Adaptive Downlink and Uplink Channel Split Ratio Determination for TCP-Based Best Effort Traffic in TDD-Based WiMAX Networks

Chih-He Chiang, Wanjiun Liao, Tehuang Liu, Iam Kin Chan, and Hsi-Lu Chao

***Abstract*—In this paper, we study the determination of downlink (DL) and uplink (UL) channel split ratio for Time Division Duplex (TDD)-based IEEE 802.16 (WiMAX) wireless networks. In a TDD system, uplink and downlink transmissions share the same frequency at different time intervals. The TDD framing in WiMAX is adaptive in the sense that the downlink to uplink bandwidth ratio may vary with time. In this work, we focus on TCP based traffic and explore the impact of improper bandwidth allocation to DL and UL channels on the performance of TCP. We then propose an Adaptive Split Ratio (ASR) scheme which adjusts the bandwidth ratio of DL to UL adaptively according to the current traffic profile, wireless interference, and transport layer parameters, so as to maximize the aggregate throughput of TCP based traffic. Our scheme can also cooperate with the Base Station (BS) scheduler to throttle the TCP source when acknowledgements (ACKs) are transmitted infrequently. The performance of the proposed ASR scheme is validated via ns-2 simulations. The results show that our scheme outperforms static allocation (such as the default value specified in the WiMAX standard and other possible settings in existing access networks) in terms of higher aggregate throughput and better adaptivity to network dynamics.**

***Index Terms*—IEEE 802.16, WiMAX, TDD, bandwidth allocation, channel allocation ratio.**

I. Introduction

IEEE 802.16 (WiMAX) is an emerging last mile technology for broadband wireless access [1]–[6]. A typical IEEE 802.16 network consists of base stations (BSs) and subscriber stations (SSs). The IEEE 802.16 standard [7] specifies two modes of operations, namely, point-to-multipoint (PMP) and mesh (optional) modes. In the PMP mode, the transmissions from SSs to BS are centrally coordinated by the BS, and the transmissions in both directions take place directly between BS and SSs. In the mesh mode, the transmissions can occur between SSs, and can also be relayed via other SSs. IEEE 802.16 supports both Frequency Division Duplex (FDD) and Time Division Duplex (TDD). In an FDD system, the uplink and downlink channels are located on separate frequencies and data can be transmitted simultaneously. With TDD, the uplink and downlink transmissions usually share the same frequency at different time intervals. In this paper, we focus only on the TDD system.

Each 802.16 TDD frame is slotted in time and is of a fixed duration. Each frame can be decomposed into one downlink (DL) sub-frame and one uplink (UL) sub-frame, as shown in Fig. 1. The TDD framing is adaptive in the sense that the downlink to uplink bandwidth ratio may vary with time so as to optimize the network performance. The downlink, i.e., from BS to SS, is a broadcast channel. At the start of each frame, BS broadcasts MAP control messages (i.e., DL-MAP and UL-MAP for downlink and uplink, respectively) to inform the SSs which are located within the same antenna sector about the usage and allocation of time slots in the DL and UL sub-frames. Upon receiving the MAP messages, all the SSs know when and how long they are allowed to receive or send their data. SSs in the same antenna sector share the uplink to BS on an on-demand basis. IEEE 802.16 defines several uplink scheduling mechanisms, including unsolicited bandwidth grants, polling, and contention to support four types of uplink scheduling services, namely, Unsolicited Grant Service (UGS), Real-time Polling Service (rtPS), Non-real-time Polling Service (nrtPS), and Best Effort (BE), for different types of applications [4]. The BS scheduler must provide a poll and/or grant to each connection at the appropriate time such that the QoS associated with the scheduling service for each connection can be satisfied. In this paper, we focus on the BE scheduling service, in which each SS asks bandwidth for a connection by sending a request to BS within the bandwidth request contention slots in the uplink sub-frame. If the request is corrupted due to a collision, the SS enters the contention resolution process based on a truncated binary exponential backoff. With the BE scheduling service, SSs can also request bandwidth via Piggyback Bandwidth Requests.

A challenge in 802.16 TDD systems is in the determination of the ratio of downlink to uplink bandwidth capacities between the BS and SSs. An inappropriate ratio may significantly degrade the total system performance. This bandwidth ratio determination problem is even more complicated when the transport layer issue is taken into account. The last mile access for residential users tends to be asymmetric (i.e., more demands on downloading and less on uploading). As a result, an equal split between the uplink and downlink channels may lead to inefficient bandwidth utilization. This explains why existing broadband access technologies are mostly asymmetric in nature. Many research efforts [8]–[11] have shown that

Manuscript received 15 January, 2008; revised 4 July 2008.

Chih-He Chiang, Wanjiun Liao, and Tehuang Liu are with the Department of Electrical Engineering, National Taiwan University, Taipei, 106, Taiwan, R.O.C. (e-mail: wjliao@ntu.edu.tw).

Iam Kin Chan is with the Graduate Institute of Communication Engineering, National Taiwan University, Taipei, Taiwan.

His-Lu Chao is with the Department of Computer Science and Information Engineering, National Chiao-Tung University, Hsin-Chu, Taiwan.

Digital Object Identifier 10.1109/JSAC.2009.090209.

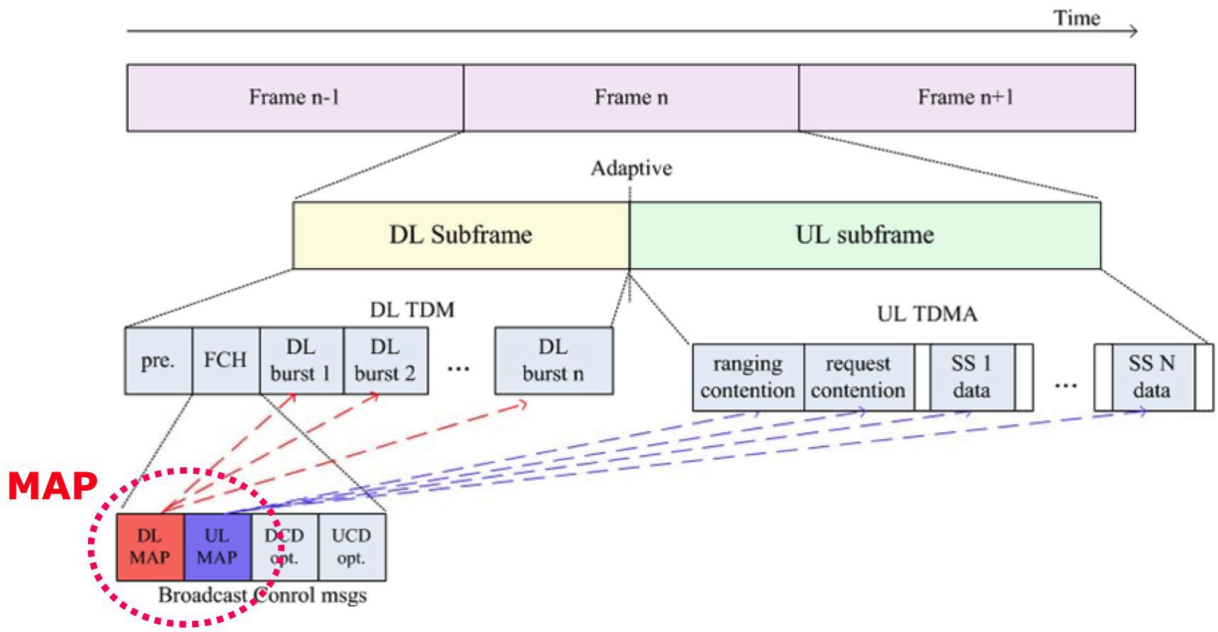


Fig. 1. The TDD framing in IEEE 802.16 (WiMAX)

bandwidth asymmetry impacts the network performance substantially, especially for TCP traffic, which relies on feedback of acknowledgments from the receiver to ensure reliability. This is because TCP is self-clocked and any disruption in the feedback process may impair the performance of the forward data transfer. Improper allocation will cause TCP ACK packets to be accumulated in the uplink queue and therefore the aggregate network throughput will degrade accordingly.

In this paper, we determine the bandwidth split ratio between uplink and downlink for TCP based best effort traffic in TDD-based WiMAX networks. Specifically, we account for the impact of bandwidth asymmetry on the performance of TCP transfers in IEEE 802.16 networks and attempt to determine the optimal channel split ratio for WiMAX TDD channels. We analyze the behavior of TCP over the WiMAX MAC mechanism, and consider the interference (e.g., location-dependent and time-variant) in wireless environments. We then propose a mechanism called Adaptive Split Ratio (ASR) scheme which allows BS to allocate bandwidths to uplink and downlink channels according to the current traffic profile, wireless interference, and transport layer parameters. In our scheme, we consider TCP flows only and allow both downloading (i.e., TCP data flows on the downlink and ACK packets on the uplink) and uploading (i.e., TCP data flows on the uplink and ACK packets on the downlink). Therefore, the result can serve as the guideline of bandwidth allocation for BE scheduling. The performance of our scheme is evaluated via ns2 simulations. The results show that ASR outperforms static allocations (i.e., with fixed downlink and uplink bandwidth allocations, such as the default value specified in the WiMAX standard and other possible settings in existing access networks) in terms of more efficient radio resource utilization and better adaptivity to network dynamics.

The rest of this paper is organized as follows. Section II presents the proposed mechanism for WiMAX wireless networks. Section III shows the simulation results via the ns-2 simulator. Finally, we conclude this paper in Section IV.

II. Adaptive Bandwidth Allocation for DL and UL Channels in WiMAX

In this section, we determine the capacity split ratio between uplink and downlink channel by jointly considering the issues of the transport layer protocol, the MAC layer operation in WiMAX networks, and the interference in the physical (PHY) layer. We focus on the point-to-multipoint (PMP) mode of IEEE 802.16, which is the primary operating mode of WiMAX for residential users. Without loss of generality, our discussion below is based on one BS serving multiple SSs within the same antenna sector.

A. System Model and Problem Specification

In accordance with IEEE 802.16 standard [7], the bandwidth request is made per connection in terms of the number of bits excluding the PHY overhead, and the bandwidth grant is made per SS. Each bandwidth request is made with the best effort (BE) scheduling service via the broadcast polling mechanism. In other words, when a poll is directed with a broadcast connection identification (CID), the connections which have data (or ACK) packets to send contend for the link by sending request messages to the BS within the bandwidth request contention slots. With BE scheduling, the connections can also use PiggyBack Requests to request for time slots.

We assume that the scheduling strategy used by BS for bandwidth requests is First Come First Served (FCFS). Upon receiving an aggregate bandwidth request from an SS[1], the BS

[1]Strictly speaking, each request comes from a connection (i.e., one TCP transfer in this paper) associated with one SS. Since the BS grants the bandwidth per SS, requests from the connections associated with the same SS tend to come in sequence and therefore can be regarded as one aggregate request from the SS. The amount of requested bandwidth is calculated as the sum of the amount of bandwidth request from each connection of the SS.

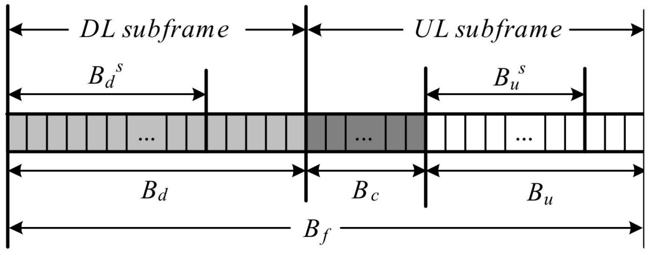


Fig. 2. The TDD frame structure

grants the SS one transmission opportunity in each uplink sub-frame and each SS can transmit up to n data frames within one granted transmission opportunity. For simplicity and without loss of generality, we set n to 1 in the rest of the paper.

Each TDD frame (as shown in Fig. 2) is of a fixed duration containing N_f time slots and each slot has a fixed length of t_{ms} seconds. Each frame is divided into a downlink (DL) sub-frame and an uplink (UL) sub-frame, which contain 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. The TDD framing is adaptive in the sense that the downlink to uplink bandwidth ratio can vary with time.

Our goal is to determine the *schedulable bandwidth* allocated to downlink and uplink TCP traffic (i.e., the portions of B_d^s and B_u^s in Fig. 2 containing N_d^s and N_u^s slots, respectively) such that the aggregate throughput of both downloading and uploading TCP transfers can be maximized.

B. Notation

The notations and symbols used in this paper are summarized in Table I.

TABLE I
NOTATION/SYMBOLS

L_{data}	The size of one TCP data packet in terms of bits
L_{ack}	The size of one TCP ACK packet in terms of bits
B_d	The downlink channel capacity in terms of bps
B_u	The uplink channel capacity in terms of bps
k	Asymmetry ratio, with a value between zero and one
n_{dSS}	The number of downloading TCP transfers in the system
n_{uSS}	The number of uploading TCP transfers in the system
η_d	Asymmetry ratio for downloading users
η_u	Asymmetry ratio for uploading users
T_{usv}	Average delay between two consecutive MAC layer data transmissions in the queue
d	TCP delay ACK parameter
N_c	Number of time slots for the contention period in a frame
N_{data}	Number of time slots for transmitting one TCP data packet in a frame
N_{ack}	Number of time slots for transmitting one TCP ACK packet in a frame
t_m	One slot time in terms of sec in a WiMAX TDD frame
B	Total capacity available for BE scheduling service in the system
N_u	Total number of time slots allocated for uplink channel
N_d	Total number of time slots allocated for downlink channel
N_f	Total number of time slots in one TDD frame

C. Bandwidth Asymmetry Ratio for TCP over a Direct Link

Asymmetric networks are networks with different downlink and uplink capacities. The main impact of bandwidth asymmetry on TCP operation 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 [8], [9]. Denote the sizes of a data packet and an ACK packet (including IP and MAC headers) by L_{data} and L_{ack}, respectively. The asymmetry ratio k for one-way TCP transfers [8] is defined as follows.

$$k = \frac{\text{Rate of TCP data packets}}{\text{Rate of ACK packets}} = \frac{B_d}{B_u} \times \frac{L_{ack}}{L_{data}}, \quad (1)$$

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 duration between consecutive ACKs arriving at the sender and also an increase in the dropping rate of the ACK packets to the buffer. As a result, the sender will slow down the growth of the congestion window, leading to throughput degradation.

D. Bandwidth Asymmetry Ratio for TCP over WiMAX

The asymmetry ratio k in (1) can be applied only to point-to-point links (and for one-way TCP transfers only), but is not applicable to multi-access links such as WiMAX which needs MAC layer protocols [12]. To address this issue, we derive an asymmetry ratio for WiMAX networks which employ a dynamic TDMA MAC protocol to arbitrate channel access among nodes, similar to what we have done in [11]. We consider two-way TCP transfers, i.e., data and ACK packets can go in both directions of the link. For ease of explanation, we first assume that each SS has at most one TCP transfer in either uplink or downlink direction and the channel condition between each SS and the BS is identical. Later we will generalize our discussion to allow each SS to perform multiple TCP transfers (i.e., multiple connections) and the channel conditions of different SSs may not be the same.

Suppose that there are n_{dSS} downloading TCP transfers and n_{uSS} uploading TCP transfers which simultaneously co-exist in the network and each of which is issued by a different SS. Now, SSs share the downlink and uplink bandwidths to transmit their data and ACK packets. Thus, the asymmetry ratio for downloading users, denoted by η_d, can be expressed by (2).

$$\eta_d = \frac{\alpha \times B_d}{L_{data} \times n_{dSS}} \times T_{usv}, \quad (2)$$

where $\alpha = \frac{n_{dSS} L_{data}}{n_{dSS} L_{data} + \frac{n_{uSS} L_{ack}}{d}}$ and $T_{usv} = \frac{L_{ack}}{d \times B_u}$.

This new ratio is obtained as follows. Considering the uplink channel in WiMAX networks being a TDMA-like channel, we replace $\frac{L_{ack}}{B_u}$ in (1) with T_{usv}, where T_{usv} is the average delay for an SS sending two consecutive packets out from the queue (i.e., the duration that an SS must wait

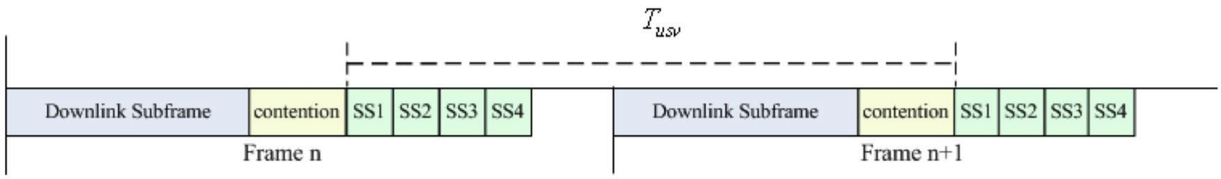


Fig. 3. T_{usv} in Case 1

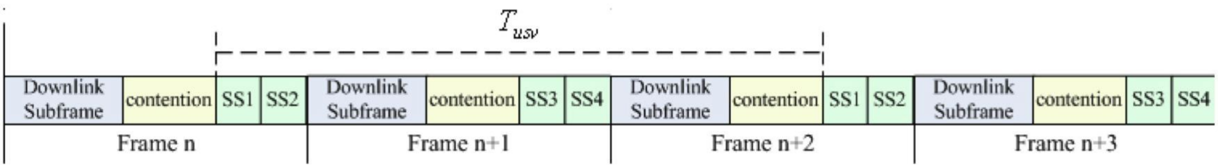


Fig. 4. T_{usv} in Case 2

for its scheduled time, as shown in Figs. 3 and 4). We further consider the delayed ACK policy, thus $T_{usv} = \frac{L_{ack}}{d \times B_u}$, where d is the delayed ACK parameter (i.e., sending one ACK packet per d data packets). Since only a portion of the downlink bandwidth is shared by TCP transfers, namely, $\alpha = \frac{n_{dSS} L_{data}}{n_{dSS} L_{data} + \frac{n_{uSS} L_{ack}}{d}}$, the effective downlink bandwidth for TCP data packet transmissions is $\alpha \cdot B_d$. Moreover, since the effective downlink capacity is shared by all n_{dSS} TCP flows, $\frac{B_d}{L_{data}}$ in (1) is replaced by $\frac{\alpha \cdot B_d}{L_{data} \times n_{dSS}}$. By putting them all together, we obtain η_d in (2).

The value of T_{usv} can be derived in two cases:

• Case 1 (Light Traffic Load):
$N_f > N_c + n_{dSS} \cdot N_{ack} + n_{uSS} \cdot N_{data} + n_{dSS} \cdot N_{data} + n_{uSS} \cdot N_{ack}$, where N_c denotes the number of time slots for the contention period, and N_{data} and N_{ack} denote the numbers of time slots for transmitting one TCP data packet and one ACK packet, respectively.

In this case, the total bandwidth requirement for all TCP transfers in both directions is less than the link capacity (see Fig. 3 for example). Therefore, each SS must wait for one frame time between sending two consecutive packets out from the queue, i.e.,

$$T_{usv} = N_f \times t_m. \tag{3}$$

• Case 2 (Heavy Traffic Load):
$N_f < N_c + n_{dSS} \cdot N_{ack} + n_{uSS} \cdot N_{data} + n_{dSS} \cdot N_{data} + n_{uSS} \cdot N_{ack}$.

In this case, the number of SSs in the network is relatively large such that the resource is insufficient to allocate each requesting SS an uplink transmission opportunity in every TDD frame (see Fig. 4 for example). Thus, each SS cannot be granted one transmission opportunity in every frame. On the average, each SS is granted once every $\frac{n_{dSS} N_{ack} + n_{uSS} N_{data}}{N_u - N_c}$ frames. Therefore, T_{usv} can be expressed by

$$T_{usv} = \begin{bmatrix} \frac{n_{dSS} N_{ack} + n_{uSS} N_{data}}{N_u - N_c} \times (N_d + N_c) \\ + n_{dSS} N_{ack} + n_{uSS} N_{data} \end{bmatrix} \times t_m, \tag{4}$$

where t_m is a slot time duration in a TDD frame.

Similarly, we can obtain the asymmetry ratio η_u for uploading users in WiMAX networks by (5). Again, with the ratio of the aggregate sending rate of data packets to the aggregate sending rate of ACK packets on the links in both directions, we have

$$\eta_u = \frac{1}{T_{usv}} \times \frac{L_{ack} \times n_{uSS}}{\beta \times B_d} \times \frac{1}{d}, \tag{5}$$

where $\beta = \frac{\frac{n_{uSS} L_{ack}}{d}}{n_{dSS} L_{data} + \frac{n_{uSS} L_{ack}}{d}}$, because there is now only a portion of downlink capacity used to carry ACK packets for uploading SSs, and the data packets of the SSs go in the uplink direction.

Interestingly, from (2) and (5), we obtain

$$\eta_d \eta_u = 1. \tag{6}$$

Our goal is to make $\eta_d \leq 1$ and $\eta_u \leq 1$. From (2), (5), and (6), the only possible solution is

$$\eta_d = \eta_u = 1. \tag{7}$$

E. Adaptive Channel Split Ratio Adjustment Mechanism

Our Adaptive Split Ratio (ASR) Scheme is then developed according to (7). Note that since the traffic pattern in the network changes with time, our scheme can be performed periodically. In the following, we elaborate on the operation of ASR step by step.

Step 1: BS retrieves the information about n_{dSS} and n_{uSS} from its connection table. To determine whether a BE connection is a downloading or an uploading TCP transfer, BS can investigate the bandwidth request sent by the SS or trace the size of the packet sent.

Step 2: To ensure that the bandwidth asymmetry ratios for both downloading and uploading TCP transfers are each equal to one, the allocated schedulable downlink and uplink bandwidths must satisfy (7).

We first consider the heavy-load case, i.e., $N_f < N_c + n_{dSS} \cdot N_{ack} + n_{uSS} \cdot N_{data} + n_{dSS} \cdot N_{data} + n_{uSS} \cdot N_{ack}$. Each frame (i.e., N_f) is composed of two parts: a DL sub-frame (i.e., N_d) and a UL sub-frame (i.e., N_u). The UL sub-frame can be further decomposed into a contention part (i.e., N_c) and Data Information Elements (IEs) (i.e., $N_u - N_c$). Note that here we consider only those portions of a frame used for Best Effort service. Thus, N_f can be regarded as the total number of time slots used for the BE scheduling service. Let the number of time slots allocated to N_d, $N_u - N_c$, and N_c

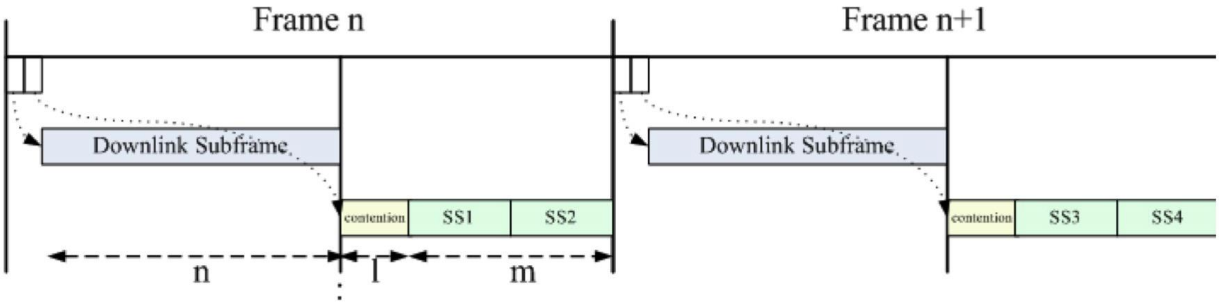


Fig. 5. An example for the numbers of time slots allocated to DL and UL sub-frames in each frame

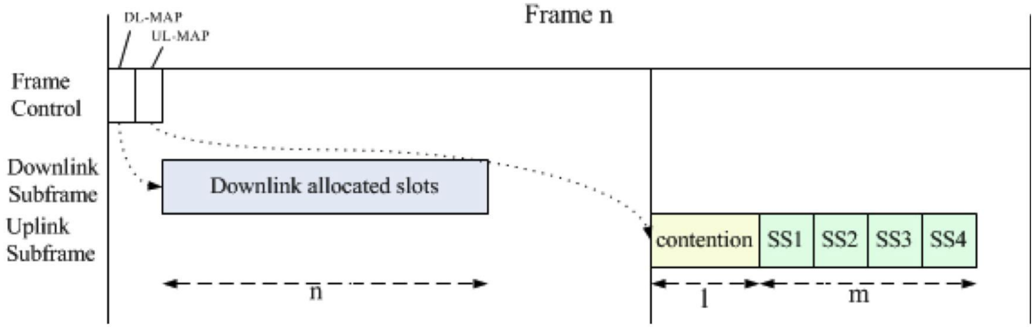


Fig. 6. The slot allocations if $N_f \geq N_c + n_{dSS} \cdot N_{ack} + n_{uSS} \cdot N_{data} + n_{dSS} \cdot N_{data} + n_{uSS} \cdot N_{ack}$

in N_f be n, m, and l, respectively, as shown in Fig. 5. To satisfy (7), asymmetry ratio η_d can be rewritten as follows.

$$\frac{\frac{n_{dSS}L_{data}}{n_{dSS}L_{data}+\frac{n_{uSS}L_{ack}}{d}} \times \frac{B_d}{d \times L_{data}} \times \left[\frac{n_{dSS}N_{ack}+n_{uSS}N_{data}}{N_u-N_c} \times (N_d+N_c) + n_{dSS}N_{ack}+n_{uSS}N_{data}\right] \times t_m}{n_{dSS}} = 1.$$

Let B denote the total capacity available for the BE scheduling service in the system. By substituting $B_d = \frac{n}{m+n+l}B$, $N_u - N_c = m$, and $N_d + N_c = n + l$ into the equation above, we obtain the ratio of n to m for η_d as

$$\frac{n}{m} = \frac{d \times n_{dSS}L_{data} + n_{uSS}L_{ack}}{n_{dSS}N_{ack} + n_{uSS}N_{data}} \times \frac{1}{d \times B \times t_{ms}}. \quad (8)$$

Similarly, the asymmetry ratio η_u can be rewritten as follows.

$$\frac{1}{\left[\frac{n_{dSS}N_{ack}+n_{uSS}N_{data}}{N_u-N_c} \times (N_d+N_c) + n_{dSS}N_{ack}+ n_{uSS}N_{data}\right] \times t_{ms}} \times \frac{n_{uSS}L_{ack}}{B_d} \times \frac{n_{dSS}L_{data}+\frac{n_{uSS}L_{ack}}{d}}{\frac{n_{uSS}L_{ack}}{d}} \times \frac{1}{d} = 1.$$

Again, by substituting $B_d = \frac{n}{m+n+l}B$, $N_u - N_c = m$, and $N_d + N_c = n + l$ into the equation above, we obtain the ratio of n to m for η_u as follows.

$$\frac{n}{m} = \frac{d \times n_{dSS}L_{data} + n_{uSS}L_{ack}}{n_{dSS}N_{ack} + n_{uSS}N_{data}} \times \frac{1}{d \times B \times t_{ms}}. \quad (9)$$

Note that the ratios of n to m in (8) and (9) are expressed identically.

Step 3: After the ratio of n to m is determined at Step 2, we can adjust the split between uplink and downlink as follows.

$$N_u = \frac{m}{m+n}(N_f - N_c) + N_c \quad (10)$$

$$N_d = \frac{n}{m+n}(N_f - N_c) \quad (11)$$

Note that (10) and (11) are obtained in the case that $N_f < N_c + n_{dSS} \cdot N_{ack} + n_{uSS} \cdot N_{data} + n_{dSS} \cdot N_{data} + n_{uSS} \cdot N_{ack}$. If $N_f \geq N_c + n_{dSS} \cdot N_{ack} + n_{uSS} \cdot N_{data} + n_{dSS} \cdot N_{data} + n_{uSS} \cdot N_{ack}$, i.e., as shown in Fig. 6, then we have

$$N_u = m = n_{dSS} \cdot N_{ack} + n_{uSS} \cdot N_{data}, \quad (12)$$

$$N_d = n = \frac{d \cdot n_{dSS} \cdot L_{data} + n_{uSS} \cdot L_{ack}}{d \cdot B \cdot t_{ms}}. \quad (13)$$

Together with (8) or (9), and (10)-(13), we can determine the exact number of time slots (or bandwidths) allocated to UL and DL channels in both cases.

Step 4: BS then informs SSs of the adjustment results, i.e., the durations of DL and UL sub-frames, via the *Allocation Start Time* field indicated in the MAP message. BS then informs the scheduler module about the new values of the allocated schedulable bandwidths in the downlink and uplink directions for proper operations.

F. Discussions

The discussion above is based on the assumption that each SS has at most one TCP transfer in either direction and the channel condition of each SS is identical. We now relax these assumptions. Let $S_k = \{CID_1, CID_2, \ldots, CID_k\}$ denote the set of CID processes for SS_k. For each SS per frame, say SS_k, the bandwidth requirements for TCP data and ACK packets in both directions can be expressed as follows.

- $\sum_{i=1}^{k} SS_k DLqueue\_data\_CID_i$ is the amount of bandwidth required for the downlink TCP data packet (i.e., downloading users' data transfers), corresponding to the sum of the downlink queue size of TCP data packets for each downlink connection of SS_k.
- $\sum_{i=1}^{k} SS_k DLqueue\_ack\_CID_i$ is the amount of bandwidth required for the downlink ACK packets (i.e., uploading users' ACK packets), corresponding to the sum of the downlink queue size of ACK packets for each downlink connection of SS_k.
- $\sum_{i=1}^{k} SS_k BWreq\_ack\_CID_i$ is the amount of bandwidth request for the uplink ACK packets (i.e., downloading users' ACK packets) , corresponding to the sum of the amount of the bandwidth request for ACK packets from each connection CID of SS_k.
- $\sum_{i=1}^{k} SS_k BWreq\_data\_CID_i$ is the amount of bandwidth request for the uplink TCP data packets (i.e., uploading users' data transfers), corresponding to the sum of the amount of the bandwidth request for TCP data from each connection CID of SS_k.

On the downlink, the total amount of bandwidth required for all downloading SSs, i.e., $n_{dSS}L_{data}$, can be replaced with $\sum_{k=1}^{n_{dSS}} \sum_{i=1}^{k} SS_k DLqueue\_data\_CID_i$, and the total bandwidth required for all uploading SSs, i.e., $n_{uSS}L_{ack}$, can be replaced with $\sum_{k=1}^{n_{uSS}} \sum_{i=1}^{k} SS_k DLqueue\_ack\_CID_i$. The value of α in (2) is then obtained accordingly. Similarly, on the uplink, the total amount of bandwidth required for all downloading SSs, i.e., $n_{dSS}L_{ack}$, can be replaced with $\sum_{k=1}^{n_{dSS}} \sum_{i=1}^{k} SS_k BWreq\_ack\_CID_i$, and the total amount of bandwidth required for all uploading SSs, i.e., $n_{uSS}L_{data}$, can be replaced with $\sum_{k=1}^{n_{uSS}} \sum_{i=1}^{k} SS_k BWreq\_data\_CID_i$. The value of β in (3) is then obtained accordingly.

We next examine the impact of different channel conditions between different SSs and the BS on the result. In IEEE 802.16, a set of burst profiles (i.e., a set of coding and modulation schemes) is defined to adapt to different channel conditions between the BS and SS. At the start of each connection and/or during each connection, each SS (based on its channel condition) will negotiate with the BS in order to determine the most appropriate burst profile for data transmission. Let R_{DLSS_k} and R_{ULSS_k} denote the transmission rates of downlink burst and uplink burst, respectively, for SS_k used in the next frame, and t_m is the duration of a time slot. For a downloading SS_k, the number of time slots required to transmit the total data packets on the downlink is

$$\frac{\sum_{i=1}^{k} SS_k DLqueue\_data\_CID_i}{R_{DLSS_k} \cdot t_m},$$

and the number of time slots required to transmit the total ACK packets on the uplink is

$$\frac{\sum_{i=1}^{k} SS_k BWreq\_ack\_CID_i}{R_{ULSS_k} \cdot t_m}.$$

The numbers of time slots required for an uploading SS on downlink and uplink can be found similarly. Therefore, the total numbers of time slots for TCP data and ACK packets on the uplink, i.e., $n_{uSS}N_{data}$ and $n_{dSS}N_{ack}$ in (4), respectively, can be expressed by

$$\sum_{k=1}^{n_{uSS}} \frac{\sum_{i=1}^{k} SS_k BWreq\_data\_CID_i}{R_{ULSS_k} \cdot t_m}, \text{ and}$$

$$\sum_{k=1}^{n_{dSS}} \frac{\sum_{i=1}^{k} SS_k BWreq\_ack\_CID_i}{R_{ULSS_k} \cdot t_m},$$

respectively.

TABLE II
PARAMETERS USED IN THE SIMULATIONS

Frame length	5 ms
Bandwidth	7 MHz
Modulation	QAM64 3/4
CP duration	16

III. Performance Evaluation

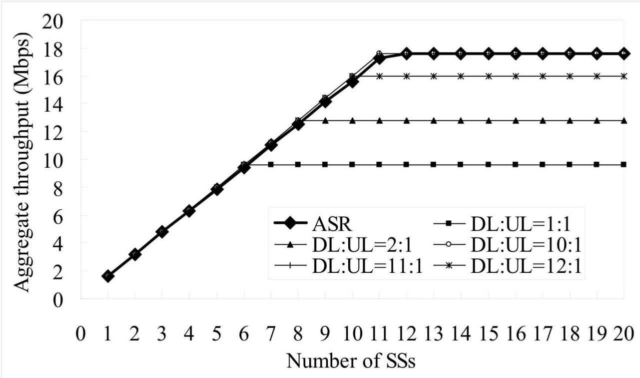
In this section, we show the effect of our proposed scheme ASR on the system performance via ns2 simulations. The IEEE 802.16 network parameters used in the simulation include: the link capacity is 25.4 Mbps, a time slot is 0.5 μs, $N_c = 650$, and $N_f = 10000$. Each TCP packet has a fixed length of 1000 bytes, and ACK packet is of 40 bytes. The setting of the WiMAX system in the simulation follows the specification defined in IEEE 802.16d [7]. Specifically, the values of the key parameters are summarized in Table II.

In this simulation, we compare the performance of the proposed adaptive ratio determination scheme with static ratios as defined in IEEE 802.16 standards. The curve "ASR" represents the result of our proposed scheme and the curves "DL:UL=x:y" represent the results of static downlink to uplink bandwidth ratio x:y. Specifically, we consider three static ratio settings. The ratio of "1:1" refers to the case of an equal split of the DL and UL bandwidth; the ratio of "2:1" is the default value specified in the WiMAX standard [7]; the ratio of "10:1" is the typically setting in DOCSIS-based CATV access networks. We simulate different split ratios in two scenarios, all with bulk FTP transfers. In the first scenario, there are only downloading TCP transfers in the system. We measure the aggregate throughput as a function of the number of TCP transfers in the network. In the second scenario, both downloading and uploading TCP transfers coexist simultaneously in the network. We fix the total number of TCP transfers in the network but vary the ratio of downloading to uploading TCP transfers in the network.

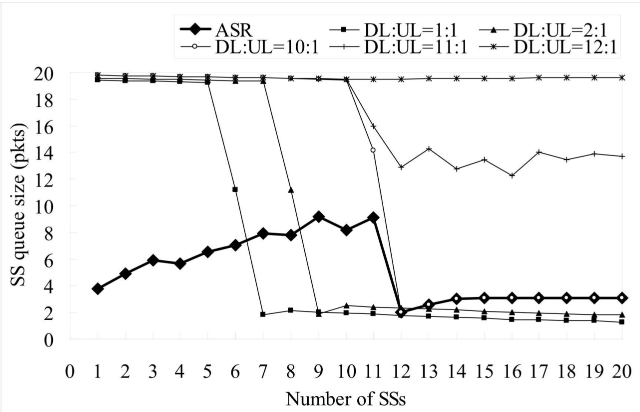
A. Scenario 1

In Scenario 1, we vary the number of TCP transfers in the network from 1 to 20. All TCP transfers are in downloading. Fig. 7 shows the aggregate downlink throughput versus the number of transfers in the network (i.e., Fig. 7(a)), together with the uplink buffer sizes at SSs (i.e., Fig. 7(b)) and the mean access delay for each TCP connection (i.e., Fig. 7(c)).

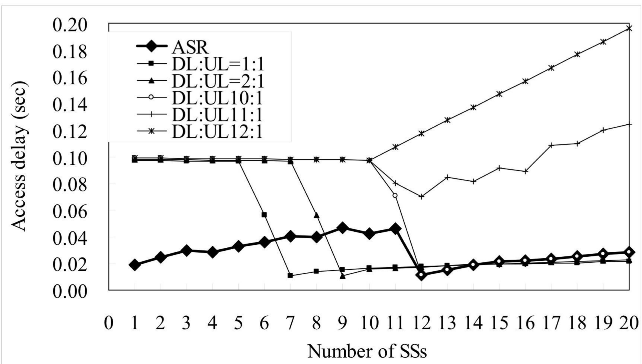
As can be seen, our scheme generates higher aggregate downstream throughput regardless of the number of TCP transfers in the network in most cases. The reasons are



(a) Aggregate throughput

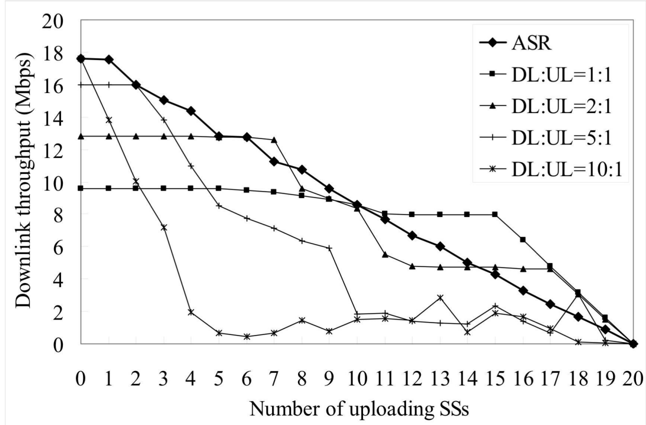


(b) Buffer sizes at SSs

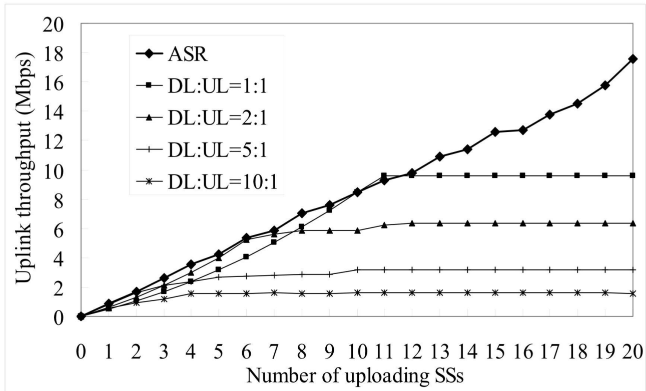


(c) Access delay

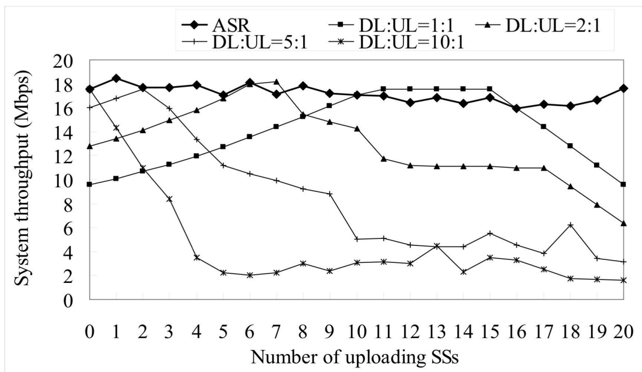
Fig. 7. Downlink aggregate throughput in Scenario 1



(a) Aggregate downlink throughput



(b) Aggregate uplink throughput



(c) System aggregate throughput

Fig. 8. Aggregate throughput in Scenario 2

described as follows. With an inappropriate bandwidth allocation ratio, either downlink or uplink is likely to become the bottleneck and degrades the network throughput. On the contrary, ASR computes the most appropriate ratio of DL to UL bandwidths and therefore improves the bandwidth utilization. In addition, since only limited transmission opportunities is granted to each SS in each TDD frame, when the number of SSs is small, the uplink ACK packets may not be able to get back to the sender in time. Our scheme helps the BS scheduler control the downlink schedulable bandwidth such that the phenomenon of uplink buffer overflow can be prevented. Therefore, ASR produces higher aggregate downlink throughput and smaller average uplink access delay. Note that in Scenario 1, the DL to UL bandwidth split ratio determined by our scheme is 10:1 because the number of time slots required to transmit a data packet and an ACK packet are 10 and 1, respectively. However, compared with the static approach using the DL/UL bandwidth ratio fixed at 10:1, our scheme generates lower uplink access delay while achieving the same level of aggregate throughput.

B. Scenario 2

In Scenario 2, we fix the total number of TCP transfers at 20 and vary the ratio of downloading to uploading TCP transfers in the network. Fig. 8 shows the aggregate downlink throughput (Fig. 8(a)), the aggregate uplink throughput (Fig. 8(b)), and the system aggregate throughput (i.e., the sum of downlink and uplink throughputs in Fig. 8(c)) in Scenario 2. Obviously, with ASR, the aggregate throughput in each direction is proportional to the number of existing TCP transfers in that direction. This implies that our scheme provides better fairness among TCP connections regardless of their directions. In addition, for most of the time, ASR generates higher aggregate throughputs in both directions than the static approaches. This is because ASR adaptively adjusts the DL/UL bandwidth ratio according to the numbers of downloading and uploading TCP transfers. Most importantly, ASR can cooperate with the scheduler to throttle the TCP source when ACK packets are infrequently transmitted, thus preventing either direction from becoming the bottleneck. On the contrary, the static approaches use a fixed downlink to uplink bandwidth ratio, which is very likely to create a bottleneck in either direction. Consequently, the network throughput is reduced and the bandwidth utilization degrades. The performance of static allocation is comparable with that of ASR only when the number of downloading and uploading TCP transfers in the network is close to the allocation ratio. For example, the curve 2:1 behaves well only when there are about 7 uploading transfers and 14 downloading transfers in the network.

IV. Conclusion and Future Work

IEEE 802.16 (WiMAX) is a promising technology for last-mile broadband wireless access. A performance challenge in 802.16 TDD systems is in determining the ratio of downlink to uplink capacities. An inappropriate ratio may significantly degrade the system performance due to poor bandwidth utilization. In this paper, we study adaptive channel split ratio of uplink to downlink capacities in TDD-based IEEE 802.16 (WiMAX) wireless networks. We focus on the BE scheduling service which aims to provide efficient service for the majority of existing Internet applications, such as Web browsing, FTP, P2P file sharing, etc. By investigating the network bandwidth asymmetry ratio, we develop an Adaptive Split Ratio (ASR) scheme which adjusts the downlink to uplink capacity ratio adaptively according to the current traffic profile, wireless interference, and transport layer parameters. ASR can also cooperate with the BS scheduler to throttle the TCP source when acknowledgements are transmitted infrequently, thus preventing either direction (i.e., downlink or uplink) from becoming the bottleneck. The simulation results show that our adaptive scheme outperforms static allocations in terms of higher aggregate throughput and better adaptivity to network dynamics.

In the future, we will extend our results to WiMAX wireless networks with relay stations. We will also consider the impact of cooperative communications on the split ratio determination problem for WiMAX relay networks.

Acknowledgment

This work was supported in part by the Excellent Research Projects of National Taiwan University, under Grant Number 97R0062-06, and in part by National Science Council (NSC), Taiwan, under Grant Number NSC96-2628-E-002-003-MY3.

References

[1] C. Eklund, R. B. Marks, K. L. Stanwood, and S. Wang, “IEEE standard 802.16: A technical overview of the WirelessMAN^{TM} air interface for broadband wireless access,” *IEEE Communications Magazine*, pp. 98–107, June 2002.

[2] A. Ghosh, D. R. Wolter, J. G. Andrews, and R. Chen, “Broadband wireless access with WiMAX/802.16: Current performance benchmarks and future potential,” *IEEE Communications Magazine*, pp. 129–136, 2005.

[3] B. Bennett, P. Hemmings, and B. A. Hamitton, “Operational considerations of deploying WiMAX technology as a last-mile tactical communications system,” *Proc. IEEE MILCOM*, 2006.

[4] M. Chatterje, S. Sengupta, and S. Ganguly, “Feedback-based real-time streaming over WiMAX,” *IEEE Wireless Communications*, no. 1, pp. 64–71, Feb. 2007.

[5] B. Rong, Y. Qian, and H.-H. Chen, “Adaptive power allocation and call admission control in multi-service WiMAX access networks,” *IEEE Communications Magazine*, no. 5, pp. 14–19, May 2007.

[6] G. Goth, “Is WiMAX’s future up in the air?” *IEEE Internet Computing*, no. 4, pp. 7–10, July-Aug. 2007.

[7] *IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems*, IEEE Std. 802.16-2004, 2004.

[8] V. N. P. H. Balakrishnan and R. H. Katz, “The effects of asymmetry on TCP performance,” *ACM Mobile Networks and Applications*, pp. 219–241, 1999.

[9] Y.-D. Li and W. Liao, “Improving TCP performance for asymmetric networks,” *Proc. IEEE ICC*, 2001.

[10] W. Liao and H.-J. Ju, “Adaptive slot allocation in DOCSIS-based CATV networks,” *IEEE Transactions on Multimedia*, no. 3, pp. 479–488, June 2004.

[11] W. Liao, “The behavior of TCP over DOCSIS-based CATV networks,” *IEEE Transactions on Communications*, pp. 1633–1642, June 2004.

[12] C.-H. Chiang, W. Liao, and T. Liu, “Adaptive uplink/downlink bandwidth allocation for IEEE 802.16 WiMAX networks: A cross-layer approach,” *Proc. IEEE Globecom*, 2007.

Chih-He Chiang received his BS degree in Computer Science and MS degree in Electrical Engineering from National Taiwan University, Taipei, Taiwan, in 2004 and 2006, respectively. He is now an engineer in HTC Corporation, Taiwan. His research interests focus mainly on broadband access networks.

Wanjiun Liao (M'97-SM'05) received her Ph.D. degree in Electrical Engineering from the University of Southern California, Los Angeles, CA, USA, in 1997. She joined the Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan, as an Assistant Professor in 1997, where she is now a full professor. Her research interests include wireless networks, multimedia networks, and broadband access networks.

Dr. Liao is currently an Associate Editor of IEEE Transactions on Wireless Communications, and was on the editorial board of IEEE Transactions on Multimedia. She served as the Technical Program Committee (TPC) chairs/co-chairs of many international conferences, including the Tutorial Co-Chair of IEEE INFOCOM 2004, the Technical Program Vice Chair of IEEE Globecom 2005 Symposium on Autonomous Networks, a TPC Co-Chair of IEEE Globecom 2007 General Symposium, and a TPC Co-Chair of IEEE ICC 2010 Next Generation Networks and Internet Symposium. Dr. Liao has received many research awards. Papers she co-authored with her students received the Best Student Paper Award for IEEE ICME 2000, and the Best Paper Award for ICCCAS 2002. Dr. Liao was the recipient of K. T. Li Young Researcher Award honored by ACM in 2003, and the recipient of Distinguished Research Award from National Science Council in Taiwan in 2006. She is a Senior member of IEEE.

Tehuang Liu received his BS and Ph.D. degrees in Electrical Engineering from National Taiwan University, Taiwan in 2004 and 2008, respectively. His research interests include performance analysis of multi-channel wireless mesh networks, routing protocols for multi-channel wireless mesh networks, and radio resource management in WiMAX networks. He is currently an engineer in MediaTek, Inc, Taiwan.

Iam Kin Chan , a member of the Phi Tau Phi scholastic honor society, was born in Macau. He received his BS degree in Electronics Engineering from National Chiao Tung University, Hsinchu, Taiwan, and MS degree in Communication Engineering from National Taiwan University, Taipei, Taiwan, in 2006 and 2008 respectively. He is now an engineer in Mediatek Inc., Taiwan. His research interests focus on wireless communication and networking.

Hsi-Lu Chao received her PhD degree in Electrical Engineering from National Taiwan University, Taiwan, in 2004. She is now an assistant professor of the Computer Science Department at National Chiao Tung University. Her recent research interests focus on power control for mobile ad hoc networks and WiMAX networks, quality of service over wireless links, relay selection in cooperative networks, and channel sensing in cognitive radio networks. She served as a technical program committee for many IEEE conferences and as a reviewer for many IEEE journals. She is members of IEEE and ACM.