Balancing Spectral Efficiency, Energy Consumption, and Fairness in Future Heterogeneous Wireless Systems with Reconfigurable Devices

Rahul Amin, *Student Member, IEEE,* Jim Martin, *Member, IEEE,* Juan Deaton, *Member, IEEE,* Luiz A. DaSilva, *Senior Member, IEEE,* Amr Hussien, *Student Member, IEEE,* and Ahmed Eltawil, *Member, IEEE*

***Abstract*—In this paper, we present an approach to managing resources in a large-scale heterogeneous wireless network that supports reconfigurable devices. The system under study embodies internetworking concepts requiring independent wireless networks to cooperate in order to provide a unified network to users. We propose a multi-attribute scheduling algorithm implemented by a central Global Resource Controller (GRC) that manages the resources of several different autonomous wireless systems. The attributes considered by the multi-attribute optimization function consist of system spectral efficiency, battery lifetime of each user (or overall energy consumption), and instantaneous and long-term fairness for each user in the system. To compute the relative importance of each attribute, we use the Analytical Hierarchy Process (AHP) that takes interview responses from wireless network providers as input and generates weight assignments for each attribute in our optimization problem. Through Matlab/CPLEX based simulations, we show an increase in a multi-attribute system utility measure of up to 57% for our algorithm compared to other widely studied resource allocation algorithms including Max-Sum Rate, Proportional Fair, Max-Min Fair and Min Power.**

***Index Terms*—heterogeneous wireless networks, reconfigurable radios, scheduling, resource allocation, network efficiency.**

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 wireless technologies at various lifecycle stages. Until recently, technology was the primary impediment to achieving universal, broadband wireless services that involve multiple radio access technologies (RATs). Today, arguably one of the most significant impediments to achieving highly efficient unified wireless systems are the after-effects of antiquated spectrum allocation regulatory policies. The effect is that in many geographic areas licensed spectrum is likely to be underutilized even though demand is expected to exceed current spectrum capacity as early as 2014 [1, 2]. This has sparked renewed interest in techniques to improve spectral efficiency, including cognitive radios and networks that can adapt their behaviors to make efficient use of open or unused spectrum. At the same time, environmental concerns and user device requirements have elevated the importance of energy efficient networks and devices. As wireless operators have learned, a handheld device's battery efficiency is a very visible attribute of an operator's services [3]. Unfortunately, in many situations, methods for improving spectral efficiency directly lead to an increase in energy consumption. A third conflicting dimension to the resource allocation problem is ensuring fair allocation across users. The wireless system design and the subsequent performance modeling and evaluation that are presented in this paper directly address this challenging dilemma through the use of joint optimization.

A large amount of prior work has explored a number of issues surrounding the delicate relationship between transmit power and subsequent effects on competing devices [4–6]. Recent work involving heterogeneous cellular systems based on pico-cell and femto-cell deployments also consider the power issue to mitigate inter and intra channel interference [7–9]. Prior joint optimization-based resource allocation approaches for wireless networks consider trade-offs between spectral efficiency and fairness objectives subject to maximum transmit power constraints [10–12]. Recently, due to a renewed interest in 'green communications' and end-users' increased expectations from mobile device battery life, researchers have started focusing on minimizing overall energy consumption subject to fairness constraints and other network-efficiency requirements, such as throughput and delay [13–15]. However, to the best of our knowledge, none of these works looked at the trade-offs surrounding power, spectral efficiency, and fairness in large-scale heterogeneous wireless networks that involve reconfigurable end-user devices.

Current cellular systems deploy a hierarchy of resource controllers. Each device, along with its assigned base station (BS), independently tries to optimize the resource allocation process within its own domain, generally ignoring impacts of co-located heterogeneous wireless networks. Localized resource allocation decisions will usually not lead to optimal resource

Manuscript received April 10, 2012; revised November 2, 2012.

Rahul Amin is with the Department of Electrical and Computer Engineering, Clemson University, Clemson, SC, 29634, USA (email: ramin@clemson.edu).

Jim Martin is with the School of Computing, Clemson University, Clemson, SC, 29634, USA (email: jmarty@clemson.edu).

Juan Deaton is with Idaho National Lab, Idaho Falls, ID, 83415, USA and a recent Ph.D. graduate from Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, 24061 (email: juan.deaton@gmail.com).

Luiz A. DaSilva is with the Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, 24061 (email: ldasilva@vt.edu). He is also with CVTR, Trinity College Dublin.

Amr Hussien and Ahmed Eltawil are with the Department of Electrical and Computer Engineering, University of California, Irvine, CA, 92617 (email: ahussien@uci.edu, aeltawil@uci.edu).

usage. In fact, [16] 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. Significant improvements in efficiency result when the resource management process jointly considers the distribution of resources across network technologies, reaping the benefits of multi-access network diversity. At the network level, several architectures and frameworks to support hybrid or heterogeneous networks have been suggested, which include IEEE 802.21, IEEE P1900.4, 3GPP's Common Radio Resource Management, Joint Radio Resource Management and Multi-access Radio Resource Management, and IETF's 'flow mobility' standards [17–22]. A common attribute of all these frameworks is that the local resource managers of different RATs interact with a centralized entity to jointly optimize the resource allocation process.

In our prior work involving a heterogeneous wireless system based on a global (i.e. centralized) resource controller, we studied the spectral efficiency and energy consumption trade-offs as the reconfiguration capabilities of devices were varied [23]. Results from a particular scenario of this prior work suggests that it is possible to increase spectral efficiency by up to 75% but at the cost of twice the energy requirement of devices. In the research presented in this paper, we extend our prior work to include a joint optimization-based resource allocation process that provides an operator with a 'control knob' to allow the conflicting demands of spectral efficiency, energy consumption, and fairness to be tailored to meet specific performance goals or policies. We use utility theory, and in particular a weighted sum multi-attribute optimization technique, to set up our network optimization problem [24]. The operator can then use economic incentives to align user perceived utility with an operator's financial utility, which is a separate problem and not addressed in the work presented in this paper.

We introduce a multi-attribute resource allocation algorithm that accounts for cost associated with the use of reconfigurable devices. Based on the well-known utility-based optimization technique [25], we use the weighted sum method in our algorithm, which maximizes utility functions related to multiple network-efficiency attributes. To provide a concrete illustration of the proposed algorithm, we assume the network provides a best-effort data service with extensions to support real-time traffic. The extensions include two components: 1) admission control; and 2) minimum instantaneous data rate per scheduling interval. The service definition matches the needs of latency-sensitive applications such as 'over-the-top' VoIP and video conferencing that periodically require small amounts of data to be delivered by the network with tight delay bounds. The main contribution of the research is an approach for managing resources in large-scale heterogeneous wireless networks that involve reconfigurable end-user devices. Moreover, the proposed resource allocation procedure provides a mechanism that allows an operator to configure the relative importance given to spectral efficiency, energy consumption, and fairness objectives in a manner that leads to predictable service levels.

The paper is organized as follows. Section II presents relevant background and motivations for our work. We describe the system model, provide a detailed problem formulation and discuss the research methodology in Section III. We present and discuss the results in Section IV. Section V provides conclusions and identifies possible future work.

II. Background and Motivations

Network selection algorithms for optimal service delivery over user devices capable of connecting with several RATs can be categorized into several strategies: (i) decision function-based strategies; (ii) user-centric strategies; (iii) multiple attribute decision making strategies; and (iv) fuzzy logic and neural networks-based strategies. All these strategies use a set of attributes in the decision making process which are either related to the user or to the service provider. Some of the user-related attributes include achieved throughput by each individual user, battery lifetime of each mobile terminal, and QoS parameters such as packet delay, jitter and loss. Service provider-related attributes include load-balancing, throughput fairness amongst users, incurred cost per transmitted data byte, and overall revenue. The decision function-based strategies use a weighted utility function that incorporates both user-related and service provider-related network selection attributes [26, 27]. The user-centric strategies focus on one or more needs of the user to decide on the choice of current access network [28]. Multiple Attribute Decision Making (MADM) deals with the problem of choosing from a set of alternatives that are characterized in terms of their attributes. The most popular classical MADM methods are Simple Additive Weighting (SAW), Technique for Order Preference by Similarity to Ideal Solution (TOPSIS), and Grey Relational Analysis (GRA). A comparison of these models was established in [29] with bandwidth, delay, jitter, and BER attributes. It showed that SAW and TOPSIS provide similar performance under all traffic classes examined. GRA provides slightly higher bandwidth and lower delay to interactive and background traffic classes. Fuzzy Logic (FL) and Neural Networks (NN) concepts are applied to choose when and to which network to hand-off among different available access networks when a decision problem contains attributes with imprecise information [30, 31].

All the strategies for network selection algorithms described in the previous paragraph make use of multiple user-related or network provider-related attributes. The method to determine the relative importance of each attribute under consideration has significant impact on the solution space and the implementation complexity of the algorithm. Several papers have looked at multiple weight combinations of the attributes based on imprecise end-user preferences [27, 32, 33]. Other papers have selected the attribute weights based on simulation results by determining the difference in magnitude of each attribute and then assigning each attribute equal importance [26, 31]. We reduce the solution space of our algorithm by using responses from network provider interviews and the Analytical Hierarchy Process (AHP) [34] to determine the relative weights of each attribute in our optimization function.

Game theory has also been employed to model the network selection problem. The authors of [35] propose a network selection scheme to accommodate current demand and minimize

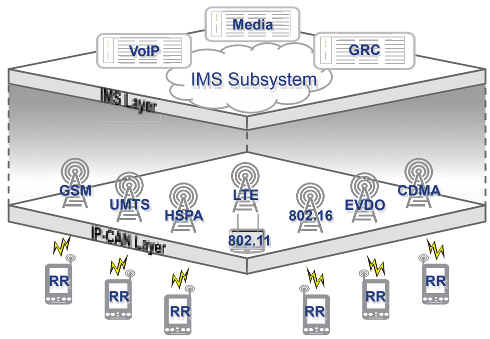

Fig. 1. System Model

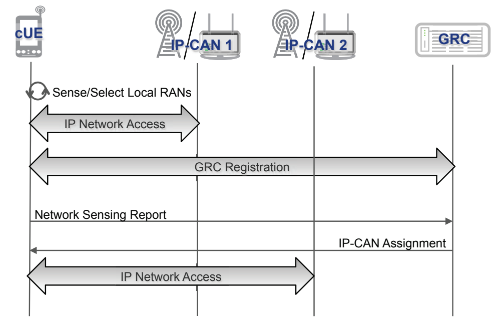

Fig. 2. Resource Allocation Procedure

handoff while meeting QoS requirements in a heterogeneous wireless network, comparing the proposed scheme to TOPSIS. The model in [36] consists of a game between access networks in a converged 4G environment, to decide which service requests should be accommodated by each access network. In [37], the authors study the dynamics of network selection in a heterogeneous wireless environment using evolutionary games. Game theory formulations model decision-making by autonomous independent agents, while in this paper we focus on a central global resource controller. Still, the extensive literature on game theory for telecommunications (e.g., [38]) provides rich ideas on network and user utility when considering multiple attributes for optimization.

III. System Description

A. System Model

With the major network operators migrating to IP Multi-Media Service (IMS) core networks [39], we consider an architecture based on the 3GPP IMS architecture [40] shown in Fig. 1. In our architectural model, we consider cognitive User Equipment (cUE) as an end-user device with reconfigurable and cognitive capabilities, which is able to access multiple IP Connectivity Access Networks (IP-CANs) individually or simultaneously. Resources are managed through the Global Resource Controller (GRC) according to the objectives of the network operator. Through these objectives, the GRC calculates cUE-IP-CAN mappings and the rate assignment per mapping. The base IMS framework naturally extends to support multi-radio, adaptive devices.

From an operational perspective, the cUE first must sense IP-CANs and register with the GRC before transmitting any data. We show the procedural flow example of this process in Fig. 2. First, the cUE senses and scans for available networks and their utilization. Selecting one of the available IP-CANs, the cUE obtains an IP Network connection, which it uses to communicate with external hosts. We assume that each end-user tries to use the most efficient IP-CAN available initially and follows the following preference order: Wi-Fi, 4G (LTE/WiMAX), 3G (HSPA/EVDO). If the user cannot establish a connection to his/her first preference due to reasons such as very high network load or interference, then he/she tries to connect to his/her second preference and this procedure continues until the cUE can establish an initial IP Network connection. Next, the cUE discovers, registers, and communicates with the GRC application server, which we assume uses standard discovery and registration procedures as described in [40][1]. After a connection with the GRC is established, the cUE delivers periodic sensing information of available networks to the GRC. Upon receiving this periodic sensing information from the cUE, the GRC is able to calculate the cUE-IP-CAN mappings and the rate assignment per mapping. This information is then relayed to each cUE, which uses it to tune its Reconfigurable Radios (RRs) to the appropriate IP-CANs.

After each RR is configured according to the cUE-GRC mapping, radio links are established with the associated IP-CANs for data transmission. A pictorial representation of the transmission plane is shown in Fig. 3. From the perspective of the cUE applications, one TCP/IP stack is managed and scheduled over one or more radio links. The Radio Link Aggregation function is used for packet resequencing and reordering data from each of the RRs. Each RR manages its own radio link and associated protocol with the IP-CANs, which provides a unique IP anchor through the corresponding IP-CAN's core network. In addition, each RR uses the rate assignment per cUE-IP-CAN mapping information calculated by the GRC in its resource demand requests to the corresponding IP-CAN's BS. The IP-CAN BSs use the resource demand requests from each cUE as guidance in coming up with their own local scheduling decisions. Note that the GRC makes periodic decisions on large time scales (seconds or minutes), while the BSs of each IP-CAN make scheduling decisions on small time scales (milliseconds) to account for short-term fluctuations in connectivity conditions. While some of the settings are customizable for LTE and WiMAX, generally these IP-CANs generate a schedule every 10 milliseconds. HSPA typically generates a schedule every 2 milliseconds and EVDO generates a schedule every 26.67 milliseconds. Wi-Fi typically assigns a channel to the user for 0.5 milliseconds to send one data frame (which includes the DIFS, Data, SIFS, ACK mechanism). The intent of the proposed solution is to let independent schedulers of each IP-CAN work at their defined schedule interval durations. The GRC performs global-level optimization (re-associations) and has to operate at larger

[1]Registration with an IMS application server involves a combination of DNS lookups with Diameter authentication procedures (RFC 3588), and SIP signaling.

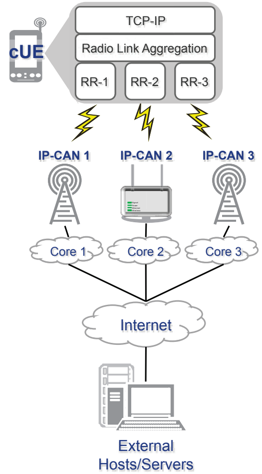

Fig. 3. Data Transmission Plane

time scales to account for issues such as overhead/result propagation delay. So, to minimize actual overhead and to make sure that the cUEs and BSs of various IP-CANs can use the decisions made by the GRC, a scheduling interval of 1 second is proposed for the GRC and is used throughout the paper.

B. Optimization Attribute Utility Functions

The GRC uses a multi-attribute resource allocation algorithm to determine the cUE-IP-CAN mappings and the rate assignment per mapping for each scheduling time step t. The attributes considered in this algorithm are system spectral efficiency, both instantaneous and long-term fairness in terms of data rate allocated to each user in the system, and battery lifetime of each user in the system. We describe the utility function for each attribute next using system parameters presented in Table I. The notation extends the notation presented in [16].

1) Spectral Efficiency: The achievable system spectral efficiency for time interval [t, t+1], denoted γ^t, is represented by (1) as the ratio of the aggregate rate allocated to each user in the system at time t to the total spectrum used. The rate allocated to user $u \in U$ at time t, denoted r_u^t, is represented by (2) and depends on two parameters: (i) x_{ua}^t - the cUE-IP-CAN assignment variable at time t, (ii) r_{ua}^t - the rate allocated to user $u \in U$ by BS/AP $a \in A$ at time t. For each time step t, the multi-attribute resource allocation algorithm implemented by the GRC computes both these parameters. Note that r_{ua}^t is a function of the resource blocks assigned to user $u \in U$ by BS/AP $a \in A$ at time t and the supported modulation and coding scheme (MCS). A resource block is a minimal resource allocation unit. Different IP-CANs use different terminology when defining a minimal resource allocation unit (for example, Wi-Fi lets users compete for the wireless medium using the CSMA/CA mechanism and lets the contention winner hold the wireless medium for the time necessary to transmit a data frame and ACKs plus any optional control frames associated with virtual carrier sensing, OFMDA-based LTE and WiMAX group twelve consecutive subcarriers in the frequency domain and six or seven symbols in the time domain to form a minimal resource allocation unit). The minimal resource allocation unit used by 3GPP based networks (LTE, HSPA) is called a resource block. To unify terminology across all IP-CANs, this term is chosen to represent a minimal resource allocation unit for all RATs in the paper.

TABLE I
SYSTEM PARAMETERS

Parameter	Description
A	Set of BSs/APs for all IP-CANs
U	Set of Users
BU^t	Set of users that are blocked by the admission control procedure at time t
x_{ua}^t	Assignment variable - Determines whether radio $a \in A$ of user $u \in U$ is on or off at time t
r_{ua}^t	Rate (bits/s) allocated to user $u \in U$ by BS/AP $a \in A$ at time t
$r_{ua,max}^t$	Maximum achievable rate (bits/s) for user $u \in U$ through BS/AP $a \in A$ at time t
$r_{ua,norm}^t$	Normalized $rate \in [0,1]$ allocated to user $u \in U$ by BS/AP $a \in A$ at time t
r_u^t	Total rate allocated to user $u \in U$ at time t
γ^t	Achievable system spectral efficiency (bits/s/Hz) for time interval [t, t+1]
κ	Total spectrum (Hz) used by the system
η_{ua}^t	Maximum data (in bits) that can be transferred by radio $a \in A$ of user $u \in U$ during time interval [t, t+1]
T^t	Vector containing minimum data rate requirement of each user $u \in U$ to support real-time traffic for time interval [t, t+1]
ω_{ua}^t	Total energy consumed (in Joules) by radio $a \in A$ of user $u \in U$ during time interval [t, t+1]
ω_u^t	Total energy consumed (in Joules) by cUE of user $u \in U$ during time interval [t, t+1]
m_u	Maximum number of usable radios for user $u \in U$ for each time step

$$\gamma^t = \frac{\sum_{u \in U} r_u^t}{\kappa} \tag{1}$$

$$r_u^t = \sum_{a \in A} x_{ua}^t * r_{ua}^t \tag{2}$$

Since we assume that the amount of spectrum allocated to each IP-CAN is constant, the total spectrum, κ, used by our system remains constant. So, to maximize the achievable system spectral efficiency, the objective of any network optimization problem is to maximize the sum of the rates allocated to each user subject to total resource usage constraints. This optimization problem has been well studied as the Max-Sum Rate (MSR) optimization problem. The idea behind the MSR optimization objective is to assign each resource block to the user that can make the best use of it. The drawback of the MSR optimization objective is that it is likely that a few users close to the BS, and hence having excellent channels, will be allocated all the system resources. As a result, the MSR optimization objective cannot be used as the only objective in any resource allocation problem. Fairness of resource distribution also has to be taken into account.

However, since the MSR optimization objective results in the highest achievable system spectral efficiency, it can be used as an upper bound in computing the spectral efficiency utility function. Let γ^t_{max} represent the achievable system spectral efficiency for time interval [t, t+1] obtained by solving the MSR optimization problem. Similarly, assuming each available resource block is allocated to some user, the minimum achievable system spectral efficiency results when each resource block is assigned to the user with worst connectivity conditions. Let γ^t_{min} represent this minimum achievable system spectral efficiency for time interval [t, t+1]. Then, γ^t_{max} can be used as a lower bound in computing the spectral efficiency utility function. The normalized system utility γ^t_{util} is then computed using (3). If the achievable system spectral efficiency equals γ^t_{max}, the spectral efficiency utility function corresponds to a value of 1, and if the achievable system spectral efficiency equals γ^t_{min}, the spectral efficiency utility function corresponds to a value of 0.

$$\gamma^t_{util} = \frac{\gamma^t - \gamma^t_{min}}{\gamma^t_{max} - \gamma^t_{min}} \tag{3}$$

2) Fairness: The fairness metric relates to the difference in data rates allocated to each user. The fairness metric can be computed for each scheduling time step to ensure instantaneous fairness or it can be computed over long time-scales to ensure long-term fairness. Since we study heterogeneous wireless systems that support both real-time and best-effort traffic, we analyze our algorithm's performance for both instantaneous fairness that deals with minimum data rate support for real-time traffic and long-term fairness that is related to best-effort traffic.

The first step in our resource allocation procedure is an iterative admission control procedure that satisfies the minimum data rate requirement, $T^t = [T^t_1 ... T^t_{|U|}]$, per scheduling interval of as many users as possible to support real-time traffic. For each GRC scheduling interval t (say, 1 second), any user $u \in U$ that is allocated resources achieves a data rate of at least T^t_u bits/s and can satisfy the needs of his/her real-time applications. The users blocked by the admission control procedure are the dissatisfied users that cannot support the needs of their real-time applications. For each time step, the proportion of satisfied users is used to compute the instantaneous fairness utility function, denoted θ^t_{util}, as described by (4). If no users are blocked, the instantaneous fairness utility equals 1 and if all users are blocked, the instantaneous fairness utility equals 0. The instantaneous fairness utility helps address the needs of real-time traffic such as VoIP and video streaming. For example, the G.729 codec used by VoIP traffic needs to periodically send out variable-sized packet bursts approximately every 20 ms. Since GRC operates on a 1 second interval, the GRC uses average data rate requirements of these variable-sized packet bursts over a period of 1 second to compute the minimum data rate requirement per scheduling interval.

$$\theta^t_{util} = 1 - \frac{|BU^t|}{|U|} \tag{4}$$

TABLE II
ENERGY CONSUMPTION NUMBERS FOR CURRENT ACCESS TECHNOLOGIES (IN JOULES)

	802.11g	**802.16e**	**LTE**	**HSDPA**	**EVDO**
$P_{t,a}$(x KB)	0.007(x)	0.018(x)	0.018(x)	0.025(x)	0.025(x)
$P_{rec,a}$	7.25	12.4	4.68	12.4	11.7
$P_{assoc,a}$	5.9	3.2	3.2	3.5	3.5
$P_{o,a}$	13.15	15.6	7.88	15.9	15.2

The second step in the resource allocation process accounts for long-term fairness by using mechanisms from a Proportional Fair scheduler. Let ϕ_{util} represent the long-term fairness utility function. Then we apply a direct mapping of Jain's Fairness Index [41] to the long-term fairness utility function as presented in (5). In general, since best-effort traffic such as FTP has very lenient or no delay constraints and because the instantaneous fairness utility already computes the fairness metric for small time-scales (or each scheduling step), the long-term fairness utility is computed using rates allocated to each user for all time steps under consideration for any study (thousands of seconds). As in the previous two utility components, the long-term fairness utility is normalized into the interval [0,1].

$$\phi_{util} = \frac{\left(\sum_{u \in U} \sum_t r^t_u\right)^2}{|U| * \sum_{u \in U} \left(\sum_t r^t_u\right)^2} \tag{5}$$

3) Battery Lifetime: The final metric in our multiple-attribute resource allocation algorithm is battery lifetime. We use two components in modeling our energy consumption, which is similar to a linear energy consumption model proposed in [42, 43]. The energy consumption for user $u \in U$ during time interval [t, t+1], denoted as ω^t_u, is computed using (6). The first energy consumption component, $P_{t,a}(x)$, relates to the transfer energy component described in [43] and it depends on η^t_{ua}, the maximum number of data bytes that can be transferred by radio $a \in A$ of user $u \in U$ during time interval [t, t+1]. The second energy component, $P_{o,a}$, represents the overhead energy and has two sub-components. The first sub-component, $P_{rec,a}$, represents the extra energy that is spent by RRs in reconfiguring the hardware to support a particular IP-CAN. We assume an FPGA platform as our RR platform [44] and use the reconfiguration energy numbers that we derived in our previous work [23]. The second sub-component, $P_{assoc,a}$, represents the extra energy that is spent associating with a new IP-CAN and is similar to the ramp energy concept used in [43]. We provide the energy consumption numbers used in this study in Table II.

$$\omega_{util} = \sum_{a \in A} [P_{t,a}(\eta^t_{ua}) + (x^t_{ua} - x^{t-1}_{ua}) * (1 - x^{t-1}_{ua}) * P_{o,a}] \tag{6}$$

The goal of the battery lifetime optimization criterion is to prolong the time each user remains active in the system. To this end, we want to ensure that for each scheduling interval [t, t+1], the overall energy consumed by each user in the system is minimized. We use the same maximum and minimum achievable system spectral efficiency concepts adopted in the spectral efficiency utility function in computing the battery lifetime utility function. Let ω_{max}^t represent the maximum achievable overall energy consumption and let ω_{min}^t represent the minimum achievable overall energy consumption for time interval [t, t+1]. Then the battery lifetime utility function for time t, denoted as ω_{util}^t, is computed using (7). If the achievable overall energy consumption equals ω_{min}^t, the battery lifetime utility function is 1 and if the achievable overall energy consumption equals ω_{max}^t, the battery lifetime utility function is 0. Alternatively, a proportional energy consumption objective could have been used as the battery lifetime optimization criterion in the problem formulation, allowing users with higher battery levels to consume more energy than users with lower battery levels. This approach would lead to approximately equal battery levels for each user over a long time period. However, we do not consider this approach in this study and identify it as a part of future work.

$$\omega_{util}^t = 1 - \frac{\sum_{u \in U} \omega_u^t - \omega_{min}^t}{\omega_{max}^t - \omega_{min}^t} \tag{7}$$

C. Resource Allocation Procedure

In this section, we present the resource allocation procedure that is used by the GRC to come up with cUE-IP-CAN mappings and the rate assignment per mapping. Since our heterogeneous wireless system supports both real-time and best effort traffic, the resource allocation problem follows a two-step approach. In the first step, an iterative admission control policy is implemented to satisfy minimum data rate requirements (for real-time traffic) of as many users in the system as possible. In the second step, the weighted spectral efficiency, long-term fairness, and battery lifetime utility functions (related to best-effort traffic) are maximized, subject to minimum data rate requirements. Algorithm 1 describes the complete resource allocation procedure that is used during each time step t.

Each step (Step 1 and 2) in the algorithm uses a mixed integer linear program (MILP) presented by (9) and (11) respectively. The objective of both MILPs is to determine x_{ua}^t (the assignment variable) and r_{ua}^t (the rates allocated to each radio of each user). The spectral efficiency, long-term fairness and battery lifetime utility functions are then computed using these x_{ua}^t and r_{ua}^t variables using (3), (5) and (7) respectively. Note that the battery lifetime utility function presented in (7) depends on (6) which uses an additional variable η_{ua}^t, the maximum amount of data (in bits) that can be transferred by radio $a \in A$ of user $u \in U$ during the scheduling interval t. Since the GRC scheduler operates on a 1 second basis, η_{ua}^t equals r_{ua}^t in our study.

The goal of the admission control procedure, described by Step 1 in the algorithm, is to determine when a user is blocked and maximize the instantaneous fairness utility metric presented in (4) by minimizing the number of blocked users. The admission control procedure first initializes the list of blocked users at time t (BU^t) to null and sets z, the variable that determines the feasibility of satisfying real-time traffic demands of each user, to be infeasible. Next, it

Algorithm 1 Multi-Attribute Resource Allocation

Step 0: Initialization
if $t == 1$
  $A_u^t \leftarrow 1 \quad \forall u \in U$
  $x_{ua}^{t-1} \leftarrow 0 \quad \forall u \in U, \forall a \in A$
  $\rho \leftarrow 0.10$
end if

Step 1: Admission Control
$BU^t \leftarrow \emptyset, z \leftarrow infeasible$
while z is $infeasible$
  $select$ any $u \in U,\ u \notin BU^t$
  $z \leftarrow solve$ P^* $using$ (9)
  if z is $infeasible$
    $r_{u,max}^t = \sum_{a \in A} r_{ua,max}^t / T_u^t \quad \forall u \in U, u \notin BU^t$
    $u_{drop} \leftarrow u \in arg\ min\{r_{u,max}^t\}$
    $BU^t \leftarrow BU^t \bigcup \{u_{drop}\}$
  end if
end while

Step 2: Multiple-Attribute Optimization
$solve$ MA^* $using$ (11)
$A_u^{t+1} = (1-\rho)A_u^t + \rho r_u^t$

recursively solves optimization problem P^*, using (9), in an effort to find a feasible solution that tries to satisfy the real-time traffic demand of each user using constraint (9b). Note that in formulating P^*, $r_{ua,norm}^t$ is used rather than r_{ua}^t in constraints (9c)-(9f) to avoid non-linear problem formulations. The relationship between r_{ua}^t and $r_{ua,norm}^t$ is described by (8). This relation removes the dependence of r_u^t on two variables, x_{ua}^t and r_{ua}^t as presented in (2). Now, r_u^t only depends on r_{ua}^t, as presented by (9a), as constraint (9d) makes sure that $r_{ua,norm}^t$ (and consequently r_{ua}^t) is greater than zero only if x_{ua}^t equals one. After solving one iteration of P^*, the admission control procedure checks whether a feasible solution is produced. If the solution to P^* is infeasible, the user with the worst achievable data rate to demand ratio is dropped and this user is added to the list of blocked users (BU^t) that are assigned no resource blocks (or are assigned rate 0 as described by constraint (9e). The admission control procedure keeps solving P^* and dropping the user with worst achievable data rate to demand ratio until all users that are to be allocated resources ($u \in U, u \notin BU^t$) can achieve a data rate of at least T_u^t bits/s. This mechanism enables the admission control procedure to block as few users as possible. Once a feasible solution is produced for P^*, the resource allocation procedure moves to Step 2 of the algorithm.

$$r_{ua,norm}^t = \frac{r_{ua}^t}{r_{ua,max}^t} \tag{8}$$

$$P^*: \ max\ r_u^t = \sum_{a \in A} r_{ua}^t \tag{9a}$$

$$s.t. \quad r_u^t \geq T_u^t \qquad \forall u \in U, u \notin BU^t \tag{9b}$$

$$\sum_{u \in U} r_{ua,norm}^t \leq 1 \quad \forall a \in A \tag{9c}$$

$$r_{ua,norm}^t \leq x_{ua}^t \qquad \forall u \in U, u \notin BU^t, \forall a \in A \tag{9d}$$

$$r_{ua,norm}^t = 0 \qquad \forall u \in BU^t, \forall a \in A \tag{9e}$$

$$r^t_{ua,norm} \geq 0 \qquad \forall u \in U, u \notin BU^t, \forall a \in A \quad (9f)$$
$$\sum_{a \in A} x^t_{ua} \leq m_u \qquad \forall u \in U, u \notin BU^t \quad (9g)$$
$$x^t_{ua} \in \{0,1\} \qquad \forall u \in U, u \notin BU^t, \forall a \in A \quad (9h)$$

The final step (Step 2) in the algorithm comes up with cUE-IP-CAN mappings and the rate assignment per mapping based on an optimization function, MA^*, described by (11), that optimizes the weighted spectral efficiency, long-term fairness and energy consumption utility functions subject to the minimum data rate requirements confirmed by the admission control procedure. The utility functions described in (3) and (7) are used in MA^* to maximize system spectral efficiency and minimize overall energy consumption, respectively. For long-term fairness, the utility function described by Jain's fairness index in (5) is non-linear and hard to solve for a large-scale heterogeneous wireless system. As a result, an alternative formulation that uses the ratio of instantaneous to average data rate described in (10) is used to maximize long-term fairness utility[2]. It has been shown that allowing the user with maximum achievable ratio of instantaneous to average data rate to transmit during each time step results in maximizing fairness over long time scales [45]. Again, the maximum and minimum achievable ratios of instantaneous to average data rate are used in (10) to scale the long-term fairness utility between 0 and 1. The algorithm initializes the average data rate of each user $u \in U$, denoted as A^t_u, to 1 during the first time step as described in the initialization step in Algorithm 1. After solving the MA^* optimization problem, the algorithm updates the average data rate of each user over a time window that is dictated by the scalar ρ. The value of this window is commonly set between 0.05 and 0.10 [46]. We set ρ = 0.10 in our work as noted in the initialization step in Algorithm 1.

$$\phi\prime^t_{util} = \frac{\sum_{u \in U}\left[\frac{r^t_u}{A^t_u} - \left(\frac{r^t_u}{A^t_u}\right)_{min}\right]}{\sum_{u \in U}\left[\left(\frac{r^t_u}{A^t_u}\right)_{max} - \left(\frac{r^t_u}{A^t_u}\right)_{min}\right]} \quad (10)$$

$$MA^* : max\ (\alpha * \gamma^t_{util}) + (\beta * \phi\prime^t_{util}) + (\tau * \omega^t_{util}) \quad (11a)$$
$$s.t. \quad r^t_u \geq T^t_u \qquad \forall u \in U, u \notin BU^t \quad (11b)$$
$$\sum_{u \in U} r^t_{ua,norm} \leq 1 \quad \forall a \in A \quad (11c)$$
$$r^t_{ua,norm} \leq x^t_{ua} \qquad \forall u \in U, u \notin BU^t, \forall a \in A \quad (11d)$$
$$r^t_{ua,norm} = 0 \qquad \forall u \in BU^t, \forall a \in A \quad (11e)$$
$$r^t_{ua,norm} \geq 0 \qquad \forall u \in U, u \notin BU^t, \forall a \in A \quad (11f)$$
$$\sum_{a \in A} x^t_{ua} \leq m_u \qquad \forall u \in U, u \notin BU^t \quad (11g)$$
$$x^t_{ua} \in \{0,1\} \qquad \forall u \in U, u \notin BU^t, \forall a \in A \quad (11h)$$

Note that as stated earlier, we assume a cUE can use multiple radios concurrently. The maximum number of radios that a cUE can concurrently use is limited by the m_u variable presented in (9g) and (11g). In our problem formulation, we assume m_u to be the same for each user. There might be cases where the value of m_u can vary for different users. For example, if a user does not have enough power to support more than one physical link (i.e. the user is operating at a low battery level), then a policy-based addition can be included in the algorithm that limits such a user to use only one of its radios. These policy-based decisions represent a possible extension to our current model.

[2]Note that $\phi\prime^t_{util}$ presented in (10) is only used in solving MA^*. ϕ_{util} representing Jain's fairness index in (5) is still used in computing long-term fairness utility.

TABLE III
AHP MATRICES DERIVED FROM EXPERT INTERVIEWS

Expert 1				
	Battery Life (BL)	Long-Term Fairness (LTF)	Spectral Efficiency (SE)	Weights
Battery Life	1.0	5.0	0.333	0.279
Long-Term Fairness	0.2	1.0	0.143	0.072
Spectral Efficiency	3.0	7.0	1.0	0.649
Expert 2				
	Battery Life	Long-Term Fairness	Spectral Efficiency	Weights
Battery Life	1.0	5.0	0.500	0.333
Long-Term Fairness	0.2	1.0	0.143	0.075
Spectral Efficiency	2.0	7.0	1.0	0.592

The scalars α, β and τ provide the relative importance of each optimization attribute in MA^* and act as 'control knobs' that allow network operators to achieve the desired performance objectives. The values for these scalars are obtained through AHP [34]. AHP is a decision analysis technique to determine weights of different utility attributes from decision stakeholders through pairwise comparisons and ratings. Using AHP, we interviewed two experts from the cellular industry to perform pairwise comparisons between our utility attributes[3]. After determining which attribute is more important, the more important attribute receives a score from 1-9, with 1 indicating that the two attributes are equally important. These pairwise comparisons are placed in matrix $\mathbf{A}$, with $a_{ji} = 1/a_{ij}$, where each row and column represents a specific attribute. Using the following equation: $\mathbf{Aw} = \lambda_{max}\mathbf{w}$, and solving for λ_{max}, the principal eigenvalue of $\mathbf{A}$, and $\mathbf{w}$, the principal right eigenvector of $\mathbf{A}$, we can normalize the entries of $\mathbf{w}$ by dividing by their sum and recover the weighted values for our utility function.

We asked each expert to compare the relative importance of battery life, fairness, and efficiency [49]. The results of the interview are placed in a comparison matrix, from which the principal eigenvector is calculated. The results from this calculation and resulting weight values are shown in Table III. From Table III, we note that results of AHP show that both experts had relatively similar weight preferences. Consequently, we use results derived from Expert 1's responses in the remainder of the paper.

Since the weight used for each attribute in this paper is based on responses from only two experts, we performed

[3]While in this paper we only examine two viewpoints, we also note that group decision-making and viewpoint aggregation has been studied in [47, 48].

TABLE IV
WEIGHT SENSITIVITY ANALYSIS

Criteria Pair \ Criterion k	BL	LTF	SE
BL-LTF	0.834	0.516	1.234
BL-SE	0.918	2.842	0.968
LTF-SE	0.894	1.291	1.027
Original Weights	0.279	0.072	0.649

TABLE V
CRITERIA PREFERENCE VALUES

Criteria Pair \ Preference Range	BL	LTF	SE
BL-LTF	[0.227, 0.310]	[0.053, 0.095]	[0.608, 0.711]
BL-SE	[0.227, 0.369]	[0.053, 0.095]	[0.552, 0.711]
LTF-SE	[0.227, 0.310]	[0.053, 0.095]	[0.608, 0.711]

sensitivity analysis to determine how much change in weights is required to change the relative criticality of our three optimization criteria (BL, LTF, SE). The criticality degree of a criterion is the smallest percent amount by which the current value of weights must change, such that the existing preference ranking of the criteria will change [50]. The sensitivity coefficient of a criterion is the reciprocal of its criticality degree. Using the methodology in [50] and Expert 1's data presented in Table III, the preference, P, of each criterion is obtained as follows: $P[BL] = 0.2405$, $P[LTF] = 0.0817$, $P[SE] = 0.6778$. So the preference ranking follows the order: $P[SE] > P[BL] > P[LTF]$. The minimum required change in weight of the kth criterion to reverse the criticality (preference order) of two criteria $A_i - A_j$ is defined as $\delta_{k,i,j}$ such that $\delta_{k,i,j} = (P_j - P_i)/(a_{jk} - a_{ik})$. The results of this calculation are presented in Table IV. In order to state that a certain criterion is sensitive to weight change, the change in the weight of the k^{th} criterion, w_k, should satisfy $\delta_{k,i,j} \leq w_k$. However, the results obtained in Table IV show that none of $\delta_{k,i,j}$ is smaller than the corresponding weights (presented in the last row of Table IV), which means that the three criteria are robust and not sensitive to weight change. In other words, any change in the weights will maintain the criticality (or preference order) of each criteria pair. This claim has been verified by changing w_k, for each criterion k, in the range $[w_k - \delta_{k,i,j}, w_k + \delta_{k,i,j}]$ for each $A_i - A_j$ criteria pair and the corresponding results are presented in Table V. As seen from Table V, there is no overlapping between the preference values of the different criteria, and so the criticality is preserved. This result is as expected intuitively as each network expert clearly distinguishes the relative importance of each criterion, as presented in Table III.

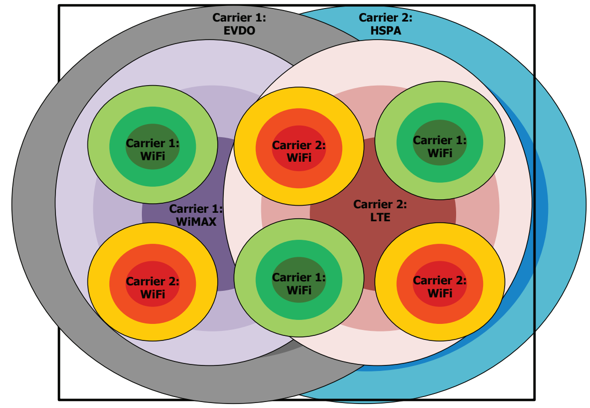

Fig. 4. Coverage of the simulation topology

D. Simulation Description

We developed a MATLAB-based simulation model with sufficient fidelity to demonstrate the properties of our multi-attribute resource allocation algorithm for a heterogeneous wireless system. We consider the presence of two major cellular carriers that operate multiple IP-CANs in a 2 x 2 km^2 area. Carrier 1 operates EVDO (3G), WiMAX (4G), and IEEE 802.11g (Wi-Fi) IP-CANs and Carrier 2 operates HSPA (3G), LTE (4G) and IEEE 802.11g (Wi-Fi) IP-CANs in our experiments. For the cellular based IP-CANs (EVDO, HSPA, WiMAX, LTE), we assume that a single base station serves all users in the simulation topology. These base stations are located near the center of the 2 x 2 km^2 grid. The EVDO and HSPA base stations have a coverage radius of 1.50 km and the WiMAX and LTE base stations have a coverage radius of 1.0 km. The IEEE 802.11g APs are spread throughout the topology and have a coverage radius of 0.15 km each. There are 6 802.11g APs in the topology, with 3 belonging to each carrier. We assume that the IEEE 802.11g APs that the cellular carriers deploy are specialized to participate in the centralized global resource allocation process by using centralized controller mechanisms proposed in the IEEE 802.11e MAC. To account for the overhead associated with the transfer of messages between BSs/APs and the centralized GRC, we deduct 25% resource blocks from each IP-CAN compared to their corresponding theoretical maximum. While we do not model the exact delay and signaling overhead for the log-on process each time the access network changes, our overhead estimate captures the reduction in efficiency that would result from frequent changes in the selection of an access network. The overall simulation topology is presented in Fig. 4.

For the proposed heterogeneous wireless system, we study two different use cases: 1) Use case 1 involves users that can connect only to their own carrier's cellular and Wi-Fi network. 2) Use case 2 allows any user to make use of both carriers' cellular and Wi-Fi networks. While there are economic and public policy obstacles surrounding use case 2, we assume these obstacles will eventually be overcome. For both use cases, the simulation involves 100 nomadic users, 50 of which subscribe to Carrier 1 and the other 50 to Carrier 2. All users are equipped with 3 reconfigurable radios (using notation presented in Table I, $m_u = 3$) to support up to 3 IP-CANs for use case 1 and 5 IP-CANs for use case 2. They move in the simulation topology using a random walk mobility model at a constant speed of 2 mph. Based on the location of a user in the topology, the user can connect to any available IP-CAN using one of the corresponding MCSs provided in our prior work [51]. The closer the user is to a BS/AP, the better the signal reception the user experiences at that location. This translates into a better MCS mapping for the specific IP-CAN under consideration, and this governs the maximum achievable rate for each user via the corresponding BS/AP. The different color shades in Fig. 4 represent an example MCS mapping for various IP-CANs, where the darker the MCS, the higher the order of MCS any user can use in a given location. Furthermore, since we do not use a detailed channel model in our studies to determine the MCS, we model the fluctuations in connectivity conditions by randomly turning off each reconfigurable radio of each user for 5% of the simulation time.

In our simulation, each use case scenario is simulated for 10,000 seconds. The GRC implements a scheduler that comes up with cUE-IP-CAN mappings and the corresponding rate assignment per mapping every one second. The GRC scheduler follows the two-step approach described in Algorithm 1. In the second step, while solving the MA^* optimization problem described by (11), the weights used for spectral efficiency utility (α), long-term fairness utility (β) and battery life utility (τ) are 0.649, 0.072 and 0.279, respectively, which correspond to the weights derived from Expert 1's interview responses presented in Table III.

IV. Results And Analysis

We first present results for when wireless data networks only support best-effort traffic. For this case, there is no minimum data rate requirement for any user. In other words, T_u^t=0 for all users in the system. Since T_u^t=0, the admission control procedure does not block any user for any scheduling time step and is not needed. As a result, the instantaneous fairness utility metric is not computed for this case. The overall utility function only depends on the spectral efficiency utility (γ_{util}), long-term fairness utility (ϕ_{util}) and energy consumption utility (ω_{util}), averaged over the entire simulation run, and is calculated using (12) where α = 0.649, β = 0.072, and τ = 0.279. We provide the overall utility results with each of the three utility components for use case 1 in Fig. 5. Since the overall utility is a normalized value in range [0,1], the results for use case 2 are very similar to the results for use case 1 and are thus omitted in this paper. Optimization problems presented in (9) and (11), which are parts of the proposed algorithm, are solved using AMPL modeling language and CPLEX optimization solver [52, 53].

$$Overall_{util,BE} = (\alpha * \gamma_{util}) + (\beta * \phi_{util}) + (\tau * \omega_{util}) \quad (12)$$

In addition to the utility results for our multi-attribute resource allocation algorithm, we provide results for four commonly used scheduling algorithms for wireless data networks:

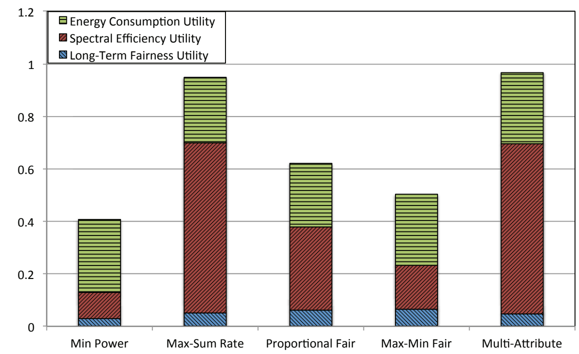

Fig. 5. Overall Utility for Use Case 1, T_u^t = 0

(i) Min Power (ii) Max-Sum Rate (iii) Proportional Fair and (iv) Max-Min Fair. Note that the first three algorithms reduce to our MA^* optimization problem if we set (i) α = 0, β = 0, τ = 1 (ii) α = 1, β = 0, τ = 0 and (iii) α = 0, β = 1, τ = 0 respectively in (11a). The Max-Min Fair results are obtained using the progressive filling algorithm [54]. Furthermore, the Max-Sum Rate algorithm always achieves the highest system spectral efficiency and as a result its γ_{util} = 1. However, because of more connectivity options for use case 2, the average spectral efficiency for use case 2 is 4.35 bits/s/Hz compared to 3.52 bits/s/Hz for use case 1. Similar to the Max-Sum Rate algorithm, the Min Power algorithm always produces the minimum possible energy consumption and therefore its ω_{util} = 1. But the average energy consumption per user is 9.6 mJ/s for use case 1 compared to 10.4 mJ/s for use case 2. All other algorithms compute their spectral efficiency utility relative to the Max-Sum Rate algorithm's spectral efficiency utility as described by (3) and their energy consumption utility relative to the Min Power algorithm's energy consumption utility as described by (7).

The overall utility of our multi-attribute resource allocation algorithm is very similar to the overall utility of the Max-Sum Rate algorithm (0.967 compared to 0.948) as seen from Fig. 5. Since the spectral efficiency utility is given the highest weight in our overall utility function, this result follows expectations. In comparison to the Max-Sum Rate algorithm, our algorithm improves the energy consumption utility (0.269 compared to 0.247) at the cost of a slight degradation in spectral efficiency utility (0.648 compared to 0.649). The long-term fairness utility is almost the same for our algorithm and the Max-Sum Rate algorithm (approximately 0.050). All other algorithms (Min Power, Proportional Fair, Max-Min Fair) sacrifice spectral efficiency in trying to achieve other objectives, as seen in Fig. 5, and as a result their overall utility is much lower than the one obtained by our algorithm.

We now consider the case of next-generation heterogeneous wireless networks that are expected to support both real-time and best-effort traffic. In this case, the overall utility function depends on utility attributes that apply to real-time traffic and the attributes that apply to best-effort traffic. We equally weigh the utilities of both traffic types to compute the overall utility function. The best-effort traffic utility, denoted $Overall_{util,BE}$, depends on spectral efficiency, long-term fairness and energy consumption utilities as presented in (12). The real-time traffic depends on the instantaneous fairness utility averaged over the entire simulation run, denoted θ_{util}, and is calculated using (4). Hence, the overall utility function is computed using (13).

$$Overall_{util,BE+RT} = \frac{1}{2} * Overall_{util,BE} + \frac{1}{2} * \theta_{util} \quad (13)$$

For the mix of real-time and non-real-time traffic, we present results for both use case 1 and use case 2 for varying levels of minimum data rate requirement of each user to support real-time traffic, T_u^t, using Fig. 6 and Fig. 7 respectively. Considering the case when T_u^t = 512K, the overall utility of our algorithm for both use cases is significantly higher than that of any other algorithm. For both use cases, the overall energy consumption utility and long-term fairness utility of all algorithms are similar. But the difference in overall utility is obtained due to instantaneous fairness and spectral efficiency utilities. For use case 1, in terms of overall utility performance, our algorithm outperforms the next closest algorithm, Max-Sum Rate, by 56.7% (0.818 compared to 0.522). The spectral efficiency utility of our algorithm for best-effort traffic decreases compared to Max-Sum Rate algorithm (0.224 compared to 0.325). But this happens as a result of satisfying more real-time traffic users. The instantaneous fairness utility of our algorithm is significantly higher than that of the Max-Sum Rate algorithm (0.437 compared to 0.048). For use case 2, our algorithm outperforms the next closest algorithm, Max-Min Fair, in terms of overall utility by 24.0% (0.975 compared to 0.786). The instantaneous fairness utility of both algorithms is 0.5. But the spectral efficiency utility of our algorithm is significantly higher compared to the Max-Min Fair algorithm's spectral efficiency utility (0.310 compared to 0.115). The result suggests that for future heterogeneous wireless systems that must support both real-time and best-effort traffic, our algorithm obtains the best of both worlds by applying the right trade-offs in terms of achieved spectral efficiency and instantaneous fairness.

Also note that for all different levels of T_u^t for both use cases, our algorithm outperforms any other algorithm. None of the other algorithms is suited to support both best-effort and real-time traffic. While Max-Sum Rate and Proportional Fair algorithms are well suited for achieving good spectral efficiency for best-effort traffic, they do not provide acceptable levels of instantaneous fairness. On the other hand, the Max-Min Fair algorithm provides good instantaneous fairness, but its spectral efficiency suffers significantly. Our algorithm achieves a balance in both instantaneous fairness and spectral efficiency utilities. Apart from this, there are two additional observations of interest in Fig. 6 and Fig. 7. First, while most traditional algorithms provide constant overall utility levels and then possibly experience sudden drops in performance (for example, Max-Min Fair algorithm for use case 2), our algorithm degrades gradually as the available resources cannot satisfy the demands. Second, since use case 2 represents more connectivity options for each user, the resulting overall utility of our algorithm is considerably higher (by up to 39.4%) compared to use case 1 for higher levels of T_u^t ($T_u^t \geq$ 512 kbps). So increasing the number of connectivity options (possibly through peering agreements among several network service providers) has significant performance benefits.

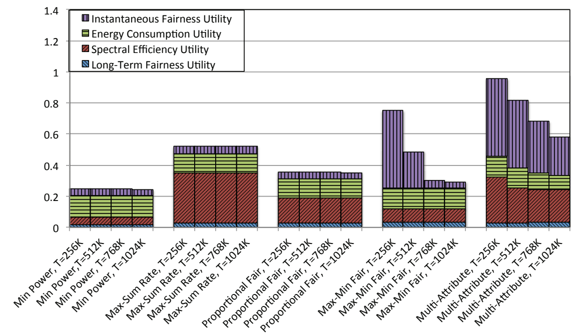

Fig. 6. Overall Utility for Use Case 1, Variable T_u^t

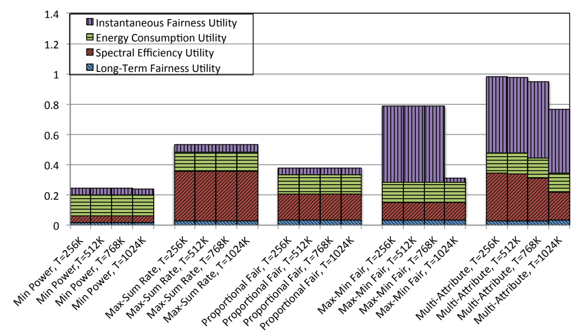

Fig. 7. Overall utility for Use Case 2, Variable T_u^t

V. Conclusions and Future Work

We presented an approach to managing resources in a heterogeneous wireless network based on the 3GPP IMS architecture that supports reconfigurable devices. We analyzed our multi-attribute scheduling algorithm implemented by a centralized GRC that considered the network-efficiency measures of system spectral efficiency, both instantaneous and long-term fairness in terms of data rate allocated to each user in the system, and battery lifetime of each user in the system. Through Matlab/CPLEX based simulations, we showed an increase in overall utility of up to 57% for our algorithm compared to the next best algorithm. By following a two-step resource allocation procedure, depending on the situation, our algorithm improves the overall system performance by achieving the right trade-offs in terms of system spectral efficiency and energy consumption (for best-effort traffic) or by achieving the best trade-offs in terms of system spectral efficiency and instantaneous fairness (for real-time traffic).

As a part of future work, we intend to look at various incentive and economic models that foster network provider cooperation to achieve an increase in overall system performance using the same amount of resources that are used today. Also note that the $P_{rec,a}$ numbers used in this study were based on the assumption that the reconfigurable radio is completely manufactured using the Field Programmable

Gate Array (FPGA) technology. As described in our previous works [23, 55], these numbers are not absolute and can vary, perhaps due to hardware technology advancements or different Application Specific Integrated Circuit (ASIC) vs. FPGA implementation percentages for the reconfigurable radio. A scalar $\lambda \in [0, 1]$, which we termed as 'impact of reconfiguration' in our previous works, can be multiplied to $P_{rec,a}$ to capture the effects of this variation. Evaluating differences in energy consumption as a result of the introduction of 'impact of reconfiguration' scalar in our model also remains a part of our future work. In addition, we are extending the work described in this paper by considering more sophisticated mobility and channel models.

References

[1] FCC, "ET Docket No. 03-222, Notice of Proposed Rule Making and Order," Dec. 2003.

[2] R. Research, "HSPA to LTE-Advanced," Sep. 2009.

[3] I. Mansfield, "Smartphone battery life has become a significant drain on customer satisfaction and loyalty," April 2012. [Online]. Available: http://www.cellular-news.com/story/53523.php

[4] F. Haider, C.-X. Wang, X. Hong, H. Mischak, D. Yuan, E. Hepsaydir, and H. Haas, "Spectral-energy efficiency tradeoff in cognitive radio networks with peak interference power constraints," in *Proceedings of IEEE ICCT*, Sep. 2011, pp. 368–372.

[5] C.-X. Wang, X. Hong, H.-H. Chen, and J. S. Thompson, "On capacity of cognitive radio networks with average interference power constraints," *IEEE Transactions on Wireless Communications*, vol. 8, no. 4, 2009.

[6] X. Kang, Y.-C. Liang, and A. Nallanathan, "Optimal power allocation for fading channels in cognitive radio networks under transmit and interference power constraints," in *Proceedings of IEEE ICC*, 2008.

[7] D. López-Pérez, I. Güvenç, G. de la Roche, M. Kountouris, T. Q. S. Quek, and J. Zhang, "Enhanced intercell interference coordination challenges in heterogeneous networks," *IEEE Wireless Communications Magazine*, vol. 18, no. 3, pp. 22–30, June 2011.

[8] M. Sawahashi, Y. Kishiyama, A. Morimoto, D. Nishikawa, and M. Tanno, "Coordinated multipoint transmission/reception techniques for LTE-advanced," *Wireless Communications*, vol. 17, no. 3, June 2010.

[9] J. Deaton, M. Benonis, L. DaSilva, and R. Irwin, "Supporting Dynamic Spectrum Access in heterogeneous LTE+ networks," *IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, 2012.

[10] Y. J. Zhang and K. B. Letaief, "Multiuser adaptive subcarrier-and-bit allocation with adaptive cell selection for OFDM systems," *IEEE Transactions on Wireless Communications*, vol. 3, no. 5, Sep. 2004.

[11] W. Rhee and J. M. Cioffi, "Increase in capacity of multiuser OFDM system using dynamic subchannel allocation," in *Proceedings of IEEE Vehicular Technology Conference*, May 2000, pp. 1085–1089.

[12] D. Tse, "Multiuser diversity in wireless networks," Wireless Communications Seminar, Stanford University, April 2001.

[13] J.-M. Liang, J.-J. Chen, C.-W. Liu, Y.-C. Tseng, and B.-S. P. Lin, "On tile-and-energy allocation in OFDMA broadband wireless networks," *IEEE Communications Letters*, vol. 15, no. 12, Dec. 2011.

[14] F. Meshkati, H. V. Poor, S. C. Schwartz, and R. V. Balan, "Energy-efficient resource allocation in wireless networks with quality-of-service constraints," *IEEE Transactions on Communications*, vol. 57, no. 11, pp. 3406–3414, Nov. 2009.

[15] P. Serrano, M. Hollick, and A. Banchs, "On the trade-off between throughput maximization and energy consumption minimization in IEEE 802.11 WLANs," *Journal of Communications and Networks*, vol. 12, no. 2, pp. 150–157, April 2010.

[16] T. Bu, L. Li, and R. Ramjee, "Generalized proportional fair scheduling in third generation wireless data networks," in *Proceedings of IEEE INFOCOM*, 2006.

[17] L. Gavrilovska and V. Atanasovski, "Resource management in wireless heterogeneous networks (WHNs)," *Telsiks*, Oct. 2009.

[18] IEEE, "Std. 802.21-2008: IEEE Standard for Local and Metropolitan Area Networks, Part 21: Media Independent Handover Services," Jan. 2009.

[19] S. Buljore, H. Harada, S. Filin, P. Houze, K. Tsagkaris, O. Holland, K. Nolte, T. Farnham, and V. Ivanov, "Architecture and enablers for optimized radio resource usage in heterogeneous wireless access networks: the IEEE 1900.4 working group," *IEEE Communications Magazine*, vol. 47, no. 1, Jan 2009.

[20] A. de la Oliva, C. J. Bernardos, M. Calderón, T. Melia, and J. C. Zúñiga, "Ip flow mobility: smart traffic offload for future wireless networks," *IEEE Communications Magazine*, vol. 49, no. 10, Oct 2011.

[21] C. Bernardos, "Proxy mobile IPv6 extensions to support flow mobility," IETF, draft-ietf-netext-pmipv6-flow-mob-01, work in progress, Sep. 2011.

[22] 3rd Generation Partnership Project, "3GPP TS 23.402," Architecture enhancements for non-3GPP accesses, 2011.

[23] J. Martin, R. Amin, A. Eltawil, and A. Hussien, "Using reconfigurable devices to maximize spectral efficiency in future heterogeneous wireless systems," in *Proceedings of IEEE ICCCN*, Aug. 2011.

[24] R. Marler and J. Arora, "Survey of multi-objective optimization methods for engineering," *Structural and Multidisciplinary Optimization*, vol. 26, pp. 369–295, 2004.

[25] F. Kelly, A. Maulloo, and D. Tan, "Rate control for communication networks: shadow prices, proportional fairness and stability," *Operational Research Society*, vol. 49, no. 3, 1998.

[26] S. Lee, K. Sriram, K. Kim, Y. H. Kim, and N. Golmie, "Vertical handoff decision algorithms for providing optimized performance in heterogeneous wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 2, pp. 865–881, 2009.

[27] P. Kosmides, A. Rouskas, and M. Anagnostou, "Network selection in heterogeneous wireless environments," in *Proceedings of IEEE International Conference on Telecommunications*, 2011.

[28] A. Calvagna and G. Di Modica, "A user-centric analysis of vertical handovers," in *Proceedings of the 2nd ACM international workshop on Wireless mobile applications and services on WLAN hotspots*, 2004.

[29] E. Stevens-Navarro and V. W. S. Wong, "Comparison between vertical handoff decision algorithms for heterogeneous wireless networks," in *Proceedings of IEEE Vehicular Technology Conference*, 2006.

[30] P. M. Chan, R. E. Sheriff, Y. F. Hu, P. Conforto, and C. Tocci, "Mobility management incorporating fuzzy logic for a heterogeneous IP environment," *IEEE Communications Magazine*, vol. 39, no. 12, pp. 42–51, Dec. 2001.

[31] K. Radhika and A. V. Reddy, "Network selection in heterogeneous wireless networks based on fuzzy multiple criteria decision making," *International Journal of Computer Applications*, vol. 22, no. 1, May 2011.

[32] W. Luo and E. Bodanese, "Optimising radio access in a heterogeneous wireless network environment," in *Proceedings of IEEE ICC*, 2009.

[33] B. Bakmaz, Z. Bojkovic, and M. Bakmaz, "Network selection algorithm for heterogeneous wireless environment," in *Proceedings of PIMRC*, 2007.

[34] T. Saaty, "How to make a decision: the analytic hierarchy process," *European Journal of Operational Research*, vol. 48, 1990.

[35] C. J. Chang, T. L. Tsai, and Y. H. Chen, "Utility and game-theory based network selection scheme in heterogeneous wireless networks," in *Proceedings of IEEE WCNC*, 2009.

[36] J. Antoniou and A. Pitsillides, "4G converged environment: modeling network selection as a game," in *Proceedings of 16th IST Mobile and Wireless Communications Summit*, 2007.

[37] D. Niyato and E. Hossain, "Dynamics of network selection in heterogeneous wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 4, pp. 2008–2017, May 2009.

[38] E. Altman, T. Boulogne, R. El-Azouzi, T. Jimenez, and L. Wynter, "A survey on networking games in telecommunications," *Computers and operations Research*, vol. 33, pp. 286–311, 2006.

[39] J. Nelson, "Verizon wireless fosters global LTE ecosystems," Verizon Press Release, Feb. 2009.

[40] 3rd Generation Partnership Project, "3GPP TS 23.228 V11.3.0," IP Multimedia Subsystem (IMS); Stage 2 (Release 11), Dec. 2011.

[41] R. Jain, D. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," DEC Research Report TR-301, Sep. 1984.

[42] L. M. Feeney and M. Nilsson, "Investigating the energy consumption of wireless network interface in an ad hoc networking environment," in *Proceedings of IEEE INFOCOM*, 2001.

[43] N. Balasubramanian, A. Balasubramanian, and A. Venkataramani, "Energy consumption in mobile phones: a measurement study and implications for network applications," in *Proceedings of the 9th ACM SIGCOMM Internet Measurement Conference*, 2009.

[44] A. Saha and A. Sinha, "An FPGA based architecture of a novel reconfigurable radio processor for software defined radio," in *Proceedings of International Conference on Education Technology and Computer*, 2009.

[45] P. Viswanath, D. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Transactions on Information Theory*, vol. 48, June 2002.
[46] J. Andrews, A. Ghosh, and R. Muhamed, *Fundamentals of WiMAX: Understanding Broadband Wireless Networking*. Prentice Hall, NJ., 2007.
[47] E. Forman and K. Peniwati, "Aggregating individual judgments and priorities with the analytic hierarchy process," *Elsevier European Journal of Operational Research*, vol. 108, no. 1, 1998.
[48] R. Ramanathan and L. Ganesh, "Group preference aggregation methods employed in AHP: An evaluation and an intrinsic process for deriving members weightages," *Elsevier European Journal of Operational Research*, vol. 79, no. 2, 1994.
[49] J. D. Deaton, C. Wernz, and L. A. DaSilva, "Decision analysis for Dynamic Spectrum Access rules," in *Proceedings of IEEE Globecom*, Dec. 2011.
[50] E. Triantaphyllou and A. Sanchez, "A sensitivity analysis approach for some deterministic multi-criteria decision making methods," *Decision Sciences*, vol. 28, no. 1, pp. 151–194, 1997.
[51] J. Martin, R. Amin, A. Eltawil, and A. Hussien, "Limitations of 4G wireless systems," in *Proceedings of Virginia Tech Wireless Symposium*, June 2011.
[52] R. Fourer, D. M. Gay, and B. W. Kernighan, *AMPL: A Mathematical Programming Language*. Brooks/Cole-Thomson Learning, 2003.
[53] CPLEX, "ILOG CPLEX 10.0 User's Manual and Reference Manual," ILOG, S.A., 2006.
[54] T. Cover and J. Thomas, *Elements of Information Theory*. Wiley, New York, NY, 1991.
[55] R. Amin, J. Martin, A. Eltawil, and A. Hussien, "Spectral Efficiency and Energy consumption tradeoffs for reconfigurable devices in heterogeneous wireless systems," in *Proceedings of IEEE WCNC*, April 2012.

Rahul Amin is a PhD Candidate in the Holcombe Department of Electrical and Computer Engineering at Clemson University, USA. Since 2009, he has been a Research Assistant in the Networking Lab for the School of Computing at Clemson University and has focused on radio resource management strategies for next generation heterogeneous wireless networks. Before his PhD work, he worked as a Research Assistant in Department of Electrical and Computer Engineering at Clemson University as a Masters student and collaborated on a research project with BMW IT research center in Greenville, SC where he focused on vehicular broadband connectivity project using a WiMAX network deployment. Before graduate school, he held an internship position at Adtran where he worked in the design verification testing and strategic software groups that focused on the development of Adtran's switches, routers and multiplexer products.

Dr. Jim Martin is an Associate Professor in the School of Computing at Clemson University. His research interests include broadband access, wireless networks, Internet protocols, and network performance analysis. Current research projects include heterogeneous wireless systems and DOCSIS 3.x cable access networks. He has received funding from NSF, NASA, the Department of Justice, Cisco, IBM, CableLabs, and BMW. Dr Martin received his Ph.D. from North Carolina State University. Prior to joining Clemson, Dr Martin was a consultant for Gartner, and prior to that, a software engineer for IBM.

Juan Deaton is a Wireless Systems Researcher at the Idaho National Lab (INL), a recent Ph.D. Graduate from Virginia Tech, and is currently investigating techniques for spectrum modeling and prediction to develop a cognitive MAC. His most recent contributions propose: the Spectrum Accountability Framework, which enables DSA in 4G wireless networks and beyond. Results were published in the prestigious 2011/2012 International Symposium on New Frontiers in Dynamic Spectrum Access Networks. Before graduate school, Juan researched options for wireless airborne emergency communications, vulnerabilities of VoIP applications, and security issues with industrial wireless systems for the INL. Before the INL, Juan worked for Motorola's CDMA network division where he wrote system requirements, tested, and coordinated new equipment deployments into commercial markets and was the recipient of the 2003 CEO award for volunteerism. His patented work includes: mobile advertising, network upgrades, and emergency communication systems.

Luiz A. DaSilva is a Professor in the Bradley Department of Electrical and Computer Engineering at Virginia Tech, USA. He also holds the Stokes Professorship in Telecommunications in the Department of Electronic and Electrical Engineering at Trinity College Dublin. His research focuses on distributed and adaptive resource management in wireless networks, and in particular cognitive radio networks and the application of game theory to wireless networks. Prof. DaSilva is currently a principal investigator on research projects funded by the National Science Foundation in the United States, the Science Foundation Ireland, and the European Commission under Framework Programme 7. He is a co-principal investigator of CTVR, the Telecommunications Research Centre in Ireland. He has co-authored two books on wireless communications and in 2006 was named a College of Engineering Faculty Fellow at Virginia Tech.

Amr Hussien received his B.Sc. and M.Sc. degrees from the Electronics and Communications Department, Cairo University, Egypt in 2006 and 2008 respectively. He is currently a PhD Candidate at Electrical Engineering and Computer Science (EECS) Department, University of California, Irvine. He has been a Research Assistant at the Wireless Systems and Circuits Laboratory (WSCL). He has held internship positions at Mindspeed, Newport Beach, CA where he worked in the PHY implementation of dual mode 3G/4G project. His current research interests include low power reconfigurable architectures for future wireless systems and hardware error resilient low power architectures for wireless systems.

Ahmed Eltawil is an Associate Professor at the University of California, Irvine. He received the Doctorate degree from the University of California, Los Angeles, in 2003. Since 2005, he has been with the Department of Electrical Engineering and Computer Science, University of California, Irvine. He is the founder and director of the Wireless Systems and Circuits Laboratory (http://newport.eecs.uci.edu/ aeltawil/). His current research interests are in low power digital circuit and signal processing architectures for wireless communication systems. He received several distinguished awards, including the NSF CAREER award in 2010 supporting his research in low power systems.