Design of QoS and Admission Control for VoIP Services Over IEEE 802.11e WLANs[*]

PEI-YEH WU[1], JEN-JEE CHEN[1], YU-CHEE TSENG[1,2] AND HUNG-WEI LEE[3]
[1]*Department of Computer Science*
National Chiao Tung University
Hsinchu, 300 Taiwan
E-mails: {phwu; chencz; yctseng}@csie.nctu.edu.tw
[2]*Department of Information and Computer Engineering*
Chung Yuan Christian University
Chungli, 320 Taiwan
[3]*ZyXEL Communications Corp.*
Hsinchu, 300 Taiwan
E-mail: hewitt.lee@zyxel.com.tw

Supporting telephone services using wireless LAN as the access network is an important application nowadays. The SIP and IEEE 802.11e are the two most promising protocols to support such services. In this paper, we show how to integrate SIP and 802.11e to conduct call admission control and resource reservation to support VoIP's QoS in IEEE 802.11e WLANs. Besides, we also suggest some adjustments and MAC enhancements to 802.11e to facilitate VoIP traffics over WLANs.

Keywords: call admission control, IEEE 802.11e, quality of service (QoS), voice over IP (VoIP), wireless network

1. INTRODUCTION

Recently, we have seen two major trends in the area of communications. First, IEEE 802.11 WLANs have been widely deployed worldwide. Second, due to the growth of Internet bandwidth, real-time audio and video applications have become more mature and popular. The combined effect has made VoIP (voice over IP) over WLANs possible. For example, to support VoIP, new products have appeared, such as Wi-Fi phones and dual-mode cellular-WiFi phones (*e.g.*, Cisco Wireless IP Phone 7920 [2] and Motorola MPx [3]).

The above observation has raised an interesting issue: how do WLANs support QoS (Quality of Service) and CAC (Call Admission Control) for VoIP traffics. The *IEEE 802.11 Task Group E* (*802.11e*) [4] has been formed to expand the current 802.11 MAC protocol to support applications with QoS requirements. In addition, the *Session Initiation Protocol* (*SIP*) [5, 6] has been widely accepted as the signaling protocol for VoIP to handle the setup, modification, and teardown of VoIP sessions.

In this work, we consider the cross-layer protocol design problem to facilitate VoIP

traffics over IEEE 802.11e WLANs. We show how 802.11e can cooperate with VoIP SIP signaling to conduct QoS and CAC over wireless channels. We also propose enhancements to 802.11e MAC part to improve its performance in delivering VoIP traffics.

Several prior works have focused on improving VoIP traffics in WLAN environments. Reference [7] claims that admission control is critical to protect VoIP traffics because resources in a WLAN is limited. A *CUE* (*Channel Utilization Estimation*) is proposed to determine whether to accept a new call. Alternatively, giving priority to VoIP traffics helps improving performance too [8]. References [9, 10] point out that the bottleneck is at the access point (*i.e.*, down link traffic). The *BC-PQ* (*Backoff Control and Priority Queue*) protocol proposed in [9] gives priority to voice traffics over data traffics and assigns zero backoff time to voice packets in access points. It is proposed in [10] to separate real-time and non-real-time packets into two queues, and an AP always processes the real-time queue first whenever it is not empty. The number of concurrent VoIP sessions that can be supported in a WLAN is evaluated in [11, 12]. It is reported that besides the bandwidth limitation in the physical layer, the codec, packetizaion interval, and delay budget may all influence the number of VoIP sessions that can be supported. It is further pointed out that the selection of packetization interval (PI) has more impact than the selection of codec.

On the standardization side, the IEEE 802.11 working group R (802.11r) is currently developing fast roaming mechanisms. Reference [13] proposes a structure to integrate Mobile IP with SIP to assist VoIP mobility, while [14] suggests using ad hoc-assisted handoff to meet the Qos requirement of VoIP during handover. The IEEE 802.11e aims at enhancing its MAC mechanism to support QoS. Several works [15-18] have studied using 802.11e to improve multimedia transmission under WLAN. QoS schedulers based on HCCA of IEEE 802.11e are proposed in [15, 16]. Some works discuss how to ameliorate EDCA in IEEE 802.11e to facilitate multimedia transmission. Reference [17] extends the basic EDCA by using an adaptive fast backoff mechanism along with a window doubling mechanism at busy period. In [18], it is suggested that in EDCA each mobile station must conduct admission control on real-time traffic streams to protect existing real-time sessions, and that AP must adjust lower-priority Access Categories' contention windows, Arbitration Inter-frame Spacing (AIFS), and so forth, to protect real-time sessions from being collided by those that do not require admission control.

From the above reviews, it can be seen that cross-layer design between VoIP and MAC protocols is needed to pave the gap and maintain the QoS of VoIP services. In this paper, we show how to support VoIP services over WLAN environments. In particular, we show how to integrate IEEE 802.11e with SIP to conduct call admission control over WLAN environments and support QoS for VoIP calls. Moreover, we also show how to increase the number of VoIP sessions being supported under an AP with compromising QoS and how to improve the MAC mechanism of IEEE 802.11e to facilitate the transmission of VoIP traffics.

The rest of this paper is organized as follows. Background and related work are given in section 2. The proposed QoS architecture for VoIP service is introduced in section 3. Section 4 presents several MAC enhancements to IEEE 802.11e for VoIP services and section 5 gives our simulation results. Finally, section 6 draws our conclusions.

2. BACKGROUND AND RELATED WORK

2.1 802.11e MAC Protocol

The current IEEE 802.11 MAC has no means of differentiating *TSs* (*traffic streams*) or sources. All packets are treated equally in both DCF and PCF. As a result, no consideration can be made for the service requirements of traffics. The IEEE 802.11 Working Group E has proposed a *HCF* (*Hybrid Coordination Function*) for both ad-hoc and infrastructure modes. Several enhancements are introduced in 802.11e. First, a concept called *TXOP* (*Transmission Opportunity*) is introduced, which is a period of time during which a QSTA (a station that supports 802.11e) can exclusively use the wireless medium. A TXOP is defined by a starting time and a maximum duration and it can be obtained by contention or by assignment from the HC (Hybrid Coordinator). Second, IEEE 802.11e supports traffic differentiation by giving traffic streams priorities. Third, it allows a TS to specify its traffic characteristic.

HCF supports two access methods, a contention-based mechanism called *Enhanced Distributed Channel Access* (*EDCA*) and a contention-free mechanism called *HCF Controlled Channel Access* (*HCCA*). Since HCCA is enhanced from PCF and PCF is seldom implemented, we only discuss EDCA in the following.

EDCA of IEEE 802.11e

To differentiate services, IEEE 802.11e adopts the eight user priorities in 802.1D and maps them to four Access Categories (ACs) (refer to Fig. 1). EDCA supports these ACs by four separated queues in both QAP (an AP that supports 802.11e) and QSTA, as illustrated in Fig. 2. Each queue operates as an independent entity and conducts backoff as in the original IEEE 802.11 DCF. The ith AC, $i = 0, \ldots, 3$, has its own arbitration interframe space ($AIFS[i]$), initial window size ($CWmin[i]$), and maximum contention window size ($CWmax[i]$). If multiple queues finish their backoff simultaneously, the *virtual collision handler* will choose the AC with the highest priority to send and the lower priority AC(s) will back off as experiencing an external collision.

The EDCA_Parameter_Set information element (Fig. 3) can be sent in beacon frames. It also contains the TXOP limit of each AC, which bounds the amount of burst transmission of a QSTA after it successfully contends the medium. If TXOP limit equals zero, a QSTA can transmit only one packet each time it gains the TXOP.

802.1D Priority	802.1D Designation	802.11e AC Index	802.11e AC Designation	Comment
1 (low)	BK	01	AC_BK	Background
2	-		AC_BK	Background
0	BE	00	AC_BE	Best Effort
3	EE		AC_BE	Best Effort
4	CL	10	AC_VI	Video
5	VI		AC_VI	Video
6	VO	11	AC_VO	Voice
7 (high)	NC		AC_VO	Voice

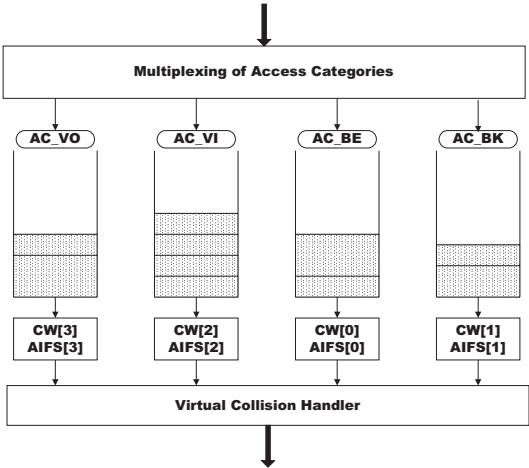Fig. 1. The mappings of 802.1D priorities to IEEE 802.11e ACs.
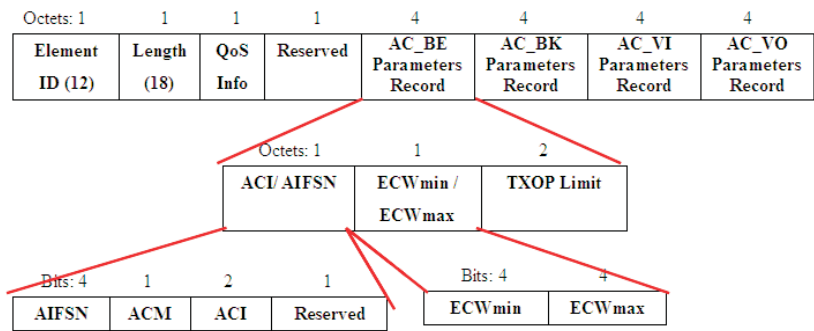
Fig. 2. Management of access categories in EDCA.

Fig. 3. Structure of the IEEE 802.11e EDCA_Parameter_Set information element.

Fig. 4. TSPEC information element of IEEE 802.11e.

Admission Control in EDCA

A QAP uses the *ACM* (*admission control mandatory*) subfield advertised in EDCA_Parameter_Set to indicate whether admission control is required for each AC. A QSTA can send an *ADDTS* (*add traffic stream*) request to the QAP to request adding a new traffic stream by specifying its direction (uplink, downlink, bidirectional, or direct) and providing a *TSPEC* (*traffic specification*) information element as shown in Fig. 4. Some important fields in TSPEC are discussed below:

- Minimum Data Rate: the lowest data rate (in bits per second) to transport MSDUs.
- Mean Data Rate: the average data rate (in bits per second) to transport MSDUs.
- Peak Data Rate: the maximum allowable data rate (in bits per second) to transport MSDUs.
- Minimum PHY Rate: the desired minimum physical rate for this traffic stream.
- Medium Time: the amount of time admitted to a stream to access the medium. This field is not used in the ADDTS request frame, but will be set by the HC in the ADDTS response frame.

On receiving an ADDTS request, the QAP may decide to accept or reject it. In the former case, the QAP will calculate a *MT* (*Medium Time*) for this traffic stream per beacon interval and reply an ADDTS response containing this information; otherwise, an ADDTS response including rejection information is replied. For QAP and each QSTA, they keep the total MT and consumed MT of each AC. Only when the former is lager than the latter, can packets in the corresponding AC be transmitted. After each beacon interval, the consumed MT will be reset to zero. The QAP can identify a traffic stream by its TSID and direction. This information is available in the TS_info field in TSPEC. In this paper, we will use bidirectional reservation for VoIP sessions.

2.2 SIP and SDP

SIP is a signaling protocol, which is considered as an attractive alternative to H.323 to support VoIP. SIP is an application-layer control protocol that can establish, modify, and terminate multimedia sessions. It often cooperates with other protocols, such as *SDP* (*Session Description Protocol*)[1] to describe session characteristics and *RTP* (*Real-time Transport Protocol*)[2] to send traffic after call setup.
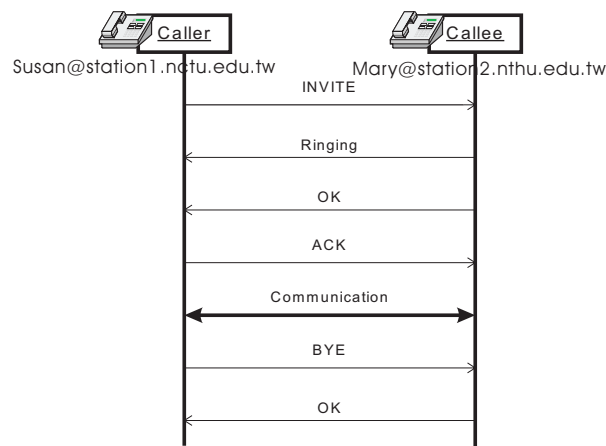
Fig. 5. An example of SIP call setup and tear-down.

---

[1] SDP is specified in RFC 2327. It does not provide a means for transporting or advertising. RFC 3264 describes how SDP co-works with SIP.

[2] RTP is often accompanied with *RTCP* (*RTP Control Protocol*) to provide transport services to support real-time applications.

```
INVITE sip:Mary@station2.nthu.edu.tw SIP/2.0
From: Caller<sip:Susan@station1.nctu.edu.tw>; tag = abc123
To: Callee<sip:Mary@station2.nthu.edu.tw>
CSeq: 1 INVITE
Content-Type: application/sdp
Content-Disposition: session

v = 0
o = Susan 123 001 IN IP4 station1.nctu.edu.tw
s =
c = IN IP4 station1.nctu.edu.tw
t = 0 0
m = audio 123 RTP/AVP 2 4 15
a = rtpmap 2 G726-32/8000
a = rtpmap 4 G723/8000
a = rtpmap 15 G728/8000
```

```
SIP/2.0 200 OK
From: Caller<sip:Susan@station1.nctu.edu.tw;> tag = abc123

To: Callee<sip:Mary@station2.nthu.edu.tw>
CSeq: 1 INVITE
Content-Type: application/sdp
Content-Disposition: session

v = 0
o = collee 456 001 IN IP4 station2.nctu.edu.tw
s =
c = IN IP4 station2.nctu.edu.tw
t = 0 0
m = audio 0 RTP/AVP 2
m = audio 0 RTP/AVP 4
m = audio 888 RTP/AVP 15
```

(a) INVITE signal.   (b) OK signal.

Fig. 6. An example of SIP with SDP message bodies.

SIP is designed to keep signaling as simple as possible. Fig. 5 shows one example of call establishment. When a caller wants to make a VoIP connection with a callee, it sends an **INVITE** including the codecs that the caller supports in a SDP message body. Fig. 6 (a) is an example, with G.726(2), G.723(4), and G.728(15) as the selections (numbers in parentheses are payload types) and 123 the receive port. If the callee decides to accept the request, it replies a Ringing and an OK signals to the caller. The OK signal will contain the callee's choice of codec. In Fig. 6 (b), the callee chooses G.728(15), using the receive port of 888. A port number of 0 indicates a rejection.

2.3 Related Work

To decide whether or not a new call should be admitted into a wireless LAN, some studies have been reported. In [11, 12], the voice capacity is evaluated. If the wireless networks support only voice traffic, the obtained results can be applied for admission control directly. However, when the WLAN supports heterogeneous voice and data traffic, [11, 12] still can not provide a good answer for CAC. Moreover, the capacity is derived by assuming all calls use the same codec, PI, and transmission rate. This is not applicable to the real world. In [7], a measurement based CAC scheme is proposed, where the channel utilization estimation (CUE) is used to perform admission control. CUE is the required fraction of time per time unit for a call or data flow. When a new call arrives, if its CUE plus the total amount of current CUE is less then the maximum CUE, the new call is admitted; if not but the data traffic inside the WLAN can be curtailed, the restrictions for data traffic are calculated and enforced, and then the new call is admitted; otherwise, the new call is blocked. However, voice capacity will be underestimated when the WLAN supports the IEEE 802.11e standard, where the voice traffic has a higher priority than the data by using smaller AIFS and initial/maximum contention window sizes. Furthermore, the proposed scheme does not take mobility into account. There are also model-based admission control schemes [19-21] proposed for CAC. They introduce analytical models for EDCA to evaluate the expected bandwidth, packet delay, collision probabilities, and voice capacity. However, they all consider a saturated condition, where

each station always has packets to transmit. In fact, this is not always true since we often experience non-saturated conditions.

3. THE PROPOSED QOS ARCHITECTURE FOR VOIP SERVICES

We consider an IEEE 802.11e wireless network operating in the infrastructure mode to support VoIP applications. We adopt SIP for call setup and management. We also assume that a VoIP session can dynamically adjust its *packetization interval* (*PI*) even during communication, where PI represents how frequently voice data should be encapsulated into packets. Our purpose is to guarantee high QoS for admitted VoIP sessions when the network load is not heavy, and to support as many VoIP connections with acceptable QoS as possible when the network load is heavy. RFC 3312 (Integration of Resource Management and SIP) discusses how QoS can be made a precondition for sessions initiated by SIP. These preconditions require that participants reserve network resources before continuing. Inspired by this, we propose architecture for IEEE 802.11e to incorporate with SIP to conduct resource reservation and admission control.

3.1 Call Establishment

Fig. 7 shows the proposed QoS architecture after integrating SIP with IEEE 802.11e. When a caller under QAP1 wants to establish a VoIP connection with a callee at QAP2, it can send an INVITE signal with a SDP message containing necessary codec information to the callee. QAP1 and QAP2, on receiving this INVITE signal (refer to boxes A and B in Fig. 7), will do pre-resource reservation and possibly filter out some codecs that they cannot support due to bandwidth constraints. When the callee receives this INVITE signal (refer to boxes C and D), it will exchange 802.11e ADDTS request and response with QAP2. These steps can prevent ghost rings.[3] After exchanging ADDTS messages, the
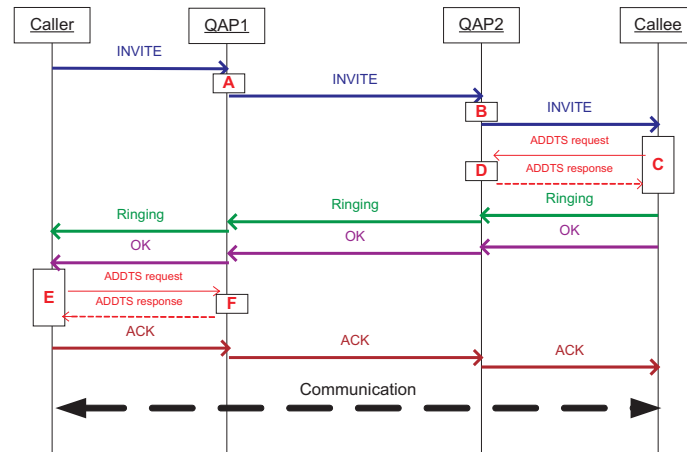

Fig. 7. The proposed QoS architecture for SIP call establishment in 802.11e networks.

---

[3] A ghost happens when a user can not communicate with the other side as he/she picks up a ringing phone. The shortage of bandwidth is often a reason for ghost rings in VoIP applications.

callee can send Ringing and OK signals to the caller. The OK signal will contain the codec being selected by the callee. After receiving the OK signal, the caller will exchange ADDTS request and response with QAP1 (refer to boxes E and F). If these steps successfully go through, an ACK signal will be replied to the callee. In the following, we will explain the detail actions to be taken in boxes A, B, C, D, E, and F.

A. Pre-resource Reservation at the Caller

A QAP has to broadcast the PHY rates that it can support in its beacon frames. When a QSTA is associated with a QAP, it also registers with the QAP its supported rates. In IEEE 802.11e, a QSTA can specify its minimum PHY rate when adding a new traffic stream. When the QSTA can transmit/receive at this rate, the requested QoS should be guaranteed; otherwise, the requested QoS is not necessarily guaranteed. To conduct pre-resource reservation, we propose that each QAP keeps a *Packet Size Table* (*PST*) as in Fig. 8, which contains the packet sizes when different codecs and packetization intervals (PI) are used. For example, in G.726 with a sampling rate of 32 kbps, if a packetization interval of 20 ms is used, then each packet is of size 154 bytes (which contains 80 bytes of voice payload, 40 bytes of IPv4/UDP/RTP/error-checking overhead, and 34 bytes of MAC/error-checking overhead). The payload sizes generated by different codecs can be inferred from [6]. Note that the calculation does not include the PLCP preamble and header, which are 24 bytes and must be sent at the lowest rate of 1 Mbps. Therefore, given a codec and its packetization information, QAP1 can compute a *medium time* (*MT*) that should be reserved for the traffic stream per beacon interval (BI):

$$
\begin{aligned}
MT &= \text{(total time needed per BI)} \\
&= \text{(time to send one packet)} \times \text{(no. packets per BI)} \times \text{(surplus bandwidth allowance)} \\
&= [AIFS + \text{(PLCP preamble and header)} + \text{payload} + SIFS + ACK] \times (BI/PI) \\
&\quad \times \text{(surplus bandwidth allowance)} \\
&= [AIFS + \text{(packet size)}/(\text{min\_PHY\_rate}) + 2 \times PLCP/Mbps + (ACK/\text{min\_PHY\_rate}) \\
&\quad + SIFS] \times (BI/PI) \times \text{(surplus bandwidth allowance)}. \quad\quad (1)
\end{aligned}
$$

Codec	Data Rate (kbps)	Packitization Interval (ms)				
		5	10	20	30	40
G.711	64	114	154	234	314	394
G.726	16	84	94	114	134	154
	32	94	114	154	194	234
G.728	16	84	94	114	134	154
G.723.1	5.3				94	
	6.3				98	

Fig. 8. The packet size table, which gives the packet sizes (in bytes) when different codecs and packetization intervals are used.

According to 802.11 and 802.11e, AIFS for *AC_VO* is 50μs, SIFS is 10μs, ACK packet is 14 bytes, and PLCP preamble and header are 24 bytes. The surplus bandwidth allowance is a value slightly larger than 1 to take into account the excess time required

for possible contentions and retransmissions (in statistical sense). For example, when BI is 1 sec, min_PHY_rate is 11 Mbps and the value of surplus bandwidth allowance is 1.1, if we use G.726 with 32 kbps and PI of 20 msec, then $MT = [50\mu s + 154/11$ (bytes/Mbps) $+ 2 \times 24$ (bytes/Mbps) $+ 14/11$ (bytes/Mbps) $+ 10\mu s] \times (1000/20) \times 1.1 = 31.14$ ms.

For each codec in the INVITE signal, if its MT exceeds the remaining MT of QAP1, we will remove the codec from the codec list. In case that the remaining resource in QAP1 does not allow it to support any codec, QAP1 can drop the INVITE silently or reply a SIP response to the caller with a status code of 480, which means "temporarily not available". Also note that since voice communications are bi-directional, the AP should reserve $2 \times MT_{max}$, where MT_{max} is the maximum time required by all codecs in the list.

B. Pre-resource Reservation at the Callee

The calculation of MT at the callee when receiving the INVITE signal is similar to the above discussion. QAP2 will also filter out those codecs that it cannot support from the INVITE signal and reserve the maximum required bandwidth. The INVITE signal will then be forwarded to the callee if at least one codec can be supported.

C. ADDTS Request by the Callee

After deciding the codec, the callee can send a bidirectional ADDTS request to QAP2 by including a TSPEC element. We suggest to convey VoIP service requirements with the following fields in TSPEC:

- Minimum_Data_Rate = the acceptable longest PI of the corresponding codec.
- Mean_Data_Rate = the PI selected by the callee.
- Maximum_Data_Rate = the acceptable shortest PI.
- Medium_Time = the codec selected by the callee.

With these information, QAP2 can do CAC and resource reservation as described in the following part D.

D. Call Admission Control in QAP2

According to the callee's ADDTS request and the Packet Size Table, QAP2 can compute the required MT following the Eq. (1). Note that with a bidirectional request, the same MT should be applied to both uplink and downlink directions. In order to conduct CAC, each QAP should keep the following variables:

- TXOPBudget[AC_i]: The remaining bandwidth that can be allocated by AC_i, $i = 0\ldots3$.
- TxAdDn[AC_i][TSID]: The admitted MT for stream TSID of AC_i in the downlink direction.
- TxAdUp[AC_i][TSID]: The admitted MT for stream TSID of AC_i in the uplink direction.
- TxAdDn[AC_i]: This value is set to $\sum_{\forall TSID}$TxAdDn[AC_i][TSID], to record the overall resource allocated to AC_i in the downlink direction.
- TxAdUp[AC_i]: This value is set to $\sum_{\forall TSID}$TxAdUp[AC_i][TSID], to record the overall resource allocated to AC_i in the uplink direction.
- TxUsedDn[AC_i]: The summation of used MT of all downlink streams of AC_i.

Initially, TXOPBudget[AC_i] contains all the bandwidth (in terms of MT) that is reserved for AC_i. Whenever a new stream is added, the corresponding resource is subtract from TXOPBudget[AC_i], and the resource is assigned to TxAdDn[AC_i][TSID] and/or TxAdUp[AC_i][TSID]. Also, each QSTA should keep the following variables:

- TxAdUp[AC_i][TSID]: The admitted MT for stream TSID of AC_i in the uplink direction in this STA.
- TxAdUp[AC_i]: This value is set to $\sum_{\forall TSID}$ TxAdUp[AC_i][TSID], to record the overall resource allocated to AC_i of this STA in the uplink direction.
- TxUsedUp[AC_i]: The summation of used MT of all uplink streams of AC_i in this STA.

Resource reservation at QAP2 is done as follows.

(a) We compute the value of TXOPBudget[AC_i] $- 2 \times MT$, where MT is computed from Eq. (1) based on the information of codec, PI, min_PHY_rate, *etc.*, provided by the TSPEC. If the value is non-negative, there is sufficient resource to support this call and go to (b). If there is no sufficient resource, go to (c).
(b) Set

TXOPBudget[AC_i] = TXOPBudget[AC_i] $- 2 \times MT$;
TxAdDn[AC_i][TSID] = MT;
TxAdUp[AC_i][TSID] = MT;
TxAdDn[AC_i] = TxAdDn[AC_i] + TxAdDn[AC_i][TSID];
TxAdUp[AC_i] = TxAdUp[AC_i] + TxAdUp[AC_i][TSID];

and then go to (d).
(c) QAP2 chooses the next larger PI (if possible) and then go back to (a). If there is no sufficient resource for all acceptable PIs of the callee, QAP2 replies an ADDTS response to the callee with Medium_Time = 0 and the resource reservation steps complete.
(d) QAP2 replies an ADDTS response to the callee with the Mean_Data_Rate = PI and Medium_Time = MT in the TSPEC.

At the callee's side, if an ADDTS response with a positive Medium_Time is received, then the QSTA sets its

TxAdUp[AC_i][TSID] = Medium_Time,
TxAdUp[AC_i] = TxAdUp[AC_i] + TxAdUp[AC_i][TSID],

retrieves the PI in the Mean_Data_Rate field, and passes it to the upper layer VoIP application program. Otherwise, the call is considered rejected. In both cases, the callee should reply a response signal with the proper status code to the caller.

E. ADDTS Request by the Caller

When the caller receives the OK signal with codec information from the callee, it will send an ADDTS request to QAP1. The operations are similar to step C.

F. Call Admission Control in QAP1

The action is similar to step D. If the caller receives a successful ADDTS response, it will send an ACK signal to the callee. Then, the voice communication can be started.

Because of the pre-resource reservation in steps A and B, a lot of potential ghost rings can be avoided. Also, voice quality can be guaranteed because of the CAC in steps D and F. Finally, we remark that although we assume that both the caller and the callee are under WLANs, the above procedure should work well if any side is not under a WLAN.

3.2 Resource Readjustment during Transmission

The above steps are for the setup of new calls. However, during transmissions, a stream may dynamically change its bandwidth requirement. In this subsection, we will introduce the steps to be taken to alleviate such problems.

3.2.1 Estimation of downlink PI by QAPs

We note that the PI selected by a codec is not conveyed via SIP signals to the codec at the other side. Therefore, although the resource reservation mentioned above in the uplink direction (from QSTA to QAP) is accurate, the MT reserved for the downlink direction is only an approximation. To solve this problem for each stream TSID, we require a QAP to observe packets from the other side for several beacon intervals and estimate the actual PI being used. After estimating the actual PI, the QAP should calculate the MT according to Eq. (1) for this stream and then update TxAdDn[AC_VO][TSID] and TxAdDn[AC_VO].

3.2.2 Adjustment for PHY rate change at QSTAs

When a traffic stream finds that its admitted MT is not enough to send all of its packets because its physical rate drops below its specified min_PHY_rate, we suggest that the QSTA can send an update ADDTS request to its QAP with the min_PHY_rate field equal to its current PHY rate or below. The operations are similar to the above steps C and D. The QAP may respond in two ways: to allocate more MT for the stream if it still has more resource available, or to suggest a longer PI to reduce the required MT of the corresponding traffic stream. If the request succeeds, a new MT will be replied; otherwise, the QAP will reply with the stream's original MT. In the latter case, the call may suffer from lower quality.

3.2.3 Mechanisms to support more VoIP sessions

When a WLAN is very congested or when there are more new VoIP calls intending to join the WLAN, we may ask current calls to reduce their resource consumption. On finding such a situation, a QAP can send a beacon frame by carrying such a notification to its QSTAs. A QSTA may respond in two ways:

- The QSTA may change the PI of one of its streams by notifying the corresponding co-dec as well as sending a new ADDTS request to the QAP with a longer PI. The QAP should grant the ADDTS request.
- The QSTA may decide to ask one of its streams to change to a lighter-load codec. This can be achieved by the RE-INVITE or UPDATE signal of SIP.

4. MAC ENHANCEMENTS FOR VOIP TRAFFICS

In this section, we discuss several enhancements to improve the performance of 802.11 medium access to support VoIP traffics.

- Bookkeeping: Some control mechanisms are needed to achieve the CAC mentioned above. Only admitted VoIP sessions can drop packets to the AC_VO queue. Besides, for each AC_i that requires admission control, we must keep the medium time that it can use in TxAdUp[AC_i] and TxAdDn[AC_i] and the amounts of time that have been used in TxUsedUp[AC_i] and TxUsedDn[AC_i] in current beacon interval. This bookkeeping work must be done for every data frame being transmitted. Only when TxUsedDn[AC_i] < TxAdDn[AC_i] (resp., TxUsedUp[AC_i] < TxAdUp[AC_i]) can the corresponding access category AC_i contend for the medium in the QAP (resp., QSTA). Also, the value of TxUsedUp[AC_i] in each station should be reset to zero at the end of each BI.
- Adjusting Access Parameters: When a station finds that its dropping rate is higher than a threshold, it can check the receive signal strength of its current QAP. If the signal quality is poor, it may consider switching to a new QAP of better signal quality. If the signal is good, then the cause might be an unexpected high contention from other AC_is. In this case, the STA may ask the QAP to increase the CW and AIFS of other access categories. Afterward, when the network is not so highly congested, the QAP may de-cide to ask other STAs to return to their original CW and AIFS. This is similar to what is suggested in [18].
- Favoring Downlink: We observe that in many cases the performance bottleneck of the network is at the QAP side, because it is in charge of delivering packets for multiple streams. So, a higher priority should be given to QAP. In our design, we will facilitate its transmission as follows: whenever the QAP gains the channel, it is allowed to allo-cate a TXOP to transmit up to N packets, where N is the number of calls admitted to the QAP. In this way, downlink traffics will not suffer lower priority than uplink traffics.

5. SIMULATION RESULTS

An event-driven simulator is developed to evaluate the performance of the proposed CAC and MAC enhancements. Unless otherwise stated, the following assumptions are made in our simulation. The network contains one QAP and multiple QSTAs under the infrastructure mode. Parameters of IEEE 802.11b are used. We set TXOP limit to zero for four ACs, which means that a QSTA can only transmit one packet in each successful con-tention. The communication channel is assumed to be error-free. No RTS/CTS is used. New calls arrive by the Poisson distribution with a rate of λ. Call holding time is expo-nentially distributed with mean of $1/\mu$. So, the offered network load is defined as $\rho = \lambda/\mu$.

VoIP calls, when being first initiated, are supported by G.726 with 32 kbps with PI equal to 20ms and are classified as AC_VO. For AC_VO traffics, we set CWmin to 7, CWmax to 15, and AIFSN to 2. In QSTAs, each AC has a queue of size 50, which means that at most 50 packets can be buffered. In our simulation, a voice packet will be dropped when it is buffered in a queue for more than 1 second. A packet can be re-transmitted at most 3 times. We use *the average voice packet dropping rate*, P_d, as the metric to measure the performance of the system. The quality of a voice stream is considered acceptable if its $P_d < 2\%$; otherwise, it is considered unacceptable.

5.1 Influence of Admission Control

In this scenario, we want to verify the importance of admission control. Voice calls already admitted to the system can only contend to transmit when TxUsedDn[AC_VO] (resp., TxUsedUp[AC_VO]) is smaller than TxAdDn[AC_VO] (resp., TxAdUp[AC_VO]). QSTAs are assumed to be static and can transmit at the rate of 11 Mbps.

With our CAC scheme, the QAP can accept 16 simultaneously on-going VoIP calls. The rest of the calls will be rejected. On the other hand, without CAC, calls are always accepted to the system. Fig. 9 shows the average voice packet dropping rates with and without CAC. When $\rho \leq 10$, both schemes can support voice calls with an acceptable voice quality ($P_d < 2\%$). Without CAC, the P_d curves increase rapidly when $\rho > 12$ for downlink voice traffics and when $\rho > 18$ for uplink voice traffics. With CAC, because the resource usage is under control, the value of P_d is always less than that without CAC. However, for downlink voice traffics, P_d still exceeds the threshold of 2% when $\rho > 14$. Fig. 9 also indicates that P_d for downlink traffics is always greater than the P_d for uplink traffics. This is because the QAP is responsible for delivering packets for multiple streams, but it can only transmit one packet in each successful contention. Therefore, as the network load increases, the QAP will become the performance bottleneck earlier than QSTAs. This will be resolved by our MAC enhancement mechanisms (refer to section 5.3).
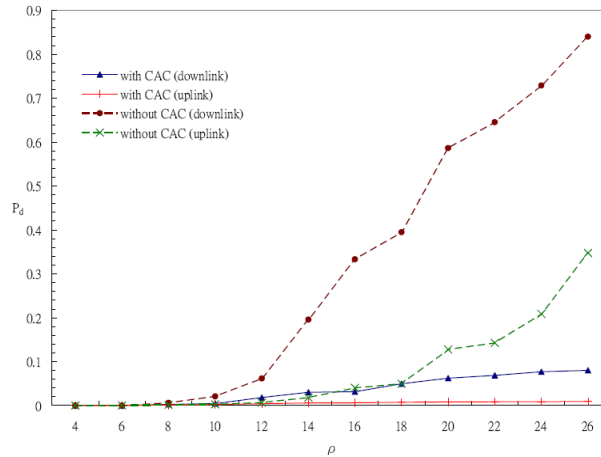

Fig. 9. Impact of CAC: the average voice packet dropping rates (P_d) with and without CAC (surplus = 1.1).
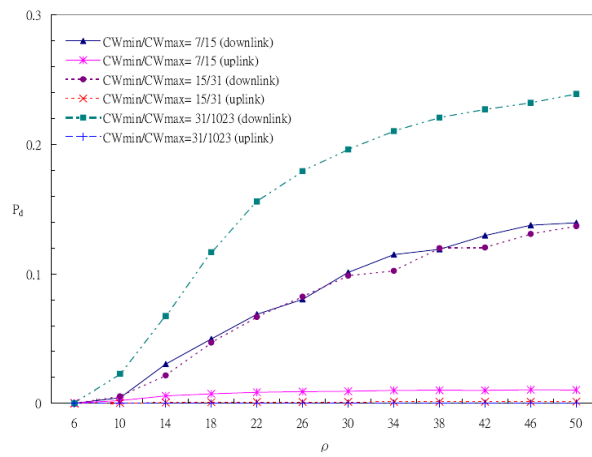
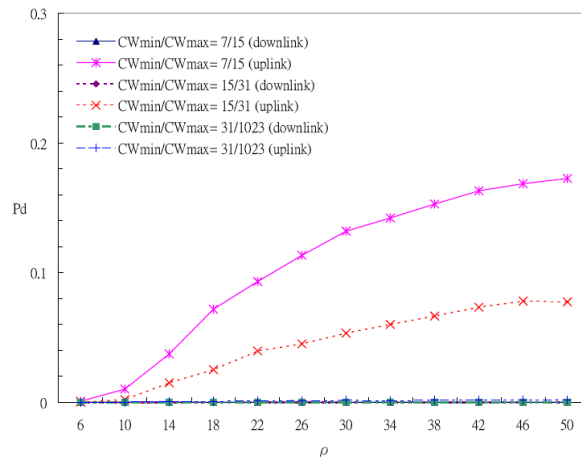Fig. 10. Impact of CWmin and CWmax values (surplus = 1.1, with CAC).

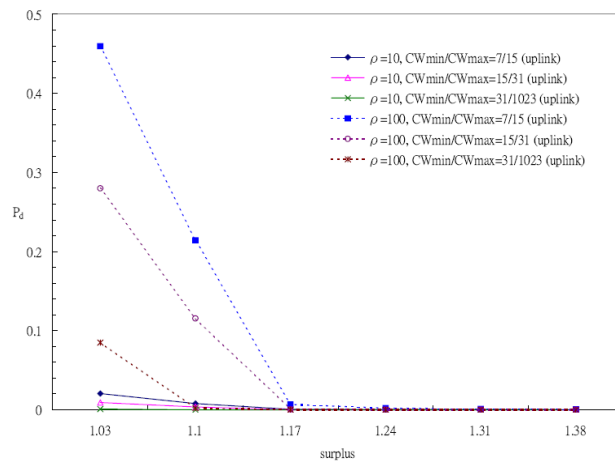Fig. 11. Impact of favoring downlink traffics (surplus = 1.1, with CAC).

Fig. 12. Impact of surplus (CAC is applied and priority is given to QAP).

5.2 Influence of CWmin and CWmax

Fig. 10 plots P_d against ρ with various CWmin and CWmax values. When CWmin and CWmax values are small, it is very likely that multiple QSTAs have the same back-off time. As a result, the collision probability and P_d increase. For uplink voice traffics, we see that the average voice packet dropping rates decrease as CWmin and CWmax increase. For downlink traffics, however, the trend is different from that of uplink traffics. This is because larger *CWmin*/*CWmax* values will increase longer backoff time and thus reduce collision. However, since all downlink packets are buffered at the QAP side waiting for transmission, longer queuing time may be incurred, which implies a higher dropping probability once a packet stays in the QAP more than 1 second. Since collision avoidance and queuing delay are competing factors, we see that the case of *CWmin*/*CWmax* = 15/31 is only slightly better than the case of *CWmin/CWmax* = 7/15, and sharp increase of P_d is incurred in the case of *CWmin/CWmax* = 31/1023. So queuing delay needs to be properly controlled for real-time traffics.

5.3 Influence of Giving Priority to QAP

In this experiment, we verify the effectiveness of favoring downlink traffics (refer to section 4). As Fig. 11 shows, by giving priority to QAP, the downlink packet drop rate is significantly improved even if CWmin/CWmax are set to large values. This is because the QAP can transmit multiple packets in one successful contention. This shows the effectiveness of our MAC enhancement scheme in reducing packet dropping in the downlink direction. From Fig. 11, we see that *CWmin*/*CWmax* = 31/1023 can keep P_d within 2%.

5.4 Influence of Surplus

The value of surplus will affect the amount of extra resource reserved for communication overhead and retransmission. Larger surplus value means less calls being simultaneously admitted. Fig. 12 illustrates P_d against surplus values with different network loads. When ρ is small ($\rho = 10$), the QAP always has enough bandwidth for new calls. Therefore, P_d is not affected much by the surplus value. When ρ is large ($\rho = 100$), for both *CWmin*/*CWmax* = 7/15 and *CWmin*/*CWmax* = 15/31 cases, $P_d < 2\%$ when *surplus* ≥ 1.17; for *CWmin*/*CWmax* = 31/1023, $P_d < 2\%$ when *surplus* ≥ 1.1. This implies that smaller *CWmin*/*CWmax* value needs more extra time being reserved for packet collisions and retransmissions. This gives a guideline in choosing a proper surplus value.

5.5 Networks with Multiple Types of Traffics

The above experiments all assume that VoIP is the only traffic in the network. In this experiment, we add several static QSTAs, each generating best-effort traffic (AC_BE) with a rate of 480kbps to compete with VoIP traffics. We follow the IEEE 802.11e by setting *CWmin*/*CWmax* = 31/1023 and AIFSN = 3 for AC_BE traffics, but varying *CWmin*/*CWmax* of AC_VO. As can be seen from Fig. 13, for both small and large ρ, P_d quickly exceeds the threshold of 2% as the number of AC_BE flows increases. For *CWmin*/
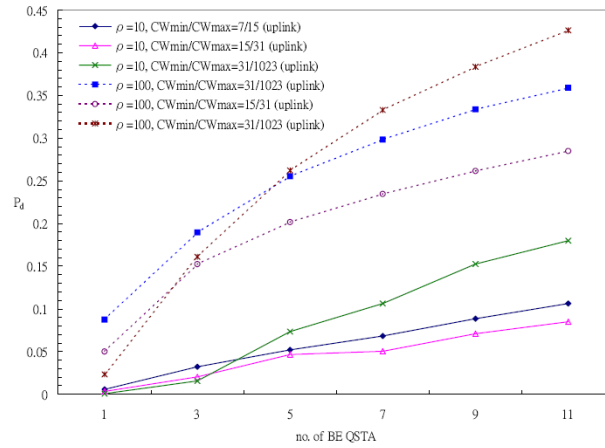
Fig. 13. Impact of interference from AC_BE traffics (CAC is applied, surplus = 1.17, and priority is given to QAP).
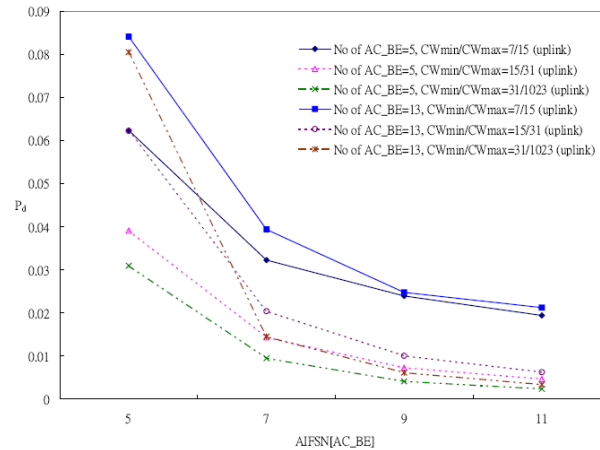

Fig. 14. Impact of interference from AC_BE traffics with various AC_BE's AIFSN values (CAC is applied, AIFSN[AC_VO] = 2, surplus = 1.17, ρ = 100, and priority is given to QAP).

CWmax = 31/1023 curves, their P_d values quickly exceed those of both *CWmin*/*CWmax* = 7/15 and 15/31 cases. This is because when *CWmin*/*CWmax* = 31/1023, AC_BE traffics have almost the same priority as AC_VO traffics. So more voice packets are dropped.

The above effect can be reduced by enlarging the value of AIFSN. As can be observed in Fig. 14, with a slightly larger AIFSN for AC_BE, the dropping rates for voice packets are significantly reduced. So the impact from AC_BE can be alleviated, which implies that the values of AIFSN can effectively differentiate the priorities of voice and best-effort packets.

5.6 Influence of Host Mobility

Host mobility will affect the transmission rate of stations. In this scenario, we as-

sume that QSTAs will move within the transmission range of the QAP and their transmission rates will change every second according to Fig. 15. For example, state 5.5Mbits/s has a probability γ to transit to state 2Mbits/s and a probability γ to transit to state 11Mbits/s. Our goal is to observe the impact of host mobility and how our scheme can adapt to it.

Without any enhancement, Fig. 16 shows that host mobility does significantly decrease performance. When the network load is high ($\rho = 100$), the system bandwidth is almost fully utilized and no free bandwidth is available when calls' transmission rates decrease. In this case, the performance degrades when γ increases.

In the mobility enhancement (denoted by MBen in the figure), we allow a call to relax its PI up to 100ms. Fig. 16 indicates that the impact of host mobility is effectively alleviated by using MBen. Even in a highly mobility case ($\gamma = 0.5$, which means a QSTA changes its transmission rate every second in average), P_d is still far below 2%.

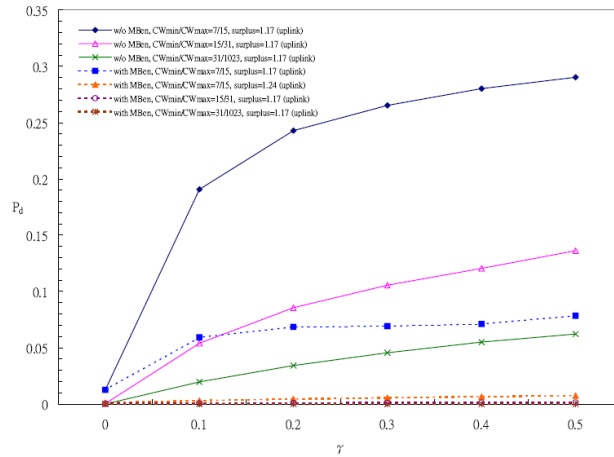Fig. 15. The state transition diagram of QSTAs' transmission rates.

Fig. 16. Impact of host mobility with or without mobility enhancement (CAC is applied, $\rho = 100$, and priority is given to QAP).

6. CONCLUSIONS

In this paper, we have described several schemes to enhance the performance of VoIP calls by integrating the SIP and 802.11e. IEEE 802.11e is expected to be a well accepted standard, so the cross-layer design proposed in this paper is an essential issue. In addition, we present several MAC enhancements to facilitate VoIP traffics under WLANs. By simulation, we investigate the performance of our CAC scheme with multiple factors.

To guarantee the quality of voice calls, a $P_d < 2\%$ condition should be held. Our study shows that

- Combined with giving priority to QAP, our CAC scheme can perform well and conform to $P_d < 2\%$ constraint for voice calls.
- To differentiate priorities for different types of traffic, IEEE 802.11e assigns small AIFSN and CWmin/CWmax values to AC_VO. However, this may result in more collisions and thus waste wireless resources. Simulation results show that using larger CWmin/CWmax values can prevent collisions and get better performance.
- Although AC_BE traffics have lower priority, they may significantly affect the performance of real-time traffics in a mixed traffic environment. By applying a larger AIFSN value to AC_BE, simulation results show that the contention intervals of real- and non-real-time traffics are divided and the quality of real-time traffics can be ensured.
- Host mobility is another important factor that critically decreases performance. Our mobility enhancement can reduce its impact.

For future work, it deserves to investigate the handoff problem where QoS is a concern.

REFERENCES

1. P. Y. Wu, Y. C. Tseng, and H. Lee, "Design of QoS and admission control for VoIP traffics over IEEE 802.11e WLANs," in *Proceedings of National Computer Symposium*, 2005.
2. Cisco Wireless IP Phone 7920, http://www.cisco.com/en/US/products/hw/phones/ps379/ps5056/index.html, Technical Report.
3. Motorola MPx, http://www.motorola.com/mediacenter/news/detail.jsp?globalObjectId =3869_3244_23, News Release.
4. IEEE Std 802.11e-2005, Part 11: Wireless LAN Medium Access Control (MAC) and Physical layer (PHY) specifications, Amendment 8: Medium Access Control (MAC) Quality of Service (QoS) Enhancements, 2005.
5. J. Rosenberg, H. Schulzrinne, E. Schooler, M. Handley, G. Camarillo, A. Johnston, J. Peterson, and R. Sparks, *SIP: Session Initiation Protocol*, IETF RFC3261, 2002.
6. D. Conllins, *Carrier Grade Voice over IP Second Edition*, McGraw-Hill Companies, Inc., 2003.
7. S. Garg and M. Kappes, "Admission control for VoIP traffic in IEEE 802.11 networks," in *Proceedings of IEEE GLOBECOM*, Vol. 6, 2003, pp. 3514-3518.
8. T. Hiraguri, T. Ichikawa, M. Iizuka, and M. Morikura, "Novel multiple access protocol for voice over IP in wireless LAN," in *Proceedings of the 7th IEEE International Symposium on Computers and Communications*, 2002, pp. 517-523.
9. F. Anjum, M. Elaoud, D. Famolari, A. Ghosh, R. Vaidyannthan, A. Dutta, P. Ageawal, T. Kodama, and Y. Katsube, "Voice performance in WLAN networks − An experimental study," in *Proceedings of IEEE GLOBECOM*, Vol. 6, 2003, pp. 3504-3508.

10. J. Yu, S. Choi, and J. Lee, "Enhancement of VoIP over IEEE 802.11 WLAN via dual queue strategy," in *Proceedings of IEEE International Conference on Communication*, 2004, pp. 3706-3711.
11. S. Garg and M. Kappes, "Can I add a VoIP call?" in *Proceedings of IEEE International Conference on Communications*, Vol. 2, 2003, pp. 779-783.
12. D. P. Hole and F. A. Tobagi, "Capacity of an IEEE 802.11b wireless LAN supporting VoIP," in *Proceedings of IEEE International Conference on Communications*, Vol. 1, 2004, pp. 196-201.
13. J. W. Jung, R. Mudumbai, D. Montgomery, and H. K. Kahng, "Performance evaluation of two layered mobility management using mobile IP and session initiation protocol," in *Proceedings of IEEE GLOBECOM*, Vol. 3, 2003, pp. 1190-1194.
14. T. D. Todd, M. He, D. Zhao, and V. Kezys, "Ad hoc assisted handoff for real-time voice in IEEE 802.11 infrastructure WLANs," in *Proceedings of IEEE Wireless Communications and Networking Conference*, Vol. 1, 2004, pp. 201-206.
15. A. Grilo, M. Macedo, and M. Nunes, "A scheduling algorithm for QoS support in IEEE 802.11e networks," *IEEE Wireless Communications*, Vol. 10, 2004, pp. 36-43.
16. L. W. Lim, R. Malik, P. Y. Tan, C. Apichaichalermwongse, K. Ando, and Y. Harada, "A QoS scheduler for IEEE 802.11e WLANs," in *Proceedings of IEEE Consumer Communications and Networking Conference*, 2004, pp. 199-204.
17. M. Malli, Q. Ni, T. Turletti, and C. Barakat, "Adaptive fair channel allocation for QoS enhancement in IEEE 802.11 wireless LANs," in *Proceedings of IEEE International Conference on Communications*, Vol. 6, 2004, pp. 3470-3475.
18. Y. Xiao, H. Li, and S. Choi, "Protection and guarantee for voice and video traffic in IEEE802.11e wireless LANs," in *Proceedings of IEEE INFOCOM*, Vol. 3, 2004, pp. 2152-2162.
19. D. Pong and T. Moors, "Call admission control for IEEE 802.11 contention access mechanism," in *Proceedings of IEEE GLOBECOM*, Vol. 1, 2003, pp. 174-178.
20. A. Banchs, X. Perez-Costa, and D. Qiao, "Providing throughput guarantees in IEEE 802.11e wireless LANs," in *Proceedings of the 18th International Teletraffic Conference*, 2003.
21. N. Hegde, A. Proutiere, and J. Roberts, "Evaluating the voice capacity of 802.11 WLAN under distributed control," in *Proceedings of the 14th IEEE Workshop on Local and Metropolitan Area Networks*, 2005.
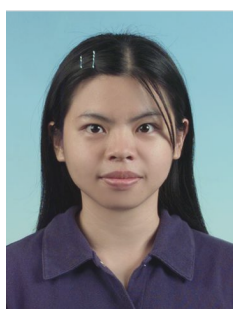
Pei-Yeh Wu (吳佩曄) received the B.S. and M.S. degrees in Computer Science and Information Engineering from National Chiao Tung University in 2003 and 2005, respectively. As a software engineer, she is currently working in Realtek Semiconductor Corp. for the development of the UWB-related products.

Jen-Jee Chen (陳建志) received the B.S. and M.S. degrees in Computer Science and Information Engineering from National Chiao Tung University in 2001 and 2003, respectively. He is currently pursuing his Ph.D. degree in Computer Science at National Chiao Tung University. His research interests are primarily in wireless networks, network performance modeling, and Internet communication service.

Yu-Chee Tseng (曾煜棋) received his B.S. and M.S. degrees in Computer Science from the National Taiwan University and the National Tsing Hua University in 1985 and 1987, respectively. He obtained his Ph.D. in Computer and Information Science from the Ohio State University in January of 1994. He was an Associate Professor at the Chung Hua University (1994-1996) and at the National Central University (1996-1999), and a Professor at the National Central University (1999-2000). Since 2000, he has been a Professor at the Department of Computer Science, National Chiao Tung University, Taiwan, where he is currently the Chairman. Dr. Tseng received the Outstanding Research Award, by National Science Council, R.O.C., in both 2001-2002 and 2003-2005, the Best Paper Award, by International Conference on Parallel Processing, in 2003, the Elite I. T. Award in 2004, and the Distinguished Alumnus Award, by the Ohio State University, in 2005. His research interests include mobile computing, wireless communication, network security, and parallel and distributed computing.

Hung-Wei Lee (李弘威) received his B.S. and M.S. degrees in Computer Science and Information Engineering from the National Chiao Tung University in 2001 and 2003, respectively. His research interests include wireless communication and mobile computing. He is working at ZyXEL Communications Corp. in Hsinchu Science Park, Taiwan as a software engineer in the Network Connectivity product division since 2004.