ISSN:0975-9646

Scheduling in Wi-MAX Networks

A. P.Vikram , B. T.Lakshmi Deepak ,and C. S.Deepthi
K L University, Electronics and Computer Engineering

Abstract: In wireless LANs, WiMAX networks incorporate several quality of service (QoS) mechanisms at the Media Access Control (MAC) level for guaranteed services for data, voice and video. The problem of assuring QoS is basically that of how to allocate available resources among users in order to meet the QoS criteria such as delay, delay jitter and throughput requirements.The key issues and design factors to be considered for scheduler designers. We classify the proposed mechanisms based on the use of channel conditions. The goals of scheduling are to achieve the optimal usage of resources, to assure the QoS guarantees, to maximize goodput and to minimize power consumption while ensuring feasible algorithm complexity and system scalability.

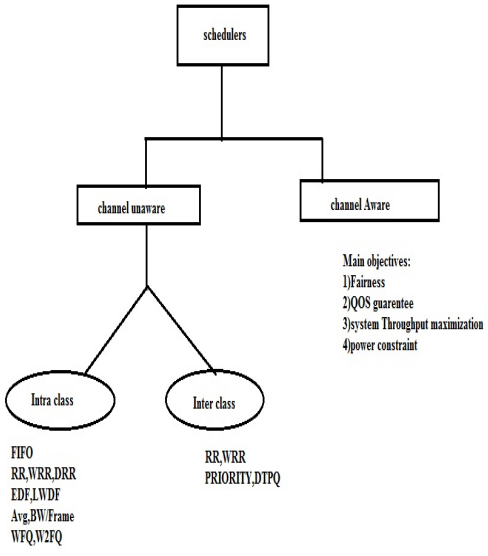
Keywords: WiMAX, QoS, Resource Allocation, Scheduling.

INTRODUCTION

Classification of schedulers

Mainly we focus on the scheduler at BS, especially DL-BS scheduler. For this scheduler, the queue length and packet size information are easily available. To guarantee the QoS for MS at UL-BS scheduler, the polling mechanism is involved. Once the QoS can be assured, how to split the allocated bandwidth among the connections depends on the MS scheduler.

scheduling techniques for WiMAX can be classified into two main categories: channel-unaware schedulers and channel-aware schedulers Basically, the channel-unaware schedulers use no information of the channel state condition in making the scheduling decision.



Classification of WiMAX schedulers

Channel-unaware schedulers generally assume error-free channel since it makes it easier to prove assurance of QoS. However, in wireless environment where there is a high variability of radio link such as signal attenuation, fading, interference and noise, the channel-awareness is important. Ideally, scheduler designers should take into account the channel condition in order to optimally and efficiently make the allocation decision.

Channel-Unaware Schedulers

This type of schedulers makes no use of channel state conditions such as the power level and channel error and loss rates. These basically assure the QoS requirements among five classes mainly the delay and throughput constraints. Although, jitter is also one of the QoS parameters, so far none of the published algorithms can guarantee jitter. and also the mappings between the scheduling algorithms and the QoS .

Intra-class Scheduling

Intra-class scheduling is used to allocate the resource within the same class given the QoS requirements.

Round Robin (RR) algorithm: A side from FIFO, round robin allocation can be considered the very first simple scheduling algorithm. RR fairly assigns the allocation one by one to all connections. The fairness considerations need to include whether allocation is for a given number of packets or a given number of bytes. With packet based allocation, stations with larger packets have an unfair advantage.

Moreover, RR may be non-work conserving in the sense that the allocation is still made for connections that may have nothing to transmit. Therefore, some modifications need to be made to skip the idle connections and allocate only to active connections. However, now the issues become how to calculate average data rate or minimum reserved traffic at any given time and how to allow for the possibility that an idle connection later has more traffic than average? Another issue is what should be the duration of fairness? For example, to achieve the same average data rate, the scheduler can allocate 100 bytes every frame for 10 frames or 1000 bytes every 10^{th} frame.

Since RR cannot assure QoS for different service classes, RR with weight, Weighted Round Robin (WRR), has been applied for WiMAX scheduling .The weights can be used to adjust for the throughput and delay requirements. The weights are in terms of queue length and packet delay or the number of slots. The weights are dynamically changed over time. In order to avoid the issue of missed opportunities, variants of RR such as Deficit Round Robin(DRR) or Deficit Weighted Round Robin (DWRR) can be used for the variable size packets. The main advantage of these variations of RR is their simplicity. The complexity is $O(1)$ compared to $O(log(N))$ and $O(N)$ for other fair queuing algorithms. Here, N is the number of queues.

Weighted Fair Queuing algorithm (WFQ):
WFQ is a General Processor Sharing (GPS). WFQ does not make the assumption of infinitesimal packet size. Basically, each connection has its own FIFO queue and the weight can be dynamically assigned for each queue. The resources are shared in proportion of the weight. For data packets in wired networks with leaky bucket, an end-to-end delay bound can be provably guaranteed. With the dynamic change of weight, WFQ can be also used to guarantee the data rate. The main disadvantage of WFQ is the complexity, which could be *O(N).*
To keep the delay bound and to achieve worst-case fairness property, a slight modification of the WFQ, Worst-case fair Weighted Fair Queueing (WF2Q) was introduced. Similar to WFQ, WF2Q uses a virtual time concept. The virtual finish time is the time GPS would have finished sending the packet. WF2Q looks for the packet with the smallest virtual finishing time and whose virtual start time has already occurred instead of searching for the smallest virtual finishing time of all packets in the queue.
In achieving the QoS assurance, procedure to calculate the weight plays an important role. The weights can be based on several parameters. Aside from queue length and packet delay we mentioned above, the size of bandwidth request can be used to determine the weight of queue (the larger the size, the more the bandwidth). The ratio of a connection's average data rate to the total average data rate can be used to determine the weight of the connection The minimum reserved rate can be used as the weight . The pricing can be also used as a weight .Here, the goal is to maximize service provider revenue.
Delay-based algorithms: This set of schemes is specifically designed for real-time traffic such as UGS, ertPS and rtPS service classes, for which the delay bound is the primary QoS parameter and basically the packets with unacceptable delays are discarded. Earliest Deadline First (EDF) is the basic algorithm for scheduler to serve the connection based on the deadline. Largest Weighted Delay First (LWDF) chooses the packet with the largest delay to avoid missing its deadline.
Delay Threshold Priority Queuing (DTPQ) was proposed for use when both real-time and non real-time traffic are present. A simple solution would be to assign higher priority to real-time traffic but that could harm the non realtime traffic. Therefore, urgency of the real-time traffic is taken into account only when the head-of-line (HOL) packet delay exceeds a given delay threshold. This scheme is based on the tradeoff of the packet loss rate performance of rtPS with average data throughput of nrtPS with a fixed data rate. Rather than fixing the delay, the author also introduced an adaptive delay threshold-based priority queuing scheme which takes both the urgency and channel state condition for real-time users adaptively into consideration.
Note that variants of RRs, WFQs and delay based algorithms can resolve some of the QoS requirements. However, there are no published papers considering the tolerated delay jitter in the context of WiMAX networks. Especially for UGS and ertPS, the simple idea is to introduce a zero delay jitter by the fragmentation mechanism. Basically, BS transfers the last fragmented packet at the end of period. However, this fragmentation increases the overhead and also requires fixed buffer size for two periods. Compared to EDF, this simple technique may require more bursts. This needs to be investigated further.

INTER-CLASS SCHEDULING

As shown in Fig. 8, RR, WRR and priority-based mechanism have been applied for inter-class scheduling in the context of WiMAX networks. The main issue for inter-class is whether each traffic class should be considered separately, that is, have its own queue. For example, in rtPS and nrtPS are put into a single queue and moved to the UGS (highest priority) queue once the packets approach their deadline. Similarly in UGS, rtPS and ertPS queues are combined to reduce the complexity. Another issue here is how to define the weights and/or how much resources each class should be served. There is a loose bound on service guarantees without a proper set of weight values.
Priority-based algorithm (PR): In order to guarantee the QoS to different classes of service, priority-based schemes can be used in a WiMAX scheduler . For example, the priority order can be: UGS, ertPS, rtPS, nrtPS and BE, respectively. Or packets with the largest delay can be considered at the highest priority. Queue length can be also used to set the priority level, e.g., more bandwidth is allocated to connections with longer queues .
The direct negative effect of priority is that it may starve some connections of lower priority service classes. The throughput can be lower due to increased number of missed deadlines for the lower service classes' traffic. To mitigate this problem, Deficit Fair Priority Queuing (DFPQ) with a counter was introduced to maintain the maximum allowable bandwidth for each service class . The counter decreases according to the size of the packets. The scheduler moves to another class once the counter falls to zero. DFPQ has also been used for inter-class scheduling .
To sum up, since the primary goal of a WiMAX scheduler is to assure the QoS requirements, the scheduler needs to support at least the five basic classes of services with QoS assurance. To ensure this, some proposed algorithms have indirectly applied or modified existing scheduling disciplines for each WiMAX QoS class of services. Each class has its own distinct characteristics such as the hard-bound delay for rtPS and ertPS. Most proposed algorithms have applied some basic algorithms proposed in wired/wireless networks to WiMAX networks such as variations of RR and WFQ. For example, to schedule within a class, RR and WFQ are common approaches for nrtPS and BE and EDF for UGS and rtPS , The priority-based algorithm is commonly used for scheduling between the classes. For example, UGS and rtPS are given the same priority which is also the highest priority .
Moreover, "two-step scheduler is a generic name for schedulers that try first to allocate the bandwidth to

meet the minimum QoS requirements - basically the throughput in terms of the number of slots or subcarrier and time duration and delay constraints. Then, especially in WiMAX networks (OFDMA-based) in the second step, they consider how to allocate the slots for each connection. This second step of allocating slots and subcarriers is still an open research area. The goal should be to optimize the total goodput, to maintain the fairness, to minimize the power and to optimize delay and jitter.

Channel-Aware Schedulers

The scheduling disciplines we discussed so far make no use of the channel state condition. In other words, they assume perfect channel condition, no loss and unlimited power source. However, due to the nature of wireless medium and the user mobility, these assumptions are not valid. For example, a MS may receive allocation but may not be able to transmit successfully due to a high loss rate. In this section, we discuss the use of channel state conditions in scheduling decisions.

The channel aware schemes can be classified into four classes based on the primary objective: fairness, QoS guarantee, system throughput maximization, or power optimization.

Basically, the BS downlink scheduler can use the Carrier to Interference and Noise Ratio (CINR) which is reported back from the MS via the CQI channel. For UL scheduling, the CINR is measured directly on previous transmissions from the same MS. Most of the purposed algorithms have the common assumption that the channel condition does not change within the frame period. Also, it is assumed that the channel information is known at both the transmitter and the receiver.

In general, schedulers favor the users with better channel quality since to exploit the multiuser diversity and channel fading, the optimal resource allocation is to schedule the user with the best channel or perhaps the scheduler does not allocate any resources for the MS with high error rate because the packets would be dropped anyway.

However, the schedulers also need to consider other users' QoS requirements such as the minimum reserved rate and may need to introduce some compensation mechanisms. The schedulers basically use the property of multi-user diversity in order to increase the system throughput and to support more users.

Consider the compensation issue. Unlike the wireless LAN networks, WiMAX users pay for their QoS assurance. Thus, in the argument of what is the level of QoS was brought on due to the question whether the service provider should provide a fixed number of slots. If the user happens to choose a bad location (such as the basement of a building on the edge of the cell), the provider will have to allocate a significant number of slots to provide the same quality of service as a user who is outside and near the base station. Since the providers have no control over the locations of users, they can argue that they will provide the same resources to all users and the throughput observed by the user will depend upon their location. A generalized weighted fairness (GWF) concept, which equalizes a weighted sum of the slots and the WiMAX equipment manufacturers can implement generalized fairness. The service providers can then set a weight parameter to any desired value and achieve either slot fairness or throughput fairness or some combination of the two. The GWF can be illustrated as an equation below.

TOTAL_SLOTS= $\sum_{i=1}^{N} Siw S_i+(1- w) B_i/ M= wS_j+(1-w)B_j/ M$ for all subscriber i and j in N $B_i = b_i S_i$

Here, S_i and B_i are total number of slots and bytes for subscriber i. b_i is the number of bytes per slot for subscriber i. N is the number of active subscribers. M is the highest level MCS size in bytes. w is a general weight parameter.

It has been observed that allowing unlimited compensation to meet the QoS requirements may lead to bogus channel information to gain resource allocations The compensation needs to be taken into account with leading/lagging mechanisms. The scheduler can reallocate the bandwidth left-over either due to a low channel error rate or due to a flow not needing its allocation. It should not take the bandwidth from other well-behaved flows. In case, there is still some left-over bandwidth, the leading flow can also gain the advantage of that left-over. However, another approach can be by taking some portion of the bandwidth from the leading flows to the lagging flows. When the error rate is high, a credit history can be built based on the lagging flows and the scheduler can allocate the bandwidth based on the ratio of their credits to theirs minimum reserved rates when the error rate is acceptable [60]. In either case, if and how the compensation mechanism should be put into consideration are still open questions.

1) Fairness

This metric mainly applies for the Best Effort (BE) service. One of the commonly used baseline schedulers in published research is the Proportional Fairness Scheme (PFS) [61, 62]. The objective of PFS is to maximize the long-term fairness. PFS uses the ratio of channel capacity (denoted as $W_i(t)$) to the long-term throughput (denoted as $R_i(t)$) in a given time window T_i of queue i as the preference metric instead of the current achievable data rate. $R_i(t)$ can be calculated by exponentially averaging the *ith* queue's throughput in terms of T_i. Then, the user with the highest ratio of $W_i(t)/R_i(t)$ receives the transmission from the BS. Note that defining T_i affects the fluctuation of the throughput. There are several proposals that have applied and modified the PFS. For example, T_i derivation with delay, given 5 ms frame duration, setting T_i to 50 ms is shown to result in an average rate over 1 second instead of 10 seconds with T_i = 1000 ms. In [64], the moving average was modified to not update when a user queue is empty. A starvation timer was introduced in to prevent users from starving longer than a predefined threshold.

2) QoS Guarantee

Modified Largest Weighted Delay First (M-LWDF) can provide QoS guarantee by ensuring a minimum throughput guarantee and also to maintain delays smaller than a predefined threshold value with a given

probability for each user (rtPS and nrtPS). And, it is provable that the throughput is optimal for LWDF . The algorithm can achieve the optimal whenever there is a feasible set of minimal rates area. The algorithm explicitly uses both current channel condition and the state of the queue into account. The scheme serves the queue *j* for which "$\rho_i W_j(t) r_j(t)$" is maximal, where ρ_i is a constant which could be different for different service classes (the difficulty is how to find the optimal value of ρ_i). $W_i(t)$ can be either the delay of the head of line packet or the queue length. $r_i(t)$ is the channel capacity for traffic class *i*.

There are several proposals that have used or modified MLWDF. For example, in , the scheduler selects the users on each subcarrier during every time slot. For each subcarrier *k*, the user selection for the subcarrier is expressed by

i=max[$channelgain(i,k)\times HOL\_delay(i)\times\{a(i)/\ d(i)\}$]

In this equation, *a* is the mean windowed arrival and *d* is mean windowed throughput. *"a"* and *"d"* are averaged over a sliding-window. HOL_delay is the head of line delay. The channel state information is indirectly derived from the normalized channel gain. Note that the channel gain is the ratio of the square of noise at the receiver and the variance of Additive White Gaussian Noise (AWGN). Then, the channel gain and the buffer state information are both used to decide which subcarriers should be assigned to each user. The buffers state information consists of HOL_delay, *a* and *d*. Similar to M-LWDF, Urgency and Efficiency based Packet Scheduling (UEPS) was introduced to make use of the efficiency of radio resource usage and the urgency (timeutility as a function of the delay) as the two factors for making the scheduling decision. The scheduler first calculates the priority value for each user based on the urgency factor expressed by the time-utility function (denoted as $U'_i(t)$) × the ratio of the current channel state to the average (denoted as $R_i(t)/R'_i(t)$). After that, the subchannel is allocated to each selected user *i* where:

i= max$|U'_i(t)|\times[R_i(t)/R'_i(t)]$

Another modification of M-LWDF has been proposed to support multiple traffic classes The UEPS is not always efficient when the scheduler provides higher priority to nrtPS and BE traffic than rtPS, which may be near their deadlines. This modification handles QoS traffic and BE traffic separately. The HOL packet's waiting time is used for QoS traffic and the queue length for BE traffic.

3) System Throughput Maximization

A few schemes, e.g.., focus on maximizing the total system throughput. In these, Max C/I (Carrier to Interference) is used to opportunistically assign resources to the user with the highest channel gain.

Another maximum system throughput approach is the exponential rule in that it is possible to allocate the minimum number of slots derived from the minimum modulation scheme to each connection and then adjust the weight according to the exponent (p) of the instant modulation scheme over the minimum modulation scheme. This scheme obviously favors the connections with better modulation scheme (higher *p)*. Users with better channel conditions receive exponentially higher bandwidth. Two issues with this scheme are that additional mechanisms are required if the total slots are less than the total minimum required slots. And, under perfect channel conditions, connections with zero minimum bandwidth can gain higher bandwidth than those with non-zero minimum bandwidth.

Another modification for maximum throughput was proposed in using a heuristic approach of allocating a subchannel to the MS so that it can transmit the maximum amount of data on the subchannel. Suppose a BS has *n* users and *m* subchannels, let λ_i be the total uplink demand (bytes in a given frame) for its UGS connections, R_{ij} be the rate for MS_i on channel *j* (bytes/slot in the frame), N_{ij} be the number of slots allocated to MS_i on subchannel j, the goal of scheduling is to minimize the unsatisfied demand, that is,

Minimize

$$\sum_{1\le j\le m}[\lambda_i-(\sum_{1\le j\le m}R_{ij}N_{ij})]$$

subject to the following constraints:

$$\sum_{1\le j\le n}N_{ij}<N_j^{'}$$

And

$$\sum_{1\le j\le m}R_{ij}N_{ij}\le\lambda_i$$

Here, N'_j is the total number of slots available for data transmission in the *j*th subchannel. A linear programming approach was introduced to solve this problem, but the main issue is the complexity, which is $O(n^3m^3N)$. Therefore, a heuristic approach with a complexity of only $O(nmN)$, was also introduced by assigning channels to MSs that can transmit maximum amount of data.

CONCLUSION:

The main goal of the schedulers are basically to meet QoS guarantees for all service classes, to maximize the system goodput, to maintain the fairness, to minimize power consumption, to have as less a complexity as possible and finally to ensure the system scalability. We classified recent scheduling based on the channel awareness in making the decision. Well-known scheduling discipline can be applied for each class such as EDF for rtPS and WFQ for nrtPS and WRR for inter-class. With the awareness of channel condition and with knowledge of applications, schedulers can maximize the system throughput or support more users.

REFERENCES

[1] IEEE P802.16Rev2/D2, "DRAFT Standard for Local and metropolitan area networks," Part 16: Air Interface for Broadband Wireless Access Systems, Dec. 2007, 2094 pp.

[2] WiMAX Forum, "WiMAX System Evaluation Methodology V2.1," Jul. 2008, 230 pp. Available: http://www.wimaxforum.org/technology/documents/

[3] S. Lu, V. Bharghavan, and R. Srikant, "Fair scheduling in wireless packet networks," *IEEE/ACM Trans. Netw.*, vol. 7, pp. 473-489, Aug. 1999.
[4] N. H. Vaidya, P. Bahl, and S. Gupta, "Distributed fair scheduling in a Wireless LAN," *IEEE Trans. Mobile Comput*., vol. 4, pp. 616-629, Dec. 2005.
[5] L. Tassiulas and S. Sarkar, "Maxmin fair scheduling in wireless networks," in *Proc. IEEE Computer Communication Conf.,* 2002, New York, NY, vol. 2, pp. 763-772.
[6] P. Bhagwat, P. Bhattacharya, A. Krishna, and S. K. Tripathi, "Enhancing throughput over Wireless LANs using channel state dependent packet scheduling," in *Proc. IEEE Computer Communication Conf.,* San Francisco, CA, 1996, vol. 3, pp. 1133-1140.
[7] S. Shakkottai and R. Srikant, "Scheduling real-time traffic with deadlines over a wireless channel," *ACM/Baltzer Wireless Networks*., vol. 8, pp. 13-26, Jan. 2002.
[8] E. Jung and N. H. Vaidya, "An energy efficient MAC protocol for Wireless LANs," in *Proc. IEEE Computer Communication Conf.,* New York, NY, 2002, vol. 3, pp. 1756-1764.
[9] X. Zhang, Y. Wang, and W. Wang, "Capacity analysis of adaptive multiuser frequency-time domain radio resource allocation in OFDMA systems," in *Proc. IEEE Int. Symp. Circuits and Systems.,* Greece, 2006, pp. 4-7.
[10] T. Ohseki, M. Morita, and T. Inoue, "Burst Construction and Packet Mapping Scheme for OFDMA Downlinks in IEEE 802.16 Systems," in *Proc. IEEE Global Telecomunications Conf.,* Washington, DC, 2007, pp. 4307-4311.
[11] Y. Ben-Shimol, I. Kitroser, and Y. Dinitz, "Two-dimensional mapping for wireless OFDMA systems," *IEEE Trans. Broadcast.*, vol. 52, pp. 388-396, Sept. 2006.
[12] A. Bacioccola, C. Cicconetti, L. Lenzini, E. A. M. E. Mingozzi, and A. A. E. A. Erta, "A downlink data region allocation algorithm for IEEE 802.16e OFDMA," in *Proc. 6th Int. Conf. Information, Communications & Signal Processing.*, Singapore, 2007, pp. 1-5.
[13] C. Desset, E. B. de Lima Filho, and G. Lenoir, "WiMAX Downlink OFDMA Burst Placement for Optimized Receiver Duty-Cycling," in *Proc. IEEE Int. Conf. Communications.*, Glasgow, Scotland, 2007, pp. 5149-5154.
[14] C. So-In, R. Jain, and A. Al-Tamimi, "Capacity Estimations in IEEE 802.16e Mobile WiMAX networks," Submitted for publication, *IEEE Wireless Comm. Mag.*, April 2008.
[15]J. Shi, G. Fang, Y. Sun, J. Zhou, Z. Li, and E. Dutkiewicz, "Improving Mobile Station Energy Efficiency in IEEE 802.16e WMAN by Burst Scheduling," in *Proc. IEEE Global Telecomunications Conf.,* San Francisco, CA, 2006, pp. 1-5.
[16] L. Tian, Y. Yang, J. Shi, E. Dutkiewicz, and G. Fang, "Energy Efficient Integrated Scheduling of Unicast and Multicast Traffic in 802.16e WMANs," in *Proc. IEEE Global Telecomunications Conf.,* Washington, DC, 2007, pp. 3478-3482.
[18] H. Choi and D. Cho, "Hybrid Energy-Saving Algorithm Considering Silent Periods of VoIP Traffic for Mobile WiMAX," in *Proc. IEEE Int. Conf. Communications.*, Glasgow, Scotland, 2007, pp. 5951-5956.
[19] A. Sayenko, O. Alanen, and T. Hamalainen, "ARQ Aware Scheduling for the IEEE 802.16 Base Station," in *Proc. IEEE Int. Conf. Communications.,* Beijing, China, 2008, pp. 2667-2673.