

# Efficient Request Mechanism Usage in IEEE 802.16

Alexey Vinel

Information Systems Department  
State University of Aerospace Instrumentation  
Saint-Petersburg, Russia  
vinel@inbox.ru

Ying Zhang

Future Radio Concepts  
Siemens AG  
Munich, Germany  
ying.zhang@mytum.de

Qiang Ni

Electronic & Computer Engineering  
School of Engineering and Design, Brunel University  
London, United Kingdom  
qiang.ni@ieee.org

Andrey Lyakhov

Laboratory of Information Transmission Networks  
Institute for Information Transmission Problems, RAS  
Moscow, Russia  
lyakhov@iitp.ru

**Abstract**—IEEE 802.16 protocols for metropolitan broadband wireless access systems have been standardized recently. According to the standard, a subscriber station can deliver bandwidth request messages to a base station by numerous methods. This paper provides both the simulation and analytical models for the investigation of specified random access method, which is compared with centralized polling and station-grouping mechanisms. Based on the assumptions of Bernoulli request arrival process and ideal channel conditions, the mean delay of a request transmission is evaluated for varying number of transmission opportunities and different arrival rates.

**Keywords**—random access algorithm; medium access control protocol; bandwidth request; IEEE 802.16; WiMaX

## I. INTRODUCTION

The IEEE 802.16 WiMaX standard [1] is supposed to play an important role in rapid and ubiquitous adoption of broadband wireless access systems worldwide. In centralized point-to-multipoint architecture of WiMaX, subscriber stations (SSs) share the uplink to a base station (BS) on demand basis. This means that if an SS needs some amounts of bandwidth, it informs the BS by means of transferring a request message. The BS scheduler accepts the requests from different SSs and grants them the transmission opportunities in time slots by using some scheduling algorithms, which should take into account the requirement of each SS and the available channel resources. These grants are made based on the negotiated quality of service (QoS) agreements. Two main methods to provide transmission opportunities are suggested: centralized polling and contention-based random access. According to the IEEE 802.16, the latter may be implemented in two different ways. The first and most typical solution is to give all stations an opportunity to access all available contention slots; another typical solution is to group some stations together and assign disjoint subsets of all contention slots to each group.

The performance comparison of the two polar access methods, namely polling and random access has been widely discussed under the framework of IEEE 802.11 MAC layer protocols: i.e., Distributed Coordination Function (DCF) and

Point Coordination Function (PCF) recently ([2]-[4]). DCF is a carrier sense multiple access (CSMA) scheme with binary exponential backoff (BEB) collision resolution, and works well under low load conditions. However, its performance degrades essentially when the number of terminals increases. Waste of bandwidth caused by collisions and increasing backoff times becomes very high in this case. In the latter case, PCF, which actually represents a Time Division Multiple Access (TDMA) style polling scheme, is preferable.

As for the IEEE 802.16 WiMaX networks, by now few scientific results have been obtained to evaluate its performance as well as to propose vendors some efficient ways to use different available tools and methods. One noticeable IEEE 802.16 MAC investigations line is devoted to efficient mesh-topology network operation [5]-[7], where this system is analyzed mainly by means of simulations. An analytical model for CSMA based Point-to-Multipoint (PMP) broadband wireless network was considered in [8]. However, according to the WiMaX standard [1] a mandatory random access method in IEEE 802.16 is based on a truncated BEB without any carrier-sense features. Hence, this model is not suitable for the analysis of 802.16 WiMaX networks.

BEB algorithm within the framework of IEEE 802.16 was first investigated in [9] under a “saturation” condition assumption by using an analytical approach similar to [10]. However, this assumption is too strong since networks do not typically operate in saturation conditions. Most Internet applications exhibit bursty traffic characteristics. A reasonable model for the fair analysis in more general case of an arbitrary request arrival rate should consider stations in two states (active/inactive). Similar idea was introduced in the analysis of wireless centralized networks, for instance, wireless ATM in [11]. Following this approach, we do not consider data packet transmission, but limit our scope to the delay analysis of bandwidth request during the reservation process. First of all, this allows us to focus attention only on the efficiency of the bandwidth request mechanisms. Secondly, packet transmission delay is also affected by the implementation of various BS scheduling algorithms, which is vendor-dependent and will be investigated in our future work.

---

This work was supported in part for Alexey Vinel by Russian Federal Agency for Education under project PHII.2.2.2.3.9612 (research work No 53-02-672-5).

In this paper, the random access mechanism of IEEE 802.16 WiMaX networks is investigated under the assumption of finite number of stations and Bernoulli request arrival process. After a simplified explanation of IEEE 802.16 protocol and its request mechanisms in Section II, we introduce in section III both the simulation and analytical models for the specified random access method. In section IV, the efficiency of the simplest cyclic shift polling scheme, random access scheme, and station-grouping scheme under different conditions, namely with various arrival rates and different numbers of transmission opportunities, are investigated. Finally, we draw conclusions in Section V.

## II. IEEE 802.16 MAC PROTOCOL

### A. Frame Structure

Let us consider the network with a PMP architecture, which consists of one BS managing multiple SSs. Transmissions between the BS and SSs are realized in fix-sized frames by means of time division multiple access (TDMA) / time division duplexing (TDD) mode of operation. The frame structure consists of a downlink sub-frame for transmission from the BS to SSs and an uplink sub-frame for transmissions in the reverse direction. The Tx/Rx transition gap (TTG) and Rx/Tx transition gap (RTG) shall be inserted between the sub-frames to allow terminals to turn around from reception to transmission and vice versa. In the downlink sub-frame the Downlink MAP (DL-MAP) and Uplink MAP (UL-MAP) messages are transmitted by the BS, which comprise the bandwidth allocations for data transmission in both downlink and uplink directions, respectively.

Another important management message, which is interconnected with UL-MAP, is called an Uplink Channel Descriptor (UCD). It can be periodically transmitted in the downlink sub-frame. The values of the minimum backoff window,  $W_{\min}$ , and the maximum backoff window,  $W_{\max}$ , are defined in this message, which are used for the collision resolution algorithm. The uplink sub-frame contains transmission opportunities scheduled for the purpose of bandwidth requests, in which Bandwidth Request (BW-REQ) messages can be transmitted, which serve for SSs to indicate to the BS that how much uplink bandwidth allocation they need. The BS manages the number of transmission opportunities through the UL-MAP message.

### B. Request Mechanisms in IEEE 802.16

Each transmission opportunity may be assigned by the BS either to exactly one subscriber station or to a group of stations. In the first case, considered station is provided a so-called unicast opportunity for BW-REQ transmission. In other words, the BS polls an SS to allow it to transmit the request in a contention-free manner. In the latter case random access algorithm is used by the group of SSs to contend for the common transmission opportunities and resolve possible collisions.

As mentioned earlier, the mandatory method of contention resolution, which shall be supported in the standard, is based on a truncated binary exponential backoff, with the initial

backoff window and the maximum backoff window controlled by the BS. This algorithm is described in detail in the next section as a part of our simulation model for the random access.

The information whether BW-REQ message is successfully transmitted or distorted (because of collision or noise) is not explicitly transmitted by the BS in the downlink. It is not specified in the standards how the SS knows the result of its transmission. It might be based on the correspondence between the amount of the resources assigned to the given SS and the amount of the resources it has asked for in the transmitted BW-REQ message.

## III. MODEL FOR RANDOM ACCESS

For the purpose of investigating transmission opportunities usage in IEEE 802.16 the following model for the random access is introduced.

### A. Simulation Model for Random Access

Let us consider  $n$  SSs, simply denoted by stations in the sequel, having a buffer sufficient to store exactly one request. A station, which has a request in the considered moment of time is referred to as “active”, otherwise it is called a “non-active” one. According to the 802.16 standard [1], each SS may potentially establish multiple connections with different negotiated QoS parameters with the BS, and a bandwidth request can be issued per-connection or per-station based. In this work, we assume that each station has only one connection at a given time. In the case of multiple connections per SS,  $n$  is referred to the total number of connections in the system.

We define the nominal request arrival rate  $\lambda$  as the mean number of the requests, appearing in the system for a frame duration, given that there are no active stations. During a frame duration, each non-active station generates a request with a probability  $\pi = \lambda/n$ . Thus, more active stations in the system we have, less actual request arrival rate is. This new request is put into the buffer, and will be transmitted not earlier than a next frame transmission. Each frame comprises  $K$  equal slots for the random access. The duration of a slot corresponds to the time needed for a bandwidth request transmission. The BS chooses the value of  $K$  in order to make a trade-off between the frame duration of contending BW request period and that of a data payload period within the whole frame duration, which is fixed. Therefore, in the following discussion,  $K$  is assumed to be a fixed value.

We assume ideal channel conditions, i.e., if exactly one station transmits in a slot, the transmission is successful, otherwise a collision occurs. No channel transmission error is considered. Furthermore, we assume that stations receive a feedback from the BS at the beginning of the next frame whether their transmission was successful or not. According to [1] this is not the case. Feedback information is not explicitly transmitted to the SS. A special request timeout called T16 is used by a SS to wait for the data grant from the BS, and if it is expired, the request transmission was considered corrupted.

For collision resolution a binary exponential backoff algorithm is used. Before each transmission attempt, a station uniformly chooses an integer number from the interval  $[0, W_i -$

1], where  $W_i$  is the current value of its backoff window. The chosen value, referred to as a backoff counter, indicates the number of slots the station has to wait before the transmission of a request. For the first transmission attempt, the backoff window size is set to  $W_{\min}$ . In the case of a collision a station doubles its backoff window value, and so the backoff window after  $i$  collisions,  $W_i$ , becomes  $2^i W_{\min}$ . The window is not doubled if it reaches the maximum value  $W_{\max} = 2^m W_{\min}$ , where  $m$  is referred to the maximum backoff stage. In the case of the successful transmission the backoff window is set to the minimum value  $W_{\min}$ .

The standard does not define any relationship between the parameters  $W_{\min}$ ,  $W_{\max}$  and  $K$ . Notice that if  $W_{\min} < K$ , then some time slots will never be used during the first transmission attempt. For this reason, we set  $W_{\min} = lK$ , where  $l$  is a natural number ( $l \geq 1$ ), in order to uniformly distribute the transmission attempts over the available random access slots.

### B. Analytical Model for Random Access

In this system, the number of active stations at the beginning of each frame is a stochastic process. It can be modeled by a discrete-time Markov chain: Let us calculate its transition probabilities as follows. Assuming the request arrival is modeled by a Bernoulli process, the number of new requests appearing in the system for the frame duration has a binomial distribution with a parameter  $\pi$ . Thus, the probability  $\psi_{j,i}$ , that  $j$  new requests appear in the system for the frame duration, under the condition, that there are  $i$  active stations is

$$\psi_{j,i} = \binom{n-i}{j} \pi^j (1-\pi)^{n-i-j}. \quad (1)$$

Let us temporarily assume, that there are constantly  $i$  active stations in the system. For this case, following the approach of [10], we suppose that the behaviour of an arbitrary station does not depend on the behaviour of the other  $i-1$  stations, and the conditional collision probability  $p$ , that a station transmits and falls into collision, is constant.

Under such an assumption a two-step procedure corresponding to the subscriber operation is modelled [9]. In the first step, a station uniformly chooses one of the  $L_w$  frames to transmit, where  $L_w = 2^w$ ,  $w = 0, \dots, m$ , and  $w$  describes the current backoff stage. In the second step, one out of  $K$  slots is uniformly chosen in the given frame.

A discrete and integer time scale is adopted, where  $t$  and  $t+1$  correspond to the beginning of two consequent frames. Let  $c(t)$  be the stochastic process representing the integer number of frames a station has to wait before the transmission at time  $t$ . So the station transmits in a frame, which starts at the moment  $t$ , if  $c(t)$  equals to zero. Let  $b(t)$  be the backoff stage of a station at the moment  $t$ . It is possible to describe the two-dimensional process  $\{b(t), c(t)\}$  by the Markov chain introduced in [10] with the following transition probabilities:

$$\begin{aligned} P\{w, v | w, v+1\} &= 1, & v &= 0, \dots, L_w - 2, w = 0, \dots, m \\ P\{0, v | w, 0\} &= (1-p)/L_0, & v &= 0, \dots, L_0 - 1, w = 0, \dots, m \\ P\{w, v | w-1, 0\} &= p/L_w, & v &= 0, \dots, L_w - 1, w = 1, \dots, m \\ P\{m, v | m, 0\} &= p/L_m, & v &= 0, \dots, L_m - 1 \end{aligned}$$

where we adopt the notation  $P\{w_1, v_1 | w_0, v_0\} = P\{b(t+1)=w_1, c(t+1)=v_1 | b(t)=w_0, c(t)=v_0\}$ .

We omit the detailed mathematical manipulations as the above chain is similar to the one in [10]. Summarizing the probabilities of the states when  $c(t)$  equals to zero, the following equation for the probability  $x$  of a station to transmit in a frame can be obtained:

$$x = \frac{2(1-2p)}{(1-2p)(l+1) + pl(1-(2p)^m)}. \quad (2)$$

Let us consider that one particular station transmits in a frame. Under this condition, the probability  $y_u$ , that  $u$  stations from the remaining  $i-1$  that transmit in the same frame, is equal to

$$y_u = \binom{i-1}{u} x^u (1-x)^{i-1-u},$$

and the probability that all of them transmit in the slots different from the one chosen by the considered station is  $(1-1/K)^u$ . Thus, the conditional collision probability  $p$  is

$$p = 1 - \sum_{u=0}^{i-1} \binom{i-1}{u} x^u (1-x)^{i-1-u} \left(1 - \frac{1}{K}\right)^u. \quad (3)$$

So, the non-linear system is represented by equations (2) and (3) with two unknown variables  $p$  and  $x$ . It can be proven that this system has a unique solution and thus probability  $x$  could be calculated numerically for any number of active stations in the system  $i$ .

Coming back to the model with variable number of active stations, let us assume, that if we have  $i$  active stations in the beginning of a frame, then each of them transmits in this frame with probability  $x(i)$ , which is equal to the corresponding value obtained for the network analogous to the considering one, but constantly having  $i$  active stations. Intuitively this assumption is valid, when the number of active stations in the systems changes slowly. Therefore, probability  $Q_{k,i}$ , that  $k$  from  $i$  stations choose this frame equals to

$$Q_{k,i} = \binom{i}{k} (x(i))^k (1-x(i))^{i-k}.$$

Compute the probability  $R_{r,k}$ , that  $r$  stations out of total  $k$  active ones successfully transmit in a frame. Denote by  $N(r, k, K)$  the total number of ways to put  $k$  balls into  $K$  boxes, under the condition, that exactly  $r$  boxes contain one ball. This function is computed recursively using the following rule

$$\begin{aligned}
N(0,0,K) &= 1; N(0,k,0) = 0 \\
N(0,k,K) &= K^k - \sum_{s=1}^{\min(k,K)} N(s,k,K), k > 0 \\
N(r,k,K) &= \binom{k}{r} \binom{K}{r} \cdot r! N(0,k-r,K-r), 0 < r \leq \min(k,K)
\end{aligned}$$

and the conditional probability  $R_{r,k}$  equals to

$$R_{r,k} = \frac{N(r,k,K)}{K^k}.$$

Thus, the probability, that  $r$  stations successfully transmit in a frame, under the condition, that there are  $i$  active stations, equals to

$$\xi_{r,i} = \sum_{k=r}^i R_{r,k} Q_{k,i}. \quad (4)$$

Now we are ready to write out the transition probabilities matrix for the Markov chain, representing the number of active stations in the system. Note that the chain has  $(n+1)$  states and non-zero probabilities  $\Pi\{i_2 | i_1\}$  of transition from state  $i_1$  into state  $i_2$  can be expressed (applying probabilities from (1) and (4)) as

$$\Pi\{i_2 | i_1\} = \begin{cases} \psi_{i_2,0} & i_1 = 0, i_2 = 0 \\ \sum_{r=r_{\min}(i_1,i_2)}^{r_{\max}(i_1,i_2)} \xi_{r,i_1} \psi_{i_2+r-i_1,i_1} & i_1 \neq 0, i_1 \neq n \\ \xi_{i_2,n} & i_2 \geq n - K \end{cases}$$

where  $r_{\min}(i_1, i_2) = \max(0, i_1 - i_2)$ ,  $r_{\max}(i_1, i_2) = \min(i_1, n - i_2, K)$  and  $r_{\min}(i_1, i_2) \geq r_{\max}(i_1, i_2)$  condition holds.

Applying iteration method one may compute stationary distribution  $(\rho_0, \rho_1, \rho_2 \dots \rho_n)$  of this ergodic Markov chain, the mean number of the requests in the system is

$$\bar{N} = \sum_{i=0}^n i \rho_i.$$

Using the Little's law and taking into account the waiting time till the beginning of frame, we finally obtain the mean delay for the request transmission

$$\bar{d} = 0.5 + \bar{N} \frac{n}{(n - \bar{N})\lambda}.$$

#### IV. PERFORMANCE COMPARISON OF REQUEST MECHANISMS IN IEEE 802.16 WiMAX

In this section we use the above simulation and analytical models to compare different request mechanisms in the framework of IEEE 802.16, namely centralized polling, random access and station-grouping mechanisms.

##### A. Random Access with Different Parameters

Let us consider the scenario with  $n$  ( $= 25$ ) stations,  $K$  ( $= 4$ ) contention slots and different BEB parameters  $(l, m)$  as shown in Fig. 1. The comparison between the analytical results (lines) and the simulation results (symbols) shows that the analytical results meet the simulation results quite well for varied system parameters. Note that there is a certain gap between the simulation and analytical results for small number of stations when the back-off stage is large. In the case when  $l = 1$  and  $m = 1$ , the mean delay increases rapidly. This is because that the relative high arrival rate results in more access trials and hence high risk of collisions. By enlarging  $m$ , the risk of collisions is reduced and so will be the mean delay. On the other hand, increasing the value of  $l$  leads to higher mean delay for small arrival rate values. From the practical point of view, for simplicity of implementation, we recommend to set up  $l = 1$  and sufficiently large  $m$ , which allows to change transmission probability adaptively and provide efficient performance for different arrival rates.

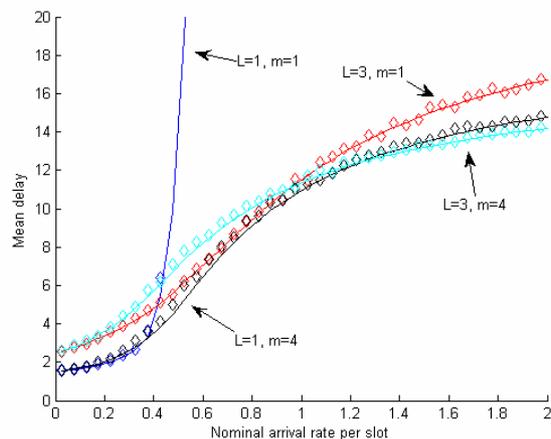


Figure 1. Performance of random access

##### B. Random Access vs. Simplest Polling

Let us consider the following simplest polling scheme: each of  $K$  slots are consequently assigned to all of  $n$  stations in the system ("round-robin" polling style). Assuming, that  $n \gg K$ , mean delay for the request transmission is an increasing function of arrival rate, with minimal value equals to  $0.5 + n/(2K)$  for small arrival rates and maximal one  $- 0.5 + n/K$  for saturated conditions. This heuristic statement can be proven and the system can be modeled by means of Markov chain to compute exact expression for this function. However, for simplicity reasons and paper space constraints we assume linear delay dependence of nominal arrival rate.

One can see in Fig. 2, that for typical scenario  $n = 30$  and  $K = 5$ , random access is preferable in term of mean delay as far as nominal arrival rate per slot is less than approximately 0.5.

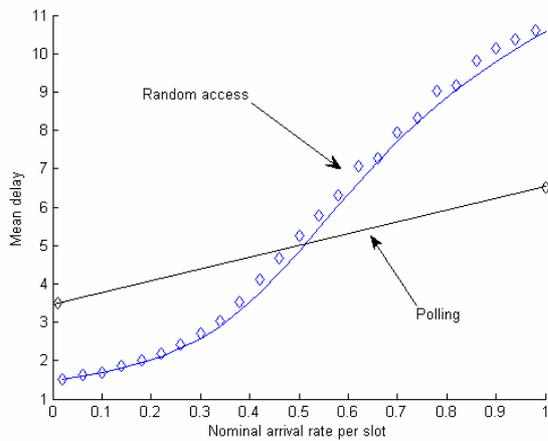


Figure 2. Performance comparison of random access and polling.

### C. Grouping of stations

As noted in the introduction, it is possible to group the stations and perform random access within group in WiMaX. For example, instead of letting all  $n$  stations compete for  $K$  slots, we can equally divide the stations as well as the available access slots into  $g$  groups and let  $n/g$  stations in the same group compete for  $K/g$  slots. In this section, we study how the grouping influences the performance of request transmission. Suppose  $n = 48$ ,  $K = 12$  and we also select  $l = 1$ ,  $m = 4$  as typical values for BEB. Firstly, we do not apply grouping, and so the total  $n$  stations share all the  $K$  access slots. The corresponding performance in terms of mean delay is depicted in Fig. 3. If we divide them into 2 groups,  $g = 2$ ,  $n/2$  users share  $K/2$  access slots in each group. As we can see in this case the mean delay significantly increases. If we use more groups, the performance degradation is still rather noticeable, as shown in Fig.3. Indeed, we gain from statistical multiplexing by allowing several stations to share the available access slots, which is the reason that random access outperforms polling at low request arrival rate. However, on the other hand, allowing more stations to share the access slots increases the risk of potential collisions within stations. By dividing all stations into groups, because of less number of stations per group, we lose some statistical multiplexing gain but reduce the risk of collisions as well. The loss on the statistical multiplexing gain can be compensated by the gain from less collision probability only if there are few stations in each group and the arrival rate is high enough. This observation can be obtained when  $g = 12$ . This is the situation, when there are 4 stations in each group and only one slot per frame is allocated for each group. Generally speaking, we recommend applying grouping in the case of high traffic density and keeping the number of Ss in each group be small.

### V. CONCLUSION

In this paper, efficient bandwidth request mechanisms for IEEE 802.16 are investigated. The contribution of this work is twofold: Firstly, we have introduced a new analytical model for the random access scheme standardized in IEEE 802.16.

Secondly, based on this model we have compared the efficiency of the simplest polling scheme, random access and station-grouping mechanisms in different scenarios. Using our model, base station can decide which method is preferable to increase the overall system performance. Future work will be dedicated to the performance evaluation of the whole IEEE 802.16 MAC scheme including data packet transmission.

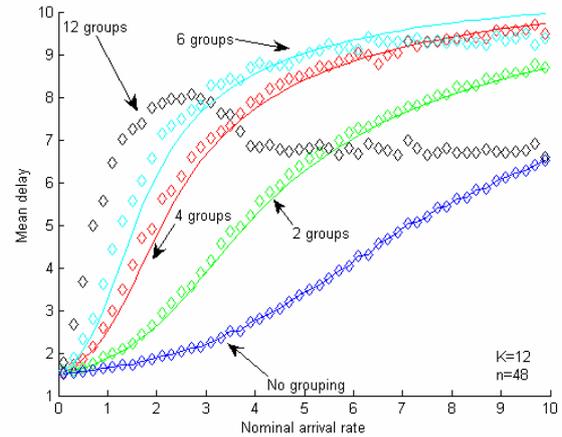


Figure 3. Performance of random access with station-grouping.

### REFERENCES

- [1] IEEE Std. 802.16-2004, IEEE standard for local and metropolitan area networks, Air Interface for Fixed Broadband Wireless Access Systems.
- [2] Kopsel A., Ebert J.-P., Wolisz A. "A Performance Comparison of Point and Distributed Coordination Function of an IEEE 802.11 WLAN in the Presence of Real-Time Requirements," in Proc. of 7th Int. Workshop MoMuc2000, Waseda, Japan, October 2000.
- [3] Choi S. "PCF vs. DCF: Limitations and Trends," IEEE 802.11-01/154, Jan. 2001.
- [4] Vishnevsky V.M., Lyakhov A.I., Guzakov N.N., "Evaluation of the Maximal Carrying Capacity of Wireless Internet Access," Automation and Remote Control, vol.65, no.9, part 2, September 2004.
- [5] Redana S., Lott M. "Performance Analysis of IEEE 802.16a in Mesh Operation Mode," The 13th IST SUMMIT, Lyon, France, June 2004.
- [6] Redana S., Lott M., Capone A. "Performance Evaluation of Point-to-Multi-Point (PMP) and Mesh Air-Interface in IEEE Standard 802.16a," IEEE VTC Fall' 04, Los Angeles, USA, 26.-29. Sept. 2004.
- [7] Lott M., Redana S., Carozzo M. "Reducing the delay of IEEE 802.16a in multi-hop scenarios" World Wireless Congress, USA, May 2005.
- [8] Guo S., Sydor J., Brandao A. "Performance Analysis of the P2MP Metropolitan Area Broadband Wireless Access Network," IEEE PIMRC 2003.
- [9] Vinel A., Zhang Y., Lott M., Turlikov A. "Performance Analysis of the Random Access in IEEE 802.16," in Proc. of the IEEE PIMRC'05, Berlin, Germany, September 2005.
- [10] Bianchi G. "Performance Analysis of the IEEE 802.11 Distributed Coordination Function," IEEE Journal on Selected Areas in Communications, V18, N3, 2000, pp 535-548.
- [11] Cortizo D. V., Garcia J., Blondia C., VanHoudt B. "FIFO by Sets ALOHA (FS-ALOHA): A Collision Resolution Algorithm for the Contention Channel in Wireless ATM Systems," IFIP WG 7.3 Performance, Istanbul, 1999.