Throughput Analysis of Best Effort Traffic in IEEE 802.16/WiMAX

Andrea Bacioccola[1], Claudio Cicconetti[1], Alessandro Erta[2], Enzo Mingozzi[1]

[1] Dipartimento di Ingegneria dell'Informazione
University of Pisa, Italy
{a.bacioccola|c.cicconetti|e.mingozzi}@iet.unipi.it

[2] IMT Lucca Institute for Advanced Studies, Lucca, Italy
alessandro.erta@imtlucca.it

***Abstract*—IEEE 802.16 is a standard for Broadband Wireless Access networks, where a number of Subscriber Stations (SSs) are centrally controlled by a Base Station (BS). The Medium Access Control (MAC) protocol is connection-oriented and time is divided into frames of fixed duration. Moreover, MAC signaling is in-band, i.e. any frame carries both data and control messages. Therefore, the capacity available for data transmission is affected by the overhead due to control messages whose size, in turn, depends on several factors including the number of SSs that are scheduled in a frame. In this paper, we derive a set of formulae to compute the maximum throughput of best effort traffic in an IEEE 802.16 network. To this end, we only assume that the BS fairly shares the available resources among the connections. We validate our formulae through simulation, and, finally, we exploit the proposed formulae to investigate the throughput performance of IEEE 802.16 in a set of relevant scenarios.**

I. INTRODUCTION

The IEEE 802.16 [8] is a mature standard for Broadband Wireless Access (BWA), which enables the delivery of last mile wireless broadband access as an alternative to wireline broadband like cable and DSL. In June 2001 the Worldwide Interoperability for Microwave Access (WiMAX) forum has been established to promote conformance and inter-operability of the IEEE 802.16 devices.

An IEEE 802.16 network consists of a number of Subscriber Stations (SSs) whose transmissions and receptions are centrally controlled by a Base Station (BS). The Medium Access Control (MAC) protocol is connection-oriented and explicitly supports Quality of Service (QoS) by defining five different QoS scheduling services: namely, Unsolicited Grant Service (UGS), real-time Polling Service (rtPS), extended real-time Polling Service (ertPS), non real-time Polling Service (nrtPS), and Best Effort (BE). Each scheduling service is designed to support a specific class of applications. Specifically, the BE scheduling service is tailored to elastic applications, which can adapt their behavior to the resources available in the network. As such, BE applications need not be guaranteed a specific amount of resources, but also benefit from any bandwidth that is not used by QoS applications. However, BE flows expect themselves to be treated fairly, i.e. to receive the same share of available resources.

In this paper, we derive a set of formulae which allows to estimate the maximum throughput available to any BE connection, in saturation conditions, under the assumption that connections are served fairly. The main issue in this task is that, in IEEE 802.16, signaling is in-band, i.e. control and the data messages share the same transmission resources. Therefore the available capacity for data transmissions incurs the overhead of control messages, which depends on several factors including the number of SSs that are scheduled in a frame. The formulae are validated against simulation results obtained by means of a detailed model of the IEEE 802.16 system, and prove themselves to be very accurate. While the formulae may not be able to capture all the details of IEEE 802.16, nonetheless, they allowed us to derive the impact of several system and network configuration parameters on the throughput, without the burden of running complex time-consuming packet-level simulations.

Despite the growing interest of industry and research in the IEEE 802.16 as a viable solution for BWA, the literature still lacks a substantial amount of work on the performance evaluation of IEEE 802.16 systems. An early work [15] proposed a packet scheduling algorithm to provide traffic flows with different classes of service with QoS, whose effectiveness has been assessed via simulation with TDD. More recently Chen et al. [2] devised a scheduling algorithm to jointly serve both uplink and downlink traffic flows, so as to exploit the dynamic variation of the uplink/downlink ratio in a TDD system. Preliminary simulation results showed that their contributed scheduling algorithm achieves better performance than a pure priority-based scheduler among flows that belong to different classes of service. In both cases, the simulation model was based on simplistic assumptions, which do not depend on any specific air interface. Another solution for packet scheduling with rtPS/nrtPS/BE scheduling services has been proposed in [7], while UGS is considered in [16, 17]. Hoymann [9] instead focused on the OFDM air interface, and performed a hybrid analytical-simulative analysis of the effect on the system performance of several MAC mechanisms, such as the packing and fragmentation of MAC Service Data Units (SDUs). The performance of a TDD system operated with the OFDM air interface was also analyzed in [6], where the authors compared the performance of two well-known scheduling algorithms, namely Weighted Round-Robin (WRR) and Weighted Fair Queueing (WFQ). Finally, in [3] we described a system architecture to support QoS and reported preliminary simulation results with FDD, which have been extended in [4] to evaluate

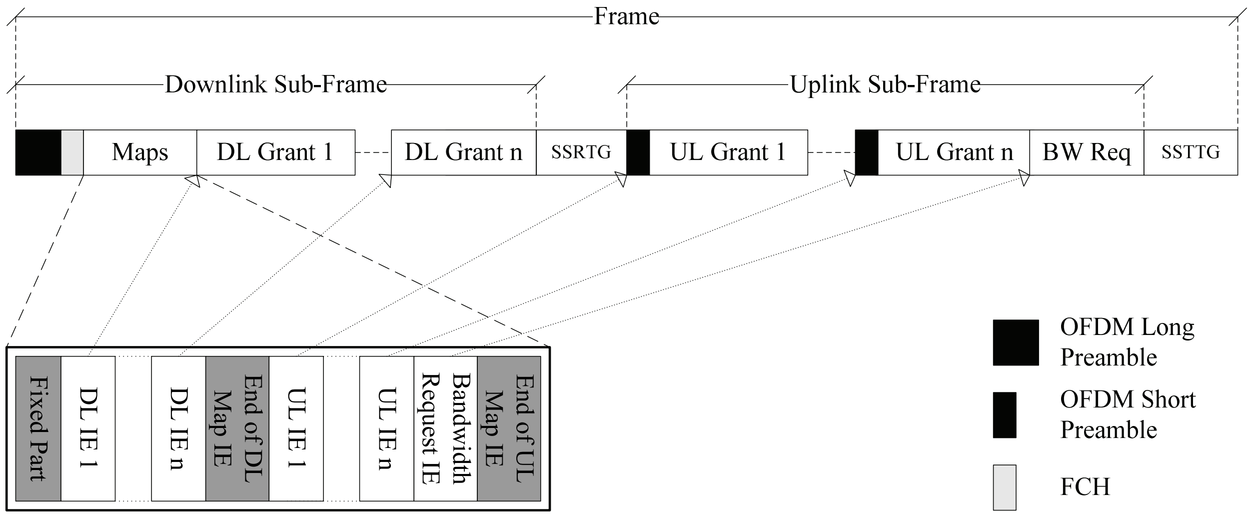

Figure 1. TDD frame structure with MAC and physical overhead.

the impact of several system parameters on the performance of QoS and non-QoS traffic.

The paper is organized as follows. In Section II we briefly introduce the IEEE 802.16 MAC, focusing on those aspects that are specifically relevant to this work. In Section III we devise the analytical formulae, and validate them through simulation. In Section IV we exploit our model to analyze the performance of the IEEE 802.16 MAC, in terms of the overall throughput, as a function of several system and network configuration parameters. Finally, we draw conclusions in Section V.

II. IEEE 802.16

The IEEE 802.16 MAC protocol is connection-oriented: all data communications, for both transport and control, are in the context of a uni-directional connection. Time is partitioned into frames of fixed duration. Each frame contains two sub-frames: the downlink sub-frame, where the BS transmits to SSs, and the uplink sub-frame, where the SSs transmit to the BS in a Time Division Multiple Access (TDMA) manner. In order to support bi-directional transmissions, the IEEE 802.16 specifies two duplexing techniques: Frequency Division Duplex (FDD), where uplink and downlink transmissions occur simultaneously on separate frequencies, and Time Division Duplex (TDD), where uplink and downlink transmissions alternate in the time frame and share the same frequencies. In this paper, we focus on the TDD mode and we consider the OFDM air interface, which is the most promising air interface for supporting Non Line of Sight (NLOS) operation in fixed BWA networks [14].

Fig. 1 reports the structure of a sample frame. Each sub-frame consists of a fixed number of OFDM symbols. Details on the OFDM symbol structure can be found in [5]. The downlink sub-frame begins with a well-known sequence, called long preamble, to synchronize the SS receivers of the SSs. The duration of a long preamble is two OFDM symbols. A synchronization preamble is also used in the uplink direction, where SSs transmit a short preamble, with duration of one OFDM symbol, to synchronize the BS's receiver, before any data transmission. Immediately after the long preamble, the BS transmits the Frame Control Header (FCH), which consists of one OFDM symbol and is used by the SSs to decode the forthcoming MAC control messages transmitted by the BS. Specifically, two MAC control messages are used by the BS to control medium access[1]: DL-MAP and UL-MAP messages (or maps, for short). The DL-MAP (UL-MAP) notifies each SS of the start and end times of the downlink (uplink) grants. Specifically, as illustrated in Fig. 1, a downlink grant is a time interval where the BS transmits to one or more SSs, while an uplink grant is a time interval where one SS can transmit to the BS. To notify a downlink (uplink) grant, a downlink (uplink) information element (IE) is included into the DL-MAP (UL-MAP). The IE specifies the start and the end times of the grant and the SSs that the grant is addressed to. Uplink grants always include the short preamble. It must be pointed out that grants are issued per SS. This means that, even though resources are scheduled per connection, all grants in a sub-frame addressed to different connections belonging to the same SS are summarized by a single IE. Thus, the SS has in turn to implement a scheduling function to select which of its connection will transmit in the assigned grant.

At the end of each downlink (uplink) sub-frame the BS provides the SSs with enough time to switch from reception (transmission) to transmission (reception). The reception to transmission time is referred as SS Receive to Transmit Gap (SSRTG), and the transmission to reception time is referred as SS Transmit To receive Gap (SSTTG).

[1] There are several MAC control messages that the BS and SSs need to exchange in a real environment, e.g. for the purpose of maintenance and network entry. However, we do not consider their contribution to the overhead in the analysis since they happen on a time scale much larger than that of a single frame and thus consume a negligible amount of resources with respect to the overall bandwidth available.

The OFDM air interface supports several burst profiles, each consisting of a combination of symbol modulation and Forward Error Correction (FEC) type. Thus, the amount of data carried in an OFDM symbols is variable and depends on the transmission burst profile. This allows both the BS and the SSs to adapt their transmission rate to the actual quality of the wireless link. In fact, whenever the physical layer detects performance degradation (improvement), a more robust (less robust) burst profile can be used, which entails a lower (higher) efficiency, in terms of the transmission rate.

As stated above, the BS is responsible for defining both uplink and downlink sub-frames. Furthermore, the IEEE 802.16 does not specify any scheduling algorithm that should be used by the BS to define the sub-frames. With regard to the downlink direction, the BS schedules downlink grants based on the exact status of the local connection queues. On the other hand, uplink grants are scheduled by the BS according to the amount of bandwidth requested by the SSs for each of their connections. SSs transmit bandwidth requests for their BE connections in response to broadcast polls. These are issued by the BS by means of an uplink grant, namely BandWidth Request (BW Req) in Fig. 1, which is accessed by the SSs in a random access manner. Therefore, the downlink and uplink available capacities are further reduced because of the bandwidth request IE and the BW Req grant, respectively.

III. Numerical Analysis

In this section we derive the maximum throughput of an IEEE 802.16 wireless network operated with the OFDM air interface in TDD mode. The computation is validated through packet-level simulation at the end of the section.

A. *Throughput computation*

In order to compute the user throughput of an IEEE 802.16 network, the MAC and physical layer overhead need be removed from the raw available bandwidth. The latter only depends on the *physical* bandwidth, in MHz, and the burst profiles employed by the SSs to receive/transmit data. On the other hand, the amount of bandwidth consumed by the MAC and physical layer overhead increases with the number of users served per frame, as detailed below. The notation used in this section is summarized in Tables 1 and 2.

As far as downlink is concerned, the MAC and physical layer overhead (O_{DL}, in OFDM symbols), as a function of the number of users served in the downlink (n_{DL}) and uplink (n_{UL}) sub-frames, can be computed as follows:

$$O_{DL}\left(n_{DL}, n_{UL}\right) = 2 + \left\lceil \frac{MAP_{fixed} + n_{DL} \cdot C_{DL\_IE} + n_{UL} \cdot C_{UL\_IE}}{B_M} \right\rceil . \quad (1)$$

More specifically, each SS served in downlink contributes C_{DL_IE} bytes, which is the size of an IE of the DL-MAP, while each user served in uplink increases the UL-MAP size by C_{UL_IE} bytes. In addition to the IEs of the downlink and uplink grants, the DL- and UL-MAP messages incur a fixed overhead (MAP_{fixed}), which takes into account the transmission of the FCH, the MAC headers and CRC, and the mandatory IEs described in Section II. Finally, the DL-MAP is pre-pended a long preamble, which consumes two OFDM symbols.

On the other hand, the uplink MAC and physical layer overhead (O_{UL}, in OFDM symbols) consists of the capacity reserved as broadcast/multicast polls for the transmission of bandwidth requests (BW) and the short preambles transmitted by the users served in the uplink sub-frame, respectively. Since a short preamble consumes one OFDM symbol, O_{UL} can be expressed as follows:

$$O_{UL}\left(n_{UL}\right) = n_{UL} + BW . \quad (2)$$

Finally, the raw capacity available for both downlink and uplink (F, in OFDM symbols) is equal to the frame duration, expressed in terms of OFDM symbols, minus the gaps that need be employed to allow the SS transceivers to switch from transmission to reception (*SSTTG*) and from reception to transmission (*SSRTG*), as illustrated in Fig. 1.

We now compute the throughput of any SS k, in downlink (TH_D^k) and uplink (TH_D^k). Without loss of generality, we assume that any SS k has the same number of uplink and downlink connections, equal to N_k^{conn}. The system throughput (TH) can then be computed as the sum, for each direction, of the throughput of each SS k in the network:

$$TH = \sum_{k=1}^{N}\left(TH_D^k + TH_U^k\right). \quad (3)$$

Table 1. System Constants

Symbol	Description	Value range
T	Frame duration (s)	{5 ms, 10 ms, 20 ms}
BW	Capacity reserved in the uplink sub-frame as broadcast/multicast polls (OFDM symbols)	2
MAP_{fixed}	Fixed MAC overhead of the DL- and UL-MAP messages (bytes)	{13, 13}
C_{DL_IE}	DL-MAP IE size (bytes)	4
C_{UL_IE}	UL-MAP IE size (bytes)	6
N	Total number of SSs	10 ÷ 100
N_k^{conn}	Number of connections of SS k	1÷10
D_F	Maximum service delay (frames)	4÷40
B_i^{DL}, B_i^{UL}	Bytes per OFDM symbol of SS i in downlink and uplink, respectively	{36, 72, 108}
B_M	Bytes per OFDM symbol to transmit the DL- and UL-MAP messages	12

Table 2. System Variables

Symbol	Description
N_f	Average number of SSs served in a frame
F_U, F_D	Downlink and uplink sub-frame durations, respectively (OFDM symbols)
n_{DL}, n_{UL}	Number of SSs served in the downlink and uplink sub-frames, respectively
TH	Aggregate system throughput (bytes/s)
TH_k^D, TH_k^U	Downlink and uplink throughput of SS k, respectively (bytes/s)
O_{DL}, O_{UL}	MAC and physical layer overhead in the downlink and uplink sub-frames, respectively (OFDM symbols)

First, since we are interested in the computation of the maximum throughput, we assume that all connections of any SS always have a packet to transmit. Second, since the IEEE 802.16 MAC protocol is connection-oriented, even though we do not focus on any specific algorithm, we can assume that resource scheduling on the BS works on a per connection basis, and therefore each connection receives a fair share of service, irrespectively of the particular SS it belongs to. In order to set a specific time scale over which considering fairness, we assume that each connection is provided with a fixed amount of service over a period of DF frames. This assumption allows us to derive the "fair" throughput share of any connection over DF, whose value will depend on the specific scheduling algorithm. As stated above, even though scheduling is performed on a per-connection basis, grants are provided on a per-SS basis, i.e. a single IE is used for all the connections scheduled in an uplink (downlink) sub-frame belonging to the same SS. This means that if more than one connection belonging to an SS is scheduled to be served in a sub-frame, this results in a (equivalent) single grant addressed to the SS, irrespective of the number of scheduled connections. Therefore, the average capacity available per sub-frame depends on the average number of SSs served per sub-frame (denoted as Nf), which in turn depends on how the BS scheduler distributes the service grants over DF. In the following we assume that the service grants to connections are uniformly distributed over the DF frames, i.e. each connection has an independent probability, equal to 1/DF, of being served in any particular frame out of DF. This assumption states that all connections are treated independently of one another, regardless of the SS to which they belong.

To summarize, we assume that the BS scheduler decides which connection to serve so as to provide it with a fair share of the throughput over DF frames, without taking into account the SS the connection belongs to. Our choice is inspired by the BS schedulers that have been proposed so far in the literature for the IEEE 802.16, e.g. [2, 6, 15].

Based on (1) and (2), the downlink and uplink throughput per connection, respectively, can then be computed as

$$TH_D^k = \frac{F_D - O_{DL}\left(N_f, N_f\right)}{T} \cdot \frac{N_k^{conn}}{\sum_{j=1}^{N} N_j^{conn} / B_j^{DL}}, \tag{4}$$

$$TH_U^k = \frac{F_U - O_{UL}\left(N_f\right)}{T} \cdot \frac{N_k^{conn}}{\sum_{j=1}^{N} N_j^{conn} / B_j^{UL}}, \tag{5}$$

where F_D (F_U) is the downlink (uplink) sub-frame size, in OFDM symbols, and B_j^{DL} (B_j^{UL}) is the number of bytes that SS j transmits in an OFDM symbol, depending on its currently employed downlink (uplink) burst profile. The first factor of each equation above is the overall system throughput, in bytes/s, purged from the MAC and physical layer overhead. On the other hand, the second factor specifies that the system throughput is shared by the SSs proportionally to their number of connections. Note that N_j^{conn} is divided by B_j^{DL} (or B_j^{UL}) because the connections whose SSs have a more robust burst profile (i.e. lower transmission rate) than others consume more capacity than the connections whose SSs have a less robust burst profile (i.e. higher transmission rate). The only unknown quantity in (4) and (5) is the average number of SSs served per frame N_f, whose computation can be expressed as a modification of the "generalized birthday problem".

Let us first consider the classical "birthday problem", which consists of computing the probability $p_2(d, n)$ that two or more people out of a group of n will have matching birthdays out of d equally possible birthdays. This probability can be easily found to be equal to:

$$p_2\left(d,n\right) = 1 - \frac{d!}{\left(d-n\right)!d^n}.$$

The more general version of the problem above is to compute the probability $p_j(d, n)$ that a birthday is shared exactly by j (and no more) people. This probability can be expressed as a recurrence relation [1]. In IEEE 802.16, the N_k^{conn} connections of SS k play the role of the group of people, while the D_F frames are the possible birthdays. Therefore $p_j(D_F, N_k^{conn})$ is the probability that exactly j connections of the same SS k are placed into the same frame out of D_F.

Actually, to compute N_f we need the probability $s_k(j)$ that exactly j frames are occupied by the connections of SS k, which is equivalent to the probability that there are exactly j birthdays. This probability is

$$s_k\left(j\right) = \frac{MC_j\left(N_k^{conn}\right) \cdot D_F!}{\left(D_F - j\right)! \cdot \left(D_F\right)^{N_k^{conn}}}, \tag{6}$$

where $MC_j(N_k^{conn})$ is the multinomial coefficient[2] (called M_3 in [1]), which yields:

$$N_f = \frac{1}{D_F} \sum_{k=1}^{N} \left(\sum_{j \in J_k} j \cdot s_k\left(j\right) \right), \tag{7}$$

where J_k is the set of all possible numbers of frames occupied by SS k. J_k can be enumerated with a constant average computation time by means of the algorithm proposed in [18].

B. Validation

We now validate our hypothesis that the service grants of connections are uniformly distributed over D_F through simulation. The simulations were carried out by means of an event-driven prototypical simulator of the IEEE 802.16 protocol, developed using C++. The method of independent replications [11] has been used to analyze the simulation output data. However, confidence intervals[3] are not reported because they are negligible.

The MAC layer of the BS and SSs have been implemented, including all procedures and functions for uplink/downlink data transmission, and uplink bandwidth request/grant. The simulator thus provides an accurate abstraction of the IEEE

[2] $MC_j(N_k^{conn}) = \left(N_k^{conn}!\right) / (\prod_{t=1}^{N_k^{conn}} (t!)^{\alpha} \cdot \alpha!)$ where α denotes the number of frames with exactly t connections scheduled over j frames.

[3] 95% confidence interval have been estimated.

TABLE 3. SYSTEM PARAMETERS.

Parameter	Value(s)
OFDM symbol duration	34 μs
Channel bandwidth	7.0 MHz
Frame duration	5 ms, 10 ms, 20 ms
Data modulation	QPSK, 16-QAM, 64-QAM
Maps modulation	BPSK
FEC code	RS-CC
SSTTG/SSRTG duration	1 OFDM symbol

802.16 MAC layer and has been used for evaluating the performance of the IEEE 802.16 MAC layer in our recent works [3, 4]. Since the IEEE 802.16 standard clearly states that the scheduling algorithm running on the BS is left up to the manufacturer, we selected Deficit Round Robin (DRR) [12] and Weighted Round Robin (WRR) [10] as the BS's downlink and uplink schedulers, respectively. The former is also used by the SSs to share the uplink grants provided by the BS among their connections.

The system parameters are those specified in the standard for use in a 7 MHz licensed band and are reported in Table 3. We assume that the frame duration (T) is 5 ms, and is shared by the downlink and uplink sub-frame evenly. Furthermore, the burst profile employed is QPSK with a code rate 3/4. We also carried out simulations with different combinations of T, R and burst profile. However, we did not report these results because they all lead to the same conclusions as below.

For the purpose of validation, each connection carries an offered load such that its buffer never becomes empty. Finally, throughput is computed as the number of bytes that are transmitted by the BS or SSs in the context of a connection. Therefore, we ignore the impact of the wireless channel, which might introduce errors due to fading/multi-path/shadowing in a real environment.

Table 4 reports the throughput obtained with both the numerical formulae (A) and through simulation (S) in a scenario with 20 SSs. Here the number of connections per SS increases from 1 to 8 and D_F is set to 4 frames (20 ms). On the other hand, in Table 5 we show the results of a scenario with 10 SSs, each provided with a different number of connections. Specifically, SS 1 has 1 connection, SS 2 has 2 connections, and so on. As can be seen, the analytical results are very accurate as they exhibit negligible differences with the simulation ones (smaller than 5%) both in uplink and downlink. This confirms that our hypothesis that the service grants of connections are uniformly distributed over D_F is not too restrictive and our formulae are able to effectively describe the impact on the throughput of a DRR and WRR grant scheduling algorithms in an IEEE 802.16 system which has multiple connections per SS.

As a further confirmation, we also computed the throughput obtained in the assumption that all connections of the same SS are served in the same frame, i.e. by removing our hypothesis. In this case, the throughput obtained through numerical analysis diverges from that obtained via simulation when the number

TABLE 4. ANALYSIS VERSUS SIMULATION THROUGHPUT COMPARISON (20 SSS, D_F = 4 FRAMES)

#conn per SS	UL (kB/s)		DL (kB/s)	
	A	***S***	***A***	***S***
1	23.59	23.63	22.72	21.56
2	22.24	22.82	21.59	21.44
3	21.22	21.96	20.75	21.25
4	20.46	20.61	20.11	20.97
5	19.89	20.15	19.64	20.52
6	19.47	20.10	19.28	20.11
7	19.15	19.72	19.02	19.59
8	18.91	19.08	18.82	19.20

TABLE 5. ANALYSIS VERSUS SIMULATION THROUGHPUT COMPARISON (10 SSS, D_F = 4 FRAMES)

#conn per SS	UL (kB/s)		DL (kB/s)	
	A	***S***	***A***	***S***
1	8.29	8.42	8.02	8.02
2	16.58	16.84	16.05	16.04
3	24.88	25.26	24.07	24.06
4	33.17	33.68	32.10	32.08
5	41.47	42.10	40.12	40.10
6	49.76	50.52	48.15	48.12
7	58.06	58.94	56.17	56.14
8	66.35	67.36	64.20	64.16
9	74.65	75.78	72.22	72.18
10	82.94	84.21	80.25	80.21

of connections per SS increases. For instance, in the scenario of Table 4, the simplified approach always yields 23.59 kB/s and 22.72 kB/s, in uplink and downlink, respectively.

IV. PERFORMANCE ANALYSIS

In this section we analyze the IEEE 802.16 MAC protocol operating in TDD mode by exploiting the formulae derived in Section III. Specifically, we investigate how different combinations of the following relevant system parameters impact on the system performance: frame duration, number of SSs, number of connections per SS, boundary offset. In the following, the overall throughput per direction is considered. We show the results for a single burst profile consisting of QPSK with 3/4 code rate only. Results obtained with different burst profiles scale according to the robustness of the burst profile, i.e. the more robust the burst profile, the lower the throughput.

In Fig. 2 we show the throughput with an increasing number of SSs ranging from 10 to 100, where each SS has one connection in both the uplink and downlink direction, and the connection service period, D hereafter, is set to 40 ms, i.e. D_F is equal to 8, 4 and 2 in the case of frame duration equal to 5 ms, 10 ms and 20 ms, respectively. As can be expected, the throughput decreases as the number of connections increases. In fact, the increasing number of SSs that have to be served over D leads to increasing downlink and uplink overhead, in terms of the size of maps and the number of preambles, respectively. Moreover, the throughput depends on the frame duration: the higher the frame duration, the higher the throughput.

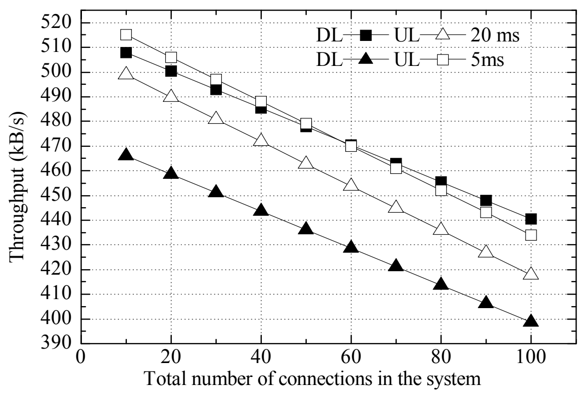

Figure 2. Throughput of BE traffic, with one connection per SS.

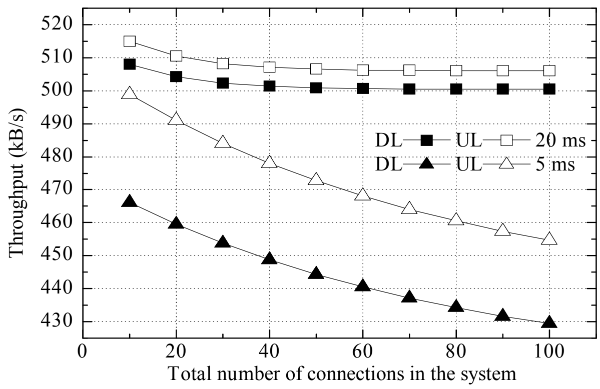

Figure 3. Throughput of BE traffic, with ten SSs.

This is because of the amount of MAC and physical layer overhead per frame which does not depend on the number of connections served, which includes basic maps control information and switching gaps. Note that the gap between uplink and downlink curves is larger in the case of 5 ms, where the number of SSs served per frame is smaller. In fact, when the number of SSs is below a given threshold, the capacity available in the downlink sub-frame is less than the capacity available in the uplink sub-frame. As the number of SSs increases the uplink and downlink curves tend to get closer, because of the increasing uplink preambles overhead. With a frame duration of 20 ms the downlink curve reaches the uplink one when the number of connections is equal to 60. From that point on, the downlink curve lies above the uplink one.

In order to evaluate the effect of the bandwidth partitioning, we now analyze the throughput with an increasing number of connections from 1 to 10, when the number of SSs is 10 and D is set to 40 ms. Results are reported in Fig. 3. As in the previous scenario, the throughput increases when the frame duration increases. However, in this case, the throughput does *not* decrease linearly when the number of connections increases. On the contrary, the throughput tends to a constant value, which depends on the number of SSs scheduled per frame on average. The latter, in turn, is bounded by the total number of SSs. Specifically, as the number of connections per SS increases, the probability that an SS is scheduled in a particular frame (i.e. the probability that at least one connection belonging to the SS is scheduled in a frame within the period D_F) tends to 1. In the case of 20 ms, the throughput reaches an almost constant value sooner than with 5 ms frames. In fact, with 20 ms frames, there are only two frames in which the connections may be partitioned, thus the probability that each SS is served in each frame soon becomes close to 1. As for the previous scenario, the gap between the uplink and the downlink curves depends on the amount of MAC and physical overhead per frame irrespective of the number of connections served. Finally, the gap between the uplink and downlink curves is constant because the number of SSs in the system is constant as well.

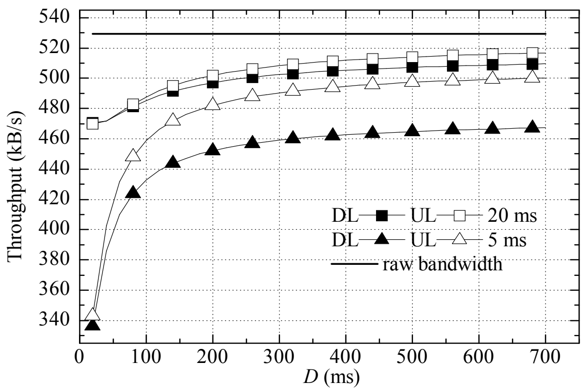

Figure 4. Throughput of BE traffic with 30 SSs, 5 connections per SS.

To conclude our analysis, we report in Fig. 4 the throughput as D increases, with 30 SSs, where each SS has five connections per direction. As a reference, we also report the system *raw bandwidth*, i.e., the total available bandwidth with the current system configuration without considering the MAC and physical layer overhead. As can be seen, the throughput increases with D. In fact, as the service period increases, the probability that a connection is served in a particular frame tends to 0. Therefore, the number of SSs served per frame decreases and so does the overhead due to maps and preambles. Even though the throughput gain is significant for high values of D, the throughput increase rate is higher with small values of D, i.e. when D is smaller than 200 ms. From 200 ms up to 700 ms, the throughput increases very slowly, since the probability that an SS is scheduled in a particular frame decreases very slowly.

It is worth noting that, unlike the other parameters, D strongly depends on the specific scheduling algorithm employed by the BS to serve BE connections. In other words, the period over which each connection receives an equal amount of bandwidth varies depending on the scheduling policies. In general, the longer this period, the smaller is the number of SSs served over a short time scale. This leads to a trade-off between the throughput achieved by each SS and the scheduling timeliness of the SSs' service grants. This relation between the throughput and the service period might be exploited from a scheduling algorithm design standpoint (e.g. to find the best trade-off between throughput and short-term fairness).

V. Conclusions

In this paper we have proposed numerical formulae to derive the throughput obtained by any best effort connection in an IEEE 802.16 network under heavy loaded traffic conditions. We have shown that the throughput depends on several factors, including the average number of SSs served in the frame. To compute the latter we have assumed that the service grants to connections are uniformly distributed over a fixed number of frames. We have then derived the connection throughput exploiting existing results in the probability theory. We have validated our hypothesis through simulation, and we have found that our numerical results are accurate with respect to those obtained through simulation.

Then, we have analyzed the system performance in terms of throughput in a set of relevant scenarios by means of our formulae. Specifically, we have shown that the throughput depends on several factors, including both system and network configuration parameters. For example, the longer the frame duration, the higher the system throughput and the lower the number of connections per SS, the higher the throughput. Finally, we have shown that the service period of connections significantly affects the system throughput.

References

[1] M. Abramovitz, I. Stegun. "Handbook of mathematical functions," *Dover Publications*, June 1964.

[2] J. Chen, W. Jiao, H. Wang. "A service flow management strategy for IEEE 802.16 broadband wireless access systems in TDD mode," *Proc. of ICC 2005*, Seoul, Korea, May 16-20, 2005, pp. 3422-3426.

[3] C. Cicconetti, C. Eklund, L. Lenzini, E. Mingozzi. "Quality of service support in IEEE 802.16 networks," *IEEE Network*, Vol. 20, No. 2, March 2006.

[4] C. Cicconetti, A. Erta, L. Lenzini, E. Mingozzi. "Performance evaluation of the IEEE 802.16 MAC for QoS support," *IEEE Transactions on Mobile Computing*, Vol. 6, No. 1, pp. 26-38, 2007.

[5] A. Ghosh, D. R. Wolter, J. G. Andrews, R. Chen. "Broadband wireless access with WiMax/802.16: current performance benchmarks and future potential," *IEEE Communications*, Vol. 43, No. 2, February 2005, pp. 129-136.

[6] O. Gusak, N. Oliver, K. Sohraby. "Performance evaluation of the 802.16 Medium Access Control layer," in Lecture Notes on Computer Science, Vol. 3280, pp. 228–237, *Springer-Verlag*, Berlin (Germany), 2004.

[7] M. Settembre, M. Puleri, S. Garritano, P. Testa, R. Albanese, M. Mancini, V. Lo Curto. "Performance analysis of an efficient packet-based IEEE 802.16 MAC supporting adaptive modulation and coding," *Proc. of IEEE ISCN 2006*, Istanbul, Turkey, June 16–18, 2006, pp. 11–16.

[8] IEEE 802.16-2004. "IEEE standard for Local and Metropolitan Area Networks – Part 16: air interface for fixed Broadband Wireless Access systems," October 1, 2004.

[9] C. Hoymann. "Analysis and performance evaluation of the OFDM-based Metropolitan Area Network IEEE 802.16," *Computer Networks*, Vol. 49, No. 3, October 2005, pp. 341-363.

[10] M. Katevenis, S. Sidiropoulos, C. Courcoubetis. "Weighted round-robin cell multiplexing in a general-purpose ATM switch chip," *IEEE Journal on Selected Areas in Communications*, Vol. 9, No. 8, 1991, pp. 1265-1279.

[11] A. M. Law, W. D. Kelton. "Simulation modeling and analysis," Third edition, *McGraw-Hill*, 2000.

[12] D. Stiliadis and A. Varma, "Latency-rate servers: A general model for analysis of traffic scheduling algorithms," *IEEE/ACM Transactions on Networking*, Vol. 6, October 1998, pp. 675 689.

[13] WiMAX forum, http://www.wimaxforum.org/.

[14] WiMAX forum. "Initial certification profiles and the European regulatory framework," WiMAX forum regulatory Working Group, September 2004.

[15] K. Wongthavarawat, A. Ganz. "Packet scheduling for QoS support in IEEE 802.16 broadband wireless access systems," *International Journal of Communication Systems*, Vol. 16, No. 1, February 2003, pp. 81-96.

[16] D.-H. Cho, J.-H. Song, M.-S. Kim, K.-J. Han. "Performance analysis of the IEEE 802.16 wireless Metropolitan Area Network," *Proc. of International Conference on Distributed Frameworks for Multimedia Applications*, Besançon, France, February 2005, pp. 130–137.

[17] Y. Yao, J. Sun. "Study of UGS grant synchronization for 802.16," *Proc. of ISCE 2005*, pp. 105–110, Macau, Hong Kong, June 14–16, 2005.

[18] A. Zoghbi, I. Stojmenovic. "Fast algorithms for generating integer partitions", *International Journal of Computer Mathematics*, 70, 1998, 319-332.