Quality of Service Management in IEEE 802.16 Wireless Metropolitan Area Networks

Daniele Tarchi, Romano Fantacci, and Marco Bardazzi
Department of Electronics and Telecommunications
University of Florence
50139 Firenze, Italy
Email: tarchi@lenst.det.unifi.it, romano.fantacci@unifi.it, bardazzi@lenst.det.unifi.it

Abstract—In the last years the Metropolitan Area Networks (MAN) have increased their popularity and kept the interest of the most important research groups all over the world. Several standards have been published which represent the first step for developing metropolitan networks: IEEE 802.16 (WiMAX) has taken a relevant role in reaching the goal of realizing a full-service network all over a urban and suburban area. This standard provide high data rate in a big covering range with low implementation costs, multi-traffic communication and the possibility of creating broadcast, multicast and mesh networks. The WiMAX networks can have a pre-defined structure, with a central Base Station (BS) covering a cell in which a variable number of Subscriber Stations (SS) could work, or a mesh distribution with SSs communicating together without BS participation. In this paper a QoS management model is proposed for a centralized structure in which the BS schedules the Best-Effort and VoIP traffics of several SSs.

I. INTRODUCTION

Wireless technologies characterize our way of life deeper and deeper in the modern culture: this has lead to a sensible improvement of research, development and investments in communications that did not need a physical wired link. Starting from the Wireless Personal Area Networks (WPAN), through Wireless Local Area Networks (WLAN) we’re actually facing the problem of covering a big area (like a metropolitan or a country scenario) with a unique wireless network: the use of radio-link communications is due to the very low costs of implementation and the lack of natural obstacles to overcome, while covering an entire city means a newer and easier way of human communication. For these reasons the Wireless Wide Area Networks (WWAN) have been created with the aim of covering an entire city, achieving good reliability and acceptable data rates. In order to improve these rates Wireless Metropolitan Area Networks (WMAN) standards have been realized, introducing a smaller way to view the net and a medium to carry informations faster than the ones included in WWAN.

The IEEE 802.16 standard [1], supported by the WiMAX commercial consortium, concerns the Physical and MAC (Medium Access Control) layers specifications for a Broadband Wireless Access (BWA) communication protocol [2]: its characteristics provide WiFi-like data rate (near 100 Mbps) on a 50 Km average range; for these reasons IEEE 802.16 will become the best way to carry BWA connections in remote areas, where the wired links would be too expensive, realizing the so-called Wireless ADSL last mile.

The IEEE 802.16 is a universal standard comprehending different kinds of link and communication [3]: single carrier, OFDM and OFDMA transmission systems could be implemented; the link physical structure could be organized either TDD or FDD (half and full).

The traffics and services made by the devices composing a WMAN are strongly specialized and they have to be scheduled respecting transmission time and used bandwidth bonds. In order to provide the service’s parameters respect and the QoS, a traffic management model is needed: IEEE 802.16 divides all services in four different groups, distinguished by traffic parameters, bandwidth request and resources allocation modalities. However, WiMAX doesn’t specify any uplink or downlink scheduling algorithm.

In this paper, the problem of dynamic resources allocation in IEEE 802.16 has been analyzed: in particular we refer to the uplink radio system in a scenario with a central BS managing several SSs (either single devices or WiFi-Ethernet isles) that use VoIP and Best-Effort services. In order to manage the bandwidth allocations and the resources’ requests toward the guaranteed QoS, a scheduling algorithm has been proposed, according to the characteristics of each interested service, and its reliability has been proved by implementing the scenario above-cited in the NePSing [4] simulating environment.

The paper is organized as follows. In the Section II it’s introduced the IEEE 802.16 standard and its QoS managing peculiarity. The proposed managing model, with the scheduling algorithm and the assumptions made, is presented in the Section III, while in the Section IV the numerical results of the simulations are produced and explained. Finally Conclusions are drawn.

II. IEEE 802.16 STANDARD SPECIFICATIONS

The IEEE 802.16 standard defines the Medium Access Control (MAC) and Physical (PHY) layers specification of the
WirelessMAN air interface for Broadband Wireless Access. A wireless MAN provides network access to buildings and residential areas through an external antennas system based on a central radio station (i.e. Base Station - BS) and a variable number of user devices (i.e. Mobile Stations - MS), and it offers a very useful alternative to cable connection, realizing less-expensive high-speed radio link communications.

IEEE 802.16 standard was first developed in order to create the so-called wirelessADSL: in the last few years the availability of broadband home connections has become more and more requested and in this way we have attended to a great expansion of DSL covering range in easy to reach areas. Unfortunately, some residential areas are difficult to reach and realizing a cable DSL cover is too expensive: in these cases, the possibility of obtaining high data rate, with certain quality of service requirements and security bounds through a low cost connection easy to install and maintain, is the best way to solve the problem. For all these reasons WiMAX have kept the interests of all the major Internet Service Providers (ISPs) and networks developers of the world and it has grown from a simple way of realizing LMDS-like links [5] to a brand new protocol for all kinds of WMAN with the purpose of introducing devices mobility inside the covered area (competing then with WWAN networks standards like UMTS and WCDMA).

The original standard, i.e. IEEE 802.16-2001 [1], addressed frequency from 10 to 66 GHz, where extensive spectrum is available in all the world countries, even though the realization costs and the highest frequencies introduce serious problems. The other versions from the IEEE 802.16a one [6] to the last one, i.e. IEEE 802.16d [7], instead, cover either the 10-66 GHz (LMDS and, generally, LOS links) band or a 2-11GHz band, including license-exempt frequencies and enabling NLOS communication (this characteristic fits best a metropolitan scenario).

As we said before IEEE 802.16 standard defines MAC and PHY layers of the air interface for a BWA. The PHY layer can be divided into two significant parts: the first including 10-66 GHz frequencies and the second regarding the 2-11 GHz band. For the 10-66 GHz the Line-Of-Sight (LOS) propagation is a necessity due to the high working frequencies: in this case a single-carrier modulation is used, realizing the so-called WirelessMAN-SC. The access protocol can be either Time Division Multiplex (TDMA) or Frequency Division Multiplex (FDMA) (with both Time Division Duplex (TDD) and Frequency Division Duplex (FDD)). The 2-11 GHz band is designed for Non-Line-Of-Sight (NLOS) operations and can be divided in:

- WirelessMAN-OFDM: using an Orthogonal Frequency Division Multiplexing with a 256-point FFT.
- WirelessMAN-OFDMA: using an Orthogonal Frequency Division Multiple Access with a 2048-point FFT.

In Tab. 1 the physical parameters considered for our purposes are shown: WiMAX uses channel bandwidth of 20, 25 or 28 MHz, a Nyquist square-root raised-cosine pulse shaping with a roll-off factor of 0.25, the possibility of QPSK, 16QAM and 64QAM modulations and frame durations of 0.5, 1 and 2 ms.

The MAC layer can be divided in three parts: Convergence Sublayer (CS), for the specifications of ATM and IP networks interfaces; Common Part Sublayer (CPS), that is the kernel of all the MAC characteristics; Privacy Sublayer (PS), that manages the authentication and crypting procedures.

For the purposes of this paper we concentrate on the Common Part Sublayer specifications and we can recognize some important sectors in the entire CPS. First of all, the different ways of managing Packet Data Units from the upper layers are described: specifically, a data unit, arriving from the upper layers of an ATM or IP network protocol, can be fragmented, packed and concatenated. After that the PHY support is defined: TDD or FDD (half and full) can be adopted and the framing structure is determined, including the synchronization between uplink (UL) and downlink (DL) sub-frames.

In this paper the TDD framing is used as represented in Fig. 1: the UL and the DL sub-frame are firmly separated by an adaptive threshold, and every sub-frame is divided in a finite number of Physical Slot (PS); the DL sub-frame comes first because it contains the bandwidth requests and the transmission informations directed from MSs to the BS, which has to schedule the UL resources between all the users.

In the following, the scheduling services are described: it’s important to underline that IEEE 802.16 does not specify any scheduling policy, neither in UL nor in DL. WiMAX only divides traffics in four service classes:

1) **UGS** (Unsolicited Grant Services): Constant bit-rate (CBR) and CBR-like flows like VoIP. This kind of applications needs constant bandwidth allocation, without request.
2) **rtPS** (real-time Polling Services): real-time variable bit-
rate (VBR) and VBR-like flows such as MPEG video or teleconferences. These applications need minimum bandwidth granted and have to request transmission resources by polling (contention and piggybacking are not allowed).

3) **nrtPS** (non-real-time Polling Services): non-real-time flows like bandwidth intense FTP. For these traffic polling bandwidth requests are allowed when minimum bandwidth requirements are needed; otherwise contention and piggybacking are allowed.

4) **BES** (Best Effort Services): best effort flows like HTTP, e-mail and short length FTP. These applications can make bandwidth request only with contention and no minimum resources allocation is granted.

As said before, each service type has its own requirements of bandwidth and bit-rate and to each flows different way of requesting band are associated. For these reasons the IEEE 802.16 determines how to make bandwidth request: every SS has to use a specific data unit, composed only of the header (that has a particular structure), and the request can be forwarded using polling, contention and piggybacking. The BS, once received all the requests from SSs, applies an *ad hoc* scheduling algorithm in order to establish how many resources have to be allocated to a defined user terminal. Finally, dealing with SS without addictive explanations could be misleading: in fact the IEEE 802.16 can manage traffic either from single device user or terminals grouped in a LAN; we refer to both of them, considering a diverse metropolitan scenario.

### III. Proposed Scheduling Approach

The IEEE 802.16 standard represent one of the best way to realize a WMN, but in this realization QoS requirements have to be considered and granted. As said before, the WiMAX standard comprehend four traffic classes, but it doesn’t introduce a specific scheduling politic: in order to manage every service flow properly, granting an appropriate QoS, we can proposed a scheduling scheme which assigns an *ad hoc* scheduling algorithm to each flow, enabling a proper control on QoS and resources allocation. The aim of this study is to propose a scheduling algorithm which manages the resource allocation and grants an appropriate QoS per connection. In Fig. 2 the proposed scheduling scheme is shown.

During working time, inside the BS a scheduling algorithm is enabled, defining a strict semi-preemptive priority between the four traffic classes, assigning the highest to UGS and the lowest to BES. Every Connection IDentifier (CID) is grouped inside a flow type and, after strict priority evaluation, each traffic group is scheduled separately: the UGS one, like VoIP flow, is scheduled by a Packet Based Round Robin (PBRR - which grants fixed-bandwidth resource allocation with periodical frequency); the rtPS is managed with an Earliest Deadline First (EDF [8]) algorithm, which is particularly reliable for real-time traffic, sensible to deadline; nrtPS and BES are scheduled with the same algorithm, i.e. Weighted Fair Queueing (WFQ - one of the PFQ disciplines derived from GPS [9]), assigning more privileges to nrtPS with priority. This is the global UL scheduling policy for IEEE 802.16 standard; in order to evaluate the reliability of this algorithm, we have chosen two traffic classes: UGS, considering some VoIP flows inside the network, and BES, adopting Web and internet basic applications (in fact these services are very significant in this first stage of WiMAX evolution).

The computer simulations were made inside the NePSing simulating environment [4] and the realized scenario was composed of central BS inside a covering cell of the WMAN to which a variable number of terminals refer: with terminals we mean both single user devices (like PDA or laptop) and WiFi/Ethernet island, because IEEE 802.16 refers not to the device but to the CID of every flow (for example, a user with a laptop could have a VoIP conference while surfing the Internet); the central BS have to schedule available resources between all CIDs.

Three fundamental procedures have been implemented to guarantee WiMAX emulation: contention request algorithm for BES connections, data units transmission and dynamic resources allocation. The first one has been realized by an ALOHA-like protocol: in every UL frame the contention zone is delimited establishing a static threshold (see Fig. 3) and in
this space BES CID can transmit bandwidth request messages choosing one of the available slot with a uniform probability; the BS is the only one that can detect collisions because the system is broadcast, so, after collision detection and scheduling operation, the BS send ACK (with the allocated bandwidth) or NACK messages to every BES CID. This procedure interests only BES CIDs because the VoIP ones have granted bandwidth allocated without request. After contention, BES CIDs transmit a number of packets based on the real allocation, made by the BS from the request messages of the previous frame; VoIP CIDs, instead, transmit a group of data units basing on the fixed bandwidth allocation established in the connection-setup time. Data transmission procedures comprehend fragmentation, packing and concatenation (as mandatory in the IEEE 802.16 standard), optimizing efficiency. Dynamic resource allocation is the central aspect of the proposed approach and it follows the scheduling algorithm of Fig. 2. BES CIDs are scheduled using a WFQ-like policy: to the $i$-th connection, with $Q_i$ bytes in the transmission queue, the $W_i$ weight is assigned by the next formula:

$$W_i = \frac{Q_i}{\sum_{j=0}^{N-1} Q_j}$$

where $N$ is the contention-winners’ number.

VoIP CIDs, instead, are scheduled using the PBRR policy: once defined the per-frame needed bytes, to reach a specific data rate the resource allocation proceeds assigning fixed-length transmission opportunity to each connection, considering frame duration, VoIP source codec and data units’ time delay.

IV. NUMERICAL RESULTS

In this section numerical results of the simulations are shown: analyzing these results and comparing them each other will be useful to determine how the proposed resource allocation technique works. As said before, in order to realize a protocol-faithful simulation without loosing implementation simplicity, two traffic models have been chosen: VoIP and Best-Effort. Both of them are taken from the UTRAN [10] specifications for 3GPP mobile networks: for the VoIP model a two state Markovian chain has been chosen, with exponential interarrival times with an average value of 3 seconds and with a voice codec with 66 Kbps maximum bit-rate; the Best-Effort one, instead, is based on the joint of a Poisson statistic for the message interarrival times and a truncated Pareto for the message length, having a p.d.f.:

$$f(x) = \begin{cases} \frac{\alpha x^{\alpha-1}}{m^{\alpha}}, & k \leq x < m \\ \frac{\alpha}{m^{\alpha}}, & x = m \\ 0, & x < k, \ x > m \end{cases}$$

where $\alpha = 1.1$ is the shape parameter, $k = 81.5$ bytes is the location parameter (corresponding to the minimum message length) and $m = 66666$ byte is the maximum message length (obtaining a 480 bytes packet average length).

The IEEE 802.16 numerical parameters used to achieve significant results without introducing extreme working situations or empty load scenarios are listed below:

- TDD frame structure
- 1 ms frame duration
- 25 MHz channel
- 5000 physical slot per frame
- 16-QAM modulation scheme
- 80 Mbps bit-rate
- 4 modulation symbols per physical slot

First of all the contention protocol has been performed: the scenario was composed of a fixed number of VoIP connections, with a granted bandwidth of 64 Kbps, and a variable number of Best-Effort flows. In order to test the ALOHA-like contention algorithm several simulations were made, increasing or decreasing the contention zone extension and varying the sources data-rate. In Fig. 4(a) the number of byte in Best-Effort transmitting queues is represented, for 50 BES and 20 VoIP connections with 64 Kbps guaranteed: we notice that increasing sources data-rates causes a growth of bytes in the transmitting queues, till the number of contention slots is less than half BES connection. After that, a bigger contention zone causes lower performances; this could be explained underlining that more contentions slot are used more transmission resources are stolen to BES connections, once established VoIP allocations. Observing Fig. 4(b) it’s possible to confirm that 25 contention slots are the best solution: a few slots mean an high number of collision, but after 25 slots the algorithm does not perform. In Fig. 4(c) it’s shown how many BES requests reach BS scheduler. These numerical results could let us say that the contention algorithm works best with a number of contention slots equal to the half of BES connections number. Fig. 5 confirms our assumptions. As shown in the figures above, IEEE 802.16 standard allows a very high sources data-rate. In several cases 700 Kbps are reached, but this results are not strange if we think that we are working with BES terminals like WiFi-Ethernet isle and that we only reach those speeds with a light-weight system.

In all the presented simulations 64 Kbps guaranteed are used for VoIP connections: this could be wrong if packet delay per connection would be too high. In Fig. 6, instead, is shown that 64 Kbps is enough for system reliability: usually 200-250 ms delay is taken as the maximum packet delay for an intelligible signal after information reconstruction and the proposed scheduling algorithm grants lower delays.

In order to evaluate how VoIP flows influence BES traffics, simulations with a variable number of VoIP stations have been made (we have chosen 50 BES with 25 contention slots corresponding to the best performances shown in Fig. 4). In Fig. 7 simulations results are shown: the constant bandwidth allocation for VoIP SSs is not too heavy for the system’s performance, as the average queue length and the collision percentage don’t vary when we change VoIP stations number, also working with very high speed sources.

Finally, considering UGS reliability constraints, we can say
Fig. 4. 50 BES performances (20 VoIP with 64 Kbps).

Fig. 5. 20 and 75 BES performances (20 VoIP with 64 Kbps).
that PBRR scheduling algorithm grant both the respect of the transmitting parameter for a VoIP flows and the maximum fairness in managing resources allocation, allowing complete flows separation.

V. CONCLUSIONS

In this paper the evaluation of a QoS managing algorithm for IEEE 802.16 standard is presented. The considered scenario is composed of a variable number of Subscriber Station inside a network’s cell with a central Base Station: every terminal could represent either a single user device (like a laptop or a PDA) or a WiFi-Ethernet isle. The Base Station has to schedule uplink traffic, organizing transmission resources between all the connections composing the network: we have chose a contention ALOHA-like algorithm for Best-Effort connections requests and a constant bandwidth allocation without request for the VoIP ones.

From the simulations we notice that IEEE 802.16 allow very high speed flows for a big number of Best-Effort Subscriber Stations, without losing performances. The proposed scheduling algorithm is reliable for both service flows, but, while it suites perfectly VoIP QoS requirements (packet delay), it could be a little expensive for BES resources: in fact, in several cases the WFQ-like policy may cause a bytes loss.

REFERENCES