3

BASIC WIRELESS SENSOR TECHNOLOGY

3.1 INTRODUCTION

In this chapter we look at basic sensor node systems technology at several levels. First, we focus on the sensor node technology itself (Section 3.2), providing a survey of sensor technology, including a taxonomy that classifies devices in families, such as large sensors (e.g., radar sensors), microsensors (tiny sensors), nanosensors, tag-reading sensors, and other sensors (Section 3.3). As already noted, WSNs are characterized by the fact that they need to operate in resource-constrained environments; in turn, this fact imposes strict design guidelines and limitations on the WNs; to this end, we address sensor functionality and components, including the sensing and actuation unit, processing unit, communication unit, power unit, and other application-dependent units. Second, we look at fundamental networking and topological issues (Section 3.4). Building on the introduction provided herein, these issues are revisited in more detail in subsequent chapters. Finally, we look at some current research trends in sensor technology (Section 3.5).

The terms *sensor node*, *wireless node* (WN), *Smart Dust*, *mote*, and *COTS* (commercial off-the-shelf) *mote* are used somewhat interchangeably in the industry; the most general terms used here are *sensor node* and *WN*.

Wireless Sensor Networks: Technology, Protocols, and Applications, by Kazem Sohraby, Daniel Minoli, and Taieb Znati

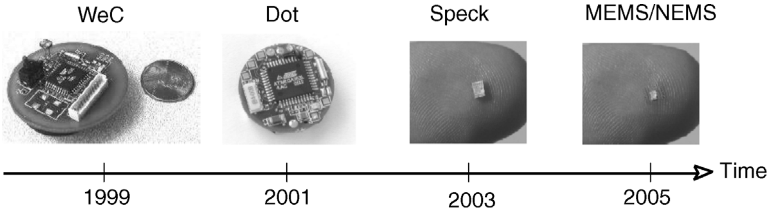


Figure 3.1 Progression of sensor technology (motes) over time (partial sample). (WeC and Dot motes: from Seth Hollar, Kris Pister, and James McClurkin, UC–Berkeley; speck motes: from SpeckNet Consortium/Scottish Higher Education Funding Council; MEMS/NEMS: authors' synthesis.)

3.2 SENSOR NODE TECHNOLOGY

3.2.1 Overview

As we saw in earlier chapters, a WSN consists of a group of dispersed sensors (motes) that have the responsibility of covering a geographic area (the sensor field) in terms of some measured parameter (also known as the measurand); alternatively, a sensor supports a point-to-point link in which the "reader" end is attached to a wireline network (e.g., a stationary tag reader sensing a mobile tag). Sensor nodes have wireless communication capabilities and some logic for signal processing, topology management (if and where applicable), and transmission handling (including digital encoding and possibly encryption and/or forward error correction). Figure 3.1 depicts the progression of sensor technology over time during the past few years. WSNs that combine physical sensing of parameters such as temperature, light, or seismic events with computation and networking capabilities are expected to become ubiquitous in the future [3.3]. Successful development of low-cost robust miniaturized sensors and detection equipment (such as mass spectrometers and chromatographs) will be of benefit; design of such systems is now being encouraged by U.S. research agencies (e.g., the National Science Foundation) [3.5]. Some sensor applications also support e-money purchases at point-of-sale locations such as from soft-drink machines, kiosks, gas stations, and checkout counters.

At the design level a WSN sits at the confluence of research in disciplines such as database query processing, networking, algorithms, and distributed systems [3.3]; hence, a lot of thought and engineering go into the development of both WNs and WSNs. The basic functionality of a WN generally depends on the application, but the following requirements are typical [3.4]:

1. Determine the value of a parameter at a given location. For example, in an environment-oriented WSN, one might need to know the temperature, atmospheric pressure, amount of sunlight, and the relative humidity at a number of locations. This example shows that a given WN may be connected to different types of sensors, each with a different sampling rate and range of allowed values.

2. Detect the occurrence of events of interest and estimate the parameters of the events. For example, in a traffic-oriented WSN, one would like to detect a vehicle moving through an intersection and estimate the speed and direction of the vehicle.
3. Classify an object that has been detected. For example, is a vehicle in a traffic sensor network a car, a minivan, a light truck, a bus?
4. Track an object. For example, in a military WSN, track an enemy tank as it moves through the geographic area covered by the network.

Naturally, the data collected must be transmitted to the appropriate data-consumption entity in a timely fashion. In many cases there are real-time or near-real-time requirements; for example, the detection of an intruder should be communicated to the police in real time so that relevant action can be taken promptly.

As noted in Chapter 1, sensors are either passive or active devices. Passive sensors in single-element form include, among others, seismic-, acoustic-, strain-, humidity-, and temperature-measuring devices. Passive sensors in array form include optical- (visible, infrared 1 μm, infrared 10 μm) and biochemical-measuring devices. Arrays are geometrically regular clusters of WNs (e.g., following some topographical grid arrangement). Passive sensors tend to be low-energy devices. Active sensors include radar and sonar; these tend to be high-energy systems.

Sensing principles include, but are not limited to, mechanical, chemical, thermal, electrical, chromatographic, magnetic, biological, fluidic, optical, ultrasonic, and mass sensing. WNs may be exposed to hostile environments; the environment may include high temperatures, high vibration or noise levels, or corrosive chemicals. WNs may be incorporated in mobile robotic systems; they could also be integral to manufacturing systems. As discussed in Chapter 1, *embedded sensing* refers to the synergistic incorporation of microsensors in structures or environments; embedded sensing enables spatially and temporally dense monitoring of the system under consideration (e.g., an environment, a building, a battlefield). In biological systems, the sensors themselves must not affect the system or organism adversely [3.5]. The technology for sensing and control includes electric and magnetic field sensors; radio-wave frequency sensors; optical-, electrooptic-, and infrared sensors; radars; lasers; location and navigation sensors; seismic and pressure-wave sensors; environmental parameter sensors (e.g., wind, humidity, heat); and biochemical national security–oriented sensors. Typical sensor parameters (measurands) include:

- *Physical measurement.* Examples include two-axis magnetometers; light and ultraviolet intensity (photo resistor); radiation levels, radio, and microwave; humidity, temperature (thermistor), atmospheric pressure, fog, and dust; sound and acoustics; two-axis accelerometers, shock wave, seismic, physical pressure, and motion; video and image (visible or infrared); and location (GPS) and locomotion measurements.
- *Chemical and biological measurements.* Examples include the presence or concentration of a substance or agent at specified concentration levels (there are no less than 50 biological agents of interest [3.9]).

- *Event measurement.* Examples include determination of the occurrence of human-made or natural events, including cyber-level events; tracking of internal and external events.

Small, low-cost, robust, reliable, and sensitive sensors are needed to enable the realization of practical and economical sensor networks. Although a large number measurands are of interest for WSN applications, commercially available sensors exist for many of these measurands; one prominent exception is that a wide range of appropriate chemical sensors is not yet broadly available [3.8].

Sensor nodes come in a variety of hardware configurations: from nodes connected to a LAN and attached to permanent power sources, to nodes communicating via wireless multihop RF radio powered by small batteries [3.3]. The trend is toward very large scale integration (VLSI), integrated optoelectronics, and nanotechnology; in particular, work is under way in earnest in the biochemical arena. The goal of recent research and engineering is to build cubic millimeter (mm^3)–scale advanced WNs and motes. As shown in Figure 3.1, motes developed in the early 2000s were on the order of a cubic inch (this is approximately 16,387 mm^3). By 2007, researchers expect to have 1-mm^3 nodes able to operate in a functional network (e.g., SpeckNet research [3.1]).

3.2.2 Hardware and Software

Related to WN design, the following functionality typically needs to be supported: intrinsic node functionality; signal processing, including digital signal processing (e.g., FFT/DCT), compression, forward error correction, and encryption; control and actuation; clustering and in-network computation; self-assembly; communication; routing and forwarding; and connectivity management. To support this functionality, the hardware components of a WN include the sensing and actuation unit (single element or array), the processing unit, the communication unit, the power unit, and other application-dependent units. Figure 3.2 (which builds on Figure 1.3) shows hardware and software components of a typical sensing node.

As we noted in Chapter 1, the following are important sensor-node issues (refer to Table 1.1): sensor type, sensor power consumption, operating environment, computational and sensing capabilities, signal-processing capabilities, connectivity, and telemetry and control of remote devices. Clearly, the sensor node architecture, scope, and complexity depend on the application. Table 2.1 identified over 200 applications, many of which probably have their own sensor technology.

Sensors, particularly Smart Dust and COTS motes [3.2], have four basic hardware subsystems:

1. *Power.* An appropriate energy infrastructure or supply is necessary to support operation from a few hours to months or years (depending on the application).

2. *Computational logic and storage.* These are used to handle onboard data processing and manipulation, transient and short-term storage, encryption, forward

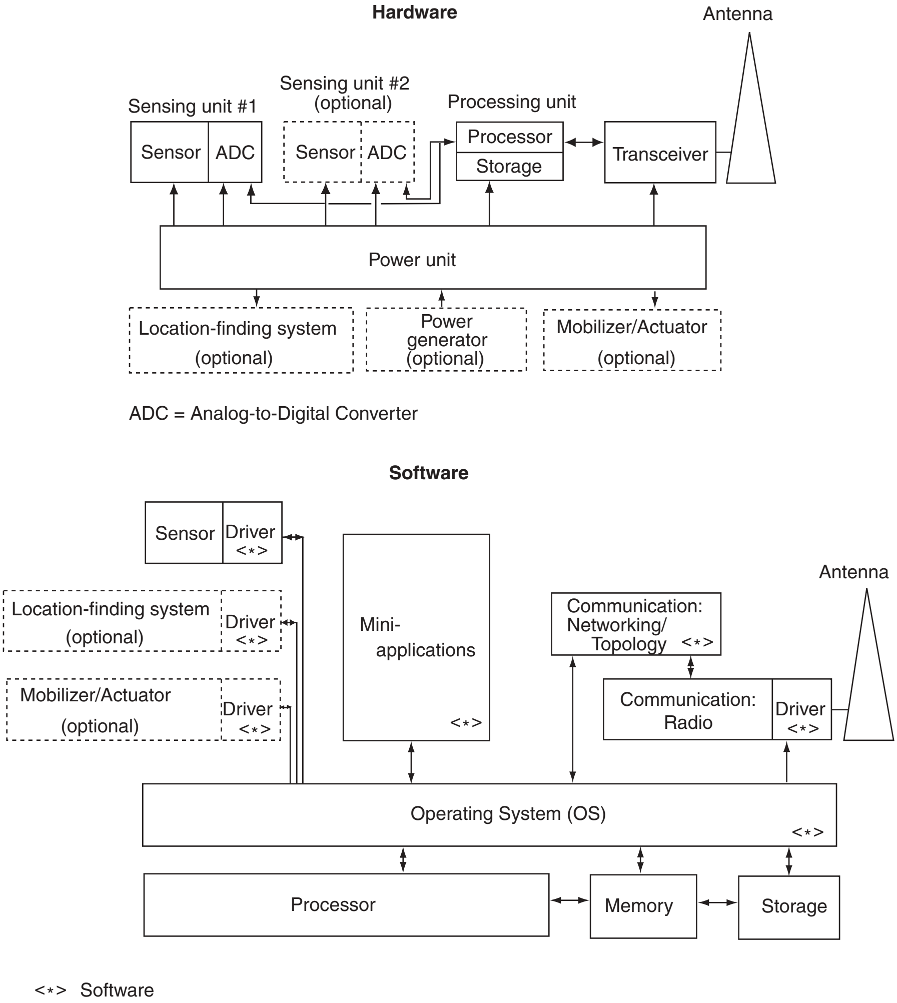


Figure 3.2 Hardware and software components of WNs.

error correction (FEC), digital modulation, and digital transmission. WNs have computational requirements typically ranging from an 8-bit microcontroller to a 64-bit microprocessor. Storage requirements typically range from 0.01 to 100 gigabytes (GB).

3. *Sensor transducer(s).* The interface between the environment and the WN is the sensor. Basic environmental sensors include, but are not limited to, acceleration, humidity, light, magnetic flux, temperature, pressure, and sound.

4. *Communication.* WNs must have the ability to communicate either in C1WSN arrangements (mesh-based systems with multihop radio connectivity among or between WNs, utilizing dynamic routing in both the wireless and wireline portions

of the network), and/or in C2WSN arrangements (point-to-point or multipoint-to-point systems generally with single-hop radio connectivity to WNs, utilizing static routing over the wireless network with only one route from the WNs to the companion terrestrial or wireline forwarding node). Researchers have developed many protocols specifically for WSNs. Transmission range, transmission impairments, modulation techniques, routing, and network topologies are issues of interest. Distances range from a few meters to a few kilometers; lower-layer communication protocols tend to be of the IEEE 802.11/802.15/802.16 class, although other methods have also been used. Throughput ranges from 10 to 256 kbps in most applications (some of the video-based application may require more bandwidth).

Sensors typically have five basic software subsystems:

1. *Operating system (OS) microcode* (also called *middleware*). This is the board-common microcode that is used by all high-level node-resident software modules to support various functions. As is generally the case, the purpose of an operating system is to shield the software from the machine-level functionality of the microprocessor. It is desirable to have *open-source operating systems* designed specifically for *WSNs*; these OSs typically utilize an architecture that enables rapid implementation while minimizing code size. TinyOS is one such example of a commonly used OS.

2. *Sensor drivers.* These are the software modules that manage basic functions of the sensor transceivers; sensors may possibly be of the modular/plug-in type, and depending on the type and sophistication, the appropriate configuration and settings must be uploaded into the sensor (drivers shield the application software from the machine-level functionality of the sensor or other peripheral).

3. *Communication processors.* This code manages the communication functions, including routing, packet buffering and forwarding, topology maintenance, medium access control (e.g., contention mechanisms, direct-sequence spread-spectrum mechanisms), encryption, and FEC, to list a few (e.g., see Figure 3.3).

4. *Communication drivers* (encoding and the physical layer). These software modules manage the minutia of the radio channel transmission link, including clocking and synchronization, signal encoding, bit recovery, bit counting, signal levels, and modulation.

5. *Data processing mini-apps.* These are numerical, data-processing, signal-value storage and manipulations, or other basic applications that are supported at the node level for in-network processing.

3.3 SENSOR TAXONOMY

Because of the variety of sensor types (sensor systems) that exist, a taxonomy is useful. The taxonomy in Table 3.1 is, in effect, an elaboration of Table 1.1. This

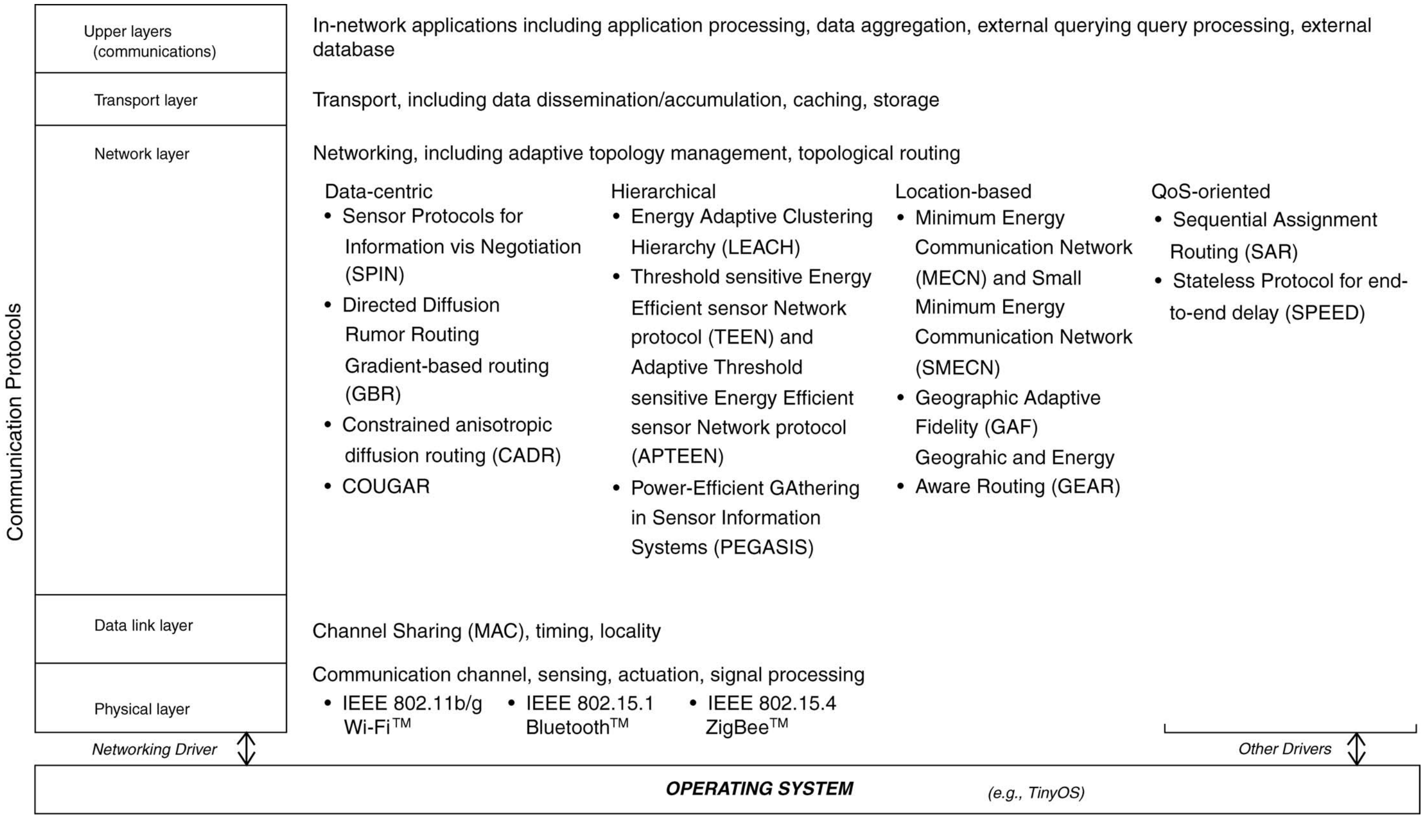

Communication Protocols
Upper layers (communications)
In-network applications including application processing, data aggregation, external querying query processing, external database
Transport layer
Transport, including data dissemination/accumulation, caching, storage
Network layer
Networking, including adaptive topology management, topological routing
Data-centric
• Sensor Protocols for Information vis Negotiation (SPIN)
• Directed Diffusion
Rumor Routing
Gradient-based routing (GBR)
• Constrained anisotropic diffusion routing (CADR)
• COUGAR
Hierarchical
• Energy Adaptive Clustering Hierarchy (LEACH)
• Threshold sensitive Energy Efficient sensor Network protocol (TEEN) and Adaptive Threshold sensitive Energy Efficient sensor Network protocol (APTEEN)
• Power-Efficient GAthering in Sensor Information Systems (PEGASIS)
Location-based
• Minimum Energy Communication Network (MECN) and Small Minimum Energy Communication Network (SMECN)
• Geographic Adaptive Fidelity (GAF)
Geograhic and Energy
• Aware Routing (GEAR)
QoS-oriented
• Sequential Assignment Routing (SAR)
• Stateless Protocol for end-to-end delay (SPEED)
Data link layer
Channel Sharing (MAC), timing, locality
Physical layer
Communication channel, sensing, actuation, signal processing
• IEEE 802.11b/g Wi-Fi™
• IEEE 802.15.1 Bluetooth™
• IEEE 802.15.4 ZigBee™
Networking Driver
Other Drivers
OPERATING SYSTEM
(e.g., TinyOS)

Figure 3.3 Some of the networking protocols supported by WNs.

TABLE 3.1 Basic Taxonomy of Sensor Nodes

Size of Sensor	Mobility of Sensor	Power of Sensor	Computation Logic; Storage Capability of Sensor	Sensor Mode	Communication Apparatus; Lower-Layer Protocols	Communication Apparatus; Upper-Layer Protocols
Very large (10^3 mm^3)	Fully mobile at deployment; fully mobile postdeployment	Self-replenishable, continuous	High-end processor (e.g., 64-bit micro); high-end storage (e.g., 100 GB)	High-end multimodal; physics	Multihop/mesh; hops in 10^1–10^2 m; IEEE MAC	Dynamic routing; data-centric
Large (10^2 mm^3)	Fully mobile at deployment; semimobile postdeployment	Self-replenishable, sporadic	Midrange processor (e.g., 16- or 32-bit micro); high-end storage	High-end multimodal; chemistry–biology	Multihop/mesh; hops in 10^2 to 10^4 m; IEEE MAC	Dynamic routing; hierarchical
Medium (10^1 mm^3)	Fully mobile at deployment; immobile postdeployment	Battery, 10^1 hours	Low-end processor (e.g., 8-bit micro); high-end storage	High-end multimodal; physics–chemistry–biology	Multihop/mesh; hops in 10^4 or more meters; IEEE MAC	Dynamic routing; location-based
Small (10^0 mm^3)	Semimobile at deployment; fully mobile postdeployment	Battery, 10^2 hours	High-end processor (e.g., 64-bit micro); midrange storage (e.g., 1 GB)	Midrange multimodal; physics	Multihop/mesh; hops in 10^1 to 10^2 m; special MAC	Dynamic routing; QOS-based

Very small (10^{-1} mm^3)	Semimobile at deployment; semimobile postdeployment	Battery, 10^3 hours	Midrange processor (e.g., 16- or 32-bit micro); midrange storage	Midrange multimodal; chemistry–biology	Multihop/mesh; hops in 10^2 to 10^4 m; special MAC	Static routing (single hop)
Ultrasmall (10^{-2} mm^3)	Semimobile at deployment; immobile postdeployment	Battery, 10^4 hours	Low-end processor (e.g., 8-bit micro); Midrange storage	Midrange multimodal; physics–chemistry–biology	Multihop/mesh; hops in 10^4 or more meters; special MAC	
Microscale (10^{-3} mm^3)	Immobile at deployment; fully mobile postdeployment	Battery, 10^5 hours	High-end processor (e.g., 64-bit micro); low-end storage (e.g., 0.01 GB)	Single function; physics	Single hop; hops in 10^1 to 10^2 m; IEEE MAC	
Nanoscale ($<10^{-4}$ mm^3)	Immobile at deployment; semimobile postdeployment		Midrange processor (e.g., 16- or 32-bit micro); low-end storage	Single function; chemistry–biology	Single hop; hops in 10^2 to 10^4 m; IEEE MAC	
	Immobile at deployment; immobile postdeployment		Low-end processor (e.g., 8-bit micro); low-end storage	Single function; physics–chemistry–biology	Single hop; hops in 10^4 or more meters; IEEE MAC	
					Single hop; special MAC	

TABLE 3.2 Reduced-Complexity Taxonomy of Sensor Nodes

Size of Sensor, s	Mobility of Sensor, m	Power of Sensor, p	Computation Logic and Storage Capability of Sensor, cp	Sensor Mode, md	Communication Apparatus or Protocols of Sensor, cm
1 Large	1 Mobile	1 Self-replenishable	1 High-end processor and storage	1 Multimodal, physics	1 Multihop/mesh with dynamic routing
2 Small	2 Static	2 Battery, hours–days	2 Midrange processor and storage	2 Multimodal, chemistry/biology	2 Single hop with static routing
3 Microscopic		3 Battery, weeks–months	3 Low-end processor and storage	3 Single function, physics	
4 Nanoscopic		4 Battery, years		4 Single function, chemistry–biology	

taxonomy is somewhat daunting since there are $8 \times 9 \times 7 \times 9 \times 9 \times 10 \times 5 =$ 2,041,200 cases or combinations. However, the classification "buckets" are reasonable, and a large majority of the combinatorial combinations are, in fact, valid. To reduce the scope of the taxonomy, we suggest the use of the modified classification shown in Table 3.2; here one has only $4 \times 2 \times 4 \times 3 \times 4 \times 2 = 768$ cases or combinations. For example, a s(2)m(2)p(3)cp(2)md(1)cm(1) WN is a system that is small, static, battery-powered, has multiple measurands, and supports multihop networking.

3.4 WN OPERATING ENVIRONMENT

As we saw in Chapter 1, networking implies a need to support physical and logical connectivity. In WSNs, physical connectivity is supported over a wireless radio link of one or more hops, at a distance of tens, hundreds, or thousand of meters. Logical connectivity has the goal of supporting topology maintenance and multihop routing (when present). The design and engineering of WNs clearly needs to take into account all the issues described in Section 3.2 as well as in this section.

Sensor nodes have to deal with the following resource constraints [3.3] (see also Table 3.3):

- *Power consumption.* Almost invariably, WNs have a limited supply of operating energy; it follows that energy conservation is a key system design consideration.
- *Communication.* The wireless network usually has limited bandwidth; the networks may be forced to utilize a noisy channel; and the communication channel may be relegated to an unprotected frequency band. The implications

TABLE 3.3 Design Constraints or Requirements for WSNs and WNs

WSN/WN Requirement	Motivation
Collaborative data processing	A factor that distinguishes WSNs from simple ad hoc networks is that the goal in WSNs is detection or estimation of specified events, not just communications. One needs to provide scalable, fault-tolerant, flexible data access and intelligent data reduction [3.3]. This drives the overall architecture because detection and estimation often require fusing data from multiple sensors; data fusion requires the transmission of data and control messages. Quantification of sensor data, including limits of detection, calibration, interferences, sampling, and verification of accuracy, also needs to be taken into account [3.5].
Constrained energy use	In many applications the WNs are deployed in remote areas; in these cases, the lifetime of a node may be determined by the battery life; this in turn requires a minimization of energy consumption.
Large topology support	Networks of 10,000 or even 100,000 nodes are envisioned for some applications. Fortunately, most WSNs/WNs are stationary (aside from the deployment of sensors on the ocean surface or the use of mobile, unmanned, robotic sensors in military operations).
Querying capabilities	A data-consumption entity may need to query an individual node or group of nodes for information collected in the region. Because it may not be feasible to transmit a large amount of the data across a network, various local sink nodes need to collect the data from a given area and create summary messages to reply to the query.
Self-organization	It is typically a requirement that WSNs be able to self-organize: Given the large number of nodes and their potential placement in hostile locations, manual configuration is typically not feasible. Also, nodes may fail (from lack of energy or from physical destruction), and new nodes may join the network: the network must be able to reconfigure itself so that it can continue to operate properly and support reliable connectivity.

Source: Adapted from [3.4].

are limited reliability, poor quality of service (e.g., high latency, high variance, high frame loss), and security exposure (e.g., denial of service, jamming, interference, high bit-error rates).

- *Computation.* WNs typically have limited computing power and memory resources. The implications are restrictions on the types of data-processing algorithms that can run on a sensor node. This also limits the scope and volume of intermediate results that can be stored in the WNs. Research aims at developing a distributed data management layer that scales with the growth of sensor interconnectivity and computational power on the sensors; the goal is to deploy mechanisms that reside directly on the sensor nodes and create the abstraction of a single processing node without centralizing data or computation.

- *Uncertainty in measured parameters.* Signals that have been often have various detected or collected degrees of intrinsic uncertainty. Desired data may be commingled with noise and/or interference from the environment. Node malfunction could collect and/or forward inaccurate data. Node placement (particularly in ad hoc networks without mobility) may impair operation and bias individual readings.

Some of the intrinsic factors that the design constraints or requirements that WSNs and WNs need to take into account include the following:

- WNs may be deployed in a dense manner (close proximity), implying communication complexity (e.g., in support of packet forwarding and topology management)
- For military and/or national security applications, WNs need to support rapid deployment; the deployment must be supportable in an ad hoc fashion; and the environment is expected to be highly dynamic.
- WNs may be prone to failure. Unattended, untethered, self-powered low-duty-cycle systems are typical, yet some WSNs require sensing systems that are long-lived and environmentally resilient.
- As just noted, WNs are limited in power, computational capacity, and memory. Communication circuitry and antennas are the primary elements that use up most of the energy.
- The topology that the WNs need to maintain may change very frequently. Communication links may be expensive (not only from an electromagnetic spectrum perspective, but also in terms of the operational support of the requisite infrastructure); the bandwidth may be limited; and as just noted, the power availability at the sensor may be limited and/or expensive in reference to supporting a high-capacity, high-range link (i.e., to feed a high-power antenna).
- WNs may not have global addresses because of the potentially large number of sensors and overhead needed to support such global addresses (IPv6 could be applicable in this context).
- WNs require special routing and data dissemination mechanisms (e.g., data-centric, hierarchical, and/or location-based routing).
- WNs often require in-network processing, even while the data are being routed. One wants to be able to perform data processing in the network in the proximity of the source of the data, and then forward only summarized,

aggregated, fused, and/or synthesized results. Typical functionality involves signal processing, data aggregation, data fusion, and data analysis. There is also an interest in database management, including querying mechanisms and data storage and warehousing.

- Arrays of ultralow-power wireless nodes may be incorporated in reconfigurable networks with high-speed connectivity to processing centers for decision and responsive action [3.5].

3.5 WN TRENDS

For WSNs to achieve wide-scale deployment, the size, cost, and power consumption of the nodes must decrease considerably and the intelligence of the WNs must increase [3.6]. To meet evolving functional requirements of the various user communities, it will be necessary for sensor systems to leverage and incorporate advances in adjacent technologies, such as nanofabrication, biosystems, massively distributed networks, ubiquitous computing, broadband wireless communications, and information and decision systems [3.5].

Evolving requirements for new WSNs and WNs include, among others: (1) the ability to respond to new toxic chemicals, explosives, and biological agents; (2) enhanced sensitivity, selectivity, speed, robustness, and fewer false alarms; and (3) the ability to function, perhaps autonomously, in unusual, extreme, and complex environments. These needs can be addressed by the design and synthesis of functionalized receptors and materials, resulting in next-generation devices. The materials may be of varying porosity, enabling them to detect single toxic compounds in complex mixtures or physical configurations that have surfaces with microchannels for microfluidic discrimination. Advanced biological, chemical, and materials research can be brought to bear on this challenge, including the design of functional nano- and mesoscale complex structures (e.g., quantum dots, nanowires, gels). Robustness under anticipated manufacturing schemes is also required [3.5].

Miniaturization, manufacturability, and cost are also critical issues. Integration of sensors, processors, energy sources, and the communications network interface on a chip would facilitate the exchange of sensor data and critical information with the outside world. Information extraction may involve detection of events or objects of interest, estimation of key parameters, and human-in-the-loop or closed-loop adaptive feedback [3.5]. Some of the goals (e.g., as defined by the PicoRadio effort at UC–Berkeley [3.7]) are to develop mesoscale low-cost (i.e., <50 cents) transceivers for ubiquitous wireless data acquisition that minimize power or energy dissipation [i.e., minimize energy (<5 nJ/(correct) bit)] for an energy-limited source and minimize power (i.e., $<100\,\mu$W for a power-limited source, enabling energy scavenging) by using the following strategies: self-configuring networks, fluid trade-off between communication and computation, an integrated system-on-a-chip (SOC) approach, and aggressive low-energy architectures and circuits.

Standardization is important. As the definition of sockets has made the use of communication services on the Internet independent of the underlying protocol stack, communication medium, and even operating system, the application interface one needs for WSNs should be an abstraction that is offered to any sensor network application and supported by any sensor network platform [3.7]. Research and engineering activity now under way seeks to advance fundamental knowledge in new sensor technologies, including sensors for toxic chemicals, explosives, and biological agents; sensor networking systems in a distributed environment; the integration of sensors into commercial systems; and the interpretation and use of sensor data in decision-making processes [3.5]. Table 3.4 provides a partial list of near-term research efforts as sponsored by U.S. government agencies.

Of late, one has seen targeted efforts to develop chemical sensors for sensor networks, particularly for monitoring soil contamination and for habitat monitoring. Specifically, one needs an array of miniaturized chemical sensors to monitor the flow of contaminants accurately (e.g., see Figure 3.4). Optimally, one is interested in developing microscale liquid chromatography systems [3.8]. According to published reports, the U.S. Department of Homeland Security (DHS) is coordinating an effort for the end-of-decade deployment of a nationwide sensor network to provide a real-time early-warning system for a plethora of chemical, biological, and nuclear threats across the United States. Planners at DHS are working on developing capabilities to deal with multifaceted threats targeted at airports, subways, and buildings; they are also looking at issues related to water sources, animal herds, and flocks of birds that could spread contaminants or harmful biological agents. This type of technology is currently under development [3.9].

National research laboratories have been working on core issues in materials, sensors, networks, and electronics, and have already established field trials of prototype networks. The multifaceted nature of the global threat has led researchers to consider a system that consists of a suite of different types of sensors. Researchers are planning to use MEMSs and nanotechnology for low-cost, high-reliability, and high-accuracy biological and chemical sensors. In one approach, researchers are studying hybrid sensors that use surface-chemical detection as a first trigger, which could then use technology on the same device for more time-consuming techniques, such as DNA testing. Other researchers are studying the use of infrared or ultraviolet spectrum analysis as well as biometric sensors that mimic human cells to create test reactions. Further into the future, MEMS technology is seen as having promise for creating miniature benchtop labs on a chip. Sensors could use polymer- or gel-coated silicon devices to trap targeted chemicals, then send the agents through fluidic channels to on-chip arrays of surface-acoustic-wave detectors. A follow-on device would integrate the fluidics, surface acoustic waves, and support electronics on a single device [3.9]. Other research teams are exploring nanotechnology to deliver new sensor materials (e.g., researchers at the Pacific Northwest National Laboratory have

TABLE 3.4 Partial List of Near-Term Research Efforts as Sponsored by U.S. Government Agencies

Designs, materials, and concepts for new sensors and sensing systems	Examples include novel sensing materials and devices; the design of solid and liquid surfaces with molecular recognition, long lifetime, and regenerability of the sensing site; biomimetic sensors, including hybrids consisting of proteins, enzyme fragments and components, bioorganometallics, or other biocatalysts that can be linked to surfaces; bioMEMS; sensors for toxic agents (biological, chemical, radiation); sensors for operation in harsh environments; wireless sensors; chip-based systems incorporating multiple sensors, computation, actuation, and wireless interfaces; sensor systems capable of remote activation and interrogation; sensor power sources; novel optical imaging concepts; novel techniques for metrology at the nanoscale; new modeling and simulation tools; new techniques for on-sensor self-calibration and self-test; enhanced specificity to maximize accuracy and minimize false alarms; and new methods for sensor fabrication, manufacture, and encapsulation.
Arrayed sensor networks and networking	This area includes: Enabling networking technologies for distributed wireless and wired sensor networks Scalable and robust architectures Design Automated tasking Querying techniques Adaptive management and control of sensor nodes Design trade-offs and performance optimization in resource-constrained sensor networks Design of ultralow-power processing nodes for local information management Investigation of localized versus distributed versus centralized processing of sensor data Common building blocks and interfaces for sensor networking Strategies for using heterogeneous sensor and network nodes to enhance performance and reduce false alarms Security and authentication for resource-constrained sensor networks Embedded and hybrid systems Application-specific network and system services, including data-centric routing, attribute-based addressing, location management, and service discovery Energy-efficient media access, error control, and traffic management protocols

(*Continued*)

TABLE 3.4 (*Continued*)

	Mobile sensor networks Scalable reconfigurability and self-organization
Interpretation, decision, and action based on sensor data	Examples include decision theory for intelligent use of sensed information; detection and identification of false alarms; feedback theory; development of new statistical algorithms, sampling theories, and supervisory control systems tailored to needs; concepts for optimal sensor locations for effective process and system control; mathematical hybrid system tools for monitoring distributed networks of large arrays of sensors and actuators; handheld diagnostic kits; and pattern recognition and state estimation. System-level sensor applications include biomedical health monitoring, diagnostic, and therapeutic systems; image-guided surgery; health monitoring systems for civil structures; crisis management sensor systems; surveillance technology; robotics; mobile sensors; tracking and monitoring of mobile units (endangered species, inventory control, transportation); and sensor assessment (reliability, verification, validation).

Source: National Science Foundation materials [3.5].

developed nanosized preconcentrators for nerve agents, botulism, and other toxins) [3.9].

Sandia has been testing handheld sensors designed to detect chemical-weapons agents on the battlefield with high sensitivity; the detection window is 2 minutes or less. The lab has been asked to explore adding networking and GPS capability

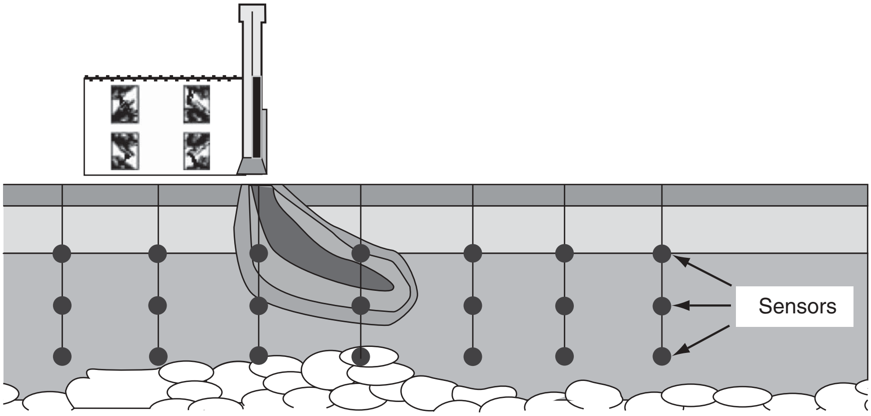


Figure 3.4 Sensor array for chemical contamination analysis.

to those sensors so that they could be mounted on military vehicles, creating a mobile battlefield sensor network. The expectation is that by the turn of the present decade, a bio smoke alarm detector will be ready for commercial deployment [3.9].

On the networking front, researchers are considering peer-to-peer network with multilevel security and quality-of-service guarantees, spanning terrestrial wireless, wireline, and satellite links. The underlying network architecture for a national sensor network has been studied at Oak Ridge National Labs. The aim is to use off-the-shelf technology as much as possible and to leverage existing infrastructure, such as the 30,000 cellular towers and 100,000 cellular base stations in the United States today. However, developing quality-of-service guarantees and multilevel security for a hybrid wired, wireless, and satellite network is a challenge [3.9]. Several pilot sensor network projects are being field tested, including systems developed by Los Alamos and UC–Berkeley researchers to safeguard crops. Trial sensor networks are also in place in Boston subways, at the San Francisco airport, and on the Miami docks. The Washington subway recently went operational with a chemical-sensor system developed by Sandia and Argonne National Laboratories in Chicago [3.9].

3.6 CONCLUSION

In this chapter we looked at basic sensor node technology along with a taxonomy of sensor types. Some current trends were also discussed.

REFERENCES

[3.1] D. Roman, "Scottish Universities Plan Speckled Computing Net," *EE Times*, Oct. 27, 2003.

[3.2] *The Scientist and Engineer's Guide to TinyOS Programming*, University of California–Berkeley, http://tinyos.org. This book was developed as an open source, freely available manuscript on the TinyOS Documentation Project.

[3.3] "Cougar: The Sensor Network Is the Database," Cornell University, Ithaca, NY, http://www.cs.cornell.edu/database/cougar/.

[3.4] "Smart Sensor Networks," National Institute of Standards and Technology, Gaithersburg, MD, http://w3.antd.nist.gov/wahn_ssn.shtml.

[3.5] Sensors and Sensor Networks, Program Solicitation, NSF 03-512, Mar. 6, 2003, National Science Foundation, Directorate for Engineering, http://www.nsf.gov/cgi-bin/getpub?gpg; also, NSF Publications Clearinghouse, pubs@nsf.gov.

[3.6] J. M. Rabaey, M. J. Ammer, J. L. da Silva Jr., D. Patel, S. Roundy, "PicoRadio Supports Ad Hoc Ultra-low Power Wireless Networking," *Computer*, July 2000; wireless sensor network research at the Berkeley Wireless Research Center, http://bwrc.eecs.berkeley.edu/Research/Pico_Radio/.

4

WIRELESS TRANSMISSION TECHNOLOGY AND SYSTEMS

4.1 INTRODUCTION

In this chapter we look at radio-channel-related issues. It should immediately be noted that to maximize the opportunity for widespread and cost-effective deployment of WSN, one needs to make use of existing and/or emerging commercial off-the-shelf (COTS) wireless communications and infrastructures rather than having to develop an entirely new, specially designed apparatus. WSNs can use a number of wireless COTS technologies, such as Bluetooth/Personal Area Networks (PANs), ZigBee, wireless LANs (WLAN)/hotspots, broadband wireless access (BWA)/ WiMax, and 3G.

Given this pragmatic perspective, we focus here less on the science of radio transmission per se as a discrete system component and more on an integrated system-level view of the field. In other words, we explore the use of the just-named technologies as a plug-and-play system integration opportunity more than looking at the fundamentals of modulation, transmission, encoding, radio impairments, and so on. Stated differently, the developer of WSN systems should not be required to have a deep understanding of radio science (beyond basic issues such as power, range and coverage, bandwidth, performance, security, and a few other factors), but rather, which off-the-shelf wireless systems already defined by various standards bodies (e.g., Bluetooth, Wi-Fi, WiMax, ZigBee/IEEE 802.15.4) can be used by way of employing and/or integrating preconfigured chipsets and ICs (integrated circuits), antennas, drivers, and protocol machinery.

Wireless Sensor Networks: Technology, Protocols, and Applications, by Kazem Sohraby, Daniel Minoli, and Taieb Znati

TABLE 4.5 Wireless Protocol Comparison

	IEEE Standard		
Property	802.11	802.15.1/Bluetooth	802.15.4/ZigBee
Range (m)	~100	~10 to 100	~10
Data throughput (Mbps)	~2 to 54	~1 to 3	~0.25
Power consumption	Medium	Low	Ultralow
Battery life measured in:	Minutes to hours	Hours to days	Days to years
Size relationship	Large	Smaller	Smallest
Cost/complexity ratio	>6	1	0.2

Source: [4.4].

The IEEE 802.15.4 standard supports a maximum data rate of 250 kbps, with rates as low as 20 kbps (slower than most telephone modems); however, it has the lowest power requirement of the group. ZigBee devices are designed to run several years on a single set of batteries, making them ideal candidates for unattended or difficult-to-reach locations. Bluetooth is a short-range communication protocol widely used in cellular-type phones and PDAs (has a range of about 10 m, or a maximum of 100 m with power boost); it operates in the 2.4-GHz ISM band and has a bandwidth of approximately 1 to 3 Mbps. IEEE 802.11a/b/g/n is a collection of related technologies that operate in the 2.4-GHz ISM band, the 5-GHz ISM band, and the 5-GHz U-NII bands; it provides the highest power and longest range of the common unlicensed wireless technologies. Transmission data rates can reach 54 Mbps (twice as much with the latest IEEE 802.11n protocol). Typically, hardware implementation of some or all of 802.11 protocols comes preinstalled on most new laptop computers; the technology is often also available for PDAs and cellular phones. RFID is the one form of wireless sensing that requires no power in the tag; it is a passive technology used for labeling and tracking. The RFID tag is the sensor; the sensor responds when power is beamed to it through the reading device. Current RFID tags can hold only 96 bits of information, but newer tags that support 128 and 256 bits are becoming available [4.4]. Most RFID tags have *integrated circuits* (ICs), microelectronic semiconductor devices with a large number of interconnected transistors and other components. Although the topic of RFIDs is not covered further in this book, a glossary of basic terms is included in Table 4.6 for completeness [4.41].

The subsections that follow provide additional details on these standardized wireless technologies. We partition the discussion into campus and MAN/WAN application spaces.

4.3.1 Campus Applications

Campus sensor communications can occur over Bluetooth, wireless LAN (WLAN), ZigBee, or WiMax/hotspot systems.

TABLE 4.6 RFID Glossary[a]

Active tag	An RFID tag that comes with a battery that is used to power the microchip's circuitry and transmit a signal to a reader. Active tags can be read from 100 ft or more away, but they are expensive—more than $20 each. Tags are used for tracking expensive items over long ranges. For instance, the U.S. military uses active tags to track containers of supplies arriving in ports.
Automatic identification	(a.k.a. automatic data capture) A method of collecting data and entering them directly into computer systems without human involvement. Technologies normally considered part of auto-ID include bar codes, biometrics, RFID, and voice recognition.
Backscatter	A method of communication between tags and readers. RFID tags using backscatter technology reflect back to the reader a portion of the radio waves that reach them. The signal reflected is modulated to transmit data. Tags using backscatter technology can be either passive or active, but either way, they are more expensive than tags that use inductive coupling.
Chipless RFID tag	An RFID tag that does not depend on an integrated microchip. Instead, the tag uses materials that reflect back a portion of the radio waves beamed at them. A computer takes a snapshot of the waves beamed back and uses it like a fingerprint to identify the object with the tag. Companies are experimenting with embedding RF reflecting fibers in paper to prevent unauthorized photocopying of certain documents. But chipless tags are not useful in the supply chain because even though they are inexpensive, they cannot communicate a unique serial number that can be stored in a database.
Closed-loop systems	RFID tracking systems set up within a company. Since the item being tracked never leaves the company's control, the company does not need to worry about using technology based on open standards.
Contactless smart card	A credit card or loyalty card that contains an RFID chip to transmit information to a reader without having to be swiped through a reader. Such cards can speed checkout, providing consumers with more convenience.
EEPROM (electrically erasable programmable read-only memory)	A nonvolatile storage device on microchips. Usually, bytes can be erased and reprogrammed individually. RFID tags that use EEPROM are more expensive than factory-programmed tags, but they offer more flexibility because the end user can write an ID number to the tag at the time the tag is going to be used.
Electromagnetic compatibility (EMC)	The ability of a system or product to function properly in an environment where other electromagnetic devices are used and not itself be a source of electromagnetic interference.
Electromagnetic interference (EMI)	Interference caused when the radio waves of one device distort the waves of another. Cells phones, wireless computers, and even robots in factories can produce radio waves that interfere with RFID tags.

TABLE 4.6 (*Continued*)

Electronic article surveillance (EAS)	Simple electronic tags that can be turned on or off. When an item is purchased (or borrowed from a library), the tag is turned off. When someone passes a gate area holding an item with a tag that has not been turned off, an alarm sounds. EAS tags are embedded in the packaging of most pharmaceuticals.
Electronic product code (EPC)	A 96-bit code created by the auto-ID center that will one day replace bar codes. The EPC has digits to identify the manufacturer, product category, and the individual item. It is backed by the Uniform Code Council and the European Article Numbering Association the two main bodies that oversee bar code standards.
Error-correcting code	A code stored on an RFID tag to enable a reader to determine the value of missing or garbled bits of data. It is needed because a reader might misinterpret some data from the tag and think that a Rolex watch is actually a pair of socks.
Error-correcting mode	A mode of data transmission between the tag and the reader in which errors or missing data are corrected automatically.
Error-correcting protocol	A set of rules used by readers to interpret data correctly from the tag.
Excite	A reader is said to "excite" a passive tag when the reader transmits RF energy to wake up the tag and enable it to transmit back.
Factory programming	The process of writing the identification number into a silicon microchip at the time the chip is made, as is necessary for some read-only tags.
Field programming	Tags that use EEPROM, or nonvolatile memory, can be programmed after being shipped from the factory.
GTAG (global tag)	A standardization initiative of the Uniform Code Council and the European Article Numbering Association for asset tracking and logistics based on RFID. The GTAG initiative is supported by Philips Semiconductors, Intermec, and Gemplus, three major RFID tag makers.
High-frequency tags	Tags that operate typically at 13.56 MHz. They can be read from about 10 ft away and transmit data faster, but they consume more power than do low-frequency tags.
Inductive coupling	A method of transmitting data between tags and readers in which the antenna from the reader picks up changes in a tag's antenna.
Low-frequency tags	Tags that typically operate at 125 kHz. The main disadvantages of low-frequency tags are that they have to be read from within 3 ft and the rate of data transfer is slow. But they are less expensive than high-frequency tags and less subject to interference.
Memory	The amount of data that can be stored on a tag.
Microwave tags	Radio-frequency tags that operate at 5.8 GHz. They have very high transfer rates and can be read from away as far as 30 ft, but they use a lot of power and are expensive.
Multiple-access schemes	Methods of increasing the amount of data that can be transmitted wirelessly within the same frequency spectrum. RFID readers use time-division multiple access (TDMA), meaning that they read tags at different times to avoid interfering with one another.

(*Continued*)

TABLE 4.6 (*Continued)*

Nominal range	The read range at which a tag can be read reliably.
Null spot	Area in the reader field that does not receive radio waves. This is essentially the reader's blind spot. It is a phenomenon common to ultrahigh-frequency systems.
Object name service (ONS)	An auto-ID center–designed system for looking up unique electronic product codes and pointing computers to information about the item associated with the code. ONS is similar to the domain name service, which points computers to sites on the Internet.
Passive tag	An RFID tag without a battery. When radio waves from the reader reach the chip's antenna, it creates a magnetic field. The tag draws power from the field and is able to send back information stored on the chip. At this juncture simple passive tags cost from about 50 cents to several dollars.
Patch antenna	A small square antenna made from a solid piece of metal or foil.
Power level	The amount of RF energy radiated from a reader or an active tag. The higher the power output, the longer the read range, but most governments regulate power levels to avoid interference with other devices.
Programming	Writing data to an RFID tag.
Proximity sensor	A device that detects the presence of an object and signals another device. Proximity sensors are often used on manufacturing lines to alert robots or routing devices on a conveyor to the presence of an object.
Read	The process of turning radio waves from a tag into bits of information that can be used by computer systems.
Read range	The distance from which a reader can communicate with a tag. Active tags have a longer read range than passive tags because they use a battery to transmit signals to the reader. With passive tags, the read range is influenced by frequency, reader output power, antenna design, and method of powering up the tag. Low-frequency tags use inductive coupling (see above), which requires the tag to be within a few feet of the reader.
Read rate	The maximum rate at which data can be read from a tag, expressed in bits or bytes per second.
Reader (also called an interrogator)	The reader communicates with an RFID tag via radio waves and passes the information in digital form to a computer system.
Reader field	The area of coverage. Tags outside the reader field do not receive radio waves and cannot be read.
Read-only tag	A tag that contains data that cannot be changed unless the microchip is reprogrammed electronically.
Read–write tag	An RFID tag that can store new information on its microchip. San Francisco International Airport uses a read–write tag for security. When a bag is scanned for explosives, the information on the tag is changed to indicate that it has been checked. The tag is scanned again before it is loaded on a plane. Read–write tags are more expensive than read-only tags and therefore are of limited use for supply chain tracking.

TABLE 4.6 (*Continued*)

RFID tag	A microchip attached to an antenna that picks up signals from and sends signals to a reader. The tag contains a unique serial number but may have other information, such as a customer's account number. Tags come in many forms, such as smart labels that are stuck on boxes, smart cards and keychain wands for paying for things, and a box that you stick on your windshield to enable you to pay tolls without stopping. RFID tags can be active tags, passive tags, or semipassive tags.
RFID tags' frequency	RFID tags use low, high, ultrahigh, and microwave frequencies. Each frequency has advantages and disadvantages that make them more suitable for some applications than for others.
Scanner	An electronic device that can send and receive radio waves. When combined with a digital signal processor that turns the waves into bits of information, the scanner is called a reader or interrogator.
Semipassive tag	Similar to active tags, but the battery is used to run the microchip's circuitry but not to communicate with the reader. Some semipassive tags sleep until they are woken up by a signal from the reader, which conserves battery life. Semipassive tags cost $1 or more.
Sensor	A device that responds to a physical stimulus and produces an electronic signal. Sensors are increasingly being combined with RFID tags to detect the presence of a stimulus at an identifiable location.
Silent commerce	This term covers all business solutions enabled by tagging, tracking, sensing, and other technologies, including RFID, which make everyday objects intelligent and interactive. When combined with continuous and pervasive Internet connectivity, they form a new infrastructure that enables companies to collect data and deliver services without human interaction.
Smart label	A label that contains an RFID tag. It is considered "smart" because it can store information, such as a unique serial number, and communicate with a reader.
Tag antenna	The antenna is the conductive element that enables the tag to send and receive data. Passive tags usually have a coiled antenna that couples with the coiled antenna of the reader to form a magnetic field. The tag draws power from this field.
Time-division multiple access (TDMA)	A method of solving the problem of the signals of two readers colliding. Algorithms are used to make sure that readers attempt to read tags at different times.
Transponder	A radio transmitter–receiver that is activated when it receives a predetermined signal. RFID tags are sometimes referred to as transponders.
Ultrahigh frequency (UHF) tag	Typically, tags that operate between 866 and 930 MHz. They can send information faster and farther than can high- and low-frequency tags. UHF tags are also more expensive than low-frequency tags, and they use more power.

(*Continued*)

TABLE 4.6 (*Continued)*

Uniform Code Council (UCC)	The nonprofit organization that overseas the Uniform Product Code, the bar code standard used in North America.
Uniform Product Code (UPC)	The bar code standard used in North America. It is administered by the Uniform Code Council.
Write rate	The rate at which information is transferred to a tag, written into the tag's memory and verified as being correct.

Source: [4.41].

[a] RFID is a method of identifying unique items using radio waves. Typically, a reader communicates with a tag, which holds digital information in a microchip; however, there are chipless forms of RFID tags that use material to reflect back a portion of the radio waves beamed at them.

Bluetooth Bluetooth is a specification for short-range RF-based connectivity for portable personal devices. It is a short-range wireless data exchange protocol designed for a small variety of tasks, such as synchronization, voice headsets, cell modem calls, and mouse and keyboard input. The specification began as a de facto industry standard; more recently, IEEE Project 802.15.1 developed a wireless PAN standard based on the Bluetooth v1.1 Foundation Specifications. The IEEE 802.15.1 standard was published in 2002. Bluetooth is directed principally to the support of personal communication devices such as telephones, printers, headsets, and PC keyboards and mice. The technology has restricted performance characteristics by design; hence, its applicability to WSN is rather limited in most cases. For these same environments, ZigBee is probably a better solution; however, given the popularity and longevity of the standard, it is given some coverage here.

As part of its effort, the IEEE has reviewed and provided a standard adaptation of the Bluetooth Specification v1.1 Foundation media access control (MAC) (L2CAP, LMP, and baseband) and the physical layer (PHY) (radio). Also specified is a clause on service access points (SAPs), which includes a LLC–MAC interface for the ISO/IEC 8802-2 LLC. A normative annex that provides a protocol implementation conformance statement (PICS) proforma has been developed. Also specified is an informative high-level behavioral ITU-T Z.100 specification and description language (SDL) model for an integrated Bluetooth MAC sublayer [4.6].

The Bluetooth specification defines a low-power, low-cost technology that provides a standardized platform for eliminating cables between mobile devices and facilitating connections between products. The system uses omnidirectional radio waves that can transmit through walls and other nonmetal barriers. Unlike other wireless standards, the Bluetooth wireless specification includes both link layer and application layer definitions for product developers. Radios that comply with the Bluetooth wireless specification operate in the unlicensed, 2.4-GHz ISM radio spectrum, ensuring communication compatibility worldwide.

Bluetooth radios use a spread-spectrum, frequency-hopping, full-duplex signal. While point-to-point connections are supported, the specification allows up to seven simultaneous connections to be established and maintained by a single radio [4.7]. AFH (adaptive frequency hopping), available with newer versions, allows for more graceful coexistence with IEEE 802.11 WLAN systems. The signal hops among 79 frequencies at 1-MHz intervals to give an acceptable degree of interference immunity between multiple Bluetooth devices and between a Bluetooth device and a WLAN device (at least in the case where not all the available frequencies are used by the WLAN—this is probably the case in a SOHO environment, where only one or two access points are used at a location). To minimize interference with other protocols that use the same band, the protocol can changes channels up to 1600 times per second. If there is interference from other devices, the transmission does not stop, but its speed is downgraded.

Bluetooth version 1.2 allowed a maximum data rate of 1 Mbps; this results in an effective throughput of about 723 kbps. In late 2004, a new version of Bluetooth known as Bluetooth version 2 was ratified; among other features it included enhanced data rate (EDR). With EDR the maximum data rate is able to reach 3 Mbps (throughput of 2.1 Mbps) within a range of 10 m (up to 100 m with a power boost). Older and newer Bluetooth devices can work together with no special effort [4.8]. Because a device such as a telephone headset can transmit the same information faster with Bluetooth 2.0 + EDR, it uses less energy, since the radio is on for shorter periods of time. The data rate is improved by more efficient coding of the data sent across the air; this also means that for the same amount of data, the radio will be active less of the time, thus reducing the power consumption [4.7]. Newer Bluetooth devices are efficient at using small amounts of power when not actively transmitting: for example, the headset is able to burst two to three times more data in a transmission and is able to sleep longer between transmissions. Noteworthy features of Bluetooth core specification version 2.0 + EDR include:

- Three times faster transmission speed than that of preexisting technology
- Lower power consumption through a reduced duty cycle
- Simplification of multilink applications due to increased available bandwidth
- Backwardly compatible to earlier versions
- Improved bit-error-rate performance

Hardware developers were shifting from Bluetooth 1.1 to Bluetooth 1.2 in the recent past; Bluetooth 2.0 products were being introduced at press time. To be exact, version 2.0 devices have a higher power consumption; however, the fact that the transmission rate is three times faster (thereby reducing the transmission burst times) effectively reduces consumption to half that of 1.x devices. Devices are able to establish a trusted relationship; a device that wants to communicate only with a trusted device can authenticate the identity of the other device cryptographically. Trusted devices may also encrypt the data that they exchange over the air.

A Bluetooth device playing the role of "master" can communicate with up to seven devices playing the role of "slave" (groups of up to eight devices are called *piconets*). At any given instant in time, data can be transferred between the master and one slave; but the master switches rapidly from slave to slave in a round-robin fashion. (Simultaneous transmission from the master to multiple slaves is possible but is not used much in practice.) The Bluetooth specification also makes it possible to connect two or more piconets to form a *scatternet*, with some devices acting as a bridge by simultaneously implementing the master role in one piconet and the slave role in another piconet.

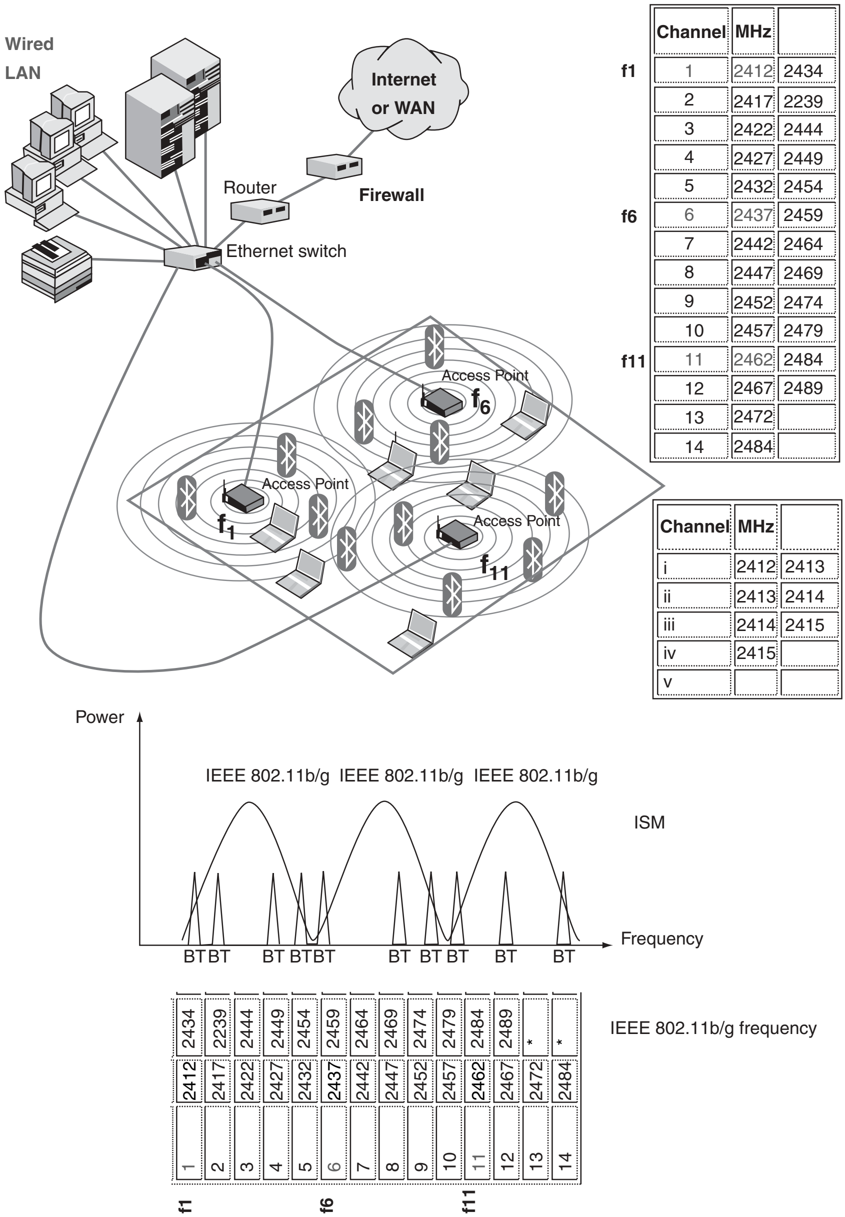
The Bluetooth SIG recently established a road map for future improvements to Bluetooth. Priorities for 2005 included quality of service (QoS), security, and power consumption; priorities for 2006 were to include multicast, additional security, and long-range performance. The Bluetooth SIG is also working with developers of UWB to ensure backward compatibility with the new standard. UWB is a short-distance wireless protocol capable of transmitting up to 100 Mbps of data a distance of about 10 m; Bluetooth is only capable of 1 to 3 Mbps over the same distance. It is conceivable that Bluetooth could be supplanted by this faster technology, so the Bluetooth SIG is working to make sure that UWB is backwardly compatible with current Bluetooth devices (at present, two groups are competing for their technology to be ratified as the UWB standard). Depending on the usage cases, technologies such as ZigBee and UWB can be either complementary or overlapping [4.7]. It is hypothetically possible that Bluetooth wireless technology and UWB could converge, but work and agreements will need to take place to make this happen. The immediate problems for UWB—the two competing standards and the lack of the international regulatory approval—need to be resolved for the idea of convergence to be interesting for Bluetooth wireless technology.

WLAN The following are areas where advances in wireless LAN (WLAN) is taking place:

1. Higher WLAN speeds to support an adequate number of users in high-density environments and also voice over IP (VoIP) users. The transition to an IEEE 802.11g and/or 802.11n environment is a basic necessity in a high-density and/or high-bandwidth context.
2. Support of QoS over the wireless (and also core intranet) infrastructure. The deployment of IEEE 802.11e QoS-supporting technology is another basic necessity.
3. Secure communications is highly desirable. The deployment of IEEE 802.11i security capabilities is yet another requirement.
4. Roaming between access points, floors, and subnets is needed, as is a handoff to a cellular service when corporate WLAN service is no longer available or generally, for WN mobility situations. The deployment of IEEE 802.11r roaming capabilities addresses this requirement (capabilities not expected to be available and/or implemented until sometime in the future). Roaming also brings up the question of whether a traditional IP solution is adequate or if one needs to utilize Mobile IP (MIP) (IETF RFC 3344) [4.9]; this is a fairly complex issue.

The IEEE 802.11b and 802.11g specifications postulate a partitioning of the spectrum into 14 overlapping staggered channels whose center frequencies are

5 MHz apart; within this partitioning of the ISM spectrum, channels 1, 6, and 11 (and if available in the regulatory domain, channel 14) do not overlap. These channels (or other sets with similar gaps) can be used so that multiple networks can operate in close proximity without interfering with each other (see Figure 4.6).



	Channel	MHz	
f1	1	2412	2434
	2	2417	2239
	3	2422	2444
	4	2427	2449
	5	2432	2454
f6	6	2437	2459
	7	2442	2464
	8	2447	2469
	9	2452	2474
	10	2457	2479
f11	11	2462	2484
	12	2467	2489
	13	2472	
	14	2484	

Channel	MHz	
i	2412	2413
ii	2413	2414
iii	2414	2415
iv	2415	
v		

2434	2239	2444	2449	2454	2459	2464	2469	2474	2479	2484	2489	*	*	IEEE 802.11b/g frequency
2412	2417	2422	2427	2432	2437	2442	2447	2452	2457	2462	2467	2472	2484	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	
f1					f6					f11				

Figure 4.6 IEEE 802.11b/g frequency bands, typical topology, and bluetooth interaction.

TABLE 4.7 IEEE WLAN-Relevant Frequencies in Various Parts of the World

Channel	MHz	U.S.	Canada	Europe (ETSI)	Spain	France	Japan
1	2412	×	×	×		×	×
2	2417	×	×	×		×	×
3	2422	×	×	×		×	×
4	2427	×	×	×		×	×
5	2432	×	×	×		×	×
6	2437	×	×	×		×	×
7	2442	×	×	×		×	×
8	2447	×	×	×		×	×
9	2452	×	×	×		×	×
10	2457	×	×	×	×	×	×
11	2462	×	×	×	×	×	×
12	2467			×		×	×
13	2472			×		×	×
14[a]	2484						

[a] Channel 14, where available, is restricted to 802.11b operation.

The spectral mask for 802.11b requires that the signal be at least 30 dB down from its peak energy at ±11 MHz from the center frequency and at least 50 dB down from its peak energy at ±22 MHz from the center frequency. Note that if the transmitter is sufficiently powerful, the signal can be relatively strong even beyond the ±22-MHz point (e.g., a powerful transmitter on channel 6 can easily overwhelm a weaker transmitter on channel 11); in most situations, however, the signal in a given channel is sufficiently attenuated to interfere only minimally with a transmitter on any other channel.

The channels that are available for use in a particular country differ according to the regulations of that country. Table 4.7 identifies IEEE-relevant frequencies in various parts of the world. In the United States, for example, FCC regulations allow only channels 1 to 11 to be used. Channels 10 and 11 are the only channels that work in all parts of the world, because Spain has not licensed channels 1 to 9 for 802.11b operation.

The UNII band used in the IEEE 802.11a context is in the range 5.15 to 5.85 GHz. The 802.11a standard uses 300 MHz of bandwidth; the spectrum is divided into three *domains*, each having restrictions imposed on the maximum output power allowed. The first 100 MHz in the lower-frequency portion is restricted to a maximum power output of 50 mW; the second 100 MHz has a higher maximum, 250 mW; and the third, 100 MHz, intended primarily for outdoor applications, has a maximum power output of 1.0 W. It is generally recognized that the higher-frequency UNII band is limited intrinsically to shorter ranges than the ISM band, due to higher path loss, limiting the utility of 802.11a relative to that of 802.11b/g in the WSN context, except for within-building applications. In particular, there is an increase of excess path loss with frequency. Table 4.8 provides a comparison

TABLE 4.8 A Comparison of IEEE 802.11b/g and IEEE 802.11a

	802.11b/802.11g	802.11a
Available bandwidth	83.5 MHz	300 MHz
Unlicensed frequencies of operation	2.4–2.4835 GHz	5.15–5.35 GHz, 5.725–5.825 GHz
Number of non-overlapping channels	3 (indoor–outdoor)	4 (indoor–outdoor)
Data rate per channel	1, 2, 5.5, 11, 54 Mbps	6, 9, 12, 18, 24, 36, 48, 54 Mbps
Modulation	DSSS	OFDM

between IEEE 802.11b/g and IEEE 802.11a. The IEEE 802.11a protocol uses a complex digital modulation method: specifically, orthogonal frequency-division multiplexing (OFDM); this digital modulation method requires more linearity in amplifiers because of the higher peak-to-average power ratio of the OFDM signal transmitted. In addition, better phase noise performance is required because of the closely spaced overlapping carriers. These issues tend to add to the implementation cost of 802.11a products. Although IEEE 802.11a was approved in the late 1990s, new product development has proceeded much more slowly than with 802.11b/g, due to the cost and complexity of implementation.

Frequency-division multiplexing (FDM) is a multiplexing technology that transmits multiple signals from or for different users simultaneously over a single transmission path, such as a cable or wireless system (commercial FM radio is an example). Each signal occupies its own unique frequency range (carrier), which is modulated by the data (text, voice, video, etc.). The OFDM spread-spectrum technique distributes the data over a large number of carriers that are spaced apart at precise frequencies. This spacing provides the orthogonality, which prevents the demodulators from seeing frequencies other than their own. The benefits of OFDM are high spectral efficiency, resiliency to RF interference, and lower multipath distortion. This is useful because in a typical terrestrial broadcasting scenario there are multipath channels (i.e., the signal transmitted arrives at the receiver using various paths of different length). Since multiple versions of the signal interfere with each other [intersymbol interference (ISI)] it becomes difficult to extract the original information. OFDM is the modulation technique used for digital television in Europe, Japan, and Australia [4.10].

As stated previously, a drawback of 5 GHz is that higher-frequency signals experience more difficulties propagating through physical obstructions encountered in an office (walls, floors, and furniture) than do those at 2.4 GHz. There is an intrinsic degradation in throughput as the distance between the transmitter and receiver increases. See Figure 4.7 for a comparison of the two standards or bands with regard to propagation or performance and distance. An advantage of 802.11a is its ability to deal with delay spread and multipath reflection effects: The slower symbol rate and placement of significant guard time around each symbol reduces

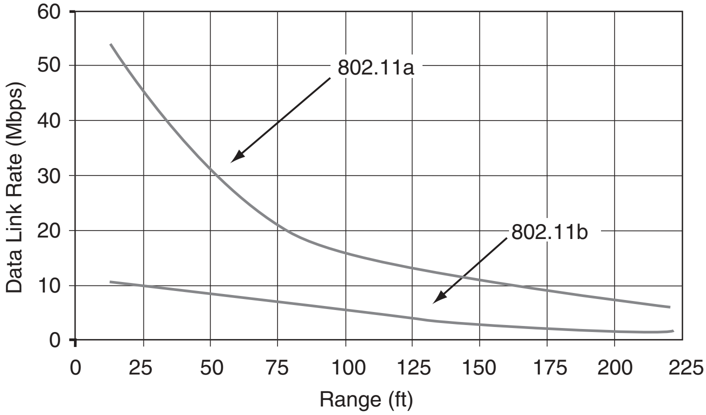


Figure 4.7 Performance characteristics of IEEE 802.11a: throughput comparison versus distance (indoor applications).

the ISI caused by multipath interference; by contrast, 802.11b networks are generally range limited by multipath interference rather than the loss of signal strength over distance.

Now-emerging multiple-input, multiple-output (MIMO) systems use multiple antennas to transmit and receive radio signals. MIMO methods increase the throughput and quality of the signals received. IEEE 802.11n uses MIMO techniques. For example, MIMO–OFDM will allow service providers to deploy a broadband wireless access (BWA) system that has non-line-of-sight (NLOS) functionality. Specifically, MIMO–OFDM takes advantage of the multipath properties of environments using base station antennas that do not have LOS. As noted, in multipath environments the original signal and the individual echoes each arrive at the receiver antenna at slightly different times, causing the echoes to interfere with one another, thus degrading signal quality. The MIMO system uses multiple antennas to transmit data simultaneously in small segments to the receiver, which can process the data flows and put them back together. This process, called *spatial multiplexing*, increases the data-transmission speed proportionally by a factor equal to the number of antennas transmitting. In addition, since all data are transmitted both in the same frequency band and with separate spatial signatures, this technique utilizes the spectrum fairly efficiently [4.10].

ZigBee In this section we provide a brief description of ZigBee. ZigBee is the only standards-based technology designed to address the unique needs of low-cost, low-power WSNs for remote monitoring, home control, and building automation network applications in the industrial and consumer markets [4.11]. The wireless systems discussed in previous subsections provide high data rates at the expense of power consumption, application complexity, and cost. However, there are many wireless monitoring and control applications for industrial and home markets that require longer battery life, lower data rates, and less complexity than those made available by existing wireless standards. For commercial success one needs a standards-based wireless technology that has performance characteristics that closely meet the requirements for reliability, security, low power, and low cost [4.12], [4.40].

For such wireless applications a targeted standard has been developed by the IEEE [4.13]. The IEEE 802.15 Task Group 4 was chartered to investigate a low-data-rate solution with multimonth to multiyear battery life and very low complexity. The standard is intended to operate in an unlicensed international frequency band. Potential applications fot this standard are home automation, wireless sensors, interactive toys, smart badges, and remote controls. The scope of the task group has been to define the physical layer (PHY) and the media access control (MAC) [4.14]. This standards-based interoperable wireless technology is optimized to address the specific needs of low-data-rate wireless control and sensor-based networks [4.12]. Functionality defined by the ZigBee Alliance is used at the upper layers.

The ZigBee Alliance ratified the first ZigBee specification in 2004, making the development and deployment of power-efficient, cost-effective, low-data-rate monitoring, control, and sensing networks a reality. ZigBee/IEEE 802.15.4 is expected to become the leading wireless technology for a plethora of uses, ranging from building automation to industrial and residential applications. Developers were anticipating ZigBee-compliant consumer products as quickly as early 2005 [4.11]. A graphical representation of the areas of responsibility between the IEEE standard, ZigBee Alliance, and user is presented in Figure 4.8; the definition of the application profiles is organized by the ZigBee Alliance [4.11], [4.42].

Hotspot/WiMax In recent years service providers have deployed IEEE 802.11b/11g-based hotspot services to support Internet access and VoIP applications [4.15]. Furthermore, there is interest in delivering metro-wide Internet/VoIP services using WiMax (IEEE 802.16-based) connectivity. Since WiMax is newer, we focus here on this technology (see Table 4.9 for a technical comparison of WiMax to Wi-Fi [4.16]).

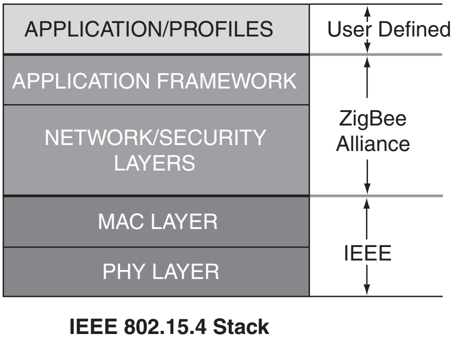


Figure 4.8 ZigBee protocol stack.

TABLE 4.9 Comparative Overview of Wi-Fi and Mobile WiMax Technology

	Wi-Fi Based on 802.11	WiMax Based on 802.16e-2005
Spectrum	Unlicensed, 2.4 GHz (ISM band) and 5.8 GHz (UNII band)	Licensed less than 6 GHz
Range and coverage	Typically less than 100 meters	PMP, NLOS typically 1 to 10 km depending on frequency and terrain characteristics. Point-to-point, LOS up to 50 km
Applications	Indoor WLAN, fixed and nomadic usage model	Outdoor WMAN, fixed, nomadic, portable, and mobile applications
Peak downstream data rate	Up to 54 Mbps in 20 MHz channel BW	Up to 50 Mbps in 10 MHz channel BW
QoS	802.11e provides better QoS support than 802.11a,b,g but only set traffic priorities (up to 8). It is not deterministic, so it is possible for one connection or traffic type to override and starve another connection	Provides guaranteed service levels for specific types of traffic on a connection-by-connection basis. 802.16 uses priority, committed, and peak information rates and can meet specific latency and jitter requirements for specific types of traffic. Can also regenerate network clocks over the air
Privacy and security	WEP uses a repetitive key and is easily defeated. This has been upgraded to WPA and WPA2 with 802.11i	802.16 has two data encryption modes: mandatory 56-bit DES and 128 AES. Also supports device base station, subscriber station, and user authentication. Has secure key exchange and is 802.1x compliant
Latency	CSMA/CA approach for scheduling increases latency with multiple connections. Latency is not deterministic and therefore, adversely affects QoS	Uses a grant-request mechanism as opposed to CSMA/CA; this eliminates delays with multiple users sharing the same channel. This is necessary to support latency sensitive traffic such as VoIP
System gain	System gain is limited by transmit power limits in the unlicensed 2.4 and 5 GHz bands, thus limiting the range capability of 802.11. Support for MIMO in 802.11n will improve this somewhat. Lack of support for subchannelization limits uplink system gain with battery-operated laptops as subscriber stations	Licensed frequency bands permits higher base station Tx power. Subchannelization provides increased system gain in the uplink direction. Adaptive antenna systems including MIMO, beamforming, space-time coding (STC), and spatial multiplexing (SM) also enhance system gain and range
Support for battery-operated handsets	Subchannelization is not supported so subscriber station Tx power must be sufficient to transmit full channel. This is satisfactory for a laptop with a large battery or	Uplink subchannelization reduces Tx power requirements for battery-operated subscriber devices. Various sleep mode options are available to conserve battery life

TABLE 4.9 (*Continued*)

	access to AC power, but not acceptable for mobile handhelds, PDAs, etc. Also no sleep mode	
Multipath immunity	OFDM with a FFT size of 64 provides some immunity to multipath	S-OFDMA with FFT size of 512 to 2048 FFT for channel BWs from 5 to 20 MHz
Interference immunity	802.11 does not have support for transmit power control (TPC) or dynamic channel selection (DCS). Some of these issues are addressed with 802.11h	Aided by transmit power control, subchannelization and support for adaptive antenna systems

Source: WiMax Forum.

The IEEE 802.16 Working Group has developed a point-to-multipoint (PMP) broadband wireless access standard for systems in the frequency ranges 10 to 66 GHz and sub-11 GHz. This technology is targeted to metropolitan area environments. The IEEE 802.16 standard covers both the MAC and PHY layers. A number of PHY considerations were taken into account for the target environment. At higher frequencies, line of sight (LOS) is a must. This requirement eases the effect of multipath, allowing for wide channels, typically greater than 10 MHz in bandwidth. This gives the IEEE 802.16 protocol the ability to provide very high capacity links on both the uplink and downlink. For sub-11 GHz, non-line-of-sight (NLOS) capability is a requirement. The original IEEE 802.16 MAC was enhanced to accommodate different PHYs and services, which address the needs of different metropolitan environments. The standard is designed to accommodate either time-division duplexing (TDD) or frequency-division duplexing (FDD) deployments, allowing for both full- and half-duplex terminals in the FDD case [4.16]. IEEE 802.16a has a LOS radius of 50 km and an NLOS of 10 km or thereabouts, depending on the types of obstacles in the topography. WiMax is the marketing name of the IEEE 802.16 standard.

The MAC was designed specifically for the PMP wireless access environment. It supports higher layer or transport protocols, such as ATM, Ethernet, and IP, and is designed to accommodate easily future protocols that have not yet been developed. The MAC is designed for high bit rates (up to 268 Mbps each way) and operates on a broadband physical layer, while delivering ATM-compatible QoS, UGS (unsolicited grant service), rtPS (real-time polling service), nrtPS (non-real-time polling service), and best effort services. The frame structure allows terminals to be dynamically assigned uplink and downlink burst profiles according to their link conditions. This allows a trade-off between capacity and robustness in real time and provides roughly a two-fold increase in capacity on average compared to nonadaptive systems while maintaining appropriate link availability. The 802.16 MAC uses a variable-length protocol data unit (PDU) along with a number of other concepts that greatly increase the efficiency of the standard. Multiple MAC PDUs may be concatenated into a single burst to save PHY overhead. Additionally, multiple service data units (SDUs) for the same service may be concatenated into a single MAC PDU, saving on MAC header overhead. Fragmentation allows large SDUs to be sent across frame boundaries to guarantee the QoS of competing services. Payload header suppression can be used to reduce the overhead caused by the redundant portions of SDU headers. The MAC uses a self-correcting bandwidth request–grant scheme that eliminates the overhead and delay of acknowledgments while allowing better QoS handling than that of traditional acknowledgment schemes. Terminals have a variety of options for requesting bandwidth, depending on the QoS and traffic parameters of their services. Terminals can be polled individually or in groups; they can steal bandwidth already allocated to make requests for more; they can signal the need to be polled, and they can piggyback requests for bandwidth [4.16].

A typical WiMax network consists of a base station supported by a tower- or building-mounted antenna. The base station connects to the appropriate terrestrial network (PSTN, Internet, etc.) Applications include, but are not limited to, point-to-point communication between stations, point-to-multipoint communication between the base station and clients, backhaul services for Wi-Fi (802.11) hotspots, broadband Internet services to home users, private-line services for users in remote locations, and metro-wide WSN applications.

4.3.2 MAN/WAN Applications

MAN/WAN sensor communications can occur over WiMax/hotspots or 3G systems. After a brief discussion of a brand-new (but speculative) technology, cognitive radios (CRs), in the remainder of the section we focus on the evolution of cellular networks in terms of the desire to provide a lateral data channel that supports any number of applications, including WSNs.

Cognitive Radios and IEEE 802.22 With the plethora of wireless services that are becoming available, stakeholders believe that the limiting factor at this time is the scarcity of radio spectrum. Studies have shown that most of this spectrum scarcity is concentrated in the unlicensed bands; this is where the major advancements in spectrum use have taken place (e.g., Wi-Fi, cordless phones). Licensed bands, however, typically experience considerable underutilization. CR-based approaches represent a new paradigm in wireless communications that aims at utilizing the large amount of underused spectrum in an intelligent way while not interfering with other incumbent devices in frequency bands already licensed for specific uses [4.43].

The IEEE 802.22 wireless regional area network (WRAN) standard is the first worldwide project to employ CR concepts for dynamically sharing spectrum with television broadcast signals. IEEE 802.22 seeks to develop a standard for a cognitive radio-based PHY–MAC–air interface for use by license-exempt devices on a noninterfering basis in spectrum allocated to the television broadcast service. This standard specifies the air interface, including the MAC and PHY, of fixed point-to-multipoint wireless regional area networks operating in the VHF–UHF TV broadcast bands between 54 and 862 MHz. This standard is intended to enable deployment of interoperable IEEE 802 multivendor wireless regional area network products, to facilitate competition in broadband access by providing alternatives to wireline broadband access and extending the deployability of such systems into diverse geographic areas, including sparsely populated rural areas, while preventing harmful interference to incumbent licensed services in the TV broadcast bands [4.44].

There is a large untapped market for broadband wireless access in rural and other unserved or underserved areas where wired infrastructure cannot be deployed economically. Products based on this standard will be able to serve those markets and increase the efficiency of spectrum utilization in spectrum currently allocated to, but unused by, the TV broadcast service. WRAN supports an approach for operation over large, potentially sparsely populated areas (e.g., rural areas), taking advantage of the favorable propagation characteristics in the VHF and low-UHF TV bands. The unique requirements of operating on a strict noninterference basis in spectrum assigned to, but unused by, the incumbent licensed services requires a new approach using purpose-designed cognitive radio techniques that will permeate both the PHY and MAC layers [4.44]. In principle, this wireless service can also be used to support metro-area WSNs.

Cognitive radio—where a device can sense its environment and location and then alter its power, frequency, modulation, and other parameters so as to dynamically reuse available spectrum—is now just emerging. CR can, in theory, allow multidimensional reuse of spectrum in space, frequency, and time, obliterating the spectrum and bandwidth limitations that have slowed broadband wireless development in the United States and elsewhere. This new technology is in a way similar to *software-defined radio* (SDR). With SDR the software embedded in a radio cell phone, for example, can define the parameters under which the phone should operate in real time as its user moves from place to place; traditional cell phone parameters, by contrast, are relatively fixed in terms of frequency band and protocol. A SDR is a flexible wireless communications device that implements its signal processing entirely in software: Software radios can easily change such features as modulation, bandwidth, and coding, which are fixed in more traditional radios. The basic technology of software radio is now being deployed in military and commercial applications. CR is even more advanced than SDR: CR, as noted, can sense its environment and learn from it [4.45]. The FCC is currently investigating commercial applications, and the Defense Advanced Research Projects Agency is proposing military applications (under the XG—or next-generation communications—program). DARPA's aim is to develop technology that allows multiple users to share spectrum in a way that coexists with, and complements, sharing protocols included in today's Wi-Fi technologies. Work on CR and IEEE 802.22 is currently under way.

3G Cellular Networks Over the past decade, mobile communications technology has evolved from first-generation (1G) analog voice-only communications to second-generation (2G) digital, voice, and data communications. The demand for more cost-effective and feature-enhanced mobile applications has led to the development of new-generation wireless systems (or simply 3G). State-of-the-art 3G handsets are designed to provide multimegabit Internet access with an "always on" feature and data rates of up to 2.048 Mbps [4.17].

In reference to cellular applications, the core network of traditional cellular systems is typically based on a circuit-switched architecture similar to that utilized in wireline networks. Wireless service providers are now in the process of evolving their core networks to IP technology. Wireless telecommunications started as a subdiscipline of wireline telephony, and the absence of global standards resulted in regional standardization. Two major mobile telecommunications standards have emerged: time-division multiple access/code-division multiple access (TDMA/CDMA) developed by the Telecommunications Industry Association (TIA) in North America, and Global System for Mobile Communications (GSM) developed by the European Telecommunications Standards Institute (ETSI) in Europe. As one moves toward third-generation (3G) wireless services, there is a need to develop standards that are more global in scope [4.18].

In the late 1990s there were discussions on the development of standards for a 3G mobile system with a *core network* based on evolutions of the GSM and an *access network* based on all the radio access technologies (i.e., both frequency- and time-division duplex modes) supported by the plethora of different carriers (in different countries). This project was called the Third Generation Partnership Project (3GPP) [4.19]. Around the turn of this decade, the American National Standards Institute (ANSI) decided to establish the Third Generation Partnership Project 2 (3GPP2), a 3G partnership initiative for evolved ANSI/TIA/Electronics Industry Association (EIA) networks [4.20]. In addition, there also was the establishment of a strategic group called International Mobile Telecommunications-2000 (IMT-2000) within the International Telecommunication Union (ITU) [4.21], which focused its work on defining interfaces between 3G networks evolved from GSM on the one hand and ANSI on the other, with the goal of enabling seamless roaming between 3GPP and 3GPP2 networks. Because of the worldwide ("universal") roaming characteristic, 3GPP started referring to 3G mobile systems as the Universal Mobile Telecommunication System (UMTS) [4.22]. Since then, there has been advocacy for and progress toward an *all-IP UMTS network architecture*. The all-IP UMTS specifications replaced the earlier circuit-switched transport technologies by utilizing packet-switched transport technologies, and introduce multimedia support in the UMTS core network [4.22].

Figure 4.9 depicts some basic industry transition paths to 3G wireless. As implied in the preceding paragraph, currently the 3G world is split into two camps: the cdma2000, which is an evolution of the IS-95 standard, and the wideband code division multiple access (W-CDMA)/time-division synchronous CDMA (TD-SCDMA)/enhanced data rates for GSM evolution (EDGE) camp, whose standards are improvements of GSM, IS-136, and packet data cellular (PDC)—these are all second-generation standards. In the United States, Verizon Wireless and Sprint PCS were the first two carriers to develop 3G networks. The other major carriers have already advanced to the 2.5G technology, with the vision to soon join the 3G community [4.17].

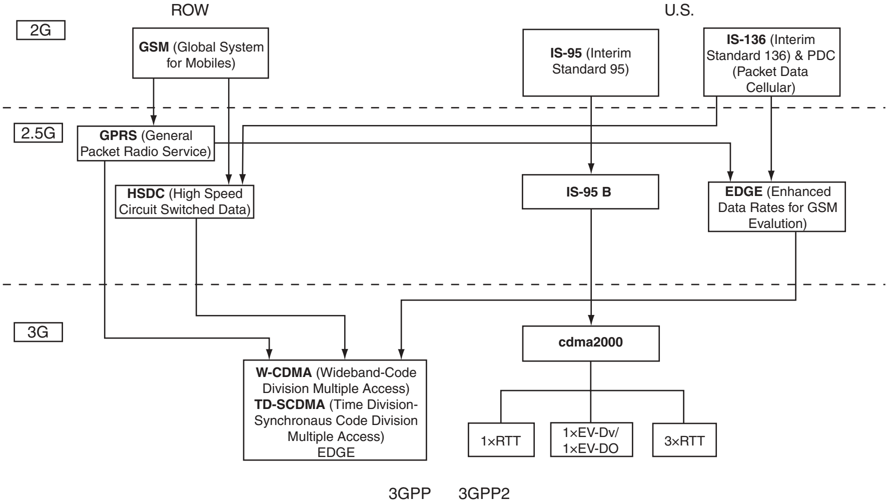

ROW
U.S.
2G
GSM (Global System for Mobiles)
IS-95 (Interim Standard 95)
IS-136 (Interim Standard 136) & PDC (Packet Data Cellular)
2.5G
GPRS (General Packet Radio Service)
HSDC (High Speed Circuit Switched Data)
IS-95 B
EDGE (Enhanced Data Rates for GSM Evalution)
3G
cdma2000
W-CDMA (Wideband-Code Division Multiple Access)
TD-SCDMA (Time Division-Synchronaus Code Division Multiple Access)
EDGE
1×RTT
1×EV-Dv/
1×EV-DO
3×RTT
3GPP
3GPP2

Figure 4.9 Migration path(s) to 3G wireless networks.

The original scope of 3GPP was to produce globally applicable technical specifications and technical reports for a 3G mobile system based on evolved GSM core networks and the radio access technologies that they support [i.e., universal terrestrial radio access (UTRA), both FDD and TDD modes]. The scope was subsequently amended to include maintenance and development of the GSM technical specifications and technical reports, including evolved radio access technologies [e.g., general packet radio service (GPRS) and EDGE] [4.23]. 3GPP and 3GPP2 also address the issue of the limited data throughput capabilities of 2G/2.5G systems, motivating providers to start work on 3G wideband radio technologies that can provide higher data rates (e.g., for Internet access, messaging, location-based services). This work resulted in 3G wireless radio technologies that provide data rates of 144 kbps for vehicular, 384 kbps for pedestrian, and 2 Mbps for indoor environments, and meet the ITU IMT-2000 requirements. Clearly, these channels can be utilized for WSN applications. Now that the radio technology standards to support higher data rates have been developed, the providers are focusing on development of standards for all-IP networks [4.18].

3GPP The basic characteristics of an all-IP network are end-to-end IP connectivity, distributed control and services, and gateways to legacy networks [4.18]. As noted earlier in the chapter, there are two major protocol suites for supporting VoIP: session initiation protocol (SIP), standardized by the IETF, and H.323, standardized by the ITU. It was decided in 3GPP to use only SIP as the call control protocol between terminals and the mobile network. Interworking with other H.323 terminals (e.g., fixed H.323 hosts) is performed by a dedicated server in the network. New elements in this architecture, compared to a traditional 2G cellular network, are as follows (see also Figure 4.10) [4.22]:

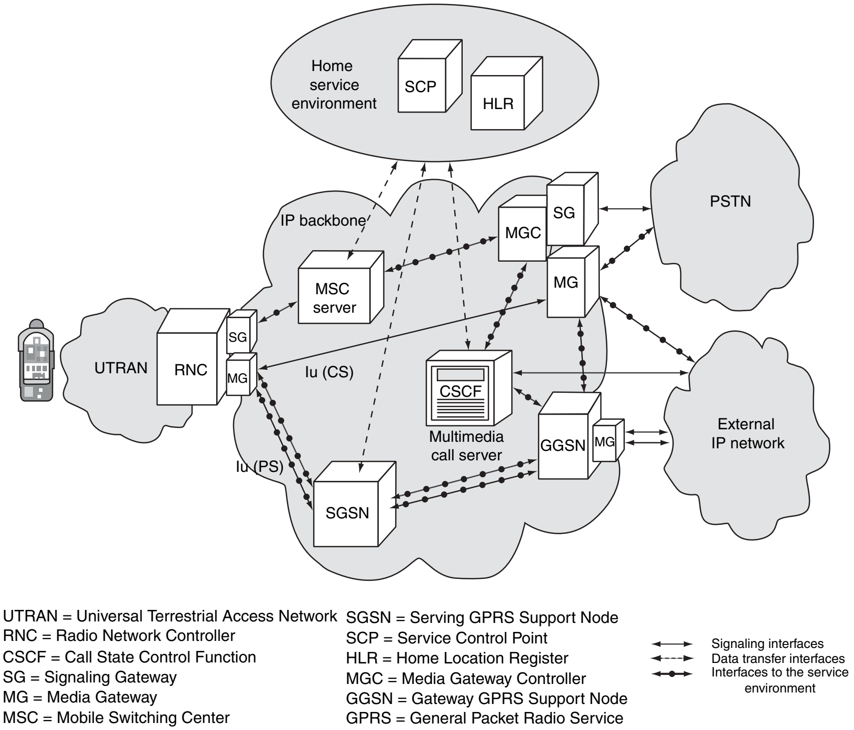

Home service environment
SCP
HLR
IP backbone
PSTN
SG
MGC
MG
MSC server
UTRAN
RNC
SG
MG
Iu (CS)
CSCF
Multimedia call server
GGSN
MG
External IP network
Iu (PS)
SGSN
UTRAN = Universal Terrestrial Access Network
RNC = Radio Network Controller
CSCF = Call State Control Function
SG = Signaling Gateway
MG = Media Gateway
MSC = Mobile Switching Center
SGSN = Serving GPRS Support Node
SCP = Service Control Point
HLR = Home Location Register
MGC = Media Gateway Controller
GGSN = Gateway GPRS Support Node
GPRS = General Packet Radio Service
Signaling interfaces
Data transfer interfaces
Interfaces to the service environment

Figure 4.10 All-IP 3G cellular service. (From [4.22].)

1. *Mobile switching center* (*MSC*) *server.* The MSC server controls all calls coming from circuit-switched mobile terminals and mobile-terminated calls from a PSTN/GSM network to a circuit-switched terminal. The MSC server interacts with the media gateway control function (MGCF) for calls to and from the PSTN. There is a functional split of the MSC, where the call control and services part is maintained in the MSC server, and the switch is replaced by an IP router [Media Gateway (MG)]. This functional split reduces the deployment cost and guarantees the support of all existing services.
2. *Call state control function* (*CSCF*). The CSCF is an SIP server that provides or controls multimedia services for packet-switched (IP) terminals, both mobile and fixed.
3. *MG at the Universal Terrestrial Access Network* (*UTRAN*) *side*. The MG transforms VoIP packets into UMTS radio frames. The MG is controlled by the MGCF by means of Media Gateway Control Protocol ITU H.248. The media gateway is added to fulfill the second requirement. In Figure 4.10 the MG is drawn at the UTRAN side of the Iu interface, hence the Iu interface between the core network and UTRAN is IP-based. The MG can also be located at the core network side of the Iu interface (without impact on the UTRAN).

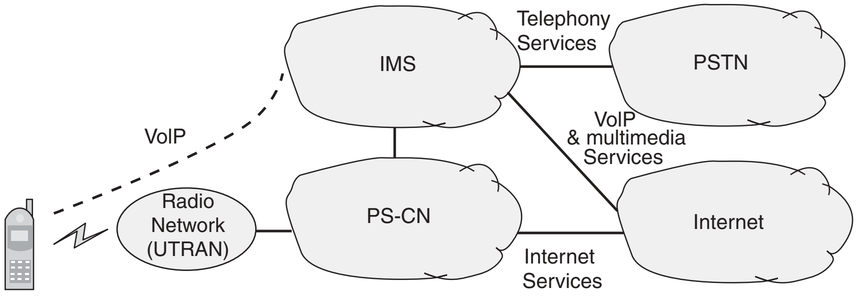
4. *MG at the PSTN side*. All calls coming from the PSTN are translated to VoIP calls for transport in the UMTS core network. This MG is controlled by the MGCF using the ITU H.248 protocol.
5. *Signaling gateway* (*SG*). An SG relays all call-related signaling to and from the PSTN and UTRAN on an IP bearer and sends the signaling data to the MGCF. The SG does not perform any translation at the signaling level.
6. *MGCF*. The first task of the MGCF is to control the MGs via H.248. Also, the MGCF performs translation at the call control signaling level between ISDN user part (ISUP) signaling used in the PSTN and SIP signaling used in the UMTS multimedia domain.
7. *Home subscriber server* (*HSS*). The HSS is the extension of the home location register (HLR) database with the subscribers' multimedia profile data.

For the transport of data traffic, UMTS uses the General Packet Radio Service (GPRS) network. For voice calls there are two options: for packet-switched mobile terminals, voice data are transported over the GPRS network using the GPRS tunneling protocol (GTP) on top of IP; all mobility is addressed by the GPRS protocols. For circuit-switched mobile terminals, voice samples are transported over IP between the MGs using the Iu frame protocol; in the latter case there is no tunneling; hence mobility has to be solved in a different way, by media gateway handovers.

An essential architectural principle of the 3GPP framework is to provide separation of service control from connection control. 3GPP started with GPRS as the core packet network and overlaid it with call control and gateway functions required for supporting VoIP and other multimedia services. The functions are provided via IETF-developed protocols to maintain compatibility with the industry direction in all-IP networks. These new networks also provide VoIP capabilities; the same capabilities that support VoIP can also support WSNs. To support VoIP, call control functions are provided by the call state control function (CSCF) (refer to Figure 4.10). The mobile terminal communicates with the CSCF via SIP protocols. The CSCF performs call control functions, service switching functions, address translation functions, and vocoder negotiation functions. For communication to the public-switched telephone network (PSTN) and legacy networks, PSTN gateways are utilized. To support roaming to 2G wireless networks, roaming gateway functions are also provided. The serving GPRS support node (SGSN) uses existing GSM registration and authentication schemes to verify the identity of the data user. This makes the SGSN access-technology-dependent. The GPRS HLR is enhanced for services that use IP protocols. The data terminal makes itself known to the packet network by doing a *GPRS-attach*. The IP address is anchored in the GPRS gateway node, GGSN, during the entire data session. This limits the mobility of the data terminal to within GPRS-based networks. To provide mobility with other networks, a MIP foreign agent can be incorporated in the GGSN [4.18].

3G Release 1999 was the first release of the 3GPP specifications; it was essentially a consolidation of the underlying GSM specifications and the development of the new UTRAN radio access network. The foundations were laid for future high-speed traffic transfer in both circuit- and packet-switched modes. That release was followed over the years by Releases 4, 5, and 6 [4.23]. Release 1999 was an introductory specification on the architecture of the UMTS network. According to Release 1999, UMTS comprises a UTRAN and two core networks [circuit-switched core network (CS-CN) and packet-switched core network (PS-CN)], which link up to services networks such as the PSTN and the Internet. Thus, using both traditional circuit- and modern packet-switched networks, UMTS Release 1999 supports various services, including voice, data (fax, SMS), and Internet access. Later, Release 4 adapted to the same architecture added more services to the UMTS network. The coexistence of two core networks, however, signified many limitations compared to competitive 3G systems, especially in video and multimedia services. Release 5 was a solution to the limitations that came along to modernize the UMTS architecture currently employed in 3G networks around the world. In this final phase, the PS-CN dominates the CS-CN and takes responsibility for telephony services. Systems based on UMTS Release 5 have much lower infrastructure and maintenance costs and provide enhanced services. Release 6 added additional capabilities [4.17].

As seen at the macro level in Figure 4.11, a new component is added to the basic UMTS architecture: the supplementary IP Multimedia Subsystem (IMS). IMS aims at supporting both telephony and multimedia services. IMS's role in UMTS architecture is to interact with both the PSTN and the Internet to provide all types of multimedia services to users. The CSCF element in the IMS infrastructure is responsible for signaling messages between all IMS components in order to control multimedia sessions originated by the user. Consequently, there is a proxy-CSCF (P-CSCF), an interrogating-CSCF (I-CSCF), and a serving-CSCF (S-CSCF), all responsible for particular signaling functions using SIP. The P-CSCF's responsibility



UTRAN = UMTS Terrestrial Access Network
PS-CN = Packet Switched Core Network
IMS = IP Multimedia Subsystem

Figure 4.11 UMTS Release 5 basic architecture.

is to act as the QoS enforcement point and to provide local control for emergency services. I-CSCF is an optional component that interacts with the HSS to find the location of the S-CSCF (it is optional because the P-CSCF can be set up to negotiate directly with the S-CSCF). The S-CSCF controls all the session management functions for the IMS. Depending on the capabilities of the IMS and the capacity requirements, there may be more than one S-CSCF node, and others can eventually be added to the system. The function of the HSS is to handle all user information, such as subscription and location queries. The HSS communicates with the CSCFs via an IP-based protocol called Cx interface; all other IMS components interact with each other via SIP. The media gateway control function (MGCF) is in charge of controlling one or more MGs; the MGCF interacts with the S-CSCF and the transport signaling gateway (T-SGW). MGs are bit processors for end-to-end users; their function is to convert PCM in the PSTN to IP-based formats, and vice versa. Finally, the T-SGW is included in the IMS because of the need to convert signaling system number 7 (SS7) to IP since the PSTN is only SS7-compatible [4.17].

3GPP2 3GPP2 has also undertaken work to enhance the IP architecture for multimedia services (including voice). The approach here is to capitalize on the synergies of Internet technologies and to use a single network for all services. 3GPP2 has created a new packet data architecture building on the CDMA 2G and 3G air interface data services. 3GPP2 has taken advantage of 3G high data rates and existing work in IETF on MIP to enhance the network architecture to provide IP capabilities. One advantage of using IETF protocols is ease in interworking and roaming with other IP networks. The other major advantage is that it can provide private network access (virtual private networking) via a MIP tunnel with IP security [4.18].

In the 3GPP2 architecture, IP connectivity reaches all the way to the base station transceiver (BTS). Both the base station controller (BSC) and BTS are contained in the IP-based radio access network node. This means that the BSC will be a router-based IP node containing some critical radio control functions (e.g., power control, soft handoff frame selection). The remaining control functions, such as call and session control, mobility management, and gateway functions, are moved out to the managed IP network. This allows for a distributed and modular control architecture. Since much of the communication will be between wireless and legacy terminals, gateway functions are provided for roaming to 2G wireless networks and interworking with the PSTN. In the 3GPP2 architecture, the mobile terminal uses mobile-IP-based protocols to identify itself. The packet data serving node (PDSN) contains a MIP foreign agent (FA) functionality. When the mobile terminal attaches to the FA, the FA establishes a mobile IP tunnel to the home agent (HA) and sends a registration message to the HA. The HA accesses the authorization, authentication, and accounting (AAA) server to authenticate the mobile terminal. The IP address of the mobile terminal is now anchored in the HA for the duration of the data session. The data device connected to the mobile terminal can be handed over to any other access device that supports mobile IP. Thus, this approach can provide mobility across different access networks (wireless, wireline, etc.). However, since it essentially uses address translation to provide mobility, it cannot do fast handoff, due to the latency of address updates from distant agents [4.18].

Comparison of Services The 3GPP and 3GPP2 architectures are different because of the underlying base networks and evolution strategies. In 3GPP, GPRS-based mobility was already defined, so the IP network enhancements were considered on top of GPRS. On the other hand, 3GPP2 needed to develop a mobility mechanism for packet data since one did not exist previously. As noted, 3GPP2 has decided to use MIP as the basis for packet data mobility [4.18].

To illustrate the similarities and differences of the two approaches, mobility needs to be addressed at three levels: air-interface mobility, link-level mobility, and network-level mobility. Air-interface mobility supports cell-to-cell handoff within a radio access network. Link-level mobility maintains a point-to-point protocol (PPP) context across multiple radio access networks. Network-level mobility provides mobility across networks. In both approaches, air-interface mobility is handled in the radio access network. Air-interface mobility is specific to the radio technology, therefore harmonization of the two depends on the harmonization efforts under way for global CDMA. In 3GPP, link-level mobility is handled by GTP; this protocol is used to provide mobility to other 3GPP-defined networks. The 3GPP architecture also provides an option in which an FA may be located in the GGSN. This allows roaming from GPRS-based networks to other IP access networks. In 3GPP2, link-level mobility is provided by defining a tunneling protocol as an extension of MIP. The MIP architecture allows the mobile device to have a point of presence and to roam across any IP network. Registration and authentication in the 3GPP architecture for access and data networks are integrated and utilize the schemes used for wireless. In the 3GPP2 architecture, the registration and authentication for access and data networks are performed separately. For a data network, authentication and registration as defined in MIP are used; hence, the data architecture is access-independent [4.18].

3G Operators After many delays, 3G networks are now being rolled out. 3G wireless networks offer all the normal mobile telephony services plus high-speed data access. 3G operators may initially limit data access to their own branded data services or at least price open Internet access significantly higher than access to their own traditional data services. The mobile market, however, is competitive, and there are consumer and business requirements for access to the open Internet. In fact, flat-rate bundles for data access services are already available in some markets. This data-channel access can be used to support VoIP services [4.24]. Wireless operators that are looking to continue to displace wireline voice revenues as their business posture need to reduce their overall delivery costs as users move from 2G TDM to 3G VoIP [4.25]. Below we look briefly at the VoIP possibilities because a successful commercial "play" in this space would accelerate the deployment (and ubiquity) of 3G services, thereby indirectly opening up an opportunity for WSN applications.

For example, equipment upgrades can introduce high-speed data capabilities to UMTS networks. Specifically, new technologies now becoming available enable carriers to provide new "blended lifestyle services" via any wireline, wireless, or Wi-Fi/WiMax endpoint by providing a variety of 3GPP IMS functional elements (as discussed previously), including the call session control functions, media resource function controller, policy decision function, and breakout gateway control function. Because this equipment expands the data channel on 3G cellular networks, these upgrades also lay the foundation for operators to introduce VoIP and more advanced multimedia services on their mobile networks (here one can transmit IP-voice datagrams over the data channel). VoIP over 3G gives operators the ability to support a greater number of voice users at a lower cost, in turn helping to ensure that voice services can continue to be delivered profitably. Some researchers estimate that 3G wireless can deliver voice by way of VoIP for a quarter of the cost per minute compared to 2G TDM methods [4.25].

For mobile operators that have invested heavily in 2G and 3G cellular networks, there may be relatively little incentive to offer VoIP services according to observers (their existing networks already deliver better-quality voice services at lower cost than VoIP can achieve today). However, VoIP may look more attractive to those service providers seeking to bypass mobile operators' traditional voice tariffs, particularly if an opportunity to undercut those tariffs using VoIP arises due to significant drops in 3G data pricing. A number of mobile operators have launched unlimited-use data tariffs that could make them vulnerable to customers using VoIP to cut their spend [4.26]. 3G service-provider VoIP offerings could appear in the United States in the 2008 or 2009 time frame. That would come after operators upgrade their 2.5G/3G networks. For example, upgrades to 1xEV-DO provide peak data rates of about 1.8 Mbps compared to typical rates of 300 to 400 kbps for the current generation of 1xEV-DO [4.27].

Calculations of the threat to 3G revenues from broadband wireless (WiMax) have focused mainly on data, but as some 3G carriers start to put VoIP in a more central position in their strategies, they could find that this service segment is also affected. The 3G UMTS and CDMA technologies may have been the first to promise both voice and broadband-class data on one network and device, but the emergence of usable VoW has also moved formerly data-only approaches into this space. A potential early limit on VoIP over 3G data access could be the limited upstream capability of the initial 3G services. W-CDMA can deliver up to 384 kbps downstream but only 64 kbps upstream; it is preferable to have data rates exceeding 64 kbps, but if that is all that is available, one can make do for most VoIP services [4.24]. Road maps for data networks such as CDMA EVDO (evolution—data only) and UMTS's data-only strand, TDD,[2] now include VoIP [4.25].

[2] UMTS TDD mobile broadband technology is a packet data implementation of the international 3GPP UMTS standard. Unlike W-CDMA, which uses FDD (frequency division), UMTS TDD is designed to work in a single unpaired frequency band. One of the largest benefits of using TDD is that it supports variable asymmetry, meaning that an operator can dictate how much capacity is allocated to downlink versus uplink. As the traffic patterns for data typically heavily favor the downlink, this results in better use of spectrum assets and higher efficiency [4.23].

1. The shift is already visible in the CDMA market, even without taking into account challenges from broadband wireless. New EV-DO equipment aims at peak data rates of 3.1 Mbps and supports VoIP. As such, it could perhaps make a further upgrade to the next CDMA generation, EV-DV (evolution—data and voice) unnecessary. This equipment was expected to start shipping in 2006, and although EV-DO with VoIP will take advantage of the spectral efficiencies of CDMA less well than EV-DV, this will be outweighed by early availability and lower prices [4.25].
2. In the UMTS space, manufacturers have already developed a TDD mobile handset offering VoIP as well as the usual broadband packet-based services, and providers have completed the first successful transmission of a call from a mobile VoIP handset over UMTS TDD and claim that the network is ideal for voice because it features high capacity, low latency, and low power requirements. Their services will be more compelling if they can offer voice, and therefore they will be less likely to opt for a pure IP solution such as 802.16 instead of TDD. TDD-ready handsets are currently becoming commercially available [4.25].

Hotspot/WiMax Operators For operators considering deployment of broadband wireless access technologies (e.g., WiMax), being able to offer VoIP could strengthen the business case for investing in such networks by moving operators beyond a focus on low-margin Internet access. Fixed/wireline operators have shown interest in use of wireless VoIP in trying to defend against fixed mobile substitution by developing services that combine VoIP over WLAN/hotspot/WiMax with cellular voice elsewhere [4.24,4.26]. Again, a successful VoIP application would drive deployment, which can be advantageous to WSN applications.

Fixed-Mobile Convergence Operators Recently, there has been interest in fixed-mobile convergence (FMC). Mobile network operators plan to leverage emerging IMS service platforms to deliver "one phone, one number" telephony over both fixed and mobile infrastructure. This means that a mobile handset will use 2G/3G mobile infrastructure when the user is outdoors and VoIP over Wi-Fi when the user is at work or at home. Mobile operators see IMS and FMC as an opportunity to take additional market share from traditional fixed-line operators. However, once high-speed Internet access becomes available on mobile phones, a plethora of VoIP services will follow [4.24].

Most telephone calls originate from inside buildings, where cellular mobile coverage is poorest. As such, residential users are often forced to keep their fixed-line services for use when they are at home; the same applies in office buildings, with the added problem that wireless operators have not been in a position to offer the Centrex or PBX features that enterprises require. In theory, however, that could change with the advent of IMS and FMC [4.24].

To enable converged handsets, FMC relies on broadband Internet access for the fixed portion and WLANs now and WiMax in the future for the mobile portion. WLANs are deployed at a large percentage of enterprises, and home-based Wi-Fi setups are spreading rapidly. Broadband Internet access is also available in thousands of public hotspots. The first round of convergence depends on handsets that support 2G, 3G, and Wi-Fi connections on the same phone. Mobile operators then use an IMS platform to transparently combine regular mobile service on their 2G or 3G mobile network with VoIP services over Wi-Fi and/or fixed broadband access. Because of the fact that the mobile portion of FMC uses the existing mobile number and the existing mobile switching network elements, mobile operators have an advantage [4.24].

Without broadband Internet access, VoIP service providers are less of a threat to mobile operators' FMC services. The business proposition of fixed-mobile convergence is to hit the sweet spot of high convenience and low cost [4.24]. VoIP vendors will be in a better position to provide their own FMC if WiMax delivers on its promise of wireless broadband Internet access; however, widespread WiMax deployment is expected to take a number of years. Instead, the VoIP competitive threat may be enabled by the mobile operators' own data services [4.24]. A successful VoIP penetration could indirectly drive WSN applications by building out the infrastructure.

4.4 CONCLUSION

In this chapter we looked at radio transmission issues. To maximize the opportunity for widespread and cost-effective deployment of WSN, plans are to use existing and/or emerging COTS wireless communications and infrastructures rather than having to develop an entirely new, specially designed apparatus. WSNs can use a number of wireless COTS technologies, such as Bluetooth, ZigBee, WLAN/hotspots, WiMax, and 3G.

APPENDIX A: MODULATION BASICS

Modulation Capsule We have indicated that WSNs implementers will probably use off-the-shelf radio technology such as ZibBee, WiMax, Wi-Fi, or 3G; this means that they do not necessarily have to worry about the fundamental aspects of radio science and modulation. However, a brief discussion of modulation is in order. Table 4A.1 lists some key terms related to modulation, from various sources, including [4.34], [4.37], and [4.38]. In the context of digital transmission and modulation, the related topic of digital encoding is also of interest.

A basic technique used in radio transmission is phase-shift keying (PSK), mentioned in the body of the chapter (Table 4A.1 lists a number of approaches, but PSK is a fundamental methodology). In PSK the frequency and amplitude of the carrying signal are both kept constant. Phase-coherent PSK utilizes two defined signals: A logic 0 is represented by a π-degree phase shift, and a logic 0 is represented by a 0-degree phase shift; this is, however, a complex situation for the