A Flexible Resource Allocation and Scheduling Framework for Non-real-time Polling Service in IEEE 802.16 Networks

Fen Hou, *Student Member, IEEE,* James She, *Student Member, IEEE,*
Pin-Han Ho, *Member, IEEE,* and Xuemin (Sherman) Shen, *Fellow, IEEE*

***Abstract*—This paper proposes an efficient yet simple design framework for achieving flexible resource allocation and packet scheduling for non-real-time polling service (nrtPS) traffic in IEEE 802.16 networks. By jointly considering the selective Automatic Repeat reQuest mechanism at the media access control layer as well as the adaptive modulation and coding technique at the physical layer, the proposed framework enables a graceful tradeoff between resource utilization and packet delivery delay while maintaining the minimum throughput requirements of nrtPS applications. An analytical model is developed for parameter manipulation in the proposed framework, where some important performance metrics, such as inter-service time, delivery delay, goodput, and resource utilization, are investigated for performance evaluation. Simulation results are given to demonstrate the efficiency of the proposed framework and verify the accuracy of the analytical model.**

***Index Terms*—Non-real-time polling service (nrtPS), scheduling, automatic repeat request, IEEE 802.16, adaptive modulation and coding.**

I. Introduction

As a promising broadband wireless access standard, IEEE 802.16 has attracted extensive attentions from both the industry and academia due to its strong capability of providing broadband data and real-time multimedia services with quality of service (QoS) satisfaction. IEEE 802.16 standard defines four types of services [1]: Unsolicited Grant Service (UGS), Real-time Polling Service (rtPS), Non-real-time Polling Service (nrtPS), and Best Effort (BE). UGS aims to support real-time constant-bit-rate (CBR) applications, such as T1/E1 classical pulse coded modulation (PCM) phone signal transmission and voice over Internet protocol (VoIP) without silence suppression, which are subject to stringent delay and delay jitter constraints. rtPS is designed to support variable-bit-rate applications, such as Internet protocol TV, online gaming, and video conferencing, where delay, minimum throughput, and maximum sustained throughput are defined and constrained. nrtPS is to support delay-tolerant applications, such as file transfer protocol (FTP), where minimum throughput is defined. BE service is subject to no QoS requirement.

Manuscript received September 7, 2007; revised December 6, 2007; accepted February 15, 2008. The associated editor coordinating the review of this paper and approving it for publication was X. Zhang.

This work has been supported in part by a Strategic Grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada.

The authors are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario, Canada (e-mail: {fhou, james, pinhan, xshen}@bbcr.uwaterloo.ca).

Digital Object Identifier 10.1109/TWC.2009.070995

Many resource allocation and scheduling schemes have been developed to deal with the four service types based on their intrinsic characteristics and QoS requirements. For UGS, a common solution is to periodically grant a fixed amount of resources since the service is designed for the CBR applications. For the rtPS applications, scheduling schemes, such as the largest weighted delay first and strict priority scheduling, were proposed to deal with the stringent delay requirement [2],[12]. For BE service, proportional fairness scheduling is an effective approach to balance the system throughput and fairness. To the best of our knowledge, very few research efforts have dedicatedly dealt with the resource allocation and scheduling for nrtPS in spite of its ultimate importance. It has been reported that transmission control protocol (TCP) traffic (which exclusively takes nrtPS as the carrier in an IEEE 802.16 network) may take up to 80% of the total Internet bandwidth [3], while an average share of 83% is observed in the High-Performance Computing (HPC) networks. With such predominant bandwidth consumption, it is crucial to develop a dedicated strategy for dealing with nrtPS traffic in IEEE 802.16 networks.

The key concerns on the resource allocation and scheduling for nrtPS traffic are the fulfillment of its minimum throughput requirement and improvement of resource utilization with acceptable delay. Efforts on improving resource utilization and reducing experienced delay are in general contradictory since high resource utilization can be achieved by assigning the resource to the subscriber stations (SSs) with good channel conditions, while leaving the SSs with poor channel conditions starved and experienced a long delay. Although nrtPS connections are delay-tolerant, they should not be starved for too long since otherwise the flows will suffer considerable performance degradation. For instance, based on the 3GPP standard, the delay for the low delay data service and the long constrained delay data service should be less than 50ms and 300ms, respectively [4]. Therefore, how to compromise the resource utilization and experienced delay of each SS is a challenging and important issue.

In this paper, we propose a flexible and effective design framework for achieving efficient resource allocation and packet scheduling of nrtPS traffic in IEEE 802.16 networks such that the minimum bandwidth requirement is satisfied while a nice compromise between the resource utilization and delivery delay for each SS can be achieved. Furthermore, an analytical model is developed to investigate some important performance measures, such as inter-service time, delivery delay, goodput, and resource utilization, and is positioned as a guideline for the parameter manipulation in the proposed framework. The proposed analytical model is characterized by jointly considering the Adaptive Modulation and Coding (AMC) technique and the selective Automatic Repeat reQuest (ARQ) mechanism. Extensive simulation results are given to demonstrate the efficiency of the proposed framework and verify the analytical model.

The remainder of the paper is organized as follows. Section II introduces the related work on the scheduling and ARQ. Section III describes the proposed resource allocation and packet scheduling framework, followed by the analysis of some important performance metrics. Section IV provides analytical and simulation results to validate the efficiency of the proposed framework and the accuracy of the analytical model. Finally, we conclude the paper in Section V.

II. Related Work

A. Scheduling in Wireless Networks

There have been several studies on scheduling issue in wireless networks, but most of them focused on either best-effort applications by improving the system throughput and fairness performance, or real-time applications with the delay constraint being the main concern [5]–[14].

Opportunistic scheduling [5] was devised to maximize the resource utilization by assigning resources to the user with the best channel condition at each timeslot. Nonetheless, it is not suitable for nrtPS applications since long-term starvation could occur for the SSs with bad channel conditions. A credit-based code-division generalized process sharing scheme was proposed in [6], where the scheduler assigned the resources based on both the general processor sharing discipline and each user's credit to achieve high resource utilization as well as long-term fairness. Proportional fairness (PF) scheduling [7][8] was designed to balance the fairness and system throughput. An opportunistic fair scheduling, αPFS , was proposed in [9] to flexibly adjust the fairness performance by manipulating the parameter α. In [10], the selective relative best (SRB) scheduling was proposed for BE traffic, in which a proper tradeoff between the fairness and system throughput was fulfilled by jointly considering short-term channel gain and long-term channel gain. An opportunistic fair scheduling was proposed in [11] to maintain the long-term fairness and achieve a smooth service rate. In general, all the aforementioned schemes focused on improving fairness performance, rather than providing the guarantee for the minimum throughput requirement, which is essential for nrtPS.

The Largest Weighted Delay First scheduling scheme in [12] focused on real-time applications by mainly considering the delay requirements while leaving the throughput and resource utilization aside, which is not suitable for nrtPS. Adaptive EXP/PF scheduling in [13] focused on BE traffic and real-time service, where the service priority was evaluated based on the experienced delay and channel condition at each queue.

A few studies have considered non-real-time applications, but they only focused on satisfying either the throughput ratio or the acceptable delay [14]–[17]. The Channel-Aware Round Robin (CARR) scheme [14][15] provided a soft bandwidth guarantee by compromising the conventional round robin and opportunistic scheduling. However, this scheme fails to guarantee the minimum bandwidth since the allocated resource is entirely determined by the corresponding long-term channel condition of each end user instead of the bandwidth demand. A minimum-bandwidth guaranteed proportional fairness scheduling (mGPFS) was proposed in [16], where each user's weight was adaptively adjusted by solving a linear program. This scheme provides a lower bound for the throughput ratio between the users with the best channel and the worst channel, but it cannot guarantee the achieved throughput for a specific end user. An adaptive exponential scheduling scheme was proposed in [17] for non-real-time service, where the priority of a non-real-time flow was adaptively adjusted based on its maximum acceptable delay, experienced delay, and channel conditions. Although taking the acceptable delay into account, it does not consider the bandwidth requirement.

B. Automatic Repeat reQuest

Automatic Repeat reQuest is an important close-loop error control mechanism to achieve a low packet loss rate in error-prone wireless environments. Extensive research efforts have been reported on the analysis of ARQ protocols in terms of different performance metrics, such as average delay of protocol data units (PDUs), service data units (SDUs), and throughput efficiency [18]–[21].

Although ARQ schemes have been extensively studied, some new challenges have emerged with the launching of the IEEE 802.16 standard. The IEEE 802.16 standard specifies some advances in physical layer techniques and media access control protocol, such as AMC, orthogonal frequency division multiplexing (OFDM), flexible retransmission of lost packets, etc. These characteristics have posed fundamental differences in the efforts of performance analysis of ARQ in IEEE 802.16 networks compared with that in many previous works. Firstly, most of former studies simply assumed that the time taken to transmit a packet is a constant, generally defined as a timeslot. Such an assumption does not hold in IEEE 802.16 networks due to the adoption of AMC, which aims at providing a better tradeoff between the system capacity and coverage. With AMC, the link capacity varies with the channel conditions, leading to the fact that the time taken to retransmit a packet varies with the current channel condition, and may be different from that taken in the previous transmission attempt.

Secondly, most of the former studies commonly assumed that the lost packets are retransmitted immediately in the very next timeslot [19]–[22]. This is a very strong assumption because the time consumed by data decoding, processing, and feedback is not negligible when the round-trip time is considered. In addition, the assumption may result in low system capacity due to varying channel conditions. For instance, when the channel condition of the next timeslot is poor, the retransmission could experience a much higher packet loss rate. In IEEE 802.16 networks, the retransmission of lost packets is flexible and does not just occur within the very next timeslot. Therefore, a fundamental difference can

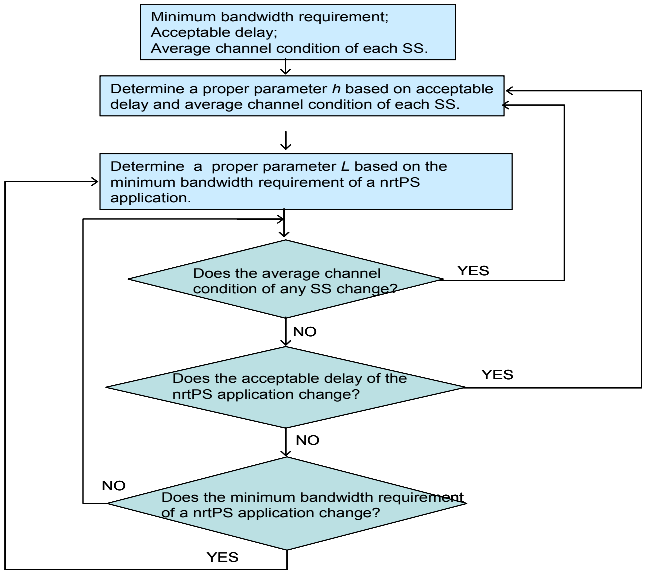


Fig. 1. The flowchart of the proposed framework for nrtPS service.

be identified in the design and performance analysis in IEEE 802.16 networks compared with that in many previous works. Thus, it is critical to develop an efficient resource allocation and scheduling framework that not only can maintain the minimum bandwidth requirements for nrtPS flows by jointly considering the selective ARQ mechanism at the media access control (MAC) layer and AMC technique at the physical layer, but also initiate a graceful compromise between the resource utilization and delivery delay. To the best of our knowledge, such design requirements and network environments have never been considered in the previous studies.

III. The Proposed resource allocation and packet scheduling framework

In this paper, we propose a simple yet efficient resource allocation and packet scheduling framework for nrtPS traffic in IEEE 802.16 networks such that the minimum bandwidth requirements can be satisfied while a flexible tradeoff between the packet delivery delay and the resource utilization is initiated. The flowchart of the proposed framework is shown in Fig. 1. h and L are two operational parameters in the proposed framework. h is the number of SSs selected at each MAC frame, which is the parameter controlling the tradeoff between the resource utilization and the packet delivery delay of each SS, while L (in the unit of PDUs) is the bandwidth granted to an SS when it is served, which is the parameter to fulfill the satisfaction of the minimum bandwidth requirement of a nrtPS application. By manipulating these two parameters, the framework can satisfy the minimum bandwidth requirement of each nrtPS flow and fulfill the flexible tradeoff between the delivery delay and the resource utilization. Furthermore, the proposed framework takes the channel condition of each SS into account. By exploiting the multi-user channel diversity, the proposed scheme can obtain a high system capacity.

Due to the error-prone wireless channel, there may exist much gap between the assigned bandwidth and achieved goodput. The proposed framework takes this situation into account by jointly considering the selective ARQ mechanism at the MAC layer. With the use of selective ARQ, the information about whether or not the PDUs transmitted at a downlink sub-frame (DL sub-frame) are successfully received is feedbacked to the base station (BS) at the following uplink sub-frame (UL sub-frame). Based on the feedback information, the failed PDUs are retransmitted next time when this queue obtains the transmission opportunity, instead of being retransmitted immediately. By taking the impact of selective ARQ on the retransmission of lost PDUs, we analyze the performance of the proposed framework in terms of the achieved goodput and the packet delivery delay for each SS. Furthermore, the proposed framework also considers the AMC technique at the physical layer which is specified in the IEEE 802.16 standard. With the adoption of AMC, the system dynamically adjusts the modulation level according to the channel condition of each SS. In IEEE 802.16 networks, different modulation levels lead to different numbers of information bits carried by an OFDM symbol. Thus, resource utilization, which is defined as the average number of information bit carried by an OFDM symbol, is one of the most important metrics to evaluate the performance of the proposed framework. In this paper, we also analyze the resource utilization of the proposed framework with the consideration of AMC.

In specific, the proposed framework works as follows. At the beginning of each MAC frame, h SSs with better channel conditions are selected and granted with transmission opportunities at this DL sub-frame. Each selected SS is assigned with a specific amount of resources, which is denoted as L, according to the minimum throughput requirements of the nrtPS flows destined to these SSs and the channel conditions of all SSs. By manipulating the parameters h and L properly, the proposed resource allocation and scheduling framework can fulfill any possible design requirement, such as resource utilization, throughput requirements, and delivery delay requirements. When h is set to 1, it is degraded to the opportunistic scheduling, which can obtain the maximum resource utilization at the expense of possibly long delivery delay of the SSs with poor channel conditions. When h equals to the total number of SSs associated to the BS, the resource utilization is low, but the delivery delays of SSs are small. Meanwhile, the setting of parameter L depends not only on the minimum bandwidth requirement of a nrtPS flow, but also on the channel conditions of all SSs associated to the BS.

In order to study the performance of the proposed framework and the impact of the two parameters on the design requirements, such as resource utilization, minimum bandwidth requirement, and packet delivery delay, an analytical model is developed to evaluate some key performance metrics, including inter-service time, PDU delivery delay, SDU delivery delay, goodput, and resource utilization. For simplicity, the SS under consideration is referred to as the tagged SS, and at the BS, the queue that buffers the nrtPS PDUs destined to the tagged SS is referred to as the tagged queue. When the tagged queue obtains the transmission opportunity, L PDUs buffered at this queue are transmitted at this DL sub-frame. The following assumptions are made in the performance analysis:

(1) a link layer SDU corresponds to an IP packet;
(2) each SDU is fragmented to F PDUs with an equal size of B bits;

TABLE I
STATE BOUNDARIES AND CORRESPONDING AMC LEVELS FOR IEEE 802.16 NETWORKS.

State ID	Modulation Level and Coding	Information (bit/OFDM symbol)	b_n (dB)
0	*silent*	0	0
1	BPSK(1/2)	96	3
2	QPSK(1/2)	192	6
3	QPSK(3/4)	288	8.5
4	16QAM(1/2)	384	11.5
5	16QAM(3/4)	576	15
6	64QAM(2/3)	768	18.5
7	64QAM(3/4)	864	21

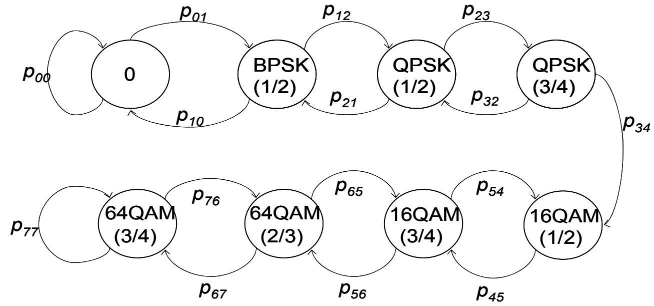


Fig. 2. Finite states Markov channel (FSMC) model.

(3) feedback information of PDUs transmitted at a DL sub-frame is sent back to the BS at the following UL sub-frame using the UL-ACK channel, which has been defined in IEEE 802.16e standard;
(4) resources are available for nrtPS traffic admitted into the network at each MAC frame. This assumption is reasonable provided with a well-defined connection admission control strategy adopted in the network;
(5) when a queue is scheduled, it has PDUs waiting for transmission.

The proposed design framework is based on the average throughput requirement and the tradeoff between the delivery delay and resource utilization. In terms of packet arrival, the only assumption in the analytical model is that when a queue is visited, it has packets waiting for transmission. For the bursty traffic, this assumption is usually satisfied.

A. Wireless Channel Model

Wireless channels suffer from deep fading that occurs randomly in the time span with a random duration and depth. Numerous studies have shown that such channels can be described by a Markov model to capture the bursty error nature. In this paper, a finite state Markov channel (FSMC) model [23] is used to describe the time-varying channel condition of each SS. In addition, when the channel model is constructed, the discrete AMC architecture defined in the IEEE 802.16 standard is taken into account, where the received signal-to-noise ratio (SNR) is divided into several disjoint regions. According to the perceived SNR, the BS selects a proper modulation level and coding scheme for each SS. The boundaries of SNR for the $(N+1)$-state FSMC is denoted as a row vector $B = [b_0, b_1, \cdots, b_N, b_{N+1}]$, where $b_0 = 0$ and $b_{N+1} = \infty$. When the received SNR is located in the set $[b_n, b_{n+1})$, the channel state is represented by state n . Based on the IEEE 802.16 standard, an 8-states Markov channel model is used in the paper. The values of the modulation and coding levels corresponding to each channel state are specified in Table I.

The channel states of a FSMC model are abstracted as shown in Fig. 2 . State 0 represents the state with no transmission permitted, which happens when the channel condition is very poor. In this case, the corresponding queue should not transmit any data in order to improve the system throughput. For simplicity, states $BPSK(1/2)$, $QPSK(1/2)$,$\cdots$, and $64QAM(3/4)$ are represented by symbol 1, 2,$\cdots$, and 7, respectively.

The probability of staying at state n (denoted as $\pi(n)$) is given by [23]

$$\pi(n) = \frac{\Gamma(m, m \cdot b_n/\bar{\gamma}) - \Gamma(m, m \cdot b_{n+1}/\bar{\gamma})}{\Gamma(m)} \ , \quad (1)$$

where $\bar{\gamma}$ is the average SNR, m is Nakagami fading parameter, $\Gamma(m)$ is the Gamma function, and $\Gamma(m,\gamma)$ is the complementary incomplete Gamma function. The channel becomes a Rayleigh fading channel when m = 1. For a slow fading channel (i.e., state transition occurs only between adjacent states), the state transition matrix for the FSMC can be expressed as follows

$$P = \begin{bmatrix} p_{00} & p_{01} & 0 & \cdots & 0 & 0 & 0 \\ p_{10} & p_{11} & p_{12} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & p_{N-2,N-1} & p_{N-1,N-1} & p_{N-1,N} \\ 0 & 0 & 0 & \cdots & 0 & p_{N,N-1} & p_{N,N} \end{bmatrix} \quad (2)$$

The transition probability from state n to k (denoted as p_{nk}) is obtained as follows:

$$p_{n,n+1} = \frac{L_{n+1} \cdot T}{\pi(n)} \qquad n = 0, 1, \cdots, N-1 \quad (3)$$

$$p_{n,n-1} = \frac{L_n \cdot T}{\pi(n)} \qquad n = 1, 2, \cdots N \quad (4)$$

$$p_{n,n} = \begin{cases} 1 - p_{n,n+1} - p_{n,n-1}, & 0 < n < N \\ 1 - p_{0,1}, & n = 0 \\ 1 - p_{N,N-1}, & n = N \ , \end{cases} \quad (5)$$

where T is the time duration of a MAC frame, and L_n is the level crossing rate at b_n corresponding to state n, which can be estimated by

$$L_n = \sqrt{2\pi \frac{m \cdot b_n}{\bar{\gamma}}} \cdot \frac{f_d}{\Gamma(m)} \cdot (\frac{m \cdot b_n}{\bar{\gamma}})^{m-1} \cdot exp(\frac{m \cdot b_n}{\bar{\gamma}}) \quad (6)$$

where f_d is the Doppler frequency.

Although the slow fading channel is considered in this study, the following analysis on the delivery delay, goodput, and resource utilization is also valid for other types of channels. The only difference is the derivation of the state transition matrix given in (2). Similar Markovian models for other types of fading channels have been discussed in [24]–[26].

B. Analysis of the Service Probability for an SS

The service probability is defined as the probability for an SS to obtain the service at a DL sub-frame. Let the SS under discussion be referred to as the tagged SS. Firstly, we classify all SSs into three groups based on the channel state of the tagged SS at a specific MAC frame. Given the channel state of the tagged SS is at state n, the three groups are composed of the SSs with channel conditions better than, same as, and worse than the state n , respectively, which is denoted as the groups G_1, G_2, and G_3, respectively. Let k_1, k_2 and k_3 denote the number of SSs in the groups G_1, G_2, and G_3, respectively. The tagged SS, which belongs to the group G_2, obtains the chance of transmission only when the condition $k_1 < h$ holds. Otherwise, all the selected SSs should come from the group G_1. When the condition is satisfied, the probability that the tagged SS obtains the chance of transmission can be derived based on the values of k_1 and k_2. Since the total number of selected SSs is h, and k_1 SSs are at the channel states better than state n, $h - k_1$ is the quota left for the SSs at G_2 and G_3. When $h - k_1$ is larger than or equal to k_2, all the SSs at the group G_2 are selected, i.e., the tagged queue obtains the chance of transmission at this DL sub-frame with a probability 1. On the other hand, when $k_2 > h - k_1$, the BS randomly selects $h - k_1$ out of k_2 SSs in the group G_2. Therefore, the tagged queue obtains the chance of transmission with a probability $(h-k_1)/k_2$. The function $Q(\cdot)$ given in (7) takes these situations into account. k_1 is a value between the set $[0, h-1]$ such that the tagged SS can obtain the chance of transmission. Let M be the total number of SSs in the network. Under a specific value of k_1, k_2 is in the set of $[1, M-k_1]$. The set begins with 1 since at least the tagged SS is in the group G_2. When the values of k_1 and k_2 are given, k_3 is $M - k_1 - k_2$.

$$Q(\frac{h-k_1}{k_2}) = \begin{cases} 1, & h-k_1 \geq k_2 \\ \frac{h-k_1}{k_2}, & h-k_1 < k_2 \end{cases} \quad (7)$$

Let $\Omega(j_1)$, $\Omega(j_2)$, and $\Omega(j_3)$ denote the sets of SSs in groups G_1, G_2, and G_3, respectively. Thus, given that the channel condition of the tagged SS is at state n along with the specific $\Omega(j_1)$, $\Omega(j_2)$, and $\Omega(j_3)$, the probability that the tagged SS obtains the transmission opportunity can be expressed as (8), where the function Q is defined as (7), while $\prod_{i_1 \in \Omega(j_1)} pr(S_{i_1} > n)$, $\prod_{i_2 \in \Omega(j_2)} pr(S_{i_2} = n)$, and $\prod_{i_3 \in \Omega(j_3)} pr(S_{i_3} < n)$ denote the probability that all SSs in groups G_1, G_2, and G_3 are at the channel states better than, same as, and worse than state n, respectively. For instance, the system is composed of SS_1, SS_2, SS_3, SS_4, SS_5, and SS_6. The tagged SS is SS_1, which stays at channel state $n = 3$. Let $h = 3$, $k_1 = 2$, and $k_2 = 3$. Given that $\Omega(j_1)$, $\Omega(j_2)$, and $\Omega(j_3)$ are $\{SS_2, SS_3\}$, $\{SS_4, SS_5\}$, and $\{SS_6\}$, respectively, the probability that SS_1 obtains a transmission opportunity is given by (9).

Equation (8) is for the specific $\Omega(j_1)$, $\Omega(j_2)$, and $\Omega(j_3)$. In the following, the number of all possible $\Omega(j_1)$, $\Omega(j_2)$, and $\Omega(j_3)$ is taken into consideration. Let a_1 represent the number of possible combinations for selecting k_1 SSs out of $M - 1$ SSs to construct the group G_1, where M is the total number of SSs in the network. After SSs in G_1 are selected, there are $M - k_1$ SSs left. Let a_2 represent the number of possible combinations for selecting $k_2 - 1$ SSs out of the left $M - k_1 - 1$ SSs to construct the group G_2. At last, the left $M - k_1 - k_2$ SSs consist of the group G_3. We have $a_1 = \binom{M-1}{k_1}$ and $a_2 = \binom{M-k_1-1}{k_2-1}$. In other words, given k_1, the total number of possible $\Omega(j_1)$ is a_1, and the set of all possible $\Omega(j_1)$ is represented by $\{\Omega(j_1), j_1 = 1, 2, \cdots a_1\}$. Given $\Omega(j_1)$, the total number of possible $\Omega(j_2)$ is a_2, and the set of all possible $\Omega(j_2)$ is represented by $\{\Omega(j_2), j_2 = 1, 2, \cdots a_2\}$. Note that given $\Omega(j_1)$ and $\Omega(j_2)$, the number of possible $\Omega(j_3)$ is 1 since G_3 is composed of all the left SSs that belong to neither G_1 nor G_2. In other words, j_3 is always 1. Thus, the service probability of the tagged SS at the channel state n is given by (10).

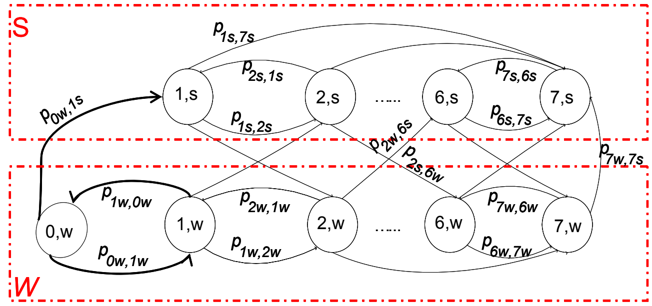


Fig. 3. The Markov model for the tagged SS.

Note that $\sigma_{S_0} = 0$ since the tagged queue is not allowed to transmit when the channel state of the tagged SS is at state 0, considering a high error bit rate at such a poor channel condition.

C. Analysis of Inter-service Time

An integrated Markov model is constructed to describe the channel states of the tagged SS, where each state represents the current channel state of the tagged SS and whether the tagged queue obtains the chance of transmission in the current DL sub-frame. It consists of $2N + 1$ ($N = 7$ in the study) states as shown in Fig. 3, where (n, s) and (n, w) represent that the tagged SS obtains and loses the chance of transmission, respectively, while its channel is at the state n. The transmission probability matrix is given by (11).

In order to derive the inter-service time, we group the states in Fig. 3 into two states denoted as S and W, which represent the states where the tagged queue obtains and loses the chance of transmission, respectively.

The transition probabilities of the grouped states are given by

$$p_{sw} = \frac{\sum_{n=1}^{7} [\theta(n,s) \sum_{j=0}^{7} p_{ns,jw}]}{\sum_{n=1}^{7} \theta(n,s)}, \qquad p_{ss} = 1 - p_{sw}, \quad (12)$$

$$p_{ws} = \frac{\sum_{n=0}^{7} [\theta(n,w) \sum_{j=1}^{7} p_{nw,js}]}{\sum_{n=0}^{7} \theta(n,w)}, \qquad p_{ww} = 1 - p_{ws}, \quad (13)$$

where $\theta(n, s)$ is the steady-state probability of the state (n, s), and $p_{ns,jw}$ is the one-step transition probability from the state (n, s) to the state (j, w) $(n = 1, 2, \cdots, 7, j = 0, 1, \cdots, 7)$.

$$Q(\frac{h-k_1}{k_2})[\prod_{i_1\in\Omega(j_1)} Pr(S_{i_1} > n) \prod_{i_2\in\Omega(j_2)} Pr(S_{i_2} = n) \prod_{i_3\in\Omega(j_3)} Pr(S_{i_3} < n)] \tag{8}$$

$$\frac{1}{3}[\prod_{i_1\in\{SS_2,SS_3\}} Pr(S_{i_1} > 3) \prod_{i_2\in\{SS_4,SS_5\}} Pr(S_{i_2} = 3) \prod_{i_3\in\{SS_6\}} Pr(S_{i_3} < 3)] \tag{9}$$

$$\sigma_{S_n} = \begin{cases} \sum_{k_1=1}^{h-1} Q(\frac{h-k_1}{k_2}) \sum_{j_1=1}^{a_1} \prod_{i_1\in\Omega(j_1)} Pr(S_{i_1} > n) \sum_{k_2=1}^{M-k_1} \sum_{j_2=1}^{a_2} \prod_{i_2\in\Omega(j_2)} Pr(S_{i_2} = n) \prod_{i_3\in\Omega(j_3)} Pr(S_{i_3} < n) & n = 1,\cdots,7 \\ 0 & n = 0 \end{cases} \tag{10}$$

$$\underline{\underline{Q}} = \begin{bmatrix} p_{00} & p_{01} & p_{01} & \cdots & p_{0N} & p_{0N} \\ p_{10} & p_{11} & p_{11} & \cdots & p_{1N} & p_{1N} \\ p_{10} & p_{11} & p_{11} & \cdots & p_{1N} & p_{1N} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{N0} & p_{N1} & p_{N1} & \cdots & p_{NN} & p_{NN} \\ p_{N0} & p_{N1} & p_{N1} & \cdots & p_{NN} & p_{NN} \end{bmatrix} \cdot \begin{bmatrix} 1-\sigma_{S_0} & 0 & 0 & \cdots & 0 & 0 \\ 0 & \sigma_{S_1} & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1-\sigma_{S_1} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_{S_N} & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1-\sigma_{S_N} \end{bmatrix} \tag{11}$$

Let m denote the inter-service time, which is defined as the duration (in unit of frame) between two successive transmission chances for the tagged queue. The mean of m is given by

$$E[m] = \sum_{i=1}^{\infty} i p_{sw} (p_{ww})^{i-1} p_{ws} \ . \tag{14}$$

D. Analysis of Goodput

Let η denote the number of PDUs successfully launched by the tagged queue during each transmission opportunity. The mean of η is given by

$$E[\eta] = \sum_{i=0}^{L} [i \binom{L}{i} (1-p)^i p^{L-i}] = L \cdot (1-p) \ , \tag{15}$$

where L is the number of PDUs transmitted by the tagged queue each time it obtains the transmission opportunity.

Goodput achieved at the tagged queue is defined as the average data rate (in unit of bit per second) successfully launched by the tagged queue, and is given by

$$G = \frac{E[\eta] \cdot B}{T \cdot (E[m]+1)} = \frac{L \cdot (1-p) \cdot B}{T \cdot (E[m]+1)} \ bps \tag{16}$$

where $E[m]$ is the mean of the inter-service time, $E[\eta]$ is derived from (15), T is the time duration of a MAC frame, and B is the size of a PDU in unit of bit.

E. Analysis of Resource Utilization

The resource utilization (denoted as RU) is defined as the number of information bits carried by an OFDM symbol. When a high modulation level is used, the number of information bits carried by an OFDM symbol is large, which yields a high resource utilization, and vise versa. RU is an important metric to evaluate the performance of the proposed framework, which can be obtained as

$$RU = \sum_{i=1}^{M} [\frac{\sum_{n=1}^{7} \theta_i(n,s)}{\sum_{j=1}^{M} \sum_{n=1}^{7} \theta_j(n,s)} \sum_{k=1}^{7} I_k \frac{a_k \theta_i(k,s)}{\sum_{n=1}^{7} a_n \theta_i(n,s)}] \ , \tag{17}$$

where M is the total number of SSs in the network, $\theta_i(n,s)$ is the steady-state probability of the state (n,s) for SS_i, I_k is the information bits carried by an OFDM symbol when the channel state is at k, which is given in Table I, and a_n is the required OFDM symbols to transmit L PDUs at the tagged queue when the channel state is n.

F. Analysis of Delivery Delay of a PDU

The average number of transmission/retransmission for a PDU is given by

$$\sum_{n=1}^{\infty} n p^{n-1} (1-p) = 1 + \frac{p}{1-p} \tag{18}$$

where p is the error probability of each transmission of a PDU.

Thus, the average delivery delay of a PDU (in unit of frame) is given by

$$E[D_P] = \frac{p}{1-p}(E[m]+1) \tag{19}$$

where $E[m]$ is the mean of the inter-service time.

G. Analysis of Delivery Delay of an SDU/Packet

The delivery delay of an SDU is defined as the number of MAC frames counted from the launching of the first PDU of the SDU until the successful recipient of the last PDU of the SDU. Let the random variable N denote the number of transmission opportunities experienced by a queue to successfully transmit an SDU, and m_i be the ith inter-service time. The mean of the delivery delay of an SDU is

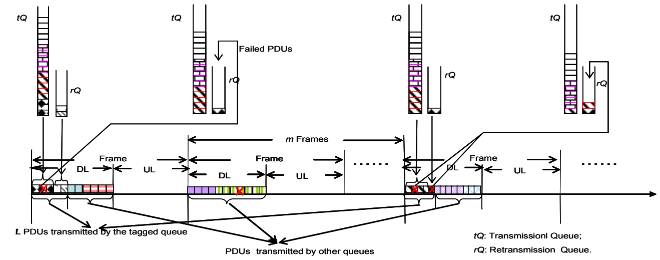


Fig. 4. The two-queue model for the delivery of SDUs.

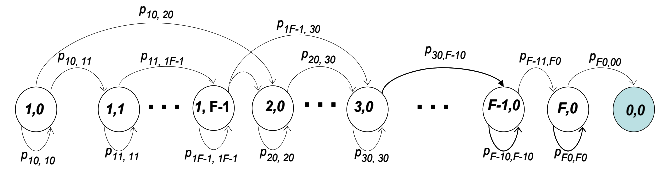


Fig. 5. The state transition diagram of the absorbing Markov chain.

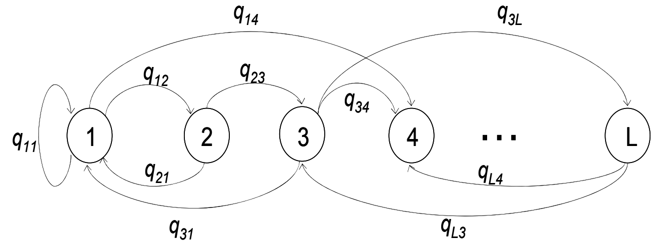


Fig. 6. The state transition diagram of HOL PDUs at the tQ queue.

given by

$$E[D_S] = E[\sum_{i=1}^{N-1}(m_i+1)] = (E[N]-1)(E[m]+1) \quad (20)$$

where $E[N]$ is the expected number of transmission opportunities experienced by the queue to successfully transmit an SDU, and $E[m]$ is the average inter-service time, which has been derived from (14). Thus, we need to obtain $E[N]$ in order to derive $E[D_s]$.

The Calculation of $E[N]$ *:* We propose a two-queue model to evaluate the delivery of SDUs at the tagged queue of the BS, where two logic queues, called *transmission queue* (tQ) and the *retransmission queue* (rQ), are devised as shown in Fig. 4. The tQ buffers the PDUs while the rQ is used to buffer the failed PDUs. The total number of launched PDUs in each transmission opportunity by the tagged queue, L, is composed of two parts: (1) the PDUs from the rQ due to the transmission failure in the previous transmission opportunity; (2) the PDUs from the tQ. We assume that the retransmitted PDUs get a higher priority than the PDUs in the tQ. In other words, the SS that obtains the transmission opportunity will launch all the PDUs in the rQ queue first, followed by transmitting the PDUs in the tQ queue by the number of leftover quota out of L.

Let an arbitrary SDU in the *tagged* queue be referred to as the *tagged* SDU, all PDUs belonging to the *tagged* SDU be referred to as the *tagged* PDUs, τ_1 be the time instant at which the tagged queue wins a transmission opportunity and launches the first tagged PDU, the subsequent instants at which the tagged queue obtains the transmission opportunity be denoted as $\{\tau_n : n > 0\}$, and the random variables $A_n \in \{0, 1, \cdots F\}$ and $f_n \in \{0, 1, \cdots F\}$represent the number of the tagged PDUs in the tQ and the rQ observed at the instants $\{\tau_n : n > 0\}$, respectively. The process $\{A_n, f_n : n = 1, 2, \cdots\}$ forms an absorbing embedded Markov chain on the state space $\{(0, 1, 2, \cdots F) \times (0, 1, 2, \cdots F)\}$, as shown in Fig. 5, which represents the state transition of the tagged PDUs. The state $(0, 0)$ is the absorbing state, representing that all the tagged PDUs are launched successfully. When the system reaches the absorbing state $(0, 0)$, the tagged SDU is completely transmitted. Thus, the one-step transition probability matrix of this Markov chain is given by

$$P = [p_{ij,\ i'j'}] \quad i, j, i', j' \in \{0, 1, 2, \cdots , F\} \quad (21)$$

where P is a $(F + 1)(F + 1) \times (F + 1)(F + 1)$ matrix, and the element $p_{ij,i'j'}$ denotes the transition probability from the state (i, j) to the state (i', j').

Let x_{in} represent the number of the tagged PDUs at the tQ that will be transmitted at the next transmission opportunity, and x_{out} denote the number of the tagged PDUs successfully launched in the current transmission opportunity. Therefore, the transition from state (i, j) to state (i', j') specifically determines a pair of (x_{in}, x_{out}), which is given by

$$x_{in} = j - j' \ , \qquad x_{out} = (i + j) - (i' + j') \quad (22)$$

Hence, the transition probability $p_{ij,i'j'}$ is given by (23), where T_1, T_2,and T_3 are the subsets of Q, T_4 is the complementary set of Q in T, Q is a subset of S, and S and T are the state space and transition space of the absorbing Markov chain, respectively. Hence, the expected number of transmission opportunities required to successfully transmit the tagged SDU is equivalent to the average number of steps to absorption for the tagged SDU, which is given by:

$$E[N] = \Pi_0(I - R)^{-1}e \ , \quad (24)$$

where Π_0 is the initial state vector, I is the identity matrix, R is the matrix derived from the one-step transition probability by deleting the row and column corresponding to the absorbing state $(0, 0)$, and e is a column vector with all elements equal to 1. In order to calculate $E[N]$, we need to know the initial state vector Π_0, which is analyzed as follows. First, the PDUs at the tQ is indexed with the mod L. Let the successive time instant the tagged queue obtains the transmission opportunities be denoted as $\{\tau_n : n > 0\}$ and random variable $\phi_n \in \{1, 2, \cdots L\}$ be the index of the head-of-line (HOL) PDU at the tQ observed at $\{\tau_n : n > 0\}$. The process $\{\phi_n : \ n = 1, 2, \cdots\}$ forms a embedded Markov chain on the state space $\{1, 2, \cdots , L\}$, as shown in Fig. 6.

The state transition probability of this Markov chain is determined by the number of PDUs at the tQ being launched at each transmission opportunity, which is equivalently the number of PDUs failed in the previous transmission opportunity. Therefore, the one-step transition probability from the

$$p_{ij,\ i'j'} = \begin{cases} [\binom{i}{x_{out}}(1-p)^{x_{out}}p^{i-x_{out}}][\binom{L-i}{x_{in}-x_{out}}(1-p)^{x_{in}-x_{out}}p^{(L-i)-(x_{in}-x_{out})}] & (i,j,i',j') \in T_1 \\ [\binom{i}{x_{out}}(1-p)^{x_{out}}p^{i-x_{out}}][\sum_{m=x_{in}-x_{out}}^{L-i}\binom{L-i}{m}(1-p)^{m}p^{(L-i)-m}] & (i,j,i',j') \in T_2 \\ \binom{i}{x_{out}}(1-p)^{x_{out}}p^{i-x_{out}} & (i,j,i',j') \in T_3 \\ 0 & (i,j,i',j') \in T_4 \end{cases}, \quad (23)$$

$$T_1 := \{(i,j,i',j') \in Q \mid j, j' \neq 0, x_{in} \geq x_{out}\}, \qquad T_2 := \{(i,j,i',j') \in Q \mid j \neq 0, j' = 0, x_{in} \geq x_{out}\},$$
$$T_3 := \{(i,j,i',j') \in Q \mid x_{in} < x_{out}\}, \qquad T_4 := \overline{Q}$$
$$Q := \{(i,j,i',j') \in T \mid 0 \leq x_{in} < j, 0 \leq x_{out} \leq min(x_{in}, j)\},$$
$$T := \{(i,j,i',j') \mid (i,j) \in S, (i',j') \in S\}, \qquad S := \{(i,j) \mid 0 \leq i \leq F, 0 \leq j \leq F-i\}$$

TABLE II
AVERAGE SNR OF SSs.

Index of SSs	1–5	6–10	11-15	16-20
Average SNR (dB)	10	15	20	25

state i to state j, q_{ij}, is given by:

$$q_{ij} = \begin{cases} \binom{L}{j-i}(1-p)^{j-i}p^{L-(j-i)} & j > i \\ (1-p)^L & j = i \\ \binom{L}{L+j-i}(1-p)^{L+j-i}p^{i-j} & j < i \end{cases} \quad (25)$$

Based on the one-step transition probability, the steady-state probability $h_i = \lim_{n\to\infty} Pr(\phi_n = i)(i = 1, 2, \cdots, L)$ can be derived from the balance equations. The state transition from state i to state j determines the occurrence of some specific initial states. For instance, the transition from state 1 to the state L, which is due to successful transmission of $(L-1)$ PDUs at the previous transmission opportunity, implies that the initial state $(F, 0)$ occurs $\lfloor (L-1)/F \rfloor$ times, and the initial state $(F-1, 1)$ occurs once concurrently. Since the transition from state i to state j occurs with the probability $h_i q_{ij}$, the probability that the corresponding initial states occur can be obtained accordingly. Therefore, the initial state vector Π_0 can be derived based on the derived steady-state probability h_i and the one-step transition probability matrix q_{ij}. When the initial state probability vector Π_0 is obtained, the average number of steps for an SDU to the absorption and the average delivery delay of an SDU can be derived from (24) and (20), respectively.

IV. NUMERICAL RESULTS

Extensive simulations are conducted to demonstrate the efficiency of the proposed framework and illustrate the impacts of two parameters, h and L, on the performance metrics in terms of PDU delivery delay, SDU delivery delay, achieved goodput, and resource utilization. We repeat the simulation 50 times with different random seeds and calculate the average value. In the simulation, a Rayleigh fading channel model is adopted, and the total number of SSs is 20, which are divided into four groups. The average SNR of each SS is given in Table II, and the other simulation parameters are given in Table III.

Fig. 7 shows the impacts of the parameter h on the SDU delivery delay for different SSs. It can be seen that the SDU

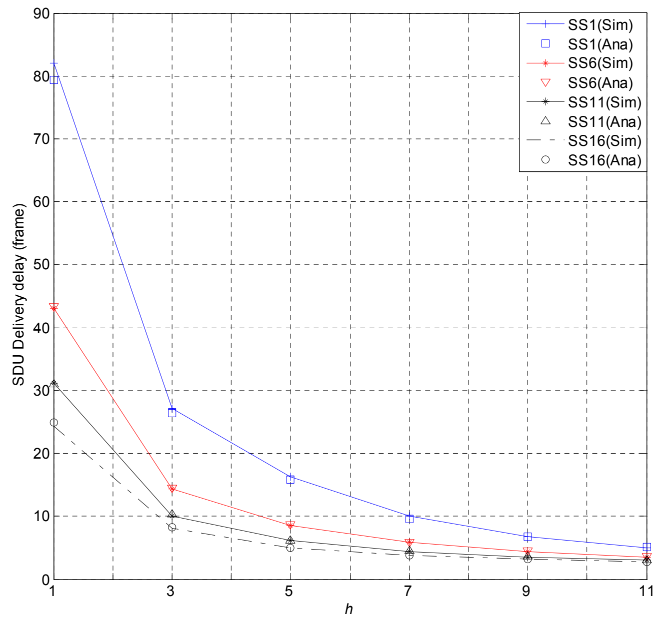


Fig. 7. The SDU delivery delay versus h.

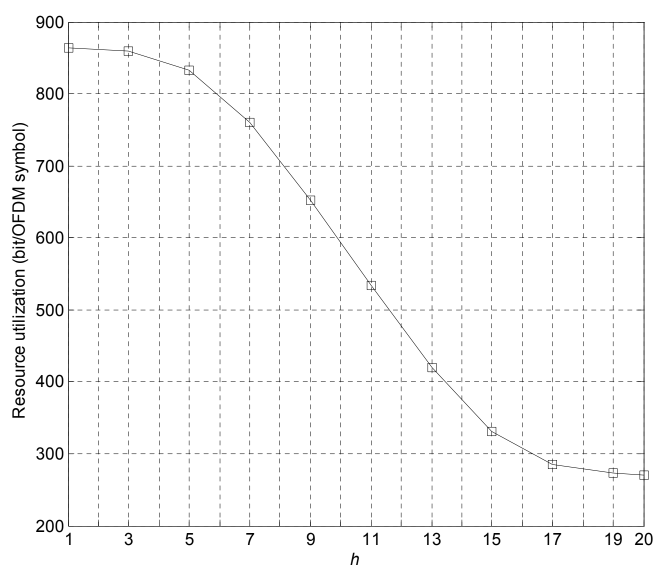


Fig. 8. The resource utilization versus h.

TABLE III
SIMULATION PARAMETERS.

DL/UL sub-frame duration	OFDM symbol duration	Channel bandwidth
1.25/1.25 ms	23.8us	10MHz
F	Doppler frequency	p
5	15Hz	0.01

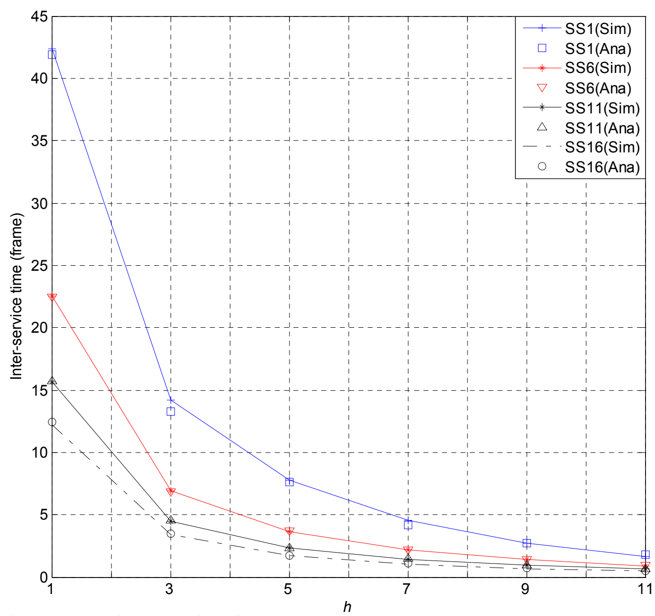


Fig. 9. The inter-service time versus h.

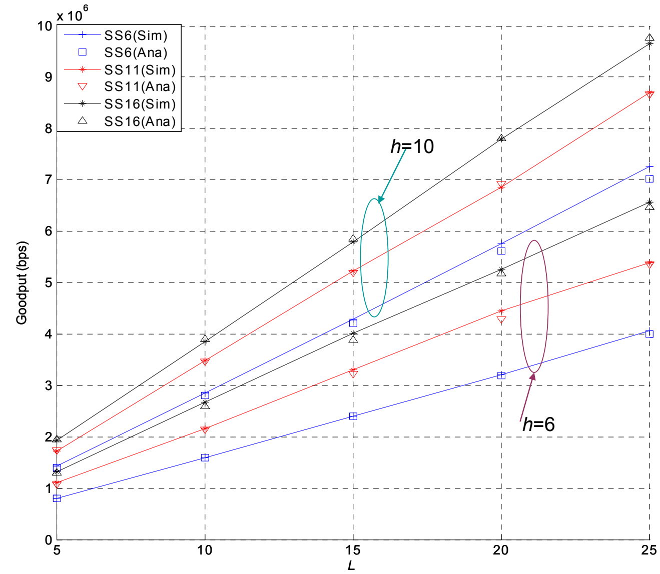


Fig. 10. The goodput versus L.

delivery delay decreases with the increase of h. When h is 1, one SS is selected to obtain the transmission opportunity at each DL sub-frame. Thus, SS1, which is subject to the worst channel condition, experiences a quite long delivery delay. With the increase of h, SS1 gets more chances of transmissions, therefore, its SDU delivery delay decreases accordingly.

Fig. 8 shows the impacts of h on the resource utilization. It is observed that the resource utilization decreases with the increase of h. With a large h, some SSs with poor channel conditions can achieve the chance of transmission, which leads to the adoption of a low modulation and coding level. As a result, a low resource utilization is obtained. From Figs. 7–8, it can be seen that the parameter h plays a key role in manipulating the SDU delivery delay and the resource utilization. A small h leads to a high resource utilization, but it also leads to a long SDU delivery delay for the SSs with poor channel conditions, and vice versa.

Fig. 9 shows the relation between the inter-service time and h. It is observed that the inter-service time of each SS decreases with the increase of h. When h is 1, the inter-service time of SS1 is almost 4 times as that of SS16.

Fig. 10 shows the relation between the achieved goodput and assigned bandwidth L with different h. It is observed that given a specific h, the goodput requirement can be achieved by manipulating a proper L. With the increase of L, the achieved goodputs of SSs increase accordingly. With a fixed L, an SS with better channel condition achieves a higher throughput since it can get more chances of transmissions. Meanwhile, the achieved goodput is also affected by the selection of h. With a larger h, SSs can obtain a higher goodput.

Figs. 7–10 illustrate the impacts of h and L on several important performance metrics. h and L are critical for the performance of the proposed framework. The setting of h is based on the requirement of the tradeoff between the SDU delivery delays and the resource utilization. The possible h is a value in the set of $[1, 2, \cdots, M]$, where M is the total number of SSs in the networks. Since the number of possible h equals M, the SDU delivery delay and resource utilization corresponding to each possible h can be obtained by using the analysis given in Section III. When the sets of possible SDU delivery delays and resource utilization are obtained, the parameter h can be selected based on the requirement of the SDU delivery delay and the resource utilization. After h is set, L can be set by using the analytical result given in (16), based on the throughput requirements of each SS. Furthermore, the parameters h and L decide the amount of resource assigned to nrtPS applications at each MAC frame, which can provide a useful guideline for the connection admission control such that the system will not be overloaded.

From Figs. 7–10, it is also observed that the simulation results match very well with the analysis results, which verifies the accuracy of the developed analytical model.

V. CONCLUSIONS

A simple yet efficient resource allocation and packet scheduling framework has been proposed for nrtPS applications in IEEE 802.16 networks, where selective ARQ at the MAC layer and AMC technique at the physical layer are jointly considered. An analytical model has been developed to provide practical guidelines to select appropriate parameters for satisfying the throughput requirement and initiating a graceful compromise between the delivery delay and the resource utilization for nrtPS applications. Extensive simulations have been conducted to demonstrate the effectiveness and efficiency of the proposed framework and verify the accuracy of the analytical model. In this paper, we have focused on unicast scheduling for nrtPS applications. Another important and promising application in IEEE 802.16 networks is the multimedia multicast application, such as IPTV, which is expected to serve as a killer application provided by the Internet Service Providers (ISPs) in the future. Our further research on the multicast scheduling for providing high quality IPTV service in IEEE 802.16 network is under way.

Acknowledgment

This work is partially supported by a Strategic Research Grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada.

References

[1] IEEE 802.16a$^{\mathrm{TM}}$-2003, IEEE Standard for local and metropolitan access network part 16: Air Interface for Fixed Broadband Wireless Access Systems - Amendment 2: Medium Access Control Modifications and Additional Physical Layer Specifications for 2-11 GHz, Apr. 2003.

[2] K. Wongthavarawat and A. Ganz, "Packet scheduling for QoS support in IEEE 802.16 broadband wireless access systems," *Int. J. Commun. Syst.*, vol. 16, no. 1, pp. 81-96, 2003.

[3] C. Fraleigh, S. Moon, B. Lyles, *et al*., "Packet-level traffic measurements from the sprint IP backbone," *IEEE Network*, vol. 17, no. 6, pp. 6-16, 2003.

[4] "Universal Mobile Telecommunications System (UMTS): Selection procedures for the choice of radio transmission technologies of the UMTS (UMTS 30.03 version 3.2.0)." URL:http://www.3gpp.org/ftp/Specs/archive/30 series/30.03U/3003U-320.zip

[5] X. Liu, E. K. P. Chong, and N. B. Shroff, "A framework for opportunistic scheduling in wireless networks," *Int. J. Computer and Telecommun. Networking*, vol. 41, no. 4, pp. 451-474, 2003.

[6] L. Xu, X. Shen, and J. W. Mark, "Dynamic fair scheduling with QoS constraints in multimedia wideband CDMA cellular networks," *IEEE Trans. Wireless Commun.*, vol. 3, no. 1, pp. 60-73, 2004.

[7] Y. Liu and E. Knightly. "Opportunistic fair scheduling over multiple wireless channels," in *Proc. IEEE INFOCOM'03*, San Francisco, CA, June 2003.

[8] H. J. Kushner and P. A. Whiting, "Asymptotic properties of proportional-fair sharing algorithms," in *Proc. Allerton Conf. 2002*, University of Illinois Press, Champaign-Urbana, IL, 2002.

[9] D. Yang, D. Shen, W. Shao, and V. O. K. Li, "Towards opportunistic fair scheduling in wireless networks," in *Proc. IEEE Int. Conf. on Commun. (ICC'06)*, vol. 11, pp. 5217-5221, Istanbul, Turkey, June 2006.

[10] D. I. Kim, "Selective relative best scheduling for best-effort downlink packet data," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1254-1259, 2006.

[11] M. Mehrjoo, M. Dianati, X. Shen, and K. Naik, "Opportunistic fair scheduling for the downlink of IEEE 802.16 wireless metropolitan area networks," in *Proc. Qshine'06*, Waterloo, Ontario, Canada, Aug. 2006.

[12] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "Providing quality of service over a shared wireless link," *IEEE Commun. Mag.*, vol. 39, no. 2, pp. 150-154, 2001.

[13] J. H. Rhee, J. M. Holtzman, and D. K. Kim, "Performance analysis of the adaptive EXP/PF channel scheduler in an AMC/TDM system," *IEEE Commun. Lett.*, vol. 8, no. 8, pp. 497-499, 2004.

[14] H. Wei and R. Izmailov, "Channel-aware soft bandwidth guarantee scheduling for wireless packet access," in *Proc. IEEE Wireless Commun. and Networking Conference (WCNC'04)*, Atlanta, Georgia, Mar. 2004.

[15] V. Hassel, M. R. Hanssen, and G. E. Øien, "Spectral efficiency and fairness for opportunistic round robin scheduling," in *Proc. IEEE Int. Conf. on Commun. (ICC'06)*, vol. 2, pp. 784-789, Istanbul, Turkey, June 2006.

[16] S.-I. Hahm, H.-J. Lee, and J.-W. Lee, "A minimum-bandwidth guaranteed scheduling algorithm for data services in CDMA/HDR system," *Lecture Notes in Computer Science*, no. 2713, pp. 757-763, 2003.

[17] J. H. Lee, S. K. Kim, S. C. Lee, Y. J. Tcha, and C. G. Kang, "Packet scheduling algorithm for non-real-time service with soft QoS requirement in mobile broadband wireless access system," in *Proc. IEEE Int. Conf. on Commun. (ICC'06)*, vol. 11, pp. 5240-5245, Istanbul, Turkey, June 2006.

[18] W.-C. Wu, S. Vassiliadis, and T. Y. Chung, "Performance analysis of multi-channel ARQ protocols," in *Proc. 36th Midwest Symposium on Circuits and Systems*, Detroit, MI, USA, Aug. 1993.

[19] J. G. Kim and M. M. Krunz, "Delay analysis of selective repeat ARQ for a Markovian source over a wireless channel," *IEEE Trans. Veh. Technol.*, vol. 49, no. 5, pp. 1968-1981, 2000.

[20] H. Shen, L. Cai, and X. Shen, "Performance analysis of TFRC over wireless links with truncated link level ARQ," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1479-1487, June 2006.

[21] F. Hou, P. H. Ho, and X. Shen, "A novel differentiated retransmission scheme for MPEG video streaming over wireless links," *Int. J. Wireless and Mobile Computing*, vol. 1, nos. $3/4$, pp. 260-267, 2006.

[22] W. Wang, Z. Guo, X. Shen, and C. Chen, "Performance analysis of ARQ scheme in IEEE 802.16," in *Proc. IEEE GLOBECOM'06*, San Francisco, USA, Nov. 2006.

[23] D. Niyato and E. Hossain, "A queuing-theoretic and optimization-based model for radio resource management in IEEE 802.16 broadband wireless networks," *IEEE Trans. Computers*, vol. 55, no. 11, pp. 1473-1488, Nov. 2006.

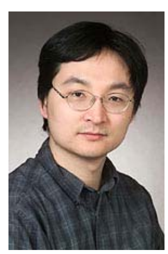
[24] F. Babich and G. Lomardi, "A Markov model for the mobile propagation channel," *IEEE Trans. Veh. Technol.*, vol. 49, no. 1, pp. 63-73, Jan. 2000.

[25] H. Wang and N. Moayeri, "Finite-state Markov channel–a useful model for radio communication channels," *IEEE Trans. Veh. Technol.*, vol. 44, no. 1, pp. 163-171, Feb. 1995.

[26] C. Pimentel, T. H. Falk, and L. Lisbôa, "Finite-state Markov modeling of correlated Rician-fading channels," *IEEE Trans. Veh. Technol.*, vol. 53, no. 5, pp. 1491-1501, Sept. 2004.

Fen Hou is currently a doctoral student in the Department of Electrical and Computer Engineering, University of Waterloo, Canada. Her research interests include resource allocation, connection admission control, and scheduling scheme in broadband wireless access networks.

James She received a B.Eng. degree in computer engineering and an M.Phil. in electrical and electronic engineering from the Hong Kong University of Science and Technology, China. He is currently a Ph.D. candidate in the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His current research interests include wireless IPTV systems and networks, wireless sensor devices and networks, and media content delivery networks.



Pin-Han Ho received his B.Sc. and M.Sc. Degree from the Electrical and Computer Engineering department in the National Taiwan University in 1993 and 1995. He started his Ph.D. study in the year 2000 at Queen's University, Kingston, Canada, focusing on optical communications systems, survivable networking, and QoS routing problems. He finished his Ph.D. in 2002, and joined the Electrical and Computer Engineering department in the University of Waterloo, Waterloo, Canada, as an assistant professor at the same year. Professor Pin-Han Ho is the first author of more than 60 refereed technical papers and book chapters, and the co-author of a book on optical networking and survivability. He is the recipient of the Best Paper Award and the Outstanding Paper Award from SPECTS'02 and HPSR'02, respectively.

Xuemin (Sherman) Shen received the B.Sc. (1982) degree from Dalian Maritime University (China) and the M.Sc. (1987) and Ph.D. degrees (1990) from Rutgers University, New Jersey (USA), all in electrical engineering. From September 1990 to September 1993, he was first with the Howard University, Washington D.C., and then the University of Alberta, Edmonton (Canada). Since October 1993, he has been with the Department of Electrical and Computer Engineering, University of Waterloo, Canada, where he is a Professor. Dr. Shen's research focuses on mobility and resource management in interconnected wireless/wired networks, UWB wireless communications systems, wireless security, and ad hoc and sensor networks. He is a coauthor of three books, and has published more than 200 papers in wireless communications and networks, control and filtering.