Publication P7

Jani Lakkakorpi and Alexander Sayenko. 2009. Measurement-based connection admission control methods for real-time services in IEEE 802.16e. In: Proceedings of the Second International Conference on Communication Theory, Reliability, and Quality of Service (CTRQ 2009). Colmar, France. 20-25 July 2009, pages 37-41.

2009 Second International Conference on Communication Theory, Reliability, and Quality of Service

Measurement-Based Connection Admission Control Methods for Real-Time Services in IEEE 802.16e

Jani Lakkakorpi
Department of Communications and Networking
Helsinki University of Technology
Espoo, Finland
jani.lakkakorpi@iki.fi

Alexander Sayenko
Research, Technology & Platforms
Nokia Siemens Networks
Espoo, Finland
alexander.sayenko@nsn.com

***Abstract*—In this paper, we present two simple-to-implement measurement-based admission control methods for real-time services in IEEE 802.16e networks. The first method simply utilizes the averaged number of free slots as input in admission decisions while the second method is more advanced and it tunes the admission thresholds according to current traffic load. Our simulations show that the proposed methods lead into more efficient use of scarce radio resources – without sacrificing the Quality of Service (QoS) requirements – than using parameter-based, i.e., static, connection admission control.**

Keywords-connection admission control, IEEE 802.16, ns-2, quality of service, WiMAX

I. INTRODUCTION

IEEE 802.16d/e, also known as WiMAX (Worldwide Interoperability for Microwave Access) is an IEEE standard for wireless broadband access network [1, 2]. Its main advantages are long range and extensive support for Quality of Service (QoS) at the MAC level. The standard defines two operational modes: mesh and point-to-multipoint (PMP). In the mesh mode, subscriber stations (SS) can communicate to each other and to the base station (BS) whereas in the PMP mode, the SSs always communicate through the BS. It is anticipated that providers will use the PMP mode to connect their customers to the Internet. Thus, the provider can control the environment to ensure the QoS requirements of its customers.

At the WiMAX BS, all downlink (DL) connections have dedicated buffers and resources are allocated per connection. (Even though there can be multiple connections per SS.) In uplink (UL) direction, however, the BS grants slots per SS (GPSS) and not per connection (GPC). The SS then decides how the slots are shared between different connections. The effective air interface bandwidth that a connection gets may vary a lot because there are no dedicated radio channels.

IEEE 802.16e has five different QoS classes. In unsolicited grant service (UGS), the BS allocates fixed-size grants periodically; UGS connections do not send any bandwidth requests. In real-time polling service (rtPS), the BS periodically polls the SS by granting one slot for sending a bandwidth request, while the goal of extended real-time polling service (ertPS) is to combine the advantages of UGS and rtPS. In ertPS, the BS continues granting the same amount of bandwidth (by default, the size of this allocation corresponds to maximum sustained traffic rate of the connection) until the ertPS connection explicitly requests a change in polling size. Extended piggyback request field of the grant management subheader can be used for this purpose. If the bandwidth request size is zero, the BS may provide allocations for bandwidth request header only or nothing at all. In the latter case contention request opportunities may be used. Non-real time polling service (nrtPS) is similar to rtPS except that connections are polled less frequently and they can also use contention request opportunities. Best Effort (BE) connections are never polled and they can receive resources only through contention.

IEEE 802.16 standard [1, 2] does not specify any connection admission control (CAC) mechanisms. However, CAC is needed at least for all real-time (i.e., UGS, ertPS and rtPS) connections – otherwise, we cannot guarantee any delay bounds or packet loss rates for these connections.

Two CAC schemes for orthogonal frequency-division multiplexing (OFDM) wireless networks are studied in [3]. The first scheme sets a threshold to limit the number of ongoing connections, and new connections are admitted as long as the total number of connections (including the incoming one) does not exceed the threshold. The second scheme admits a connection with a certain probability based on the queue status.

Another new CAC scheme, called Quadra-Threshold Bandwidth Reservation (QTBR), is proposed in [4]. In QTBR, different threshold values are used for different service classes. The threshold values are determined by the number of calls of the corresponding service class in the system and the number of free channels.

Our approach to the connection admission control problem in IEEE 802.16e is somewhat similar to the schemes proposed in [3] and [4]. However, we have a lot more pragmatic viewpoint; the proposed techniques take into account all the details of a real IEEE 802.16e system. This paper studies connection admission control of real-time flows in a WiMAX network. For these purposes, we run different simulation scenarios and apply different CAC algorithms. The first algorithm utilizes the averaged number of free slots as input in admission decisions while the second method is more advanced and it tunes the admission thresholds according to current traffic load.

978-0-7695-3696-5/09 $25.00 © 2009 IEEE
DOI 10.1109/CTRQ.2009.9

The rest of this paper is organized as follows: section II presents the proposed algorithms, sections III and IV present our simulator and the simulation results, respectively, while section V concludes the paper.

II. CONNECTION ADMISSION CONTROL METHODS

As in the case of other wireless technologies, it is likely that the radio interface [1, 2] will be the biggest system bottleneck with WiMAX, too. Thus, the amount of real-time connections[1] has to be controlled in order to guarantee the QoS. This can be done at the BS, for example, by monitoring the DL queuing delays, virtual UL queue sizes and the number of free (DL and UL) slots for real-time traffic. However, it seems that only the number of free slots is a reasonable choice for UL admission control as the virtual (bandwidth request based) queue sizes may not always be that accurate.

A. Measurement-Based Admission Control (MBAC): Monitoring the Number of Free Slots

In each frame *i*, when all real-time connections have been served, we check the number of remaining DL and UL slots for real-time connections ($freeSlots_i$) and update their exponentially weighted moving averages ($freeSlotsAv_i$). These averages are used in admission control and they are compared to our "safety margin" (e.g., 10 slots). If the averaged number of free DL/UL slots for real-time traffic is above the safety margin, we can admit the connection. w_S is a configurable averaging weight that determines how fast the average changes over time

$$freeSlotsAv_i = (1 - w_S) * freeSlotsAv_{i-1} + w_S * freeSlots_i \,. \quad (1)$$

B. Measurement-Aided Admission Control (MAAC): Adjusting the Limits Based on Measurements

Since using the aforementioned averaged number of free slots for real-time traffic as such offers no protection against connections arriving in large batches, we can choose a more conservative approach instead and exercise bookkeeping with dynamically updated reservation limits for DL and UL traffic. A similar method has been presented in [5].

Whenever a connection arrives, we check if the sum of currently reserved DL/UL bandwidth and the MRTR (Minimum Reserved Traffic Rate) of the connection is below the corresponding limit. If this is the case, the connection is admitted and the MRTR is added to the reserved DL/UL bandwidth. Naturally, the MRTR is subtracted from the reserved bandwidth when the connection is torn down.

```
if (freeSlotsAv > highTh)&&(limit < maxBw)
     limit := limit + increment
if (freeSlotsAv < lowTh)
     limit := limit * coefficient
     if (limit < reservedBw)
             limit := reservedBw
```

Figure 1. Reservation limit updating algorithm for DL and UL.

[1] As explained later, the amount of active SSs has to be controlled, too.

The reservation limits are updated (additive increase with parameter *increment*, multiplicative decrease with parameter *coefficient*) periodically and they are based on the averaged number of free slots for real-time traffic. Fig. 1 illustrates the updating algorithm. If the averaged number of free slots is larger than *highTh*, we adjust *limit* upwards. Similarly, if the averaged number of free slots is smaller than *lowTh*, we adjust *limit* downwards. We cannot set *limit* higher than *maxBw* or lower than currently reserved bandwidth, *reservedBw*.

III. IEEE 802.16E NETWORK SIMULATION MODEL

The basic implementation of our IEEE 802.16e module is described in [6]. The module includes the following features: OFDM and orthogonal frequency-division multiple access (OFDMa) PHY levels, hybrid automatic repeat request (HARQ), transport and management connections, fragmentation, packing, ranging and bandwidth request contention periods, code division multiple access (CDMA) codes for ranging and bandwidth requests, support for the most important MAC level signaling messages and the ARQ mechanism that allows retransmitting dropped PDUs. Additionally, the module includes several different BS schedulers and it has a simple, trace-based model for link adaptation. These features are described in more detail in the following sections.

A. MCS, Link Adaptation and Errors

Modulation and coding scheme (MCS) defines how many bits can be sent in a single slot. The BS can dynamically change both the DL and UL MCS of an SS. Link adaptation is modeled with signal-to-noise ratio (SNR) traces that are obtained from our system simulations, where the focus has been on the lower protocol layers.

B. Base Station Scheduler

The BS scheduler grants slots for the SSs according to the QoS parameters and bandwidth request sizes of the individual connections. Uplink virtual queue sizes are updated whenever slots are granted and every time a bandwidth request arrives. For DL connections, we use the BS queue sizes and the QoS parameters. In our scheduler, slots are assigned in deficit round-robin [7] fashion. Quantum size is a configuration parameter (default: 17 slots for all connections); a bigger quantum decreases the map overhead as we then serve fewer connections per frame.

We have implemented support for three IEEE 802.16e QoS classes: ertPS, rtPS and BE. ertPS and rtPS connections are served before BE connections; they are assigned slots until all ertPS and rtPS requests have been satisfied or until there are no more slots left for this traffic. We do not let the ertPS connections participate in contention, because that might introduce large medium access delays. In order to avoid the starvation of BE connections, we can reserve a portion of all slots exclusively for these connections. Connection admission control should take care that there are always enough slots for real-time connections. Moreover, rate limiters are used at the BS to enforce the MRTR of real-time connections; excess real-time traffic gets BE treatment.

However, before any connection can be granted slots, we have to serve the management traffic in every frame, i.e., we need to grant slots for the UL-MAP, DL-MAP, channel quality indicator channel (CQICH) reports, HARQ acknowledgements, HARQ retransmissions and CDMA uplink allocations[2]. Since all SSs contribute to this overhead, admission control for real-time connections alone is not sufficient but we have to limit the number of active SSs, too. In the case of MBAC, this is rather simple: the arrival of a new SS is treated in a similar fashion as the arrival of a new real-time connection from an SS that is already registered to the BS. If the averaged number of (UL or DL) slots is too low, the new SS is rejected. In the case of MAAC, however, we need to come up with a suitable "MRTR" for the SSs. How much resources are reserved for control traffic of a single SS should depend mostly on the CQICH report interval.

TABLE I. IEEE 802.16E RELATED SIMULATION PARAMETERS

Parameter	Value
PHY	OFDMa
Bandwidth	10 MHz
FFT size	1024
Cyclic prefix length	1/8
TTG (Transmit-receive Transition Gap)	296 PS
RTG (Receive-transmit Transition Gap)	168 PS
Duplexing mode	TDD
Frame length	5 ms
DL/UL ratio	35/12 OFDM symbols
DL/UL permutation zone	FUSC/PUSC
Channel report type and interval	CQICH, 20 ms
MAP MCS	QPSK-1/2, REP 2
Compressed MAP	Yes
Number of ranging opportunities	1
Ranging backoff start/end	0/15
Number of request opportunities	3
Request backoff start/end	3/15
CDMA codes for ranging and bw requests	64/192
HARQ (CC)	For VoIP connections only
Number of HARQ channels	16
HARQ buffer size	2048 B per channel
HARQ shared buffer	Yes
Max. number of HARQ retransmissions	4
HARQ ACK delay	1 frame
PDU SN	With HARQ (no ARQ)
Fragmentation/Packing	Yes/Yes
Maximum MAC PDU size	100 bytes
ARQ	For FTP connections only
ARQ feedback types	All
ARQ block size / window size	64 bytes / 1024
ARQ block rearrangement	No
ARQ feedback frequency	5 ms
ARQ retry timer	50 ms
ARQ block lifetime	1500 ms
ARQ rx purge timeout	2000 ms
MRTR for VoIP connections	11800 bps
Max. SS/BS queuing delay for VoIP SDU	150 ms

[2] This is not something explicitly required by the standard [1, 2] but rather something that makes sense.

If ARQ is enabled for a connection, we apply the following connection-internal scheduling order: 1) ARQ feedback messages, 2) retransmissions and 3) all other PDUs [8]. We simulate only cases with one DL transport, one UL transport and one management connection per SS.

IV. PERFORMANCE EVALUATION

We use a modified version of the ns-2 simulator [9]. The WiMAX related modifications have been described in the preceding section. Six simulations are run in each case in order to obtain 95% confidence intervals. Simulation time is always 200 seconds. One-way core network delay between a server and the BS is set to 31 ms, using a few links with latencies ranging from 1 ms to 10 ms. This is mainly done in order to have realistic round-trip times for TCP connections, and thus realistic throughput. The only bottleneck in our system is the air interface. The most important WiMAX network parameters are listed in Table I.

We simulate the following traffic mix: a variable number of VoIP connections and 10 file downloading connections. New VoIP connections arrive to the system, according to Poisson process, with an intensity of 6.7 connections per second. However, 200 first connections arrive with an intensity of 20 connections per second and without admission control. VoIP connection duration is exponentially distributed with a mean of 60 seconds, while the file downloading connections are active during the whole simulation run.

Our VoIP traffic source is a simple On-Off Markov model. On and Off state durations are exponentially distributed with mean durations of 1.2 s and 0.8 s, correspondingly. 24 bytes of payload is sent every 30 ms during the active periods. Altogether, RTP, UDP and IP add 40 bytes of overhead, which results in a total packet size of 64 bytes. Packet header compression (from 40 bytes to 4 bytes) is applied at the BS and the SS. VoIP connections are given ertPS or rtPS treatment. Our file downloading user model is a simple FTP model, where a single 250 kB file is downloaded over and over again. Time between two downloads is uniformly distributed between 1 and 5 seconds. A single NewReno TCP connection is utilized. File downloading traffic is given BE treatment.

Depending on the simulated connection admission control method, a new VoIP connection is admitted to the network only if:

1. Average number of free DL/UL real-time slots is bigger than 25, 20, 15 or 10 (MBAC, averaging weight, w_S = 0.001).
2. The number of VoIP connections is less than 140, 125, 110 or 95. This is an example of static parameter-based admission control (PBAC).
3. The dynamic reservation limits allow a new connection to be admitted to the system. Here we use MAAC with the following parameters: w_S = 0.001, *highTh*/*lowTh* = 27/23, 22/18, 17/13 or 12/8, *increment* = 3 kbps, *coefficient* = 0.9, *maxBw* = 3 Mbps. Limit update frequency is set to 100 ms.

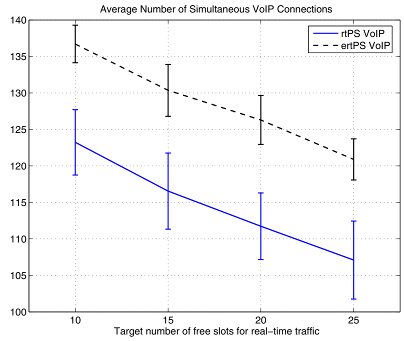

Figure 2. MBAC: rtPS vs. ertPS, number of VoIP users.

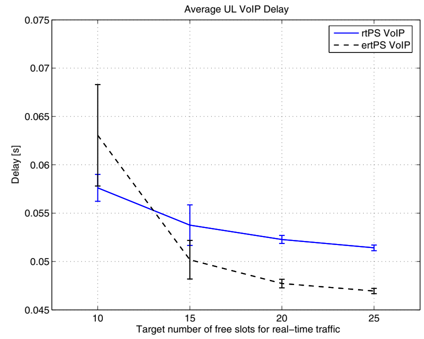

Figure 3. MBAC: rtPS vs. ertPS, average UL VoIP delay.

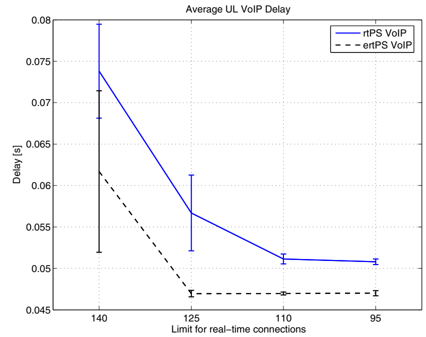

Figure 4. PBAC: rtPS vs. ertPS, average UL VoIP delay.

A. *rtPS VoIP vs. ertPS VoIP*

Fig. 2 illustrates that when VoIP is given ertPS treatment instead of rtPS treatment, we can have more connections as there is no need for polling during the talkspurt. However, during the silence period, one polling slot is granted every 30 ms. We use silence suppression detection at the BS: when more than 60 ms has passed from the last UL frame, we go into silence mode (where only polling slots are granted) and back to active mode when the next UL frame is received.

UL delays are also lower with ertPS (see Fig. 3). When the limit for the number of real-time connections is fixed (in parameter-based admission control, PBAC, see Fig. 4) the difference is more visible: ertPS VoIP will result in lower VoIP delay, since the ertPS VoIP connections consume fewer resources. Naturally, TCP goodput of BE connections is then better, too.

B. *MBAC vs. MAAC*

Fig. 5 and Fig. 6 illustrate the benefit of measurement-aided admission control. With Poisson connection arrivals, there is no real difference between MBAC and MAAC. However, when batch arrivals are introduced, MBAC cannot reject all the connections it should. Thus, too many connections are admitted and UL delays grow intolerable (see Fig. 6). MAAC does not have this problem as it utilizes bookkeeping with adaptive reservation limits instead of the number of free slots as such.

Fig. 7–9 illustrate the dynamics of MBAC and MAAC. Fig. 8 shows that the number of connections as a function of time does not follow a “saw tooth” pattern as in Fig. 7 (MBAC) but the curve is more stable. Fig. 9 shows that uplink was the bottleneck all the time.

V. CONCLUSIONS

In this paper, we have presented two simple-to-implement measurement-based connection admission control methods for real-time services in IEEE 802.16e networks. Our simulations show that the proposed methods lead into more efficient use of scarce radio resources than purely parameter-based connection admission control mechanisms with conservative limits for the number of admitted connections. For example, if dedicated resources were reserved for each VoIP connection, assuming that all connections used the most robust MCS (QPSK-1/2), we could only admit about 20 simultaneous VoIP connections[3]. However, since less robust MCSs are also used and since we do not have to serve all VoIP connections in a single frame, we can actually admit close to 140 VoIP connections when measurement-based (or measurement-aided) admission control is deployed.

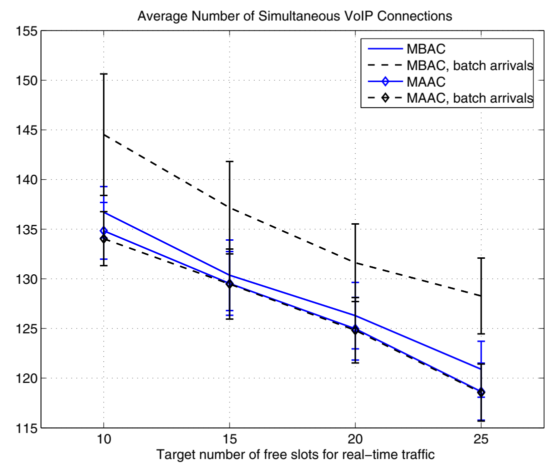

Figure 5. MBAC vs. MAAC (ertPS VoIP): number of VoIP users.

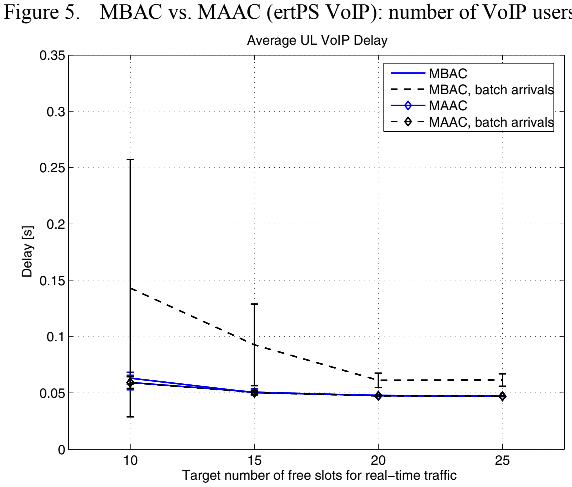

Figure 6. MBAC vs. MAAC (ertPS VoIP): average UL VoIP delay.

[3] The size of our VoIP packet (with overhead and HARQ) is 36 bytes, which translates into six slots with QPSK-1/2. According to Table I, we have circa 140 UL slots per frame.

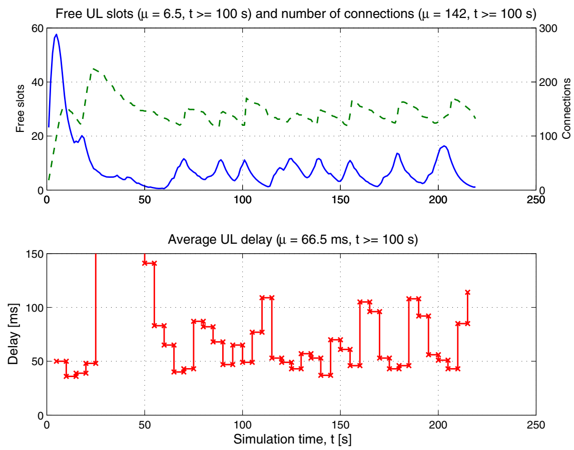

Figure 7. MBAC, ertPS, batch arrivals, target number of free slots: 10.

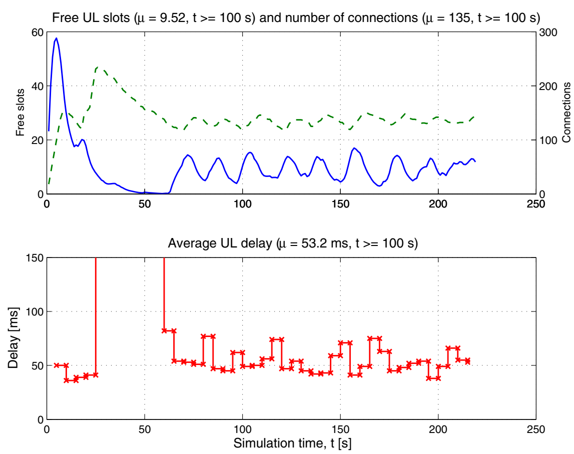

Figure 8. MAAC, ertPS, batch arrivals, target number of free slots: 10.

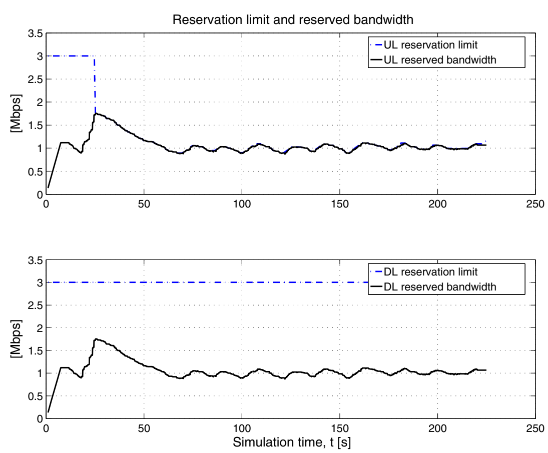

Figure 9. MAAC, ertPS, batch arrivals, target number of free slots: 10.

Since we cannot assume that connections always arrive to the system according to Poisson process, we should combine the benefits of parameter-based and measurement-based admission control in order to manage batch arrivals. We call this method measurement-aided admission control. If the connections arrive according to Poisson process, MAAC can admit as many VoIP connections as MBAC. However, if there are batch arrivals, MAAC blocks connections more aggressively than MBAC and thus delays are better controlled. Based on our results, we can conclude that it would be unwise to use MBAC instead of MAAC.

Moreover, we have shown that a well-designed ertPS implementation is a more bandwidth efficient solution for silence suppression capable VoIP than rtPS. However, if the VoIP client does not support silence suppression and packet size varies a lot (this is the case with, e.g., Skype), rtPS might be a better choice.

Our future research includes connection dropping, because even the most advanced admission control methods can sometimes admit too many connections. Dropping decision can be based on priority but it is also possible to utilize, e.g., DL queuing delays and UL virtual queue sizes. Thus, we could drop those connections first that are having the worst QoS. Moreover, it should be investigated how to prevent bandwidth stealing, i.e., how to prevent CQICH and HARQ acknowledgement slots from being granted to SSs hosting BE connections before serving real-time connections.

Acknowledgment

This work was performed, when Jani Lakkakorpi was still with Nokia Devices R&D. The authors would like to thank Kalevi Kilkki, Pasi Lassila (Helsinki University of Technology), Sandro Grech, Jani Moilanen (Nokia Siemens Networks) and everyone involved in the ns-2 simulator development at the Telecommunication laboratory of University of Jyväskylä.

References

[1] Air interface for fixed broadband wireless access systems. IEEE Standard 802.16, Jun. 2004.

[2] Air interface for fixed broadband wireless access systems – amendment for physical and medium access control layers for combined fixed and mobile operation in licensed bands. IEEE Standard 802.16e, Dec. 2005.

[3] D. Niyato and E. Hossain, "Connection Admission Control Algorithms for OFDM Wireless Networks," Proceedings of IEEE Globecom 2005, pp. 2455–2459, St. Louis, MO, USA, Nov. 2005.

[4] K.F. Tsang, L.T. Lee, H.Y. Tung, R. Lam, Y.T. Sun and K.T. Ko, "Admission Control Scheme for Mobile WiMAX Networks," Proceedings of IEEE International Symposium on Consumer Electronics, ISCE 2007, Dallas, TX, USA, Jun. 2007.

[5] J. Lakkakorpi, O. Strandberg and J. Salonen, "Adaptive Connection Admission Control for Differentiated Services Access Networks," IEEE Journal on Selected Areas in Communications, vol. 23, pp. 1963–1972, Oct. 2005.

[6] A. Sayenko, O. Alanen, H. Martikainen, V. Tykhmyrov, A. Puchko and T. Hämäläinen, "WINSE: WiMAX NS-2 Extension", Proceedings of 2nd International Conference on Simulation Tools and Techniques, Rome, Italy, Mar. 2009.

[7] M. Shreedhar and G. Varghese, "Efficient fair queueing using Deficit Round-Robin," IEEE/ACM Transactions on Networking, vol. 4, pp. 375–385, Jun. 1996.

[8] A. Sayenko, V. Tykhomyrov, H. Martikainen and O. Alanen, "Performance Analysis of the 802.16 ARQ Mechanism," Proceedings of IEEE/ACM MSWiM 2007, pp. 314–322, Chania, Greece, Oct. 2007.

[9] UCB/LBNL/VINT, "Network Simulator – ns (version 2)," Dec. 2006.