The Next Decade of Wireless

Dr. Hossein Eslambolchi

March 1, 2012

Just as the Internet liberated communications for everyone, the rapidly emerging broadband technology continues the revolution. With its potential for sharing virtually any type of information instantaneously, broadband fuels the creation of new applications almost on a daily basis.

As a result, the *sociology of information* itself – how we view and use information – is being transformed. Information is no longer passively confined to libraries and other repositories. Information resides, thrives, grows everywhere and can be accessed from anywhere.

The Internet – now enhanced with broadband technology – enables anyone to access the mass of information represented by everyone's knowledge and experiences. The combination is likely to propel the Internet into being the most useful network ever devised. But reaching that potential demands that the technology and the environment assert greater control of the information – retrieving, packaging, and directing the right information, at the appropriate time, to the expectant recipients wherever they are.

Wireless demonstrated that the benefit of having more information is proportional to how conveniently and effectively users can integrate that information with their daily lives. Simply receiving massive information may not be as valuable as receiving the appropriate information when and where it is most applicable, and in the most convenient, most intuitively used form. Accordingly, machine-to-machine communication can be anticipated to increase, freeing human beings to concentrate on the essence of information rather than the detail.   Cost and simplicity dictate than much of this machine communication will be conducted wirelessly as well.

Real-time and location context is the ultimate in information asset value, and wireless networks are unique in their ability to recognize the presence of communications devices and their owners. Wireless systems can "find" users – if they desire – to provide information with or without active user participation. Based on the immediacy and convenience of untethered mobility, there is little doubt that the last link to individuals will be broadband wireless.

Networks of the past, such as the Public Switched Telephone Network (PSTN), were built top-down at enormous expense, with devices dictated by the network provider and the universal service concept. Broadband networks of the future are more likely to evolve as partnerships of users and service providers.

Think of the broadband process as a handshake between users and networks. The users reach toward the network and network reaches toward users. Today's demanding shoppers, presented with a myriad of choices, are overwhelmingly choosing what works for them, what fits their lives. They are voting with their pocketbooks for wireless access devices. They are investing their own capital in completing the connection. Examples include Wi-Fi devices, business WLANs, home networks, and, of course, cellular.

Meanwhile, networking is becoming a competitive enterprise, which implies that construction of newer networks will be based on user demand in specific areas. Competition, user preference, and bandwidth accessibility are likely to fuel consolidation of wireless systems into a few types: a dominant small-call broadband access alternative, a large cell medium bandwidth alternative, and a low bandwidth nationwide (or global) alternative.

Wireless networks will evolve to become "tiered" networks of networks, like the Internet, with Personal Area networks, connected to Local Area Networks, connected to Metropolitan Area Networks, and hence to high speed backbones. Separate networks will also persist, such as cellular systems, providing a wide variety of value propositions and service price points for consumers and businesses::

- High Bandwidth Wireless Networks – These networks will fulfill the need for high bandwidth, full-multimedia communications. Small cells will dominate, except for directional, long distance links due to limitations of spectrum, battery, and radio frequency (RF) propagation. Enhancements will be driven by increasing radio sophistication based on the 802.11/4G framework. Largely automatic network formation, maintenance, and upgrades will minimize staffing and balance the cost of many more cells than today's large-cell systems. Wireless devices will probably be consolidated. Multi-mode devices will offer a window to each access alternative. Detecting and switching between systems will be done first by multiple radios and eventually by "software" radio that "computes" receiver / transmitter functions, while providing enhanced multi-service awareness. Able to operate on all spectrum bands and complementing cognitive radio algorithms, such digital radio processing will make it easier to replicate radio architectures so signals can be sent and received simultaneously by multiple antennas. Such multiple-input, multiple-output systems (MIMO) can extend range and throughput of 4G small-cell and large-cell systems, improving performance and economy of services.

- Medium Bandwidth Wireless Networks – A fusion of 3G cellular and wide area packet systems such as IEEE 802.16e will evolve to a packet-oriented, vehicular-speed mobile 3G successor that will undoubtedly expand to "fill in" areas not covered by 4G service, but probably not everywhere. These systems will fulfill the applications envisioned for 3G, but at lower cost, improved performance, and full compatibility with IP backbone networks. Their ability to also provide longer distance fixed links as well as mobility will enable use for backhaul of 4G small-cell systems and other broadband metropolitan trunking and distribution applications.
- Narrow Bandwidth Wireless Networks – Will probably continue as improved versions of 2.5G, such as high-speed ANSI-95 and GSM extensions, providing low-cost, lower bandwidth services to voice handsets and enriched "slow multimedia" devices. These 2.5G extensions will provide coverage almost everywhere, and will continue offer basic communication services until medium bandwidth and 4G systems are fully established.

What Is Cognitive Radio?

Like many modern concepts, "Cognitive Radio" is not entirely new, but rather an amalgam of manual radio resource management techniques invented and proven earlier for enhancing the performance of wireless communication systems, now automated by a computer.

Simply put, cognitive radio is the combination of a radio transceiver with computerized intelligence to automate coordination of devices, networks, and services for improved functionality, interoperation, and spectrum utilization. Recently a focal point has become coexistence of radios in common spectrum.
It is termed "cognitive" because the actions it is envisioned to take are analogous to the way a biological entity reacts to its environment based on a regime of goals, sensory inputs, and behaviors.

Cognitive radio has become a topic of interest lately because of two important changes in the radio application space:

1. The growth of personal portable communications services
2. The need to provide enough suitable spectrum for such services

Personal communications has changed the face of two-way radio from a microwave relay, push-to-talk, and mobile telephone world to mainstream, anytime, anywhere wireless for the masses. Cellular telephone started this revolution in the 1980s, but today's broadband multimedia wireless networks have carried the banner forward. Even at its inception, the very nature of cellular radio dictated some level of cognition: The systems depended upon devices to detect an available wireless network, interact with it to allow a connection to be set up, and to maintain the connection as the user moves about. It is fortunate indeed that cellular and microcomputer technology arose contemporaneously, as much of what cellular telephony has become would have been impossible otherwise. In today's advanced radio networks, even more is expected: wireless devices may actually be parts of the network as well as clients (as with forwarding mesh topologies), they may have to accommodate a variety of media, rates, and QoS needs, and they may have to operate in spectrum that is shared with other services while minimizing interference. They may even have to recognize exactly where they are and to notify other devices or networks.

The Origins of Cognitive Radio

Having begun almost 100 years after the development of wired technologies, radio has traversed several stages of maturation, the latest of which is exploitation of the signal processing techniques that had proved so successful in making broadband transmission possible using wires. However adoption of these techniques has also forced radio engineers to reconcile with Shannon: Information theory increases consciousness of the bandwidth resource, link attenuation, power, and noise floor as fundamental limits of reach and rate. This entry into the “Shannon Zone” has also

sparked the realization that spectrum is finite and not all of it is necessarily appropriate for every application.

The realization that spectrum is finite is easily visualized if one thinks of a wireless “cell” as a coaxial cable of the same diameter. Individuals within the cell/cable must share the RF medium efficiently with others while minimizing interference. With wires, one can always add more cables to fixed-use locations because the spectrum is contained and can be reused almost infinitely; with wireless the spectrum is much less contained and can be reused only to the extent smaller cells can be used. Even with small cells, the spectrum resource may have to be divided among a significant number of users within the coverage area. With appetite for wireless broadband connections increasing, cells must inevitably become smaller with better resource reuse and higher channel spectral efficiency absent vast amounts of new spectrum. In the end, one might envision a “Shannon Communication Volume” (SCV) expressing the maximum number of users at a particular rate that can be sustained over a geographic area in much the same way as the Shannon limit is now routinely used to bound the “reach” and “rate” of transmission over wires.

The Communications Act of 1934 established the FCC as the manager of spectrum for citizens of the United States and a regime of licenses, regulations, and process thereafter ensued to ensure that spectrum use was organized, productive, and efficient. In a sense, the FCC’s goal has been to maximize the SCV of the United States by building rules for various radio services one-at-a-time into a large body of regulations.

Smart Radios

Over the past quarter-century the marriage of wireless and inexpensive computing has allowed “smart” radios to operate within the “one-at-a-time” FCC rules to make spectrum work harder. Since smart radios can do more than simply receive and transmit radio signals---they can also measure them and respond to propagation conditions or availability of known services---they are at the very heart of modern radio resource reuse concepts. Many of these capabilities were pioneered by 1st generation analog cellular systems, wherein phones would search for an “overhead” channel, adjust their power level according to information provided by the base station, and then “register” with the system to use it. Because intelligence was costly then, much of it was centralized at the MSC.

As microcomputer technology became increasingly capable and economical, more and more of the intelligence migrated into the terminal. Soon, transmit power was adjusted dynamically according to received signal strength. Still later, Mobile Assisted Channel Assignment (MACA) and Mobile Assisted Hand-Off (MAHO) used signal strength information collected by the mobile radio itself to determine the best serving channel or handoff target. Likewise, early digital cordless telephones could detect interference at the handset or base and retune or change hopping sequences to improve operation.

In spite of the ability to react to signal strength, even 2G digital cellular systems were designed for only one connection type: voice telephony. Accordingly, they could deliver only one “lowest common denominator” service level, that which could be sustained at the edges of a cell. This constraint is the basis of today’s 2.5G cellular service, even though “data” packets may now be transmitted on a “voice” channel. ›G” cellular and 802.16/WiMAX smart radios have moved to the next step: Link adaptation. Increasing experience with digital transmission and use of more sophisticated modulation, coding, and antenna techniques have taught that if users are close to the base station they can probably sustain higher rates than those at the cell edge. Using improved link measurements, the radio can adjust or “gearshift” its rate, coding, and spatial signal combining to optimize performance for the particular propagation environment in which it finds itself. This improved “cognition” is an integral part of 3rd generation systems, and the adaptation has lately extended to advanced antenna technologies such as MIMO and beam forming, providing a new kind of intelligence---“smart antennas”--- for smart radios.

The Smart MAC Radio

In the 1980s, a new standards creation strategy was popularized: the ISO OSI transmission model. This model sought to break a communications path into individually specified functional “layers”, each with identified inputs and outputs to other layers or to the transmission medium itself.

Prior to the ISO model, most radios were designed holistically with all functions interoperating broadly and capable of sharing information freely. Most mainstream wireless communication systems (e.g. cellular and cordless) also used frequency-division multiple access and frequency duplexing to share the radio resource in the spectral domain. The “paired” nature of spectrum allocated for these services made use of arbitrary spectrum impractical.

The ISO model “compartmentalized” intelligence, and specified the format of all messages that traversed layers, in effect producing “smart layers”. It simplified standards setting and device interoperability, but limited the ability to communicate the products of each layer’s intelligence to other layers. With physical layer (PHY) transmission “machinery” in place and control messages specified, most intelligence eventually rose to the MAC layer. The flexibility of packet transmission combined with the ISO model, the “smart MAC”, and the reduction-to-practice of time-division duplexing opened the field for communication services in non-paired spectrum. Now it was practical to move services to new bands on an opportunistic basis.

802.11 was the first standardized packet radio framework to capitalize on the opportunity and simultaneously address the problem of allowing users to share a common channel asynchronously, and its appearance was just in time to “catch the wave" of Internet popularity. It borrowed a technique for mediating transmissions adopted from Ethernet, which had previously capitalized on the shared-medium Packet upsurge: Carrier Sense Multiple Access – Collision Detection (CSMA-CD). However since 802.11 targeted not a wired application, but rather wireless, collisions could not always be detected as with Ethernet. Accordingly, the protocol was augmented to implement Collision Avoidance (CSMA-CA) instead. CSMA-CA uses Request-to-Send/Clear-to-Send messages to effect "reservation" of the medium that discourages access by other stations until the sending station has completed its Packet transmission.

CSMA itself arose from earlier ALOHA protocols developed to allow packet transmissions within a network of radio stations with completely decentralized control (very unlike cellular systems). With the simplest ALOHA protocol a station completely transmits its entire packet when it arrives. If a collision happens with another station that has also transmitted (signaled by not hearing a repeat of the message from the intended receiver), the station retransmits the entire collision-damaged frame after a random delay. ALOHA could be termed "strongly challenged" on the cognition scale, and its efficiency was low. Various schemes (including "slotted" ALOHA) subsequently improved its performance somewhat at the cost of synchronization. However with many active users, ALOHA networks experience "congestion collapse" where throughput (and system efficiency) may be reduced to near zero.

Simply put, CSMA extended ALOHA by implementing true spectral etiquette for packets including "congestion awareness": if a device has a packet to send, it listens to see if the channel is in use. If not, it sends the packet. If it fails to receive an acknowledgement, it tries again after a congestion-modulated random delay to send a duplicate of the damaged packet.

In adopting CSMA for Media Access Control, 802.11 developed not only a reasonably efficient means of utilizing an isolated cell's capacity, but also created a simplistic method of sharing the radio resource between groups of ‛embedded" cells using the same frequency whose coverage areas might overlap. The multi-cell sharing property was enabled by the so-called "Clear Channel Assessment" function that determines whether the MAC can transmit with reasonable assurance that it will not harm transmissions already in progress. CSMA enabled easy assembly of networks that could cover larger areas than a single WLAN cell, and paved the way for "do-it-yourself" networking. Use of CSMA for cognitive spectrum use is now lumped under the heading of "contention based sharing techniques".

Private do-it-yourself wireless LANs capitalized on the contention-based protocol advantage, while contributing another biological-like behavior to the cognitive radio vision: the "organic" growth model where a network of small cells expands coverage and capacity by "birth" of new "smart" base stations that "bond" to the old ones fueled by the presence of nutrients (e.g. users, utility, investment, etc.).   The potential of such “self-organizing” networks is a fundamental driver of 4th generation broadband wireless thinking, since the sheer number of cells required to provide true Ethernet rates and contiguous coverage to large areas mandates a distributed-intelligence control architecture for scalability and cost reasons. But perhaps equally importantly, the small-cell precept also makes very dense radio resource reuse practical, particularly in spectrum that is either unlicensed or licensed in less-rigid ways. Although WiFi networks covering large areas (even entire cities) have lately become fashionable, there is an important limitation that must be overcome before such networks can truly be considered broadband multimedia systems. This is because 802.11’s implementation of CSMA works well for bursty, non-time-sensitive data packets, but not so well for streaming real-time services such as VoIP. The reason is that the latency CSMA systems develop as they become more heavily loaded can cause packets to be received too late for decoders to process them as part of the stream reconstruction process, causing loss of quality.

802.11 addressed this shortcoming by creating the 802.11e standard, which establishes two new operating modes to support QoS-dependent applications. The most simplistic conceptually is EDCA, which allows QoS clients to use shorter back-off times. The approach provides “prioritized” QoS in lightly-loaded systems, but cannot guarantee QoS when systems become more heavily loaded. The second mode, HCCA, is capable of providing scheduled radio resource management by suspending CSMA operation in part of the 802.11 super-frame, while instituting a QoS specification, scheduling, and polling process that provides reserved “Transmit Opportunities” for clients requesting “parameterized”, carrier-grade, QoS. Since operation in the contention-controlled portion of the frame provides a “protected” radio resource, a “smart scheduler” can materially improve system performance, not only for clients, but for better reuse in embedded cell environments.

The Smart-Scheduling Radio

The simplest definition of a wireless MAC is a protocol that tells radios when they can transmit. For networks where no centralized admission or management of clients is used --- so-called peer-to-peer operation---clients all transmit packets according to CSMA-like rules. However for “infrastructure-mode” networks in which Aps connect to a backbone, MACs operate differently according to whether the device is a client or an AP. As has been discussed above, 802.11e HCCA depends on strong scheduling coordination to minimize collisions, allocate TXOP lengths sufficient to maintain QoS for individual clients, and maintain high network resource efficiency. One important way this can be of value is associated with the so-called gear-shifting transmission behaviors already touched upon in previous sections.

Gear-shifting is a term used to describe adaptive behavior that tailors transmission rate to link quality. In a smart radio, the need to shift “gears” is usually triggered by examination of error rate and signal strength as a session proceeds. If error rate increases or signal strength decreases, the MAC responds by transmitting data more dependably, but also more slowly. If conditions improve, the MAC "shifts" to a higher transmission rate. Gear-shifting can be viewed as "local" adaptation, because it improves the performance of a particular client being served by the system. The consequence of the relationship between energy-per-bit and throughput with fixed transmitter power and a given link attenuation means that a "downshifted" client's application gets fewer packets over a period of time. The optimization allows that client to receive some packets as opposed to getting few or none at all. With bursty, best-effort data applications, the slower behavior causes applications to appear less responsive (e.g. Web pages will "paint" slower), as those who routinely use dial-up modems know. Absent gear-shifting however, long outages due to errors could cause TCP/IP protocol timeouts, possibly closing sessions.

But streaming real time traffic is a different situation. These streams connect applications that must produce synchronized time-critical media outputs at each end.
In QoS enabled systems, clients are admitted according to a TSPEC
that attempts to furnish a minimum rate with a maximum latency and jitter. Since longer bursts are required to maintain the same throughput following a "downshift", more time must be allocated for these clients to allow them to stay synchronized in real time.
In a distributed coordination system (such as EDCA) we may illustrate the effect of local optimization of one link upon other links, showing that gear-shifting can become a global resource allocation problem: A large number of QoS-bound streaming clients is being supported under stable conditions. One client moves away from the serving AP toward the edge of the coverage area.

The signal strength at the client and AP drop, and the error rate increases. Since at the edge of coverage, radio error rates rise strongly as the signal at the receiver becomes noisier, the AP attempts to compensate. It gear-shifts to accommodate the poorer conditions, transmitting longer bursts to the more distant client in an attempt to maintain TSPEC throughput.

The throughput for other clients must now be accommodated in less time, raising apparent system loading. Contention builds due to the higher loading rate, causing packet queues to lengthen and latency to rise. If more clients force downshifts or more clients enter the network, the AP may then not be able to support new streams, accept incoming handoffs, or continue to honor admitted QoS specifications.
With a point coordination system using a strong scheduler such as that associated with HCCA operation, clients are admitted only to a certain resource-use level.

Time is held in reserve to allow some QoS clients to downshift their sessions as circumstances require, while still allowing other clients to be admitted either as new sessions or due to handoffs from other cells. If the aggregate throughput in process reaches a pre-set policy limit, it is possible for an entering client to experience a "spectrum busy" indication. The client can respond by waiting and trying again, or generating a TSPEC with lower rate or higher latency.

In such an HCCA system, the smart scheduler becomes the heart of a smart MAC radio: it dynamically adjusts session TXOPs, admissions and the reserved resources to globally optimize network resource efficiency as well as locally optimizes individual client performance---up to a point. Since wireless is by definition untethered, it is possible that a client could request downshift that might require many times more TX-OP length than could reasonably be accommodated at the current network loading. If the amount of best-effort traffic is insufficient to support “capacity borrowing” without violating policy minimums for best-effort service, the downshift may have to be disallowed. Under such conditions, the scheduler may trigger a notification that the end of range has been reached, and that the stream cannot be supported if the link were to deteriorate further. Seen in this way, the smart MAC becomes the globally-aware arbiter of when downshifts are needed, if they can be accommodated, and the guarantor of QoS when the downshift is complete.

Dr. Hossein Eslambolchi

March 1, 2012