1 The concept of cognitive radio

Narayan B. Mandayam and Chandrasekharan Raman

1.1 Motivation for cognitive radios: spectrum is underutilized

Wireless spectrum is one of the most important resources required for radio communications. Throughout the world, spectrum utilization is regulated so that essential services can be provided and also protected from harmful interference. Traditional spectrum governance across the world has tended toward static long-term exclusivity of spectrum use in large geographic areas, often based on the radio technologies employed at the time of decision making. In particular, until recently spectrum regulatory bodies such as the Federal Communications Commission (FCC) in the US or the European Telecommunications Standards Institute (ETSI) in Europe have always allocated spectrum frequency blocks for specific uses, and assigned licenses for these blocks to specific groups or companies.

While the more or less static spectrum allocation strategy has led to many successful applications like, for example, broadcasting and cellular phones, it has also led to almost all of the prime available spectrum being assigned for various applications (see [63]). It may thus seem that there is little or no spectrum available for emerging wireless products and services.

On the other hand, there have been several studies and reports over the years that show that spectrum is in fact vastly underutilized. A report presenting statistics regarding spectrum utilization showed that even during the high demand period of a political convention such as the one held in 2004 in New York City, only about 13% of the spectrum opportunities were utilized [59]. Further, measurement on radio frequency bands from 30 MHz to 3 GHz, collected during 2009 in Vienna, Virginia (a dense suburb of Washington DC) revealed a number of bands with low spectrum occupancy [82].

These findings suggest that devices using advanced radio and signal processing technology should be able to exploit underutilized spectrum. Much of the early motivation for cognitive radio technology was indeed to accomplish such opportunistic spectrum use and to also alleviate the artificial scarcity of prime spectrum. If successful, this technology could revolutionize the way spectrum is allocated worldwide. Among other benefits, it would yield added bandwidth to support the demand for higher-quality and higher-data-rate wireless products and services well into the future.

1.2 What is cognitive radio?

1.2.1 Agile radios and dynamic spectrum access

Cognitive radio is a generic term used to describe a radio that is aware of the environment around it and can adapt its transmissions according to the interference it sees. In their simplest embodiments, cognitive radios can recognize the available systems around them and adjust their frequencies, waveforms and protocols to access those systems efficiently. Conceptually, cognitive radios include multiple domains of knowledge, model-based reasoning and negotiation [44, 62]. The knowledge and reasoning can include all aspects of any radio etiquette such as RF bands, air interfaces, protocols, and spatial as well as temporal patterns that moderate the use of the radio spectrum. An important feature that differentiates cognitive radios from normal radios is their *agility* along the following lines.

Spectrum agility or frequency agility refers to the discovery strategies for available spectrum as well as opportunistic transmission in the identified spectrum. Such operation requires the design of good algorithms and protocols for appropriate selection of transmission frequencies, coordination and cooperation. Spectrum agility also refers to advanced sensing capabilites.

Technology agility refers to operation of a single radio device across various access technologies. Such seamless interoperability can be enabled by multiplatform radios that are realized as a system-on-a-chip (SoC) and can operate for example as WiFi, Bluetooth, FM, and GPS transceivers.

Protocol agility refers to constituting a dynamically reconfigurable protocol stack on radio devices so that they can proactively and reactively adapt their protocols depending on the devices they interact with.

Cognitive radios equipped with such agility would be a first step towards making radios follow an etiquette in a society of radios. Such cognitive behavior could extend to networks of radios so that they mimic human behavior in civilized society [37]. This opens up interdisciplinary approaches to the study of cognitive radio networks that touch upon cognitive and neurosciences, economics, and sociology. Such interdisciplinary methodologies can be applied in modeling the behavior and dynamics of complex networks with cognitive radios.

While these developments offer exciting possibilities for the future, there are many fundamental engineering issues that need to be addressed before such a vision can be realized. Cognitive radio networks basically extend the software defined radio (SDR) framework to the development of dynamic spectrum access (DSA) algorithms that exploit temporal and spatial variability in the spectrum via: (a) initial cooperative neighbor discovery and association; (b) spectrum quality estimation and opportunity identification; and (c) radio bearer management. These, in turn, imply a framework that senses neighborhood conditions to identify spectrum opportunities for communication by building an awareness of spectrum policy, local network policy, and the capability of local nodes (including noncooperative or legacy nodes).

The fundamental issues that need to be addressed include understanding the information theoretic limits of such networks, constructing propagation models for such networks, devising efficient algorithms for spectrum sensing, as well as fostering mechanisms for spectrum coexistence, all of which are addressed in this book. Chapter 2 develops fundamental capacity limits and associated transmission techniques for different cognitive network paradigms. Chapter 3 comprehensively describes the propagation channel models used for design and analysis of cognitive radio systems. Spectrum sensing and basics of white space detection are discussed in Chapter 4 and Chapter 5 discusses how a cognitive radio network can optimize its spectrum sensing function to efficiently exploit the availability of spectrum.

1.2.2 User hierarchy in cognitive radio networks

Since cognitive radios and networks of these involve opportunistic use of spectrum and the associated rights of users to transmit over such spectrum, it is only natural to classify users according to multiple hierarchies. One classification is based on the ownership (license) of spectrum across users. In this scenario, the users with cognitive radios desirous of opportunistic use of the spectrum are usually referred to as the *secondary users*. The incumbent (licensed) users occupying the spectrum are referred to as *primary users*. The secondary users communicate either with infrastructure or other secondary users without interfering with the active primary users. An example of this hierarchy relates to operation of cognitive-radio-enabled secondary users in the TV bands. The primary user is the television receiver in the licensed TV band. But the spectrum may be intermittently used depending on the programming broadcast schedule of the television channel. A secondary user could sense the spectrum and use the band of frequencies in the television channel if the spectrum is unused. This paradigm can be extended to a network of secondary users coexisting with the primary users. Another classification of users may arise due to the differences in technology capability of the radio devices themselves. *Capable users* are those users that may have access to side information regarding the transmission of the noncapable users. The capable users can then make use of the side information to avoid interfering with the less capable users.

1.2.3 Usage scenarios for cognitive radio

Depending on the usage scenario, cognitive radio network operation can be classified as interweave, overlay, and underlay paradigms, as will be discussed further in Chapter 2. The *interweave* paradigm of operation was the original motivation for the idea of cognitive radio. The stringent requirement is that the secondary users should not interfere with the communication between the already active primary users. This mandates the secondary users to be able to detect (sense), with very high probability, the primary user transmissions in the network. Once the cognitive radio successfully detects the primary user transmissions, it can opportunistically communicate only if it is able to do so without harming the primary transmissions. This requires spectrum agility or the ability to transmit at different frequencies. The temporary space–time–frequency void in the transmission of primary users is referred to as a *spectrum hole* or a *white space*.

In the *overlay* paradigm, the secondary user needs to know the channel between the primary transmitter and the primary and secondary receivers as well as the channel between the secondary transmitter and the primary receiver. With the channel knowledge of both the primary and secondary users, the secondary user can then choose appropriate transmission strategies so that the communication in the secondary network causes least interference to the primary network. This paradigm represents an advanced operation by a highly sophisticated radio and associated architecture and poses many challenges. In the *underlay* paradigm, the secondary transmitter keeps the interference levels below a certain threshold. The primary receiver sees a higher noise level if the primary and secondary transmission overlap in the same band. An example system for the underlay paradigm is the ultrawideband (UWB) transmissions [22], where the transmissions are spread over a wider band to achieve power spectrum densities below the noise floor. While the primary transmission is decoded with an enhanced noise floor, the secondary receiver despreads the data to decode the transmissions.

1.2.4 Cognitive radio bands

In order to design efficient networks of cognitive radios, the propagation environment of these networks must be well understood. The knowledge of propagation will also help in identification, design, implementation, and analysis of transmission strategies for cognitive radios. A quantity that needs to be kept in mind by the cognitive radio system designer is the amount of interference the cognitive radio creates at the primary (or victim) receiver. This quantity depends on the transmission power of the cognitive radio as well as the wireless channel through which the cognitive radio signals propagate. The channel characteristics depend on the frequency band of operation as well as other properties of the environment, as discussed in detail in Chapter 3.

Cognitive radios may be deployed over a wide range of the frequency spectrum. The bands below about 3.5 GHz have lower propagation loss and are sought after by all services. These bands are therefore ideal (but not necessarily exclusive) candidates for the deployment of cognitive radio networks. They have different incumbent (primary) systems, each with its own mix of service type, architecture, bandwidth, and resilience to interference. Some typical candidate bands for cognitive radio systems are:

- *UHF bands*. These bands are currently used by broadcast television, though some conversion to wireless broadband services is in progress. Terrestrial broadcasting transmitters tend to have high antennas (hundreds of meters) and large powers (kilowatt range). In this service, the transmission is one-way, the transmitting antenna may be outside the area containing the cognitive radios and the TV customers are generally fixed. In 2010, the FCC adopted rules to allow unlicensed radio transmitters to operate in the broadcast television spectrum at locations where the spectrum is not being used by the licensed services [36]. The unused TV spectrum can be used as white spaces. This might be one of the first spectrum ranges where innovative products and services using cognitive radio systems may appear.

- *Cellular bands*. Typical cellular bands are centered near 800/900 MHz, 1.8/1.9 GHz, 2.1 GHz, 2.3 GHz, and 2.5 GHz. The International Mobile Telecommunications-Advanced (IMT-Advanced) systems (fourth-generation cellular systems as defined by the International Telecommunications Union (ITU)) may also be deployed in the 3.5 GHz band. Cellular networks have ubiquitous coverage, with cell site antennas mounted typically at rooftop or lamp-post height. This service is two way, with the cell sites generally in the same region as the cognitive radios, and the cellular customers can be mobile.
- *Fixed wireless access bands*. These bands provide two-way broadband service and are centered near 2.5 and 3.5 GHz. Fixed wireless systems are similar in layout to cellular networks, with the customers at fixed locations, like homes and businesses.

1.3 Spectrum policy: present and future

1.3.1 Role of spectrum policy

Efficient regimes for spectrum management have been a research focus since the earliest days of radio communications, but the mix of technologists, lawyers, and economists that has emerged in recent years has produced a new and lively debate. The long employed command-and-control policy [17] of the FCC has thus tended toward static long-term exclusivity of spectrum use via licenses in large geographic areas, often based on the radio technologies employed at the time of decision making. This has led to many successful applications like television broadcasting and cellular networks, which can be cited as evidence by the proponents of spectrum property rights, but has also been criticized as inefficient in the overall use of spectrum. In addition to the static nature of such licensed spectrum allocations, the inherent political and nonpolitical inefficiencies of government controllers also play a role in the poor spectrum utilization achieved [14].

In addition to the licensed bands, some spectrum is set aside in specific frequency bands that can be used without license by radios following certain rules of operation, such as a maximum power per Hertz of a shared channel access mechanism. The FCC mandates strict operational requirements to be followed in the unlicensed bands like the industrial, scientific, and medical (ISM) radio bands, where devices must conform to the FCC Part 15 regulation [66]. The purpose of these unlicensed bands is to encourage innovation without the high cost to entry associated, for example, with purchasing licensed spectrum through auctions. The success of applications in the unlicensed bands (cordless telephony and WiFi being well-known examples) has sparked a hot debate regarding the spectrum governance employed by the FCC, and how spectrum policy can be improved to alleviate artificial spectrum scarcity, promote efficiency, and also encourage innovation.

The proposals for new governance regimes fall into two broad categories: spectrum property rights and spectrum commons. In its broadest sense, spectrum property rights refers to a governance mechanism in which portions of spectrum are owned by individuals (or companies). Such portions can be traded to other parties through monetary transactions, or used exclusively, in a flexible manner, with not many technical constraints. The spectrum commons approach, on the other hand, advocates that spectrum should be considered common property, shared by all communicating parties, based on predefined but minimal rules or standards.

The spectrum property rights approach is motivated by the landmark work of R. H. Coase [26], in which it is suggested that spectrum can be treated like land, and private ownership of spectrum is viable. The proponents of spectrum ownership believe that the spectrum should be allocated to the prospective spectrum holders through market forces. The spectrum holders would then have exclusive use of the spectrum portion they possess, without the potential of harmful interference from other parties. Alternatively, they would be able to trade their spectrum in a secondary market. The use of spectrum would be flexible, in that the authorized party could use the spectrum portion for any purpose. Thus, the focus in this approach is on transferring ownership of the spectrum from the government to private parties and substituting market forces for traditional spectrum regulation, overcoming two sources of inefficiency in the status quo regime. The common view is that, since the early 1990s, the FCC has chosen a partial implementation of this approach by employing spectrum auctions as a means of licensing.

The spectrum commons approach, encouraged by the unlicensed spectrum band experiments, argues that as smart technologies evolve, communicating devices will become able to avoid interference through mutual cooperation and coexistence, and the spectrum will become less scarce. The emergence of cognitive and software defined radio concepts, multiple antenna and multicarrier techniques, UWB technologies, and mesh network topologies provide a technology panacea that proponents of this approach use to support their arguments. Communicating devices will be able to efficiently share a specified spectrum band through the enforcement of technical restrictions and multiple access protocols, without requiring exclusive access or private ownership.

The analogy often articulated is that of a highway, which the motorists treat as a common property and can efficiently share as long as they abide by the traffic rules. The highway analogy also illustrates that in spite of all the smart radio technologies, there is still a need for a controller or enforcer. Thus, even the commons regime [43] is a form of lightly controlled shared access [14]. Even though the generic descriptions of the two proposals seem clear, the lack of precise modeling creates many unanswered questions regarding the details of implementation. The exact nature of the controller or enforcer mechanisms in both models is vaguely defined. The government's role in managing controlled access in a spectrum commons regime is not clear. This lack of clarity also pervades the many issues related to transferability and duration of transmission rights, transactions costs, and the specific mechanisms involved in allocating the spectrum when needed. It is not clear, for example, how often transmission rights are anticipated to change hands in a spectrum property rights model.

Cognitive radio technology, along with appropriate architecture and infrastructure support, holds the promise of bridging the gap between the above regimes of spectrum governance. As discussed in [49], cognitive radios can enable spectrum governance regimes that support both exclusivity of property rights and the dynamic nature of shared managed access to a spectrum commons. Thus the advent of cognitive radio technology and its innovative ability to use spectrum opportunistically suggests a renewed look at and possibly guidelines for shaping spectrum governance in the future.

1.4 Data explosion: future spectrum implications

There is yet another reason for evolving spectrum policy in the context of cognitive radio technology. The convergence of voice and data in wireless communications, triggered by the convergence of wireless and the Internet, has led to an explosion in the number of bits transmitted over the air in the first decade of the twenty-first century. With the number of mobile users steadily increasing, the number of wireless devices has increased along with mobile traffic in the past few years. With more applications being developed every day and convergence of applications in one mobile device, the mobile traffic is expected to grow exponentially over the next generation of wireless devices. According to a study by Cisco [23], global mobile data traffic in 2010 (237 petabytes per month)[1] was over three times greater than the total global Internet traffic in 2000 (75 petabytes per month). Future projections as shown in Fig. 1.1 suggest that this explosive growth will continue well into the foreseeable future. According to Cisco reports, laptops and netbooks will continue to generate a disproportionate amount of traffic, but new device categories such as machine-to-machine and tablets will begin to account for a significant portion of the traffic by 2015. In fact wireless communications devices themselves have seen an explosive growth from about 1 billion devices in 2005 to about 10 billion devices in 2010 with projections of about 100 billion devices by 2020. In order to address the bandwidth demand of these mobile data services, wireless

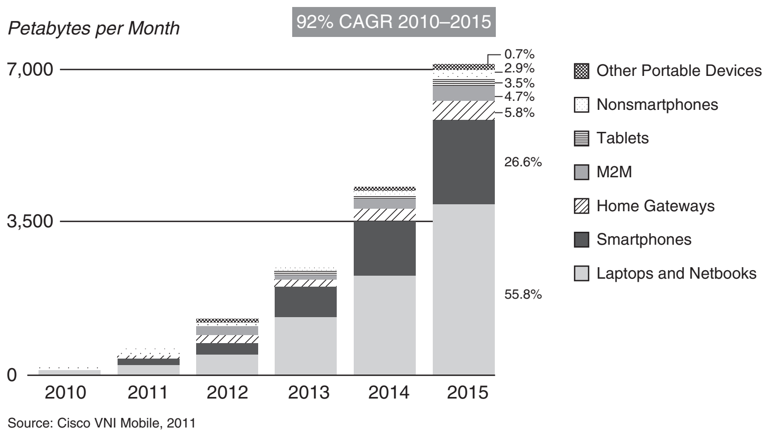

Figure 1.1 Forecast of growth of mobile traffic in the five years from, 2010 to 2015.

[1] A petabyte is 10^{15} bytes.

networks need to evolve so that they can handle bandwidth-hungry applications. Cognitive radio technology has the potential to enable opportunistic spectrum access in these devices and can mitigate the spectrum sharing problems posed in dense environments. Some example applications of cognitive radios are given in Section 1.5.

In future wireless systems, cognitive radio technology could serve as the foundation on which pervasive systems that solve a myriad problems can be built. Consider for example, that vehicular collisions kill forty thousand people in the US annually, and injure almost three million more. Almost one hundred thousand people die every year from medical mistakes, with a large number of these instances occurring in emergency rooms. Terror attacks and natural disasters stress our response systems, leading to unnecessary casualties and property loss and often putting responders at risk. Each of these distinct problems is complex in both technological and human terms; yet surprisingly, engineered solutions to these problems share several common attributes and challenges. Vehicles could communicate with each other and with roads and infrastructure to improve safety and manage traffic. Embedded sensors could remotely monitor patients and help medical staff manage critical information in hospital emergency rooms. Emergency workers could increase their effectiveness and personal security through wirelessly connected networks of robots and unmanned aerial vehicles. All these solutions require embedded wireless devices, allowing machines and people to interact within a local area and across the global Internet.

These solutions will be implemented in the form of pervasively deployed wireless ecosystems, which are large-scale, heterogeneous, and decentralized physical world deployments of networked cognitive radio devices including sensors, actuators, and machines, with humans in the loop. While tapping the potential of pervasive wireless ecosystems will require solutions involving general-purpose network architectures for heterogeneous mobile/sensor devices, robust computing models that include programming of the physical world, cyber-physical control, and human-centered design, fundamentally they will have to rely on spectrum coexistence in dense environments.

Wireless ecosystems require efficient and decentralized spectrum sharing in extremely dense populations of wireless devices, embedded and otherwise. Cognitive radio technology will be a cornerstone for successful spectrum sharing in these situations. Such radio nodes could sense the usage of radio spectrum across a wide range of frequencies and then make decisions about how to transmit. These decisions may be as simple as frequency band selection, or they may be considerably more complex, involving transmitter waveform design, modulation and coding formats, and even cooperation methods. The power of this transforming radio technology is even a point of agreement between the two sides of the contentious spectrum debate among regulators, economists, and engineers over spectrum policy [49].

1.5 Applications of cognitive radio

The most common application of cognitive radio is in the TV white spaces, where cognitive-radio-enabled secondary users opportunistically utilize the unused spectrum without interfering with primary incumbents of spectrum, namely TV transmitters. Cognitive radio devices can also form a secondary network in such bands where all secondary users are required to detect white spaces or spectrum holes. Cognitive radio technology finds many applications beyond TV white spaces. Over short ranges, cognitive radios find applications in wireless LANs, e.g., IEEE 802.11 family of standards, vehicle-to-vehicle communications over a range of 100 meters where data rates can be on the order of a few tens of Mbps. In the mid-range of distances spanning orders of a kilometer, cognitive radios can be used to deploy an extended range wireless LAN spanning hundreds of meters [8]. In the long range scenario, cognitive radio networks can use white spaces to meet the spectrum requirements of networks for public safety. Each of these scenarios can be classified into one of the three usage paradigms discussed earlier.

1.5.1 Dynamic spectrum access in cellular systems

With the advent of smart phones, cellular networks are undergoing rapid growth and are striving to improve spectral efficiency required to cope with the high bandwidth demands. While physical layer techniques have matured a lot in enabling spectrally efficient bit pipes, at the system level, the need for spectrally efficient transmission techniques still remains. Next-generation cellular systems are going to be more heterogeneous, with small cells embedded in large macrocells. These embedded picocells or femtocells are miniaturized versions of base stations transmitting at low power. They can form small coverage islands inside a macrocell coverage region to provide coverage and capacity to users in those coverage islands. Such coverage islands could be hot-spot areas like a shopping mall, or a sports arena. Small cells provide more cell-splitting gains, when they share the same spectrum as the entire macrocell [30]. Although the above methods require infrastructure support, schemes like DSA can exploit the unused spectrum to provide better spectral efficiency gains. Cognitive radio technology can be used to estimate the spatio-temporal characteristics of the traffic and spectrum usage in the macrocell and opportunistically use the available spectrum during periods of high data demand. Dynamic spectrum access networks can work in the interweave paradigm where the cognitive radios can detect white spaces to opportunistically transmit in that spectrum. It can also work in the overlay paradigm if overhead signaling through the backhaul can provide side information to the cognitive transmitters.

1.5.2 Cellular data boost

As a further extension of DSA in cellular networks, cognitive radios can possibly be used to opportunistically offload traffic to white spaces to provide a "data boost" to the existing cellular network. For an example scenario shown in Fig. 1.2, cognitive radios can help alleviate high loads on the cellular network. In the example scenario, an overlay mesh network of white space hot-spots can carry non-real-time or delay tolerant data like mail, content, file transfer, etc. The hot-spots can operate either in the licensed or unlicensed band. Off-loading part of the delay tolerant data traffic from the cellular network helps operators to meet the quality of service (QoS) requirements of delay

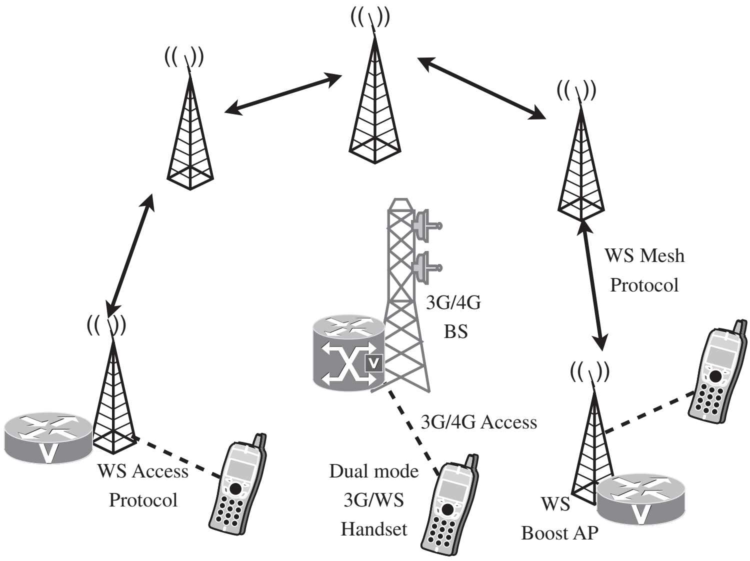

Figure 1.2 A cellular data boost network for off-loading fast-growing cellular traffic using dual-mode radio.

sensitive traffic like voice, streaming video, etc. This approach could lead to capacity boost of two to five times the existing cellular capacity, depending on the percentage coverage and service mix of the data generated in the area. The boost in the capacity comes at no additional cost, but from the detection of available white spaces and the opportunistic transmission over white spaces using the hot-spot mesh network. The end user handsets are required to support dual-mode radios equipped with the capability to transmit in the cellular band as well as in white spaces. This is an example of the interweave paradigm of cognitive radio operation.

1.5.3 Machine-to-machine communications

Machine-to-machine (M2M) communications [90] enable mechanical automation of ubiquitous sensors. The demand for wireless personal communications for exchange of voice, audio, video, emails, and photos over ad hoc networks and infrastructure-based cellular networks have empowered full automation of sensor networks. Cognitive sensor networks can be used for M2M communications in many ways. The communication could be between the sensor and the decision maker (as in meters/monitors reporting data to a decision maker), among multiple calculation agents within the decision maker (as in cloud computing) or between the decision maker and the action executer. In the future, multiple radio access networks may be integrated and managed as part of a single hierarchical network. Cognitive radios can enable efficient air interface design of M2M networks. Cognitive radios could opportunistically exploit the additional spectrum and connectivity available in different networks to improve the system capacity and device quality of service. Cognitive-radio-enabled devices can substitute