Wireless Pers Commun (2015) 82:2673–2693
DOI 10.1007/s11277-015-2372-8

A QoS-Aware CAC with Bandwidth Reservation and Degradation Scheme in IEEE 802.16e Networks

Ibrahim Saidu · Shamala Subramaniam · Azmi Jaafar · Zuriati Ahmad Zukarnain

Published online: 18 April 2015

Abstract Call admission control (CAC) schemes play a critical role in providing quality-of-service (QoS) guarantees for various classes of traffic with diverse QoS requirements in IEEE 802.16e networks. The dynamic CAC and bandwidth reservation scheme is one of the current schemes that simultaneously provides efficient utilization of network resources and guarantees QoS for admitted connections. However, its admission criteria starved high and low service classes due to its linear adaptation policy to accommodate more users into the network under moderate-to-heavy traffic load conditions. Its adaptive threshold for handoff connections is adjusted based on the arrival of new and handoff connections, that results in a waste of resources when the new and handoff connection arrival rate occurs frequently. In this paper, a new CAC scheme for Mobile WiMAX networks is proposed to prevent starvation of service classes and enhance the efficient utilization of network resources. The scheme determines a new admission criteria based on a scheduling service class. In the admission criteria, bandwidth degradation policy is used to admit more users when there is no available bandwidth to admit a more users. The adaptive threshold has been introduced dynamically to adjust the quantity of reserved bandwidth for handoff connections based on the traffic intensity of handoff requests. In addition, an analytical model for the proposed scheme is also developed. Extensive simulation experiments have been conducted to evaluate the performance of our proposed approach. The simulation results illustrate that the proposed scheme significantly improves the network efficiency compared to other schemes in terms of

I. Saidu (✉) · S. Subramaniam · A. Jaafar · Z. A. Zukarnain
Faculty of Computer Science and Information Technology, Universiti Putra Malaysia (UPM), 43400 Serdang, Selangor D.E., Malaysia
e-mail: ibrasaidu.@gmail.com

A. Jaafar
e-mail: azmij@upm.edu.my

Z. A. Zukarnain
e-mail: zuriati@upm.edu.my

S. Subramaniam
Sports Academy, Universiti Putra Malaysia (UPM), 43400 Serdang, Selangor D.E., Malaysia
e-mail: shamala@upm.edu.my

accepting more connections into the network and assuring QoS for all service classes. The numerical results have shown similar performance to the simulation results.

Keywords Mobile WiMAX · Handoff connection · New connection · Admission criteria · Adaptive threshold

1 Introduction

The IEEE 802.16e standard [1] is a wireless technology that provides ubiquitous wireless access at high bandwidth, wide coverage, and low cost of deployment to residential users and small businesses. The standard sets out two specifications: the physical (PHY) and media access control (MAC) layers. It has several advantages, including ease and cost of deployment, first-mile/last-mile access, and QoS support for multimedia applications at the MAC layer [2–5]. Because multimedia applications must support different types of traffic simultaneously, each of which has different QoS requirements from the network, such as bandwidth, delay, jitter, and packet loss, providing QoS to these traffic classes represents a challenge. Therefore, CAC schemes are essential in providing the required level of network QoS [6,7].

The CAC scheme is a process in which a new or handoff connection is accepted into the network only if its QoS can be satisfied while ensuring the QoS of existing connections is not degraded [8] according to availability of network resources. Handoff is the process of transferring a mobile station (MS) connection from its connected base station (BS) that suffers worse signal quality to a neighbouring BS with the best quality signal. The scheme prioritizes a handoff user to a new network user in order to provide better user-perceived satisfaction. The design of a CAC mechanism must consider the need for available bandwidth to achieve the QoS requirements for the handoff connections. Therefore, the BS should reserve a certain quantity of bandwidth for upcoming handoff connections only and allot the remaining to new connections. The CAC scheme is considered effective if it is able to simultaneously provide the QoS to the admitted connections and efficiently utilize the network resources. Since 802.16e is an extension of the original standard 802.16, the earlier CAC schemes proposed in 802.16 [9–12] support only new connections in order to improve the bandwidth utilization as well as QoS of various service classes. Nonetheless, the schemes ignore consideration of handoff connections because the standard does not support mobility. Because the mobile standard supports mobility, the later schemes are also proposed [13,15, 16] with consideration of both new and handoff connections. Among these, dynamic CAC and bandwidth reservation scheme in [16] is one of the recently used schemes that adaptively defines the admission criteria based on traffic loads and adopts an adaptive QoS technique. The scheme efficiently utilized network resources and assured QoS for admitted new connections and handoff connections. However, its admission criteria starved both high and low priority service classes to accommodate more users into the network under moderate-heavy loads because of the linear adaptation policy. The scheme also adjusts the amount of reserved bandwidth for handoff connections based on the arrival of handoff and new connections by considering a fixed maximum reserved bandwidth threshold for handoff connections that may lead to a waste of resources during frequent arrival of new and handoff connections.

In this paper, a new CAC scheme is proposed called, *QoS-Aware connection admission controls with bandwidth reservation (BR) and bandwidth degradation (BD)*, for 802.16e networks to improve upon the efficiency of dynamic CAC and BR [16]. First, the scheme determines new admission criteria based on a scheduling service class. In the admission criteria, a bandwidth degradation strategy is introduced to create more access opportunities to new connections when the available network resource is insufficient under moderate-to-heavy load. Then, it employs an adaptive threshold that adaptively reserves the quantity of bandwidth for handoff connections based on traffic intensity of the handoff connections. Moreover, the paper also presents an analytical model for the proposed scheme. The algorithm is evaluated through extensive simulations conducted by discrete event simulation. The performance of the proposed algorithm is compared and evaluated by means of simulation results with dynamic threshold (DT) CAC [15] and dynamic CAC and BR [16]. The results show that the proposed scheme significantly improves network resource utilization compared to other schemes in terms of granting more connections to the network and guaranteeing QoS of service classes. Finally, numerical results have shown similar performance compared to the simulation results with negligible variations.

This paper is organized as follows: Section 2 briefly reviews the IEEE 802.16e standard, and Sect. 3 addresses related work on CAC schemes. In Sect. 4, describe the proposed scheme. In Sect. 5 presents an analytical model. Section 6 presents the performance evaluations, and conclusion in Sect. 7.

2 Overview of IEEE 802.16e Standard

The standard [1] defines two layers to support both fixed and mobile wireless metropolitan area networks (WMANs). The MAC layer sits atop the PHY layer and mediates between it and the layers above. The protocol that operates the MAC layer performs the main tasks of the standard, such as QoS provisioning, connection admission control (CAC), bandwidth allocation, and scheduling. It supports two modes of operation: point-to-multipoint (PMP) and mesh. The former is a cellular-like structure that supports communication between a BS and a set of subscriber stations (MSs) in broadcast fashion. The BS is the central controller, regulating all communications between itself and a set of MSs. Each MS can represent a single or multiple users. The two paths of communication between the BS and the MSs are the uplink (UP; from MS to BS) and downlink (DL; from BS to MS) directions. In contrast, mesh mode supports multihop communication between MSs. In this paper, PMP is considered as the main operational mode.

The PHY layer is responsible for transmitting bits over the wireless channel by means of the adaptive modulation and coding (AMC) technique. AMC supports two transmission modes, frequency-division duplexing (FDD) and time-division duplexing (TDD). In FDD mode, uplink and downlink data are sent on different frequencies. In contrast, in TDD mode both UP and DL data are sent using the same frequency but in different time slices. Both duplexing modes operate in a frame format. Each frame is partitioned into DL and UP subframes. The DL subframe is used by the BS to transmit data and manage messages to an MS, while the UP subframe is used by all MSs to transmit data.it's uses an orthogonal frequency multiple access (OFDMA) slot as its smallest unit of resource.

The MAC layer is a connection-oriented protocol that has the advantage of controlling network resource sharing among individual connections. The protocol maps both connected and connectionless traffic to a unique connection identifier (CID). If traffic coming from an upper layer arrives at the MAC layer, the MS attempts to establish a connection with the BS. The BS employs a CAC scheme that checks whether the resources available can ensure the QoS of the new connection while maintaining the QoS guarantees for the existing users. With the acceptance of a new connection, the BS responds to the MS with a CID to use for the UP and DL directions. Once a connection is set, the MS can request bandwidth from the BS. The

BS grants bandwidth using the grant per subscriber station (GPSS) approach. Once the MS receives its bandwidth from the BS, its packet scheduler distributes the bandwidth among its own active connections. The CAC and request grant bandwidth allocation components of the BS provide support to different applications with various QoS requirements. The 802.16 standard partitions applications into service classes as follows.

The *unsolicited grant service* (UGS) periodically generates constant-size data packets for real-time traffic such as VoIP without silence suppression. UGS is sensitive to transmission delays, and the BS allocates grants to the MS in an unsolicited fashion using the maximum sustained traffic rate (MSTR), traffic priority, and maximum latency tolerance as its QoS requirements.

The *real-time polling service* (rtPS) generates variable-size data packets for real-time traffic such as MPEG video. It has less stringent delay requirements and is periodically polled by the BS for each MS to individually determine its bandwidth requirement. Its mandatory QoS specifications are the minimum reserved traffic rate (MRTR), MSTR, traffic priority, and maximum latency tolerance.

The *extended real-time polling service* (ertPS) generates variable-size data packets for real-time traffic such as VoIP with silence suppression. It combines features of both UGS and rtPS and has strict, guaranteed delay requirements and provides unicast grants in an unsolicited manner by the BS, as with UGS. Because UGS grants are of constant size whereas ertPS grants vary in size, an MS can request a change of its bandwidth grant to suit its requirements. The ertPS QoS requirements are MRTR, MSTR, traffic priority, maximum latency tolerance, and delay jitter tolerance.

The *non–real-time polling service* (nrtPS) generates variable-size data packets for non–real-time traffic such as FTP. It has minimum bandwidth requirements that are delay tolerant. It is polled by the BS in order for each MS to state its desired bandwidth. The QoS requirements are MRTR, MSTR, and traffic priority.

The *best-effort service* (BE) is designed to support traffic for which delay and throughput are not guaranteed, such as HTTP. It requests bandwidth through contention request opportunities and unicast request opportunities.

3 Related Work

Several CAC schemes have been proposed to address the problems that affect service class applications in IEEE 802.16e. This section provides an overview of some major CAC schemes that have been applied in 802.16 and 802.16e networks.

In [9], a dynamic admission control scheme is proposed according to the scheduling services defined in a fixed standard. It uses bandwidth reservation and degradation in order to provide QoS to the service classes. The bandwidth reservation is used to give higher priority to the UGS service class because it is widely used by people for their daily communication. The degradation is supported by rtPS and nrtPS service because of the variable generation of packets. The two-service classes have varying bandwidth requirements between the MSTR and MRTR. In this approach, only the nrtPS class is degraded. The paper illustrates that the scheme provides maximum priority to the UGS service class and maximum bandwidth utilization by bandwidth degradation as well as minimization of the blocking probabilities of the service classes. However, handover to mobile users is ignored because the standard does not support mobility. A similar approach has also been proposed in [10], with the only difference being the giving of a reservation to the rtPS service class.

In [11,12], there were proposed CAC and packet scheduling schemes for QoS provisioning in 802.16 networks. The CAC scheme is a token bucket based CAC that only admits new connection by ensuring that the QoS of the existing connections will not be degraded, and the newly admitted connections will be provided with the required QoS. The scheme reserves bandwidth to UGS and rtPS and uses the available bandwidth for other service classes. It also ensures a delay guarantee to the rtPS service class when the bandwidth requirements are satisfied. Therefore, the scheme provides bandwidth guarantees to real time traffic classes as well as ensures a delay guarantee for rtPS. However, the algorithms starve the low priority service class and ignore considerations of the handoff connections as well as the efficient utilization of network resources.

In [13], there is proposed an efficient admission control scheme which considers handoffs to the mobile users based on the characteristics of an adaptive multimedia service. It uses the available bandwidth and bandwidth degradation of the active service flow when there is insufficient bandwidth to admit a handoff service flow during a heavy network condition, while when the network load reduces, the degraded service flows are upgraded in the reverse order. The authors showed that the method efficiently utilizes the resources and reduces the new blocking probability and handoff dropping probability. However, the method is unfair because it degrades the bandwidth requirements of the lowest priority service class to its minimum bandwidth requirement in a step-wise manner.

In [14], a CAC scheme is proposed for the non-provisioned service flow in mobile WiMAX, which supports new arrival and handoff connections. The scheme used guard channel and proportional bandwidth-borrowing policy in order to assign high priority to handoff connections than new originating connections. The two policies used provide a reasonable priority order to a new and handoff connections to various service classes.The proposed scheme maximized the bandwidth utilization and reduced the connection-blocking probability as well as the connection-dropping probability of the two connections. However, if handoff connections do not consume the entire reserved resources, a certain portion of may be wasted because it can never be used to admit a new connections.

In [15], a CAC scheme is proposed in order to assure QoS to different scheduling service classes in mobile WiMAX. The scheme groups the scheduling service classes into real time and non-real time service classes. The scheme uses a dynamic guard channel that is dynamically changed between a minimum bandwidth reserved threshold and a maximum bandwidth reserved threshold, based on the arrival and departure of handoff connections. This dynamic adjustment gives handoff connections the opportunity to use the maximum reserved bandwidth when the handoff requests arrive heavily. The scheme reduces handoff dropping and new call blocking probabilities. However, when the level of handoff requests is low, the amount of reserved bandwidth will be almost fixed. As a result, the scheme behaves like the fixed guard channel policy, which may also lead to a waste of network resources.

In [16], dynamic CAC and BR schemes are proposed for mobile WiMAX that both improve the efficient utilization of network resources and assure the QoS for admitted new and handoff connections. The scheme dynamically adjusts the admission criteria based on the network loads and an adaptive QoS approach. Linear adaptation is used to regulate the admitted connections when there is no available available. If the new and handoff connections are admitted into the network based on this criterion, then the bandwidth reservation scheme is dynamically adjusted to the quantity of the reserved bandwidth for handoffs based on the arrival of new and handoff connections. The authors showed that the schemes enhance the efficiency of the resource utilization as well as assuring QoS for new and handoff connections. However, the admission criteria used starved the high and the low service classes when the traffic load is moderate or heavy. It also uses a fixed maximum reserved bandwidth threshold based on a handoff or new arrival connection to adjust its adaptive threshold, and this may lead waste of network resources under frequent arrival of the new and handoff connections.

In order to address the aforementioned problems, a QoS-Aware CAC scheme for mobile WiMAX with BR and BD is proposed to improve the efficient utilization of resources and ensure QoS to all the service classes. The CAC scheme that is proposed differs from the dynamic CAC and BR [16] that it dynamically adjusts the admission criteria based on the scheduling services . It also adjust the amount of bandwidth reservation for handoff connections according to the traffic intensity of the handoff connections.Moreover, an analytical model for the scheme is also developed.

4 Proposed Scheme

In this section, a QoS-Aware scheme that is an improvement on the joint CAC and BR scheme in [16] is described. First, however, the shortcomings of the joint CAC and BR scheme are presented. That scheme defined its admission criteria based on the traffic loads and an adaptive QoS strategy to improve the efficiency of the bandwidth utilization as well the QoS guarantee of the various service classes. The loads are classified according to low, moderate, and heavy. The adaptive QoS strategy is a strategy in which the scheduled service class is assigned MSTR or MRTR requirements. The admission criteria are dynamically determined according to the three classifications of the network loads. Firstly, if the network load is low, the QoS is adjusted to MSTR. Then, if the network load becomes moderate or heavy, a linear adaptation approach is used to reduce the bandwidth requirement of the accepted service classes to the MRTR. During moderate-to-heavy network loads, the QoS strategy adapts the UGS class from the MSTR to the MRTR while the BE service class goes from MSTR to its MRTR, which is zero. The two-service classes have no MRTR QoS parameter as defined by the IEEE 802.16e standard. If the bandwidth requirement of the UGS service class is reduced to MRTR during a moderate-heavy traffic load, then its QoS requirement is degraded because of the inadequate bandwidth available to this traffic class. As a result, UGS starvation occurrs, and consequently leads to a waste of resources because the available bandwidth may not be sufficient to admit the traffic class and cannot be used to admit other service classes. While the BE service class will not be admitted into the network, as a result of there being no minimum bandwidth requirement associated to its MRTR, this may also lead to its starvation. Therefore, the starvation of these service classes leads to an increase in handoff blocking and new dropping probabilities. The scheme also used an adaptive threshold that is dynamically changed based on the arrival of handoffs and new connections by employing a fixed maximum bandwidth reservation for handoffs. Since the adaptive threshold is updated based on the minimum and maximum reserved bandwidth thresholds, and when both the new and handoff connection arrival rate occurs frequently, then some resources may be left unused and hence lead to an inefficient utilization of network resources.

For example, let assume the total bandwidth (B) in the network be 100 and the network is empty in the beginning. Suppose that 80 new connections, and 10 handoff connections arrived sequentially in the network. Each of the connection arrived is assumed to request a unit of bandwidth. The scheme also assumed the minimum and maximum reserved threshold for bandwidth to be 0 and 90, respectively. The initial adaptive threshold is computed to be 45 units. From Fig. 1, it shows the first 90 connections (80 new and 10 handoff) are accepted into the network. However, the last 10 new connections are rejected because the dynamic adjustments of the adaptive threshold uses the arrival of both new and handoff connections.

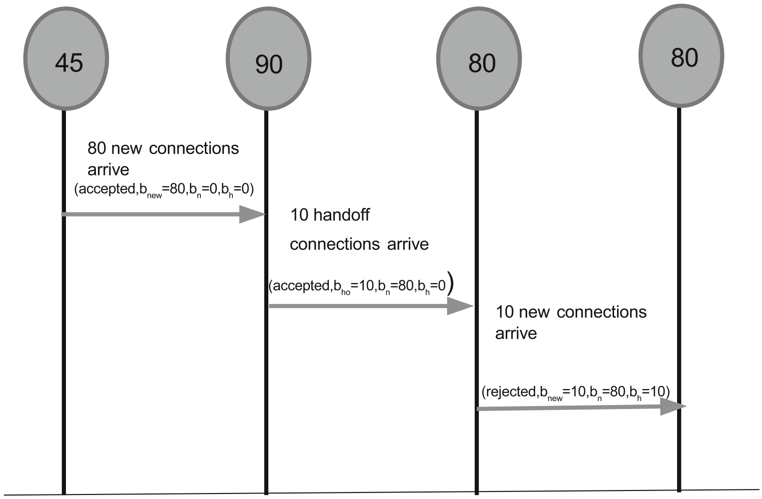


Fig. 1 The process of connection admissions with the dynamic CAC and BR scheme

It also shows that 20 units of bandwidth are left unused and can never be used to admit a new connections. Therefore, the bandwidth resource is inefficiently utilized.

To address the shortcomings of the CAC and BR described above, firstly, a new admission criterion is proposed:

The admission criterion proposed in [15] is modified that uses scheduling service classes to admit new or handoff connections. Because most of the scheduling services are adapted to the QoS parameters, the scheduling service is classified into real time and non-real time service classes. On the one hand, the real-time service class is a class that has MSTR as one its QoS parameters. On the other hand, the non-real time class is a class with an MRTR as one of its QoS parameters. However, the BE service class does not have an MRTR as defined by the standard. But assumed an MRTR in order to prevent its being starved [15]. Three cases are defined in our admission criterion.

Case 1:

If a new or handoff connection belongs to the UGS, rtPS, or ertPS service classes, the MSTR requirement will be assigned. For each class, the required bandwidth can be derived as

$$b_i = b_i^{max}$$

where b_i is the admission criteria for connection i and b_i^{max} is the maximum bandwidth requirements corresponding to the highest QoS for connection i.

Case 2:

If a new or handoff connection belongs to the nrtPS or BE service classes, the MRTR requirement will be assigned. For each class, the required bandwidth can be derived as

$$b_i = b_i^{min}$$

where b_i^{min} is the minimum bandwidth requirements corresponding to the lowest QoS for connection i.

Case 3:

If there is no more available bandwidth to admit a new connection, then a bandwidth degradation mode will be used. The degradation is only performed on rtPS and ertPS, which are already assigned a maximum rate since they can adapt their QoS requirements from MSTR to MRTR. This contrasts with the UGS class, which can only support MSTR. To save the UGS class from starvation, its MSTR already assigned is not degraded. In this mode, the admission criterion for rtPS and ertPS will be decreased to MRTR. The bandwidth degradation for each class $j\,(j = 1$ or $2)$ is computed as follows:

$$b_j^d = b_j^{max} - l_j^n \delta_j \tag{1}$$

where b_j^d is bandwidth degradation of class j; b_j^{max} is the maximum bandwidth available on the existing connection in class j, l_j^n and δ_j are the current degradation level and the quantity of degraded bandwidth for every degradation step in class j, respectively.

The above equation needs to satisfy the equation below,

$$b_j^{max} - l_j^n \delta_j \geq b_j^{min}. \tag{2}$$

where b_j^{max} is the minimum bandwidth available on the existing connection in class j

The maximum degradation step size is calculated below,

$$l_j^{max} = \frac{b_j^{max} - b_j^{min}}{\delta_j}. \tag{3}$$

where l_j^{max} is the maximum degradation step size of class j.

To accept a handoff or a new connection based on the proposed admission criterion, the bandwidth allocated to an admitted handoff and new connection is denoted as $b_j^h(t)$ and $b_j^{nw}(t)$, respectively, over time.

A handoff connection $h_{con-accepted}(t)$ is accepted into the network when

$$h_{con-accepted}(t) = \left(b_i^{hof} + \sum_{t=0}^{n} \left(b_i^h\,(t) + b_i^{nw}(t) \right) \leq B \right) \tag{4}$$

where b_i^{hof} is the handoff admission criterion of the handoff connection shown in the admission criteria defined above and B is the total bandwidth. The term in the RHS ensures the maximum bandwidth is not exceeded.

The system capacity B may vary over time because the mobile standard supports multiple transmission rates. The formula in [16] is adopted to compute the data transmission rate. That formula is based on various MCSs defined in the standard by

$$R_{MCS_i} = \left(N_{Data\_SC} / T_S \right) * B_{MCS_i} \tag{5}$$

where N_{Data_SC} is the number of data sub-carriers, T_S is the symbol period, and B_{MCS_i} is the amount of information in bits/symbol with respect to the i_{th} MCS.

Then, introduced an adaptive threshold, adopted from [18], to dynamically change the bandwidth reservation threshold for handoff connections based on the traffic intensity of the handoff connections.

$$th_{adap} = \lfloor \rho_{hof} \times \beta \rfloor \times b_i^{hof} \tag{6}$$

where $\rho_{hof} = \frac{\lambda_{hof}}{\mu_{hof}}$ denotes the traffic intensity, λ_{hof} is the arrival rate for handoff connections, and μ_{hof} is the mean service rate. In addition, b_i^{hof} is the bandwidth required for each handoff connection, while $\beta \in [0, 1]$ represents the bandwidth reservation factor.

A new connection is accepted $n_{con-accepted}(t)$ based on the admission criterion defined earlier, when the condition below holds:

$$n_{con-accepted}(t) = \left[\left(b_i^{new} + \sum_{t=0}^{n} \left(b_i^{h}(t) + b_i^{nw}(t) \right) \leq B - th_{adap} \right) \bigcup \left(b_i^{new} \leq b_j^{d} \right) \right] \tag{7}$$

where b_i^{new} is the new bandwidth admission criterion. The first term in the RHS ensures that new and existing connections do not exceed the maximum available bandwidth for new connections. The second term increases the chances of receiving new users when there is no available bandwidth.

The pseudocode for the QoS-Aware CAC With BR and BD scheme is in Algorithm 1.

Algorithm 1: A QoS-Aware CAC With BR and BD.

At time epoch t
for *(all pending connection c)* **do**
    if *(service class i of c is in* $(UGS, rtPS, ertPS)$*)* **then**
        $b_i \leftarrow b_i^{max}$
    else
        $b_i \leftarrow b_i^{min}$
    if *(type is handoff and service class i)* **then**
        if $(b_i^{hof} + \sum_{t=0}^{n}(b_i^{h}(t) + b_i^{nw}(t)) \leq B$ **then**
            accept handoff
            $th_{adap} \leftarrow \lfloor \rho_{hof} \times \beta \rfloor \times b_i^{hof}$
        else
            reject handoff
    else
        if $(b_i^{new} + \sum_{t=0}^{n}(b_i^{h}(t) + b_i^{nw}(t)) \leq B - th_{adap})$ **then**
            accept new connection
        else
            if *(*$b_{rtPS}^{max} - l_{rtPS}^{n}\delta \geq b_{rtPS}^{min}$*)* **then**
                $b_{rtPS}^{d} \leftarrow b_{rtPS}^{max} - l_{rtPS}^{n}\delta$
                if $b_{rtPS}^{d} \geq b_i^{new}$ **then**
                    accept new
            else
                if $b_{ertPS}^{max} - l_{ertPS}^{n}\delta \geq b_{ertPS}^{min}$*)* **then**
                    $b_{ertPS}^{d} \leftarrow b_{ertPS}^{max} - l_{ertPS}^{n}\delta$
                    if $b_{ertPS}^{d} \geq b_i^{new}$ **then**
                        accept new
        else
            reject new
end

5 Analytical Model

In this section, to develop an analytical model, a system is considered consisting of a single BS and several MSs, as shown in Fig. 3: some of the MSs are in the cell that requests new connections, while others are around the cell that requests handoff connections. The MSs request bandwidth from the BS whenever there is a new or a handoff connection. All the WiMAX service classes are used at the BS, such as UGS, rtPS, ertPS, nrtPS, and BE;

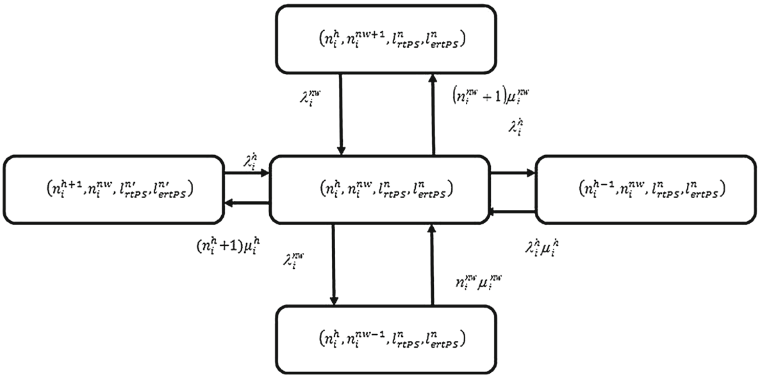


Fig. 2 State transition diagram of the BS representing an ergodic Markov chain

each of the service classes requires QoS guarantees and request either a new admission or a handoff connection. In the BS, a CAC scheme is used that reserves bandwidth for handoff connections and admits a new connection whenever there is sufficient bandwidth left, or degrades the service of the already admitted rtPS and ertPS to their minimum. Otherwise, it rejects either of the connections.The new connection blocking probability and the handoff connection dropping probability are used as the performance measures for the proposed CAC scheme. To obtain these probabilities, the diagram described in Fig. 2 is used to model the proposed scheme using Markov chain for each class i of traffic for $i = 1, 2, 3...k$. In each class i, the model uses the assumptions below:

1. The new and handoff connections have an arrival rate that follows a Poisson process, while their mean service rates follow exponential distributions.
2. λ_i^{nw} is the mean arrival rate of new connections of class i traffic.
3. λ_i^h is the mean arrival rate of handoff connections of class i traffic.
4. $\lambda_i^t = \lambda_i^h + \lambda_i^{nw}$ is the mean total arrival rate of class i traffic.
5. $\frac{1}{\mu_i^{nw}}$ is the mean service rate of new connections of class i traffic.
6. $\frac{1}{\mu_i^h}$ is the mean service rate of handoff connections of class i traffic.
7. $\frac{1}{\mu_t^i} = \frac{1}{\mu_i^h + \mu_i^{nw}}$ is the mean total service rate of connections of class i traffic.

The threshold limit for handoff connection i (λ_i^h) is

$$\lambda_i^h = \lambda_i^t \times \alpha_i^h. \tag{8}$$

The threshold limit for new connection i (λ_i^{nw}) is

$$\lambda_i^{nw} = \lambda_i^t \times \alpha_i^{nw} \tag{9}$$

where λ_i^t denote the total arrival of both handoff and new connections; α_h and α_{nw} represent the threshold parameter settings for the handoff and the new connections, respectively.

The BS is model in the form of a four dimensional continous Markov chain with a state $S = \{n_i^h, n_i^{nw}, l_{rtPS}^n, l_{ertPS}^n\}$ according to the number of connections accepted and the degraded level of the two service classes. The set of all possible states of BS is denoted as:

$$S = \{n_i^h, n_i^{nw}, l_{rtPS}^n, l_{ertPS}^n\} \tag{10}$$

Table 1 State transitions of the Markov chain

Current state	Next state	Transition state
$\left(n_i^h, n_i^{nw}, l_{rtPS}^n, l_{ertPS}^n\right)$	$\left(n_i^h+1, n_i^{nw}, l_{rtPS}^{n'}, l_{ertPS}^{n'}\right)$	λ_i^h
	$\left(n_i^h-1, n_i^{nw}, l_{rtPS}^{n'}, l_{ertPS}^{n'}\right)$	$\left(n_i^h+1\right)\mu_i^h$
	$\left(n_i^h, n_i^{nw}+1, l_{rtPS}^{n'}, l_{ertPS}^{n'}\right)$	λ_i^{nw}
	$\left(n_i^h, n_i^{nw}-1, l_{rtPS}^{n'}, l_{ertPS}^{n'}\right)$	$\left(n_i^{nw}+1\right)\mu_i^{nw}$

where n_i^h and n_i^{nw} are the current number of handoff and new connections, respectively. While l_{rtPS}^n and l_{ertPS}^n is the current degradation level of rtPS and ertPS, respectively.

The steady probability of the state S is defined as $\pi_{\{n_i^h, n_i^{nw}, l_{rtPS}^n, l_{ertPS}^n\}}$

The state space S for all possible states according to the proposed scheme is defined as

$$S = \left\{s = \left(n_i^h, n_i^{nw}, l_{rtPS}^n, l_{ertPS}^n\right) \middle| n_i^h b_i^h + n_i^{nw} b_i^{nw} + l_{rtPS}^n \leq B \right. \\ \left. \bigwedge n_i^{nw} b_i^{nw} \leq B - th_{adap} \bigwedge l_{rtPS}^n \leq l_{rtPS}^{max} \bigwedge l_{ertPS}^n \leq l_{ertPS}^{max} \right\} \quad (11)$$

A state is considered valid if and only if

$$n_i^h b_i^h + n_i^{nw} b_i^{nw} \leq B \bigwedge n_i^{nw} b_i^{nw} \leq B - th_{adap} \\ \left. \bigwedge l_{rtPS}^n \leq l_{rtPS}^{max} \bigwedge l_{ertPS}^n \leq l_{ertPS}^{max} \right\} \quad (12)$$

For a given state $s = \{n_i^h, n_i^{nw}, l_{rtPS}^n, l_{ertPS}^n\}$, a state transition takes place when a new or handoff connection is admitted or terminated. The state transition is described in Table 1, derived from Fig. 2. $l_{rtPS}^{n'}$ and $l_{ertPS}^{n'}$ are the degradation levels of rtPS and ertPS, respectively, after the state transition, and may take different values under certain conditions. Hence, the state transition can be different for different states.

From each state transition diagram, the corresponding transition balance equation, can be seen below.

$$n_i^h > 0 \bigwedge n_i^{nw} > 0 \bigwedge l_{rtPS}^n \leq l_{rtPS}^{max} \bigwedge l_{ertPS}^n \leq l_{ertPS}^{max} \\ \bigwedge n_i^h b_i^h + n_i^{nw} b_i^{nw} + b_i^h \leq B \\ \bigwedge n_i^h b_i^h + n_i^{nw} b_i^{nw} + b_i^{nw} \leq B - th_{adap} \\ \left. \bigwedge n_i^h b_i^h + n_i^{nw} b_i^{nw} + \left(l_{rtPS}^{max} - l_{rtPS}^n\right) + \left(l_{ertPS}^{max} - l_{ertPS}^n\right) \leq B - th_{adap} \right\} \quad (13)$$

The state balance equation for the above equation can be derived as

$$\left(\lambda_i^h + \lambda_i^{nw} + n_i^h \mu_i^h + n_i^{nw} \mu_i^{nw}\right) \pi_{\left(n_i^h, n_i^{nw}, l_{rtPS}^{n'}, l_{ertPS}^{n'}\right)} \\ = \lambda_i^h \pi_{\left(n_i^h - 1, n_i^{nw}, l_{rtPS}^{n'}, l_{ertPS}^{n'}\right)} + \lambda_i^{nw} \pi_{\left(n_i^h, n_i^{nw} - 1, l_{rtPS}^{n'}, l_{ertPS}^{n'}\right)} \\ + \left(n_i^h + 1\right) \mu_i^h \pi_{\left(n_i^h + 1, n_i^{nw}, l_{rtPS}^{n'}, l_{ertPS}^{n'}\right)} \\ \left. + \left(n_i^{nw} + 1\right) \mu_i^{nw} \pi_{\left(n_i^h, n_i^{nw} + 1, l_{rtPS}^{n'}, l_{ertPS}^{n'}\right)} \right\} \quad (14)$$

Here, $\pi_{n_i^h, n_i^{nw}, l_{rtPS}^n, l_{ertPS}^n}$ represents the steady state probability of the system and $\phi_{(n_i^h, n_i^{nw}, l_{rtPS}^n, l_{ertPS}^n)}$ represents characteristics function by

$$\phi_{(n_i^h, n_i^{nw}, l_{rtPS}^n, l_{ertPS}^n)} = \begin{cases} 1, & \left(n_i^h, n_i^{nw}, l_{rtPS}^n, l_{ertPS}^n\right) \in S \\ 0, & otherwise \end{cases} \tag{15}$$

The equation above is used to avoid a transition into an invalid state. In addition, a normalizing condition is also used:

$$\sum\nolimits_{n_i^h, n_i^{nw}, l_{rtPS}^n, l_{ertPS}^n \in S} \pi_{(n_i^h, n_i^{nw}, l_{rtPS}^n, l_{ertPS}^n)} = 1 \tag{16}$$

The above equation is used to obtain the steady state probability for each state. The handoff connection dropping probability (*HCDP*) and the new connection blocking probability (*NCBP*) connections are

$$HCDP = \sum\nolimits_{s \in S'} \pi_{(n_i^h, n_i^{nw}, l_{rtPS}^n, l_{ertPS}^n)} \tag{17}$$

Here, $S' = \left\{s = \left(n_i^h, n_i^{nw}, l_{rtPS}^n, l_{ertPS}^n\right) | (n_i^h + 1) b_i^h + n_i^{nw} b_i^{nw} > B\right\}$

$$NCBP = \sum\nolimits_{s \in S'} \pi_{(n_i^h, n_i^{nw}, l_{rtPS}^n, l_{ertPS}^n)} \tag{18}$$

where $S' = \left\{s = \left(n_i^h, n_i^{nw}, l_{rtPS}^n, l_{ertPS}^n\right) | n_i^h b_i^h + \left(n_i^{nw} + 1\right) b_i^{nw} > B\right\}$.

6 Performance Evaluation

This section evaluates the performance of three CAC schemes: the dynamic threshold (DT) [15], dynamic CAC and BR [16], and the new proposed scheme. The schemes are evaluated in terms of the throughput of the two service classes UGS and BE, new connection blocking rate and the handoff connection dropping rate, by means of simulations. The discrete event simulation is developed via C++ programming language. In the experiments, two scenarios are examined for the performance evaluation of the proposed scheme. For each of the scenario,the arrival rates of new connections and the handoff connections is assume to be Poisson arrivals: the mean departure rate has also been assumed to be $\frac{1}{10th}$ of the arrival rate. In addition, BS is assumed to know the bandwidth requirements of each connection with respect to its current MCS. The MCS parameters are shown in Table 2 and, the simulation used a 2×2 MIMO mechanism. The simulation time is 1000 s and the results are taken from an average of 20 different simulations performed for each scenario. The simulation parameters are presented in Table 3, while the simulation topology is shown in Fig. 3.

6.1 The Achieved Throughput for UGS and BE Traffic Class with Different CAC Schemes

In the this section,the performance of the proposed scheme and the other schemes are compared in terms of the throughput achieved for either UGS or BE service class. the simulation setup for this scenario consider only the two types of traffic classes : UGS and BE. Here,the reserve bandwidth for handoff users is not considered. Hence, each of the in coming connection will be accepted into the network when the available bandwidth can achieve the admission criteria determined by different CAC schemes. To study the achieved QoS performances regarding the two types of service classes with the proposed scheme and the other

Table 2 Modulation and coding parameters for 10 MHz channel [17]

Modulation	Coding rate	Downlink	Uplink
QPSK	$\frac{1}{2}CTC, 6\times$	1.06	0.78
	$\frac{1}{2}CCT, 4\times$	1.58	1.18
	$\frac{1}{2}CTC, 2\times$	3.17	2.35
	$\frac{1}{2}CTC, 1\times$	6.34	4.70
	$\frac{3}{4}CTC$	9.50	7.06
16 QAM	$\frac{1}{2}CTC$	12.67	9.41
	$\frac{3}{4}CTC$	19.01	14.11
64 QAM	$\frac{1}{2}CTC$	19.01	14.11
	$\frac{2}{3}CCT$	25.34	18.82
	$\frac{3}{4}CTC$	28.51	21.17
	$\frac{5}{6}CTC$	31.68	23.52

Table 3 Parameters for IEEE 802.16e PHY data rates [17]

Parameter	Downlink	Uplink
System bandwidth	10 MHz	
FFT size	1024	
Null subcarriers	184	
Pilot subcarriers	120	280
Data subcarriers	720	560
Symbol period	102.9 μs	
Frame duration	5 ms	
OFDM symbols/frame	48	
Data OFDM symbols	44	

schemes, the performance of each is evaluated separately with its specific QoS requirements shown in Table 4.

Figure 4 shows the throughput of UGS traffic class for DT CAC, dynamic CAC and BR, and the proposed scheme. It shows that when the network arrival rate is low, i.e., from 0 to 0.3, all the schemes have similar throughput. But as the network arrival rate increases, the proposed scheme performs better than the other two schemes. The proposed scheme increases the throughput by 14.93 % from the DT scheme, and by 28.81 % from the dynamic CAC and BR scheme. The increase in the throughput is attributed to the new admission criteria introduced, that prevent the starvation of the UGS traffic and allows it to transmit at maximum rate as well as the used of degradation scheme to admit more UGS traffic class when there is no available bandwidth.

Fig. 5 shows the throughput of BE traffic class for DT CAC, dynamic CAC and BR, and the proposed scheme. The figure shows that the proposed scheme has a performance similar to that of the other two schemes when the arrival rate is within 0.1–0.4. However, as the network arrival rates increase, the proposed scheme has a better performance compared to the other two schemes. The proposed approach has a 5.85 and 34.04 % improvements as compared to the DT CAC and dynamic CAC and BR schemes,respectively. The reason for this is the proposed scheme prevent the starvation of BE service class during moderate-heavy traffic

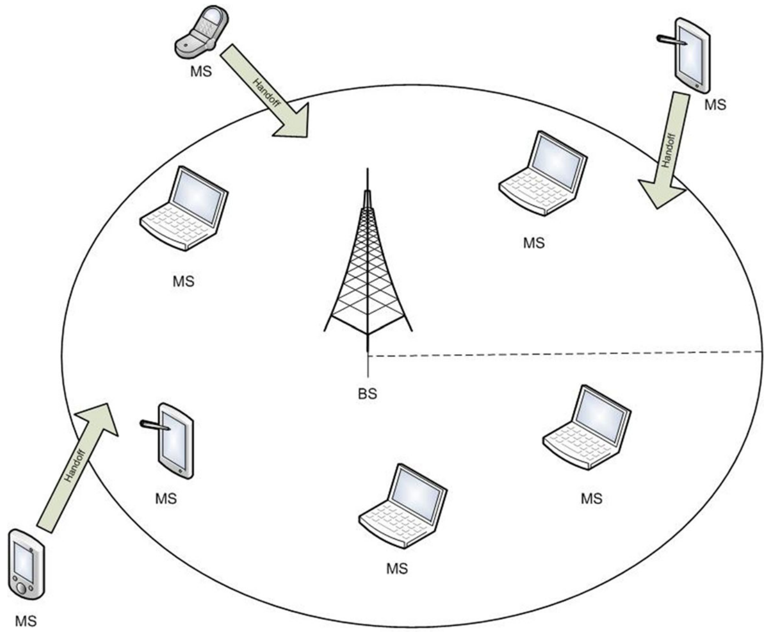


Fig. 3 Simulation topology

Table 4 Parameters for IEEE 802.16e PHY data rates

Class	Maximum rate (kbps)	Minimum rate (kbps)
UGS	32	32
rtPS	128	32
ertPS	128	32
nrtPS	64	32
BE	32	0

load and when there is no available bandwidth,degradation scheme is also used to allow the BE service class into the network at minimum rate.

In Figs. 4 and 5, similar observations regarding the UGS and BE traffic classes is achieved. The results show that the proposed algorithm can efficiently improve utilization of network resources and also assure QoS guarantee for admitted connections

6.2 The New Connection Blocking Rate and Handoff Dropping Rate with Different CAC Schemes

In the section, the simulation scenario considers new connections and handoff connections to investigate the performance of the proposed scheme. The scheme is compared with the other schemes. The simulation setup in this scenario considers all the scheduling service classes in Mobile WiMAX networks, each service class is associated with the QoS requirements shown in Table 4. Although a lower priority BE class has zero as its minimum rate, 32 kbps is assumed as its minimum reserved rate in order to avoid starvation under moderate and heavy network loads. The scheduling services are assumed to have uniform probabilities of

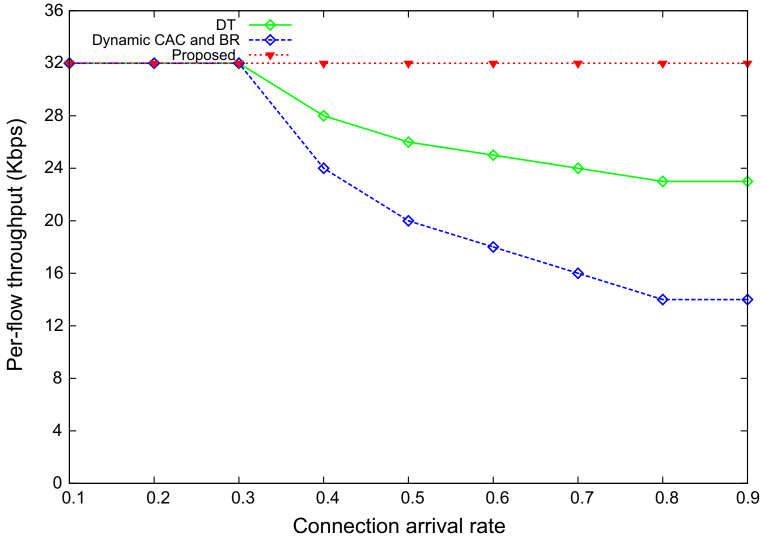


Fig. 4 Throughput of UGS traffic class with different CAC schemes

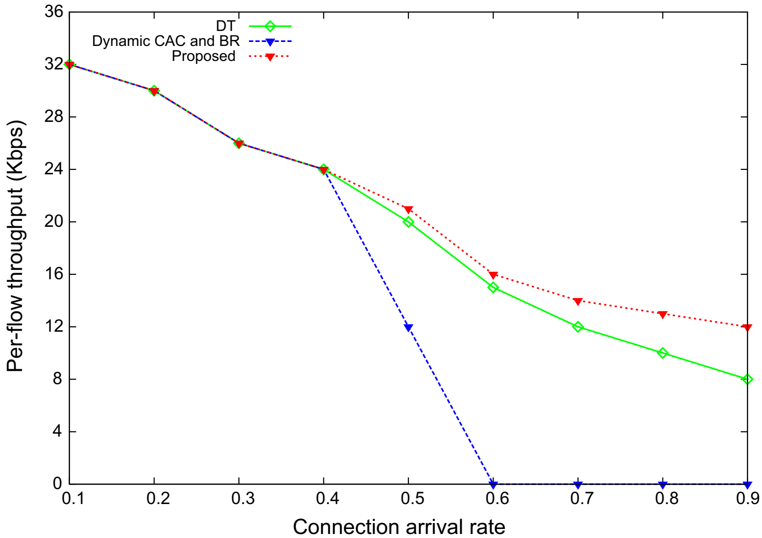


Fig. 5 Throughput of UGS traffic class with different CAC schemes

occurrence, with which the arrival rates of new connections and handoff connections are λ_n and λ_h, respectively. The ratio of handoff arrival to new connection arrival rates is considered to be 1:1.

In the first experiment, the impact of an equal traffic ratio between new and handoff connections is evaluated.

Figure 6 shows the blocking rates of new connections with the traffic rates of 1:1 for DT CAC, dynamic CAC and BR, and the proposed scheme. It shows that from 0.05 to 0.15 traffic arrival parameter, the schemes have similar performance because of the low traffic load, but

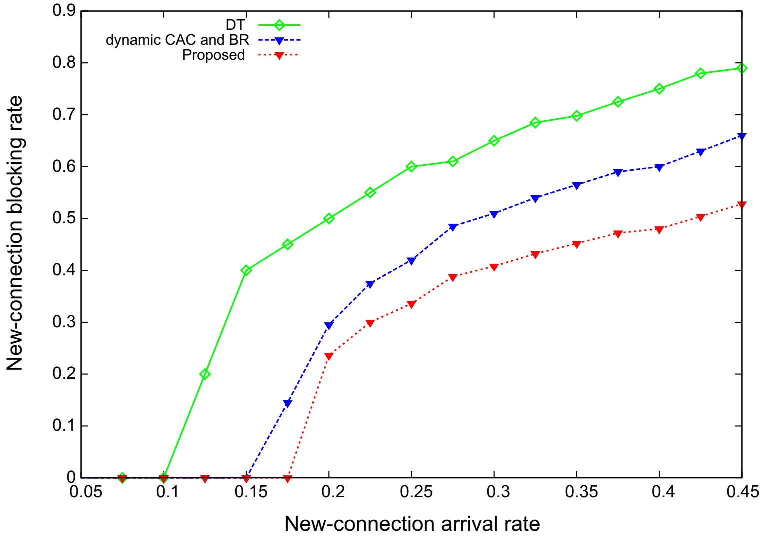


Fig. 6 Blocking rates of new connections with the traffic rates of 1:1

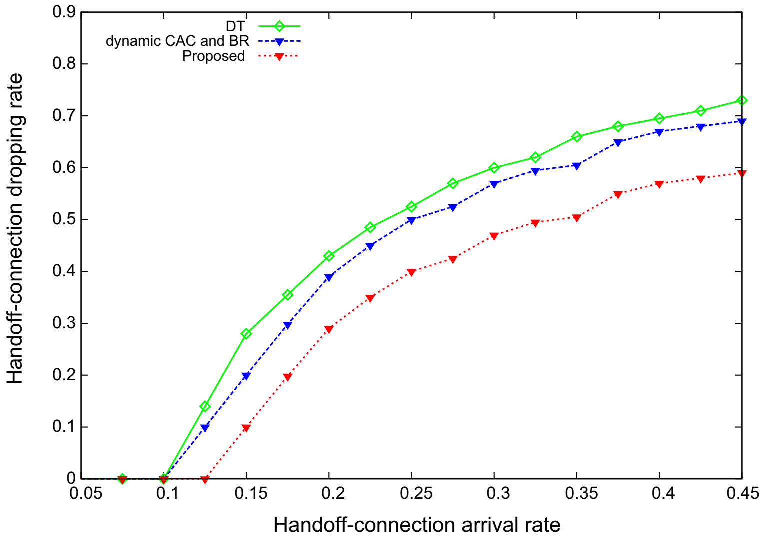


Fig. 7 Dropping rates of handoff connections with the traffic rates of 1:1

as the intensity of the traffic increases, the proposed scheme performs better than the other two schemes. The proposed scheme reduces the new connection blocking rate by 45.92 % compared to DT and 22 % compared to dynamic CAC and BR. The increase in network efficiency is attributed to the new admission criteria introduced, that prevent the starvation of the UGS and BE service classes, and avoid reserving too much resources for handoff connections that remain unused as a result of the fixed maximum bandwidth threshold, but are now used to admit more new connections because of the appropriate computation of an

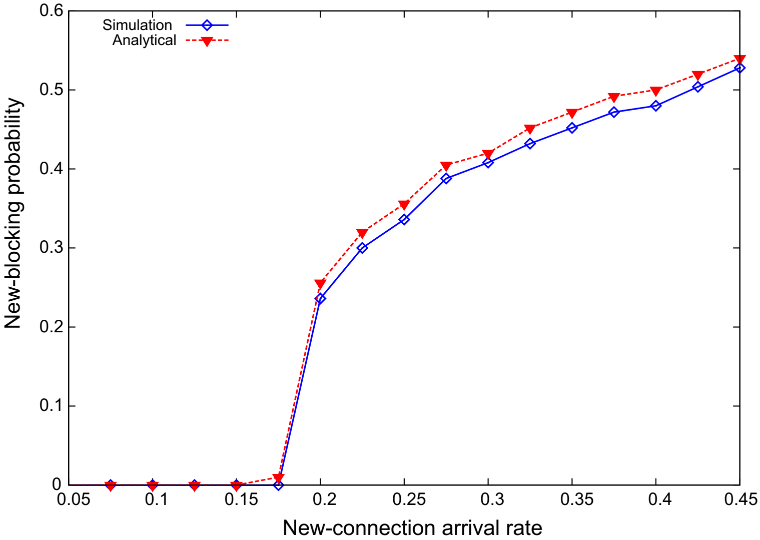


Fig. 8 Analytical and simulation results for new connection blocking probability

adaptive threshold according to the traffic intensity. It also allows more new connections to be accepted into the network during degradation mode.

Figure 7 shows the dropping rates of handoff connections with the traffic rates of 1:1 for DT CAC, dynamic CAC and BR, and the proposed scheme. The figure shows that the proposed scheme has a similar performance to the others when the arrival rate is within the range 0–0.125. Based on this range, the traffic intensity is considered to be low. The figure shows that within the range 0.125–0.45, the proposed scheme has superior performance than the other schemes. The proposed scheme has reduced the connection dropping rate by 26.16 % as compared to the DT scheme, and by 20.22 % from that of the dynamic CAC and BR scheme. The reason for this is that during the frequent arrival of handoff connections the proposed scheme appropriately adjust an adaptive reserved bandwidth threshold for handoffs according to the traffic intensity of the handoff connections. Unlike the other two schemes that adjusted an adaptive reserved bandwidth threshold for handoffs based on arrival of new and handoff connections as well as the departure of handoff connections. It is also attributed to the new admission criterion introduced, which ensure the QoS requirements of each class of service.

In Figs. 6 and 7, the results show the effectiveness of the proposed approach in terms of increase in the network efficiency by accepting more connections into the network, and also assure QoS guarantee for admitted new and handoff connections.

6.3 The Comparison of Numerical and Simulation Results

To verify the analytical model, the numerical results are compared with the simulation results already obtained in Figs. 6 and 7. As illustrated in Figs. 8 and 9, the numerical results conform to the simulation results with only small deviations. In Fig. 8 there is only a 4.56 % deviation, while Fig. 9 has 3.0 %.

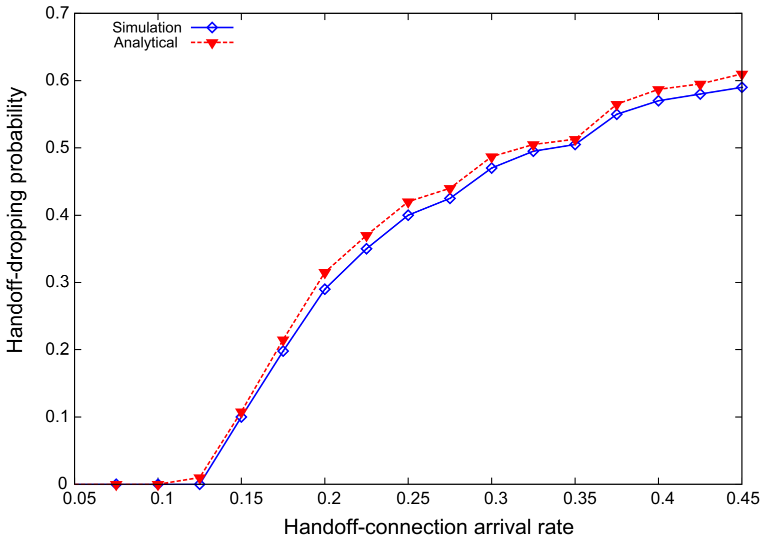


Fig. 9 Analytical and simulation results for handoff connection dropping probability

7 Conclusion

In this paper, a new CAC scheme called QoS-Aware connection admission control with bandwidth reservation and degradation is proposed to address the starvation problem and enhance the efficient utilization of network resources in IEEE 802.16e networks. The new admission criteria has been determined based on the scheduling service class in order to prevent starvation of service classes. The criteria used bandwidth degradation to admit more users into the network when there is no available resources to accommodate a new user. The adaptive threshold has been introduced to adjust the reserved bandwidth for handoff connections using the handoff traffic intensity to improve the efficient utilization of network resources. A number of simulation experiments were conducted to evaluate its performance. The simulation results have illustrated that the proposed scheme significantly outperform the other schemes in terms of accommodating more connections into the network and assuring QoS to all the service classes. Finally, an analytical model is also developed based on new-connection blocking probability and handoff-dropping probability to validate the scheme. The accuracy of the model has been validated by comparing it with a simulation result. The numerical results show a close relation to the simulation results with an insignificant variation.

References

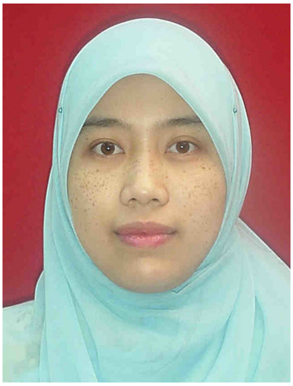
1. IEEE 802.16 Working Group. (2005). IEEE standard for local and metropolitan area networks. Part 16: Air interface for fixed broadband wireless access systems. Amendment 2: Physical and medium access control layer for combined fixed and mobile operation in licensed bands. IEEE Std 802.
2. Borin, J. F., & da Fonseca, N. L. S. (2008). Simulator for WiMAX networks. *Simulation Modelling Practice and Theory*, *16*(7), 817–833.
3. Othman, Jalel Ben, Mokdad, Lynda, & Ghazal, Sahar. (2014). Performance analysis of WiMAX networks AC. *Wireless Personal Communications*, *74*(1), 133–146.

4. Naghshbandi, D., Subramaniam, S., & Moghadam, S. R. (2013). An event-driven simulation for highest urgency first (HUF): A latency and modulation aware bandwidth allocation algorithm for WiMAX base stations. *Wireless Personal Communications*, *69*(1), 39–52.
5. Saidu, I., Subramaniam, S., Jaafar, A., & Zukarnain, Z. A. (2015). *An efficient battery lifetime aware power saving (EBLAPS) mechanism in IEEE 802.16 e networks, wireless personal communications* (pp. 1–21). Berlin: Springer.
6. Lu, J., & Ma, M. (2011). Cross-layer QoS support framework and holistic opportunistic scheduling for QoS in single carrier WiMAX system. *Journal of Network and Computer Applications*, *34*(2), 765–773.
7. Ali, Mohammed A., Borhanuddin, Mohd A., Nor, Kamariah N., & Hafizal, Mohamad. (2013). Fair uplink bandwidth allocation and latency guarantee for mobile WiMAX using fuzzy adaptive deficit round robin. *Journal of Network and Computer Applications*, *39*, 17–25.
8. Yang, S.-F., Wu, J.-S., & Hwang, B.-J. (2013). Performance evaluation of priority based adaptive multi-guard channel call admission control for multiclass services in mobile networks. *International Journal of Communication Systems*, *26*(5), 597–609.
9. Wang, H., Li, W. & Agrawal, D.P. (2005). Dynamic admission control and QoS for 802.16 wireless MAN. In *Wireless Telecommunications Symposium, 2005* (pp. 60–66) IEEE.
10. Hou, F., Ho, P.-H. & Shen, X. (2006). WLC17-1: Performance analysis of a reservation based connection admission scheme in 802.16 networks. In *Global Telecommunications Conference, 2006. GLOBECOM'06. IEEE* (pp. 1–5), IEEE.
11. Wongthavarawat, K., & Ganz, A. (2003). Packet scheduling for QoS support in IEEE 802.16 broadband wireless access systems. *International Journal of Communication Systems*, *16*(1), 81–96.
12. Jiang, C-H. & Tsai, T-C. (2006). Token bucket based CAC and packet scheduling for IEEE 802.16 broadband wireless access networks. In *Consumer Communications and Networking Conference, 2006. CCNC 2006. 3rd IEEE* (Vol. 1, pp. 183–187), IEEE.
13. Ge, Y. & Kuo, G-S. (2006). An efficient admission control scheme for adaptive multimedia services in IEEE 802.16 e networks. In *Vehicular Technology Conference, 2006. VTC-2006 Fall. 2006 IEEE 64th* (pp. 1–5), IEEE.
14. Wang, L., Liu, F., Ji, Y., & Ruangchaijatupon, N. (2007). Admission control for non-preprovisioned service flow in wireless metropolitan area networks. In *Universal Multiservice Networks, 2007. ECUMN'07. Fourth European Conference on* (pp. 243–249), IEEE.
15. Chaudhry, S.B. & Guha, R.K. (2007). Adaptive connection admission control and packet scheduling for QoS provisioning in mobile WiMAX. In *Signal Processing and Communications, 2007. ICSPC. IEEE International Conference on, 2007* (pp. 1355–1358), IEEE.
16. Wang, C., Yan, W.-J., & Lo, H.-K. (2012). Dynamic admission control and bandwidth reservation for IEEE 802.16 e mobile WiMAX networks. *EURASIP Journal on Wireless Communications and Networking*, *1*, 1–20.
17. WiMAX, Forum. (2006). *Mobile WiMAX—Part I: A technical overview and performance evaluation* (pp. 9–24). Mobile, WiMAX Forum: WiMAX-Part.
18. Antonopoulos, A., Skianis, C. & Verikoukis, C. (2010) Traffic-aware connection admission control scheme for broadband mobile systems. In *Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE* (pp. 1–5), IEEE.

Ibrahim Saidu received the B.Sc. and M.Sc. degrees in Mathematics from Usmanu Danfodiyo University Sokoto, Nigeria, in 1998 and Bayero University Kano, Nigeria, in 2009, respectively. He also received the Postgraduate Diploma in Computer Science and the M.S. degree in Computer Science from Federal University of Technology Minna, Nigeria, in 2003 and Universiti Putra Malaysia(UPM), in 2010, respectively. Currently, he is working towards the Ph.D. degree at UPM. His research interests include Performance Evaluation, Scheduling Algorithm and Call Admission Control in WiMAX networks.

Shamala Subramaniam received the B.S. degree in Computer Science from University Putra Malaysia (UPM), in 1996, M.S. (UPM), in 1999, Ph.D. (UPM) in 2002. Her research interests are Computer Networks, Simulation and Modeling, Scheduling and Real Time System.

Azmi Jaafar is an associate professor at the Department of Software Engineering and Information Systems, Universiti Putra Malaysia. He had a Ph.D (Mathematical Programming), M.Sc and B.Sc (Computer Science and Mathematics) from Indiana State University, USA. His research interest includes Empirical Studies, Knowledge Systems, Cryptography and Mathematical Programming.



Zuriati Ahmad Zukarnain is an associate professor at the Faculty of Computer Science and information Technology, University Putra Malaysia. She is the head for high performance computing section at Institute for Mathematics and Research (INSPEM), University Putra Malaysia. She received her PhD from the University of Bradford, UK. Her research interests include: Efficient multiparty QKD protocol for classical network and cloud, load balancing in the wireless ad hoc network, quantum processor unit for quantum computer, Authentication Time of IEEE 802.15.4 with Multiple-key Protocol, Intra-domain Mobility Handling Scheme for Wireless Networks, Efficiency and Fairness for new AIMD Algorithms and A Kernel model to improve the computation speedup and workload performance. She has been actively involved as a member of the editorial board for some international peer-reviewed and cited journals. Dr. Zuriati is currently undertaking some national funded projects on QKD protocol for cloud environment as well as routing and load balancing in the wireless ad hoc network.