Study of the IEEE 802.16 Contention–based Request Mechanism *

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

Instituto de Investigación en Informática de Albacete (I^{3}A)
Universidad de Castilla–La Mancha (UCLM)
02071–Albacete, Spain
{jdelicado,franman,lorozco}@dsi.uclm.es

Abstract. Broadband wireless access systems offer a solution for broadband access and QoS–aware multimedia services through a wireless medium. The IEEE 802.16 standards specify the physical and medium access control layers for broadband wireless access systems as well as the various mechanisms to meet the quality of service (QoS) requirements of a wide variety of applications. Among the mechanisms defined by the standards, the bandwidth request and grant mechanisms play a central role on guaranteeing the QoS required by the subscriber stations.
In this paper, we undertake the study of the bandwidth request mechanisms defined in the IEEE 802.16 standards. We then propose an approach for reducing the overhead required by the signalling mechanisms. Our simulation results show that our approach outperforms the mechanisms proposed in the IEEE 802.16 standards.

Keywords: TDMA/FDD Resource Request, WiMAX.

1 Introduction

The design and adoption of broadband wireless access (BWA) systems is one of the most significant networking research and development activities nowadays. BWA systems provide fixed–wireless access to SS —Subscriber Station— (residential or business customers) to internet service provider (ISP) facilities. Their main advantages are: rapid deployment, scalability, low maintenance and upgrade costs. Their installation can be particular useful in (a) very crowded geographical areas such as urban areas or in (b) rural areas lacking broadband wired infrastructure. The IEEE 802.16 standards have been particularly designed for BWA systems. The IEEE 802.16 standards support a point–to–multipoint (PMP) architecture within the 10–66 GHz range of frequencies, achieving about

---

50 km of distance. Different products have been developed in the past few years [**?**].

The IEEE 802.16 medium access control (MAC) protocol is a centralized connection–oriented mechanism. Under this protocol, the BS is responsible of allocating the bandwidth required by the SSs. Furthermore, the BS has to classify, prioritize and schedule the SSs requests. Towards this end, the BS has to timely and efficiently manage the overall available bandwidth. The signalling protocol is therefore a central element allowing the SSs to place their requests and the BS to issue the grants to the SSs.

Regarding the IEEE 802.16 signalling protocols, the standard specifies that a SS has first to request the bandwidth according to the needs of each one of the individual connections associated to the SS. However, the requests can be granted using one of two modes: GPC (*Grants Per Connection*) or GPSS (*Grants Per Subscriber Station*). Under the GPC mode, grants are made to individual connections. The GPSS mode instead grants the bandwidth to the SS without specifying a particular connection. Under the latter mode, the responsibility of allocating the granted bandwidth is left to the SS.

In this paper, we develop and analyze a new algorithm implementing the request bandwidth primitives available to the SSs. Our proposal aims to improve the performance of the modes proposed by the standards in terms of three metrics, namely the overhead, throughput and delay.

The paper is organized as follows. Section 2 provides an overview of the IEEE 802.16 standard, which includes a brief description of the MAC protocol and request and grant bandwidth mechanisms. In Section 3, we analyze the relevant literature closely related to our proposal. Section 4 describes our proposal and in Section 5, we carry out a comparative performance study of our proposal with respect to the modes included in the standards. Finally, Section 6 concludes the paper.

2   IEEE 802.16

The IEEE 802.16 standards specify the physical (PHY) and MAC layer of the air interface of interoperable point–to–multipoint and optional Mesh topology BWA systems.

The range of frequencies supported by IEEE 802.16 is from 2 to 66 GHz, which includes the licensed and license–exempt bands. Line–of–sight (LOS) is sometimes necessary depending on the range of frequencies. 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. The bit rate and robustness in the presence of errors depend on the type of modulation and frequency being used.

2.1   MAC layer

As already mentioned, the MAC layer is a centralized and connection–oriented mechanism, that is, the BS is responsible of allocating the resources and provi-

sioning the system with QoS–aware mechanisms. To perform this task, the SSs are required to request to the BS their needs in a timely manner. That is to say, each SS has to request to the BS the required bandwidth on a frame by frame basis. In response to the resources required by all the SSs, the BS distributes the available bandwidth taking into account the requirements of all the outstanding SSs requests.

The communication between the BS and the SSs is realized by means of fixed–length frames, divided into two subframes: the downlink and uplink subframes, whose lengths are dynamically controlled by the BS. The mode of operation can be Frequency Division Duplexing (FDD) or Time Division Duplexing (TDD). The downlink and uplink communications are time multiplexed by means of a Time Division Multiple Access (TDMA) mechanism.

The downlink subframe starts with a Frame Start Preamble used by the PHY for synchronization and equalization. This is followed by the *frame control section*, which is composed of management messages. It contains the downlink and uplink maps stating the physical slots (PSs) at which bursts begin, that is to say, they comprise the bandwidth allocations for the data transmission in the downlink and uplink directions, `DL-MAP` and `UL-MAP` messages, respectively.

The `DL-MAP` contains the addresses of the first time slots being used to convey the data transmitted by the BS and its correspondent downlink burst profile (Downlink Interval Usage Code, DIUC).

The `UL-MAP` contains the specific data (Information Element, IE) which include the transmission opportunities, that is to say, the time slots at which each an every active SS can transmit. After receiving the `UL-MAP` message, the SSs transmit their data in predefined time slots as indicated in the IE. The BS scheduling module determines the transmission opportunities (IEs) using the bandwidth request (`BW-REQ` message) sent by the SSs to BS. The size of this message affects the remaining size of the downlink frame, that is to say, the longer this message, the less available bandwidth for the transmission of data over the downlink direction and vice versa. However, the `UL-MAP` size will increase as the number of grants in the uplink direction increases.

Following these maps, the `DCD` and `UCD` messages indicate the physical characteristics of the physical channels. Finally, it is introduced a TDM portion which carries data, organized into bursts depending on the negotiated burst profile between the BS and the SS, and therefore different levels of transmission robustness. Bursts are transmitted in order of decreasing robustness.

In the FDD case, the TDM portion may be followed by a TDMA portion used for transmitting data to any half–duplex SSs scheduled to transmit earlier than in the frame they are receiving data. Each burst begins with the Downlink TDMA Burst Preamble to regain synchronization. In this case, the `DL-MAP` message includes a map of, both, the TDM and TDMA bursts.

The structure of the uplink subframe is divided into three classes of bursts:

1. Contention opportunities reserved for Initial Maintenance. The `RNG-REQ` messages are transmitted by the SSs joining the network.

2. Contention opportunities defined by the Request Intervals. These are reserved to convey the replies to the multicast and broadcast polls, in which the transmitted `BW-REQ` messages are placed. These messages are issued by the SSs to indicate to the BS their needs.
3. Intervals defined by Data Grant IEs specifically allocated to individual SSs.

The bandwidth allocated for Initial Maintenance and Request contention opportunities may be grouped together and is always used with the uplink burst profiles. The remaining transmission slots are grouped by SS. The `UL-MAP` message, in the downlink subframe, grants bandwidth to specific SSs and indicates the uplink burst profile assigned to transmit. These data are used by each SS to transmit in its assigned allocation. These transmissions are separated by SS Transition Gaps in order to properly synchronize to the SS.

2.2 Requests

A request refers to the mechanism that SSs use to indicate to the BS that they need uplink bandwidth allocation. This request may come as a stand–alone bandwidth request header (`BW-request`) or as a piggyback request.

The IEEE 802.16 standard uses a random access mechanism in the uplink subframe for transmitting the opportunity requests to the SSs, during the request contention period defined into this subframe. The BS is responsible of establishing this reservation period at the beginning of each subframe. In this way, the SSs can place their reservations (`BW-requests`) for transmitting in the next subframe (or later, depending on the happening or not of collisions). The standard defines the *truncated binary exponential backoff* algorithm as the mechanism for resolving potential conflicts during this interval.

To limit the length of the contention resolution period is one of the main issues to be addressed. The longer the contention resolution period, the shorter the available capacity left for transmitting data.

According to the IEEE 802.16 standard, the requests have to be issued on an individual basis, that is, a SS has to issue a request for each connection associated to it. Each request has to be identified by a connection identifier denoted from now on as CID (Connection IDentifier).

The use of piggyback requests is out of the scope of this paper, because we are interested in improving the contention period in which it is only possible to make `BW-requests` using the backoff mechanism.

2.3 Grants

According to the IEEE 802.16 standards, the grants can be issued according to two schemes. The first version of IEEE 802.16 standard has defined a mechanism to grant bandwidth on a connection by connection basis, where each connection is associated to its corresponding SS. This mechanism has been named *Grants Per Connection* (GPC). The grant is associated to the connection by explicitly indicating on the grant the connection identifier, CID. In the second approach,

the grants are issued on a station basis (*Grants Per Subscriber Station*, GPSS). Here, the bandwidth is granted to the SS and not explicitly to an individual CID. Each SS is then responsible of allocating the received opportunities among its different types of service flows (applications).

It is obvious that the GPC method requires more bandwidth to convey the grants messages, since it is necessary to individually indicate the bandwidth to each connection. In the GPSS method, only one grant per SS is sent to the concerned SS. There is therefore a trade–off to be considered between the amount of signalling traffic and the processing required for effectively distributing the bandwidth among the competing connections.

3 Related work

Many research efforts on the performance of MAC protocols have been carried out and have been reported in the literature. In [?], authors made an analysis of different parameters involved in the truncated binary exponential backoff algorithm to resolve collisions produced in the system. The optimal contention period has been studied in [?], according to the number of users under definition of a cost function where the channel throughput and delay of the system have been considered the two essential metrics. Finally, the authors have come to the conclusion that the optimal size of the contention period is $2M - 1$, where M is the number of SSs (users). In [?], the authors have proposed a new QoS architecture for IEEE 802.16 and have obtained the size of contention period accordingly to maximize the throughput. In this case, the result is a contention period equal to the number of competing SSs. In both papers, the results highly depend on the number of SSs, so the remaining uplink subframe to send data is smaller each time the number of SSs increases. An adaptive bandwidth request mechanism implemented at the SSs for real time traffic is proposed in [?]. In this approach, each SS predicts the arrival of real time packets prior to their arrival and requests bandwidth in advance. In [?], the authors have introduced a dynamic minislot allocation scheme. The number of required contention minislots in each frame period is based on an estimate of the maximum number of data packets that can be transmitted in a frame. The study formally analyzes the proposal, but it does not include numerical results. [?] introduces a new MAC protocol to prioritize traffic according to the waiting times required by the different types of traffic. The requests of the different types of traffic are assigned priorities. In [?], the authors introduce a new algorithm, 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.

In all the studies having been reported, the authors make use of the mechanisms defined by the IEEE 802.16 standard to request and grant bandwidth. However, they do not make an in–depth analysis of the overhead. The overhead increases as the contention period increases. A simple modification can be used
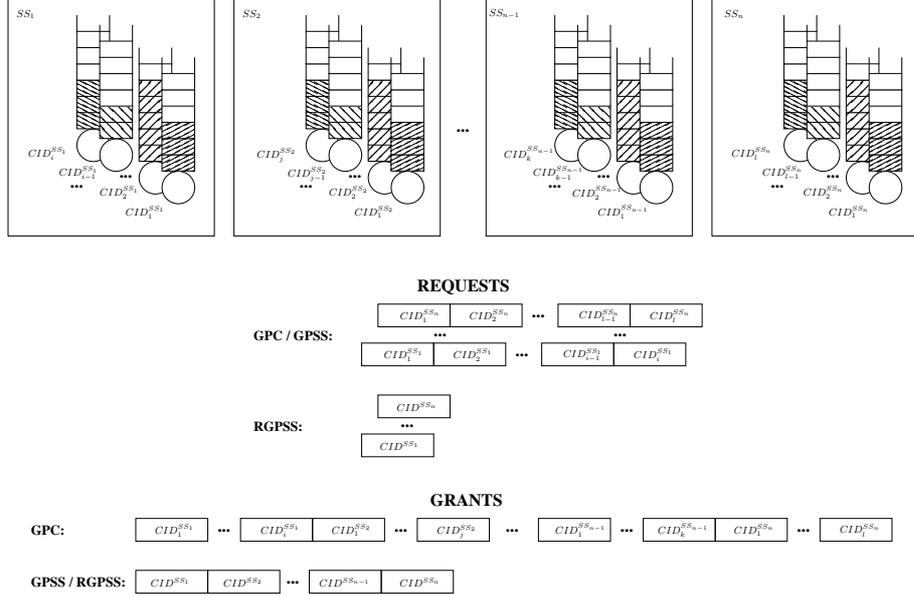
Fig. 1. Comparing GPC, GPSS & RGPSS modes regarding requests and grants.

to improve overhead in the system introduced by these mechanisms, especially requesting mechanism defined in the standard. This modification, which is explained below, will have a clear impact on the size of contention period.

4 Our proposal

As it has been described above, the IEEE 802.16 standard specifies a bandwidth request per connection made by SSs utilizing the corresponding CID. In this way, each SS has to make a request per active connection, requesting the amount of enqueued data associated to this connection. A drawback of this resource request scheme is that the internal collisions increase with the number of connections (an internal collision happens when two or more requests of different connections within a SS try to use the same contention slot). The other one is that to keep the probability of collision up when the number of connections increases, the contention period must increase too, reducing the size of uplink frame to be used for data transmission. This is derived from [?] and [?] when the number of connections per SS is bigger than one.

Regarding the grant bandwidth mechanisms, the standard describes two methods. The GPC, which grants bandwidth per connection, and the GPSS, which grants per subscriber station. Obviously the GPSS mode improves on the use of the frame, because the overhead in downlink is reduced, sending only a grant message in the DL-MAP per SS.

In the case of using the GPSS mode to grant resources, the BS does not need to know the individual request of each connection, since the grants are made by SS. In this case, the BS only needs to know the aggregated bandwidth request of all connections of each SS. To obtain this information, the BS can add up the all connection requests belonging to each SS. Obviously, this requires each connection to send its request, increasing the amount of request messages and therefore the overhead in the uplink.

In order to reduce the request messages in the uplink, we propose that each SS should send a single aggregated request containing the total amount of resources needed by all the active connections in this SS. This requesting scheme, which is called RGPSS (*Request and Grants Per Subscriber Station*), tries to minimize the number of requests to be sent, reducing the overhead in the uplink phase. A direct result of the reduction in the number of request messages is a decrease in the collision probability in the contention period, increasing the efficiency of the resource request mechanism based on a random access method. This operation mode reduces the length of the contention period since it only depends on the number of SSs.

The RGPSS scheme could be adapted to support classifications of connections in order to provide QoS. In this case, each SS will make a request per type of service flow, aggregating the resources that all the connections of a service flow need. So the BS will know bandwidth needs of each type of service flow of all SSs, and its scheduler could prioritize some service flows or SSs against another one.

5 Performance Evaluation

In this section, we carry out a performance analysis of our proposal. Throughout our study, all simulation are conducted using a model of IEEE 802.16 implemented in the OPNET Modeler v11.5 tool [?].

5.1 Scenarios

In our simulations, we consider an IEEE 802.16 wireless network consisting of several SSs and a BS describing a point–to–multipoint 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 Mbits/s. Accordingly to the standard a frame duration of 1 ms is used. The mode of operation is FDD. Ideal channel conditions are assumed, i.e., no packet corruption is due to the wireless channel. The system is assumed to operate in a steady–state, where the number of connections does not change over time. Each SS runs voice, video, background and best–effort applications, which are modelled as follows.

Constant bit–rate voice sources at a rate of 16 Kbits/s according to the G.728 standard are used. The voice packet size has been set to 384 bytes including RTP/UDP/IP headers, and all voice traffic are randomly activated within

the interval [0,0.024] seconds, corresponding to a generating frequency of two consecutive packets. Each SS runs 37 voice connections.

For video applications, we assume H.264 variable bit–rate video traffic. This one is generated from the sequence *funny* encoded on CIF format at a frame rate of 30 frames/s are used. The average video transmission rate is around 1.1 Mbits/s with a packet size equal to 1,064 bytes (including RTP/UDP/IP headers). This type of traffic starts within the interval [0,0.5] seconds following a uniform random distribution. For each SS, we limit to one the number of video connection per SS.

For the background and best–effort traffics, we consider a *Pareto* distribution traffic model with an average bit rate for both types of traffic of 256 Kbits/s. The packet size has been set to a 552 bytes packet, including the TCP/IP headers. The sources are activated randomly within the interval [1,1.5] seconds. Associated to each SS, we assume six background connections and ten best–effort connections.

Since we are interested in studying the behaviour of the system with the new bandwidth request mechanism, all connections will request bandwidth utilizing the contention period. The scheduling algorithms used by the BS and SSs is FIFO (*First In, First Out*).

Throughout our study, we have simulated 10 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 28 simulation runs, and error bars indicate 95% confidence interval. The study is carried out by varying the number of SSs (2, 5, 8 and 10) in order to increase the total offered traffic. The size of the contention period (#TOpp) is increased by 2 from 4 to 10 transmission opportunities. We evaluate the performance of the GPC, GPSS and RGPSS mechanisms allowing us to fairly compare them.
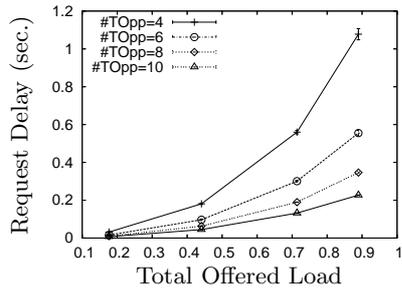
5.2   Results

In the case of the GPC mechanism, the collision probability is shown in Figure 2.(a). As expected, the collision probability decreases as the number of transmission opportunities increases. Moreover it increases as the network load increases. As a result, the request delay, defined as the time elapsed between the creation of a request in a SS and its arrival to the BS, is higher when the contention period size is shorter or when the traffic load offered increases. This is clearly depicted in Figure 2.(b). Since in GPC, the requests are sent per connections and the bandwidth grants are granted in the same manner; the grant delay (Figure 2.(c)), which is defined as the time elapsed between the first request of a connection and its associated grant, behaves in a similar manner to the request delay.

The results for the metrics described above in GPSS mode are shown in Figure 3. From the results, it is clear that the collision probability increases as a function of the traffic load offered. In the same way than in the GPC mode, if the contention period size decreases, the probability increases (Figure 3.(a)). The request delay is shown in Figure 3.(b), similar to the GPC mechanism,
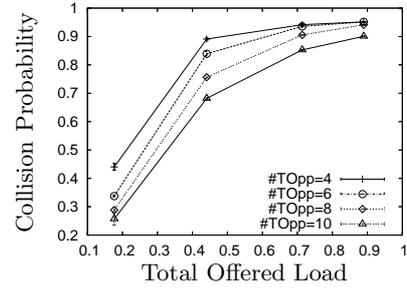
(a) Collision Probability.

(a) Collision Probability.
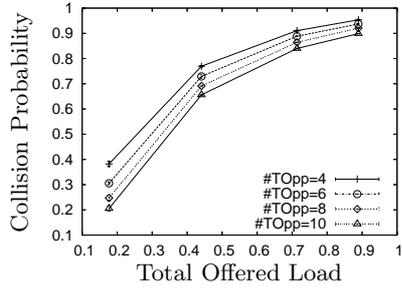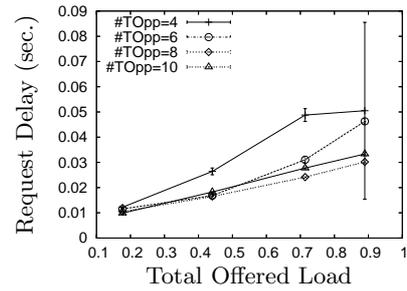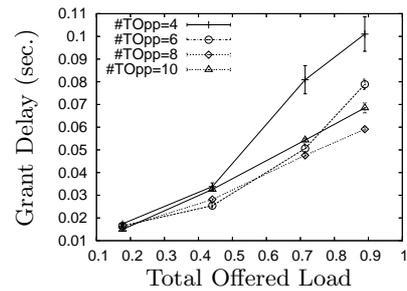

(b) Request Delay.

(b) Request Delay.
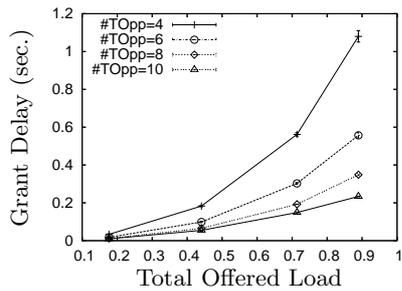
(c) Grant Delay.

(c) Grant Delay.

Fig. 2. Results in GPC mode.   **Fig. 3.** Results in GPSS mode.

it is smaller if the number of contention opportunities increases, except in the case of 8 and 10 transmission opportunities where this trend changes. This can be explained by the fact that for a longer contention period, more capacity is used by the resource request messages resulting in a reduction on the available capacity dedicated to data transmission. This means that the BS spends more time to serve a request. This can be clearly seen in Figure 3.(c). This increase on the time required to send a grant results on the expiration of the request timer. This event results in turn on the retransmission of the request. As a result of this retransmission, the request delay increases because upon receiving the request,
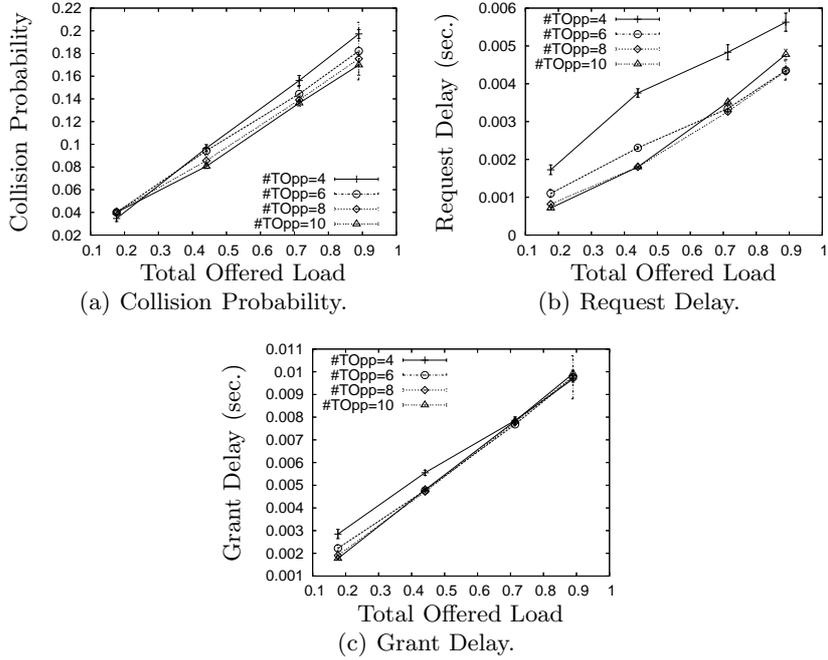
(a) Collision Probability.

(b) Request Delay.

(c) Grant Delay.

Fig. 4. Results in RGPSS mode.

the BS considers that it is the original request. Obviously, its creation time is the creation time of the original request.

In the RGPSS mechanism, the collision probability increases linearly as a function of the load (Figure 4.(a)), regardless of the contention period values. This is because in RGPSS only a request is sent per SS, aggregating all the connection bandwidth requests of the SS. Similar behavior is depicted in Figure 4.(b), where the request delay is represented. Not big difference could be appreciated in the case of 6, 8 and 10 contention opportunities, and only in the case of using 4 opportunities for contention, the delay is higher than 1 ms, which could be considered as acceptable since it matches the period of a MAC frame. The grant delay, shown in Figure 4.(c), presents a similar trend. It increases with the injected traffic and there are not big differences in the use of longer contention periods.

We now compare the performance of the three request and grant mechanisms. According to the previous results, the best results limiting the length of the contention period for each mechanism were: 10 transmission opportunities for the GPC; 8 for GPSS and 6 for RGPSS.

Figure 5.(a) depicts the collision probability. In it, the overall best result is provided by the RGPSS mechanism, proving effective the reduction of requests sent by the SSs.
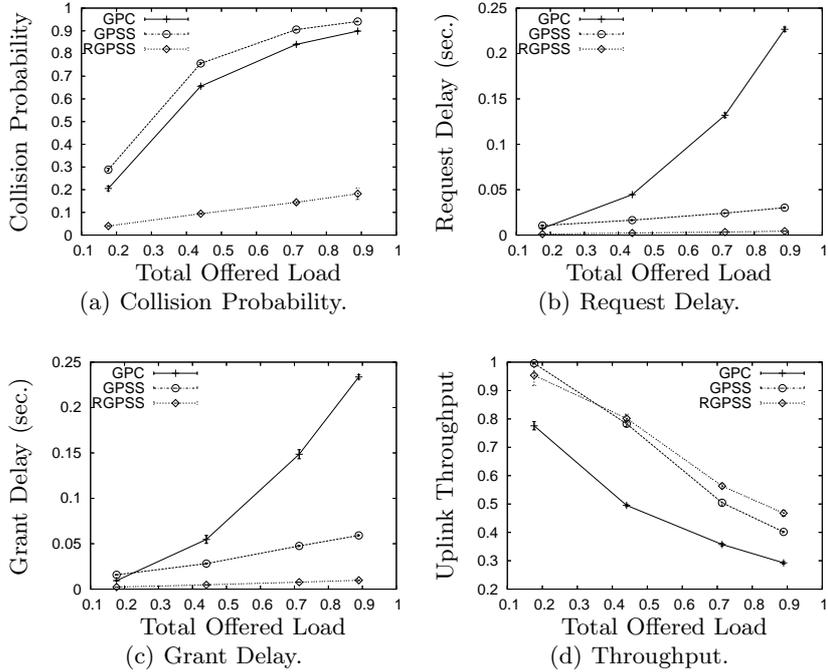
Fig. 5. Comparison between GPC, GPSS & RGPSS.

Regarding the request delay and grant delay (Figures 5.(b) and 5.(c), respectively), the difference between the mechanisms under study is clearly appreciated. This is particularly true as the network load is increased. The RGPSS exhibits the best behavior due to the limited number of requests.

As a result on the reduction in the grant delay in RGPSS, this mechanism presents a higher throughput in uplink direction, as seen in Figure 5.(d). This results in a lower collision probability, more requests arrive to the BS resulting on a shorter request delay. The BS is able to grant more connections in a shorter period of time. Other reason for this behavior is that contention period size in the RGPSS mechanism is shorter. In conclusion, more capacity is available to deliver actual data.

6 Conclusions

In this paper, a new mechanism to request in IEEE 802.16 network, called RGPSS (*Requests and Grants Per Subscriber Stations*), has been proposed. It sends only one request per SS, carrying the aggregated bandwidth requirements of all its associated connections.

It has been shown that the proposal exhibits a lower collision probability in the contention period and a higher throughput in the uplink direction than

the schemes defined in the standard. These results demonstrate that with the RGPSS mechanism the network can support more traffic. It has also been shown that the contention period size used by RGPSS is shorter than the one required by the other two mechanisms.

This mechanism could be easily adapted to support different types of traffic (connections). It could be used as a means to provision a network with QoS support. In future works, we plan to evaluate the performance of such mechanisms and the impact of other parameters heavily impacting the length of the contention period.