IEEE Africon 2011 - The Falls Resort and Conference Centre, Livingstone, Zambia, 13 - 15 September 2011

Connection Admission Control (CAC) for QoS Differentiation in PMP IEEE 802.16 Networks

Samuel K. Falowo and Neco Ventura
Department of Electrical Engineering
University of Cape Town, Private Bag, Rondebosch 7701, South Africa
Email: (sfalowo, neco)@crg.ee.uct.ac.za

***Abstract*—Connection admission control (CAC) is an important element for quality of service (QoS) provisioning in wireless networks, such as IEEE 802.16 standard, which is also known as worldwide interoperability for microwave access (WiMAX). While the standard defines PHY and MAC requirements, CAC is left to the vendors to design and implement for service differentiation and QoS support. In this paper a Quadra-Threshold (QT) CAC scheme is proposed for IEEE 802.16 networks. The proposed CAC considers four different thresholds for the four connection types namely, UGS, ertPS, rtPS and nrtPS services. The QT scheme is used to prioritize each connection type for service differentiation and QoS support. A performance analysis model based on Markov chains is proposed and numerical results are presented to demonstrate the performance of the proposed scheme. The scheme when compared with complete partitioning scheme performs better in terms of blocking probability.**

Index Terms— QoS, CAC, IEEE 802.16, Bandwidth Allocation

I. INTRODUCTION

WiMAX is a promising alternative for providing last mile broadband access in Wireless Metropolitan Area Networks (WMAN) with high speed, low cost and rapid deployment where wired infrastructure is economically and technically infeasible.

The IEEE 802.16 Standard [1] defines a flexible architecture of a base station (BS) and a number of subscriber stations (SSs). The standard specifies two operation modes: Point-to-Multipoint (PMP) and Mesh. In a PMP mode a base station provides connectivity, management, control and centrally coordinates the SS under its antenna sector while in mesh mode access coordination can be distributed among the SSs. PMP mode is considered in this paper. The communication path between SS and BS has two directions: uplink (from SS to BS) and downlink (from BS to SS). Transmission in uplink and downlink is multiplexed in either frequency division duplexing (FDD) or time division duplexing (TDD). The physical (PHY) layer of IEEE 802.16 operates in a frame format. Each frame is divided into uplink and downlink sub-frames.The frame control information broadcast to all SSs contains Downlink Map (DL-MAP) and Uplink Map (UL-MAP) messages that define the transmission burst profiles, including coding and modulation schemes. The medium access control (MAC) layer is connection-oriented. This layer-2 connection must be established with BS before data transmission can take place. A unidirectional transmission between the BS and a SS is defined by a connection in the PMP mode of the IEEE 802.16. Each connection associated with a single service class within the scheduling service domain is determined by a set of QoS descriptors that quantify aspects of its behavior. MAC scheduling services and their associated QoS parameters are defined in the standard. The scheduling services are:

- Unsolicited Grant Service (UGS) such as T1/E1 and VoIP without silence suppression with maximum sustained traffic rate (MSTR), maximum latency (ML) and tolerated jitter (TL) as its QoS requirements.
- real time Polling Service (rtPS) such as streaming audio and video with minimum reserved traffic rate (MRTR), MSTR, ML and traffic priority (TP) as its QoS requirements.
- non-real time Polling Service (nrtPS) such as FTP with MRTR, MSTR, and TP as its QoS requirements,
- Best Effort (BE) service such as data transfer and web browsing with MSTR and TP as its QoS requirements, and
- extended real time Polling Service (ertPS) of VoIP without silence suppression added in 2005 [2] with the same QoS requirements as rtPS.

While PHY and MAC specifications are defined in the Standards, CAC and packet scheduling are left to the vendors to design and implement. The objective of this paper is to present a design entitled a Quadra-Threshold (QT) based CAC for PMP IEEE 802.16 networks. The QT based CAC is designed to differentiate each UGS, ertPS, rtPS and nrtPS connection according to QoS requirement by assigning a peculiar threshold limit to each connection type. The different threshold limits are used to prioritize each connection for service differentiation and QoS support. The contributions of this work are (1) develop a QT based CAC scheme for service differentiation and QoS support, (2) develop an analytical model for our proposed scheme and (3) evaluate its efficacy and compare its performance with bandwidth partitioning scheme proposed in [3].

The rest of the paper is organized as follows. Section II gives the background relevant to this work. Section III describes the proposed Quadra-Threshold CAC scheme. Section IV discusses the analytical model. Simulations results are provided in section V. Finally, section VI contains the conclusions of this paper.

II. RELATED WORK

Previous researchers in an attempt to address call admission control problem in WiMAX employed complete sharing (CS) CAC scheme, in which the BS accepts connection requests and allocates available bandwidth resource equally without considering the QoS requirements of each connection request. This scheme is effective if all connection requests belong to a single service type. In a scenario where connections of all service types are randomly soliciting for admission, the connections with high priority will be treated in like manner with the connections with low priority. This in turn will lead to poor QoS of connection with high priority.

Other CAC schemes include bandwidth reservation as proposed by authors in [3], where priority is given to unsolicited grant service (UGS) connection by allocating a predetermined value of the total bandwidth of the network. The predetermined bandwidth is to ensure that UGS connection QoS requirements are guaranteed. Moreover a degradation model is developed by the authors to reduce nrtPS connection from its maximum sustained traffic rate to minimum traffic rate so that more UGS, rtPS and nrtPS connections can be admitted into the network. Their result shows that blocking probability is reduced in degradation mode compared to non-degradation mode. However, only UGS and nrtPS connections are addressed in their proposal. In [4], the authors use bandwidth reservation for different service types and user satisfaction based on utility functions to determine allocated bandwidth to polling services but only rtPS connection is considered in their result. Dynamic bandwidth reservation is proposed in [5]. The authors focus on UGS flows by assigning them higher priority. The bandwidth reservation for UGS flows takes place dynamically with respect to arrival rate. The UGS class is reserved bandwidth only during 'busy hour' conditions when the arrival rate of the connection requests exceeds a specified threshold. The scheme divides the scheduling services to UGS and Non-UGS (ertPS, rtPS and nrtPS) service types. This type of grouping is not compliant with IEEE 802.16 standard since each service type has associated QoS requirements. Complete sharing with multiple thresholds [6] [7] has been employed in mobile WiMAX to prioritized handover connection over new connection request to fulfill the specific QoS requirements for different service classes.

III. PROPOSED QUADRA-THRESHOLD CAC

In recent years, WiMAX as one of the wireless access networks has received enormous attention in wireless communication networks. The increasing bandwidth demands of network users and the emerging bandwidth-intensive applications such as video- conferencing and video on demand necessitate efficient utilization of limited network resources for QoS guarantee.

In order to have a balance between good QoS and efficient resource utilization, an efficient admission control algorithm is essential. The main objective of CAC in WiMAX network is to improve the QoS by limiting the number of on-going connections.

CAC operates when a new connection is being initiated (see Figure 1). Before a user can start transmission in the uplink channel, a user must be assured that bandwidth resource is available to support the transmission. To ensure bandwidth availability, the user makes connection request. The CAC checks whether there is available bandwidth to establish the connection (see Figure 1). A connection is rejected if the network resources are insufficient to establish the connection otherwise, the connection is admitted. Bandwidth requests are made by newly admitted and on-going rtPS, ertPS, nrtPS and BE connections. Bandwidth request is not made by UGS connection because it generates constant bit rate data and its bandwidth requirement does not change between connection establishment and termination as defined in IEEE 802.16 standards [1].

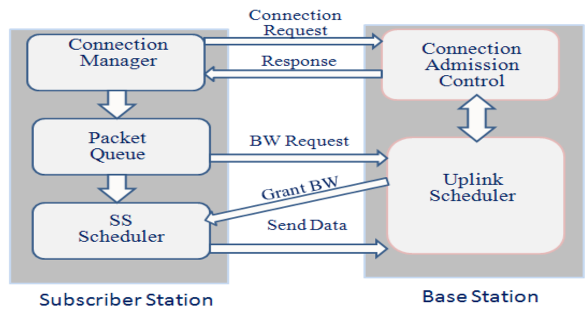

Figure 1. QoS architecture of IEEE 802.16.

In general, the uplink bandwidth request is made per connection while bandwidth allocation is performed in two modes: grant per connection (GPC) and grant per SS (GPSS). In the GPC mode, a BS scheduler handles each connection request from the SS independently and the bandwidth is explicitly granted to each connection while in the GPSS mode all connections from a SS are treated as one and the SS is granted an aggregate bandwidth. The SS scheduler allocates the aggregate bandwidth to its connections according to priority and QoS requirement of each connection.

The PMP mode is considered in this paper whereby transmission is from BS to a number of fixed SSs and vice versa. Therefore users' mobility is within the SS coverage sector. Every connection request originates from SS and BS initiates the connection creation process after a request has been made. The threshold based bandwidth sharing is employed in the proposed CAC scheme.

A. Quadra-Threshold (QT) bandwidth sharing scheme

Most of CAC schemes for WiMAX are based on complete sharing where network resources are shared between connection types without QoS differentiation. Unlike the complete sharing policy, the proposed admission control scheme is threshold based. While the complete sharing scheme can be fair to all connection types without ensuring the QoS requirements, the threshold based scheme gives priority to delay bound real time connections for QoS support and at the same time ensures fairness among different connection types.

In our threshold-based bandwidth sharing scheme, each connection type is assigned a bandwidth threshold value according to a priority given to each connection type. The order of threshold priority is given as: UGS > ertPS > rtPS > nrtPS. BE connections are not considered. In 802.16 MAC layer, the BE connections get the transmission opportunities only when other service connections do not transmit. Generally, BE connections do have long idle period and data in each transmission is relatively small, especially in the uplink direction. Therefore QoS of BE can be easily satisfied [3].

Let T_v denote the set of threshold values for connection types.

$$T_v = \{t_u, t_e, t_r, t_n\} : t_n \le t_r \le t_e \le t_u \le B \quad (1)$$

Where the parameters, t_u, t_e, t_r and t_n denote the threshold values for UGS, ertPS, rtPS, nrtPS connections and parameter B, the uplink bandwidth capacity of the network respectively (see figure 2). The parameter B is dynamically adjusted by BS according to the uplink bandwidth requirement of connections after a period of time, T which is long enough for BS to understand the behavior of uplink bandwidth requirement. IEEE 802.16 defines a dynamic frame format of uplink and downlink subframe [1].

Figure 2 illustrates the QT CAC scheme; all connection types are admitted into the network provided that network resources are available to sustain the flows.

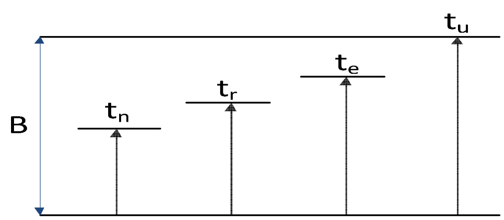

Figure 2. QT CAC sharing scheme.

All nrtPS connections are blocked after the threshold value t_n. The network could only admit rtPS, ertPS and UGS connections until the threshold point t_r after which all rtPS connections are blocked. The network admits ertPS and UGS connections until t_e threshold point. All ertPS connections are blocked after this point and the network admits only UGS connection until the network can not admit any other new UGS connections without violating the quality of service of ongoing connections. With Quadra-Threshold scheme QoS is guaranteed for each service type. The QT scheme does not only give priority to each connection type by assigning different thresholds but also ensures fairness to the connection types. In bandwidth partitioning scheme, even though short term fairness is guaranteed, in the long run the scheme is not fair to each connection type and the QoS of connections with high priority cannot be guaranteed.

IV. ANALYTICAL MODEL

Our proposed scheme can be modeled as a four dimensional Markov chain where each dimension is modeled as a M/M/∞ queue. In a M/M/∞ model there is no buffering of arriving connection. A connection is blocked if there are no network resources to establish it. Queuing model has been used in some literatures to model call arrival into networks [8] and [9]. The following assumptions are made for our model: (i) Connection arrival into the system follows Poisson distribution. (ii) Inter- arrival and service time are exponentially distributed (iii) the arrival process is independent of each other.

Let the set M of connection types be given as:

$$M = [UGS\ (u), ertPS\ (e), rtPS\ (r), nrtPS\ (n)] \quad (2)$$

The basic bandwidth unit (bbu) requirement of each connection type is represented by a set D given as:

$$D = [b_u, b_e, b_r, b_n] \quad (3)$$

Where the integers b_u, b_e, b_r, b_n denote the bbu requirement of each UGS, ertPS, rtPS and nrtPS connection respectively. Bandwidth requirements of service flows are prioritized according to their respective service type and priority is assigned to service types based on their QoS requirements.

For rtPS and nrtPS connections the reserved bandwidth is between the minimum and maximum bandwidth requirements. This is in accordance with IEEE 802.16 standards that define minimum reserved traffic rate and maximum sustained traffic rate for rtPS and nrtPS services. For rtPS connections the requested bandwidth, b_r which may be equal or greater than the minimum reserved bandwidth is given as:

$$b_r^{min} \le b_r \le b_r^{max} \quad (4)$$

For nrtPS the requested bandwidth, b_n which may be equal or greater than the minimum reserved bandwidth is given as:

$$b_n^{min} \le b_n \le b_n^{max} \quad (5)$$

UGS and ertPS connections are allocated maximum bandwidth requirements according to IEEE 802.16 standards so that delay requirements can be met.

Let α_i denote the threshold setting parameter of connection type-i. Connection type-i is an element of a set M (2). The threshold value of each connection will be calculated by using the equations below:

$$\begin{aligned} t_u &= \alpha_u B \\ t_e &= \alpha_e B \\ t_r &= \alpha_r B \\ t_n &= \alpha_n B \end{aligned} \quad (6)$$

Where the parameters α_u, α_e, α_r and α_n denote the threshold setting parameter of UGS, ertPS, rtPS and nrtPS connections respectively. The threshold setting parameters are calculated using the equations below:

978-1-61284-993-5/11/$26.00 ©2011 IEEE

$$\alpha_i = K * [b\_r_i] * [Q_f * p\_w_i] \forall i \in M$$
$$b\_r_i = \frac{H - b_i}{H}, H = b_u + b_e + b_r + b_n \quad (7)$$
$$Q_f = \frac{H^2}{4\sum (b_i)^2} \forall i \in M$$

The parameter b_r_i denotes the bandwidth ratio factor, Q_f is the fairness quotient factor derived from Jain's fairness index. Since the connection type with small basic bandwidth unit (bbu) requirement will have definitely have low blocking probability, a predefined traffic priority weight denoted as p_w_i is used to protect the connections with big bbu from small bbu connections. Equation (7) is bounded by the condition given as:

$$0.7 \le \alpha_i \le 1 \quad (8)$$

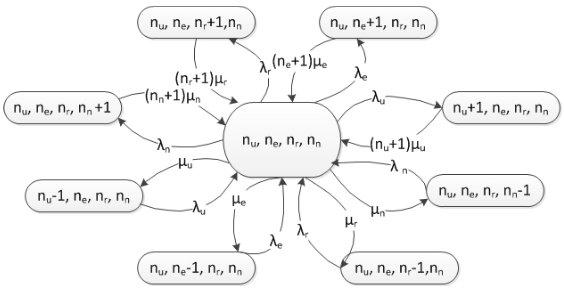

Figure 3. 4-dimentional Markov model's transition diagram.

Let λ_i and μ_i denote connection type-i arrival rate and service rate respectively. The state of the system is represented by vector s. The vector s given as:

$$s = (n_u, n_e, n_r, n_n) \quad (9)$$

where the non-negative integer n_i denotes the number of connection type-*i* in the network. For a given state $s = (n_u, n_e, n_r, n_n)$, state transition occurs when a new connection request is admitted or when an on-going connection completes. The transition diagram is given in Figure 3 which depicts the initial state, the transition state and the transition rate of the Markov chain for the proposed CAC. The arrival of a new connection type i into the network increases the number of the connection type in the network when admitted and the service of a connection type i reduces the number of the connection type in the network when completed.

Let S denote the state space of all possible states. The state S of all possible states is given as:

$$S = \{s = (n_u, n_e, n_r, n_n) \mid (n_n b_n \le t_n) \wedge (n_r b_r \le t_r) \wedge (n_e b_e \le t_e) \wedge (n_u b_u \le t_u) \wedge (\sum_{i \in m} n_i b_i \le B \forall i \in M)\} \quad (10)$$

Let ρ_i denote the load generated by a connection type-i. The load generated is given as:

$$\rho_i = \frac{\lambda_i}{\mu_i} \quad (11)$$

Let $P(h)$ denote the steady state probability that the system is in state h. State h is the state of the system in which the combination of number of connections in each class can be simultaneously supported without violating the QoS requirement of the connections. The steady state probability is given as:

$$P(h) = \frac{1}{\pi_0} \prod_{i \in m} \frac{\rho_i^{n_i}}{n_i!} \forall i \in M$$
$$\pi_0 = \sum_{h \in S} \prod_{i \in m} \frac{\rho_i^{n_i}}{n_i!} \forall i \in M \quad (12)$$

π_0 is the normalization constant.

From the steady state solution of the Markov model, performance measures of interest can be determined by summing up appropriate state probabilities.

A. *Connection blocking probability of nrtPS*

Let S_n denote the set of states in which a new nrtPS connection is blocked in the system. The set of states is given as:

$$S_n = \{h \in S : (b_n + \sum_{i \in m} n_i b_i) > t_n \forall i \in M \quad (13)$$

The blocking probability of a new nrtPS connection, P_n in the system is given as:

$$P_n = \sum_{h \in S_n} P(h) \quad (14)$$

B. *Connection blocking probability of rtPS*

Let S_r denote the set of states in which a new rtPS connection is blocked in the system. The set of states is given as:

$$S_r = \{h \in S : (b_r + \sum_{i \in m} n_i b_i) > t_r \forall i \in M \quad (15)$$

The blocking probability of a new rtPS connection is given as:

$$P_r = \sum_{h \in S_r} P(h) \quad (16)$$

C. *Connection blocking probability of ertPS*

Let S_e denote the set of states in which a new ertPS connection is blocked in the system. The set of states is given as:

$$S_e = \{h \in S : (b_e + \sum_{i \in m} n_i b_i) > t_e \forall i \in M \quad (17)$$

The blocking probability of a new ertPS connection is given as:

$$P_e = \sum_{h \in S_e} P(h) \quad (18)$$

D. *Connection blocking probability of UGS*

Let S_u denote the set of states in which a new UGS connection is blocked in the system. The set of states is given as:

$$S_u = \{h \in S : (b_u + \sum_{i \in m} n_i b_i) > t_u \forall i \in M \qquad (19)$$

The blocking probability of a new UGS connection is given as:

$$P_u = \sum_{h \in S_u} P(h) \qquad (20)$$

V. SIMULATION RESULTS

In order to evaluate the proposed QT CAC scheme, we compare our result with the scenario of bandwidth partitioning scheme (denoted as PS), the commonly used admission control scheme in IEEE 802.16 networks. The simulation program is conducted in MATLAB [10].

Table I, Parameter settings

Parameters	Values
Total uplink bandwidth, B (bbu)	100
UGS bandwidth requirement, b_u	1
ertPS bandwidth requirement, b_e	2
rtPS bandwidth requirement, b_r	3
nrtPS bandwidth requirement, b_n	4
Threshold limit of UGS connections, t_u	100
Threshold limit of ertPS connections, t_e	88
Threshold limit of rtPS connections, t_r	82
Threshold limit of nrtPS connections, t_n	80

The service rate of each connection type is set to 5conn/sec for UGS, 4conn/sec for ertPS, 3conn/sec rtPS, and 2conn/sec for nrtPS. The connection arrival rate is the same for all connection types and ranges from (2-20) conn/sec. Threshold setting parameters are calculated using equations (6) and (7). Other parameters used are provided in Table I.

In bandwidth partitioning scheme (PS), the uplink bandwidth capacity is partitioned into four parts and each part can only be used by a connection type. This method has been used by authors in [3] and [5] to partition the uplink capacity into two parts and each part can only be accessed by a designated group of connection types. In bandwidth partitioning scheme, UGS connections are allocated 14% of the uplink bandwidth, ertPS connections are allocated 20%, rtPS are allocated 30% and nrtPS connections are allocated 36% of the uplink bandwidth capacity. The total bandwidth allocated to the connection types is 100% when added together.

Figure 4 shows the connection blocking probability (BP) of UGS, ertPS, rtPS and nrtPS connections of PS scheme and the proposed QT scheme against connection arrival rate. The Figure shows that QT CAC scheme retains lower BP when compared with PS scheme. For the comparison between the two schemes to be clearly shown, the four connection types are separated in Figures 5-8. Figure 5 shows the BP of UGS connections against connection arrival rate. The BP of UGS-PS increases linearly from 0 after the 8th arrival rate to the value of 0.22 after 18th arrival. The UGS-QT maintains almost zero BP until after 12th arrival rate when the BP increases gradually to 0.11 after 18th arrival rate. UGS-QT scheme achieve lower BP as a result of sufficient uplink bandwidth available to its connections through the UGS threshold setting parameters. When other connections types do not solicit for bandwidth usage, the UGS connections make use of the total uplink bandwidth capacity. For UGS-PS, the connections only make use of the bandwidth within the set partition and the unused bandwidth of other connection types cannot be accessed by the UGS-PS connections. Figure 6 shows the BP of ertPS connections. In QT CAC scheme, ertPS connections can access up to a bandwidth capacity of 88bbu out of 100bbu through the threshold setting. If all the connections soliciting for admission are ertPS connections, 88% of the total bandwidth units are used by the connection type. In case of PS scheme, ertPS connections only made use of the allocated partitioned bandwidth of 30% irrespective of the types of connection present. Therefore, with increase in connection arrival rate, the BP of ertPS-PS increases from 0 after 4th connection arrival rate to 0.41 after 18th connection arrival. In ertPS-QT scheme all ertPS connections are admitted until after 12th arrival rate when the BP increases from 0 to 0.20 with 12th connection arrival rate. The BP of rtPS connections of QT scheme and that of PS in Figure 7 is similar to BP of ertPS connections in Figure 6. The similarity is as a result of close threshold values and bandwidth requirement allocated to the connections. The BP of nrtPS connection is presented in Figure 8. The nrtPS-PS connections suffer the highest BP when compared to other connection types in the same scheme despite the high bandwidth value allocated to its partition. The bandwidth requirement of each rtPS connection is four times that of UGS connection and two times that of ertPS connection. With proposed nrtPS-QT scheme, connections of nrtPS are admitted until the threshold value of 80 after when all nrtPS connections are rejected and if all the connections present are nrtPS connections 80% of the total uplink bandwidth are used.

The BP of nrtPS-PS increases from 0 after 4th connection arrival rate to the maximum value of 0.46 after 18th arrival rate. The QT scheme maintains 0 BP until after 12th connection arrival rate when the BP increases to 0.27 after 18th arrival rate.

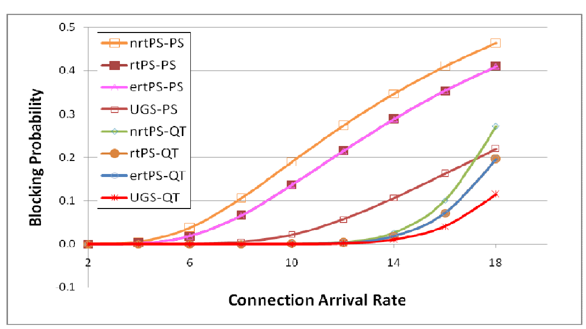

Figure 4. Blocking probability of QT and PS Schemes

978-1-61284-993-5/11/$26.00 ©2011 IEEE

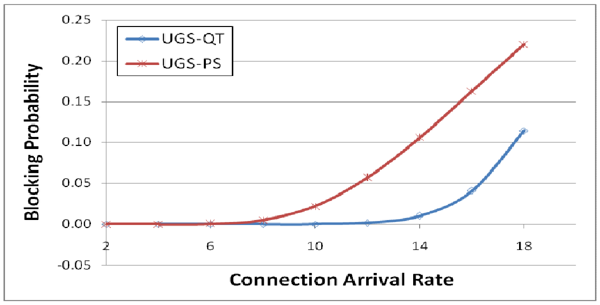

Figure 5. Blocking probability of UGS connections.

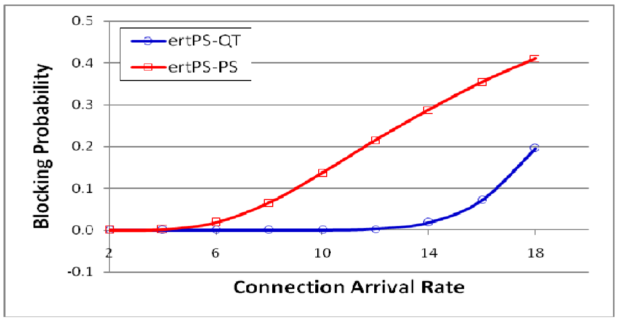

Figure 6. Blocking probability of ertPS connections

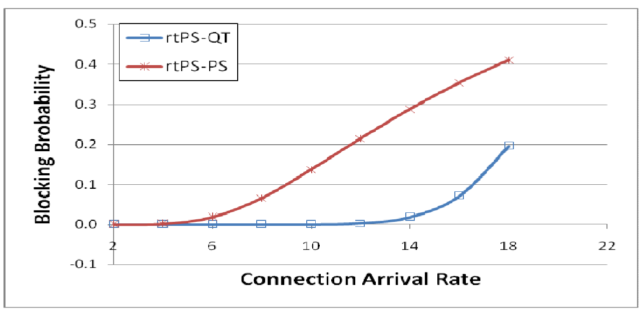

Figure 7. Blocking probability of rtPS connections.

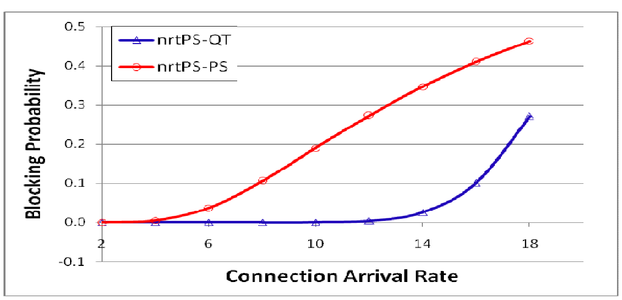

Figure 8. Blocking probability of nrtPS connections.

VI. CONCLUSIONS

The major contribution of this paper is that a Quadra-Threshold (QT) based CAC for service differentiation and QoS support in PMP IEEE 802.16 networks has been proposed. The QT based CAC is designed to differentiate each UGS, ertPS, rtPS and nrtPS connection type according to its QoS requirements by assigning it a threshold limit different from others. The different threshold limits are used to prioritize the connections for service differentiation and QoS support when making admission decisions. The efficacy of the proposed scheme has been proved by analytical model with numerical results. The evaluation shows that the proposed solution can improve CAC decision with lower blocking probability when compared with generic bandwidth partitioning scheme.

REFERENCES

[1] IEEE Standard 802.16-2004, Part 16. Air Interface for Fixed Broadband Wireless Access System, 2004.

[2] IEEE Standard 802.16-2005, Part 16. Air Interface for Fixed and Mobile Broadband Wireless Access System, 2005

[3] Haitang Wang, Wei Li and D.P, Agrawal, “Dynamic admission control and QoS for 802.16 Wireless MAN,” in Wireless Telecommunications Symposium, April 2005, pp.60-66.

[4] Chen, Jenhui, Chang and Chiang-Wei, “Traffic-Variation-Aware Connection Admission Control Mechanism for Polling Services in IEEE 802.16 Systems,” IFIP International Conference on Wireless and Optical Communications Networks, 2007. 2-4 July 2007 pp. 1-5.

[5] Angelos Antonopoulos and Christos Verikoukis, “Traffic-Aware Connection Admission Control Scheme for Broadband Mobile Systems” IEEE Communication Letters, Vol. 14, No. 8, August 2010.

[6] HY Tung, KF Tsang, “QoS for Mobile WiMAX Networks: Call Admission Control and Bandwidth Allocation,” In Proceedings of CCNC, 2008

[7] T. Kwon and Y. Choi, “ Call Admission Control for Adaptive Multimedia in Wireless/Mobile Networks, Proc. First ACM Int’l Workshop Wireless Mobile Multimedia, pp. 111-116, Oct. 1998.

[8] N. Nasser and H. Hassanein, “Adaptive call admission control for multimedia wireless networks with QoS provisioning,” Proceedings of the 2004 International Workshop on Mobile and Wireless Networking, Montreal, Canada, August 2004, pp. 30-37.

[9] O.E. Falowo and H.A. Chan, "Joint Call Admission Control Algorithm to Enhance Connection-level QoS in Heterogeneous Cellular Networks," Proceedings of IEEE Africon 2007, Windhoek, Namibia, 26 - 28 September 2007..

[10] The MathWorks, “MATLAB User’s Guide”, version 7.9, Natick, USA, 2009