Adaptive Bandwidth Allocation for TCP Traffic in IEEE 802.16j Wireless Networks with Transparent Relay Stations

Iam Kin Chan and Wanjiun Liao
Graduate Institute of Communications Engineering, National Taiwan University, Taipei, Taiwan
Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan
Email: {ikchan@kiki.ee.ntu.edu.tw, wjliao@ntu.edu.tw}

Abstract—In this paper, we study adaptive bandwidth allocation for TCP-based best effort traffic in IEEE 802.16j networks with relay stations. We consider the effect of bandwidth asymmetry on the performance of TCP and take into account the current traffic profile and network channel condition for each subscriber station. We then propose an adaptive scheme which enhances the system aggregate throughput and ensures fairness among TCP flows. Our scheme also cooperates with the scheduler to throttle the TCP source when acknowledgements are infrequent. The performance of our scheme is validated via ns-2 simulations. The results show that our scheme outperforms static allocation in terms of higher aggregate throughput.

I. INTRODUCTION

IEEE 802.16 (WiMAX) is an emerging last mile technology for broadband wireless access to the Internet [1-3]. A typical 802.16 network consists of base stations (BSs) and subscriber stations (SSs) [4]. In the newly defined IEEE 802.16j standard [5], relay stations (RS) are introduced to enhance coverage, throughput and system capacity. IEEE 802.16j specifies two types of relay stations, namely, transparent and non transparent relay stations. Relay networks with distributed resource allocation use non transparent RSs. BS and non transparent RSs allocate the bandwidth for their sub-ordinate stations in a distributed manner. Transparent RSs only forward the data packets and control massages between BS and SSs. Relay networks with centralized resource allocation use transparent RSs. The bandwidth allocation of all links in the cell is centralized at the BS. IEEE 802.16j standards support both Frequency Division Duplex (FDD) and Time Division Duplex (TDD). In an FDD system, the uplink (UL) and downlink (DL) channels are located on separate frequencies and data can be transmitted simultaneously. With TDD, the uplink and downlink transmissions share the same frequency at different time intervals. In this paper, we focus on the TDD-based relay system with centralized resource allocation at the base station using transparent RSs.

Each 802.16j TDD frame has a fixed duration and consists of one DL sub-frame and one UL sub-frame, as shown in Fig. 1. The DL to UL bandwidth ratio can vary with time to enhance network performance. At the start of each frame, BS broadcasts control messages, called MAPs (i.e., DL-MAP and UL-MAP for downlink and uplink, respectively), to inform SSs of the time slot allocation in the DL and UL sub-frames. Upon receiving the MAP messages, SSs know when to receive or send data. SSs share the uplink to BS on an on-demand basis. DL and UL sub-frames are divided into access zones and relay zones. The DL sub-frame shall include one access zone for BS to RS and SS transmissions and may include one transparent zone for RS to sub-ordinate station transmissions. The UL sub-frame may include one access zone and may include one relay zone for RS to super-ordinate station transmissions. IEEE 802.16j supports four uplink scheduling services. In this paper, we focus on the best effort (BE) scheduling service for TCP-based applications, in which SS asks for bandwidth for a connection by sending a request to BS within the bandwidth request contention slots in the uplink sub-frame.

A performance challenge in IEEE 802.16j TDD systems is to determine the bandwidth allocation for SSs at different locations with different network condition. An inappropriate allocation may significantly degrade the total system performance. Many research efforts [6-9] show that bandwidth asymmetry on a link impacts the network performance substantially, especially for TCP, which relies on feedback of acknowledgments from the receiver to ensure reliability. In [10], a combined uplink resource reservation and power control framework for relay-based systems is presented. That work focuses on providing fair uplink scheduling. However, it is not designed for enhancing TCP performance. In [11], we propose a scheme which adjusts the DL to UL bandwidth ratio according to the current traffic profile by a cross layer approach. However, [11] does not take into account locations and channel conditions of different SSs. As a result, it is only applicable to the system supporting single-hop operation. In this paper, we further propose an adaptive scheme which allocates bandwidths to each SS and RS in IEEE 802.16j network. The design objective is to enhance system aggregate throughput and ensure fairness among TCP flows. The performance of our scheme is validated by ns2 simulations [12]. The results show that our scheme outperforms static allocations, offering more efficient bandwidth utilization and better adaptability to network dynamics.
The rest of this paper is organized as follows. Section II presents the proposed bandwidth allocation scheme for IEEE 802.16j wireless networks with transparent relay stations. Section III shows the simulation results via the ns-2 simulator. Finally, we conclude this paper in Section IV.

II. ADAPTIVE BANDWIDTH ALLOCATION FOR IEEE 802.16J WIRELESS NETWORK

A. System Model and Problem Specification

We consider an IEEE 802.16j network consisting of one BS, one RS and multiple SSs. Each SS requests bandwidth for its uplink data connection. We assume that the scheduling strategy used by BS is First Come First Served (FCFS) due to its simplicity. Our scheme performs very well even though the simplest scheduling strategy is used.

Each TDD frame is of a fixed duration containing $N_f$ time slots and each slot has a fixed length of $t_m$ seconds. Each frame is divided into a DL sub-frame and an UL sub-frame, which have $N_d$ and $N_u$ slots, respectively. In the UL sub-frame, $N_c$ slots are designated as bandwidth request contention slots, which are used for initial ranging and broadcast polling. Within each DL sub-frame and UL sub-frame, the access zone has $N_{az}^d$ slots and $N_{az}^u$ slots, respectively, while the relay zone has $N_{rz}^d$ slots and $N_{rz}^u$ slots, respectively.

IEEE 802.16j networks with centralized resource allocation use transparent RSs. The BS is responsible to grant each SS's both DL and UL transmissions in the access zone and relay zone. Our goal is to determine the bandwidth allocation for each portion in Fig. 1 such that for both downloading and uploading TCP transfers, the aggregate network throughput is enhanced.

B. Bandwidth Asymmetry Ratio for TCP over a Direct Link

Asymmetric networks are networks with different DL and UL capacities. The main impact of bandwidth asymmetry on TCP performance is that TCP's self-clocking may be disturbed. An important parameter $k$, called the bandwidth asymmetry ratio, is frequently employed to investigate this phenomenon [6-9]. Denote by $L_{data}$ and $L_{ack}$, respectively, the sizes of a data packet and an ACK packet. The asymmetry ratio $k$ for one-way TCP transfers [6] is defined as follows.

$$k = \frac{\text{Rate of TCP data packets}}{\text{Rate of ACK packets}} = \frac{B_d \times L_{ack}}{B_u \times L_{data}},$$  

where TCP data packets are transmitted on the downlink channel and acknowledged on the uplink channel. When $k$ is less than or equal to one, TCP operates normally. However, when $k$ exceeds one, ACK packets arrive on the bottleneck link in the reverse direction at a rate faster than what the link can support. Thus, the ACK packets may fill up the sending buffer on the reverse bottleneck link rapidly. This causes an increase in the time between consecutive ACKs received by the sender and also an increase in the dropping rate of the ACK packets for the buffer. As a result, the sender slows down the growth of the congestion window, leading to throughput degradation.

C. Bandwidth Asymmetry Ratio for TCP over IEEE 802.16

The asymmetry ratio $k$ above can only be applied to point-to-point links (and for one-way TCP transfers only), but is not applicable to IEEE 802.16 multi-access links. In addition, when considering different burst profiles for different SSs, the time slots needed for different data transfers are different. Thus, the formula of asymmetry ratio should be independent of the burst profile. In IEEE 802.16, data transfer is maintained at the granularity of “connection” by connection ID (CID). Therefore, TCP traffic can be handled separately from other BE applications. The bandwidth request is made per connection while the bandwidth grant is made per SS. The bandwidth request is in terms of the number of bytes excluding the PHY overhead. The downlink queue size in bytes can be calculated by BS. Therefore, the downlink and uplink traffic information can be gathered at BS by checking the downlink queue and the bandwidth request. To determine whether a BE connection is a downloading or an uploading TCP transfer, BS can investigate the bandwidth requests sent by the SS or trace the sizes of the packets sent.

Let $S_{kd} = \{CID_1, CID_2, \cdots, CID_{dn}\}$ be the set of downloading CID SS processes. In the one-way downloading TCP case, for each SS per frame, (1) can be rewritten as
follows.

\[ k_d = \frac{\sum_{i=1}^{n_{CID}} SS_i DLqueue_data_CID \times L_{data}}{L_{data} \times \sum_{i=1}^{n_{CID}} SS_i BWreq_ack_CID}, \]  

where \( \sum_{i=1}^{n_{CID}} SS_i DLqueue_data_CID \) is the sum of the downlink queue size of each connection of SS that has packet size of the order of a data packet. \( \sum_{i=1}^{n_{CID}} SS_i BWreq_ack_CID \) is the sum of the amount of bandwidth request from each connection that has packet size of the order of an ACK packet. Similarly, in the one way uploading TCP case, for each SS per frame, (1) can be rewritten as follows.

\[ k_u = \frac{\sum_{i=1}^{n_{CID}} SS_i BWreq_data_CID \times L_{data}}{L_{data} \times \sum_{i=1}^{n_{CID}} SS_i DLqueue_ack_CID}, \]  

where \( \sum_{i=1}^{n_{CID}} SS_i DLqueue_ack_CID \) is the sum of the downlink queue size of each connection of SS that has packet size of the order of an ACK packet, and \( \sum_{i=1}^{n_{CID}} SS_i BWreq_data_CID \) is the sum of the amount of bandwidth request from each connection that has packets size of the order of a data packet.

When \( k \) is less than or equal to one, TCP operates normally. Therefore, our goal is to adjust our DL and UL allocations so that \( k_d \) and \( k_u \) are greater than one only when the numerators are larger than the denominators. Therefore, we should adjust the amount of transmissible data in downlink and uplink. \( k_d \leq 1 \) only when \( \sum_{i=1}^{n_{CID}} SS_i DLqueue_data_CID \leq \frac{L_{data} \times \sum_{i=1}^{n_{CID}} SS_i BWreq_ack_CID}{L_{ack}} \).

For downlink, when \( k_d > 1 \), instead of allowing \( \sum_{i=1}^{n_{CID}} SS_i DLqueue_data_CID \) amount of data to be transmitted, BS should allocate downlink bandwidth so that only \( \frac{L_{data} \times \sum_{i=1}^{n_{CID}} SS_i BWreq_ack_CID}{L_{ack}} \) can be transmitted. This causes \( k_d = 1 \) and TCP can operate normally. A similar operation can also be used for uplink bandwidth allocation so that \( k_u \leq 1 \). Note that the expressions above (i.e., (2) and (3)) can also be applied to two-way TCP transfers. For two-way TCP transfers, both uploading and downloading TCP flows co-exist in the network. Both \( k_d \) and \( k_u \) have to be computed for each SS. Since data transfer in IEEE 802.16 is maintained at the granularity of connection, different uploading and downloading TCP flows can be identified by their CIDs. Therefore, the \( k_d \) and \( k_u \) of each SS can be calculated accordingly.

**D. Adaptive Bandwidth Allocation Scheme**

Our Adaptive Bandwidth Allocation Scheme is developed according to (2) and (3) and is summarized in Table I. The rationale behind this algorithm is to calculate the downlink and uplink allocation in bytes for each SS so that TCP can operate normally. The exact number of downlink time slots required to transmit the allowed total downlink packets for \( SS_k \) is

\[ N_{d_{-SS_k}} = \frac{downlink\_allocation\_in\_bytes}{R_{DLSS_k} \times t_m}, \]  

and the exact number of uplink time slots required to transmit the allowed total uplink packets for \( SS_k \) is

\[ N_{u_{-SS_k}} = \frac{uplink\_allocation\_in\_bytes}{R_{ULSS_k} \times t_m}, \]  

where \( R_{DLSS_k} \) and \( R_{ULSS_k} \) denote the transmission rates of downlink burst and uplink burst, respectively, for \( SS_k \) in the next frame, and \( t_m \) is the duration of a time slot. With (4) and (5), we can determine the bandwidth (the exact number of time slots) allocated to each \( SS_k \).

In IEEE 802.16j centralized resource allocation networks, the BS has to know the path from the RS and the processing time at each RS in the cell. When BS allocates bandwidth to forward a packet to or from a given SS, it shall allocate bandwidth on all links (relay and access) that make up the path. The BS has the information of all connections on the path. Both
the access link and relay link have to satisfy \( k_d \leq 1 \) and \( k_u \leq 1 \). Our algorithm can calculate the downlink and uplink allocation in bytes of each SS so that TCP can operate normally. The allocations in bytes are the same on all links. The difference is the exact number of time slots required for each link to transmit the allowed packets.

The exact number of downlink time slots required for \( RS_j \) is

\[
N_{d\_RS_j} = \frac{RS_{\_downlink\_allocation\_in\_bytes}}{R_{ULRS_j} \cdot t_m},
\]

and the exact number of uplink time slots required for \( RS_j \) is

\[
N_{u\_RS_j} = \frac{RS_{\_uplink\_allocation\_in\_bytes}}{R_{ULRS_j} \cdot t_m}.
\]

The exact number of downlink and uplink time slots for \( SS_k \) on the access link can be obtained by (4) and (5), respectively.

After \( N_{d\_SS_k} \), \( N_{u\_SS_k} \), \( N_{d\_RS_j} \) and \( N_{u\_RS_j} \) for all SSs and RSs are determined, BS can adjust the split of two parts: (i) between uplink and downlink (ii) between access zone and relay zone as follows.

\[
\begin{align*}
N_d - N_b &= \sum_{k=1}^{N_{SS}} N_{d\_SS_k} + \sum_{j=1}^{N_{RS}} N_{d\_RS_j}, \\
N_u - N_c &= \sum_{k=1}^{N_{SS}} N_{u\_SS_k} + \sum_{j=1}^{N_{RS}} N_{u\_RS_j},
\end{align*}
\]

where \( N_{SS} \) and \( N_{RS} \) are the number of SSs such that

\[
\sum_{k=1}^{N_{SS}} N_{d\_SS_k} + \sum_{j=1}^{N_{RS}} N_{d\_RS_j} + \sum_{k=1}^{N_{SS}} N_{u\_SS_k} + \sum_{j=1}^{N_{RS}} N_{u\_RS_j} + N_c \leq N_f.
\]

Therefore, we have

\[
\begin{align*}
N_d^{\alpha} &= \sum_{k=1}^{N_{SS}} N_{d\_SS_k}, \\
N_u^{\alpha} &= \sum_{k=1}^{N_{SS}} N_{u\_SS_k}, \\
N_d^{\beta} &= \sum_{j=1}^{N_{RS}} N_{d\_RS_j}, \\
N_u^{\beta} &= \sum_{j=1}^{N_{RS}} N_{u\_RS_j}.
\end{align*}
\]

BS can use (8)-(13) to determine the exact number of time slots (or bandwidths) allocated to UL and DL channels in their frames in a centralized manner. BS can then inform its sub-ordinate stations of the bandwidth allocation results, i.e., the durations of DL and UL sub-frames, access zone, relay zone and the transmission opportunity of each sub-ordinate station, via the MAP message. BS then informs the scheduler module about the new values of the allocated schedulable bandwidths for proper operations.

### III. PERFORMANCE EVALUATION

In this section, we evaluate the performance of our proposed scheme via ns2 simulations. The network parameters used in the simulation are listed in Table II. In this simulation, we compare the performance of the proposed scheme for dynamic ratio with some three particular static ratio settings. The ratio of “1:1” is the case of equal splits of the DL and UL bandwidth; the ratio of “2:1” is the default value specified in the WiMAX standard; the ratio of “10:1” is the typically setting in DOCSIS-based CATV access networks. The aggregate throughput and the Fairness Index (i.e., \( FI = \frac{\sum_{i=1}^{n} x_i}{n} \sum_{i=1}^{n} x_i^2 \), where \( x_i \) is the throughput of the \( i \)th flow and \( n \) is the total number of flows) are evaluated. The curve “ABA” represents the result of using our proposed scheme and the curves “x:y” represent the results of using static DL to UL bandwidth ratio x:y.

In Scenario 1, we vary the number of SSs in the network. All TCP transfers are downloading. All SSs under BS use QPSK 1/2 and all SSs under RS use the QAM16 3/4 modulation and coding schemes on both relay links and access links. The results are shown in Fig. 2. In Scenario 2, we vary the ratio of downloading to uploading TCP transfers in the network and the results are shown in Fig. 3. As can be seen, our scheme achieves higher aggregate throughput regardless of the presence of RS and the number of SSs in the networks with different modulation and coding schemes (i.e., different channel conditions). In addition, the Fairness Index of our scheme is close to 1.

We observe that when most of the TCP transfers are uploading, the “1:1” allocation has higher aggregate throughput than the other static approaches (“2:1” and “10:1”) as more uplink bandwidth is allocated. On the other hand, the “10:1” allocation has the lowest aggregate throughput when all the TCP transfers are uploading. The static approaches use a fixed downlink to uplink bandwidth ratio, and each of them is appropriate only for a particular traffic profile and modulation and coding schemes used by the SSs. When all TCP transfers are uploading, all the data packets are sent in the UL sub-frame.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>data rate</td>
<td>10Mbps</td>
</tr>
<tr>
<td>( t_m )</td>
<td>0.35 ms</td>
</tr>
<tr>
<td>( N_c )</td>
<td>1700</td>
</tr>
<tr>
<td>( N_f )</td>
<td>612</td>
</tr>
<tr>
<td>( N_f )</td>
<td>22848</td>
</tr>
<tr>
<td>TCP DATA segment size</td>
<td>1000 bytes</td>
</tr>
</tbody>
</table>

**TABLE II**

**PARAMETER SETTINGS USED IN THE SIMULATION**
All flows experience the same bottleneck for the UL bandwidth. Therefore, the fairness is high. However, when a small portion of the TCP transfers is downloading, the data packers are more easily sent through the abundant downlink bandwidth. Therefore, the throughput of the downloading TCP transfers is much higher than those of the uploading TCP transfers, especially when the uplink bandwidth is very limited in the 10:1 case. Therefore, the fairness indexes of the static allocations are low in this scenario. When the percentage of downloading TCP transfers increases, less data packets are needed to be sent at the limited uplink bandwidth. The unfairness is thus reduced.

Our scheme adaptively adjusts the bandwidth allocation according to the traffic profile and channel condition so that appropriate bandwidth is allocated for each SS and RS, thus improving the bandwidth utilization, and offering higher aggregate throughput with reasonable fairness.

IV. CONCLUSION

In this paper, we have proposed an adaptive bandwidth allocation scheme for TCP-based best effort traffic according to the current traffic profile, channel condition, and transport layer parameters. The proposed scheme can also be applied to any multi-hop relay network scenarios. The simulation results show that our scheme outperforms static allocations, offering higher aggregate throughput and better adaptability to network dynamics. Since this proposed bandwidth allocation scheme can enhance the TCP performance, the most popular Internet TCP based applications can be supported by IEEE 802.16 with good performance.

ACKNOWLEDGEMENTS

This work was supported in part by Excellent Research Projects of National Taiwan University, under Grant Number 95R0062-AE00-04, and in part by National Science Council, Taiwan, under Grant Number NSC96-2628-E-002-004-MY3.

REFERENCES