

Slow and Steady: Modelling and Performance Analysis of the Network Entry Process in IEEE 802.16

Matthias Hollick, Parag S. Mogre, Christian Schott, and Ralf Steinmetz

Multimedia Communications Lab (KOM), Department of Electrical Engineering and Information Technology, Technische Universität Darmstadt, Merckstr. 25, 64283 Darmstadt, Germany.

Email: [matthias.hollick;parag.mogre]@kom.tu-darmstadt.de

Abstract—The IEEE 802.16 standard represents the state-of-the-art for quality of service (QoS) aware broadband wireless access networks. Initially the standard provided sophisticated QoS mechanisms only for a static network (i.e. stationary subscribers), but recently the amendment IEEE 802.16e-2005 introduced mobility support for the individual subscriber stations. Network entry is the first step required for nodes to register themselves with the 802.16 network. The performance of the network entry process is crucial to support QoS, especially if node churn is high. In this paper, we develop an analytical model of the network entry process in IEEE 802.16. This model enables us to predict the effect and influence of important protocol parameters for joining the network. Using our model, we show that the selection of correct and appropriate protocol parameters is crucial to support setup-delay sensitive applications such as emergency services or IP-telephony. Our model can be used as a tool to derive the feasible range for individual network parameters in selected application scenarios.

I. INTRODUCTION

The IEEE 802.16 standard (often referred to as WiMAX™) represents the state-of-the-art for broadband wireless access technologies to support QoS. 802.16 was originally designed as a replacement for wireline technologies, which operate on the last-mile to the subscriber. Recently, the amendment and corrigendum IEEE 802.16e-2005 introduced mobility support for the individual subscriber stations.

IEEE 802.16 is a very flexible technology and enables a variety of services and applications. In particular, multimedia services, peer-to-peer applications, high-bandwidth data-services, etc. are envisioned. Moreover, the unique combination of wireless broadband technology with sophisticated QoS mechanisms enables the use in emergency response scenarios, where the wireless network is established as a replacement for possibly damaged infrastructure networks and needs to support QoS-aware voice-, video-, and data-communication (Ref. [1] discusses further QoS challenges in wireless mesh networks).

QoS in 802.16 has been researched mostly with focus on scheduling disciplines. However, the network entry process, which is the first step for nodes joining the network or handing over into the network, has not yet been analysed in detail. Network entry refers to the initial procedure subscriber stations perform to register themselves with the base station that controls the 802.16 network. We consider the performance of the initial network entry as well as the entry due to handover crucial to QoS support in 802.16. This is especially true, if we consider QoS-sensitive applications such as wireless

communication networks for emergency response, where a large number of subscribers are likely to join the network during a short period of time. In our work, we contribute

- a thorough dissection of 802.16's network entry process, a mathematical model to derive the upper and lower bounds for the network entry delay, and
- an analysis of key parameters influencing the performance of joining a 802.16 network.

The paper is structured as follows. Section II briefly introduces the IEEE 802.16 standard and surveys related work. In Section III, the network entry process is thoroughly discussed. Section IV describes our mathematical model to calculate the delay-bounds for 802.16's network entry procedure. In Section V we place the most important protocol parameters under scrutiny and present selected results. We conclude this paper in Section VI and point to future work.

II. BACKGROUND AND RELATED WORK

This section gives a brief introduction to the IEEE 802.16 standard as specified in Refs. [2], [3]. Further, we discuss related work that models performance aspects of 802.16.

A. Basic Overview of IEEE 802.16

IEEE 802.16 specifies a *Point-to-Multipoint (PMP)* mode of operation, which enables direct (one-hop) communication between the *Base Station (BS)* and the *Subscriber Stations (SS)* as well as an optional *MeSH* mode of operation that introduces multi-hop communication between the stations. Here, we focus on the PMP mode of operation.

The standard covers *Physical (PHY)* layer and *Medium Access Control (MAC)* layer mechanisms. As already mentioned, the standard has been designed to be very flexible, which results in different PHY profiles specified (for 10 – 66 GHz: WirelessMAN-SC, for below 11 GHz: WirelessMAN-SCa, Human, -OFDM, and -OFDMA). Most profiles can either operate in Time Division Duplex (TDD) and Frequency Division Duplex (FDD). In our work, we refer to the basic profiles WirelessMAN-SC and -SCa. Here, the uplink uses a combination of Time Division Multiple Access (TDMA) and Demand Assigned Multiple Access (DAMA); the downlink channel uses TDM. The time-frames (with durations between 0.5 ms and 20 ms) are divided into slots that can be assigned dynamically by the BS for management or user traffic (cf. Ref. [2], pp.307).

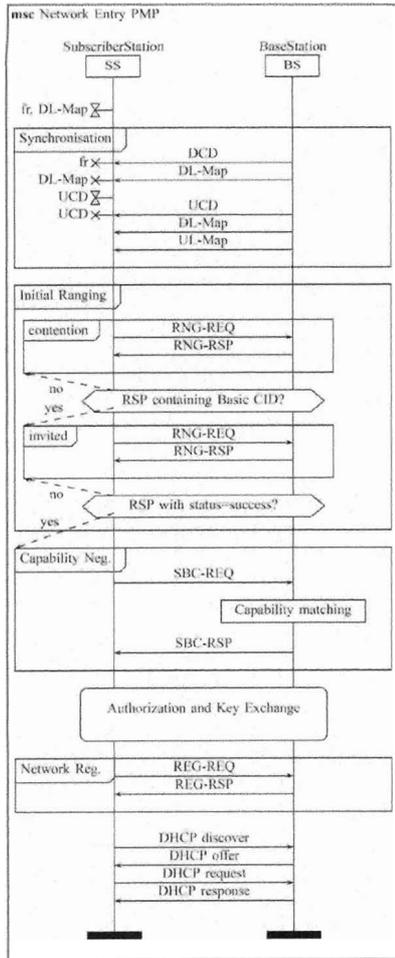


Fig. 1. Network entry procedure in PMP mode of operation.

802.16 operates using a reservation-based MAC scheme. It provides sophisticated mechanisms to support QoS. In the PMP mode the BS is solely responsible for bandwidth allocation and scheduling by allocating time-slots for transmissions of both, upstream and downstream data flows. The scheduling services provided are unsolicited grant service (UGS), real-time polling service (rtPS), extended real-time polling service (ertPS), non-realtime polling service (nrtPS), and best-effort service (BE) (see Refs. [2], [3]).

To request bandwidth, a node has to first join the network and perform an authentication and key-exchange process, the so-called network entry procedure. During network entry, the joining SS establishes different types of connections (basic connection, primary and secondary management connection), which are used for control traffic; for data communication different connections (unicast or multicast transport connections) are established on demand. For additional details the interested reader is referred to articles of tutorial nature covering 802.16's PMP [4] or the optional MeSH [5] mode of operation.

TABLE I
TIMERS/EVENTS AND CONDITIONS/ACTIONS DURING NETWORK ENTRY

Timer/Event	Condition/Action
fr	The SS waits for a DL-message to extract framing information and establish physical synchronisation (PHY-Sync).
DCD	The DCD defines the characteristics of the downlink physical channel.
DL-Map	The first received DL-Map permits the SS to extract information on the DL-Channel needed to establish MAC-Sync. Upon receipt of the UCD, the next DL-Map allows for time synchronisation with the BS.
UCD	From the UCD, the SS retrieves uplink channel parameters and decides whether the chosen uplink is usable.
UL-Map	The UL-Map contains the current bandwidth allocation scheme for the subsequent transmissions of the SS.

B. Related Work

Related work that models performance aspects of 802.16 mainly focuses on the different PHY profiles and the QoS-aware scheduling of MAC time-frames, respectively. Work of particular interest includes articles describing the QoS support in 802.16 networks [6], [7], however, the network entry process has not been included in detail for the study of the QoS capabilities of the standard. A body of work deals with scheduling algorithms to optimally schedule time-frames, Refs. [8], [9] being examples, or to propose data structures for efficient handling of the complex QoS state information [10]. With the advent of mobility support in 802.16, work has been devoted to the analysis and optimisation of the handover process [11]. To the best of our knowledge, as of today, there exists no work that mathematically models the network entry process.

III. DISSECTING THE NETWORK ENTRY PROCESS

This chapter describes the detailed working of the network entry procedure and serves as a basis for our mathematical model. In particular, we describe the individual steps a node performs upon entering the network (see Fig. 1):

- (A) Synchronisation
- (B) Initial Ranging (contention-based and invited)
- (C) Basic Capability Negotiation
- (D) Authorization and Key Exchange
- (E) Network Registration

The detailed network entry process is shown in the message sequence chart in Fig. 1; the corresponding timers/events and conditions/actions are detailed in Table I. In order to connect to a BS, the SS first scans all available downlink channels that are periodically advertised by each BS via a so-called *Downlink Channel Descriptor (DCD)* containing basic information about the specific channel. Afterwards the SS waits for the *Downlink-Map (DL-Map)*.

(A) The SS obtains physical **synchronisation** by means of the synchronisation sequence of the DCD-frame as well as MAC-synchronisation by decoding of the DL-Map. The SS next waits for an *Uplink Channel Descriptor (UCD)* informing the SS about the uplink channel properties and whether the

channel is appropriate for the SS. If not, the SS searches for another downlink (DL) and corresponding uplink (UL).

(B) As soon as an appropriate UL and DL are discovered, the SS extracts the timing information from the next DL-Map received. It then starts the **initial ranging** cycle by sending a *Ranging-Request (RNG-REQ)* to the BS using the **contention** slot (IR-slot) reserved for *Initial Ranging (IR)*. Upon receipt of the RNG-REQ message the BS answers by submitting a *Ranging-Response (RNG-RSP)* message containing information for the SS on how to adapt its transmission parameters (i.e. correct timing offset and power adjustment). The RNG-RSP can also (optionally) provide a basic and a primary management *Connection ID (CID)* by which the SS can continue ranging upon **invitation** by the BS during slots reserved for this CID. The initial ranging process continues to exchange ranging messages until both SS and BS can communicate correctly.

(C) The SS commences the **capability negotiation** in order to exchange the common capabilities of BS and SS. The SS initiates a *Send Basic Capabilities Request (SBC-REQ)* message to the BS. The BS in turn calculates the common capabilities of itself and the SS and returns the intersecting capabilities using a *Send Basic Capabilities Response (SBC-RSP)* message.

(D) The next step is the **authorization and key exchange** to enable authenticated and encrypted communication between BS and SS. All necessary cryptographic keys for securing the communication on link layer are exchanged and a security association is established.

(E) Finally the SS performs the **network registration** with the BS by sending a *Registration Request (REG-REQ)* message. The REG-REQ indicates whether the SS is managed or unmanaged. In the case of a managed SS, the answer, a *Registration Response (REG-RSP)*, contains the secondary management CID. The SS completes the network entry process by establishing IP-connectivity via DHCP and by synchronising time via NTP, if needed. During normal operation, the BS and SS may exchange further operational parameters; the SS may join further channels or the BS may instruct the SS to listen on further provisioned channels such as optional multicast channels.

The refined standard IEEE 802.16e-2005 [3] defines (optional) process optimisations for network re-entry and handover. In particular, a SS may acquire the BS-IDs of neighbour BSs, initial ranging intervals may be reserved by the BS at neighbour BSs via messages exchanged over the backhaul network, etc. The detailed procedures related to mobility are outside the scope of this paper due to space constraints and left to future work.

IV. MODELLING THE NETWORK ENTRY PROCESS

In this chapter, we develop a mathematical model to calculate the overall time to join an IEEE 802.16 network. We introduce all necessary assumptions and derive the formulae to enable the estimation of the delay-penalty for entering the network. The model variables are summarised in Table II. In

TABLE II
MODEL VARIABLES

Variable Name	Description
$T_{\text{reception}}$	Overall delay of individual packets from queuing at the sender until reception at the receiver
$T_{\text{S-queue}}$	Queuing delay at the sender's side
$T_{\text{R-queue}}$	Time from receipt of a message until it is processed at the receiver's side
T_{xmit}	Transit time from receiver to sender
P_{slot}	Probability of a slot to be chosen
$P_{\text{slot}}(x \text{none before})$	Probability of a specific slot to be chosen while skipping the ones before
i	Node state, a retransmission increases i by 1
x_i	Chosen random opportunity in state i
$N_{\text{IR},n}$	Number of opportunities in frame n
r	Backoff factor (here: 2 for binary exp. backoff)
BoS	Backoff start value
BoE	Backoff end value
W_i	Contention window size in node state i
$W_0 = r^{\text{BoS}}$	Contention window size for first transmission
$W_{\text{max}} = r^{\text{BoE}}$	Maximum contention window size
$\sigma(x)$	One-unit-step-function
	$\sigma(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$
T_{slot}	Duration of a ranging slot
T_{frame}	Duration of a time-frame
$T_{\text{UL}, T_{\text{DL}}}$	Duration of UL, DL subframe
c_0	Speed of light
T_3	Timeout (cf. Ref. [2], Table 342, pp.637)
$T_{\text{hardware-transit}}$	Delay a packet observes in hardware
$T_{\text{RNG-proc-BS},r}$	BS's processing time
$T_{\text{RNG-proc-SS},r}$	SS's processing time
k_{cont}	# of contention-based ranging message pairs
k_{inv}	# of invited ranging message pairs
P_i	Relative frequency that a node enters state i in steady state
S_i	Time a node stays in state i
p_c	Collision probability for a transmission
p_t	Probability that a node transmits in an arbitrary timeslot
M	Maximum retransmission count
n_r	Average number of retransmissions per packet
T_n	Duration until a node successfully transmits on the n -th retransmission
P_{drop}	Drop probability
\bar{D}_{Σ}	Average medium access delay
$\bar{T}_{\text{BO-delay}}$	Average overall backoff delay
$T_{\text{startPrivNeg}}$	Time until privacy negotiations start
T_{startIR}	Time until SS may start initial ranging
$T_{\text{IR-cont}}$	Time for contention-based initial ranging
$T_{\text{IR-inv}}$	Time for invited initial ranging

our calculations we are referring to the WirelessMAN-SCa profile of Ref. [2], the adaptation to the other profiles of 802.16 is straightforward.

A. Entering the Network

As has been explained in Section III, a SS willing to join an IEEE 802.16 network first establishes the basic communication channels with the BS. We, in the next subsections, derive models to describe (A) synchronisation, (B) initial ranging (contention-based and invited), and (C) capability negotiation. We then summarise our findings.

Every single message exchanged experiences a certain delay from the time where it enters the sender's MAC sending queue

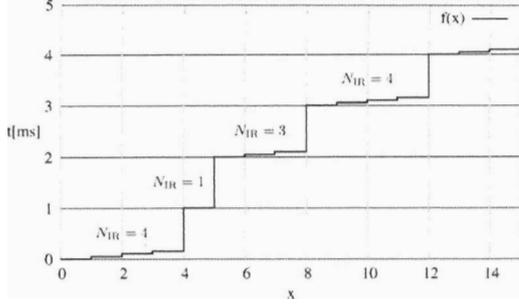


Fig. 2. Given a randomly chosen number out of the initial ranging backoff window, the SS has to wait for this opportunity until it may send its first RNG-REQ. Each SS initialises its internal counter starting with the first IR-Slot it recognises. In the figure, each frame has N_{IR} initial ranging slots advertised in its UL-Map. The frame size is set to 1 ms, the initial ranging slot size is set to 0.05 ms for better visibility.

until it is received by the receiver. The overall time to join the network needs to take this delay into account for all messages exchanged. Please note that there are different queues based on the QoS classes for the specific connections (basic, primary and secondary management, unicast-data, etc.). The time until reception $T_{\text{reception}}$ can be calculated by Eqn. (1):

$$T_{\text{reception}} = T_{\text{S-queue}} + T_{\text{xmit}} + T_{\text{R-queue}} \quad (1)$$

$T_{\text{S-queue}}$ is a function of the load of the network, which is primarily related to the number of nodes in the network and their transmission activity (in the same QoS class). The transmission time T_{xmit} , describes the time a packet is “on the air”, which depends on the distance between the two nodes and the chosen PHY profile. The queuing time on the receiver’s side $T_{\text{R-queue}}$ is a function only of the receiver’s processing delay, which can be assumed to be nearly constant.

B. Initial Ranging

Initial ranging can be divided into two phases: 1.) contention-based initial ranging and 2.) invited initial ranging. Since a node entering the network is not known a priori, it has to contend against other nodes in the first phase of the initial ranging process. After the BS has knowledge of the node, it can allocate bandwidth for the further network entry process to ensure collision freedom of the subsequent message exchange. We now discuss both phases of the initial ranging procedure.

1) *Contention-based Initial Ranging:* The BS provides uplink slots and downlink slots to the connected SSs. For initial ranging, the UL-Map reserves special transmission opportunities. The number of these slots is controlled by the BS on per-frame basis. An entering SS competes with all other entering SSs for these slots. Also, the BS can adjust the initial ranging backoff window defined in the UCD message by the ranging *backoff start* values ($\text{BoS} \in \{2^0 \dots 2^{15}\}$). A SS willing to join the network sends its first ranging request at a random slot in this backoff interval once the message is assembled and the necessary information of the channel is gathered by reception of the UCD and the DCD. Each slot has

the probability of P_{Slot} to be chosen if we assume a random distribution. The corresponding probability is given by Eqn. (2).

$$P_{\text{Slot}} = \frac{1}{W_i} \quad \text{where } W_i \in \{2^{\text{BoS}} \dots 2^{\text{BoE}}\} \quad (2)$$

Having chosen the ranging opportunity in which the SS sends its RNG-REQ, it has to wait for this specific slot.

We derive the equation for the conditional probability for the SS to send in the exact IR-slot n , passing over all slots before, to be:

$$P_{\text{Slot}}(n) = (1 - P_{\text{Slot}})^{n-1} P_{\text{Slot}} \quad (3)$$

See Fig. 2 for a visualisation of the waiting time vs. the randomly selected IR-slot number x .

We derive the average time until this first message can be sent to be:

$$\begin{aligned} T_{\text{S-queue}} = & \sigma(x_i) T_{\text{Slot}} \cdot x_i \\ & + (T_{\text{Frame}} - N_{\text{IR}0} T_{\text{Slot}}) \sigma(x_i - N_{\text{IR}0}) \\ & + (T_{\text{Frame}} - N_{\text{IR}1} T_{\text{Slot}}) \sigma(x_i - (N_{\text{IR}0} + N_{\text{IR}1})) \\ & + (T_{\text{Frame}} - N_{\text{IR}2} T_{\text{Slot}}) \sigma(x_i - (N_{\text{IR}0} + N_{\text{IR}1} + N_{\text{IR}2})) \\ & + \dots \end{aligned} \quad (4)$$

σ describes the unit-step function and $N_{\text{IR}n}$ the amount of ranging slots in Frame n . x_i denotes the random number chosen by the SS. This leads to:

$$\begin{aligned} T_{\text{S-queue}} = & \sigma(x_i) T_{\text{Slot}} \cdot x_i \\ & + \sum_{n=0}^{\infty} \left((T_{\text{Frame}} - N_{\text{IR}n} T_{\text{Slot}}) \sigma \left(x_i - \sum_{j=0}^n N_{\text{IR}j} \right) \right) \end{aligned} \quad (5)$$

Note that in Eqn. (5) the initial ranging interval (IR-Int), i.e., the “time between Initial Ranging Regions assigned by the BS” [2, p.638]) has not been taken explicitly into account. In other words, a SS might be forced to wait for maximal 2s until the next contention ranging opportunity is scheduled by its BS. Without loss of generality, we omit this characteristic for our model formulation.

Having sent the first RNG-REQ in a contention window of size $W_0 = r^{\text{BoS}}$, the SS waits for a RNG-RSP from the BS until the timeout T_3 occurs, which is set by default to its maximal value, 200 ms (cf. Ref. [2], Table 342, pp.637). If the timeout is triggered, the SS reschedules a RNG-REQ using a binary exponential backoff scheme, choosing a random number out of the doubled contention window W_1 as long as its size stays below the $W_{\text{max}} = r^{\text{BoE}}$ value provided in the UCD. At non-reception after another T_3 , the same procedure happens and the backoff window size is doubled again, until the maximal number of *Contention Ranging Retries*, (*CRR*) (minimum CRR:16, cf. Ref. [2], Table 342, pp.637) is reached. The time until receipt of the initial RNG-REQ at the BS in the case of erroneous receptions can be modelled as

$$\begin{aligned}
T_{\text{reception}} &= T_{\text{xmit}} \\
&+ \sum_{i=0}^{\text{CRR}-1} (T_3 + T_{\text{S-queue}}(x_i, N_{\text{IRn}})) + T_{\text{R-queue}} \\
\text{with } P(x_i) &= \begin{cases} \frac{1}{2^{i \cdot \text{BoS}}} & \text{for } 2^i \cdot \text{BoS} < \text{BoE} \\ \frac{1}{\text{BoE} - \text{BoS}} & \text{for } 2^i \cdot \text{BoS} \geq \text{BoE} \end{cases} \\
\text{and } T_{\text{xmit}} &= \frac{c_0}{\text{distance}} + T_{\text{hardware-transit}} \\
&+ \text{PacketLength} \cdot \text{SymbolRate}
\end{aligned} \tag{6}$$

Eqn. (6) describes the time until the BS receives the RNG-REQ. The BS processes the request ($T_{\text{RNG-proc-BS}}$), identifies the MAC-address of the SS and answers with a RNG-RSP. If these messages cannot be transferred during one frame (UL- or DL-subframe with duration of T_{UL} or T_{DL}) they have to be split. The SS next processes the response ($T_{\text{RNG-proc-SS}}$) and either continues contention-based ranging, or waits for an invited ranging opportunity. In case of an erroneous reception of the RNG-REQ at the BS, it may choose either to not answer at all, or to send a RNG-RSP specifying not the SS's MAC-address but the ranging slot and frame number used to send the request. The latter is an optional feature with only a marginal effect on the timing properties; we, without loss of generality, do not further discuss it in this paper. The overall ranging procedure measured in contention slots depends on how many ranging message pairs have to be exchanged, (k_{cont}), until the SS is provided bandwidth for invited ranging. For the value of k_{cont} , no values have been published at the time of writing but need to be obtained in deployments over time. Thus, for the overall contention-based ranging procedure, we get:

$$T_{\text{IR-cont}} = k_{\text{cont}} \cdot (T_{\text{reception,r}} + T_{\text{Frame}} + T_{\text{RNG-proc-BS,r}} + T_{\text{RNG-proc-SS,r}}) \tag{7}$$

where the values are rounded to multiples of frames as follows:

$$T_{\text{reception,r}} = T_{\text{Frame}} \cdot \left\lceil \frac{T_{\text{reception}}}{T_{\text{Frame}}} \right\rceil \tag{8}$$

$$T_{\text{RNG-proc-BS,r}} = \begin{cases} 0 & \text{for } (T_{\text{reception}} \bmod T_{\text{Frame}}) + T_{\text{RNG-proc-BS}} < T_{\text{UL}} \\ T_{\text{Frame}} \cdot \left\lceil \frac{T_{\text{RNG-proc-BS}}}{T_{\text{Frame}}} \right\rceil & \text{else} \end{cases} \tag{9}$$

$$T_{\text{RNG-proc-SS,r}} = \begin{cases} 0 & \text{for } T_{\text{RNG-proc-SS}} < T_{\text{DL}} - T_{\text{RNG-Rsp}} \\ T_{\text{Frame}} \cdot \left\lceil \frac{T_{\text{RNG-proc-SS}}}{T_{\text{Frame}}} \right\rceil & \text{else} \end{cases} \tag{10}$$

Until now, we considered only the case of a single SS entering the network. In reality, a number of SSs compete for the contention slot. 802.16 proposes a truncated exponential backoff scheme to resolve the possible collisions of SSs. Please note that our model to describe this truncated exponential backoff for network entry also holds for other contention mechanisms in 802.16. We base our model on the work of Kwak, Song, and Miller who have analysed exponential

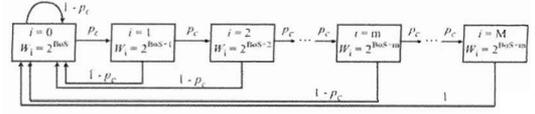


Fig. 3. State transition diagram in steady state for a truncated exponential backoff with maximum retry limit M .

backoff algorithms in [12]. In particular, we extend their model to describe the truncated backoff mechanism included in the IEEE 802.16 MAC, where the window size will not grow beyond a specified value *backoff end* (BoE) value. The model works with the transitions of node states where states are defined by i . Each node has a packet to transmit in its initial state $i = 0$ and switches states after transmission (be it unsuccessful or successful). As soon as the packet is either correctly submitted or dropped, the node switches to state $i = 0$ again. In case where the timeout occurs and the packet is considered lost, the node's state switches to state $i + 1$. For $M > m$, where M denotes the maximal retransmission count and m is the maximum window size (see Fig. 3), we can use the derivation from [12] to describe the contention window size being in a specific state i as:

$$W_i = \begin{cases} r^i W_0, & \text{for } i = 0, 1, \dots, m \\ r^m W_0, & \text{for } i = m, m + 1, \dots, M \end{cases} \tag{11}$$

For the binary exponential backoff in 802.16, the backoff factor $r = 2$. In the case where $M \leq m$, the contention window size is $W_i = r^i W_0$, $i = 0, 1, \dots, M$. We next extend the model in Ref. [12] (binary exponential backoff with maximal retry count) to match the case of IEEE 802.16, i.e. exponential backoff with a backoff window start size (BoS), a backoff window end size (BoE) and a maximal retry count. We have to adapt m as defined in Ref. [12] to describe BoE in the IEEE 802.16 standard. The relation of these variables is:

$$\begin{aligned}
r^{\text{BoE}} = W_0 r^m = r^{\text{BoS}} r^m &\Leftrightarrow \frac{r^{\text{BoE}}}{r^{\text{BoS}}} = r^m \\
&\Rightarrow m = \text{BoE} - \text{BoS}
\end{aligned} \tag{12}$$

Eqn. (13) (see Ref. [12]) describes the time a node stays in state i (P_i denotes the relative frequency that a node enters state i in steady state, p_c is the probability that a transmission experiences a collision). In order to include only transmissions where $i = m < M$ (which leads to window size $W_i = W_0 r^m$ for these transmissions) we obtain Eqn. (14):

$$S_i = \frac{P_i \bar{d}_i}{\sum_{j=0}^M P_j \bar{d}_j} = \frac{p_c^i (W_i + 1)}{W_0 \sum_{j=0}^M (r p_c)^j + \sum_{j=0}^M p_c^j} \tag{13}$$

$$\begin{aligned}
S_i &= \frac{P_i \bar{d}_i}{\sum_{j=0}^M P_j \bar{d}_j} \\
&= \frac{p_c^i (W_i + 1)}{W_0 \sum_{j=0}^m (r p_c)^j + W_0 \sum_{j=m+1}^M r^m p_c^j + \sum_{j=0}^M p_c^j}
\end{aligned} \tag{14}$$

This leads us to Eqn. (15) for the probability that a node is in state i and the backoff timer has expired ($k = 0$), i.e., the queued packet is sent. Similar to Eqn. (42) in Ref. [12], the probability of a node transmitting in an arbitrary time slot, $p_t = \sum_{i=0}^M s_{i,0}$, is obtained as shown in Eqn. (16).

$$s_{i,0} = \frac{2p_c^i}{W_0 \sum_{j=0}^m (rp_c)^j + W_0 \sum_{j=m+1}^M r^m p_c^j + \sum_{j=0}^M p_c^j} \quad (15)$$

$$p_t = \frac{2}{W_0 \left(\sum_{i=0}^m (rp_c)^i + \sum_{i=m+1}^M r^m p_c^i \right) / \sum_{i=0}^M p_c^i + 1} \quad (16)$$

As the numerical value of p_t is also “constrained by the fact that p_c can be expressed in terms of p_t ”, the determination of p_t is possible by means of Eqn. (19) of Ref. [12], where N is the number of nodes in the network (in our case the ones competing for the contention slots):

$$p_t = 1 - (1 - p_c)^{1/(N-1)} \quad (17)$$

The joint solution of Eqn. (17) and Eqn. (16) yields the value of p_c for given N , W_0 , m and M , which can be calculated numerically. This enables us to calculate the expected medium access delay for the IEEE 802.16 backoff mechanism. Consistent with [12], we define n_R as the average number of retransmissions per packet. The average required number of transmissions per packet for successful transmission is:

$$n_R = E[N_R | \text{no drop}] = \sum_{n=0}^M n \frac{T_n}{1 - P_{\text{drop}}} \quad (18)$$

$$= \frac{p_c}{1 - p_c} - \frac{(M+1)p_c^{M+1}}{1 - p_c^{M+1}} \quad (19)$$

$$n_R + 1 = \frac{1}{1 - p_c} - \frac{(M+1)p_c^{M+1}}{1 - p_c^{M+1}}$$

We have to adapt the derivation of the medium access delay \bar{D}_Σ to match the case of 802.16 and adapt Eqn. (48), Ref. [12]:

$$\bar{D}_\Sigma = E \left[\sum_{i=0}^{N_R-1} (D_i + 1) + D_{N_R} \mid \text{no drop} \right] \quad (20)$$

Using $\bar{d}_i = E(D_i + 1) = \frac{W_i + 1}{2}$ (Ref. [12], Eqn. (4)) we get

$$\bar{D}_\Sigma = E_{N_R} \left[\sum_{i=0}^{N_R-1} \frac{W_i + 1}{2} + \frac{W_{N_R} + 1}{2} - 1 \mid \text{no drop} \right] \quad (21)$$

$$= E_{N_R} \left[\sum_{i=0}^{N_R} \frac{W_i + 1}{2} \mid \text{no drop} \right] - 1$$

$$= \frac{1}{2} E_{N_R} \left[\sum_{i=0}^{N_R} W_i + \sum_{i=0}^{N_R} 1 \mid \text{no drop} \right] - 1$$

$$= \begin{cases} \frac{1}{2} E_{N_R} \left[(N_R + 1) + \sum_{i=0}^m W_0 r^i \right. \\ \quad \left. + \sum_{i=m+1}^{N_R} W_0 r^m \mid \text{no drop} \right] - 1 & \text{for } m < N_R \leq M \\ \frac{1}{2} E_{N_R} \left[\frac{W_0 (1 - r^{N_R+1})}{1 - r} + (N_R + 1) \mid \text{no drop} \right] - 1 & \text{else} \end{cases}$$

We are interested in the case for $m < N_R \leq M$ and obtain Eqn. (22). Next, Eqn. (23) models the duration T_n for a node successfully transmitting on the n -th retransmission. Eqn. (24) describes the probability P_{drop} that a packet will be dropped after reaching the maximum allowed retransmission-count (see Eqns. (43),(44) in [12]). Using these, [12] calculates the expected value for $E[r^{N_R+1} | \text{no drop}]$ as shown in Eqn. (25).

$$\bar{D}_{\Sigma,1} = \frac{1}{2} E_{N_R} \left[(N_R + 1) \right. \quad (22)$$

$$\left. + W_0 \left(\sum_{i=0}^m r^i + \sum_{i=m+1}^{N_R} r^m \mid \text{no drop} \right) \right] - 1$$

$$= \frac{1}{2} E_{N_R} \left[(N_R + 1) \right.$$

$$\left. \cdot W_0 \left(\frac{1 - r^{m+1}}{1 - r} + (N_R - (m + 1)) r^m \right) \mid \text{no drop} \right] - 1$$

$$T_n = (1 - p_c) p_c^n \quad n = 0, 1, \dots, M \quad (23)$$

$$P_{\text{drop}} = p_c^{M+1} \quad (24)$$

$$E[r^{N_R+1} | \text{no drop}] = \sum_{n=0}^M r^{n+1} \frac{T_n}{1 - P_{\text{drop}}} \quad (25)$$

$$= \frac{(1 - p_c)r}{1 - p_c^{M+1}} \sum_{n=0}^M (rp_c)^n$$

Including (19) in (21) and substituting the values in (21) by (25) and (18) yields

$$\bar{D}_{\Sigma,1} = \frac{1}{2} \{ (n_R + 1) \quad (26)$$

$$+ W_0 \left[\frac{1 - r^{m+1}}{1 - r} + (n_R + 1 - (m + 2)) r^m \right] \} - 1$$

$$= \frac{1}{2} \left\{ \left(\frac{1}{1 - p_c} - \frac{(M+1)p_c^{M+1}}{1 - p_c^{M+1}} \right) W_0 \left[\frac{1 - r^{m+1}}{1 - r} \right. \right.$$

$$\left. + \left(\frac{1}{1 - p_c} - \frac{(M+1)p_c^{M+1}}{1 - p_c^{M+1}} - (m + 2) \right) r^m \right] \} - 1$$

Finally, we obtain the expected medium access delay \bar{D}_Σ measured in transmission opportunities:

$$\bar{D}_\Sigma = \begin{cases} \frac{1}{2} \left\{ \left(\frac{1}{1 - p_c} - \frac{(M+1)p_c^{M+1}}{1 - p_c^{M+1}} \right) W_0 \right. \\ \quad \cdot \left[\frac{1 - r^{m+1}}{1 - r} + \left(\frac{1}{1 - p_c} - \frac{(M+1)p_c^{M+1}}{1 - p_c^{M+1}} - (m + 2) \right) \right. \\ \quad \left. \cdot r^m \right] - 1 & \text{for } m \leq M \\ \frac{1}{2} \left\{ \frac{1}{1 - p_c} - \frac{(M+1)p_c^{M+1}}{1 - p_c^{M+1}} + \frac{W_0}{1 - r} \right. \\ \quad \cdot \left. \left(1 - \frac{r(1 - p_c)}{1 - p_c^{M+1}} \frac{1 - (rp_c)^{M+1}}{1 - rp_c} \right) \right\} - 1 & \text{for } m > M \end{cases} \quad (27)$$

We can now calculate the time until a SS successfully sends a RNG-REQ using the just developed formula and Eqn. (5), with \bar{D}_Σ as x . Assuming the same number of contention slots

for each frame $N_{IRi} = N_{IR}$ for all $i \in \mathbb{N}$, we obtain Eqn. (28), where \div indicates a division without remainder.

$$\begin{aligned} T_{S\text{-queue}} &\approx (\bar{D}_{\Sigma} \bmod N_{IR}) \cdot T_{\text{slot}} + \text{---} \\ &= (\bar{D}_{\Sigma} \bmod N_{IR}) \cdot T_{\text{slot}} + \bar{D}_{\Sigma} \div N_{IR} \end{aligned} \quad (28)$$

Finally, we obtain the delay by multiplying Eqn. (21) with the slot duration and frame duration for each of the two cases. For the first case, for operation in contention slots $m < N_R \leq M$ we obtain Eqn. (29). The function derived is for the case where $m \geq M$ as well as for $N_R < m$ is given in Eqn. (30).

$$\begin{aligned} \bar{T}_{\Sigma,1} &= T_{\text{slot}} \left(\left(\frac{W_0(1-r^{m+1})}{2(1-r)} + \frac{m+1}{2} - 1 \right) \bmod N_{IR} \right) \\ &+ T_{\text{frame}} \left(\left(\frac{W_0(1-r^{m+1})}{2(1-r)} + \frac{m+1}{2} - 1 \right) \div N_{IR} \right) \\ &+ T_{\text{slot}} \left(\left(\left(\frac{1}{1-p_c} - \frac{(M+1)p_c^{M+1}}{1-p_c^{M+1}} - (m+1) \right) \right. \right. \\ &\quad \left. \left. \left(\frac{W_0 r^m + 1}{2} - 1 \right) \right) \bmod N_{IR} \right) \\ &+ T_{\text{frame}} \left(\left(\left(\frac{1}{1-p_c} - \frac{(M+1)p_c^{M+1}}{1-p_c^{M+1}} - (m+1) \right) \right. \right. \\ &\quad \left. \left. \cdot \left(\frac{W_0 r^m + 1}{2} - 1 \right) \right) \div N_{IR} \right) \end{aligned} \quad (29)$$

$$\begin{aligned} \bar{T}_{\Sigma,2} &= T_{\text{slot}} \left(\left(\frac{W_0}{2(1-r)} \left(1 - \frac{(1-p_c)r}{1-p_c^{M+1}} \frac{1-(rp_c)^{m+1}}{1-rp_c} \right) \right. \right. \\ &\quad \left. \left. + \frac{1}{2} \left(\frac{1}{1-p_c} - \frac{(M+1)p_c^{M+1}}{1-p_c^{M+1}} \right) - 1 \right) \bmod N_{IR} \right) \\ &+ T_{\text{frame}} \left(\left(\frac{W_0}{2(1-r)} \left(1 - \frac{(1-p_c)r}{1-p_c^{M+1}} \frac{1-(rp_c)^{m+1}}{1-rp_c} \right) \right. \right. \\ &\quad \left. \left. + \frac{1}{2} \left(\frac{1}{1-p_c} - \frac{(M+1)p_c^{M+1}}{1-p_c^{M+1}} \right) - 1 \right) \div N_{IR} \right) \end{aligned} \quad (30)$$

For every collision and subsequent retransmission, we have to add the timeout T_3 of 200 ms (default and maximum value, cf. [2], pp.638) until the SS realises the packet loss. Using the average number of retransmissions required for a packet to be transmitted without collision in Eqn. (19), the summation of the timeouts T_3 , T_R is given by:

$$T_R = \sum_{i=0}^{n_R} T_3 = \left(\frac{p_c}{1-p_c} - \frac{(M+1)p_c^{M+1}}{1-p_c^{M+1}} \right) T_3 \quad (31)$$

The complete formula to calculate the delay added by the IEEE 802.16 exponential backoff mechanism including collision probabilities is hence given as:

$$\begin{aligned} \bar{T}_{\text{BO-delay}}(p_c, W_0, M, m, r = 2, T_3, T_{\text{Frame}}, N_{IRi}) \\ = T_{S\text{-queue}}(D_{\Sigma}(m, M, W_0, r, p_c), N_{IRi}, T_{\text{slot}}, T_{\text{Frame}}) \\ + \bar{D}_R(M, p_c, T_3) \end{aligned} \quad (32)$$

2) *Invited Ranging*: The receipt of a RNG-RSP by the SS specifying its primary CID marks the begin of the invited ranging procedure, which refers to non-contention timeslots (in the further derivation we assume that BS and SS respond to the ranging messages in the next frame after their receipt, thus $T_{\text{RNG-proc-SS}}, T_{\text{RNG-proc-BS}} = 1 \cdot T_{\text{Frame}}$). The SS sends a RNG-REQ in a reserved slot for this primary CID. The BS waits for the SS to adjust to the parameters given in the RNG-RSP, before reserving a ranging slot for the SS. This interval is at least $T_{\text{RNG-proc-SS}} = 10$ ms long (cf. Ref. [2], Table 342, pp.637). The SS tries up to *Invited Ranging Retries* (min: 16) times to get a response for the RNG-REQ it sends during these slots. For the number of RNG-REQ/RNG-RSP pairs exchanged, k_{inv} until the SS and the BS have achieved synchronisation and receive each other correctly, no assumptions can be made, as was already said in the last subsection for the value of k_{cont} .

C. Capability Negotiation

Upon reception of a RNG-RSP with status "success", the SS starts the basic capabilities negotiation. It sends a SBC-REQ to the BS via the just established basic connection, which is answered with a SBC-RSP. The process starts with the next slot reserved by the BS for the SS's basic CID. The time to match the capabilities of BS and SS should be negligible, i.e., the BS sends its response directly during the subsequent DL-subframe. As a result, we can assume the best-case delay of the basic capabilities negotiation to be $T_{\text{SBCnego}} \approx 2T_{\text{Frame}}$. In reality, the negotiation might be deferred until the BS is granting sufficient bandwidth to the basic CID of the SS.

D. Summary

We have derived a model to describe the duration of the network entry process until the privacy negotiation starts. The duration is the sum of the time until the SS gathers all network information (T_{StartIR}), the time spent for initial ranging, and the time for the capability negotiation as shown in Eqn. (33). The time for initial ranging on contention slots as well as invited ranging is given in Eqns. (35), (36), where the exponential backoff mechanism contributes to $T_{\text{reception}}$.

$$T_{\text{StartPrivNeg}} = T_{\text{StartIR}} + T_{\text{IR-cont}} + T_{\text{IR-inv}} + T_{\text{SBCnego}} \quad (33)$$

$$\begin{aligned} T_{\text{IR-cont}} &= k_{\text{cont}} (T_{\text{reception}} + T_{\text{Frame}} \\ &\quad + T_{\text{RNG-proc-SS}} + T_{\text{RNG-proc-BS}}) \end{aligned} \quad (34)$$

$$\begin{aligned} T_{\text{IR-inv}} &= 2k_{\text{inv}} T_{\text{Frame}} \\ &\quad + k_{\text{inv}} (T_{\text{RNG-proc-SS}} + T_{\text{RNG-proc-BS}}) \end{aligned} \quad (35)$$

V. ANALYSIS OF RESULTS

The formulae developed in the last section allow to compute the delay to join the network. Here we provide an upper and lower bound for the network entry process time. Moreover, we analyse the influence of selected variables. An Octave [13] implementation of the model is available from the authors upon request, it allows to use the model as a tool to calculate the expected network entry behaviour for different scenarios.

TABLE III
PARAMETERS AND CONSTANTS RELATED TO NETWORK ENTRY

Variable	Name	Time reference	Max. value
T_{frame}	Frame Duration	Time between two UL-Maps (WirelessMAN-SC & -SCa)	0.5; 1; 2 & 4 ms
$T_{\text{IR-int}}$	Initial Ranging Interval	Time between Initial Ranging Regions assigned by the BS	2 s
$T_{\text{DCD}}, T_{\text{UCD}}$	DCD, UCD Interval	Time between transmission of DCD, UCD Messages	10 s
$T_{\text{DCD,UCD-Transition}}$	DCD, UCD Transition	Time BS shall wait before applying new DCD, UCD parameters	$2 T_{\text{frame}}$
$\text{CRR}=\text{M}$	Contention Ranging Retries	Number of retransmitted ranging requests for ranging in contention slots	16 (minimal)

A. Worst-case Calculation

The worst-case delay depends on the choice of protocol parameters, which vary for the different profiles of the standard (see Table III for an overview of the parameters governing network entry for the profiles defined by [2]).

We first calculate the maximum time until the SS is allowed to send the first initial ranging message to the BS. In the worst case scenario, a SS starts into the process having just missed the reception of the DCD and UCD message blocks. Thus, the SS has to defer queuing its RNG-REQ until it has received the subsequent DCD and UCD and after the transition time of $2 T_{\text{frame}}$ has passed. Also, between two initial ranging intervals, there may be a BS-assigned pause of length of $T_{\text{IR-int}}$. We obtain $T_{\text{StartIR}} = 2T_{\text{frame}} + T_{\text{DCD}} + T_{\text{UCD}} + T_{\text{IR-int}}$.

The time until the first IR-request is sent depends on the choice of BoS and BoE. Let us assume a BoS of 2^{10} , and BoE of 2^{15} . Let, in average, the SS always choose half of its current window size for x_n . We further assume that the retry count of (minimal) 16 unanswered retries is reached. As $T_{\text{Slot}} \ll T_{\text{Frame}}$, T_{Slot} can be neglected in Eqn. (5) for this calculation. Let the BS provide one ranging opportunity per timeframe (4 ms in WirelessMAN-SCa). The thus simplified Eqns. (5) and (6) yield:

$$T_{\text{reception}} = T_{\text{xmit}} + T_{\text{R-queue}} + \sum_{i=0}^{16-1} \left(T_3 + \sum_{n=0}^{\infty} \left(T_{\text{frame}} \sigma \left(x_i - \sum_{j=0}^n N_{\text{IR},j} \right) \right) \right) \quad (36)$$

with $x_i = \{2^9, 2^{10}, 2^{11}, 2^{12}, 2^{13}, 2^{14}, 2^{14}, 2^{14}, \dots\}$. We can neglect T_{xmit} and $T_{\text{R-queue}}$ and obtain:

$$T_{\text{reception}} = 15T_3 + 4 \text{ ms} \cdot (2^9 + 2^{10} + 2^{11} + 2^{12} + 2^{13} + 10 \cdot 2^{14}) \approx 12 \text{ min.}$$

B. Best-case Calculation

The best case for joining the network can be described as follows. The joining SS receives UCD and DCD at once and the following frame bears a sufficient amount of available ranging slots. If the first IR-Request is received correctly by the BS, it invites the SS to range using an invited ranging opportunity. If we further assume that the stations process the ranging messages instantly, they may answer directly in the subsequent time-frames. This sums up to one frame for the receipt of the UCD and DCD, two frames until these may be applied (UCD/DCD-Transition, cf. Table III). We obtain the

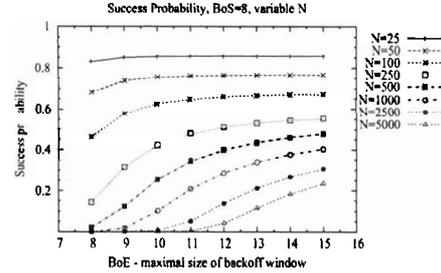


Fig. 4. Success probability on contention slots for BoS = 8 and varying N .

overall time until the security negotiations may start by means of Eqn. (33) to be $T_{\text{StartPrivNeg}} = 4 \cdot 2T_{\text{Frame}} = 8T_{\text{Frame}}$.

For our example (WirelessMAN-SCa with $T_{\text{Frame}} = 4 \text{ ms}$), the procedure thus minimally takes 32 ms.

C. Analysis of Protocol Parameters

Our example has shown that the choice of network parameters is indeed crucial to allow for reasonable QoS guarantees for the network entry procedure. We next investigate the sensible parameter ranges using our model formulae. Since the contention-based ranging is the main factor influencing the delay, we study the influence of the number of nodes trying to join the network simultaneously and the variation of BoS and BoE. As response variables, we investigate the probability of successfully winning the contention as well as the delay.

Fig. 4 shows the average success probability for a BoS = 8. The success probability drops significantly with an increase in nodes. However, choosing large values for BoS obviously leads to an increase in delay. Fig. 5(a) shows the average success probability of individual ranging attempts for $N = 25$ nodes and varying BoS and BoE. Fig. 5(b) shows similar results for $N = 100$. The results indicate that BoS should be set sufficiently high to avoid excessive collisions for the nodes trying to join.

Fig. 6(a) and (b) show the average delay accumulated during contention-based ranging for $N = 25$ and $N = 100$, respectively. A small BoE leads to an increase in delay, because the backoff algorithm produces excessive collisions (until a packet is dropped). A large BoS lead to even higher delays, because of the large initial contention window size. For the analysed network, the optimum is reached if the initial contention window size is able to resolve collisions during the first or second contention attempt.

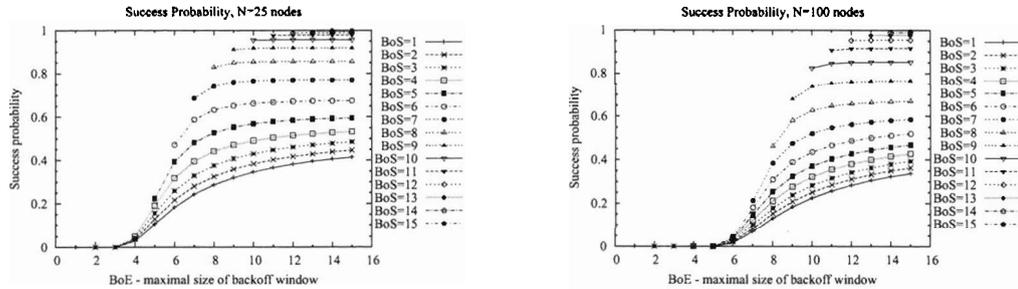


Fig. 5. Success probability on contention slots for varying BoS and BoE; (a) $N = 25$ nodes and (b) $N = 100$ nodes.

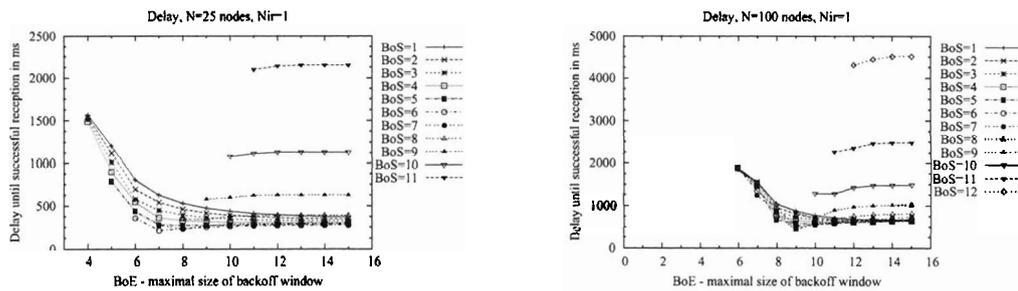


Fig. 6. Expected delay for $N_{IR} = 1$ and varying BoS and BoE; (a) $N = 25$ nodes and (b) $N = 100$ nodes.

VI. CONCLUSION

IEEE 802.16 is a very promising standard for applications such as QoS-aware voice-, video-, and data-communication. In this paper, we have presented a mathematical model to describe the network entry procedure of 802.16. In particular, as a basis of our work, we have presented a thorough dissection of 802.16's network entry process. We then derived our model to describe the delay-characteristics of the individual steps of the network entry procedure. We identified the contention slots for initial ranging to be of high importance for the overall delay observed by SAs joining the network. Using our model, we analysed key parameters influencing the performance of the network entry procedure. As future work, the analysis can be extended to include the delay of the privacy negotiation or to refine the model to include the optional protocol optimisations introduced in the latest version of the standard.

REFERENCES

- [1] P. S. Mogre, M. Hollick, and R. Steinmetz, "QoS in Wireless Mesh Networks: Challenges, Pitfalls, and Roadmap to its Realization," in *Proceedings of Network and Operating System Support for Digital Audio and Video (NOSSDAV 2007)*, 2007.
- [2] "802.16 IEEE Standard for local and metropolitan area networks, Part 16: Air Interface for Fixed Broadband Wireless Access Systems," IEEE Computer Society, IEEE Microwave Theory and Techniques Society, New York, NY, USA, Oct. 2004.
- [3] "802.16 IEEE Standard for local and metropolitan area networks, Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems. Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands and

- Corrigendum 1," IEEE Computer Society, IEEE Microwave Theory and Techniques Society, New York, NY, USA, Feb. 2006.
- [4] C. Eklund, R. Mark, K. Stanwood, and S. Wang, "IEEE standard 802.16: a technical overview of the WirelessMAN air interface for broadband wireless access," *IEEE Communications Magazine*, vol. 40, no. 6, pp. 98–107, June 2002.
- [5] P. S. Mogre, M. Hollick, and R. Steinmetz, "The IEEE 802.16-2004 MeSH Mode Explained," Multimedia Communications Lab, Technische Universität Darmstadt, Germany, Tech. Rep. KOM-TR-2006-08, 2006, <http://ftp.kom.tu-darmstadt.de/pub/TR/KOM-TR-2006-08.pdf>.
- [6] C. Cicconetti, L. Lenzini, E. Mingozzi, and C. Eklund, "Quality of service support in IEEE 802.16 networks," *IEEE Network*, vol. 20, no. 2, pp. 50–55, Mar./April 2006.
- [7] J. Chen, W. Jiao, and Q. Guo, "An integrated QoS control architecture for IEEE 802.16 broadband wireless access systems," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM'05)*, 2005.
- [8] H. Shetiya and V. Sharma, "Algorithms for routing and centralized scheduling to provide QoS in IEEE 802.16 mesh networks," in *Proceedings of the 1st ACM workshop on Wireless multimedia networking and performance modeling (WMuNeP'05)*, 2005.
- [9] K. Wongthavarawat and A. Ganz, "Packet scheduling for QoS support in IEEE 802.16 broadband wireless access systems," *International Journal of Communication Systems*, vol. 16, pp. 81–96, 2003.
- [10] S. Xergias, N. Passas, and L. Merakos, "Flexible resource allocation in IEEE 802.16 wireless metropolitan area networks," in *Proceedings of the 14th IEEE Workshop on Local and Metropolitan Area Networks (LANMAN 2005)*, 2005.
- [11] K. Doo H. L. and Kyamakya and J. Umondi, "Fast handover algorithm for IEEE 802.16e broadband wireless access system," in *Proceedings of the 1st International Symposium on Wireless Pervasive Computing, 2006*, 2006.
- [12] B. Kwak, N. Song, and L. Miller, "Performance analysis of exponential backoff," *IEEE/ACM Transactions on Networking (TON)*, vol. 13, no. 2, pp. 343–355, April 2005.
- [13] J. W. Eaton et al., "GNU Octave," Dec. 2006. [Online]. Available: <http://www.gnu.org/software/octave/octave.html>