19

Mobile WiMAX Performance Investigation*

Alessandro Bazzi, Giacomo Leonardi, Gianni Pasolini and Oreste Andrisano
WiLab, IEIIT-BO/CNR, DEIS-University of Bologna
Italy

1. Introduction

The IEEE802.16-2004 Air Interface standard (IEEE Std 802.16-2004, 2004), which is the basis of the WiMAX technology, is the most recent solution for the provision of fixed broadband wireless services in a wide geographical scale and proved to be a real effective solution for the establishment of wireless metropolitan area networks (WirelessMAN).
On February 2006, the IEEE802.16e-2005 amendment (IEEE Std 802.16e-2005, 2006) to the IEEE802.16-2004 standard has been released, which introduced a number of features aimed at supporting also users mobility, thus originating the so-called Mobile-WiMAX profile. Currently IEEE802.16 Task Group (TG) and WiMAX Forum are developing the next generation Mobile-WiMAX that will be defined in the future IEEE802.16m standard (Ahmadi, 2009; Li et al., 2009).
Although the Mobile-WiMAX technology is being deployed in the United States, Europe, Japan, Korea, Taiwan and in the Mideast, there are still ongoing discussions about the potential of this technology. What is really remarkable, in fact, with regard to the Mobile-WiMAX profile, is the high number of degrees of freedom that are left to manufacturers. The final decision on a lot of very basic and crucial aspects, such as, just to cite few of them, the bandwidth, the frame duration, the duplexing scheme and the up/downlink traffic asymmetry, are left to implementers. If follows that the performance of this technology is not clear yet, even to network operators.
This consideration motivated our work, which is focused on the derivation of an analytical framework that, starting from system parameters and implementation choices, allows to evaluate the performance level provided by this technology, carefully taking all aspects of IEEE802.16e into account. In particular, the analysis starts from the choices to be made at the physical layer, among those admitted by the specification, and "goes up" through the protocol pillar to finally express the application layer throughput and the number of supported voice over IP (VoIP) users, carefully considering "along the way" all characteristics of the the medium access control (MAC) layer, the resource allocation strategies, the overhead introduced, the inherent inefficiencies, etc.
Let us remark that the analytical framework described in the following can be used not only as a mean to gain an insight into the IEEE802.16e performance, but, above all, to drive the choices of network operators in terms of system configuration. This is particularly true considering that beside the model derivation, here we provide criteria, equations and algorithms to make the best choices from the viewpoint of the system efficiency.

2. IEEE802.16 overview

Before starting our analysis let us introduce the most relevant characteristics of the IEEE802.16 technology, that are recalled hereafter.
The result of the IEEE802.16 TG/WiMAX Forum activity is a complete standard family (IEEE Std 802.16-2004, 2004; IEEE Std 802.16e-2005, 2006) that specifies the air interface for both fixed and mobile broadband wireless access systems, thus enabling the convergence of mobile and fixed broadband networks through a common wide area broadband radio access technology and a flexible network architecture.
The IEEE802.16 standard family supports four transmission schemes:

- WirelessMAN-SC, which has been mainly developed for back-hauling in line-of-sight (LOS) conditions and operates in the 10 GHz - 66 GHz frequency range adopting a single carrier modulation scheme;
- WirelessMAN-SCa, which has the same characteristics of WirelessMAN-SC but operates even in non-LOS conditions in frequency bands below 11 GHz;
- WirelessMAN-OFDM, which has been developed for fixed wireless access in non-LOS conditions and adopts the orthogonal frequency division multiplexing (OFDM) modulation scheme in frequency bands below 11 GHz;
- WirelessMAN-OFDMA, which has been conceived for mobile access and adopts the orthogonal frequency division multiple access (OFDMA) scheme in the 2 GHz - 6 GHz frequency range.

Since we are interested in the mobility enhancement provided by the IEEE802.16e amendment, here we focus our attention on the WirelessMAN-OFDMA transmission scheme.
WirelessMAN-OFDMA is based on the OFDMA multiple-access/multiplexing technique which is, on its turn, based on an N_{FFT} subcarriers OFDM modulation scheme (Cimini, 1985; Van Nee & Prasad, 2000) with N_{FFT} equal to 128, 512, 1024 or 2048.
The N_{FFT} subcarriers form an OFDM symbol and can be further divided into three main groups:

- data subcarriers, used for data transmission;
- pilot subcarriers, used for estimation and synchronization purposes;
- null subcarriers, not used for transmission: guard subcarriers and DC subcarrier.

Considering sequences of OFDM symbols, it is easy to understand that transmission resources are available both in the time domain, by means of groups of consecutive OFDM symbols, and in the frequency domain, by means of groups of subcarriers (subchannels); it follows that a given mobile station can be allocated one or more subchannels for a specified number of symbols.
Several different schemes (in the following, permutation schemes) for subcarriers grouping are provided by the specification, with different possibilities for the downlink and uplink phases: among them we can cite DL-FUSC (downlink full usage of subchannels), DL-PUSC (downlink partial usage of subchannels), DL-TUSC (downlink tile usage of subchannels), or UL-PUSC (uplink partial usage of subchannels) (see the first column of table 1 for a complete list).
The minimum OFDMA time-frequency resource that can be allocated is one OFDMA-slot, which corresponds to 48 data subcarriers that can be accommodated in one, two or three OFDMA symbols, depending on which kind of permutation scheme (DL-FUSC, DL-PUSC, UL-PUSC, ...) is adopted; in particular:

Permutation scheme	Available subchannels N_{Ch}				N_{GS}
	N_{FFT} 128	N_{FFT} 512	N_{FFT} 1024	N_{FFT} 2048	
DOWNLINK					
DL-FUSC	2	8	16	32	1
DL-PUSC	3	15	30	60	2
DL-OptFUSC	2	8	16	32	1
DL-TUSC1	4	17	35	70	3
DL-TUSC2	4	17	35	70	3
UPLINK					
UL-PUSC	4	17	35	70	3
UL-OptPUSC	4	17	35	70	3

Table 1. Permutation schemes' parameters.

- with DL-FUSC a subchannel is constituted by 48 subcarriers in each OFDM symbol (hence, one OFDMA-slot covers one symbol);
- with DL-PUSC a subchannel is constituted by 24 subcarriers in each OFDM symbol (hence, one OFDMA-slot covers two symbols);
- with UL-PUSC a subchannel is constituted by 12 subcarriers in the first OFDM symbol, 24 subcarriers in the second OFDM symbol, 12 subcarriers in the third OFDM symbol and so on, according to the sequence 12-24-12-12-24-12....In this case one OFDMA-slot covers three symbols.

Since seven fixed combinations of modulation scheme and coding rate R_c, hereafter denoted as transmission modes, are provided by the IEEE802.16e physical layer, it follows that a single OFDMA-slot allows to transmit differently sized payloads (see table B in figure 1).

Both time division duplex (TDD) and frequency division duplex (FDD) are supported. However, the initial release of Mobile-WiMAX certification profiles only includes TDD, since it makes resource allocation more flexible (the downlink/uplink ratio can be easily adjusted to support asymmetric DL/UL traffic); for this reason, here we only consider the TDD duplexing scheme.

The TDD frame structure is depicted in the bottommost part of figure 1. Each TDD frame is divided into downlink and uplink subframes, separated by transmit/receive and receive/transmit transition gaps (TTG and RTG).

Each subframe may include "multiple zones", which means that the permutation method can be changed, thus moving, for instance, from DL-PUSC to DL-FUSC.

Focusing on the TDD frame, the first OFDM symbol of the DL subframe always carries a preamble, while a number of subsequent OFDM symbols are necessarily allocated to accomodate MAC layer control messages (FCH, DL-MAP and UL-MAP) adopting the DL-PUSC permutation scheme; similarly, a number of OFDM symbols are necessarily allocated in the UL subframe to accomodate several common signalling channels (e.g. UL Ranging, UL CQICH, UL ACK CH) adopting the UL-PUSC permutation scheme (see figure 1).

Finally, in order to correctly manage each data flow giving an acceptable quality of service to the end user, IEEE802.16e-2005 provides five different scheduling services for traffic delivery:

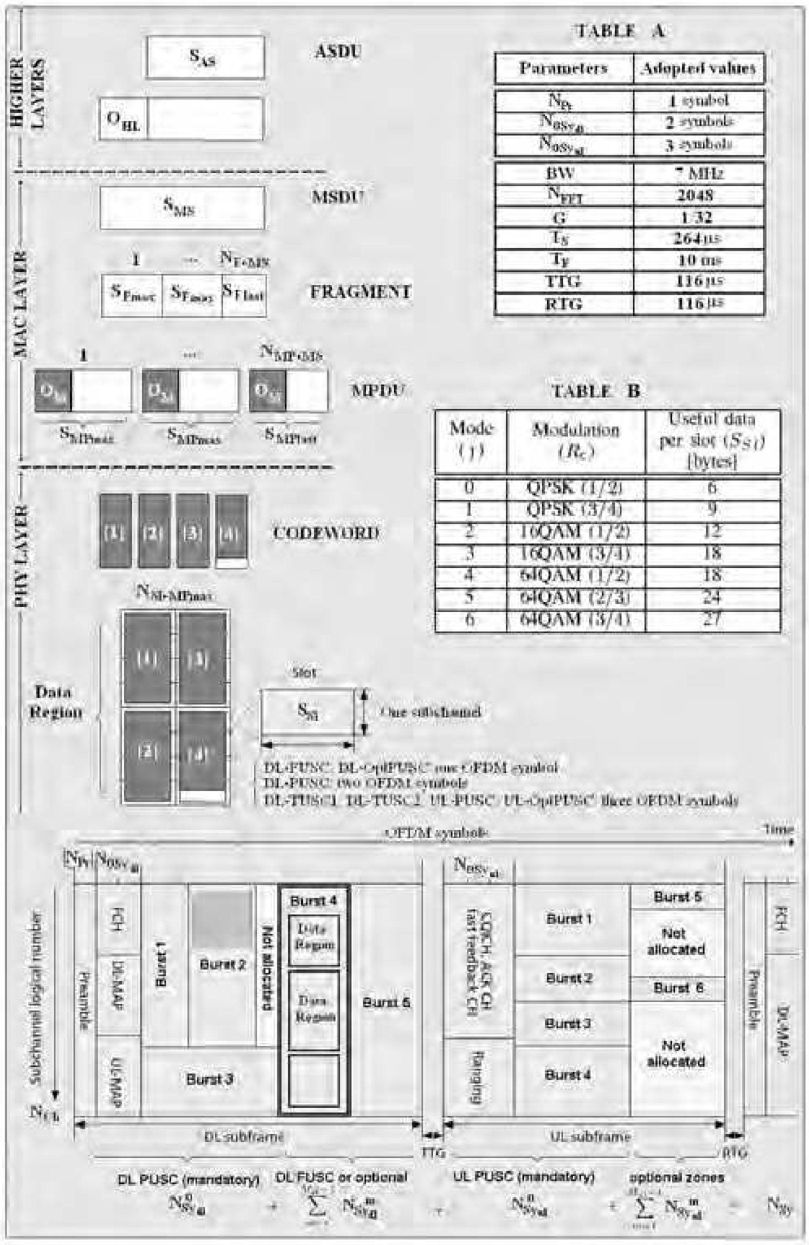


Fig. 1. IEEE802.16e WirelessMAN-OFDMA data processing and parameters setting.

- Unsolicited Grant Service (UGS),
- Real-Time Polling Service (rtPS),
- Extended Real-Time Polling Service (ertPS),
- Non-Real-Time Polling Service (nrtPS),
- Best Effort (BE).

Each scheduling service is associated with a set of quality of services (QoS) parameters: (a) maximum sustained rate, (b) minimum reserved rate, (c) maximum latency tolerance, (d) jitter tolerance and (e) traffic priority. These are the basic inputs for the service scheduler placed in the base station, which is aimed at fulfilling service specific QoS requirements. The main differences among these services are on the uplink resource allocation; resource allocation is, in fact, defined by the base station, which cannot have a perfect knowledge of all uplink buffers in any instant. The interested reader can find a detailed description of the scheduling services in (IEEE Std 802.16-2004, 2004) and (IEEE Std 802.16e-2005, 2006).

3. Transmission resources: OFDMA-slots

In this section the amount of resources that are available for data transmission is evaluated as a function of all parameters that can be chosen by system implementers. In particular, the OFDMA-slot, which is the minimum resource available at the physical layer for data allocation, is focused and the amount of available OFDMA-slots is derived as a function of the physical layer configuration.

In order to ease the reader's task, the scheme reported in figure 2 summarizes the analytical framework outlined in this section, which lead to the assessment of the amount of available OFDMA-slots.

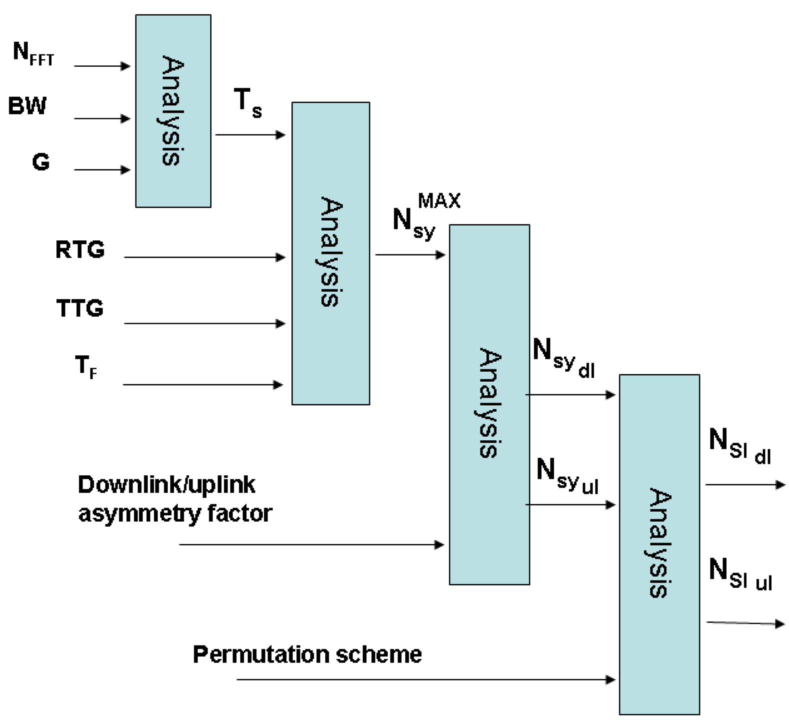


Fig. 2. Analytical framework for the derivation of the amount of OFDMA-slots per downlink/uplink subframe.

3.1 OFDM symbol duration

In order to assess the amount of resources available for data transmission at the physical layer, the OFDM symbol duration T_s must be obtained at first. T_s depends on the transmission bandwidth BW (it is typically a multiple of 1.75 MHz or 1.25 MHz), the number N_{FFT} of OFDM subcarriers (equal to 128, 512, 1024 or 2048) and the normalized (to the useful symbol duration) guard interval G (equal to 1/4, 1/8, 1/16 or 1/32).

Given BW, the value of an auxiliary parameter n (called sampling factor), introduced by the specification, can be immediately derived: in particular, $n = 28/25$ if BW is a multiple of 1.25 MHz, 1.5 MHz, 2 MHz or 2.75 MHz, otherwise $n = 8/7$.
Once BW and n are known, we can derive the sampling frequency F_s, which is defined by the specification as follows:

$$F_s = \left\lfloor n \cdot \frac{BW}{8000} \right\rfloor \cdot 8000, \tag{1}$$

having denoted with $\lfloor x \rfloor$ the highest integer not greater than x.
Given the number N_{FFT} of OFDM subcarriers, the subcarriers spacing Δf and the useful OFDM symbol duration T_u can be immediately derived from the knowledge of F_s:

$$\Delta f = \frac{F_s}{N_{FFT}}, \qquad T_u = \frac{1}{\Delta f}. \tag{2}$$

The guard time interval T_g and, finally, the OFDM symbol duration T_s follow:

$$T_g = G \cdot T_u, \qquad T_s = T_u + T_g. \tag{3}$$

3.2 Number of OFDM symbols per frame

Once T_S has been obtained, the second step to derive the amount of OFDMA-slots available at the physical layer is to assess the number of useful OFDM symbols in a frame.
Since we are interested in the TDD version of IEEE802.16e WirelessMAN-OFDMA, we have to consider a frame structure consisting of two parts, that represent the downlink and uplink subframes, separated by the TTG and RTG time intervals (see figure 1).
In order to derive the number of useful OFDM symbols in a frame, let us recall that the first symbol of the frame is used to transmit the preamble (thus the number of preamble symbols is $N_{Pr} = 1$) and that both TTG and RTG cannot be smaller than 5 μs ($RTG_{min} = TTG_{min} = 5\ \mu s$).
Thus, once the frame duration T_F has been chosen (possible values admitted by the specification are $2, 2.5, 4, 5, 8, 10, 12.5$ and 20 ms), and given the previously derived value of T_s, the maximum number N_{Sy}^{MAX} of OFDM symbols per frame (excluding the preamble) can be derived:

$$N_{Sy}^{MAX} = \left\lfloor \frac{T_F - TTG_{min} - RTG_{min}}{T_s} \right\rfloor - N_{Pr}\ , \tag{4}$$

as depicted in figure 2.

3.3 Number of OFDMA-slots per frame

Given the value of N_{Sy}^{MAX}, it is now possible to assess the number of OFDMA-slots that can be allocated in the downlink and uplink subframes. Since OFDMA-slots extend both in the time and in the frequency domains, the derivation of their amount requires considerations on both domains.
As for the time domain, let us recall that, depending on the adopted permutation scheme (DL-FUSC, DL-PUSC, UL-PUSC, ...), an OFDMA-slot is spread over $N_{GS} = 1, 2$ or 3 consecutive OFDM symbols, as reported in the last column of table 1, whereas, as far as the frequency domain is concerned, the amount of available subchannels N_{Ch} depends on the permutation scheme and the number N_{FFT} of OFDM subcarriers (as reported in the second column of table 1).
Please note that different permutation schemes are provided for the downlink and uplink subframes and that more than one scheme can be used in a single subframe. Each permutation

scheme (denoted in the following with the superscript m) requires the allocation of a multiple of an integer (1, 2 or 3, depending on the permutation scheme) number of OFDM symbols N_{Sy}^{m} ($N_{Sy_{dl}}^{m}$ in the downlink and $N_{Sy_{ul}}^{m}$ in the uplink, respectively) and the sum N_{Sy} of uplink and downlink symbols allocated for each permutation scheme is bounded by the above assessed N_{Sy}^{MAX}:

$$\sum_{m=0}^{M_{dl}-1} N_{Sy_{dl}}^{m} + \sum_{m=0}^{M_{ul}-1} N_{Sy_{ul}}^{m} = N_{Sy} \leq N_{Sy}^{MAX}, \tag{5}$$

where M_{dl} (M_{ul}) is the amount of permutation schemes adopted in the downlink (uplink) subframe.

As recalled in section 2, at least two OFDM symbols are allocated with DL-PUSC in the downlink subframe, in order to carry frame management messages, while three OFDM symbols are allocated with UL-PUSC in the uplink subframe, in order to carry signalling common channels; here we assume that the entire first two OFDM symbols in downlink (with DL-PUSC) and the entire first three OFDM symbols in uplink (with UL-PUSC) are used for this scope, denoting the related overhead with $N_{OSy_{dl}} = 2$ and $N_{OSy_{ul}} = 3$.

Moreover, the rest of the sub-frame is supposed to be transmitted adopting only one permutation scheme. Thus, the superscript correspondent to the adopted permutation scheme will be omitted in the following and (5) is rearranged as follows:

$$N_{OSy_{dl}} + N_{Sy_{dl}} + N_{OSy_{ul}} + N_{Sy_{ul}} = N_{Sy} \leq N_{Sy}^{MAX}, \tag{6}$$

where $N_{Sy_{dl}}$ and $N_{Sy_{ul}}$ now represent the amount of downlink/uplink symbols available for user data.

Let us observe that the choice of $N_{Sy_{dl}}$ and $N_{Sy_{ul}}$ is not only constrained to fulfill (6), but is also a consequence of the desired asymmetry between the downlink and uplink phases of the TDD frame, hereafter referred to as "desired asymmetry factor" and denoted as AF_{in}.

Let's keep in mind, in this regard, that the minimum resource that can be allocated is the OFDMA-slot and that the number of slots in a subframe is related not only to the number of OFDM symbols within the frame, but also to the adopted permutation scheme; as an example, with $N_{FFT} = 2048$ subcarriers two OFDM symbols adopting the DL-PUSC permutation scheme carry 60 slots while three OFDM symbols adopting UL-PUSC carry 70 slots (refer to (8) and table 1). Thus, defining AF_{Sl} as the asymmetry factor in terms of ratio between the amounts of downlink and uplink slots:

$$AF_{Sl} = \frac{N_{Sl_{dl}}}{N_{Sl_{ul}}}, \tag{7}$$

and deriving the amount of OFDMA-slots available for data transmission in the downlink/uplink subframe through the equation (where dl/ul denotes downlink or uplink as alternatives):

$$N_{Sl_{dl/ul}} = N_{Ch_{dl/ul}} \cdot \left\lfloor \frac{N_{Sy_{dl/ul}}}{N_{GS_{dl/ul}}} \right\rfloor, \tag{8}$$

the desired asymmetry AF_{in} can be approached finding the values $N_{Sy_{dl}}$ and $N_{Sy_{ul}}$ that make AF_{Sl} as near as possible to AF_{in}. In general, a perfect matching between AF_{in} and AF_{Sl} will be not possible, due to the system constrains.

The detection of $N_{Sy_{dl}}$ and $N_{Sy_{ul}}$ in such a way to minimize the resource wasting (that is, OFDM symbols within the frame that are unused because unable to accomodate entire

OFDMA-slots) for a given AF_{in} can be carried out by means of the algorithm provided in appendix I.
Equation (8) represent the final outcome of this section since, jointly with the constraints given by the desired asymmetry factor and system choices (bandwidth, guard interval, number of subcarriers, frame duration, ...) accounted for by the previous equations, allows to derive the amount of OFDMA-slots available in each subframe for data transmissions.
Please refer to figure 2 for a pictorial representation of the whole methodology described in this section.

4. From application layer packets to subcarriers allocation

In the previous section the amount $N_{SI_{dl}}$ and $N_{SI_{ul}}$ of OFDMA-slots available for data allocation have been derived, taking into account all the physical and MAC layers parameters. The next step is to understand how packets to be transmitted, coming from the higher protocol layers, are mapped onto these resources. A brief overview on the packet processing is given hereafter, followed by an analytical evaluation of the number of OFDMA-slots that are finally needed to allocate each packet.

4.1 Packet processing overview

The data mapping process, starting from the application layer data unit down to the physical layer, is illustrated step by step hereafter. Please refer to figure 1, where each step is depicted, to better understand the whole process.
Let us denote as $ASDU$ the application level data fragment of S_{AS} bytes that is allocated into the payload of a TCP/IP packet. Each $ASDU$ is firstly added with O_{HL} bytes, where O_{HL} represents the overhead added from the application to the network layer, and then mapped, at the MAC layer, onto a MAC service data unit ($MSDU$) of S_{MS} bytes. Each $MSDU$ is then partitioned into fragments of S_{Fmax} bytes, whose value is negotiated during the connection setup phase; obviously the last fragment of each $MSDU$ may be smaller (S_{Flast} bytes). If the ARQ mechanism is active fragments are also called ARQ blocks.
One or more fragments are then allocated into a MAC protocol data unit (MPDU), with some overhead: in particular, a MAC header will be added plus either (a) one fragmentation subheader if all fragments are contiguous and related to the same MSDU or (b) a packetization subheader per each group of contiguous fragments belonging to the same MSDU; a CRC (cyclic redundancy check) tail of 32 bits will be added at the end of the MPDU, in order to check its integrity at the receiver side.
At the physical layer, MPDUs are partitioned into groups of bytes that are subject to the forward error correction coding process, giving birth to a certain number of codewords. One or more OFDMA-slots can be combined in order to convey each codeword. Adjacent slots, both in the time and subchannels domain, are grouped into OFDMA data regions, which are two-dimensional (squared o rectangular) allocations of a group of contiguous subchannels in a group of contiguous OFDM symbols (see figure 1).

4.2 From application layer data to MPDUs

After the data processing overview provided above, in this subsection we analytically derive the amount of OFDMA-slots needed to deliver an $ASDU$. Having derived (section 3) the amount of OFDMA-slots available in the uplink/downlink subframes of a TDD Mobile-WiMAX system, this is the second step along the path that leads to the Mobile-WiMAX performance assessment.

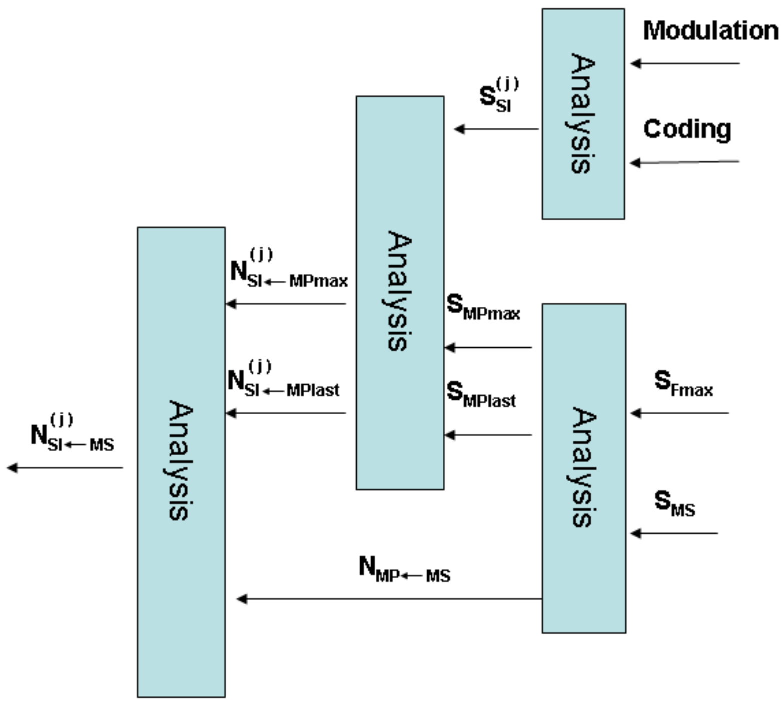


Fig. 3. Diagram of the calculation from the packet size to the number of OFDMA-slots that are needed to accomodate it.

Also in this case, in order to help the reader, the analytical framework outlined in the following has been summarized in a pictorial fashion, reported in figure 3.
Let us consider the aforementioned *ASDUs* of S_{AS} bytes; as represented in figure 1 they eventually arrive at the MAC layer with the addition of the higher layers overheads of O_{HL} bytes, thus originating MAC layer service data units (*MSDU*s) of S_{MS} bytes:

$$S_{MS} = S_{AS} + O_{HL}. \tag{9}$$

The *MSDUs* are then fragmented into a fixed number $N_{F\leftarrow MS}$ of fragments, each of them of size S_{Fmax} except, in case, the last one. This one has a size of $S_{Flast} \leq S_{Fmax}$ bytes; thus:

$$N_{F\leftarrow MS} = \left\lceil \frac{S_{MS}}{S_{Fmax}} \right\rceil, \tag{10}$$

$$S_{Flast} = S_{MS} - [(N_{F\leftarrow MS} - 1) \cdot S_{Fmax}] \tag{11}$$

where $\lceil x \rceil$ indicates the lowest integer not less than x.
It follows that each *MSDU* is carried at the MAC layer by:

- $(N_{F\leftarrow MS} - 1)$ fragments of size S_{Fmax},
- 1 fragment of size S_{Flast}.

Of course, if S_{MS} is a multiple of S_{Fmax}, then S_{Flast}=S_{Fmax}.
Let us assume now, for the sake of simplicity, that no packetization is performed at the MAC layer; each fragment is therefore mapped onto one *MPDU* with the addition of the MAC layer overhead of O_M bytes. It follows that the number $N_{MP\leftarrow MS}$ of *MPDUs* needed to carry a single *MSDU* is equal to $N_{F\leftarrow MS}$. Since all but (in case) the last *MPDU* have the same size, we have:

- $(N_{MP\leftarrow MS} - 1)$ $MPDUs$ of size S_{MPmax},
- 1 $MPDU$ of size S_{MPlast},

where:

$$\begin{aligned} N_{MP\leftarrow MS} &= N_{F\leftarrow MS}, \\ S_{MPmax} &= S_{Fmax} + O_M, \\ S_{MPlast} &= S_{Flast} + O_M. \end{aligned} \tag{12}$$

Of course, if $N_{MP\leftarrow MS} = 1$, each $MPDU$ carries a complete $MSDU$ and its size is S_{MPlast}.

4.3 MPDUs into OFDMA-slots

Starting from the results obtained in subsection 4.2, we can now derive the number of OFDMA-slots needed to carry any $MPDU$ and, as a consequence, any $MSDU$.

Let us recall that every transmission mode j can convey a different amount $S_{Sl}^{(j)}$ of data bytes into a single OFDMA-slot (see table B in figure 1); since there are two possible sizes for $MPDUs$ (S_{MPmax} and S_{MPlast}), we can derive, for every transmission mode j, the minimum number of slots needed to carry each of them:

$$\begin{aligned} N_{Sl\leftarrow MPmax}^{(j)} &= \left\lceil \frac{S_{MPmax}}{S_{Sl}^{(j)}} \right\rceil, \\ N_{Sl\leftarrow MPlast}^{(j)} &= \left\lceil \frac{S_{MPlast}}{S_{Sl}^{(j)}} \right\rceil. \end{aligned} \tag{13}$$

These equations show that when the $MPDU$ size is not a multiple of the amount of bytes carried by a single OFDMA-slot, some padding bits have to be added in order to fill the last slot, wasting some resources.

Thus, a single $MSDU$ is transmitted with the generic transmission mode j through:

- $(N_{MP\leftarrow MS} - 1)$ $MPDUs$ of size S_{MPmax}, accommodated into $(N_{MP\leftarrow MS} - 1) \cdot N_{Sl\leftarrow MPmax}^{(j)}$ slots,
- 1 $MPDU$ of size S_{MPlast}, accommodated into $N_{Sl\leftarrow MPlast}^{(j)}$ slots.

Assuming no resource wastage due to data regions' allocations [1], the number $N_{Sl\leftarrow MS}^{(j)}$ of OFDMA-slots needed to carry a complete $MSDU$ adopting transmission mode j is given by:

$$N_{Sl\leftarrow MS}^{(j)} = ((N_{MP\leftarrow MS} - 1) \cdot N_{Sl\leftarrow MPmax}^{(j)}) + N_{Sl\leftarrow MPlast}^{(j)} \,. \tag{14}$$

Recalling that the scope of the analysis reported in this section was to derive the amount of OFDMA-slots needed to accomodate an $ASDU$, we can state that (14) is the final outcome of this section since, jointly with the equations reported in subsection 4.2, it achieves our end.

5. System performance: throughput of a TCP connection

Having derived the resources (that is, OFDMA-slots) available for data allocation at the physical layer (in section 3) and the amount of resources needed to carry each $ASDU$ (in section 4), we can now assess the performance level provided by IEEE802.16e for a given configuration.

[1] Data regions must be squared or rectangular.

In this section, in particular, a TCP connection is considered, requiring a best effort service. In order to evaluate the maximum end-to-end throughput at the application layer (hereafter simply denoted as throughput), a single connection is supposed to be active either in the downlink or in the uplink. Furthermore, the return link (the uplink when considering a downlink TCP flow and vice versa) is considered to be always sufficient for the transmission of TCP acknowledgments; this assumption implies that the analysis is focused on the direction of the data flow.

Since a single link (uplink or downlink) is here considered for throughput derivation, in the following the subscript *dl* or *ul* will be omitted. Similarly, since a single (generic) transmission mode is considered, the superscript *j* will also be omitted.

In section 3.3 we evaluated the number N_{Sl} of OFDMA-slots available in a given direction (uplink or downlink) and in section 4.3 we derived the number $N_{Sl \leftarrow MS}$ of OFDMA-slots needed to carry a complete $MSDU$, that is, a complete $ASDU$; it follows that the number N_{MS} of complete $MSDUs$ that can be allocated in the considered subframe will be, therefore:

$$N_{MS} = \left\lfloor \frac{N_{Sl}}{N_{Sl \leftarrow MS}} \right\rfloor . \quad (15)$$

Since, in general, N_{Sl} is not a multiple of $N_{Sl \leftarrow MS}$, it follows that R_{Sl} slots will remain available for the allocation of $MPDUs$ that do not form a complete $MSDU$, where $R_{Sl} = mod(N_{Sl}, N_{Sl \leftarrow MS})$, having denoted with $mod(A, B)$ the remainder of the division $\frac{A}{B}$.

In particular, the number of $MPDUs$ of size S_{MPmax} (which occupy $N_{Sl \leftarrow MPmax}$ slots each) that can be accommodated in R_{Sl} free slots is given by:

$$N_{MPinR_{sl}} = \left\lfloor \frac{R_{Sl}}{N_{Sl \leftarrow MPmax}} \right\rfloor . \quad (16)$$

In general, after the allocation of $N_{MPinR_{sl}}$ $MPDUs$ of size S_{MPmax}, a residual amount RR_{Sl} of slots will remain available, with $RR_{Sl} = mod(R_{Sl}, N_{Sl \leftarrow MPmax})$; of course, another $MPDU$ of size S_{MPmax} cannot be allocated, but it could be possible to accommodate an $MPDU$ of size S_{MPlast} occupying $N_{Sl \leftarrow MPlast}$ slots.

If we denote with $N_{MSextra}$ the fraction of $MSDU$ that, in average, can be allocated in the residual region of size R_{Sl}, it results:

$$N_{MSextra} = \begin{cases} \frac{N_{MPinR_{sl}}}{N_{MP \leftarrow MS}} & if\ RR_{Sl} < N_{Sl \leftarrow MPlast} \\ \frac{N_{MPinR_{sl}}}{N_{MP \leftarrow MS} - 1} & else. \end{cases} \quad (17)$$

Obviously, if $N_{MPinR_{sl}} = 0$ it results $N_{MSextra} = 0$.

Let us recall that the amount of application layer data conveyed by a single $MSDU$ is given by S_{AS} bytes; it follows that the average amount of application layer data accommodated in a given subframe is:

$$S = (N_{MS} + N_{MSextra}) \cdot S_{AS}, \quad (18)$$

The application layer throughput can be thus immediately derived as:

$$Thr_A[bit/s] = \frac{8 \cdot S[bytes]}{T_F[s]}, \quad (19)$$

where T_F represents the frame duration (see section 3.2).

5.1 Numerical results

In this section some numerical results obtained through (19) are given. ASDUs of $S_{AS} = 1460$ bytes were chosen, since this is the payload size of a typical TCP/IP packet. Considering 20 bytes for the IP overhead, 20 bytes for the TCP overhead and neglecting the overhead introduced by the upper layers, we assumed that each MSDU has a size of $S_{MS} = S_{AS} + O_{HL} = 1500$ bytes.
A further overhead of $O_M = 10$ bytes is introduced by the MAC layer, following the assumption of no packetization.
The OFDM modulation parameters were set to $N_{FFT} = 2048$, $BW = 7$ MHz, $G = 1/32$; moreover a frame duration of $T_F = 10\ ms$ has been chosen and $RTG = TTG = 116\ \mu s$ were considered; all other physical layer parameters are consequently derived (e.g. $T_S = 264\ \mu s$).
The impact of the remaining parameters affecting the throughput will be investigated in the following. In particular, different values of S_{Fmax} and all transmission modes and permutations schemes will be considered.
In figure 4, a comparison between physical layer and application layer throughput is given varying the fragments maximum size S_{Fmax}. The physical layer throughput Thr_P has been evaluated considering the total amount of bits carried over the medium by all available OFDMA-slots, as follows:

$$Thr_P[bit/s] = \frac{N_{Sl} \cdot 8 \cdot S_{Sl}[bytes]}{T_F[s]}, \tag{20}$$

where the same notation introduced in section 4.2 has been adopted (please note that the previous equation considers only those resources available for data transmission, thus excluding the preamble symbol and the subcarriers used for signalling and control messages).

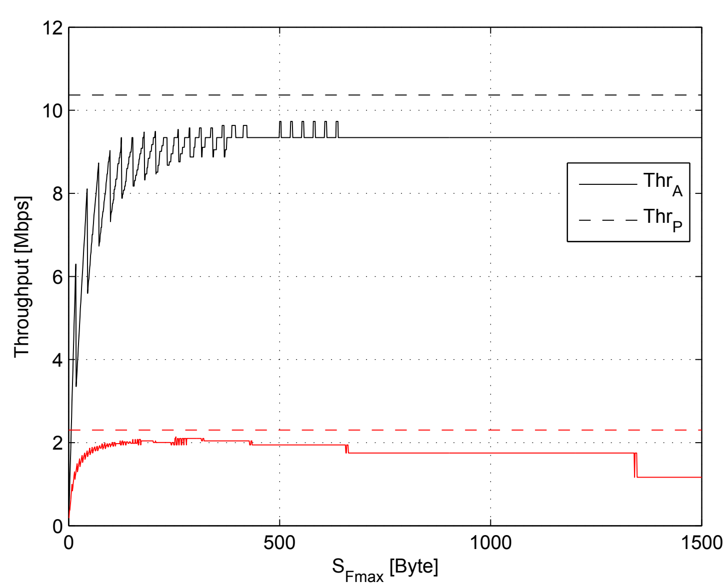


Fig. 4. Comparison between physical layer and application layer throughput (Thr_P and Thr_A) varying the fragment (ARQ block) maximum size S_{Fmax}. Transmission modes 0 and 6.

DL-PUSC has been considered in the downlink and UL-PUSC in the uplink, with $AF_{in} = 1$ and $AF_{Sl} = 1.37$ (following the equations reported in section 3.3, we obtain $N_{Sy_{dl}} = 16$, $N_{Sy_{ul}} = 15$, $N_{Sl_{dl}} = 480$ and $N_{Sl_{ul}} = 350$). Transmission modes 0 and 6 are considered.

The comparison between the dashed and solid curves highlights the reduction of throughput due to both the allocation procedure of $ASDUs$ and the overhead. As can be noted, small variations in the choice of S_{Fmax} may affect the system performance, due to the slot granularity in the physical resource allocation and the impossibility to further divide a fragment (i.e., an ARQ block).

These curves also show that a too small value of S_{Fmax} should not be chosen (this is mainly a consequence of the presence of the overheads O_{HL} and O_M). However, although considering error prone transmissions is out of the scope of the present work, it should be clear that a large value of S_{Fmax} should be avoided too, since each ARQ block must be entirely retransmitted if not correctly received.

Figure 5 deepens the previous results focusing the attention on the application layer throughput and considering all transmission modes.

A direct comparison of the throughput perceived adopting the different transmission modes as a function of S_{Fmax} shows that the choice of S_{Fmax} is a tricky task, since there is no optimal value providing the maximum throughput for all transmission modes.

As can be observed, a number of choices for S_{Fmax} are highlighted through vertical lines and the correspondent throughput values with circles. These values are somehow suboptimal and have been chosen according to the following steps:

1. for each value of S_{Fmax} in the range [1, 500 bytes] the throughput values achieved by each transmission mode normalized to the peak value for that mode were summed into $SUM_{Thr}(S_{Fmax})$;
2. the values of S_{Fmax} that brought to relative maximum of $SUM_{Thr}(S_{Fmax})$ were found, neglecting those values of S_{Fmax} that do not give an absolute value of $SUM_{Thr}(S_{Fmax})$ higher than the previous one.

The values of S_{Fmax} derived as previously described ($S_{Fmax} = 98, 125, 134, 152, 206, 254, 422$ bytes) allow to reduce resource wasting when a single fragment (ARQ block) is transmitted in a single $MPDU$.

In figure 6 the value of AF_{Sl} is compared to AF_{App}, which is defined as the ratio between the maximum application layer throughput in downlink and the one in uplink. The DL-PUSC and UL-PUSC permutation schemes have been considered in the downlink and in the uplink, respectively; $AF_{in} = 1$, $AF_{in} = 2$ and $AF_{in} = 3$ have been assumed as desired asymmetry factors. Transmission mode 6 only.

As can be noted, a good match between AF_{Sl} and AF_{App} is achieved for all the considered AF_{in} (avoiding to consider too large values for S_{Fmax}). On the contrary, it is quite hard to exactly respect the desired AF_{in} with no wasting: note, in fact, that the cases $AF_{in} = 1$ and $AF_{in} = 2$ bring to the same result (that is, the need to minimize the resource wasting brings, in both cases, to the same choice of $N_{Sy_{dl}}$ and $N_{Sy_{ul}}$).

In figure 7 the throughput is shown as a function of the number of OFDM symbols available for data transmission in the downlink subframe for all possible permutation schemes (refer to table 1). In this case, $N_{Sy_{dl}}$ is set and the correspondent AF_{Sl} follows as a consequence (see section 3.3). Transmission mode 6 and $S_{Fmax} = 206$bytes have been considered. This figure also highlights that an increase (reduction) in $N_{Sy_{dl}}$ has an effect only if it involves at least N_{GS} symbols.

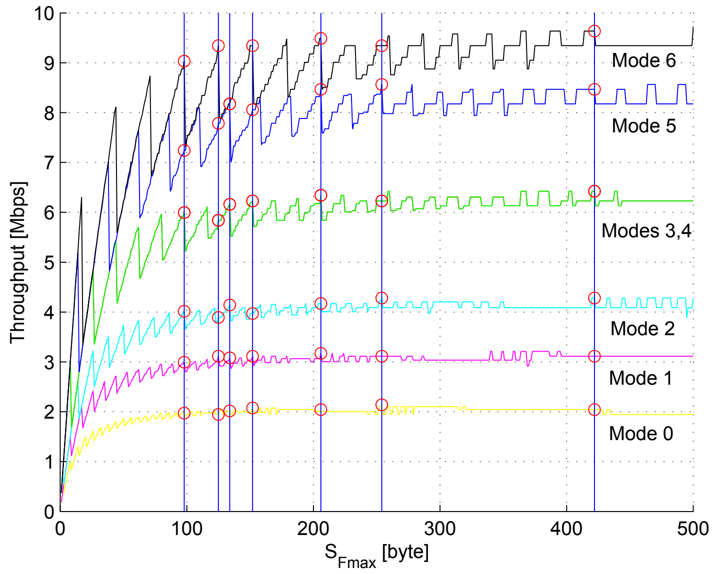


Fig. 5. Application layer throughput (Thr_A) varying the fragments (ARQ blocks) maximum size S_{Fmax} for all transmission modes. The values of S_{Fmax} that allow to have a good occupation adopting any possible transmission mode are marked with vertical lines and small circles (o); they correspond to $S_{Fmax} = 98, 125, 134, 152, 206, 254, 422$ bytes

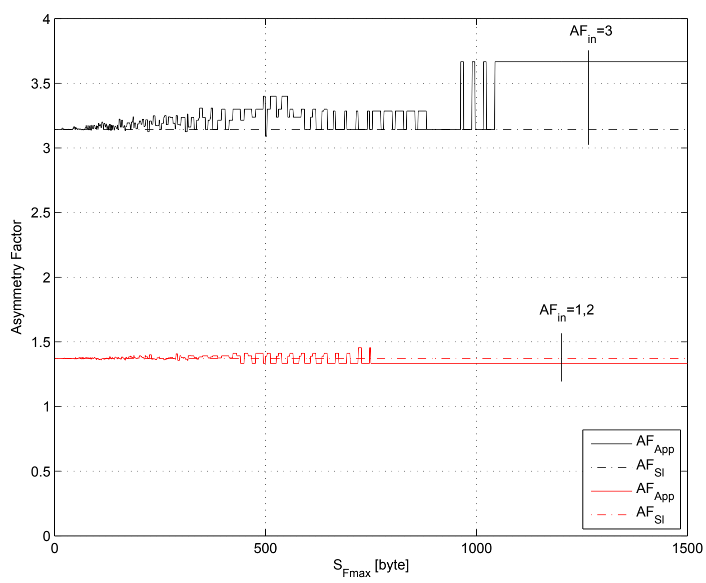


Fig. 6. Comparison between AF_{SI} and AF_{App}, given $AF_{in} = 1$, $AF_{in} = 2$ and $AF_{in} = 3$. Transmission mode 6.

6. System performance: VoIP capacity on UGS or ertPS

In this section the maximum number of users performing a VoIP call that can be served by IEEE802.16e is evaluated, following the analysis described in sections 3 and 4.

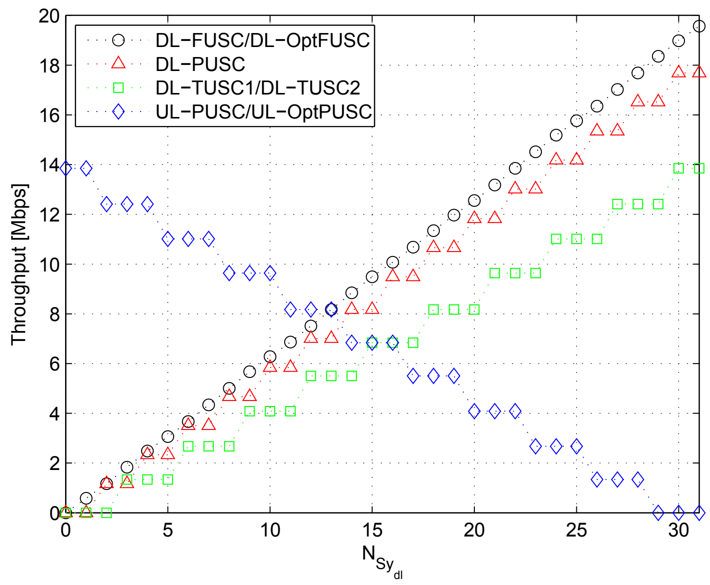


Fig. 7. Comparison of application layer throughput (Thr_A) adopting the various permutation schemes, varying the number of downlink useful symbols $N_{Sy_{dl}}$. Transmission mode 6. $S_{Fmax} = 206$ bytes.

A description of the considered VoIP codecs and scheduling services is given before entering into the details of the analytical model.

6.1 UGS and ertPS scheduling services

As already mentioned in section 2, five scheduling services are provided by the IEEE802.16e specification for traffic delivery. Since our attention is now focused on real-time VoIP traffic, UGS and ertPS are the only possible choices, due to latency constraints, and are therefore considered in the following:

- ***Unsolicited grant service (UGS)*** is designed to support real-time uplink service flows that generate transport fixed-size data packets on a periodic basis, such as T1/E1 and VoIP without silence suppression. The service offers fixed size grants on a real-time periodic basis, which eliminate the overhead and latency of user's requests and assure that grants are available to meet the flowŠs real-time needs.
- ***Extended real-time polling service (ertPS)*** (Lee et al., 2006) improves UGS when the application layer rate varies in time. The base station (BS) shall provide unicast grants in an unsolicited manner like in UGS, thus saving the latency of a bandwidth request. However, whereas UGS allocations are fixed in size, ertPS allocations are dynamic. The BS may provide periodic uplink allocations that may be used for requesting the bandwidth as well as for data transfer. By default, size of allocations corresponds to current value of maximum sustained traffic Rate at the connection. Users may request changing the size of the uplink allocation by either using an extended piggyback request field of the grant management subheader or using BR field of the MAC signaling headers, or sending a specific codeword over the signalling channel CQICH. The BS shall not change the size of uplink allocations until receiving another bandwidth change request from the user.

6.2 VoIP codecs

The most important and mainly adopted voice codecs have been considered:

1. ***ITU G.711*** (ITU-T Rec. G.711, 1988), the well known constant bit rate PCM at 64 kbps; this is the codec used in PSTN networks, with no compression, neither during the speech nor during silences of a conversation;
2. ***ITU G.729*** (ITU-T Rec. G.729, 1996a), the most used codec for VoIP, at 8 kbps; when an active speech period is detected, it produces one packet of 80 bits every 10 ms, but more than one packet may be concatenated in order to reduce protocols overheads (Goode, 2002); obviously, this process enlarges the average delivery delay of packets. Hereafter, we will consider the concatenation of a couple of packets (Goode, 2002) and we will denote this codec as *G*.729. 160 bits packets are thus generated every 20 ms.
3. ***AMR*** (adaptive multi rate (3GPP TS 26.071, 2008)), standardized by 3GPP and used for voice in second and third generation cellular radio access; it generates one packet every 20 ms with a variable data rate, going from a minimum of 4.75 kbps (95 data bits plus 18 overhead bits for each packet) to a maximum of 12.2 kbps (244 data bits plus 18 overhead bits for each packet). The variation of the data rate is given in order to better select the appropriate tradeoff between resource usage and speech quality (obviously, a data rate reduction leads to a quality degradation). In order to investigate the capacity of the IEEE802.16e WirelessMAN-OFDMA system with the AMR codec, we will consider both the minimum data rate (4.75 kbps, denoted as $AMR4.75$) and the maximum data rate (12.2 kbps, denoted as $AMR12.2$).

ITU G.729 and *AMR codecs* have been designed, in particular, with the specific goal to reduce the resources occupation: when no voice activity is detected the *silence suppression* procedure is activated. As a consequence small and less frequent packets are transmitted, which convey the information for a "comfortable noise" generation at the receiver side.
In particular, adopting the *AMR* codec, the detection of a silence period (3GPP TS 26.092, 2008) gives rise to the following steps:

- eight full voice packets are normally transmitted in the first interval, called *hangover period*;
- then, a SID (silence insert descriptor) packet (36 data bits plus 18 overhead bits) is transmitted after 60 ms;
- further SID packets are transmitted every 160 ms until a new speech activity is detected.

This process is depicted in the topmost part of figure 8.
As far as the *G.729* codec is concerned, the silence suppression procedure is defined in the Annex B (Benyassine et al., 1997; ITU-T Rec. G.729, 1996b). In this case, each SID packet has a length of 15 bits. However, the time interval between two successive SID packets is not fixed: in fact, for a good quality at the receiver, a lower or a greater transmission rate may be needed depending on the specific background noise observed in each environment.
In order to allow the derivation of meaningful results, SID packets are here supposed to be generated with a fixed rate during silences, as with *AMR*. In particular, the rate of *G.729* SID packets is assumed to be twice the one adopted by *AMR*, following the results provided in (Estepa et al., 2005).
In the following, the *activity factor* ν is defined as the ratio between the time during which full packets are generated and the total duration of the conversation. Thus, in particular, we

Codec	B [bits]	ρ [pack/s]	B_{SID} [bits]	ρ_{SID} [pack/s]	ν	Maximum number of allowed users Mode 0	Mode 6
$G.711$	1280	50	-	-	1	20 / __	87 / __
$G.729$	160	50	-	-	1	58 / __	233 / __
$G.729_{ss}$	160	50	15	12.5	0.45	58 / 90	233 / 482
$AMR4.75$	113	50	-	-	1	63 / __	233 / __
$AMR4.75_{ss}$	113	50	54	6.25	0.45	63 / 111	233 / 506
$AMR12.2$	262	50	-	-	1	50 / __	175 / __
$AMR12.2_{ss}$	262	50	54	6.25	0.45	50 / 90	175 / 374

Table 2. Codec parameters and analytical calculation of maximum number of allowed users. Multiple values refer to $UGS/ertPS$.

will adopt $\nu = 1$ when no silence suppression is activated and $\nu < 1$ in the case of silence suppression.

In the latter case, in order to derive a realistic value of ν, we simulated the dynamic of a conversation according to a detailed model that takes into account also periods of simultaneous talks or silences of the two parties, and the short gaps through the speeches (Stern et al., 1996). This model gives a 33% of voice activity over the total conversation duration for each of the two parties. It must be noted, however, that since 8 full packets are still transmitted during the hangover period, the activity factor ν is approximately 0.45; please note that this value corresponds to both our simulations and the experimental results given in (Estepa et al., 2005). As for the less sophisticated *G.711* codec, here it was considered only as a reference, and no silence suppression is introduced.

The parameters of all considered codecs, with and without silence suppression, are given in the first four columns of table 2. In particular: the first column indicates the considered codecs (the subscript ss indicates that silence suppression is considered); the second column defines the number of bits B of full packets and the number of full packets per second ρ that are transmitted when the voice is detected; similarly, the third column defines the number of bits B_{SID} of SID packets and the number ρ_{SID} of SID packets per second that are transmitted when silences are detected; finally, the fourth column represents the activity factor ν; the rest of the table will be illustrated in the following.

6.3 Amount of supported VoIP users

In this section, the number of VoIP users that can be served will be evaluated as a function of the adopted codec x, the scheduling service k and the transmission mode j (all users are supposed to be served adopting the same mode). Also in this case we need to consider the whole packet processing from the application layer down to the transmission over the medium.

Before being transmitted, each packet generated by the codec must pass through the whole protocol pillar, thus increasing its size owing to the overheads introduced by each protocol layer. In particular, the RTP, UDP, IP and MAC layers overheads are added, which are, respectively, $O_{RTP} = 12$ bytes, $O_{UDP} = 8$ bytes, $O_{IP} = 20$ bytes and $O_M = 10$ bytes (thus, $O_{HL} = O_{RTP} + O_{UDP} + O_{IP} = 40$ bytes). Assuming that no fragmentation is carried out from the application to the MAC layer, the size (in bytes) of full packets and SID packets generated by the codec x is given by:

$$S_{MS}^{(x)} = B^{(x)}/8 + O_{HL} + O_M , \tag{21}$$

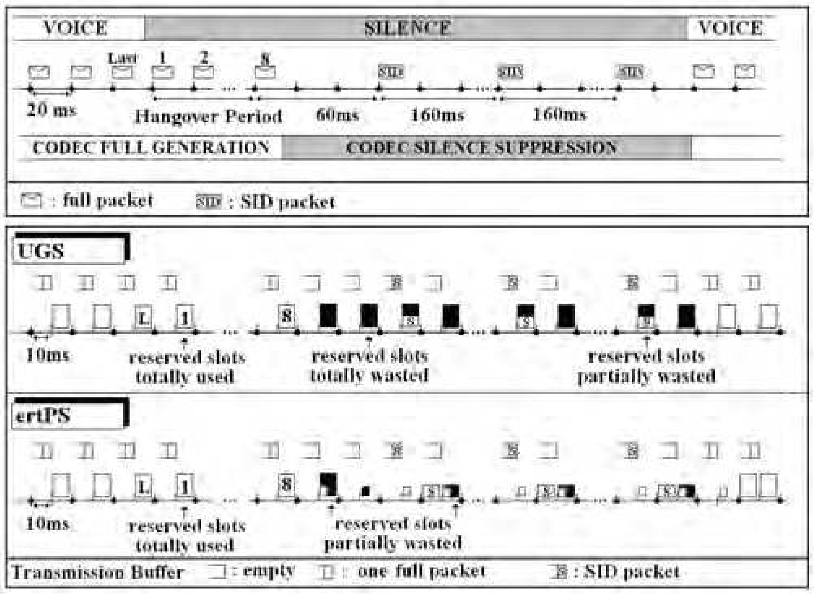


Fig. 8. Topmost part: AMR codec full packets and SID packets generation, with $\rho = 50 pack/s$. Rest of the figure: buffer state and resource allocation of a generic uplink voice traffic flow with UGS and ertPS, following the packets generation depicted in the topmost part.

$$S^{(x)}_{MS_{SID}} = B^{(x)}_{SID}/8 + O_{HL} + O_M \; . \tag{22}$$

Of course, the ARQ mechanism is assumed inactive, since we are considering a real time service.
Let us recall (from section 6.2), now, that for a given codec x:

$$\nu^{(x)} = \frac{Time\ with\ full\ packets}{Total\ conversation\ duration} \tag{23}$$

is the activity factor, while

$$\rho^{(x)} = \frac{Number\ of\ full\ packets}{Time\ with\ full\ packets} \tag{24}$$

is the rate of transmission of full packets (of $B^{(x)}$ bits) during voice activity and

$$\rho^{(x)}_{SID} = \frac{Number\ of\ SID\ packets}{Time\ without\ full\ packets} \tag{25}$$

is the rate of transmission of SID packets (of $B^{(x)}_{SID}$ bits) during silences, if silence suppression is considered.
As already recalled each packet to be transmitted is mapped onto OFDMA-slots, thus, in order to assess the maximum number of users that can be served, the number of slots that are needed for each packet must be firstly calculated, also considering that the number $S^{(j)}_{Sl}$ of bytes carried by one slot depends on the adopted mode j (see table B of figure 1).
In particular, for a given codec x and a given transmission mode j, the number $N^{(x,j)}_{Sl \leftarrow MS}$ of slots needed to carry a full packet and the number $N^{(x,j)}_{Sl \leftarrow MS_{SID}}$ of slots needed to carry a SID

packet are:

$$N_{Sl\leftarrow MS}^{(x,j)} = \left\lceil \frac{S_{MS}^{(x)}}{S_{Sl}^{(j)}} \right\rceil , \quad (26)$$

$$N_{Sl\leftarrow MS_{SID}}^{(x,j)} = \left\lceil \frac{S_{MS_{SID}}^{(x)}}{S_{Sl}^{(j)}} \right\rceil . \quad (27)$$

Depending on the considered scheduling service k and the specific VoIP codec x, the average number of slots required in a given (DL/UL) subframe by a single user adopting mode j, is given by:

$$M_{Sl\leftarrow U}^{(x,k,j)} = f^{(x,k)}\left(N_{Sl\leftarrow MS}^{(x,j)}, N_{Sl\leftarrow MS_{SID}}^{(x,j)}\right), \quad (28)$$

where the analytical expression of $f^{(x,k)}(\cdot)$ will be provided in the following for both the considered scheduling services.
As a general consideration, please note that all difficulties in resource allocation are on the uplink, since the base station has a perfect knowledge of all buffers in the downlink and no resource requests are needed. For this reason, and assuming that $AF_{sl} \geq 1$ (that is, we have more downlink slots than uplink slots), all evaluations will be done focusing on the uplink direction. Thus, given an amount $N_{Sl_{ul}}$ of available slots (in the uplink subframe), the maximum number $N_{U_{MAX}}$ of users can finally be evaluated:

$$N_{U_{MAX}}^{(x,k,j)} = \left\lfloor \frac{N_{Sl_{ul}}}{M_{Sl\leftarrow U}^{(x,k,j)}} \right\rfloor . \quad (29)$$

Performance on UGS. Since UGS resources are statically allocated, as negotiated during connection setup, the silence suppression procedure does not provide any benefit in the uplink direction.
Denoting with T_F the frame duration, (28) becomes:

$$M_{Sl\leftarrow U}^{(x,UGS,j)} = T_F\left(\rho^{(x)} \cdot N_{Sl\leftarrow MS}^{(x,j)}\right), \quad (30)$$

Please note that there is no dependence on the activity factor (some resources will be wasted if $\nu^{(x)} < 1$).
Combining (30) and (29) the maximum number $N_{U_{MAX}}^{(x,UGS,j)}$ of VoIP users supported by the UGS scheduling service can be easily derived.
Performance on ertPS. In the case of ertPS scheduling service, besides the adopted codec x and transmission mode j, also the activity factor $\nu^{(x)}$ must be considered in order to derive the amount of supported VoIP users. After the hangover period, in fact, the transmission rate is reduced and a single slot is allocated ($N_{Sl\leftarrow MH} = 1$) in order to allow the transmission of a stand-alone MAC signaling header for a quick modification of the resources request (see figure 8). Please note, by the way, that, although reduced, this allocation entails a resource wasting, since no transmission is performed in the most of cases. Resource wasting will occur also at any rate reduction (refer to figure 8): before a rate decreases, in fact, the request is sent in the uplink using an oversized resource.

Let us observe, furthermore, that following a rate increase request, the first (larger) resource is allocated in the subsequent frame, without respecting the normal rate of allocation, in order to reduce the latency.

In the case of ertPS scheduling service, therefore, in order to calculate $M_{Sl \leftarrow U}^{(x,ertPS,j)}$ we must take into account:

- the slots needed for full packets, that are generated with rate $\rho^{(x)}$ during active voice periods: $R \cdot N_{Sl \leftarrow MS}^{(x,j)}$ average slots per second, where $R = \nu^{(x)} \cdot \rho^{(x)}$ indicates the average number of full packets per second;
- the slots needed for SID packets, that are generated with rate $\rho_{SID}^{(x)}$ during silent periods: $R_{SID} \cdot N_{Sl \leftarrow MS_{SID}}^{(x,j)}$ average slots per second, where $R_{SID} = (1 - \nu^{(x)}) \cdot \rho_{SID}^{(x)}$ indicates the average number of SID packets per second;
- the slots needed for stand-alone MAC headers, allocated with rate $\rho^{(x)}$ when neither full packets nor SID packets are generated: $R_{MH} \cdot N_{Sl \leftarrow MH}$ average slots per second, where $R_{MH} = 1 - R - R_{SID}$; indicates the average number of slots left for MAC headers per second, transmitted at the same rate;
- the slots that are wasted during variations from full packets allocation to stand-alone MAC headers allocations: $R_{CFG} \cdot N_{Sl \leftarrow MS}^{(x,j)}$ average slots per second, where $R_{CFG} = \frac{1}{T_{CFG}^{av} + T_{CSS}^{av}}$ indicates the average number of uninterrupted periods of full packets per second, that depends on the average duration of a period of continuous full packets generation by the codec (T_{CFG}^{av}) and on the average duration of a period of silence suppression with SID packets generation (T_{CSS}^{av});
- the slots that are wasted during variations from SID packets to stand-alone MAC headers: $R_{SID} \cdot N_{Sl \leftarrow MS_{SID}}^{(x,j)}$ average slots per second.

Thus, in this case, $M_{Sl \leftarrow U}^{(x,ertPS,j)}$ is given by:

$$M_{Sl \leftarrow U}^{(x,ertPS,j)} = T_F \left((R + R_{CFG}) \cdot N_{Sl \leftarrow MS}^{(x,j)} + 2 \cdot R_{SID} \cdot N_{Sl \leftarrow MS_{SID}}^{(x,j)} + R_{MH} \cdot N_{Sl \leftarrow MH} \right), \tag{31}$$

Concerning the parameters T_{CFG}^{av} and T_{CSS}^{av}, please note that they are not only related to the adopted codec, but also to the characteristics of the specific conversation, also including, as an example, the language; in order to derive meaningful results, hereafter we adopted the values reported in (Stern et al., 1996) ($T_{CFG}^{av} = 0.17$s and $T_{CSS}^{av} = 0.3428$s), although they are related to the voice-silence intervals rather than to codec full generation-codec silence suppression (they do not consider the hangover periods); for this reason, a slight underestimation of the maximum number of VoIP users is expected in the numerical results.

Finally, combining (31) and (29), the maximum number $N_{U_{MAX}}^{(x,ertPS,j)}$ of VoIP users supported by the ertPS scheduling service can be easily derived.

6.4 Numerical results

For the numerical results derivation, the amount of OFDMA-slots available for data transmission in the downlink subframe has been assumed equal to $N_{Sl_{ul}} = 350$; this value is a consequence of the same assumption reported in section 5.1: $BW = 7$ MHz, N_{FFT}=2048 OFDM subcarriers, normalized OFDM guard interval $G = 1/32$, frame duration $T_F = 10$ ms (which is the most suited value for VoIP traffic allocation); 15 of the 31 OFDM symbols of each frame

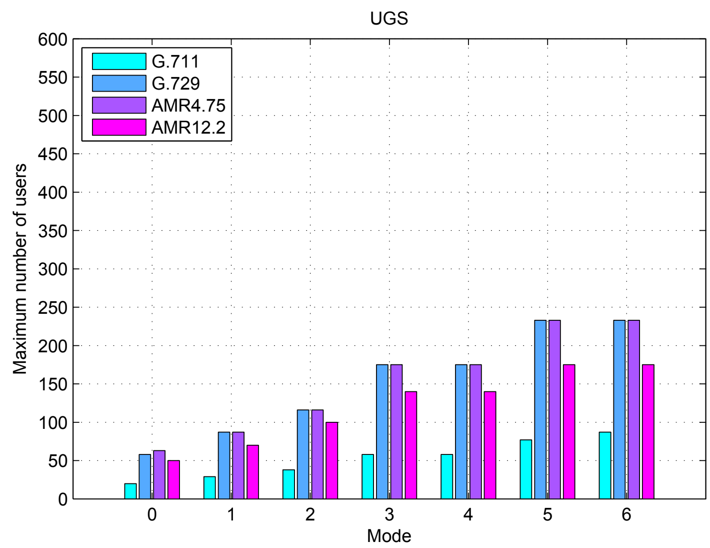


Fig. 9. Maximum number of users using UGS scheduling service. All modes. All VoIP codecs without silence suppression.

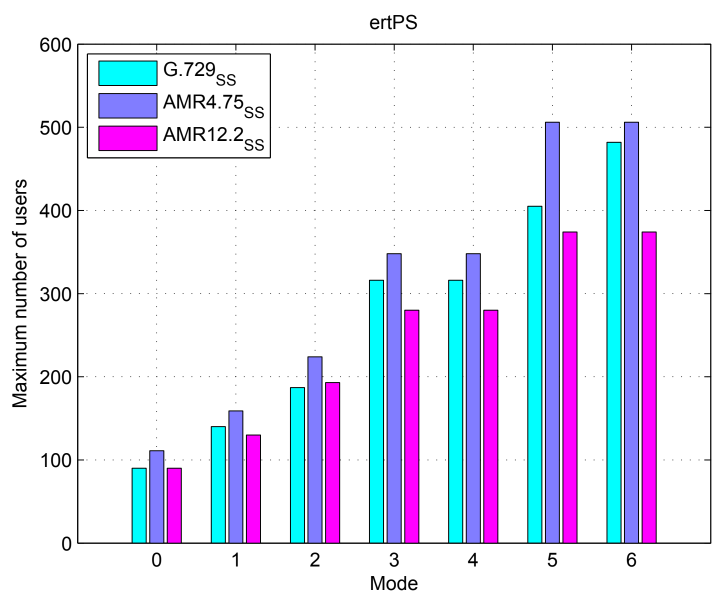


Fig. 10. Maximum number of users using ertPS scheduling services. All modes. All VoIP codecs with silence suppression.

are left for uplink data, adopting UL-PUSC (the rest of the symbols are used for the uplink common channels and the downlink subframe).
In the last column of table 2 the maximum number of VoIP users that can be served with *UGS* and *ertPS* (separated by the symbol "/") are reported for each codec; modes 0 and 6 are considered. Two underscores are typed when the adoption of that scheduling service makes no sense with that codec (i.e., ertPS with no silence suppression).
The maximum amount of VoIP users that can be supported is shown for all modes in figures 9 and 10 focusing on *UGS* and *ertPS*, respectively. Obviously, the VoIP capacity adopting *UGS*

with and without silence suppression is always the same; for this reason the results related to the case of silence suppression are not reported in figure 9.
It can be observed that, as expected, VoIP capacity is strictly related to the average data rate generated by the codec.
Comparing figures 9 and 10 we can also appreciate the significant benefit provided by the adoption of the *ertPS* instead of *UGS*.

7. Conclusions

The performance of IEEE8021.16e WirelessMAN-OFDMA depends on a large number of system parameters and implementation choices, such as, among the others, the available bandwidth, the frame duration, the uplink/downlink traffic asymmetry and the fragmentation policy.
In the previous sections we provided an analytical framework that allows to evaluate the throughput achievable with TCP/IP connections as well as the amount of supported VoIP users as a function of the most significant parameters characterizing this technology. Beside the analytical model derivation, here we provided criteria, equations and algorithms to make the best choices from the viewpoint of system efficiency.
Furthermore, some numerical results were given, showing the impact of some specific parameters on the system performance. With reference to TCP/IP connections, for instance, the troublesome choice of the maximum size of ARQ blocks has been discussed as well as the potential resource wastage entailed by a wrong choice of the asymmetry factor between the downlink and uplink subframes. The main outcome of this analysis is given by a set of criteria to be followed in order to maximize the throughput provided to the final user.
As far as VoIP connections are concerned, here we assessed the maximum number of users that can be supported, carefully considering the voice codec characteristics and the adopted scheduling service. The outcomes of this investigation provide an indication about the capacity of this technology to be alternative to other technologies such as UMTS and LTE for the provision of the voice service.
As a final remark, let us observe that the analytical framework proposed in this chapter provides a tool to evaluate the upper limits of the throughput and the maximum amount of VoIP users that can be supported by IEEE802.16e WirelessMAN-OFDMA. However, in order to investigate the actual performance of such a complex technology in a given scenario considering the degradation due, for instance, to fading, shadowing, noise and interference, the only feasible way is to adopt a simulation tool able to carefully reproduce all aspects of communications, with particular reference to the physical layer behavior.
This kind of investigation has been carried out at WiLab (Italy) by means of the simulation platform SHINE, that has been developed in the last years to assess the performance of wireless networks in realistic scenarios. The interested reader may refer to (Andrisano et al., 2007; 2009; Bazzi et al., 2006).

Appendix I

Here we illustrate the algorithm that allows to maximize symbols usage starting from the number of useful symbols $N_{USy} = N_{Sy} - N_{OSy_{dl}} - N_{OSy_{ul}}$.
The following steps must be followed:

1. derive $J = Res_{ul} = mod(N_{USy}, N_{GS_{ul}})$;
2. find, if possible:

$$a = \min_{a \in \left[0, \lfloor \frac{N_{USy}}{N_{GS_{ul}}} \rfloor\right]} \ so\ that\ \ mod(J + a \cdot N_{GS_{ul}}, N_{GS_{dl}}) = 0. \tag{32}$$

3. if a was not found, then reduce J by one and return to step 2, else exit.

The obtained value a and the parameter J allow the derivation of the minimum value for $N_{Sy_{dl}}$ ($N_{Sy_{dl}}^{MIN}$) that maximally reduces the symbols wasting:

$$N_{Sy_{dl}}^{MIN} = J + a \cdot N_{GS_{ul}}, \tag{33}$$

All possible solutions will be:

$$N_{Sy_{dl}}^{OPT}(b) = N_{Sy_{dl}}^{MIN} + b \cdot mcm(N_{GS_{ul}}, N_{GS_{dl}}), \tag{34}$$

where $b \in \left[0, \lfloor \frac{N_{USy} - N_{Sy_{dl}}^{MIN}}{N_{GS_{dl}}} \rfloor\right]$ and $mcm(x, y)$ is the minimum common multiplier of x and y.
It follows:

$$N_{Sy_{ul}}^{OPT}(b) = N_{USy} - (Res_{ul} - J) - N_{Sy_{dl}}^{OPT}(b). \tag{35}$$

and

$$AF_{Sl}(b) = \frac{N_{Sl_{dl}}^{OPT}(b)}{N_{Sl_{ul}}^{OPT}(b)}. \tag{36}$$

Finally, we can choose the value of b that brings to the $AF_{Sl}(b)$ nearer to AF_{in}.

8. References

3GPP TS 26.071 (2008). Mandatory speech codec speech processing functions; amr speech codec; general description.

3GPP TS 26.092 (2008). Mandatory speech codec speech processing functions; adaptive multi-rate (amr) speech codec; comfort noise aspects.

Ahmadi, S. (2009). An overview of next-generation mobile wimax technology, *IEEE Communications Magazine* **Vol.47**(n.6): pp.84–98.

Andrisano, O., Bazzi, A., Leonardi, G. & Pasolini, G. (2007). Ieee802.16e best effort performance investigation, *Proceedings of IEEE International Conference on Communications, 2007 (ICC 2007)*, IEEE, Glasgow, Scotland, pp. 4837–4842.

Andrisano, O., Bazzi, A., Leonardi, G. & Pasolini, G. (2009). Ieee802.16e simulation issues, *Proceedings of IEEE Mobile WiMAX Symposium 2009 (MWS 2009)*, IEEE, Napa Valley, California, pp. –.

Bazzi, A., Gambetti, C. & Pasolini, G. (2006). Shine: Simulation platform for heterogeneous interworking networks, *Proceedings of IEEE International Conference on Communications, 2006 (ICC 2006)*, IEEE, Istanbul, Turkey, pp. 5534–5539.

Benyassine, A., Shlomot, E. & Su, H. (1997). Itu-t recommendation g.729 annex b: a silence compression scheme for use with g.729 optimized for v.70 digital simultaneous voice and data applications., *IEEE Communications Magazine* **Vol.35**(Issue 9): pp.64–73.

Cimini, J. (1985). Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing., *IEEE Trans. Comm.* **Vol.COM-33**(n.7): pp.665–675.

Estepa, R., Vozmediano, J. & Estepa, A. (2005). Accurate prediction of voip traffic mean bit rate., *ELECTRONICS LETTERS* **Vol.41**(n.17): pp.985–987.

Goode, B. (2002). Voice over internet protocol (voip), *Proceedings of the IEEE* **Vol.90**(n.9): pp.1495–1517.

IEEE Std 802.16-2004 (2004). Ieee standard for local and metropolitan area networks part 16: Air interface for fixed broadband wireless access systems.

IEEE Std 802.16e-2005 (2006). Ieee std 802.16eŹ-2005 and ieee std 802.16Ź-2004/cor1-2005 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.

ITU-T Rec. G.711 (1988). Pulse code modulation (pcm) of voice frequencies.

ITU-T Rec. G.729 (1996a). Coding of speech at 8kbit/s using conjugate-structure algebraic-code-excited linear-predictive (cs-acelp)coding.

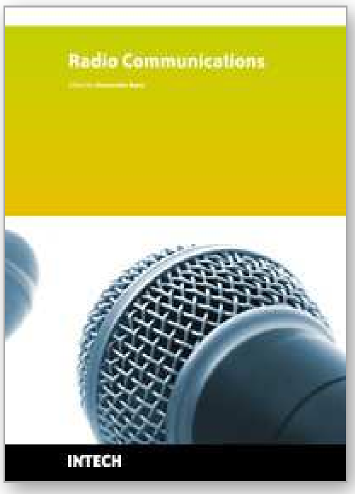
ITU-T Rec. G.729, A. (1996b). A silence compression scheme for g.729 optimized for terminals conforming to itu-t v.70.

Lee, H., Kwon, T., Cho, D., Lim, G. & Chang, Y. (2006). Performance analysis of scheduling algorithms for voip services in ieee 802.16e systems, *Proceedings of IEEE Vehicular Technology Conference (VTC 2006-Spring).*, IEEE, Melbourne, Australia, pp. 1231–1235.

Li, Q., Lin, X., Zhang, J. & Roh, W. (2009). Advancement of mimo technology in wimax: from ieee 802.16d/e/j to 802.16m, *IEEE Communications Magazine* **Vol.47**(n.6): pp.100–107.

Stern, H., Mahmoud, S. & Wong, K. (1996). A comprehensive model for voice activity in conversational speech-development and application to performance analysis of new-generation wireless communication systems, *Wireless Networks* **Vol.2**(n.4): pp.359–367.

Van Nee, R. & Prasad, R. (2000). *OFDM for Wireless Multimedia Communications*, Artech House.



Radio Communications

Edited by Alessandro Bazzi

ISBN 978-953-307-091-9
Hard cover, 712 pages
Publisher InTech
Published online 01, April, 2010
Published in print edition April, 2010

In the last decades the restless evolution of information and communication technologies (ICT) brought to a deep transformation of our habits. The growth of the Internet and the advances in hardware and software implementations modified our way to communicate and to share information. In this book, an overview of the major issues faced today by researchers in the field of radio communications is given through 35 high quality chapters written by specialists working in universities and research centers all over the world. Various aspects will be deeply discussed: channel modeling, beamforming, multiple antennas, cooperative networks, opportunistic scheduling, advanced admission control, handover management, systems performance assessment, routing issues in mobility conditions, localization, web security. Advanced techniques for the radio resource management will be discussed both in single and multiple radio technologies; either in infrastructure, mesh or ad hoc networks.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Alessandro Bazzi, Giacomo Leonardi, Gianni Pasolini and Oreste Andrisano (2010). Mobile WiMAX Performance Investigation, Radio Communications, Alessandro Bazzi (Ed.), ISBN: 978-953-307-091-9, InTech, Available from: http://www.intechopen.com/books/radio-communications/mobile-wimax-performance-investigation

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821