Efficient Resource Allocation through Bandwidth Degradation of Non-Real-Time Polling Service in IEEE 802.16 Network

Samuel K. Falowo and Neco Ventura
Department of Electrical Engineering
University of Cape Town, Private Bag, Rondebosch 7701, South Africa
Tel: +27 78 0577078, Fax: +27 21 6503465
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 resource allocation with bandwidth degradation mechanisms for IEEE 802.16 network is proposed. The proposed CAC employs a threshold-based mechanism to allocate bandwidth to four connection types namely, UGS, ertPS, rtPS and nrtPS services. Bandwidth degradation scheme is used to admit more connection requests in order to increase connection throughput. A performance analysis model based on Markov decision process is used and numerical results are presented to demonstrate the performance of the proposed scheme. The scheme when compared with partitioning scheme and scheme without CAC performs better in terms of connection throughput.

Index Terms—QoS, IEEE 802.16, Admission Control, WiMAX,

I. INTRODUCTION

Emerging telecommunications services and applications are the strong drivers of increasing bandwidth demands for last mile broadband access. They pose new requirements to the existing network access technologies [1]. The demand for Internet Protocol (IP) connectivity is yielding rapid development in the wireless access network domain due to the proliferation of portable multimedia application devices such as laptop computers, Smartphone, hand-held computers and tablet personal computers. While the deployment of these wireless devices tends to address issues like network accessibility and portability, they also have their challenges. Although subscribers are willing to pay more for more bandwidth to obtain better quality of service (QoS) from network operators, limited bandwidth and allocation of the available bandwidth among different subscribers pose a big challenge.

The insufficient throughput support for broadband IP traffic in the existing wireless radio access technologies motivated the 3rd Generation Partnership Project (3GPP) to introduce WiMAX as a complementary broadband wireless access (BWA) technology. Although, cable and digital subscriber line (DSL) are already deployed on a large scale, IEEE 802.16 [2] is emerging as an access technology with several advantages. These include wireless connectivity, flexible architecture, high data rate and low cost of deployment.

The IEEE 802.16 Standard [2] 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. Figure 1 shows the PMP operation mode of IEEE 802.16 networks.

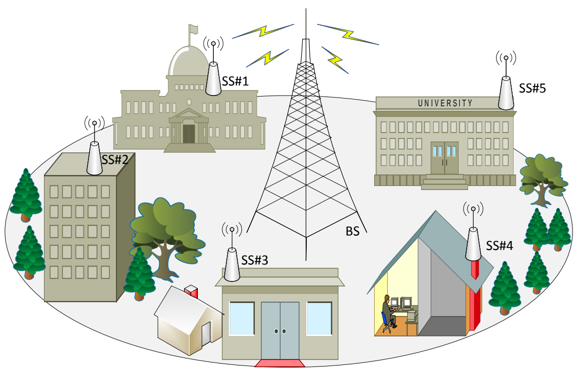

Figure 1: The PMP Operation Mode

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), extended real time polling service (ertPS), real time polling service (rtPS), and non-real time polling service (nrtPS). The QoS parameters associated with these scheduling classes are maximum sustained traffic rate (MSTR), minimum reserved traffic rate (MRTR), maximum latency (ML), tolerated jitter (TJ) and traffic priority (TP).

The problem of ensuring QoS is basically that of how to allocate available resources among users in order to meet QoS requirements [3]. While PHY and MAC specifications are defined in the Standards, Connection Admission Control (CAC) scheme is left to vendors to design and implement. CAC provides users with access to a wireless network with the objective of providing services to users with guaranteed QoS by limiting the number of users accessing the network and at the same time achieving efficient resource utilization.

The contributions of this work are (1) develop an efficient CAC scheme for service differentiation and QoS support, (2) develop an analytical model for our proposed scheme and (3) evaluate the performance of the proposed scheme.

The rest of the paper is organized as follows. Section II gives the background relevant to this work. Section III describes the proposed CAC framework. Section IV discusses the analytical model. Simulation results are presented in section V and finally, section VI concludes the paper.

II. RELATED WORK

Connection admission control is a vital part in QoS provisioning process. Some studies have focused on development of CAC since the introduction of IEEE 802.16 Standard. Different algorithms have been employed in making admission decisions for connection requests by service flows. The simplest of these algorithms is the complete sharing scheme. Complete sharing (CS) scheme assumes the base station accepts all connections until it runs out of resources. CS is easy to implement and it works efficiently when BS is handling a single type of service. However, a classical approach is required when handling multiple services.

Classic approach to CAC in wireless networks assumes allocation of dedicated resources like bandwidth reservation, service degradation to admit new connection request and fixed/dynamic guide channel or threshold to make provision for varying traffic.

Wang et al [4] proposed a CAC scheme that assigns highest priority to UGS flows and aims to maximize bandwidth utilization by using bandwidth borrowing and reduction methods. While bandwidth borrowing and reduction ensure that more connections are accepted, the QoS of the ongoing rtPS and nrtPS connections must be guaranteed. In addition, the authors allocated a predefined value of the total network capacity to UGS connections which could result in bandwidth wastage when not in use. The same approach was used in [5]. In the approach, the authors did not pre-allocate bandwidth capacity to UGS service as done in [4] but allowed all the service types to fully access the total bandwidth capacity which is an equivalence of complete sharing scheme.

In [6], the authors proposed a traffic aware Connection Admission Control scheme for broadband mobile systems. The scheme is based on bandwidth reservation concept, which is basically designed for 'busy hour' of a typical day. The scheme provides higher priority to VoIP calls (UGS connections) compared to other types of traffic (ertPS, rtPS and nrtPS connections) in the network. Although this classification simplifies the scheme and assures UGS connections a lower blocking probability, the service types of ertPS, rtPS and nrtPS cannot be classified as one, because their QoS requirements are different.

Dynamic bandwidth reservation is proposed in [6]. The authors focus on UGS flows by assigning higher priority to the flows. 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.

III. PROPOSED CAC

In this paper, a novel connection admission control algorithm with degradation mechanism for reducing connection blocking probability and increasing connection throughput is proposed. Figure 2 is the block diagram of the proposed admission control scheme. As shown in Figure 2, the connection admission control system consists of three main components: Connection bandwidth allocation (CBA), Bandwidth Unit Degradation (BRD) and Class Threshold Update, all being coordinated by CAC MANAGER.

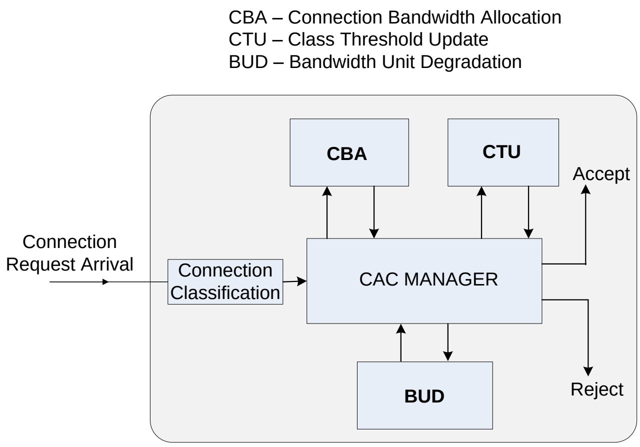

Figure 2: Connection Admission Control Framework

A. *Connection Bandwidth Allocation (CBA)*

When a connection request arrives, this component checks for bandwidth availability in the network, the threshold limit of the connection type requesting admission, estimates the achievable rate that can be offered to the connection and determines if the connection has not violated its admission requirements. This component is triggered by arrival of a new connection into the network. Each connection type has a threshold limit beyond which no connection would be admitted.

B. *Bandwidth Unit Degradation (BUD)*

Bandwidth unit degradation is a vital component of the system. This component is used to reduce the bandwidth of active connections to the minimum required bandwidth so that more connections can be admitted into the system thereby reducing connection blocking and increasing connection throughput. The component is triggered by a connection request blocking or rejection due to insufficient bandwidth to admit the newly arriving connection.

C. Class Threshold Update (CTU)

Each connection type is assigned a bandwidth threshold value according to the priority given to each connection type. The order of threshold priority is given as: $UGS > ertPS > rtPS > nrtPS$. The BE connections are not considered. In 802.16 MAC layer, BE connections get 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 [4], [7].

Class threshold update is responsible for dynamically updating the threshold values of the connection service types. Based on the changes in traffic condition of the connection types and in order to react to the changing traffic patterns and to allocate the scarce resources efficiently; the admission threshold of each service types is recalculated periodically. The time between two consecutive threshold updates is referred to as Class Threshold Update Period (CTUP). The time is fixed and threshold update takes place at the end of each CTUP. The time is chosen to be relatively long compared to connection service time so that it will not constitute another processing overhead. The new threshold values are used by CBA in the next CTUP.

Figure 3 shows the connection admission control operation flow chart. The flow chart explains how the CAC scheme operates.

IV. ANALYTICAL FRAMEWORK

This section deals with the task of developing an efficient analytical framework for performance evaluation of the proposed connection admission control for IEEE 802.16 networks.

A. Markov Decision Process

The Markov Decision Process is used to model the connection admission control of different service types defined in IEEE 802.16 Standards.

The CAC scheme is modeled as a four dimensional Markov chain where each dimension represents each service type of UGS, ertPS, rtPS and nrtPS and is modeled as an M/M/∞ queue. In an M/M/∞ queuing system, we have a situation where network resources are always available for each arriving connection into the system. If we perform call admission control and limit the number of connections admitted into the network to a value which can only be supported by the available network bandwidth, then, the state space of the system is a truncation of M/M/∞ open queuing network. Queuing model has been used in some literatures to model call arrival into networks [8] and [5]. 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 service types be represented by a set M. The set M is given as:

$$M = \{UGS\ (u), ertPS(e), rtPS(r), nrtPS(n)\} \quad \textbf{(1)}$$

Let the bandwidth requirement of each connection type be represented by a set D. The set D is given as:

$$D = \{\ b_u, b_e, b_r, b_n\ \} \quad \textbf{(2)}$$

Where the integers b_u, b_e, b_r, and b_n denote the basic bandwidth unit (bbu) requirements offered to UGS, ertPS, rtPS and nrtPS connections respectively. The offered bandwidth units to rtPS, b_r and nrtPS, b_n are given as:

$$b_r^{min} \le b_r \le b_r^{max} \\ b_n^{min} \le b_n \le b_n^{max} \quad \textbf{(3)}$$

Where b_i^{min} and b_i^{max} are the MRTR and MSTR of service type *i* respectively.

In the proposed CAC scheme, each connection type is assigned a bandwidth threshold value according to the priority given to each connection type. The order of threshold priority is given as: $UGS > ertPS > rtPS > nrtPS$.

Let T_v denote the set of threshold values for connection types

$$T_v = \{[t_u, t_e, t_r, t_n]: t_n \le t_r \le t_e \le t_u \le B\} \quad \textbf{(4)}$$

Where parameters, t_n, t_r, t_e and t_u are the set threshold limits for nrtPS, rtPS, ertPS and UGS connections and the parameter B is the uplink bandwidth capacity of the IEEE 802.16 network respectively.

Let λ_i and μ_i denote the connection arrival rate and service rate of connection type-*i*, for $i \in M$. The state of the system in the base station is represented by the vector s. The vector s is given as:

$$s = (n_u, n_e, n_r, n_n) \quad \textbf{(5)}$$

Where the non-negative integers $n_u, n_e, n_r\ and\ n_n$ denote the number of UGS, ertPS, rtPS and nrtPS connections in the network respectively. The state s represents the group service of number of connections of service type-*i* in the base station. The maximum number of connections, N_i of a service type-*i* that can be present in the network at a given time is given as the ratio of the bandwidth threshold of the service type and its required bandwidth. The parameter N_i is given as:

$$N_i = {}^{t_i}/_{b_i} \quad \forall\ i \in M \quad \textbf{(6)}$$

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

$$S = \{s = (n_u, n_e, n_r, n_n) | (n_n b_n \le t_n) \wedge (n_r b_r \le t_r) \wedge (n_e b_e \le t_e) \wedge (n_u b_u \le t_u) \wedge (\textstyle\sum_{i \in M} n_i b_i \le B\ \forall\ i \in M)\} \quad \textbf{(7)}$$

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

$$\rho_i = \frac{\lambda_i}{\mu_i} \quad \textbf{(8)}$$

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 service class can be simultaneously supported by the capacity of the network without violating the QoS requirement of the connections. The steady state probability that the system is in state h is given as:

$$P(h) = \frac{1}{\pi_0}\prod_{i\in M}\frac{\rho_i^{n_i}}{n_i}\ \forall\, i \in M$$
$$\pi_0 = \sum_{h\in S}\prod_{i\in M}\frac{\rho_i^{n_i}}{n_i}\ \forall\, i \in M \qquad \textbf{(9)}$$

The parameter 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.

can be borrowed from each nrtPS connection after degradation is $(b_n^j - b_n^{min})$ and the total bandwidth that can be borrowed from nrtPS connections is given as $\sum_{j=1}^{n_n}(b_n^j - b_n^{min})$. A new connection request, b_k is accepted if $b_k + n_i b_i + \sum_{j=1}^{n_n}(b_n^j - b_n^{min}) \le t_k$ and the request is reserved a minimum required bandwidth, otherwise, the connection is rejected. The parameter b_k is a connection request belonging to nrtPS and rtPS, and $n_i b_i$ is the bandwidth occupied by all ongoing connections of the

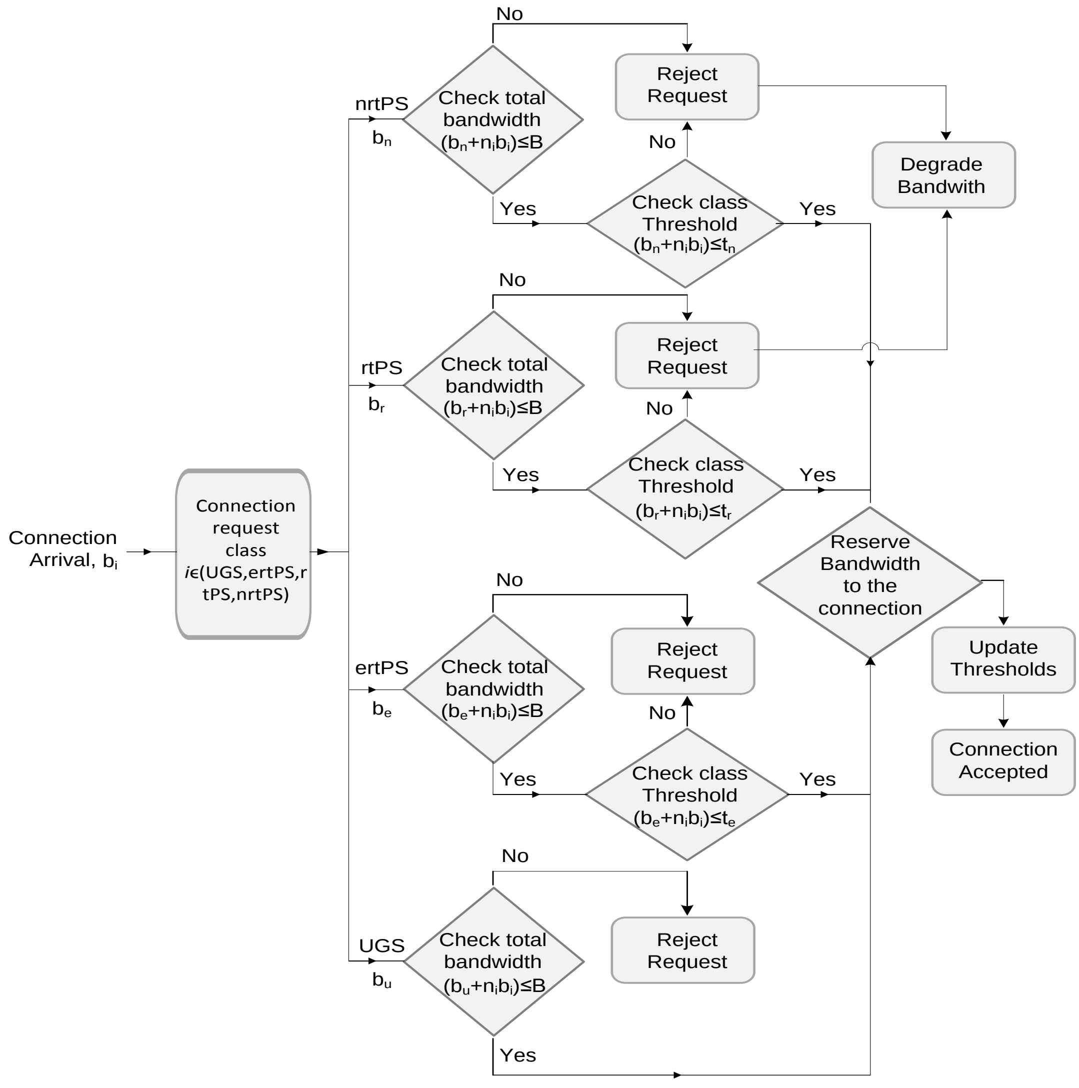

Figure 3: Connection Admission Control Operation Flow Chat

B. Bandwidth Degradation

When a new nrtPS and rtPS connection request is blocked, a bandwidth borrowing process is performed by degrading the bandwidth of ongoing nrtPS and rtPS connections to their minimum bandwidth requirement, thereby more connection requests can be admitted.

The number of ongoing nrtPS and rtPS connections has earlier been given as n_n and n_r respectively.

The reserved bandwidth for each nrtPS connection is b_n^j $(0 \le j \le n_n)$. The reserved bandwidth for each rtPS connection is b_r^j $(0 \le j \le n_r)$. The amount of bandwidth that

four service types while t_k is the admission threshold of the service type-k. After possible bandwidth degradation has been made on all on-going nrtPS connections and a new connection request is rejected due to insufficient available bandwidth, a degradation process is also performed on rtPS connections.

C. Class Threshold Determination

In this section, an expression is formulated for determining the threshold limit for each of the four service types. Let the total bandwidth requests of a connection type i

for i is an element of M be limited by some function f of the total uplink bandwidth capacity, B; connection type i with a total bandwidth requests below the control threshold T can have its bandwidth requests admitted.

At time t, let $T_i(t)$ be the control threshold of connection type-i. The control threshold is given as:

$$T_i(t) = f(B) \quad \textbf{(10)}$$

The simplest approach is to set the control threshold to a multiple $\propto$ of the total uplink bandwidth capacity. Therefore, it is given as:

$$T_i = \propto_i B \quad \textbf{(11)}$$

The threshold limit of each connection type is given as:

$$\begin{aligned} t_u &= \propto_u B \\ t_e &= \propto_e B \\ t_r &= \propto_r B \\ t_n &= \propto_n B \end{aligned} \quad \textbf{(12)}$$

D. *Performance Metric – connection throughput*

Let S_i denote the set of states in which a new connection request of service type i is blocked in the system. A new connection request of service type i is blocked when the set threshold is reached or the total bandwidth capacity of the network is used up after possible degradation has been made.

The set of states S_i is given as:

$$S_i = \{h \in S : (b_i + \sum_{i \in M} n_i b_i) > t_i \; \forall \; i \in M \quad \textbf{(13)}$$

The blocking probability of a new connection of service type i, P_i in the system is given as:

$$P_i = \sum_{h \in S_i} P(h) \quad \textbf{(14)}$$

In a system where a connection request can be blocked, the throughput cannot be defined as connection arrival rate because not all connection requests are admitted. The probability that an arriving connection request of service type i is blocked in the system is P_i for $i \in M$. The probability that the system is not full and an arriving connection requests is accepted into the system is $1 - P_i$. Thus the throughput Th_i of a connection type-i is given as:

$$Th_i = \lambda(1 - P_i) \quad \textbf{(15)}$$

where the parameter λ is the total connection arrival rate.

V. RESULT AND ANALYSIS

In this section the performance of the proposed scheme is evaluated by using MATLAB [9]. Connection requests of each service type arrive to the base station according to Poisson process. The connection arrival rate is the same for each service type and ranges from 2-18 connections per second. The limiting threshold of each connection type is calculated by using equations 10 to 12. More explanation about threshold calculation and assignment is given in [10]. Other simulation parameters are provided in Table 1. The table shows the parameters used in the performance evaluation.

The proposed CAC scheme is compared with a bandwidth partitioning scheme [10] and Non-CAC scheme as shown in Figure 4. In the bandwidth partitioning scheme, the uplink bandwidth capacity is partitioned into four parts and each part can only be used by a unique connection type. This method has been used by authors in [4] and [6] to partition the uplink capacity into two parts and each part can only be accessed by a designated group of connection types.

Table 1: Parameters used for Performance Evaluation

Parameter Settings			
Service Class	Min bandwidth (bbu)	Max Bandwidth (bbu)	Service Class Threshold
UGS		1	100
ertPS		2	88
rtPS	3	5	82
nrtPS	3	5	80

In the Non-CAC, the number of connection requests of each service type admitted into the system is not limited to the capacity that can be handled under the controlled threshold of each service type, thereby affecting the performance of the system.

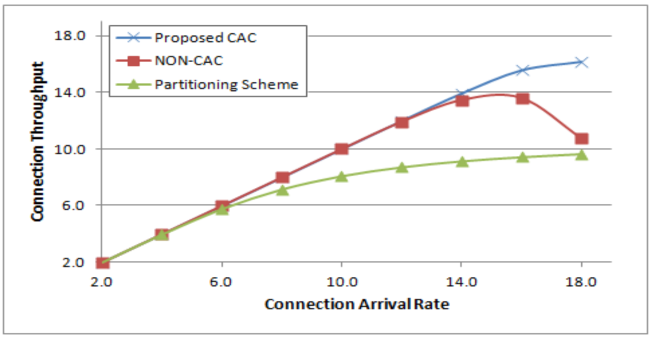

Figure 4: Connection Throughput of the Proposed Scheme and other CAC Schemes

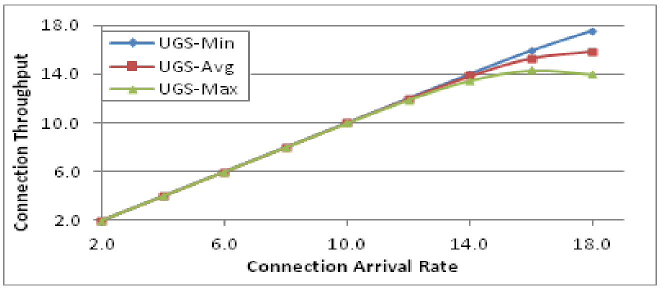

Figure 5: Connection Throughput of UGS Connections under Bandwidth Degradation of nrtPS Connections.

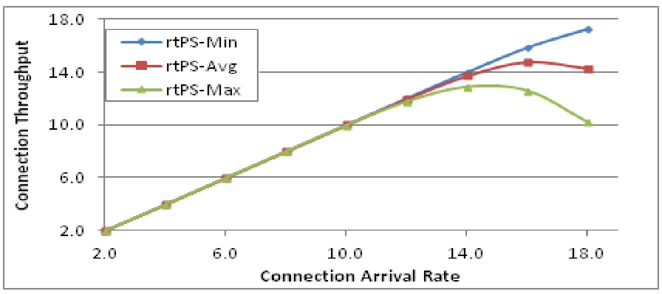

Figure 6: Connection Throughput of rtPS Connections under Bandwidth Degradation of nrtPS Connections.

It is ensured in the proposed scheme that the system should not admit connection request beyond the set threshold for each service type, otherwise, a connection request of a service type is blocked when the set threshold is reached.

Figure 4 shows that the proposed scheme performs better in terms of connection throughput when compared to partitioning scheme and the Non-CAC scheme. Unlike the partitioning CAC scheme where a connection type cannot use bandwidth beyond the allocated partition even though other service types are not using their allocated partition, the proposed scheme is threshold-based and a service type can make use of the available bandwidth to a set threshold in the absence of other connection types.

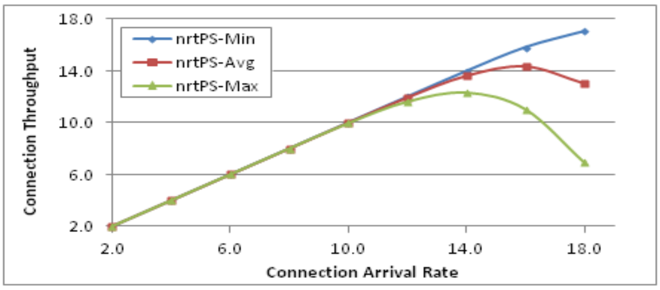

Figure 7: Connection Throughput of nrtPS Connections under Bandwidth Degradation of nrtPS Connections.

Having compared the proposed scheme to the existing schemes found in literatures in terms of connection throughput, the effect of bandwidth degradation of nrtPS connections from the maximum bandwidth requirement to the minimum bandwidth requirement is presented in Figure 5 to Figure 7. The legends Max represents when nrtPS connection request is offered maximum bandwidth for data transmission, The Avg. represents average offered bandwidth while the Min represents minimum offered bandwidth.

In Figure 5, more connection requests of UGS service type are admitted when nrtPS connections are offered minimum bandwidth (UGS-Min) compared to when they are offered maximum required bandwidth (UGS-Max), thereby an increase in connection throughput. Since the connection admission of each service type takes place simultaneously, offering nrtPS connection minimum required bandwidth increases the connection throughput of other service types.

Figure 6 shows the effect of degrading the bandwidth of nrtPS on rtPS. The connection throughput of rtPS-Min is almost constant throughout the arrival rate. At the 18th arrival rate the throughput dropped to 10connection/sec when nrtPS is offered maximum bandwidth.

The nrtPS connection type suffers the lowest connection throughput as shown in Figure 7. This is due to the high required bandwidth unit and the set threshold which is the lowest among other service types. Nevertheless the achieved connection throughput is still acceptable since there is no delay requirement for nrtPS connections. With maximum bandwidth unit, the connection throughput starts to decrease after 14th arrival rate.

VI. Conclusions

The major contribution of this paper is that a resource allocation and efficient bandwidth degradation has been proposed. A threshold-based connection admission control has been used to allocate bandwidth to the UGS, ertPS, rtPS and nrtPS service types of IEEE 802.16 network. Bandwidth degradation has been performed to admit more connection requests and increase connection throughput of all service types. The performance of the proposed scheme has been proved by analytical model with numerical results. Result evaluation shows that the proposed CAC with degradation mechanism can increase connection throughput and when compared with the generic partitioning scheme and scheme without CAC, it performs better.

References

[1] Radha, Krishna, Rao, G.S.V. and R. Radhamani. *WiMAX: A Wireless Technology Revolution* 2008.

[2] Anonymous "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),* pp. C1-2004, 2009.

[3] Chakchai So-In, R. Jain and A. -. Tamimi, "Scheduling in IEEE 802.16e mobile WiMAX networks: key issues and a survey," *Selected Areas in Communications, IEEE Journal on,* vol. 27, pp. 156-171, 2009.

[4] H. Wang, W. Li and D. P. Agrawal, "Dynamic admission control and QoS for 802.16 wireless MAN," in *Wireless Telecommunications Symposium, 2005,* 2005, pp. 60-66.

[5] S. Kalikivayi, I. S. Misra and K. Saha, "Bandwidth and delay guaranteed call admission control scheme for QOS provisioning in IEEE 802.16e mobile WiMAX," in *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE,* 2008, pp. 1-6.

[6] A. Antonopoulos, C. Skianis and C. Verikoukis, "Traffic-aware connection admission control scheme for broadband mobile systems," in *GLOBECOM 2010, 2010 IEEE Global Telecommunications Conference,* 2010, pp. 1-5.

[7] S. Chandra and A. Sahoo, "An efficient call admission control for IEEE 802.16 networks," in *Local & Metropolitan Area Networks, 2007. LANMAN 2007. 15th IEEE Workshop on,* 2007, pp. 188-193.

[8] Ke Yu, Xuan Wang, Songlin Sun, Lin Zhang and Xiaofei Wu, "A statistical connection admission control mechanism for multiservice IEEE 802.16 network," in *Vehicular Technology Conference, 2009. VTC Spring 2009. IEEE 69th,* 2009, pp. 1-5.

[9] The Math Works, "MATLAB User's Guide ," 2009.

[10] S. K. Falowo and N. Ventura, "Connection admission control (CAC) for QoS differentiation in PMP IEEE 802.16 networks," in *AFRICON, 2011,* 2011, pp. 1-6.

Samuel K. Falowo received his undergraduate degree in 2008 from The Federal University of Technology, Akure and is presently studying towards his Master of Science degree at the University of Cape Town. His research interest is in radio resource management.