Abstract — The proliferation of wireless broadband usage over the last decade has led to the development and deployment of multiple broadband wireless radio access technologies (RATs) such as EVDO, WiMAX, HSPA, and LTE. To support the ever-increasing wireless traffic demand, researchers have worked on the concept of an integrated heterogeneous wireless environment that encompasses several of these RATs which makes the resource allocation process more efficient by assigning each user in the system to the best RAT/RATs. In this paper, for such an integrated heterogeneous wireless system, we show the possible gains in spectral efficiency at the cost of increased energy consumption for an unbalanced heterogeneous wireless network deployment scenario. In prior work, based on the assumption that all cellular operators under study were equally well-provisioned, we showed an increase in spectral efficiency of up to 75%. In the research presented in this paper, we assume the coverage of each operator might differ significantly in a given area. With this ‘unbalanced’ scenario, we show that an ideal, centralized allocation strategy provides an almost linear tradeoff between gain in spectral efficiency (554%) and worst-case increase in energy consumption (615%) for users supporting elastic traffic.

Keywords - heterogeneous wireless networks; reconfigurable radios; next generation wireless systems; cyberinfrastructure

I. INTRODUCTION

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. Until recently, technology was the primary impediment to achieving universal, broadband wireless services that involve multiple radio access technologies. Today, the most significant impediments to achieving highly efficient unified wireless systems are the after-effects of antiquated government spectrum allocation policies. The effect is that in many geographic areas licensed spectrum is likely to be underutilized [1]. Meanwhile, mobile broadband traffic demand is experiencing unprecedented growth due to the proliferation of smart phones over the last decade. The demand is expected to exceed current operator spectrum capacity as early as 2014 [2].

Current cellular systems deploy a centralized scheduler at the base-station. Each base-station ‘independently’ tries to optimize the resource allocation process within its own network, generally ignoring impacts of co-located wireless networks. This selfish behavior will usually not lead to optimal resource usage. Significant improvements in resource usage are possible compared to this scenario when jointly considering network-wide resource allocation process due to the benefits of multi-user diversity. In fact, [3] shows that the selfish approach can result in non-Pareto optimal bandwidth allocation as compared to the case where a centralized entity performs network-wide resource allocation process.

At the network level, architectures and frameworks to support hybrid or heterogeneous networks have been suggested [4]-[8]. A survey of these architectures has been provided in [5]. Recent proposals have been based on the Media Independent Handover function defined by the IEEE 802.21 standard which provides a framework to support seamless mobility through networks based on different radio access technologies without the need to restart the radio connection every time the mobile moves to a new network [6]. 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 [7]. 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. The overhead associated with a centralized hierarchical wireless system is studied in [8]. 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.

In prior work [9], we studied the impacts of using reconfigurable devices in a heterogeneous wireless system in terms of an increase in spectral efficiency and the corresponding increase in energy consumption. In this previous work, we used a balanced network topology where cellular 3G/4G and Wi-Fi resources of two cellular carriers in a 2 * 2 km² area were intentionally arranged so that both carriers had roughly equivalent capacity within the grid. As observed in our prior results as well as recent studies that are similar in motivations with our work [10], system performance varies greatly with topology and user behaviors. Our previous work showed moderate improvements in spectral efficiency largely because of the ‘balanced’ nature of the study (i.e., the carriers...
had similar deployments and the random waypoint mobility pattern caused all user’s to smoothly flow over the majority of the grid area). In this paper, we compare the results of our previous work with an unbalanced network topology where cellular and Wi-Fi resources of one carrier significantly exceed those of the other carrier. In addition, we cluster the users of both carriers and limit their movement in a restricted area. The use of unbalanced topology and clustered node movement pattern helps realize the maximum benefit of using reconfigurable devices in a heterogeneous wireless system by converting a highly unfavorable situation to a favorable situation. Furthermore, in our previous study, we provided results based on reconfigurable devices that were implemented using FPGA technology. In this paper, we analyze the differences in energy consumption when using devices made up of (i) only ASIC components, (ii) only FPGA components, or (iii) a combination of ASIC and FPGA components. The main contribution of this paper is the insight it provides with the tradeoffs involved between two conflicting objectives of increasing spectral efficiency and minimizing energy consumption in a heterogeneous wireless network with reconfigurable client devices.

This paper is organized as follows. Section II presents and relates the relevant background and provides motivations for the work. We describe the system model and the research methodology in Section III. We discuss the results in Section IV. We conclude the paper in Section V.

II. BACKGROUND AND MOTIVATIONS

Once wireless operators have completed their 4G deployments of licensed spectrum (to the point where it is difficult for carriers to differentiate themselves just by connectivity), they will need to differentiate themselves through services. With possible prodding from FCC to decouple the physical access provider from service providers, it is not unreasonable to see carriers form peering relationships so that they can each better deliver more robust services to their subscribers. In addition to the mandate passed by the FCC that requires the cellular carriers to establish roaming agreements for their 3G/4G data services [11], we point out two additional examples that demonstrate steps in this direction. Wi-Fi offloading refers to how cellular systems cooperate with 802.11 Wi-Fi networks. While innovative economic models are being developed, currently cellular systems leave the choice of access to the end user. However, cellular operators would prefer their customers to use Wi-Fi when available. The benefits from this approach are multi-fold; firstly, the carrier saves the expensive macro-cell capabilities for the truly mobile members of the cell. Secondly, the performance for low mobility, indoor members of the cell is improved due to avoiding indoor penetration issues, thus significantly improving overall network performance. While some commercial carriers such as T-mobile have already experimented with this approach, maintaining a seamless transition between the networks has proven elusive to-date. In the case of T-mobile, although the company does support Wi-Fi calling, they explicitly warn that there is no handoff between the macro cell and the Wi-Fi call, which limits the applicability of this approach [12]. Femtocells have been proposed as a method to increase spectral efficiency by supplementing the macrocell with an overlay of smaller, cooperative networks [13-14]. The two examples are similar in spirit as they attempt to improve the connectivity of handheld devices and to offload traffic from the macrocell.

To fully benefit from such emerging network concepts, there is a need for efficient design time and run time reconfigurable platforms. Numerous reconfigurable architectures have been proposed spanning different technologies including application specific instruction set processors (ASIPs), field programmable gate arrays (FPGAs), and digital signal processors (DSPs). Recently, multiprocessor systems on chip (MPSoC) architectures have evolved rapidly in the race of high performance embedded computing [15], especially in applications that require a flexible computing structure that can be reconfigured to handle various applications. A common design metric among all platforms is reducing energy consumption that restricts both the capabilities of the device and the design choices that are available. Towards that end, numerous techniques have been developed to optimize energy consumption at different levels including algorithm, system, architecture, and circuit levels [16-17].

III. SYSTEM DESCRIPTION

A. System Model

Figure 1 illustrates the system model. The system consists of mobile devices (also referred to as nodes) that have connectivity to one or more Autonomous Wireless Systems (AWSs). Each AWS comprises of one or more Base Stations (BSs) or Access Points (APs) of the same RAT. There are two resource-controlling entities in our proposed system: 1) Global Network Resource Controller (GRC) present in carrier’s backend network 2) BS/AP of each RAT. The GRC makes decisions on large time-scales (seconds) using average statistics assigning each node one or more BS/AP to use for connectivity. The BSs/APs operate over small time-scales (milli-seconds) to manage the resources of their corresponding RAT and account for fluctuations in connectivity parameters due to the effects of fading. The GRC calculates user-technology mappings (and supported throughput per mapping) and relays the results to each BS/AP as well as the Local
Resource Controller (LRC) of each mobile device. The BS/AP uses this information to establish active connections with the corresponding devices and in making its own scheduling decisions. The LRC uses this information to configure its radios to the specified RAT(s).

Users are presented with a unified network. Depending on node capabilities, users can operate over more than one RAT concurrently at any given time. A node’s TCP/IP stack sees a single IP link. The wireless virtual link layer (WVLL) handles packet scheduling over one or more radio links. Packet resequencing, error recovery using ARQ and/or FEC can optionally be implemented over the tunnel. The radio link block pictured in Figure 1 represents the MAC and physical layers that operate 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), digital signal processors (DSPs) or multi-core Application Specific Integrated Circuits (ASIPs). User data is tunneled over the unified network cloud. Entities such as Packet Data Network Gateway (PGW) for 3GPP LTE’s System Architecture Evolution represent the termination point for the tunnel. Additional entities such as Mobility Management Entity (MME) in carrier’s backend adhere to the functionality described in 3GPP LTE’s standards document [18] and help in managing information related to each user in the system while the user is transitioning from one RAT to another.

B. Network Topology

We articulate the presence of two major cellular carriers in a 2X2 km2 area that operate multiple RATs. We use EVDO (3G), HSPA (3G), WiMAX (4G), LTE (4G) and IEEE 802.11g (Wi-Fi) in our experiments as the representative RATs that current cellular carriers support. The 3G base stations (EVDO, HSPA) have a coverage radius of 1.50 km. The 4G base stations (WiMAX, LTE) have a coverage radius of 1.0 km. The Wi-Fi APs have a coverage radius of 0.15 km. In our pervious work [9], we created a balanced network topology where each carrier deployed a 3G technology (EVDO – Carrier 1, HSPA – Carrier 2), a 4G technology (WiMAX – Carrier 1, LTE – Carrier 2) and 3 Wi-Fi APs in the 2X2 km2 grid. We placed the 3G/4G base-stations at the center of the grid and the Wi-Fi APs were spread throughout the topology to give each carrier equal network coverage. In this study, we create a scenario where Carrier 1 only has 3G network coverage (EVDO) from two base-stations, whereas Carrier 2 has network coverage from one 3G base-station (HSPA), one 4G base-station (LTE) and six Wi-Fi APs that are spread throughout the topology. The resulting unbalanced network topology used in this study is shown in Figure 2.

C. 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 energy consumption due to its custom nature. For example, a non-reconfigurable radio is able to support only one RAT, but it can operate using various modulation and coding schemes (MCS) supported by that RAT. In the network topology shown in Figure 2, the color shades represent various MCS levels supported by each RAT. The darker the shade, the higher the MCS level a radio can use in a specific area. A reconfigurable radio is fully adaptive, but consumes comparatively higher energy. If a reconfigurable radio moves from the coverage of one RAT to the coverage of another RAT, the GRC will instruct the radio to reconfigure itself to support the new RAT. We define this event as a ‘reconfiguration handoff’. A reconfiguration handoff is a vertical handoff that requires a radio to reconfigure itself.

The radios (static or reconfigurable) are made of either ASIC, FPGA or a combination of ASIC and FPGA components. Depending on these components (ASICs or FPGAs), the energy consumption of radios will vary. For static radios, the energy consumption is categorized as dynamic power ($P_{\text{dyn,ASIC}}$) and FPGA ($P_{\text{dyn,FPGA}}$) based radios for RATs used in this study is presented in Table 1 and was provided in our previous work [9]. The ratio of $P_{\text{dyn,FPGA}}$: $P_{\text{dyn,ASIC}}$ is 12:1 as recommended by an analysis performed in [19]. For reconfigurable radios, in addition to the dynamic energy consumption, we use reconfiguration power ($P_{\text{rec}}$), defined as the energy that is consumed when the circuit is reconfigured from one RAT to another, that is also presented in Table 1.

Table 1. Implementation statistics for current technologies [9]

<table>
<thead>
<tr>
<th>Component</th>
<th>No. of Kgates</th>
<th>Mobile WiMAX</th>
<th>LTE</th>
<th>HSPA</th>
<th>EVDO</th>
<th>EVD (est from UMTS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{\text{dyn,ASIC}}$ (Watts) @100 MHZ</td>
<td>1.76</td>
<td>3</td>
<td>1.13</td>
<td>3</td>
<td>2.83</td>
<td></td>
</tr>
<tr>
<td>$P_{\text{dyn,FPGA}}$ (Watts) @100 MHZ</td>
<td>0.15</td>
<td>0.25</td>
<td>0.09</td>
<td>0.25</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td>$P_{\text{rec}}$ (Watts) @50 MHZ</td>
<td>7.25</td>
<td>12.4</td>
<td>4.68</td>
<td>12.4</td>
<td>11.7</td>
<td></td>
</tr>
</tbody>
</table>

To determine the portion of the radio implemented using FPGA and ASIC technology, we define a scalar $\beta \in [0,1]$ that represents the percentage of radio components made using FPGA technology. $1 - \beta$ represents the percentage of radio...
components made using ASIC technology. Note that the exact combination of FPGA vs. ASIC implementation is radio vendor specific [21]. The goal of this paper is not to argue what the correct \( \beta \) value is, but rather to show the trend in energy consumption as different percentages of FPGA vs. ASIC combinations are tried. In addition, we use another scalar \( \lambda \in [0,1] \) which we define as ‘impact of reconfiguration’. \( \lambda = 1 \) represents using reconfiguration energy consumption numbers presented in Table 1. But as hardware evolves and these costs go down as a result of innovative radio design architectures, the reconfiguration power will only be a fraction of that presented in Table 1. The scalar \( \lambda \) is supposed to capture this effect and as \( \lambda \to 0 \), the reconfiguration power becomes almost negligible. Using these scalars and the energy consumption values presented in Table 1, the total energy consumption of each node can be calculated as follows:

\[
P_{\text{total}} = \alpha_{\text{run}}[\beta P_{\text{dyn,FPGA}} + (1-\beta) P_{\text{dyn,ASIC}}] + \alpha_{\text{rec}}[\lambda P_{\text{rec,FPGA}} + (1-\lambda) P_{\text{rec,ASIC}}]
\]

where \( \alpha_{\text{run}} \) and \( \alpha_{\text{rec}} \) represent the percentage of time the system operates in regular mode and reconfiguration mode respectively. \( P_{\text{rec,FPGA}} \) represents the energy consumption when the circuit is reconfigured from one RAT to another. \( P_{\text{rec,ASIC}} \) represents the increase in energy consumption when ASIC components of a radio are turned ‘on’ from an ‘off’ state. The \( P_{\text{rec,FPGA}} \) and \( P_{\text{rec,ASIC}} \) numbers used in this study are the same as represented by \( P_{\text{rec}} \) in Table 1.

D. Simulation Description

We developed a MATLAB-based simulation model with sufficient fidelity to demonstrate the tradeoffs between increase in spectral efficiency and decrease in energy consumption in a heterogeneous wireless network environment. We study two use cases: 1) Use case 1 involves nomadic nodes that can connect only to its own carrier’s cellular and Wi-Fi network. 2) Use case 2 allows any nomadic node to make use of the other carrier’s cellular and Wi-Fi network if there is excess capacity on other carrier’s networks. While there are economic and political obstacles surrounding use case 2, we assume these obstacles will eventually be overcome. Furthermore, for use case 1, since a node can only connect using RATs of its own carrier, the node is equipped with up to three static radios. For use case 2, each node is equipped with three reconfigurable radios and is capable of supporting all available RATs.

The unbalanced 2 * 2 km² grid network shown in Figure 2 is created as the simulation topology. The simulation involves 100 nomadic nodes, 50 of which are subscribed to Carrier 1 and the other 50 subscribed to Carrier 2. The users of Carrier 1 are clustered in 2 groups. The first group is located on the left side of the grid (centered at [0 m, 1000 m]) and the second group is located on the right side of the grid (centered at [2000 m, 1000 m]). The users of Carrier 2 form the third cluster and are located in the center of the grid (centered at [1000 m, 1000 m]). All users are allowed to move in a restricted space of 500*500 m² using random walk mobility model based on the cluster they belong to. This leaves each user with relatively bad coverage for use case 1 and significantly improves performance when they start using resources of the other carrier under policies of use case 2.

The Global Resource Controller implements a scheduler that comes up with user-technology association decisions every one second and is presented in [9, 20]. Each node uses radios according to the decisions made by the GRC. We assume each node supports elastic traffic and has infinite number of backlogged flows to use up all assigned bandwidth each scheduling period. 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 temporary downtime and an increase in energy consumption. Because the scheduler operates on a 1 second allocation basis, we approximate the communication downtime cost by not allocating any bandwidth to the radio for 1 second. If we assume that downtime also includes the time required to establish the new physical and logical link connections, a downtime cost of 1 second seems reasonable based on vertical handover times claimed to be between few hundred milliseconds to a few seconds by the work presented in [4]. The energy consumption cost during reconfiguration is presented in Table 1. Since both communication downtime and increase in energy consumption are hardware and implementation specific, we multiply the communication downtime (1 second) and reconfiguration energy cost (Table 1) with an experimental parameter, \( \lambda \in [0,1] \) defined as ‘impact of reconfiguration’, to capture the effects of evolution as better hardware and seamless heterogeneous network architectures are developed.

We do not include a detailed channel model in our studies, but rather introduce an artificial degradation in network quality. We use a parameter which we refer to as ‘network outage’ to model the percentage of time the network is unavailable to the users. An outage might occur as a result of a number of situations including congestion due to increased network load, increased RF interference levels, AP/BS malfunction/software upgrades, or even network attacks such as Denial of Service. The network outage is an experimental parameter that controls the percentage of slots of a channel that are effectively not used. The outage percentage ranges from 0% to 25% in increments of 5% in our simulation. Each AWS suffers independent random outages with the probability determined by the network outage percentage.

IV. RESULTS AND ANALYSIS

Each simulation is run for 10,000 seconds. The results from the simulations include spectral efficiency and the average energy consumption per node that are observed as the two experimental parameters (network outage and the relative impact of reconfiguration) are 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 managed by all AWSs. The total energy consumption of each node is calculated using Equation (1). At the end of the simulation, the aggregate energy consumption of all nodes is divided by the number of nodes and the simulation time resulting in the average energy consumption per node.

A. Spectral Efficiency

The spectral efficiency observed in the simulations is visualized in Figure 3. As expected, use case 2 utilizes the spectrum more efficiently than use case 1. Reconfiguration
allows the global and local controllers to assign each node to the most efficient RATs. To get a lower bound, for no network outage and impact of reconfiguration of 1, the spectral efficiency gain for use case 2 (1.43 bits/sec/Hz) when compared to use case 1 (0.34 bits/sec/Hz) is around 314.3%. The maximum spectral efficiency gain for use case 2 (1.79 bits/sec/Hz) when compared to use case 1 (2.73 bits/sec/Hz) is around 553.7% when there is 25% network outage and the impact of reconfiguration is 0. The increase in spectral efficiency range [314.3%, 553.7%] is quite high for unbalanced network topology when compared to a balanced network topology [14.3%, 75.5%] derived in our previous study [9]. This is mainly due to the fact that in the unbalanced topology, for use case 1 all the users connect to RATs supported by their own carrier at very low rates since they are at the edge of their carrier’s network coverage. But for use case 2 when all the users connect to any available RAT, they connect with RATs supported by the other carrier at very high rates since they are very close to the base-stations of those RATs. This shows the tremendous gains that are possible in a realistic unbalanced network deployment scenario where resources of one carrier exceed those of another if a truly heterogeneous wireless system is created where all available resources in a given area are managed at a global level.

B. Energy Consumption

Energy consumption in our heterogeneous wireless system depends on two factors: 1) The number of reconfigurations 2) The type of hardware fabric used (ASIC/FPGA). The number of reconstructions depends on the number of connectivity options for each node when GRC comes up with node-RAT mapping every scheduling period. Since nodes have more connectivity options under use case 2, there are a greater number of reconstructions for use case 2 compared to use case 1. A reconfiguration for use case 1 is equivalent to switching one radio ‘off’ and turning another radio ‘on’. For use case 2, a reconfiguration requires the circuitry of one radio to be switched to support a different RAT. From the hardware perspective, since we assume static multi-modal radios for use case 1, the radios are made up of complete low-power consuming ASIC components, or using our energy consumption model described in (1), \( \beta = 0 \). For use case 2, since the radios require reconfigurable components, we investigate three architectural models for wireless devices, each differing in the amount of custom circuitry (e.g., ASIC components) and the reconfigurable capabilities (e.g., FPGA components). The model involves radios that are: i) Made up of completely FPGA components, i.e. \( \beta = 1 \) ii) Made up of 50% FPGA and 50% ASIC components, i.e, \( \beta = 0.5 \) iii) Made up of completely ASIC components, i.e. \( \beta = 0 \).

![Figure 3. Spectral Efficiency](image)

![Figure 4. Average Energy Consumption for Hardware Settings 3 versus 1 (ASIC - \( \beta = 0 \) vs. FPGA - \( \beta = 1 \))](image)

![Figure 5. Average Energy Consumption for Hardware Settings 3 versus 2 (ASIC - \( \beta = 0 \) vs. Hybrid - \( \beta = 0.5 \))](image)

![Figure 6. Average Energy Consumption for Hardware Setting 1 (ASIC - \( \beta = 0 \)) under different use cases (Use Case 1 vs. 2)](image)

The result for the first hardware setting is provided in Figure 4. As can be seen, the increase in the energy consumption lies in the range [104.9%, 614.9%] when the
radios are implemented completely using FPGA fabric as compared to a complete ASIC implementation. The increase in energy consumption of over 600% might be too costly even though the gain is spectral efficiency is about 550% for that setting. As an alternative, when 50% of the radio fabric is made using ASIC components, the increase in energy consumption lies in the [70.0%, 355.4%] range as shown in Figure 5. The increase in spectral efficiency of about 550% at the cost of increase in energy consumption of about 350% would be a better choice to implement reconfigurable radios. The average number of radios used per node for use case 2 is 1.23. At each time step, none of the nodes use more than 2 radios. So, it might suffice to limit the number of reconfigurable radios per node implemented using FPGA fabric and have some static radios that use low-power custom built circuitry (ASIC fabric). How much of this hybrid architecture is possible today is still an open research question. The third hardware setting is not feasible today in building a completely reconfigurable device, but is studied to provide an intuition on how much extra energy is consumed if the only difference between the two use cases is the number of connectivity options available to each node per scheduling interval. As can be seen from Figure 6, the increase in energy consumption lies in the range [35.1%, 98.8%]. This setting matches our results obtained in our previous study [9] where we claimed an increase in energy consumption of about 100% when the gain in spectral efficiency was maximum (75%). So just based on an increase in number of connectivity options, which results in a higher rate of reconfiguration for use case 2, and using the same hardware components in constructing radios for both use cases results in a huge increase in spectral efficiency (553.7% for unbalanced topology) at twice the amount of energy consumption, a result we highlighted in our previous work.

V. CONCLUSIONS

Global allocation of resources in an integrated heterogeneous wireless environment that encompasses several RATs makes the resource allocation process more efficient by assigning each user in the system to the best RAT/RATs. The gains in spectral efficiency are even greater for an unbalanced network topology when compared to a balanced network topology (from 75.5% to 553.7%). However, the gain in spectral efficiency comes at the cost of increased energy consumption. Based on the hardware choices, the increase in energy consumption can range from 98.8% to 614.9% for the corresponding increase in spectral efficiency of 553.7%. Depending on the number of possible modalities supported by devices, it might be possible to attain a tradeoff in terms of lower energy consuming ASIC radios at the cost of decreased reconfigurable options. In the worst case, our results show a more or less linear trade-off between spectral efficiency and power consumption. This result is an artifact of our workload assumptions that assume traffic flows are always backlogged. In future work, we will explore more realistic scenarios that involve on/off traffic flows. In this case, for certain traffic conditions, we would expect to find an inverse relationship between spectral efficiency and energy consumption.

REFERENCES