A HYBRID CAC SCHEME FOR EFFECTIVE RESOURCE ALLOCATION IN IEEE 802.16 WIRELESS NETWORKS

CHANDRASEKARAN V\textsuperscript{1}, NAGARAJAN N\textsuperscript{2},
\textsuperscript{1}\textsuperscript{2}Anna University, Chennai, India
\textsuperscript{1,2}Network System Design Laboratory,
Coimbatore Institute of Engineering and Technology, Coimbatore, TN, INDIA
E-mail: vccpt@yahoo.in, swekalnag@gmail.com

ABSTRACT

Call admission control (CAC) plays a vital role in providing the desired Quality of Service (QoS) in broadband wireless networks. The aim of this paper is to propose an efficient Greedy Shaper based Call Admission Control (GSCAC) scheme for WiMAX networks to manage the number of admitted user connections and maximize the utilization of the wireless resources. The proposed algorithm is based on bandwidth reservation and greedy shaper concepts. It has been developed considering the problem of “busy hour” in communications traffic variation during a typical day. The success of the proposed scheme depends on the revenue it generates to the service providers and service quality it provides to its subscribers. The proposed scheme is evaluated and compared to a traditional CAC scheme using OPNET simulator.

Key words: IEEE 802.16; CAC; Throughput; QOS; Bandwidth Utilization; Greedy Shaper.

1. INTRODUCTION

The WiMAX is based on IEEE 802.16 wireless Metropolitan Area Network standard, which focuses on solving the problems associated with point-to-multipoint broadband outdoor wireless networks [1],[2],[3]. The IEEE 802.16 technology (WiMAX) continues to gain momentum as a promising alternative to 3G or wireless LAN to provide last-mile internet access because of its large coverage area, low cost of deployment and high speed data rates. The IEEE 802.16 technology will likely be the first available alternative to fourth generation (4G) networks but they are not synonymous [1],[2],[3],[4],[5]. The high data rates offered by the WiMAX networks make it very popular among mobile data network operators and allows them to stream new converged services such as audio and video streaming, mobile Internet browsing and Voice over IP (VoIP), to name a few. To achieve such objectives, multicarrier transmission based on Orthogonal Frequency-division Multiplexing (OFDM) is used and is combined with a connection-oriented approach at the Medium Access Control (MAC) layer. Such connections, or service flows, are dynamically established using a three-way handshake protocol called the Dynamic Service Addition (DSA) protocol [6]. The services necessitate the support of different classes of traffic with different QoS requirements, which need to be guaranteed by the wireless network.

The IEEE 802.16 technology aims to support maximum number of Mobile Stations (MS) in the network. The total available bandwidth or spectrum is divided into number of channels and they are shared among the mobile stations. The shared channels, unlike circuit switching facilitate the support for higher number of MS, hence improving the performance utilization of the network [7]. QoS provisioning can be carried out at three levels in the IEEE 802.16 systems, namely admission level, class level, and packet level 1. Admission - level QoS provisioning is typically realised by employing a proper Call Admission Control procedure. The CAC algorithm is responsible for accepting or rejecting a new call request by the MS. The CAC aims at maximizing the number of admitted users, satisfying the QoS requirements of the existing MS at the same time. The new MS are given access only if there is enough bandwidth in the network to meet the QoS requirements of the new calls.

CAC has been extensively studied in the literature [8], [9], [10], [11], [12] and [13]. The Call Admission Control schemes in mobile WiMAX systems can be broadly classified into distributed schemes [8], [9] and non-distributed schemes[10], [11], [12], [13]. The CAC employing the
distributed schemes consider status information of other base stations apart from the base station to which the connection requests are made. The non-distributed schemes interact with only with single base stations. The distributed schemes and the non-distributed schemes proposed so far make their admission decisions based on the actual current network traffic load. A scheduler with Call Admission Control based on Latency-Rate server theory and with system characteristics as specified by the system standard using the Wireless MAN-OFDM air interface was proposed in [14]. A ring-based CAC model is proposed in [15]. The CAC algorithm proposed in [15] makes decisions based on mobility-related parameters such as call dropping probability. [16] Proposes an analytical model to analyze the performance of the Dynamic Service Addition protocol in terms of signal blocking, admission control blocking, and latency in an IEEE 802.16 network. The analytical model based on queuing theory is combined with the quality estimation of the OFDM-based physical layer, using a cross-layer approach. A CAC algorithm based on traffic modeling is introduced in [17]. The algorithm proposed in [17] is proposed to handle bursty traffic. Existing CAC schemes have shown to be very efficient in improving the packet-level QoS of ongoing connections amidst congestion periods. However, the proposed schemes are not very efficient in improving admission-level QoS. This is because these schemes do not have any provisions to provide incentives to users to share the system resources rationally and efficiently. Therefore, the connection blocking and dropping probabilities can reach high levels during peak hour traffic. Hence an efficient call admission scheme that can efficiently allocate wireless system resources especially during peak hour traffic is proposed in this paper.

In this paper, we introduce a GSCAC algorithm for IEEE 802.16 broadband wireless access standard to handle Peak hour traffic. The algorithm is based on bandwidth reservation and greedy shaper concept with dynamic pricing concept to support “busy hour traffic” variation during a typical day. IEEE defines the “busy hour” as “the uninterrupted period of 60 minutes during the day when the traffic offered is the maximum” [18]. The problem becomes more intense if we take under consideration the variation of daily traffic volume, where there is a peak during the “busy hour”. The algorithm is implemented to reduce peak load traffic and maximize support for UGS services. Recent studies have shown that the proportion of VoIP users will continue to grow tremendously [19]. Due to that fact, our proposed CAC focuses in UGS flows, giving them higher priority comparing to the other three types of flows of the standard. During peak load all UGS calls which are below or equal to SLA are supported. If UGS calls exceed SLA they are served but they are charged higher. So only the customers who want to use the service will go ahead. Others may opt to divert when the load is low. Care should be taken that the UGS calls do not eat all the resources. Hence, the delay and vertical deviation experienced by the other services are continuously monitored. If the delay and vertical deviation of the ertPS and rtPS schemes exceed a threshold then in the next time frame the algorithm may not entertain some of the UGS calls from the customer who have exceeded the SLA and will start serving other services i.e., only a portion of all the services that have exceeded the SLA will be entertained but this condition rarely happens because the customers will intend to divert the traffic during non peak hours.

2. GSCAC ALGORITHM

Initially the CAC scheme follows statistical assignment of bandwidth. The UGS service schemes are assigned about 40 percent of the total bandwidth. 50 percent of the bandwidth resources are allocated to other QoS guaranteed services namely ertPS, rtPS, and nrtPS and the remaining 10 percent is allocated to Best Effort Services. Service provider does not charge different revenue rates for streaming different service classes as long as the algorithm remains in the statistical assignment mode as the network is underutilized in this mode. But with varying traffic conditions the CAC algorithm is expected to readjust the bandwidth assignment procedure for better bandwidth management.

The CAC algorithm follows the statistical assignment strategy as long as there is no starvation for resources but when starvation of resources happens i.e., when the number of required slots for a particular service supporting QoS guaranteed services exceeds the number of pre allotted slots for five continuous windows, then the CAC jumps to a dynamic mode. In the dynamic mode there is no fixed pre allocation of resources. However, similar to the static mode, higher priority is given to streaming UGS and ertPS service flows compared to rest of the flows with the highest priority assigned to UGS service flows. This may penalize the other service flows and prevent the MS from accessing the BS at appropriate time to complete
the task. So, the bandwidth allocation process carried out by the GSCAC algorithm should ensure a fair resource allocation process to all the service class schemes.

As the CAC algorithm detects an increase in demand of the number of required slots for a particular service for five continuous windows the algorithm moves to the dynamic mode. The CAC in the dynamic mode is better explained after introducing the necessary parameters and definitions.

We consider ‘Ki’ classes of traffic, class ‘i’ has priority than i + 1. Each class includes a number of services, where service ‘x’ of class ‘i’ requires bw i x units of bandwidth and service ‘x’ has higher priority than x + 1. The same class can stream audio and video each requesting different units of bandwidth. The Monitoring component in the proposed algorithm is very simple. The main objective of the algorithm is to trigger the dynamic CAC component when a change in the available bandwidth of next time window is detected.

Let I i represent the next time window. LT gives rate in units of time of next time window. TS i indicates the total numbers of services in class ‘i’ and T i is the number of admitted user that request service ‘x’ in class ‘i’. The total number of users admitted in class class ‘i’ is given by (1)

\[ T_i = \sum_{x=1}^{TS_i} T_i^x \] (1)

Let S indicate the system capacity and bw i x represents the bandwidth request per unit time of service ‘x’ in class ‘i’. BWav is the total available bandwidth in next time window. \( \theta_i^x \) represents the number of connection requests for service ‘x’ in class ‘i’ in next time window.

The total demand for bandwidth by class ‘i’ in the next time window is given by

\[ \sum_{x=1}^{T_i} T_i^x \cdot \theta_i^x \cdot LT + \sum_{x=1}^{T_i} (\theta_i^x)(bw_i^x \cdot LT) \] (2)

Where \( \sum_{x=1}^{T_i} T_i^x \cdot LT \) is the demand of class ‘i’ already admitted user connections and \( \sum_{x=1}^{T_i} (\theta_i^x)(bw_i^x \cdot LT) \) is the maximum demand of new incoming users provided that they are admitted to the system.

\( \theta_{total} \) is the total number of users who could make connection requests in the next time window and \( \theta_{total} \) is equal to the total number of admitted users subtracted from the total number of users that could make connection requests at the cell, where dynamic pricing is implemented. \( r_i^x \) gives rate in terms of units of money per unit of bandwidth and \( W_i^x \) the rate of users who have sufficient willingness to pay, \( W_i^x \) is a function of price given by \( W_i^x = \frac{f_i^x}{r_i^x} \)

The pricing scheme works well with any demand model. We have evaluated the pricing scheme based on well known demand model [20], [21]. The main objective of the CAC component is to find the optimal number of connection requests for each service in each class in the next time window so that the utilization of available bandwidth is maximized. To achieve this objective, the CAC component should ensure that the next maximum demand for next time window does not exceed the total available bandwidth which is given by (3)

\[ \max \{\theta_i\}_{i=1}^{K_i} \sum_{i=1}^{K_i} \sum_{x=1}^{T_i} \theta_i^x \cdot (bw_i^x \cdot LT) \] (3)

The CAC should also ensure that the total number of connection requests does not exceed the total number of subscribers given by (4) and (5)

\[ \sum_{i=1}^{K_i} \sum_{x=1}^{T_i} \theta_i^x \cdot (I_i) \cdot (bw_i^x \cdot LT) \leq BW_{av} \] (4)

\[ \sum_{i=1}^{K_i} \sum_{x=1}^{T_i} \theta_i^x \leq \theta_{total} \] (5)

Where \( \sum_{i=1}^{K_i} \sum_{x=1}^{T_i} \theta_i^x \) is the total number of connection requests to the system. The optimal number of connection requests can be estimated using ICP.

During peak hours, all classes of traffic should be entertained, but highest priority is given to UGS scheme followed by eRTS, rRTS, nRTS, BE. During peak load all calls which are below or equal to SLA are supported. During peak hour, if the calls from a customer exceed the SLA then first the UGS calls are entertained. This is done because the UGS applications are sensitive to delay. The only constraint is that the UGS bandwidth should not exceed the total available bandwidth \( BW_{av} \) in the next time window. However, the UGS calls that exceed the SLA are charged higher. So only the customers who want to use the service will go ahead. Others may opt to transmit when the load is low. Care should be taken that the UGS calls do not eat all the resources and in order to ensure that the algorithm invokes the greedy shaper.

The greedy shaper estimates the delay and vertical deviation experienced by other service classes. If the delay and vertical deviation of the eRTS and rRTS schemes exceed a threshold then in the next time frame the algorithm may not entertain some of the UGS calls from the customer who have exceeded the SLA and will start serving other services. But this condition rarely happens because the customers will intend to divert the traffic during non peak hours.
To estimate the delay experienced by the various service flows we use a EWMA estimator. We introduce the following notations

- \( i \) = differentiation parameter
- \( d_i(m) \) = queuing delay of the \( m^{th} \) packet in class 'i'
- \( w_i(m) \) = normalized head waiting time of class 'i' when \( m \) packets have departed
- \( g = \) constant
- \( \gamma_i = \) filtering coefficient for class 'i'
- \( s_i = \) mean packet size of class 'i'
- \( q_i = \) maximum queue size for class 'i'
- \( C = \) link capacity

The delay experienced by the particular service flow is given by (6)

\[
d_i(m) = \gamma_i d_i(m) + (1 - \gamma_i)d_i(m - 1) \tag{6}
\]

Separate filtering coefficient is employed for each class.

\[
\gamma_i(q_i) = \frac{1}{N * \sqrt{q_i} * \ln(q_i)}
\]

The Vertical deviation experienced by the packets from other services is given by

\[
v = \alpha(T) - \alpha(T - 1) \tag{7}
\]

Where \( \alpha(T) \) is the maximum total demand of bandwidth of a particular class 'i' in the next time window and \( \alpha(T - 1) \) is the Bandwidth allotted to the particular class in the previous time window.

If the delay and vertical deviation of other service flows that have exceeded the SLA is greater than the threshold, entertain only a percentage of calls from services that have exceeded SLA given by (8). The new scheme introduces a time factor \( \eta \), in which the user occupies the time slots by specifying the service after which more slots or ratio of slots are allotted for other service flows \( \frac{\eta N_2}{\eta N} \).

For each class 'i' in service, the above expression will vary. The user permissible limit for the requests is given by \( \eta_i \) where is the class 'i' over the time slots.

The percentage of traffic that is allocated to the UGS flow depends on the time factor and the number of admitted users for a class is given by

\[
\frac{N_2}{\eta} * s * 100 \tag{8}
\]

Where 's' is the weighting factor. The above equation gives the percentage of traffic entertained in a particular class, which helps to determine the traffic for rest of the class and service. The new design is again in such a way that the highest percentage of calls is entertained for UGS scheme followed by erTPS, rTPS, nrtPS and BE respectively [23],[24]. This is achieved by changing the weighting factor 's'. All the serviced calls exceeding the SLA will be charged higher.

The optimum price for service 'x' in class 'i' should be computed such that

\[
W_{i}^{x} = f_i^{x}(r_{i}^{x}) = \frac{\eta_i N}{\eta} * s \quad \forall x, 1 \leq x \leq TS_i.
\]

The proposed algorithm serves two purposes. First, the connection request for a particular service exceeds the SLA they are charged higher which would generate revenue for the service provider. Second, during network under utilization periods they are charged lower, thereby encouraging customers to transmit more during under utilization periods.

3. SIMULATION MODEL

Figure 1 illustrates the network model. The performance of the proposed GSCAC scheme is evaluated and compared with a Conventional CAC scheme for various class of service. The scheme was tested on WiMAX networks using OPNET modeler. The performance of the algorithm was implemented on a homogeneous system in statistical equilibrium where all the cells are statically assumed same as any other cell. The mean arrival and departure rates in each cell are same. Therefore, a cell can be decoupled from the rest of the system and the system performance can be evaluated based on the performance of the cell. The BS and MS attributes specified in the Suburban Fixed path loss model, as specified in Figure 2 and Figure 3, are applied to the signals received at the WiMAX Media Access Control (MAC). Terrain Type A, which is characterised by hills with moderate-to-heavy tree densities, is chosen, and a shadow fading term is introduced to account for site-specific departures from the generic path loss model due to obstructions in the signal path [25],[26],[27],[28]. OFDMA 2048-FFT mode, which specifies the number of sub channels available throughout the channel spectrum, is chosen for analysis of the algorithm. It is assumed that the sub channel separation is fixed for a given symbol duration and that an increase in the number of sub channels must correspond proportionally to an increase in the channel bandwidth, or, in other words, more slots are to be allotted as the result of an increase in channel bandwidth. Various applications have various bandwidth requirements, and thus the number of slots or sub channels allotted will vary based on the application. The performance of the algorithm is tested using various MSs, but the analysis presented in this paper involves results that are obtained for one node (mobile_node_15).
4. RESULTS AND DISCUSSIONS:

The scheme was evaluated using the following performance metrics.
1. System Throughput
2. Percentage of Bandwidth utilization
3. Connection Blocking probability

4.1. Average Throughput: It is the average number of successfully delivered bits over the lifetime of the user’s connection. In the first case, the average throughput of the various class of service for the GSCAC scheme is evaluated and compared with the Conventional CAC scheme. Figure 4 demonstrates the throughput of UGS scheme. It may be observed that the proposed algorithm maintains the throughput during the entire day (0600-1500 hours). It may also be observed that as the full peak load approaches, (during evening hours) the throughput rate drops in the conventional algorithm by approximately 20 percent compared to the algorithm presented in this paper. Similarly, Figure 5 illustrates that the throughput of rtPS scheme is improved by approximately 20 percent in the GSCAC algorithm during full peak period. Figure 6 demonstrates the throughput of rtPS scheme and it has been shown that the throughput of the GSCAC scheme is much higher than the conventional scheme. The rtPS scheme supports video calls and it may be noted that the conventional CAC scheme exploited the resources of BE services to provide enhanced throughput of other services like UGS, rtPS and rtPS schemes as illustrated in Figure 7, whereas, the GSCAC scheme achieves better and constant throughput for UGS, rtPS and rtPS in spite of improving the throughput of BE service by 100 percent.
4.2 Percentage Of Bandwidth Utilization: It is defined as the percentage of the utilized bandwidth to the total bandwidth. As mentioned earlier, the bandwidth allocation is designed in such a way that the highest priority is assigned to UGS service flow followed by other service flows but care is taken that the other service flows are not penalized. Figure 8 illustrates that the GSCAC algorithm achieves much better bandwidth utilization than the conventional algorithm. The reason for the ineffective bandwidth utilization in the conventional algorithm is that, in the conventional algorithm the bandwidth is equally divided between the different services and usually UGS services request more channels than other services and this results in a higher bandwidth share for UGS scheme. Similarly, ertPS services will request for more bandwidth share than rtPS and rtPS will request more bandwidth share than BE. Hence higher service classes will drain the bandwidth of the network, thereby penalizing BE services. The GSCAC scheme, on the other hand, uses a dynamic bandwidth allocation procedure for the different services to achieve the maximum possible bandwidth utilization while maintaining a certain fairness level. Hence, the proposed scheme achieves better bandwidth than the conventional scheme as illustrated Figure 8.

4.3 Blocking Probability (%)

Connection blocking probability: The probability that a users’ connection is blocked due to insufficient bandwidth to meet his requirements. It may be noted that the call blocking probability in
UGS and rtPS services is zero. This is because the GSCAC algorithm ensures that all the high priority calls are entertained. The Connection blocking probability for BE service is just 15 percent and given that BE services are less sensitive to delay the connection blocking probability is totally acceptable. On the other hand the connection blocking probability for rtPS calls is also less than 5 percent (Figure 9) which is a significant achievement of our algorithm. In addition to the efficient bandwidth management achieved by GSCAC, the scheme can efficiently prevent network congestion, and therefore, achieving 0% blocking probabilities as shown in Figure 9. The results confirm the superiority of the proposed scheme compared to the conventional CAC scheme, where static provisioning schemes regulate the usage of the network. Such a scheme can result in very high blocking probabilities, which would lead to dissatisfaction of customers and revenue loss to the service provider.

Voice Environment

Voice Packet End to End Delay: The GSCAC outperforms the conventional schemes in terms of System Throughput, Percentage of Bandwidth utilization and Connection Blocking probability. However all these achievements should not be at the cost of compromising on the quality of the voice or video signals. The performance of the voice is analysed in this section. Figure 10 illustrates the voice end-to-end delay. The total voice packet delay, called "analog-to-analog" or "mouth-to-ear" delay = network_delay + encoding_delay + decoding_delay + compression_delay + decompression_delay. In mobile WiMAX, voice application is usually transmitted over the UGS service flow. Voice requires a reserved throughput and the inefficiency of the conventional CAC scheme to provide it increases the discrepancy between load and throughput which causes unacceptable packet end to end delays for voice as explained in Figure 10.

Figure 9. Connection Blocking Probability

Figure 10. Video Packet End to End Delay

Voice Jitter: If two consecutive packets leave the source node with time stamps t1 & t2 and are played back at the destination node at time t3 & t4, then: jitter = (t4 - t3) - (t2 - t1). Negative jitter indicates that the time difference between the packets at the destination node was less than that at the source node. Figure 11 illustrates the jitter experienced by voice for both environments. Jitter value has dropped to zero because the end to end delay has been reduced considerably.
MOS: Mean Opinion Score is a measurement of quality of the reconstructed voice signal. MOS is a subjective quality score that ranges from 1 (worst) to 5 (best). The MOS score on a VoIP network is further reduced when there is packet loss, excessive delays, etc. Figure 12 illustrates that the MOS value of the voice environment has been improved by changing the throughput of the UGS service class.

Web Browsing: Object response time: Specifies response time for each inclined object from the HTML page. It may be noted that the GSCAC algorithm provides BE service which is comparable to that of the conventional scheme (Figure 14) which is a significant achievement.
5. CONCLUSION

This paper presented a Greedy shaper based CAC scheme to manage the wireless resources efficiently and rationally, allowing efficient bandwidth management at the admission level. Simulation results show that the scheme can significantly improve the utilization of the wireless system and increase the revenues of network operator. In addition, the scheme can guarantee zero blocking probabilities. The proposed GSCAC scheme was evaluated using OPNET and as part of the future work we plan to implement the scheme in a real-time WiMAX switch and it would be amazing to see how the scheme performs in a real time WiMAX switch.

REFERENCES:


