Scheduling in Multihop WiMAX Networks

Debalina Ghosh
debghosh@ucdavis.edu
Ashima Gupta
ashgupta@ucdavis.edu
Prasant Mohapatra
pmohapatra@ucdavis.edu
Department of Computer Science, University of California at Davis, Davis, CA 95616

IEEE 802.16, popularly known as WiMAX, is at the forefront of the technology drive because of the growing demand for high-speed wireless broadband networks. Multihop WiMAX networks are particularly useful as they increase the coverage area without the need to deploy expensive base stations. There are two kinds of multihop WiMAX networks - WiMAX mesh and Mobile Multihop Relay Networks. Scheduling is very important in both of these multihop WiMAX networks. The links have to be scheduled in such a way so that they do not interfere with each other while maximizing the throughput of the networks. As WiMAX networks are geared towards broadband applications, any scheduling scheme should accomodate the rate, latency and jitter requirements of the applications. This article provides an insight to the scheduling framework presented in the IEEE 802.16 standard. It also presents a few representative research proposals for centralized scheduling in WiMAX networks. We discuss some of the research issues and challenges that need to be addressed for multihop WiMAX networks to realize their full potential.

I. Introduction

WiMAX (World Interoperability for Microwave Access) technology is based on IEEE 802.16 standard that provides wireless access to metropolitan area networks (MANs). IEEE 802.16 provides an inexpensive and easily deployable alternative to wired technologies like DSL, T1 or cable that are used for backhauling Personal Area Networks (PANs) and Local Area Networks(LANs). This is particularly useful in regions with little existing wired infrastructure. Adding multihop technology increases the coverage area of WiMAX networks without the need to deploy expensive base stations. A multihop WiMAX network may also lead to increased user throughput as more efficient modulation techniques can be used over shorter links [1].

The first IEEE 802.16 standard, published in April 2002, defined the Medium Access Control (MAC) and Physical (PHY) layers, operating in licensed spectrum between 10 and 66 GHz. It requires Line of Sight (LOS) connectivity and supports data rate up to 134 Mb/s. A later amendment, IEEE 802.16a, published in April 2003, defines additional PHYs for the 2-11 GHz licensed and unlicensed spectrum and provides enhancements to the MAC to support a mesh topology. IEEE 802.16-2004 (also known as IEEE 802.16-d) incorporates IEEE 802.16a into the original standard. IEEE 802.16-2005 adds mobility extensions to the fixed IEEE 802.16 standard. This is popularly known as Mobile WiMAX. A new working group, IEEE 802.16j, is currently adding multihop capabilities to Mobile WiMAX by introducing new entities called Relay Stations.

IEEE 802.16 supports a number of parameters related to the MAC and physical layers. In order to ensure that IEEE 802.16-based products from different vendors are interoperable, an industry consortium called the WiMAX forum provides guidelines known as profiles, which specify the frequency band of operation, the PHY to be used and a number of other parameters. We use the terms IEEE 802.16 and WiMAX interchangeably in this article.

The rest of the article is organized as follows. Section II gives a brief overview of IEEE 802.16 Point-to-Multipoint (PMP) mode. We briefly describe WiMAX mesh in Section III and Mobile Multihop Relay Networks in Section IV. Sections V and VI describe centralized and distributed scheduling respectively. Section VII discusses some of the research proposals that address scheduling in multihop WiMAX Networks. We describe some research issues and challenges in Section VIII. Finally we conclude in Section IX.

II. Overview of IEEE 802.16

The basic architecture of IEEE 802.16 is similar to that of cellular networks. In a particular region, there is a base station (BS) and multiple subscriber stations (SSs). IEEE 802.16 Point-to-MultiPoint (PMP) mode is a star-shaped network where every SS communicates directly with the BS. IEEE 802.16 standard de-

Table 1: List of Acronyms

WiMAX	World Interoperability for Microwave Access
MAN	Metropolitan Area Network
WMN	Wireless Mesh Network
TDMA	Time Division Multiple Access
CSMA/CA	Carrier Sense Multiple Access with Collision Avoidance
OFDMA	Orthogonal Frequency Division Multiple Access
PMP	Point-to-MultiPoint
BS	Base Station
SS	Subscriber Station
LOS	Line of Sight
FDD	Frequency Division Duplex
TDD	Time Division Duplex
PHY	Physical Layer
MAC	Medium Access Control
PDU	Protocol Data Unit
SDU	Service Data Unit
CRC	Cyclic Redundancy Check
CID	Connection Identifier
BRH	Bandwidth Request Header
GMH	Generic MAC Header
MMR	Mobile Multihop relay
MS	Mobile Station
RS	Relay Station

fines the physical layer (PHY) and the data link layer which are the bottommost layers of the protocol stack ([2], [3]).

The physical layer of IEEE 802.16 consists of a number of air interfaces such as WirelessMAN-OFDM and WirelessMAN-OFDMA. Both frequency-division duplex (FDD) and Time-Division Duplex (TDD) are supported for communication between BS and SS. A frame consists of a downlink subframe and an uplink subframe. IEEE 802.16 also supports adaptive burst profile, that enables the transmission parameters to be modified on a frame-by-frame basis for each SS.

The MAC is connection-oriented, which means that all services are mapped to a connection identified by a 16-bit connection identifier (CID). Scheduling of data transfer is done by the Base Station in the PMP mode. The downlink subframe sent by the BS contains the DL-MAP and UL-MAP. The DL-MAP specifies the downlink channel access and the associated burst profile. The UL-MAP defines the uplink channel access, that is, the time slot in which the SS can transmit in the uplink subframe and the uplink data burst profiles.

IEEE 802.16 standard defines five different service classes of traffic as illustrated in Table 2.

Table 2: Service Classes in WiMAX

Class	**Application**	**QoS parameters**
Unsolicited Grant Service (UGS)	VoIP,E1; fixed-size packets on periodic basis	max rate, latency and jitter
Real-Time Polling Service (rtPS)	Streaming audio/video	minrate, maxrate and latency
Enhanced Real-Time Polling Service (ertPS)	VoIP with activity detection	minrate, maxrate, latency and jitter
Non Real-Time Polling Service (nrtPS)	FTP	minrate and maxrate
Best Effort (BE)	Data transfer, Web	maxrate

II.A. OFDM and OFDMA

IEEE 802.16d (fixed WiMAX) uses Orthogonal Frequency Division Multiplexing (OFDM) whereas IEEE 802.16e (mobile WiMAX) uses Orthogonal Frequency Division Multiplexing Access (OFDMA). Instead of a single carrier, OFDM uses multi-carrier modulation that increases the data throughput and eliminates problems with multi-path signal and spectral interference. OFDM allows only one user on the channel at a time. Time Division Multiple Access (TDMA) is used to accomodate multiple users where interfering users are assigned different timeslots.

OFDMA is a multi-user OFDM that allows multiple users to access the channel at the same time. Interfering users are assigned different subchannels on the same timeslot. Thus OFDMA can be viewed as a combination of FDMA and TDMA.

In our discussion, a slot refers to a timeslot in OFDM-based fixed WiMAX (WiMAX mesh) networks and a slot refers to a slot in the time-frequency

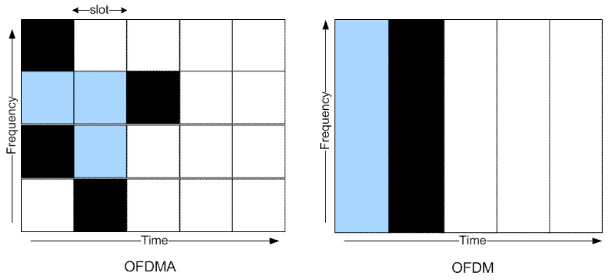

Figure 1: Comparison of OFDM and OFDMA

grid in OFDMA-based mobile WiMAX (MMR) networks as depicted in Figure 1.

III. WiMAX mesh

A mesh network consists of a mesh BS and multiple SSs. In contrast to the PMP mode, a SS can be multiple hops away from the BS in the mesh mode and may communicate with the BS with the help of intermediate nodes. Figure 2 shows a mesh network.

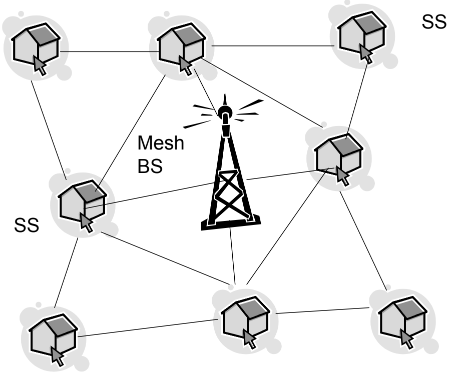

Figure 2: IEEE 802.16 mesh network

IEEE 802.16a defines a new mesh frame format. A Mesh frame is addressable by a 12-bit frame number and is divided into a number of minislots. Figure 3 shows the mesh frame where each frame consists of a control subframe and data subframe. The control subframe may be a network control subframe or a scheduling control subframe. The network control subframe is used to send network control messages like MSH-NENT and MSH-NCFG that enable new nodes to join a mesh. The scheduling control subframe is used for sending centralized scheduling messages like MSH-CSCF and MSH-CSCH. Scheduling control subframe may also be divided into transmission opportunities where distributed scheduling messages (MSH-DSCH) are sent. The network control subframe is repeated periodically after a few scheduling control subframes. The data subframe is used for transmitting data packets and some distributed scheduling messages. There are two mechanisms by which data transmissions can be scheduled in the mesh mode - centralized and distributed. Scheduling is done so that there are no collisions during the transfer of data in the data subframe. In centralized scheduling, the BS determines how the SSs should share the channel in different time slots. The scheduling procedure is simple, however the connection setup delay is significant. Hence centralized scheduling is not suitable for occasional traffic needs. In distributed scheduling, the nodes themselves determine the schedule of data transmissions without the help of the BS. Therefore, distributed scheduling is more flexible and efficient for connection setup and data transmission. Distributed scheduling may be coordinated or uncoordinated. In coordinated distributed scheduling, every node competes for channel access using a pseudo-random election algorithm based on scheduling information of two-hop neighbors. MSH-DSCH messages are sent during the scheduling control subframe to set up the schedule. Uncoordinated distributed scheduling uses “idle” slots in the data subframe to determine the schedule of data transmissions. In both cases, data subframe is allocated based on request-grant-confirm three-way handshaking among the nodes.

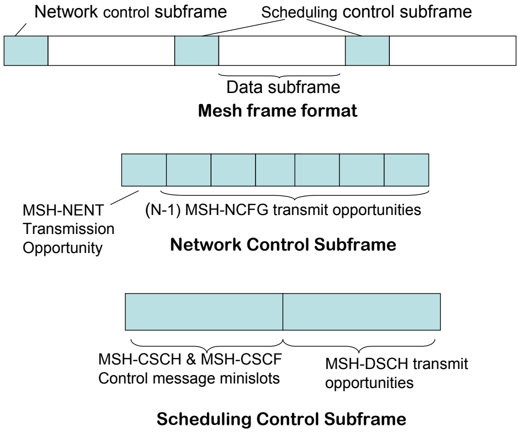

Figure 3: IEEE 802.16 mesh frame format

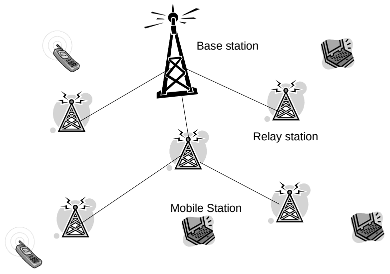

Figure 4: IEEE 802.16 mobile multihop relay mode

IV. Mobile Multihop Relay

The goal of IEEE 802.16j is to add multihop capabilities to IEEE 802.16e so that the throughput and coverage area of mobile WiMAX networks is increased while ensuring compatibility with the PMP mode. The network topology of a Mobile Multihop Relay (MMR) network is a tree with the BS at the root of the tree. New network entities called Relay Stations (RSs) are introduced. RSs relay information between a subscriber station(SS)/mobile station(MS) and a BS or between other RSs or between an RS and a BS. A RS does not provide backhaul functionality and hence it is much simpler than a BS. We refer to a BS in a MMR network as MR-BS from now on. Figure 4 shows a typical relay network.

The differences between mesh and relay mode is illustrated in Table 3.

Table 3: Comparision of Mesh and Relay Mode

Mesh	Relay
Not compatible with PMP mode	Compatible with PMP mode
Fixed broadband access	Mobile broadband access
Network architecture is a mesh	Network architecture is a tree
Network entities are Mesh BS and SS	Network entities are BS, SS and RS

IEEE 802.16j only focuses on the OFDMA PHY mode of IEEE 802.16e-2005. There are two types of relay stations:

1. Transparent Relays - These relays serve mobile stations that receive control information from the base station. Thus the mobile station can receive signals from both the base station and the relay and hence can achieve higher throughput. Transparent relays, therefore lead to increased network capacity. All transparent relays must operate in centralized scheduling mode, relying on the MR-BS to allocate its resources.

2. Non-transparent relays - These relays serve mobile stations that cannot decode control information from the MR-BS. These relay stations act as the base stations for the mobile stations and must transmit control information at the beginning of the frame. Non-transparent relays increase the coverage area and may operate in both centralized and distributed scheduling mode.

V. Centralized Scheduling

In addition to communication between SS and BS, WiMAX mesh networks allow communication between SSs. Similarly, MMR networks allow communication between RSs, between a MS and a RS and between a RS and a BS. Hence scheduling in multihop WiMAX networks has to accomodate all these different communications.

In centralized scheduling scheme, the BS determines the schedule and the scheduling packets are transmitted in a collision-free way within scheduling control subframes. Two control messages are used in centralized scheduling: MSH-CSCF and MSH-CSCH. MSH-CSCF message delivers the information of channel configuration and routing tree, while MSH-CSCH message delivers bandwidth request and grant information.

During a downstream MSH-CSCF or MSH-CSCH frame, a BS transmits first, followed by all its children with hop-count 1. This is followed by the nodes with hop-count 2 and so on until all the nodes in the routing tree have transmitted. The nodes with a given hop count transmit depending on the their order in the most recent MSH-CSCF or MSH-CSCH frame. The upstream messages follow the reverse order.

The MSH-CSCH:Request message is used by SSs to send bandwidth demands to the BS. When a node sends a MSH-CSCH:Request message to its parent, it includes the estimate of its own upstream and downstream traffic demand, along with the demands reported by its children. After the BS receives this message, it estimates the bandwidth to grant to each node and issues a MSH-CSCH:Grant message. This MSH-CSCH:Grant message then propagates down the tree.

The SSs use the bandwidth assignment to determine the starting times and link durations in the frame. There is no spatial reuse; that is, the links transmit one after another.

VI. Distributed Scheduling

There are two types of distributed scheduling: coordinated and uncoordinated. In coordinated distributed scheduling, the scheduling packets(MSH-DSCH) are transmitted in a collision-free manner within the control subframe. In uncoordinated distributed scheduling, the scheduling is performed in a partially, contention-based manner while avoiding any conflicts with the schedules established using the coordinated methods. Uncoordinated distributed scheduling is best suited for links with occasional or brief traffic needs [1]. Distributed scheduling involves a three-way handshake - request, grant and confirm messages during which the slots in which data transfer takes place are selected. The different scheduling techniques are compared in Table 4.

Table 4: Comparison of Scheduling Techniques

Centralized	**Coordinated Distributed**	**Uncoordinated Distributed**
BS determines the schedule for all the nodes in the network	Data transfer between neighboring SSs/RSs is done without BS involvement	Data transfer between neighboring SSs/RSs without BS involvement
Connection setup overhead is high	Connection setup overhead is lower than centralized scheduling but higher than Uncoordinated distributed scheduling	Conenction setup overhead is lowest
Data transfer is completely contention free	Data transfer is contention free	Data is transferred based on an estimation of idle slots and may involve collisions
MSH-CSCF and MSH-CSCH messages are used to deliver routing information and bandwidth request and grant information	Nodes compete for sending MSH-DSCH messages that are used to deliver the request-grant and confirm information are sent in the scheduling control subframe using a distributed election algorithm	Nodes compete for sending MSH-DSCH messages in the idle slots of the data subframe using a random access algorithm
Suitable for consistent continuous traffic	Suitable for intermittent and bursty traffic	Suitable for occasional and brief traffic

VI.A. Coordinated Distributed Scheduling

Coordinated distributed scheduling sets up a schedule for data transfer between neighbouring nodes without the help of the BS. Nodes compete for transmitting scheduling packets (MSH-DSCH) in the scheduling control subframe so that there is no contention in the data time slots. The scheduling control subframe is divided into transmission opportunities and nodes contend for transmitting MSH-DSCH messages in these transmission opportunities using a distributed election algorithm. If a node wins the election algorithm, it sets the temporary transmission opportunity as its transmission time and broadcasts it to the neighbors in the MSH-DSCH packet. There are two important parameters which are used in this algorithm [1]. These are *NextXmtMx* and *XmtHoldOffExponent*. The *NextXmtTime*, that is, the time when a node can transmit again is given by:

$$2^{XmtHoldOffExponent}.NextXmtMx$$

$$\le NextXmtTime$$

$$\le 2^{XmtHoldOffExponent}.(NextXmtMx+1)$$

So the *eligibility interval* of a node spans a duration of $2^{XmtHoldOffExponent}$ transmission opportunities. The station can transmit in any slot during this interval. After one eligibility interval, a station must hold off at least $2^{XmtHoldOffExponent+4}$ transmission opportunities before the next transmission. The holdoff exponent value decides the channel contention time of node and so it is an important parameter that can affect the system performance.

A node calculates its *NextXmtTime* during the current transmission time using the *distributed election algorithm*. The node sets the first transmission slot after the holdoff time as the next transmission opportunity (*CandidateXmtOppurtunity*) and then competes for this slot with other competing nodes in the two-hop neighborhood.

For a given *CandidateXmtOpportunity*, the eligible competing nodes of a node within the local node's extended neighborhood are those nodes whose:

- *NextXmtTime* interval includes the *CandidateXmtOpportunity*.
- *EarliestSubsequentXmtTime* (equal to *NextXmtTime + XmtHoldoffTime*) is $\leq$ the *CandidateXmtOpportunity*
- *NextXmtTime* is not known.

For example, in Figure 5, node A competes for the next tranmission opportunity with nodes B, C and D. Node B's *NextXmtTime* includes node A's *CandidateXmtOppourtunity*, node C's *EarliestSubsequentXmtTime* is $\leq$ than node A's *CandidateXmtOpportunity* and node D's *NextXmtTime* is not known. The algorithm takes the slot number and the IDs of all the competing nodes as inputs and uses a pseudo-random function to generate mixing values. If the current node ID and the slot number generate the largest mixing value, it wins; otherwise it loses and it chooses the next transmission opportunity as *CandidateXmtOpportunity* and repeats the same procedure.

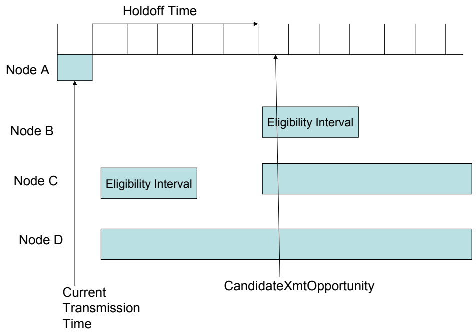

Figure 5: Nodes competing for a CandidateXmtOpportunity

VI.B. Uncoordinated Distributed Scheduling

Uncoordinated distributed scheduling is used for fast setup of new, temporary data "bursts" between neighboring nodes [1]. The scheduling messages are sent during data subframe and the connection setup uses a three-way request-grant-confirm handshake.

The node sending the MSH-DSCH:Request message first observes the "idle" slots of the current schedule and uses a random-access algorithm to determine when it should send the request message. If there is a collision or after a nearby "burst" is completed, random backoff is used before scheduling a request message. The request message contains a list of the neighbors with whom it wants to exchange data. The message also lists the "idle" slots in its neighborhood. On receiving a request, a requestee sends a MSH-DSCH:Grant message in one of the "idle" slots. The "idle" slot in which the requestee sends the grant message depends on its order of appearance in the Request message. The Requestee must also make sure that transmission of the grant message does not cause any collision in its own neighborhood. The requestee determines jointly available slots from the "idle" slots in the request message and its own schedule. Upon receiving this grant message, the requester confirms the schedule by sending another MSH-DSCH:Grant message. Data transfer can then take place in the selected slots.

VII. Centralized Scheduling Techniques

Many researchers have proposed centralized scheduling techniques for WiMAX mesh networks. We present some centralized scheduling techniques in this section. However, to the best of our knowledge, no distributed scheduling techniques have been proposed so far and hence we do not discuss distributed scheduling schemes in this article.

Scheduling techniques can be broadly divided into two categories - with no spatial reuse and those enabling spatial reuse. Schemes that allow spatial reuse try to allocate the same slots to non interfering links whereas those schemes that do not allow spatial reuse only allocate one link in one slot. Note that slot may be a timeslot in TDMA based OFDM networks or slot may be a slot in the time-frequency domain in OFDMA networks. Other differentiating factors between various scheduling schemes include whether or not they provide any QoS guarantees, whether the schemes consider fairness etc. Some centralized scheduling algorithms also suggest routing schemes since a good route might lead to efficient scheduling.

Shetiya and Sharma design routing and centralized scheduling algorithms that provide per flow QoS guarantees to real and interactive data applications [4] for IEEE 802.16 mesh networks. They use fixed routing based on a tree structure and adopt shortest path routing algorithms that work well for both type of applications. They propose separate scheduling algorithms for UDP(Constant bit rate and Variable bit rate) and/or TCP connections. The resource allocation in terms of slots is done assuming an OFDM physical layer and does not support spatial reuse. They compute the number of slots required per flow along the path and hence at each node per frame. The number of slots is computed depending on derived flow characteristics such as end to end packet drop probability. However, the slot ordering per node or per flow is not discussed in the paper. Traffic is divided into different queues at every node depending on the bandwidth requirements. Once the Mesh BS assigns the computed number of slots to the nodes, the nodes provide the required slots to its different queues in a weighted round robin manner. For TCP traffic, the slot allocation for nodes is done in a manner that is proportionally fair to the minimum bandwidth requirements of the nodes. The authors consider link rates as well as the traffic requirements while performing slot allocation. This could lead to starvation of nodes with bad link connections. To avoid this problem they employ an adaptive fixed allocation scheme where allocation is done depending on link quality and the time that a node has not been allocated slots as well as traffic requirements. The proposed algorithms are efficient and can be implemented in real time. However, there is no spatial reuse.

Bandwidth requirements of flows are satisfied in [5] where the number of slots required is computed based on the minimum rate, maximum rate and the bandwidth request of the flows. The authors present an algorithm on how to allocate free slots. They also discuss how the order in which the slots are assigned can be changed to reduce jitter. The algorithm is for an OFDM physical layer for a single hop network. The algorithm accounts for MAC overhead in both the cases when the BS allows packing and when it does not.

In [6] the authors introduce and compare scheduling algorithms for 802.16d OFDMA/TDD based systems that provide fair and efficient allocation to all users. The algorithms are for PMP (point to multipoint) systems. The algorithms schedule flows based on service class priority. SSs with larger amount of data to transmit are given priority in the heuristic algorithm. The authors use proportional fairness such that bandwidth is provided to a user in proportion to its throughput requirement. They present algorithms that determine the proportional constant value. Fairness is not applied when scheduling UGS flows as the bandwidth demand for UGS flows is low.

Cohen and Katzir [7] divide the OFDMA scheduling problem into a macro and a micro scheduling problem. The authors assume that the association between PDUs and their Phy-pofiles has been already been determined. The macro scheduling problem decides which Phy-profiles will be accomodated in this frame and which PDUs will be transmitted for every selected Phy-profile. The micro scheduling decision determines the number of bursts that will be used for each Phy-profile and the location of each rectangle within the frame. However, the algorithm proposed consider only one burst per Phy-profile. Also, they address this problem only for single-hop networks.

An optimization problem that minimizes the transmit power of the BS and RSs in OFDMA-based relay networks is formulated in [8]. In the network considered an SS may have a direct connection or a two-hop connection via an RS to the BS. The authors use instantaneous channel state information and address the problem of dynamic resource allocation in terms of subcarriers, bits and power to the links in a relay network. The resource allocation is done such that the transmit power is minimised subject to the constraint that the requested data rate on each link is met. The RSs are assumed to be half duplex and can transmit simultaneously separated in the frequency domain. First an algorithm that allocates subcarriers is used, then a greedy approach for bit and power loading is applied for each link.

In [9] the authors address resource allocation for QoS traffic in OFDMA based wireless networks. The considered network comprises of a base station transmitting to a number of mobile users. The authors propose a technique for determining the rate requirement for delay constrained sessions. Then, based on the rate requirements, they propose algorithms for resource allocation that achieve proportional fairness for users and short term rate guarantees for real-time applications subject to power and bandwidth constraints.

The authors in [10] do joint routing and scheduling so as to maximize spatial use. The routing algorithm constructs routes with minimum interference. Two parameters are defined- blocking value of a node and the blocking metric of a route. Blocking value $b(n)$ of a node n is the number of blocked nodes when node n is transmitting. The blocking metric B(k) of a multihop route indicates the total number of blocked (interfered with) nodes along the route from the root node toward the destination node k. Therefore, the blocking metric of a route is a summation of the blocking values of the nodes that transmit along the route. An example of how blocking metric is calculated is shown in Figure 6.

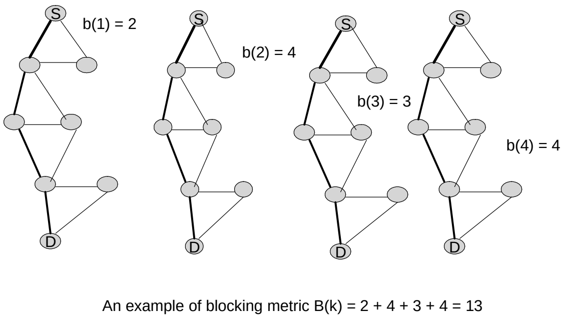

Figure 6: Blocking value and Blocking metric

The algorithm starts with a single mesh BS node and adds one SS into the mesh at a time. When an SS node joins the mesh, it selects the node with minimum blocking as its parent node. The goal of the scheduling algorithm is to maximize the number of concurrent transmissions, without creating exceeding interference for other simultaneous transmissions. The capacity request of an SS node is converted into link demands. The scheduling algorithm iteratively determines the set of active links at any time t. In each allocation iteration t, a link with the highest unallocated traffic demand is selected for next allocation of a unit traffic. The algorithm ensures the set of active links excludes interfering links. The iterative allocation continues until there is no unallocated capacity request. The drawback of this scheme is that the fairness among the SSs is not considered and SSs farther away from the BS may not get fair share of the resources.

The authors in [11] propose a simple and generalized even-odd framework for link activation. Their main goal is to satisfy bandwidth and delay requirements of flows. Each node is alternately labeled even or odd; even nodes transmit in even timeslots and odd nodes transmit in odd timeslots. Heuristics for construction of efficient backhaul routes are provided that construct routes with alternate even and odd links. They provide a mapping between the half-idle even-odd framework with an imaginary wireless system and show by analysis that when a multihop wireline scheduler with worst case delay bounds is implemented over a wireless backhaul, the even-odd framework guarantees approximately twice the delay compared to the corresponding wireline strategy. Their algorithm takes subchannelization into account and so it is suitable for OFDMA networks. Also, bandwidth and delay guarantees are provided. However, the routing algorithm has to ensure that no two interfering even or odd links are present in the same route. Thus, this scheme may not be able to find a feasible routing even if a route exists between two nodes. Also as each link is scheduled only half the time, the bandwidth requirement of each link must not exceed half of the total capacity of the link.

Flow based heuristic admission control and scheduling algorithms for multi-hop WiMAX networks are proposed and compared in [12]. Flows are scheduled by service class priority and within each service class by arrival time first. The deadline of a flow is computed from its arrival time, latency and number of hops. A flow is scheduled close to its deadline and the slot allocation then proceeds towards the beginning of the frame. A flow is admitted only if its end to end QoS requirement is guaranteed. The algorithms discussed in the paper are for an OFDMA physical layer and the time-frequency slot allocations are done in spectrum efficient manner while considering interference constraints. The authors define the “Schedule Efficiency” metric that is the proportion of the weighted measure of the admitted flows to the weighted measure of all flows seeking admission. Based on this metric the ”schedule flow subchannel” algorithm outperforms the other algorithms. However, the authors do not consider fairness in their algorithms and also do not account for the overhead in the uplink map.

In [13] a SS is assigned service tokens based on its traffic demand. In each timeslot a link is selected based on a certain criterion and the service token of the transmitter is decreased and the service token of the receiver is increased by one. The algorithm then finds another non-interfering link that can be scheduled in the same timeslot. However, the channel utilization rate of this algorithm is very low.

The objective of [14] is to formulate a scheduling problem which maximizes the system throughput under the fairness model defined by the authors. The fairness is considered at the granularity of a SSaggregated flow. Each SS node i is assigned a weight f_i that is determined according to pricing or some other criteria. The authors define the uplink capacity region C of the WiMAX mesh. If the traffic demand vector s is within the capacity region, then the traffic demand can be met even though the bandwidth request s_i may not be proportional to f_i. If the traffic demand vector is outside the capacity region, then fairness constraints are imposed for those SS nodes whose demands cannot be met without violating the fairness constraints relative to other nodes. The optimal fair rate allocation (OFRA) problem is then formulated and a fair uplink scheduling algorithm is proposed that solves the OFRA problem and finds the optimal rate allocation vector.

Tang, Xue and Zhang [15] studied bandwidth allocation in multi-channel multihop wireless mesh networks. They tried to maximize network throughput and enhance fairness at the same time. They use a max-min fairness model to achieve fairness and provide an LP that maximizes the minimum bandwidth allocation. The authors also propose a heuristic to solve the lexicographical max-min bandwidth allocation problem in polynomial time.

Agrawal et al in [16] discuss scheduling and resource allocation for an OFDMA-based wireless network. They address the problem of subcarrier allocation to selected users while determining the transmission power and modulation scheme used for each subcarrier. They formulate the scheduling and resource allocation problem as a convex problem and characterize the solution using a dual formulation.

Another centralized scheduling scheme using multiple channels and single transceivers in a WiMAX Mesh Network is discussed in [17]. The goal in this paper is to minimize the length of scheduling defined as the number of timeslots needed to complete all the data transmissions.

Djukic and Valee [18] discuss three centralized scheduling algorithms for 802.16 networks. All the algorithms take the number of OFDM slots each link should transmit in the frame as the input and produces a transmission schedule. The algorithms ranks the links. Links with lower rank transmit before higher-ranked links. The algorithms differ in the link ranking schemes and the assignment of transmission opportunities to the links.

The first algorithm, based on the IEEE 802.16 standard, does link ranking based on a breadth-first traversal of the routing tree. The links are then scheduled one after another and hence there is no spatial reuse.

The second algorithm, which is a link scheduling version of the node load-balancing algorithm proposed in [19] works in iterations. At the beginning of each iteration, a link is ranked based on its satisfaction with the schedule in the previous iteration. The satisfaction s_j of a link e_j can be defined as:

$$s_j = \frac{\hat{r}_j}{r_j}$$

where $\hat{r}_j$ is the rate achieved with the previous schedule and r_j is the required bandwidth of the link. In the next iteration, links with the lowest satisfaction index transmit before links with high saisfaction index which ensures links with low satisfaction index get higher bandwidth than the other links. Based on the ranks, a link is allowed to transmit in the first transmission opportunity that does not overlap with any conflicting links.

The third algorithm, which is based on [20], finds a ranking with good TDMA delay properties. A bandwidth optimal ranking is found using branch-and-bound search techniques. A ranking is bandwidth optimal if link bandwidths resulting from its schedule cannot be increased with a schedule from another ranking. A conflict graph is created with the links as vertices and conflicts between the links as edges. A transmission schedule is determined with the Bellman-Ford algorithm based on minimum distances in the conflict graph.

VIII. Research Issues and Challenges

It can be inferred from the above discussion that scheduling in WiMAX networks is a challenging problem. Some of the key issues involved are:

1. Routing: In a mesh network, multiple routes may exist between two nodes. Hence algorithms should evaluate different routing strategies like minimum hop, minimum packet loss, minimum intererence, to ascertain their affect on scheduling.

2. QoS: None of the algorithms except [12] and [11] consider delay. Also none of the algorithms consider jitter. Delay and jitter are very important for real-time audio and video applications and hence algorithms that address delay and jitter needs to be proposed.

3. Fairness: New fairness metrics need to be defined that takes into account Service Level Agreements between the service providers and

custmomers. The scheduling algorithm should ensure that BE traffic is not starved while giving higher priority to UGS, ertPS and rtPS flows. Fairness has to be implemented at the SS level as well as the user level.

4. Distributed scheduling: A performance analysis of the distributed scheduler was done by [21]. However, more research is required to determine slot assignment for data transfer and holdoff exponent values. Also, different distributed scheduling techniques need to be proposed that can do effficient scheduling while taking into account the interference and half-duplex constraints of the nodes.
5. Adaptive burst profile: WiMAX networks support adaptive burst profile. It is therefore important to study how burst profiles can be modified to increase the scheduling efficiency.
6. Overheads: A scheduling scheme is incomplete if the various overheads and constraints are not considered. Overheads include size of the uplink and downlink map, schedule propagation and signaling overhead and interference constraints in multi-hop networks. Hence, algorithms that minimize wasted bandwidth, scheduling overhead and interference for multihop WiMAX networks need to be developed.

While many of the issues listed above have been studied in isolation or in conjunction with a few more, however a complete solution with realistic assumptions is still lacking especially in the area of scheduling for OFDMA based WiMAX networks.

IX. Conclusion

Multihop WiMAX networks increase the capacity and coverage area of single-hop WiMAX networks by either allowing subscriber stations to communicate with each other or by using relay stations that act as ”virtual” base stations to mobile stations. However scheduling become an even more challenging problem in multihop networks. An efficient and fair scheduling algorithm that maximizes network throughput and minimises overhead while considering bandwidth, delay and jitter requirements of the various flows needs to be developed. The algorithm should also take into account interference constraints, channel conditions and adaptive modulation rates. As WiMAX technology becomes more and more popular, these issues should be addressed in an efficient to ensure long term commercial viability of WiMAX products.

References

[1] D. Bayer, N. v. Waes, and C. Eklund, “Tutorial:802.16 mac layer mesh extensions overview,” Website, 2002, ”http://grouper.ieee.org/groups/802/16/tga/contrib/S80216a-02_30.pdf”.

[2] C. Eklund, R. B. Marks, K. L. Stanwood, and S. Wang, “IEEE Standard 802.16: A Technical Overview of the WirelessMAN Air Interface for Broadband Wireless Access,” *IEEE Communications Magazine.*, 2002.

[3] D. Johnston and J. Walker, “Overview of IEEE 802.16 security,” *IEEE Security and Privacy Magazine*, 2004.

[4] H. Shetiya and V. Sharma, “Algorithms for Routing and Centralized Scheduling to Provide QoS in IEEE 802.16 Mesh Networks,” in *WMuNeP*, 2005.

[5] A. Sayenko, O. Alanen, J. Karhula, and T. Hämäläinen, “Ensuring the QoS requirements in 802.16 scheduling,” in *MSWiM*, 2006.

[6] V. Singh and V. Sharma, “Efficient and Fair Scheduling of Uplink and Downlink in IEEE 802.16 OFDMA Networks,” in *IEEE WCNC*, 2006.

[7] R. Cohen and L. Katzir, “Computational Analysis and Efficient Algorithms for Micro and Macro OFDMA Scheduling,” in *IEEE Infocom*, 2008.

[8] C. Muller, A. Klein, F. Wegner, M. Kuipers, and B. Raaf, “Dynamic Subcarrier, Bit and Power Allocation in OFDMA-Based Relay Networks,” in *Proceedings of 12th International OFDM Workshop*, 2007.

[9] T. Girici, C. Zhu, J. R. Agre, and A. Ephremides, “Practical Resource Allocation Algorithms for QoS in OFDMA-based Wireless Systems,” in *2nd IEEE International Broadband Wireless Access Workshop*, 2008.

[10] H.-Y. Wei, S. Ganguly, R. Izmailov, and Z. J. Haas, “Interference-aware IEEE 802.16 Wimax mesh networks,” in *Vehicular Technology Conference*, 2005.

[11] G. Narlikar, G. Wilfong, and L. Zhang, “Designing multihop wireless backhaul networks with

delay guarantees," in *Proceedings of Infocom*, 2006.

[12] D. Ghosh, A. Gupta, and P. Mohapatra, "Admission control and interference-aware scheduling in multi-hop wimax networks," in *IEEE MASS*, 2007.

[13] B. Han, W. Jia, and L. Lin, "Performance evaluation of scheduling in IEEE 802.16 based wireless mesh networks," *Computer Communications.*, 2007.

[14] M. Cao, V. Raghunathan, and P. Kumar, "A Tractable Algorithm for Fair and Efficient Uplink Scheduling of Multi-hop WiMAX Mesh Networks," in *WiMesh*, 2006.

[15] J. Tang, G. Xue, and W. Zhang, "Maximum throughput and fair bandwidth allocation in multi-channel wireless mesh networks," in *Proceedings of IEEE Infocom*, 2006.

[16] R. Agarwal, R. Berry, J. Huang, and V. Subramanian, "Optimal Scheduling for OFDMA Systems," in *Proceedings of 40th Annual Asilomar Conference on Signals, Systems and Computers*, 2006.

[17] P. Du, W. Jia, L. Huang, and W. Lu, "Centralized Scheduling and Channel Assignment in Multi-Channel Single-Transceiver WiMax Mesh Network," in *WCNC*, 2007.

[18] P. Djukic and S. Valaee, "Scheduling algorithms for 802.16 mesh networks," *WIMAX/MobileFi: Advanced Research and Technology*, 2007.

[19] D. Kim and A. Ganz, "Fair and efficient multi-hop scheduling for IEEE 802.16 BWA systems," in *Broadnets*, 2007.

[20] P. Djukic and S. Valaee, "Link scheduling for minimum delay in spatial re-use TDMA," in *Infocom*, 2007.

[21] M. Cao, W. Ma, Q. Zhang, X. Wang, and W. Zhu, "Modelling and performance analysis of the distributed scheduler in IEEE 802.16 mesh mode," in *Mobihoc*, 2005.