QoS requirements for bandwidth request and allocation in WiMAX Networks

Arthur Joseph Kweku
St. Petersburg State university of Telecommunication
Russia, St. Petersburg
E-mail: ajkweku@yahoo.com

Korotin Vladimir Evgenevich
St. Petersburg State university of Telecommunication
Russia, St. Petersburg
E-mail: vekorotin@sut.ru
Abstract. Allocation of contention slots (CS) is an open area for research in WiMAX networks. This is because the wimax standard does not specify the amount of bandwidth to be used in the uplink for contention. In this article’s model for evaluating the effect of CS allocation on access delay has been proposed using normalized throughput and number of subscriber stations to evaluate how the number of contention slots affect access delay. It has been clearly shown that increase in number of contention slots has an effect on access delay at the expense of system throughput, therefore the number of CS used for granting bandwidth request affects the QoS of the network. In the analysis, it has been shown that increase of CS more than 2000CS has no meaningful influence on access delay. This was validated by simulation using Matlab 2012. Network operators could use these values in the planning of WiMAX networks.

Keywords: WiMAX; Bandwidth request; contention slot; access delay; Quality of service; scheduling algorithm; subscriber station.
1 Introduction

The WiMAX standard specifies a metropolitan area broadband wireless access air interface.

In order to support quality of service (QoS) for multimedia applications, various bandwidth request and scheduling mechanisms are suggested in WiMAX, in which a subscriber station can send request messages to a base station, and the base station can grant or reject the request according to the available radio resources (Ni, Vinel, Xiao, Turlikov, & Jiang, 2007). 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 describes devices which are the origin and destination of connections (Ahmadi, 2010). Because the BS allocates the resources, it has to know the needs of each connection in the system to allocate 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 (Delicado, Ni, Delicado, & Orozco–Barbosa, 2009). At peak times, the requesting process determines the network performance, because when the number of request is large, the probability of the requests colliding is higher and this affects the data being transmitted. The WiMAX standard specifies a fixed uplink frame but does not specify the amount of physical slots to be used for contention or data transmission (Nuaymi, 2007). This leaves a dilemma, because when the number of contention slots is increased there will be lesser collisions of bandwidth request (BR) meaning more request will be received by the base station. However, because the uplink frame is fixed, this will affect the available bandwidth to transmit data packets. Therefore, the question arises: Which ratio of contention slot to data slots makes use of the radio resources efficiently? The objective of this paper is to investigate and propose a ratio that will enable an efficient use of the networks radio resources.

2 Literature Review

2.1 UL PHY

2.1.1 UL subframe

The structure of the UL subframe used by the SS to transmit to the BS is shown in Figure 1. The following three classes of bursts may be transmitted by the SS during the UL subframe:

a) Those that are transmitted in contention opportunities reserved for initial ranging.

b) Those that are transmitted in contention opportunities defined by Request Intervals reserved for response to multicast and broadcast polls.

c) Those that are transmitted in intervals defined by Data Grant IEs specifically allocated to individual SSs.
Figure 1. UL subframe structure

Any of these burst classes may be present in any given frame. They may occur in any order and any quantity (limited by the number of available PSs) within the frame, at the discretion of the base station uplink scheduler (BS UL) as indicated by the UL_MAP in the frame control section (part of the DL subframe).

The bandwidth allocated for initial ranging and request contention opportunities may be grouped together and is always used with the uplink (UL) burst profiles specified for initial ranging intervals (UIUC = 2) and request intervals (UIUC = 1), respectively. The remaining transmission slots are grouped by the SS. During its scheduled bandwidth, an SS transmits with the burst profile specified by the BS. SSTGs separate the transmissions of the various SSs during the UL subframe. The gap allows for ramping down of the previous burst, followed by a preamble allowing the BS to synchronize to the new SS. The preamble and gap lengths are broadcast periodically in the UL channel descriptor (UCD) message. (The Institute of Electrical and Electronics Engineers, 2012).

2.1 Bandwidth allocation and request mechanisms

During network entry and initialization, every SS is assigned up to three dedicated connection identifier (CIDs) for the purpose of sending and receiving management messages. These connection pairs are used to allow differentiated levels of QoS to be applied to the different connections carrying medium access control (MAC) management traffic. Increasing (or decreasing) bandwidth requirements is necessary for all services except unsolicited grant service (UGS) connections (Taha, Ali, & Hassanein, 2011). The needs of UGS connections do not change between connection establishment and termination. When an SS needs to ask for bandwidth on a connection with best effort (BE) scheduling service, it sends a message to the BS containing the immediate requirements of the connection. QoS for the connection was established at the connection setup and is looked up by the BS. The additional methods are described in the following sections.

2.2.1 Requests

Requests refer to the mechanism that SSs use to indicate to the BS that they need UL bandwidth allocation. A Request may come as a stand-alone BR header or it may come as a PiggyBack Request (e.g., Grant management subheader). The use of Grant management subheader is optional. Because the UL burst profile can change dynamically, all requests for bandwidth shall be made in terms of the number of bytes needed to carry the MAC PDU excluding PHY overhead. The BR message may be transmitted during any UL allocation, except during any initial ranging interval. An SS shall not request bandwidth for a connection if it has no protocol data unit (PDU) to transmit.
on that connection. BRs may be incremental or aggregate. When the BS receives an incremental BR, it shall add the quantity of bandwidth requested to its current perception of the bandwidth needs of the connection. When the BS receives an aggregate BR, it shall replace its perception of the bandwidth needs of the connection with the quantity of bandwidth requested. The Type field in the BR header indicates whether the request is incremental or aggregate. Since Piggybacked BRs do not have a type field, Piggybacked BRs shall always be incremental. The self-correcting nature of the request/grant protocol requires that SSs may periodically use aggregate BRs as a function of the QoS of a service and of the link quality. Due to the possibility of collisions, contention-based BRs shall be aggregate requests except in the OFDMA PHY. In the OFDMA PHY, the SS may respond to the CDMA Allocation information element (IE) with either aggregate or incremental BR.

Capability of incremental BRs is optional for the SS and mandatory for the BS. Capability of aggregate BRs is mandatory for SS and BS.

In OFDMA, the bandwidth request is to be interpreted by the BS as the amount of data that the SS requires for a connection after the SS has sent the data that is in the current burst.

The bandwidth request message, mechanism, and capability defined above for the SS and BS shall be applicable for the RS and the scheduling station respectively. Capability of incremental BRs is only mandatory if the RS is a scheduling RS (Korowajczuk, 2011).

2.3 MAC support of PHY

Several duplexing techniques are supported by the MAC protocol. The choice of duplexing technique may affect certain PHY parameters as well as impact the features that can be supported.

2.3.1 Frequency division duplexing (FDD)

In an FDD system, the UL and DL channels are located on separate frequencies and the DL data can be transmitted in bursts. A fixed duration frame is used for both UL and DL transmissions. This facilitates the use of different modulation types. It also allows simultaneous use of both full-duplex SSs (which can transmit and receive simultaneously) and optionally half-duplex SSs (which cannot). If half-duplex SSs are used, the bandwidth controller shall not allocate UL bandwidth for a half-duplex SS at the same time that it is expected to receive data on the DL channel, including allowance for the propagation delay, SS transmit/receive transition gap (SSTTG) and SS receive/transmit transition gap (SSRTG).

Figure 2 describes the basics of the FDD mode of operation. The fact that the UL and DL channels utilize a fixed duration frame simplifies the bandwidth allocation algorithms. A full-duplex SS is capable of continuously listening to the DL channel, while a half-duplex SS can listen to the DL channel only when it is not transmitting in the UL channel.
2.3.2 Time division duplexing (TDD)

In the case of TDD, the UL and DL transmissions occur at different times and usually share the same frequency. A TDD frame (see Figure 3) has a fixed duration and contains one DL and one UL subframe. The frame is divided into an integer number of physical slots (PSs), which help to partition the bandwidth easily. The TDD framing is adaptive in that the bandwidth allocated to the DL versus the UL can vary. The split between UL and DL is a system parameter and is controlled at higher layers within the system (The Institute of Electrical and Electronics Engineers, 2012).

**Figure 2. Example of burst FDD bandwidth allocation**

**Figure 3. TDD frame structure**

[*] for single carrier, the Rate is the Symbol Rate; for OFDM, OFDMA, the Rate is the nominal sampling frequency (Fs)
2.4 Contention resolution

The BS controls assignments on the UL channel through the UL-MAP messages and determines which minislots are subject to collisions. Collisions may occur during initial ranging and request intervals defined by their respective IEs. The potential occurrence of collisions in request intervals is dependent on the CID in the respective IE. This subclause describes UL transmission and contention resolution. For simplicity, it refers to the decisions an SS makes. Since an SS can have multiple UL service flows (each with its own CID), it makes these decisions on a per CID or per service QoS basis. The mandatory method of contention resolution that shall be supported is based on a truncated binary exponential backoff, with the initial backoff window and the maximum backoff window controlled by the BS. The values are specified as part of the UCD message and represent a power-of-two value. For example, a value of 4 indicates a window between 0 and 15; a value of 10 indicates a window between 0 and 1023. When an SS has information to send and wants to enter the contention resolution process, it sets its internal backoff window equal to the request (or ranging for initial ranging) backoff start defined in the UCD message referenced by the UCD Count in the UL-MAP message currently in effect. The SS shall randomly select a number within its backoff window. This random value indicates the number of contention transmission opportunities that the SS shall defer before transmitting. An SS shall consider only contention transmission opportunities for which this transmission would have been eligible. These are defined by Request IEs (or Initial Ranging IEs for initial ranging) in the UL-MAP messages. Note that each IE may consist of multiple contention transmission opportunities. Using BRs as an example, consider an SS whose initial backoff window is 0 to 15 and assume it randomly selects the number 11. The SS shall defer a total of 11 contention transmission opportunities. If the first available Request IE is for 6 requests, the SS does not use this and has 5 more opportunities to defer. If the next Request IE is for 2 requests, the SS has 3 more to defer. If the third Request IE is for 8 requests, the SS transmits on the fourth opportunity, after deferring for 3 more opportunities. After a contention transmission, the SS waits for a Data Grant Burst Type IE in a subsequent map (or waits for a RNG-RSP message for initial ranging). Once received, the contention resolution is complete. (Mavromoustakis, Pallis, & Mastorakis, 2014).

The SS shall consider the contention transmission lost if no data grant has been received in the number of subsequent UL-MAP messages specified by the Contention-Based Reservation Timeout parameter (or no response within T3 for initial ranging). The SS shall now increase its backoff window by a factor of two, as long as it is less than the maximum backoff window. The SS shall randomly select a number within its new backoff window and repeat the deferring process described above. This retry process continues until the maximum number (i.e., request retries for BRs and contention ranging retries for initial ranging) of retries has been reached. At this time, for BRs, the PDU shall be discarded. Note that the maximum number of retries is independent of the initial and maximum backoff windows that are defined by the BS.

For BRs, if the SS receives a unicast Request IE or Data Grant Burst Type IE at any time while deferring for this CID, it shall stop the contention resolution process and use the explicit transmission opportunity.

The BS has much flexibility in controlling the contention resolution. At one extreme, the BS may choose to set up the request (or ranging) backoff start and request (or ranging) backoff end to emulate an Ethernet-style backoff with its associated simplicity and distributed nature as well as its fairness and efficiency issues. This would be done by setting request (or ranging) backoff start = 0 and request (or ranging) backoff end = 10 in the UCD message. At the other end, the BS may make the request (or ranging) backoff start and request (or ranging) backoff end identical and frequently update these values in the UCD message so that all SS are using the same, and hopefully optimal, backoff window.
2.4.1 Transmission opportunities

A transmission opportunity is defined as an allocation provided in a UL-MAP or part thereof intended for a group of SSs authorized to transmit BRs or initial ranging requests. This group may include either all SSs having an intention to join the cell or all registered SSs or a multicast polling group. The number of transmission opportunities associated with a particular IE in a map is dependent on the total size of the allocation as well as the size of an individual transmission. The size of an individual transmission opportunity for each type of contention IE shall be published in each transmitted UCD message. The BS shall always allocate bandwidth for contention IEs in integer multiples of these published values. (Wei, Rykowski, & Dixit, 2013).

As an example, consider contention-based BRs for a WirelessMAN-SC system where the PHY protocol has a frame duration of 1 ms, 4 symbols for each PS, 2 PSs for each minislot, an UL preamble of 16 symbols (i.e., 2 minislots), and an SS transition gap (SSTG) of 24 symbols (i.e., 3 minislots). Thus, assuming quadrature phase-shift keying (QPSK) modulation, each transmission opportunity requires 8 minislots: 3 for the SSTG, 2 for the preamble, and 3 for the BR message. This payload requirement would be specified as a value of 16 assigned to the UCD TLV “Bandwidth request opportunity size”.

If the BS schedules a Request IE of, for example, 24 minislots, there will be three transmission opportunities within this IE. Details of the three transmission opportunities are shown in Figure 4.

![Figure 4. Example of Request IE containing multiple transmission opportunities](image-url)
3 Methodology

3.1 Flowchart of the algorithm

The flowchart (figure 5) explains the algorithm used to investigate the effective allocation of contention slots. At a time $t=0$ the number of contention slots is set to 1000 i.e. the ratio of contention slots to data slot is 1:3 respectively. Taking into account the frame size used for this analysis of the uplink frame. A delay access threshold is then set to maintain an allowable access delay. This access delay constant is set to maintain the QoS in the system.

The number of BR is then increased by a margin of one. This is done by increasing the number of SS. It is assumed that each SS is sending a bandwidth request. At each instance delay access is calculated to make sure the delay in the system is not above the threshold. Throughput is also calculated at each instance. Immediately the delay in the system reaches the threshold the ratio of contention slot to data slot is changed to 1:1 by increasing the number of contention slots by a thousand. The process of calculating the access delay and throughput is repeated until the delay in the system reaches its threshold. At this stage, the ratio of contention slot to data slot is changed to 1:3 respectively and the process of calculating the access delay and throughput is repeated.
3.2 Access delay

It can be calculated using a statistical model proposed in (Arthur & Korotin, 2014). In this model, if $BR_i$ was not successfully transmitted in frame $j$, the delay experienced at the beginning of frame $j + 1$ will equal $T_{UL-subframe}$. At that point, and since contention results in a frame is independent of the contention results in previous frames, the process statistically starts over and mathematically formulated in equation 1 as:

$$D_{access}^1 = \frac{n_j}{2} * T_{CS} + \frac{(1-Np(1-p)^{N-1})^n_j}{1-(1-Np(1-p)^{N-1})^n_j} * T_{UL-subframe}$$

(1)

Where:
$D_{access}^1$ – Access delay
$n_j$ – Number of contention slots (CS)
$T_{CS}$: CS duration
$N$ – Number of SS
$P$ – Probability a SS transmits a packet in the beginning of CS.

3.3 Throughput

In the uplink frame there are fixed number of transmission opportunities for BR set by the BS. If there is, only one request submitted to a request slot, the request is successful. On the other hand, if there are two or more SSs sending their requests in the same request slot, collision will happen and truncated binary exponential backoff (TBEB) is used to solve this contention problem (Liu, Chan, & Vu, 2012). Let $W_i$ be the contention window for backoff state $i$, and each SS randomly selects a backoff time in the range $[0, W_i - 1]$. With TBEB, $W_i$ is given by equation 2:

$$W_i = \begin{cases} 2^i W & 0 \leq i \leq R \\ 2^r W & r < i < R \end{cases}$$

(2)

Where:
$r$ is referred to as the truncation value,
$W$ is the initial contention window and
$R$ is the maximum allowable number of attempts.

If the request still fails after $R$ attempts, the packet will be discarded. The throughput of each SS is given by $\lambda * (1 - p^R)$. Since the network provides a capacity of $d$ data slots in each frame with duration $\Delta$, the normalized network throughput is thus given by equation 3:

$$Thru = \frac{N*\lambda*(1-p^R) + \Delta}{d}$$

(3)

Where:
d – Number of data slots
$\Delta$ – Duration of the data slots
$N$ – Number of subscriber stations
$p$ – Failure probability
$R$ – Maximum allowable number of attempts
4 Results

Figure 6. Effect of contention slot allocation on Access delay

Figure 7. Effect of number of SS on Access delay

Fig 6 shows the relationship between the access delay and the number of contention slots. It can be observed that access delay is at its minimum from 0-2000 CS. Then steadily increases up to 2500 CS. This shows that as the number of CS increases it reduces the access delay in the system. This is because there are enough communication resources for the SS to transmit a BR. However it can be seen that from 2000CS the access delay starts increasing exponentially meaning a lot of BR are not being granted because data channels are not readily available or there are a lot of collision of the BR. In the simulation set-up the number of SS were increased by a factor of 10 at every instance. It can be observed from fig 7 that when the number of SS is around 230 the access delay in the system is 0.1833. This is below the access delay threshold of 0.2. At the next instance the access delay in the system gets to 0.3270, which is above the threshold so the number of CS is automatically increased to 2000 CS nevertheless it has little impact on the access delay. It only reduces access delay by 10.70% i.e from 0.3270 to 0.2920. This is still above the threshold so the number of contention slots is increased by a factor of 1000 (3000 CS) i.e a ratio of 3:1, contention slots to data.
slots respectively. This had absolutely no impact on access delay. From fig 13, it can be seen that from 1000CS to 2000CS access increased 10.7%. But from 2000CS to 3000CS it increased exponentially by 687.02% i.e. from 230 SS an increase in CS couldn’t have any meaningful impact on access delay. It can therefore be concluded that the ratio of CS to data slots should be 1:1 for optimum performance.

Figure 8. Effect of contention slot allocation on Normalized Throughput

From fig 8 it is noticed that at a time t=0 when the number of CS is 0, the system’s throughput is 0. This means no BR was being granted so obviously no request can be granted. System throughput increases steadily from the moment the number of CS 100. This shows that requests are being granted. As the number of SS and CS increases throughput also increases steadily. It peaks when the number of CS is about 500 and number of SS is about 45 (fig 9). It starts reducing from 500 to 1000CS when the ratio of CS to data slots is 1:3 and the number of SS reaching 90(fig 16). When the SS is further increased, the throughput in the system approaches 0 at a ratio of CS to data slots being 1:1. This is in line with the fact that access delay also increases by about 10% from 1000CS to 2000CS. This means collisions of BR has drastically increased, so few BR gets to the BS.

Figure 9. Effect of number of SS on Normalized Throughput
for allocation of bandwidth for data transmission. At a ratio of 3:1 CS to data slots respectively, throughput in the system approaches 0. This is because the physical slot for data has become too small so it takes a long time for BR to be answered by the BS. This explains why access delay increases by about 627% within this period.

5 Conclusion

In this article, a model for evaluating the effect of CS allocation has been proposed using normalized throughput and number of SS to evaluate how the number of contention slots affect access delay. It has been clearly shown that increase in contention slot has an effect on access delay at the expense of system throughput, therefore the number of CS used for granting BR affects the QoS of the network. In our analysis, it has been shown that increase of CS more than 2000CS has no meaningful influence on access delay. This was validated by simulation using Matlab 2012.
REFERENCES


УДК 621.396.1

Артур Джозеф Квеку
ФГБУ ВПО «Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича»
Россия, Санкт-Петербург
Аспирант
E-mail: ajkweku@yahoo.com

Коротин Владимир Евгеньевич
ФГБУ ВПО «Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича»
Россия, Санкт-Петербург
Декан
Кандидат технический наук
E-mail: vekorotin@sut.ru

Влияние распределения состязательных слотов на задержку доступа в сети WiMAX

Аннотация. Распределение состязательных слотов (периодов конкуренции) - это открытая область для исследований в сетях WiMAX. К настоящему времени в опубликованных технических требованиях WiMAX не определены алгоритмы и количественные параметры выделения состязательных слотов, и соответственно, в известных научных работах не исследовано влияние числа и механизмов выделения СС на эффективность процесса распределения канального ресурса. Данная статья предлагает модель оценки влияния количества СС на задержку доступа в зависимости от числа абонентов с использованием нормализованной пропускной способности. Было четко показано, что увеличение числа СС оказывает влияние на задержку доступа за счет изменения пропускной способности системы, поэтому количество СС, используемых для предоставления запрашиваемой полосы влияет на качество обслуживания в сети. В анализе было показано, что увеличение СС свыше 2000 не имеет значимого влияния на задержку доступа. Это было подтверждено путем моделирования с использованием пакета Matlab 2012 года. Сетевые операторы могут использовать полученные результаты при планировании сетей WiMAX.

Ключевые слова: WiMAX; запрашиваемая полоса; состязательный слот; средняя задержка; качество обслуживания; планирование алгоритма; абонентское оборудование.

Ссылка для цитирования этой статьи:
Артур Джозеф Квеку, Коротин В.Е. Влияние распределения состязательных слотов на задержку доступа в сети WiMAX // Интернет-журнал «НАУКОВЕДЕНИЕ» Том 7, №1 (2015)

193232, Санкт-Петербург, Пр. Большевиков, д. 22, корп. 1, К. 400
**ЛИТЕРАТУРА**


**Рецензент:** Томашевич Сергей Викторович, заведующий кафедрой технической электродинамики и антенн, профессор, д.т.н., Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича.