ISSN: 2277-9655
Impact Factor: 1.852

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

Efficiency of Resource Allocation for Enhanced Quality of Service in MobileWiMAX Networks

L.Naha Iswari Kiruba[*1] , Dr. T.V.U. Kiran Kumar[2] Ms. S. Arulselvi[3]

[*1] Department of Electronics and Communication Engineering, Bharath University, Selaiyur, Chennai - 600 073, Tamilnadu, India

[2] Professor, & HOD, Department of Electronics and Communication Engineering, Bharath University, Selaiyur, Chennai - 600 073, Tamilnadu, India

[3]Asst. Prof, Department of Electronics and Communication Engineering, Bharath University, Selaiyur, Chennai - 600 073, Tamilnadu, India

kiruba.iswari@gmail.com

Abstract

The mobile wireless networks have to enable full mobility for users with non degraded Quality of Service (QoS). The QoS is closely related to a mobilityof users. Resource allocation is a vital component of call-admission control that determines the amount of resource to assign to new and handoff connections for quality-of-service (QoS) satisfaction. In this paper, we present approximate analytical formulations of virtual partitioning resource-allocation schemes for handling multiclass services with guard channels in a cellular system. Resource-allocation models for best effort and guarantee access with preemption for best effort traffic and virtual partition with preemption for all classes are investigated. The analytical models, derived using a -dimensional Markov chain, are solved using preemption rules for these schemes. Call-level grade of service, such as new-call-blocking probability, handoff-call-blocking probability, and system utilization, and packet-level QoS, such as packet-loss probability, are used as performance metrics.

Introduction

Broadband wireless access networks have rapidly been growing in these years to support the increasing demands of wireless multimedia services, like streaming audio/video, Internet Protocol TV, and video conferencing. Mobile Worldwide Interoperability for Microwave Access (WiMAX), which has been standardized by IEEE 802.16e [1], is one of the most promising solutions to provide ubiquitous wireless access with high data rates, high mobility, and wide coverage. The IEEE 802.16e Media Access Control (MAC) layer provides differential Quality of service (QoS) for various classes of scheduling services, which are Unsolicited Grant Service (UGS), Extended Real-Time Polling Service (ertPS), Real-Time Polling Service (rtPS), Non-real-time Polling Service (nrtPS), and Best Effort (BE). Each scheduling class is associated with a set of QoS parameters for quantifying its bandwidth requirement, e.g., maximum/minimum data rates and maximum delays. The radio resources (i.e., time slots and frequency spectrums) for different scheduling services are centrally controlled by the base station (BS). To provide QoS for data transmissions in WiMAX networks, BS generally applies a Connection Admission Control (CAC) scheme which determines whether a new connection should be established according to the available network resources. Essentially, the effectiveness of CAC schemes can be critical to both the performances of QoS for admitted connections and the utilization efficiency of network resources. However, the IEEE 802.16e standards do not specify how to implement CAC mechanisms and remain that as open issues.

On the other hand, a resource allocation mechanism is also important to the provisioning of QoS for some prioritized users like users in a handoff process. Handoff occurs when mobile station (MS) transfers its connection from the original serving BS with worse and worse link qualities to a neighboring BS with better qualities. In general, a handoff user will be prioritized over a new incoming user in order to provide better user-perceived satisfaction especially when it is with real-time applications which have specific QoS requirements, e.g., throughput demands and delay/jitter constraints. Since the reserved bandwidth cannot be taken by a new coming user, the design of BR mechanisms can significantly affect the performance of handoff QoS and also the utilization efficiency of network resources.

 ISSN: 2277-9655
Impact Factor: 1.852

The WiMAX Forum has defined a two-tiered mobility management: ASN Anchored Mobility and CSN Anchored Mobility:

The ASN Gateway supports connection and mobility management across cell sites and inter-service-provider network boundaries through processing of subscriber control and bearer data traffic between WiMAX networks and Code Division Multiple Access (CDMA), Universal Mobile Telecommunications Standard (UMTS), Wi-Fi, and femtocell access networks.

Connectivity Service Network (CSN) is a set of functions related to network offering IP services for connectivity to Wimax clients. A CSN may include network fundamentals such as AAA, server, routers, and user database and gateway devices that support validation for the devices, services and user. The Connectivity Service Network also handled different type of task such as management of IP addresses, support roaming between different NSPs, management of location, roaming, and mobility between ASNs. The WiMAX architecture is offering a flexible arrangement of functional entities when constructing the physical entities, Because AS may be molded into BTS, BSC, and an ASNGW, Which are equivalent to the GSM model of BSC, BTS and GPRS Support (SGSN)

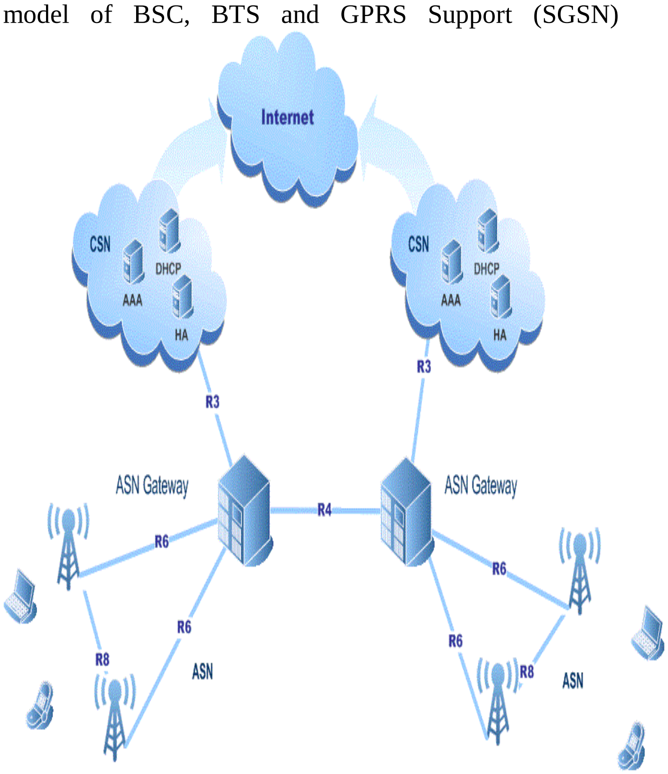

Fig. 1. Mobile Wimax Network Architechture

QOS ARCHITECTURE

The IEEE 802.16e MAC layer provides QoS differentiation for various categories of scheduling services. The IEEE802.16e uplink scheduling framework is shown in Fig. 2. The scheduling of uplink packet transmissions is centrally controlled in the BS. The IEEE 802.16e standards adopt a connection-oriented MAC protocol, i.e., each connection is associated with a connection ID. When a service flow generated at the application layer arrives at the MAC layer, the MS first sends a connection establishment request to the BS. The admission control mechanism at BS then estimates whether the remaining bandwidth can support the QoS requirements of new connections without violating existing users' QoS. If the connection request is accepted, the BS replies with a connection response which indicates the connection IDs for each direction of this connection. After the process of connection establishment is finished, the MS can issue a bandwidth request. The connection classifier then classifies the service data units into different scheduling classes according to their service flow identifier and connection identifier. The uplink bandwidth requests by users are performed on a per connection basis, whereas the BS grants bandwidth on a per subscriber station basis (GPSS). After the BS allocates a certain amount of bandwidth to each of the MSs, the packet scheduler at each MS will redistribute the bandwidth to the corresponding connection. By means of the connection-admission-control mechanism and request-grant bandwidth-allocation scheme, QoS for different scheduling classes can be guaranteed.

The IEEE 802.16e standard divides all service flows into five scheduling classes, each of which is associated with a set of QoS parameters for quantifying its bandwidth requirement. The five scheduling classes are described as follows.

UGS: UGS is designed to support real-time service flows with fixed-size packets generated at periodic intervals (i.e., constant bit rate--CBR), such as T1 services and voice-over-Internet-Protocol (VoIP) applications without silence suppression. This service can grant a fixed amount of bandwidth for CBR real-time applications without any requests.

rtPS: rtPS is designed to support real-time service flows with variable-size packets generated at periodic intervals (i.e., variable bit rate--VBR), such as Motion Pictures Experts Group (MPEG) video.

Based on a polling mechanism to request bandwidth periodically, this service can guarantee QoS such as the minimum data rate and maximum latency for VBR real-time applications.

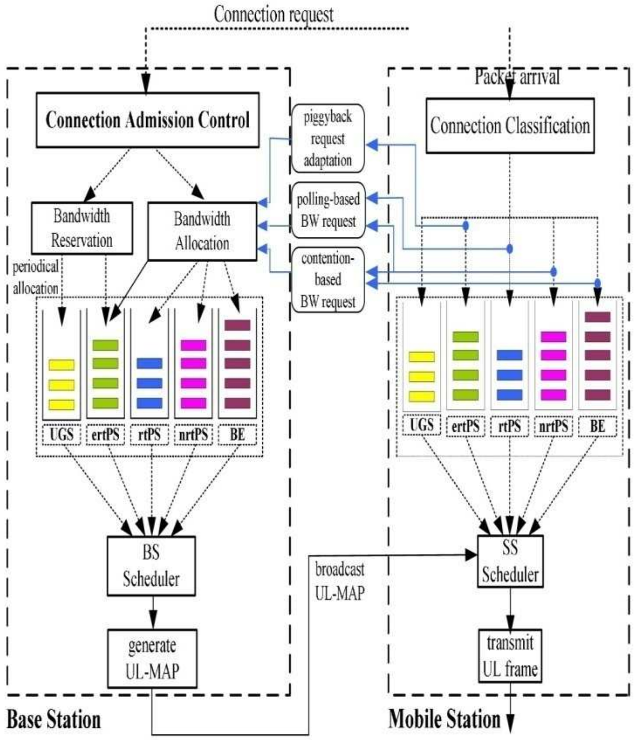

Fig. 2. Uplink scheduling Framework

ertPS: The characteristic of this service class is between UGS and rtPS. On detecting that the allocated bandwidth is either insufficient or excessive, ertPS can send a request to change the amount of allocated bandwidth like rtPS does. Otherwise, if the bandwidth demand remains unchanged, ertPS behaves as UGS. ertPS is designed to support VBR real-time data services such as VoIP applications with silence suppression.

nrtPS: This service class is to support non-real-time VBR services which require minimum-data-rate guarantees but can be tolerant to delay, such as File-Transfer-Protocol (FTP) applications.

BE: The BE service is designed for best-effort applications which have no explicit QoS requirements, e.g., web services or e-mail.

The QoS parameters and the supporting application types associated with each of the IEEE 802.16e scheduling classes are shown in Table 1.

QoS classes	**Applications**	**QoS parameters**
UGS	T1 services, VOIP without silence suppression	Max Rate Min Rate Jitter
rtPS	Video Streaming	Max Rate Min Rate Max Latency
ertPS	VOIP with silence suppression	Max Rate Min Rate Max Latency Jitter
nrtPS	FTP	Max Rate Min Rate
BE	Web browsing, e-mail	Max Rate

Table. 1. IEEE802.16e QoS classes

Resource Allocation

The IEEE 802.16e physical layer (PHY) adopts an Orthogonal Frequency Division Multiple Access (OFDMA) slot as the minimum possible resource. The IEEE 802.16e PHY supports Frequency Division Duplex (FDD) and Time Division Duplex (TDD) for bandwidth allocation mechanisms. In FDD mode, the uplink (UL) and downlink (DL) channels are located on split frequencies, with which a fixed duration frame is used for both UL and DL transmissions. In TDD mode, the UL and DL transmissions are arranged at different time periods using the same frequency. In this paper, we focus on the TDD mode for the IEEE 802.16e resource allocation mechanism.In TDD mode, Time Division Multiplexing (TDM) is used for DL transmissions and Time Division Multiple Access (TDMA) is used for UL transmissions. As shown in Fig. 3, a TDD frame has a fixed duration and contains one DL subframe and one UL subframe whose durations can adapt to the traffic loads of UL and DL transmissions. The DL subframe consists of a preamble, Frame Control Header (FCH), and a number of data bursts. The FCH specifies the profiles of the DL bursts that immediately follow it. The broadcast messages including downlink map (DL-MAP),

ISSN: 2277-9655
Impact Factor: 1.852

uplink map (UL-MAP), DL Channel Descriptor (DCD), UL Channel Descriptor (UCD), etc., are sent at the beginning of these DL bursts. The UL subframe contains a contention interval for initial ranging and bandwidth request and UL PHY protocol data units (PDUs) from different MSs. The DL connections are scheduled by BS in a broadcast manner, while the UL connections apply a request-grant mechanism for bandwidth allocation in a shared manner. The UL bandwidth requests are performed on a per connection basis, whereas the BS grants bandwidth on a per subscriber station basis (GPSS). After the BS allocates a certain amount of bandwidth to each of the MSs, each MS will redistribute the bandwidth to the corresponding connection. The information about bandwidth allocations for DL and UL transmissions is broadcast to the MSs through DL-MAP and UL-MAP messages at the beginning of each frame. Therefore, each MS can receive from and transmit data to BS in the predefined OFDMA slots.

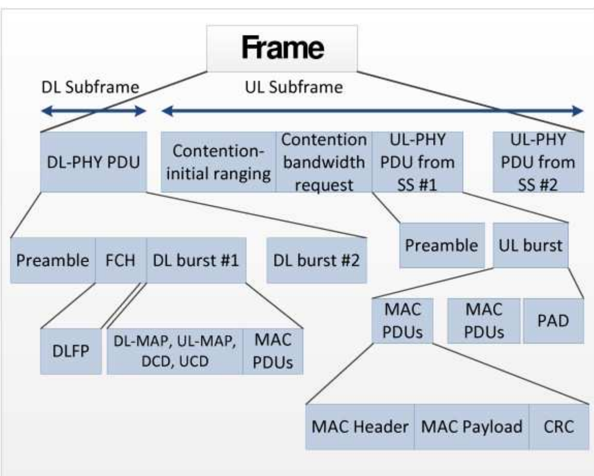

Fig. 3. TDD Frame Architechture

Admission Control (AC): is one of the resource management techniques to limit maximum amount of traffic in the network to guarantee service quality for subscribers. In wireless and mobile networks, the AC algorithms are much more complicated due to the movement of MSs. An MS served in current network maymove to another network. The connection of the MS may be dropped if the required resources in the target network cannot be supported. It is generally agreed that keeping an ongoing connection unbroken is more important than admitting a new MS. Therefore, a handover MS is given higher priority to access the network resources. For this purpose, the overall resources are partitioned and some resources are preserved for the handover MSs only. This is called priority-based AC. Call admission control (CAC) plays a significant role in providing the desired quality of service in wireless networks. Many CAC schemes have been proposed. Analytical results for some performance metrics such as call blocking probabilities are obtained under some specific assumptions. It is observed, however, that due to the mobility, some assumptions may not be valid, which is the case when the average values of channel holding times for new calls and handoff calls are not equal.

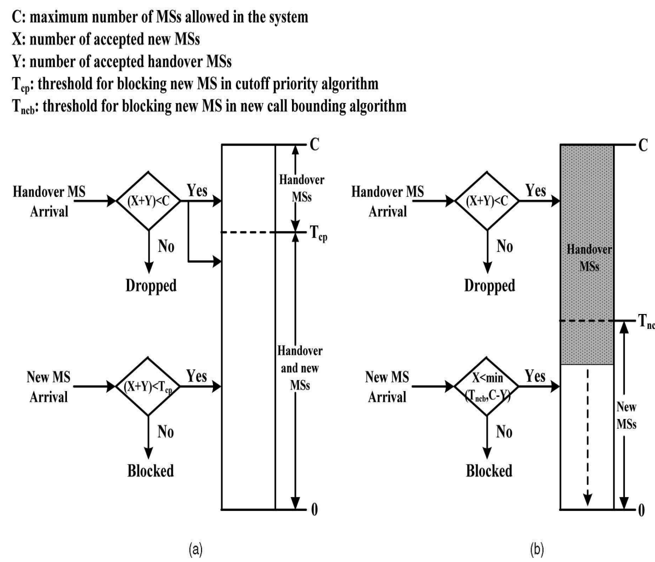

Fig. 4. Resource allocation. (a) Cutoff priority algorithm. (b) New call bounding algorithm.

Fig. 4. illustrates the resource allocation in the cutoff priority algorithm and new call bounding algorithm. In the cutoff priority algorithm, both new MS and handover MS can be admitted if the total number of new MSs and handover MSs in the network is equal to or less than a predefined threshold, Tcp, which is less than the total capacity C. Once the number of new MSs and handover MSs in the network reaches Tcp, new MSs are blocked. Only handover MSs are admitted. Once the total number of MSs exceeds C, handover MSs are dropped. In the new call bounding algorithm, there is a limit, Tncb, for the number of new MSs admitted into the network, which is also less than the total capacity C. The handover MSs use the resources in C -Tncb first. If the number of new MSs is less than Tncb, handover MSs can use more resources than C -Tncb.

However, the number of new MSs is always less than Tncb or the remainder resources the handover MSs have not used. This is shown as X <minðTncb; C _ Y Þ in Fig. 2b. To show the difference between the two algorithms, we assume C equals 50, and both Tcp and

Tncb are 30. We also assume in both algorithms, there are now 20 new MSs and
10 handover MSs. In the cutoff priority algorithm, a newly arrived MS will be blocked and a handover MS will be admitted. In the new call bounding algorithm, however, both a new MS and a handover MS will be admitted. There are still many other AC algorithms.
The ideas are similar although they may have different names. Nevertheless, they cannot be applied to WiMAX networks directly.
As aforementioned discussion, due to the specific mobility management techniques in WiMAX, an MS may be served by two ASN GWs simultaneously. Hence, the required resources of an Anchored MS are reserved in both ASN GWs. Besides, the Anchored MS will be counted twice in two ASN GWs in the AC algorithm. Thus, when many MSs are served by two ASN GWs in the system, a newly arrived MS or handover MS may be easily blocked or dropped by the AC algorithm. Without considering ASN GW relocation in the AC algorithm, the network performance will be degraded significantly.

Proposed Gateway Relocation Admission Control (GRAC)

The ASN GW relocation may be initiated at different times with different reasons. For example, as aforementioned discussion, an MS may perform ASN Anchored Mobility without performing CSN Anchored Mobility to reducehandover latency. After the handover is completed (i.e., the handover delay has been reduced), the MS may perform ASN GW relocation immediately so the number of Anchored MSs can be kept small. However, it may not be a good strategy always to relocate an Anchored MS so quick. For example, an MS may move fast and keep changing its Serving ASN GW. In this example, it might be better to keep the Anchored ASN GW unchanged. In some other examples, if the system load is light, there is no emergent need to perform ASN GW relocation. However, when more and more MSs are served by two ASN GWs, the system load will become heavy. New users may be blocked. Handover users may be dropped as well. The network performance may be reduced significantly. Therefore, performing ASN GW relocation is essential.

In WiMAX standards [4], [5], it is specified that ASN GW can decide when to perform ASN GW relocation. In this paper, we consider that the system load is heavy so Anchored MSs are forced to perform ASN GW relocation. The proposed GRAC determines when to request Anchored MSs to perform ASN GW relocation and how many Anchored MSs should be relocated. After ASN GW relocation, resources are released and system performance is improved.

Because WiMAX is based on all-IP network architecture, a variety of services, including voice and data services, can be deployed. Unlike voice traffic, data traffic tends to be bursty. Therefore, it is hard to estimate the resource required in an ASN GW to fulfill the requirements of the MSs the ASN GW is currently serving.
If the resource in one ASN GW is overprovisioned, the ASN GW may become a performance bottleneck.

Another approach is that the number of BSs controlled by each ASN GW can be scaled down to prevent the resource overprovision. However, because the number of BSs controlled by each ASN GW is reduced, this will cause many inter-ASN handovers. As a result, this approach will incur high cost. In [22], the authors discuss the flat mobile WiMAX network architecture. The paper shows that the resource management problem in the ASN GW has a great impact on the performance of WiMAX network architecture. Besides, in WiMAX, the AC algorithm can be deployed in each ASN GW to limit the maximum number of MSs to ensure network service quality. Our goal is to design a stand-alone algorithm such that each ASN GW can determine when to request Anchored MSs to perform ASN GW relocation. The corresponding parameters used in this section are listed in Table 2.

C	Maximum number of MSs in one ASN GW
T_{ncb}	Threshold for blocking a new MS
T_{wnr}	Threshold for carrying out WP-based prediction
$W(t)$	Number of MSs in one ASN GW at time t
$N_S(t)$	Number of serving MSs in one ASN GW at time t
$N_A(t)$	Number of anchored MSs in one ASN GW at time t
$N_H(t)$	Number of handover MSs in one ASN GW at time t
α	Standard normal random variable
Δt	Prediction time interval
τ	Sampling time interval
k	Number of latest samples
λ_n	Arrival rate of new MSs
λ_h	Arrival rate of handover MSs
$1/\mu_c$	Average connection holding time for new MSs
$1/\mu_n$	Average network residence time for new MSs and handover MSs
p^u_{nb}	Blocking probability of new MSs in upper-bound analysis
p^l_{nb}	Blocking probability of new MSs in lower-bound analysis
p^u_{hd}	Dropping probability of handover MSs in upper-bound analysis
p^l_{hd}	Dropping probability of handover MSs in lower-bound analysis
Θ_u	Average serving rate in upper-bound analysis
Θ_l	Average serving rate in lower-bound analysis
Λ_u	Average signaling overhead in upper-bound analysis
Λ_l	Average signaling overhead in lower-bound analysis

Table. 2. List Of Parameters

New Call Bounding AC with ASN GW Relocation

Basically, the overall resources are partitioned and some resource are preserved for the handover MSs only. The proposed GRAC can work with any AC algorithm. In this section, we simply pick up the new call bounding algorithm. For simplicity, here we assume that the resource assigned to each MS in one ASN GW is equal. The main point is not on a specific AC algorithm. The focus is on how to modify an AC algorithm for the two-tier mobility management in WiMAX.

The proposed GRAC with the new call bounding algorithm is presented in the Algorithm. In the Algorithm, we limit the number of Serving MSs and Anchored MSs in one ASN GW. As shown in Fig. 4, C is the maximum number of MSs in the network and Tncb is the limit for the number of new MSs, which have been admitted into the network. Let W(t) denote the total number of MSs in the ASN GW at time t. W(t)consists of NS(t), NA(t), and NH(t), which represent the number of Serving MSs, the number of Anchored MSs, and the number of handover MSs, respectively, at time t. As aforementioned discussion, a new MS admitted into the ASN GW is regarded as a Serving MS. After the MS performs inter-ASN handover to a neighboring ASN, the MS becomes an Anchored MS of the ASN GW. Thus, NA(t) is increased by 1 but NS(t)is decreased by 1

Algorithm: New call bounding AC with ASN GW Relocation.
Require: A new or handover MS is requesting to connect
with the ASN GW at time t.

<u>If a New MS arrives</u>
if Ns(t)+NA(t)<min(T*ncb*, C-NH(t)) **then**
NS(t)=NS(t)+1
else if Ns(t)+NA(t)=min(T*ncb*, C-NH(t)) **then**
if NA(t)>0 **then**
NA(t)=NA(t)-1
NS(t)=NS(t)+1
else
end if
end if

<u>If a Handover MS arrives</u>
if W(t)<C **then**
NH(t)=NH(t)+1
else
The handover MS is dropped.
end if
end if

To adapt the new call bounding algorithm into WiMAX networks, the algorithm is modified as:

If $N_s(t) + N_A(t) < \min(T'_{ncb}, C' - N_H(t))$ *and a new MS arrives; the new MS is accepted.*
where $T'_{ncb} \leq T_{ncb}$, $C' \leq C$.How to choose the value of and will be discussed later. However,
When $N_s(t) + N_A(t) = \min(T'_{ncb}, C' - N_H(t))$ *and* $N_A(t) > 0$,*one anchored MS is requested to perform ASN GW relocation*.

Because one Anchored MS is relocated, the new MS can be accepted. Otherwise, the new MS is blocked. Furthermore, if a handover MS arrives at time t, it is always accepted unless W(t) = C'

As aforementioned discussion, in this paper, we considerthat the system load is heavy. Therefore, Anchored MSs are forced to perform ASN GW relocation to accommodate new coming users. Based on this principle, we can set T'_{ncb} as T_{ncb} and C' as C. Thus, an Anchored MS is requested to perform ASN GW relocation only when no more resource is available for a new coming MS. The proposed GRAC does not limit the selection of other parameters for other conditions.

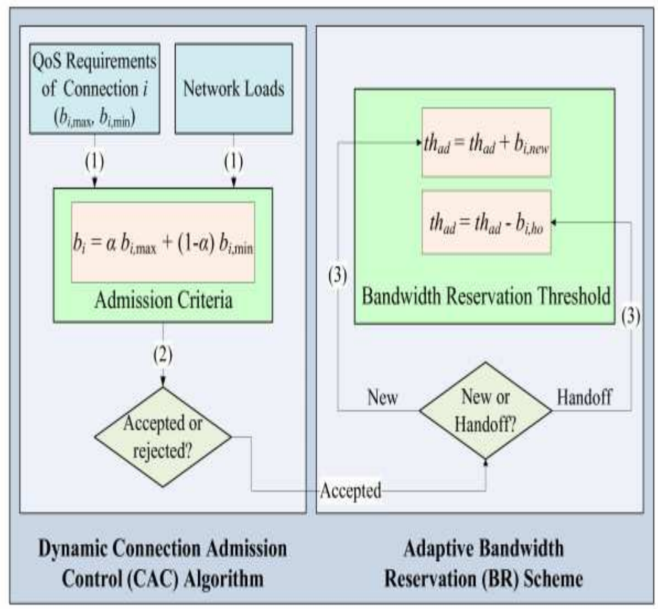

Fig.5. Proposed dynamic CAC algorithm and adaptive BR scheme

When a new connection or handoff connection arrives, it will inform the BS of its highest and lowest bandwidth requirements (i.e., $bi_{,\ max}$ and $bi_{,\ min}$). The proposed dynamic CAC scheme will adjust the admission criterion according to the currently estimated system capacity and network load. When the admission criterion is determined, the proposed adaptive BR scheme will accept or reject this handoff or new connection depending on the criterion .If a handoff connection is established, th_{ad} will be decreased with the amount of allocated resources, if a new connection is granted, th_{ad} will be increased with the amount of allocated resources.

ISSN: 2277-9655
Impact Factor: 1.852

The implementation of the proposed CAC and BR schemes in practice can involve the overheads as follows.

The estimate of system capacity**:** The system capacity can be evaluated at the BS with the specific PHY characteristics like channel spectrum, the amount of data sub-carriers, supported MCSs, used MIMO mechanisms, etc. The estimation of system capacity can be obtained in the initial phase of a network built-up.

The estimate of network loads**:** In general, the network loads can be evaluated at the BS with the information of currently adopted MCSs and the number of supported users with respect to each MCS. This part may need the exchanges of some context information between BS and SSs periodically, e.g., currently channel condition and used modulation.

The determination of admission criteria for incoming connections**:** When a connection arrives and requests for an admission, it will inform the BS of its specific QoS requirements, e.g., maximum and minimum data rates. Based on the estimated system capacity and network loads and QoS parameters, the BS will compute the admission criteria for the incoming connection with its specific QoS parameters.

The adaptation of BR for handoff connections**:** If a connection is admitted in the network, the BS will therefore adapt the BR threshold depending on the type of connection, i.e., new or handoff.

Performance Analysis

In this section, we propose an analytical model to investigate the performance of the proposed algorithm. In the analysis, the connection holding time is defined as the time from an MS connects to the network until it is disconnected. The network residence time is the time an MS is served by an ASN GW.

We assume each ASN GW has two arrival processes which are Poisson distributed with rate λn and λh for new MSs and handover MSs, respectively. If a new MS is admitted into the network, we assume the connection holding time and network residence time follow exponential distribution with mean 1/μc and 1/μn, respectively. For a handover MS, only network residence time is required. It is also assumed to be exponentially distributed with mean 1/μn. The corresponding parameters are also listed in Table 2.

Dropping Probability:

Fig.6 illustrates the dropping probability of handover MSs when λn is varied from 0.01 (1/s) to 0.1 (1/s). As that in Section 5.1. we set λh = 0.04 (1/s) and 1/μn = 400 (s). When λn increases, i.e., there are more MSs in the system, the dropping probability increases too. The handover MS is dropped when C in the AC algorithm is reached. In the proposed GRAC, however, the WP-based prediction is sensitive to the variation of the samples. The Anchored MSs are requested to perform ASN GW relocation when the system is expected to be overloaded. Thus, the dropping probability of handover MSs is reduced significantly.

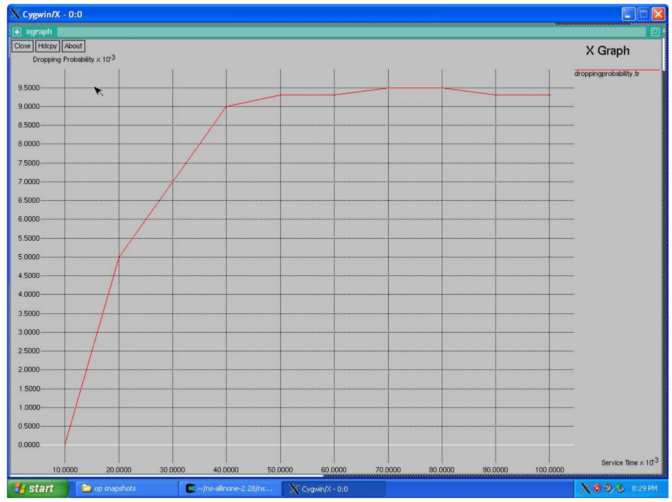

Fig. 6. Dropping probability versus new MS arrival rate

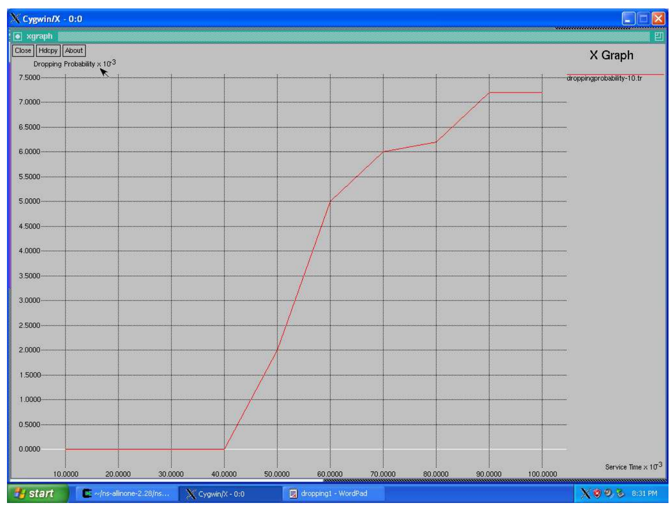

Fig. 7. Dropping probability versus network residence time

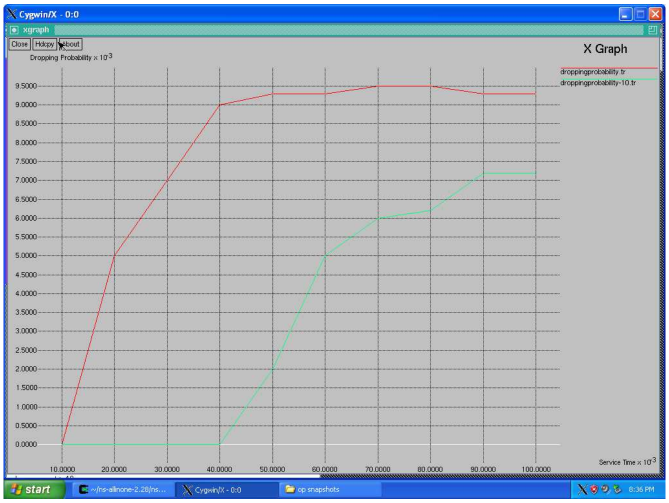

Fig. 8.Dropping probability versus new MSs arrival rate and network residence time

Fig. 9. Throughput versus new MS arrival rate.

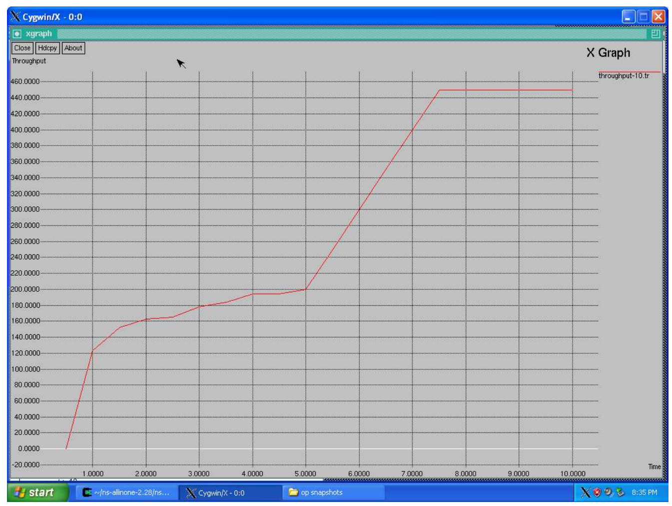

Fig. 10. Throughput versus network residence time.

Average Signaling Overhead:

Fig. 9. illustrates the average signaling overhead per minute versus λn, where λn is varied from 0.01 (1/s) to 0.1(1/s). We set λh =0.04 (1/s) and 1/µn= 400 (s). The amount of signaling traffic generated by executing CSN Anchored Mobility can be measured by the number of ASN GW relocation performed in the system. As shown in the figure, the signaling overhead of the upper-bound case is 0, because new MSs never perform ASN GW relocation in the upper-bound case. In the lower-bound case, the signalling overhead is increased when λn increases.The proposed GRAC can request ASN GW relocation only when the system is expected to be overloaded

Furthermore, we also investigate the average signalling overhead with different mean network residence time, 1/µn, as shown in Fig. 10. We still set λn=0.04 (1/s) and λh = 0.04 (1/s). Again, the signaling overhead of the upperbound case is 0. For the lower-bound case, when 1/µn is small, the signaling overhead is relatively high because the MSs are more likely to perform inter-ASN handover. However, regardless of the variation of 1/µn, the average signaling overhead of the proposed GRAC almost remains
constant.

Fig. 11. Throughput versus new MSs arrival rate and network residence time.

Conclusion

In WiMAX standards, an ASN GW can decide when to perform ASN GW relocation. In this paper, we consider that the system load is heavy, so Anchored MSs are forced to perform ASN GW relocation. We propose GRAC which considers admission control and ASN GW relocation jointly to improve the performance of WiMAX networks. The traditional AC algorithms cannot be used directly when thetwo-tiered mobility management is deployed in WiMAX because some MSs may be served by two ASN GWs. If there are many Anchored MSs, new incoming users will likely be rejected due to the lack of resources. In the proposed GRAC, the AC algorithm cooperates with the ASN GW relocation. When a new MS arrives and there is no resource for the newly arrived MS, the proposed GRAC will request an Anchored MS to perform ASN GW relocation.

References

[1] IEEE 802.16-2004 Std., Air Interface for Fixed Broadband Wireless Access Systems, IEEE, Oct. 2004.

[2] IEEE 802.16e-2005 Std., Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands, IEEE, Feb. 2006.

[3] WiMAX Forum, http://www.wimaxforum.org, 2011.

[4] WiMAX Forum Std. 1.0, Rev. 4, WiMAX Forum Network Architecture (Stage 2: Architecture Tenets, Reference Model and Reference Points), WiMAX, Feb. 2009.

[5] WiMAX forum Std. 1.0, Rev. 4, WiMAX Forum Network Architecture(Stage 3: Detailed Protocols and Procedures), WiMAX, Feb. 2009.

[6] L. Nuaymi, WiMAX: Technology for Broadband Wireless Access. John Wiley, 2007.

[7] K. Etemad, “Overview of Mobile WiMAX Technology and Evolution,” IEEE Comm. Mag., vol. 46, no. 10, pp. 31-40, Oct. 2008.

[8] Z.-H. Liu, S.-Y. Pan, and J.-C. Chen, “Access Service Network (ASN) Gateway Relocation Algorithms in WiMAX Networks,” Proc. IEEE Int’l Conf. Comm. (ICC ’08), pp. 2674-2679, May 2008.

[9] Y. Fang and Y. Zhang, “Call Admission Control Schemes and Performance Analysis in Wireless Mobile Networks,” IEEE Trans.Vehicular Technology, vol. 51, no. 2, pp. 371-382, Mar. 2002.

[10] D. Hong and S.S. Rappaport, “Traffic Model and Performance Analysis for Cellular Mobile Radio Telephone Systems with Prioritized and Nonprioritized Handoff Procedures,” IEEE Trans.Vehicular Technology, vol. 35, no. 3, pp. 77-92, Aug. 1986.

[11] Y.-B. Lin, S. Mohan, and A. Noerpel, “Queueing Priority Channel Assignment Strategies for PCS Hand-Off and Initial Access,” IEEEB Trans. Vehicular Technology, vol. 43, no. 3, pp. 704-712, Mar. 1994.

[12] B. Li, S.T. Chanson, and C. Lin, “Analysis of a Hybrid Cutoff Priority Scheme for Multiple Classes of Traffic in Multimedia Wireless Networks,” Wireless Networks, vol. 4, no. 4, pp. 279-290, 1998.

[13] R. Ramjee, D. Towsley, and R. Nagarajan, “On Optimal Call Admission Control in Cellular Networks,” Wireless Networks, vol. 3, no. 1, pp. 29-41, 1997.

[14] M.D. Kulavaratharasah and A.H. Aghvami, “Teletraffic Performance Evaluation of Microcellular Personal Communication Networks (PCN’s) with Prioritized Handoff Procedures,” IEEE Trans. Vehicular Technology, vol. 48, no. 1, pp. 137-152, Jan. 1999.

[15] R. Garg and H. Saran, “Fair Bandwidth Sharing Among Virtual Networks: A Capacity Resizing Approach,” Proc. IEEE INFOCOM, pp. 255-264, Mar. 2000.