Performance Analysis of the Random Access in IEEE 802.16

Alexey Vinel[1*], Ying Zhang[1], Matthias Lott[1], Andrey Tiurlikov[2]

[1]Siemens AG, Communications, Munich, Germany
[2]State University of Aerospace Instrumentation, St.-Petersburg, Russia

E-mail: vinel@inbox.ru, ying.zhang@mytum.de, matthias.lott@siemens.com, tiurlikov@mail.ru

Abstract—**Recently, the IEEE has standardized the 802.16 protocol for metropolitan broadband wireless access systems. According to the standard, the random access scheme based on the slotted binary exponential backoff algorithm is used in this system for initial ranging and bandwidth requests transmission. This paper provides both the simulation and analytical models for the investigation of the random access in IEEE 802.16. Based on the assumption of finite number of subscriber stations and ideal channel conditions the delay is evaluated for varying number of transmission opportunities and different backoff window sizes.**

I. INTRODUCTION

The design and adoption of broadband wireless access systems is the most significant direction of telecommunication technologies development at the moment. The IEEE 802.16 standard 0 defines a very complex system of this type, having a great number of different modes and technical solutions. Though these solutions individually are investigated well enough, their efficiency in a setting of a whole system has not been analyzed, yet. For instance, a collision resolution algorithm called "binary exponential backoff" (BEB) is introduced in the standard. It is widely known and modifications have been applied and investigated in quite different data transmission systems, for example, in traditional IEEE 802.3 local area networks or IEEE 802.11 wireless systems. However, the efficiency of BEB algorithm is strongly dependent on the system it is used in. That is why special analysis is needed in each specific case.

Furthermore, according to the standard, there is an opportunity to dynamically tune the parameters of the collision resolution algorithm, like, the minimum and maximum backoff window sizes. However, the substantial algorithms for this purpose are not defined.

In this paper the random access protocol defined in the IEEE 802.16 standard is investigated. The attempt of solving the similar task is made in [2]. However, the carrier sense multiple access (CSMA) scheme is considered there and does not correspond to the access scheme defined in the IEEE 802.16 standard for that purpose 0. After a simplified description of the IEEE 802.16 protocol in Section II we introduce in Section III a simulation and an analytical model for the random access. In Section IV we evaluate the performance of the random access in IEEE 802.16. A fixed number of active stations and ideal channel conditions are assumed. The mean delay for request transmissions is computed. The analytical model gives extremely accurate results for different parameters, which has been proven by means of simulations. Based on this model an optimal fixed backoff window size to minimize the delay is derived, assuming the knowledge the number of active stations. Finally, main conclusions and remarks are contained in the Section V.

II. IEEE 802.16

This section provides the simplified description of IEEE 802.16 medium access control protocol.

A. Frame Structure

Let us consider the network with a point-to-multipoint (PMP) architecture, which consists of one base station (BS) managing several subscriber stations (SS). Transmissions between the BS and SSs are realized in fixed frames by means of time division multiple access (TDMA) / time division duplexing (TDD) mode of operation. The frame structure, shown in Figure 1, 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, which comprise the bandwidth allocations for data transmission in downlink and uplink direction, respectively.

Another important management message, which is interconnected with UL-MAP is an Uplink Channel Descriptor (UCD), which can be periodically transmitted in the downlink sub-frame. The values of the minimum backoff window, W_{min}, and maximum backoff window, W_{max}, are defined in this message, which are used for the collision resolution algorithm.

The uplink sub-frame consists of transmission opportunities scheduled for initial ranging and bandwidth requests purposes, in which respectively the Ranging Request (RNG-REQ) and Bandwidth Request (BW-REQ) message can be transmitted. The RNG-REQ message is meant for, e.g., new SSs for initialization to join the

* A. Vinel is a student of St.-Petersburg State University of Aerospace Instrumentation currently doing an internship at Siemens AG.

978-3-8007-2909-8/05/$20.00 ©2005 IEEE

network, whereas the BW-REQ message serves for SSs to indicate to the BS that they need UL bandwidth allocation. The BS manages the number of transmission opportunities through the UL-MAP message.

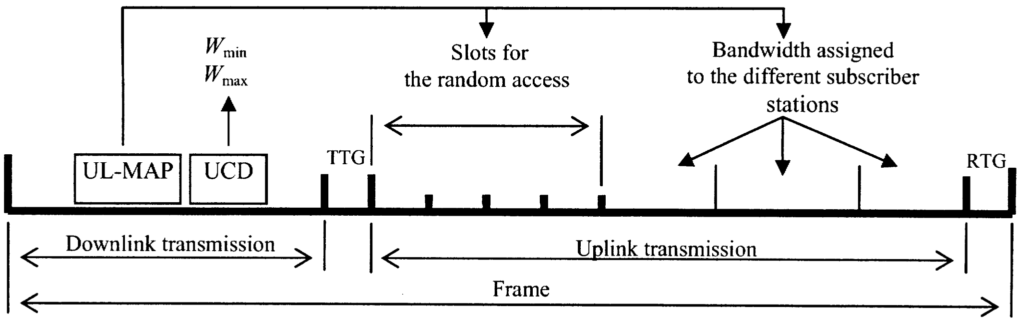

Figure 1: The simplified frame structure

B. Random Access in IEEE 802.16

According to the standard, BW-REQ messages can be transmitted in dedicated transmission opportunities managed by the BS or in contention by means of random access. In this paper we focus on the random access where all SSs are contending for the available transmission opportunities.

The decision whether a transmission of RNG-REQ is successful or not is made by the SS according to the information contained in the Ranging Response (RNG-RSP) message explicitly transmitted by the BS in the downlink. In case of contention-based transmission of a BW-REQ message, the information whether a collision has occurred is not transmitted by the BS and 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.

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 details in the next section as a part of our simulation model, which has been implemented in MatLab.

III. Model for the Random Access

For the purpose of investigating the random access in IEEE 802.16 the following model is introduced.

A. Simulation Model for the Random Access

Let us consider n "active" SSs, simply denoted by stations in the sequel, always having a request ready to transmit. Each frame comprises K equal slots for random access. The duration of a slot is chosen to be sufficient for the transmission of one request. The BS chooses K in order to make a trade-off between the duration of contention period and the duration of payload transmission 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. Furthermore, we assume that stations receive a feedback by the BS at the beginning of the next frame whether their transmission was successful.

For collision resolution a binary exponential backoff algorithm is introduced. Before each transmission attempt, a station uniformly chooses an integer number from the interval $[0, W_i\text{-}1]$, 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 W_0 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. Let us notice, that if $W_{min}<K$, then some time slots will never been used during the first transmission attempt. Furthermore, we set $W_{min} = lK$, where l is a natural number, in order to uniformly distribute the transmission attempts over the available random access slots.

B. Analytical Model for the Random Access

Following the approach of [3], let us assume that the behaviour of an arbitrary station does not depend on the behaviour of the other $n-1$ stations, and the conditional collision probability p, that a station transmits and fells into collision, is constant. Under such an assumption, we present two ways to analytically describe the above simulation model for the random access in the following.

1. Approach with Markov chain and binomial distribution

A two-step procedure is proposed: In the first step, a station uniformly chooses one of the L_i frames to transmit, where $L_i = 2^i l$, $i = 0,\dots,m$, and i 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 $c_1(t)$ 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 $c_1(t)$ equals to zero. Let $b(t)$ be the backoff stage of a station at the moment t. It is possible to model the two-dimensional process $\{c_1(t),b(t)\}$ by the Markov chain introduced in [3] with the following transition probabilities:

$$\begin{array}{ll} P\{i,k \mid i,k+1\} = 1, & k = 0,\ldots,L_i-2, i = 0,\ldots,m \\ P\{0,k \mid i,0\} = (1-p)/L_0, & k = 0,\ldots,L_0-1, i = 0,\ldots,m \\ P\{i,k \mid i-1,0\} = p/L_i, & k = 0,\ldots,L_i-1, i = 1,\ldots,m \\ P\{m,k \mid m,0\} = p/L_m, & k = 0,\ldots,L_{mi}-1 \end{array},$$

where we adopt the notation $P\{i_1,k_1 \mid i_0,k_0\}=P\{c_1(t+1)=i_1, b(t+1)=k_1 \mid c_1(t)=i_0, b(t)=k_0\}$

We omit the mathematical calculations and the detailed explanations as the chain is similar to the one contained in [3]. Summarizing the probabilities of the states when $c_1(t)$ equals to zero, the following equation for the probability of a station to transmit in a frame π_1 can be obtained:

$$\pi_1 = \frac{2(1-2p)}{(1-2p)(l+1)+pl(1-(2p)^m)}. \quad (1)$$

Let's consider that one particular station transmits in a frame. Then under this condition, the probability v_i, that i stations from the remaining n-1 transmit in the same frame, is equal to

$$v_i = \binom{n-1}{i}\pi_1{}^i(1-\pi_1)^{n-1-i}, \quad (2)$$

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

$$p = 1-\sum_{i=0}^{n-1}\binom{n-1}{i}\pi_1{}^i(1-\pi_1)^{n-1-i}\left(1-\frac{1}{K}\right)^i. \quad (3)$$

So, the non-linear system is represented by equations **(1)** and **(3)** with two unknowns p and π_1.

2. Approach with Markov chain with additional idle states

Here a discrete and integer time scale is also adopted, but with t and t+1 corresponding to the beginning of two consequent slots. Let $c_2(t)$ be the stochastic process representing the backoff counter of a station, i.e., integer number of slots it has to wait before transmission, at time t and $b(t)$ still is the backoff stage of a station at time t.

Note that according to the protocol rules, after a transmission attempt the station does not immediately starts the backoff process for the next transmission attempt like in the model in [3], but waits till the beginning of the next frame. That is, suppose the transmission happens in the k^{th} slot of the totally K, the station will wait for $(K - k)$ slots before it continues. Since the station always uniformly chooses one of K slots in a frame to transmit, the distribution of the number of slots that the station will wait is uniformly distributed over [0, K-1]. Denote $a(t)$ as the waiting time counter after a transmission.. Thus, the system can be modelled by the same Markov chain as the one described in [3] but in addition with K-1 idle states corresponding to the waiting time. The additional states are shown in Figure 2, where the transition probabilities are

$$\begin{array}{ll} P\{a(t+1) = j-1 | a(t) = j\} = 1 & j = 1,\ldots,K-1 \\ P\{a(t+1) = j | b(t) = 0\} = 1/K & j = 1,\ldots,K-1 \end{array}. \quad (4)$$

Let $b_{i,j} = \lim_{t\to\infty}P\{c_2(t) = i, b(t) = k\}$, $a_j = \lim_{t\to\infty}P\{a(t) = j\}$, i=0..m, k=0,…,W_i-1, j=1,…,K-1 be the stationary distribution of this ergodic chain.

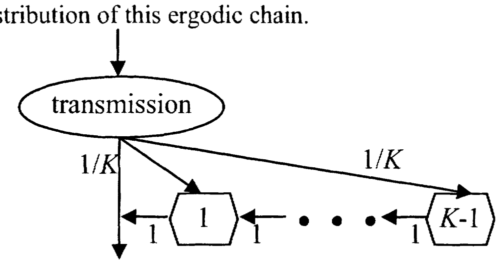

Figure 2: Additional idle states in the Markov chain

Since the probability of transmission is $\sum_{i=0}^{m} b_{i,0}$, from **(4)** we have

$$a_j = \frac{j}{K-j}\sum_{i=0}^{m} b_{i,0} \Rightarrow \sum_{j=1}^{K-1} a_j = \frac{K-1}{2}\sum_{i=0}^{m} b_{i,0} = \frac{K-1}{2}\cdot\frac{b_{0,0}}{1-p}$$

Then, from the normalization condition

$$1 = \sum_{i=0}^{m}\sum_{k=1}^{W_i} b_{i,k} + \sum_{j=1}^{K} a_j,$$

where the first item has been calculated in [3], $b_{0,0}$ is obtained. Thus, the probability π_2 that a station transmits in a randomly chosen slot time can be expressed as

$$\pi_2 = \sum_{i=0}^{m} b_{i,0} = \frac{2(1-2p)}{(1-2p)(Kl+1)+pKl(1-(2p)^m)+(K-1)(1-2p)}. \quad (5)$$

Since the conditional collision probability p is equal to the probability that at least one of the n-1 remaining stations transmit, it yields,

$$p = 1-(1-\pi_2)^{n-1}. \quad (6)$$

Now, the non-linear system is represented by equations **(5)** and **(6)** with two unknowns π_2 and p.

It can be proven that both systems, which are represented by equations **(1)** and **(3)** and by equations **(5)** and **(6)**, have a unique solution and the conditional collision probability p could be calculated numerically.

C. Derivation of the Delay

Now let us calculate the mean delay for the request transmission $\bar{d}$ (measured in number of frames). First, notice, that when the backoff window of a subscriber station is $W_j = 2^j Kl$ (j = 0,…,m), the average number of frames it has to wait is

$$\overline{N}_j = \frac{1}{2^j l}\sum_{i=1}^{2^j l} i = \frac{1+2^j l}{2}. \quad (7)$$

Then the mean delay has a geometrical distribution and can be calculated as

$$\bar{d} = (1-p)\left(\sum_{i=0}^{m} p^i \sum_{j=0}^{i} \bar{N}_j + \sum_{i=m+1}^{\infty} p^i \left(\sum_{j=0}^{m} \bar{N}_j + (i-m)\bar{N}_m\right)\right) \quad (8)$$

From **(8)** and taking into account **(7)** it can be derived (we omit the algebraic simplifications) the following equation for the mean delay:

$$\bar{d} = \frac{1}{2(1-p)} + \frac{1}{2(1-2p)} - l\frac{2^{m-1}p^{m+1}}{(1-2p)(1-p)}. \quad (9)$$

It is easy to show (see the next section) that for $m = 0$, both of the considered approaches give the same analytical result for the mean delay. When $m \neq 0$ it can be shown numerically, that both approaches lead to the same result, too. The difference between the approaches consists in computing the probabilities of different random events: the transmission in a frame and the transmission in a slot.

IV. Performance Evaluation of the Random Access in IEEE 802.16

In this paper, when values for the mean delay are plotted, the analytical results are represented in lines and the relative simulation results are represented as symbols on the lines. The analytical results meet the simulation ones quite well for varied system parameters. The cases, when there is a mismatching between the simulation and the analytical results are not revealed.

A. Optimization of window size for fixed n, K

The system performance is evaluated in terms of mean delay $\bar{d}$. Actually, minimizing it consequently maximizes the throughput of request transmission. For fixed n and K, the mean delay derived by equation (9) is a function of l and m. The pair of (l_{opt}, m_{opt}) that minimizes the mean delay exists and can be found numerically. In Table the mean delays obtained with different pairs of (l_{opt}, m) are listed, where l_{opt} is the value that minimize the mean delay for the given m. The difference of the mean delay for the different pairs are so slightly that can be ignored. Hence, it is reasonable to conclude that the optimum pair (l_{opt}, m_{opt}) is not unique, but for any value of m, there exists a certain l that achieves the minimum delay.

Table: The mean delay obtained with different pairs of (l_{opt}, m), $K = 6$

n	M	l_{opt}	$\bar{d}_{min}$, analytical
6	0	1	2.488
12	0	3	5.208
	1	2	5.214
	2	1	5.253
18	0	5	7.927
	1	3	7.928
	2	2	7.929
	3	2	8.002
	4	1	7.934

Now let us focus on the particular case when no backoff is considered (i.e. $m = 0$), the probability that a station transmits in a randomly chosen slot time π_2 is $2/(Kl+K)$ according to **(5)**, and so the conditional collision probability p can also be written out explicitly, which is

$$p = 1 - \left(1 - \frac{2}{Kl+K}\right)^{n-1}. \quad (10)$$

And the corresponding mean delay is

$$\bar{d} = \frac{1+l}{2(1-p)} = \frac{1+l}{2\left(1-\frac{2}{Kl+K}\right)^{n-1}}. \quad (11)$$

Taking the derivative of (11) with respect to l, and imposing it equal to 0, after some simplifications we obtain $(1 + l)(Kl - 2n + K) = 0$. The optimum value of l, l_{opt}, only depends on the ratio between n and K. When n/K is an integer, optimum l is $2n/K$-1 and the minimum delay is

$$\bar{d}_{\min,m=0} = \frac{n/K}{(1-1/n)^{n-1}}, n = iK, i = 1,2,\ldots \quad (12)$$

When $n < K$, the minimum possible value of l, which is 1, is chosen and the corresponding delay is

$$\bar{d}_{\min,m=0} = \frac{1}{(1-1/K)^{n-1}}, n < K. \quad (13)$$

It is obvious that increasing the number of slots K leads to the decreasing of the delay, since the delay is measured in unit of frame. Recall that each frame contains K slots, from (12), it can be seen that the delay only depends on n when it is measured in units of slots. In case that n/K is not an integer, either the larger closest integer or the smaller one is the optimum value. In Figure 3, it is shown the minimum delay as a function of n for different K.

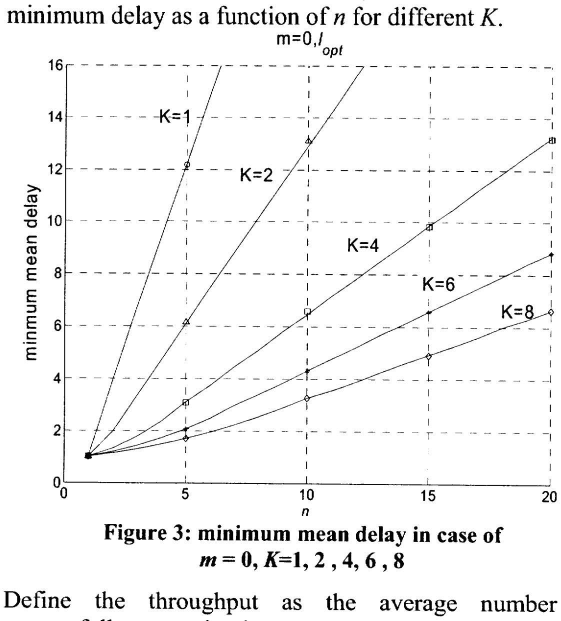

Figure 3: minimum mean delay in case of $m = 0$, K=1, 2, 4, 6, 8

Define the throughput as the average number of successfully transmitted requests per contention slot. It is equal to the probability of the successful transmission in one slot, which can be written as

$$\rho = n\pi_2(1-\pi_2)^{n-1}. \quad (14)$$

Since minimizing the mean delay is the same as maximizing the throughput, with $m = 0$ and the optimum value of l for the minimum mean delay, the maximum throughput can be achieved. π_2 is $2/(Kl+K)$ when m=0, and according to **(14)**, substituting l_{opt} we have the expression for the maximized throughput as

$$\rho_{\max,m=0} = \begin{cases} (1-1/n)^{n-1} & ,n = iK \\ \frac{n}{K}(1-1/K)^{n-1} & ,n < K \end{cases} \quad (15)$$

where i is any positive integer number .

B. Sensitivity of delay for variable number of stations

From above, it is concluded that once the number of stations n is known, there are many pairs of (l, m) for a fixed number of K that minimizes the delay. In reality, we might not know the actual number of active stations in the system but only an estimation, denoted as $\hat{n}$. And we use l and m that are optimized for the estimated n. Hence, it is important to check the sensitivity of delay with respect to the estimation error $\Delta n = n - \hat{n}$. We define the delay degradation as the difference between the mean delay and the lower bound, which is the value when using the optimized l and m for the actual n. In Figure 4 it is shown that the delay degradation is smaller for the pair with larger m (or smaller l). So when sensitivity is under consideration, it is better to start with a small contention window (choose the pair with l =1) and resolve the collisions in the succeeding frames.

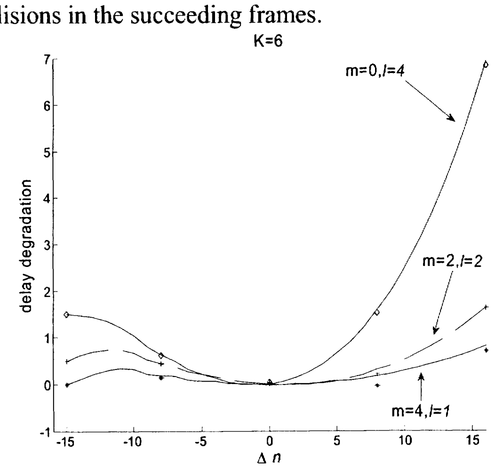

Figure 4: delay degradation in case of $K = 6$, $\hat{n} = 16$

C. Performance for different n with (l = 1, m_{opt})

From the previous discussion it becomes clear that for a stable operation the first transmission should take part in the same frame (l = 1) and the maximum backoff, m_{opt}, should be appropriately chosen as function of the number of stations, n to minimize the delay. The optimized performance of the system for this selection is shown as solid curves in Figure 5. Since the performance with pair (l = 1, m_{opt}) is the same as the one with pair (l_{opt}, m = 0), the optimized performance is expressed by **(12)**, **(13)** and **(15)**.

On the one hand it can be noticed, that the difference between the performance with m_{opt} and the one with m>m_{opt} is almost negligible. For instance, when n = 5, m_{opt} is 0. But even setting m to 4, the throughput decreases a bit and the mean delay remains almost the same. However, on the other hand, if m is selected to be smaller than m_{opt}, the performance degrades significantly, which leads to rapidly decreased throughput and increased delay.

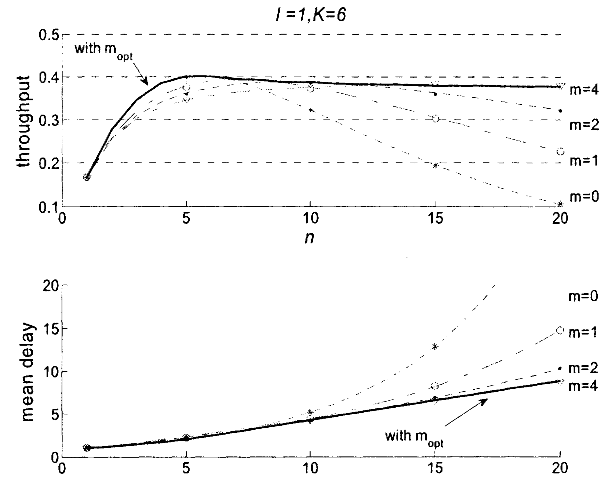

Figure 5: Throughput and mean delay with (l =1, m_{opt}), K=6

V. CONCLUSIONS

In this paper the performance of the random access scheme specified in IEEE 802.16 is investigated. Two approaches to analytically model the considered system are introduced. Using the simulation model it is shown that these approaches give very accurate estimation of the mean delay for request transmission. The mean delay can be minimized by selecting proper values of l and m when the number of station n is known. Using any optimized pair of these parameters leads to the same value of minimum delay. Further investigations show that if only an estimation of n is given, larger m or equivalently smaller l is preferred to reduce the sensitivity with respect to estimation errors.

ACKNOWLEDGMENT

This work has been performed in the framework of the IST project IST-2003-507581 WINNER [4], which is partly funded by the European Union. The authors would like to acknowledge the contributions of their colleagues, although the views expressed are those of the authors and do not necessarily represent the project.

REFERENCES

[1] IEEE Std. 802.16-2004, IEEE standard for local and metropolitan area networks, Part 16: Air Interface for Fixed Broadband Wireless Access Systems.

[2] *Guo S., Sydor J., Brandao A.* Performance Analysis of the P2MP Metropolitan Area Broadband Wireless Access Network, IEEE PIMRC, 2003.

[3] *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.

[4] WINNER - Wireless World Initiative New Radio, IST-2003-507581, https://www.ist-winner.org.