An efficient Bandwidth Demand Estimation for Delay Reduction in IEEE 802.16j MMR WiMAX Networks

Fath Elrahman Ismael fathi_sml@yahoo.com

UTM MIMOS CoE Telecommunication Technology,
Faculty of Electrical Engineering (FKE)
Universiti Teknologi Malaysia (UTM)
Johor, 81300, Malaysia

Sharifah K. Syed Yusof kamilah@fke.utm.my

UTM MIMOS CoE Telecommunication Technology,
Faculty of Electrical Engineering (FKE)
Universiti Teknologi Malaysia (UTM)
Johor, 81300, Malaysia

Norsheila Fisal Sheila@fke.utm.my

UTM MIMOS CoE Telecommunication Technology,
Faculty of Electrical Engineering (FKE)
Universiti Teknologi Malaysia (UTM)
Johor, 81300, Malaysia

Abstract

IEEE 802.16j MMR WiMAX networks allow the number of hops between the user and the MMR-BS to be more than two hops. The standard bandwidth request procedure in WiMAX network introduces much delay to the user data and acknowledgement of the TCP packet that affects the performance and throughput of the network. In this paper, we propose a new scheduling scheme to reduce the bandwidth request delay in MMR networks. In this scheme, the MMR-BS allocates bandwidth to its direct subordinate RSs without bandwidth request using Grey prediction algorithm to estimate the required bandwidth of each of its subordinate RS. Using this architecture, the access RS can allocate its subordinate MSs the required bandwidth without notification to the MMR-BS. Our scheduling architecture with efficient bandwidth demand estimation able to reduce delay significantly.

Keywords – IEEE 802.16j, Grey prediction, Delay, throughput, wireless networks.

1. INTRODUCTION

The IEEE 802.16 WiMAX is a standard for the wireless broadband access networks and is recognized as one of the strongest contenders of the wired networks to provide internet connection to the end users. However, the existing WiMAX products have limited coverage and provide poor signal strength for indoor users as well as users at cell boundaries [1].

Mobile Multi-hop Relay (MMR) WiMAX systems have the potential to offer improved coverage and capacity over single-hop radio access systems. Standards development organizations are considering how to incorporate such techniques into new standards. One such initiative is the

IEEE 802.16j standardization activity, adding relay capabilities to IEEE 802.16e systems [2]. The Relay Station (RS) significantly reduces the installation and operation cost [3]. Moreover, the RS increases the overall network capacity as a result of enhancing CINR in a poor coverage area.

Transmission Control Protocol (TCP) is the most commonly used transport protocol on the Internet [16]. But conventional TCP requires acknowledgement from the receiver side to know that the sent packet is received correctly or not.

The bandwidth request- grant mechanism in the standard WiMAX when applied to multi- hop systems affects the performance of the TCP traffic. It introduces much delay to the transmitted data in the uplink and also to the TCP acknowledgement, which increases the Round Trip Time (RTT) and leads to unnecessary retransmission and reduces the throughput.

The paper is organized as follows. Section 2 introduces general overview on MMR network architecture. In 3 scheduling algorithm of WiMAX networks are discussed. The Problem definition is stated in 4. The literature and related work is given in 5. The proposed scheduling architecture is explained in 6, the grey prediction algorithm is discussed in 7, the performance of the proposed scheme is evaluated in 8, and the conclusion in 9.

2. Mmr networks architecture

The MMR networks consist of three types of station: the MMR-BS, a fixed (RS), and the (MS) as shown in FIGURE 1. The MS can communicate with a MMR-BS either directly or over a two-hop via the RS, and also more than two hops can be exploited via multiple RSs. The direct link between the MS and the MMR-BS or RS is referred to as an access link, while the backhaul link between the RS and the MMR-BS referred to as a relay link.

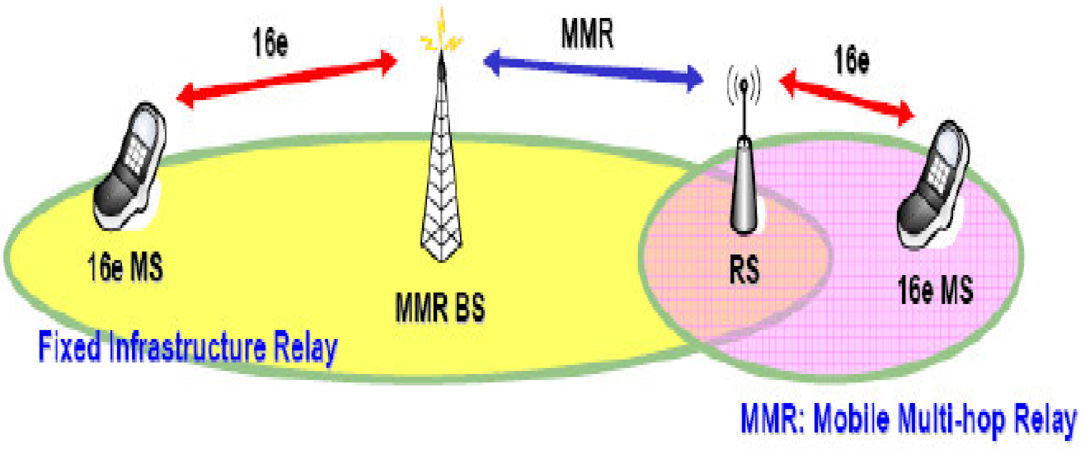

FIGURE 1: MMR network architecture

In WiMAX OFDMA the most common duplexing mode is frame structure TDD, in which the frame is divided into downlink and uplink sub frame that are separated by a guard interval. At the beginning of each Downlink (DL) frame there are several symbols that used for frame control as shown in FIGURE 2. The Preamble enables MS to synchronize to the MMR-BS. DL MAP maps each MS the location of its pertinent burst in the DL sub frame. UL MAP that define for the MS the burst location and size in the coming UL sub frame. Frame Check Header (FCH) which is a special management burst that is used for the MMR-BS to advertize the system configuration The relay operation is classified in two different modes: transparent and non-transparent relay (NTR): The transparent relay is an element that is located in the network between the MMR-BS and the MS, without the MS being aware of its existence. It does not transmit any control data. The MS receives directly all the control data such as preamble, FCH, DL MAP and UL MAP directly from the MMR-BS. The transparent relay usually does not extend coverage; its main function is channel quality improvement that as a result increases the link capacity [1] [2].

NTR is a network element that the MS is totally aware of it. The MS which is IEEE 802.16e compatible is not aware of any relaying operation, from the MS point of view the NTR is its conventional serving BS. So, the NTR should support most of the capabilities of a plain 802.16e BS like sending frame control data that includes preamble, UL/DL MAP, FCH. Furthermore, the NTR supports handover, network entry and MS ranging. The NTR serves MSs that are beyond reach of the MMR-BS. As so the NTR is perfect tool for achieving extension of cell and coverage. In addition it can also act as a capacity enhancer because of significant improvement in the channel quality of both access and relay links [1] [2].

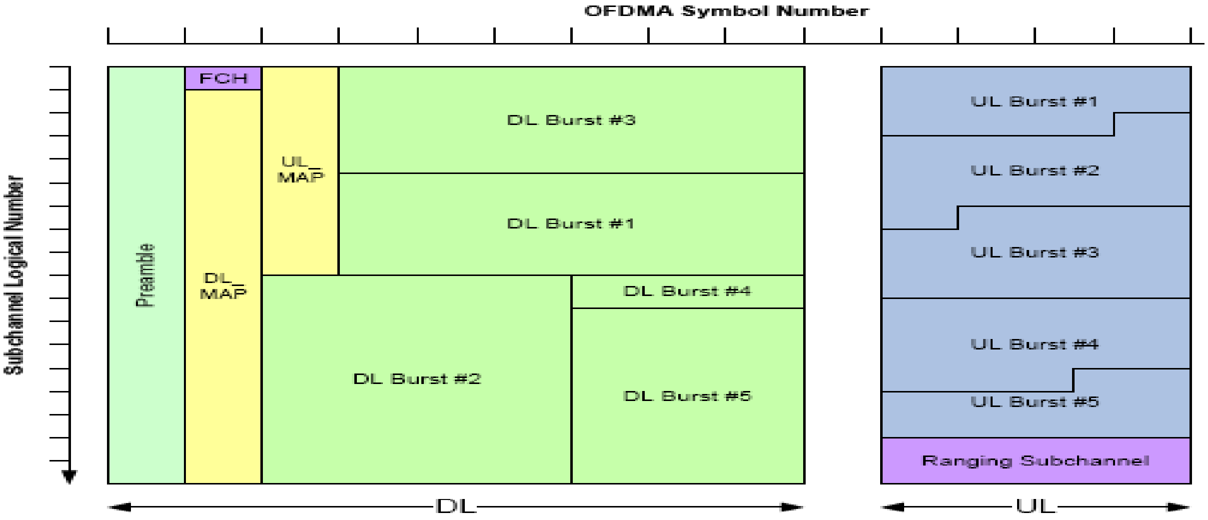

FIGURE 2: OFDMA frame structure

3. Scheduling algorithms of wimax

Scheduling is a sequence of allocating fixed-length time slots to users, where each possible transmission is assigned a time slot. The scheduling challenges are how to develop scheduling architecture that overcomes the problems of increased delay and signaling overhead, and design scheduling algorithm that guarantees the QoS requirements of each service flow [13].

All transparent relays must (and NTRs may) operate in centralized scheduling mode. In this option the MMR-BS is doing all scheduling tasks as it is defined in IEEE 802.16e. It supports optimal centralized radio resource management (RRM), and enables network load balancing among the subordinate relays. However, it requires heavy management traffic within the network, and it is slow in granting bandwidth, which make it not feasible in heavy multi-hop network [4][5].

In distributed scheduling, some MAC intelligence can be given to the RSs. IEEE 802.16j allows NTR relays to operate in distributed scheduling mode, where they make decisions about resource allocation to their subordinate stations most always in coordination with the MMR-BS. In distributed scheduling there is no heavy management traffic within the network and it allows fast bandwidth request grants for high priority requests, which it a suitable technique in heavy multi-hop architecture. But it does not give option for load balancing between relays, and there is no option for centralized RRM [4] [5].

Five scheduling services are defines in the IEEE 802.16j, these are Unsolicited Grant Service (UGS), Real Time Polling Service (rtps), Extended Real Time Polling Service (ertps), Non Real

Time polling Service (nrtps), and Best Effort (BE). The selected network service should allocate sufficient amount of bandwidth in order to actually guarantee network QoS provisioning [17].

4. Problem definition

In WiMAX network when the MS wants to send data, it first sends a bandwidth request message to the corresponding base station (BS), and then the BS grants the appropriate amount of bandwidth to the MS based on the uplink scheduling algorithm. In MMR networks, the control messages used for network entry or bandwidth request are doubled for every RS introduced between MS and MMR-BS, as shown in FIGURE 3. The main reason of the increasing overhead and latency is because all service flow management decisions should be made by MMR-BS in spite of centralized or distributed scheduling. In the centralized scheduling all the functions of network entry and bandwidth request are performed by the MMR-BS. The signaling messages received by the RS from the MSs are forwarded to the MMR-BS, which decides and sends a response messages to the MSs through the intermediate RSs, this process increases the signaling messages exchanged and the delay. Although in the distributed scheduling the RS have an ability to respond to the bandwidth request from their subordinate stations, the relay station requires also a bandwidth request to relay received data to the MMR-BS, which also increase the signaling overhead and delay. Considering that, a signaling message from one hop to another cannot be transmitted in one frame's duration, the delay of control messages transfer from MMR-BS to the MS will increase significantly due to the multi-hop added.

The increased delay affects the performance of the services offered by the MMR WiMAX, because in UGS, rtps, and ertps services the lower latency is required as a QoS metric. Furthermore, even for the BE using TCP as transport protocol, the increased delay increases Round Trip Time (RTT) which degrades the throughput.

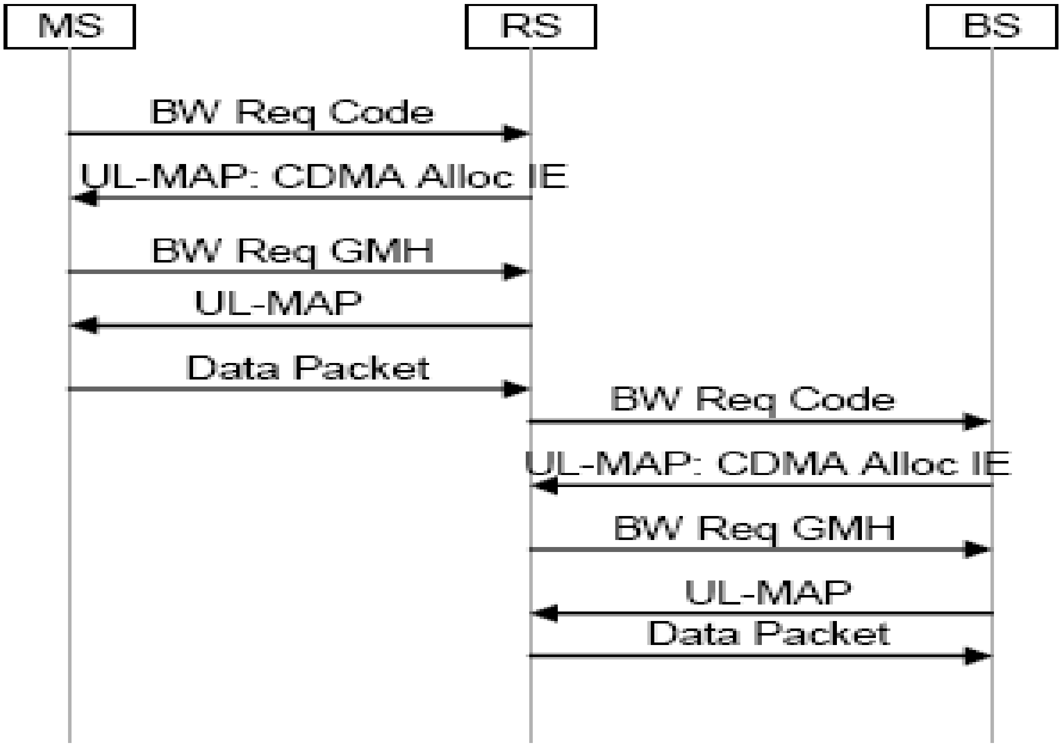

FIGURE 3: Bandwidth request procedure

5. Related works

In the literature there is some works done to reduce the signaling overhead and delay in one hop WiMAX networks such as IEEE 802.16e.

The authors of [8, 9] suggest bandwidth grant mechanism without bandwidth request to fully utilize the available bandwidth for the BE traffic. The BE traffic has a lower priority among services supported by WiMAX; this will result in small amount of bandwidth for this type of service. In addition the BE uses contention based bandwidth request, in which BE connections contend limited number of shared slots for sending their request, and if there is more than one SS attempts to send their requests via a slot, the requested are corrupted.

In this scheme, the BS allocate bandwidth to the SSs without bandwidth request, the amount of granted bandwidth is determined from the current sending rate, this scheme is suitable for large scale network. On the other hand, it may waste the bandwidth when the sending rate is highly fluctuated, and it’s hard to estimate the required bandwidth accurately.

In the downlink TCP flow, the standard bandwidth request- grant mechanism in single hop networks introduces much delay to transmit uplink TCP acknowledgement, which increase the Round Trip Time (RTT) and causes unnecessary retransmission that reduce the TCP traffic performance. The authors of [11] introduce new bandwidth scheme that simultaneously allocate bandwidth for both downlink TCP data and uplink TCP acknowledgement.

In the uplink TCP traffic, since the BS controls the uplink transmission and the MS needs to get permission before transmission, these steps severely limit the performance of the TCP by introducing much delay and degrade the throughput. The authors of [12] propose scheme allocating uplink bandwidth in advance. When the BS receive acknowledgement to be sent to the MS, it relays the TCP ACK and at the same time allocate bandwidth for the uplink TCP data packet without bandwidth request from the SS. These schemes for bidirectional bandwidth allocation require more information about the actual SS needs to not waste the bandwidth.

For MMR networks there are a little work done to reduce the signaling overhead and delay. In [6] MSs send their bandwidth request to their super ordinate RS at their respective polling intervals. The RS collects all bandwidth requests from MSs and generates an aggregated bandwidth request to the MMR-BS. The MMR-BS, instead of allocating the bandwidth to the MSs directly, it grants bandwidth to the intermediate RS, which then allocates bandwidth to individual MSs. This scheme decreases the signaling overhead, but not decreases the delay.

The authors of [7], has defined an operation called concatenation, whereby multiple packet data units (MPDUs) can be concatenated into a single transmission burst in either uplink or downlink directions, regardless whether these MPDUs belong to the same connection or not. Also multiple service data units (MSDUs) belong to the same connection can be aggregated to form one MPDU. These schemes reduce the signaling overhead but there is still much delay.

In [10], the authors propose Extra Resource Reservation (ERR) scheme in which the MMR-BS pre-allocate some resources to the RSs for purpose of transporting control messages or data traffic. So the RS can admit its MSs without MMR-BS notification. This scheme can reduce the delay but it wastes the channel resources if the ERR is used ineffectively in the RS.

Grey theory has been widely applied to many disciplines such as economics, sociology, engineering, and so forth. The authors of [14] use the grey prediction to estimate the demand value of the electricity on line. In [15] it is used for internet access forecasting.

6. The proposed scheduling algorithm

The IEEE 802.16j support tree architecture in which all the traffic go through the MMR-BS, so it should be managed in a centralized manner. But to reduce the signaling overhead and delay its better to distribute the management of the bandwidth request in the access RS. So, our scheme manages the relay links centrally in the MMR-BS, and the bandwidth request for the end users distributed in the RS to take the advantages of both scheduling architectures.

The aims of the proposed scheduling scheme are to fully utilize the available bandwidth, and effectively respond to the demand changes of the subordinate Relay Stations.

In the proposed scheduling architecture in FIGURE 4, the MMR-BS manages its subordinate RSs as one hop stations. The same procedure is applied in the intermediate RS in the case of more than two hops, and in the access RS the conventional bandwidth request can be deployed.

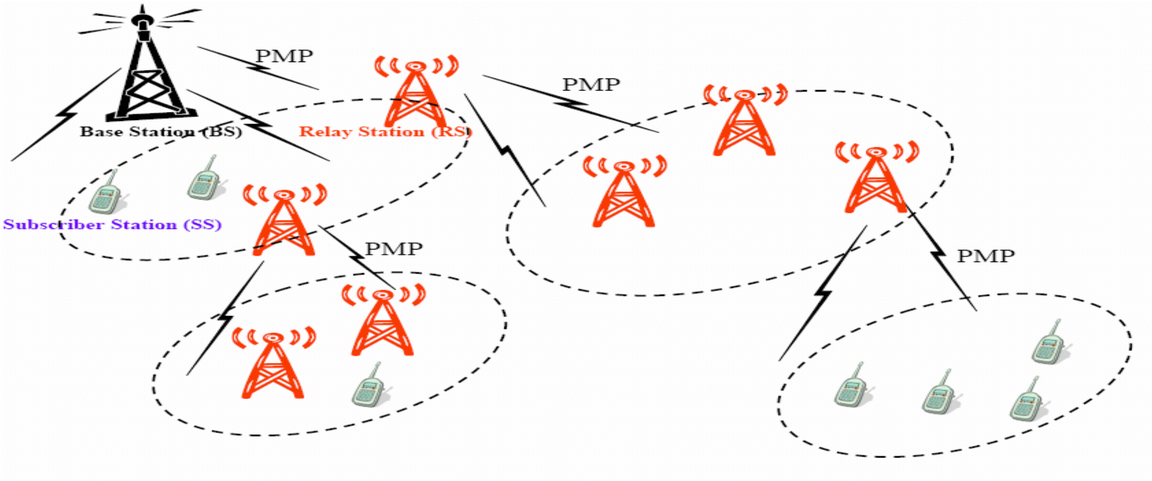

FIGURE 4: Proposed Scheduling Architecture

In this scheme the bandwidth request is not needed, instead MMR-BS measures the existing sending rate of its subordinate RSs, estimates the future demand of each RS using Grey prediction algorithm and grants them a suitable bandwidth without request.

Grey prediction theory [14] aims to find the optimized system parameters of a Grey differential equation such that the dynamic behavior of the traffic demand could be best fitted with the differential equation.

One of the major advantages of Grey prediction theory is that only small amounts of data are needed to describe the demand behavior and reveal the continuous changing in the required bandwidth.

7. The grey prediction algorithm

The grey prediction algorithm is used in the proposed scheme to estimate the required bandwidth of each RS from the previous values. In this paper, the GM (1, 1) model is adopted to perform the prediction of the required bandwidth for each RS. The standard procedure is as follows:

Step 1: Collecting the original data sequence, those are at least four consequence values of the measured traffic for the RS that is used in the estimation process.

$$x^{(0)} = \left\{x^{(0)}(1), x^{(0)}(2), \ldots\ldots\ldots x^{(0)}(m)\right\} \qquad m \geq 4 \quad (1)$$

Step 2: Calculating an accumulated generation operation, AGO, on the original data sequence in order to diminish the effect of data uncertainty.

$$x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i) \qquad k = 1,2,......, m \qquad (2)$$

Step 3: Establishing Grey differential equation and then calculating its background values: First we define z (1) as the sequence obtained by the MEAN operation to x (1) as:

$$z^{(1)}(k) = \left[\alpha * x^{(1)}(k) + (1-\alpha)x^{(1)}(k-1)\right] \quad (3)$$

$$for \qquad k = 2,3,...m$$

Secondly Grey differential equation can be obtained the as follow:

$$x^{(0)}(k) + az^{(1)}(k) = b \qquad (4)$$

Where the parameters a, b are called the development coefficient and the grey input, respectively. The equation (5) is called the whitening equation corresponding to the grey differential equation.

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b \qquad (5)$$

Step 4: calculate the value of a, b by means of the least square method.

$$\begin{bmatrix} a \\ b \end{bmatrix} = (B^T B)^{-1} B^T x^N \qquad (6)$$

Where

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ . & \\ . & \\ . & \\ -z^{(1)}(m) & 1 \end{bmatrix} \qquad (7)$$

And $x^N = \left[x^{(0)}(2)\, x^{(0)}(3) ...\, x^{(0)}(m)\right]^T \qquad (8)$

Step 5: Deriving the solution to the Grey difference equation:

$$\hat{x}^{(1)}(k+1) = \left[x^{(0)}(1) - \frac{b}{a}\right]e^{-a(k)} + \frac{b}{a} \qquad (9)$$

Where the parameter (k) is the forecasting step size and the up script “^” means the value $\hat{x}$ is a forecasting value of traffic demand.

Step 6: conducting the inverse accumulated generation operation (IAGO) on $\hat{x}^{(1)}$ to obtain a prediction value as follow:

$$\hat{x}^{(0)}(k) = \hat{x}^{(1)}(k) - \hat{x}^{(1)}(k-1) \qquad (10)$$

$$K = 2,3,......, n$$

Combining the equations (9) and (10) gives the predicted value of the traffic demand as stated in (11)

$$\hat{x}^{(0)}(k) = (1-e^{a})\left[x^{(0)}(1) - \frac{b}{a}\right]e^{-a(k-1)} \qquad (11)$$

Step 7: The parameters are changed and the accuracy of the estimated value is calculated using the formula:

$$e(k) = \frac{\hat{x}_i^{(0)}(k) - x_i^{(0)}(k)}{\hat{x}_i^{(0)}(k)} * 100\ \% \qquad (12)$$

$$K <= n$$

The average error percentage can be calculated as below:

$$e = \frac{1}{(n-m)} * \sum_{k=4}^{n-1} \left\| e(k+1) \right\| * 100\ \% \qquad (13)$$

$$m = 4,5 or\ 6$$

The accuracy of the traffic demand prediction can be given by:

$$\varepsilon = (1 - e) * 100\ \% \qquad (14)$$

8. Performance evaluation

In the Grey prediction, the parameter settings are highly relevant to accuracy of prediction. To get more accurate predicted value, there are three parameters should be optimized, these are: the number of modeling points , those are the previous values used in the estimation (m), the parameter α used to construct the matrix z in equation (3), and the time difference between successive estimations or the number of estimated values in a given time (n).

To evaluate the Grey prediction algorithm, we write a MATLAB program, in which the traffic sequence is generated using random function that generates traffic between 10 Kbps and 40 Kbps. Initially we begin with minimum values, in this case it's taken as 10 Kbps for all modeling points, then after each cycle we shift the old values left and insert the actual measured traffic value, and so after m cycles all the modeling points should be replaced by the actual traffic values.

In the simulation program we use the equations from (1) to (11) to estimate the next traffic value and then we used the equations from (12) to (14) to calculate the accuracy and the error percentage

We changed the time difference between the successive samples, the modeling points, and the parameter α to select the suitable values of these parameters that give best estimation accuracy.

Our experimental results shows that the optimal parameters settings of the Grey prediction are that it is better to take the sampling time as short as possible or n maximum, the Modeling points maximum (m=6), and the Parameter α=0.5. The selection of this parameters give an accuracy of (ε=97.8992 %) and error of (e=2.1008 %).

The optimum parameters of Grey prediction are chosen to give the highest accuracy and then we plot the FIGURE 5 for the increasing traffic and the decreasing one.

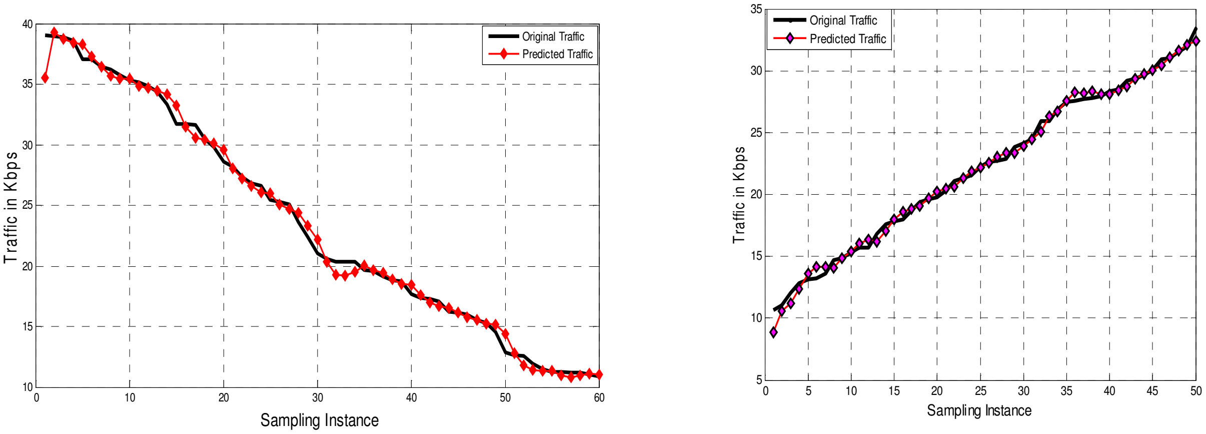

FIGURE 5: Predicted RS traffic using Grey prediction algorithm

The gained prediction accuracy makes the deployment of our scheduling algorithm to overcome the effects of the delay in MMR networks applicable. To evaluate the performance of the proposed scheduling algorithm, we assume that the RS uses the next frame to relay the data received in the current frame and there are no errors in the transmission, and the frame duration is τ =10 ms, and the number of hops is n<=10. The delay associated a data packet a cross n hops with a conventional bandwidth request scheme given by:

$$D = 4 * \tau * n \qquad (15)$$

The delay for n hops using our proposed scheduling architecture can be given by:

$$D = 4 * \tau + (n - 1) * \tau \qquad (16)$$

FIGURE 6 indicates that, as the number of hops increased the delay rapidly increased in the conventional bandwidth allocation (CBR) scheme, but in our enhanced bandwidth allocation (EBA) scheme it increased slightly. This figure shows that our algorithm can be more useful to reduce the signaling overhead and the delay when the number of hops is high.

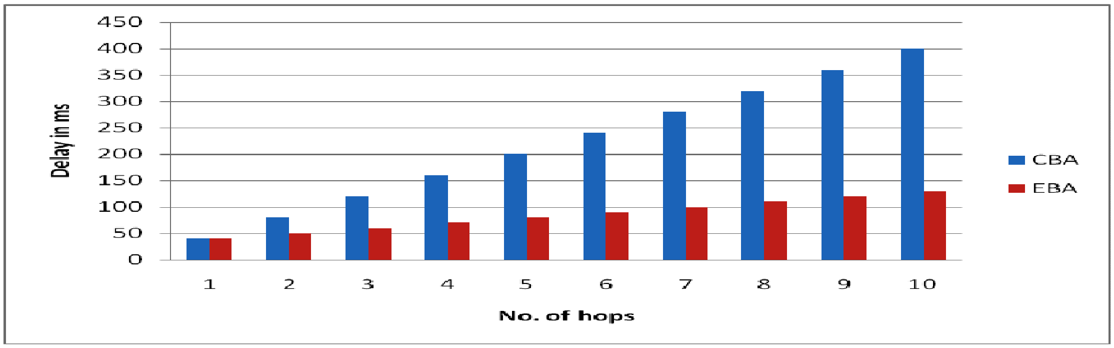

FIGURE 6: Delay in MMR network

9. Conclusion

In this paper we have presented an efficient bandwidth demand estimation to estimate the required bandwidth of the RS in IEEE 802.16j using the Grey prediction algorithm. The high accuracy offered by Grey prediction algorithm make the use of hybrid centralized and distributed scheduling architecture applicable solution of the increased signaling overhead and delay in IEEE 802.16j MMR WiMAX networks.

In this scheme, the MMR-BS allocates bandwidth to its direct subordinate station without bandwidth request. Using this scheduling architecture, each RS can admit its subordinate SSs the required bandwidth without notification to the MMR-BS. Our scheme can reduce the delay in MMR WiMAX networks significantly.

10. References

1. Steven W. Peters and Robert W. Heath, “The future of WiMAX: Multihop relaying with IEEE 802.16j”, IEEE communication magazine, January 2009.
2. Vasken Genc, Sean Murphy, Yang Yu, and John Murphy, “IEEE 802.16j Relay Based Wireless Access Networks: An overview”, IEEE wireless communications, October 2008.
3. Bharathi Upase, and Mythri Hunukumbure, “Dimensioning and cost analysis of Multihop relay enabled WiMAX networks”, FUJITSU Sci. Tech. J., 44, 3, p.303-317(July 2008).
4. Masato Okuda, Chenxi Zhu, and Dorin Viorel, “Multihop relay extension for WiMAX Networks overview and benefits of IEEE 802.16j standard”, Fujtsu Sci. Tech. J., 44, 3, p.292-302(July 2008).
5. REWIND project, " D4.1: Summary of Network Architecture Analysis and Selected Network Architecture”, December 2008.
6. Chun Nie, Thanasis Korakis, and Shivendra Panwar, “ A Multihop polling service with bandwidth request aggregation in IEEE 802.16j networks”, 978-1- 4244-1645-5/08/2008 IEEE.
7. Zhifeng Tao, Koon Hoo Teo, and Jinyun Zhang,” Aggregation and Concatenation in IEEE 802.16j Mobile Multihop Relay (MMR) networks”, 1-4244-0957-8/07, 2007 IEEE.
8. Seungwoon Kim, Minwook Lee, and Ikjun Yeom, “ Impact of bandwidth request schemes for Best Effort traffic in IEEE 802.16 networks”, Science Direct, Computer Communications 32 (2009) 235-245.
9. Seungwoon Kim, and Ikjun Yeom, “TCP aware uplink scheduling for IEEE 802.16”, IEEE communication letters (February) (2) (2007) 146-148.
10. Junkai Zhang, Suili Feng, We Ye, and Hongcheng Zhung,” Reducing signaling overhead and latency of 802.16j service flow management”, 978-1-4244-2108-4/08, 2008 IEEE
11. Eun Chan Park, Jae Young Kim, Hwangnam Kim, and Han Seok Kim, “Bidirectional bandwidth allocation for TCP performance enhancement in IEEE 802.16 broadband wireless access networks”, WiBro system Lab, Telecommunication and networking devision, Samsung Electronics Co., LTD, Korea.
12. Noureddine Lasfar, Jeonghoon Mo, and Byongkwon Moon, “ TCP ACK Triggered bandwidth request scheme in IEEE 802.16e systems”, School of Engineering, Information and Communications University, Korea.
13. Carlos G. Bilich, “TCP over WiMAX Networks”, Wireless Access networks project number II, University of Trento, Trento, TN 38100 italy.
14. Albert W.L. Yao, S.C. Chi, “Analysis and design of a Taguchi–Grey based electricity demand predictor for energy management systems”, Elsevier, Energy Conversion and Management 45 (2004) 1205–1217.
15. Wann-Yih Wu, Shuo-Pei Chen, “A prediction method using the grey model GMC(1,n) combined with the grey relational analysis: a case study on Internet access population forecast”, Elsevier, Applied Mathematics and Computation 169 (2005) 198–217.
16. Sachin Shetty, Ying Tang & William Collani, “A Cross-Layer Packet Loss Identification Scheme to Improve TCP Veno Performance”, International Journal of Computer Networks (IJCN), Volume (1): Issue (1), November 2009.
17. Qiang Duan, Enyue Lu, “Network Service Description and Discovery for the Next Generation Internet”, International Journal of Computer Networks (IJCN), Volume (1): Issue (1), November 2009.

18. Syed S. Rizvi, Aasia Riasat, & Khaled M. Elleithy, “Deterministic Formulization of End-to-End Delay and Bandwidth Efficiency for Multicast Systems”, International Journal of Computer Networks (IJCN), Volume (1): Issue (1), November 2009.