Adaptive Scheduling Strategy for WiMAX Real-time Communication

Zhu Peng, Zhu Guangxi, Lin Hongzhi, Shi Haibin
Dept. of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan, 430074, China
E-mail: pengzhu2k@163.com, gxzhu@mail.hust.edu.cn, linhz@mail.hust.edu.cn, s29952@tom.com

Abstract

Defined by the IEEE 802.16 standard, WiMax system is also called broadband wireless metropolitan area network. It supports real time and non-real time services. There is no packet-scheduling algorithms defined, which was need for the requirement of quality of service (QoS) supporting for real-time services. Normally, the service station (SS) requests for bandwidth from the base station (BS), The BS then allocates the bandwidth to SS according to priority-based request. In this paper we propose an adaptive bandwidth scheduling strategy for real time polling services (rtPS), the SS estimate the arrival of rtPS packets prior to the arrival and requests the BS for bandwidth in advance, based on the SS's request and the actual bandwidth request by past. The analytical model and simulation show that, this adaptive algorithm provided better results with respect to the number of packets waiting at SS and average delay as compared to the widely accepted weighted scheduling algorithm and other adaptive bandwidth request mechanism.

Keywords: WiMax, Adaptive, QoS, Bandwidth scheduling

1. Introduction

Broadband wireless access system is becoming very popular, in which, WiMax, the broadband wireless metropolitan area network (WMAN), with open standard and support quality of service (QoS) for different categories of services. The BS of a single cell schedules the traffic flow. The communication between BS and SS are bidirectional, through downlink channel (from BS to SS) and uplink channel (from SS to BS). The downlink channel is in broadcast mode. The uplink channel is shared by various SSs through time division multiple access (TDMA). The subframes consist of a number of time slots. BS scheduler determines the duration of sub frames, and the number of slots [1].

WiMax supports all categories of services under real and non-real time communications. Real time caters to two types of traffic flow, unsolicited grant service (UGS) for TI/El type telecom voice and real-time-polling services (rtPS) for video streaming, teleconferencing as well as video conferencing. Similarly, the non-real time services are divided into two categories of traffic, i.e., non-real-time-polling services (nrtPS) to cater to file transfer and best effort (BE) services for non-delay sensitive communication, over Internet as e-mail, etc. The QoS is vendor specific and flexible for futuristic different services. WiMax provides only signaling mechanisms and standard for UGS's QoS. It does not specify the admission control and packet scheduling algorithms at the BS as well as SS for rtPS, nrtPS and BE traffic. Out of these three undefined services, rtPS scheduling is more important, because it will cater to the real-time traffic, whose packet delay is very sensitive. It has been left to the traffic-designer to employ a particular admission control and packet-scheduling algorithm for better performance of rtPS traffic [1-2].

There have been some studies on WiMax QoS [2-6] and its improvements depend on packet scheduling algorithms. GuoSong have discussed uplink and downlink packet schedulers for bandwidth allocation process and admission control for QoS [3]. Dong-Hoon has suggested changes in the MAC architecture to improve the throughput of services and have not discussed about delay [4]. Kitti presented packet arrival and QoS issues for multimedia system [5]. Howon have done throughput evaluation for real-time communication but have not discussed the effect on delay due to the algorithm [6]. In these packet-scheduling algorithms SS sends a bandwidth request to BS after arrival of packet at SS. Mukul presented an adaptive bandwidth request mechanism [9], but have not considered about the SS's bandwidth request and the queue length factor, so it can not get the best performance. In this paper, we proposed an effective adaptive bandwidth mechanism, which can decrease the delay and needs lesser buffer size, we have presented by analysis and simulation that, the proposed algorithm is better than the existing mechanism on WiMax, in terms of delay and buffer requirement at SS.

This paper is organized as follows. Section 2, presents the existing bandwidth request mechanism for rtPS and the proposed adaptive bandwidth request mechanism. In Section 3, mathematical analysis of the proposed solution is given. Section 4, the results of simulation and discussion, Section 5 contains the concluding remarks on contribution of this work as well as futuristic problems.

2. Current rtPS Scheduler

In 802.16 the rtPS bandwidth request scheduler at SS demands time slots for data transmission. In response to the request by SS the BS will provide some time slots equal to or less than the requested time slot [1]. The time slots requested at a particular time depend on the total data in the rtPS queue at that time. Typically if the request for time slot is made at time t_0, and request is granted at time t_l, then $t_l > t_0$. This means that there will be always delay incurred by the data packets because of request grant process. As the request was made at t_0, the BS will allocate bandwidth for rtPS traffic waiting till t_1. However, between t_0 & t_1 new rtPS data might have arrived since the time last request was made. For this new set of data, which have arrived between t_0 and t_1, the SS has to send fresh request for bandwidth allocation. This will incur extra delay. As rtPS services are very delay sensitive, we can alleviate these problems by scheduling an adaptive bandwidth allocation according to data flows in the rtPS queue. So we propose an rtPS bandwidth request model, which has following features.

The rtPS queue at SS will demand time slots from BS, based on amount of data departures to be facilitated. And it will be adaptive in nature i.e., a SS will request time slot not only for present data in the queue but also on the data which will arrive in the queue in between t_0 and t_1. As the data arrival is random, this needs a stochastic prediction. Here, we use the of "Differential time grant" method [9], to estimating average rate of incoming data and duration of above time interval. These calculations will be used to estimate amount of the data that has arrived in this time interval.

3. Analysis of Adaptive Scheduler

In this section, the analytical model shows below for our proposed adaptive algorithm of bandwidth allocation mechanism.

Here, we follow the strategy of [9] to estimate the cumulative data flow at SS, and the evaluation of data arrival rate at SS.

A(t), which means the cumulative data arrivals into the rtPS scheduler. The time slots for which rtPS queue output link will be ON-OFF is stochastic in nature. OFF time interval of rtPS output is (t_0, t_1), (t_2, t_3), (t_4, t_5)… (t_{2n-2}, t_{2n-1}), denoted as $Q_1, Q_2, Q_3 \ldots Q_n$, Similarly the ON time of the rtPS will be, (t_1, t_2), (t_3, t_4), (t_5, t_6) … (t_{2n-1}, t_{2n}), denoted as $P_1, P_2, P_3 \ldots P_n$. The data arrival rate is assumed to be 1Mbps. The maximum packet size is assumed to be 200 bytes. We assumed that the time origin of the system is at t_0. And we are assuming the system is a causal one. That is $A(t_{0-}) = D(t_{0-}) = C(t_{0-}) = 0$. Then using Network Calculus [8] model and Mukul's Equation [9] we can get the estimation of 'ON' time interval,

$$X(t) = \sup\,(F(t) - F(s) - C(t-s)) \quad \text{where } s \in (t_{2n-1}, t_{2n}) \tag{1}$$

$$D(t) = \inf\,(F(s) - C(t-s)) \quad \text{where } s \in (t_{2n-1}, t_{2n}) \tag{2}$$

Where, X(t) means the queue length in rtPS scheduler of SS. D(t), the cumulative data departure from rtPS queue. C means maximum transmission output rate of rtPS. F(t) defined as

$$F(t) = A(t) - D(\max(P_{n-1})) \tag{3}$$

This considers net data flown out of rtPS queue at SS in last 'ON' time. Estimation of 'OFF' time interval

$$X(t) = F(t) \tag{4}$$

$$D(t) = D(\max(P_{n-1})) \tag{5}$$

Let's define the estimation function first. Here, we use the LAGRANGE interpolating function to estimate the value of time width and the data arrival rate. For the Runge phenomenon, n should not very big, so, we let n=2, that is, the estimation equation shows below,

$$E(t) = \frac{(t-t_1)(t-t_2)}{(t_0-t_1)(t_0-t_2)}E_0 + \frac{(t-t_0)(t-t_2)}{(t_1-t_0)(t_1-t_2)}E_1 + \frac{(t-t_0)(t-t_1)}{(t_2-t_0)(t_2-t_1)}E_2 \tag{6}$$

A. Estimation of time width

The times at which request are made be $\alpha_1, \alpha_2, \alpha_3 \ldots \alpha_x$. The time at which the corresponding is fetched, called $\beta_i = t_{i-1} - \alpha_i$, estimation of β_i is to be made in term of $\beta_{i-1}, \beta_{i-2}, \beta_{i-3}, \ldots \beta_{i-x}$, called β_i'', let $E_0 = \beta_{i-1}, E_1 = \beta_{i-2}, E_2 = \beta_{i-3}$, we can get the estimate of β_i'' by equation (6).

B. Estimation of data arrival rate

Similarly, v_i, which gives the data arrival rate

$$v_i = \frac{(A(t_{i+1}) - A(\alpha_i))}{\beta_i} \tag{7}$$

Let $E_0 = v_{i-1}, E_1 = v_{i-2}, E_2 = v_{i-3}$, we can get the estimate of v_i, called $v_i^{''}$, by equation (6).

C. Calculation of adaptive time slot

The calculation for time slot will be estimate as,

$$x_{ei} = a_i * \frac{(X(\alpha_i) + v_i^{''} * \beta_i^{''})}{C} + (1 - a_i)x_{ri} \quad \text{where } 0 \le a_i \le 1 \tag{8}$$

Where, the variable x_{ri} means the request of time slot that SS send to BS according to the queue length, variable a_i means the proportion of estimation factor, the value of a_i should in (0, 1), the initial value of a_i set to a_i=0.5.

We define an evaluation function of estimate as

$$e_i = \begin{cases} \left| \dfrac{x_i - x_{ri}}{x_i - x_{ei}} \right| & , x_{ei} \ne x_i \\ 1 & , x_{ei} = x_i \end{cases} \tag{9}$$

Where, variable x_i means the actual time slot that allocated, so, if $e_i > 1$, means x_i is more approach to the estimate value x_{ei}, and if $e_i < 1$, means x_i is more approach to the request value x_{ri}. So, we can define a tiny correction variable as s, here, we let s=0.05, then modify the a_i adaptively as bellows

$$a_i = \begin{cases} \min[a_{i-1} + s, 1] & , e_i > 1 \\ a_{i-1} & , e_i = 1 \\ \max[a_{i-1} - s, s] & , e_i < 1 \end{cases} \tag{10}$$

For $0 \le a_i \le 1$, we should use min and max to keep a_i legal, Then, repeat the estimate procedure at next request time slot.

4. Results and Discussions

We simulated the network traffic to input link of rtPS queue at a given SS. Based on this simulation we evaluated cumulative data entry A(t), D(t), and the corresponding requests for time slots. The input link of rtPS queue is assumed to receive data at an average rate of 1Mbps. The output link of rtPS queue is served at a rate of C = 2Mbps. The average packet size is assumed to be 200 bytes. Simulation is done for 2000 packet arrival instants.

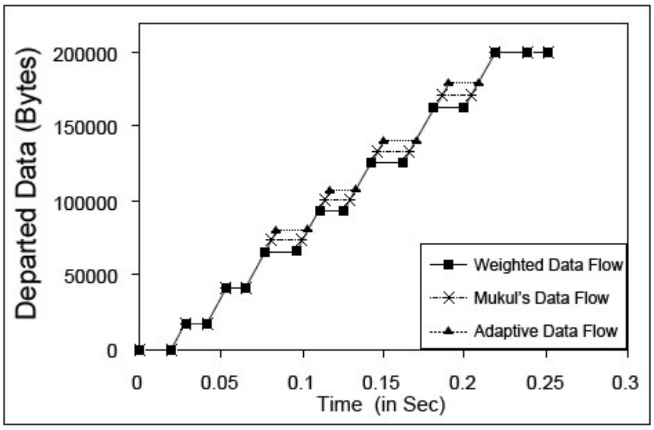

Fig 1. Departed data packets versus time

Figure 1 presents departed data packets versus time in a SS for adaptive and weighted (referred as non-adaptive algorithm) and Mukul's adaptive scheme. It is observed that our proposed adaptive algorithm (referred as adaptive flow) takes less time to transmit the entire arrived packet from a SS than weighted and Mukul's for same packet arrivals. Thus it provides a better QoS for rtPS services with respect to delay. This is because; the departure in weighted flow considers only the amount of data in that rtPS queue at request time, and the Mukul's only consider about the past request, but have not consider about the current bandwidth request of upper layers. However the departure in case of adaptive flow considers not only the request, but also the pattern of data that enters the rtPS queue at the time of request and additionally data between times of request to bandwidth allocation information arrival to SS from BS. Thus effectively, more data will depart from rtPS queue in case of adaptive flow, as data will flow out in larger chunks in former case.

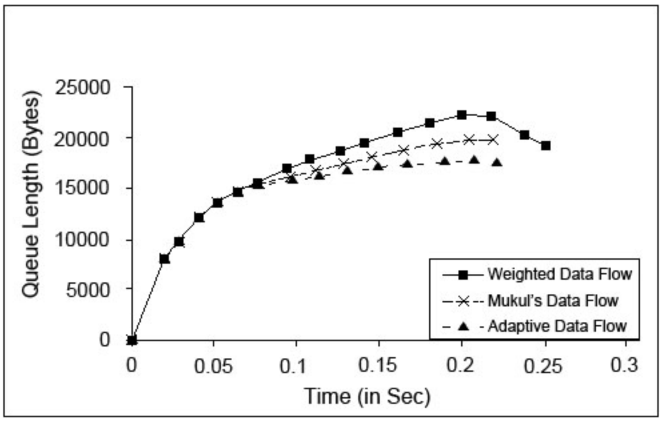

Fig 2. Average queue length versus time

Figure 2, presents average queue length (in bytes) versus time (Sec) for rtPS queue at SS, for weighted and Mukul's as well as adaptive flows. This shows that queue length at

SS is smaller in our adaptive request scheduler. This is attributed to fast data flow out of rtPS queue due to adaptive flow mechanism as mentioned above. It can also be seen that data remains for longer time in case of weighted flow and Mukul's adaptive flow, for the same amount of data entering the rtPS queue. Moreover, adaptive algorithm needs less buffer size, as the queue length is less.

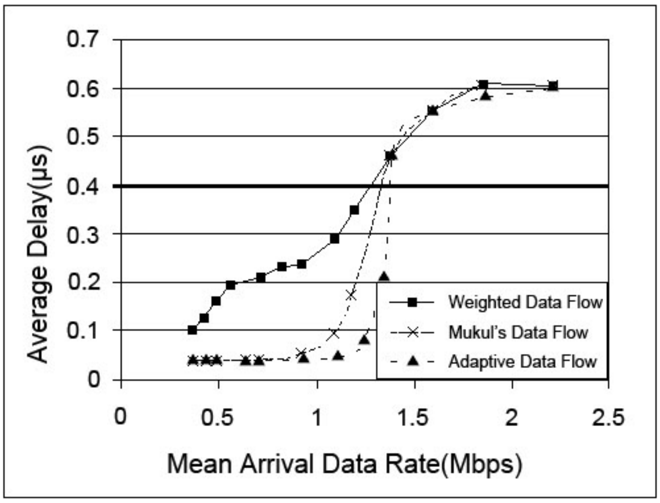

Fig 3. Average delay versus mean packet arrival rate

Figure 3, presents average delay (in micro-seconds) versus mean packet arrival rate in Mbps at SS, for adaptive, weighted and Mukul's adaptive flows. The mean data arrival rates range from 0.5 Mbps to 2 Mbps. At data arrival rate less than (say about 3/4 times) output link speed (here output link speed 2Mbps) the adaptive curve has less delay than the other two curves. It has been observed that, at 1 Mbps input rtPS packets the average delay in adaptive flow presents decrease of about 70% of delay in comparison to weighted flow. When mean arrival rate approaches output link speed the performance of weighted and Mukul's and adaptive request schemes become nearly equal because the output link has been saturated.

5. Conclusion

The convergence of all kinds of traffic, will lead to a complex design-scheduling algorithm for real-time communication. Our adaptive scheduling algorithm for WiMax's rtPS packets has shown remarkable decrement in delay. This is particularly suited for real time services. Moreover, this algorithm needs lesser size buffer at SS in comparison to conventional scheduling algorithms and Mukul's adaptive scheme. Based on these calculations, further optimization can be done for multi-session rtPS service; Multi-service can also be optimized by this scheme.

6. References

[1] Draft of WiMax Standard, "IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems", IEEE 802.16 forum.

[2] Carl Eklund, Roger B. Markas, and Kenneth L. Stanwood, "IEEE Standard 802.16: A technical overview of the wireless MAN Air Interface for Broadband Wireless Access", IEEE Communications Magazine, pp 98-107, June 2002.

[3] GuoSong Chu, Deng Wang and Shunliang Mei, "A QoS architecture for the MAC protocol of IEEE 802.16 BWA system" Communications, Circuits and Systems and West Sino Expositions, pp 435-439, IEEE 2002.

[4] Dong-Hoon Cho, Jung-Hoon Song, Min-Su Kim and Ki-Jun Han, "Performance analysis of the IEEE 802.16 wireless Metropolitan Area Network", Proceedings of the first International Conference on Distributed Frameworks for Multimedia Applications, IEEE Computer Society, 2005.

[5] Kitti Wongthavarawat and Aura Ganz, "IEEE 802.16 based last mile wireless military networks with quality of service support", pp 779-784, 2003

[6] Howon Lee, Taesoo Kwon and Dong-Ho Cho, "An efficient uplink scheduling algorithm for VoIP Services in IEEE 802.16 BWA Systems", IEEE, pp 3070-3074, 2004.

[7] Anurag Kumar, D. Manjunath and Joy Kury, Communication Networking : An analytical approach, Elsevier, 1st Edition, 2004.

[8] Jean-Yves Le Boudec and Patrick Thiran, A theory of deterministic queuing systems for the Internet, Open book, May 2004

[9] Reetesh Mukul, Pradyumna Singh and D. Jayaram, An adaptive bandwidth request mechanism for QoS enhancement in WiMax real time communication, IEEE Wireless and Optical Communications Networks International Conference, April 2006