Moving Beyond 4G Wireless Systems

Jim Martin, Rahul Amin
School of Computing
Clemson University
Clemson, SC USA
Jim.martin@cs.clemson.edu

Ahmed Eltawil, Amr Hussien
Department of Electrical and Computer Engineering
University of California
Irvine, CA USA
aeltawil@uci.edu

Abstract—As broadband data further blends with cellular voice, mobile devices will become the dominant portals to the connected world. However current design practices still involve building independent networks that each make their own resource decisions. To move beyond 4G, wireless networks will need to cooperate to provide better performing and more robust wireless services. In this paper, we present an approach to wireless networking that requires several possibly contentious assumptions. First we assume that growing user demand and expectations for wireless data services will require wireless providers to cooperate with each other to improve coverage and performance. Second, we assume devices will have to be highly reconfigurable and capable of supporting multiple radio links concurrently. With these liberating assumptions, we present an architecture for a heterogeneous wireless system that moves beyond 4G systems. We present results from a MATLAB-based simulation designed to illustrate cost-benefits tradeoffs that are likely to exist as device reconfiguration capabilities increase. Finally, we identify the key research challenges that must be addressed to move the idea from a concept to a viable network system.

Keywords-heterogeneous wireless networks; reconfigurable radios; 4G wireless systems; cyberinfrastructure;

I. INTRODUCTION

While advances at both the network layer and the physical layer are prompting unprecedented mobile broadband growth, the means by which mobile terminals are offered services have created inefficiencies in both spectrum usage as well as mobile physical resources. The premise behind this issue is that while emerging devices support a multitude of wireless access methods, the current access methods require the user to select the active access network either by purchasing an appropriate handset (and service) or, in the case of multimodal smartphones, by manually selecting the access network. While this approach has worked fairly well for current 2G cellular networks, upcoming 3G and 4G systems will face significant challenges as wireless operators are held accountable for poor performance or inadequate coverage. It is well understood that individual systems that manage blocks of spectrum independently are inevitably operating at suboptimal performance.

Until recently, technology was the primary impediment to achieving universal, broadband wireless services that involve multiple radio access technologies. Today, the most significant impediments are the after-effects of antiquated government spectrum allocation policies and the resulting economic forces that drive the wireless industry. The effect is that in many geographic areas licensed spectrum is likely to be underutilized [1]. Techniques or paradigms such as cooperative communications, symbiotic networking, cognitive networking, and dynamic spectrum access attempt to improve spectral efficiency through cooperation at the radio level [2]-[8]. At the network level, architectures and frameworks to support hybrid or heterogeneous networks have been suggested [10]-[14]. Recent proposals have been based on the IEEE 802.21 standard which provides a framework to support vertical handoffs transferring a mobile user between two networks that are based on different radio access technologies [9].

Although emerging 4G networks embody many of these recent advances, current design practices still involve building independent networks. The economic forces that are driving the cellular industry are reducing the number of cellular providers but causing their wireless networks to become large, heterogeneous systems based on numerous cellular data technologies at various lifecycle stages. At the same time, users are requiring more bandwidth intensive services that might not exist in all coverage areas. On the other hand, 802.11 networks have proliferated to the point that they are now considered to be a part of the computing (cyber) infrastructure. To augment wireless connectivity, some organizations or agencies have deployed site-wide, campus-wide, or city-wide broadband wireless coverage. These trends and the subsequent impacts on end users are becoming commonplace. Using Clemson University as an example, depending on the specific location on campus, one can find multiple 802.11 networks (some operated by the University, some operated by local public safety, others operated by the adjoining city of Clemson or nearby businesses whose coverage spills out over areas of the campus), coverage from an experimental campus WiMAX network, and multiple 2G/3G networks from cellular providers including AT&T, Sprint, T-Mobile, and Verizon. Over the next year, the spectrum on campus will become further cluttered as cellular providers deploy 4G networks. It is important to note that while wireless technology is advancing rapidly, research that facilitates cooperation across independent networks is still immature.

In this paper we present an architecture for a wireless system that allows for a ‘clean-slate’ approach to the problem under the following assumptions:

1. Incentives are in place motivating independent, autonomous wireless systems (AWSs) to cooperate to
provide users a unified network with enhanced coverage and performance than could be achieved by any single AWS.

2. In any given geographic area, a handset might have access to a large number of AWSs. The use of multiple, concurrent radio links will become standard practice at least for a certain class of future smartphones.

3. Due to cost, size, and battery constraints, mobile nodes will be limited to a small number of adaptive radios that must be capable of operating in a relatively large number of supported communication modes and over a range of spectrum.

We consider a carrier-centric scenario where a service provider offers services for specific markets (e.g., commercial wireless access, solutions for public safety etc.). The carrier might own and operate portions of the physical network and possibly broker ‘peering’ arrangements with other wireless providers. The optimal selection of the current active mode (or modes) is performed at a co-operative network level. A centralized algorithm manages bandwidth and user resources at a global level to optimize network–wide metrics, while localized algorithms based on standards-based MAC protocols manage bandwidth over small time scales. The carrier initiates vertical handoffs that might reflect load balancing or cost optimization.

This paper is organized as follows. Section II presents and relates the relevant background and provides motivations for the work. We provide preliminary results from a MATLAB-based simulation in Section III. We then describe the key technical challenges that must be addressed to pave the way for future ‘beyond 4G’ wireless networks in Section IV. We conclude by describing our current project status and possible future directions in Section V.

II. BACKGROUND

Significant advances have been made in the last decade in radio technology. Recent work in cross-layer optimization has shown that breaking the rigid OSI layered stack model can enhance both spectral efficiency and application performance [2]. Increased efficiency can be obtained through the use of methods such as multiple input, multiple output (MIMO) techniques, relays and other forms of MAC/PHY layer cooperative communications [3]. Dynamic spectrum access (DSA) is another form of cooperation that is likely to be a foundation of future wireless systems [4]. DSA requires devices to be frequency agile giving rise to software defined radios (SDR) and cognitive radios (CR) as possible implementations. While SDRs and CRs have been under development for many years, they are limited to relatively narrowband systems due to multiple issues including power consumption and available computing cycles [5]. The challenge in handset design is managing the tradeoff between flexibility in how spectrum is used and in space/power requirements of the platform.

An extension to CR is cognitive networking which is defined in [6] as “a network with a cognitive process that can perceive current network conditions, and then plan, decide, and act on those conditions” [6]-[7]. One step further, symbiotic networking observes that current wireless networks try to achieve the best performance within their own network and generally neglects the impact on co-located wireless networks [8]. In other words, radio behaviors are generally selfish and based only on information observed locally. Symbiotic networks extend the scope of cooperation across all layers and network boundaries.

While SDRs, CRs, cognitive networking and symbiotic networking focus on improving efficiency from the bottom up, heterogeneous wireless networks represent methods for cooperation driven from the top down. The IEEE 802.21 standard provides a framework to support seamless mobility through networks based on different radio access technologies (RATs) without the need to restart the radio connection every time the mobile moves to a new network [9]. Another relevant standard, IEEE P1900.4, defines building blocks for enabling coordinated network-device distributed decision making which will aid in the optimization of radio resource usage, including spectrum access control, in heterogeneous wireless access networks [10].

A number of architectures to support heterogeneous wireless networks have been proposed in the literature. Hierarchical resource managers have been proposed by the Common Radio Resource Management, Joint Radio Resource Management and Multi-access Radio Resource Management schemes studied by the 3GPP group [11]-[13]. The overhead associated with a centralized hierarchical wireless system is studied in [14]. In these hierarchical schemes, and also in our proposed system, the local resource managers of different wireless technologies interact with a centralized entity to jointly optimize the process of resource allocation. To the best of our knowledge, we have found that the work in [12] is the only work that shares our design requirement that future heterogeneous wireless systems will involve reconfigurable devices. We take this a step further and explore a system that blends local and global resource management of wireless devices that can possibly operate over multiple wireless links at any given time.

III. SYSTEM DESCRIPTION

A. System Model

Figure 1 illustrates the system model. The system consists of devices (also referred to as nodes) that have connectivity to one or more AWSs. Each AWS will have a controller (referred to as an AWS controller) that represents all nodes in the AWS and that serves as a gateway connecting the AWS with other AWSs or external networks. A global resource controller (GRC) manages resources in a manner that supplements decisions made by local AWSs.

Users are presented with a unified network. Depending on node capabilities, users can operate over more than one AWS at any given time. The radio link block pictured in Figure 1 represents the MAC and physical layer that operates over a portion of the spectrum. A radio would be implemented using a combination of custom hardware along with programmable hardware based on technologies such as field programmable gate arrays (FPGAs) or digital signal processors (DSPs). User data is tunneled over the unified network cloud. The GRC (or
another located in the backend network) represents the termination point for the tunnel. Nodes must maintain periodic contact with the GRC by sending status update information periodically. The GRC sends network level status and/or resource management control information periodically to the AWSC (and possibly to individual devices).

Figure 1 System Model

B. Radio Capabilities

Radios are either static or capable of reconfiguration. Static radios are equipped with one or more non-reconfigurable radios. A non-reconfigurable radio supports a limited level of adaptive capability, but provides the lowest power consumption due to its custom nature. The position of mobile devices will be tracked by the GRC as location-based management is required. An AWS will likely consist of multiple wireless networks, generally of the same radio access technology. For example, a particular 3G network or a campus-wide 802.11 network would be considered an AWS. The GRC is necessary to facilitate cooperation across AWSs. For example, a node that concurrently supports 802.11 and 3G might be told by the GRC to use the 3G link as the control channel and the 802.11 network as the data channel. Alternatively, if the same node moves to a location that has 802.16e coverage, the GRC might initiate a ‘reconfiguration handoff’. For example, as the node moves out of coverage of the 802.11 network and the GRC determines that the node is in coverage of the 802.16e network, the network could issue a reconfiguration command to the node instructing it to reconfigure the radio to 802.16e. A reconfiguration handoff is a vertical handoff that requires a radio to reconfigure itself. The implicit assumption here is that due to the multimodal nature required, radios will support reconfiguration rather than have several chipsets, each dedicated to a specific standard.

Reconfigurable architectures span the gamut from general purpose processors to multicore DSPs, with application specific instruction set processors presenting a good compromise between processing power and computational resources. Chip integration densities allow integrating a large number of cores on a single die, however, the issue of programming and managing these multi-cores in an efficient manner that meets real time processing requirements is still an open research issue. On the other hand, having dedicated chipsets for each individual standard becomes both cost and power prohibitive when the number of supported standards increases. An alternative approach that is gaining momentum due to the massive integration of transistors in advanced technology nodes is the potential use of FPAGs as programmable mobile platforms. Reconfiguration of the FPGA fabric allows for multi-modal support, while real-time performance metrics are easy to achieve due to the hardware acceleration of computational intensive tasks, as well as the massive parallelism achievable by the FPGA architecture. In fact, several FPGA manufacturers are already proposing low power versions of their current FPGA platforms specifically for that reason.

In the remainder of the paper, we investigate using an FPGA platform as a platform for reconfigurable radios. While currently available FPGA’s are still considered too power hungry to be used as mobile chips, the preceding argument indicates that there is strong push towards making this a reality in the near future. The intention is not to argue that FPAGs are the best platform for reconfigurable computing, but rather to use them as an exemplar to demonstrate the impact of reconfiguration in terms of network performance improvement, as well as estimate the impact on throughput and power consumption. Table 1 presents implementation and performance statistics for common wireless access technologies. The table is by no means representative of the vast amount of architectures available in literature but is intended to extract a measure of the complexity in terms of area (measured in Kilo gate equivalents of a simple 2 input, drive one, NAND gate), as well as power consumption.

Table 1 Implementation statistics for current access technologies

<table>
<thead>
<tr>
<th></th>
<th>802.11a</th>
<th>Mobile WiMAX</th>
<th>LTE</th>
<th>RISDPA</th>
<th>EVDO (est. from UNTS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference</td>
<td>[17]</td>
<td>[18]</td>
<td>[19]</td>
<td>[20]</td>
<td>[21]</td>
</tr>
<tr>
<td>No of K gates</td>
<td>416</td>
<td>728</td>
<td>270</td>
<td>723*</td>
<td>684</td>
</tr>
<tr>
<td>No of DPCs</td>
<td>31</td>
<td>53</td>
<td>50</td>
<td>53</td>
<td>50</td>
</tr>
<tr>
<td>$P_{dio}(W)$ @ 100 MHZ</td>
<td>1.76</td>
<td>3</td>
<td>1.13</td>
<td>3</td>
<td>2.83</td>
</tr>
<tr>
<td>$P_{rec}(W)$ @ 50 MHZ</td>
<td>7.25</td>
<td>12.4</td>
<td>4.68</td>
<td>12.4</td>
<td>11.7</td>
</tr>
<tr>
<td>$T_{dio}(ms)$</td>
<td>19.53</td>
<td>33.4</td>
<td>12.6</td>
<td>33.4</td>
<td>31.5</td>
</tr>
</tbody>
</table>

*(est. from area and technology)

The power consumption is categorized as dynamic power consumption ($P_{dio}$), which is consumed during regular circuit operation, and reconfiguration power ($P_{rec}$) which is consumed when the circuit is reconfigured to a new standard. The reconfiguration power and time ($T_{dio}$) are estimated based on the complexity of the standard, where the minimum reconfigurable block is defined as a data path container (DPC=13.5 Kilo gate equivalent). Based on [15], the average reconfiguration power for each DPC (for a Xilinx Virtex II) is 234 mW and the average reconfiguration time is 0.63 ms, while the dynamic power is estimated from [16]. Thus, the average total power consumption can be calculated as follows:
\[ P_{\text{total}} = \alpha_{\text{run}} P_{\text{dyn}} + \alpha_{\text{rec}} P_{\text{rec}} \]

where \( P_{\text{dyn}} \) and \( P_{\text{rec}} \) represent the running and reconfiguration power respectively, and \( \alpha_{\text{run}} \) and \( \alpha_{\text{rec}} \) represent the percentage of time the system operates in regular operation versus reconfiguration mode.

C. Use-Case Scenarios

Two ways to increase the coverage and capabilities of a wireless network are to use more spectrum and to increase the efficiency of how allocated spectrum is utilized. Government regulations usually dictate the spectrum that is available in a given geographic area. Increasing spectral efficiency is possible under two conditions: (i) the cellular carrier deploys significant amount of higher throughput, lower cell radius access technologies (such as 4G and Wi-Fi); (ii) the mobile terminals are capable of reconfiguration which allows them to access additional resources through roaming agreements between cellular carriers. We formulate two use cases based on these conditions, both of which assume that two cellular wireless providers (we refer to each as Carrier 1 and Carrier 2) provide coverage within the same geographic area. The two use cases differ in the level of cooperation that exists between the two carriers. Use case 1 involves \( x \) nodes that can connect to only Carrier 1’s cellular and WiFi network and \( y \) nodes can connect to only Carrier 2’s cellular and WiFi network. Use case 2 allows any node to make use of the other carrier’s network if there is excess capacity.

D. Simulation Description

We developed a MATLAB-based simulation model with sufficient fidelity to demonstrate the possible benefits of the proposed wireless system. We base the analysis methodology loosely on that used in [22] which demonstrates the potential increase in spectral efficiency when femtocells augment the reach of a wireless provider’s macrocell network. While our underlying assumptions allow us to consider more complex heterogeneous wireless systems, our methodology is similar. We use simulation to model an approach for managing resources taking into account the benefits and possible costs of reconfiguration. As in [22], we are interested in showing the improvements observed by each node and also at the global network-wide level. There are many aspects of the analysis that we do not take into account, including modeling the impacts of the messaging overhead that is required by a global resource control strategy and modeling the impacts of signal fading and channel interference.

The simulation topology that was used for the results reported in this paper consists of a 2 \( \times \) 2 km\(^2\) grid. Any number of wireless access technologies can be used for connectivity within the grid. The AWS model is parameterized by transmit power, a propagation model, and a mapping between effective receive power at a given location and effective data rate. The latter capability allows, for example, the modulation and coding of specific radio access technologies (e.g., WiMAX16e, 802.11g) to be taken into account. An example simulation configuration involving three AWSs per carrier is illustrated in Figure 2. Using this example, a simulation involves any number of nodes, each of which can be assigned a maximum of three radios. Further, each radio might be reconfigurable (i.e., can be instructed to operate over any of the AWSs) or non-reconfigurable (statically set to one AWS). Periodically, a GRC (which is not shown in the figure) allocates bandwidth using a scheduling algorithm that selects channels for each reconfigurable radio and that assigns bandwidth to nodes in a manner that attempts to maximize both resource utilization and fairness.

We present simulation results from a scenario involving two wireless providers, each of which has three RATs: EVDO Rev 0 (3G Carrier 1), HSPA (3G Carrier 2), IEEE 802.16e (4G Carrier 1), LTE (4G Carrier 2) and IEEE 802.11g (Wi-Fi Carrier 1 and Carrier 2). For the cellular based AWSs (3G/4G AWS) we assume that a single base station serves all nodes in the simulation topology. These base stations are located near the center of the grid. The 3G base stations have a coverage radius of 1.25 km and the 4G base stations have a coverage radius of 0.75 km. The IEEE 802.11g APs are spread throughout the topology as illustrated in Figure 2. There are two Wi-Fi AWSs, each having three APs, and each AWS belongs to a different carrier. The coverage radius of Wi-Fi AP is 0.15 km. The scenario is based on an urban location that has two large service providers in the area and is designed to show the possible tradeoffs surrounding varying levels of device reconfiguration and varying levels of sharing between the two wireless providers.

Figure 2: Coverage of an example simulation topology

The simulation involves 100 mobile nodes that are initially positioned randomly within the grid. 50 nodes represent users subscribed to Carrier 1 and the other 50 represent users subscribed to Carrier 2. Nodes move using a random walk mobility model. The node speed is an experimental parameter, however for a given simulation, all nodes move at the same speed. Due to this movement, link performance varies over time depending on the node’s location in the grid as well as on local and global resource allocation decisions. To strike a balance between update frequency and overhead, the model assumes that the mobile nodes send location and connectivity information to the GRC every one second. The GRC implements a scheduler that assigns each node’s radio the most efficient access technology that is in range and that allocates bandwidth in a manner that seeks fairness while maximizing total throughput. The scheduler follows a two step approach. In the first step, the scheduler allocates a minimum required

\[ \begin{align*}
\text{Coverage of an example simulation topology}
\end{align*} \]
throughput of 100 kbps to each node using its best cellular (3G/4G) radios. In the second step, the scheduler distributes unused cellular access technology resources to a window of 10 mobile nodes with best connectivity parameters in increments of 100 kbps. Nodes are limited to a maximum allocation of 1 Mbps by the cellular networks. Finally, Wi-Fi resources are distributed evenly amongst connected users by each Wi-Fi AP.

Each node uses radio configurations according to the decisions made by the GRC. When the GRC instructs a node to switch/reconfigure the radio to be used, there is a cost associated with this operation in terms of communication downtime and an increase in power consumption. The model assumes that each node will consume all of the bandwidth that it is allocated. Since the traffic loads are static, the reconfiguration rate experienced by nodes is determined primarily by the speed at which the mobile nodes move. The simulation is run for mobile node speeds in the range of [2 mph, 40 mph] in increments of 2 mph. The reference time and power cost values are presented in Table 1. The reference power estimates from the table are used in the simulation. Because the scheduler operates on a 1 second allocation basis, we approximate the communication downtime cost by not allocating any bandwidth to the radio 1 second. This significantly exceeds the reconfiguration times shown in Table 1. However if we assume the downtime also includes the time required to establish the new physical and logical link connections, a downtime cost of 1 second seems reasonable. To better understand the impact of the cost, we multiply the cost of reconfiguration by a scalar in the range [0, 1] which we define as the “impact of reconfiguration”. The scalar value of 0 implies there is no switch/reconfiguration cost. The scalar values of 1 represents the reconfiguration power costs provided in Table 1 and the reconfiguration time cost of 1 second.

Each simulation is run for 10,000 seconds. The results from the simulations include the spectral efficiency and the average power consumption per node that were observed as the two experimental parameters (node speed and the relative impact of reconfiguration) were varied. We compute the spectral efficiency at the end of a simulation run by summing the throughput achieved by each node and dividing the sum by the total spectrum bandwidth (48.25 MHz) managed by all AWSs. The total power consumption of each node is calculated using Equation (1). At the end of the simulation, the aggregate power of all nodes is divided by the number of nodes and the simulation time resulting in the average power consumption per node.

E. Results and Discussion

The results are dependent on node mobility, access technology coverage range, allocation resources of each access technology and scheduler implementation. The scheduler implemented in this study attempts to maximize network efficiency and maintain fairness across users. Other forms of schedulers might result in different results. We will explore in detail the scheduling problem in future work. For the purposes of this study, our objective was a simple allocation strategy that is sufficient to illustrate the potential benefits of the proposed ideas.

The spectral efficiency observed in the simulations is visualized in Figure 3. As expected, use case 2 utilizes the spectrum more efficiently than in use case 1. Reconfiguration allows the global and local controllers to use the most efficient RAT and modulation and coding. On average, the spectral efficiency gain for use case 2 (1.70 bits/sec/Hz) when compared to use case 1 (1.45 bits/sec/Hz) is around 17%. The average throughput achieved by each user is around 700 kbps for use case 1 and 820 kbps for use case 2. Other results that are not included in this paper show that use case 2 exhibits roughly three times the reconfiguration rate as in use case 1. Figure 3 shows that spectral efficiency does not dramatically change as the node speed increases. It turns out that the rate of reconfiguration increases by only 10% on average as the node speed increases from 2 mph to 40 mph. This is an artifact of the topology and of the relatively small number of AWSs. We conjecture that time varying workloads would have a more significant impact on the effective rate of reconfgurations. Figure 3 also shows that the impact of reconfiguration is greater with use case 2. Again, this is expected since use case 2 experiences a much higher rate of reconfiguration than use case 1.

Figure 3: Spectral Efficiency

The average power consumption depicted in Figure 4 suggests that the impact of reconfiguration has a far greater effect on average power consumption than the speed of the mobile user. For use-case 2, since the reconfiguration rate is significantly high, the power consumption almost doubles (from 3.4 Watts to 7.0 Watts) with the reference power reconfiguration specs used from Table 1 at impact of reconfiguration value of 1. This suggests that the power cost needs to be carefully examined before the GRC issues a reconfiguration command. For an average power consumption of 7 Watts as compared to 3.4 Watts, a mobile battery powering a mobile device such as an iPhone 4G would last 1 hour 20 minutes in comparison to 2 hours 50 minutes. This decreases the battery life of a mobile terminal by almost half. But as better hardware innovations are made, the actual battery life reduction is going to decrease since the impact of

---

1Additional results and discussion are available in a longer version of this paper [22].
reconfiguration is going to move farther away from 1 and move closer towards 0.

![Figure 4: Average Power Consumption](image)

IV. TECHNICAL CHALLENGES

We identify the following technical challenges that guide our next phases of this work:

- Develop resource allocation methods that support desired services and that achieve fairness objectives.
- Establish optimization techniques to manage the tradeoffs between power, degree and frequency of reconfiguration, channel depth, and spectrum efficiency.
- Identify and develop an appropriate device platform and architecture in terms of power consumption, reconfigurability, programming model and scalability.

V. CONCLUSIONS AND FUTURE WORK

In spite of recent advances in wireless technology, wireless networks continue to be designed as independent networks that make resource decisions without considering co-located networks. Moving beyond 4G networks requires cooperation at the network layers. The results presented in this paper provide a snapshot of one possible cost-benefits tradeoff that exists as the reconfiguration capabilities of devices increases.

REFERENCES