

Request Mechanisms to Reduce the Contention Period in 802.16: A Comparison

Jesús Delicado, Francisco M. Delicado and Luis Orozco-Barbosa

Instituto de Investigación en Informática de Albacete (I³A)
Universidad de Castilla-La Mancha (UCLM)
02071-Albacete, Spain
e-mail: {jdelicado,franman,lorozco}@dsi.uclm.es

Abstract Nowadays, a large number of applications characterized by stringent end-to-end requirements are on high demand. The development of Broadband Wireless Access (BWA) should enable the ubiquitous and cost-effective deployment of such applications. The IEEE 802.16 standard is currently one of the promising BWA technology under development. The IEEE 802.16 defines the principles of the Physical and Medium Access Control mechanisms to be implemented by the standard. Under the proposed MAC protocol, the Base Station (BS) is responsible of allocating the bandwidth required by the Subscriber Stations (SS). A SS issues its bandwidth request during a pre-defined contention period. Since the user data and signalling data are transmitted over the same band, the design of efficient signalling mechanism aiming to reduce the time and space length of the contention period is nowadays the main focus of a large number of studies. In this paper, we introduce and evaluate two novel bandwidth request mechanisms and compare their performance with the signalling mechanisms proposed by the standard.

1 Introduction

Nowadays, there is an increasing demand for high-speed Internet wireless access mainly due to the use of applications characterized by stringent time requirements. The IEEE 802.16 standard is currently one of the most promising proposals Broadband Wireless Access technologies [7]. An upgrade from fixed BWA systems to mobile service to vehicular speeds was ratified in December 2005, IEEE 802.16e [8].

The MAC protocol of the IEEE 802.16 Standard defines a set of connection-oriented services and distinguishes two types of devices: the Base Station (BS) and the Subscriber Stations (SSs). The former is responsible of allocating on-demand the bandwidth required by the all active SSs. The SSs enable the connectivity between the end-user equipments and the BS. According to the MAC protocol, a SS should request to the BS the required bandwidth before actually transmitting a data unit over the wireless channel. The request mechanisms to be employed by the SS depend on the QoS requirements of the application. It is therefore clear that the signalling protocol plays a central element on the QoS guarantees offered by the

IEEE 802.16 Standard. The signalling protocol should allow the SSs to request the bandwidth according to the needs of active connections associated to the SSs. In turn, the BS should comprise the signalling mechanisms to respond to the various SSs requirements.

In this work, we present two signalling proposals whose main aim is to reduce the number of signalling messages. We also conduct a comparative study of these two proposals and the signalling mechanisms defined by the standard. The paper is organized as follows: Section 2 briefly describes the IEEE 802.16 Standard. Section 3 overviews previous efforts reported in the literature related to our proposals. Our proposals are described in Section 4. In Section 5, we conduct a comparative performance study of the three different signalling schemes. Section 6 draws our conclusions and future work plans.

2 IEEE 802.16

In this section, we provide an overview of the IEEE 802.16 physical (PHY) and the medium access control (MAC) layers. The IEEE 802.16 proposes a PMP (Point-to-MultiPoint) topology, and optionally a mesh topology [4]. Under the PMP mode, the standard supports a range of frequencies from 2 to 66 GHz, including the licensed and license-exempt bands. Depending on the range of frequencies there may be necessary line-of-sight (LOS) or non-line-of-sight (NLOS) between the communicating entities. Three types of modulation can be used: QPSK (Quadrature Phase-Shift Keying), 16-QAM (Quadrature Amplitude Modulation) and 64-QAM, but only QPSK is mandatory. Through the different releases of the standard, three different physical air interface specifications are proposed: SC/SCa (Single Carrier), OFDM (256-carrier Orthogonal-Frequency Division Multiplexing) and OFDMA (2048-carrier OFDM scheme).

The MAC is a centralized and connection-oriented mechanism, i.e., the BS allocates the resources and provides the system with QoS-aware mechanism according to the required needs of the various applications while the SSs request the needed bandwidth on a frame by frame basis.

The communication between the BS and the SSs is carried out by means of fixed-length frames and is time multiplexed by means of TDMA (*Time Division Multiplexing Access*), divided into two subframes: the downlink subframe (from the BS to the SSs) and the uplink subframe (from the SSs to the BS). Both subframes are multiplexed using time-division duplexing (TDD) or frequency-division duplexing (FDD).

The downlink subframe begins with a Frame Start Preamble (FSP), which is used for synchronization and equalization at the PHY layer. This is followed by the Frame Control Section (FCS), which is composed of management messages. Two of these messages are the downlink and uplink maps (DL-MAP and UL-MAP, respectively), which are comprised by the bandwidth allocations for SSs, and their corresponding burst profiles, in both downlink and uplink directions, respectively. Following these

maps, other important management messages are the DCD and UCD, which indicate the physical characteristics of the physical channels in both directions. Finally, a time-division multiplexing (TDM) portion is introduced, organized into bursts on a decreasing robustness.

The uplink subframe is divided into three periods. The initial period is used by the SSs requiring access to the network. This part is followed by another contention period reserved to convey the *BW-request* messages sent by each active connection to request their bandwidth needs to the BS. These two contention periods are divided into slots. Finally, the last period is used by the SSs to send their user data to the BS (uplink). Each transmission is separated by SS Transition Gaps in order to properly synchronize the channel activity.

The sizes of the downlink and uplink subframes are determined dynamically by the BS on a frame to frame basis. This information is broadcasted by the BS to the SSs through *DL-MAP* and *UL-MAP* messages, respectively. The BS is also responsible node of controlling the number of transmission opportunities given to the SSs during the first two periods and the data transmission through the *UL-MAP*. The Tx/Rx Transition Gap (TTG) and the Rx/Tx Transition Gap (RTG) are used to switch from transmission to reception and vice versa.

2.1 Grants

According to the first release of the IEEE 802.16 Standard [6], the bandwidth grants to be issued by the BS should be associated to a specific connection by explicitly indicating its CID (Connection IDentifier) in the grant message. This mechanism is called GPC (*Grants Per Connection*) and it requires that the BS issues a grant in response to a request.

The 2004 release of the Standard [7] defines the GPSS (*Grants Per Subscriber Station*) mechanism. Under this mechanism, the bandwidth is granted to the SS through its Basic CID and not explicitly to each and every individual CID associated to the SS. In this way, the number of needed signalling messages is reduced to one grant per SS. Under this mechanism, the SSs are charged of distributing the granted bandwidth among their backlogged connections.

2.2 Requests

According to the IEEE 802.16 Standard, the bandwidth requests have to be issued by the SSs on a connection by connection basis. Since potential channel access conflicts may arise during the contention period reserved in the uplink subframe for this purpose, the standard defines a conflict resolution mechanism, namely the *truncated binary exponential backoff* algorithm. This mechanism is put in place to resolve potential conflicts due to simultaneous access to this period by different

SSs. Under this mechanism, the initial time slot to be used by a SS to transmit a request message is randomly selected from a set of consecutive slots, referred as a window. Once having issued its request, the SS activates a timer and waits for its confirmation for a maximum time period defined by the initial value of the timer. If the timer expires before the SS receives the confirmation, the SS resends the message during a time slot determined from a window size. In this second attempt, the size of the window is set to twice the value of the one having been used in the previous transmission. This procedure is repeated as long as a conflict arises, i.e., the SS does not get a confirmation within the time period defined by the timer and up to a given number of attempts defined by the standard, whatever it arises first. It is clear that as the number of active connections increases the collision probability also increases. This condition will adversely affect the network performance. In the following, we will review the various proposals being put forward to limit the number of signalling messages being used to convey the bandwidth requests.

3 Related work

Many research efforts on the performance of IEEE 802.16 contention algorithm have been carried out through the past few years. In [5], a dynamic mini-slot allocation scheme has been introduced based on an estimate of the maximum number of data packets that can be transmitted through a frame. In [14], the authors try to minimize the delay for transmitting the requests, by identifying the number of active SSs and by uniformly distributing the transmission attempts over the available random access slots. The authors of [12] have concluded that the optimal size of the contention period is $2M - 1$, where M is the number of SSs (users). In [1], the authors have determined the size of the contention period aiming to maximize the throughput. In this case, the result is a contention period equal to the number of competing SSs. In the previous papers, the length of the contention period is fixed by the number of active SSs. However, they do not take into account the statistics of the various applications which may potentially be exploited to improve the overall system performance. The authors of [15] have developed a novel random access method. They have compared the random access method with centralized polling and station-grouping mechanisms, and introduce an adaptive switching mechanism between both methods depending on the request arrival rate. In [11] a new algorithm has been introduced, called Multi-FS-ALOHA, which divides the contention period into two parts: the first is used by the SSs issuing bandwidth request for the first time while the second part is used by the SSs having previously attempted to transmit without success. These two parts are dynamically fixed on a frame by frame basis. A cross-layer design has been developed in [16] to design the optimal size of the contention period with various classes of services. As conclusion, a contention period size, which is approximately equal to the number of the SSs, is used to optimize all classes of service performance. In [3] the authors have conducted a general study on the contention period in IEEE 802.16. They have analyzed the impact of

the contention period resolution mechanisms over the data transmission period, in terms of the achievable data throughput and packet delay.

In all the previous studies reported, the IEEE 802.16 standard request mechanism has been used. However, they do not make an in-depth analysis of the bandwidth required to convey the request messages. Obviously, this bandwidth increases as the contention period increases, limiting the amount of bandwidth available for data transmission. In this work, we propose some simple enhancements to the request mechanism defined by the standard, aiming to reduce the length of the contention period.

4 Our Proposals

In order to reduce the introduced overhead during the contention period, we proposed a new request bandwidth mechanism in [2], called RGPSS (*Requests and Grants Per Subscriber Station*). In it, each SS only sends one request with the needed amount of bandwidth of all the active connections belonging to this SS.

Although the RGPSS reduces the contention period, the received information by the BS about the needed resources for each service flow is null. So, the method defined by the standard (GPSS) gives better information about the needs of each connection individually, allowing that the BS allocates the resources of the system using QoS criteria. However, the information sent by the GPSS method is too much, because it uses a request per connection while the grants are made per SS. In this way, a lot of requests are sent by each SS, depending on its active connections, increasing the collision probability. On the other hand, as the bandwidth is granted to each SS and not to individual connections, the BS only needs to know the requirements of each service flow in each SS (as aggregation of the requirements of all connections of each service flow, associated with a SS) to implement a scheduler, which could classify and prioritize service flows applying QoS criteria.

So, we have proposed a new requesting mechanism, called RPSF (*Requests Per Service Flow*). In it, each SS sends one request per service flow aggregating the required amount of bandwidth to transmit all enqueued data in all connections which belong to this service flow. In this way, the RPSF reduces the number of requests with regard to the GPSS, introducing more overhead than the RGPSS, but giving the necessary information to the BS to allocate the resources of the system using QoS criteria. At the same time, the RPSF gives the same information to the BS as the GPSS, but reducing the number of requests when more than one active connection belongs to the same service flow. That is, the GPSS uses more BW-request messages to give the same information to the BS as the RPSF does.

To use this mechanism according to the standard, each SS will select one fixed CID among every active connection which belongs to this service flow to send its request, and this fixed CID will always be used to request bandwidth of this kind of connections. When the BS receives a request with this fixed CID, it will be interpreted like a request according to the service flow and not the individual connec-

Table 1 Traffic characterization

	BE - BG	Voice	Video
Start (sec.)	0	uniform(0, 0.024)	uniform(0, 0.5)
ON period (sec.)	pareto(1, 1.9)	always	
OFF period (sec.)	pareto(1, 1.25)	never	
Interarrival time (sec.)	0.0215	0.024	
Packet size (bytes)	138	384	396
Average rate	21 Kbps	16 Kbps	75 Kbps
Trace			Jurassic Park
Frame rate			25 fps

tion. Another possibility is to assign a specific CID to each service flow, which will always be used to request bandwidth like the Basic CID. In both cases, an interpretation of CIDs inside the request messages is only required.

5 Performance Evaluation

In this section, we carry out a performance analysis of our proposals jointly with the standard method (GPSS). Throughout our study, every simulation is conducted using an IEEE 802.16 model implemented in the OPNET Modeler v11.5 tool [13].

In our simulations, we consider an IEEE 802.16 wireless network describing a point-to-multipoint (PMP) system consisting of a BS and a different number of SSs in each simulation. Each SS runs voice, video, best-effort (BE) and background (BG) applications, which are modelled as described in Table 1. These four applications are assigned to four different service flows whose priorities in descending order are: voice, video, best-effort and background. The rate corresponding to each kind of service flow is constant in each scenario and corresponds to a quarter of the overall demand.

All nodes operate at 28 MHz, with a symbol rate of 22.4 MBaud. All transmissions are done using QPSK modulation with a bit rate of 44.8 Mbps. According to the standard, a frame duration of 1 ms is used. The mode of operation is FDD. We assume ideal channel conditions and a system operating in a steady-state, where the number of connections does not change over time.

The BS and each SS use a priority scheduling discipline. The highest priority is assigned to the voice and video applications, followed by the best-effort applications and finally the background applications. Within a given service class, the connections are served in a FCFS order (*First Come First Server*).

The initial values for initial and maximum backoff window in the contention resolution algorithm are 4 and 10, respectively.

Throughout our study, we have simulated 15 seconds of operation of each particular scenario, collecting statistics after a warm-up period of 4 seconds. Each point in our plots is an average over 32 simulation runs, and error bars indicate 95% confidence interval.

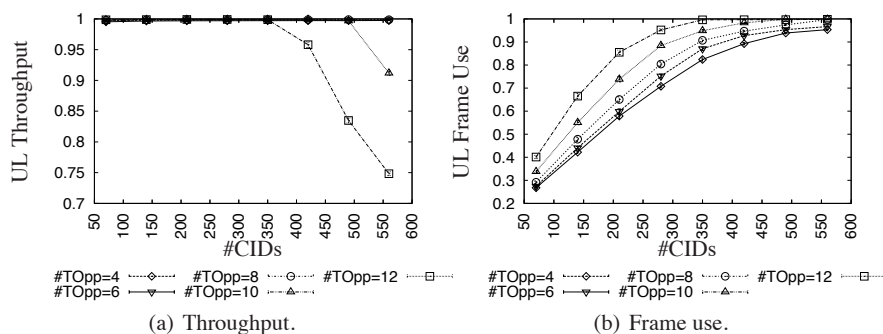


Fig. 1 Throughput and frame use in RPSF mode (*Campus Connectivity* scenario).

Due to the nature of the WiMAX standard, it is possible to consider a wide range of applications and usage scenarios, over a broad range of markets and geographies [9]. We consider two scenarios: the *Campus Connectivity* scenario and the *Individual Connectivity* scenario.

5.1 *Campus Connectivity* scenario

In this scenario, there are a limited number of SSs with a large number of connections per SS and all kind of applications are used. Each SS is usually used to provide buildings with connectivity on the Internet through the BS.

This scenario is composed of a fixed number of SSs. Five in our case. In each SS runs all kind of applications as described before. Initially, 5 voice, 1 video, 4 BE and 4 BG applications are running, and they are increased in the same proportion, obtaining 14, 28, 42, 56, 70, 84, 98 and 112 connections per SS.

Every connection requests bandwidth through the contention period independently from their service flow. This mainly responds to our interest in reducing the size of contention period. The size of the contention period (#TOpp) is set, according to [2], to 8 and 4 in the case of GPSS and RGPSS, respectively. For RPSF, the #TOpp value is elected after an evaluation, in which #TOpp is increased by 2 from 4 to 12 transmission opportunities.

5.1.1 RPSF parametrization

To select the optimum #TOpp value in the RPSF case, a study increasing its value from 4 to 12 is done.

Figure 1 shows the uplink throughput and frame use in the RPSF mechanism. Regarding the throughput, in all data except for the largest contention period, all the

data is effectively transmitted. This is mainly due to reduced available capacity for data transmission.

Figure 1.b. We can see as the frame use increases with the #TOpp, because the uplink subframe for data is reduced, and so, the frame is totally used with less connections.

Another important aspect to analyze is the end-to-end delay (Ete delay) of the critical applications, voice and video. They are depicted on Figure 2. Both types of applications keep their respective deadlines higher than the obtained values in the simulations. These deadlines are 10 ms and 100 ms for voice and video, respectively, in line with specified values by standards and in literature [10]. A decrease of the delay is appreciated when the number of connections is increased. This is because when a grant is received by a SS, it allocates resources in descending order of priority. As the number of connections increases, it is possible that new incoming high priority connections takes advantages of the allocated resources to the SS regardless of their class.

With the study of these previous figures, we can conclude that, for the RPSF mechanism, if we select the smallest value from transmission opportunities (#TOpp=4) the traffic load of the network can totally be transmitted and the critical applications keep their deadlines.

5.1.2 GPSS, RGPSS and RPSF comparative

In this section, we will compare the performance of these different mechanisms under study using the best values. In Figure 3.a shows the uplink throughput. The GPSS mechanism is not capable of keeping the network in good conditions when the traffic load is high. Regarding the collision probability, see Figure 3.b, the RGPSS mechanism exhibits the best results. This is not surprising since this mechanism requires the least number of signalling packets to operate.

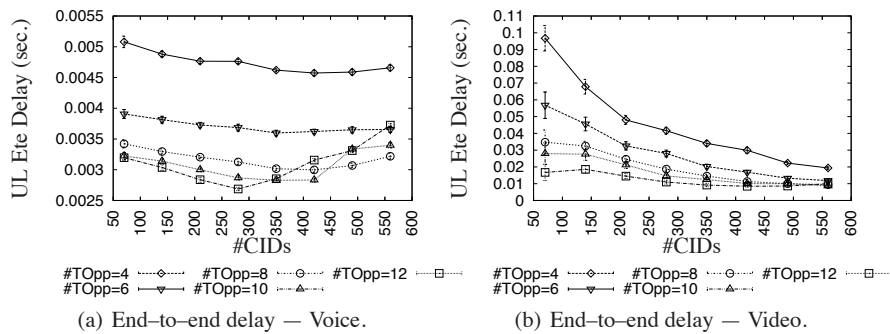


Fig. 2 End-to-end delay in RPSF mode (*Campus Connectivity* scenario).

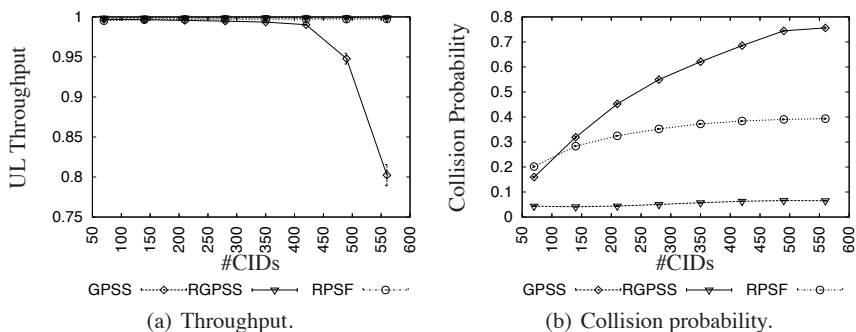


Fig. 3 Throughput and collision probability of GPSS, RGPSS and RPSF (*Campus Connectivity* scenario).

Figure 4 shows the mean of the end-to-end delay of the two time critical applications under study, voice and video. Initially, the RPSF mechanism shows the best results for voice while the RGPSS mechanism proves to be better for video applications. However, a closer look to the Cumulative Distribution Function (CDF) of the end-to-end delay of these applications in the case of the largest number of connections (Figure 5), the RPSF mechanism is considered to be the best, that is, it reduces the size of the contention period and at the same time it keeps the deadlines of these applications.

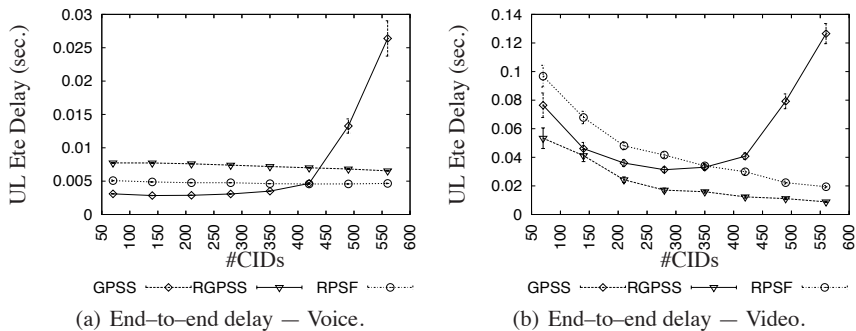


Fig. 4 Uplink end-to-end delay of voice and video applications of GPSS, RGPSS and RPSF (*Campus Connectivity* scenario).

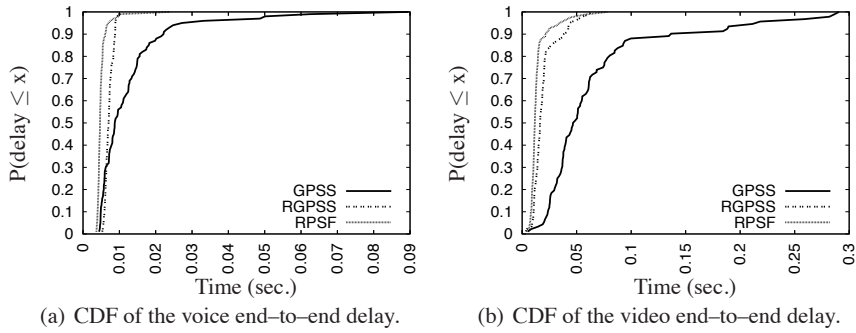


Fig. 5 CDF of end-to-end delay of GPSS, RGPSS and RPSF — #CID=112 (*Campus Connectivity* scenario).

5.2 Individual Connectivity scenario

In this case, the BS is located in the control center or head office, allowing each SS (or user) to connect directly to the BS. We consider a variable number of SSs in this scenario. This number is increased by 4. One of these four SSs only runs a voice application, which is an aggregating application, that is, the resulting voice application is composed of 15 voice applications similar to the voice applications describe above. For a video application, the aggregation is 16 video applications. BE application is composed of 12 original BE applications, and finally 13 BG applications compose the final BG application. In this way, every SS only has one application with independence of its type, and the total amount of needed bandwidth is the same for each service flow (kind of application).

This scenario is composed of a variable number of SSs, which is increased by 4 from 20 to 48 SSs. Each one runs a different kind of application. In this scenario, we

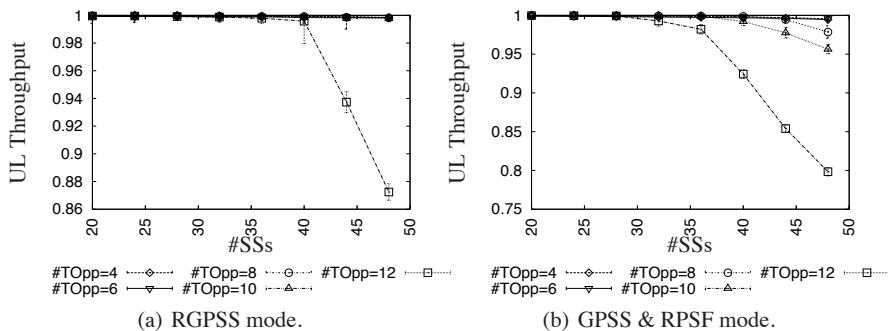


Fig. 6 Throughput (*Individual Connectivity* scenario).

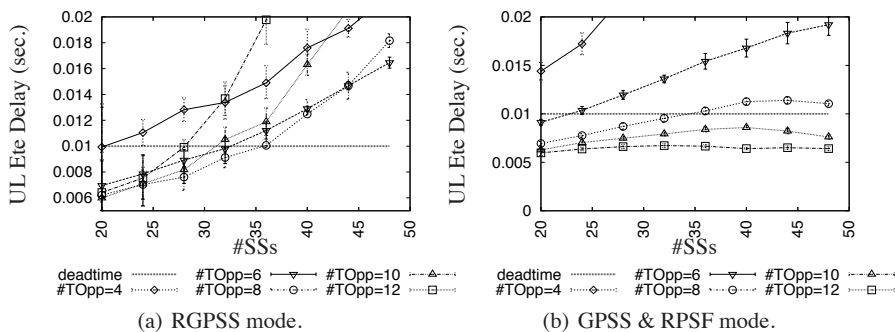


Fig. 7 End-to-end delay — Voice (*Individual Connectivity* scenario).

increase the #TOpp by 2 from 4 to 12, in order to evaluate the performance of each request method (GPSS, RGPSS and RPSF) as function of the contention period.

The behavior of the GPSS and RPSF mechanism is similar through the simulations. This result is expected, since each SS supports only one connection. In other words, the number of request to be sent by service flow and by connections are the same.

Regarding the uplink throughput, Figures 6.a and 6.b, the best behavior is exhibited by the RGPSS mechanism, because only with the highest value of the contention period this mechanism is not able to transmit all data. However, if we analyze the end-to-end delay of the two critical applications (Figures 7 and 8), we observe that the RGPSS mechanism is unable to meet their deadlines, which are depicted on the figures with a horizontal line in their maximum values. However, the GPSS and RPSF mechanisms can meet these voice and video deadlines of voice and video if #TOpp values are higher than 8. This is due to the fact that in the RGPSS method, the bandwidth request does not inform of the service flow associated to each connection. The requests are served in a FCFS order. On the other hand, the GPSS and

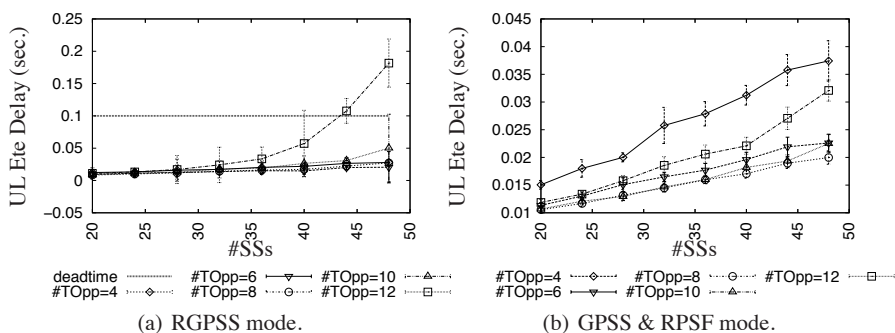


Fig. 8 End-to-end delay — Video (*Individual Connectivity* scenario).

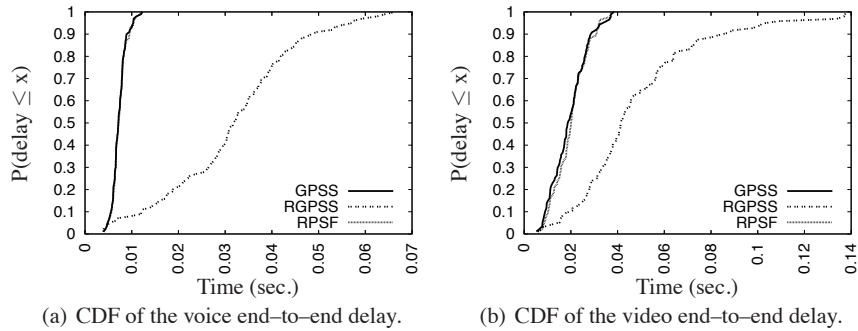


Fig. 9 CDF of end-to-end delay of GPSS, RGPSS and RPSF — #SS=48 (*Individual Connectivity scenario*).

RPSF methods provide the BS with information on the service flow of the connections which request bandwidth. For that, the BS can prioritize the requests, and then, it firstly grants resources to the most important connections.

In Figure 9, the CDF of the end-to-end delay for voice and video are represented. When #Topp=10 is used and the number of SSs is equal to 48, we can see that the GPSS and RPSF mechanisms maintain the video and the voice end-to-end delays below of their deadline values, and their variations are minimal. In the case of the RGPSS method, we can see that every data voice received exceeds the deadline of this type of service flow.

6 Conclusions

In this paper, we have presented a comparative study of various to bandwidth request mechanisms for the IEEE 802.16 networks. It has been shown that our proposals exhibit better behavior than the mechanism proposed by the standards: shorter contention period to request the same amount of bandwidth.

Out of our two proposals, the RGPSS mechanism provides the lowest collision probability by limiting the number of bandwidth requests per SS. However, the RPSF method provides the smallest end-to-end delay for the time critical applications. We conclude that the RPSF mechanism is the best mechanism to provide IEEE 802.16 networks with QoS support. Moreover, our proposals are compatible with the IEEE 802.16 standard, providing higher QoS criteria.

Our future work plans, we will undertake the study of other bandwidth request methods defined by the standard: piggyback and polling unicast. We are also thinking on providing networks with QoS support through the use of a more robust scheduling algorithm.

Acknowledgements This work has been jointly supported by the Spanish MEC and European Commission FEDER funds under the grants “Consolider Ingenio–2010 CSD2006–00046” and “TIN2006–15516–004–02”; by JCCM under the project PAI06–0106 and the grant 05/112.

References

1. Cho, D.H., Song, J.H., Kim, M.S., Han, K.J.: Performance Analysis of the IEEE 802.16 Wireless Metropolitan Area Network. In: 1st International Conference on Distributed Frameworks for Multimedia Applications (DFMA'05), pp. 130–137 (2005)
2. Delicado, J., Delicado, F.M., Orozco-Barbosa, L.: Study of the IEEE 802.16 Contention-based Request Mechanism. In: Personal Wireless Communications (PWC'07), vol. 245, pp. 87–98 (2007)
3. Doha, A., Hassanein, H., Takahara, G.: Performance Evaluation of Reservation Medium Access Control in IEEE 802.16 Networks. In: IEEE International Conference on Computer Systems and Applications, pp. 369–374 (2006)
4. Eklund, C., Marks, R.B., Stanwood, K.L., Wang, S.: IEEE Standard 802.16: A Technical Overview of the WirelessMANTM Air Interface for Broadband Wireless Access. IEEE Communications Magazine **40**(6), 98–107 (2002)
5. Hawa, M., Petr, D.W.: Quality of Service Scheduling in Cable and Broadband Wireless Access Systems. 10th IEEE International Workshop on Quality of Service pp. 247–255 (2002)
6. IEEE: IEEE 802.16 Standard—Local and Metropolitan Area Networks—Part 16: Air Interface for Fixed Broadband Wireless Access Systems (IEEE Std 802.16–2001) (2002)
7. IEEE: IEEE 802.16 Standard—Local and Metropolitan Area Networks—Part 16: Air Interface for Fixed Broadband Wireless Access Systems (IEEE Std 802.16–2004) (2004). Revision of IEEE Std 802.16–2001
8. IEEE: IEEE 802.16 Standard—Local and Metropolitan Area Networks—Part 16: Air Interface for Fixed Broadband Wireless Access Systems—Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands and Corrigendum 1 (IEEE Std 802.16e–2005) (2006). Amendment and Corrigendum to IEEE Std 802.16–2004
9. Inc., W.C.: Can WiMAX Address Your Applications? WiMAX Forum (2005)
10. Karam, A., Tobagi, F.: On the Traffic and Service Classes in the Internet. In: In Proc. of IEEE GLOBECOM'00 (2000)
11. Kobliakov, V.A., Turlikov, A.M., Vinel, A.V.: Distributed Queue Random Multiple Access Algorithm for Centralized Data Networks. In: 10th International Symposium on Consumer Electronics (ISCE'06), pp. 1–6 (2006)
12. Oh, S.M., Kim, J.H.: The Analysis of the Optimal Contention Period for Broadband Wireless Access Network. In: Proceedings of the 3rd International Conference on Pervasive Computing and Communications Workshops (PerCom'05 Workshops), pp. 215–219 (2005)
13. OPNET Technologies, I.: OPNET Modeler 11.5 (2005)
14. Vinel, A., Zhang, Y., Lott, M., Tiurlikov, A.: Performance Analysis of the Random Access in IEEE 802.16. In: Proceedings of the 16th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'05), vol. 3, pp. 1596–1600 (2005)
15. Vinel, A., Zhang, Y., Ni, Q., Lyakhov, A.: Efficient Request Mechanism Usage in IEEE 802.16. In: Proceedings of the IEEE GLOBECOM'06 (2006)
16. Yan, J., Kuo, G.S.: Cross-layer Design of Optimal Contention Period for IEEE 802.16 BWA Systems. In: IEEE International Conference on Communications (ICC'06), vol. 4, pp. 1807–1812 (2006)