1

Introduction

Dipankar Raychaudhuri and Mario Gerla

1.1 Background

Over the next ten-to-fifteen years, it is anticipated that significant qualitative changes to the Internet will be driven by the rapid proliferation of mobile and wireless computing devices. Wireless devices on the Internet will include laptop computers, personal digital assistants, cell phones (more than 3.5 billion in use as of 2009 and growing!), portable media players, and so on, along with embedded sensors used to sense and control real-world objects and events (see Figure 1.1). As mobile computing devices and wireless sensors are deployed in large numbers, the Internet will increasingly serve as the interface between people moving around and the physical world that surrounds them. Emerging capabilities for opportunistic collaboration with other people nearby or for interacting with physical-world objects and machines via the Internet will result in new applications that will influence the way people live and work. The potential impact of the future wireless Internet is very significant because the network combines the power of cloud computation, search engines, and databases in the background with the immediacy of information from mobile users and sensors in the foreground. The data flows and interactions between mobile users, sensors, and their computing support infrastructure are clearly very different from that of today's popular applications such as email, instant messaging, or the World Wide Web.

As a result, one of the broad architectural challenges facing the network research community is that of evolving or redesigning the Internet architecture to incorporate emerging wireless technologies – efficiently, and at scale.[1] The Internet's current TCP/IP protocol architecture was designed for static hosts and routers connected by wired links. Protocol extensions such as mobile IP have been useful for first-generation cellular mobile services involving single-hop radio links from mobile devices to base stations or access points.[2] However, incremental solutions based on IP are inadequate for dealing with the

https://doi.org/10.1017/CBO9780511921117.002 Published online by Cambridge University Press

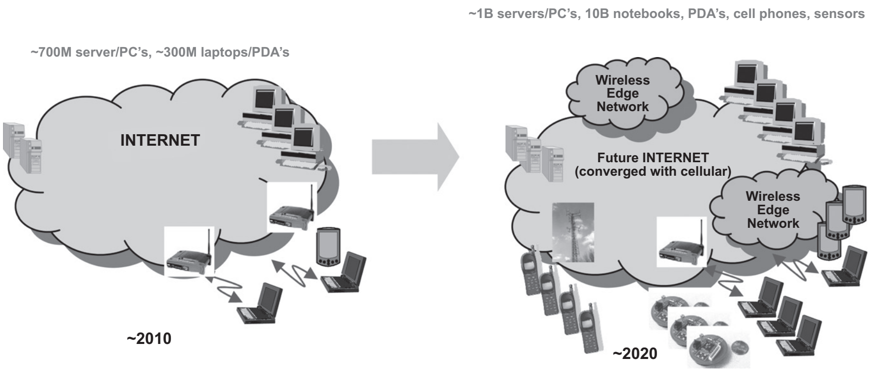

Figure 1.1. Migration of Internet usage from fixed PCs and servers to mobile devices and sensors.

requirements of fast-growing wireless usage scenarios such as multihop mesh,[3] peer-to-peer,[4] disruption-tolerant networks (DTN),[5] sensor systems,[6] and vehicular applications.[7] These emerging wireless scenarios motivate us to consider "clean-slate" network architectures and protocols capable of meeting the needs of these and other emerging wireless scenarios. In the next section (1.2), we present an overview of these emerging wireless networking scenarios, identifying new architecture and protocol requirements for each of these usage cases. These mobile network architecture requirements will then be aggregated into a number of key protocol features in Section 1.3 that follows. Technical challenges associated with implementing these new wireless/mobility requirements into a unified comprehensive future Internet architecture protocol will then be discussed briefly in Section 1.4. Each of the emerging wireless technology scenarios identified in this introductory chapter will then be discussed in greater depth in each of the chapters that follows. In the concluding chapter, we will review the overall challenge of evolving the current Internet to meet these mobile networking needs, and provide a brief view of the road ahead.

1.2 Wireless Technology Roadmap

Wireless and mobile networks represent an active research and new technology development area. The rapid evolution of core radio technologies, wireless networks/protocols, and application scenarios is summarized for reference in the technology roadmap given in Figure 1.2. It can be seen from the chart that in addition to 2.5G/3G cellular data and WLAN systems developed during the 1990s, emerging wireless scenarios include personal-area networks, wireless peer-to-peer (P2P), ad hoc mesh networks, cognitive radio networks, sensor networks, RFID systems, and pervasive computing.

Each of the previously mentioned wireless technologies or usage scenarios is associated with unique network architecture and service requirements that

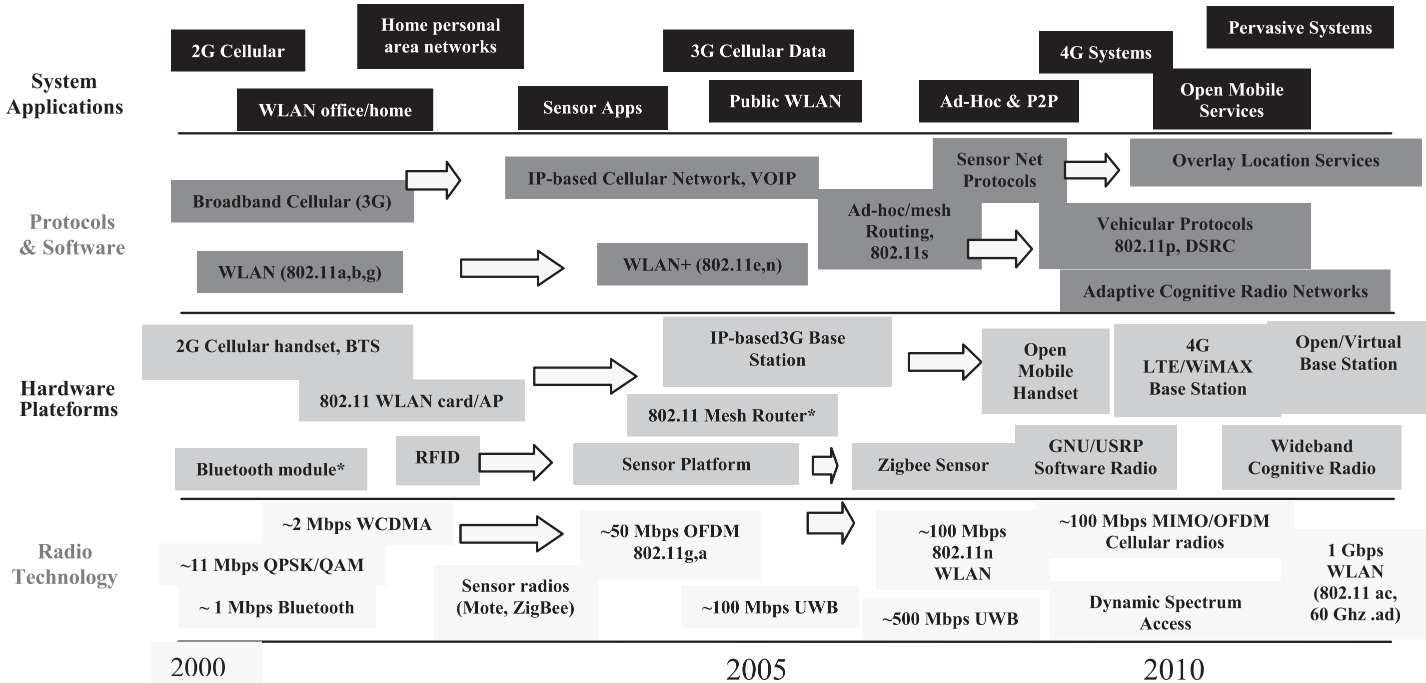

Figure 1.2. Wireless technology roadmap.

affect both the access and infrastructure portions. The default approach adopted by most of the research community is to treat the wireless access portion as a "layer 2" local area network connecting to the Internet (i.e., layer 3 IP) through a gateway. This approach is pragmatic, but it precludes uniform dissemination of control and routing information through the entire network and creates a potential processing bottleneck at the gateway. A more integrated end-to-end control and routing architecture is important for optimizing mobile/wireless service features such as location management, dynamic handoff, quality-of-service (QoS) or cross-layer transport. Also, a local-area wireless network may contain one or more routing elements, which can create inconsistencies in protocol layering and addressing. If compatibility with the current IP network is not viewed as an essential constraint, it may be possible to develop a clean-slate network architecture that can accommodate emerging wireless networks in a single unified protocol structure.

1.3 Wireless Networking Scenarios

The most important wireless technology in use today is the cellular network that provides mobile phone and data services on handheld devices. Cellular networks are ubiquitous in all parts of the world, with almost 4 billion cell phones in use worldwide at the time of writing of this book. Cellular networks have evolved from first-generation analog systems (such as the AMPS system used in the United States prior to 1990) to second-generation digital systems (such as GSM and CDMA[8] used in most parts of the world between 1990 and 2005), and then to third-generation, or 3G, systems such as CDMA2000 and UMTS/WCDMA in use since about 2005. Second-generation cellular systems such as GSM are capable of supporting packet data services at bit-rates of ~100 Kbps, whereas 3G systems such as UMTS or CDMA2000 can deliver between ~300 Kbps – 2 Mbps, depending on signal quality. Further evolution from 3G to 4G cellular systems with the goal of supporting service bit-rates in the range of ~10–100 Mbps is planned by the industry over the next three to five years. Examples of 4G systems are LTE and WiMAX/IEEE 802.16.

From a network architecture point of view, cellular has always been built as a separate custom network with its own set of protocols for key interfaces, such as mobile terminal to base station and base station to mobility service gateways such as the MSC and GGSN. These networks were initially built for integration with the telephone network that was based on a set of signaling protocols defined by the ITU. More recently, 3G networks have been migrating toward integration with the IP network using voice-over-IP (VoIP) protocols such as SIP[9] for signaling and mobility protocols such as mobile IPv6.[2] As data services for mobile devices continue to grow, this may be expected to lead to a gradual migration of mainstream cellular services to the Internet. However, gradual migration of cellular networks to the Internet involves the use of overlays and gateways for interfacing between mobile network features such as authentication, addressing, and mobility – an approach that has scalability and performance limitations.

In addition to cellular, a number of short-range wireless data technologies such as WiFi, Bluetooth, and Zigbee have started to penetrate the market for enterprise and home networks starting in the late 1990s. Of these technologies, WiFi (based on the IEEE 802.11 standard) is the most ubiquitous as an Internet access link, with more than 500 million devices in use today, with the number expected to grow to a billion by 2012.

Most of these WiFi devices are used as wireless local area networks (WLAN) that connect to the Internet as "layer 2" networks similar to the widely used Ethernet LANs. When WiFi is used as a home or office LAN, it is the last hop for Internet access, but does not provide mobility or global roaming features associated with the cellular network. As we will see in later chapters, 802.11 WLAN technology is also being used in the ad hoc mode to build new kinds of networks such as peer-to-peer (P2P), vehicular networks (V2V and V2I), and mesh networks. In addition to 802.11 radios, there are several short-range radio standards such as Bluetooth and Zigbee that are used to provide short-range access to devices such as wireless speakers and sensors/actuators. Power and size limitations on the sensor devices imply the need for a more general wireless network architecture that provides connectivity to a range of heterogeneous radios with different transmission ranges. In contrast to the cellular network, the emerging wireless network will incorporate multiple radio technologies operating under a decentralized control framework.

This is illustrated in Figure 1.3, which shows that the overall network architecture is evolving from the separate special-purpose cellular and WiFi networks toward a more general, heterogeneous wireless access network with multiple radio technologies, opportunistic ad hoc association, self-organization, multihop routing, and so on. The long-term architectural goal would be to evolve the Internet architecture to seamlessly meet all the requirements associated with the general wireless "network of networks" shown in the right-hand side of the figure.

Next, let us consider some of the key wireless networking scenarios of importance to the future Internet architecture. The first and most well-understood emerging wireless service scenario is that of anytime, anywhere access to the Internet from personal mobile devices. As shown in Figure 1.4, this scenario implies the need for a network addressing and routing scheme capable of handling roaming and continuous mobility across multiple points of attachment.

User mobility of this sort is handled quite effectively in today's cellular network using the concepts of a "home network" and "visited network." In particular, users of the network have a permanent address to which all communication is initially addressed, and a forwarding (or visiting) address used to temporarily forward connections during mobility outside the home area. A modified form of this approach has been used in the mobile IP specification

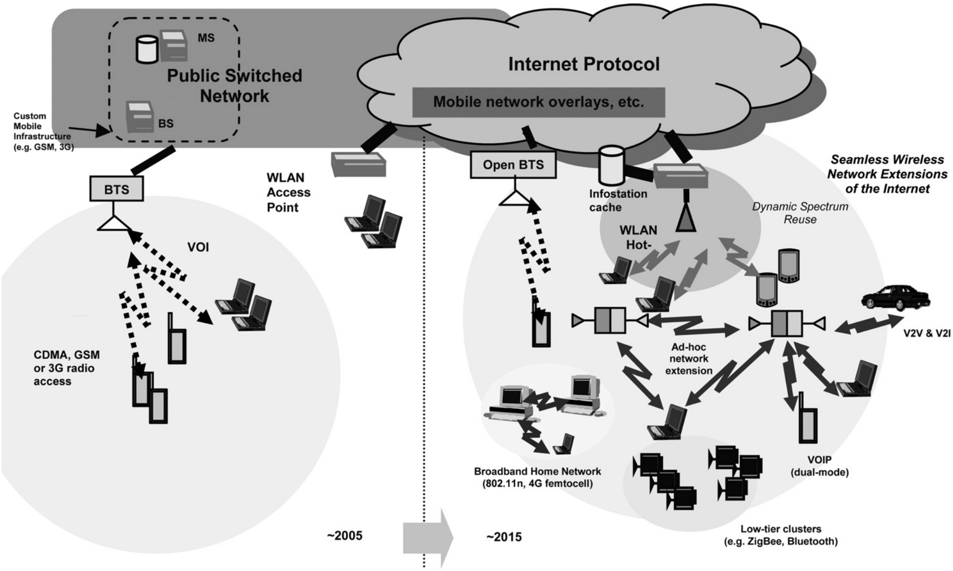

Figure 1.3. Anticipated evolution of wireless network architecture from special-purpose networks to heterogeneous "wireless network of networks."

that is part of IPv6, but is not widely implemented in the Internet today. For connection-oriented traffic, an additional requirement that is also met by today's cellular networks is that of dynamic handoff by which an existing connection can be smoothly migrated from one point of radio attachment to another without setting up a new connection. Clearly, end-user roaming and dynamic mobility support is a key requirement for the future Internet given the rapid increase in mobile data devices. Although mobile IPv6 does provide a solution for this requirement, it may be appropriate to consider alternative approaches toward achieving this functionality in the future network. Mobility also involves security

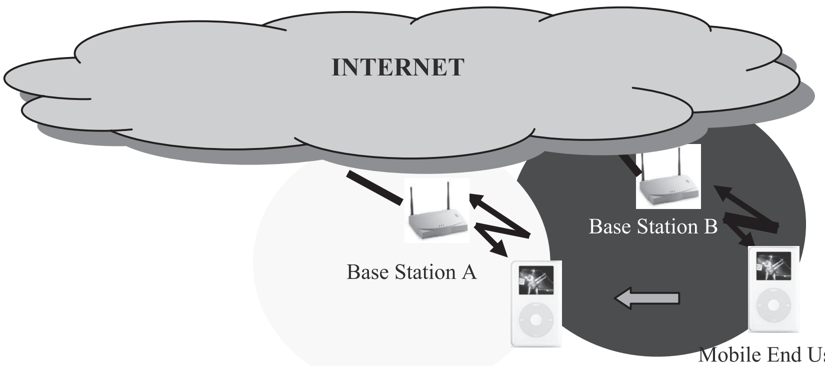

Figure 1.4. Mobile data service scenario.

https://doi.org/10.1017/CBO9780511921117.002 Published online by Cambridge University Press

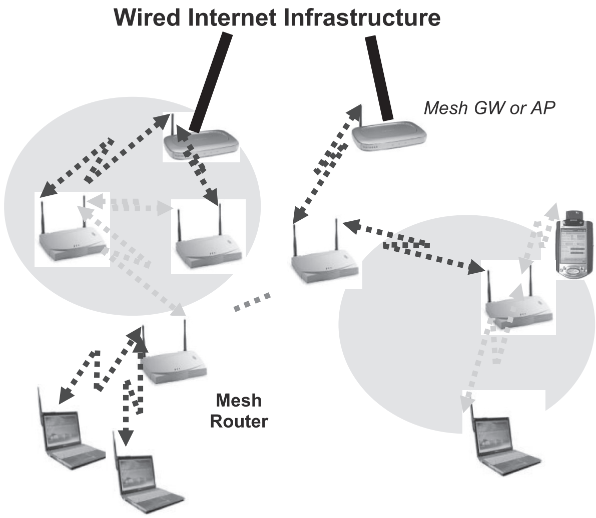

Figure 1.5. Wireless mesh network with multihop routing of data packets between radio nodes.

considerations such as user authentication, which will need to be an integral part of any solution.

A second emerging wireless usage scenario is that of an ad hoc or mesh network in which multiple wireless devices with short-range radios form a multihop network with increased coverage and connectivity. Ad hoc networks were first proposed to support tactical communications between small groups of mobile radio nodes. More recently, multihop mesh architectures (illustrated in Figure 1.5) have been used to extend wireless access network coverage in both urban and rural areas using low-cost short-range radios such as WiFi. In these ad hoc and mesh scenarios, each radio node serves as a router with the capability of forwarding packets to their destination across multiple wireless hops.

Traffic to or from the Internet must pass through one or more gateways or access points that are designed to have both wired and wireless network interfaces. Specialized ad hoc network routing protocols (such as the MANET specification from IETF[10]) have been devised for this purpose, and there is a considerable body of research on this class of routing protocols. Routing in mesh and ad hoc networks generally requires an awareness of cross-layer parameters from the radio links that make up a potential path. Given the growing importance of multihop wireless routing, it may be useful for the future Internet protocol to provide seamless routing across both wired and wireless portions of the network. As for the mobile data service scenario in Figure 1.3, the network needs to support end-user roaming and dynamic mobility as part of the basic transport service.

https://doi.org/10.1017/CBO9780511921117.002 Published online by Cambridge University Press

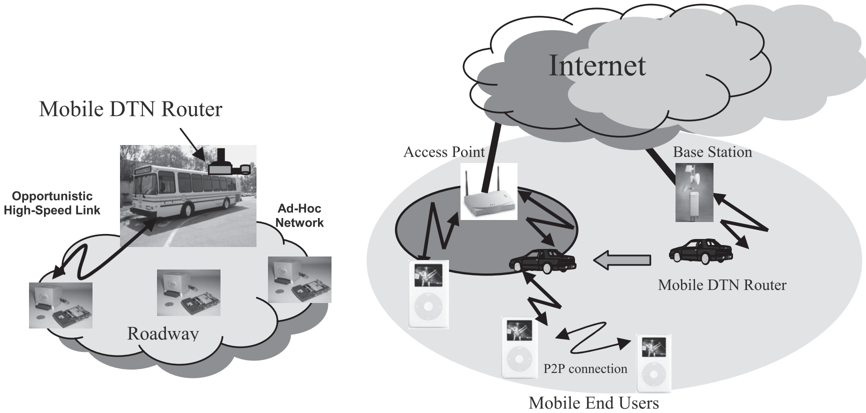

Figure 1.6. P2P wireless networking scenarios involving sensor pickup by a mobile device, or opportunistic content delivery to a passing vehicle.

The third scenario of current significance is the P2P network model in which short-range radios associate with each other opportunistically for content delivery or some type of machine-to-machine (M2M) interaction. This scenario is also sometimes referred to as delay-tolerant networking (DTN), because intermittent opportunistic connectivity implies the need for delay-tolerant applications designed to wait for transmit/receive opportunities. Figure 1.6 shows two kinds of P2P or DTN applications, one in which a bus is picking up data from sensors in the roadway and storing this data for later delivery to the wired network core (perhaps using WiFi or other short-range radios once parked inside its regular garage). The second part of the figure shows the P2P and "Infostations" service models in which users associate opportunistically with each other to exchange content, or when users associate for short periods with wireless data caches (or Infostation) to download popular or personal content. Both these scenarios are important because of the fact that opportunistic short-range radio access is fundamentally faster and more efficient than continuous cellular-type connectivity. Moreover, continuous long-range wireless access may not be feasible for small, low-power sensor devices such as those shown in Figure 1.6. It is noted that the TCP/IP protocol stack used in the Internet today was not designed to support discontinuous or opportunistic connectivity of this type, indicating the need to consider this requirement further when designing future Internet protocols.

Another emerging wireless scenario of importance is that of vehicular networking, involving both V2V (vehicle-to-vehicle) and V2I (vehicle-to-infrastructure) modes. In vehicular networking, cars on the highway may exchange safety information with those in proximity, or might download content (such as navigational maps or audio/video files) from infrastructure access points placed along the highway. The vehicular scenario shares some common elements with the ad hoc and P2P cases considered earlier, but have the additional property of location or geographic awareness. Referring to Figure 1.7, it is observed that

https://doi.org/10.1017/CBO9780511921117.002 Published online by Cambridge University Press

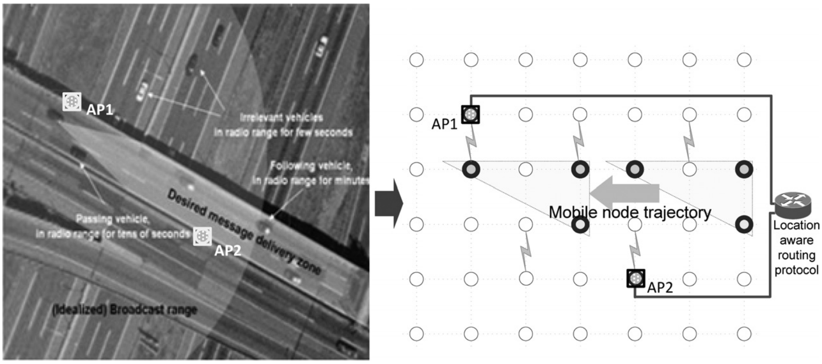

Figure 1.7. Vehicular networking scenario (figure courtesy of Prof. Marco Gruteser).

a typical transmission in a V2V situation is a "geographic multicast" in which a message is propagated to all receivers along a certain section of roadway, but not to those outside that region. This requirement motivates a new service, called geocasting, in which a message is forwarded to all radio nodes within a defined geographic area. This type of network routing is very different from device-address-based routing currently used in the Internet. Given the fact that there are approximately 500 million vehicles worldwide and growing, it would be desirable to consider this geographic routing capability as a requirement when designing future Internet protocols.

Another important wireless scenario is that of sensor networks and pervasive computing (see Figure 1.8). The sensor network scenario generally involves a hierarchical network structure with clusters of low-power sensors connected as

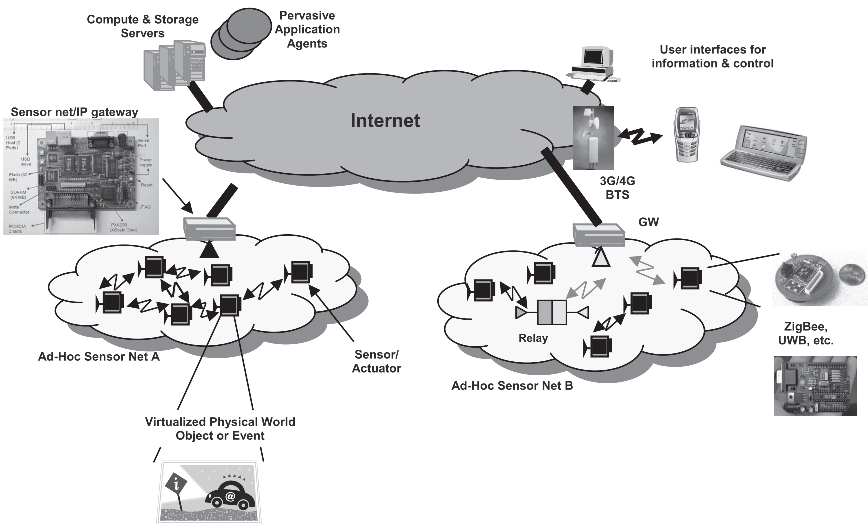

Figure 1.8. Wireless sensor network scenario.

https://doi.org/10.1017/CBO9780511921117.002 Published online by Cambridge University Press

ad hoc multihop networks at the lowest tier. The function of a sensor (or actuator) is to provide a virtualized representation of a physical-world object or event, thus making it possible to design "pervasive computing" applications that allow us to observe and interact with the physical world. The sensor network clusters connect into the Internet cloud through multiple gateways that convert from the localized sensor network protocol to the global Internet protocol.

Within the sensor network cluster, there may also be a tiering of nodes including low-power sensors, relays, forwarding nodes, and gateways. Of course, ad hoc routing considerations similar to those discussed earlier for the ad hoc/mesh case continue to apply. However, there is an additional requirement of energy efficiency because of severe power constraints at each sensor, and there may also be unique data aggregation requirements involving processing and aggregation of data at each transit node. Sensor network applications involve computing and storage servers in the network cloud as shown in Figure 1.8, and there are many related issues of how to architect the computing and networking system given the greater importance of content and location over the physical address itself. Applications will also have an end-user interface, typically a mobile device such as a cellular handset or PDA. Currently sensor systems are built as special-purpose networks with gateways to the Internet, but a long-term goal is to improve scalability and performance by using a single unified protocol across both sensor and Internet clouds.

In concluding this subsection, it is noted that core radio technology itself is going through a fundamental change, moving from hardware radios to cognitive software-defined radios. Examples of early cognitive radio prototypes are the USRP (Universal Software Radio Prototype), WARP from Rice University, the Microsoft Research Software Radio, and the WINLAB WiNC2R software radio platform. Cognitive radios are motivated by the need to use radio spectrum more efficiently to accommodate rapidly increasing wireless traffic. The use of cognitive radios as network elements will enable dynamic spectrum sharing and adaptive networking methods that are inherently more flexible than the radio access technology standards in use today. This implies the need for extensions to control and resource management protocols in the access network, providing for features such as dynamic spectrum coordination, cross-layer awareness, and the ability to set and control radio parameters based on networking requirements.

1.4 Classifying Wireless Networking Scenarios

The NSF Wireless Mobile Planning Group report[1] written in 2005 provides a useful classification for the full range of future wireless networking scenarios, some of which were individually discussed earlier, in Section 1.3. In that report, three distinct clusters of usage scenarios are identified as summarized below.

1.4.1 Scenario A – Individual Wireless Devices Interfacing with the Internet ("Mobile Computing")

The simplest scenario involves a single wireless device that interfaces with the broader Internet. The mobile device may be a cellular phone, a PDA, a media player, a digital camera, or some type of combination consumer device. Mobile computing devices may connect through a wireless local area network, a mesh-style wireless network, or a wide-area wireless technology (such as cellular 3G or WiMAX). Service models to be considered include mobile services, hot-spot services with limited mobility, as well as cached content delivery via opportunistic wireless links. High mobility, the potential for intermittent connectivity, and heterogeneity of radio access are key characteristics of this scenario.

A typical example of this mode of operation is that of a mobile customer downloading a real-time video stream (e.g., a live sporting event) to a portable media player from the Internet. Seamless connectivity should be maintained as the customer moves from a shopping mall (WiFi coverage) to outdoors (2.5G or 3G cellular connectivity), and then to the car (Bluetooth within the car, WiMAX radio to the Internet). At each step, the wireless media player needs to be aware of available connectivity options and then select the best service. The multimedia server must also be aware of current connectivity constraints so that it can deliver a stream with parameters (data rate, format, etc.) consistent with the configuration. The same mobile customer should be efficiently tracked by the network and reachable by VoIP calls if he/she so chooses. Location- or context-aware queries (such as "where is the nearest pharmacy?") and delay-tolerant services (e.g., seamless suspension and resumption of a large file transfer when the user walks or drives through areas without coverage) should be supported. Caching of files for rapid downloading within a hot spot may also be useful in this scenario.

1.4.2 Scenario B – Constellations of Wireless Devices ("Ad hoc Nets")

The second type of wireless scenario is motivated by a variety of settings in which multiple radio devices may be in close physical proximity and can collaborate by forming an ad hoc network. For example, wireless devices in an office or home environment can set up an ad hoc network between themselves to improve coverage and communications quality. Another popular application involving constellations is that of community mesh networks formed by rooftop radios for the purpose of shared broadband access. In the important emerging application of vehicular communication, clusters of cars on the highway may participate in an ad hoc network for the purpose of collision avoidance and traffic flow management. Constellations may include heterogeneous radio and computing devices with different capabilities and resource levels. Emerging cognitive radio

https://doi.org/10.1017/CBO9780511921117.002 Published online by Cambridge University Press

technologies also offer the capability of highly adaptive wireless ad hoc networks with physical layer negotiation between nodes, scavenging unused spectrum at low cost to support a private ad hoc network. Opportunistic association, changing network topologies, varying link quality, and potentially large scale (in terms of number of nodes) are some of the characteristics of this scenario.

A simple example of opportunistic constellations is the formation of an ad hoc network between several user laptops in a meeting room with limited Internet access coverage. The ad hoc network enables high bandwidth communication between participants at the meeting and allows them to use a favorably positioned (e.g., with good cellular network throughput) node as a forwarding relay to the Internet. Another example is the cooperative downloading of popular files from the Internet by drivers on a highway, when hot-spot “Infostations” with WiFi service are spaced by several miles on the highway, and a car traveling at 60 miles per hour may not be able to download an entire file through short V2I (vehicle-to-infrastructure) mode access. If several drivers are interested in the same file, it is possible for the cars to collaborate and exchange segments in a P2P opportunistic networking arrangement similar to that used in Bit Torrent (see Chapter 7). This allows the download to be completed without requiring a car to stop at a hot spot, saving time for the end-user and avoiding traffic congestion problems. The same ad hoc networking capability can also be used by cars to exchange control information necessary for traffic flow management or collision avoidance.

Ad hoc radio constellations also apply to civilian disaster recovery and in tactical defense environments. These applications usually involve communications between a number of first responders or soldiers who work within close proximity of each other. The response team may need to exchange text messages, streaming media (e.g., voice or video), and use collaborative computing to address a shared task such as target recognition or identification of a spectral jammer. Individual nodes may also need to access the Internet for command-and-control purposes or for information retrieval. This application has similarities with the ad hoc mesh network for suburban or rural broadband access mentioned earlier.

1.4.3 Scenario C – Pervasive Systems and Sensor Networks (“Sensor Nets”)

Sensor nets refer to a broad class of systems involving embedded wireless devices connected to the Internet. The first generation of sensor networks involves collecting and aggregating measured data from large numbers of sensors in a specified geographic area. In the near future, sensor net applications will also include closed-loop sensor/actuator systems for real-time control of physical world objects. Current sensor net applications are in science (ecology, seismology, ocean and atmospheric studies, etc.) and engineering (water quality

https://doi.org/10.1017/CBO9780511921117.002 Published online by Cambridge University Press

monitoring, precision agriculture, livestock tracking, structural monitoring), as well as consumer-oriented applications (home security and energy management, hobbyist and sports enthusiast applications of distributed imaging, eldercare, pet monitoring, etc.). Sensor networks share several characteristics of ad hoc scenarios but are differentiated by the fact that tiny sensor devices have more stringent processing power, memory, and energy constraints. These constraints generally imply the need for a hierarchical ad hoc network structure in which low-tier sensor nodes connect to the Internet via one or more levels of repeating wireless gateways. Other important characteristics of this scenario are the data-centric nature of applications, potential for large scale (in terms of numbers of sensors), and geographic locality.

Traditionally, large "sensor fabrics" such as those installed to monitor the environment have been designed as vertically optimized systems, with an ad hoc network designed to meet specific energy and processing constraints and optimized to support specialized queries dictated by the application at hand. The interface to the Internet has been via edge nodes that isolate the Internet stack from the sensor fabric architecture. However, more recent trends indicate an increased need for sensor networks that provide open access via the Internet, in a more extensive and capillary way that can be supported via edge nodes. For instance, scientists interested in the correlation between data found in different data bases (e.g., soil characteristics, pollutants carried in the local water supplies, productivity of local vineyards, production and sale of local wines) can be permitted to access specific regions within a sensor fabric directly from the Internet to extract the required data rather than overburdening the access gateways. Moreover, new types of sensor networks based on "mobile" sensor platforms are becoming available – for example, vehicles in the urban grid or firefighters in a disaster recovery operation equipped with a variety of sensors (video, chemical, radiation, acoustic, etc.). These sensor platforms have practically unlimited storage, energy, and processing resources. The vehicle grid then becomes a sensor network that can be accessed from the Internet to monitor vehicle traffic congestion and to help investigate accidents, chemical spills, and possible terrorist attacks. Likewise, firefighters carry cameras and several other sensors, allowing the commander to be aware of the conditions in the field and to direct the operations to maximize the use of the forces while preserving the life of his responders. These latter examples also show that the gap between sensor networks and ad hoc networks tends to diminish in mobile sensor systems at least in terms of communications capabilities and Internet access. In the longer term, pervasive systems involving personal mobile devices, smart offices/homes, and densely deployed multimodal sensors/actuators will serve as a platform for development of various new applications ranging from tracking and inventory control to personal productivity, public safety, and resource management.

1.5 Future Network Requirements

Before moving to more detailed discussions of future wireless scenarios and their networking protocols in the following chapters, let us briefly consider the general future network design requirements that arise from the scenarios introduced in this chapter.

Considering the wide range of future wireless network usage scenarios (4G cellular/mobile, WLAN, mesh, P2P, DTN, sensor networks, vehicular networks, sensor/pervasive systems), it is important to extract a set of common requirements general enough to meet these needs, as well as those of future applications that cannot easily be predicted today. We suggest an approach for decomposing these requirements into two major categories, the first reflecting the intrinsic properties of the radio medium and the second reflecting the needs of future mobility and pervasive services. It is important to note that these requirements should apply to future access networks and the Internet protocol stack as a whole in view of the increasingly predominant role of wireless end-user devices. The current approach of designing specialized networking solutions for cellular systems, ad hoc nets, sensor applications, and so on leads to undesirable fragmentation (and hence poor scalability, lack of interoperability, inefficiencies in application development, etc.) among different parts of the network, and needs to be replaced by a unified end-to-end protocol architecture that supports emerging requirements of both wired and wireless networks.

To elaborate further, basic transport services of future Internet protocols should reflect intrinsic radio properties such as spectrum use, mobility, varying link quality, heterogeneous PHY, diversity/MIMO, multihop, multicast, and so on, and the capabilities of emerging radio technologies such as LTE, next-generation WLAN, Bluetooth, Zigbee, vehicular standards such as 802.11p, and of course, cognitive software-defined radio (SDR). In addition, Internet protocol service capabilities should be designed to serve emerging uses of wireless technology, not only for conventional mobile communications, but also for content delivery, cloud computing, sensing, M2M control, and various other pervasive system applications.

Here, we briefly identify some of the key requirements for a future network designed to support the range of wireless usage scenarios discussed in Sections 1.3 and 1.4. Of course, it might not be feasible to achieve the full set of requirements in a single networking architecture, but it is still instructive to understand all the needs in a top-down manner before considering implementation issues. Examples of specific mobile network protocol features that may be useful are:

1. *Dynamic spectrum coordination capability*: Historically, network protocols have been designed to support resource management in terms of wired network concepts such as link bandwidth and buffer storage. As radios

https://doi.org/10.1017/CBO9780511921117.002 Published online by Cambridge University Press

become an increasingly important part of the network, it will be useful to be able to specify and control radio resources within the networking protocol itself. For example, control protocols should be able to support dynamic assignment of spectrum to avoid conflicts between multiple radio devices within the network. Just as current IP networks incorporate protocols such as dynamic host control protocol (DHCP) for address assignment, future networks could incorporate a distributed repository of spectrum usage information that could then be used to assign nonconflicting spectrum to radio devices when they join the network.

2. *Dynamic mobility for end-users and routers*: As more and more end-user devices become wireless, networks will need to be designed to support mobility as a normal mode of operation rather than as a special case. This means that end-user devices should be able to attach to any point in the network (i.e., global roaming), with the network providing for fast authentication and address assignment at a very large scale. Currently available mechanisms such as DHCP and mobile IP represent a first in this direction, but a more general solution could involve a clean separation of naming and addressing where each device would have a unique name, but would only be assigned a routable address local to the network with which it is currently associated (and this routable address may be as general as a geographic location, i.e., geo-address). The main challenge is to provide a distributed global name resolution and address assignment service that scales to the level of billions of mobile devices. Because wireless devices may also serve as routers in some of the ad hoc environments discussed earlier, the network should be able to support dynamic migration of subnetworks. In addition, dynamic handoff of traffic from one point of attachment to another may be required for certain connection-oriented services.
3. *Fast discovery and ad hoc routing*: Because several wireless usage scenarios involve ad hoc associations and continuously changing network topology, it is important for the network to support fast discovery of neighboring network elements. Discovery protocols for ad hoc networks should support efficient topology formation in multihop wireless environments taking into account both connectivity requirements and radio resources. Multihop wireless scenarios further require efficient ad hoc routing between network elements with dynamically changing topologies and radio link quality. The ad hoc routing protocol used in wireless access networks should seamlessly integrate with the global routing protocol used for end-to-end connectivity.
4. *Cross-layer protocol stack for adaptive networks*: Routing in multihop wireless networks requires a greater awareness of radio link layer parameters to achieve high network throughput and low delay. This means that the network's control plane should include information about radio link parameters to be used for algorithms that support topology discovery and routing. A

https://doi.org/10.1017/CBO9780511921117.002 Published online by Cambridge University Press

key architectural issue is that of determining the appropriate granularity and degree of aggregation with which this cross-layer information is exchanged across different parts of the network (i.e., access, regional, core, etc.).

5. *Incentive mechanisms for cooperation*: Ad hoc mobile networking scenarios typically involve cooperation among independent wireless devices. It will be important for future Internet protocols to include protocols that enable such cooperation, first by advertising resources and capabilities to neighboring radios and second by providing mechanisms for exchange of credits or barter of resources in return for services such as relaying or multihop forwarding.
6. *Routing protocols for intermittent disconnection*: Today's Internet routing and transport protocols are designed under the assumption of continuous connectivity. However, this assumption is no longer valid for mobile devices that frequently experience disconnection due to radio signal fading and/or service unavailability. Future protocols should be designed for robustness in presence of occasional disconnection. In order to achieve this, the network generally needs to be able to store in-transit data during periods of disruption, while forwarding messages opportunistically when a path becomes available.
7. *Transport protocols for time-varying link quality*: Reliable delivery of data on the Internet is currently accomplished using transport control protocol (TCP) for end-to-end flow control and error control. TCP is known to perform poorly in wireless access networks that are characterized by higher packet error rates than wired links, along with time-varying bandwidth caused by variations in radio channel quality and medium access control (MAC) layer contention. Future transport layer protocols should be designed to work efficiently in presence of packet errors and varying end-to-end bandwidth – this will require the ability to distinguish between congestion in the network and channel quality variations.
8. *Efficient multicasting and multipath routing*: The wireless channel has inherent multicast capabilities, that is, a single packet sent by a radio is simultaneously received by all receivers within the transmission range. This property can be exploited to improve network performance in various scenarios, but the routing and transport protocols have to be enhanced to support multicast operation as a core capability. Radio multicast also opens up the possibility of multipath routing in which multiple independent paths are used for routing a single packet to improve end-to-end reliability and delay.
9. *Location awareness and geographic routing*: As discussed earlier, emerging pervasive computing applications (i.e., vehicular, sensor, M2M) often require the ability to delivery packets to an entire geographic region rather than to a specific IP address. Also, for mobility services, knowledge of the current geographic location is central to providing various new services such as navigation and geographic search. This means that future networks should provide location information as a basic control plane capability. In

https://doi.org/10.1017/CBO9780511921117.002 Published online by Cambridge University Press

addition, it would be desirable to optionally offer geographic multicast and routing modes by which packets can be delivered to a specified geographic region.

10. *Content- and context-awareness*: A number of future network service scenarios involve content addressability or content routing. For example, an M2M application might involve a query for a particular functionality (such as "printer"), and it would thus be useful if the network protocol can resolve a content query to one or more specific network addresses. Another network capability to be considered is that of content routing by which network routers forward traffic based on content attributes of the data being carried in the packet rather than the IP address in the header.
11. *In-network storage for content caching*: A network with content addressability capabilities can also be enhanced to provide in-network storage and caching services in an integrated manner. Caching of popular or personal content can provide significant improvements in both end-user application performance and network throughput. Although these capabilities can be provided above the network as an "overlay," it is worth considering whether content caching should be fully integrated with the network layer protocol to minimize control overheads and delay.
12. *Programming model for in-network processing*: Emerging sensor and pervasive applications may involve in-network computation for functions such as data aggregation, data-dependent routing or local content search. Whereas these functions are typically implemented above the networking layer as overlays, it is worth considering basic computing features for a future mobile network in which an increasing proportion of applications would benefit from in-network computation. A key issue is the design of a programming model by which to specify optional computational functions at each network element.
13. *Enhanced security and privacy for radio medium*: Because the wireless channel is open to eavesdroppers and potential denial-of-service attackers, it is important to consider enhanced security and privacy features for emerging mobile networks. User mobility implies the need for stronger authentication features as a baseline for any device joining the network, while the open radio medium means that transmissions should generally use strong encryption. In addition, if the network has information about location or content, it would be important to build in privacy guarantees that prevent tracking of users or their content.

1.6 Discussion

In Section 1.5, we have used a top-down approach to identify a number of new network protocol capabilities that would be desirable for the future mobile Internet. Clearly, it is very difficult to incorporate all or most of these features

https://doi.org/10.1017/CBO9780511921117.002 Published online by Cambridge University Press

into a single network architecture even if we start from a so-called clean slate. Moreover, clean-slate design of an existing network as large and complex as the Internet is not really a practical option, and any practical attempt to upgrade functionality must eventually consider factors such as backward compatibility, evolutionary upgrade of equipment, equipment cost, software complexity, and so on. However, the top-down clean-slate design methodology described in this book is expected to be beneficial because it exposes key requirements and design issues without being constrained by current practices. Although a single new Internet protocol is unlikely to emerge from this methodology, it may be expected that researchers will design and validate several of the key network capabilities outlined in Section 1.5, and eventually some of these ideas will migrate into the mainstream Internet protocol. In the chapters that follow, we will explore the details of protocol design for each of the emerging wireless service scenarios outlined in this introductory chapter. In the concluding chapter, we will briefly discuss a roadmap to the future, including some strategies for how to put all these ideas together into a unified network architecture.

References

[1] Raychaudhuri, D. and Gerla, M. 2005. New Architectures and Disruptive Technologies for the Future Internet: The Wireless, Mobile and Sensor Network Perspective. *Report of NSF Wireless Mobile Planning Group (WMPG) Workshop*. http://www.winlab.rutgers.edu/WMPG

[2] Perkins, C. 1998. Mobile IP. *IEEE Communications Magazine*, 35(5), 84–86, 91–99.

[3] Akyildiz, I., Wang, X., and Wang, W. 2005. Wireless Mesh Networks: A Survey, Computer Networks. *Computer Networks and ISDN Systems*, March.

[4] Lamming, M. and Bohm, D. 2003. SPECs: Another Approach to Human Context and Activity Sensing Research, Using Tiny Peer-to-Peer Wireless Computers. *Lecture Notes in Computer Science*, October.

[5] Fall, K. 2003. A Delay Tolerant Network Architecture for Challenged Internets. *ACM SigComm*.

[6] Culler, D., Estrin, D., and Srivastava, M. 2004. Guest Editors' Introduction: Overview of Sensor Networks. *IEEE Computer Magazine*, August.

[7] Zanella, A., Fasolo, E., Padova, C., Chiasserini, F., Meo, M., Torino, M., Franceschinis, M., and Spirito, A. 2006. Inter-vehicular communication networks. *Second Internal NEWCOM Workshop*.

[8] Rappaport, T. 2002. *Wireless Communications: Principles and Practices*. Second ed. Prentice Hall.

[9] Rosenberg, J. et al. 2002. SIP: Session Initiation Protocol. *Internet Engineering Task Force RFC 3261*.

[10] Perkins, C., and Belding-Royer, E. 2003. Ad hoc On-Demand Distance Vector (AODV) Routing. *Internet Engineering Task Force, RFC 3561*.