

A two Layer Guaranteed and Sustained Rate based Scheduler for IEEE 802.16-2009 based WiMAX Networks

Volker Richter, Rico Radeke, and Ralf Lehnert

Technische Universität Dresden
Dresden, Mommsenstrasse 13 01062
richter@ifn.et.tu-dresden.de,
rico.radeke|ralf.lehnert@tu-dresden.de
www.ifn.et.tu-dresden.de/tk

Abstract. Worldwide Interoperability for Microwave Access (WiMAX) based on the IEEE 802.16-2009 standard is a serious competitor of Long Term Evolution (LTE) for cellular networks providing high data rates to mobile users. To support various multimedia applications the Quality of Service (QoS) support is a crucial function. While the IEEE 802.16-2009 standard defines five service classes to support the QoS requirements of various multimedia applications, the algorithms and especially the scheduling algorithms to enforce these requirements have not been specified for vendor differentiation. We propose a scheduling algorithm allocating resources according to the guaranteed and sustained rates of each connection, in contrast to common two layer scheduling approaches differentiating between service classes. Simulation results show that our scheduler is able to fulfill the QoS requirements of connections defined by the IEEE 802.16 standard.

Keywords: WiMAX; IEEE 802.16; QoS; Scheduling; Traffic Shaping

1 Introduction

The IEEE 802.16 standard family allows different vendors to build WiMAX systems compatible with each other. Since the introduction of the standard amendment IEEE 802.16e [1] the physical layer WirelessMAN-OFDMA allows the deployment of cellular networks supporting mobile users with high data rates. Therefore, WiMAX is a serious competitor to Long Term Evolution (LTE), if no existing 3GPP infrastructure is present.

On top of different variants of the physical layer, the standard defines a mostly common Medium Access Control (MAC) layer, which is responsible for functions such as connection management and resource allocation. One substantial property of this MAC layer is the All-Internet Protocol approach. All services like Voice over Internet Protocol (VoIP), video streaming or web surfing are provided through Internet Protocol (IP) connections. In order to ensure the functionality

of diverse applications, the IEEE 802.16-2009 standard provides a QoS concept consisting of various QoS parameters including five service classes.

Figure 1 illustrates the WiMAX resource allocation process according to our studies of the standard document [2]. In the first step, packets from higher layers

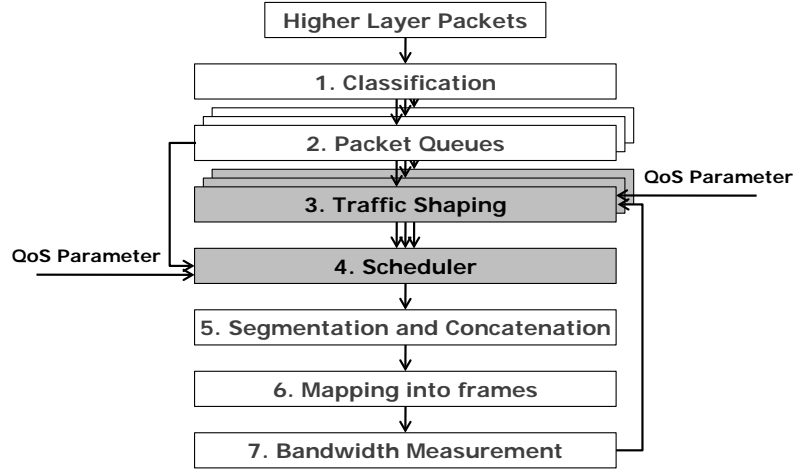


Fig. 1. WiMAX Resource Allocation

are mapped to their corresponding MAC connections based on a certain rule set. All header information such as the IP source and destination addresses and the port numbers can be used for packet classification. Afterwards, packets belonging to one connection are stored in a queue. To enforce QoS parameters as the maximum data rate, negotiated during connection establishment, a traffic shaping algorithm is applied on each connection. Our traffic shaping algorithm has been described in detail in [8] and will be summarized in section 3.1. Based on the results of the traffic shaping, the current queue size and the QoS parameters the scheduling algorithm distributes the available radio resources to all connections in the fourth process step of figure 1. Packets are segmented and concatenated to MAC Protocol Data Units (PDUs) to fit into assigned radio resources. Finally, the transmitted data volume is fed back to the traffic shaping algorithm in order to calculate the correct data rates for the next frame.

However, the IEEE 802.16-2009 standard neither defines the scheduling nor the traffic shaping algorithm to ensure the adherence to the admitted QoS parameters. Therefore, various scheduling algorithms have been proposed for WiMAX.

One class of scheduling algorithms for WiMAX are optimization approaches. For example, the authors of [5] and [6] define a objective function for the scheduling problem and solve it by applying heuristic methods. Another common approach is to divide the scheduling problem into two layers. The first layer sched-

uler distributes the available resources to groups of connections belonging to one service class. The second layer schedulers allocate these resources to specific connections. An overview of this scheduler class can be found in [4]. The author of [3] uses a Deficit Fair Priority Queue Scheduler for the first layer scheduling. For the second layer scheduling the algorithms Earliest Deadline First (EDF), Weighted Fair Queuing (WFQ) and Round Robin (RR) are applied according to the delay requirements of the service classes.

In contrast to the common two layer scheduling, our proposed scheduler does not differentiate between service classes. It only differentiates between guaranteed and sustained rates defined for each connection in the standard. Simulation results show that our scheduler is able to fulfill the QoS requirements of connections defined by the IEEE 802.16-2009 standard [2].

In the first section, the IEEE 802.16-2009 QoS concept is introduced based on our studies of the current standard document [2]. Subsequently, we describe our proposal for the WiMAX resource allocation including a summary of our traffic shaping approach, which is closely related to the proposed scheduling algorithm. The focus of this paper lies on our Two Rate based Scheduling algorithm, which will be described in detail. These algorithms have been implemented in a network simulator ns-2 based WiMAX model. Our simulation results are presented in section 4. Finally, we conclude our paper and give an outlook on future work.

2 The IEEE 802.16-2009 QoS Concept

The central element of the IEEE 802.16-2009 connection oriented MAC layer is the Service Flow (SF). To each SF belongs a QoS parameter set. When a SF is activated a transport connection is assigned to it ([2], p. 366).

As mentioned before, the standard defines five service classes, which are part of the QoS parameter set. For the Downlink (DL) transmission they are denoted as Data Delivery Services ([2], p. 422). In contrary, the service classes are named Scheduling Services for the Uplink (UL) channel. They provide different UL bandwidth grant and request mechanisms optimized to specific traffic characteristics and requirements.

The service class Unsolicited Grant Service (UGS) supports real time connections with fixed interarrival times and packet sizes. Typically VoIP without silence suppression generates such a traffic pattern. Currently, VoIP with silence suppression is used by many applications. To adapt to the different packet size in active and silence periods, the service class Extended Real-Time Variable Rate (ERT-VR) has been introduced with the IEEE 802.16e standard [1]. Another realtime service class represents Real-Time Variable Rate (RT-VR). It is designed for video transmissions with variable interarrival times and packet sizes. Non realtime service classes are Non Real-Time Variable Rate (NRT-VR) and Best Effort (BE). BE does not provide any guarantees for connections. In contrast to BE, NRT-VR includes a minimum data rate, which is necessary to avoid time outs of File Transfer Protocol (FTP) connections.

The IEEE 802.16-2009 standard defines for each service class a set of mandatory QoS parameters. The most important in terms of data rates are the Maximum Sustained Traffic Rate (MSTR) and Minimum Reserved Traffic Rate (MRTR). The MSTR defines the average maximum data rate which can be allocated for the corresponding connection. It is measured in bit per second, excluding MAC and physical overhead ([2], p. 1285). This parameter allows the network operator to limit the connection to a certain speed, often used for mobile data flat rate offers. On the other hand, the MRTR describes the average minimum data rate which must be allocated for connections. Also, it is expressed in bit per second without MAC and physical overhead ([2], p. 1285). This parameter guarantees the functionality of user applications, also in cases of a highly loaded cell. Furthermore, it is defined by the standard, that only resources should be assigned for data currently stored in the queues. Therefore, it is possible, that a connection receives less resources than the MRTR allows.

Figure 2 depicts the functionality of the MRTR and MSTR QoS parameters. In an idle system the traffic shaping and the scheduler are allowed to allocate

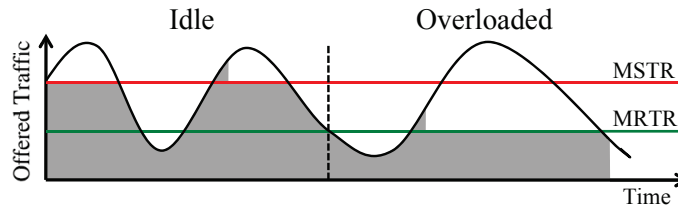


Fig. 2. System behavior in idle and overload state

resources up to the MSTR boundary. Since the MSTR is defined as an average over time, unused data volumes can be used later to exceed this limit. If the system is in an overloaded state, the resource allocation process has primarily to fulfill the guaranteed data rate. Therefore, firstly data volumes are assigned to reach the MRTR. Also, the MRTR is specified as an average over time, which permits a short overshoot of the average value.

In order to configure the time interval used for calculation of the averaged rate, the IEEE 802.16-2009 standard specifies the QoS parameter Time Base (TB) given in *ms* ([2], p. 1306). The combination of TB and MSTR or respectively MRTR defines the amount of data sending permissions, which can be shifted from idle to busy periods of a connection.

Table 1 summarizes the service classes for DL and UL including their mandatory QoS parameters. Since the service class UGS was designed for connections with a constant data rate, only the MRTR and TB parameters are present. In contrast, BE does not provide any guarantees. Therefore, the MRTR is zero or absent. All other service classes include both data rate QoS parameters to support fluctuating data rates.

Service classes DL	UGS	ERT-VR	RT-VR	NRT-VR	BE
Service classes UL	UGS	ertPS	rtPS	mrtPS	BE
MSTR		X	X	X	X
MRTR	X	X	X	X	
TB	X	X	X	X	X

Table 1. Service classes and their mandatory (X) QoS parameters

3 WiMAX Resource Allocation

In the following section, we will firstly summarize our traffic shaping algorithm, which was presented in [8]. This algorithm is applied on both the MSTR and MRTR of each connection in order to calculate the data volume necessary in the next frame to obtain these rates within the current Time Base (TB) interval. Based on these values, our scheduling algorithm distributes the radio resources. This scheduling algorithm will be presented in section 3.2.

3.1 WiMAX Traffic Shaping Algorithm

The traffic shaping will be described for the MSTR calculation denoted as R_{MS} . It is equivalent for the MRTR calculation.

The data volume S , which has to be transmitted to reach the R_{MS} within the interval T_B is:

$$S = R_{MS} \cdot T_B \quad (1)$$

This data volume has to be transmitted within all frames n belonging to one T_B .

$$n = \left\lceil \frac{T_B}{T_F} \right\rceil \quad (2)$$

Here, T_F denotes the frame duration, which is typically 5 ms [9]. To calculate the maximum schedulable data volume S'_k of the next frame k , all previous transmitted data volumes S_i within the current T_B period have to be subtracted from S . Therefore, all allocations made in frames from $k - n + 1$ to $k - 1$ have to be considered. This can be expressed as:

$$S'_k = R_{MS} \cdot T_B - \sum_{i=k-n+1}^{k-1} S_i \quad (3)$$

S'_k is the maximum data volume which can be scheduled to satisfy R_{MS} . In the next execution of the algorithm, to calculate the data volume S'_{k+1} for the following frame $k + 1$, the considered time interval T_B is shifted by one frame duration T_F . Therefore, allocations made in the frames from $k - n + 2$ to k are subtracted.

As described in section 2, the IEEE 802.16-2009 standard requires that only resources for available data in the queues shall be assigned. Therefore, a minimum function between S'_k and the current queue size S_Q of a connection is introduced resulting in maximum schedulable data volume S''_k .

$$S''_k = \min(S'_k, S_Q) \quad (4)$$

Based on the current system state and further QoS parameters such as maximum latency and jitter, the scheduler assigns radio resources to transmit the data volume S_k in the corresponding frame.

$$0 \leq S_k \leq S''_k \quad (5)$$

In conclusion, our traffic shaping algorithm fulfills the requirements of the IEEE 802.16-2009 standard in an exact manner and can be described with the following equation:

$$S_k \leq S''_k = \min \left(S_Q, R_{MS} \cdot T_B - \sum_{i=k-n+1}^{k-1} S_i \right) \quad (6)$$

As mentioned before, the traffic shaping algorithm calculates the schedulable data volumes based on MSTR and MRTR, which will be denoted as S''_{k-MS} and S''_{k-MR} in the following.

3.2 WiMAX Rate Based Scheduling

The aim of our scheduling algorithm is to primarily provide the agreed QoS parameter by giving strict priority to guaranteed MRTR based data volumes S''_{k-MR} before allocating MSTR based demands S''_{k-MS} . Secondary, a Min-Max fairness [7] based on radio resources is realized in order to balance between the two opposite goals of fairness based on user throughput and maximum cell throughput. Therefore, we equally increment the radio resource assignments of each connection with MRTR or respectively MSTR demands until the desired data volume is reached.

When designing a WiMAX DL scheduling algorithm, the WirelessMAN-OFDMA frame format has to be taken into consideration. OFDMA permits to distribute the radio resources in time and frequency domain. The time domain is counted in OFDM symbols, while the frequency domain is divided into subchannels consisting of 28 subcarriers ([2], p. 938). In the mandatory Partial Usage of Subchannels (PUSC) frame structure, the minimum allocation unit called slot comprises 2 OFDM symbols on 1 subchannel. In one slot from 6 *Byte* to 27 *Byte* can be transmitted, dependent on the used modulation and coding scheme.

Figure 3 shows a simplified flowchart of our Two Rate based Scheduling algorithm. N_F describes the number of free slots. N_C denotes the maximum number of slots which can be assigned to a connection in the current iteration of the connection list.

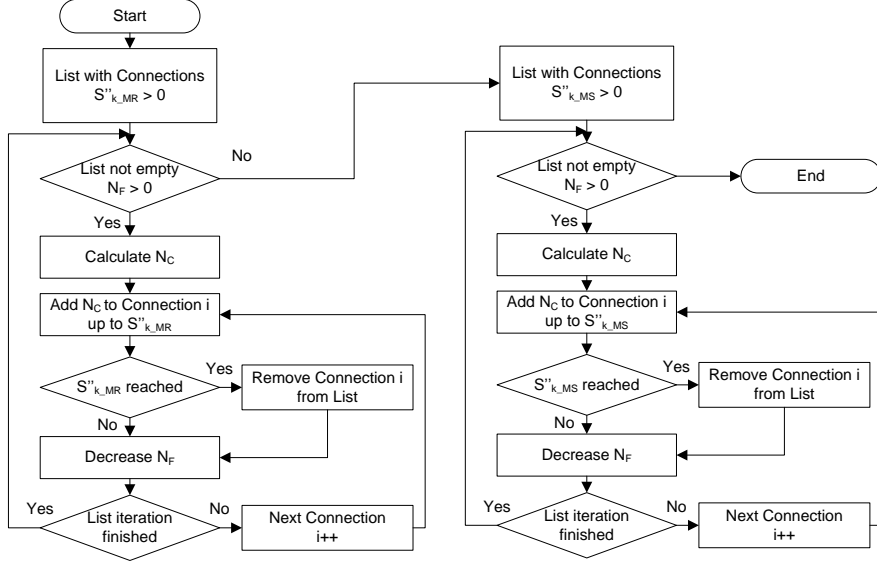


Fig. 3. Flowchart Two Rate based Scheduler

At the beginning of the scheduling algorithm, a list of all connections with $S''_{k_MR} > 0$ is created. These connection can receive allocations to fulfill their MRTR requirements. If this list is not empty and the number of free slots N_F is greater than zero, the maximum number of slots for each connection N_C is calculated according to equation 7.

$$N_C = \max \left(\left\lfloor \frac{N_F}{N_L} \right\rfloor, 1 \right) \quad (7)$$

Here, N_L denotes the number of connections in the list. If the number of connections in the list N_L is higher than the number of free slots N_F , one slot will be assigned. N_C is added to previous assignments for the corresponding connection. When the transmission capacity of all assigned slots exceeds the permitted data volume S''_{k_MR} according to the current modulation and coding scheme, the number of added slots is reduced to the necessary number of slots fulfilling S''_{k_MR} . If S''_{k_MR} has been reached the connection is removed from the list for the next iteration. Finally, the number N_F of free slots is decreased by the number of newly assigned slots and the next connection of the current list iteration is processed. After all connections of the current iteration have been served and there are still demands and free slots the next iteration of the connection list begins with an updated N_C value.

Once, all MRTR based demands are served, a list of MSTR connections with $S''_{k_MS} > 0$ is prepared and the assignment is conducted in the same manner. Because S''_{k_MS} is always greater than or equal to S''_{k_MR} , all previously assigned

slots are included in the comparison with the maximum permitted data volume S''_{k_MS} .

With this equal filling approach, we achieve our goal of a Min-Max fairness based on radio resources. Furthermore, we obtain a fair allocation in case of a high number of connections, when only one slot is assigned per iteration. We always start the assignment process with the first connection not being served in the last execution of the scheduler for the previous frame.

4 Simulation Results

The scheduling as well as the traffic shaping algorithm have been implemented in our WiMAX simulation environment. It consists of the network simulator ns-2 and an improved version of the WiMAX Add-On for ns-2 [10], which was originally developed by the Application Working Group of the WiMAX Forum.

To investigate the behavior of our algorithm, we built up a scenario with two Mobile Stations (MSs) placed next to a Base Station (BS). Therefore, the highest modulation and coding scheme QAM 64 3/4 was used without transmission errors. We used typical system parameters deduced from [9]. A WirelessMAN-OFDMA TDD system was simulated with 10 *Mhz* channel bandwidth and a cyclic prefix of 1/8. The applied 5 *ms* long frame consisted of 47 OFDM symbols, after subtracting the gap periods [9]. The DL subframe comprises 27 OFDM symbols including preamble.

A Constant Bit Rate (CBR) foreground traffic of different service classes and data rates was sent to the first Mobile Station (MS1). For the second Mobile Station (MS2) a 25 *Mbit/s* CBR background traffic was offered, which is higher than the overall link capacity. Both traffic sources generated packets with a size of 576 *Byte*. The MSTR and MRTR were set to 14 *MBit/s* and 10 *MBit/s* in order to stay below the system capacity of approximately 17 *Mbit/s*. To focus on the scheduling algorithm and not on the traffic shaping algorithm, we configured the TB with 20 *ms* or respectively 4 frame durations. Therefore, the influence of the averaging of the data rates during the traffic shaping remains low.

We implemented six test cases, where the offered foreground traffic alternates in all possible relations around the MSTR and MRTR boundaries. Figure 4 shows the achieved goodput of Mobile Station (MS) 1 over time without traffic shaping and background traffic. Therefore, the goodput corresponds to the offered traffic. This sequence has been also used in all following simulations.

Figure 5 gives the simulation result of BE foreground traffic sent to MS1. It is represented by the dark gray curve. Because the service class BE does not provide any data rate guarantees, the MRTR limit is absent. It can be seen, that background traffic always uses the remaining resources from the foreground traffic. The system goodput depicted with the black line reaches always 17 *MBit/s*. If only MS2 is active, the system goodput is slightly higher, due to the less signaling overhead for one station. This proves, that the system capacity is always used by our scheduler. The equilibrium between foreground and background traffic

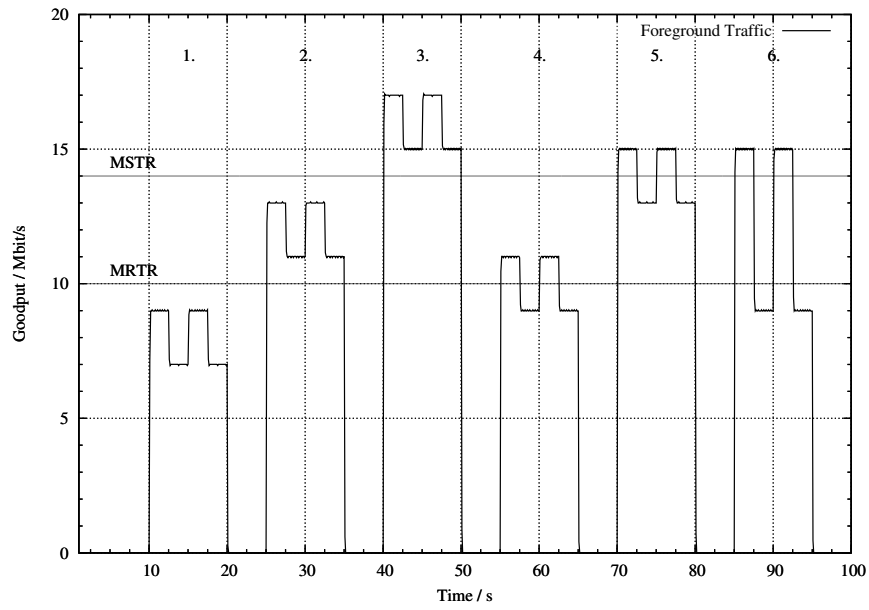


Fig. 4. Offered Foreground Traffic

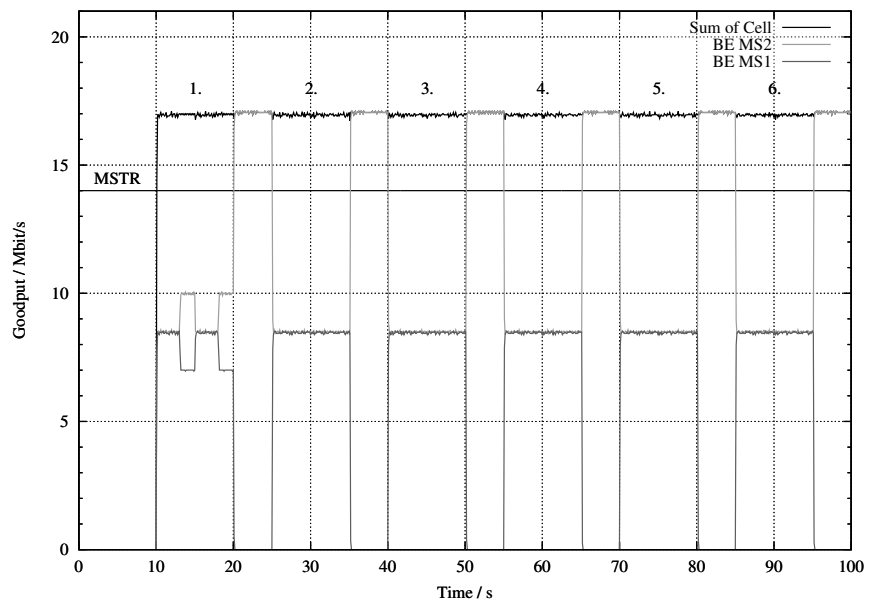


Fig. 5. Goodput of two flows using Service Class Best Effort (BE)

at a goodput of 8.5 MBit/s shows, that the scheduler distributes the resources fairly.

In the second simulation result depicted in figure 6 the service class UGS was chosen for the foreground traffic. Therefore, the MRTR is 14 MBit/s and equal to the MSTR. Similar to figure 5, all resources are used by both data

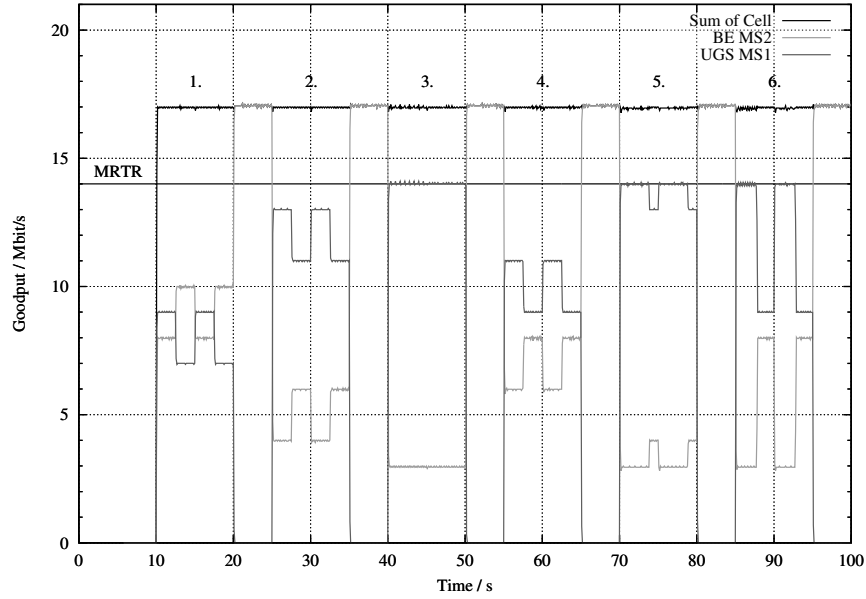


Fig. 6. Goodput of MS1 using Service Class Unsolicited Grant Service (UGS)

flows. Therefore, the system goodput is almost constant. When the offered UGS foreground traffic exceeds the MRTR threshold of 14 MBit/s shown in the test cases 3, 5 and 6, the traffic shaping permits only the transmission up to the MRTR. In contrast, the offered UGS traffic is entirely transmitted, when it is below this boundary. Especially in test case 5, it can be seen, that the duration of the high data period is longer than compared to the offered traffic in figure 4. This effect results from packets stored in the queue, which are transmitted after 75 s simulation time.

The realtime service class RT-VR is designed for video streaming using both QoS parameters MRTR and MSTR. Therefore, the influence of both boundaries can be studied in the simulation result presented in figure 7. The figure shows, that the offered RT-VR traffic is not influenced up to the MRTR limit. In the test cases 3, 5 and 6, when the foreground traffic is high, an equilibrium of 13.5 MBit/s can be observed. Due to the strict priority of MRTR assignments before MSTR allocations, a data volume of 10 MBit/s is given to the RT-VR data flow. The remaining resource of 7 MBit/s is fairly distributed between

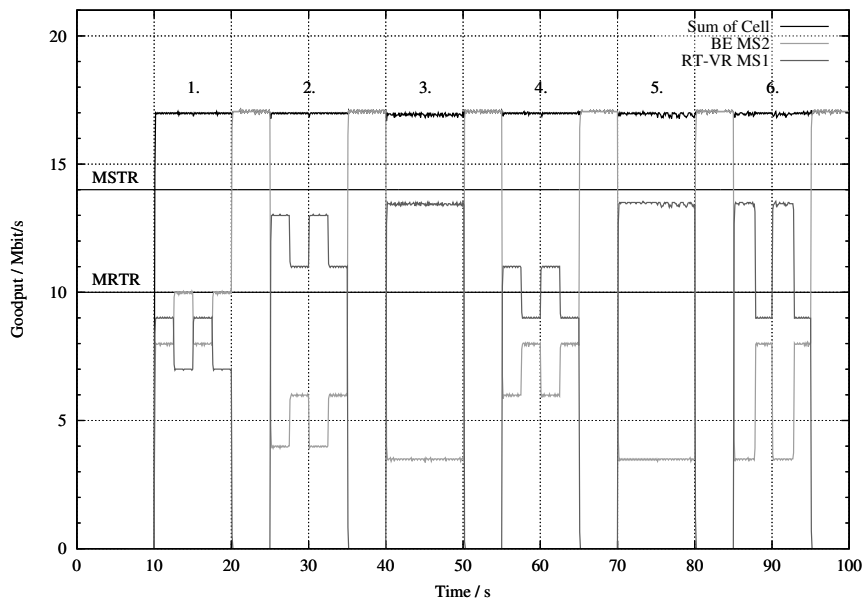


Fig. 7. Goodput of MS1 using Service Class Real-Time Variable Rate (RT-VR)

both flows resulting in a goodput of 13.5 MBit/s of the foreground traffic and respectively 3.5 MBit/s of the background traffic.

5 Conclusion and Future Work

In this paper we proposed a new Two Rate based Scheduling algorithm, which does not differentiate between service classes, but between guaranteed and sustained data rates. In combination with our traffic shaping algorithm, this resource allocation approach corresponds to the requirements of the IEEE 802.16-2009 standard. The simulation results prove that our algorithm is able to provide different service classes with their required QoS constraints. Future work includes the investigation of the algorithms under more complex scenarios and more realistic traffic models. Special interest will be given to the influence of our algorithm on other metrics such as delay and jitter. Furthermore, the algorithms will be compared with widely used scheduling algorithms applied in mobile networks to determine their practical relevance.

Acknowledgments We would like to thank our project partner Detecon Consulting, Germany for sponsoring this work. Special thanks are devoted to Dr. Petry and Dr. Knospe from Detecon Bonn for their great support.

References

1. IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands and Corrigendum 1. IEEE Std 802.16e-2005 and IEEE Std 802.16-2004/Cor 1-2005 (Amendment and Corrigendum to IEEE Std 802.16-2004) (2006)
2. IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Broadband Wireless Access Systems. IEEE Std 802.16-2009 (Revision of IEEE Std 802.16-2004) (2009)
3. Chen, J., Jiao, W., Wang, H.: A Service Flow Management Strategy for IEEE 802.16 Broadband Wireless Access Systems in TDD Mode. IEEE International Conference on Communications (2005)
4. Dhrona, P., Abu Ali, N., Hassanein, H.: A performance study of scheduling algorithms in Point-to-Multipoint WiMAX networks. In: Local Computer Networks, 2008. LCN 2008. 33rd IEEE Conference on (2008)
5. Ermini, D., Bonuccelli, M.A.: Realtime Traffic Scheduling for Missed Deadline Minimization in Point-to-Multipoint OFDMA Systems. In: PM2HW2N '10 Proceedings of the 5th ACM workshop on Performance monitoring and measurement of heterogeneous wireless and wired networks (2010)
6. J.Song, Li, J., Li, C.: A Cross-Layer WiMAX Scheduling Algorithm Based on Genetic Algorithm. In: Communication Networks and Services Research Conference, 2009. CNSR '09. Seventh Annual (2009)
7. Le Boudec, I.: Rate adaptation, Congestion Control and Fairness: A Tutorial. Ecole Polytechnique Federale de Lausanne (December 2006)
8. Richter, V., Tuerk, S.: New IEEE 802.16-2009 compliant Traffic Shaping Algorithms for WiMAX Networks. In: EUNICE 2011 Workshop on Energy-aware communications (September 2011)
9. WiMAX Forum: WiMAX System Evaluation Methodology Version 2.1 (Juli 2008)
10. WiMAX Forum - Application Working Group: NS-2 MAC + PHY Add-On for WiMAX (IEEE 802.16) (May 2009)