Cross-layer Uplink Scheduler for the IEEE 802.16 Standard

Flavio A. Kubota, Juliana F. Borin and Nelson L. S. da Fonseca
Institute of Computing, University of Campinas
Campinas 13089-971, SP, Brazil
Email: kubota@lrc.ic.unicamp.br, {juliana,nfonseca}@ic.unicamp.br

Abstract—The IEEE 802.16 standard for broadband wireless access is a low cost solution for Internet access in metropolitan and rural areas. Although it defines five service levels to support real-time and bandwidth demanding applications, scheduling mechanisms are not specified in the standard. Due to the wireless channel variability, scheduling mechanisms widely studied for wired networks are not suitable for IEEE 802.16 networks. This paper proposes an uplink cross-layer scheduler which makes bandwidth allocation decisions based on information about the channel quality and on the Quality of Service requirements of each connection. Simulation results show that the proposed scheduler improves the network performance when compared with a scheduler which does not take into account the channel quality.

Index Terms—IEEE 802.16, Quality of Service, Scheduling, Cross-layer.

I. INTRODUCTION

The IEEE 802.16 standard [1], also known as WiMAX (Worldwide Interoperability for Microwave Access), and its amendment, the IEEE 802.16e [2], are expected to provide Quality of Service (QoS) for stationary and mobile users. To reach this goal, five types of service flows and mechanisms for bandwidth requests by connections in the uplink direction are defined. The Unsolicited Grant Service (UGS), periodically receives fixed size grants without the need to request them. The extended real time Polling Service (ertPS) uses a grant mechanism similar to the one used to support UGS connections, but periodically allocated grants can be used to send bandwidth requests to inform the required grant size. The real-time Polling Service (rtPS) offers periodic unicast bandwidth request opportunities to subscribers; these opportunities ensure latency bound and minimum traffic rate guarantees. The non-real-time Polling Service (nrtPS) provides periodic unicast bandwidth request opportunities, but using more spaced intervals than rtPS, as well as minimum traffic rate guarantee. The Best Effort (BE) shares contention bandwidth request opportunities with the nrtPS service flow.

A complete solution for QoS provisioning requires a scheduling mechanism, however, it is not defined in the standard. The scheduling mechanism aims at guaranteeing the bandwidth required by the subscriber stations as well as enabling the efficient wireless link usage.

Scheduling mechanisms have been widely investigated for wired networks, however, wireless systems pose extra challenges to the development of this mechanism due to the wireless link variability. When not considered, link variability can result in poor bandwidth allocation performance, specially in urban areas where the high amount of obstacles can affect the signal propagation.

In [3], Borin and Fonseca propose a standard-compliant scheduling solution for the uplink traffic in IEEE 802.16 networks. The proposed scheduler provides maximum latency and minimum rate guarantees without violating the maximum sustained traffic rate and the maximum traffic burst values, as required by the standard. However, wireless channel characteristics are not considered in this solution.

This paper extends the scheduler proposed in [3], by using a cross-layer approach, so that bandwidth allocation decisions also take into account information about the channel quality of each mobile subscriber station (MSS). Simulation results show that the proposed approach improves the network performance in suburban areas without affecting the QoS provision required by the IEEE 802.16 standard.

The rest of this paper is organized as follows. Section II discusses related work. Section III describes the channel model considered in this paper. Section IV presents the proposed scheduling mechanism. Section V describes the simulation environment used to evaluate the scheduler. Section VI presents numerical results and Section VII concludes the paper.

II. RELATED WORK

Due to characteristics such as location-dependent and time-varying wireless link capacity, scheduling algorithms developed for wired networks are not suitable for wireless networks. To deal with the wireless link variability problem as well as energy consumption restrictions while providing Quality of Service (QoS), fairness and high channel utilization, many scheduling mechanisms have been proposed [4], [5] such as Channel State Dependent Packet Scheduling (CSDPS), Idealized Wireless Fair Queuing (IWQ), Channel-Condition-Independent Fair Queueing (CIF-Q), Server-Based Fairness Approach (SBFA) and Wireless Fair Service (WFS). However, none of them is able to support the QoS requirements of the five types of service flow defined by the IEEE 802.16e standard [2].

In order to guarantee latency requirements to real time applications, the scheduling mechanism proposed in [6] uses...
a history of packets delays to classify packets in four classes: Extremely urgent, Moderately urgent, Not urgent, and Best Effort. The scheduler also gives higher priority to packets destined to users whose instant channel conditions are better to promote the efficient use of the wireless channel bandwidth.

Iera et al [7] propose a scheduler based on the Worst-Case Fair Weighted Fair Queueing (WFQ$^2$Q+) [8] to support the five types of service defined by the IEEE 802.16e standard. In the proposed scheme, packets can be blocked when the user channel conditions are not satisfactory and a credit counter is updated every time a flow is not able to transmit so that the flow can be compensated for its lag. The main disadvantage of this scheduler is that it is based on the virtual-time approach, which is known as being computationally complex.

An approach with fuzzy controls for the provisioning of fairness and QoS in IEEE 802.16 OFDMA networks is proposed in [9], however, the rtPS service and the minimum reserved traffic rate requirement of the rtPS service are not considered. In [10], the authors propose a two-stage cross-layer QoS support framework with a scheduling algorithm. The scheduler includes latency guarantee and a mechanism to avoid starvation, but fails to provide maximum rate guarantee.

In [11], it is introduced a set of algorithms for the BS to allocate channels/slots to different MSSs in an IEEE 802.16 OFDMA/TDD network, but the authors do not evaluate the performance for their proposal. Schedulers can use different metrics to estimate the channel condition. In [7], the channel quality is measured as a signal-to-noise ratio (SNR), while in [6], [12], and [13] it is estimated according to the instantaneous transmission rate.

III. CHANNEL MODEL

This work uses the OFDM PHY layer model and the COST 231 Hata Propagation Model [14] to simulate the channel of an IEEE 802.16 network in a suburban environment. Although the channel operates in the 500MHz to 2000MHz frequency range, the empiric model has low execution complexity and uses correction factors for suburban, urban, and rural environments.

Based on the distance between the antenna of the base station and the receptor of the mobile subscriber station (MSS), the channel model estimates a path loss value and, hence, computes the signal to noise ratio (SNR) of each packet arriving at the receptor. Depending on the loss probability, the packet can be dropped due to errors. Therefore, the larger the distance between the base station (BS) and the MSS, the higher the packet loss rate and, consequently, the lower the MSS goodput.

IV. SCHEDULING MECHANISM

The scheduler proposed in this section extends the work of Borin and Fonseca [3] by accounting for the variability of the wireless link during scheduling decisions.

The scheduler proposed in [3] is fully-standard compliant and it guarantees the QoS requirements of the five types of service defined in the IEEE 802.16 standard in an error-free channel environment. It uses three queues: low, intermediate and high priority queues. The scheduler serves the queues according to their level of priority. The low priority queue stores the BE bandwidth requests. The intermediate queue stores the bandwidth requests sent by both rtPS and nrtPS connections. These requests can migrate to the high priority queue to guarantee that their QoS requirements are met. In addition to the requests migrating from the intermediate queue, the high priority queue stores periodic grants and unicast bandwidth request opportunities that must be scheduled in the following frame. The BS executes the uplink scheduler at every frame and it broadcasts the scheduling agenda to the MSSs in the UL-MAP message. This scheduler is referred in this paper as MBQoS - Migration-Based Scheduler for QoS provisioning.

At each frame, the MBQoS scheduler generates periodic grants and inserts them into the high priority queue at predefined intervals. In this way, UGS and rtPS grants for data transmission, and rtPS and nrtPS unicast opportunities for bandwidth request transmission are provided as specified by the standard. The maximum latency requirement of rtPS connections is guaranteed by assigning deadline values to their bandwidth requests in the intermediate queue. Requests with deadlines expiring two frames ahead and associated with connections which have not received the minimum reserved traffic rate in a window with duration $T$ migrate from the intermediate queue to the high priority queue. Maximum latency is guaranteed for the flows that do not exceed the minimum reserved traffic rate, as specified by the standard. The minimum reserved traffic rate requirement of rtPS and nrtPS connections is guaranteed in windows of duration $T$. The scheduler computes a priority value for each request at the intermediate queue, considering the per connection: minimum reserved traffic rate, backlogged requests, and traffic rate received in the current window. Requests of connections which have already reached the minimum reserved traffic rate in the current window receive low priority values, while, for the remaining requests, the lower the rate received by the connection, the higher is the priority value associated to them.

A dual leaky bucket is used for maximum burst and maximum rate policing. The capacity of the first bucket (bucket1) is infinite and its leaky rate is equal to the maximum sustained traffic rate. The second bucket (bucket2) has capacity equal to the maximum traffic burst. Therefore, before allocating a grant, the MBQoS scheduler checks whether the grant size is less or equal to the bucket2 size.

To use the network resources more efficiently in a suburban area, where the quality of the channel is location-dependent and time-varying, the uplink scheduler proposed in this paper uses a cross-layer approach to estimate the average channel quality experienced by each MSS. The channel quality is given by the signal to noise ratio (SNR) provided by the physical layer. The channel quality is considered good for an individual MSS when the average value for its SNR is higher than a certain threshold. The average SNR value is calculated using the Exponential Weighted Moving Average (EWMA) method. The threshold value used corresponds to the SNR value when the packet loss probability is 0.5. The OFDM PHY model takes into account the SNR value to infer the correct reception of a packet.

MSSs with bad channel conditions are prevented to transmit by a blocking mechanism in order to reduce the packet loss rate.
and, consequently, the bandwidth waste. To avoid the violation of the maximum latency requirement, rtPS connections are allowed to transmit packets with approaching deadline even when the MSS is blocked. UGS and ertPS connections are not subject to the blocking constraint.

The blocking mechanism can impact the bandwidth fairness provided to the connections, since different MSSs can be blocked for periods with different durations. This problem is generally solved by using a compensation scheme. In the MBQoS scheduler, the minimum reserved traffic rate requirement of rtPS and nrtPS services is guaranteed by assigning a priority value for each connection. One of the parameters used to derive the priority value is the number of bytes requested by the backlogged requests sent by the corresponding connection. Connections hosted by unblocked MSSs are served according to the priority values. In this way, when a MSS is unblocked after a blocking period, its connections are served with high priority in case they have more backlogged data than connections in other MSSs. This scheme provides the expected compensation among connections.

The Algorithm Scheduling presents the proposed scheme. After inserting periodic grants in the high priority queue, the algorithm checks which rtPS and nrtPS requests should migrate from the intermediate queue to the high priority queue (lines 2 and 3). In line 4, the scheduler distributes the non-allocated bandwidth among the BE connections. At the final step, it serves all the requests at the high priority queue.

In the CheckDeadline procedure, the scheduler tries to migrate the rtPS requests from the intermediate queue to the high priority queue using the MigrateBWRequest procedure to guarantee the maximum latency requirement. Although the proposed approach is based on the idea of delaying packets transmission in MSSs with low channel quality, rtPS requests with approaching deadlines may be served even when the average SNR of the corresponding MSS is lower than the expected threshold since, for real time applications, dropping packets is preferable to waiting for delayed packets.

The CheckMinimumBandwidth procedure checks the channel conditions of the MSSs hosting rtPS and nrtPS connections and blocks the ones with low channel quality (lines 12-15). After that, it calculates a priority value for each request in the intermediate queue (lines 16-23) and the intermediate queue is sorted according to the priority values (line 24). Finally, the scheduler tries to migrate requests sent by unblocked MSSs to the high priority queue using the MigrateBWRequest procedure.

The DistributeFreeResources procedure distributes the available bandwidth among the BE requests by migrating some

### Algorithm Scheduling

1. insert, in the high priority queue, the periodic data grants and unicast bandwidth request opportunities that must be scheduled in the next frame;
2. CheckDeadline;
3. CheckMinimumBandwidth;
4. DistributeFreeResources;
5. schedule the requests in the high priority queue starting from the head of the queue;

### Procedures

**CheckDeadline:**

1. for each request i at the intermediate queue do
2. if availableBw = 0 then
3. break
4. if service[CID] = rtPS then
5.  "frame[i] = [(deadline[i] - currentTime) ÷ frameDuration];"
6. if frame[i] = 3 and TwndTR[CID] < minTR[CID] then
7. MigrateBWRequest(i)

**CheckMinimumBandwidth:**

8. for each connection of type rtPS or nrtPS do
11.  bucket2_tmp[CID] = bucket2[CID]
12.  if not haveGoodQuality[CID] then
13.    isBlocked[CID] = true
14.    else
15.    isBlocked[CID] = false
16. for each request i at the intermediate queue do
17.  if minTR[CID] ≤ TwndTR_tmp[CID] or bucket2_tmp[CID] = 0 then
18.    priority[i] = 0
19.  else
24. sort the intermediate queue
25. for each request i at the intermediate queue do
26.  if availableBw = 0 then
27.    break
28.  if isBlocked[CID] = false then
29.    MigrateBWRequest(i)

**DistributeFreeResources:**

30. for each connection of type BE do
31.  if not haveGoodQuality[CID] then
32.    isBlocked[CID] = true
33.  else
34.    isBlocked[CID] = false
35. for each request i at the low priority queue do
36.  if availableBw = 0 then
37.    break
38.  if isBlocked[CID] = false then
39.    MigrateBWRequest(i)
40. for each request i at the intermediate and low priority queue do
41.  if availableBw = 0 then
42.    break
43.  if isBlocked[CID] = true and protocol = TCP then
44.    MigrateBWRequest(i)

**MigrateBWRequest(i):**

45. if BR[i] > availableBw then
46.  grantSize = availableBw
47.  else
48.  grantSize = BR[i]
49. if grantSize > bucket2[CID] then
50.  grantSize = bucket2[CID]
51.  if 0 < grantSize < BR[i] then
52.  create a new request j for connection CID with
53.  BR[j] = BR[i] − grantSize
54.  insert request j in the end of the intermediate queue
55.  BR[i] = grantSize
56.  move request i to high priority queue
57.  TwndTR[CID] = TwndTR[CID] + grantSize
58.  bucket2[CID] = bucket2[CID] − grantSize
59.  availableBw = availableBw − grantSize
of them from the low priority queue to the high priority queue. Only connections hosted by MSSs with good channel quality may have their requests migrated to the high priority queue. The remaining capacity, if any, is distributed among blocked users with TCP connections in order to reduce the impact of the TCP congestion control mechanism on the transmission rate when timeouts occur.

V. SIMULATION EXPERIMENTS

The scheduler proposed in this paper was implemented in the IEEE 802.16 module for the ns-3 simulator [15].

The simulated network uses a point-to-multipoint topology with a centralized BS and the MSSs distributed around it. The initial distance between the MSSs and the BS ranges from 1600 to 1800 meters. In this range, the signal starts to deteriorate for the assumed modulation. To generate the signal variation, the MSSs move randomly within this distance range.

The physical layer uses the OFDM technology with QPSK 1/2 modulation and coding mode for data packets transmission and BPSK 1/2 modulation and coding for signaling packets transmission. The distance between the MSSs and the BS always guarantees that management packets are not lost. The frame duration is set to 5 ms with 1:1 downlink-to-uplink TDD split.

To eliminate the impact of packet scheduling at the MSSs on the proposed uplink scheduling, each MSS has only one uplink service flow.

The first scenario simulated includes three rtPS flows, three nrtPS flows and three best effort flows. The packet size generated is 540 bytes. In the second scenario, different types of traffic are considered: voice, voice with silence suppression, and video, which are associated with UGS, ertPS, and rtPS services, respectively, and FTP traffic for nrtPS and BE services.

The voice model used is an exponential “on/off” model with mean duration of the “on” and of the “off” periods equals to 1.2 s and 1.8 s, respectively. During the “on” periods, 66-byte packets are generated at every 20-ms [16]. The voice with silence suppression model used the Enhanced Variable Rate Codec (EVRC) [17], with packets generated every 20 ms employing Rate 1 (171 bits/packet), Rate 1/2 (80 bits/packet), Rate 1/4 (40 bits/packet) or Rate 1/8 (16 bits/packet). Video traffic is generated by real MPEG-4 traces [18] encoded at a low quality level. Twenty-minutes long videos with 200 Kbps bit rate are used in the simulations. FTP traffic is generated by an on-off source with infinite residence time on the “on state”, 1000 Kbps transmission rate and 512 packet size.

The unsolicited grant interval for the UGS and for the ertPS service is 20 ms. The unsolicited polling interval for the rtPS service is 20 ms and for the nrtPS service is 1 s.

The maximum latency requirement of the rtPS service is 400 ms and each connection has its own minimum reserved traffic rate and maximum sustained traffic rate requirements (which vary according to the mean rate of the transmitted video). The nrtPS service has minimum reserved traffic rate requirement of 800 Kbps, and maximum sustained traffic rate requirement of 1500 Kbps. The BE service does not have any QoS requirement.

To capture the path-loss, the COST 231 Hata Propagation Model[14] is used while the Free-space model is used for an ideal channel.

Each result was produced by running the simulation five times with different seeds. The mean values and the 95% confidence intervals are shown in the figures.

VI. NUMERICAL RESULTS

This section discusses the results found in the simulations for the two scenarios described. The aim is to assess the scheduler capability of avoiding packet losses, thus providing enhanced Quality of Service. UGS and ertPS receive periodic grants and consequently do not have their quality enhanced by the scheduler.

A. Experiment 1

In the first scenario, there are nine clients, three of each class rtPS, nrtPS and BE. The transmission rate varies from 100 Kbps to 400 Kbps. UDP was the transport protocol used to avoid the impact of the TCP congestion window.

![Fig. 1. Mean goodput for nrtPS service Flow](image1)

![Fig. 2. Mean goodput for BE service Flow](image2)

Figures 1 and 2 show the mean goodput for the nrtPS and BE classes, respectively. The figures compare the goodput given by the cross-layer MBQoS scheduler with those of the MBQoS scheduler operating in an ideal channel as well as in a realistic channel. The difference of the goodput given by the MBQoS scheduler operating in a realistic channel shows the benefits of the adoption of a cross-layer approach. Such difference can be of about 25% of the cross-layer approach. Moreover, the difference between the goodput given by the
cross-layer scheduler and that of MBQoS when considering an ideal channel can be interpreted as the transmission that would probably be lost due to the poor signal quality at the receiver. Furthermore, the difference between these goodput values shows that the cross-layer scheduler is capable to produce goodput which approaches the ideal goodput. Figure 2 shows that the cross-layer scheduler promotes increased goodput to the BE class which takes advantage of the non used bandwidth by the classes with delay requirements. However, it is not possible to provide such advantage under high loads causing a sharp decrease of the goodput of this class.

Although the graphics are not presented due to space limitation, simulation results show that the cross-layer scheduler is able to guarantee the minimum reserved traffic rate requirement.

### B. Experiment 2

In the second scenario, the number of MSSs increases from 5 to 25 in steps of 5 units (one for each type of service). This experiment aims at checking the impact of the TCP congestion window on nrtPS and BE goodput results as well as whether or not the proposed scheduler is able to guarantee the maximum latency requirement for rtPS connections.

![Fig. 3. Mean goodput for nrtPS service Flow](image1)

![Fig. 4. Mean goodput for BE service Flow](image2)

Figures 3 and 4 show the mean goodput for the nrtPS and BE classes, respectively. The figures compare the goodput given by the cross-layer MBQoS scheduler with those of the MBQoS scheduler operating in a realistic channel. The low performance of the cross-layer scheduler is due to the impact of the TCP congestion control mechanism. Since for the TCP protocol timeouts indicate that packets are being dropped due to a congested network, blocked users have their transmission rates reduced by the congestion control mechanism. For the MBQoS scheduler, the low performance results are expected, since it allows transmission from users under low channel quality conditions resulting in high packets loss rate.

![Fig. 5. Mean delay for rtPS service Flow](image3)

Figure 5 shows that the proposed scheduler is able to guarantee the maximum latency requirement of rtPS connections. Although not shown here due to space limitations, the cross-layer scheduler is able to provide periodic grants for UGS and erTPS connections as specified by the IEEE 802.16 standard.

### VII. Conclusions

This paper introduced a standard-compliant cross-layer uplink scheduler for IEEE 802.16 networks which uses information about channel quality conditions to efficiently allocate bandwidth while furnishing the expected QoS. Simulation results show that the proposed scheduler improves the goodput of nrtPS and BE connections when compared to a non cross-layer approach and is able to guarantee maximum latency for real time connections.

As a future work, the interaction between the proposed scheduling mechanism and the TCP congestion control mechanism shall be investigated to counteract the resource underutilization side effect.

### REFERENCES


[17] 3GPP2 C.S0014-0, “Enhanced Variable Rate Codec (EVRC).”