

Hindawi Publishing Corporation
EURASIP Journal on Wireless Communications and Networking
Volume 2009, Article ID 635304, 10 pages
doi:10.1155/2009/635304

Research Article

Service Differentiation in OFDM-Based IEEE 802.16 Networks

Yi Zhou,¹ Kai Chen,² Jianhua He,³ Haibin Guan,⁴ Yan Zhang,⁵ and Alei Liang¹

¹ School of Electronic, Information and Electrical Engineering, Shanghai Jiaotong University, Shanghai 200240, China

² School of Information Security and Engineering, Shanghai Jiaotong University, Shanghai 200240, China

³ Institute of Advanced Telecommunications, Swansea University, Swansea SA2 8PP, UK

⁴ Department of Computer Science, Shanghai Jiaotong University, Shanghai 200240, China

⁵ Simula Research Laboratory, 1325 Lysaker, Norway

Correspondence should be addressed to Kai Chen, kchen@sjtu.edu.cn

Received 1 August 2008; Accepted 1 December 2008

Recommended by Lingyang Song

IEEE 802.16 network is widely viewed as a strong candidate solution for broadband wireless access systems. Various flexible mechanisms related to QoS provisioning have been specified for uplink traffic at the medium access control (MAC) layer in the standards. Among the mechanisms, bandwidth request scheme can be used to indicate and request bandwidth demands to the base station for different services. Due to the diverse QoS requirements of the applications, service differentiation (SD) is desirable for the bandwidth request scheme. In this paper, we propose several SD approaches. The approaches are based on the contention-based bandwidth request scheme and achieved by the means of assigning different channel access parameters and/or bandwidth allocation priorities to different services. Additionally, we propose effective analytical model to study the impacts of the SD approaches, which can be used for the configuration and optimization of the SD services. It is observed from simulations that the analytical model has high accuracy. Service can be efficiently differentiated with initial backoff window in terms of throughput and channel access delay. Moreover, the service differentiation can be improved if combined with the bandwidth allocation priority approach without adverse impacts on the overall system throughput.

Copyright © 2009 Yi Zhou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

In today telecommunications, networking and services are changing in a rapid way to support next generation Internet user environment. Broadband wireless access (BWA) is one of the most promising solutions for broadband access and will play an important role in the next generation Internet. BWA systems are being increasingly deployed and used in the last mile for extending or enhancing Internet connectivity for fixed and/or mobile clients located on the edge of the wired network. IEEE 802.16 standard has been developed as one of the technical solutions for BWA systems [1]. The physical (PHY) layer and MAC layer specifications are defined in 802.16d for fixed BWA, and enhanced with low to moderate mobility support in 802.16e [1]. World-wide Interoperability for Microwave Access (WiMAX) was founded to promote the compatibility of 802.16 products. IEEE 802.16 networks can have a wide variety of applications, including high-speed Internet access, backhaul for WiFi

hotspot and cellular networks [2–7]. It can provide a cost-effective alternative to the existing solutions for these applications.

Flexible bandwidth request and allocation mechanisms have been specified in the 802.16 standards to support different scheduling services for uplink traffic, namely, unsolicited grant services (UGSs), real-time polling services (rtPSs), nonreal-time polling services (nrtPSs), and best-effort (BE) services [1–3, 8, 9]. UGS connections will periodically receive bandwidth grant for the uplink traffic from the base station without sending bandwidth request. rtPS connections periodically receive bandwidth grants to send bandwidth request to the base station to indicate bandwidth demands. On the other hand, nrtPS and the best-effort (BE) service need transmit bandwidth requests via random access or by piggybacking the requests to already granted data transmissions. nrtPS connections may receive sporadic exclusive bandwidth request opportunities to request bandwidth from the base station.

UGS and rtPS scheduling services can be used for the applications with stringent QoS requirements, for example, VoIP and VoD, and so on. However, one problem of the advanced scheduling services (UGS and rtPS) is that a good knowledge of the traffic characteristics (e.g., packet arrival interval and packet size) is required to provide satisfactory QoS and efficiently utilize bandwidth [10]. Since the traffic characteristics of many applications are not known in advance, traffic transported via the random bandwidth request scheme will achieve better bandwidth utilization, which makes the random access very important for the applications over IEEE 802.16 networks.

Several papers have investigated the IEEE 802.16 random access scheme by simulation, theoretic analysis, or both. Vinel et al. accurately analyzed the truncated binary exponential backoff (TBEB) algorithm specified in the 802.16 standard [11]. The performance of the 802.16 contention-based CDMA requesting mechanism are analyzed with orthogonal frequency-division multiple address (OFDMA) physical layer [12–15]. The attempts of analyzing the random access protocol and finding the optimal access parameters are made in [16, 17]. The 802.16 bandwidth request scheme is analyzed with saturated traffic and limited bandwidth in [14, 18]. Random access and polling mechanisms for WiMAX networks are compared in [9]. The OFDM- and OFDMA-based random access schemes are compared by simulation in [10].

In this paper, we investigate several SD approaches for the bandwidth request scheme with limited bandwidth. As the bandwidth request scheme is anticipated to support applications with diverse QoS requirements, for example, emergency services with bursty traffic and real-time voting for live entertainments, it is important to implement SD over the bandwidth request scheme to support diverse QoS requirements [19]. A QoS differentiation scheme for IEEE 802.16 mesh network was investigated in [20]. However, to the best of our knowledge, no papers have been published so far, with a comprehensive study of the SD with the 802.16 bandwidth request scheme. Moreover, most of the existing work has not considered the constraints of available bandwidth [11, 13, 16, 17]. In addition, we also propose an analytical model to study the impacts and effectiveness of the SD approaches. The analytical model can be used as a tool for the adaptive configuration of the SD approaches. In the remainder of the paper, the bandwidth request scheme is introduced in Section 2. SD approaches are also described in Section 2. Analytical model is presented in Section 3. Simulation results are presented and analyzed in Section 4. Finally, Section 5 concludes the paper.

2. Bandwidth Request Scheme

2.1. Overview of MAC Layer Protocol. The 802.16 standard specifies two modes for sharing wireless medium: point-to-multipoint (PMP) or mesh modes. A base station (BS) serves multiple subscriber stations (SSs) in the PMP mode. The downlink (from the BS to the user) operates on a PMP basis and is generally broadcast. All the transmissions in the uplink

from SSs are directed to and centrally coordinated by the BS. Mesh mode is an optional configuration, in which SSs can communicate with the BS over multihops. The IEEE 802.16 standard specifies four physical layers: SC and SCa for single carrier transmission in line-of-sight (LoS) and nonline-of-sight (NLoS) environments, OFDM and OFDMA for multicarrier transmission in NLoS environments. A common MAC layer is defined for all physical layers with small adaptations to the different physical layers. In the following, we focus on the bandwidth request scheme in the PMP networks with OFDM physical layer. But it worths noting that the proposed SD approaches and analytical model can be easily extended to the PMP networks with OFDMA physical layer, as the bandwidth request scheme is very similar for both OFDM and OFDMA physical layers.

The MAC layer supports both time-division duplexing (TDD) and frequency-division duplexing (FDD) modes. In the TDD mode, each MAC frame consists of a downlink subframe followed by an uplink subframe, which is investigated in this paper. The investigation can be applied to FDD model as well. The downlink subframe starts with a preamble and a frame control header. The frame control header specifies the presence of control information within the downlink subframe. Downlink MAP (DL-MAP) and uplink MAP (UL-MAP) follow the frame control header and specify the usage of the downlink and uplink subframes, respectively. Particularly, the DL-MAP defines the starting times, destinations and the burst profiles (modulation and coding) of the data bursts within the downlink subframe. The UL-MAP allocates resources in the uplink subframe to the different subscriber stations, for the purpose of ranging, bandwidth requesting, and data burst transmission, and so on.

2.2. Basic Bandwidth Request Scheme. The bandwidth request scheme in 802.16 networks consists mainly of procedures of bandwidth requesting and granting. A general method of bandwidth requesting is that SSs send stand-alone bandwidth request header to indicate required bandwidth (in bytes) in the transmission opportunities (TOs), which are allocated by the BS in the uplink subframes. The BS will process the successfully received bandwidth request headers from the SSs and allocate available bandwidth to the SSs. The bandwidth allocation will be indicated in the downlink subframes. If bandwidth is granted to an SS, the SS can send collision-free transmission burst in the allocated bandwidth. Each burst is associated with a physical burst profile (i.e., modulation and error control coding schemes). In addition, a collision-free bandwidth request header can be sent together with data in the allocated bandwidth to request further bandwidth if the SS has more traffic to send. The above general requesting method can be used for different services. Alternative bandwidth requesting mechanisms include contention-based focused bandwidth requests (for WirelessMAN-OFDM physical specification only) and contention-based CDMA bandwidth requests (for WirelessMAN-OFDMA only) [1]. In this paper, we focus on

the general bandwidth requesting method without request piggyback.

As the traffic from the nrtPS and BE connections can be bursty and is hard to predict, the TBEB algorithm has been specified in the 802.16 standard for collision resolution. When an SS has a packet to send to the BS, it uses the TBEB algorithm to determine which frame and which TO to transmit their bandwidth requests [1]. A backoff process is initiated, with backoff counter uniformly chosen in $[0, W_0 - 1]$, where W_0 is the initial backoff windows. The backoff counter decreases by one for each eligible TO in the frames without the need of sensing the channel. Once the backoff counter reaches zero, the SS can transmit its bandwidth request header to the BS. After the request header is transmitted, the SS sets a timer (with value of N_w in the unit of frames) to wait for the bandwidth grant from the BS. If the transmitted request header does not collide with other request headers and can be decoded by the BS, the BS may allocate the requested bandwidth to the SS provided that bandwidth is available. In the case that the SS receives bandwidth grant before the timer expires, it can use the granted bandwidth to transmit data burst in the subsequent frame [1]. However, if the SS fails to receive bandwidth grant before the timer time out, the SS will retransmit the bandwidth request header to the BS. The bandwidth request header can be transmitted up to a maximal number of m retries. For every retransmission purpose, the backoff window is doubled until it reaches the maximal backoff window (denoted by W_x), which means the backoff counter for the i th retransmission of a bandwidth request header will be uniformly chosen in $[0, W_i - 1]$ with $W_i = \min(2^i W_0, W_x)$. If the bandwidth request header does not result in a successful bandwidth grant after m retries, the SS will stop this bandwidth request attempt and discard the packet associated with this request attempt. If there are still packets left in the buffer, the SS can start the TBEB algorithm again to request bandwidth.

It is noted that the BS will determine and broadcast the channel access parameters, such as the initial backoff window W_0 , the maximal backoff window W_x , the maximal number of retries m , and the timeout value N_w . It is also the BS's responsibility to determine the number of TOs for the SSs in each uplink subframe.

2.3. Service Differentiation Schemes. From the bandwidth request scheme, we can find that the quality of service for the nrtPS and BE connections will be affected directly by the number of eligible TOs in a frame, the random channel access parameters, and the policy of granting available bandwidth to the successfully received bandwidth request headers by the BS. To differentiate the services for the connections, we can use the following intuitive methods.

- (1) Allocating different TOs for the connections.
- (2) Setting different channel access parameters.
- (3) Giving priorities on allocating bandwidth to SSs with successfully received bandwidth request headers.

The differentiation methods can be used separately or jointly.

In this paper, we will consider the approaches of differentiating channel access parameters and priority-based bandwidth allocation. In the 802.16 standard, the maximal number of retries m is required to be no less than 16. Therefore, setting different m for the connections will not be an efficient SD approach. With regard to the priority-based bandwidth allocation, equal priority and absolute priority-based schemes are considered. In the absolute priority-based scheme, the available bandwidth is allocated to the higher priority connections first. The lower priority connections will receive bandwidth grant only after all the higher priority connections receive bandwidth grants. For the connections with the same priority and insufficient bandwidth, the BS randomly choose some connections to serve.

In general, we will assume that a BS serves N_u independent SSs in the PMP network. There are K service classes with each service class associated with $N_{u,k}$ SSs, and $N_u = \sum_{k=1}^K N_{u,k}$. Denoted the duration of an OFDM symbol by T_{sym} in seconds and the frame duration by T_f in seconds. Each REQ transmission opportunity (TO) consists of T_r OFDM symbols. Assume a fixed number (N_r) of TOs are assigned in each uplink subframe, and the bandwidth (in bytes) requested by each request header can be accommodated by one transmission burst. Each REQ is assumed to request bandwidth for a data packet of a fixed length of L bytes. Each burst consists of T_d OFDM symbols. Bandwidth for N_d bursts is allocated for the bandwidth requesting scheme in each uplink subframe.

In principle an SS can have traffic from multiple connections. For convenience, we assume each SS has only one connection with BS, and the connection belongs to one of the two predefined service classes in the network. Each service class is associated with a set of channel access parameters. A priority policy is used for the service classes on bandwidth allocation. Without loss of generality, we assume class i connections (SSs) has higher priority over class j connections (SSs) with $1 \leq i < j \leq K$. Therefore, the channel parameters of W_0 and W_x for class 1 connections are not larger than those for class 2 connections, and class 1 connections have higher or equal priority if compared to class 2 connections on receiving bandwidth grant from the BS.

3. Analytical Model

3.1. Model Assumption. For simplicity, we model 2 service classes in the PMP networks, with $N_{u,1}$ and $N_{u,2}$ SSs associated with the first and the second service classes, respectively. Each SS has saturated traffic to send to the BS. If the number of successfully received bandwidth request headers (simply abbreviated as REQ) in a frame is larger than N_d , BS will simply serve N_d randomly chosen SSs and drop the REQs received from the other SSs. It is trivial to extend the model to more than 2 services classes and unequal priority on the bandwidth allocation. Let $W_{0,k}$ and $W_{x,k}$ denote the initial and maximal backoff windows of TBEB, for $k \in [1, 2]$, respectively. Denote the maximal number of retries as m_k for service class k . The contention window $W_{i,k}$

for backoff stage i ($i \in [0, m_k]$) is $\min(2^i W_{0,k}, W_{x,k})$, where function $\min(\cdot)$ calculates the minimum of the variable set. If a class k SS is unsuccessful in a TO transmission due to either collision or insufficient bandwidth, the SS waits N_w frames before retransmission.

3.2. REQ Transmission Probability. We assume that each SS transmits independently of the other SSs in steady state. Let τ_k denote the probability that a tagged class k SS transmits REQ over TOs in a general frame. Let p_k denote the probability that a TO transmission from the tagged class k SS fails due to either collision or insufficient bandwidth in BS. τ_k and p_k are assumed to be constant in the steady state.

To derive the expression for τ_k , we define a contention resolution process (CRP) as the process of the tagged SS from the initialization of TBEB to the end of TBEB for a packet transmission attempt. At the end of a TBEB, the packet is either discarded or successfully sent to BS. τ_k is therefore computed for the tagged class k SS by the ratio of the average number of TO transmissions in a CRP for class k SS (denoted by $N_{t,k}$, in frames) and the average period of a CRP (denoted by $N_{\text{crp},k}$, in frames) for class k SS. We can obtain τ_k ($k \in [1, 2]$) by

$$\tau_k = \frac{N_{t,k}}{N_{\text{crp},k}}. \quad (1)$$

Let $p_{\text{req},k}(i)$ denote the probability that the number of class k REQ transmissions is exactly i in an EBRP, for $i \in [1, m+1]$. We have $p_{\text{req},k}(i) = (1-p_k)p_k^{i-1}$, for $i \in [1, m]$ and $p_{\text{req},k}(m+1) = p_k^m$. The average number of REQ transmissions in an EBRP ($N_{t,k}$) can be computed by

$$\begin{aligned} N_{t,k} &= \sum_{i=1}^{m+1} i p_{\text{req},k}(i) \\ &= \sum_{i=0}^m (i+1)(1-p_k)p_k^i + (m+1)p_k^{m+1}. \end{aligned} \quad (2)$$

The average period of a CRP depends on the number of TO transmissions in a CRP by an SS and the outcomes of the transmissions. Let $N_{z,k}$ denote the number of frames required for the tagged class k SS to transmit a data packet after a successful TO transmission. Denote by $N_{b,k}(i)$ the average of the total number of frames that the tagged class k SS spends on the backoff processes for TO transmissions in a CRP, under the condition of exactly i retries in the CRP. Assume that $W_{0,k}$ is dividable by $N_r \cdot N_{b,k}(i)$, for $i \in [0, m]$, is given by

$$N_{b,k}(i) = \sum_{j=0}^i \frac{1}{W_{j,k}} \sum_{k=0}^{W_{j,k}-1} \left\lceil \frac{k+1}{N_r} \right\rceil = \sum_{j=0}^i \frac{N_r + W_{j,k}}{2W_{j,k}}, \quad (3)$$

where $\lceil x \rceil$ obtains the minimal integer n no smaller than x .

The average period $N_{\text{crp},k}$ of a CRP for the tagged class k SS can be computed by

$$\begin{aligned} N_{\text{crp},k} &= (1-p_k) \sum_{i=0}^m p_k^i [i(N_w-1) + N_{b,k}(i) + N_{z,k}] \\ &\quad + p_k^{m+1} [(m+1)(N_w-1) + N_{b,k}(m+1)]. \end{aligned} \quad (4)$$

3.3. Probability of Unsuccessful REQ Transmission. Note that here an unsuccessful REQ means a collided REQ transmission or a uncollided REQ but not receiving bandwidth allocation from the BS due to limited bandwidth. Let $p_{q,k}(l)$ denote the probability of exactly l TO transmissions from N_k class k SSs in a frame, and $N_{\text{tr},k}$ denote the average number of TO transmissions from class k SSs in a frame. We can calculate them with given τ_k by

$$p_{q,k}(l) = \binom{N_k}{l} \tau_k^l (1-\tau_k)^{N_{u,k}-l}, \quad (5)$$

and $N_{\text{tr},k} = \sum_{l=1}^{N_{u,k}} l p_{q,k}(l)$.

A collision will happen if more than one REQ is transmitted in the same TO of a frame. Denote by $p_{\text{uc}}(s_1, s_2 | l_1, l_2)$ the probability that exactly s_1 REQs among the l_1 REQs from class 1 SSs and s_2 REQs among l_2 REQs from class 2 SSs are uncollided in a frame, where $s_1 \leq l_1$, $s_2 \leq l_2$ and $s_1 + s_2 \leq N_r$. We can compute $p_{\text{uc}}(s_1, s_2 | l_1, l_2)$ by (6) with a similar method for a classic occupancy problem given in [21]:

$$\begin{aligned} p_{\text{uc}}(s_1, s_2 | l_1, l_2) &= \frac{1}{N_r^{l_1+l_2}} \binom{l_1}{s_1} \binom{l_2}{s_2} \frac{N_r!}{(N_r - s_1 - s_2)!} \\ &\quad \times \sum_{v=0}^{\min(N_r - s_1 - s_2, l_1 + l_2 - s_1 - s_2)} \\ &\quad \times (-1)^v \binom{N_r - s_1 - s_2}{v} \\ &\quad \times (N_r - s_1 - s_2 - v)^{l_1 + l_2 - s_1 - s_2 - v}. \end{aligned} \quad (6)$$

for $0 \leq s_1 \leq l_1$, $0 \leq s_2 \leq l_2$ and $s_1 + s_2 \leq N_r$; otherwise $p_{\text{uc}}(s_1, s_2 | l_1, l_2) = 0$. In (6), the first factor $1/N_r^{l_1+l_2}$ is the probability of each arrangement of random transmission of $l_1 + l_2$ REQs over N_r Tos; the second, the third, and the fourth factors calculate the number of ways that s_1 of l_1 class 1 REQs and s_2 of l_2 class 2 REQs are chosen and each of the chosen REQs is transmitted over one of the N_r TOs without collision. The summation calculates the number of ways that the left $l_1 + l_2 - s_1 - s_2$ REQs are transmitted over the left $N_r - s_1 - s_2$ TOs and none of the $l_1 + l_2 - s_1 - s_2$ REQ transmissions is successful.

Then we can compute the average number of successful REQ transmitted from class k SSs (denoted by $N_{\text{suc},k}$ in a frame by (7) for $k = 1$:

$$\begin{aligned} N_{\text{suc},1} &= \sum_{l_1=0}^{N_{u,1}} \sum_{l_2=0}^{N_{u,2}} p_{q,1}(l_1) p_{q,2}(l_2) \\ &\quad \times \sum_{s_1=1}^{l_1} \sum_{s_2=0}^{l_2} \frac{p(s_1, s_2 | l_1, l_2) s_1 \min(N_d, s_1 + s_2)}{(s_1 + s_2)}, \end{aligned} \quad (7)$$

and by (8) for $k = 2$, resp.):

$$N_{\text{suc},2} = \sum_{l_1=0}^{N_{u,1}} \sum_{l_2=0}^{N_{u,2}} p_{q,1}(l_1) p_{q,2}(l_2) \times \sum_{s_1=0}^{l_1} \sum_{s_2=1}^{l_2} \frac{p(s_1, s_2 | l_1, l_2) s_2 \min(N_d, s_1 + s_2)}{(s_1 + s_2)}. \quad (8)$$

The probability of unsuccessful REQ transmission p_k can be simply computed from $N_{\text{suc},k}$ and the average number of transmitted REQ in one frame:

$$p_1 = 1 - \frac{N_{\text{suc},1}}{\sum_{l=1}^{N_{u,1}} p_{q,1}(l)l}, \quad (9)$$

$$p_2 = 1 - \frac{N_{\text{suc},2}}{\sum_{l=1}^{N_{u,2}} p_{q,2}(l)l}.$$

The values of τ_k and p_k ($k \in [1, 2]$) can be obtained by solving nonlinear equations (1) and (9) with numeric techniques.

3.4. Throughput and Delay. With the transmission probability τ_k and the probability of unsuccessful REQ transmission p_k , we can calculate the performance metrics of interest for service differentiation, including throughput and channel access delay for each service class. In addition, the overall system bandwidth efficiency can also be computed.

Define throughput of a single SS (denoted by θ_k for service class k , $k \in [1, 2]$) as the average number of bits transmitted from an SS to the BS in one second. The throughput of a single SS depends largely on the physical burst profile. Let B_{sym} denote the number of uncoded bits that can be transmitted with an OFDM symbol for an SS. Then the throughput θ_k of a class k SS can be computed for $k \in [1, 2]$:

$$\theta_k = \frac{N_{\text{suc},k} T_d B_{\text{sym}}}{T_f}, \quad (10)$$

where T_f is the duration of a frame. Define channel access delay D_k for service class k ($k \in [1, 2]$) as the average time elapsed between the beginning of the frame in which the first backoff process is initiated for a packet from a class k SS to the end of the frame in which the last backoff process ends in a bandwidth request attempt. Then we can simply get $D = N_{\text{crp}} T_f$.

We define the bandwidth efficiency (denoted by η) of a bandwidth request scheme as the ratio of the data packet length (denoted by T_d in OFDM symbols) to the average bandwidth (in OFDM symbols) consumed to successfully transmit a data packet. Let T_0 denote the bandwidth (in OFDM symbols) needed for a TO, which is two for the general OFDM-based bandwidth request scheme. The bandwidth efficiency η for the whole system can be calculated by

$$\eta = \frac{T_d (N_{\text{suc},1} + N_{\text{suc},2})}{N_r T_r + N_d T_d}. \quad (11)$$

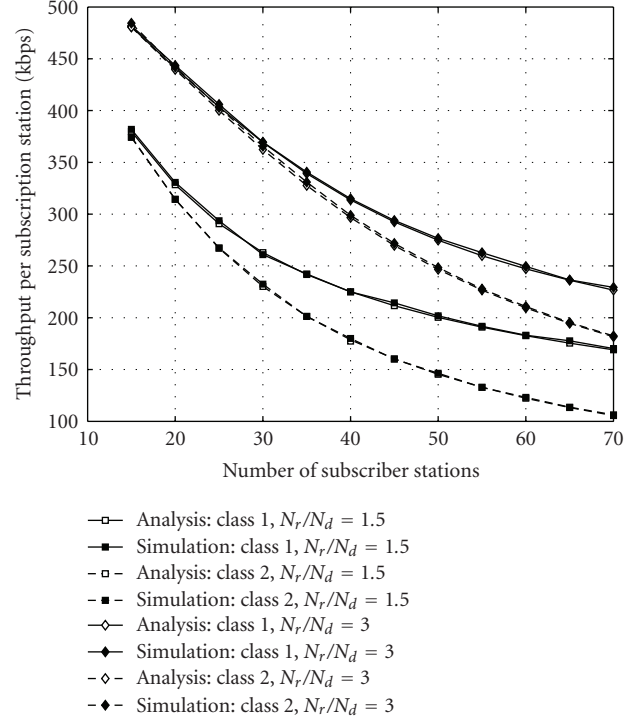


FIGURE 1: Throughput of single SS versus the number of SSs with equal bandwidth allocation priority and differentiated channel access parameters: $W_{0,1} = W_{0,2} = N_r$, $W_{x,1} = 2^4 W_{0,1}$, and $W_{x,2} = 2^8 W_{0,2}$.

4. Numerical Results

A discrete event driven simulator for the IEEE 802.16 bandwidth request scheme has been implemented. The simulator can be configured with various system and channel access parameters. With the simulator, we can investigate the effectiveness of the analytical model and how the network performances can be differentiated under the conditions of changing number of SSs and limited bandwidth.

For the results presented in the paper, channel bandwidth is set to 20 MHz in the physical layer. We assume an ideal channel, where frames will be successfully received unless collision happens. Frame duration is set to 10 milliseconds. There are 256 independent subcarriers in an OFDM symbol and 844 OFDM symbols in a frame. Assume each TO takes 2 OFDM symbols ($T_r = 2$) and each transmission burst takes 16 OFDM symbols ($T_d = 16$). The number of uncoded bits B_{sym} that can be transmitted in an OFDM symbol is set to 384, obtained with the burst profile of 16QAM modulation and 1/2 coding rate. The value of timer T_w is set to 2. For the bandwidth configuration, the number of bursts N_d that can be transmitted in a frame is set to 12. The number of TOs N_r in a frame has two configurations: $N_r = 1.5N_d$ and $N_r = 3N_d$. The maximal number of retries m is set to 16 for both service classes in all the tests.

4.1. Differentiation with Channel Access Parameters. In this section, we investigate the accuracy of the analytical model

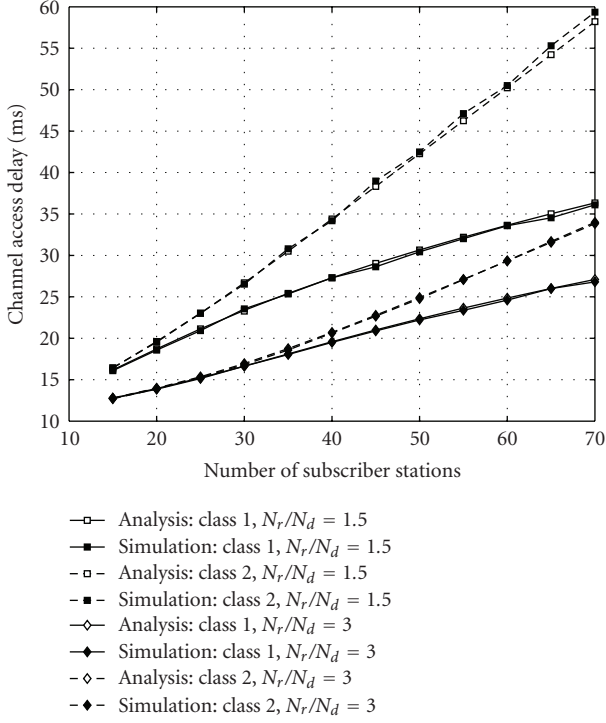


FIGURE 2: Delay of SS versus the number of SSs with equal bandwidth allocation priority and differentiated channel access parameters: $W_{0,1} = W_{0,2} = N_r$, $W_{x,1} = 2^4 W_{0,1}$ and $W_{x,2} = 2^8 W_{0,2}$.

and the achievable SD performances by tuning channel access parameters. In this case, only the channel access parameters are used for SD and the priority on bandwidth allocation is the same for the two classes of SSs.

We first test the impact of maximal backoff window on SD. The initial backoff windows for both classes are set to $W_{0,1} = W_{0,2} = N_r$. The SD is achieved by setting $W_{x,1} = 2^4 W_{0,1}$ and $W_{x,2} = 2^8 W_{0,2}$. The throughput of single SS θ , REQ transmission probability τ , and channel access delay D is plotted in Figures 1, 2, and 3, respectively. As the REQ unsuccessful probability is very close for both service classes, the corresponding results are not presented. The number of SSs in the figures is the sum of class 1 and class 2 SSs, and the number of class 1 SSs is fixed as 10. The throughput and channel access delay of class 1 SSs are represented by solid lines, and class 2 SSs by dashed lines in the figures. We also use symbols “square” and “diamond” to denote the performances with $N_r = 1.5N_d$ and $N_r = 3N_d$, respectively. The throughput, channel access delay, and REQ transmission probability for each class of SSs are averaged over the corresponding SSs in the classes. For comparison, the symbols corresponding to analytical results are not filled with any color and the symbols corresponding to simulation results are filled with black color. Each simulation result is obtained by averaging 30 simulations. We can observe from Figures 1, 2, and 3 that the analytical model has very high accuracy. It is also observed that differentiating services with only the maximal backoff window is not so effective.

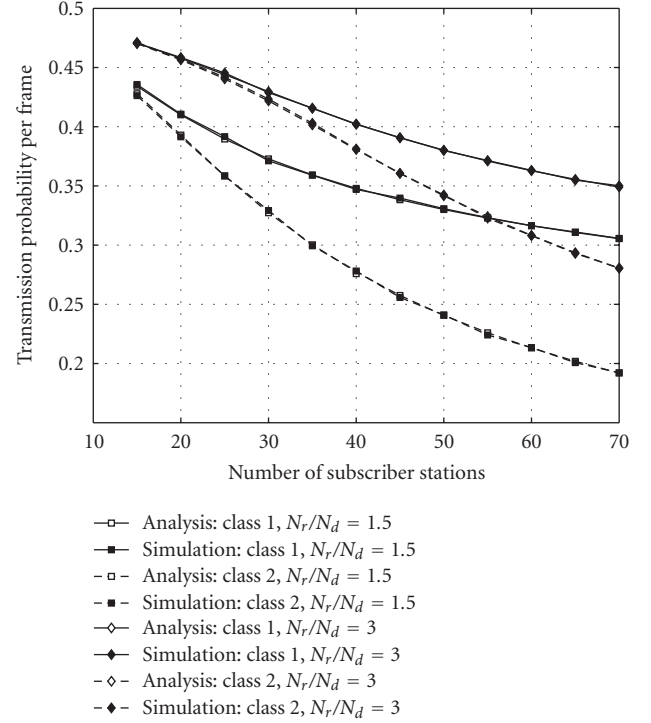


FIGURE 3: REQ transmission probability of SS versus the number of SSs with equal bandwidth allocation priority, and differentiated channel access parameters: $W_{0,1} = W_{0,2} = N_r$, $W_{x,1} = 2^4 W_{0,1}$, and $W_{x,2} = 2^8 W_{0,2}$.

Especially when the network is not congested (e.g., less than 40 SSs in the network), the SD is not obvious in Figure 1.

Next, we investigate the SD performances with different initial backoff windows and maximal backoff windows. We set different initial backoff windows and maximal backoff windows for the two classes of SSs. We keep $W_{0,1} = N_r$, $W_{x,1} = 2^4 W_{0,1}$, and $W_{x,2} = 2^8 W_{0,2}$, but change $W_{0,2}$ to $2N_r$ and $3N_r$. The corresponding throughput associated with $W_{0,2} = 2N_r$ and $W_{0,2} = 3N_r$ is plotted in Figures 4 and 5, respectively. Compared to the results in Figure 1, again it can be observed that the analytical model has very high accuracy. We can find that the initial backoff window is much more effective than the maximal backoff window for SD, for the whole range of the number of SSs in the network. The SD is more obvious with increased differentiation on the initial backoff window. The class 1 SS throughput is more than twice that of the class 2 SS throughput in most of the investigated cases for $W_{0,2} = 3N_r$. Class 1 SS throughput is proportional to the throughput of class 2 SS with differentiated initial backoff window, which can help to configure the channel access parameters for SD.

Similarly, we plotted the results of channel access delay in Figures 6 and 7, and the REQ transmission probability in Figures 8 and 9, for $W_{0,2} = 2N_r$ and $W_{0,2} = 3N_r$, respectively. It can be observed from Figures 8 and 9 that the REQ transmission probability of class 1 SSs does not change much as their initial backoff window is unchanged. The REQ transmission probability of class 2 SSs reduces with the increased initial backoff window. Consequently, the

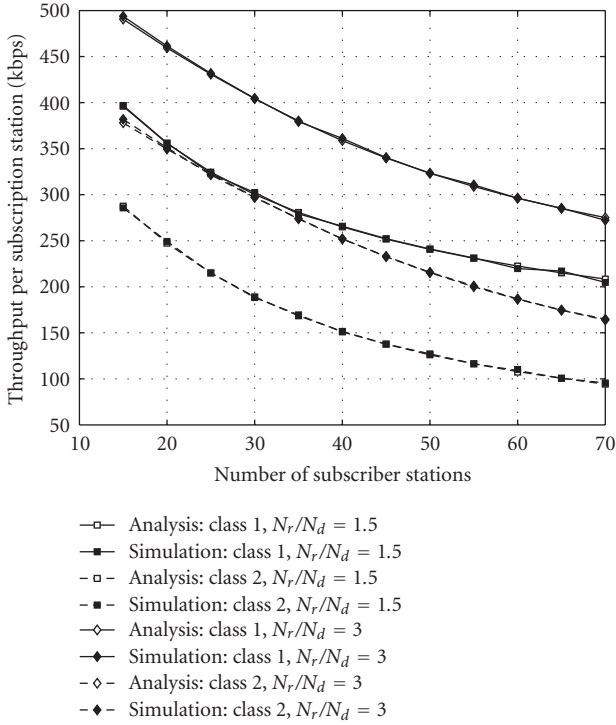


FIGURE 4: Throughput of single SS versus the number of SSs with equal bandwidth allocation priority and differentiated channel access parameters: $W_{0,1} = N_r$, $W_{0,2} = 2 N_r$, $W_{x,1} = 2^4 W_{0,1}$, and $W_{x,2} = 2^8 W_{0,2}$.

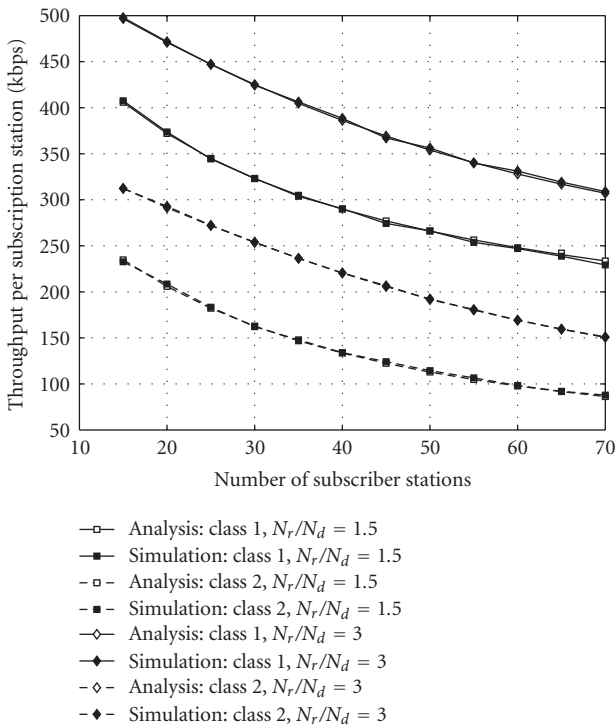


FIGURE 5: Throughput of single SS versus the number of SSs with equal bandwidth allocation priority and differentiated channel access parameters: $W_{0,1} = N_r$, $W_{0,2} = 3 N_r$, $W_{x,1} = 2^4 W_{0,1}$, and $W_{x,2} = 2^8 W_{0,2}$.

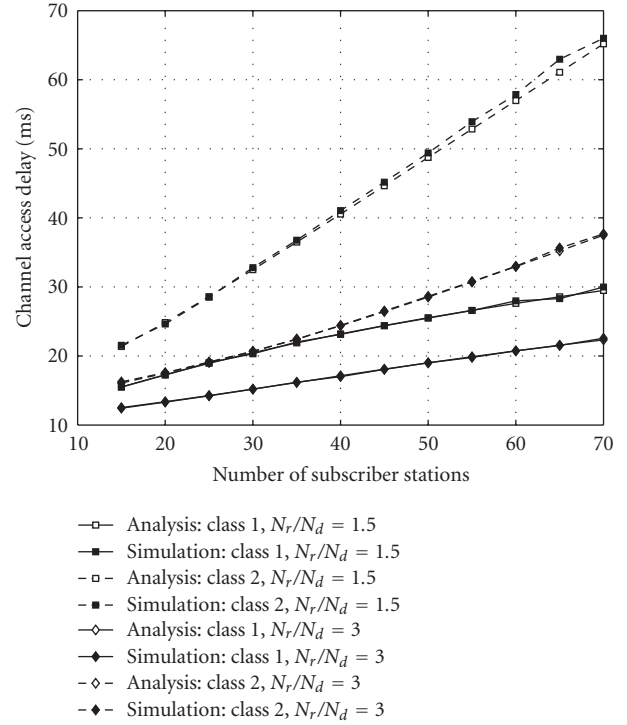


FIGURE 6: Delay of SS versus the number of SSs with equal bandwidth allocation priority and differentiated channel access parameters: $W_{0,1} = N_r$, $W_{0,2} = 2 N_r$, $W_{x,1} = 2^4 W_{0,1}$, and $W_{x,2} = 2^8 W_{0,2}$.

throughput of class 2 SSs largely reduces and the channel access delay increases.

4.2. Bandwidth Allocation Priorities. Next, we will investigate the impact of bandwidth allocation priority on SD performances. We keep the configurations of channel access parameters the same as those in Section 4.1, except for changing the equal bandwidth allocation priority to absolute priority. Only simulation results are presented for the investigation on bandwidth allocation priority performance.

The simulation results of throughput and channel access delay with absolute bandwidth allocation priority and differentiated maximal backoff window are shown by the symbols filled with black color in Figures 10 and 11, but the initial backoff window for both classes of SSs is set to the same. As observed previously, the maximal backoff window has minor contribution to SD. Therefore, it is easy to understand the impact of bandwidth allocation priority on SD from Figures 10 and 11. It is observed that in the case of $N_r = 1.5N_d$, bandwidth allocation priority has almost no impact on SD. However, in the case of $N_r = 3N_d$, bandwidth allocation priority is shown effective and can make more significant contribution than the initial backoff window, especially when the number of SSs is large. The reason can be explained by that when the number of TOs is small, channel access is the bottleneck and the initial backoff window is more critical than bandwidth allocation priority. In contrast, when the number of TOs is large in a frame, the bottleneck is no longer

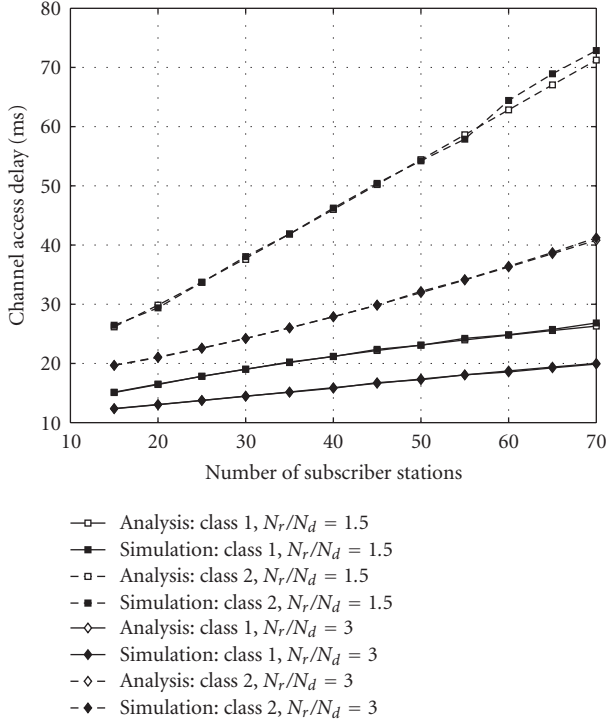


FIGURE 7: Delay of single SS versus the number of SSs with equal bandwidth allocation priority and differentiated channel access parameters: $W_{0,1} = N_r$, $W_{0,2} = 3 N_r$, $W_{x,1} = 2^4 W_{0,1}$, and $W_{x,2} = 2^8 W_{0,2}$.

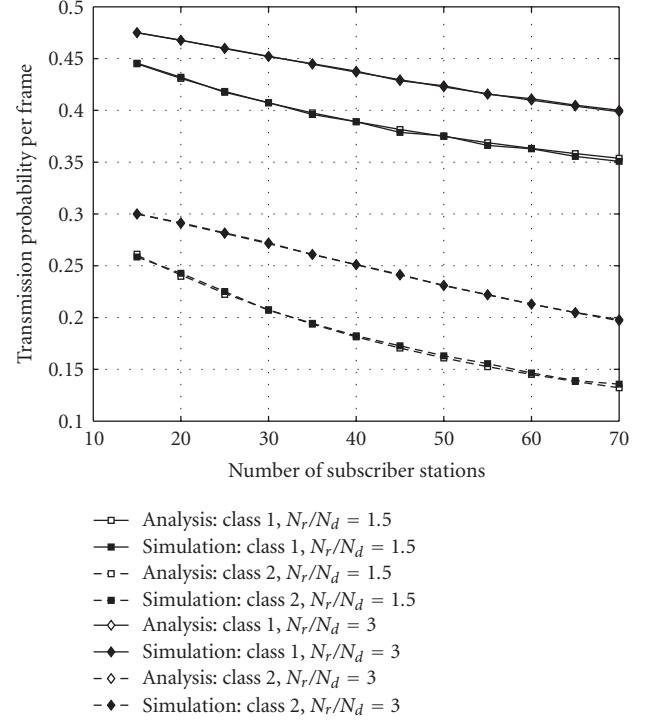


FIGURE 9: REQ transmission probability of SS versus the number of SSs with equal bandwidth allocation priority, and differentiated channel access parameters: $W_{0,1} = N_r$, $W_{0,2} = 3 N_r$, $W_{x,1} = 2^4 W_{0,1}$ and $W_{x,2} = 2^8 W_{0,2}$.

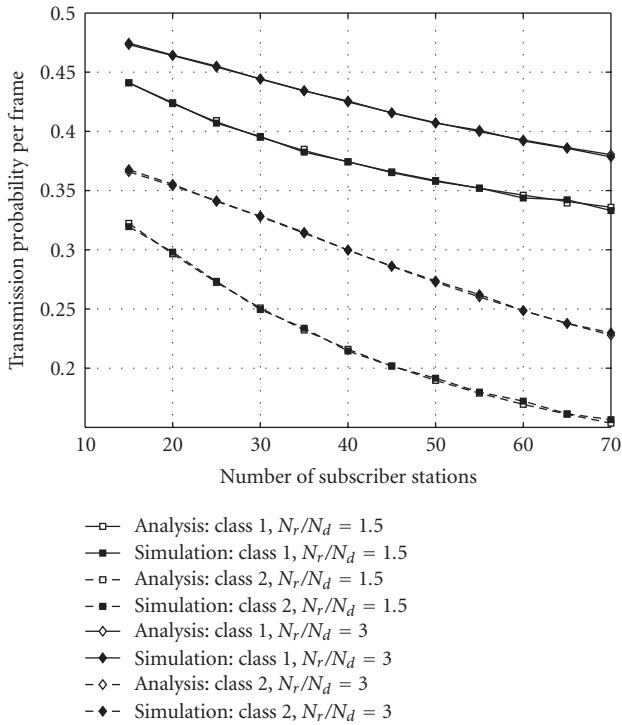


FIGURE 8: REQ transmission probability of SS versus the number of SSs with equal bandwidth allocation priority, and differentiated channel access parameters: $W_{0,1} = N_r$, $W_{0,2} = 2 N_r$, $W_{x,1} = 2^4 W_{0,1}$ and $W_{x,2} = 2^8 W_{0,2}$.

in the channel access, and bandwidth allocation plays an important role. In this case, bandwidth allocation priority can be more effective. But the SD performance obtained with bandwidth allocation priority is more nonlinear than that with channel access parameter with increasing number of SSs in the networks.

The joint impact of channel access parameters and bandwidth allocation priority on SD is illustrated in Figures 10 and 11 by the symbols without filling. It is observed that in the case of small number of TOs ($N_r = 1.5N_d$), the contribution of initial backoff window is dominated in SD. However, the joint impact of initial backoff window and bandwidth allocation priority is larger than that can be achieved when they are separately used for SD.

5. Conclusion

In this paper, we investigated the service differentiation (SD) for the bandwidth request scheme specified in the IEEE 802.16 standard. Several SD approaches including differentiating channel access parameters (mainly initial and maximal backoff windows) and bandwidth allocation priority are studied in detail. An analytical model is proposed to understand the impact of the SD approaches, which can be used to adaptively configure and optimize the system performances. Simulation validates the analytical model. Through the numerical results, it was observed that using maximal backoff window cannot effectively differentiate

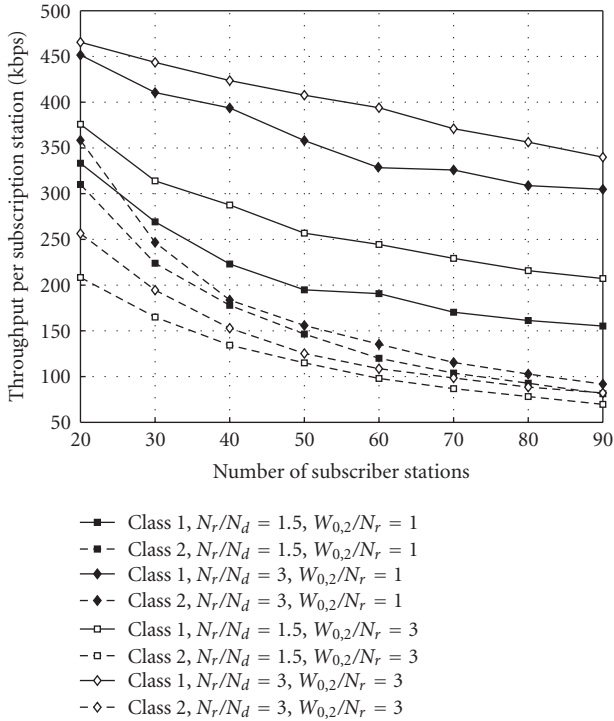


FIGURE 10: Throughput of single SS versus the number of SSs with absolute bandwidth allocation priority, and differentiated channel access parameters: $W_{0,1} = W_{0,2} = N_r$, $W_{x,1} = 2^4 W_{0,1}$ and $W_{x,2} = 2^8 W_{0,2}$.

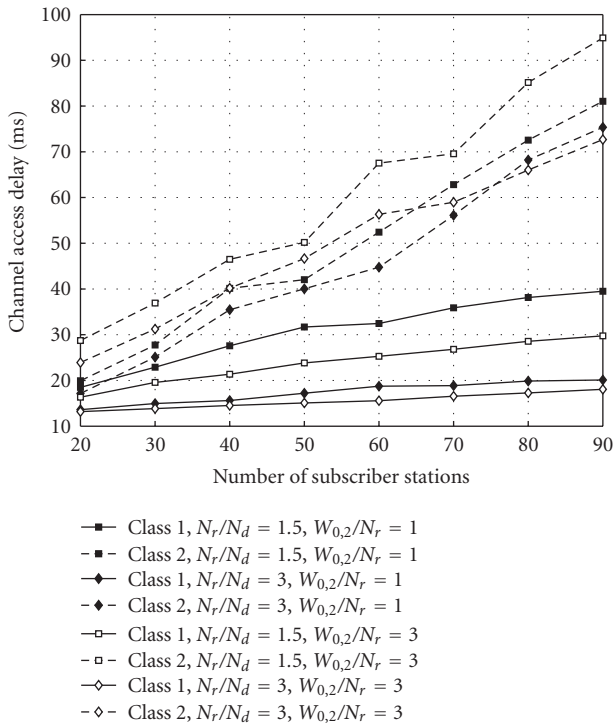


FIGURE 11: Channel access delay versus number of SSs with absolute bandwidth allocation priority, and differentiated channel access parameters: $W_{0,1} = W_{0,2} = N_r$, $W_{x,1} = 2^4 W_{0,1}$, and $W_{x,2} = 2^8 W_{0,2}$.

service. Instead, initial backoff window-based SD approach is very effective, and it is slightly worse than the priority-based bandwidth allocation scheme in SD when the number of REQ transmission opportunities (TOs) in a frame is large. When the number of TOs is small, the initial backoff window-based SD approach is much more effective than the bandwidth allocation priority-based approach. The service can be better differentiated when the initial backoff window and bandwidth allocation priority are jointly used.

Acknowledgments

This work is supported by the National Grand Fundamental Research 973 Program of China under Grant no. 2007CB316506, 863 Program of China under Grant no. 2006AA01Z169, and by the European Union through the Welsh Assembly Government.

References

- [1] "IEEE Standard for Local and metropolitan area networks—part 16: air interface for fixed broadband wireless access systems," 2004.
- [2] J. He, K. Yang, K. Guild, and H.-H. Chen, "Application of IEEE 802.16 mesh networks as the backhaul of multihop cellular networks," *IEEE Communications Magazine*, vol. 45, no. 9, pp. 82–90, 2007.
- [3] C. Cicconetti, A. Erta, L. Lenzi, and E. Mingozzi, "Performance evaluation of the IEEE 802.16 MAC for QoS support," *IEEE Transactions on Mobile Computing*, vol. 6, no. 1, pp. 26–38, 2007.
- [4] J. He, K. Yang, and K. Guild, "A dynamic bandwidth reservation scheme for hybrid IEEE 802.16 wireless networks," in *Proceedings of IEEE International Conference on Communications (ICC '08)*, pp. 2571–2575, Beijing, China, May 2008.
- [5] T.-C. Chen, Y.-Y. Chen, and J.-C. Chen, "An efficient energy saving mechanism for IEEE 802.16e wireless MANs," *IEEE Transactions on Wireless Communications*, vol. 7, no. 10, pp. 3708–3712, 2008.
- [6] M. P. Anastasopoulos, P.-D. M. Arapoglou, R. Kannan, and P. G. Cottis, "Adaptive routing strategies in IEEE 802.16 multihop wireless backhaul networks based on evolutionary game theory," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 7, pp. 1218–1225, 2008.
- [7] H. Zhu, Y. Tang, and I. Chlamtac, "Unified collision-free coordinated distributed scheduling (CF-CDS) in IEEE 802.16 mesh networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 10, pp. 3889–3903, 2008.
- [8] J. Borin and N. Fonseca, "Scheduler for IEEE 802.16 networks," *IEEE Communications Letters*, vol. 12, no. 4, pp. 274–276, 2008.
- [9] Q. Ni, A. Vinel, Y. Xiao, A. Turlikov, and T. Jiang, "Wireless broadband access: WiMax and beyond—investigation of bandwidth request mechanisms under point-to-multipoint mode of WiMAX networks," *IEEE Communications Magazine*, vol. 45, no. 5, pp. 132–138, 2007.
- [10] D. Staehle and R. Pries, "Comparative study of the IEEE 802.16 random access mechanisms," in *Proceedings of the International Conference on Next Generation Mobile Applications, Services and Technologies (NGMAST '07)*, pp. 334–339, Cardiff, UK, September 2007.

- [11] A. Vinel, Y. Zhang, M. Lott, and A. Tiurlikov, "Performance analysis of the random access in IEEE 802.16," in *Proceedings of the 16th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '05)*, vol. 3, pp. 1596–1600, Berlin, Germany, September 2005.
- [12] J.-B. Seo, H.-W. Lee, and O.-H. Cho, "Queueing behavior of IEEE802.16 random access protocol for sporadic data transmissions," in *Proceedings of the 15th International Conference on Computer Communications and Networks (ICCCN '06)*, pp. 351–357, Arlington, Va, USA, October 2006.
- [13] H.-W. Lee and J.-B. Seo, "Queueing performance of IEEE 802.16 random access protocol with bulk transmissions," in *Proceedings of IEEE International Conference on Communications (ICC '07)*, pp. 5963–5968, Glasgow, UK, June 2007.
- [14] J. He, Z. Tang, and H. Chen, "Performance comparison of OFDM bandwidth request schemes in fixed IEEE 802.16 networks," *IEEE Communications Letters*, vol. 12, no. 4, pp. 283–285, 2008.
- [15] S. Kwon and D. Cho, "CDMA code-based bandwidth request mechanism in IEEE 802.16j mobile multi-hop relay (MMR) systems," in *Proceedings of the 68th IEEE Vehicular Technology Conference (VTC '08)*, pp. 1–5, Calgary, Canada, September 2008.
- [16] J. Yan and G.-S. Kuo, "Cross-layer design of optimal contention period for IEEE 802.16 BWA systems," in *Proceedings of IEEE International Conference on Communications (ICC '06)*, vol. 4, pp. 1807–1812, Istanbul, Turkey, July 2006.
- [17] A. Doha, H. Hassanein, and G. Takahara, "Performance evaluation of reservation medium access control in IEEE 802.16 networks," in *Proceedings of the 4th ACS/IEEE International Conference on Computer Systems and Applications (AICCSA '06)*, pp. 369–374, Dubai, UAE, March 2006.
- [18] J. He, K. Guild, K. Yang, and H.-H. Chen, "Modeling contention based bandwidth request scheme for IEEE 802.16 networks," *IEEE Communications Letters*, vol. 11, no. 8, pp. 698–700, 2007.
- [19] K. Chen, Y. Zhou, J. He, and Z. Tang, "Service differentiation for the bandwidth request scheme in fixed IEEE 802.16 networks," in *Proceedings of the 4th IEEE International Conference on Circuits and Systems for Communications (ICCSC '08)*, pp. 718–722, Shanghai, China, May 2008.
- [20] H. Hu, Y. Zhang, and H.-H. Chen, "An effective QoS differentiation scheme for wireless mesh networks," *IEEE Network*, vol. 22, no. 1, pp. 66–73, 2008.
- [21] W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. 1, John Wiley & Sons, New York, NY, USA, 1957.