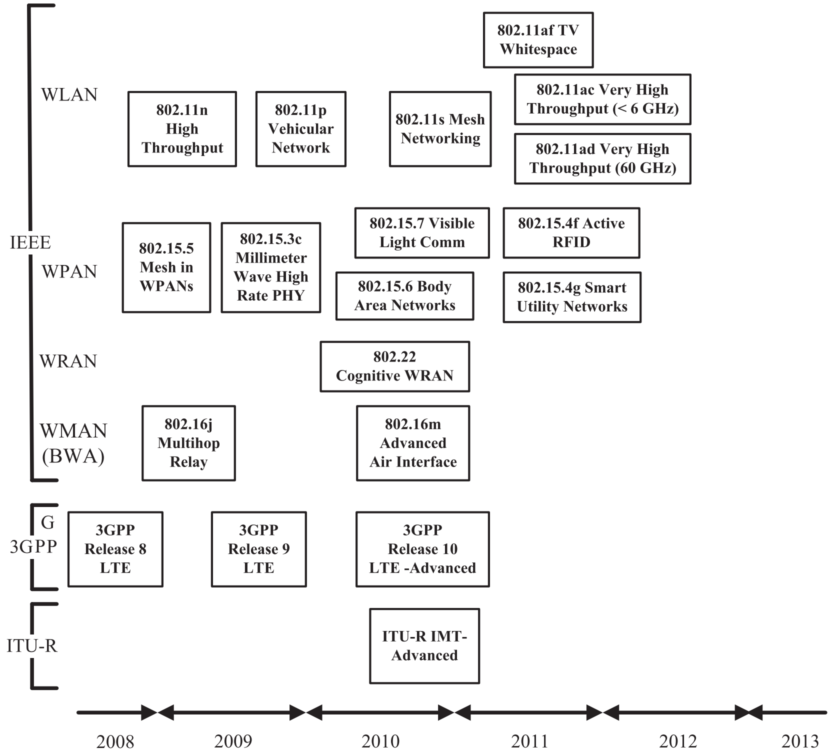

Figure 2.1. Major standards processes for next-generation wireless networks.

encompasses radio access, core network, and service architecture. Figure 2.1 illustrates major standards processes for next-generation wireless networks in IEEE, 3GPP, and ITU-R.

2.1 Technology and Service Trends of Emerging Wireless Standards

The standardization efforts for future wireless networks focus on both new radio access interfaces and improved network architectures. The standardization work on new radio interfaces aims at increasing network capacity to match or shorten the gap with wireline broadband access, and improving bandwidth efficiency and coverage range by employing advanced physical and MAC layer techniques such as multiple-input and multiple-output (MIMO), orthogonal frequency-division multiple access (OFDMA), and space-division multiple access (SDMA), as well as extending battery life and reducing latency for real-time communications. As shown in Table 2.1, future WLAN and WPAN standards will support up to 1 Gbps data rate, and future WMAN and cellular standards can support a peak

https://doi.org/10.1017/CBO9780511921117.003 Published online by Cambridge University Press

Table 2.1. *Emerging Wireless Interfaces*

Standard	Maximum PHY Rate	PHY Technology	MAC Technology	Operating Frequency
802.11n WLAN	600 Mbps (4×4 MIMO, 4 spatial streams, 40 MHz bandwidth); 200 Mbps (3×3 MIMO, 3 spatial streams, 20 MHz bandwidth)	MIMO and OFDM	EDCA and HCCA	<6 GHz, typical 2.4 GHz and 5 GHz
802.11ac WLAN	>1 Gbps for multi-station; >500 Mbps for a single link	MU-MIMO and OFDM	SDMA	<6 GHz, typical 2.4 GHz and 5 GHz
802.11ad WLAN	>1 Gbps	TBD	TBD	60 GHz
802.15.3c high rate WPAN	5 Gbps on 2 GHz bandwidth	Single carrier and OFDM	TDMA and CSMA-CA	60 GHz
802.15.4/4a low rate WPAN	250 kbps with 802.15.4; 27 Mbps with 802.15.4a UWB PHY; 1 Mbps with 802.15.4a spread spectrum PHY	Spread spectrum and UWB	TDMA and CSMA-CA	Spread spectrum PHY: typical 2.4 GHz, 915 MHz, 868 MHz; UWB PHY: 3 GHz to 5 GHz, 6 GHz to 10 GHz, and less than 1 GHz
802.16m WMAN	300 Mbps for downlink (4×4 MIMO, 20 MHz bandwidth); 135 Mbps for uplink (2×4 MIMO, 20 MHz bandwidth)	MU-MIMO and OFDM	OFDMA in downlink and uplink	<6 GHz
3GPP LTE E-UTRAN	300 Mbps downlink (4×4 MIMO, 20 MHz bandwidth); 75 Mbps uplink for a user (SC-FDMA, 20 MHz bandwidth)	MU-MIMO and OFDM	OFDMA in downlink and SC-FDMA in uplink	<6 GHz

downlink rate of several hundred Mbps and a peak uplink rate of ∼100 Mbps under high mobility.

It is critical to utilize the spectrum efficiently and ensure the coexistence of different wireless systems. Cognitive and dynamic spectrum access schemes provide a promising solution. In addition, new FCC regulations for unlicensed devices to operate in the TV whitespace requires that the secondary whitespace

https://doi.org/10.1017/CBO9780511921117.003 Published online by Cambridge University Press

devices have cognitive radio and dynamic spectrum access capabilities and shall not interfere the operation of primary users. Several standard working groups and committees such as IEEE 802.22, IEEE SCC41, IEEE 802.19, and IEEE 802.11 are developing or plan to develop the standards for radio systems to operate in TV whitespace using cognitive radio technology.

The standardization work on the mobile network architecture aims at optimizing network performance, improving cost efficiency, facilitating the fixed-mobile convergence and mass-market IP-based services with seamless mobility and global roaming capability, as well as enhanced network QoS and security. New network architecture to integrate various radio access technologies under IP is defined in 3GPP to support seamless global roaming, interworking, and vertical handover between different access systems. In addition, IEEE 802.21 also defines a layer 2 solution to support mobility and media independent handover.

Multihop wireless networks are emerging as a promising architecture to extend wireless coverage in a flexible and cost-effective way. They have broad applications in Internet access, emergency networks, public safety, and so forth. Technical solutions for multihop wireless networks are being specified in IEEE 802.11s, 802.16j, 802.16m, 802.15.5, and 3GPP LTE-advanced. IETF has also defined routing protocols for mobile ad hoc networks.

2.2 Radio Technologies in Next-Generation Wireless Standards

2.2.1 Emerging IEEE WLAN Standards

The throughput of wireless LANs[1] keeps increasing with advances in radio technologies. The new IEEE 802.11n standard[2] is able to achieve up to 600 Mbps data rate when operating on 40 MHz bandwidth by using advanced physical layer techniques including MIMO and channel bonding. 802.11n supports backward compatibility with 54 Mbps 802.11a/g radios. At the MAC layer, it is still based on carrier-sensing multiple access with collision avoidance (CSMA/CA) contention-based media access, called enhanced distributed channel access (EDCA) and polling-based content-free media access, called hybrid coordination function controlled channel access (HCCA). To take advantages of high physical layer data rate and reduce protocol overhead, 802.11n defines two levels of aggregation at MAC layer. MAC Service Data Unit (MSDU) aggregation is processed at the top of MAC by packing multiple MSDUs into an aggregated MSDU, and MAC Protocol Data Unit (MPDU) aggregation is processed at the bottom of the MAC by packing multiple MPDUs into an aggregated MPDU. Block acknowledgment mechanism defined in 802.11e is also enhanced in 802.11n for better performance. These MAC features reduce the overhead, thus increasing the user-level data rate.

As wireless usage grows, there exists an increasing need for additional capacity. To provide comparable throughput as gigabit per second wired LAN products, a new task group (TG), 802.11ac[3] Very High Throughput for Operation in Bands below 6GHz, was formed in September 2008 to develop the enhancements to both the 802.11 PHY and MAC that enable modes of operation capable of supporting a maximum multistation (STA) throughput of at least 1 Gbps and a maximum single-link throughput of at least 500 Mbps while ensuring backward compatibility and coexistence with legacy IEEE 802.11 devices in the 5 GHz unlicensed band. 802.11ac will also provide enhancements over 802.11n on a set of other interdependent performance indicators including range of operation, spectrum efficiency, and power consumption.

In order to provide higher throughput than IEEE 802.11n, Space-Division Multiple Access (SDMA) has been proposed in the 802.11ac TG to handle multiple simultaneous communications between an access point and its associated stations. In general, SDMA employs multiuser MIMO (MU-MIMO) as a channel access method and allows a station to transmit (or receive) signal to (or from) multiple other stations in the same band simultaneously. Compared to point-to-point MIMO or single-user MIMO used in 802.11n, MU-MIMO leverage the availability of multiple independent stations and their diverse channel conditions to create parallel spatial channels using beam forming for superior communications performance in radio multiple access systems. Other techniques proposed to 802.11ac include backward compatibility and coexistence with 802.11n and other WiFi systems, support of more than 40 MHz channel bonding, and more than 4 MIMO antenna elements. The projected timeline for this task group is to have an initial draft by November 2010 and the approved standard in 2012.

For wireless access in vehicular environments (WAVE), IEEE 802.11 TGp[6] is specifying amendments to 802.11 to support Intelligent Transportation Systems (ITS) applications, which include data exchange between high-speed vehicles and between the vehicles and the roadside infrastructure in the licensed ITS band of 5.9 GHz. It specifies the functions and services that allow WAVE-conformant 802.11 stations to operate in a rapidly varying environment and to establish communications quickly each other. IEEE 1609 Family of Standards for Wireless Access in Vehicular Environments is a higher layer standard on which IEEE 802.11p is based.

IEEE 802.11ad[4] is developing technology to enable WLAN operation in the 60 GHz frequency band (typically 57–66 GHz). Due to high available bandwidth at 60 GHz band, multi-gigabit per second throughput can be achieved to support high throughput applications such as simultaneous streaming of multiple HDTV video streams or less compressed/uncompressed video streams, very-high-speed Internet access, wireless data bus for cable replacement, and so forth. It is expected that future mobile devices can be equipped with multiband WLAN access capabilities, short-range multi-Gbps throughput using 60 GHz band, and middle-range Gbps throughput operating at 5GHz band with seamless session transfer. 802.11ad is investigating the fast session transfer techniques between 60 GHz and 2.4/5 GHz. It is also studying the mechanisms that enable coexistence with other systems in the band, including IEEE 802.15.3c[14] systems.

2.2.2 Emerging IEEE WPAN Standards

Unlike wireless LANs, WPANs are used to convey information over relatively short distances, generally up to 10 meters, among a relatively few participants via power efficient and inexpensive networks. WPAN involves little or no infrastructure. IEEE 802.15 Task Group 3c (TG3c) is developing a millimeter-wave-based high-rate WPAN. The 802.15.3c WPAN will operate in the 60 GHz unlicensed band. The standard defines three PHY modes with different modulation and channel coding techniques, which can achieve a data rate up to 5 Gbps on the 2.16 GHz channel bandwidth.

802.15.3c MAC is based on 802.15.3 piconet with enhancements. A piconet is an ad hoc network that allows a number of devices to communicate with each other. One device acts as a piconet coordinator (PNC) that provides the basic timing for the piconet with beacons, and manages the QoS requirements, power save modes, and access control to the piconet. A piconet is formed without preplanning and as long as the piconet is needed.

Timing in the 802.15.3 piconet is based on the superframe composed of beacon, contention access period (CAP), and channel time allocation period (CTAP). The beacon is used to set the timing allocations and to communicate management information for the piconet. The CAP uses CSMA/CA as the medium access mechanism for commands and asynchronous data. CTAP is composed of channel time allocations (CTAs) that can be used for commands, isochronous streams, and asynchronous data connections.

Sensor networks will become part of Internet to provide various types of information. The IEEE 802.15 TG4 has defined the PHY and MAC specifications for low data rate, low complexity, and low power consumption WPANs for inexpensive devices. The 802.15 TG4 and its later enhancements TG4a, TG4c, and TG4d have defined various physical layer modes. These PHYs use different techniques such as spread spectrum or ultra-wideband (UWB), support different data rates from 20 Kbps to 27.24 Mbps, operates at different frequency band to meet different country's regulations – for example, 2.4 GHz ISM band, 915 MHz, 3 GHz to 5 GHz, and the like – and targets different applications such as sensors, interactive toys, smart badges, remote controls, and automation. The IEEE 802.15.4 standard is the basis for the ZigBee, WirelessHART, and MiWi specifications, each of which further offers a complete networking solution by developing the upper layers not covered by 802.15.4.

https://doi.org/10.1017/CBO9780511921117.003 Published online by Cambridge University Press

Depending on the application requirements, an IEEE 802.15.4 low-rate WPAN (LR-WPAN) may operate in either the star topology or the peer-to-peer (P2P) topology. It can be formed automatically. At the MAC layer, 802.15.4 LR-WPAN can use unslotted CSMA-CA or a superframe structure. A superframe contains contention free period (CFP) with guaranteed time slot for low-latency applications or applications requiring specific data bandwidth, as well as CAP with slotted CSMA-CA. The standard was developed with limited power supply availability of the devices in mind. A device may spend most of its operational life in a sleep state, only periodically listening to the channel in order to determine whether a message is pending.

Moreover, TG4f[10] is currently defining the new PHY layer and enhancements to the 802.15.4 MAC layer for active radio-frequency identification (RFID) systems. TG4g[11] is defining an amendment to 802.15.4 to facilitate very large scale process control applications such as the utility smart-grid networks, capable of supporting large, geographically diverse networks with minimal infrastructure, and potentially millions of nodes. The IEEE 802.15 TG6[12] is developing a standard for body area networks, and the IEEE 802.15 TG7[13] is defining a PHY and MAC standard for visible light communications (VLC). The low-power and low-cost sensor networks are expected to connect to the Internet in certain ways to provide various types of information.

2.2.3 Emerging 3GPP and IEEE Mobile Broadband Access Standards

Regarding cellular networks, the ITU-R has commenced the process of developing the International Mobile Telecommunications-Advanced (IMT-Advanced) systems standards[26,27,28] for next-generation (4G) mobile networks. The first invitation for the submission of proposals for candidate radio interface technologies (RITs) or a set of RITs (SRITs) for the IMT-Advanced was issued in March 2008. Under the current schedule, the deadline for submission of candidate RIT and SRIT proposals was October 2009, and it is anticipated that the development of radio interface specification recommendations will be completed in 2011.

According to ITU-R requirements, IMT-Advanced provides enhanced data rates to support advanced services and applications (100 Mbps for high mobility and 1 Gbps for low mobility were established as target peak downstream rates), as well as improved spectrum efficiency and battery life. It will be fully IP-based system with voice carried by VoIP, which is different from hybrid circuit-switching and packet-switching IMT-2000 (3G) mobile communications systems. IMT-Advanced also has capabilities for supporting high-quality multimedia applications in a cost-efficient manner, providing a significant improvement in performance, quality of service, and security. It has key features such as worldwide roaming capability, compatibility of services within IMT and with

https://doi.org/10.1017/CBO9780511921117.003 Published online by Cambridge University Press

fixed networks, capability of interworking with other radio access systems, and high-quality mobile services.

Both IEEE 802.16m[9] and 3GPP LTE-Advanced projects are developing advanced air interfaces to meet the cellular layer requirements of ITU-R IMT-Advanced. They are based on MIMO and OFDMA radio technologies with enhanced QoS and security. This reflects the technology trend from code division multiple access (CDMA) based hybrid circuit/packet switching 3G wireless systems to OFDMA-MIMO-based packet-switching 4G systems.

OFDMA employs orthogonal frequency-division multiplexing (OFDM) digital modulation scheme as a multiuser channel access strategy. It allows assigning subsets of subcarriers to individual users and simultaneously transmits to or receives signals from multiple users, achieving even better system spectral efficiency by leveraging channel frequency selectivity of multiple users and adaptive subcarrier assignment.

Compared to CDMA, OFDMA can better combat multipath and achieve a higher MIMO spectral efficiency because it can have flatter frequency channels than a CDMA RAKE receiver. In addition, OFDM is more flexible in the use of spectrum than CDMA. CDMA requires a wide bandwidth to maintain high chip rates and high spectral efficiency, and it is complex to implement radios with capability of different chip rates and spectrum bandwidths. 3G radio interface such as wideband CDMA (W-CDMA) thus defines the fixed 5 MHz channel spectrum bandwidth. However, this limits the flexibility in system deployment and the maximum bandwidth per handset. OFDMA can easily control the data rate and error probability of each individual user by dynamically allocating resources in the time and frequency domains. It offers a cost-efficient solution for wide bandwidth communications with high peak rates. Therefore, it is considered as more suitable for next-generation broadband wireless networks.

Evolved Universal Mobile Telecommunications System (UMTS) Terrestrial Radio Access (E-UTRA) was introduced in 3GPP Release 8 in 2009. E-UTRA aims at significantly increasing data rates for mobile stations, lowering end-to-end latency for real-time communications, and reducing setup times for new sessions. It uses OFDMA for the downlink and Single Carrier Frequency Division Multiple Access (SC-FDMA) for the uplink and employs MIMO with up to four antennas per station. It supports both single-user MIMO and multiuser MIMO for downlink, and SDMA for uplink. Both frequency-division duplexing (FDD) mode and time-division duplexing (TDD) mode with a number of defined channel bandwidths between 1.25 and 20 MHz are supported to provide system deployment flexibility. The E-UTRA provides a peak downlink rate of 300 Mbps with 4×4 MIMO antennas and a peak uplink rate of 75 Mbps for a mobile user over 20 MHz channel, which greatly improves network capacity over 3G systems. MIMO enables ten times as many users per cell as 3GPP's original W-CDMA radio access technology. E-UTRA also increased spectral

https://doi.org/10.1017/CBO9780511921117.003 Published online by Cambridge University Press

efficiency by two-to-four times compared to 3GPP CDMA-based UTRA interface. Improvements in architecture and signaling further reduce round-trip latency. It also enhances multicast service capability with single-frequency network support. In addition, E-UTRA improves coverage and battery life. However it is an entirely new air interface and incompatible with W-CDMA. E-UTRA is designed only to connect to 3GPP's new IP-based evolved packet core network.

3GPP is developing further advancements for E-UTRA, also called LTE-advanced, to meet all the IMT-advanced requirements for 4G, which is compatible with E-UTRA and expected to be included in 3GPP Release 10. 3GPP's proposal to ITU-R IMT-Advanced will be based on the LTE-Advanced. Multiple techniques including air interface optimization, scalable system bandwidth up to 100 MHz, enhanced precoding and forward error correction, hybrid OFDMA and SC-FDMA in uplink, relay nodes, advanced inter-eNodeB coordinated MIMO, and so forth are under investigation.

IEEE 802.16m is amending the IEEE 802.16 OFDMA specification to meet the cellular layer requirements of IMT-Advanced, while providing continuing support and upgrade path for IEEE 802.16–2005 based WiMAX OFDMA system. It supports scalable bandwidths from 5 to 40 MHz, with a normalized peak data rate of 15.0 bps/Hz for downlink (4×4 MIMO) and 6.75 bps/Hz for uplink (2×4 MIMO). Both TDD and FDD modes are supported. IEEE 802.16m aims to be the IEEE candidate radio interface for IMT-Advanced 4G mobile networks and compete with 3GPP LTE-Advanced.

Although 802.16m and E-UTRA adopts similar technologies such as OFDMA and MIMO, the differences in detail MAC and PHY layer design make them incompatible. 802.16m will be in conformance with the IEEE 802 architecture defined in 802.1 and provide seamless interworking with other IEEE 802 wired and wireless systems.

2.3 Spectrum Management and Cognitive Radio Networks

Cognitive radio technology allows either a network or a wireless node to dynamically change its transmission or reception parameters to communicate efficiently and to avoid interference with licensed or unlicensed users based on the active monitoring of its operation environment. In general, a cognitive radio system is reconfigurable and can take various external and internal radio environments such as radio frequency spectrum, user behavior, and network state into account to make decision, and adapts various parameters such as frequency spectrum, transmit power, transmit mode, media access method, and so on. More specifically, cognitive radios intelligently access and share radio spectrum by obtaining and sensing spectrum operating environment for efficient usage of licensed/unlicensed spectrum.

https://doi.org/10.1017/CBO9780511921117.003 Published online by Cambridge University Press

The radio frequency spectrum is a limited and valuable resource, but its usage is unbalanced. Some frequency bands are heavily used, for example, cellular network bands. However, a lot of frequency bands are inefficiently utilized, for example, amateur radio and paging frequencies. Furthermore, spectrum utilization depends strongly on time and place. Fixed spectrum allocation prevents the frequency spectrum unused by primary users from being used by unlicensed secondary users. Spectrum utilization can be improved significantly by allowing secondary users to access spectrum holes in the licensed band whenever it would not cause any interference to primary users. Cognitive radio has been proposed as the means for secondary users to utilize the spectrum holes, share the spectrum among them, and avoid the spectrum whenever primary users present.

In November 2008, the Federal Communications Commission (FCC) issued its report and order for unlicensed use of the TV white spaces. The TV white spaces are the frequencies that allocated to TV broadcasting, wireless microphones, and the like, but not used locally. Especially after full-power analog television broadcasts ceased operating in June 2009, many channels had freed up. The new FCC rules allow unlicensed devices to operate in the broadcast television spectrum at locations where that spectrum is not being used, given the secondary white space devices have cognitive radio and dynamic spectrum access capabilities, and shall not interfere the operation of primary users. The FCC currently requires that secondary devices must consult a frequently updated geo-location database to determine which channels are available for use at a given location. Other regulatory bodies such as ITU, European Radio Spectrum Policy Group (ERSPG), U.K. Ofcom, and Japan's Ministry of Internal Affairs and Communication (MIC) are also considering similar regulations.

Various proposals have advocated using TV white spaces to provide different services. The IEEE 802.22[16] working group is developing a standard for wireless regional area network (WRAN) that will operate in unused television channels. 802.22 WRAN mainly aims at providing wireless broadband access in rural areas using vacant TV channels in the VHF and UHF bands while avoiding interference to the broadcast incumbents in these bands. It typically operates with a coverage radius of 17 km to 30 km. 802.22 WRAN system uses TDMA/OFDMA similar to WiMAX networks, but it does not support MIMO because of the large antenna separation requirement at its low operating frequency.

Especially 802.22 specifies cognitive radio capability at the MAC/PHY air interface for dynamic frequency access. It can adjust to the location-dependent and time-variable spectrum availability to avoid interference to incumbents on a real-time basis. Specifically, 802.22 includes two new modules, namely Spectrum Sensing Function (SSF) and Geo-location module. The spectrum-sensing function monitors the RF spectrum of the television channels for a set of signal types and reports the results. The location information is important to protect

TV incumbent transmissions. The TV contours to be protected from interference are stored in a database. The base station (BS) controls the maximum allowed transmit power at individual CPEs using the collective knowledge of channel sensing, the CPE location, the TV operation database information, and so on. The standard also specifies the protocols for coexistence of multiple 802.22 cells.

Several other working groups in IEEE 802 are also studying TV white space. 802.11 has formed a task group 802.11af[7] for WLAN operation in TV white space 802.19, which has started studying coexistence of two or more unlicensed wireless networks such as WLANs, WMANs, WRANs, and ad hoc networks when they operate in the TV white space. Possible coexistence mechanisms under consideration include dynamic frequency selection and transmit power control, listen-before-talk media access or time division multiplexing of different wireless technologies, message-based on-demand spectrum contention based on coexistence beaconing or backhaul, as well as control through a centralized coexistence manager, coexistence database, or spectrum broker.

IEEE Standards Coordinating Committee (SCC) 41 is also developing standards related to dynamic spectrum access networks. The focus is on spectrum management, coexistence, reconfiguration, and dynamic spectrum access for cognitive radio. ITU and ETSI have also started the standard activities related to cognitive radio. In particular, ETSI's Reconfigurable Radio System (RRS) technical committee is defining the system functionalities related to spectrum management and joint radio resource management across heterogeneous access technologies, developing a functional architecture, and studying the concept of a Cognitive Pilot Channel (CPC) as an enabler to support the management of the reconfigurable radio systems.

2.4 All IP Mobile Networks

As part of LTE/System Architecture Evolution (SAE) effort, 3GPP defined the Evolved Packet System (EPS), an IP-based flat mobile network, to meet the increasing user and service demands, and to conform to Internet protocols for converging mobile and fixed network services. It aims at providing improved experience for users and increased performance and reduced cost for network operators. 3GPP All IP Network (AIPN) architecture represents its vision that next-generation mobile networks are based on core Internet protocols.

The existing 2G/3G networks consist of two subdomains: circuit switching for voice and packet switching for data, as shown in Figure 2.2.[18] The EPS unifies these two subdomains into a single end-to-end AIPN, in which voice calls are handled by VoIP using IP Multimedia Subsystem (IMS). EPS is able to integrate and support different radio access systems such as 3GPP radio access (LTE, 3G, and 2G) and non-3GPP radio access (CDMA 2000, WLAN, WiMAX),

https://doi.org/10.1017/CBO9780511921117.003 Published online by Cambridge University Press

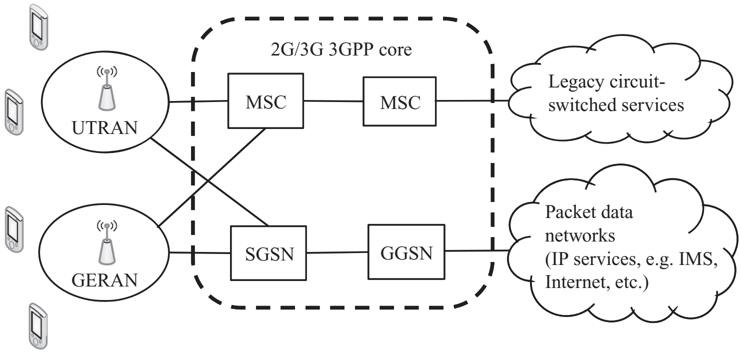

Figure 2.2. Simplified architecture of the 2G/3G 3GPP network.

as well as fixed access (Ethernet, DSL, cable, and fiber) with one common packet core network. It provides diversified mobile services with convergence to IP and enables the introduction of new business models and services, for example, partnering and revenue sharing with third-party content and application providers. It also supports incremental deployment because at the beginning, LTE may be only deployed at most needed areas and coexistence with legacy networks.

The IMS was originally standardized by the 3GPP to deliver IP multimedia services over cellular access networks (UMTS/GPRS networks). It was later enhanced to support other network accesses including Wireless LAN, CDMA 2000, and fixed networks. The IMS includes various control function components such as call session control functions (CSCF) and application servers, for example, the session initiation protocol (SIP) application server, service centralization and continuity (SCC) application server, with standard interfaces based on SIP and many related protocols. It controls the services with user registration, origination, termination, transfer, and release of multimedia sessions. The IMS provides a horizontal control layer that isolates the access networks from the service layer, and is able to maintain the services even when the user is moving across different access networks and terminal types. The user can connect to an IMS system from any access network through IP connectivity as long as it runs a SIP agent. The 2G or 3G circuit-switched network can also be supported as an access network to the IMS through gateways.

As shown in Figure 2.3,[18,19] the flat EPS architecture consists of two parts: the access network and the core network. 3GPP LTE specifies a new access network, E-UTRAN, which offers higher bandwidth, better spectrum efficiency, and better coverage. The core network is called evolved packet core (EPC), which consists of several major elements, including Serving Gateway (S-GW), Packet Data Network (PDN) Gateway (P-GW) and Enhanced Packet Data Gateway (ePDG), Mobility Management Entity (MME), Policy and Charging Rules

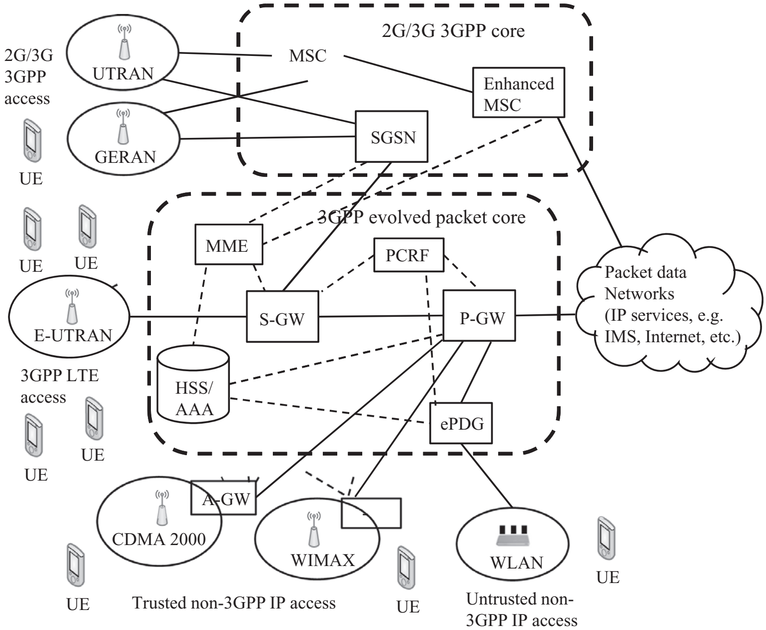

Figure 2.3. Simplified architecture of the 3GPP evolved packet system.

Function (PCRF), Home Subscriber Server (HSS), and Authentication, Authorization, and Accounting (AAA) Server.

The S-GW and P-GW are user data plane elements. S-GW manages user-planes mobility and also serves as a layer 2 mobility anchor as the User Equipment (UE) moves within 3GPP access. It maintains the IP data paths between eNodeBs and P-GW and separates the radio access network (RAN) and the core network. The eNodeBs are the LTE base stations in E-UTRN and combine the functionality of Node-B and RNC in 3G RAN. The P-GW provides access to different external packet data networks (PDNs) such as Internet, a corporate network, or an operator private network. It assigns an IP address to the UE from the address space of the PDN that can be an IPv4 address, an IPv6 prefix, or both. The P-GW performs policy and charging enforcement (PCEF) function such as packet filtering, service flow detection, dynamic policy and QoS enforcement, and flow-based charging control. It also serves as the mobility anchor point for the UE as the UE moves between access technologies. A UE may connect to multiple PDNs through EPC.

MME, PCRF, and HSS/AAA are control plan entities. MME is responsible for the signaling and control functions for UE access to network, session and mobility state management, authentication, and security by interacting with HSS. PCRF makes policy decisions for a user's IP data service flow and provides the

https://doi.org/10.1017/CBO9780511921117.003 Published online by Cambridge University Press

QoS policy and charging rules to the enforcement entities at P-GW and S-GW. HSS maintains user's subscription information, and AAA server supports authentication, authorization, and accounting.

Various access networks can connect to the EPC via different interfaces. Mobility management is an integrated part of the EPS. It provides seamless mobility at IP layer for intra- and interaccess system handover, and ensures the service continuity and QoS for a user's session as the user moves from one access technology to another as described in next section.

In EPS, data plane traffic is carried over virtual connections called service data flows (SDFs). SDFs are carried over bearers, that is, virtual IP transmission containers with unique QoS characteristics such as capacity, delay, packet loss rate, and so forth. A data path between a UE and a PDN, an end-to-end bearer, consists of three segments: a radio bearer between UE and eNodeB, a data bearer between eNodeB and S-GW, and a data bearer between S-GW and P-GW. A bearer exists per combination of QoS class and IP address of the terminal and identified by a unique identifier. The terminal may have multiple IP addresses, for example, when it is connected to multiple IP networks, each assigning it an IP address. It is possible that a terminal has multiple separate bearers associated with the same QoS class to multiple different PDNs. A packet flow is typically specified by an IP quintuple packet filter, that is, the source and destination IP addresses, source and destination port number, and protocol ID. Other filters can also be set up. The terminal (for uplink traffic) and the P-GW (for downlink traffic) classify the packets and map them into the corresponding bearers based on the packet filters. All the packets mapped into the same bearer receive the same packet-forwarding treatment such as scheduling, queuing management, rate shaping, and the like, in the network. The GPRS tunneling protocol (GTP) or proxy mobile IP can be used to implement the bearers in the EPC. Each IP packet entering the network is provided with a tunnel or proxy mobile IP header to route the packet to the destination and provide proper QoS.

The bearer-level QoS control enables network operators to manage the QoS for the different services, for example, mobile TV, telephony, Internet access, and the like, with different QoS requirements, and for each of its subscriber groups, for example, post- versus prepaid subscribers, home versus roaming subscribers. There are two types of bearers: guaranteed bitrate (GBR) and nonguaranteed bitrate (non-GBR) bearers. A GBR bearer typically is established on demand and may require for admission control and resource reservation. A non-GBR bearer can remain established for a long period of time because it does not reserve the resource. Once a terminal attached to the network, one default non-GBR bearer is set up per terminal IP address and remains as long as the terminal retains this IP address. The default bearer provides the basic connectivity and its QoS level is assigned based on the user subscription.

https://doi.org/10.1017/CBO9780511921117.003 Published online by Cambridge University Press

A set of policy and charging control (PCC) procedures has been specified in 3GPP release 8 to manage bearers, provide QoS to subscribers, and charge for the provided resources. PCC in EPS supports multiple-access technologies, IMS and non-IMS services, roaming, and mobility. In the PCC architecture, the application function (AF) – for example, a call-state control function in the IMS architecture – extracts the service-related information for a session by interacting with the applications that requires dynamic policy and charging control. It provides the PCRF with the service information, including traffic parameters such as IP addresses and port numbers, and QoS parameters such as type of traffic, data rate, and the like. The PCRF also obtains user-specific policies and information from the subscription profile repository, as well as user access information from S-GW and P-GW. The PCRF then makes the session policy decisions and provides the charging and policy rules to the policy and charging enforcement function (PCEF) at P-GW, and the policy rules to the bearer-binding and event-reporting function (BBERF) at S-GW. The PCC rules contain uplink and downlink packet filters to identify the service data flow, the gate control information to block or allow the IP flow, and its QoS behavior to be enforced such as QoS class, guaranteed bitrate, and so on.

The PCEF and BBERF are responsible for enforcing the PCC rules to ensure appropriate QoS for a service data flow. Once the PCEF or BBERF receives new or modified PCC rules for a service data flow, it creates or modifies the bearer and initiates resource reservation in the network. The PCEF also interacts with online charging system (OCS) for service access such as prepaid charging and reports the resource usage to the offline charging system.

The PCC provides seamless roaming support. The operators can apply the same dynamic policy and charging control and provide the same QoS to the user no matter whether the user accesses the home or visited networks. There are two different roaming scenarios in the LTE/SAE, namely home-routed access and visited access. In the home-routed roaming scenario, an IP connection with the outside PDN is established through a P-GW in the home public-land mobile network (H-PLMN) and the S-GW in the visit PLMN (V-PLMN). The home PCEF is responsible for the PCC enforcement. In the visited access, an IP connection with the outside PDN is established through a P-GW in the V-PLMN and a S-GW in the V-PLMN. The user data packets are routed through the visited P-GW between the outside PDN and the visited S-GW. The visited P-GW is connected to H-PCRF through V-PCRF to receive the PCC rules. It is also possible to use AFs in the V-PLMN for the visited-access roaming in which the signaling is proxied through the V-PRCF to the H-PRCF. Online charging is also connected to the home OCS through a proxy OCS.

3GPP LTE/SAE also specifies new security mechanisms to handle more diverse architecture with multiple access technologies and improves security robustness. EPS specifies four levels of security. Network access security (level I) protects the radio link and provides users with secure access to the EPC and the backend networks. Network domain security (level II) protects the wireline networks using the IPSec. User domain security (level III) provides the mutual authentication of the Universal Subscriber Identity Module (USIM) and the UE. Application domain security (level IV) enables the applications in the UE and the network to exchange data in a safe manner. The enhancements over UMTS include, among other things, increased security functions for protecting the confidentiality and integrity of signaling messages in access network, more secure key management and identity protection, and better interworking security between 3GPP and non-3GPP networks.

2.5 Mobility and Vertical Handover

It is expected that multiple access technologies will be seamlessly integrated into the global Internet. Both 3GPP and IEEE are defining standards to support mobility and vertical handover, that is, the handover from one network access technology to another. Vertical handover can greatly enhance the user experience. For mobile users, handovers can occur when wireless link conditions change due to the users' movement. For the stationary user, handovers become imminent when the surrounding network environment or application changes, making one network more attractive than another. In the handover, service continuity should be maintained. As an example, when making a network transition during a phone call, the handover procedures should be executed in such a way that any perceptible interruption to the conversation is minimized. Handover can be classified as hard and soft; hard handover is "break before make" regarding the exchange of data packets between the mobile terminal and the network, whereas the soft handover can achieve "make before break."

Generic Access Network (GAN), also called Unlicensed Mobile Access (UMA) defines a secure, managed connection from the 3GPP mobile core network to different devices/access points over IP, which was initially introduced in 3GPP Release 6. It allows extending the services and applications in an operator's mobile core (voice, data, IMS) over IP and Internet to other access technologies. One of applications of GAN is that with a dual-mode mobile phone, users can seamlessly roam and hand over between wireless LANs and cellular networks. When the mobile phone detects a wireless LAN, it establishes a secure IP connection to a GAN Controller (GANC) on the carrier's network. The GANC presents itself to the mobile core network as a standard cellular base station. The handset communicates with the GANC over the secure connection using existing GSM/UMTS protocols. Thus, when a mobile device moves from a GSM to an 802.11 network, it appears to the core network as if it simply attaches to a different base station. Femtocells, analog terminal adaptor for fixed line phone services, and UMA-enabled mobile VoIP clients for PCs are other GAN applications.

3GPP LTE/SAE further advances mobile networking technology by integrating various radio access networks under a single mobile core network. It specifies various interworking and mobility mechanisms based on all IP architecture to enable seamless handover between different access technologies and maintain IP services and voice calls continuity, which facilitate different deployment scenarios and support a flexible evolution path toward 4G. Multiple 3GPP or non-3GPP access networks can connect to the EPC through various access gateways. The EPS specifies different IP mobility mechanisms depending on the access technologies. For 3GPP-defined access technologies such as UTRAN, GERAN, E-UTRAN, either the GPRS tunneling protocol (GTP) or proxy mobile IPv6 (PMIPv6) can be used. For other accesses to connect to the EPC, any of PMIPv6, dual stack mobile IPv6 (DSMIPv6), or Mobile IPv4 (MIPv4) can be used.

PMIPv6 provides a network-based mobility mechanism. The mobile access gateway (MAG), that is, the 3GPP S-GW or the non-3GPP mobile access gateway (A-GW) acts as the proxy/foreign agent for the UE and handles the mobility management signaling. Once the UE has changed its point of attachment to a new mobile access gateway, the new MAG provides the UE with the same IP address as the UE had at its previous point of attachment. The new MAG also handles updating the mobility anchor in the network so that the packets arrive at the new point of attachment. The UE is not aware of the mobility management signaling. On the other hand, DSMIPv6 and mobile IPv4 are client-based. The UE obtains a new care-of address when it moves to a new point of attachment. The UE is responsible for updating its home agent about the binding between the care-of address and the home address of the UE. Compared to the client-based mobility management, the network-based mobility management reduces the UE involvement in mobility management, leading to better UE battery life, less wireless resource usage, and reduced latency in handover.

When terminals move across areas served by eNodeB elements within E-UTRAN, the S-GW serves as a local mobility anchor. The S-GW also serves as an anchor for the mobility within other 3GPP-specific access technologies. S-GW relays packets among eNodeB, P-GW, and legacy SGSN for intra E-UTRAN mobility and mobility with other 3GPP technologies, such as 2G GSM and 3G UMTS.

All data paths from the access networks are combined at the P-GW and routed to the external packet networks. Mobility management between 3GPP and non-3GPP access systems are involved by multiple data plane and control plane entities, including P-GW, S-GW, non-3GPP access gateway, PCRF, and MME based on mobile IP technology. For interaccess handover, 3GPP defines nonoptimized handover and optimized handover procedures, depending on whether the source network is involved in preparing resource in the target network during the handover. Optimized handover is more suitable when the UE cannot simultaneously access the source network and the target network.

Figure 2.4 illustrates the high-level message flow for nonoptimized handover when an UE hands over a VoIP call from a trusted non-3GPP access to a 3GPP LTE E-UTRAN access.[19] PMIPv6 is used in the EPC for this example. The UE initially decides to attach to the trusted non-3GPP access. It initiates an attachment request toward the access gateway via the base station. The UE and the network perform the mutual authentication. After the authentication, an IP address is assigned to the UE, a PMIPv6 tunnel is setup between the P-GW and A-GW, the default access bearer is established, and the UE attaches to the trusted non-3GPP access network. When the subscriber places an IMS VoIP call, the SIP protocol is used to set up the call. The end-to-end signaling is intercepted by the IMS CSCF function in the network. The CSCF extract and pass the session information to the PCRF. The PCRF makes the decisions on charging and QoS rules and sends them to the PCEF at P-GW and BBERF at A-GW. A voice bearer is then set up to carry the call. When the UE decides to hand over to the 3GPP access from the trusted non-3GPP access, it initiates the handover attach procedure to the 3GPP S-GW through eNodeB using its E-UTRAN interface. The 3GPP S-GW obtains the QoS rules for both the default traffic and the VoIP traffic from the PCRF and prepares the resource with the appropriate QoS in the 3GPP access network. Through the proxy binding update between the 3GPP S-GW and P-GW, the P-GW provides the same IP address used by the UE in the non-3GPP access to the S-GW. Meanwhile, the P-GW also updates the PCRF with the UE's handover request and obtains the corresponding charging rules. The default bearer and the dedicated bearer are established in the 3GPP access network, and a PMIPv6 tunnel is set up between the P-GW and 3GPP S-GW. The UE then completes the attachment to the 3GPP. The tunnel is then switched and the traffic is routed through the 3GPP access between the UE and the P-GW. The resource in the non-3GPP access is released.

Dual-radio-capable UEs can simultaneously access both the source and the target networks, and seamless handover can be achieved using the previously mentioned nonoptimized handover. However, if the UE cannot access the two networks simultaneously, a "make-before-break" handoff cannot be achieved with nonoptimized handover, leading to service interruption during interaccess handover. Therefore, optimized handover is specified in 3GPP LTE/SAE to enable seamless handover even for single-radio UEs. In the optimized handover, a tunnel is established between the source system and the target system so the UE can communicate with the target system through the source system and prepare for the handover before the real handover occurs.

In E-UTRAN, voice calls are carried with VoIP technologies and offered as IMS-based services. However, in legacy 2G/3G networks, voice calls are

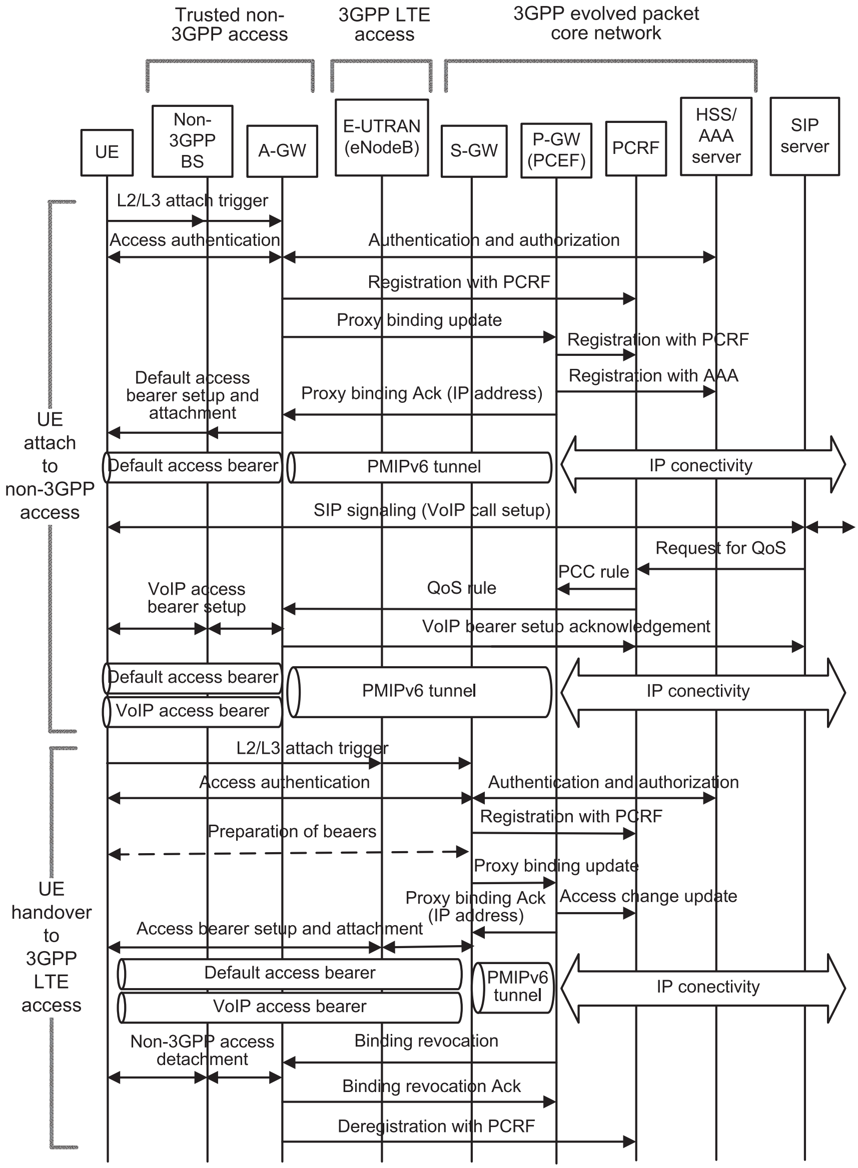

Figure 2.4. Message flow diagram for the scenario when a UE attaches to non-3GPP access and then hands over to 3GPP LTE E-UTRAN access.

carried with traditional circuit switching (CS) technologies. Mobile IP itself cannot meet the voice call continuity requirement. 3GPP LTE/SAE also specifies seamless voice call handover mechanisms between E-UTRAN and various 2G/3G radio accesses, which transfer the call between the CS and IMS domains. It supports the call continuity for single-radio terminals and the handover of

https://doi.org/10.1017/CBO9780511921117.003 Published online by Cambridge University Press

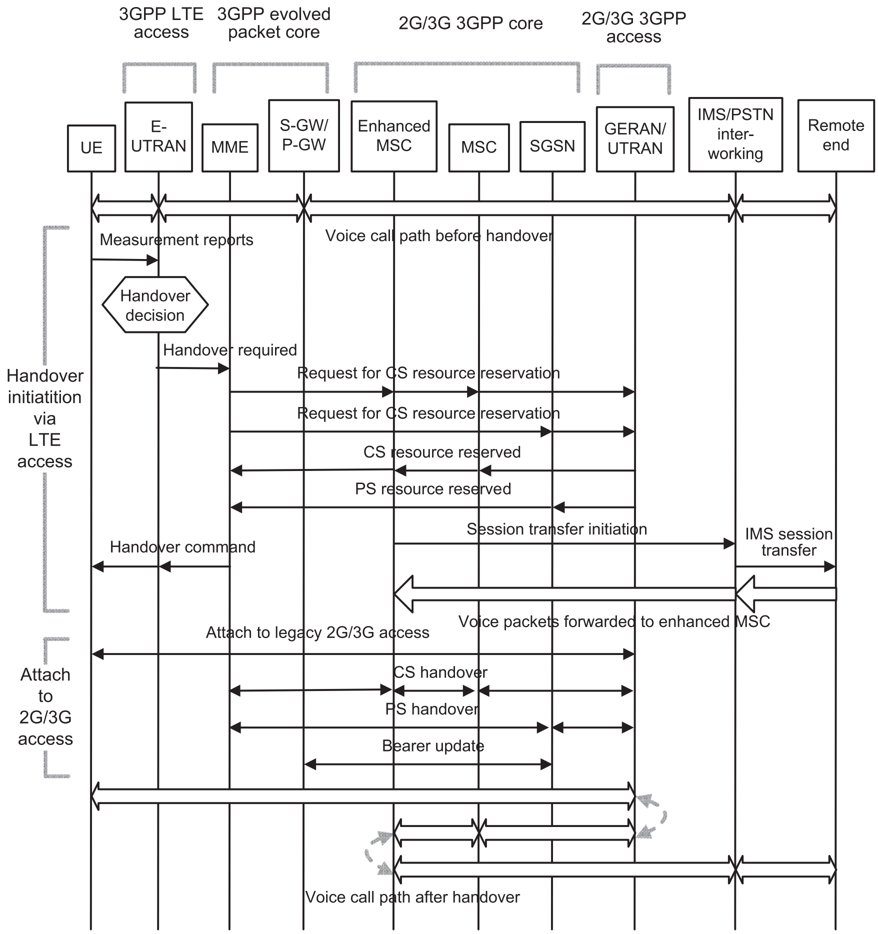

Figure 2.5. Message flow diagram for voice and nonvoice single-radio handover from 3GPP LTE E-UTRAN access to 2G/3G 3GPP access.

associated nonvoice sessions. As an example, Figure 2.5 shows the high-level message flow of the voice and nonvoice single-radio handover procedure from E-UTRAN to UTRAN/GERAN.[18] At least one MSC in the traditional CS domain is enhanced with interworking functionality and a new interface Sv. The MME in the EPC also needs additional functionality to support the Sv interface and the associated single-radio voice call continuity procedure. Due to the make-before-break approach, the voice interruption in the handover procedure is normally less than several hundreds of milliseconds, which should be imperceptible to the user. However, it cannot be guaranteed that the QoS of the nonvoice session is sustained after the handover because of the bandwidth limitation in the UTRAN/GETRAN.

It is possible that voice services are not initially supported over the E-UTRAN access due to the cost of VoIP service deployment. 3GPP also defines the fallback

https://doi.org/10.1017/CBO9780511921117.003 Published online by Cambridge University Press

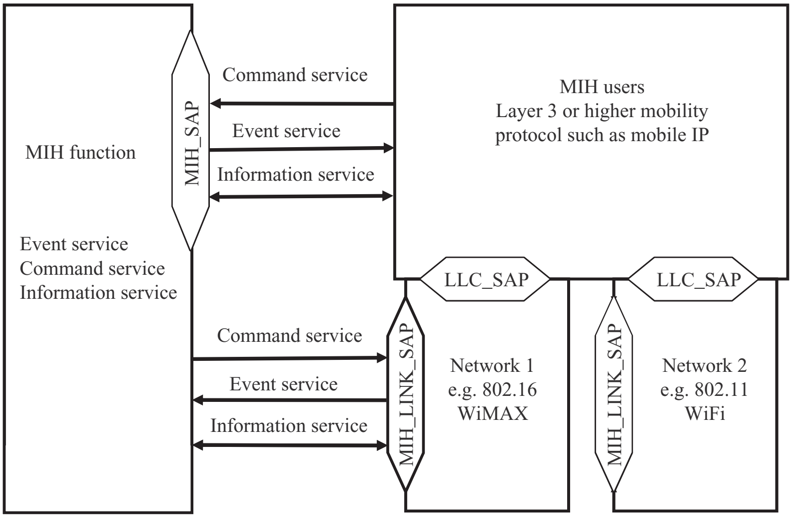

Figure 2.6. 802.21 media independent handover (MIH) services.

mechanisms to handle this case. It hands over the user to the legacy 2G/3G access when a voice service is requested. After the user falls back on the 2G/3G access, the standard CS voice call setup procedure is used to establish the call. Furthermore, 3GPP also specifies the IMS service continuity procedures to hand over a multimedia session above the transport layer based on SIP protocol. One of the benefits with the IMS is that the same service continuity procedures can be used no matter what the source and target accesses are. The IMS-based application layer handover mechanisms provide additional tools for mobility, especially when vertical handover is not supported by the network layer. Note that the pure application layer handover such as that supported by the IMS normally takes a longer time, especially for single-radio terminals, and may lead to perceptible service interruption.

IEEE 802.21[17] is also developing media independent handover (MIH) standards to enable handover and interoperability between heterogeneous networks including both wired and wireless, 802 (e.g., 802.11, 802.16, Ethernet), and non-802 networks (such as cellular). Compared to similar technologies defined by 3GPP (UMA and SAE vertical handover), 802.21 does not assume a 3GPP core network. It intends to provide generic link-layer intelligence independent of the specifics of mobile nodes or radio networks.

As shown in Figure 2.6,[17] 802.21 defined a framework and a set of control functions to facilitate the media independent handover. Specifically, it defined a new logical control entity, called the MIH function (MIHF), in the framework that locates on the mobile nodes or in the network, and provides the event, command, and information services to facilitate seamless handovers between heterogeneous networks. It also standardizes a generic MIH service access point, called the MIH_SAP, as well as associated primitives that provide MIH users with access to the MIHF services. For support of remote MIHF services and communications between the peer MIHF entities, 802.21 specifies a MIH protocol. However, the MIHF in 802.21 depends on the presence of a mobility management protocol stack, for example, mobile IP, within the network elements that support the handover. Enhancements at the higher layer and link layers are required to support the function abstractions of this standard and carry the messages defined in this standard. In addition, handover policies and other algorithms involved in handover decision making do not fall within the scope of the standard, which are left to the network operators and applications. Handover decision making involves cooperation of mobile nodes and network infrastructure. The 802.21 standard supports both hard and soft handover procedures. 802.21 WG is working on the extensions to the basic 802.21 specification to add security mechanisms and support of handover for downlink-only broadcast networks such as DVB network.

2.6 Multihop Wireless Networks

Multihop wireless mesh networks (WMNs) are a promising technology to extend wireless coverage in a flexible and cost-effective way. WMNs can be infrastructure-based or infrastructureless. In infrastructureless WMNs, client stations such as laptops, smart phones, and so on are equipped with mesh-routing functions and form a network on an ad hoc basis to forward the traffic to each other without dedicated infrastructure, in which each node is a mesh router and an end device. In infrastructure WMNs, mesh routers or mesh access points (MAPs) constitute a multihop wireless infrastructure. One or more mesh router/MAP can be connected to the other wired or wireless networks or the Internet, acting as the mesh gateway. Client stations without mesh functions do not participate in the packet relay, but associate with a MAP to obtain the network access. The MAPs forward traffic for the client stations in the mesh.

Industry standards are being developed in the IETF for mobile ad hoc networks (MANET) routing protocols, in the IEEE 802.11s[5] for WiFi-based mesh networks, and in 802.15.5[15] for wireless PAN mesh. Next-generation WiMAX networks based on 802.16j[8] and 802.16m[9] will support multihop relay. 3GPP LTE-advanced is also considering multihop relay architecture for next-generation cellular networks.

2.6.1 IETF MANET Routing Protocols

Radio nodes in a multihop WMN self-organize themselves in a mesh topology and self-heal from failures using discovery and routing protocols, which enhances the network reliability. The nodes cooperatively make forwarding

https://doi.org/10.1017/CBO9780511921117.003 Published online by Cambridge University Press

decisions based on a routing protocol. Many mesh-routing schemes have been proposed in research literatures. IETF MANET working group (WG) has standardized a few of IP routing protocols that can be applied for general wireless mesh networks consisting of mobile routers or fixed routers, or a mixture of both. IPV4 and IPv6 are both supported. The WG has developed two tracks of routing protocol specifications: reactive/on-demand MANET protocol (RMP) and proactive MANET protocol (PMP). In a proactive routing protocol, each node establishes and maintains routes to all reachable destinations at all times, whether or not there is currently any need to deliver packets to those destinations. In contrast, an on-demand routing protocol discovers and maintains routes only when they are needed.

Ad hoc On-Demand Distance Vector (AODV) routing[29] is a typical on-demand routing protocol specified by IETF MANET WG. In AODV, when a route to a new destination node is needed, the originating node floods a Route Request (RREQ) message to discover the route to the destination. The intermediate nodes propagate the RREQ hop by hop and also create a reverse route to the originator in its routing table based on the distance vector. When the target receives the RREQ, it responds with a Route Reply (RREP) sent hop by hop in unicast toward the originator over the reverse route. Each intermediate node that receives the RREP creates a route in its routing table to the destination. When the originator receives the RREP, the route has been established between the originator and the destination in both directions. It is also possible that an intermediate node with a valid route to the destination responds to the RREQ with a RREP to reduce route setup time. To maintain the active route and react to changes in the network topology, nodes monitor the link status through optional Hello messages and traffic flow over the link. If a node detects a link break for the next hop of an active route or receives a data packet for forwarding to a destination for which it does not have an active route, it sends the Route Error (RERR) toward the originator of the packet to notify the loss of the link to the other nodes that use this route. The originating node will delete the route when it receives the RERR and initiate a route discovery again if it needs to send the packet to the same destination. The WG later specified a modified version of AODV, called Dynamic MANET On-demand (DYMO) Routing. DYMO uses a more generic and flexible message format, and enables DYMO routers to perform routing on behalf of other attached nodes. The Dynamic Source Routing (DSR)[32] is another on-demand routing protocol defined by the IETF MONET WG. Unlike the AODV, DSR uses source routing to forward the packets.

The IETF MONET WG standardized two proactive routing protocols: Optimized Link State Routing Protocol (OLSR)[30] and Topology Dissemination Based on Reverse-Path Forwarding (TBRPF).[31] OLSR is a table-driven, proactive protocol. It optimizes the classic link state protocol by considering the

MONET requirements and broadcast wireless media characteristics to reduce the number of transmissions in the process of control traffic flooding. In OLSR, each router selects a subset of its neighbor routers as “MultiPoint Relays” (MPRs) to retransmit the broadcast messages from it so that the broadcast messages, retransmitted by these selected MRPs, will reach all nodes two hops away. A node only forwards the broadcast messages directly received from its MPR selectors, that is, the nodes that have selected it as an MPR. The neighbors of a node N that are not in its MPR set receive and process broadcast messages but do not retransmit the broadcast messages received from node N. Hello messages are used between neighbor nodes for link sensing, 1-hop and 2-hop neighbor detection, and MPR selection. This technique facilitates efficient flooding of control messages in the network as compared to a classical flooding mechanism, where every node retransmits each message when it receives the first copy of the message. To reduce the number of control message transmissions further, the link state information may only be generated by nodes elected as MPRs, that is, MPRs declare the link state information for their MPR selectors. In addition, an MPR node may chose to report only links between itself and its MPR selectors. Then in route calculation, the MPRs are used to form the route from a given node to any destination in the network. The WG has also specified an updated version of the OLSR, OLSR version 2, which retains the same basic mechanisms and algorithms while providing a more flexible signaling framework and some simplification of the messages being exchanged.

TBRPF is another proactive, link-state routing protocol standardized by the MONET WG. In TBRPF, each node computes and updates a source tree that provides the shortest paths to all reachable destinations, based on partial topology information stored in its topology table. Instead of disseminating the link states for all the links, each node reports only part of its source tree to neighbors, that is, the status of the links consisting of this reported subtree, to minimize the overhead.

Compared to proactive routing with on-demand routing, proactive routing protocols generally have the advantage of routes immediately available when needed because a node continuously maintains routes to all destinations in the network. The proactive protocols are beneficial for traffic patterns where a large subset of nodes are communicating with another large subset of nodes, and the source and destination pairs are changing over time. However, the proactive protocols incur more routing overhead to keep the routing information current, especially when the nodes are moving or the network topology changes frequently. On the other hand, the on-demand protocols require less routing overhead because they do not maintain the unused routes. However, they require more time to discover and establish a route when the route is needed, leading to extra route discovery delay and data buffering at the source.

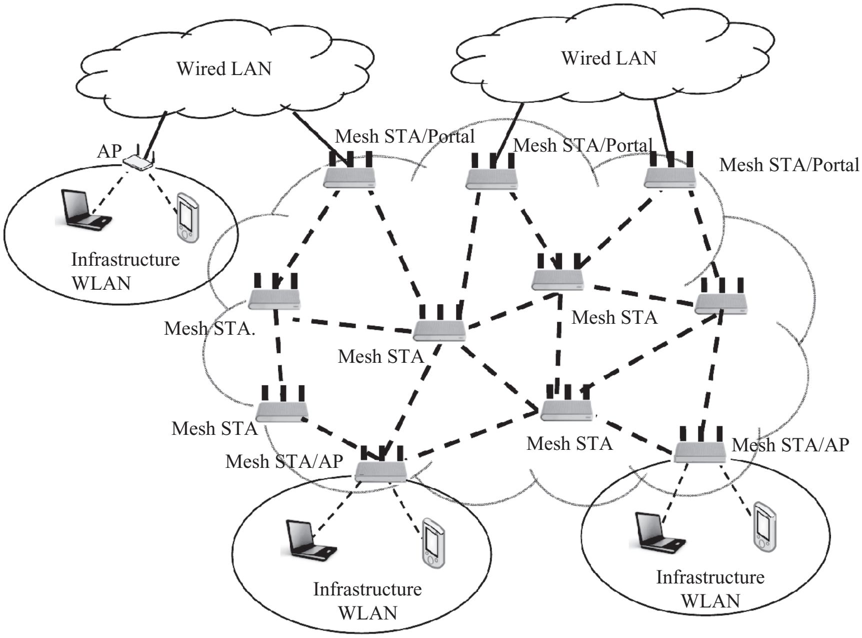

Figure 2.7. An example of a multihop wireless mesh network.

2.6.2 IEEE 802.11s WLAN Mesh Networking

Wireless mesh networks can be implemented with various radio technology including 802.11, 802.16, 802.15, cellular radios, or combinations of more than one type. IEEE 802.11s is developing a mesh networking standard using the IEEE 802.11 MAC/PHY layers. Within the scope of the IEEE 802 standards, IEEE 802.11s addresses layer 2 operations. Compared to IP routing, layer 2 routing uses the MAC address. There is no IP address assignment issue in the use of layer 2 mesh routing protocols. In terms of implementation, the layer 2 mesh software can be incorporated in the drivers and offered by the IC vendors.

The IEEE 802.11s extends the 802.11 MAC layer by defining the architecture and a set of mechanisms to support multihop mesh networking. An example of a 802.11 wireless mesh network is illustrated in Figure 2.7.[5] A mesh station (STA) may be collocated with one or more other entities (e.g., AP, portal, etc.). A mesh AP that incorporates a mesh STA with one or more access points can provide both mesh functionalities and AP functionalities simultaneously. Client STAs associate with APs to gain access to the network, and do not participate in mesh functionalities such as path selection and forwarding. Mesh portals interface the mesh network to other networks.

https://doi.org/10.1017/CBO9780511921117.003 Published online by Cambridge University Press

The main functionalities of IEEE 802.11s include mesh discovery, authentication and link security, peering management, channel selection, routing and forwarding, interworking, congestion control, synchronization and beaconing, and power save. Traditional 802.11 EDCA is the basic medium access protocol. 802.11s also defines an optional channel access mode, called Mesh Coordinated Channel Access.

IEEE 802.11s standardizes the procedures for mesh STAs to discover one another and to organize themselves into a mesh network. A mesh STA discovers candidate peer mesh STAs and their configuration by listening to the beacons sent by its neighbors or using proactive probe request/probe response message exchanges. After discovering a candidate peer mesh STA, a mesh STA can establish peering with the candidate peer mesh STA, secured or not, depending on the local policy. If the peering does not require security, the mesh STA will initiate the peering management protocol to the candidate mesh STA. The peering management protocol allows exchanging and confirming the configuration parameters such as mesh ID, active path selection protocol and metric IDs, and so on, and establishes a peer session through handshake. If the peering requires security, the mesh STA shall initiate a secure authentication protocol. IEEE 802.11s defines an authentication protocol, called Simultaneous Authentication of Equals (SAE), to provide mutual authentication between two mesh STAs using a shared key or password. If the secure authentication protocol succeeds, the two mesh STAs obtain a common pairwise master key (PMK). The IEEE 802.11s further defines two mesh link security protocols, called the authenticated peering exchange and the mesh group key handshake. The mesh link security protocols rely on the existence of the common PMK at the two mesh STAs established by executing the authentication protocol. The authenticated peering exchange protocol is used to authenticate a peering using the PMK, to establish session keys for protecting unicast traffic between two peers, and exchange the group keys. A group key is assigned by the broadcast/multicast source and is used to protect broadcast/multicast traffic from that source. The mesh group key handshake allows a mesh STA to update its group key.

IEEE 802.11s also specifies the channel switching procedures that can be used to satisfy regulatory requirements such as radar signal detection, or to reassign the mesh STA channel to ensure the network connectivity, or other reasons. A mesh STA that determines the need to switch the channel transmits a Mesh Channel Switch Announcement frame to each of its peer mesh STAs to announce its intent through unicast or broadcast. It also includes the channel switch announcement information in its beacon frames and probe response frames during the channel switch process.

IEEE 802.11s specifies an extensible routing framework to enable flexible implementation of path selection protocols and metrics. The standard includes a default mandatory path selection protocol, Hybrid Wireless Mesh Protocol

(HWMP), and default mandatory path selection metric (Airtime Link Metric) for all implementations, to ensure interoperability. Even though the extensible framework allows multiple protocol and metric implementations, only one path selection protocol and one path selection metric shall be actively used by a mesh STA at a time, which is announced by the mesh network in the beacons and probe responses.

The default HWMP protocol specified by 802.11s combines the flexibility of on-demand path selection with proactive topology tree extensions, which takes advantages of both proactive and reactive routing approaches, and enables efficient path selection in a wide variety of mesh networks. HWMP can concurrently operate in on-demand mode and proactive tree-building mode. The on-demand mode of the HWMP protocol is based on the AODV with many enhancements and adapted to the MAC address-based path selection and radio link metric awareness. It allows mesh STAs to communicate using P2P paths. The proactive tree building mode is used to establish the paths between a root mesh STA and the rest of the mesh STAs in advance, so that the communications can begin instantly without executing the path selection operation. This can be performed by configuring a particular mesh STA as a root mesh STA and periodically broadcasting its existence to the rest of the mesh so that every mesh station could create a path to the root station. Typically root STAs are the STAs that act as portal to Internet access. One example of concurrent usage of on-demand and proactive modes is that two mesh STAs begin communicating using the proactively built tree via the root, but subsequently perform an on-demand discovery for a direct path between each other. This type of concurrent usage of the proactive and on-demand modes allows communication to begin immediately while an on-demand discovery finds a more optimal path between two mesh STAs.

The default routing metric specified by 802.11s is the airtime link metric, which takes the link data rate and frame error rate of a wireless link into account. Airtime reflects the amount of channel resources consumed by transmitting a frame over a particular link. The total cost for a path is the sum of the cost of the links on the path.

A 802.11s mesh network functions like an IEEE 802 LAN segment. It can have zero or more portals that can be connected to one or more LAN segments. If two portals connect a mesh to an external LAN segment, broadcast loops may occur, and the IEEE 802.1D bridging protocol can be used to turn off the LAN port of one of the portals for preventing from traffic looping. A portal can send the portal announcement (PANN) to advertise its presence. Portal Announcements allow mesh STAs to select the appropriate portal and build a path toward it.

A mesh STA can serve as a proxy for nonmesh STAs, transmitting and receiving the frames on behalf of the proxied STAs through a tunnel. For example, a mesh AP can serve as a proxy for the client stations associated with it, and a mesh portal can serve as a proxy for an entity behind of it. 802.11s defines the signaling protocol for a proxy mesh station to send the proxy information, including the MAC address of the proxied entities, to a destination mesh STA.

802.11s does not define a multicast routing protocol. Multicast frames are forwarded as broadcast frames. To improve multicast reliability, the standard allows an implementation of multiple unicast transmissions to transmit a multicast frame, which are individually acknowledged. In such a case, the multicast frame can be converted to individually addressed frames and transmitted as individually addressed frame to each of the peer mesh STAs.

IEEE 802.11s also specifies a Congestion Control Signaling protocol. A mesh STA that detects congestion and the incoming traffic sources causing this congestion may transmit a Congestion Control Notification frame to the source mesh STAs or other neighboring mesh STAs. Of course, specific algorithms for local congestion monitoring and congestion detection, as well as local rate control, are beyond the scope of the standardization and are left to the implementers for innovation.

To detect and mitigate the collisions of beacons transmitted by different stations on the same channel within 2 hop range, a Mesh Beacon Collision Avoidance (MBCA) mechanism is specified by IEEE 802.11s. A mesh station reports the target beacon transmission time and beacon interval of its neighboring STAs in its beacon or probe response frames. Using this information, a mesh STA can select and adjust its target beacon transmission time and beacon interval so that its beacon frames do not collide with the beacon frames transmitted by other STAs in a 2-hop range. In addition, a mesh STA can send a message to request its neighbor to adjust the target beacon transmission time.

IEEE 802.11s specifies power save operation. A mesh STA has the capability to buffer frames and track the power mode of a peer mesh STA. A mesh STA uses the peer service periods for unicast frame transmissions to a neighboring peer mesh STA in power save mode. A peer service period is directional. To trigger a peer service period, a mesh STA in power save mode can send a peer trigger frame to its peer, and a mesh STA can also send a peer trigger frame to the mesh STA in power save mode during its Mesh Awake Window. The Mesh Awake Window of a mesh STA is announced in its beacon and probe response frames.

In addition to the traditional EDCA, 802.11s standardizes an optional medium access method, called Mesh Coordinated Channel Access (MCCA), which allows MCCA-capable mesh STAs to access the wireless medium at selected time periods with lower contention. MCCA can be used by a subset of mesh STAs in a mesh network. However, MCCA connections can only be set up among MCCA-enabled mesh stations and their performance may be impacted by the devices that do not respect MCCA reservations. MCCA uses management frames to determine a series of target transmission starting times and durations,

https://doi.org/10.1017/CBO9780511921117.003 Published online by Cambridge University Press

called MCCA Opportunities (MCCAOPs), between an MCCAOP owner and one (for individually addressed frames) or more (for group addressed frames) MCCAOP responders for frame transmissions. These MCCAOPs are advertised in the neighborhood around the MCCAOP owner and responders. The MCCA mesh STAs in this neighborhood that could cause interference to transmissions during these MCCAOPs, or that would experience interference from them, shall refrain from accessing the wireless medium during these MCCAOPs. The MCCAOP owner and the MCCAOP responders access the wireless medium during these MCCAOPs using contention-based channel access (EDCA) because some other stations may not respect the MCCA reservations.

Synchronization is needed between mesh STAs that use MCCA, MBCA, or operate in power save mode. IEEE 802.11s introduces an extensible framework to enable implementation of multiple synchronization protocols for mesh STAs. It also includes a default mandatory protocol, called the neighbor offset synchronization protocol, to enable minimal synchronization capabilities and interoperability. With the neighbor offset synchronization protocol, a mesh STA should maintain a timing offset value between its own time synchronization function (TSF) timer and the TSF timer of each neighbor mesh STA with which it synchronizes. A mesh STA can start its TSF timer independently of other mesh STAs, and can update the value of its TSF timer offset based on the time stamps received in the beacon or probe response frames from other mesh STAs.

2.6.3 IEEE 802.16j WMAN Multihop Relay

IEEE 802.16j has specified enhancements to the IEEE 802.16 OFDMA-based PHY and MAC layers to enable the operation of multihop relay stations. However, 802.16j only supports tree topology consisting of one or more relay stations (RS) rooted at a multihop relay base station (MR-BS). Traffic between the subscriber station (SS) and MR-BS is relayed by one or more RS. Each RS is under the supervision of the MR-BS. The RS can be fixed (e.g., attached to a building) or mobile (e.g., traveling with a transportation vehicle). The SS can also communicate directly with the MR-BS. However, it does not allow the P2P communications between relays. The standard specifies new functionality on the relay link to support the multihop relay features. But the protocols on the access link between the SS and RS/MR-BS are not changed from 802.16. 802.16m will also support multihop relay based on the techniques developed in 802.16j.

Two different modes, namely centralized and distributed scheduling, are specified for controlling the allocation of bandwidth for an SS or an RS. In centralized scheduling mode, the bandwidth allocation for an RS's subordinate stations is determined by the MR-BS; in contrast, for distributed scheduling mode, the bandwidth allocation of an RS's subordinate stations is determined by the RS, in cooperation with the MR-BS. Note that the standard only provides the signaling support for centralized and distributed scheduling. The scheduling algorithms themselves are out of the scope of this work and are left for the innovation by the implementers. Two different security modes, namely centralized security mode and distributed security mode, are also defined in 802.16j. The centralized security mode is based on the key management between an MR-BS and an SS. The distributed security mode incorporates the authentication and key management between an MR-BS and an access RS and between the access RS and an SS.

The MAC layer enhancements defined in 802.16j include signaling extensions to support functions such as network entry of an RS, and of an SS through an RS, bandwidth request, packet forwarding, connection management, and handover. The PHY enhancements include extensions to the OFDMA-PHY for transmission of data across the relay links between the MR-BS and the RS.

In 802.16, connections are identified by a 16-bit connection ID (CID), and the CID is carried in the MAC header. At a SS or RS initialization, the management connections are established between the SS/RS and the MR-BS, which are used to carry management traffics. An RS may be configured to operate either in normal CID allocation mode, where management CIDs are allocated by the MR-BS, or in local CID allocation mode where the MR-BS allocates the management CID range to a subordinate RS that assigns CIDs from this range to its subordinate stations. Data traffic connections are established dynamically. Connections may span multiple hops and may pass through one or more intermediate RSs. The CIDs will be unique within an MR cell. In addition, a tunnel connection can be established between the MR-BS and an access RS. Tunnel connections can be used for transporting relay traffic from one or more connections between the MR-BS and the access RS, and can pass through one or more intermediate RSs.

In the network entry, an RS scans the preambles transmitted by the existing MR-BS(s) or RS(s), synchronizes with the MR-BS, and selects a temporary RS/MR-BS to access the network. It then obtains transmission parameters, performs ranging, negotiates basic capabilities, and performs authorization, security key exchange, and registration with the MR-BS. Then the MR-BS obtains the neighbor station measurement reports and selects the final access station for this new RS. After that, the path is created, and the tunnel and IP connectivity are also established. The MR-BS transmits the operation parameters to the RS to configure it. The SS uses similar procedures for the network entry. The differences are the SS will select the RS/MR-BS as the access station once after scanning. It does not perform the neighbor station measurement and the second stage of access station selection, and the path and tunnel establishment.

The MR-BS can instruct the RSs to perform complete neighborhood discovery and measurement. Based on the topology information obtained from topology discovery or update process, MR-BS makes centralized calculation for the path between the MR-BS and an access RS for both the uplink and downlink direction. The path creation is subject to the constraints of tree topology, that is, an RS shall have only one superordinate station and other constraints such as the availability of radio resource, quality of the link, load condition of an RS, and so forth. The specific path calculation algorithms are out of scope of the standardization and left to the implementers for innovation.

802.16j defines two path management modes: the embedded path management and explicit path management. In the embedded path management mode, the MR-BS systematically assigns CIDs to its subordinate stations such that the CIDs allocated to all subordinate RSs are a subset of the allocated CIDs for that station. The network topology is embedded into a systematic CID structure to help RSs find routing paths without storing all CIDs of subordinate RSs in the routing table, which means the packets for a connection are routed based on the CID assignment structure. This is similar to a telephone call being routed based on the telephone number. The CIDs are assigned systematically, using either contiguous integer block or bit partitioning methods. In the explicit path management, after a MR-BS discovers the topology between a newly attached MS or RS and itself, or detects a topology update due to events such as mobility, MR-BS may remove an old path, establish a new path, and notify the new path information to all the RSs on the path. When connections are established or removed, MR-BS may distribute the mapping information between the connection and the path to all the RSs on the path. This is similar to packet routing in data networks and requires for routing table in each node. With this method, it is possible to have multiple paths between a SS/RS and the MR-BS.

In general, while a relay station is transmitting a signal, other neighboring stations do not transmit using the same time-frequency resource. However, a receiver may experience improved decoding performance through diversity gain if it receives the same information from multiple sources. 802.16j standardizes the cooperative relaying technique for downlink transmission. Either an MR-BS and one or more RSs or multiple RSs can transmit the same signal and/or space-time-code encoded signals for the same data to a subordinate subscriber station using the same time-frequency resource in cooperative manner to achieve diversity gain. Cooperative relaying can be seen as a distributed MIMO system in multihop environments. It requires for appropriate MAC scheduling of the transmissions from the MR-BS and multiple RSs, and the data needs to be sent to the cooperative RSs before the cooperative transmission can occur.

2.6.4 IEEE 802.15.5 WPAN Mesh

IEEE 802.15 TG5 has defined a recommended practice to provide the framework that enables WPAN devices for interoperable and scalable wireless mesh networking. The standard consists of two parts: low-rate (LR) WPAN mesh and high-rate (HR) WPAN mesh networks. The low-rate mesh is built on IEEE 802.15.4 MAC, whereas the high-rate mesh utilizes IEEE 802.15.3 MAC.

In LR WPAN mesh, a mesh coordinator (MC) can start a mesh network by scanning all the channels to gather the information from existing networks, deciding the channel and PAN ID, and sending beacons. To join a mesh, a device simply discovers existing channels and networks, and selects a channel, network, and parent device to associate. A logic tree rooted at the MC is first formed for both addressing and routing purposes. The logical address of a device is assigned based on its level on the tree and the number of its children. By binding logical addresses to the network topology, routing can be carried out easily without going through route discovery. After a device is assigned an address block for it and its children, it should broadcast several hello messages to its neighbors, and the number of hops that the hello message will propagate is carried in the time-to-live (TTL) field. By exchanging hello messages, mesh links are established.

Similar to an LR mesh network, an HR mesh network usually gets started by a device that is capable of operating as a mesh coordinator. A device first scans to gather information about the existing networks in its neighborhood. If there is at least one mesh network found from the scan process, the device may join one of them. Otherwise, the device determines to operate as an MC. The MC then determines the mesh operation parameters such as mesh ID, tree IDs, operating channel, and so on. It starts a piconet and sends beacons containing mesh information. A mesh network is constructed on the basis of tree topology with the MC as its root. To join a mesh, a device searches for existing channels/piconets and then selects one of the discovered mesh piconet controllers (MPNCs) to associate, request a block of TREEIDs, and create a new child piconet. When constructing a tree topology, the unique TREEID block for a MPNC is assigned from its parent MPNC in a top-down manner starting from the MC. The TREEIDs are conveyed in the beacons and enable the tree-based routing.

The HR WPAN mesh also supports an alternative routing method based on the optional topology servers to provide the optimal route between two MPNCs. With the existence of topology servers, MPNCs may consult these servers for routing information instead of forwarding the packets on the tree-based route. Every MPNC in the tree network can be allowed to play the role of a topology server. A MPNC initiates the link state registration process by broadcasting a link state request command to its descendants. When an MPNC in a tree receives a link state request command from its parent, the MPNC sends a link state registration command to its parent. Then, the MPNC forwards the link state request command to its children. Based on the received link state information, an MPNC calculates the optimal route between any source-destination pair on the subtree consisting of its descendent MPNCs. When a source MPNC wants to deliver a frame by using server routing, it can seek the help from one of its ancestors to locate the optimal route toward the destination using the route discovery command. To provide the optimal route between any source-destination pair, the common ancestor that is closest to both of them calculates the shortest path and sends the calculated explicit path to the destination MPNC. The destination MPNC sends a route formation command toward the source MPNC to update/establish the routing table entries of the relay MPNCs along the derived optimal route.

2.7 Concluding Remarks

Advanced physical and MAC layer techniques such as MU-MIMO, OFDMA, and SDMA have emerged for increasing network capacity and improving bandwidth efficiency and coverage range. These wireless technologies are viewed as the key components in improving the performance of next-generation wireless networks. However, to obtain the full benefits of these technologies, the higher-layer networking protocols should exploit their capabilities in a systematic way due to the interdependence. The standardization effort to achieve overall system optimization is important.

Wireless access and mobility will be an integrated part of the future Internet. New architecture and protocols for the future Internet will enhance network performance in terms of QoS and security, improve cost efficiency, meet increasing user demand, and facilitate various services/applications and the fixed-mobile convergence with seamless mobility and global roaming capability. The research advance in this area certainly has an impact to the standardization effort.

References

[1] IEEE 802.11 Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, June 2007.
[2] IEEE 802.11n Wireless LAN MAC and PHY Specifications – Amendment 5: Enhancements for Higher Throughput, October 2009.d
[3] IEEE 802.11ac PAR for Wireless LAN MAC and PHY Specifications – Amendment: Enhancements for Very High Throughput for Operation in Bands below 6GHz. September 2008.
[4] IEEE 802.11ad PAR for Wireless LAN MAC and PHY Specifications – Amendment: Enhancements for Very High Throughput in the 60 GHz Band, December 2008.
[5] IEEE 802.11s Wireless LAN MAC and PHY Specifications – Amendment 10: Amendment: Mesh Networking, Draft 3.0, May 2009.
[6] IEEE 802.11p Wireless LAN MAC and PHY Specifications – Amendment 7: Wireless Access in Vehicular Environments, Draft 8.0, July 2009.
[7] IEEE 802.11af PAR for Wireless LAN MAC and PHY Specifications – Amendment: TV White Spaces Operation, December 2009.

[8] IEEE 802.16j Air Interface for Broadband Wireless Access Systems – Amendment 1: Multihop Relay Specification, June 2009.
[9] IEEE 802.16m Air Interface for Fixed and Mobile Broadband Wireless Access Systems – Amendment: Advanced Air Interface, Draft 3.0, December 2009.
[10] IEEE 802.15.4f PAR for Wireless MAC and PHY Specifications for Low Rate WPANs – Amendment: Active RFID System PHY, December 2008.
[11] IEEE 802.15.4g PAR for Wireless MAC and PHY Specifications for Low Rate WPANs – Amendment: Physical Layer Specifications for Low Data Rate Wireless Smart Metering Utility Networks, December 2008.
[12] IEEE 802.15.6 Wireless MAC and PHY Specifications for WPANs Used in or around a Body, 2009.
[13] IEEE 802.15.7 PAR for PHY and MAC Layer Standard for Short-Range Wireless Optical Communication Using Visible Light, December 2008.
[14] IEEE 802.15.3c Wireless MAC and PHY Specifications for High Rate WPANs: Amendment 2: Millimeter-wave based Alternative Physical Layer Extension, Draft 13.0, July 2009.
[15] IEEE 802.15.5 Mesh Topology Capability in Wireless Personal Area Networks (WPANs), May 2009.
[16] IEEE 802.22 Draft Standard for Wireless Regional Area Networks Part 22: Cognitive Wireless RAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Policies and Procedures for Operation in the TV Bands, Draft 1.1, May 2008.
[17] IEEE 802.21 Media Independent Handover Services, January 2009.
[18] Salkintzis, A. K., M. Hammer, M., Tanaka, I., and Wong, C. 2009. Voice Call Handover Mechanisms in Next-Generation 3GPP Systems. *IEEE Communications Magazine*, **47**(2), Pages 46–56.
[19] Ali, I., Casati, A., Chowdhury, K., Nishida, K., Parsons, E., Schmid, S., and Vaidya, R. 2009. Network-Based Mobility Management in the Evolved 3GPP Core Network. *IEEE Communications Magazine*, **47**(2), Pages 58–66.
[20] Pastor Balbas, J., Rommer, S., and Stenfelt, J. 2009. Policy and Charging Control in Evolved Packet System. *IEEE Communications Magazine*, **47**(2), Pages 68–74.
[21] Ekstrom, H. 2009. QoS Control in the 3GPP Evolved Packet System. *IEEE Communications Magazine*, **47**(2), Pages 76–83.
[22] Sankaran, C. B. 2009. Network Access Security in Next-Generation 3GPP Systems: A Tutorial. *IEEE Communications Magazine*, **47**(2), Pages 84–91.
[23] Irmer, R., Mayer, H., Weber, A., Braun, V., Schmidt, M., Ohm, M., Ahr, N., Zoch, A., Jandura, G., Marsch, P., and Fettweis, G. 2009. Multisite Field Trial for LTE and Advanced Concepts. *IEEE Communications Magazine*, **47**(2), Pages 92–98.
[24] Astely, D., Dahlman, E., Furuskar, A., Jading, Y., Lindstrom, M., and Parkvall, S. 2009. LTE: The Evolution of Mobile Broadband. *IEEE Communications Magazine*, **47**(4), Pages 44–51.
[25] Larmo, A., Lindstrom, M., Meyer, M., Pelletier, G., Torsner, J., and Wiemann, H. 2009. The LTE Link-Layer Design. *IEEE Communications Magazine*, **47**(4), Pages 52–59.
[26] ITU-R M.2133 Requirements, evaluation criteria and submission templates for the development of IMT-Advanced, December 2008.
[27] ITU-R M.2134 Requirements related to technical performance for IMT-Advanced radio interface(s), December 2008.
[28] ITU-R M.2135 Guidelines for evaluation of radio interface technologies for IMT-Advanced, December 2008.
[29] IETF RFC 3561 Ad hoc On-Demand Distance Vector (AODV) Routing, July 2003.

[30] IETF RFC 3626 Optimized Link State Routing Protocol (OLSR), October 2003.
[31] IETF RFC 3684 Topology Dissemination Based on Reverse-Path Forwarding (TBRPF), February 2004.
[32] IETF RFC 4728 Dynamic Source Routing Protocol (DSR) for Mobile Ad Hoc Networks for IPv4, February 2007.