

New Contention Resolution Schemes for WiMAX

Jesús Delicado*, Qiang Ni†, 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

†School of Engineering and Design

Brunel University, West London, UK

e-mail: Qiang.Ni@brunel.ac.uk

Abstract—The use of Broadband Wireless Access (BWA) technology is increasing due to the use of Internet and multimedia applications with strict requirements of end-to-end delay and jitter, through wireless devices. The IEEE 802.16 standard, which defines the physical (PHY) and the medium access control (MAC) layers, is one of the BWA standards. Its MAC layer is centralized basis, where the Base Station (BS) is responsible for assigning the needed bandwidth for each Subscriber Station (SS), which requests bandwidth competing between all of them. The standard defines a contention resolution process to resolve the potential occurrence of collisions during the requesting process. In this paper, we propose to modify the contention resolution process to improve the network performance, including end-to-end delay and throughput.

Index Terms—Resource request, QoS, IEEE 802.16, WiMAX.

I. INTRODUCTION

Voice and video applications, which have strict requirements of bandwidth and bounded delay, jointly with the success of high-speed Internet, through wireless devices, have spurred the demand of Broadband Wireless Access (BWA) systems, due to their flexibility and low installation cost. The IEEE 802.16 WiMAX standard, which is one of the BWA standards, appeared in 2001 [1], being its actual release IEEE 802.16-2004 [2]. Later, the IEEE 802.16e [3] upgraded the standard from fixed BWA systems to mobile service at vehicular speeds. Refer to [2], [3], for a detailed description about the standard.

Two modes of operation are defined by 802.16: a point-to-multipoint (PMP) topology, and optionally a mesh topology. Moreover, two grant mechanisms can be used: the *Grants Per Subscriber Station* (GPSS) mechanism, in which the grants are issued on a station by station basis, being necessary a scheduler inside every SS to allocate the granted bandwidth. A second mechanism was defined in the first release of the standard [1], called *Grants Per Connection* (GPC), in which the bandwidth is granted to individual connections. In this paper we focus on the PMP mode, GPC grant mode and use the frequency-division duplexing (FDD).

Two different types of nodes are defined by 802.16: the Base Station (BS), which allocates the resources of the system between every active connection, and the Subscriber Stations (SSs), which describe devices which are origin and destination of connections. Because the BS allocates the resources, it has

to know the needs of each connection in the system, allocating the resources between all received requests. To resolve the potential occurrence of collisions in the requesting process, the IEEE 802.16 standard defines a mandatory method of contention resolution. In a heavy load condition, the requesting process heavily penalizes the network performance, due to the number of requests is larger, which results in transmitting less data. Thus, our objective is to reduce the number of requests by modifying the contention resolution process, and therefore improving the network performance.

The rest of this paper is organized as follows: Section II provides a description of the contention resolution process defined in the 802.16. In Section III, we review the related works. Our proposal is presented in Section IV, and evaluated in Section V. Finally, Section VI concludes this paper.

II. IEEE 802.16 CONTENTION RESOLUTION PROCESS

The medium access control (MAC) layer of WiMAX is connection-oriented, that is, every flow in the system is associated to a connection through a Connection Identifier (CID). Moreover, to fulfill the Quality of Service (QoS) requirements of each type of application, once a connection is created, it will be uniquely associated with a service flow characterized by a set of QoS parameters. The service flow not only reflects the QoS requirements but also specifies the adopting scheme to request the needed resources in the uplink direction. And the same time, the MAC layer is centralized, that is, the BS node acts as a central controller, having to know the resource requirements of each SS. The main problem is the sharing of the uplink channel by all SSs. In this way, every SS is required to request the bandwidth needed by all the active connections associated to it. The mechanisms defined by the standard operate following one of two well-known principles, namely *polling* or *contention*. In polling, an SS is polled by the BS to send a request (BW-request message) in the assigned time slots, while in contention, an SS can issue its request using a random access scheme. It is well known that the use of the latter scheme may translate into important multiplexing gains due to the bursty traffic characteristics of many applications. It is for this reason that the random access has become the focus of many proposals and studies.

In the contention process, when two or more SSs attempt to transmit their requests in the same time slot a collision is produced. To resolve potential conflicts, a mandatory contention resolution process, based on a *truncated binary exponential backoff* algorithm, is supported by the standard, being the initial backoff window and the maximum backoff window controlled by the BS. The values are specified as part of the UCD message (`RequestBackoffStart` and `RequestBackoffEnd` fields, respectively) and represent a power-of-two value, being their sizes equal to 8 bits. The aim of the algorithm is to minimize the collision probability between the requests transmitted in the contention period.

When an SS has a connection, at least one packet to transmit, and wants to enter the contention process, it sets its internal backoff window equal to `RequestBackoffStart` field. Within its backoff window, the SS shall randomly select a number, which indicates the number of contention transmission opportunities¹ the SS shall defer before transmitting, considering only contention transmission opportunities for which this transmission would have been eligible.

After transmitting a `BW-request` message, the SS waits for a grant in a subsequent frame. Once received, the contention resolution process is complete. Otherwise, the SS shall consider the request lost if no data grant has been received within a temporal interval, called T16. Then, it shall increase its backoff window by a factor of two, as long as it is lower than the maximum backoff window (`RequestBackoffEnd` field), randomly selecting a number within its new backoff window and repeating the deferring process described above. This retry process continues until the maximum number of retries has been reached. At this time, the PDU, which generated the contention resolution process, shall be discarded.

III. RELATED WORKS

Over the past years, different technologies have appeared in which a contention period is defined to request the needed resources by the different nodes of the system, being developed some studies, like [4], [5] in HFC (*Hybrid Fiber Coaxial*), and [6], [7] in cable modem. An overview and comparison of contention resolution algorithm are shown in the latter papers. An analysis about the exponential backoff algorithm is developed in [8], obtaining the saturation throughput and the medium access delay. A new scheme is proposed for 802.11 in [9], assigning a different contention window depending on the priority of the different classes of service. In [10] is presented a model to determine the optimal fraction of the uplink subframe dedicated to contention in order to minimize the response time in DOCSIS (*Data Over Cable Service Interface Specifications*).

Recently, some research efforts on the performance of the 802.16 contention process have been carried out. References [11]–[14] are focused on determining the optimal contention period size. Other studies have focused on analyzing

¹A transmission opportunity is defined as an allocation provided in a UL-MAP or part thereof intended for a group of SSs authorized to transmit Bandwidth or Initial Ranging requests.

the impact of the contention period on the performance of the overall system. In [15], the authors attempt to minimize the delay by selecting the proper values of the backoff window and the maximum backoff stage in the contention resolution algorithm. [16] analyzes the impact of the contention period over the data transmission period, showing the importance of limiting the contention period to improve the network performance. A general study about the truncated binary exponential backoff collision resolution algorithm is shown in [17].

In all the reported papers, a central node is responsible to fix the values of the backoff window for the global system. Sometimes, the values are fixed dynamically but considering the number of SSs in the system, not the number of connections. Moreover, the nature of each application is not considered, using the same values for all the nodes in the system. For this reason, we propose a new mode to assign these values, doing the assignment per connection independently.

IV. OUR PROPOSALS

A. Motivation

The IEEE 802.16 standard defines a mandatory contention resolution algorithm, which has been described above, being the BS the node which fixes the values of the initial and maximum backoff window². In this way, every SS in the system will use the same value for all their connections. In this way, when the number of SSs is going to increase, the average uplink end-to-end (Ete) delay of the applications are increased strongly, as Figure 1 shows. Values below of SS=10 obtain an end-to-end delay more or less constant, but increasing this value the delay increases to high values, because the system is congested. Thus, if the initial value is small, then the collision probability will be large, but if the initial value is large, then the time to wait will be large, so the end-to-end delay will increase, doing that the applications with strict requirements overtake their bounded delays.

²When we are talking about the backoff window, we are referencing the power value to determine the size of the backoff window, that is, if the initial backoff window is 8, then the referenced value is 3 ($2^3 = 8$).

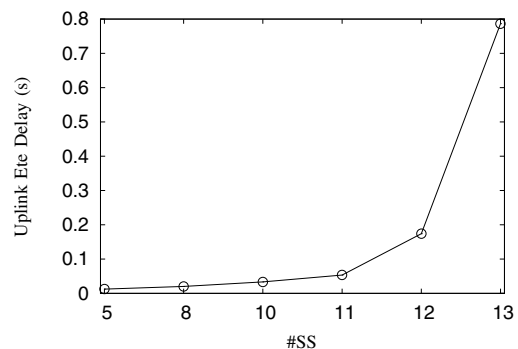


Figure 1. Uplink end-to-end delay by #SS.

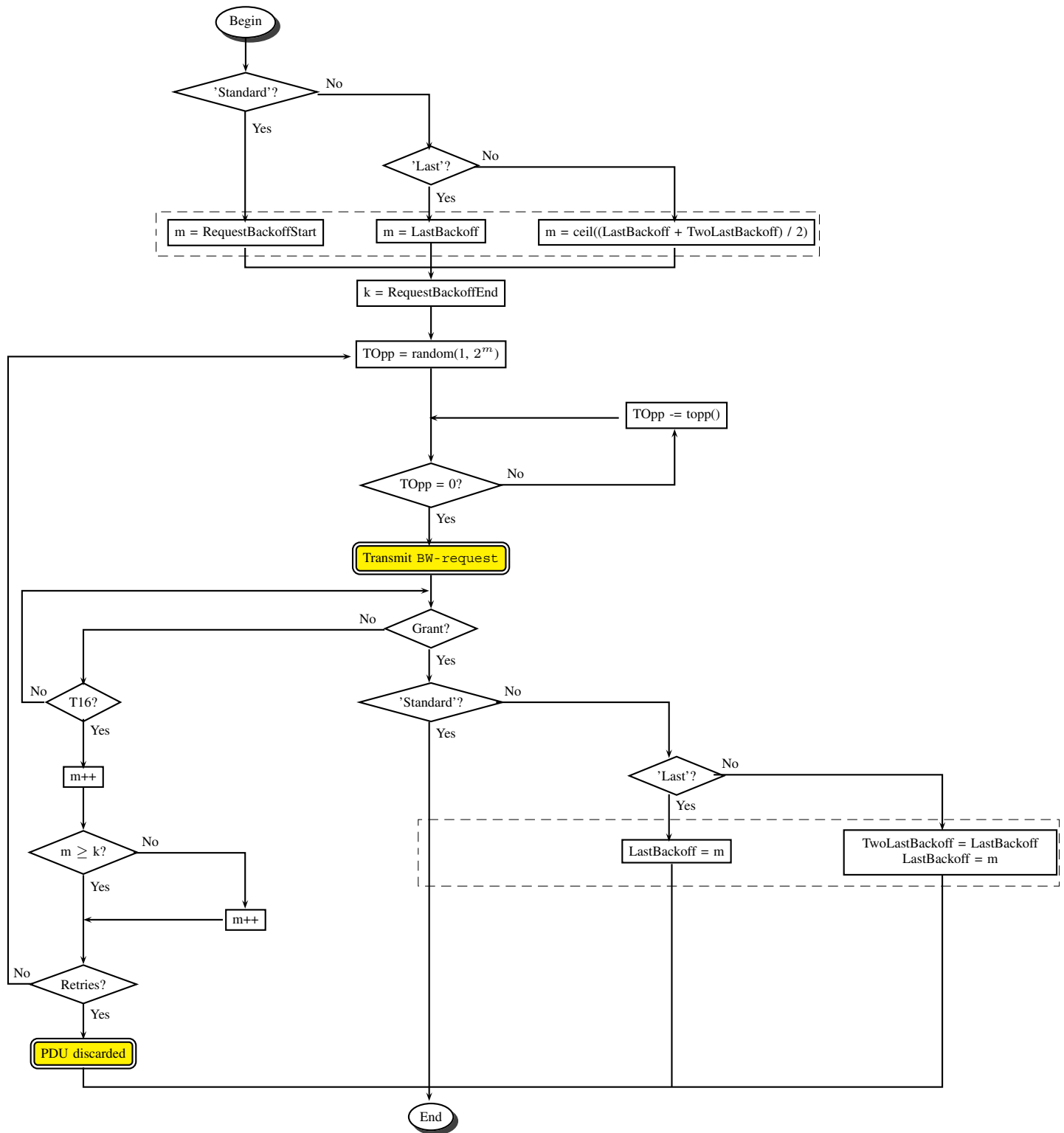


Figure 2. Flowchart of the contention resolution algorithm (standard and our two proposals).

B. Proposals

If it is used a fixed value to assign the initial backoff window, then the end-to-end delay will be increased, degrading the prioritized applications, like voice and video, which have

strict requirements with respect to their corresponding deadlines. In the case it was used a dynamic assignment of this value, then the initial value will be equal for every connection in the system independently of the type of application, causing the same problem, that is, the end-to-end delay rises.

When a system is in a steady state, a connection will need the similar waiting time to send a request than last time it was, because the conditions of the system are not so much variable through the time. So, instead assigning the initial backoff window in the same way every time and trying to minimize the waiting time to the different connections in the system, we propose a new method to assign the initial value for the initial backoff window, basing on connection (Connection IDentifier, CID). Thus, the different applications can use different initial values according to their priority, doing independence between the connections to improve the more prioritized applications, minimizing the end-to-end delay in general, like we will be able to appreciate in Section V. For this, each SS has to control the value of the backoff window when a request is successful, saving this value per CID.

However, if we only use the last value of the backoff window, then applications defined by an ON-OFF pattern can be in a different state with respect to the last time. For this, another possibility is to consider the two last values of the backoff window, that is, the next time the initial backoff window will be the average of the two previous backoff windows, resolving the situation in which the last state of the application was different from now.

Figure 2 displays the flowchart of the contention resolution algorithm defined by the IEEE 802.16 standard, jointly with our two new proposals. Through the dotted rectangles, we represent the parts of the flowchart which are different according to the method used in the contention resolution algorithm. The actions to performance (transmit and discard the PDU) are displayed using double-rounded boxes.

We use the label 'Standard' to identify the standard method, 'Last' to our proposal when it is used the last value of the backoff window, and, finally, 'TwoLast' when the two last values are considered. m and k are the power values to calculate the initial backoff window and the maximum backoff window, respectively. $RequestBackoffStart$ and $RequestBackoffEnd$ correspond to the $RequestBackoffStart$ and $RequestBackoffEnd$ fields, inside the UCD message. $LastBackoff$ is the last value used by the connection last time it sent a request, and this value has to be saved by the SS. $TwoLastBackoff$ is the penultimate value used in the backoff window. $ceil(h)$ is a function which returns the smallest integer value not less than the argument h , while the $random(i, j)$ function returns an integer number in the interval $[i, j]$, which is selected randomly. $TOpp$ is the number of transmission opportunities, being $topp()$ a function which returns the number of transmission opportunities in each frame. $T16$ indicates if the waiting temporal interval to receive a grant is reached, $Grant$ if there has been received a grant, and $Retries$ if the maximum number of retries has been reached.

V. PERFORMANCE EVALUATION

In this section, we carry out a performance analysis of our new proposals. Throughout our study, all simulations are

conducted using a model of the IEEE 802.16 implemented in the OPNET Modeler v11.5 tool [18].

In our simulations, we consider an IEEE 802.16 wireless network consisting of a BS and 7 or 12 SSs, describing a PMP system. 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. The system assumes ideal channel conditions and operates in a steady-state, where the number of connections does not change over time.

Each SS runs voice, video, best-effort (BE) and background (BG) applications, being each kind of application equal because we are interested in showing the behaviour of the two new proposals. The four kinds of applications are assigned to four different service flows, whose priorities in ascending order are: voice, video, best-effort and background. The number of applications a service flow is similar, using two applications per each service flow. Independently the service flow to which a connection belongs, every connection will request bandwidth through the contention period. This contention period size is fix through the simulation time and equal to 10 transmission opportunities.

Table I
TRAFFIC CHARACTERIZATION

	Voice - Video - BE - BG
Start (sec.)	uniform(0, 0.024)
ON period (sec.)	always
OFF period (sec.)	never
Interarrival time (sec.)	0.036
Packet size (bytes)	1250

The used mode in the simulation to request and grant bandwidth is the GPC mode, avoiding that the scheduling algorithm, inside each SS, influences on the behaviour of the sending data packet process.

The BS uses a FIFO (*First Input First Output*) scheduling algorithm to allocate the resources between every request it has received, according to its input time in the correspond queue.

Throughout our study, we have simulated 30 seconds of operation in each particular scenario. The figures are based on one simulation during the time of simulation.

We evaluate the performance of the contention resolution process defined by the standard and our proposals according to the GPC mode allowing us to fairly compare them.

Results

Figure 3 displays the uplink (UL) throughput. This throughput is equal independently of the method used when the number of SSs is 7 (see Figure 3(a)), because the system is not congested, reaching the maximum value. However, if SS=12 the difference is appreciated, like it is shown in Figure 3(b), which has been enlarged. In this case, the throughput is slightly improved with our proposals, because the system is becoming too congested and the waiting time to send a request is less

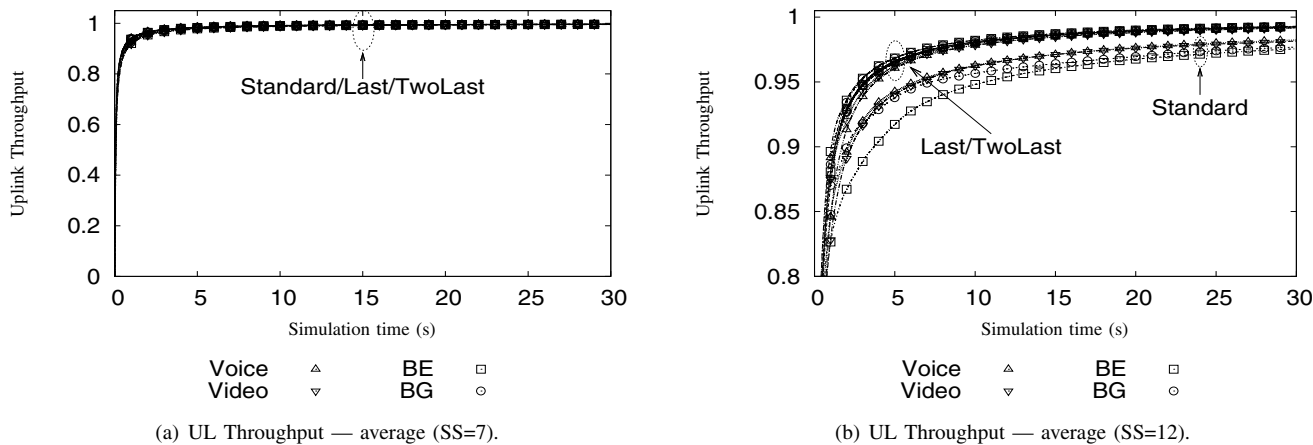


Figure 3. Uplink throughput — average.

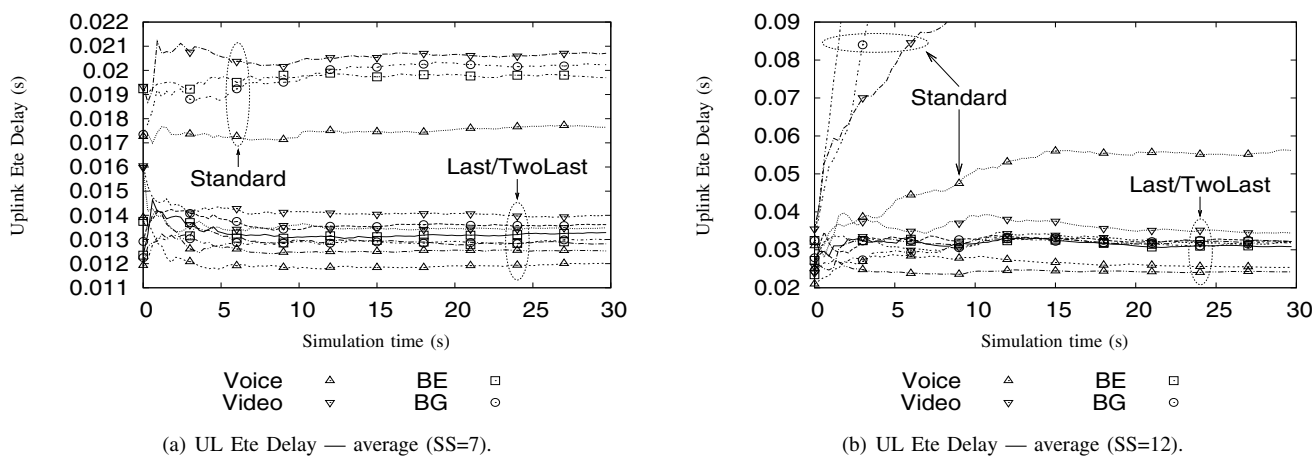


Figure 4. Uplink end-to-end delay — average.

in our proposals than in the standard method. The difference between our two proposals is indifferent.

In Figure 4 is depicted the uplink (UL) end-to-end (Ete) delay when the system is and isn't congested. For SS=7 (without congestion), Figure 4(a) shows two different zones. At the top, it is displayed the end-to-end delay for the standard method, while below this, the values of our two proposals are represented. So, our proposals minimize the end-to-end delay of all type of applications with an average of 5 ms, more or less. However, if the system is in a congested situation (SS=12), the difference increases strongly, like you can appreciate in enlarged Figure 4(b). The values for the standard method are increased quickly, while our proposals keep a constant behaviour. Between our proposals, it is appreciated a slightly decrease for the 'TwoLast' method.

Now to appreciate better the difference between the three used methods, we will display in the next figures the three methods, considering the two scenarios (SS=7 and SS=12), but it will be only represented the results to voice applications. For the rest of applications, the behaviour is the same because we

are using a FIFO scheduling algorithm to allocate the resources in the BS, without prioritizing any applications with respect to the rest of them.

In Figure 5(a) is depicted the uplink throughput, observing that the values for the standard method are always less than our proposals. The best behaviour, with a slightly difference, is for the 'TwoLast' method, although the obtained values can be considered equal. This characteristic is independently of the situation of system. According to the end-to-end delay, which is depicted on Figure 5(b), our proposals obtain a lower value with respect to the 'Standard' method. However, between our two proposal the difference values are in an interval very small, being able to consider them equal.

VI. CONCLUSION

The WiMAX standard defines a mandatory method of contention resolution based on a truncated binary exponential backoff algorithm controlled by the BS, which fixes the initial and maximum backoff windows, that is, the BS defines these values for every SS inside the network. But when the BS

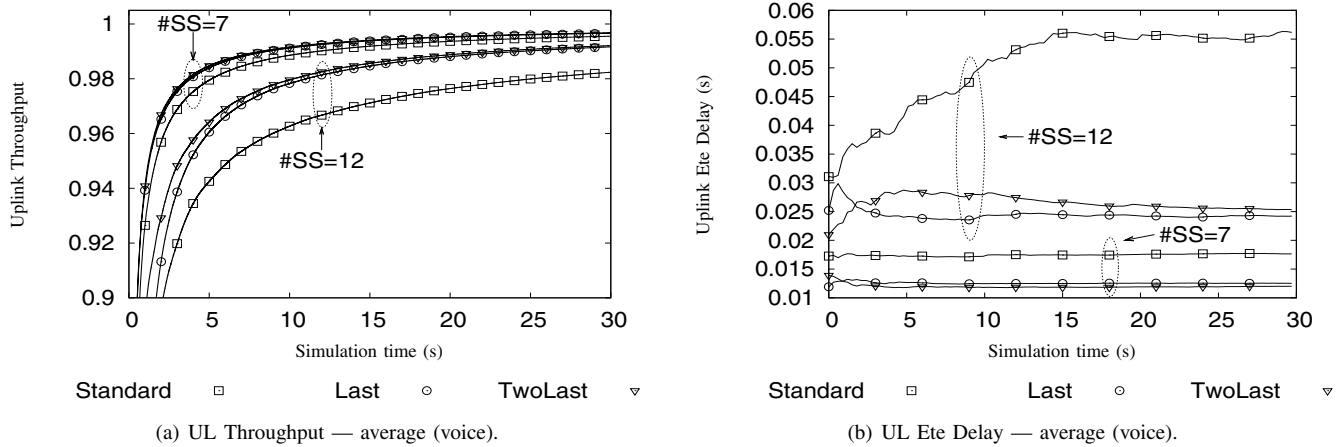


Figure 5. Uplink throughput and end-to-end delay — average (voice).

receives a request it doesn't know what the backoff window value is, which does it. For this reason, we propose to set the initial backoff window according to the previous or the two last values, which generated a successful request. The simulation results have demonstrated that our proposals improve significantly the performance of the system, reaching slightly values in the throughput, and significant differences in the end-to-end delay, independently the system was congested or not.

Our future work will check our proposals jointly with other methods of requesting bandwidth defined by the standard: piggyback and polling.

ACKNOWLEDGMENT

This work has been jointly supported by the Spanish MEC and European Commission FEDER funds under grants "Consolider Ingenio-2010 CSD2006-00046" and "TIN2006-15516-004-02"; by JCCM under project PAI06-0106 and grant 05/112; and by Brunel University BRIEF Award under grant LBK801.

REFERENCES

- [1] 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)*, IEEE Std., April 8 2002.
- [2] —, *IEEE 802.16 Standard—Local and Metropolitan Area Networks—Part 16: Air Interface for Fixed Broadband Wireless Access Systems (IEEE Std 802.16-2004)*, IEEE Std., October 1 2004, revision of IEEE Std 802.16-2001.
- [3] —, *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)*, IEEE Std., February 28 2006, amendment and Corrigendum to IEEE Std 802.16-2004.
- [4] N. Golmie, Y. Saintillan, and D. H. Su, "A Review of Contention Resolution Algorithms for IEEE 802.14 Networks," *IEEE Communications Surveys*, pp. 2–12, First Quarter 1999.
- [5] M. D. Corner, J. Liebeherr, N. Golmie, C. Bisdikian, and D. H. Su, "A Priority Scheme for the IEEE 802.14 MAC Protocol for Hybrid Fiber-Coax Networks," *IEEE/ACM Transactions on Networking*, vol. 8, no. 2, pp. 200–211, April 2000.
- [6] D. Sala, J. O. Limb, and S. Khaunte, "Performance of Contention Resolution Algorithms using Continuous-mode Operation," IEEE P802.14WG, Tech. Rep., March 1997.
- [7] D. Sala and J. O. Limb, "Comparison of Contention Resolution Algorithms for a Cable Modem MAC Protocol," in *Proceedings of International Zurich Seminar on Broadband Communications*, 1998, pp. 83–90.
- [8] B.-J. Kwak, N.-O. Song, and L. E. Miller, "Performance Analysis of Exponential Backoff," *IEEE/ACM Transactions on Networking*, vol. 13, no. 2, pp. 343–355, April 2005.
- [9] L. Romdhani, Q. Ni, and T. Turlitti, "Adaptive EDCF: Enhanced Service Differentiation for IEEE 802.11 Wireless Ad-Hoc Networks," in *Proceedings of WCNC'2003*, March 2003, pp. 1373–1378.
- [10] J. Lambert, B. V. Houdt, and C. Blondia, "Dimensioning the Contention Channel of DOCSIS Cable Modem Networks," in *Proceedings of Networking'2005*, vol. 3462, May 2005, pp. 342–357.
- [11] S.-M. Oh and J.-H. Kim, "The Analysis of the Optimal Contention Period for Broadband Wireless Access Network," in *Proceedings of PerCom'2005 Workshops*, March 2005, pp. 215–219.
- [12] —, "The Optimization of the Collision Resolution Algorithm for Broadband Wireless Access Network," in *Proceedings of ICACT'2006*, February 2006, pp. 1944–1948.
- [13] D.-H. Cho, J.-H. Song, M.-S. Kim, and K.-J. Han, "Performance Analysis of the IEEE 802.16 Wireless Metropolitan Area Network," in *Proceedings of DFMA'2005*, 2005, pp. 130–137.
- [14] J. Yan and G.-S. Kuo, "Cross-layer Design of Optimal Contention Period for IEEE 802.16 BWA Systems," in *Proceedings of ICC'2006*, vol. 4, June 2006, pp. 1807–1812.
- [15] A. Vinel, Y. Zhang, M. Lott, and A. Tiurlikov, "Performance Analysis of the Random Access in IEEE 802.16," in *Proceedings of PIMRC'2005*, vol. 3, September 2005, pp. 1596–1600.
- [16] A. Doha, H. Hassanein, and G. Takahara, "Performance Evaluation of Reservation Medium Access Control in IEEE 802.16 Networks," in *Proceedings of AICCSA'2006*, 2006, pp. 369–374.
- [17] Y. Ling and D. Meng, "Study on Improved Truncated Binary Exponential Back-off Collision Resolution Algorithm," *International Journal of Computer Science and Network Security (IJCSNS)*, vol. 6, no. 11, pp. 97–101, November 2006.
- [18] I. OPNET Technologies, "OPNET Modeler 11.5," 2005.