Call Admission Control and Scheduling Schemes with QoS Support for Real-time Video Applications in IEEE 802.16 Networks

Ou Yang
Department of Electronic Engineering, Tsinghua University, Beijing, China
yangou@wmc.ee.tsinghua.edu.cn

Jianhua Lu
Department of Electronic Engineering, Tsinghua University, Beijing, China
lujh@wmc.ee.tsinghua.edu.cn

Abstract — **IEEE 802.16 networks are going to provide broadband wireless access with quality of service (QoS) guarantee. In all of services, real-time video traffic plays an impeditive role because of its varying bit-rate and stringent delay constraint. To the best of our knowledge, no call admission control (CAC) and scheduling schemes cover throughput expectation, delay constraint and fairness requirement simultaneously. In this paper, by taking advantage of traffic periodicity and regularity, a set of CAC and scheduling schemes for real-time video traffic in IEEE 802.16 networks is proposed. Specifically, two key parameters are studied to compromise throughput and delay, as well as, delay and fairness. Simulations with real life video traces show that the proposed schemes may well bear flexibility in balancing throughput, delay and fairness, or, offering significant throughput improvement with acceptable delay and fairness.**

Index Terms — **IEEE 802.16, real-time video, QoS, CAC, scheduling, throughput, fairness.**

I. INTRODUCTION

IEEE Standard of 802.16 networks [1] is designed to provide broadband wireless access with quality of service (QoS) guarantee between residential or business customers and internet service providers (ISP) [2]. It attracts great attention nowadays in both academic and industrial communities [3][4][5]. Besides potential profit from cost-effective network deployment, IEEE 802.16 networks prospect high QoS for emerging wireless applications, especially for real-time video application. However, the standard only specifies the framework of Medium Access Control (MAC) service, but leaves the call admission control (CAC) and scheduling schemes undefined and open to discussion.

For real-time video applications, varying bit-rate and stringent delay are always impediments to improve the QoS. So far, research has been done on CAC and scheduling schemes to help real-time video traffic obtain desirable delay performance [2][6][7][8][9][10]. For example, Ref. [2] presents a CAC scheme for IEEE 802.16 networks with a token bucket. It reserves adequate bandwidth for every admitted flow. However, it is so conservative that much of bandwidth is reserved unnecessarily. Likewise, for scheduling schemes, the Earliest Deadline First (EDF) [2][6][9][10][11] is applied to real-time traffic according to different service requirements for maximizing throughput under delay constraints [12], while preemptive First In First Out (FIFO) and Weighted Fair Queuing (WFQ) are utilized for real-time video traffic to avoid flow starvation [7][8]. However, none achieve good delay performance and fairness simultaneously.

To effectively improve throughput while maintaining QoS guarantee, including especially delay performance and fairness, this paper proposes a set of CAC and scheduling schemes for real-time video applications. The proposed CAC admits an incoming flow with more flexibility via an appropriate access time, which is figured out through coordination with existing I-frames and non-I-frames. Likewise, the proposed scheduling scheme arranges real-time video packets according to individual delay performance of each flow under a loose constraint on earliest latest starting time (LST) first. Simulation results show that the proposed CAC significantly improves the throughput of network, while the scheduler minimizes the difference of individual delay performance among admitted flows, thus maintaining fairness. Furthermore, the combination of both may provide high throughput and QoS support for real-time video application simultaneously.

The remainder of this paper is organized as follows. Section II provides some background. Sections III and IV describe in detail the proposed CAC and scheduling schemes, respectively. In Section V, simulation results are presented and discussed. Finally, Section VI concludes the paper.

Based on "A New Scheduling and CAC Scheme for Real-Time Video Application in Fixed Wireless Networks", by Ou Yang, Jianhua Lu which appeared in the Proceedings of the IEEE Consumer Communications and Networking Conference 2006, Las Vegas, USA, January 2006. © 2006 IEEE.

II. BACKGROUND

A. Wireless Access Mechnism in IEEE 802.16 Networks

An IEEE 802.16 network consists of two kinds of stations, a Base Station (BS) and Subscriber Stations (SS). BS is connected to a wired backbone network, while SSs are portals of various local area networks. The IEEE 802.16 Standard defines the air interface between BS and SS, with an uplink from SS to BS and a downlink from BS to SS.

To provide broadband applications with high resource utilization, the IEEE 802.16 architecture employs Time Division Multiple Access (TDMA) as its wireless access mechanism. An MAC frame is formatted as shown in Fig. 1.

The frame begins with a frame start preamble followed by a downlink sub-frame and then an uplink sub-frame. Downlink sub-frame starts with a frame control section, containing DL-MAP and UL-MAP, indicating the resource arrangements in downlink and uplink, respectively. Followed are downlink data. Uplink transmission occurs in burst manner according to the order and quantity indicated in the UL-MAP section. BS executes CAC and uplink scheduler to determine the scheme of every UL-MAP. They are of significant importance in determining throughput, delay performance and fairness merit of the network.

B. Previous Work on CAC and Scheduling Algorithms

CAC decides whether a new connection can be established [2]. Permission is granted only when QoS requirement of the new incoming flow is met without degrading those of the existing flows [9]. Generally, every admitted real-time video flow is allocated adequate bandwidth to meet its delay requirement. In the situation of heavy load, bandwidth is distributed to a flow in proportion to its weight, the ratio of its average bit-rate to the sum of average bit-rate of all admitted flows [2]. As a result, the delay requirement of a flow is satisfied only when the allocated bandwidth within the delay bound can transmit more data than those arrive at its maximum bit-rate in one frame burst. Albeit the CAC provides every admitted flow with adequate bandwidth under the worst situation, and thus fairness is out of question, it results in low bandwidth utilization, especially when incoming flows render great burstiness (ratio of maximum bit-rate $R_{\max}$ to average bit-rate R_{avg}).

Uplink scheduler decides the starting time of every transmission within the uplink sub-frame and makes its solution broadcast in the UL-MAP. Currently, much work has been done on packet level scheduling. EDF is commonly used to handle real-time traffic. As packet with the earliest deadline goes first, EDF is naturally suited for maximizing throughput under delay constraints. In Ref. [2], BS polls admitted video flows periodically to collect bandwidth requests. A scheduling table is established with flow ID in row and time in column to record the size of the packets waiting to be transmitted before a certain time. Since the scheduler handles data one column after another, packets are transmitted as soon as possible to catch up with their deadline. However, EDF has a shortcoming of unfairness among individual flows.

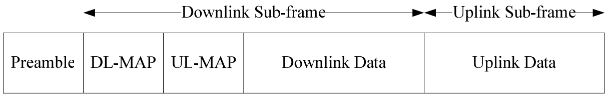

Figure 1. MAC frame format for IEEE 802.16 networks

C. Traits of Real-time Video Traffic

Real-time video traffic often leads to difficulties in scheduling strategy because of its indeterminate time-varying bit-rate and stringent delay. Usually, average bit-rate and maximum bit-rate are utilized to describe the long-term attributes of a real-time video flow. On the other hand, detailed traffic structure contributes to improve the scheduler. Fortunately, real-time video traffic renders periodicity and regularity in frame sequence, identified by I, P and B, according to its Group of Picture (GOP) structure. A GOP structure starts with an I-frame and ends before the next I-frame.

Compared with P-frames and B-frames, I-frames are typically 2-10 times bigger in size, but much less in quantity. Likely this causes problems in the CAC and the scheduler, which applies only long-term traffic traits. In particular, when several video flows happen to access to an invariable channel with respective I-frames located too close with each other, heavy load of aggregated I-frames likely results in extra delay [9]. Moreover, too much bandwidth is reserved in CAC since bulky I-frames seldom arrive. All these lead to a conservative CAC and an inefficient scheduler that accommodate low throughput. In this paper, the periodicity and regularity in real-time video sequences are well employed by a set of new CAC and scheduling schemes to overcome the problem.

III. THE PROPOSED CAC SCHEME

A. Access Pendency

Generally, the CAC decides whether an incoming flow can be admitted at the time it arrives at SS. On the other hand, it is unlikely a proper time for the flow to establish a connection because of some predictable delay violation, like aggregated I-frame overload mentioned in Section II. We therefore introduce a pending period for every incoming flow after its arrival. The flow is not admitted until the CAC finds the earliest proper time within its pending period to establish a connection. If there is no such an appropriate access time, the CAC ultimately rejects the incoming flow after the pending period expires.

To illustrate the benefit of the policy, Fig. 2 shows two video flows accessing a channel with I-frames synchronized and Fig. 3 makes a comparison of their aggregated flow without and with pending period. Obviously, when the two flows are immediately admitted, the profile of aggregated traffic is severely serrated; however, if the second flow delays a little while before the connection establishment, the profile of aggregated traffic is relatively smoother [9]. Generally, smoother the traffic is, less traffic burst causes delay violation.

B. Coordination with I-frames

Within the pending period, CAC firstly figures out all the possible access times, when the incoming flow will not cause I-frame overload with all the admitted flows. In this case, the periodicity of real-time video traffic is utilized to find out any I-frame superposition and subsequent delay violation.

Take two flows for example. The first one has been admitted to the network, while the second one requests to establish a connection. Parameters for those two flows are listed in Tab. I. Note that all the timing parameters mentioned in this paper are measured in the unit of a MAC frame.

Suppose x and y are nonnegative integers, which denote the number of GOP periods the two flows experience. $0 < \delta_1 \le \Delta_1$ and $0 < \delta_2 \le \Delta_2$. A superposition of I-frames exists when the following Diophantine equation has natural number solutions,

$$S_1 + T_1 x + \delta_1 = S_2 + T_2 y + \delta_2 . \tag{1}$$

After some manipulations, we have

$$\begin{aligned} T_1 x - T_2 y &= S_2 - S_1 + \delta_2 - \delta_1 \\ &= S_2 - S_1 + \Omega \end{aligned} , \tag{2}$$

where $-\Delta_1 < \Omega < \Delta_2$. By the theory of linear Diophantine equations [13], a necessary and sufficient condition for the existence of an integer solution of (2) is given by

$$(T_1, T_2) \mid (S_2 - S_1 + \Omega) . \tag{3}$$

That is to say, if $S_2 - S_1 + \Omega$ is divisible by the greatest common divisor of T_1 and T_2, (2) has integer solutions. Obviously, (2) has natural number solutions as long as it has an integer solution (x_0, y_0) , since $(x_0 + T_2 n, y_0 + T_1 m)$, where n and m are both integers, will be positive as n and m increase.

TABLE I.
PARAMETERS OF TWO INCOMING FLOWS

	Flow 1	Flow 2
Arriving Time	A_1	A_2
Access Time	S_1	S_2
GOP Period	T_1	T_2
Delay Bound	d_1	d_2
Pending Period	p_1	p_2
Transmission Time of Maximum I-frame	Δ_1	Δ_2

An aggregated I-frame overload occurs only when an I-frame superposition exists and the total transmission time of the maximum I-frame of the two flows exceeds either of their delay bound. Therefore, if (3) is satisfied and

$$\Delta_1 + \Delta_2 > \min(d_1, d_2) , \tag{4}$$

where $\min(x, y)$ gets the minimum value of x and y, the second incoming flow can not be admitted at S_2 because the I-frames of the two flows get too close with each other, thus cause delay violation.

Since S_2 is adjustable within its pending period $[A_2, A_2 + p_2)$, those values which dissatisfy (3) or (4) are possible access times for the incoming flow to establish a connection. Likewise, if there are more than one admitted flows, CAC measures them one by one to ultimately find out all possible access times which avoid delay violation with existing I-frames.

C. Coordination with Non-I-frames

After I-frame coordination, CAC detects delay violation between the incoming flow and the existing non-I-frames through non-I-frame coordination. It tries

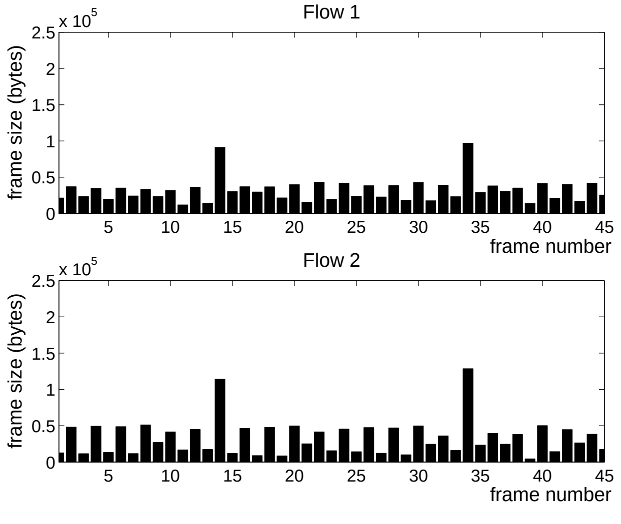

Figure 2. Two real-time video flows accessing the network [9]

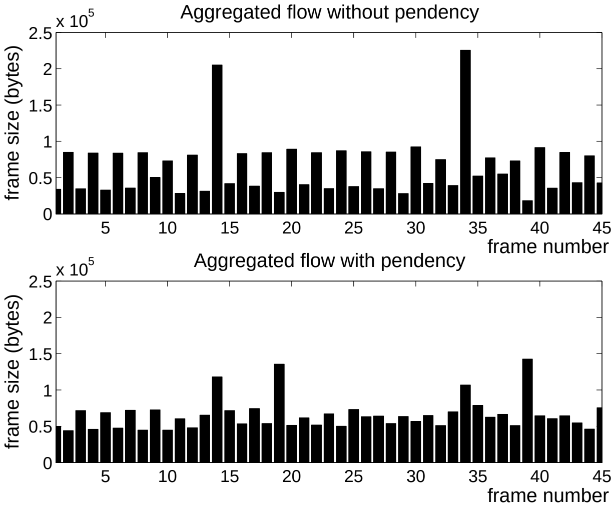

Figure 3. Aggregated flow without/with pendency [9]

all the possible access times determined before, one by one to find the earliest proper time for the incoming flow to establish a connection without QoS degradation.

Due to the periodicity of real-time video flow, the aggregated traffic of admitted flows, as well, bears periodicity,

$$T_{all} = LCM(T_1, T_2, \cdots, T_N), \tag{5}$$

where LCM represents the least common multiple and T_i denotes the GOP period of the i-th flow. Keeping the channel usage of successive T_{all} MAC frames, CAC maintains a sequence of T_{all} records to simplify the detection of delay violation upon the overall aggregated traffic.

When an incoming flow requests to establish a connection at a certain possible access time, CAC firstly updates T_{all}, which includes all admitted flows, as well as the incoming flow into calculation. Then, admitted flows mark their non-I-frame transmissions in the T_{all} sequence in order of access time. For every admitted flow, the amount of data of its non-I-frame is estimated as

$$D = R_E \times t, \tag{6}$$

where R_E is the rate estimation of non-I-frame of the flow and t is the time interval between successive frames, e.g., 40ms for traffic at 25 frames/s. The amount of data D occupies the empty records one by one in the T_{all} sequence, beginning from the record corresponding to the flow's access time. Since D is no larger than the maximum I-frame of the flow, it can be placed within successive records, the amount of which equals its delay bound.

Suppose the incoming flow has a delay bound of d. CAC permits it depending on whether its maximum I-frame can be successfully placed in the T_{all} sequence before its deadline. Within successive d records in the T_{all} sequence, beginning from the one corresponding to its possible access time, if there is enough empty space for its maximum I-frame to be placed, the flow can be admitted without negative influence on delay performance. Otherwise, CAC tries the next possible access time until it rejects the flow since its pending period expires.

In this model, R_E is estimated ranging from zero to the maximum bit-rate of the flow. Simulations are presented in the next section to guide a proper selection of R_E.

D. CAC Process and An Example

Assume the channel capacity is C. There are N admitted flows, each of which is described as a tuple $(A_i, S_i, T_i, p_i, d_i, \Delta_i, D_i/C)$, $1 \le i \le N$, defined the same as before. An incoming flow arrives at SS, described as $(A, S, T, p, d, \Delta, D/C)$. D_i/C and D/C indicate the estimation of average transmission time of non-I-frame for the i-th admitted flow and the incoming flow, respectively. The process of CAC is presented in Fig. 4.

Take two flows for a simple example. The first admitted flow bears a tuple (1,1,10,5,7,3,2). The second incoming flow is modeled as (2, S ,5,5,3,2,1). The CAC makes a decision in the following steps.

Step 1: CAC updates $T_{all} = T_1 = 10$ and marks the non-I-frame transmission for the first admitted flow by occupying $D_1/C = 2$ empty records from $S_1 = 1$ in the T_{all} sequence. See the left figure in Fig. 5.

Step 2: The two flows meet (3) since $\Delta_1 + \Delta = 5 > \min(d_1, d) = 3$.

Step 3: From $A = 2$ to $A + p - 1 = 6$, CAC finds out all possible access times in Tab. II. When $(T_1, T) = (10,5) = 5$ and $-\Delta_1 = -3 < \Omega < \Delta = 2$, it is observed that $S = 4$ and $S = 5$ are possible access times.

Step 4: When $S = 4$, CAC executes non-I-frame coordination. It places the maximum I-frame of the incoming flow into the T_{all} sequence, starting from $S = 4$ and $S = 4 + T = 9$, shown in the middle figure in Fig. 5. Obviously, the maximum I-frame can be placed within the delay bound of $d = 3$. Therefore, CAC admits the incoming flow and makes its access time $S = 4$.

The right figure in Fig. 5 shows the situation of non-I-frame transmission in T_{all} sequence after the incoming flow establishes a connection, and before the next incoming flow arrives at SS.

IV. THE PROPOSED SCHEDULING SCHEME

When the proposed CAC bears with acceptable QoS failure in delay performance to improve the throughput, the scheduling scheme turns to be critical to distribute delay violation to individual connections. In other words, the scheduler is responsible for maximizing the throughput and minimizing the difference of delay

TABLE II.
RESULTS OF I-FRAME COORDINATION

Starting Time	$(T_1,T)\,|\,(S - S_1 + \Omega)$	Access Possibility
$S = 2$	$5\,|\,(1+\Omega)$	×
$S = 3$	$5\,|\,(2+\Omega)$	×
$S = 4$	$5\,|\,(3+\Omega)$	√
$S = 5$	$5\,|\,(4+\Omega)$	√
$S = 6$	$5\,|\,(5+\Omega)$	×

```
Process CAC {
  Decision='rejection';   /* Default decision is rejection */
  /* I-frame Coordination*/
  For $i = A$ to $A + p - 1$ {        /* Find possible access times within $[A, A + p)$ */
    P[i]=true;   /* P is an array indicating possible access time */
    For $j$ =1 to $N$ {     /* Detect delay violation with existing I-frames */
      If ( $\Delta_j + \Delta > \min(d_j, d)$ ) and ( $(T_j, T) \mid (i - S_j + \Omega)$ , $-\Delta_j < \Omega < \Delta$ ) {     /* Use (3) and (4) */
        P[i]=false;        /* Rule out impossible access time*/
  };};};};
  /* Non-I-frame Coordination */
  For $i = A$ to $A + p - 1$ {
    If P[i]=true {           /* Find the final access time within all possible access times */
      $T_{all} = LCM(T_1, T_2, \cdots, T_N, T)$ ;  /* Update $T_{all}$ */
      For $j$ =1 to $N$ { UpdateT( $j$ , $S_j$ , $D_j / C$ ); };    /* Place $D_j$ of every admitted flow into $T_{all}$ sequence */
      If (PlaceI( $i$ , $\Delta$ , $d$ , $T_{all} / T$ )=true) { /* PlaceI returns whether maximum I-frame of the incoming flow */
      /* can be placed into $T_{all}$ sequence within its deadline */
        Decision='admission';   /* Grant admission */
        $S = i$ ;  /* Find out the earliest access time */
        $N = N$ +1;       /* Update the number of admitted flows */
        Return;
  };};};};
  Return;
}
```

Figure 4. CAC process

performance among individual admitted flows at the same time.

Obviously, EDF can no longer serve the purpose. Packets with the earliest deadline always have priority to obtain bandwidth, thus those in the waiting list have to bear, if any, the transmission delay caused by their previous goers. As real-time video frames arrive periodically, frames from some flows may always be scheduled before frames from some others. EDF therefore can not guarantee the fairness among individual flows.

However, the proposed scheduling scheme still takes the advantage of EDF to maximize the throughput of real-time video packets under delay constraints. The first difference is the urgency gauge of a packet. EDF takes deadline as an indication of urgency; our scheduling scheme, nevertheless, uses LST as a counterpart. Specifically, in the scheduling table, EDF puts every waiting packet in the column corresponding to its deadline, while our scheduler distributes a packet ranging from the deadline column back to the column which corresponds to its LST, shown as shadows in Fig. 6, to meet the delay constraint. This action attributes to fair service, since LST is defined as the latest starting time to schedule the packet, thus reflects the urgency of a packet more reasonable than its deadline. The second difference is the precision in gauging the earliest LST and deadline in Ref. [2]. EDF, employed in Ref. [2], schedules all the packets in the earliest nonempty column in the scheduling table, in proportion to its average bit-rate. Our proposed

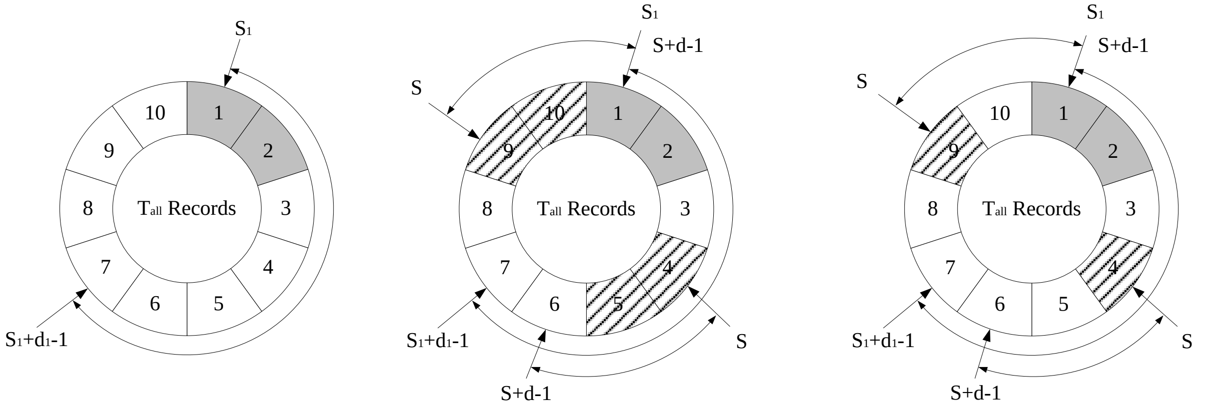

Figure 5. An example of CAC

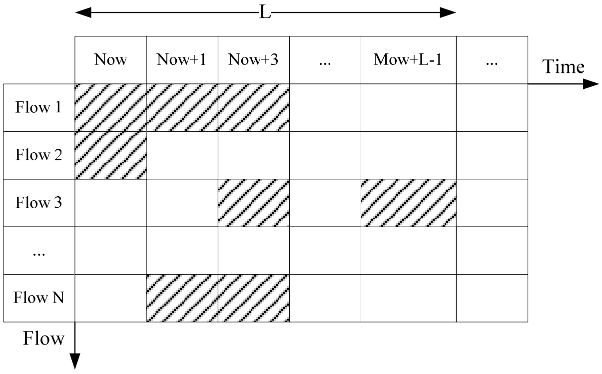

Figure 6. Scheduling table in our proposed scheduler

scheduler, on the contrary, enlarges the pool of prospective packets from only one column to L columns, where L is a natural number. The first packet of each flow in the L columns is possibly selected to send according to the flow's current delay performance, which is a scheduling parameter other than LST to support fair service.

For each admitted flow, the scheduler records tot, the amount of data which were transmitted, and err, the amount of data which missed their deadline. Prospective packets in the L columns are selected to send according to the following regulations.

1. $Packet_{err=0,tot=0}$ vs. $Packet_{err=0,tot=0}$, the one with smaller average bit-rate is selected.
2. $Packet_{err=0,tot=0}$ vs. $Packet_{err=0,tot\neq 0}$, the one with smaller tot is selected.
3. $Packet_{err=0,tot=0}$ vs. $Packet_{err\neq 0,tot\neq 0}$, the one with smaller err is selected.
4. $Packet_{err=0,tot\neq 0}$ vs. $Packet_{err=0,tot\neq 0}$, the one with smaller tot is selected.
5. $Packet_{err=0,tot\neq 0}$ vs. $Packet_{err\neq 0,tot\neq 0}$, the one with smaller err/tot is selected.
6. $Packet_{err\neq 0,tot\neq 0}$ vs. $Packet_{err\neq 0,tot\neq 0}$, the one with smaller err/tot is selected.

In the proposed scheduling scheme, packets with most urgency are considered to be scheduled under a loose constraint on earliest LST first, so that packets catch up with their deadline as more as possible. Meanwhile, selecting packets based on tot and err, the proposed scheduler minimizes the unfairness among all admitted flows.

V. SIMULATION RESULTS

In this section, we develop a simulation model in C++ to evaluate the performance of the proposed CAC and scheduling schemes, and discuss the selection of some key parameters. Three criteria are applied, including: (1) throughput, defined as the ratio of total bandwidth of admitted flows to the channel capacity; (2) delay performance, which is the ratio of total amount of data which are transmitted within their deadlines to which are transmitted; and (3) fairness, denoting the maximum difference of delay performance among all admitted flows. All input flows are real-life H.264 videos with frame rate of 25 frames/s and average bit-rate ranging from 50kb/s to 1Mb/s. Smoothed by the SS, input flows are jitter-free. It is assumed that one MAC frame takes 1ms [1] and the pending period of a flow is 40ms.

A. Throughput

Throughput is determined by the CAC. Since CAC records bandwidth occupation in a T_{all} period based on the rate estimation of non-I-frame of a flow, R_E reflects bandwidth reservation for every admitted flows. Ranging from 0 to $R_{\max}$, R_E is defined as a function of abscissa x:

$$R_E = \begin{cases} x \times R_{avg} & 0 \le x \le 1 \\ R_{avg} + (x-1)\times(R_{\max} - R_{avg}) & 1 < x \le 2 \end{cases}. \quad (7)$$

In this subsection, the effectiveness of I-frame coordination and non-I-frame coordination are examined respectively. The first experiment is deployed to test the throughput under delay violation caused by superposed I-frames from different flows. Parameters of input flows are listed in Tab. III. Flows with I-frame superposed are tagged with the same ★, ◎, △ and ◆. The second

TABLE III.
PARAMETERS OF INCOMING FLOWS IN THE FIRST EXPERIMENT

Parameters of Incoming Flows			
Flow	Average Bit-rate (kb/s)	Delay (ms)	I-frame Superposition
Claire	200	40	★
Salesman	500	80	◎
Bridge-close	50	40	★
Grandma	200	40	△
Carphone	50	40	◎
Salesman	100	40	◆
Highway	100	40	◎
News	100	40	△
Container	50	40	★
Forman	50	40	★
Trevor	200	40	△
Suzie	200	40	△
Highway	200	40	◆

TABLE IV.
PARAMETERS OF INCOMING FLOWS IN THE SECOND EXPERIMENT

Parameters of Incoming Flows		
Flow	Average Bit-rate (kb/s)	Delay (ms)
Trevor	50	40
Highway	200	40
Carphone	1000	67
Suzie	500	62
News	50	40
Salesman	100	40
Grandma	50	40
Forman	50	40

experiment examines the throughput under input flows free of I-frame superposition. Parameters of input flows are listed in Tab. IV. It is assumed that the channel capacity is 2Mb/s. We simulate the throughput as a function of R_E.

In Fig. 7 and Fig. 8, solid lines show the trend of throughput as R_E increases from 0 to $R_{\max}$ in case of I-frame superposition and that free of I-frame superposition. Obviously, throughput decreases as a result of rising R_E, which essentially reflects bandwidth reserved for every admitted flow. When R_E is approaching 0, nearly no bandwidth is reserved for flows in the non-I-frame coordination. Incoming flows have more possibility to get an admission because less delay violation, caused by its maximum I-frame and the existing non-I-frames in the network will be detected. It is observed that the throughput begins to fall when R_E is near R_{avg}. That is because the actual bit-rate of non-I-frame does not far deviate R_{avg}. Although in a given time, some non-I-frames arrive faster than their R_{avg}, some others may be slower to offset the excess. As a result, when R_E grows larger than its actual bit-rte, bandwidth reservation becomes more than necessary. When R_E gets close to $R_{\max}$, CAC makes excessive bandwidth reservation pretty much to the extreme. The throughput deteriorates. Therefore, it is recommended to attune R_E to slightly adjust the tightness of CAC [9].

In Fig. 7 and Fig. 8, dashed lines indicate the throughput of the CAC proposed in Ref. [2]. In the case of I-frame superposition, it only accommodates flows at 500kb/s, even less than the performance of our proposal at 700kb/s when R_E equals to $R_{\max}$. In the case free of I-frame superposition, the CAC proposed in Ref. [2] admits flows at 250kb/s, compared with our 1250kb/s in the worst. The simulation results show that both I-frame coordination and non-I-frame coordination are effective in detecting delay violation and controlling bandwidth assignment. Our proposed CAC significantly improves the throughput especially when R_E is not far more than R_{avg}.

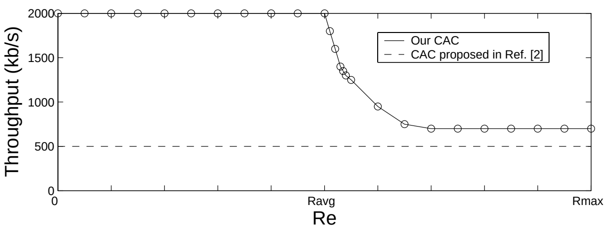

Figure 7. Throughput in case of I-frame superposition

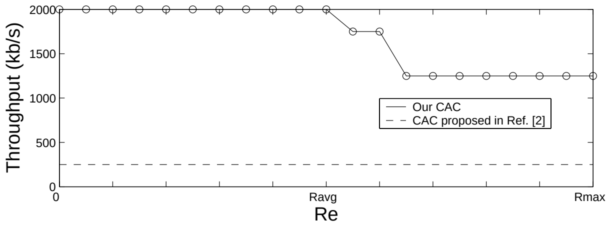

Figure 8. Throughput in case free of I-frame superposition

B. Fairness

Under a fixed CAC with acceptable QoS failure in delay performance, scheduler determines the fairness performance in serving admitted flows. Compared with EDF, which ignores the maximum difference of delay performance among individual flows, our scheduler takes two actions to mitigate the unfairness. The first is to evaluate the urgency of a flow by its LST, and the second is to enlarge the pool of prospective packets and schedule the neediest one to maintain fairness. The scale of the packets pool is reflected by the parameter L. We simulate the fairness under different L against the scheduling scheme proposed in Ref. [2]. In this experiment, the simulation environment is the same as in subsection A.

As shown in Fig. 9 and Fig. 10, dashed lines and solid lines indicate the simulation results on fairness given by the scheduler proposed in Ref. [2] and our scheduler, respectively. Obviously, the maximum difference of delay performance among individual admitted flows, provided by the former scheduler in Ref. [2], is above 20%, greatly larger than what a commercialized telecom service can put up with. Unfairness becomes severe since the scheduler proposed in Ref. [2] discriminates flows with smaller average bit-rate. As it distributes bandwidth to packets with the same deadline in proportion to its R_{avg}, flows with larger R_{avg} always obtain more bandwidth to continuously improve their delay performance. Our scheduler, on the contrary, schedules the packet belonging to the flow with most demand for delay performance improvement. It significantly brings down the difference of individual flow's delay performance at less than 3.0%, even though L maintains

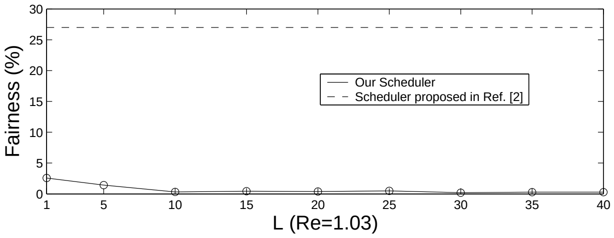

Figure 9. Fairness in case of I-frame superposition

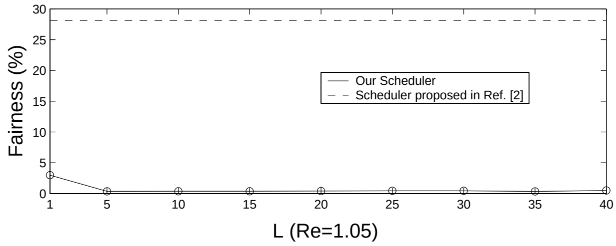

Figure 10. Fairness in case free of I-frame superposition

as 1. When L increases, the difference continues to drop until it stays around a very small value less than 0.3%. Fairness does not achieve further significant improvement as L turns larger and larger, since however large L is, the scheduler only schedules the first packet of each flow within the packet pool. After all, our scheduler provides the flexibility to attune L online to find a preferable fairness for all admitted flows.

C. Delay Performance

In this experiment, we hope to figure out what kind of influence CAC and scheduler lay on the delay performance. The simulation environment is the same as in subsection A.

Solid lines in Fig. 11 and Fig.12 represent delay performance under our CAC and scheduler. Obviously, delay performance improves along with a rising R_E , but deteriorates as L grows up. That is because a larger R_E reflects a generous bandwidth reservation and consequently less delay violation, while larger L leads to less possibility for the most urgent packet to steal transmission opportunity against the packet which demands greatest improvement in delay performance. It is observed that R_E has stronger influence on delay performance than L . Particularly, delay performance leaps up when R_E is around R_{avg} , where CAC starts making sufficient bandwidth reservation for every admitted flow. An increasing L , however, smoothly decreases the delay performance without a sudden drop. To obtain a better delay performance, it tends to adopt a larger R_E and a smaller L .

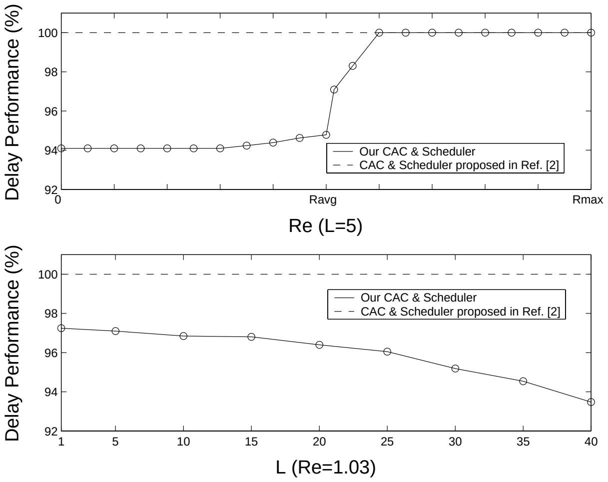

Figure 11. Delay performance in case of I-frame superposition

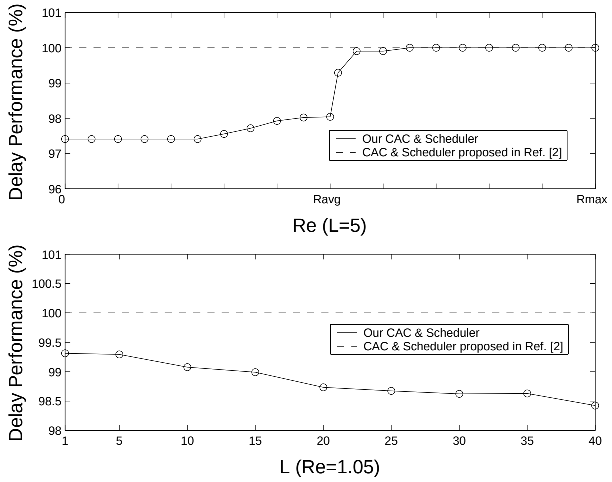

Figure 12. Delay performance in case free of I-frame superposition

Dashed lines in Fig. 11 and Fig. 12 provide delay performance obtained by the CAC and scheduler proposed in Ref. [2]. Based on the very small amount of admitted data, the CAC and scheduler meet the delay requirement of every flow. However, our CAC and scheduler can achieve high-throughput at more than 85% bandwidth capacity, good delay performance no less than 97%, and fairness at about 2% under preferable R_E and L .

D. Discussion

To sum up, throughput, delay performance and fairness are involved in two trade-offs, handled by R_E and L , simultaneously. The first trade-off is between throughput and delay performance. To obtain high throughput, a smaller R_E is expected. However, when good delay performance is required, R_E needs to be larger. The second trade-off is between fairness and delay performance. A larger L contributes to a better fair service, but a smaller L benefits delay performance. Simulation results show that to make a balance among throughput, delay performance and fairness, R_E is recommended to be a little more than R_{avg} , while it is better to attune L online a bit larger than 1.

VI. CONCLUSION

In this paper, a set of new CAC and scheduling schemes for real-time video applications in IEEE 802.16 networks is proposed. For CAC, high throughput is achieved by introducing a pending period for every incoming flow to subtly arrange its access time. A combination of I-frame coordination and non-I-frame coordination also contributes to detect, and consequently avoid delay violation. For scheduling scheme, a loose constraint on earliest LST first is applied to compromise transmission delay and fairness. Simulation results show that throughput, delay and fairness may be compromised well in the study. The proposed CAC and scheduling schemes well bear the flexibility to balance them by significantly improving the throughput with acceptable degradation in delay performance and fairness.

ACKNOWLEDGMENT

This work was supported in part by NSFC (60132010), KAIST-Tsinghua joint research funding, and Intel research funding.

REFERENCES

[1] IEEE 802.16-2004, “IEEE Standard for Local and metropolitan area networks-Part16: Air Interface for Fixed Broadband Wireless Access Systems”, 2004.

[2] K. Wongthavarawat and A. Ganz, “Packet Scheduling for QoS Support in IEEE 802.16 Broadband Wireless Access Systems”, *International Journal of Communication Systems*, 16:81-96, 2003.

[3] Steven J. Vaughan-Nichols, “Achieving Wireless Broadband with WiMax”, *Computer*, 37(6):10-13, Jun 2004.

[4] G. Nair, J. Chou, T. Madejski, K. Perycz, D. Putzolu, J. Sydir, et al., “IEEE 802.16 Medium Access Control and Service Provisioning”, *Intel Technology Journal*, 8:213-227, 2004.

[5] ALTRA White Paper, *Accelerating WiMAX System Design with FPGAs*, ver. 1.0, Oct. 2004.

[6] K. Wongthavarawat, A. Ganz, “IEEE 802.16 Based Last Mile Broadband Wireless Military Networks with Quality of Service Support”, *In MILCOM 2003 IEEE*, 2:779-784.

[7] M. Hawa, D. W. Petr, “Quality of Service Scheduling in Cable and Broadband Wireless Access Systems”, *In IWQOS 2002 IEEE*:247-255.

[8] G. Chu, D. Wang and S. Mei, “A Qos Architecture for the MAC Protocol of IEEE 802.16 BWA System”, *In ICC 2002, Circuits and Systems and West Sino Exposition Proceedings*, 1(4):435-439.

[9] Ou Yang and Jianhua Lu, “A New Scheduling and CAC Scheme for Real-Time Video Application in Fixed Wireless Networks”, *In CCNC 2006 IEEE*, 303-307.

[10] C.-H. Jiang and T.-C. Tsai, “Token Bucket Based CAC and Packet Scheduling for IEEE 802.16 Broadband Wireless Access Networks”, *In CCNC 2006 IEEE*, 183-187.

[11] Dong-Hoon Cho, Jung-Hoon Song, Min-Su Kim and Ki-Jun Han, “Performance Analysis of the IEEE 802.16 Wireless Metropolitan Area Network”, *In DFMA 2005 IEEE*, 130-136.

[12] Maria Adamou, Sanjeev Khanna, Insup Lee, Insik Shin and Shiyu Zhou, “Fair Real-time Traffic Scheduling over A wireless LAN”, *In RTSS 2001 IEEE*, 279-288.

[13] W. Narkiewicz, *Number Theory*, Polish Scientific Publishers, Warszawa, 1977.

Ou Yang was born in Changsha, China in May, 1982. She received the B.E. degree in communication engineering from Beijing University of Posts and Telecommunications, Beijing, China, in 2003. She is currently a graduate student at the Department of Electronic Engineering in Tsinghua University, Beijing, China.

Her research interests include wireless multimedia networks, network optimization, resource allocation, quality of service, routing, and congestion control.

Ms. Yang is a Student Member of IEEE Communication Society.

Jianhua Lu was born in Jiangsu, China in July, 1963. He received the B.S.E.E. and M.S.E.E. degrees from Tsinghua University, Beijing, China, in 1986 and 1989, respectively, and the Ph.D. in electrical and electronic engineering from the Hong Kong University of Science and Technology, Hong Kong, in 1998.

Since 1989, he has been with the Department of Electronic Engineering, Tsinghua University, where he currently serves as a Professor. He has been an active member of professional societies and has publishes papers in international journals and conference proceedings. He served as the Panel/Invited Session Chair of the 2003 IEEE International Symposium on Personal, Indoor and Mobile Communications, which was held in Beijing. He also delivered a keynote speech on Wireless Multimedia Communications at the CCF Young Computer Scientists and Engineers Forum of China. His current research interests include wireless multimedia communication, multimedia signal processing, satellite communication, and wireless networking.

Dr. Lu is a Member of IEEE Communication Society and the IEEE Signal Processing Society.