Assessing Performance Gains Through Global Resource Control of Heterogeneous Wireless Networks

Rahul Amin and Jim Martin

Abstract—We study the resource allocation and management issues related to heterogeneous wireless systems made up of several radio access technologies (RATs) that collectively provide a unified wireless network to a diverse set of users through co-ordination managed by a centralized global resource controller (GRC). We assume that the user devices are multimodal, which makes it possible for each device to use any available access point (AP)/base station (BS) of a RAT at any given time. Through detailed protocol level simulations performed in ns-2, we show an increase in spectral efficiency of up to 99 percent and an increase in short-term fairness of up to 28.5 percent for two greedy sort-based user device-to-AP/BS association algorithms implemented at the GRC compared to a distributed solution used in practice today where each user makes his/her own association decision. While the increase in overhead due to re-associations for a centralized solution grows only slightly (by up to 4.1 percent) compared to a distributed solution, we find the performance increase in spectral efficiency and short-term fairness attributes come at the cost of an order of magnitude increase (of up to 794 percent) in energy consumption.

Index Terms—Heterogeneous wireless networks, global resource controller, scheduling, resource allocation, network selection

1 Introduction

Due to widespread deployment of wireless access technologies, it is quite common for any geographical location to be covered by more than one wireless network. The number of wireless networks in any given area is expected to grow for at least the following reasons: the trend for ubiquitous IEEE 802.11 (Wi-Fi) access will continue; radio access technologies (RATs) involving open spectrum are likely to become available; and as 4G/5G continues to evolve, the number of cellular RATs will grow. A user device at any given location will have multiple connectivity options determined by the number of radios equipped on the device and in case of cellular networks, the network usage agreement (subscription) the user has signed up for. While emerging user devices are expected to support a multitude of wireless access methods, the current access methods *require the user to select the active access network* either by purchasing an appropriate handset (and service) or, in the case of multimodal smartphones, by manually selecting the access network. Once the user selects the access network, each network attempts to achieve the best performance within its own network, generally ignoring impacts of co-located wireless networks.

Localized resource allocation decisions will usually not lead to optimal resource usage. The work in [1] shows that the 'selfish' resource allocation approach can result in non-Pareto optimal bandwidth allocation as compared to the case where a centralized entity performs network-wide resource allocation. A non-Pareto optimal allocation is one such that, there exists another feasible allocation where at least one user gets more bandwidth, and all others get at least the same bandwidth. Significant improvements in efficiency result when the resource management process jointly considers the distribution of resources across network technologies, reaping the benefits of multiaccess network diversity.

Heterogeneous wireless networks (hetnets) that constitute multiple access networks have been widely studied. Initial research related to hetnets focused on the topic of seamless handovers when a mobile device moved from one access network to another. Mobile IP was the recommended solution, which ensures correct delivery of datagrams when a mobile node moves to foreign networks by using a care-of address that associates the mobile node with its home address. Much of the recent published research related to hetnets focuses on optimizing system operating parameters, such as transmit power, to meet objectives that are based on metrics that quantify fairness, quality-of-service (QoS) constraints, or spectral efficiency outcomes. While optimization-based methods can be used to identify system performance bounds, the methods generally require simplifying assumptions. As we show in our previous work [2], the optimization problem quickly becomes intractable when the model takes into account details that are required to model practical hetnet systems. The work in [3], [4] shows that framing the hetnets optimization problem with the constraint that multimodal devices are limited to a single active radio at any given time (referred to as an 'integral association' constraint) is NP-hard. To avoid complexity, approximation algorithms or heuristic algorithms are commonly used. However, to the best of our knowledge, all

- *R. Amin is with MIT Lincoln Laboratory, 244 Wood Street, Lexington, MA 02420. E-mail: rahul.amin@ll.mit.edu.*
- *J. Martin is with Clemson University, 211 McAdams Hall, Clemson, SC 29634. E-mail: jim.martin@cs.clemson.edu.*

Manuscript received 19 Jan. 2014; revised 6 Mar. 2015; accepted 22 Mar. 2015. Date of publication 30 Mar. 2015; date of current version 4 Jan. 2016.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TMC.2015.2417871

comparable work in the literature makes unrealistic assumptions in crucial areas such as RAT resource allocation behaviors or ignoring potential inaccuracy caused by not considering protocol effects such as overhead due to control messaging. These limitations make it difficult to evaluate the feasibility and applicability of proposed ideas for real networks.

The main contribution of this paper is the design and evaluation of a realistic hetnet system that considers conflicting network objectives involving spectral efficiency, short-term and long-term fairness and overall energy consumption. We develop a hetnets framework that is required to operate with standards-based wireless systems in realistic network environments. In particular, we evaluate two greedy sort-based user device-to-access point (AP)/base station (BS) association algorithms implemented at a centralized global resource controller (GRC) that attempt to maximize system spectral efficiency and short-term fairness. We show that the additional overhead created by a centralized GRC solution is low compared to the distributed solution used in practice today where each user makes his/her own association decision. As in [4], we are interested in quantifying the integration gain as a result of a centralized resource allocation strategy, where the integration gain comes from intelligently allocating resources of all combined RATs as opposed to individual RAT resource allocation. We analyze performance gains for our centralized GRC solution in terms of spectral efficiency, and short-term and long-term fairness compared to the distributed solution. Our results show that an allocation strategy that prioritizes system throughput over fairness will increase spectral efficiency in the range 19-99 percent while reducing long term fairness by 8-17 percent. The improvement, however, comes at the expense of increased energy consumption due to frequent network re-associations. We also identify technical challenges associated with periodic network scans required for obtaining link parameter report used by the user device-to-AP/BS association algorithms implemented at the GRC and provide possible alternatives to remedy the challenges.

The paper is organized as follows. Section 2 presents relevant background for our research. System description is presented in Section 3. Simulation methodology is presented in Section 4. We analyze simulation results in Section 5. Section 6 provides conclusions and identifies possible future work.

2 Background

The joint optimization of the resource allocation process in a cellular (or WLAN) system constrained by combinations of fairness, spectral efficiency and power requirements has been widely studied [5], [6], [7], [8]. More recent effort has gone into the ‘network selection’ problem which describes the method by which a client device or a network determines when to initiate a vertical handoff and which network should be joined. Network selection algorithms for optimal service delivery over user devices capable of connecting with several RATs can be categorized into several strategies: decision function-based strategies, user-centric strategies, multiple attribute decision-making strategies, and 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 between 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 [5], [6]. The user-centric strategies focus on one or more needs of the user to decide on the choice of current access network [7]. 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 [8] with bandwidth, delay, jitter, and bit error rate 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.

While all these optimization frameworks provide guidance on network efficiency objectives that should be considered for a hetnet system, none of them consider the complications that arise due to the heterogeneity of the underlying RATs. The prior work has not modeled different scheduling objectives for different RATs or the overhead required to set up the centralized control. The work in [4] studies the integral association problem for a Wi-Fi/WiMAX hetnet system via a proposed heuristic that achieves a two-approximate solution compared to an optimal global max-min fair solution, but it assumes that both Wi-Fi and WiMAX RATs implement a max-min fair scheduler. This work does not account for overhead related to the message exchange, such as identification of data rates achievable by a user device using all available APs/BSs in range, that is required between user device and AP/BS to solve the association problem.

Several studies have taken into account the overhead associated with a hetnet system [9], [10], [11], [12]. However, these studies focus on seamless transition between RATs rather than the optimal allocation of resources. Moreover, such studies employ a trigger-based mechanism where a user device only sends a link parameter report to the centralized controller if its current connectivity condition (usually based on the RSSI) to a certain AP/BS drops below a certain threshold. In our work, we estimate the overhead required for a centralized solution using an IEEE 802.21 framework where each user device in the system sends a link parameter report for all its radio interfaces periodically. Our modeling incorporates our prior work that developed cost models associated with network re-association in terms of increased energy consumption and communication downtime taking into account various user device assumptions [13], [14]. The hetnet system requires periodic dissemination of link parameter information to reap the maximum

TABLE 1
System Parameters

Parameter	Description
A	Set of APs/BSs for all RATs
U	Set of Users
$r_{ua,max}$	Maximum achievable rate (bits/sec) for user $u \in U$ through AP/BS $a \in A$
r_u	Actual rate (bits/sec) allocated to user $u \in U$
p_{ua}	Percentage of airtime usage user $u \in U$ gets through AP/BS $a \in A$
κ	Total Spectrum (Hz) used by the system
η_{ua}	Data (in bits) transmitted/received by radio $a \in A$ of user $u \in U$
ω_u	Total energy consumed (in Joules) by cUE of user $u \in U$

benefit of multi-access network diversity as GRC can make efficient user device-to-AP/BS association decisions more frequently based on each user's time-varying connectivity conditions.

3 System Description

3.1 System Model

The hetnet system consists of multimodal user devices (also referred to as cognitive User Equipment, or cUE) that can connect to one or more RATs. There are two management entities in our proposed system: 1) A GRC entity present in the backend network; 2) An AP/BS for each RAT. The GRC periodically computes user device-to-AP/BS associations and relays this information to the user devices, instructing them to change network associations as necessary to meet global performance objectives or policy requirements. The GRC makes decisions on large time-scales (seconds) using average statistics assigning each user device to an AP/BS to use for connectivity. The APs/BSs operate over small time-scales (milliseconds) to manage the resources of their corresponding RAT and account for short-term fluctuations in connectivity conditions. The relevant system parameters are presented in Table 1.

We consider the presence of Wi-Fi and WiMAX RATs in our hetnet system. The scheduler implemented at the MAC layer of each RAT deployed in practice has a predefined scheduling objective. The default Wi-Fi MAC achieves a max-min fairness scheduling objective by implementing the distributed coordination function (DCF) specified by the IEEE 802.11 standard with a FIFO queuing mechanism [15], [16]. Other schemes that obtain proportional fairness objective by achieving airtime fairness in Wi-Fi networks have been proposed and can be implemented on Wi-Fi APs that support IEEE 802.11e QoS enhancements [17], [18], [19]. The WiMAX standard leaves the scheduler implementation at the MAC layer up to the equipment/service provider. Max-min fairness and proportional fairness resource allocation schemes have been studied extensively in the literature as a means of sharing resources fairly among all connected users [20], [21], [22], [23]. We use a Deficit Weighted Round Robin (DWRR) scheduler at the WiMAX BS. By tuning the weights assigned to service flow queues to an appropriate value, the WiMAX BS can achieve either max-min fairness or proportional fairness objectives for both short and long time-scales. The GRC takes into account the scheduling objective (max-min fairness or proportional fairness) information for both WiMAX BSs and Wi-Fi APs while computing user device-to-AP/BS association decisions each scheduling interval.

Both WiMAX and Wi-Fi RATs in our system use an adaptive modulation and coding scheme (MCS). The signal strength at which the management/data packets are received on a user device (which is based on the distance of the user device from the corresponding AP/BS) dictate the MCS used by the radios on the device to connect to the corresponding AP/BS. The fast feedback channel quality indicator data burst in the uplink sub-frame has been implemented for the WiMAX MAC to relay the MCS update information to the WiMAX BS. An ACK piggyback mechanism has been implemented for the Wi-Fi MAC to relay the MCS update information to the Wi-Fi AP. The MCS dictates the maximum achievable data rate ($r_{ua,max}$ parameter presented in Table 1) for each radio on each user device. Data packets are transmitted using the data rate specified by the adaptive MCS algorithm. All control messages are transmitted using the most robust MCS, which is binary phase shift keying with a code rate of $^1/_2$ (BPSK 1/2) for both WiMAX and Wi-Fi RATs.

Each user device is equipped with a Wi-Fi and WiMAX radio. We consider the integral association scenario used in practice today. Extensions in the networking stack are required to support multi-radio multi-flow capability assumed with fractional association, where both radios can be used simultaneously.

3.2 Performance Objectives

The GRC uses achievable system spectral efficiency and short-term fairness as performance objectives while coming up with user device-to-AP/BS associations for each scheduling interval. We describe both these performance objectives using system parameters presented in Table 1.

3.2.1 *Spectral Efficiency*

The achievable system spectral efficiency, denoted γ, is represented by (1) as the ratio of the actual rate allocated to each user in the system to the total spectrum used. Since the spectrum (κ) used by the overall hetnet system is constant, achieved spectral efficiency depends on the rate allocated to each user in the system, denoted r_u. $r_u$is represented by (2) and depends on two parameters: (i) p_{ua}—the percentage of airtime usage user $u \in U$ gets through AP/BS $a \in A$, (ii) $r_{ua,max}$—the maximum achievable rate for user $u \in U$ through AP/BS $a \in A$. The percentage of airtime usage each user gets through an AP/BS $a \in A$ depends on the scheduling mechanism implemented by the AP/BS and is calculated using Proposition 1 and 2 for proportional fairness and max-min fairness scheduling objectives respectively

$$\gamma = \frac{\sum_{u \epsilon U} r_u}{\kappa}, \tag{1}$$

$$r_u = \sum_{a \in A} r_{ua,max} * p_{ua}. \tag{2}$$

3.2.2 Fairness

The fairness metric relates to the difference in data rates allocated to each user. The fairness metric, denoted ϕ, can be computed for each scheduling time step to ensure short-term fairness or it can be computed over long timescales to ensure long-term fairness. A direct mapping of Jain's Fairness Index [24] is used to compute both short-term and long-term fairness by considering rates allocated to users over short or long timescales respectively. The formula to calculate the fairness metric is presented in

$$\phi = \frac{\left(\sum_{u \in U} r_u\right)^2}{|U| * \sum_{u \in U} (r_u)^2}. \quad (3)$$

The short-term fairness metric is supposed to capture support for 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 long-term fairness metric captures the ability to support delay tolerant applications such as file transfer and email.

Proposition 1. For a single independent multirate network $a \in A$, proportional fairness is achieved when the percentage of air-time usage (p_{ua}) of all users $u \in U$ connected to network a (represented by $u \in U_a$) is equal, i.e. $p_{ua} = \frac{1}{|U_a|}$.

Proof. The objective function of proportional fairness resource allocation problem is as follows: $\max \sum_{u \in U_a} \ln r_{ua} = \max \sum_{u \in U_a} \ln(r_{ua,max} * p_{ua}) = \max \ln \prod_{u \in U_a} (r_{ua,max} * p_{ua})$. The constraints related to the optimization problem are as follows: $\sum_{u \in Ua} p_{ua} = 1$ and $p_{ua} \geq 0, \forall u \in U_a$. Since $r_{ua,max} > 0$ (and is a constant) and $p_{ua} \geq 0, \forall u \in U_a$, the ln in the PF objective function can be ignored due to its monotonicity. Also, the constant $r_{ua,max}$ term can be omitted from the objective function which reduces to the following: $\max \prod_{u \in U_a} p_{ua}$. For solving the optimization problem given by this objective function and the two constraints presented above, we can successfully ignore the second (or inequality) constraint since that constraint is non-binding as setting any $p_{ua} = 0$ would give an objective function value of -∞ which would clearly not provide the maximum. So, using the objective function and the first (or equality) constraint, we use method of Lagrangian Multipliers to solve the problem as follows: $L(p_{1a}, p_{2a}, \ldots, p_{|U_a|a}, \lambda) = \prod_{u \in U_a} p_{ua} + \lambda(1 - \sum_{u \in U_a} p_{ua})$. Setting the gradient $\nabla L(p_{1a}, p_{2a}, \ldots, p_{|U_a|a}, \lambda) = 0$, we get $\frac{\partial L}{\partial p_{1a}} = \prod_{u \in U_a, u \neq 1} p_{ua} - \lambda = 0$, $\frac{\partial L}{\partial p_{2a}} = \prod_{u \in U_a, u \neq 2} p_{ua} - \lambda = 0 \cdots \frac{\partial L}{\partial p_{|U_a|}} = \prod_{u \in U_a, u \neq |U_a|a} p_{ua} - \lambda = 0$. Solving these set of equations yields $\prod_{u \in U_a, u \neq 1} p_{ua} = \prod_{u \in U_a, u \neq 2} p_{ua} = \cdots = \prod_{u \in U_a, u \neq |U_a|a} p_{ua} = \lambda$ which implies $p_{1a} = p_{2a} = \cdots = p_{|U_a|a}$. Using this result along with the first (or equality) constraint $\sum_{u \in U} p_{ua} = 1$ results in $p_{1a} = p_{2a} = \ldots = p_{|U_a|a} = \frac{1}{|U_a|}$. □

Proposition 2 . *For a single independent multirate network $a \in A$, max-min fairness is achieved when the percentage of air-time usage (p_{va}) of user $v \in U$ connected to network a (represented by $v \in U_a$) is given by $p_{va} = \frac{1}{\sum_{u \in U_a} (\frac{r_{va,max}}{r_{ua,max}})}$.*

Proof. Since we consider a single independent network, there exists only one bottleneck link. For a single bottleneck link, the max-min fairness objective results in equal data rate allocation to each user. Therefore, we obtain the following objective to provide max-min fairness: $r_{1a} = r_{2a} = \cdots = r_{|U_a|a.}$ Therefore, $r_{1a,max} * p_{1a} = r_{2a,max} * p_{2a} = \cdots = r_{|U_a|a,max} * p_{|U_a|a}$. Solving $p_{2a}, \ldots, p_{|U_a|a}$ in terms of p_{1a} yields: $p_{ua} = \frac{r_{1a,max}}{r_{ua,max}} p_{1a} \forall u \in U_a$. Again, the constraints related to the optimization problem are as follows: $\sum_{u \in U_a} p_{ua} = 1$ and $p_{ua} \geq 0, \forall u \in U_a$. Using the value of p_{ua} in terms of p_{1a} obtained from the objective function and the first constraint, we get: $p_{1a}(\sum_{u \varepsilon U_a} \frac{r_{1a,\max}}{r_{ua,\max}}) = 1$ which implies $p_{1a} = \frac{1}{\sum_{u \varepsilon U_a} \frac{r_{1a,\max}}{r_{ua,\max}}}$. We obtain the required values of $x_{ua} \in \{0, 1\}$ for any user $v \in U_a$ by using the relationship between p_{ua} and p_{1a} presented. That is, $p_{va} = \frac{1}{\sum_{u \varepsilon U_a} \frac{r_{va,max}}{r_{ua,max}}}$. □

3.3 Allocation Algorithms

The baseline algorithm, referred to as the distributed algorithm, assumes that a GRC is not involved. Each user device makes its own AP/BS association decision. For this approach, each user picks a RAT according to the norm today: use a Wi-Fi network if available; otherwise, use WiMAX network. After selecting Wi-Fi or WiMAX RAT, the user connects to the AP/BS of the corresponding RAT to which it achieves the highest signal strength. When the user is connected to WiMAX network, the user performs a Wi-Fi link scan every 5 seconds using its unused radio. If a Wi-Fi network is detected, the user switches to Wi-Fi network. When the user is connected to Wi-Fi network, the user is satisfied and does not perform any link scans. If the user receives a link going down (or link down) event while using either Wi-Fi or WiMAX network, the corresponding radio goes into scan mode to search for other available Wi-Fi APs or WiMAX BSs.

We identify two centralized allocation strategies of interest: a 'Max Throughput' strategy that maximizes system spectral efficiency and a 'Max Fairness' strategy that maximizes fairness over the (short-term) five second timescale that it operates at. Each strategy can be achieved if the GRC optimizes the system to maximize (1) or (3) respectively subject to the following constraints: (i) $p_{ua} = \frac{1}{|Ua|} \forall u \in U_a, \forall a \in A_{PF}$, where A_{PF} represents RATs implementing proportional fairness scheduling (ii) $p_{va} = \frac{1}{\sum_{u \varepsilon U_a} \frac{r_{va,max}}{r_{ua,max}}} \forall v \in U_a, \forall a \in A_{MM}$, where A_{MM} represents RATs implementing max-min fairness scheduling (iii) $0 \leq p_{ua} \leq x_{ua} \;\; \forall u \in U$, $\forall a \in A$, where $x_{ua} \in \{0, 1\}$ is the assignment variable that determines if user u $\in U$ is connected to AP/BS a $\in A$ (iv) $\sum_{u \varepsilon U} p_{ua} \leq 1 \;\; \forall u \in U, \forall a \in A$, which ensures percentage of airtime allocated to users by an AP/BS does not exceed 100 percent and (v) $\sum_{a \varepsilon A} x_{ua} = 1 \; \forall u \in U, \forall a \in A$, which ensures the 'integral association' constraint and makes sure each user is connected to exactly one AP/BS. Other works [1], [4] have looked at a relaxation of this problem and have shown that the integrality constraint makes the problem NP-hard. The goal of our work is not to come up with an optimal solution, but rather to use these objectives as guidance in

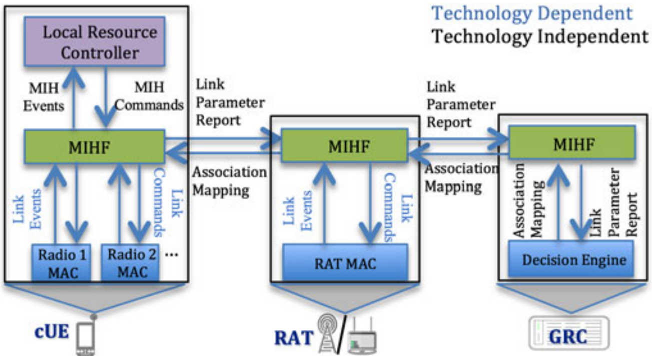

Fig. 1. MIHF implementation.

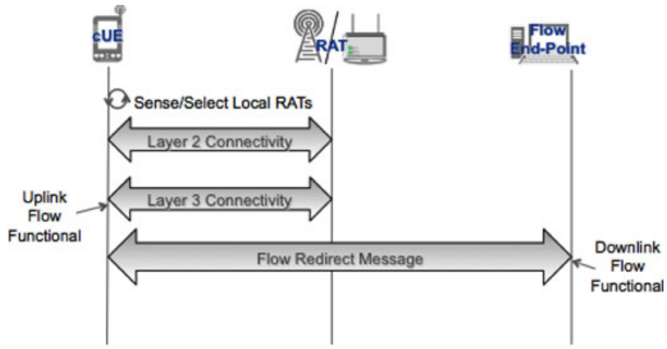

Fig. 2. Procedural flow of a re-association process.

our resource allocation heuristics and compare our proposed centralized control solution to currently used distributed solution through detailed protocol level simulation.

3.4 Centralized Control

The interaction between GRC, AP/BS of each RAT and each user device is handled via Media Independent Handover Function (MIHF) defined by the IEEE 802.21 standard [25]. The MIHF functionality is implemented at Layer 2.5 of the OSI stack as shown in Fig. 1. The MIHF defines three different services: media independent event service (MIES), media independent command service (MICS) and media independent information service (MIIS). MIES provides events triggered by changes in the link characteristic and status. MICS provides the user devices necessary commands to manage and control the link behavior of each radio to accomplish handover functions. MIIS provides information about the neighboring networks and their capabilities. We make use of MIES and MICS functionalities to manage the link-layer (Layer 2 of the OSI stack) network re-associations in our proposed hetnet solution. The messages related to each of these two services that are used in our study are summarized in Tables 2 and 3 respectively.

From the events and commands presented in Tables 2 and 3 respectively, only the association mapping event service and link parameter report command service generate actual packet overhead as messages related to these events/services are exchanged between the cUE and the GRC. The association mapping and link parameter report messages are technology independent and are sent over the radio that is active at the corresponding cUE at the time these triggers are generated. All other messages are locally generated and aid cUE in managing its local interfaces. However, as noted in Table 3, the technology dependent association/scan procedures, which follow a link connect/link scan trigger, might generate technology-specific message overhead.

The procedural flow of a re-association process is shown in Fig. 2. During a re-association process, after a radio establishes link-layer (Layer 2) connectivity, IP connectivity

TABLE 2
Media Independent Event Services

Event Trigger	Trigger Generating Entity	Description
Link Up	cUE	Generated when any radio in cUE establishes link-layer connectivity with an AP/BS.
Link Down	cUE	Generated when any radio in cUE loses connectivity with an AP/BS.
Link Going Down	cUE	Generated when any radio in cUE receives a packet whose signal strength is lower than LGD_Threshold ($= 1.1$) * Rx Threshold.
Link Detected	cUE	Generated by any radio in cUE that receives synchronization messages (beacon packets for Wi-Fi and DL-MP packets for WiMAX) from AP/BS to which it is not currently connected.
Association Mapping	GRC	Generated by the GRC after computing periodic re-associations based on the decision engine (resource allocation procedure). This message is only sent to cUEs whose current network association must change.

TABLE 3
Media Independent Command Services

Command Trigger	Trigger Generating Entity	Description
Link Connect	cUE	Generated when the MIHF in cUE wants one of its radios to establish a data connection with new AP/BS. Once this trigger is received by the corresponding radio, technology dependent association procedure follows.
Link Scan	cUE	Generated when the MIHF in cUE wants one of its radios to scan for APs/BSs. Once this trigger is received by the corresponding radio, technology dependent scanning procedure follows.
Link Parameter Report	cUE	Generated periodically by cUE to send current link parameter status information (such as achievable MCS) related to all of its radios to the GRC.

TABLE 4
Overhead Messages

Message	OSI Layer	Message Contents	Packet Size (Bytes)
Association Mapping	2	New Link Type, New Point-of-Attachment (PoA)	42
Link Parameter Report	2	Current Link Type, Current PoA, [Link Type, PoA, MCS] for all detected RATs OR Current Location	60
Router Solicitation	3	Route Request	48
Router Advertisement	3	Route Reply	96
Flow Redirect Request	3	Redirect IP Address	48

(Layer 3) has to be established before a data flow can be directed to the new connection. We use the neighbor discovery protocol for obtaining an IPv6 address to establish Layer 3 connectivity [26]. Upon establishing Layer 2 connectivity, a 'router solicitation' message is broadcasted by the radio. We assume that the neighbor discovery protocol functionality has been implemented at each AP/BS. When the AP/BS receives the router solicitation message, it sends a 'router advertisement' broadcast packet in response. Upon receiving the router advertisement packet, the cUE uses the prefix information of the router advertisement packet to determine its new IP address. Moreover, to account for cases where a router solicitation/advertisement message is lost, the AP/BS broadcasts the router advertisement packet periodically so that a radio waiting for a new IP address will eventually obtain the required address. Once IP connectivity is established, any flow in the uplink direction can start using the new radio connection. For the flow in the downlink direction, the other end-point of the flow has to be informed of the new IP address that data packets need to be sent on. A 'flow redirect request' message is sent by the cUE to the other end-point to accomplish this task. Upon receiving the flow redirect request message, the other end-point starts sending packets for the corresponding cUE using the new destination IP address. The other end-point sends an ACK packet back to the cUE to inform the cUE of the reception of flow redirect request message. The flow redirect request message is retransmitted by the cUE until an ACK packet is received from the other end-point.

From an overhead perspective, we consider router solicitation, router advertisement responses to the router solicitation messages, and flow redirect request messages as overhead messages required by the hetnet solution. The periodic retransmission of 'router advertisement' packets is defined by the neighbor discovery protocol standard and would apply to any network using this IP address discovery method. Hence, in our analysis we do not consider 'router advertisement' messages as overhead. The information related to Layer 2 and Layer 3 overhead messages used in our study is summarized in Table 4.

3.5 Heuristic Resource Allocation Procedure

We explain in detail the two heuristic algorithms that we have developed. The first, referred to as Algorithm 1, approximates the Max Throughput optimization strategy defined earlier. The second, referred to as Algorithm 2, approximates the Max Fairness strategy. The GRC computes the user device-to-AP/BS association decisions each scheduling interval by considering the independent scheduling objective (max-min fairness or proportional fairness) for both WiMAX and Wi-Fi RATs. While the independently implemented DWRR scheduler at WiMAX MAC converges to a max-min fairness or proportional fairness objective on short time-scales (milliseconds), the DCF-based Wi-Fi MAC converges to max-min fairness objective on larger time-scales (seconds). To allow the Wi-Fi MAC to converge to the max-min fairness solution and also to account for issues such as optimizations (re-association computations) every 5 seconds.

Algorithm 1. Max Throughput

```
system_throughput = −1
max_system_throughput = −1
AP_BS_to_Allocate = −1
for each AP/BS a ∈ A
  if total_throughput_AP_BS[a] != −1
    system_throughput =
        (∑_{A\{a}} achieved_throughput_AP_BS[a])
        + total_throughput_AP_BS[a]
    if system_throughput ≥ max_system_throughput
      if system_throughput = = max_system_throughput
        if lowest_user_throughput[a] >
            lowest_user_throughput[AP_BS_to_Allocate]
          max_system_throughput = system_throughput
          AP_BS_to_Allocate = a
      end if
    else
          max_system_throughput = system_throughput
          AP_BS_to_Allocate = a
    end else
   end if
end if
end for
```

The GRC uses the link parameter report and independent scheduling objective of both WiMAX and Wi-Fi RATs when computing user device-to-AP/BS association decisions each-scheduling interval. From the link parameter report, the GRC can identify the maximum data rate each user can achieve using any AP/BS (depending on the MCS/location the user reported). The information related to maximum data rate that each user can achieve via all available AP/BS is used by GRC in computing the user device-to-AP/BS associations. The pseudo-code for the resource allocation procedure is described in Table 5. The resource allocation procedure calls initialization (Table 6), virtual association (Table 7) and Max Throughput or Max Fairness subroutines described in Algorithms 1 and 2 respectively.

TABLE 5
Resource Allocation Procedure

```
1. Initialization (Table 6)
2. num_allocated_users = 0
3. while num_allocated_users != |U|
4.   for each AP/BS a ∈ A
5.     total_throughput_AP_BS[a] = −1
6.     lowest_user_throughput[a] = −1
7.     curr_index = 0;
8.     while userid(sorted_users_AP_BS[a][curr_index]).
                   association_status == true
9.       curr_index++
10.    end while
11.    if curr_index ≥ sorted_users_AP_BS[a].size()
12.      continue
13.    end if
14.    Compute total throughput and lowest user throughput
       if one more user is added to AP/BS a ∈ A depending
       on its scheduler type (Table 7)
15.  end for
16.  Perform next association decision based on Max
     Throughput(Algorithm 1) or Max Fairness (Algorithm 2)
     objective
17.  curr_index = 0;
18.  while userid(sorted_users_AP_BS[AP_BS_to_Allocate]
                 [curr_index]).association_status == true
19.    curr_index++
20.  end while
21.  userid(sorted_users_AP_BS[AP_BS_to_Allocate]
                 [curr_index]) .association_status = true
22.  users_allocated_AP_BS[AP_BS_to_Allocate]
                 .append(sorted_users_AP_BS[curr_index])
23.  achieved_throughput_AP_BS[AP_BS_to_Allocate] =
                 total_throughput_AP_BS[AP_BS_to_Allocate]
24.  num_users_AP_BS[AP_BS_to_Allocate]++
25.  num_allocated_users++
26. end while
```

Algorithm 2. Max Fairness

```
1. lowest_throughput = −1
2. AP_BS_to_Allocate = −1
3. for each AP/BS a ∈ A
4.   AP_BS_throughput[a] = −1
5.   if lowest_user_throughput[a] != −1
6.     system_throughput[a] =
              (∑_{A\{a}} achieved_throughput_AP_BS[a])
              + total_throughput_AP_BS[a]
7.     if lowest_user_throughput[a] ≥ lowest_throughput
8.       if lowest_user_throughput[a] == lowest_throughput
9.         if AP_BS_throughput[a] >
              AP_BS_throughput[AP_BS_to_Allocate]
10.          lowest_throughput = lowest_user_throughput[a]
11.          AP_BS_to_Allocate = a
12.        end if
13.      else
14.        lowest_throughput = lowest_user_throughput[a]
15.        AP_BS_to_Allocate = a
16.      end else
17.    end if
18.  end if
19. end for
```

The resource allocation procedure attempts to either maximize system spectral efficiency or global short-term fairness as depicted by Step 14 in Table 5. The resource allocation procedure initialization step (Step 1 in Table 5) first sorts each user in descending order for each AP/BS based on the maximum data rate the user can achieve via the corresponding AP/BS. In case of ties, the user with lowest achievable overall data rate over all APs/BSs (and hence having fewer options) is put ahead of the other users. Based on this sorted order, in each decision round (Steps 3-26 of Table 5), the resource allocation procedure computes the achievable total system throughput and lowest user throughput metrics (Step 14 of Table 5) under virtual association scenario (assumption that the best unassociated user for each AP/BS is associated to the corresponding AP/BS) described in Table 7. In performing these computations, the GRC uses the scheduling objective of each AP/BS (proportional fairness or max-min fairness) to determine the percentage of air-time usage (p_{ua}) user $u \in U$ gets through AP/BS $a \in A$ if the next best unassociated user $u \in U$ is associated to AP/BS $a \in A$ as described in Table 7. p_{ua} is determined according to Proposition 1 (Step 2 in Table 7) or Proposition 2 (Step 6 in Table 7) if the scheduling objective of AP/BS $a \in A$ is proportional fairness or max-min fairness respectively. Using p_{ua}, the values for total throughput through AP/BS $a \in A$ and lowest user throughput are computed (Steps 3-4, 7-8 in Table 7). Based on the achievable total throughput through AP/

TABLE 6
Initialization

```
1. for each AP/BS a ∈ A
2.   sorted_users_AP_BS[a].list = Sort(r_{ua,max})
3.   users_allocated_AP_BS[a].list = NULL
4.   num_users_AP_BS[a] = 0
5.   achieved_throughput_AP_BS[a] = 0
6.   total_throughput_AP_BS[a] = −1
7.   lowest_user_throughput[a] = −1
8. end for
```

TABLE 7
Virtual Association Total and Lowest User Throughput

```
1. if a.scheduler == Proportional_Fair
2.   p_{ua} = Calculate according to Proposition 1 where |Ua| =
           num_users_AP_BS[a] + 1
3.   total_throughput_AP_BS[a] = p_{ua} * (sorted_users_AP_BS[a]
                 [curr_index].r_{ua,max} +
                 ∑ users_allocated_AP_BS[a].r_{ua,max})
4.   lowest_user_throughput[a] = p_{ua} * sorted_users_AP_BS[a]
                 [curr_index].r_{ua,max}
5. else if a.scheduler == Max_Min_Fair
6.   p_{va} = Calculate according to Proposition 2 where Ua
           includes all users in users_allocated_AP_BS[a].list
           and v = sorted_users_AP_BS[a][curr_index]
7.   total_throughput_AP_BS[a] = (num_users_AP_BS[a] +
           1) * p_{va} * sorted_users_AP_BS[a][curr_index].
           r_{ua,max}
8.   lowest_user_throughput[a] = total_throughput_AP_BS
     [a]/
9.                               (num_users_AP_BS[a] + 1)
10. end else
```

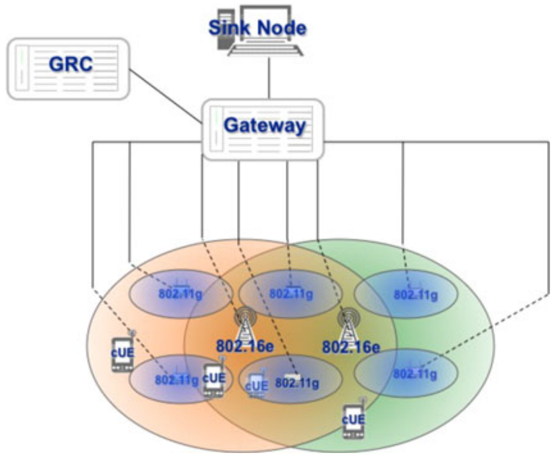

Fig. 3. Simulation Topology in ns-2.

BS $a \in A$ and lowest user throughput computations made for each AP/BS $a \in A$ under virtual associations, the resource allocation procedure makes its next user device-to-AP/BS association decision (Step 16 in Table 5). Algorithm 1 makes decisions based on maximum achievable total system throughput (Steps 7-17 in Algorithm 1) and Algorithm 2 tries to maximize short-term fairness by making decisions based on maximum achievable lowest user throughput (Steps 7-17 in Algorithm 2). In case of ties, each algorithm makes decision based on the other metric (maximum lowest user throughput metric for algorithm trying to maximize system spectral efficiency and maximum achievable total system throughput metric for algorithm trying to maximize short-term fairness as depicted by Steps 8-12 in Algorithm 1 and 2) to break ties. The process of making user device-to-AP/BS association decisions based on the computed achievable total system throughput and lowest user throughput metrics in each decision round continues until all users are associated to a RAT.

TABLE 8
Wi-Fi Simulation Parameters

Parameter	Description
CW_{min}	15
CW_{max}	1,023
Beacon Interval	100 ms
Max Acceptable Beacon Loss	10
RTS/CTS Mechanism	Off
Location of APs (x, y co-ordinates)	(650, 750); (650, 1,250); (1,000, 750); (1,000, 1,250); (1,350, 750); (1,350, 1,250)
Number of Channels	11
Channel Bandwidth	22 MHz
Supported MCS	BPSK1/2, BPSK 3/4, QPSK 1/2, 16-QAM 1/2, 16-QAM 3/4, 64-QAM 2/3, 64-QAM 3/4
MCS Feedback	ACK Piggyback
Coverage Range	150 meters
Propagation Model	TwoRayGround
Scan Duration	1.32 seconds (120 ms for each channel)
Scan Mode	Passive
Link Going Down Factor	1.1

TABLE 9
WiMAX Simulation Parameters

Parameter	Description
Channel Bandwidth	10 MHz
Frame Duration	5 ms
Location of BSs (x, y co-ordinates)	(500, 1,000); (1,500; 1,000)
Scheduler	Deficit Weighted Round Robin
Scan Duration	125 ms (25 frames)
Scan Iterations	1
DL:UL Ratio	3:2
Supported MCS	BPSK1/2, QPSK 1/2, QPSK 3/4, 16-QAM 1/2, 16-QAM 3/4, 64-QAM 2/3, 64-QAM 3/4
MCS Feedback	CQI Channel
Coverage Range	1 kilometer
Propagation Model	TwoRayGround
Link Going Down Factor	1.1

4 Simulation Methodology

We use network simulator, ns-2, to perform our simulation studies. We consider a 2 * 2 km^2 grid where six Wi-Fi APs spread evenly throughout the topology and two WiMAX BSs located near the center of the grid are available to users for data connectivity. The simulation topology we consider for this study is presented in Fig. 3. The coverage range of Wi-Fi AP is 0.15 km and the coverage range of WiMAX BS is 1 km. Note that the two WiMAX BSs have overlapping coverage area. However, both BSs operate on different frequency bands and thus avoid the interference co-ordination problem. There are 100 user devices in the topology. Each user device is equipped with a static Wi-Fi and WiMAX radio that is implemented using low energy consuming ASIC-based hardware. Each user device receives a Constant Bit Rate data flow over TCP transport layer from the sink node. The relevant simulation parameters related to the Wi-Fi RAT, WiMAX RAT and the data flow are presented in Tables 8, 9 and 10 respectively.

All users move in the network topology using one of three user movement patterns: (i) Linear movement pattern where all users move in a straight line starting from [0, 750 m] coordinate in the topology and ending at [2,000, 750 m] coordinate in the topology. Each user is located 1 meter apart from the user in front and behind that user

TABLE 10
Data Flow Simulation Parameters

Parameter	Description
Traffic Direction	Downlink (sink node to cUE)
Transport Protocol	TCP
TCP Flavor	Selective ACK (Sack)
TCP Congestion Control Mechanism	Additive Increase Multiplicative Decrease (AIMD)
Traffic Pattern	Constant Bit Rate
CBR Packet Size	500 Bytes
Packet Interval	0.160 ms
Traffic Generation Rate	25 Mbps

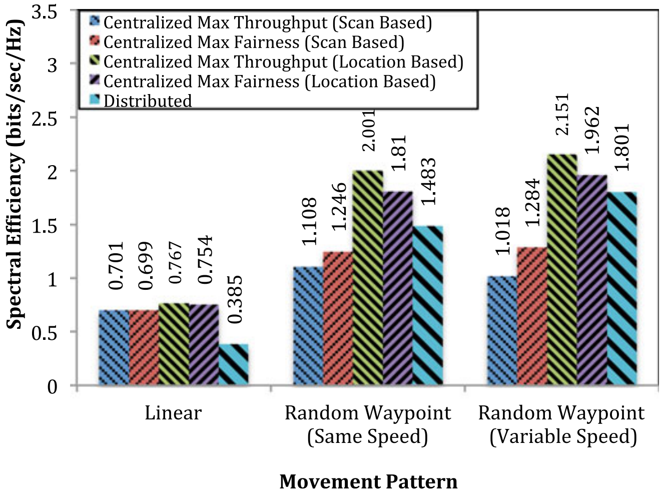

Fig. 4. Spectral efficiency comparisons.

(except for the first and last user). (ii) Random waypoint movement pattern where all users move throughout the topology by picking a destination based on generating uniformly distributed random waypoints. Each user moves at a constant speed of 2 mph. (iii) Random waypoint movement pattern where each user selects a speed in the range [2], [20] mph according to a uniform distribution while moving between the current waypoint and the next waypoint.

Extensions to the mobility package provided by NIST [27] are made to implement the MIHF functionality and the centralized control described in Section 3.4. For the proposed centralized approach, each user device connects to the AP/BS according to the decision made by the GRC. Each user device periodically (on a 5-second basis) sends link parameter report to the GRC to inform the GRC of the available APs/BSs and the associated MCS the user can use to connect to those APs/BSs. To obtain the link parameter status for each available AP/BS, the user device employs one of the following two solutions: (a) Periodic Scanning (on a 5-second basis) on both its radios to search for Wi-Fi APs and WiMAX BSs (b) Location-based solution where the user sends its current location in the link parameter report. The GRC maintains a database of information related to MCS achievable with all available AP/BS by the user device at any given location. Using the link parameter report, the GRC computes the re-association decisions on a 5-second basis based on the sort-based heuristic algorithms presented in the Section 3.5. Note that if a link parameter report packet for any user device is lost (which can happen due to collisions if the packet is transmitted via a Wi-Fi connection), the GRC uses the most recent link parameter report it obtained successfully from that user device. Also, for the centralized association decision solution, if any user device receives a link going down (or link down) trigger, it does not wait for the next GRC re-association computation (and the subsequent report) to switch APs/BS. It automatically goes into scan mode on both radio interfaces and establishes a connection with an available AP/BS.

5 Results and Analysis

We first assess the benefits of the centralized solution implemented at the GRC compared to the distributed solution in terms of achieved system spectral efficiency. Each simulation is run in ns-2 for 2,000 seconds. The pattern for achieved results in terms of all network efficiency measures for both WiMAX MAC implementations (proportional fairness and max-min fairness objectives) is similar. Hence, we only provide results using the proportional fairness WiMAX MAC implementation for our work. Note that proportional fairness WiMAX MAC achieves higher throughput (and spectral efficiency) at the expense of short-term and long-term fairness. Also note that Wi-Fi MAC implementation is set to the default IEEE 802.11g behavior and is always assumed to achieve max-min fairness in our solution. The spectral efficiency comparisons for each solution combination (centralized or distributed decision making and each resource allocation procedure) are presented in Fig. 4. The results shown in Fig. 4 present the average system spectral efficiency averaged over entire simulation duration. The average spectral efficiency metric is computed as the ratio of average throughput achieved by each user in the system to the total spectrum managed by the hetnet system.

5.1 Spectral Efficiency Results

As can be seen from Fig. 4, the location based centralized solutions with both Algorithm 1 (Max Throughput) and Algorithm 2 (Max Fairness) resource allocation procedures outperform the distributed solution in terms of spectral efficiency for all movement patterns due to the benefits of multiaccess network diversity. The gain in spectral efficiency for the centralized Max Throughput resource allocation procedure compared to the distributed solution is 99.2 percent (from 0.385 to 0.767 bits/sec/Hz), 34.9 percent (from 1.483 to 2.001 bits/sec/Hz) and 19.4 percent (from 1.801 to 2.151 bits/sec/Hz) for linear, random waypoint same speed and random waypoint variable speed movement patterns respectively. The gain in spectral efficiency for the centralized Max Fairness resource allocation procedure compared to the distributed solution is 95.8 percent (from 0.385 to 0.754 bits/sec/Hz), 22.0 percent (from 1.483 to 1.81 bits/sec/Hz) and 8.9 percent (from 1.801 to 1.962 bits/sec/Hz) for the linear, random waypoint same speed and random waypoint variable speed movement patterns respectively. As expected, the gain in spectral efficiency is higher for the resource allocation algorithm trying to maximize system spectral efficiency (as compared to short-term fairness). The highest gain in spectral efficiency for a centralized solution (for both resource allocation procedures) occurs for the linear movement pattern where all the users are grouped together and experience similar connectivity conditions and the lowest gain occurs for the random waypoint (variable speed) movement pattern where all the users (because of the randomness in their movement patterns) experience the most frequent change in connectivity conditions. For linear movement pattern, all users experience similar connectivity conditions (for example, one Wi-Fi AP and one WiMAX BS is available to all users at the same time), and as a result the distributed algorithm performs very poorly as each user for this solution selects the Wi-Fi access network for data connectivity. The centralized solution (both resource allocation algorithms) intelligently associates some users to Wi-Fi AP and other users to WiMAX BS and as a result achieves

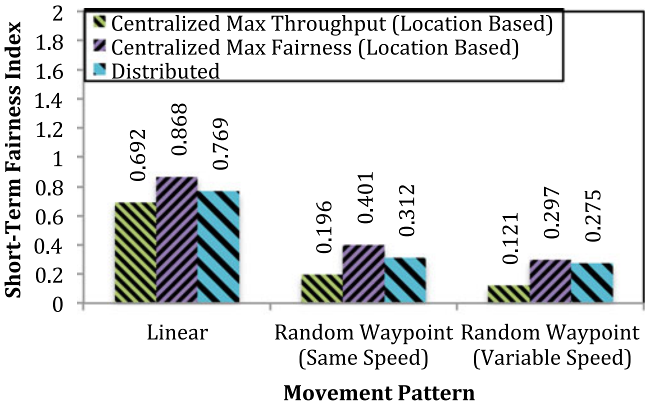

Fig. 5. Short-term fairness comparisons.

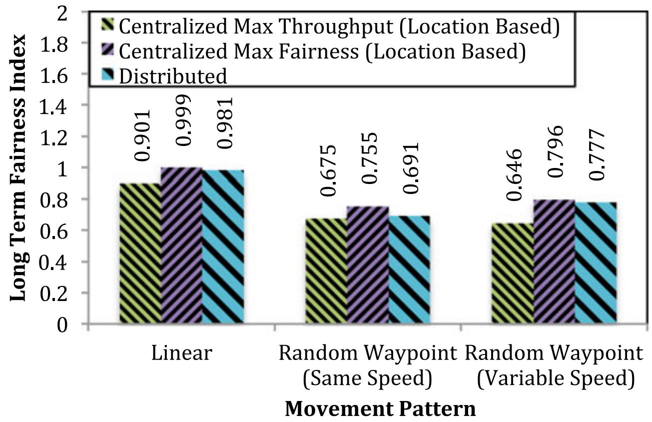

Fig. 6. Long-term fairness comparisons.

significant performance improvement. Also, for the linear movement pattern, since all users are grouped together, resources of only one Wi-Fi AP are used at any given time in addition to the two WiMAX BSs. Whereas for the random waypoint movement pattern, since all users are spread out throughout the network topology, up to six Wi-Fi APs are used at any given time in addition to the two WiMAX BSs. As a result, the overall spectral efficiency obtained for the linear movement pattern (for any association decision solution) is much lower than that of random waypoint movement pattern, which can be seen in Fig. 4.

Also note that the spectral efficiency of random waypoint variable speed (2-20 mph) scenario (2.151 bits/sec/Hz) is higher than that of random waypoint same speed (2 mph) scenario (2.001 bits/sec/Hz). This phenomenon depends on the randomness of the user device distribution in the simulation scenario since one of the requirements of our GRC user device-to-AP/BS association algorithm is to assign each user device to an AP/BS. In this particular simulation scenario, several user devices end up spending a longer amount of time in low data rate region of several APs/BSs when they all move at a slower speed of 2 mph compared to the variable speed case where user devices move out of the low data rate region a bit quicker. As a result, the overall spectral efficiency for random waypoint same speed scenario is lower than that of random waypoint variable speed scenario.

The scan based centralized solution has technical challenges associated with it. For this solution, since the radios on the user device disrupt active data connections to search for available networks on a periodic basis (5 seconds), multiple data packets sent by the sink node are either dropped or significantly delayed. This phenomenon results in a TCP timeout and resetting (halving) of the window size by the AIMD TCP congestion control mechanism every 5 seconds. As a result, there usually aren't enough data packets at the AP/BS to send to each connected user to fully utilize the radio link. Moreover, if all users connected to the AP/BS scan at the same time, no traffic is sent by the corresponding AP/BS for the scan duration resulting in further underutilization of the radio link. So, the performance achieved by this solution completely depends on the scanning process. While the centralized scan solution (both resource allocation algorithms) for linear movement pattern outperforms the distributed solution as seen from Fig. 4, this solution performs worse than the distributed solution for both random waypoint movement patterns. To remedy this challenge, a solution needs to be worked on where active data connections are not disrupted during the scanning process. However, this requires extra dedicated hardware on user devices for scanning purposes or the information related to neighbor APs/BSs needs to be broadcasted to the user devices on a periodic basis by the serving AP/BS in a smart fashion. We omit results related to the scan based centralized solution for the remainder of our work.

5.2 Fairness Results

We next present the short-term and long-term fairness results in Figs. 5 and 6 respectively. Jain's Fairness Index (Equation (3)) is used to compute both metrics. For short-term fairness metric, Jain's Fairness Index is computed using throughput achieved by each user on a one second basis and results are averaged for the entire simulation duration. Since we compute the short-term fairness metric for a one second interval, in addition to the real-time traffic such as VoIP and video, this metric also captures the ability to support applications such as web-browsing as users are not willing to wait for a page to load for too long. For long-term fairness metric, Jain's Fairness Index is computed using throughput achieved by each user for the entire simulation duration. This metric captures the ability to support delay tolerant applications such as file transfer and email. The results for both short-term and long-term fairness follow the same trend. As can be seen from Figs. 5 and 6, the location based centralized solution, which attempts to maximize short-term fairness, outperforms the distributed solution in terms of both short-term and long-term fairness. The gain in short-term fairness metric for the centralized Max Fairness resource allocation procedure compared to the distributed solution is 12.9 (from 0.769 to 0.868), 28.5 (from 0.312 to 0.401) and 8.0 percent (from 0.275 to 0.297) for linear, random waypoint same speed and random waypoint variable speed movement patterns respectively. The gain in long-term fairness metric for the centralized Max Fairness resource allocation procedure compared to the distributed solution is 1.8 (from 0.981 to 0.999), 9.3 (from 0.691 to 0.755) and 2.4 percent (from 0.777 to 0.796) for the linear, random waypoint same speed and random waypoint variable speed movement patterns respectively. Note that this improvement for the centralized Max Fairness resource allocation

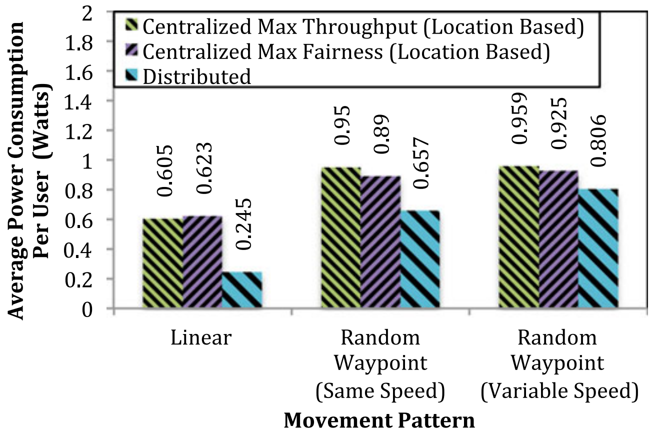

Fig. 7. Energy consumption comparisons.

procedure compared to the distributed solution for both fairness metrics is experienced in addition to the spectral efficiency improvement shown for this procedure in Fig. 4. So the centralized solution with Max Fairness resource allocation procedure improves both conflicting objectives of maximizing system throughput and (short-term and long-term) fairness compared to a distributed solution by making smart association decisions. Both short-term and long-term fairness metrics for the centralized Max Throughput resource allocation procedure suffer compared to the distributed solution for all user movement patterns as seen from Figs. 5 and 6 respectively. But the degradation in fairness metrics for the centralized solution with Max Throughput resource allocation procedure comes as a cost of achieving highest system spectral efficiency as shown in Fig. 4.

5.3 Energy Consumption Results

We finally present the energy consumption results in Fig. 7. The energy consumption computation follows the same approach presented in our earlier work [2], where the energy consumption of a user device depends on two components: (i) the number of bits transmitted/received using Wi-Fi/WiMAX RAT and (ii) the number of handovers performed by the device. For the first component ($P_{t,a}$), the energy consumption numbers for Wi-Fi and WiMAX RATs are presented in Table 11. The second component, which represents the overhead in terms of energy consumption during a handover, has two sub-components: $P_{rec,a}$—energy consumed to turn off one radio and turn on another radio, $P_{assoc,a}$—energy consumed during the association process to a new AP/BS. For a horizontal handover (HH_{ua}), only the $P_{assoc,a}$ energy costs are incurred while for a vertical handover (VH_{ua}), both $P_{rec,a}$ and $P_{assoc,a}$ energy costs are incurred. The equation to compute overall energy consumption for a user device (ω_u) during the entire simulation run is presented in (1), where η_{ua} represents the number of data bits transmitted by user $u \in U$ over AP/BS $a \in A$, $|HH_{ua}|$ represents the number of horizontal handovers experienced by user $u \in U$ within a RAT and $|VH_{ua}|$ represents the number of vertical handovers experienced by user $u \in U$ to AP/BS $a \in A$ during the entire simulation run. We summarize the $P_{t,a}$, $P_{rec,a}$, and $P_{assoc,a}$ energy consumption numbers used in this study in Table 11. The results presented in Fig. 7 represent the average energy consumption cost per user, where ω_u for each user $u \in U$ is computed using (4). The computed ω_u value for each user is summed and the sum is divided by the simulation duration to obtain average energy consumption per user

TABLE 11
Energy Consumption Components for Simulated RATs

	Wi-Fi	WiMAX
$P_{t,a}$ (Joules/x KB)	0.007(x)	0.018(x)
$P_{rec,a}$ (Joules)	0.05	0.28
$P_{assoc,a}$ (Joules)	5.9	3.2

$$\omega_u = \sum_{a\in A}[P_{t,a}(\eta_{ua}) + |\mathrm{HH}_{ua}| * P_{assoc,a} + |\mathrm{VH}_{ua}| * (P_{rec,a} + P_{assoc,a})]. \quad (4)$$

Since the average energy consumption model depends on two components (energy consumption per bit transmitted/received and the number of handovers), the average energy consumption does not give an accurate measure of increase in energy consumption resulting from frequent re-associations. As seen from Fig. 7, the average energy consumption trend generally mimics the spectral efficiency trend shown in Fig. 4. This indicates that the first energy component (energy consumption per bit transmitted/received) dominates the overall energy consumption. Since the centralized solutions achieve higher spectral efficiency (and as result transmit/receive more data bits), the overall energy consumption for centralized solutions is higher than that of the distributed solution. To quantify the energy consumption increase resulting from frequent re-associations for a centralized solution more accurately, the overhead results are presented next.

The number of horizontal (WiMAX-to-WiMAX) and vertical (WiMAX-to-Wi-Fi and Wi-Fi-to-WiMAX) handovers determines the energy consumed by the centralized and distributed solutions resulting from re-associations. The actual number of each type of handover for each simulation scenario is presented in Fig. 8. As can be seen from the figure, the horizontal handovers dominate the total number of handovers for the centralized solutions in each movement pattern. There are approximately 50×, 60× and 30× more horizontal handovers for the centralized solution with Max

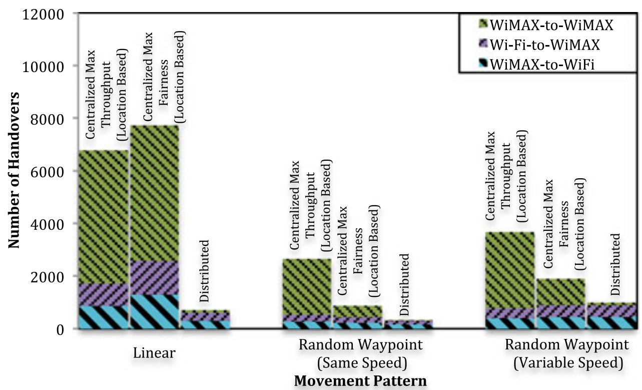

Fig. 8. Handover comparisons.

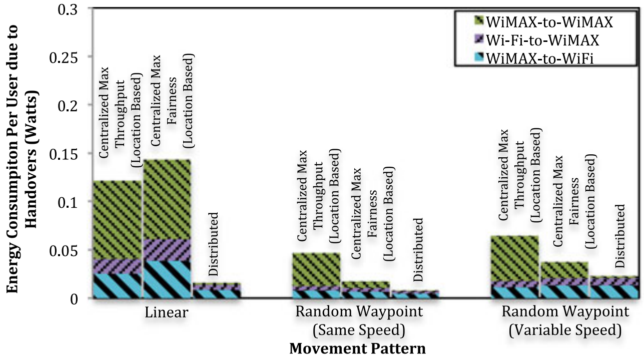

Fig. 9. Energy consumption comparisons due to handovers.

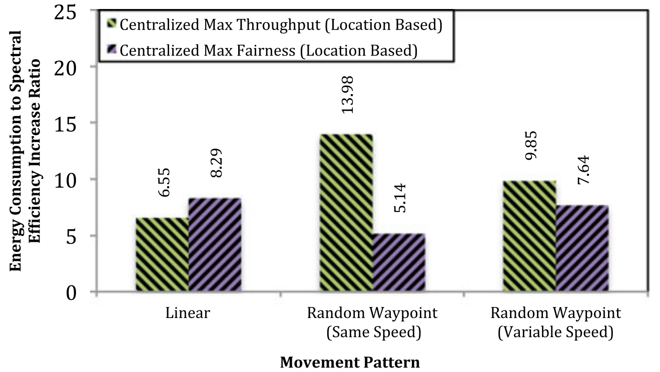

Fig. 10. Ratio of relative increase in energy consumption compared to spectral efficiency.

Throughput resource allocation procedure compared to the distributed solution for linear, random waypoint same speed and random waypoint variable speed movement patterns respectively. The increase in horizontal handovers for the centralized solution with Max Fairness resource allocation procedure are approximately 50×, 10× and 10× for linear, random waypoint same speed and random waypoint variable speed movement patterns respectively. As a result, the increase in energy consumption resulting from re-associations for the centralized solutions will be an order of magnitude greater compared to the distributed solution.

The increase in vertical handovers for the centralized solution (both Max Throughput and Max Fairness resource allocation procedure) compared to the distributed solution is all under 4×. Hence, vertical handovers do not contribute to the increase in energy consumption as much as horizontal handovers. Note, however, that vertical handovers consume more energy than horizontal handover (3.48 Joules for a Wi-Fi-to-WiMAX handover and 5.95 Joules for a WiMAX-to-Wi-Fi handover compared to 3.2 Joules for a WiMAX-to-WiMAX handover compared to). As a result, vertical handovers are equally important to consider in any future energy optimization investigation.

The average energy consumption results that consider both horizontal and vertical handovers for each solution are presented in Fig. 9. As seen from Fig. 9, the increase in average energy consumption per user resulting from frequent re-associations (handovers) for the centralized solution with Max Throughput resource allocation procedure is 650 (0.12 Watts compared to 0.016 Watts), 488 (0.047 Watts compared to 0.008 Watts) and 191 percent (0.064 Watts compared to 0.022 Watts) compared to the distributed solution for linear, random waypoint same speed and random waypoint variable speed movement patterns respectively. The increase in the same metric for the centralized solution with Max Fairness resource allocation procedure compared to the distributed solution is 794 (0.143 Watts compared to 0.016 Watts), 113 (0.017 Watts compared to 0.008 Watts) and 68 percent (0.037 Watts compared to 0.022 Watts) for linear, random waypoint same speed and random waypoint variable speed movement patterns respectively. To get an estimate of the increase in energy consumption relative to the increase in spectral efficiency (which is almost linear in the range of [8.9-99.2 percent]) shown in Fig. 4, the ratio of increase in energy consumption to spectral efficiency increase for a centralized solution (with both resource allocation procedures) compared to the distributed solution is shown in Fig. 10. As can be seen from this figure, the increase in energy consumption is a factor of 5.14 to 13.98 times greater than the increase in spectral efficiency, which indicates an almost exponential increase in energy consumption compared to the linear increase in spectral efficiency.

In addition to the system efficiency performance measures just presented, we analyze the messaging overhead required during network re-associations and compare it with achieved system throughput. We consider the technology-independent messages presented in Table 4 in our overhead modeling. These messages include periodic link parameter report, association mapping message sent by the GRC based on re-association computations, router solicitation message sent by the user device to obtain an IP address once link-layer connectivity has been established, router advertisement message sent by the AP/BS in response to the router solicitation message, and the flow redirect request sent by the user device to inform the other endpoint of the switch in interfaces. The comparison of average throughput consumed by the overhead messages vs. the actual average system (data) throughput is presented in Fig. 11. The trend in the amount of overhead created by each solution mimics the number of handovers experienced by each solution (shown in Fig. 8), which follows expectations. The highest amount of overhead throughput produced by any solution compared to the overall throughput is 18.3 percent (24.13 Mbps of data throughput and 5.41 Mbps of overhead throughput) for the Max Fairness

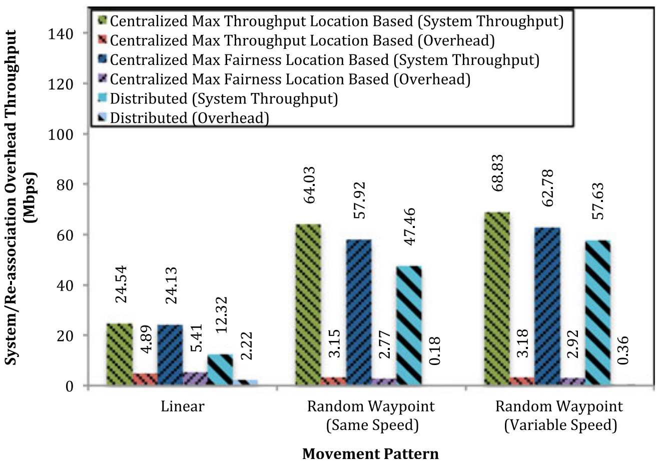

Fig. 11. Comparison of re-association overhead versus system throughput.

centralized solution for the linear movement pattern. While this is a significant amount of overhead, this happens only in extreme cases where all users are grouped together in one location where each user can use a limited set of RATs. For this movement pattern, even the distributed solution has an overhead throughput of 15.3 percent (12.32 Mbps of data throughput and 2.22 Mbps of overhead throughput). For the users that are spread throughout the topology (random waypoint mobility pattern), the highest overhead throughput is 4.7% (64.03 Mbps of data throughput and 3.15 Mbps of overhead throughput) for the Max Throughput centralized solution. For all centralized solutions for the random waypoint movement patterns (same speed and variable speed), the overhead throughput to total throughput ratio is in the range [4.4, 4.7 percent]. For the distributed solution for the random waypoint movement patterns (same speed and variable speed), the overhead throughput to total throughput ratio is in the range [0.3, 0.6 percent]. So as seen from these results, the overhead related to the centralized solution is very manageable and the increase in overhead due to network re-associations for a centralized solution compared to a distributed solution does not exceed by about 4.1 percent.

The analysis presented in this paper uses WiMAX as the 4G cellular technology. In reality, cellular operators have adopted LTE as the 4G cellular technology of choice. The reason WiMAX was chosen in our work was due to the maturity of the WiMAX simulation code compared to LTE code. Nonetheless, the results presented in this paper would be applicable to the case where LTE was used as the 4G cellular technology since both WiMAX and LTE are similar. They both use OFDMA access technology for downlink resource allocation. To help alleviate peak-to-average power ratio problem created by OFDMA, LTE uses OFDMA/SC-FDMA hybrid access technology while WiMAX still uses OFDMA access technology for uplink resource allocation. However, the mechanism by which resources are allocated even in the uplink for both LTE and WiMAX essentially remain the same. For equal amount of spectrum used by both WiMAX and LTE, the amount of resource blocks available to both WiMAX and LTE would be approximately the same and as a result, both standards would achieve similar performance. Further, because the spectral efficiency is defined as the total achieved throughput divided by the amount of used spectrum, the resulting spectral efficiency presented in this work should remain about the same. Moreover, fairness metric computed using Jain's Fairness metric that just measures the differences in throughput achieved by each individual user device should also remain the same.

Unlike WiMAX, LTE is evolving quickly. Advances available in LTE-A, including carrier aggregation and support for small cells would require our models to be updated. However, the hetnet approach, and in particular, the heuristic resource allocation algorithms that we have considered are applicable.

6. Conclusions

We studied the resource allocation and management issues related to hetnet wireless systems made up of several RATs that collectively provide a unified wireless network to a diverse set of users through co-ordination managed by a centralized GRC. Through detailed protocol level simulations performed in ns-2, we showed that that an allocation strategy that prioritizes system throughput over fairness will increase spectral efficiency in the range 19-99 percent while reducing long term fairness by 8-17 percent. For our centralized solution that uses Max Fairness resource allocation procedure, using the linear user movement pattern, we showed a spectral efficiency increase of 95.8, short-term fairness increase of 12.9 percent and long-term fairness increase of 1.8 percent compared to the distributed solution.

A centralized solution where a user device scans periodically disrupting active data connections (TCP) results in unpredictable performance results because of the TCP congestion control mechanism and the underutilization of available RATs during the scanning procedure. A location based solution such as the one we used in our study or other mechanism such as additional scanning hardware needs to be implemented at each user device to support the generation of periodic link parameter report required by a centralized hetnet solution without disrupting active data connections.

The overhead required by the centralized solution based on IEEE 802.21 framework does not exceed more than 4.1 percent compared to a distributed solution for the various user movement patterns analyzed in our work and the overhead throughput compared to overall throughput does not exceed 18.3 percent (which only happens in rare cases where all users are grouped together in the linear movement pattern. The overhead throughput accounts for fewer than 4.7 percent of overall throughput for all random movement patterns).

The centralized solution experiences a significant number of handovers compared to the distributed case, and as a result there is a significant increase in energy consumption (up to 794 percent) resulting from network re-associations for the centralized solution compared to the distributed solution. The resource allocation procedure implemented at the centralized solution needs to limit the number of handovers by using incremental policies in addition to the traditional objectives of optimizing network efficiency measures of spectral efficiency and fairness. This represents an area of future work.

Acknowledgments

Rahul Amin is the corresponding author.

References

[1] T. Bu, L. Li, and R. Ramjee, "Generalized proportional fair scheduling in third generation wireless data networks," in *Proc. IEEE INFOCOM*, 2006, pp. 1–12.

[2] R. Amin, J. Martin, J. Deaton, L. DaSilva, A. Hussien, and A. Eltawil, "Balancing spectral efficiency, energy consumption, and fairness in future heterogeneous wireless systems with reconfigurable devices," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 5, pp. 969–980, May 2013.

[3] Y. Bejerano, S. Han, and L. E. Li, "Fairness and load balancing in wireless LANs using association control," in *Proc. 10th Annu. Int. Conf. Mobile Comput. Netw.*, 2004, pp. 315–329.

[4] W. Wang, X. Liu, J. Vicente, and P. Mohapatra, "Integration gain of heterogeneous WiFi/WiMAX networks," *IEEE Trans. Mobile Comput.*, vol. 10, no. 8, pp. 1131–1143, Dec. 2010.

[5] S. Lee, K. Sriram, K. Kim, Y. Kim, and N. Golmie, "Vertical handoff decision algorithms for providing optimized performance in heterogeneous wireless networks," *IEEE Trans. Veh. Technol.*, vol. 58, no. 2, pp. 865–881, Feb. 2009.
[6] P. Kosmides, A. Rouskas, and M. Anagnostou, "Network selection in heterogeneous wireless environments," in *Proc. Int. Conf. Telecommun.*, 2011, pp. 250–255.
[7] A. Calvagna and G. Di Modica, "A user-centric analysis of vertical handovers," in *Proc. 2nd ACM Int. Workshop Wireless Mobile Appl. Services WLAN Hotspots*, 2004, pp. 137–146.
[8] E. Stevens-Navarro and V. Wong, "Comparison between vertical handoff decision algorithms for heterogeneous wireless networks," in *Proc. IEEE Veh. Technol. Conf.*, 2006, pp. 947–951.
[9] G. Fodor, A. Eriksson, and A. Tuoriniemi, "Providing quality of service in always best connected networks," *IEEE Commun. Mag.*, vol. 22, no. 41, pp. 154–163, 2003.
[10] A. Dutta, D. Famolari, S. Das, Y. Ohba, V. Fajardo, K. Taniuchi, R. Lopez, and H. Schulzrinne, "Media-independent pre-authentication supporting secure interdomain handover optimization," *IEEE Wireless Commun.*, vol. 15, no. 2, pp. 55–64, Apr. 2008.
[11] W. Shi, B. Li, N. Li, and C. Xia, "A network architecture for load balancing of heterogeneous wireless network," *J. Netw.*, vol. 6, no. 4, pp. 623–630, Apr. 2011.
[12] A. De La Oliva, A. Banchs, I. Soto, T. Melia, and A. Vidal, "An overview of IEEE 802.21: Media-independent handover services," *IEEE Wireless Commun.*, vol. 15, no. 4, pp. 96–103, Aug. 2008.
[13] J. Martin, R. Amin, A. Eltawil, and A. Hussien, "Using reconfigurable devices to maximize spectral efficiency in future heterogeneous wireless systems," in *Proc. IEEE 20th Int. Conf. Comput. Commun. Netw.*, Aug. 2011, pp. 1–8.
[14] R. Amin, J. Martin, A. Eltawil, and A. Hussien, "Spectral efficiency and energy consumption tradeoffs for reconfigurable devices in heterogeneous wireless systems," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Apr. 2012, pp. 1698–1703.
[15] M. Heusse, F. Rousseau, G. Berger-Sabbatel, and A. Duda, "Performance anomaly of 802.11b," in *Proc. IEEE INFOCOM*, 2003, pp. 836–843.
[16] E. Garcia, D. Viamonte, R. Vidal, and J. Paradells, "Achievable bandwidth estimation for stations in multi-rate IEEE 802.11 WLAN Cells," in *Proc. IEEE Int. Symp. World Wireless, Mobile Multimedia Netw.*, Jun. 2007, pp. 1–8.
[17] L. Jiang and S. Liew, "Proportional fairness in wireless LANs and ad hoc networks," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2005, pp. 1551–1556.
[18] A. Banchs, P. Serrano, and H. Oliver, "Proportional fair throughput allocation in multirate IEEE 802.11e wireless LANs," *Wireless Netw.*, vol. 13, no. 5, pp. 649–662, Oct. 2007.
[19] T. Joshi, A. Mukherjee, Y. Yoo, and D. P. Agrawal, "Airtime fairness for IEEE 802.11 multirate networks," *IEEE Trans. Mobile Comput.*, vol. 7, no. 4, pp. 513–527, Apr. 2008.
[20] J.-Y. Le Boudec, "Rate adaptation, congestion control and fairness: A tutorial," Ecole Polytechnique Federale de Lausanne (EPFL), Chapter 1, Nov. 22, 2005.
[21] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control for communication networks: Shadow prices, proportional fairness and stability" *J. Oper. Res. Soc.*, vol. 49, pp. 237–252, 1998.
[22] T. Girici, C. X. Zhu, J. R. Agre, and A. Ephremides, "Proportional fair scheduling algorithm in OFDMA-based wireless systems with QoS constraints," *J. Commun. Netw.*, vol. 12, no. 1, pp. 30–42, Feb. 2010.
[23] A. Rashwan, H. Elbadawy, and H. Ali, "Comparative assessments for different WiMAX scheduling algorithms," in *Proc. World Congr. Eng. Comput. Sci.*, San Francisco, CA, USA, Oct. 2009.
[24] R. Jain, D. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," DEC, Maynard, MA USA, DEC Res. Rep. TR-301, Sep. 1984.
[25] *IEEE Standard for Local and Metropolitan Area Networks, Part 21: Media Independent Handover Services*, IEEE Std 802.21-2008, Jan. 2009.
[26] T. Narten, E. Nordmark, W. Simpson, and H. Soliman, Neighbor discovery for IP version 6 (IPv6). RFC 4861, IETF, Sep. 2007.
[27] NIST Mobility Package, Seamless and security project: Software tools [Online]. Available: http://www.nist.gov/ctl/wireless-networks/upload/MIH_module.pdf, Apr. 2015.

Rahul Amin received the PhD degree from Clemson University in August 2013. As a PhD student, he was a research assistant in the Networking Lab for the School of Computing at Clemson University and focused on radio resource management strategies for next generation heterogeneous wireless networks. As a master's student, he was a research assistant in the Department of Electrical and Computer Engineering, Clemson University, and collaborated on a vehicular broadband connectivity project with BMW IT research center in Greenville, SC. He is a member of Technical Staff in Airborne Networks Group at MIT Lincoln Laboratories. His research interests include resource allocation, network optimization and network integration problems in both ad hoc and infrastructure-based wireless networks.

Jim Martin received the PhD degree from North Carolina State University. He is an associate professor in the School of Computing, Clemson University. His research interests include broadband access, wireless networks, Internet protocols, and network performance analysis. His current research projects include heterogeneous wireless systems and DOCSIS 3.x cable access networks. He received funding from the US National Science Foundation (NSF), NASA, the Department of Justice, Cisco, IBM, CableLabs, and BMW. Prior to joining Clemson, he was a consultant for Gartner, and prior to that, a software engineer for IBM.