

A Distributed QoS Guaranteed Scheme for IEEE 802.16 Broadband Wireless Access Systems

Jau-Yang Chang
Department of Computer Science and
Information Engineering
National Formosa University
Hu-Wei, Taiwan
jychang@nfu.edu.tw

ABSTRACT

This paper proposes a distributed quality of service (QoS) guaranteed scheme to guarantee QoS for multimedia services of connection users in IEEE 802.16 broadband wireless access (BWA) Systems. An environment of combination of the fixed wireless connections and mobile wireless connections is considered in this paper. In order to provide better QoS guarantees for connection users in the IEEE 802.16 BWA Systems, efficient QoS guaranteed schemes and corresponding algorithms must be developed in the base station (BS) and subscriber station (SS). According to the connection requests of service types, the QoS influence factors of connection-level and packet-level are analyzed in this paper. We utilize the concept of distributed load balance to share the overheads of BS and SS reasonably. Simulation results show that the proposed scheme is efficient for reducing the connection-blocking probability (CBP), connection-dropping probability (CDP), and packet delay rate.

Keywords

connection-blocking probability, connection-dropping probability, IEEE 802.16, multimedia, quality of service (QoS), packet delay rate.

1. INTRODUCTION

The IEEE 802.16 BWA, also refereed as WiMax (Worldwide Interoperability of Microwave Access), is designed to evolve as a suite of air interface for fixed, portable, and mobile wireless access systems. This standard is a promising technology to support high transmission rate and predefined QoS framework in broadband wireless networks. This technology can provide a cost-effective broadband access solution to use for connecting local area networks to the internet and to support mobile applications such as four generation mobile systems. One of the main design challenges in 802.16 networks is to deal with the resource allocation and traffic schedule. Four types of scheduling service classes are defined in the IEEE 802.16 MAC (Medium Access Control) protocol including Unsolicited Grant Service (UGS), real-time Polling Service (rtPS), non-real-time Polling Service (nrtPS), and Best Effort (BE) to support various types of traffic (e.g. constant

bit rate, real-time, nonreal-time, and best-effort). Since different types of traffic require different QoS performances, it is important to employ mechanisms for efficiently allocating available bandwidth and scheduling traffic to satisfy the QoS requirements. In this paper, we consider an environment of combination of fixed wireless connections and mobile wireless connections. This framework is one of the promising solutions for metropolitan mobile wireless network systems. However, these networks have the inherent problem of handoffs. This problem leads to higher CDP and results in bandwidth resource availability varying repeatedly. Frequent changes in the network traffic make the provision of guaranteed QoS more difficult. Hence, research in the area of QoS provisioning in the next-generation high-speed wireless networks focuses on the integration of resource allocation and admission control policies [1]-[6].

In [2], an admission control and bandwidth allocation mechanism is proposed to reduce CBP for IEEE 802.16 wireless MAN by using Gaussian model for aggregated traffic in large network and Chernoff bound method. However, the packet-level problems are not considered in this paper. In [3], a quality of service architecture for IEEE 802.16 standards is proposed to improve delay and bandwidth utilization. However, the connection-level problems are not considered in this paper. Moreover, these papers and relational papers are not dealt with QoS issues for an environment of combination of the fixed wireless connections and mobile wireless connection in IEEE 802.16 BWA Systems.

This paper introduces a novel distributed QoS guaranteed scheme to guarantee QoS for multimedia services of connection users in mobile wireless networks. We make a tradeoff between the connection-level (CBP and CDP) and packet-level (packet delay rate) QoS constraints such that the QoS can be maintained at an acceptable level. Efficient QoS guaranteed schemes and corresponding algorithms are designed in the BS and SS. Each SS makes an adaptive decision for bandwidth allocation. The amount of allocated bandwidth for each SS is dynamically adjusted, according to the on-line traffic information in the BS. Also, efficient packet scheduler mechanisms are developed in the SS and BS to reduce the packet delay rate for connections. This paper makes the system more responsive to the current traffic conditions, providing efficient bandwidth reservation for the UGS connections and handoff connections. The control of bandwidth reservation is adaptive under overloading traffic conditions, thus can effectively deal with sudden traffic surges. In order to provide the efficient bandwidth re-distribution in a BS, we employ a global borrowing mechanism to reduce the overhead of bandwidth reconfiguration and to satisfy QoS requirements of ongoing connections. Multimedia traffic is a combination of both real-time and nonreal-time traffic. Real-time traffic includes UGS and rtPS service types while nrtPS and BE service types make up nonreal-time traffic.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

QShine 2008, July 28-31, 2008, Hong Kong, Hong Kong.
Copyright 2008 ICST ISBN 978-963-9799-26-4
DOI 10.4108/ICST.QSHINE2008.3949

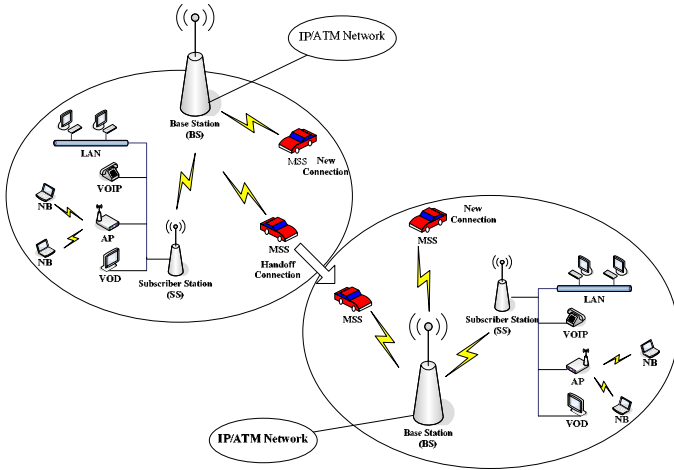


Figure 1. System model.

We design efficient call admission control algorithms for real-time and nonreal-time traffic, according to the connection processing scheme and different multimedia services.

The rest of this paper is organized as follows. In Section 2, we present our assumed model of wireless networks. In Section 3, we illustrate the proposed scheme in detail. In Section 4, we present our simulation model and analyze the comparative evaluation results of the proposed scheme through simulations. Finally, some conclusions are given in Section 5.

2. SYSTEM MODEL

We consider an environment of combination of fixed wireless connections and mobile wireless connections with an IEEE 802.16 BWA system infrastructure. Figure 1 illustrates the system model. This infrastructure contains multiple BSs for serving multiple SSs and mobile subscriber stations (MSSs). Each BS grants bandwidth to SSs (or MSSs) according to the bandwidth request, and communicates with other BSs. BSs are connected to a switching network through wired lines. The communication path between SS (or MSS) and BS has two directions: uplink and downlink. In downlink, the BS is the unique transmitter broadcasting to all the SSs and MSSs. In the other direction, the uplink channel to BS is shared by all the SSs and MSSs. Each SS is responsible for allocation the bandwidth among the corresponding connections. IEEE 802.16 MAC supports grant per SS (GPSS) class of SS. In this case, a portion of the available bandwidth is granted to each of the SSs. We consider an SS of type GPSS for which the allocated bandwidth is shared among the different service types in the SS. Once a new connection arrives at the SS, the SS attempts to set up an uplink connection with BS. An SS accepts a new connection if required bandwidth for the new connection is smaller than available bandwidth of the SS. Otherwise, the new connection is rejected. This situation is called connection blocking. Once the BS receives the MSS connection request, it makes decision to admit whether the connection can be served or not and to allocate corresponding bandwidth. Two types of MSS connections share the bandwidth of the BS: new connections and handoff connections. A BS may receive new connections of MSSs within its coverage area and handoff connections of MSSs from the

neighboring BSs. When a connection arrives at a BS where the bandwidth is not available, this situation is called connection blocking for new connections or connection dropping for handoff connections. The CBP and CDP become the two most significant QoS factors in the wireless networks. In order to support the QoS for different service, following four types of scheduling service classes are defined in the IEEE 802.16 standard. Each type of service has different QoS requirement [1]-[6]. 1) UGS: This service supports real-time constant bit-rate (CBR) traffic that generates fixed-size data packets at periodic interval, such as voice over IP (VoIP). The BS allocates a fixed amount of bandwidth to each of the connections. 2) rtPS: This service supports real-time variable bit-rate (VBR) traffic that generates variable size data packets at periodic interval, such as MPEG video. These applications have specific bandwidth requirements and tight delay latency. The amount of bandwidth required for this type of service is determined based on the required QoS performances. 3) nrtPS: This service supports non real-time VBR traffic which requires minimum data transmission rate and tolerates long-term delay. It is suitable for applications such as FTP. This service requires a QoS guarantee which is not as tight as that for rtPS and the bandwidth allocation is also adaptive. 4) BE: This service is for best-effort traffic with no minimum data transmission rate. It is suitable for applications such as telnet and WWW. The amount of bandwidth allocated to BE connections is adaptively changed depending on other types of connections in network.

3. PROPOSED DISTRIBUTED QoS GUARANTEED SCHEME

In order to provide better QoS guarantees for connection users in the IEEE 802.16 BWA Systems, efficient QoS guaranteed schemes and corresponding algorithms must be developed and designed in the BS and SSs. Figure 2 shows the block diagram of distributed QoS guaranteed scheme.

3.1 Connection Classifier Mechanism

We classify all data traffic into four types of scheduling service classes according to the IEEE 802.16 standard. Let S_1 , S_2 , S_3 , and S_4 , respectively, denote the UGS service, the rtPS service, the nrtPS service, and the BE service. Also, let Bw_{SS_k, S_i} be the allocated bandwidth for the S_i service connections in the k th SS. Let Bw_{SS_k} be the total amount of allocated bandwidth in the k th SS according to the allocation bandwidth mechanism in the BS. Let Br_{SS_k} denote the reserved bandwidth for S_1 service connections in the k th SS according to the hybrid reservation mechanism described below. When a new connection arrives at the k th SS, the SS tries to admit this request by using the algorithm shown in Figure 3, where $\sum b_{S_i}$ is the total amount of allocated bandwidth for ongoing connections of the S_i service, Nb_{S_i} is the required bandwidth for new connection of the S_i service, and the borrowing strategy is described in the reallocation bandwidth mechanism.

3.2 Reallocation Bandwidth Mechanism

We provide the efficient bandwidth allocation approach in the SS by using the attribute of multimedia traffic. Let N_{SS_k, S_i} be the total number of S_i service in the k th SS and W_{SS_k} denote the total weight value in the k th SS, which can be expressed by

$$W_{SS_k} = \sum_{i=1}^4 N_{SS_k, S_i} \cdot W_{S_i}, \quad (1)$$

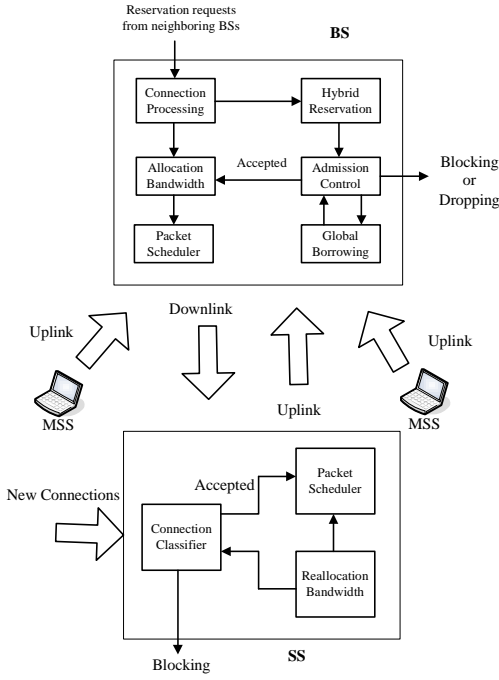


Figure 2. Block diagram of distributed QoS guaranteed scheme.

where the W_{S_i} is the weight vale of the S_i service. The initial weight value of the S_i service may be chosen based on the real operation of system (e.g., $W_{S_1} = 0.4$, $W_{S_2} = 0.3$, $W_{S_3} = 0.2$, and $W_{S_4} = 0.1$). The allocated bandwidth for the S_i service connections in the k th SS can be calculated by

$$Bw_{SS_k, S_i} = \frac{N_{SS_k, S_i} \cdot W_{S_i}}{W_{SS_k}} \times Bw_{SS_k} \quad (2)$$

We utilize the local borrowing strategy to readjust the allocated bandwidth of ongoing connections in the SS when bandwidth is borrowed or returned. When the SS does not have enough residual bandwidth to provide a connection, the borrowable bandwidths of connections will be borrowed. In the borrowing-based bandwidth allocation scheme [7], it assumes that multimedia applications can tolerate transient fluctuations in the QoS and allows for the temporary borrowing of bandwidth from existing connections in order to accommodate new connections. The parameters that must be specified in a connection request are the connection class and the required, minimum, and expected levels of bandwidth. The expected amount falls between the required and the minimum, and represents a comfortable working level for the application. The difference between the required and expected amounts of a connection is the actual borrowable bandwidth and the SS may borrow some of this bandwidth from an existing connection in order to accommodate other incoming connections. We assume that the minimum bandwidth is defined as 60% required bandwidth. The expected bandwidth is set at half way between the required bandwidth and minimum bandwidth. According to the four types of scheduling service classes in the IEEE 802.16 standard, we assume that no borrowable bandwidth can be

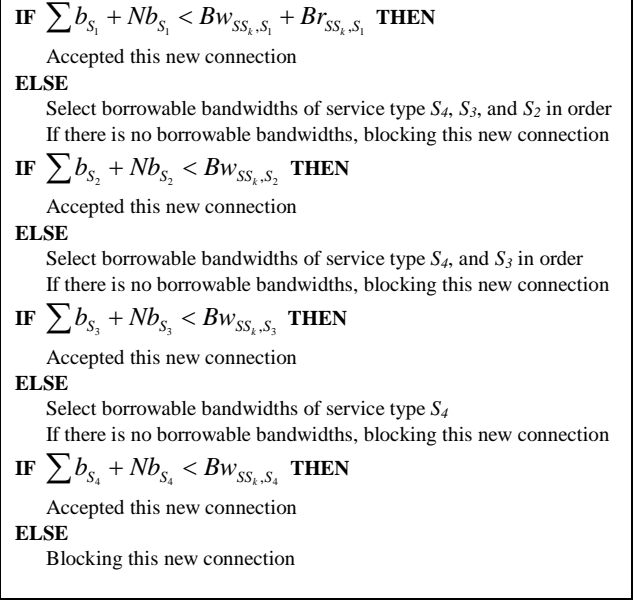


Figure 3. Connection classifier algorithm

borrowed for the S_1 service connections, 20% borrowable bandwidth can be borrowed for the S_2 service connections, 40% borrowable bandwidth can be borrowed for the S_3 service connections, and 100% borrowable bandwidth can be borrowed for the S_4 service connections. In this strategy, each SS establishes a service-queue according to the type of services S_4 , S_3 , and S_2 in order. In order to maintain a comfortable QoS for a high priority connection, when the borrowable bandwidth must be borrowed, this strategy examines the total amount of borrowable bandwidths. If the total amount of borrowable bandwidths is enough for borrowed, the borrowable bandwidth for each connection is borrowed from the service-queues in order.

3.3 Packet Scheduler Mechanism in SS

The 802.16 BWA system uses a frame size for data packet transmission. This frame is divided into subframes for downlink and uplink transmissions. The lengths of the downlink and uplink subframes for each SS are determined by the BS and broadcast to the SSs through downlink and uplink map messages (UL-MAP and DL-MAP) at the beginning of each frame. Therefore, in the downlink direction, each SS knows when and how long to receive data from the BS. In the uplink direction, each SS can request bandwidth to the BS by using BW-request PDU (Protocol data unit) and then transmit data to the BS. In this mechanism, the service-oriented packet schedule is employed to reduce the packet delay. According to the type of services, the packets are aligned in the frame appropriately. We assume that the length of a MAC PDU is fixed at m bits, the transmission frame size is n ms, and the allocated bandwidth is B Mbps in the SS. Let P_{SS_k, S_i} be the number of transmission packets of the S_i service in the k th SS. If a MAC PDU is defined as one packet, the P_{SS_k, S_i} can be calculated by

$$P_{SS_k, S_i} = \left[\frac{B \cdot n}{m} \times \frac{Bw_{SS_k, S_i}}{\sum_{i=1}^4 Bw_{SS_k, S_i}} \right]. \quad (3)$$

In order to provide better QoS, the transmission packets are scheduled, according to the type of services S_1 , S_2 , S_3 , and S_4 in order.

3.4 Connection Processing Mechanism

In order to obtain reasonable weight and reservation parameters, we monitor several traffic parameters within each time interval in the BS. Let W_{BS_j, S_i} be the dynamic weight parameter of S_i service connections in the BS j . If m SSs communicate with the BS j , the W_{BS_j, S_i} can be calculated by

$$W_{BS_j, S_i} = \frac{\sum_{k=1}^m N_{SS_k, S_i}}{\sum_{k=1}^m \sum_{i=1}^4 N_{SS_k, S_i}} \times W_{S_i}. \quad (4)$$

Let UW_{BS_j, S_i}^t and N_{BS_j, S_i}^t respectively, be the total used bandwidth and total number of connections for the S_i service connections at time t in the BS j . Let α and β , respectively, be the bandwidth rate and the numeral rate of S_i service connections in the BS, which can be obtained as follow:

$$\alpha = \frac{\sum_{t=t_{n-1}}^{t_n} UW_{BS_j, S_i}^t}{\sum_{t=t_{n-2}}^{t_{n-1}} UW_{BS_j, S_i}^t}, \quad (5)$$

$$\beta = \frac{\sum_{t=t_{n-1}}^{t_n} N_{BS_j, S_i}^t}{\sum_{t=t_{n-2}}^{t_{n-1}} N_{BS_j, S_i}^t}. \quad (6)$$

The parameter α represents bandwidth variation of S_i service connections within the time interval $[t_{n-2}, t_n]$ in the BS. The parameter β represents variation of S_i service connections within the time interval $[t_{n-2}, t_n]$ in the BS. We define α and β as reservation-factor in the BS. These two parameters make the system responsive to the UGS traffic conditions reasonably and highly affect the operation of the reservation system in the proposed scheme. For MSS, we assume that each BS receives the reservation requests from its neighboring BS. The BS receives the reservation requests including P_x , T_d , and B_x . P_x denotes the probability of MSS x moving from the neighboring BS to the current BS. T_d denotes the estimated departing time for MSS x , moving from the neighboring BS to the current BS. B_x denotes the required bandwidth for MSS x . We also monitor several traffic parameters for MSS within each time interval in the BS. We classify all data traffic into real-time (Class I, e.g., UGS and rtPS) and nonreal-time (Class II, e.g., nrtPS and BE) according to the different multimedia services. Let $mi_{BS_j, C}^t$ and $mo_{BS_j, C}^t$, respectively, be the sum of required bandwidths of hand-in connections and the sum of released bandwidths of hand-out connections for class C data traffic at time t in the BS j , where t is measured time from t_{n-1} to t_n . Let $\gamma_{BS_j, C}^t$ denote the bandwidth rate of hand-in and hand-out connections for class C data traffic at

time t_n in the BS j . This parameter can be calculated by $\gamma_{BS_j, C}^t = (mi_{BS_j, C}^t - mo_{BS_j, C}^t) / (mi_{BS_j, C}^t + mo_{BS_j, C}^t)$ and represents bandwidth variation of MSSs moving into and departing from the BS j for the class C data traffic within the time interval $[t_{n-1}, t_n]$. Let $ni_{BS_j, C}^t$ and $no_{BS_j, C}^t$, respectively, be the number of hand-in connections and the number of hand-out connections for the class C data traffic at time t in the BS j . Let $\lambda_{BS_j, C}^t$ denote the numeral rate of hand-in and hand-out connections for class C data traffic at time t_n in the BS j . This parameter can be calculated by $\lambda_{BS_j, C}^t = (ni_{BS_j, C}^t - no_{BS_j, C}^t) / (ni_{BS_j, C}^t + no_{BS_j, C}^t)$ and represents variation of MSSs moving into and departing from the BS j for the class C data traffic within the time interval $[t_{n-1}, t_n]$. These are important parameters for hybrid reservation, admission control, and allocation bandwidth mechanisms in the proposed scheme.

3.5 Hybrid Reservation Mechanism

The hybrid reservation mechanism is employed to calculate the amount of reserved bandwidth for UGS connections and handoff MSS based on the reservation-factor and network conditions. Let Br_{BS_j, S_1}^t be the amount of reserved bandwidth for S_1 service connections at time t_n in the BS j . The amount of new reserved bandwidth for S_1 service connections at the next time t_{n+1} in the BS j can be calculated by

$$Br_{BS_j, S_1}^{t_{n+1}} = (\alpha + \beta) \cdot Br_{BS_j, S_1}^t. \quad (7)$$

Let $N_{SS_k, S_1}^{t_n}$ be the total number of S_1 service at time t_n in the k th SS. If m SSs communicate with the BS j , the reserved bandwidth for S_1 service connections at the next time t_{n+1} in the k th SS can be calculated by

$$Br_{SS_k, S_1}^{t_{n+1}} = \frac{N_{SS_k, S_1}^{t_n}}{\sum_{j=1}^m N_{SS_j, S_1}^{t_n}} \times Br_{BS_j, S_1}^{t_{n+1}}. \quad (8)$$

In the mobile environments, the BSs record the requested bandwidths of MSSs within each time interval for each traffic class. Let $\omega_{BS_j, C}^{t_{n+1}}$ represent the reserved bandwidth for the class C data traffic when the time T_d is equal to the time t_{n+1} for all reserved MSSs in the BS j , which can be calculated by

$$\omega_{BS_j}^{t_{n+1}} = \sum_{m=1}^6 \sum_{k=1}^n P_{x_k, i_m, j} \cdot Bx_k, \quad (9)$$

Let $P_{x_k, i_m, j}$ be the handoff probability of x_k , moving from BS i_m to BS j , where x_k is a reserved MSS in the BS j , for $k = 1, 2, \dots, n$, and i_m , for $m = 1, 2, \dots, 6$, denote the neighboring BSs of BS j . Bx_k is required bandwidth of x_k . Let $Br_{BS_j, C}^{t_{n+1}}$ denote the amount of new reserved bandwidth for the next time t_{n+1} and $Br_{BS_j, C}^t$ the amount of reserved bandwidth at the current time for the class C data traffic in the BS j . Then, the $Br_{BS_j, C}^{t_{n+1}}$ can be calculated by

$$Br_{BS_j, C}^{t_{n+1}} = \left(\frac{1 - \varepsilon_{BS_j, C}^t}{2} \right) \cdot Br_{BS_j, C}^t + \left(\frac{1 + \varepsilon_{BS_j, C}^t}{2} \right) \cdot \omega_{BS_j, C}^{t_{n+1}}, \quad (10)$$

where $\varepsilon_{BS_j, C}^t$ is the explicit reservation-factor which can be calculated by $\varepsilon_{BS_j, C}^t = \gamma_{BS_j, C}^t + \lambda_{BS_j, C}^t$. The explicit reservation-factor $\varepsilon_{BS_j, C}^t$ controls the relative weights for the recent and past traffic histories. In order to obtain more reasonable sampling data, we use sum of $\gamma_{BS_j, C}^t$ and $\lambda_{BS_j, C}^t$ to extend the range for variation of traffic. If the mobility of MSSs is uniformly distributed, the new reserved bandwidth for the next time in the base station of BS j

can rely on the current reserved bandwidth ($Br_{BS_j,C}^n$). However, when the traffic distribution is non-uniform, due to the temporal traffic fluctuations, the new reserved bandwidth for the next time in the BS j should depend on the reserved bandwidths of handoff connections ($Br_{BS_j,C}^{n+1}$). We decide the value of $Er_{BS_j,C}^n$ by considering the current network traffic conditions. Hence the value of reservation-factor $Er_{BS_j,C}^n$ can affect the network performance significantly. In the proposed scheme, we take into account the existing network conditions. The amount of reserved bandwidth for each BS is dynamically adjusted, according to the on-line traffic information of each BS. It reserves bandwidth only when necessary. So the BSs make bandwidth reservation efficiently.

3.6 Global Borrowing Mechanism

In this mechanism, each BS establishes a service-queue according to the borrowable bandwidths of SSs and MSSs. Let Bo_{SS_k} be the amount of borrowable bandwidth for the k th SS, which can be calculated by

$$Bo_{SS_k} = \frac{Bw_{SS_k}}{\sum_{i=1}^4 N_{SS_k,S_i}} \times (N_{SS_k,S_2} \cdot \phi_{S_2} + N_{SS_k,S_3} \cdot \phi_{S_3} + N_{SS_k,S_4}), \quad (11)$$

where ϕ_{S_2} and ϕ_{S_3} denote the borrowing factor for S_2 and S_3 service connections, respectively (e.g., $\phi_{S_2} = 0.2$ and $\phi_{S_3} = 0.4$).

When the BS does not have enough residual bandwidth to provide a connection, the borrowable bandwidths of SSs or MSSs will be borrowed. Thus, the borrowed SSs or MSSs will temporarily have to give up a certain amount of allocated bandwidth. In order to provide high satisfying degree of QoS for ongoing connections and to decrease the redistribution overheads in mobile communication systems, the borrowable bandwidths are selected from the service-queue in order, according to the borrowable bandwidths of SSs and MSSs. The borrowable bandwidths are recorded in the service-queue. The arrangement of borrowable bandwidth in the service-queue is described as follows: 1) The borrowable bandwidths of SSs are aligned in front of the service-queue, according to the size of borrowable bandwidth. The larger size of borrowable bandwidth is aligned previously. 2) The borrowable bandwidths of MSSs are added to the service-queue, according to type of services S_4 , S_3 , and S_2 in order. In the same type of service, the larger size of borrowable bandwidth is aligned previously. Because the high bandwidth requests also have high borrowable bandwidths, the influence of satisfying degree of QoS for connections is slight and the number of redistribution is also small when the borrowable bandwidths are borrowed from the high bandwidth requests. In order to maintain a comfortable QoS for each connection, when the borrowable bandwidth must be borrowed, this strategy examines the half total amount of borrowable bandwidths. If the half total amount of borrowable bandwidths is enough for borrowed, the half borrowable bandwidth for each connection is borrowed from the service-queue in order. Otherwise, the total borrowable bandwidth for each MH is borrowed from the service-queue in order. When an MSS terminates, the borrowed bandwidth will be returned. It must be returned to the MSS for type of services S_2 , S_3 , S_4 and SSs in order. Hence only few connections must be re-distributed.

3.7 Admission Control Mechanism

When an MSS requests a new connection in the BS, the BS tries to admit this request by using the borrowable bandwidth of

connections in the service-queue. The new connection of MSS is rejected in the BS if its expected bandwidth is larger than the borrowable bandwidth in the service-queue. Otherwise, the BS examines the reservation factor, according to the connection processing mechanism. If there is a high reservation-factor in the BS (e.g., $Er_{BS_j,C}^n > 0.6$), then the new connection will be blocked. Otherwise, the BS accepts the connection and allocates the corresponding bandwidth for the new connection of MSS. If the borrowable bandwidth has to be used for the MSS, the BS executes the global borrowing strategy to borrow the borrowable bandwidth and to re-distribute the allocated bandwidth.

When an MSS moves from neighboring BS to the current BS, the neighboring BS releases allocated bandwidth of the MSS, and executes the global borrowing strategy to return the borrowed bandwidth and to re-distribute the allocated bandwidth. There are two classes of the call admission control algorithms for handoff connection in the current BS, according to the traffic class of MSS. If the handoff connection of MSS is real-time data traffic, the current BS checks whether the reserved bandwidth plus total amount of borrowable bandwidth is sufficient or not. If the required bandwidth for handoff connection of MSS is not sufficient, the current BS drops the handoff connection of MSS. Otherwise, the current BS accepts the connection and allocates the corresponding bandwidth for the handoff connection of MSS. If the borrowable bandwidth has to be used for MSS, the current BS executes the global borrowing strategy to borrow the borrowable bandwidth and to redistribute the allocated bandwidth. If the handoff connection of MSS is nonreal-time data traffic, it is dropped only when there is no residual bandwidth for the MSS or high reservation-factor for class I traffic in the current BS. If there is not any reserved bandwidth for class II traffic and total amount of borrowable bandwidth in the current BS, the current BS drops the handoff connection of MSS. Otherwise, the current BS accepts the connection and allocates the corresponding bandwidth for the handoff connection of MSS. If the borrowable bandwidth has to be used for MSS, the current BS executes the global borrowing strategy to borrow the borrowable bandwidth and to re-distribute the allocated bandwidth.

3.8 Allocation Bandwidth Mechanism

The connections in the SSs, new connections and handoff connections of MSSs will be allocated the required bandwidths in a BS if they are accepted for service. Let $Bw_{BS_j}^{total}$ and $Bw_{BS_j}^{avail}$, respectively, be the total amount of bandwidth and the amount of available bandwidth in the BS j . Let $Bw_{BS_j}^{resv}$ denote the total amount of reserved bandwidth ($Br_{BS_j,S_1}^n + Br_{BS_j,C}^n$) in the BS j and $Bw_{BS_j}^{alloc}$ the allocated bandwidth for MSS connections in the BS j . Then, the available bandwidth $Bw_{BS_j}^{avail} = Bw_{BS_j}^{total} - Bw_{BS_j}^{alloc} - Bw_{BS_j}^{resv}$. If m SSs communicate with the BS j , the total amount of allocated bandwidth in the k th SS can be calculated by

$$Bw_{SS_k} = \frac{\sum_{i=1}^4 N_{SS_k,S_i} \cdot W_{BS_j,S_i}}{\sum_{s=1}^m \sum_{i=1}^4 N_{SS_s,S_i} \cdot W_{BS_j,S_i}} \times Bw_{BS_j}^{avail}. \quad (12)$$

3.9 Packet Scheduler Mechanism in BS

In the downlink direction, SSs and MSSs receive data from the BS. In this mechanism, the service-oriented packet schedule is employed to reduce the packet delay and packet lose rate. According to the type of services, the packets are aligned in the

frame appropriately. We assume that the maximum number of PDUs can be transmitted in downlink subframe is N units. Let Ab_{M_k, S_i} be the allocated bandwidth of S_i service for the k th MSS and P_{BS_j, S_i} be the number of transmission packets of the S_i service in the BS j . If a MAC PDU is defined as one packet and there are N MSSs in the system and m SSs communicate with the BS j , the P_{BS_j, S_i} can be calculated by

$$P_{BS_j, S_i} = \left\lfloor N \times \frac{\sum_{k=1}^n Ab_{M_k, S_i} + \sum_{s=1}^m Bw_{SS_s, S_i}}{Bw_{BS_j}^{total}} \right\rfloor. \quad (13)$$

In order to provide better QoS, the transmission packets are scheduled, according to the type of services S_1, S_2, S_3 , and S_4 in order.

4. PERFORMANCE ANALYSIS

In this section, we present performance analysis for the proposed scheme. We describe our simulation model and illustrate the simulation results, comparing our scheme with the traditionally non-reserved bandwidth allocation mechanism. We consider a transmission scenario from multiple SSs and MSSs to BSs. The simulation environment is composed of 50 BSs and available bandwidth of each BS is from 10 to 100 Mbps, each BS keeping contact with its six neighboring BSs. The SSs work in GPSS mode. The arrival process for new connection requests is Poisson distribution with rate μ , which is uniform in all SSs. The location of each MSS is randomly distributed in all BSs at the initial state and the arrival rate for new connection of MSSs is Poisson distribution. The moving situation of each MSS is randomly selected at the initial state. Table I shows that there are four different multimedia applications in the simulation based on the traffic class, bandwidth requirement, and connection duration. The connection duration is assumed to follow a geometric distribution between the minimum and the maximum values shown in Table I. The bandwidths required for rtPS, nrtPS and BE connections are assumed to follow a geometric distribution between the minimum and the maximum values shown in Table I. The transmission frame size is 2 ms and the length of a MAC PDU is fixed at 100 bits. The PDU arrival process follows an exponentially distribution.

Fig. 4 shows the CBP and CDP when the connection arrival rate is 0.01 requests/sec and the available bandwidth of each BS is 50 Mbps. The CBPs of two mechanisms have similar trends, but due to the better bandwidth allocation approach in the proposed scheme, CBP increases slowly in heavy traffic load situations. When the number of SSs increases, it is evident that the CBPs of our proposed scheme are better than that of the non-reserved bandwidth allocation mechanism. Furthermore, as the number of SSs increases, it is obvious that the proposed scheme results in lower CDP. For example, when the number of SSs is 50, the proposed scheme achieves the CDP that is approximately 17% lower than the non-reserved bandwidth allocation mechanism. The reason for this behavior is that the proposed scheme dynamically adjusts the bandwidth allocation for each BS, providing more efficient call admission control and adaptive bandwidth reservation and borrowing for handoff connections. Fig. 5 and 6 show the CBP for UGS, rtPS, nrtPS, and BE connections for SS = 20 and the available bandwidth of each BS is 50 Mbps when the connection arrival rate is from 0.001 to 0.05 requests/sec. The

proposed scheme also results in lower CBP for different service types. Fig. 7 and 8 show the packet delay for UGS, rtPS, nrtPS, and BE connections for SS = 10 and the available bandwidth of each BS is 50 Mbps when the connection arrival rate is from 0.001 to 0.05 requests/sec. The proposed scheme results in lower packet delay for real-time service types. Also, the packet delay for nonreal-time service types of proposed scheme is close to that of the non-reserved bandwidth mechanism. This is because the packet scheduler of the proposed scheme is determined by using the concept of service-oriented packet schedule, providing more efficient packet scheduler. So the packet delay can be reduced. Fig. 9 shows the bandwidth utilization when the connection arrival rate is 0.01 requests/sec and the available bandwidth of each BS is 50 Mbps. The non-reserved bandwidth mechanism results in higher bandwidth than the proposed scheme. This is because the non-reserved bandwidth mechanism does not reserve any bandwidth for connections. However, when the number of SSs increases, the bandwidth utilization of proposed scheme is close to that of the non-reserved bandwidth mechanism. The major reason is that the amount of reserved bandwidth for each BS is dynamically adjusted, according to the on-line traffic information. It reserves bandwidth only when necessary, and thus few bandwidths are wasted.

5. CONCLUSION

The provision of QoS tends to become complicated in future mobile communication systems. In this paper, a distributed QoS guaranteed scheme is proposed as a solution to support QoS guarantees in the IEEE 802.16 mobile multimedia wireless networks. In the proposed scheme, the amount of allocation bandwidth for each BS is dynamically adjusted, according to the on-line traffic information. The simulation results show that the proposed scheme using dynamically adaptive techniques improves CBP, CDP, and packet delay significantly.

6. REFERENCES

- [1] IEEE Std 802.16e-2005 and IEEE Std 802.16-2004, "IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems," 2005.
- [2] H. Wang, B. He, and D. P. Agrawal, "Admission control and bandwidth allocation above packet level for IEEE 802.16 wireless MAN," *Proc. 12th International Conference on Parallel and Distributed Systems (ICPADS 2006)*, vol. 1, July, 2006.
- [3] H. S. Alavi, M. Mojdeh, and N. Yazdani, "A Quality of Service Architecture for IEEE 802.16 Standards," *Proc. Asia-Pacific Conference on Communications*, pp. 249-253, Oct. 2005.
- [4] Z. Abichar, Y. Peng, and J. M. Chang, "WiMax: The Emergence of Wireless Broadband," *IT Professional*, vol. 8, no. 4, pp. 44-48, July-Aug. 2006.
- [5] K. R. Santhi, and G. S. Kumaran, "Migration to 4 G: Mobile IP based Solutions," *Proc. International Conference on Internet and Web Applications and Services/Advanced International Conference on Telecommunications*, Feb. 2006.
- [6] A. Malla, M. El-Kadi, S. Olariu, and P. Todorova, "A Fair Resource Allocation Protocol for Multimedia Wireless Networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 14, no. 1, pp. 63-71, Jan. 2003.

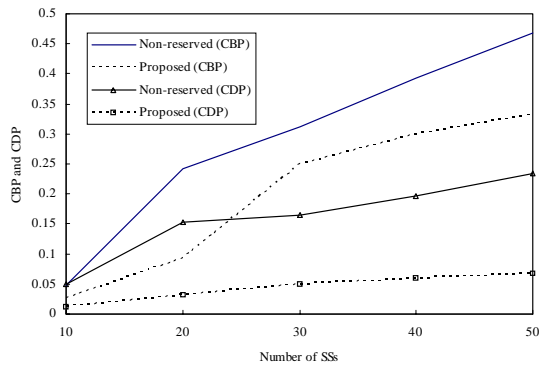


Figure 4. CBP and CDP (connection arrival rate = 0.01 requests/sec).

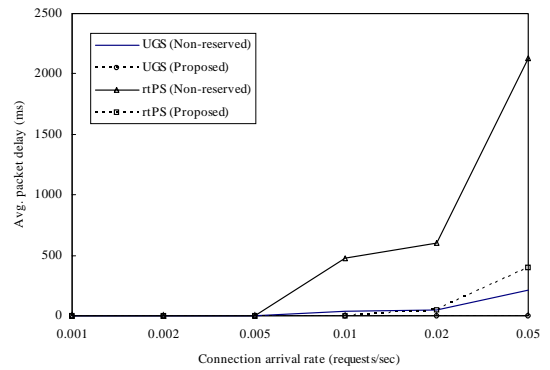


Figure 7. Packet delay for UGS and rtPS connections.

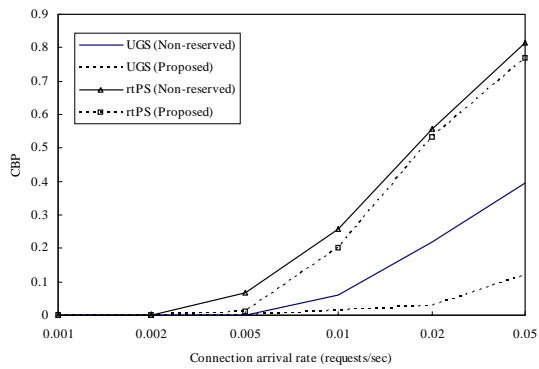


Figure 5. CBP for UGS and rtPS connections.

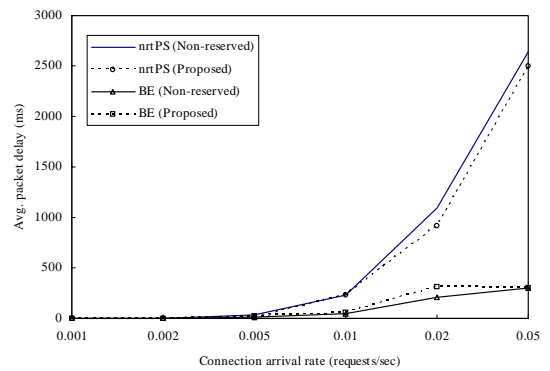


Figure 8. Packet delay for nrtPS and BE connections.

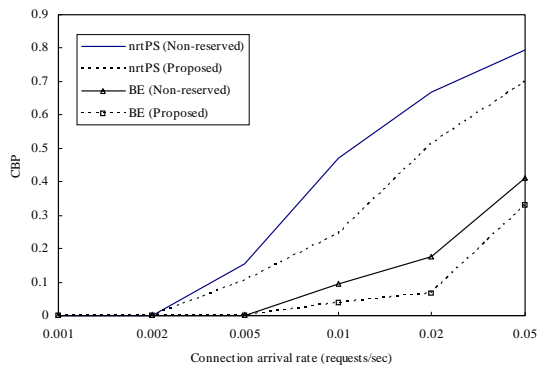


Figure 6. CBP for nrtPS and BE connections.

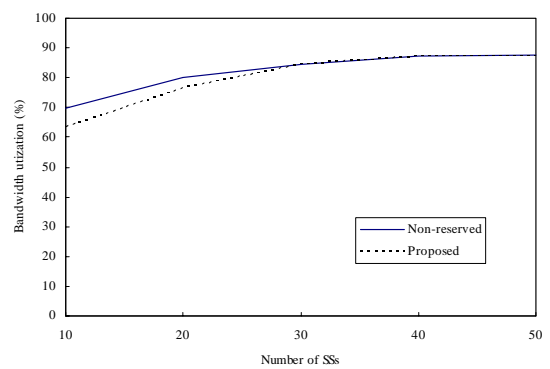


Figure 9. Bandwidth utilization (connection arrival rate = 0.01 requests/sec)

TABLE I. Multimedia traffic used in simulation mode

Application	Traffic Class	Bandwidth Requirement	Average Bandwidth Requirement	Connection Duration	Average Connection Duration
1	UGS	256 Kbps		1-30 m	5 m
2	rtPS	1-5 Mbps	2.5 Mbps	5 m-5 h	10 m
3	nrtPS	1-6 Mbps	3 Mbps	30 s-20 m	2 m
4	BE	64-512 Kbps	256 Kbps	30 s-10 h	3 m