

# PERFORMANCE ANALYSIS OF BEST EFFORT SUPPORT IN BROADBAND IEEE 802.16 NETWORKS

V. Rangel, J. Gomez, J. Chapa  
Department of Telecommunications  
UNAM, CP 04510, Mexico City  
{victor, javierg, javierc}@fi-b.unam.mx

M. Lopez-Guerrero  
Department of Electrical Engineering  
UAM-I, CP 09340, Mexico City  
milo@xanum.uam.mx

Raul Aquino  
School of Telematics  
UCOL, Colima, Mexico  
aquinor@ucol.mx

## ABSTRACT

This paper presents a performance analysis of best effort support over the MAC protocol proposed in the IEEE 802.16 standard. We focus on the uplink channel since it is the critical element for efficient delivery of services on demand to individual users (i.e., Best Effort service). We describe a modeling approach and derive the resulting analytical model. Such a model allows us to efficiently compute network throughput. We test this model by means of computer simulations which are also reported in this paper. Simulation results and theoretical computations showed a close agreement, thus validating the model.

*Index Terms*— Channel modeling; Wireless Broadband Networks; performance analysis; WiMAX.

## 1. INTRODUCTION

Broadband Wireless Access (BWA) has become the best option to meet residential and small business demand for high speed Internet access and multimedia services. However, the wide-scale adoption of BWA systems will be determined by its ability to overcome cost and performance barriers. If BWA can meet these challenges, it could easily be the next revolution in wireless systems after WLAN.

The first version of the IEEE 802.16 protocol was completed in October 2001. This version was called IEEE 802.16-2001 [1] and it defines the air interface and medium access control (MAC) protocol for a wireless metropolitan area network (WMAN). It was intended for high-bandwidth wireless voice and data for residential and enterprise use. At the beginning of its development, this protocol was oriented to fixed wireless users with line of sight (LOS) in the 11-66 GHz spectrum range.

In 2004 the aim of the 802.16 protocol was changed to support residential access and non line of sight (NLOS). The second version is called the IEEE 802.16-2004 standard [2] and it supports two operating modes of the MAC layer: 1) point to multipoint (PMP), where traffic occurs only between the Base Station (BS) and Subscriber Stations (SS), and 2) Mesh topology, where traffic can be routed through other SSs and can occur directly between SSs. In addition, IEEE 802.16-2004 also includes OFDM modulation and

supports 256 carriers, which reduces considerably multipath fading effects.

Recently, the IEEE 802.16 Task Force released a new version that enables mobility in SSs. The IEEE 802.16e [3] standard promises mobility support up to speeds of 120 km/h and an asymmetrical link structure. It will enable a SS to have a handheld form factor well suited for PDAs, phones and laptops.

Due to all potential options in BWA standards, at the present time there is a significant interest by the research community, manufactures and operators about what type of performance can really be expected from IEEE 802.16-2004 [2] and IEEE 802.16e [3] certified equipment in the near future. Recent studies found in the literature focus on system performance regarding support of Quality of Service (QoS) [4-7] and cross-layer issues [8-10] for the IEEE 802.16 MAC protocol. However, only relatively few studies [11-13] explicitly approach the analysis of the Best Effort (BE) service class in comparison to other service classes (UGS, rtPS and nrtPS).

Performance of the Best Effort service is severely affected as the network gets congested due to collisions of bandwidth requests transmitted on contention access. The aim of this paper is to provide an analytical model to assess the performance of BE in IEEE 802.16 broadband wireless networks.

The proposed model takes into account that in IEEE 802.16 MAC based systems there is no packet capture. This is due to the fact that subscriber stations always transmit with the minimum transmit power to achieve a specific Bit Error Rate (BER) at the base station. This policy avoids co-channel interference with other cells. Consequently, when two or more stations transmit a bandwidth request in the same contention minislot, a collision occurs and it is not possible to recover any of the bandwidth requests. Our model is based on a few assumptions. The system is considered to be operating in steady state and it is considered that all active SSs are continuously backlogged. It also makes use of a constant value to represent the size of the backoff window of the Exponential Backoff Algorithm (EBA) as we will explain in Section 4.

The paper is structured as follows. Section 2 presents an overview of the relevant parts of the IEEE 802.16 protocol. Section 3 derives the theoretical model for the performance of the uplink channel of the protocol. In Section 4, we present the performance analysis of the IEEE 802.16 protocol comparing analytical and simulation results. We conclude the paper in Section 5.

## 2. IEEE 802.16 MAC LAYER OVERVIEW

The IEEE 802.16 MAC layer makes use of two access techniques, frequency division duplex (FDD) and time division duplex (TDD). Bandwidth requests and data from SSs to the BS are carried in an uplink (UL) frame. Transmissions from the BS to SSs are carried by a downlink (DL) frame. Upon receiving a bandwidth request, the BS handles bandwidth allocation by assigning uplink grants based on requests from the SSs. A typical signaling frame for TDD includes an UL-frame (see Fig. 1a) and a DL-frame (see Fig. 1b) using a single channel frequency as illustrated in Fig. 1c. In FDD these frames are transmitted at the same time using different frequencies as illustrated in Fig. 1d (see [2] for further details on the frame structure).

On power-up the SSs carry out the Initialization and Registration setup as indicated in [2]. After this setup is completed, a SS can create one or more connections over which their data will be transmitted to and from the BS. SSs request transmission opportunities using the contention access period (or contention block) of the current UL-frame. The BS collects these requests and determines the number of minislots (grant size) that each SS will be allowed to transmit in the next UL-frame, using a UL\_MAP subframe, as shown in Fig. 1b. The UL\_MAP frame contains Information Elements (IE) which describe the use of the UL-frame (maintenance, contention and reservation access). The UL\_MAP is broadcasted in the DL channel by the BS in each DL-Frame. After receiving the UL\_MAP, a SS will transmit data in the predefined reserved minislots indicated in the IE. These reserved minislots are transmission opportunities assigned by a scheduling algorithm using a service class, such as UGS (Unsolicited Grant Service) for CBR (Constant Bit Rate) traffic, rtPS (real-time Polling Service) for VBR (Variable Bit Rate), nrtPS (non real-time Polling Service) for non real-time bursty traffic, and BE (Best Effort) for traffic such as Internet, email and all other non real time traffic. Regarding the UL-frame structure depicted in Fig. 1a, in this work we assume that only contention and reservation minislots are considered in the UL-frame, since maintenance minislots are included scarcely.

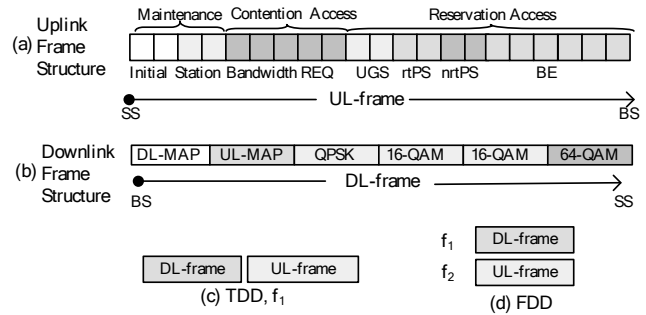


Fig. 1. Frame structure for TDD and FDD access.

## 3. ANALYTICAL MODEL

Since this work focuses on the BE service, let us start by describing the sequence of actions that take place when a SS makes use of it for data transfer. When a SS is active, (let us say  $SS_x$ ), it forms a continuous loop with the sequence of actions depicted in Fig. 2. Upon a packet arrival from an upper layer protocol,  $SS_x$  waits for the next UL-MAP containing a contention period. Then,  $SS_x$  randomly chooses one of the available contention minislots and transmits a bandwidth request (REQ) indicating the packet length. In case some other SS (let us say  $SS_y$ ) selects the same contention minislot, a collision occurs and the subscriber stations ( $SS_x$  and  $SS_y$ ) will receive neither a grant nor an acknowledgement (ACK) in the following UL-MAP. Thus,  $SS_x$  retransmits its REQ until it is successfully transmitted. Upon a successful reception of a REQ from  $SS_x$ , the BS converts the packet size to a number of minislots that should be reserved in subsequent UL-frames. In case the REQ from  $SS_x$  does not fit in the next UL-frame, the BS sends a null grant to  $SS_x$  in order to acknowledge the REQ.

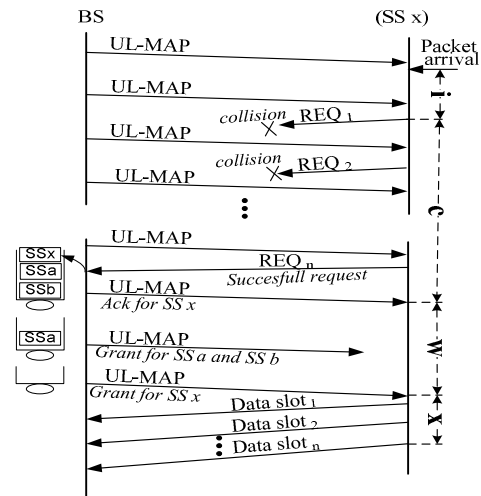


Fig. 2. Access delay components.

Modeling of such sequence of events can be carried out by breaking down a single packet transmission in its delay components. Let us denote by  $i$  the time delay, measured in minislots, from the time a packet arrives from upper layers until the beginning of the contention block where  $SS_x$  transmits its REQ. Let  $c$  represent the total time in minislots spent during contention, which starts with the beginning of the contention block where  $SS_x$  transmits the first REQ, until  $SS_x$  receives an ACK (i.e., a null grant in the IEEE 802.16 protocol). Let us define by  $w$  the time in minislots that the scheduler takes in order to grant the REQ of  $SS_x$ . It is measured from the ACK reception to the grant reception at  $SS_x$ . At the BS, the scheduler serves REQs using a FIFO discipline. If the REQ from  $SS_x$  cannot be granted in the next UL-frame, it waits until previous REQs from other SSs are served. Note that in case a REQ can be immediately served, instead of returning an ACK, the BS returns a grant indicating the number or minislots that were reserved. In this case, reception of such a grant also signals the end of the contention time  $c$  and therefore  $w$  is zero.

Finally, let us denote by  $x$  the delay component that represents the actual number of minislots spent during packet transmission from  $SS_x$ . Fig. 2 depicts the relation between the events described above and the delay components of the model. Therefore, the time to transmit a single packet ( $t$ ) can be directly obtained by adding all delay components, i.e.,  $t=i+c+w+x$ . By taking expectation on both sides of this equation we obtain

$$\bar{t} = \bar{i} + \bar{c} + \bar{w} + \bar{x}. \quad (1)$$

From these parameters, the normalized network throughput ( $\gamma$ ) of the UL channel can be computed as the fraction of minislots actually spent in data transmission per station, scaled to the total number of SSs ( $N$ ), thus,

$$\gamma = N\bar{x} / \bar{t}. \quad (2)$$

The aim of this paper is to provide a method for the computation of this metric. The notation used in this derivation is presented in Table I.

Regarding Eq. (2) it is necessary to compute  $\bar{x}$  and  $\bar{t}$ . The mean delay for packet transmission in minislots (i.e.,  $\bar{x}$ ) can be directly obtained from the packet size distribution, and taking into consideration that minislots can be of size 2, 4, 8, 16, 32 or 64 bytes. Thus,

$$\bar{x} = \left\lceil \frac{\bar{m} + MAC\_OV + PHY\_OV}{slot\_size} \right\rceil \quad (3)$$

where  $\bar{m}$  is the mean packet size at the LLC layer,  $MAC\_OV$  is the MAC overhead of the IEEE 802.16 protocol (the default value for  $MAC\_OV$  is 6 bytes). The overhead at the physical layer,  $PHY\_OV$ , depends on the coding techniques involved (see [2] Chapter 8, for further information on coding rates and modulation techniques).

Computation of  $\bar{t}$  implies computation of other delay components as described below. Estimation of the mean

initial delay  $\bar{i}$  has to consider the following three delay factors. First, a packet arrival from an upper layer protocol may occur anywhere within the current UL-frame, thus the SS must wait until the next available contention block in order to transmit a REQ. The mean value for this delay can be approximated by  $M/2$  minislots, where  $M$  is the total length of the UL-frame. Second, the SS has to wait a complete frame of size  $M$  minislots in order to transmit a REQ in the available contention block of the next UL-frame. Third, once the REQ is received, the BS may grant this REQ anywhere in the following UL-frame, which can also be approximated by  $M/2$  minislots. Thus the initial delay, can be approximated as

$$\bar{i} \cong 2M \quad (4)$$

Whereas computation of  $\bar{x}$  and  $\bar{i}$  is straightforward, computation of  $\bar{c}$  and  $\bar{w}$  is far more complicated. This procedure is described in the following sections.

### 3.1 Computation of the contention delay

For simplicity we assume that the number of failed contentions that a SS needs to succeed follows a geometric distribution. Thus, given the probability of a successful contention in a contention block ( $P_{SC}$ ), the mean number of minislots used for contention  $\bar{c}$  would be given by

$$\bar{c} = C + M \frac{(1 - P_{SC})}{P_{SC}} \quad (5)$$

where  $C$  is the size of the contention block and  $M$  is the total length of a UL-frame (both measured in minislots).

Let  $\bar{n}$  and  $\bar{s}$  be the expected number of the total and successful contenders in a contention block, respectively. It is clear that the probability  $P_{SC}$  can be estimated as

Symbol	Definition
$C$	Size of the contention block in a UL-frame
$R$	Size of the reservation period in a UL-frame
$M$	Total size of a UL-frame, $M=C+R$
$i$	Initial delay
$c$	Contention delay to transmit a REQ successfully
$w$	Time to grant a REQ for a SS
$x$	Time spent in a packet transmission
$t$	Total time in a transmission cycle, $t=i+c+w+x$
$\bar{i}, \bar{c}, \bar{w}, \bar{x}, \bar{t}$	Expectations of $i, c, w, x$ , and $t$ , respectively
$P_{SC}$	Probability of a successful contention in a UL-frame
$n$	Number of SSs contending in the current UL-frame
$s$	Number of SSs that transmitted a REQ successfully in the current UL-frame
$\bar{n}, \bar{s}$	Expectations of $n$ and $s$ , respectively
$\bar{b}$	Expected number of contention periods per cycle, $\bar{t}$
$N$	Maximum number of active SSs
$P_C$	Probability that a SS contends in a contention period
$P_S$	Probability of a successful contention in a minislot
$\bar{d}$	Available transmission minislots for a SS per cycle
$\gamma$	System throughput

$$P_{SC} = \frac{\bar{s}}{\bar{n}}. \quad (6)$$

Let us now turn our attention to the estimation of  $\bar{n}$  and  $\bar{s}$ . Let us denote by  $P_C$  the probability that a SS decides to contend in a contention block. Therefore, the expected number of contenders in a contention block is given by

$$\bar{n} = NP_C \quad (7)$$

and the number of successful contenders per contention block is

$$\bar{s} = (NP_C)P_{SC}. \quad (8)$$

Note that there are  $\bar{i}$  minislots per transmission cycle and one contention block every  $M$  minislots. Therefore, the mean number of contention blocks ( $\bar{b}$ ) per transmission cycle is given by

$$\bar{b} = \frac{\bar{i}}{M}. \quad (9)$$

Let us assume that the system is operating in steady state. Under this assumption, each active SS gets a chance to transmit every  $\bar{i}$  minislots and parameter  $\bar{s}$  can be estimated dividing the total number of SS among the mean number of contention blocks per transmission cycle as follows

$$\bar{s} = \frac{N}{\bar{b}}. \quad (10)$$

Substituting (9) in (10) we obtain

$$\bar{s} = \frac{MN}{\bar{i}}. \quad (11)$$

From (7) and (11) in (6) we can obtain

$$P_C = \frac{M}{P_{SC}\bar{i}}. \quad (12)$$

Let us define the probability of successful contention in an arbitrary slot  $P_S$  as the probability that, from the mean number of contenders in a contention block  $\bar{n}$ , only *one* contends in a minislot (packet capture is not possible),

$$P_S = \binom{\bar{n}}{1} \frac{1}{C} \left(1 - \frac{1}{C}\right)^{\bar{n}-1}. \quad (13)$$

Therefore, parameter  $\bar{s}$  can be computed as

$$\bar{s} = CP_S. \quad (14)$$

Finally, from (14) and (7)  $P_{SC}$  can be computed as

$$P_{SC} = \frac{CP_S}{NP_C}. \quad (15)$$

Note that computation of  $\bar{c}$  in (5) implies computation of  $P_{SC}$  which in turn depends on computation of  $P_C$  and  $P_S$  as indicated by (15). At this point it is fair to mention that the authors in [14] made use of these three probabilities in their analysis of IP-based services over GPRS networks. Although  $P_C$  and  $P_{SC}$  are computed here in the same way, the fact that we do not consider packet capture allows us to compute  $P_S$  in a fundamentally different way. Computation of this probability as defined by (13) is not equivalent to the

method presented in [14]. This difference allows us to go one step further as it will be shown.

The authors in [14] make use of the previously derived probabilities in order to iteratively compute  $P_{SC}$ . Although our model and theirs differ in several ways, we also made use of this procedure with good results. The procedure starts by assuming that  $\bar{w}$  is known and providing an initial estimate for  $P_{SC}$ . Then, from (5), (1), (12) and (13) we compute  $P_C$  and  $P_S$ . These values are used in (15) in order to obtain the following value for  $P_{SC}$ . With this value the process can be repeated starting from (5). This loop can be ended when the difference between two consecutive values for  $P_{SC}$  is below a certain threshold. The final value for  $P_{SC}$  allows us to compute the value for the contention delay  $\bar{c}$  using (5). Further adjustments on  $\bar{w}$  would be needed if the assumed value for  $\bar{w}$  does not satisfy all system conditions.

In what follows we derive the equation that represents the procedure just described. Let us identify with subindex  $n$  the  $n$ -th iteration. Thus, (5) and (1) become  $\bar{c}_n = C + M(1 - P_{SC_n}/P_{SC_n})$  and  $\bar{i}_n = \bar{i} + \bar{c}_n + \bar{w} + \bar{x}$ , respectively. From these equations, we have

$$\bar{i}_n = \bar{i} + C + M \left( \frac{1 - P_{SC_n}}{P_{SC_n}} \right) + \bar{w} + \bar{x}. \quad (16)$$

From (12) we know that  $P_{C_n} = M / (P_{SC_n} \bar{i}_n)$  which combined with (16) yields,

$$P_{C_n} = \frac{1}{\beta P_{SC_n} + 1} \quad (17)$$

where  $\beta = ((\bar{i} + C + \bar{w} + \bar{x})/M) - 1$ .

From (7) and (17) the estimated mean number of users at the  $n$ -th iteration is

$$\bar{n}_n = \frac{1}{\beta P_{SC_n} + 1} N. \quad (18)$$

From (13) and (18) the estimated value of the probability of successful contention in an arbitrary slot at the  $n$ -th iteration is

$$P_{S_n} = \frac{N}{\beta P_{SC_n} + 1} \left( \frac{1}{C} \right) \left( 1 - \frac{1}{C} \right)^{\frac{N}{\beta P_{SC_n} + 1} - 1}. \quad (19)$$

Finally, from (17), (19) and (15), after some algebra, the difference equation that can be used to obtain the value of  $P_{SC}$  is

$$P_{SC_{n+1}} = \left( 1 - \frac{1}{C} \right)^{\frac{N}{\beta P_{SC_n} + 1}}. \quad (20)$$

In summary, given  $\bar{w}$  and an initial estimate for  $P_{SC}$ , we iterate (20) until a consistent value for  $P_{SC}$  is obtained. With this value the contention delay  $\bar{c}$  can be computed using (5). The algorithm implied by (20) is very easy to implement and it is much simpler than the one presented in [14].



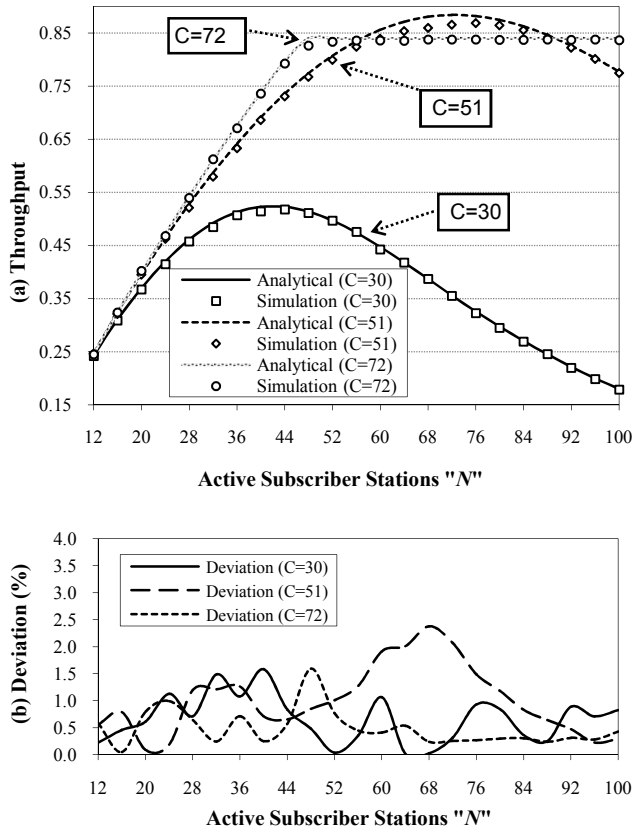


Fig. 4. Performance analysis of IEEE 802.16 based Network.

each UL-frame. With 40 SSs, the average number of grants served per UL-frame was of 11.6 of a total of  $(M-C)/\bar{x} = 21$  grants. The rest of the UL-frame was wasted, since just a few REQ could arrive to the BS due to collisions.

By increasing  $C$  to 51 minislots, the average number of grants served by a UL-frame increased considerably to 17, which nearly achieves 90% of system throughput when there are between 60 and 70 SSs. However, when there is a large number of SSs in the network, (i.e., more than 80 SSs) the system throughput cannot be maintained due to the large number of collisions reported.

With  $C = 72$  minislots, the system throughput can be sustained even on very large networks, however the maximum system throughput is just over 80% of channel capacity.

For the three values of  $C$  shown in Fig. 4b, we observed that simulation results were in good agreement with analytical results. The maximum deviation from our simulation model was less than 3%.

## 5. CONCLUSION

In this paper we presented an analytical model for the MAC layer of IEEE 802.16 based networks under best effort traffic. The results obtained from the analytical model were

verified by simulation. We found that the analytical model is sufficiently accurate to study the trends in system performance and, in almost all cases, gives numerical results within the confidence intervals derived from the simulation results. The maximum deviation between analytical and simulation results was less than 3%, with a mean deviation of 0.8%. In future work will consider performance modeling at the physical layer, different traffic scenarios and we will also take into account the dynamic behavior of the Exponential Backoff Algorithm.

## ACKNOWLEDGMENTS

This work was supported by DGAPA, UNAM under Grant PAPIIT IN104907 and PAPIME PE103807. Additional funding came from *Área de Investigación en Redes y Telecomunicaciones* (UAM-Iztapalapa, Mexico).

## REFERENCES

- [1] IEEE 802.16-2001, "IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems," April 2002.
- [2] IEEE 802.16-2004, "IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems," October 2004.
- [3] IEEE 802.16e, "IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed BWA Systems", Amendment for PHY and MAC for Combined Fixed and Mobile Operation in Licensed Bands, December 2005.
- [4] C. Cicconetti, A. Erta, L. Lensini, E. Mingozzi, "Performance Evaluation of the IEEE 802.16 MAC for QoS Support," *IEEE Trans. on Mobile Computing*, vol. 6, no.1, pp. 26-38, Jan. 2007.
- [5] R. Iyengar, P. Iyer, B. Sikdar, "Delay Analysis of 802.16 based Last Mile Wireless Networks," in *2005 Proc. GLOBECOM Conf.*, St. Louis.
- [6] Y. Yu, H. Chao, "QoS Aware Link Adaptation in IEEE 802.16 Broadband Wireless Access Systems," in *2007 Proc. PIMRC Conf.*
- [7] D.Hao, S.Zou, B. Adamou, "A QoS Guaranteed Cross-Layer Scheduling Algorithm in Wireless Networks," in *2003 Proc. ICWMC Conf.*
- [8] J. Yan, G. Kuo, "Cross-Layer Design of Optimal Contention Period for IEEE 802.16 BWA Systems," in *Proc. IEEE ICC 2006*, Istanbul.
- [9] Y. Chang, F. Chien, J. Kuo, "Cross-Layer QoS Analysis of Opportunistic OFDM-TDMA and OFDMA Networks," *IEEE Journal of Selected Areas in Communications*, vol. 25, no. 4, May 2007.
- [10] Y. Mai, C. Yang, Y. Lin, "Cross-Layer QoS Framework in the IEEE 802.16 Networks," in *Proc. ICACT 2007*, Gangwon-Do, Korea
- [11] K. Gakhar M. Achir, A. Gravey, "Dynamic Resource Reservation in IEEE 802.16 Broadband Wireless Networks", in *Proc. IEEE IWQoS*, 2006. New Haven, USA.
- [12] B. Bhandari, R. Kumar, S. Maskara, "Performance of the IEEE 802.16 MAC Protocol with Varying Request Backoff Start," in *Proc. IEEE TENCON 2006*, Hong Kong.
- [13] A. Vinel, Y. Zhang, M. Lott, A. Tiurlikov, "Performance Analysis of the Random Access in IEEE 802.16," in *Proc. IEEE PIMRC 2005*, Berlin.
- [14] V. A. Chitre and J. N. Daigle, "Performance of IP-Based Services over GPRS," *IEEE Trans. on Computers*, vol. 52, No. 6, pp. 727-741, June 2003.
- [15] V. Rangel, J. Gomez y J. Ortiz, "Performance Analysis of QoS Scheduling in Broadband IEEE 802.16 Based Networks," in *Proc. OPNETWORK Technology Conf.*, Washington 2006.