Ubiquitous Fair Bandwidth Allocation for Multimedia Traffic on a WiMAX Mesh Network with Multi-channels

Scott Fowler and Xingjun Zhang

Linköping University Post Print

N.B.: When citing this work, cite the original article.

Original Publication:
http://dx.doi.org/10.1504/IJAHUC.2012.047009
Copyright: InderScience

Postprint available at: Linköping University Electronic Press
http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-56773
Ubiquitous Fair Bandwidth Allocation for Multimedia Traffic on a WiMAX Mesh Network with Multi-channels

Scott Fowler*
Linköping University
Department of Science and Technology
SE-601 74, Norrköping, Sweden
E-mail: scott.fowler@liu.se
*Corresponding author

Xingjun Zhang
Xi’an Jiaotong University
Department of Computer Science
School of Electronic & Information Engineering
710049, Xi’an, China
E-mail: xjzhang@mail.xjtu.edu.cn

Abstract: The WiMAX, also known as IEEE 802.16, provides a mechanism for deploying high-speed wireless mesh network, (an ubiquitous wireless network) in metropolitan areas. WiMAX technology can be used as “last mile ubiquitous” broadband connections to deliver streaming audio or video to clients. Thus, Quality of Service (QoS) is very important for WiMAX networks. Providing QoS in multi-hop WiMAX networks such as WiMAX mesh networks is challenging since the WiMAX mesh networks MAC is connectionless based and does not have proper support guarantees for QoS over multiple hops. As a result, multiple links can interfere with each other when they are scheduled at the same time. The scheduling function plays a crucial role in QoS support, and various algorithms have been proposed and analyzed. These analysis basically assume a backlogged situation evenly in all queues. However, multimedia traffic is bursty in nature, and the fairness of bursty traffic relative to continuous traffic has not been fully studied yet. Therefore, in this paper, we will discuss the potential unfairness that bursty traffic may be subjected to and propose a new frame-based packet scheduling algorithm. The scheduling is evaluated with multiple channels on WiMAX mesh nodes, which are operated in a distributed coordinated scheduling mode. Each wireless node has a single radio interface which is able to switch between multiple channels. We evaluated the performance of the proposed dynamic changing scheduling method by extensive simulations, and it was shown to provide fair bandwidth allocation while increasing the traffic performance by means of throughput.

Keywords: 802.16, WiMAX, Scheduling, Wireless Mesh, Multi-Channel, Bandwidth Allocation, Quality of Server, Ubiquitous Wireless Networks.

Reference to this paper should be made as follows: Fowler, S. and Zhang, X. (xxxx) ‘Fair Bandwidth Allocation for Multimedia Traffic on a WiMAX Mesh Network with Multi-channels’, Int. J. Ad Hoc and Ubiquitous Computing, Vol. x, No. x, pp.xxx–xxx.

Biographical notes: Scott Fowler received a B.Sc. (1998) from Minot State University, Minot, ND, U.S., M.Sc. (2001) from the University of North Dakota, Grand Forks, N.D., U.S. and Ph.D. (2006) from Wayne State University, Detroit, MI, U.S., all degrees are in Computer Science. In 2006 - 2010 he was a Research Fellow in the Adaptive Communications Networks Research Group at Aston University, UK. Since 2010 he has been a staff member at Linköping University, Sweden. His research interests include, Quality of Service (QoS) support over heterogeneous networks, Computer networks (wired, wireless), Mobile Computing and Performance Evaluation of Networks.

Xingjun Zhang received a Ph.D (2003) from Xi’an Jiaotong University, X’an, China. He is an Associate Professor in the Computer Science and Technology Department of the School of Electronic and Information Engineering at Xi’an Jiaotong University. From 2006 to 2009 was a Contract Research Fellow in the School of Engineering and Applied Science at Aston University, UK. His research interests include, Multimedia Transmission over Lossy Networks (particularly H.264/AVC, erasure coding), Web and Internet, Computer Network (computer network architectures, IP network management and measurement, network security, network coding), Distributed System, High Performance Computing (Grid computing) and Wireless Sensor Networks.
1 Introduction

Wireless Mesh Networks (WMNs) is a cost-effective method for providing alternative technology for last-mile ubiquitous broadband Internet access and high speed connectivity. WMNs have emerged as a key technology for the next generation of wireless networking. However, the existing IEEE 802.11 technology for WMNs has a very limited transmission range (Anastasi et al., 2005). Unless expensive external amplified antennas are employed, the Medium Access Control (MAC) protocol could only achieve low performance for multi-hop traffic flows (Gambiroza et al., 2004). An alternative to IEEE 802.11 is the IEEE 802.16 standard (IEEE 802.16 Working Group (Revision of IEEE Std 802.16-2001) (2004), 2004) or commonly known as WiMAX. The use of WiMAX for WMNs is an efficient method large area coverage as well as last-mile ubiquitous communication since WiMAX can have a range of up to 50 km at data rates of 75 Mbps using both unlicensed and licensed frequency bands. The last-mile connection to buildings often causes problems in providing high-speed access to subscribers such as SOHO (Small Office Home Office) or larger businesses. The installation of DSL and cable solutions are likely expensive and laborious. Instead, the use of WiMAX for this last-mile ubiquitous link will guarantee a lower cost whilst offering a comparable speed.

There are two ways to deploy a WiMAX configuration. The first one (Figure 1(a)) is Point to Multipoint (PMP). The PMP configuration of WiMAX consists of subscriber stations (SSs) that communicate to the Base Station (BS) via high-speed wireless link. The BS acts as a gateway to the Internet/Network backbone.

The second deployment of the WiMAX configuration is the WiMAX Mesh Network (Figure 1(b)). A BS is surrounded by numerous SSs, and some of which are either not within the BS’s physical layer range or having physical obstructions. A SS that is unable to have a one-hop link to a BS may instead logically attach itself to a neighbouring node which is available, thereby gaining access to the network. This feature of Mesh significantly increases the coverage area of the network as illustrated.

Moreover, in the WiMAX configuration, the Mesh mode will become a more flexible and faster approach for network deployment since WiMAX is capable of providing high-speed data and telecommunication services compared to the other emerging 4G technologies (Santhi et al., 2006). This is enabled by the WiMAX MAC which features multimedia support for the next-generation wireless networks.

The coordination of nodes in a WiMAX WMN is distributed. In other words, transmissions in WMN are scheduled in a fully distributed fashion and do not require any interaction with the BS. In the distributed mode, the WiMAX standard specifies a MAC protocol to coordinate the transmission of control messages in a collision-free manner (Cao et al., 2007). On the other hand, WiMAX PMP protocol is connection oriented. Therefore, SS must register to the BS prior to sending/receiving data, and the BS is responsible for defining the schedule of transmission in the entire network (i.e. centralized coordination mode).

Despite all these advantages, WiMAX WMN is not sufficient for supporting guaranteed QoS over multiple hops because the WiMAX WMN MAC is connectionless-based. Therefore, in order to implement QoS in WiMAX WMN, several modifications have to be made. The first step would be to define the bandwidth allocation for multimedia traffic. So far, however, the bandwidth allocation problem in the distributed mode is left unsolved by the WiMAX standard. Although some control messages that may be used for this purpose, such as bandwidth requests and grants (Cicconetti C., Akyildiz et al., 2007), apparently they are not sufficient enough for multimedia.

Two major classes of queue indexing schedulers which may potentially be utilize for bandwidth allocation for QoS implementation are Timestamp-based and Fair Queue Indexing schedulers. Timestamp-based schedulers, such as Virtual Clock (VC) (Zhang, L., 1990), and Self-Clocked Fair Queuing (SCFQ) (Golestani., 1994), can achieve a good approximation of the Generalized Processor Sharing (Generalized Processor Sharing: is an ideal scheduler that serves an infinitesimal amount of data from each flow according to the flow’s reserved rate or relative bandwidth weight.) model with tight and low latency bounds. Consequently, they can provide good fairness by distributing excess bandwidth fairly among all contending flows according.
to their relative reserved rates (Keshav, 1997; Parekh et al., 1993). Timestamp schedulers are solely latency-sensitive, and unfortunately, have no consideration on the throughput.

Fair Queue Indexing schedulers, on the other hand, focus on the throughput. However, in order for utilizing this potential of improving throughput rate in multimedia traffic, the existing Fair Queue Indexing schedulers are somewhat deficient. They have known lack of consideration for multimedia traffic and proportionally distribute the service for the traffic flow (Cicconetti C., Akyildiz et al., 2007; Cicconetti et al., 2009). It is because the existing Fair Queue Indexing schedulers assume so-called “backlogged-situation” where packets arrive continuously. However, multimedia traffic like audio and video are bursty in nature and this backlogged situation cannot always be assumed in such traffic patterns. It is true for the existing Fair Queue Indexing schedulers, such as Weighted Fair Queuing (WFQ) (Demers et al., 1989; Parekh et al., 1993), Worst-case Fair Weighted Fair Queuing (WF2Q) (Bennett et al., 1996), as well as Frame-based schedulers (Frame-based schedulers use Round Robin and serve flows in rounds or frames. During each round a flow receives at least one transmission opportunity.), which includes Weighted Round Robin (WRR) (Katevenis et al., 1991), Deficit Round Robin (DRR) (Shreedhar et al., 1996), and Elastic Round Robin (ERR)(Kanhere et al., 2002; Kanhere, S. and Sethu, H., 2002). Hence, certain improvement is necessary in the Fair Queue Indexing schedulers when aiming for fully exploiting its throughput enhancing property in the bandwidth allocation for QoS implementation on the multimedia traffic.

In this paper we propose a fair end-to-end bandwidth allocation algorithm for WiMAX WMNs to negotiate bandwidth in a multi-channel environment (Cicconetti C., Akyildiz et al., 2007). Our contributions are summarized as follows.

- A scheduling method tailored for “busty traffic”, which is a fundamental property in multimedia, is proposed.
- Differentiated services are provided for multimedia traffic flows specified in the standard WiMAX MAC header.
- Unlike most solutions for TDMA MAC protocols, the proposed method is able to react promptly to bursty traffic characteristics of the traffic load for multimedia in the network.
- The proposed method employs a fully distributed manner so as not to encounter the overhead of signaling towards/from a centralized node.
- The proposed method provides a fair bandwidth allocation while improving throughput rate.
- Fairness is achieved by allocating bandwidth requests/grants at each neighbour in a modified round-robin fashion.
- The key feature of our modification is that traffic backlog situation is taken into account, and the bandwidth allocation is performed such that a flow with the highest importance value (multimedia traffic) is treated with priority.

The rest of the paper is organized as follows. Section 2 consists of the list of the related works. Section 3 provides an overview of the MAC in WiMAX mesh mode. In section 4, our proposed work is detailed. Section 5 explains the simulation testbed we used for this study. Both sections 6 and 7 present the simulation results of the topologies described in section 5 or randomly distributed nodes, respectively. Finally, in section 8, we make some concluding remarks.

2 Related Work

Despite many advantageous features of mesh configuration, when it is applied to wireless network the bandwidth is limited due to its low capacity. The wireless link quality is time and space varying, depending on the environment and interference. The existing approaches for link assessment consume substantial amount of time and thus introduce significant delay and overhead. Having a fast link assessment in WMNs is important so as to transmit a packet effectively. One of the traditional approaches for link assessment is to sequentially assess each link by having all nodes in the network transmit in a specific order (Ephremides et al., 1990; Smith., 1999). The major drawback of this approach is excessive time consumption that is proportional to the number of nodes in the network. For instance, consider a scenario where there are 5,000 nodes in the network. Let us assume that each link requires a time slot equivalent to 2 seconds. Then the total time required for link assessment will be 5,000 * 2 =10,000 seconds. Other research have proposed admission control and scheduling algorithms to provide QoS for applications in mesh networks. In (Chen et al., 2005; Wei et al., 2006), Call Admission Control (CAC) for Voice over IP (VoIP) in WMN was investigated. However, unlike Video on Demand (VoD), VoIP service cannot take advantage of multicasting. In (Kashyap et al., 2007), authors proposed an integrated QoS control architecture for WiMAX to support different types of traffics. This work elaborates the design and implementation issues and conducts simulation on VoIP, FTP and HTTP traffics. In spite of this, the problem of VoD over WiMAX mesh networks was not particularly addressed.

A cross-layer framework for multicast so as to maximize throughput in WMN was proposed in another study (Yuan et al., 2006). However, this framework is not suitable for video streaming application, where each stream requires certain data rate for continuous video
playback. A multi-source multi-path video streaming system was proposed in (Li et al., 2006). Their assumption is that different clients will have more than one copy of the video files, and accordingly, the assortment of the sources to select from will improve the quality of the VoD service. However, it is not a common case. In reality, the video server should be the one who is responsible for providing more reliable VoD.

Another solution has been proposed in (Kim et al., 2005), where resources are scheduled in two parts: first the demands of the nodes are collected by the BS and flooded throughout the WMN. In the second part individual wireless nodes execute a collision-free schedule based on the states of their neighbours and extended neighbours. However this introduces extra traffic in the network, thereby causing a decrease in the wireless bandwidth.

The authors of (Cicconetti C., Akyildiz et al., 2007) proposed a fair end-to-end bandwidth allocation (FEBA) algorithm to address the end to end throughput over multi-hops. The FEBA was implemented at the MAC layer of single-radio and multi-channels on WiMAX mesh nodes, which are operated in a distributed coordinated scheduling mode. FEBA negotiates bandwidth among neighbours to assign a fair share proportionally to a specified weight to each end-to-end traffic flow. This was done by use of the fairness index. The bandwidth requesting and granting is carried out in a round-robin fashion, where the amount of service at each round is proportional to the number of incoming or outgoing flows. Unfortunately, however, FEBA (Cicconetti C., Akyildiz et al., 2007) does not consider multimedia traffic, which is a prerequisite for providing quality of service.

3 Background

3.1 MAC overview in WiMAX mesh mode

To describe WiMAX distributed scheduling algorithm, we illustrate the frame structure in Figure 2. In the WiMAX standard, both the control messages and data packets are transmitted in the same channel, but in different time subframes. Control sub-frames are partitioned into slots of fixed duration (labeled Control slot), which are accessed by nodes based on the distributed election procedure. This will make certain that each node has the chance to transmit control messages in a regular manner. The control frame may consist up to 16 transfer occasions for data transmission, with each transmission opportunity being seven OFDM symbols. Since this study is using multi-channels, control messages are transmitted by all nodes in the network in the same channel. An example is illustrated in Figure 3. Data subframes are composed of fixed number of data mini-slots (hereafter referred to as slots) as shown in Figures 2 and 3. The number of bytes transmitted in a slot will depend on the Modulation and Coding Scheme (MCS) used by the sending node. In WiMAX, every node adjusts the MCS of its neighbour based on the measurement of the received signal on the physical layer.

The control message defined by the WiMAX forum for Mesh nodes is Mesh Distributed Schedule (MSH-DSCH). The MSH-DSCH message of each node contains the schedule and information of data subframe allocation for its one-hop neighbours as well as its own. By broadcasting MSH-DSCH messages, each node will have knowledge of its neighbours up to two hops away.

In a coordinated distributed scheduling, MSH-DSCH message is the key component in the whole scheduling process. During distributed scheduling, request and grant of channel resource are delivered
by MSH-DSCH message among nodes, while every node sends its available channel resource table to the neighbouring nodes with MSH-DSCH messages. The WiMAX standard defines a three-way handshake (Figure 4) procedure that uses specific signaling messages to request, grant, and confirm available resources (bandwidth). A two-way handshake (i.e. the absence of the confirmation step) is also possible (Figure 5), but all nodes need to be in the same transmission range.

A MSH-DSCH message should include a list of Information Elements (IEs) in the following four specific signaling fields. The requesting node sends a request IE in the MSH-DSCH to inform the receiver that there is data waiting to be transmitted. The receiver reserves slots (bandwidth) over a range of frames in a given set for the sender. The receiver then acknowledge it by sending grant IE to the sender. A grant IE is expressed by [slot range, frame range, channel]. For example, [{7,12},{3,8},3] stands for slots range from seven to twelve, data sub-frame three and eight in channel three. Next, the sender responds with confirmation IE expressed in the same format as grant IE. Finally, after completing the three-way handshake, availability IE may be used to inform the sender of the slots on available slots for transmitting or receiving data.

3.2 Problem in WiMAX mesh mode

The problem with the preceding coordinated distributed scheduling is collision. Collision will arises by means of “protocol-model” (Gupta et al., 2000). Consequently we make the same assumption as in (Cicconetti C., Akyildiz et al., 2007) where the receiving or sending node transmits on the same channel. The receiving/sending node will need to kept track of all of the [slots, frame, channel]. This will be done by grant bitmaps when a node transmits/receives an MSH-DSCH message.

As long as all nodes are in the same transmission range, a two-way handshake is sufficient. However, this is not always the case in a WMNs. A problem occurs with the three-way handshake as we consider the basic topology setup shown in Figure 6. In the figure Node C is within the radio range of node B and D, but not in the transmission range of node A. If both flows are backlogged, traffic flow from node A to B will receive significantly higher throughput than flow C to D. Using the same topology, in the Figures 7, 8 and 9, the arrow with a solid line represents a link between two nodes and the arrow with dotted line stands for a communication line received but not for the node.

The problem in relation to the Three-way handshake in WiMAX is depicted in Figure 7. Consider the case when Node A sends a bandwidth request to Node B, and more or less at the same time, Node C also sends a request to Node D. When Node B receives the request from Node A, it will grant the information of channel resource (grant:IE or GNT(X):IE). Since node D is not a neighbour to Node B, it is oblivious of the grant from Node B to Node A. As a result, the same slot(s) (GNT(X):IE) may be allocated to Node B and Node D, making the request from Node C invalid. Consequently, Node C will have to withhold from sending the confirmation (CNF(C):IE). Now, Node C needs to repeat the request for resources, resulted in additional three stages.

The same problem will occur when Node B, instead of Node A, sends a request even though Node B is in the range of Node C (Figure 8). Node B and C send bandwidth requests to Node C and D, respectively, more or less at the same time. Should Node C grant the information of channel resource (grant:IE or GNT(X):IE) for Node B, Node C will be unable to have its request fulfilled. Note that Node D is not a neighbour to Node B, therefore it will be unaware of Node B’s request to Node C. Node D can only be aware of the request from Node C. As a result, Node D might allocate the same slot(s) for Node C, which Node C already reserved for Node B. This will make the request from Node C invalid. Consequently, Node C will have to withhold from sending the confirmation. Now, Node C needs to repeat the request for resources, resulted in additional three stages.

4 Proposed Method

Our work had various extensions on the employed method proposed by (Cicconetti C., Akyildiz et al., 2007) in NS-2 (NS-2.). This enabled us to evaluate our proposed method for multi-media traffic over a WMN and a function of multi-channel quality. The traditional
FEBA multi-channel described in the previous section computes the amount of service proportionally to the number of traffic flows. The method is not sufficient enough for multimedia traffic because it does not completely exploit the maximum throughput by transmitting multiple packets. In addition, the traditional FEBA does not consider characteristics of multimedia traffic, which are bursty in nature and the identical backlogged situation cannot always be assumed in such traffic patterns. Therefore, it is necessary to make several modifications in FEBA to be applied for multimedia traffic while maintaining its advantage of throughput-enhancing feature.

Instead of assigning weights evenly to the traffic flows, our proposed method assigns different weights when establishing a connection, which allows multimedia traffic to maximize throughput to the channel. In our simulator, each wireless node has a single radio interface which is able to switch between multi-channels. Notice that the currently proposed dynamic changing fair bandwidth allocation method is not fully depending on the use of the multi-channels for increasing the traffic performance.

In most wireless network applications, fixed position of the wireless nodes and free of the channel error cannot be assumed. However, our study implemented WiMAX wireless mesh router for backhaul of a highly reliable WMN to ensure the stability of wireless routers. In this scenario, even with the quality of the link changes over the life of the connection, the wireless nodes are not added or removed from the WMN similarly to a fixed WMN. In other words, there is no need to re-establish the connection due to the link being stable enough. Furthermore, the same assumption has been made in various published studies with similar setting (Cicconetti et al., 2009; Li, Y. et al., 2009; Li, Y., Yang, Y. et al., 2009). Therefore, the above assumption could be reasonably accepted.

Our proposed method is illustrated in the Figure 9. We propose to have the sending node to transmit information on the rest of the available resources $[AVL(node\ sending)]$ at the same time as the request $[REQ(node\ sending)]$. From the information provided by the $AVL()$ the receiving node will be able to make smarter choice on channel resources [slot range, frame range, channel] available for the receiver and the other nodes. The receiver will inform the sender by sending grant $[GNT()]$. Simultaneously, the receiver will also send information to the sender on the rest of the resources available for the other nodes $[RES()]$. Because of the information sent in $AVL()$, the sender (Node C) should be able to realize another $GNT()$ would overlap with the request sent to Node D or the grant from Node D reserved for this request. The sender (Node C), then, will take one of the following two options i) wait until it receives $GNT()$ with $RES()$ from the receiving node (Node D) and then send another request to Node D as depicted in Figure 9(a), or ii) automatically resend
request \[\text{REQ}()\] with \[\text{AVL}()\] to Node D as in Figure 9(b).

The backlog of the sending node dictates whether option i) (Figure 9(a)) or ii) (Figure 9(b)) is taken. The default is the option i), and only if the sending node has large backlog the sender (Node C) will take the option ii). In the option i), the sender (Node C) will not take an action until receiving the information on the available resources \[\text{RES}()\] which arrives with \[\text{GNT}()\]. The sender now will be able to make an intelligent decision, i.e. resending a new request to Node D immediately upon reception of the grant, thereby shortening the number of the steps by one stage comparing to the traditional method.

Option ii) (Figure 9(b)) is used only if the sending node has a large backlog (Node C). In the option ii), the sender (Node C) will automatically resend request \[\text{REQ}()\] with \[\text{AVL}()\] to Node D in the following manner. Node C recognizes the request from Node A to Node B by overhearing the grant \[\text{GNT}()\] from Node B, despite that Node A and C are not neighbors. This information enables the sender (Node C) to make a smart decision, i.e. Node C will resend a <new request> immediately after the first <unsuccessful request> without waiting for the grant \[\text{GNT}()\] from the receiver (Node D). This as a result, will eliminate the extra steps. Therefore, whether option i) or ii) is taken, the proposed method will reduce the number of stages in a three-way handshake.

4.1 Selecting Channels

Any multi-channel protocol must ensure that the sender and receiver are on the same channel before
communicating each other. It may be achieved by either switching both the sender and receiver to a predetermined channel, or by using a separate channel set aside for them.

Traditionally, the range of slots in the frame was determined based on the availability (Figure 10). The grey boxes represent both the sender and receiver in the various channels which may be used, while the white boxes represent those of which may not be used. The crossed boxes represent slots that cannot be granted because they have already been allocated for data transmission by granter or the granter’s neighbour. First, a channel will be randomly selected, and the receiver then, will compare the \textit{REQ(node requesting)} with the receiver slots available in the channel. This will define the range of slots in the frame. If no slots are available in the channel the receiver will move to the next channel. Once all channels have been compared, the receiver will move on to the next frame. Consider the case, where the receiver selected slots 1 and 4 in channel 1 (Figure 10). The receiver will still allocate slot 2 in channel 2 for the sender but miss slot 1 in channel 2. In addition, channel 2 slot 4 will also be missed since this overlaps with time slot 4 in channel 1, and therefore it will not be granted. This is because the traditional \textit{REQ()} does not have information about availability of the slot.

![Figure 11](image-url) Proposed way for allocation of data slots in WiMAX technology.

On the contrary, in our method, the receiver do not have to examine the availability of each slot one by one. The available slot will be identified ahead of time because the \textit{REQ()} is sent with the recorded list of slots that have been allocated for the sender. Furthermore, the receiver will be able to realize if there are overlaps of slots. Therefore, the receiver can skip over unavailable slots, rather than visiting each single slot (Figure 11). Thus, all the requests for the receiver slots can be more efficiently fulfilled. In a real system, it takes some time for nodes to switch from one channel to another and for the overhead to recognize the overlap slots which should be taken into account by nodes for allocating slots, but this is beyond the scope of the simulator. The once the requests are fulfilled, the receiver will transmit \textit{grant:IE} followed by MSH-DSCH, which contains the following information: \{1,4,\{x,x\},1\}, \{2,\{x,x\},2\}, \{slots_{n+1},y_{n+1}\}, \{slots_{n+1},y_{n+1}\}, \{channels_{n+1}\}.

### 4.2 Assigning Fair Queue

In order to evaluate the fairness of scheduling algorithms, Shreedhar and Varghese (Shreedhar et al., 1996) used a metric called \textit{FairnessIndex}. The \textit{FairnessIndex} is defined as:

\[
\text{FairnessIndex}_i = FQ_i \frac{\sum_{j=1}^{n} f_j}{f_i}
\]

where $f_i$ is a quantity which expresses the ideal share to be obtained by flow $i$. $FQ$ denotes Fair Queue and $FQ_i$ represents the Fair Queue for flow $i$. $FQ_i$ is given by:

\[
FQ_i = \max \left( \lim_{t \to \infty} \frac{send_{i,t}}{sent_t} \right)
\]

where $send_{i,t}$ is the total number of bytes sent by flow $i$ by time $t$, and $sent_t$ is the total number of bytes sent by all $n$ flows by time $t$. The maximum is taken across all possible input packet size distributions for all flows.

Let us consider an ideal case in which the average length of a packet is known on $flow_{i}$ in every round. Packet scheduler has rounded $N$ times by time $t$. We define $s_{i,n}$ as number of bytes sent by flow $i$ at round $n^{th}$:

\[
 send_{i,t} = \sum_{n=1}^{N} s_{i,n}
\]

Now, consider the case in which the number of bytes sent at the $n^{th}$ round ends at $t_{n}$. In this scenario, the number of bytes in the flow will be corresponding to queue $i$ at time $t$, with weights $w_1,w_2...w_N$, which is expressed as:

\[
p_{i}(t_{n}) = \frac{s_{i,n}}{w_{i}(t_{n})}
\]

where $p_{i}(t_{n})$ is the number of bytes in the flow corresponding to queue $i$ at time $t_{n}$, $w_i$ represents a weight for flow $i$, and the $w_i(t_{n})$ is derived from the requested bandwidth ($BW$) for flow $i$ at time $t_{n}$.

If $N$ data flows are currently active over queue $i$, with weights $w_1,w_2...w_N$, then we have:

\[
\sum_{i=1}^{N} w_i = 1
\]
Then this means:

\[ W_i = \frac{1}{w_i} \quad (6) \]

where \( W_i \) is a fraction of \( w_i \) out of the total weight.

In this case we can express \( FQ_i \):

\[ FQ_i = \max \left( \lim_{t \to \infty} \frac{\sum_{j=1}^{N} W_{j,t}}{R \sum_{j=1}^{N} sent_{j,t}} \right) = \frac{1}{R} \quad (7) \]

where \( R \) is a link data rate at any given time over the \( N \) active data flows, \( sent_{j,t} \) represents the total number of bytes sent by flow \( j \) at time \( t \), and \( W_{j,t} \) is a fraction of the total weight for flow \( j \) at time \( t \). However, if intermittent or bursty traffic arrives, the previously unused share is not taken into account. Therefore such traffic may be treated unfairly in comparison to continuous traffic, even when two or more traffic sources are transmitted at the same rate. This will result in a situation called “backlog”. In backlog, a number of bytes are awaiting transmission from one node to another for which no bandwidth request have been issued.

Now if we add the backlog to the packets to be sent, we obtain the following equation:

\[ W'_i(t) = w_i(t) + w'_i(t) = c_i(t) \quad (8) \]

this means lower priority traffic will loose their share by \( w_i \), whereas the higher priority traffic will gain the share by \( w_i \) (Figure 12). Note that based on equation (6) it can be stated that \( W_{i}^{max} \) is present to ensure an acceptable threshold:

\[ w_i(t) \leq W_{i}^{max} \quad (9) \]

![Figure 12 Assigning Weights.](image)

In service request, all the active flows calculate their index and \( BW_{REQ} \) to form a tuple of \( BW_{demand} \). The tuple is sent to the next node (grant node) during bandwidth request burst. On receipts of \( BW_{demand} \) tuples from all the active flows, the receiving node processes these demands. The demands are then arranged in descending order based on Type of Service values. Bandwidth allocation is performed such that a flow with the highest importance (in our case multimedia traffic) value is awarded its requested bandwidth at first, followed by the flow with the second highest request (which is Best effort traffic) value and so on, until the entire bandwidth is exhausted or the allocation is completed. Shown in the pseudo code is the functioning of algorithm both at request and grant node:

**Algorithm Request:**

**input:** Active Flow ID, \( I \) 

begin  

for each active flow \( i \):

compute Type of Service, \( TS_i \)

if \( TS_i == Multimedia \) 

compute Packet bytes buffer occupancy ratio, \( BO_i \)

measure Packet bytes for Bandwidth Request, \( BW_{req} \)

else 

compute Packet bytes buffer occupancy ratio, \( BO_i \)

Formulate Bandwidth demand \( BW_{demand} \)

Forward \( BW_{demand} \)

end

**Algorithm Grant:**

**input:** \( BW_{demand} \) from all active flows 

begin  

sort \( BW_{demand} \) received in descending order of \( TS_i \)

while (link bandwidth > 0 ) 

compute available bandwidth in channels 

allocate bandwidth in channels for the flow 

remove this flow from sorted list of flows

end

**5 Simulation**

To evaluate the proposed adaptive sending rate over a WMN, all experiments were simulated with TCP and UDP connections as competing traffic for each scenario of 5 minutes. The results were average of 5 simulation runs with different random seeds. The resulting network link bandwidth for all simulations ranged from 12.19 Mbps for a QPSK-1/2 user, and up to 55.78 Mbps for 64-QAM 3/4 node. The modulation and coding scheme (MCS) for a 16-QAM-1/2 was 24.69 Mbps, which is about two-folds higher than the QPSK-1/2 throughput and about half of 64-QAM-3/4. We did not take into account for the channel errors so that we could focus on the results of the performance at the MAC layer.
5.1 Simulation Setup

All the WMN queue sizes were set to 100 kB buffer. The network covered an area of 2000 x 2000 meters, where the nodes were at least 200m apart from each other. Each node was configured to a transmission range of 250m. We also had performed simulations with different chain lengths, and obtained very similar results (data not shown).

As stated, multimedia traffic is bursty in nature. Therefore, fairness in bursty traffic relative to that in continuous traffic should be examined for multimedia traffic. Thus, the proposed method was tested against video traffic and voice packets. For the video traffic source, we used a trace of real H.264 video stream. More information on the video trace may be found in (Auwera et al., 2009; Auwera, G.V.d. et al., 2008; Seeling et al., 2004). For the voice traffic, we used a Constant Bit Rate (CBR) traffic with a packet size of 340 bytes. Voice packets were sent at a fixed rate of 352 ms, and no packets were sent for 650 ms (Ali et al., 2005; Chuah et al., 2002).

In wireless networks, multimedia traffic is identified and treated differently from other data traffic by the following two ways. One way is to use the Type of Service Field (ToS) inside of the IP header, which requires the wireless packet to be transverse up to the network layer to identify the type of traffic.

The other method is specified in the standard MAC header. The 3 bit field in the standard MAC header allows different traffic flows to be mapped for different type of service. WiMAX is equipped with the 3 bit field. Therefore, WiMAX will easily be able to accommodate the various multimedia requirements even more so than other WMN technology such as 802.11s (802.11s draft indicates the Enhanced Distributed Channel Access (EDCA) has the ability to accommodate classes of services.). Furthermore, this field enables QoS for various LANs through 802.11d (Cicconetti et al., 2009; IEEE 802.1D (2004), 2004), thus QoS through an 802.16 WMN should also be achievable. Therefore in this study, we implemented the 3 bit field in WiMAX to work with NS2 WiMAX model. Recall that the objective of this paper is to address the bottleneck in wireless, therefore the focus should be on the MAC layer rather than the IP layer. The use of IP layer may be more applicable to the wired network study.

Additionally, we repeated the same scenario using either one or two channel(s) on wireless nodes to examine the effects of increased number of channels on the throughput enhancement. Simply adding an extra channel does not double the throughput where single radio interface is applied on the wireless nodes. When three nodes are used with two channels, for example, it is not possible to make use of the both channels for increasing throughput because the middle node always occupies one of the channels to take part in the communication. As a result, only one of the channels can be utilized for throughput. Although the magnitude of throughput amplification will not be simple multiplication of to the channel number, the possibility of certain effects of the use of multi-channel should not be ignored.

5.2 Topologies

In our simulations with deterministic topologies and fixed packet streams, we considered the following scenarios as shown in Figures 13 - 14 which represent the testbeds used.

![Figure 13 Bidirectional-Chain (Chain) Wireless Mesh Topology.](image)

The first topology (Figure 13) is based on the use of one “Bidirectional-Chain” (Chain). We used two connections along this chain. In the Bidirectional-Chain, the one connection travels from the first to the last node in the chain. The second connection running in the opposite direction, from the last towards the first node, provides competing traffic which is carried by UDP and TCP.

![Figure 14 Bidirectional-Cross (Grid) Wireless Mesh Topology.](image)
We also evaluated the proposed method by simulating uplink and downlink. To achieve this, we set up a network with 19 nodes as done in (Cicconetti C., Akyildiz et al., 2007), and arranged them in a topology of a six-pointed star, as shown in Figure 15. The node in middle is the centre node. Let a traffic flow entering the centre node be the uplink (traffic entering) and leaving the centre node be the downlink (traffic leaving). Each neighbour of the centre node has a bi-directional traffic flow traversing the various paths associated with each other.

6 Performance Evaluation/Results

The focus of this research is to improve the traffic performance, in other words, to refine the average rate of successful message delivery over a wireless communication channel. We are particularly interested in the performance over a given period of time by means of throughput. Recall that our study implemented WiMAX wireless mesh router for backhaul of a highly reliable WMN to ensure the stability of wireless routers. Therefore, our simulation results are based on the reasonable assumption of fixed position of the wireless nodes and free of the channel error. Figures 16 - 27 show the average throughput results for the traffic as the number of hops increases over the WMN.

6.1 Chain and Grid

The test results for the Chain topology are depicted in the Figures 16, 18, 17 and 21, while those of the Grid topology are in the Figures 19, 22, 20 and 23. In the Grid (Bidirectional-Chain) topology, there is an increase in the number of traffic streams traveling in y direction of a plane as the number of y chains increases (TCP and UDP), while along the x plain, only one traffic flow is transverse. Thus, the traffic flows must share one common node eventually. The resulting competing background traffic is indicated by the fluctuation of transmission rates of TCP and UDP connections. This in turn, results in the decline of the throughput rate. On the other hand, the Chain testbed has less competition since it only has limited numbers of traffic flows, and therefore, TCP flows are able to adjust its sending rate accordingly. Because of its simple topology, Chain testbed should have higher throughput rate comparing to Grid topology. It was true for our simulation results. Regardless of the traditional or our proposed method, when the performance of Chain and Grid topology was compared between the same channel number, modulation scheme, and type of traffic, overall average throughput rate was higher in the Chain topology. In addition, every single data point in the series of our simulation indicated that Chain topology has higher performance.

All the graphs on the left side (Figures 16, 18, 19 and 20) are the results from traditional method, while the ones on the right (Figures 17, 21, 22 and 23) employed our proposed method. The graphs listed side by side are using the same topology (Chain or Grid) and type of traffic (VoD or VoIP), but different method (Traditional or Proposed). Types of traffic tested are VoIP (Figures 16, 19, 17 and 22) or VoD (Figures 18, 21, 20 and 23).

In all cases, regardless of the topology, type of traffic, or proposed/traditional method, there were general tendencies that: 1) the throughput is proportional to the MCS efficiency, 2) two-channel application performed
Figure 18  Chain for Traditional using VoD.

Figure 19  Grid for Traditional using VoIP.

Figure 20  Grid for Traditional using VoD.

Figure 21  Chain for Proposed using VoD.

Figure 22  Grid for Proposed using VoIP.

Figure 23  Grid for Proposed using VoD.
better than one-channel application, 3) as the number of hops in the chain increases the data contention grows, which at the end causes throughput degradation. 1) The throughput was increased proportionally as the efficiency of MCS increased. We had originally designated the modulation scheme for our simulation as QPSK-1/2 throughput (12.19 Mbps), 16-QAM-1/2 throughput (24.69 Mbps), and 64-QAM throughput (55.78 Mbps), thus the order of MCS efficiency was; QPSK-1/2 throughput < 16-QAM-1/2 throughput < 64-QAM throughput. This order was preserved in our simulation results where the throughput rate was highest in the 64-QAM, the second highest was 16-QAM, and the lowest was QPSK-1/2, regardless of the topology, proposed/traditional method, or type of traffic as expected in the MCS. 2) Two-channel application performed better than one-channel application. The relationship between the number of channels and the throughput rate is not simple multiplication, i.e. the use of two-channel does not double the throughput rate. Nonetheless, there was significant overall performance improvement in the two-channel application. The throughput rate difference between the one- and two-channel applications was rather small. However, every single data point in the series of our simulation showed that the two-channel throughput rate was higher than that of the one-channel. Even in the Chain/VoD for both traditional (Figure 18) and proposed methods (Figure 21), where it appears as if there were no effects of increased number of channels at the beginning (Hop number 2) in either QPSK, 16-QAM, or 64-QAM nodes, there was actually a slight increase in the use of two channels in both the traditional and proposed method (the two-channel effects were more pronounced in the later.) In the traditional method, with two-channel application for QPKS, 16-QAM, and 64-QAM at the Hop number 2 showed 0.09%, 0.22%, and 0.02% increases, respectively (increased by 10 KB, 40 KB, and 5 KB, respectively) comparing to the one-channel application. These two-channel effects at small scale in the beginning of traffic flow grew larger at the Hop number 10 for QPKS, 16-QAM, and 64-QAM up to 11.61%, 17.28% and 22.39%, respectively (increased by 850 KB, 1650 KB, and 3225 KB, respectively). Likewise, in the case of the proposed method, with two-channel application for QPKS, 16-QAM, and 64-QAM at the Hop number 2 showed 1.22%, 2.34%, and 0.84% increases, respectively (increased by 200 KB, 550 KB, and 300 KB, respectively) while at the Hop number 10 the magnitude of two-channel effect increased up to 6.90%, 10.20% and 16.61%, respectively (increased by 850 KB, 1500 KB, 3225 KB, respectively) compared to the one-channel application. The throughput degradation was exacerbated as the number of hops in the chain increased. This observation could be explained as following. It is well known problem in Mesh Network that the performance degrades as all flows match their throughput to those with the lowest quality channel. Furthermore, in multi-hop wireless networks, contention of neighborhood nodes and variable rate channels are contribution factors of unfairness. In (Gupta et al., 2000), it is demonstrated that the average throughput capacity per node of a wireless multi-hop network decreases as $\Theta(1/n)$, where $n$ is the number of nodes in the network. This phenomenon has been observed in the context of IEEE 802.11 WMNs as well (Gambiroza et al., 2004). This is due to fraction of the channel capacity which is used to relay packets at intermediate nodes as explained in (Cicconetti C., Akyildiz et al., 2007; Cicconetti et al., 2009).

While the all cases (regardless of the topology, proposed/traditional method, or type of traffic) showed the same trend (i.e. increased hop number with declined performance), the magnitude of performance degradation, however, differed between the proposed and traditional method. Our proposed method showed less of degradation in performance (Figure 17, 21, 22, and 23) in comparison with the respective results with traditional method (Figure 16, 18, 19, and 20), where the goal of our study is to maximize the throughput between nodes.

Taking the single-channel/ node/Grid/VoIP as an example (Figure 19 and 22), the overall average throughput rate was 3.8 Mbps for the traditional method (Figure 19), while that for the proposed method was 8.8 Mbps (Figure 22), which is 2.3-fold increase. Even in 64-QAM 3/4 node, where the efficiency of MCS was set higher to begin with, there was significant 28% overall average throughput increase in the proposed method (Figure 21) with single-channel/Chain/VoD in comparison to its counter part (Figure 18). Furthermore, every single data point in the series of our simulation showed that the proposed adaptive rate method had higher throughput rate than the traditional sending packets method, regardless of the topology, type of traffic, or single-/two-channel application.

### 6.2 Uplink and downlink

Figures 24, 25, 26 and 27 represent the simulation results using the uplink and downlink topology. The graphs listed here only depict the results from QPKS modulation scheme as a representative. We had chosen this data set since 16-QAM and 64-QAM had data patterns similar to those of QPKS, except that the scale of the throughput rate was highest in the 64-QAM, the second highest in the 16-QAM, and the lowest in the QPSK, regardless of the proposed/traditional method, type of traffic, or use of one- or two-channel (data not shown).

In this scenario, the traffic flows both in the uplink and downlink directions simultaneously, which resulted in the increased chance of congestion. Consequently, the overall throughput rate of the Uplink and downlink topology was substantially lower than that of the Chain or Grid topologies.

In the graphs, it appears that the downlink delays are less than that of uplink with either the traditional
or our proposed approach. This phenomenon is simply due to the current simulation design and should not be interpreted that the performance in downlink direction is always better. The delay in the uplink direction could be explained by the heavily loaded centre node relative to the other nodes in the network. When the network becomes overloaded, the links between the centre node and its neighbors becomes bottlenecks and slows down the overall traffic flow. In the current simulation setup, it was so designed that the uplink direction suffers more from the bottleneck effect.

Types of traffic tested are VoIP (Figures 24 and 26) or VoD (Figures 25 and 27). The graphs listed side by side were tested in the same conditions except for the number of the channels. The graphs on the left side (Figures 24 and 25) are the results of one-channel application, while the ones on the right (Figures 26 and 26) employed two-channel application. Similarly to the results from the Chain and Grid topologies, the use of two-channels significantly improved the performance in the Up- and Down-link topology regardless of the type of traffic (VoIP/VoD) or method used (Traditional/Proposed). The magnitude of throughput rate enhancement by the two-channel effect was in fact, more pronounced in this topology than in others, where the increase in the overall average throughput rate by the two-channel effect was ranged between 1.4- to 1.7-fold.

Traditional method versus our proposed method were compared in the same graphs, where circle, square, diamond, and triangle represent the data point of downlink Traditional method, Uplink Traditional method, downlink Proposed method, and uplink Proposed method, respectively. Similarly to the other topologies tested in this study, the performance was improved significantly by our proposed approach for both VoIP and VoD traffics. For example, overall average throughput rate in the downlink traffic for VoIP/one-channel with the proposed method was 6.2 Mbps which is 2-fold higher than that with the traditional method. Even in the uplink traffic for VoD/two-channel, our
proposed approach resulted in the overall average throughput of 9.5 Mbps, which is 1.2-fold improvement of performance in comparison to the result of the traditional method. Furthermore, every single data point in the series of simulation for this topology indicated that our proposed method had the higher throughput rate than the traditional approach.

7 Random Topologies

In the next set of simulations, we examined the steady-state throughput of our proposed method. The simulated networks covered an area of 1500 x 1500 square meters, where 40 nodes were placed randomly. Five random connections were set up in each scenario that continuously try to deliver as much data as possible. All experiments were simulated with TCP connections as competing traffic since majority of traffic over the internet is TCP, and TCP has an adjustable sending rate for the proposed method to be evaluated against.

Figures 28 - 31 show the results of the measurements for the five random network setups. In these figures, each bar represents throughput rate of a topology, while each segment of the bar stands for the throughput of one stream. An identical colour fill pattern indicates that the respective segment belongs to the same pair of communicating nodes. The chosen representation thus allows not only a comparison of the total throughput, but also of its distribution to the five streams.

Figures 28 and 30 represent the results of traditional approach, while Figures 29 and 31 are those of our proposed approach. The VoIP throughput rates of our proposed method (Figure 29) from Topology 1 through 5 were 21.0 Mbps, 22.0 Mbps, 21.4 Mbps, 18.5 Mbps, and 20.7 Mbps, respectively. Similarly, the proposed method in the video traffic (Figure 31) had throughput rate of 26.6 Mbps, 32.8 Mbps, 27.6 Mbps, 23.0 Mbps, and 22.8 Mbps (from Topology 1 though 5), respectively, whereas the traditional method barely reached to 14.8 Mbps, 10.6 Mbps, 12.9 Mbps, 9.8 Mbps, and 8.6 Mbps, respectively. Again, the currently proposed adaptive scheduling for bursty traffic clearly demonstrated its efficiency in enhancing throughput.

Our primary motivation of this study is to achieve better throughput rate and fair bandwidth allocation in WiMAX WMN for multimedia. In wireless networks, severe unfairness has been a persistent problem as has been reported in the literatures on IEEE 802.11 (Gambiroza et al., 2004; Gupta et al., 2000). This problem in wireless may be traced back to the medium capture problems which also induce throughput degradation. Unfortunately, WiMAX has yet to fully overcome these problems.

In order to address fairness issues, WiMAX WMN uses the coordination mode which is distributed. In the distributed mode, transmissions are scheduled in a fully distributed fashion which requires no interaction with the BS. Nonetheless, it is not sufficient to support guaranteed traffic flows over multiple hops in WiMAX Mesh networks since the WiMAX WMN MAC is connectionless-based, where multiple links can interfere with each other when they are scheduled at the same time.

Unlike the traditional static bandwidth allocation algorithms, which do not take into consideration for the burstyness of multimedia, our currently proposed method dynamically changes scheduling for multimedia traffic. Therefore, fair bandwidth allocation can be achieved without suffering throughput rate.

The results clearly demonstrated the impact of the adaptive scheduling on fair bandwidth allocation in
multimedia traffic. For the traditional method with VoIP traffic (Figure 28), the Coefficient of Variations (CV) among the flow 1 through 5 were calculated as 34.7%, 22.7%, 60.0%, 31.9%, and 42.3% from Topology 1 through 5, respectively, which signify high fluctuation of bandwidth. On the other hand, the CV of our proposed method in VoIP (Figure 29) were 23.8%, 21.5%, 19.2%, 35.4%, and 18.1% from Topology 1 through 5, respectively, which are less fluctuated as depicted in the graph as evenly distributed segments of colour fill in each bar. Likewise, in VoD traffic, traditional method (Figure 31) showed larger fluctuations with CV of 55.7%, 47.1%, 65.5%, 78.0%, and 64.1% from Topology 1 through 5, respectively, whereas those of our proposed method are 15.8%, 23.2%, 30.3%, 24.8%, and 26.9% from Topology 1 through 5, respectively, and indicates much smaller fluctuation in bandwidth distributions.

8 Conclusion

Multi-channel WMNs have been developed to perform ubiquitous wireless broadband network access. In this study, we have proposed a dynamic and adaptive distributing scheduling method for multimedia in multi-channel WiMAX WMN. The aim of the proposed method is to support guarantees for QoS over multi-hop traffic, i.e. providing fair bandwidth allocation while improving the traffic performance (throughput rate) for multimedia traffic, which is bursty in nature. The key feature of currently proposed scheduling method is that higher priority was given to multimedia traffic by adding predicted backlog traffic to the number of the packets requesting to be sent.

The simulation results demonstrated that: 1) Regardless of the simple/complex topologies tested in the study (Chain, Grid, Up/Down-link, Random) the proposed method had higher performance in terms of the throughput rate. It was true for both the voice and video traffics. 2) With our proposed method, fair bandwidth allocation was achieved without sacrificing throughput rate for both the voice and video traffics. The challenge for WiMAX Mesh networks (multiple hops) is to support guaranteed traffic flow since WiMAX WMN MAC is connectionless-based. Our proposed method improved the unfairness issues by initiating QoS scheduling techniques. 3) Additionally, we have tested the effects of the use of multi-channels in WiMAX WMN. Regardless of the simple/complex topologies tested, types of multimedia traffic, or scheduling methods applied, the multi-channel application had somewhat higher throughput rate compared to the single-channel, although the magnitude of improvement was at smaller scale.

In future work, we will further explore the potentials of WMNs in the following areas. Firstly, we will extend our research to the energy performance using this proposed model. We are seeking for a better understanding of the complex interactions between the energy requirements and techniques for improving throughput in multimedia traffic. As energy consumption in the computing is a great concern in relation to the global environmental issues, it is urgent to develop technologies for energy efficient networking for multimedia. Secondly, we will also investigate the performance of an adaptive sending rate in relation to end-to-end Quality of Service. Finally, we will examine the effects of usage of multiple channels along with a single fixed channel. These techniques could further improve network performance and increase network capacity of adaptive sending rate for WMNs.

Acknowledgements

We thank the anonymous reviewers for their constructive suggestions that helped us to improve the quality of
the paper. This work was partially supported by the Excellence Center at Linköping - Lund in information Technology (ELLIT) of Sweden.

References


NS2 [Online], http://www.isi.edu/nsnam/ns/index.html.


