Dynamic Bandwidth Reservation Scheme in 802.11 and 802.16 Interworking Networks

Li-Ping Tung[1], Yeali S. Sun[2], and Meng Chang Chen[1]
[1]Institute of Information Science, Academia Sinica, Taipei, Taiwan
[2]Dept. of Information Management, National Taiwan University, Taipei, Taiwan
lptung@iis.sinica.edu.tw, sunny@ntu.edu.tw, mcc@iis.sinica.edu.tw

Abstract— **One of the most popular applications of IEEE 802.16 network is to serve as a backhaul service for IEEE 802.11 networks. However, the traffic of an IEEE 802.16 connection aggregated from IEEE 802.11 networks fluctuates. Thus, efficient bandwidth reservation at the subscriber station (SS) is an importance issue. This study proposes a simple and flexible bandwidth reservation scheme at the SS, called multi-stage self-correction bandwidth reservation (MSBR), to make effective use of the bandwidth without violating the QoS requirements for real-time traffic under the proposed cost model. The MSBR scheme introduces the concept of Decision Period for bandwidth reservation to reduce the control message overheads. The proposed method also adopts the RLS algorithm to predict the traffic arrival and applies the MSBR method to capture the traffic dynamics for bandwidth reservation. Simulation results demonstrate that the proposed MSBR scheme utilizes the bandwidth efficiently without violating the QoS requirements of real-time services.**

Keywords- dynamic bandwidth reservation; bandwith request; real-time services; IEEE 802.16; recursive-least-squares algorithm

I. Introduction

IEEE 802.16 standards have been developed for metropolitan broadband wireless access (BWA) systems. Due to their high data rate, large network coverage, and QoS capacity, IEEE 802.16 networks offer a wide variety of applications, including backhaul services for IEEE 802.11 hotspots and high-speed Internet access. This study considers a scenario in which the Internet Service Provider provides IEEE 802.16 networks to act as a backhaul service for IEEE 802.11 networks, as Fig. 1 illustrates. The IEEE 802.16 network operates in Point-to-MultiPoint (PMP) mode. The subscriber station (SS) collects traffic from one or several access points of the IEEE 802.11 networks and delivers the collected traffic to the base station (BS). The BS then forwards traffic from the SS to the Internet. This study considers uplink real-time traffic, e.g., VoIP with silence suppression. Providing an end-to-end QoS guarantee in such interworking networks remains a challenge.

Providing a real-time traffic end-to-end QoS guarantee in an integrated IEEE 802.11 and IEEE 802.16 network is an important resource management issue. A simple way to achieve this involves one-by-one QoS mapping and per-flow resource management. In other words, one IEEE 802.11 flow maps into one IEEE 802.16 connection and bandwidth requests are made independently [1]. However, for an IEEE 802.16 network serving as a backhaul network, the frequency of activating and terminating connections can be very high, incurring a high cost in terms of bandwidth request signaling and provisioning. Therefore, this study considers an aggregate resource management approach that combines multiple IEEE 802.11 flows with the same QoS level into a single IEEE 802.16 connection. This approach reduces the overhead caused by frequent connection requests.

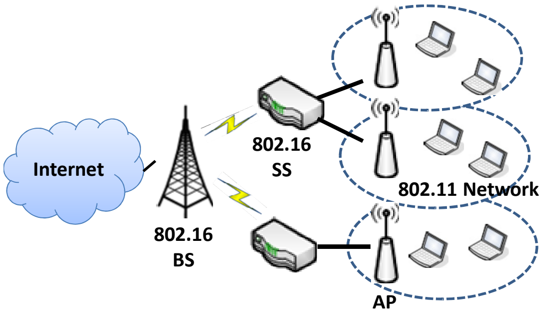


Figure 1. The network architecture

The traffic of an IEEE 802.16 connection aggregated from IEEE 802.11 networks will no doubt fluctuate. Thus, another problem that may arise is how to design an efficient bandwidth request-allocation algorithm that effectively uses bandwidth within a connection without violating QoS. To support QoS for various types of traffic, the IEEE 802.16 MAC protocol defines five types of scheduling classes: Unsolicited Grant Service (UGS), real-time Polling Service (rtPS), non-real-time Polling Service (nrtPS), Best-Effort (BE), and extended real-time Polling Service (ertPS). Both UGS and rtPS support real-time services. While UGS is suitable for constant bit rate (CBR) traffic, such as VoIP, rtPS is for variable bit rate (VBR) traffic, such as MPEG traffic. The UGS scheduling mechanism can minimize delay in bandwidth request-allocation process, but might waste bandwidth or suffer from insufficient bandwidth with VBR traffic. On the other hand, the rtPS mechanism effectively utilizes bandwidth at the cost of additional delay due to on-demand bandwidth request. IEEE 802.16-2004 introduced ertPS to strike a balance between minimizing delay and maximizing utilization. Like UGS, the ertPS scheduling mechanism allocates bandwidth periodically without any request to minimize the delay. Similar way to rtPS, ertPS can adjust the size of bandwidth allocation to maximize utilization.

Due to the lack of standards for bandwidth request-allocation algorithms, previous studies [2]-[3] focus on

This research is partly supported by National Science Council under grants NSC 97-2221-E-002-080-MY2 and NSC 98-2221-E-001-005-MY3

designing an uplink bandwidth request algorithm for the polling services. These efforts are aimed at minimizing the bandwidth wastage or PDU dropping probability. One study [2] proposes queue-aware uplink bandwidth allocation at the SS for polling services. The proposed scheme adaptively allocates bandwidth to control queue length at a target level and meet the delay and PDU dropping probability requirements. However, the deficiency of the proposed model is that allocation is performed on a frame basis, which introduces a lot of signaling overhead. Another study [3] proposes a bandwidth request algorithm to minimize bandwidth wastage in ertPS polling services. In this design, the bandwidth allocation interval equals the packetized interval of the real-time service. If the packetized interval is too large, the traffic amount during the period cannot be served in time because the allocated resources are fixed within the allocation interval for ertPS services. On the other hand, if the packetized interval is too small, this method produces the same drawbacks as [2], introducing significant signaling overhead.

This study considers an integrated IEEE 802.11 and IEEE 802.16 network that combines multiple IEEE 802.11 real-time traffic flows with the same QoS level into a single IEEE 802.16 connection. To make good use of bandwidth, the bandwidth aggregation is carried on a rtPS (real-time Polling Service) connection. Since there are no standardized bandwidth request-allocation algorithms for the rtPS service, this paper proposes a simple, flexible, and efficient uplink bandwidth request-allocation algorithm that reduces the control message exchanges between the BS and SS per frame and effectively uses bandwidth. This study introduces the concept of a Decision Period (DP), which consists of M frames. The SS issues a bandwidth request message with M reservation values per DP. This allows the SS to reserve the bandwidth for the subsequent frames in advance using a single control message rather than M control messages. The benefits of the DP are twofold; (i) it reduces the control message exchange between BS and SS per frame since the SS issues only one bandwidth request message rather than M messages; (ii) it reduces the delay incurred from the bandwidth request-allocation process since the resource allocation for multiple frames can be made in advance using one control message.

In TDMA scheduling, time slots are perishable resources [4]. In other words, the usage of resources is limited to a point of time, and time slot reservation is a time-dependent quantity. This study assumes that un-used time slot introduces the *resource reservation cost*. To decrease the resource reservation cost, it is possible to make conservative bandwidth request which may delay un-served demand to the subsequent frames. However, the delayed packet affects the QoS for real-time services, and incurs a *delay penalty cost*. Note that there is a tradeoff between the resource reservation cost and the delay penalty cost. Thus, to make good bandwidth reservation for multiple frames in a DP, the proposed method adopts a recursive-least-squares (RLS) adaptive filter as the bandwidth prediction algorithm. Furthermore, this study proposes a multi-stage self-correction bandwidth reservation (MSBR) scheme to correct the inaccuracy from the RLS algorithm by considering the tradeoff between the resource reservation cost and the delay penalty cost. Simulation results confirm that the proposed scheme utilizes bandwidth efficiently.

The rest of this paper is organized as follows. Section II reviews relevant literature. Section III presents the basic dynamic bandwidth reservation scheme and its simulation results. Section IV proposes a multi-stage self-correction bandwidth reservation scheme to amend the deficiencies of the prediction algorithm as well as simulation results. Section V provides a conclusion.

II. Related Work

There are several studies on dynamic bandwidth reservation ([2]-[3], [5]-[6]). The basic goal of dynamic bandwidth reservation is to minimize the amount of bandwidth being provisioned while keeping the cost of MAC signaling to a minimum or reducing the connection blocking probability. One study [2] proposes an uplink bandwidth request scheme that allocates bandwidth adaptively according to the queue state on a frame basis. This study uses an analytical discrete time Markov chain model of the bandwidth allocation to analyze the queueing performance, and assumes that the traffic source is a Markov Modulated Poisson Process. However, the proposed scheme makes no assumption of the traffic model in resource demand prediction. Another study [3] proposes an uplink bandwidth request-allocation algorithm for VBR real-time services. That algorithm calculates the amount of bandwidth request such that the delay is regulated around the desired level to minimize delay and delay jitter for real-time services. The bandwidth request also considers the mismatch between packet arrival and service rates. However, the bandwidth request interval is equal to the packetized interval of the real-time service. With a smaller interval, the BS must schedule more frequently. With a larger interval, the BS cannot capture the traffic dynamics fast enough since the allocated resources are fixed within the allocation interval for the ertPS service. In the proposed algorithm, the DP concept allows the BS to perform the scheduling algorithm and admission control in a more predictive way. Other studies [5] and [6] consider an integrated IEEE 802.11 and IEEE 802.16 network, and propose multi-thresholds-based dynamic bandwidth reservation schemes. The goal is to minimize the amount of bandwidth being provisioned while minimizing the cost of MAC signaling. However, this approach is a multi-threshold approach in which the reserved bandwidth does not change from one to another threshold until the number of connection being active reaches another threshold. Thus, it cannot react to the bandwidth dynamics in time. Furthermore, it also assumes that the connection arrival process is a Poisson process for analysis.

III. Dynamic Bandwidth Reservation Scheme

A. Problem Description

Consider an integrated IEEE 802.11 and IEEE 802.16 network that combines multiple IEEE 802.11 real-time traffic flows with the same QoS level into a single IEEE 802.16 rtPS connection. The traffic of an IEEE 802.16 connection aggregated from IEEE 802.11 networks is likely to fluctuate. The best way to make effective use of the bandwidth reserved is to monitor traffic and dynamically change the amount of bandwidth reserved in each frame. However, this approach may introduce large signaling overhead and delay. Therefore, this study introduces the concept of Decision Period for the bandwidth reservation. Each decision period consists of M frames. The SS issues a bandwidth request per DP, allowing the SS to reserve the bandwidth for the subsequent frames in advance using one control message rather than M control messages. If there are any un-served packets left in the end of the DP, they will be served as soon as possible at the beginning of the following DP.

There are two important factors in issuing the bandwidth request for resource reservation. The first one is that time slots are perishable resources [4]. Thus, the usage of this resource is limited to a point in time, making time slot reservation a time-dependent task. For example, if no packet is queued for transmission at a reserved time slot, this time slot resource is wasted. Thus, un-used time slots incur a *resource reservation cost*. On the other hand, demand can be accumulated. That is, the un-served demand can be served in the following frames. However, the delayed packet affects the QoS for real-time services. Thus, each delayed real-time packet incurs a *delay penalty cost*. Note that there is a tradeoff between the resource reservation cost and the delay penalty cost. On one hand, we wish to avoid over-allocation to reduce the resource reservation cost. On the other hand, the un-served demand will be delayed to the next frame and causes the delay penalty cost. In this paper, the cost of a decision period includes the signaling cost, the resource reservation cost, and the delay penalty cost for each frame within a DP. The goal is to design a bandwidth reservation scheme for the frames in a decision period that minimizes the DP cost by considering the tradeoff between the resource reservation cost and the delay penalty cost.

B. Bandwidth Reservation Model and Assumptions

In the proposed bandwidth reservation scheme, the SS issues a bandwidth request message in the last uplink subframe of DP *i*-1 to request the bandwidth allocation for the whole DP *i*, as Fig. 2 shows. The amount of bandwidth request depends on the predicted traffic arrival of frames in DP *i*. If there are any un-served packets at the end of DP *i*-1, they will be served as soon as possible at the beginning of DP *i*. This bandwidth request-allocation process can reduce packet delay since the amount of requested bandwidth is predicted in advance.

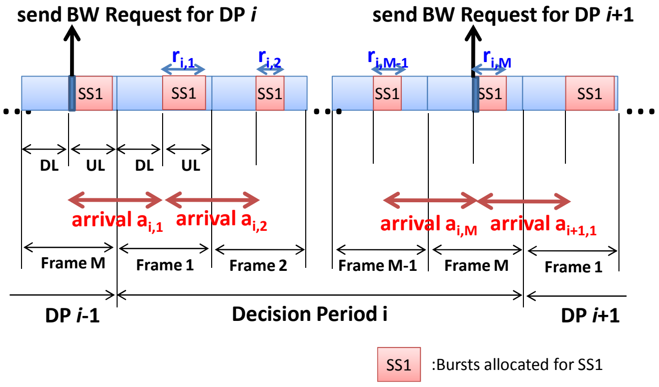


Figure 2. The structure of a decision period

The proposed bandwidth reservation model makes the following assumptions:

- The service queue employs a FIFO queueing discipline.
- Each data packet needs one time slot resource.
- Packets that have arrived in the downlink subframe of frame *j* are ready to be served in frame *j*.
- Packets that have arrived in the uplink subframe of frame *j* will be ready to be served in frame *j*+1.
- The BS grants the requested bandwidth to the SS.

The proposed bandwidth reservation scheme uses the following notations:

$\mathbf{A}_i = <a_{i,1}, a_{i,2}, \ldots, a_{i,M}>$: the new traffic arrival within DP *i*, where $a_{i,j}$ is the new traffic arrival ready to be transmitted in the uplink subframe of frame *j* of DP *i*.

$\mathbf{R}_i = <r_{i,1}, r_{i,2}, \ldots, r_{i,M}>$: the resource reservation amount within DP *i*, where $r_{i,j}$ is the resource reservation amount in the uplink subframe of frame *j* of DP *i*.

$\hat{\mathbf{A}}_i = <\hat{a}_{i,1}, \hat{a}_{i,2}, \ldots, \hat{a}_{i,M}>$: the predicted traffic arrival within DP *i*, where $\hat{a}_{i,j}$ is the predicted traffic arrival ready to be transmitted in the uplink subframe of frame *j* of DP *i*.

$e^{+}_{i,j}$: the prediction error between the over-estimation and the actual traffic arrival in frame *j* of DP *i*.

$$e^{+}_{i.j} = \hat{a}_{i,j} - a_{i.j}\,, \quad \hat{a}_{i,j} > a_{i.j} \tag{1}$$

$e^{-}_{i,j}$: the prediction error between the under-estimation and the actual traffic arrival in frame *j* of DP *i*.

$$e^{-}_{i.j} = a_{i,j} - \hat{a}_{i.j}\,, \quad a_{i,j} > \hat{a}_{i.j} \tag{2}$$

$q_{i,j}$: the backlog or queue size before the beginning of the uplink subframe in frame *j* of DP *i*.

$$q_{i,j} = \max(q_{i,j-1} - r_{i,j-1}, 0) + a_{i,j} \tag{3}$$

C_r: Unit Resource Reservation Cost. If one time slot is reserved for transmission but is not used, one unit of reservation cost is incurred.

$C^{r}_{i,j}$: the resource reservation cost in frame *j* of DP *i*. If the total demand is less than the total number of time slots reserved, it incurs the following resource reservation cost:

$$C^{r}_{i,j} = \begin{cases} (r_{i,j} - q_{i,j}) \times C_r \text{ ,if } r_{i,j} > q_{i,j} \\ \quad 0 \quad\quad\quad\quad , \quad \text{else} \end{cases} \; . \tag{4}$$

C_d: Unit Delay Penalty Cost. When a packet is delayed for one frame, it introduces one unit of delay penalty cost.

$C_{i,j}^{d}$: the delay penalty cost in frame *j* of DP *i*. If the total demand is larger than the total number of time slots reserved, the following delay penalty cost is incurred:

$$C_{i,j}^{d} = \begin{cases} (q_{i,j} - r_{i,j}) \times C_d & \text{,if } q_{i,j} > r_{i,j} \\ 0 & \text{, else} \end{cases} . \tag{5}$$

C$_s$: Unit Signaling Cost.

C_i^{DP} : DP cost. The cost of DP *i* consists of the signaling cost, the resource reservation cost, and the delay penalty cost of each frame within DP *i*.

$$\begin{aligned} C_i^{DP} &= C_S + \sum_{j=1}^{M} C_{i,j}^{r} + \sum_{j=1}^{M} C_{i,j}^{d} \\ &= C_S + \sum_{j=1}^{M} \max(r_{i,j} - q_{i,j}, 0) \times C_r \\ &\quad + \sum_{j=1}^{M} \max(q_{i,j} - r_{i,j}, 0) \times C_d \end{aligned} \tag{6}$$

According to (3) and (6), performing resource reservation (i.e., determine the r$_{i,j}$) is a challenging task in minimizing the DP cost. The major issue to be tackled here is how to predict the traffic arrival of frames within a DP in advance. Furthermore, any prediction errors that occur must be corrected to make effective use of resources.

C. Bandwidth Prediction

At the end of a decision period, the SS sends the BS a bandwidth request message specifying the forthcoming bandwidth amount required for the subsequent frames. Thus, the SS must on-line predict the traffic arrivals in advance.

Adaptive filters are a well-known class of on-line estimation techniques. One of their applications is prediction. The recursive-least-squares (RLS) algorithm is an adaptive filter algorithm that is simple and computationally efficient. The RLS algorithm has been widely used in control and communications applications, such as MPEG traffic prediction, handoff resources prediction, RTT prediction, and packet loss probability prediction ([7], [8]). Because network traffic is highly dynamic, it is desirable to have an algorithm that can adapt quickly. Thus, this paper uses the RLS algorithm to predict the traffic arrival of frames within a decision period.

Assume that the measured traffic arrival follows a random process {a$_{i,1}$, a$_{i,2}$, …}. The RLS prediction algorithm makes good use of the past measurements in forecasting future conditions. Consider a fixed amount of transmission history denoted as H-order. Then, a H-order predictor can be expressed as:

$$\hat{A}_i = W_1 A_{i-1} + W_2 A_{i-2} + \ldots + W_H A_{i-H} , \tag{7}$$

where $\mathbf{A}_i$ ($\hat{\mathbf{A}}_i$) is a M × 1 vector, and the j^{th} element a$_{i,j}$ (â$_{i,j}$) denotes the actual (predicted) traffic arrival in frame *j* of DP *i*. The predictor coefficient W$_h$ is a M × M matrix for every *h*, and the coefficients are time-varying since the predicted errors are fed back to adapt the prediction coefficients. After predicting the traffic arrival, the bandwidth request is made based on the prediction results. If there are any un-served packets remaining at the end of the DP, those packets will be served at the beginning of the following DP. The SS makes a bandwidth request as follows:

$$< r_{i,1}, r_{i,2}, \ldots, r_{i,M} > = < \hat{q}_{i,1}, \hat{a}_{i,2}, \ldots, \hat{a}_{i,M} > , \tag{8}$$

where

$$\hat{q}_{i,1} = \max(q_{i-1,M} - r_{i-1,M}, 0) + \hat{a}_{i,1} . \tag{9}$$

D. Basic Performance Evaluation

This section first evaluates the RLS prediction algorithm to assess whether the RLS algorithm is able to capture traffic dynamics. Consider a WiFi/WiMAX integrated network environment. The wireless AP is connected to the SS, while the SS is responsible for the traffic relay from the WiFi networks. All WiFi traffic is aggregated into the FIFO queue of the SS. Assume that the WiFi clients generate G.729 VoIP traffic with silence suppression. The ON/OFF period is followed exponential distribution with mean = 240ms and 400ms, respectively [9]. The packet size is 20 bytes and the sending rate in the ON period is 8 kbps. Each WiFi AP has 16 wireless clients, and the simulation time is 300 sec. Assume that each VoIP packet needs one time slot resource, and the duration of a WiMAX TDD frame is 20ms. Sample the traffic at the SS in terms of number of packets for each 20ms. In addition, for the bandwidth request model, assume C$_r$ = C$_d$ = 1 and C$_s$ = 0.

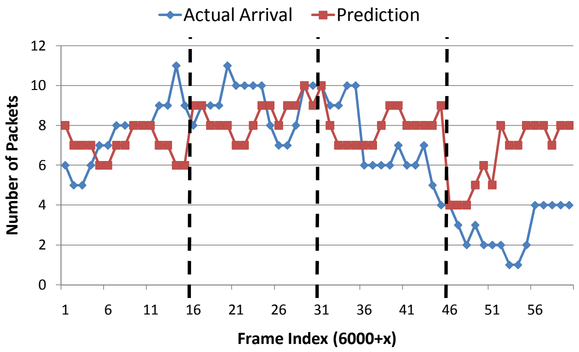
First, use the mean absolute error (MAE) as an evaluation metric. In statistics, the MAE is a measure of how close predictions are to actual outcomes.

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left| \hat{f}_i - f_i \right| \tag{10}$$

Note that $\hat{f}_i$ is the prediction while f_i is the true value. Table I summarizes the simulation results of MAE and prediction errors for different decision periods. Although this table shows that the prediction is not sufficiently accurate, Fig. 3 shows that the leftover in a DP is small enough. However, since this paper considers the real-time applications, the packets should be served as soon as possible. Although the leftover of the second decision period (i.e., frame 6016-6030) in Fig. 3 is zero, it has larger delay penalty cost. This in turn affects the QoS of real-time applications. Thus, it is still necessary to correct the prediction error and consider a better reservation amount to decrease the DP cost.

TABLE I. PREDICTION ERRORS FOR DIFFERENT DECISION PERIODS

	Mean Absolute Error	Mean Over-estimation	Mean Under-estimation
M = 15	1.78	2.11	2.25
M = 30	2.24	2.60	2.65



Leftover	4	0	0	0
C_i^r	7	1	13	55
C_i^d	65	117	41	0

Figure 3. The traffic logs of actual traffic arrival and the predictions (M = 15)

TABLE II. DP COST FOR DIFFERENT DECISION PERIODS

	Resource Reservation Cost	Delay Penalty Cost	DP Cost
M = 15	10.29	94.79	105.08
M = 30	25.37	431.44	456.81

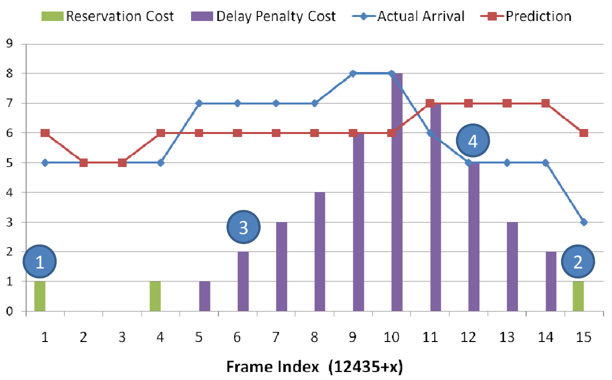


Figure 4. The relationship between the prediction errors and the induced cost.

Table II summarizes the DP cost. This table shows that the delay penalty cost is a major part of the DP cost. Although there are similar results for the mean over-estimation and the mean under-estimation as shown in Table I, the induced resource reservation cost and delay penalty cost exhibit a large disparity. Figure 4 illustrates the relationship between the prediction errors and the induced cost. When the resource reservation cost is 1, the resource reservation cost introduced in a frame should be equal to the over-estimation (see point (1) in Fig. 4). However, results show that the resource reservation cost introduced in a frame may be less than the over-estimation (see point (2) in Fig. 4). This is because the over-allocated time slots in frame *j* can serve the delayed packets from frame *j*-1, decreasing the resource reservation cost. Conversely, when the traffic arrival is under-estimated, packets will be delayed to the next frame and introduce a delay penalty cost. In addition, the delay penalty cost of a frame accumulates from the previous frame (see point (3) in Fig. 4). The accumulated delay penalty cost in frame *j* decreases when the traffic arrival in frame *j* is over-estimated (see point (4) in Fig. 4). These observations lead to the following two propositions.

Proposition 1. When C_r = 1, the resource reservation cost introduced in a frame should be equal to or less than the over-estimation, i.e., $C^r_{i,j} \le e^+_{i,j}$.

Proof. Denote the resource reservation cost in frame *j* of DP *i* as $C^r_{i,j} = (r_{i,j} - q_{i,j}) \times C_r$ as defined in (4). Assuming C_r = 1, the proposed system model leads to

$$\begin{aligned} C^r_{i,j} &= r_{i,j} - q_{i,j} \\ &= \hat{a}_{i,j} - (\max(q_{i,j-1} - r_{i,j-1}, 0) + a_{i,j}) \\ &= (\hat{a}_{i,j} - a_{i,j}) - \max(q_{i,j-1} - r_{i,j-1}, 0) \\ &= e^+_{i,j} - \max(q_{i,j-1} - r_{i,j-1}, 0) \\ &\le e^+_{i,j} \end{aligned} \qquad \square$$

Proposition 2. When C_d = 1, the delay penalty cost will accumulate until over-estimation occurs.

$$C^d_{i,j} = \begin{cases} \max(q_{i,j-1} - r_{i,j-1}, 0) + e^-_{i,j}, \text{if } (a_{i,j} - \hat{a}_{i,j}) > 0 \\ \max(q_{i,j-1} - r_{i,j-1}, 0) - e^+_{i,j}, \text{if } (a_{i,j} - \hat{a}_{i,j}) < 0 \end{cases}$$

Proof. Denote the delay penalty cost in frame *j* of DP *i* as $C^d_{i,j} = (q_{i,j} - r_{i,j}) \times C_d$ as defined in (5). Assuming C_d = 1, the proposed system model leads to

$$\begin{aligned} C^d_{i,j} &= q_{i,j} - r_{i,j} \\ &= (\max(q_{i,j-1} - r_{i,j-1}, 0) + a_{i,j}) - \hat{a}_{i,j} \\ &= \max(q_{i,j-1} - r_{i,j-1}, 0) + (a_{i,j} - \hat{a}_{i,j}). \end{aligned}$$

The definitions of (1) and (2) then lead to

$$C^d_{i,j} = \begin{cases} \max(q_{i,j-1} - r_{i,j-1}, 0) + e^-_{i,j}, \text{if } (a_{i,j} - \hat{a}_{i,j}) > 0 \\ \max(q_{i,j-1} - r_{i,j-1}, 0) - e^+_{i,j}, \text{if } (a_{i,j} - \hat{a}_{i,j}) < 0 \end{cases} \qquad \square$$

IV. MULTI-STAGE SELF-CORRECTION BANDWIDTH RESERVATION SCHEME

The simulation results of the Section III-D show that under-estimation errors have a greater effect on real-time applications, and the delay penalty cost is the major part of the DP cost. Is there any way to address inaccurate predictions and minimize the delay penalty cost? The best way to minimize the delay penalty cost in a DP is to over-allocate resources. However, over-allocation increases the resource reservation cost, which should not be too high. Note that there is a tradeoff between the resource reservation cost and the delay penalty cost.

In RLS algorithm, the prediction model is a linear combination of the previous data, and may not completely capture variations in traffic. Capturing this variation is essential to achieving better performance. Thus, according to the observations and propositions in Section III-D, the bandwidth reservation was corrected by *adding* the deviation in estimation to get better performance. This study proposes a Multi-stage Self-correction Bandwidth Reservation (MSBR) scheme. Assume there are K stages in a decision period. Define $\mathbf{B}_i = <\beta_{i,1}, \beta_{i,2}, \ldots, \beta_{i,K}>$ as the self-corrected parameter for each stage at DP *i*. Choose the standard deviation of each stage as the self-correction parameter. Thus, the bandwidth request is

$$r_{i,1} = \max(q_{i-1,M} - r_{i-1,M}, 0) + \hat{a}_{i,1} + \beta_{i,1} \tag{11}$$

and

$$r_{i,j} = \hat{a}_{i,j} + \beta_{i,\left\lceil j/\frac{M}{K} \right\rceil}, \quad j = 2 \sim M \quad, \tag{12}$$

where

$$\beta_{i,k} = (1-\alpha)\cdot\beta_{i-1,k} + \alpha \cdot StDev\,(A^k_{i-1}) \tag{13}$$

and

$$StDev\,(A^k_{i-1}) = \sqrt{\frac{\sum_{j=(k-1)\times\frac{M}{K}+1}^{k\times\frac{M}{K}} (a_{i-1,j} - E[A^k_{i-1}])^2}{\frac{M}{K}-1}}\ . \tag{14}$$

The performance of the proposed multi-stage self-correction bandwidth reservation scheme is evaluated using the same environment and traffic settings described in Section III-D. Table III summarizes the results. Compared with the results in Table II, this table shows a substantial decrease in the delay penalty cost despite a moderate increase in the resource reservation cost. Thus, the DP cost decreased as well. The individual cost per stage shown in Table IV indicates that delay penalty cost per stage increases only slightly under the MSBR scheme. This is because there are self-corrections per stage, which decreases the accumulated effect of the delay penalty cost.

TABLE III. DP COST FOR DIFFERENT DECISION PERIODS

	Resource Reservation Cost	Delay Penalty Cost	DP Cost
M = 15	43.72	10.72	54.44
M = 30	114.28	9.79	124.07

TABLE IV. THE RESOURCE RESERVATION COST AND DELAY PENALTY COST PER STAGE

	Resource Reservation Cost		Delay Penalty Cost	
M = 30	w/o Correction	with Correction	w/o Correction	with Correction
Stage 1	7.62	38.58	43.77	1.29
Stage 2	9.00	37.55	142.54	3.76
Stage 3	8.74	38.14	245.13	4.74
DP	25.37	114.28	431.44	9.79

V. CONCLUSION

This study examines the dynamic bandwidth reservation problem in an IEEE 802.11 and IEEE 802.16 integrated network. This study propose a simple, flexible, and efficient uplink bandwidth reservation scheme, called multi-stage self-correction bandwidth reservation (MSBR), that reduces the control message exchanges between BS and SS per frame and makes good use of bandwidth. The proposed design is based on the concept of decision period (DP), which consists of multiple frames. Thus, the SS issues a bandwidth request message per DP rather than per frame, reducing the control message exchanges. Second, the proposed design adopts the RLS algorithm as a bandwidth prediction algorithm, and uses a multi-stage self-correction method to reserve bandwidth by considering the tradeoff between the resource reservation cost and the delay penalty cost for real-time applications. Simulation results show that the proposed MSBR scheme utilizes bandwidth efficiently (in terms of resource reservation cost) without violating the QoS for real-time services (in terms of the delay penalty cost).

REFERENCES

[1] K. Gakhar, A. Gravey and A. Leroy, "IROISE: A New QoS Architecture for IEEE 802.16 and IEEE 802.11e Interworking," in *Proc. 2nd Intl. Conf. on Broadband Networks* (BroadNet 2005), Boston, MA, Oct. 2005, pp. 607-612.

[2] D. Niyato and E. Hossain, "Queue-Aware Uplink Bandwidth Allocation for Polling Services in 802.16 Broadband Wireless Networks," in *Proc. IEEE GLOBECOM 2005*, St. Louis, MO, Dec. 2005, pp.3702-3706.

[3] E.-C. Park, H. Kim, J.-Y. Kim and H.S. Kim, "Dynamic Bandwidth Request-Allocation Algorithm for Real-Time Services in IEEE 802.16 Broadband Wireless Access Networks," in *Proc. IEEE INFOCOM 2008*, Phenix, AZ, Apr. 2008, pp. 852-860.

[4] A. Parasuraman and P. Varadarajan, "Future Strategic Emphases in Service versus Goods Businesses," *Journal of Services Marketing*, vol. 2, no. 4, pp.57-66, 1988.

[5] K. Gakhar, M. Achir and A. Gravey, "Dynamic Resource Reservation in IEEE 802.16 Broadband Wireless Networks," in *Proc. IEEE IWQoS 2006*, New Haven, CT, June 2006, pp. 140-148.

[6] J. He, K. Yang and K. Guild, "A Dynamic Bandwidth Reservation Scheme for Hybrid IEEE 802.16 Wireless Networks," in *Proc. IEEE ICC 2008*, Beijing, China, May 2008, pp.2571-2575.

[7] S. Haykin, "Adaptive Filter Theory," Prentice-Hall Englewood Cliffs, 1986.

[8] Y. S. Sun, F.-M. Tsou and M C Chen, "Predictive Flow Control for TCP-friendly End-to-end Real-time Video on the Internet," *Computer Communications 25*,, pp. 1230-1242, 2002.

[9] D. Zhao and X. Shen, "Performance of Packet Voice Transmission Using IEEE 802.16 Protocol," *IEEE Wireless Communications*, vol. 14, no. 1, pp. 44-51, 2007.