Efficient Silence Suppression and Call Admission Control through Contention-Free Medium Access for VoIP in WiFi Networks

Irshad A. Qaimkhani and Ekram Hossain, University of Manitoba

ABSTRACT

In view of the rapidly growing trend of migrating customers from traditional wired phones to mobile phones and then to VoIP services in the recent past, there is a tremendous demand for wireless technologies to support VoIP, specially on WiFi technologies which have already matured commercially. This has put forth great research challenges in the area of wireless VoIP. In this article we have addressed two core issues, efficient silence suppression and call admission control, in QoS provisioning for VoIP services in WiFi networks. In this connection we present a QoS-aware wireless MAC protocol called Hybrid Contention-Free Access (H-CFA) and a VoIP call admission control technique called the Traffic Stream Admission Control (TS-AC) algorithm. The H-CFA protocol is based on a novel idea that combines two contention-free wireless medium access approaches, round-robin polling and TDMA-like time slot assignment, and provides substantial multiplexing capacity gain through silence suppression of voice calls. The TS-AC algorithm ensures efficient admission control for consistent delay bound guarantees and further maximizes the capacity through exploiting the voice characteristic so that it can tolerate some level of nonconsecutive packet loss. We expose the benefits of our schemes through numerical results obtained from simulations.

INTRODUCTION

Voice over IP (VoIP) technology suppresses the alternating silence periods of voice calls and increases bandwidth utilization through efficient multiplexing of voice calls. However, it does not fulfill the increasing mobility demands of users. Pervasive commercial deployment of VoIP over wireline networks, and the mobility, flexibility, and scalability provided by WiFi hotspots have attracted great research efforts recently in the area of VoIP in WiFi. Real-time voice traffic requires some parametric type of quality of service (QoS) such as maximum acceptable end-to-end delay bounds and jitters. For example, interactive voice can normally tolerate end-to-end delay up to 25 ms without echo cancellation and 150 ms with it [1]. This stringent parametric QoS requirement can only be fulfilled through end-to-end connection-oriented service. However, voice traffic can tolerate consecutive packet loss and provides the opportunity of multiplexing gains through the suppression of its alternating silence periods.

The legacy WiFi standard, IEEE 802.11a/b/g, does not provide the parametric QoS required by voice traffic. Of its two modes of medium access, distributed coordinated function (DCF) and point coordination function (PCF), DCF uses carrier sense multiple access with collision avoidance (CSMA/CA), which is contention-based and causes nondeterministic medium access delays. PCF mode, which uses a round-robin polling scheduler, is contention-free and provides connection-oriented medium access. However, the round-robin polling scheduler does not suppress the silence periods of voice calls, and the nondeterministic beacon delay in PCF mode becomes a barrier to guaranteeing strict delay bounds. The hybrid coordination function controlled channel access (HCCA) mode of the QoS-enhanced version, IEEE 802.11e, addresses most of the QoS-related problems of PCF. It also provides an outline for call admission control. However, HCCA deploys the same inefficient round-robin scheduler and does not specify any technique for admission control.

In this article we outline the research challenges in provisioning VoIP services over WiFi networks. We provide a comprehensive survey on some of the important research work that has attempted to develop efficient wireless medium access scheduling protocols for silence suppression and call admission control techniques. To this end, we present a novel medium access control (MAC) protocol and an associated call admission control scheme for providing QoS to VoIP services in WiFi systems.

MEDIUM ACCESS CONTROL ARCHITECTURE FOR WIFI AND WIRELESS VOIP CHALLENGES

WIFI MAC ARCHITECTURE

WiFi MAC architecture has two wireless medium access mechanisms. One is contention-based medium access carried out by CSMA/CA as mentioned above. This is incorporated in the DCF mode of IEEE 802.11a/b/g and enhanced DCF (EDCF) of IEEE 802.11e with some modifications for prioritizing services. DCF and EDCF work at the user end. Due to its nondeterministic nature, contention-based medium access architecture is suited only for asynchronous data transmission or best effort service.

The other is contention-free medium access carried out by a round-robin type polling method through the PCF mode of IEEE 802.11a/b/g or HCCA of IEEE 802.11e incorporated at the central access point (AP). These two functions are combined in the hybrid coordination function (HCF). HCF can work concurrently with DCF and PCF for backward compatibility, combining functions from both DCF and PCF with some enhanced QoS-specific mechanisms and frame subtypes.

The contention-based medium access carried out by DCF is the default MAC scheme in WiFi networks. In the basic architecture the contention-free medium access carried out by PCF is optional and works on top of DCF. In the enhanced MAC architecture, the HCF works on top of the DCF. In HCF mode, PCF is still optional for non-QoS contention-free service.

There is another approach for contention-free wireless MAC design: TDMA-like fixed time slot assignment to users under some defined order. In this approach polling overhead is reduced greatly through the broadcast of a single poll. However, the TDMA-like approach becomes bandwidth inefficient when the user is silent, and its time synchronization problem poses complexities in accommodating asynchronous service time requirements of individual admitted services. There is no time synchronization problem in round-robin polling since scheduling is differentiated at the individual user level. Also, round-robin polling is more bandwidth-efficient as the scheduler waits for a comparatively much smaller time (e.g., PCF interframe space [PIFS] in case of 802.11 PCF-mode) if the user is silent.

WIRELESS VOIP CHALLENGES

QoS for wireless transmission can be characterized as either parameterized or prioritized QoS. Parameterized QoS strictly promises some consistent levels of quantitative values as required by voice services such as delay bound, data rate, and jitter, while prioritized QoS is based on relative delivery priority. Accommodating time-sensitive VoIP services that require some minimum level of parameterized QoS is a great challenge.

Since DCF mode is contention-based, it does not promise parameterized or prioritized QoS, and there is no service priority categorized in it. However, since wireless medium access in PCF mode is contention-free, it is deterministic in nature and supports QoS to some extent. But PCF has some inherent design problems as follows:

- It causes bandwidth wastage due to the centralized communication.
- PCF causes unnecessary bandwidth wastage by polling voice users during their periodic silent periods. There is great motivation to suppress these silent periods by somehow not attending the user during its idle periods and multiplexing other traffic with it to reduce bandwidth wastage. The central controller (i.e., AP) can do this job through a polling list management function that should exclude a voice user from the polling list during its silent periods and include it when the user becomes active. Since the AP does not communicate with the user during its silent state, the hardest part of this process is for an AP to know when the silent to talk spurt state change occurs on the user side so that the user can be included again in the polling list.
- Nondeterministic beacon delays are barriers to guaranteeing strict delay bounds. To start CFP, beacons can only be transmitted if the channel becomes idle around the target beacon transmission time (TBTT) for at least a PIFS. If a user has not finished its transmission at TBTT, it will continue its transmission according to the PCF rules, causing delays in beacon transmission.
- Since PCF can only work in the designated CFP of the beacon interval, the size of the beacon interval becomes dependent on the delay bounds required by real-time voice service.
- It is unable to predict the transmission time of a polled user due to the physical layer (PHY) rate variation in time and space and frame size variation between 0 and 2346 bytes.

QoS for wireless transmission can be characterized as either parameterized or prioritized QoS. Parameterized QoS strictly promises some consistent levels of quantitative values as required by voice services such as delay bound, data rate, and jitter, while prioritized QoS is based on relative delivery priority.

RESEARCH ADVANCEMENTS FOR WIRELESS VOIP: CURRENT STATE OF THE ART

We categorize and summarize the selected research work as follows: one that addresses the priority and/or fairness type of QoS issues; second are the parametric and priority type of QoS enhancements within the WiFi standard; third, in connection to silence suppression in the voice traffic and call admission control, those that address parametric and/or priority type of QoS issues.

PRIORITY AND FAIRNESS TYPES OF QOS-BASED SCHEMES

The Priority Effort-Limited Fair (priority-ELF) scheme proposed in [2] checks the priority of users defined as type of service (TOS) combined with effort-limited fair scheduling. This scheme deters the point coordinator (PC) from polling in an error-prone wireless link, and also provides prioritized QoS similar to HCF that IEEE 802.11e provides.

The Distributed Deficit Round-Robin (DDRR) scheme in [3] proposes a fairness-based polling scheme that decides whether or not to poll a user based on the positive or negative value of a deficit counter (DC) defined in the PCF for that user.

There are two ways in which EDCA can work. One is by using different but comparatively small Inter-frame Spacing called Arbitration IFS instead of DIFS for different ACs, and the second is by using different contention window sizes for different ACs according to the priority.

These schemes do not propose any measure to guarantee the bounded end-to-end delays and jitters required for voice traffic, especially under high load situations.

Parametric and Priority-QoS Provisioning within the IEEE 802.11 MAC Standard

HCF in the IEEE 802.11e MAC standard [4] proposed QoS-enhancement methods within the WiFi standard to support all types of services (real-time and non-real-time) through its sub-functions EDCA and HCCA, respectively. EDCA is a contention-based mechanism that proposes prioritized QoS for four queue-based differentiated services called access categories (ACs). HCCA is a contention-free mechanism that proposes parameterized QoS to eight queue-based differentiated services called traffic streams (TSs). HCF allots transmission opportunity (TXOP) time units to users in both EDCA (EDCA-TXOP) and HCCA (polled-TXOP) modes according to AC priority or TS requirement. In this way, for bursty voice traffic, frame bursting and block acknowledgment are possible for efficient use of bandwidth.

There are two different ways EDCA can work. One is using different but comparatively small interframe spacing (IFS) called arbitration IFS (AIFS) instead of DIFS for different ACs; the second is using different contention window (CW) sizes for different ACs according to priority. All these variables such as AIFS, minimum and maximum CWs, and maximum value of TXOP (TXOPLimit) are determined by the AP and announced in the beacon frame. But the standard does not specify how the AP determines these values to fulfill the specific QoS requirement.

HCCA addresses most of the QoS-related problems of PCF. It guarantees parametric QoS through setting up virtual connections called traffic stream (TSs) following a call admission methodology of negotiating the traffic specifications (TSPEC). It can start at any time, whether in the optional CFP or in the CP, by initiating a controlled access phase (CAP) that does not restrict the size of the CFP according to the delay bound requirements. However, HCCA polls the TXOPs in the same round-robin fashion as the PCF does irrespective of whether or not the polled user has data to send.

Schemes that Address Silence Suppression Issues in Voice Services

The uplink notification scheme in [5] proposed a contention-based uplink status notification mechanism that is quite different from WiFi DCF mode. This scheme quantified some parameters such as round-trip delays and number of active users in comparison to polling under a general tree-based topology.

The Simultaneous Transmit Response Polling (STRP) scheme proposed in [6] exploits the capture phenomenon to report the uplink status by sending a weak jamming signal to the AP in response to a special control frame sent by the AP and succeeds theoretically in reducing silence periods. However, the capture effect may not be effective in a harsh fading environment.

The two schemes described above are examples of round-robin polling. In the following we describe several schemes based on TDMA-like time slot assignment.

Deterministic Access Priority of Voice during contention period (CP) as proposed in [7] suggests a modified service interval structure of HCCA with a TDMA-like time slot assignment scheme for MAC scheduling. The AP broadcasts the single super CF-poll frame to all pollable users notifying them of their respective turns. However, for uplink status notification, it suggests contention-based prioritized access through a dedicated time zone that is wasted if there is no reconnection request.

The Isochronous Coordination Function (ICF) scheme proposed in [8] is similar to the CAP cycle in the HCCA mode of IEEE 802.11e. But unlike in CAP, ICF suggests TDMA-like time slot assignment that is broadcast through a super ICF-poll frame at the start of each ICF cycle, thereby reducing the polling overhead significantly. However, ICF proposes contention-based EDCA mode for uplink status notification.

Both schemes described above suffer time synchronization problems for active users with their respective TDMA time slots.

All the above schemes suggest silence suppression, but one way or the other, they do not provide explicit contention-free medium access, the prime requirement of voice traffic. Following are a couple of schemes that have attempted to do this job.

The Cyclic Shift and Station Removal (CSSR) polling scheme proposed in [9] reduces the number of polls to voice users by not polling them for a fixed threshold time cycle during their silent periods.

The adaptive polling MAC scheme proposed in [10] suggests a talk spurt detection algorithm for silence suppression in PCF mode. As opposed to the scheme proposed in [9], this scheme divides this average idle period into decreasing value threshold time intervals. Both of the above schemes provide silence suppression, but at the cost of increased waiting time delays to the first talk spurt frames, which are not sent due to the imposition of the threshold time period.

Schemes that Address Admission Control Issues for Voice Services

The connection admission control (CAC) scheme in association with a signaling protocol proposed in [11] quantifies the maximum capacity and delay jitter in relation to the interpoll periods (super frame time) in the PCF mode and maximum delay bound requirements of the voice calls. It suggests a call admission control strategy based on the capacity of PCF mode under specified conditions. However, this scheme suggests contention-based DCF mode for signaling the arrival of the first talk spurt packet to the AP.

The measurement-based CAC strategy introduced in [12] is based on channel occupancy analysis. The channel occupancy rate, which is the ratio of the channel busy time to an observation time, is constantly measured and matched to a predetermined threshold value to learn the traffic congestion level and admit new users accordingly. But this scheme only considers contention-based EDCA mode, which is not suited for high traffic load situations.

A qualitative comparison among the different QoS-aware MAC schemes is shown in Table 1.

A Novel Hybrid Contention-Free Access MAC Protocol and a Traffic Stream Admission Control Mechanism

Our objective is to develop a smart wireless MAC protocol that can guarantee the required performance in terms of capacity enhancement, end-to-end delay bounds, and consecutive packet loss control. The schemes discussed above have specific characteristics, but none of them serves our purpose completely. While developing the Hybrid Contention-Free Access (H-CFA) protocol and a measurement-based TS admission control (TS-AC) scheme, we put forth the following targets:

- Enhancement of the capacity of APs for large-scale commercial deployment
- Capacity enhancement without incurring any cost such as increased waiting time delay for the first MAC frame(s) that arrives when the silence to talk spurt state change occurs
- Guaranteed delay bound service
- No consecutive packet loss

Protocol Overview

H-CFA defines a hybrid contention-free interval (H-CFI) similar to the CAP in HCCA or the CFI in PCF. For guaranteed delay bound service, a call admission control (CAC) scheme is developed that maximizes the capacity gain through exploiting the characteristic of bursty voice traffic that can tolerate nonconsecutive packet loss to some extent. In H-CFA the Dynamic Polling List Management (D-PLM) algorithm incorporated in the central controller completely eliminates the wastage of bandwidth caused by the round-robin polling scheduler that polls all admitted users whether in silent or talk state. D-PLM function manages three logical lists. The first is the admitted service list (ASL) of the association IDs (BSSID) of all admitted users (pollable users) under the traffic stream (TS) setup mechanism in [4] on a first-come first-served basis and under a CAC mechanism defined below. The second is the polling list (PL) of the association IDs of admitted users in talking state. The third is the idle service list (ISL) of admitted users in silent state.

The uplink status notification is carried out by a Contention-Free Activity Detection (CF-AD) algorithm that works in the contention-free activity detection interval (CF-ADI), a small portion of the H-CFI at its beginning. The CF-AD uses a TDMA-like time slot assignment scheme for this purpose.

H-CFA Protocol Description

The AP starts the H-CFI as long as there is at least one active user in the PL, in either the optional CFP of periodic beacon intervals or the CP, such as in CAP [4] or ICF [8]. However, the internal architecture and working of H-CFI are quite different, as it comprises three subintervals: CF-ADI, contention-free downlink voice interval (CF-DVI), and contention-free polling interval (CF-PI).

Contention-Free Activity DetectionAlgorithm — The CF-ADI starts simultaneously along with the H-CFI after the channel has been sensed idle for a PIFS. At the start of CF-ADI, the AP broadcasts a single contention-free activity detection poll (CF-AD-Poll) frame with a field vector (FV) that represents the AIDs of all silent users in the ISL in the order of the ISL. This FV enables silent users to notify the AP of their status change in a contention-free manner according to their respective order in the FV. In this way the user in the first position in the FV of the CF-AD-Poll gets the opportunity to respond first. A response from a silent user is also a broadcast response (RB) frame. The silent user responds only in the case of activity arrival; otherwise, it does not respond. The next silent user waits for a SIFS if it hears the RB frame; otherwise, it waits for a PIFS before it uses its own opportunity to respond. The AP moves the user from the ISL to the PL on hearing the RB frame from that user. If two consecutive silent users do not respond, the next user will respond after two PIFSs, and the AP will expect this from that user after two PIFSs. Thus, the length of the CF-ADI (say *AD_Length* in terms of bit times) is a function of the number of activity arrivals that can be calculated as follows.

Let n be the number of idle users in the FV; m out of these n idle users do not broadcast RB frames. We denote the lengths of CF-AD-Poll frame, SIFS, RB frame, and PIFS with *ADPol_Length*, *SIFS_Length*, *RB_Length*, and *PIFS_Length*, respectively, in terms of bit times. According to the CF-AD algorithm as described above,

$$AD\_Length = I_n + (n - m) \times SIFS\_Length + (n - m) \times RB\_Length + m \times PIFS\_Length \quad (1)$$

where $I_n = 0$ for $n = 0$ and $I_n = ADPol\_Length$ for $n \geq 1$.

The following two extreme cases are worth considering when $n \geq 1$.

Case 1: There is no activity arrival (i.e., no user broadcasts an RB frame), which means $m = n$, in which case

$$AD\_Length = ADPol\_Length + n \times PIFS\_Length.$$

Case 2: All n users in the ISL broadcast an RB frame (i.e., $m = 0$), in which case

$$AD\_Length = ADPol\_Length + n \times SIFS\_Length + n \times RB\_Length.$$

The AP starts the contention-free downlink voice interval (CF-DVI) after a SIFS from the end of the CF-ADI, where the AP sends frame(s), separated by SIFS, to their destined users (talking or silent) in FIFO order from its local queue. The contention-free polling interval (CF-PI) starts after a SIFS from the end of the CF-DVI. In the CF-PI the AP sends the CF-Poll

The uplink status notification is carried out by a Contention-Free Activity Detection algorithm that works in the Contention-Free Activity Detection Interval — a small portion of the H-CFI at its beginning. The CF-AD uses a TDMA-like time slot assignment scheme for this purpose.

Schemes that address priority, fairness and silence suppression issues					
	Delay bound guarantees	Silence suppression	Waiting time delays[2]	Consecutive packet loss[4]	Fairness
Priority-ELF [2]	Yes	Good	Large	Large	Very good
DDRR [3]	Yes	Good	Large	Large	Excellent[5]
Round-robin (PCF/HCCA) [4]	Yes	No	No	No	Moderate
Uplink notification scheme [5]	Partial[1]	Good	Moderate[3]	Moderate[3]	Moderate
STRP [6]	Partial[1]	Very good	Small[3]	Small[3]	Moderate
Deterministic access priority [7]	Partial[1]	Excellent	Small[3]	Small[3]	Moderate
ICF [8]	Partial[1]	Excellent	Moderate[3]	No	Good
CSSR [9]	Yes	Very good	Very large	Large	Moderate
Adaptive polling [10]	Yes	Very good	Large	Large	Moderate
H-CFA[8]	Yes	Very good	No	No	Excellent[6]
Schemes that address call admission control issues					
	Delay bound guarantees	Contention-free Access	Waiting time delays[2]	Capacity gains	Fairness/Priority
CAC [11]	Partial[1]	Partial	Moderate[3]	Moderate[3]	Moderate
WLAN CAC [12]	No	No	Probabilistic	Moderate[3]	Moderate
TS-AC[8]	Yes	Yes	No	Large[7]	Moderate

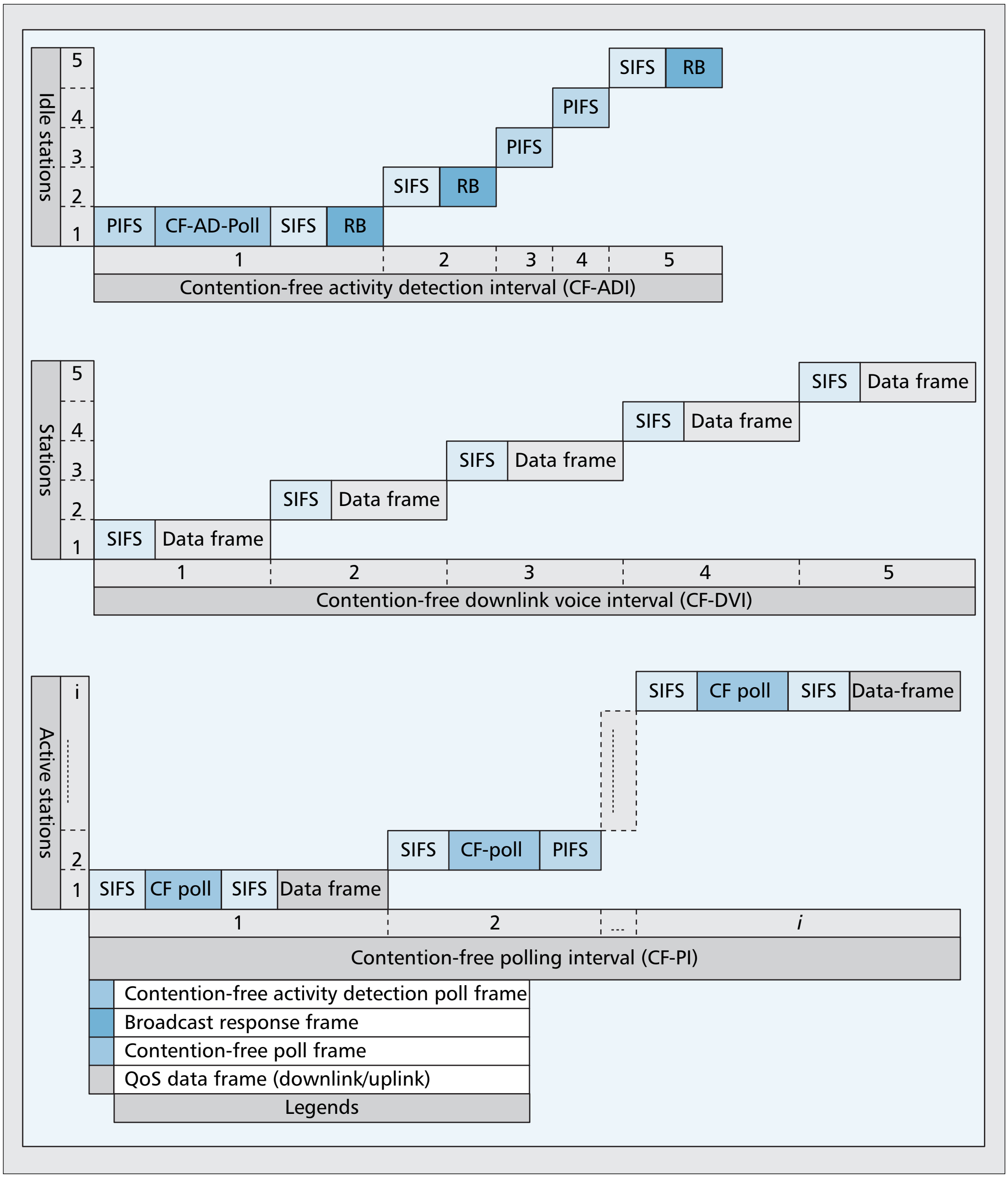
Note:
[1] If number of contending stations is large, delay bound guarantees become increasingly vulnerable.
[2] Unnecessary delays to initial frames when idle-to-active state change arrives.
[3] Number of contending stations is assumed small, and if large, it will lead to un-acceptable for delay-bounds.
[4] It is assumed that the delayed frame is lost.
[5] Fairness in terms of bandwidth allocation.
[6] Fairness according to the traffic arrival behavior at each admitted station.
[7] At the cost of accepted large non-consecutive packet loss.
[8] For H-CFA and TS-AC, see next sections.

Table 1. *Qualitative performance comparison among different QoS-aware MAC schemes.*

frame to allocate the polled-TXOP to active user according to its order in the PL. After a SIFS, the polled user sends its frame to the AP. If the polled user has more than one frame, it may send multiple frame exchange sequences through the contention-free burst (CFB), each frame sequence separated by a SIFS, as long as the TXOP-limit is not reached. The remaining frame exchange sequence(s) will either be dropped if it is only one or wait for their transmission in the next CF-PI.

If transmission of the frame(s) is complete before the TXOP-limit, the remaining time of the TXOP is surrendered to the AP. After receiving the last frame, the AP waits for a SIFS time and then sends the CF-Poll frame to the next active user. If a talk spurt to silent state change arrives during the allocated TXOP time, the user adds 0 as the more data field in its last frame to let the AP know about its status change. In such a case the AP will move the user from PL to ISL. The structure of H-CFI is demonstrated in Fig. 1.

Dynamic Polling List Management Algorithm — The order of the PL may not be the same in each H-CFI. A newly admitted voice user is placed at the top of the PL. The order of the admitted service list (ASL) is kept constant in ascending order of the AIDs of all admitted users. However, in the ISL, a user whose status has changed from talking to silent state prior to another user will get a prior position. The D-



■ **Figure 1.** *Structure of the hybrid contention-free interval (H-CFI).*

The D-PLM function calculates the silence duration of each silent user through initiation of a timer called Idle State Timer for each idle user as soon as it becomes idle. These users get their respective position at the head of the PL before the old active users in the PL.

PLM function calculates the silence duration of each silent user through initiation of a timer called an idle state timer (IS-T) for each idle user as soon as it becomes idle. These users get their respective position at the head of the PL before old active users in the PL.

The D-PLM function records the positions of such users in an idle state duration vector (ISD-V) based on the priority of a user with the shorter silence duration. In this way, a user with the shorter silence duration will get a prior position in the PL. The ISD-V determines the traffic arrival behavior at individual voice user. A user with shorter silence duration shows more frequent traffic arrival pattern than that with the longer silence period. Intuitively, a user having more frequent arrival patterns of busrty traffic will need priority over others. Therefore, the D-PLM algorithm smartly provides a fair share of bandwidth according to the requirements of users.

The D-PLM algorithm addresses the issue of status change from talk spurt to silence state through in-band signaling, as in [6–10]. This is an efficient proactive approach. However, it requires either a modification of the MAC frame format by adding a "more data field" with a logic value 0 or 1 bit (with 0 indicating the last frame and 1 indicating more frames to follow) or help from the upper layers. The sequence of actions during CF-ADI is shown in Fig. 2.

Traffic Stream Admission Control Algorithm

In view of the wireless medium's bandwidth constraints, it is imperative to have efficient control over the admission of new voice calls in order to provide a consistent level of parametric QoS, especially the delay bounds and acceptable level of packet loss, to already admitted traffic

In view of the wireless medium bandwidth constraints, it is imperative to have an efficient control over the admission of new voice calls in order to provide a consistent level of parametric QoS, specially, the delay bounds and the acceptable level of packet loss, to already admitted traffic streams.

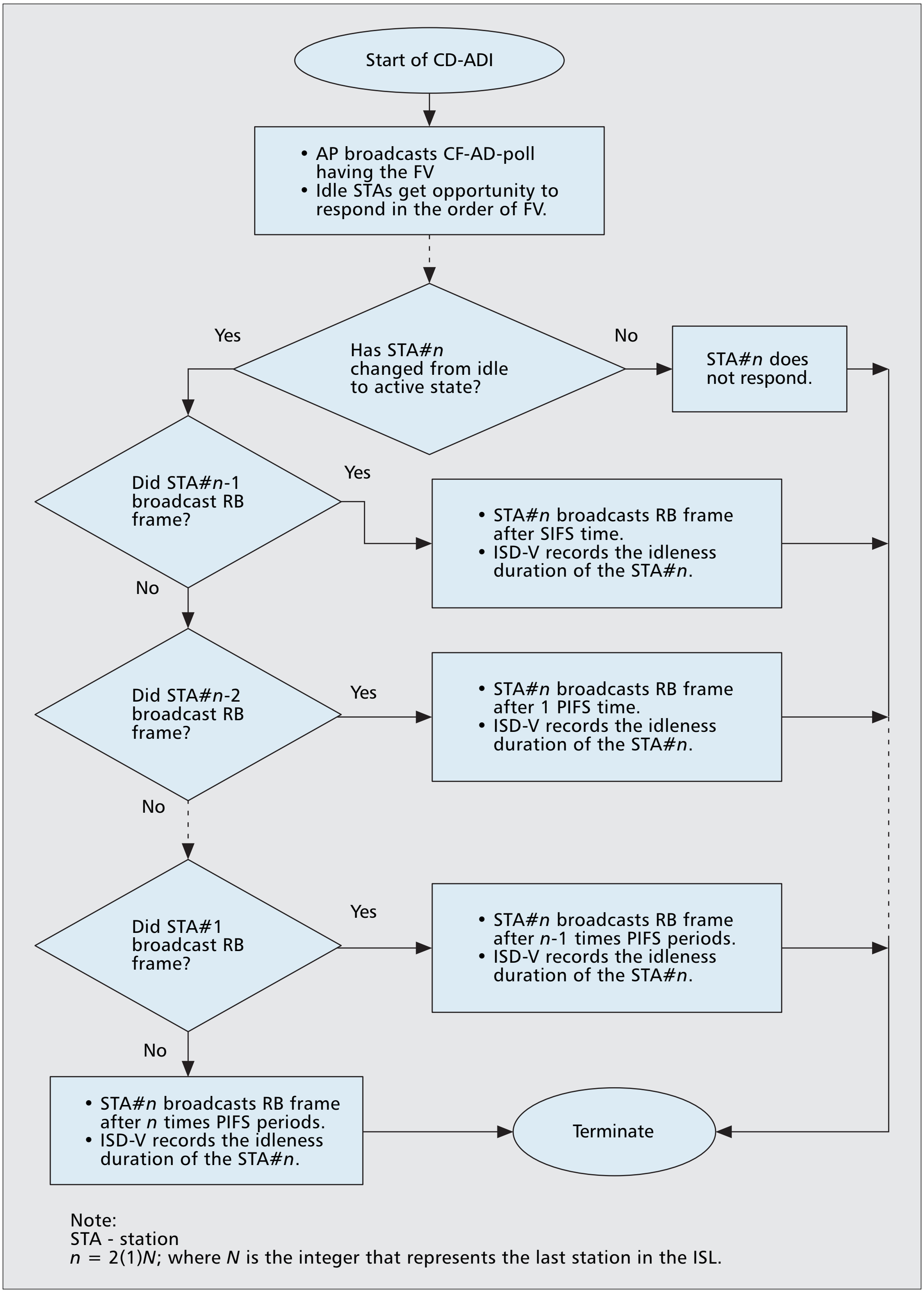
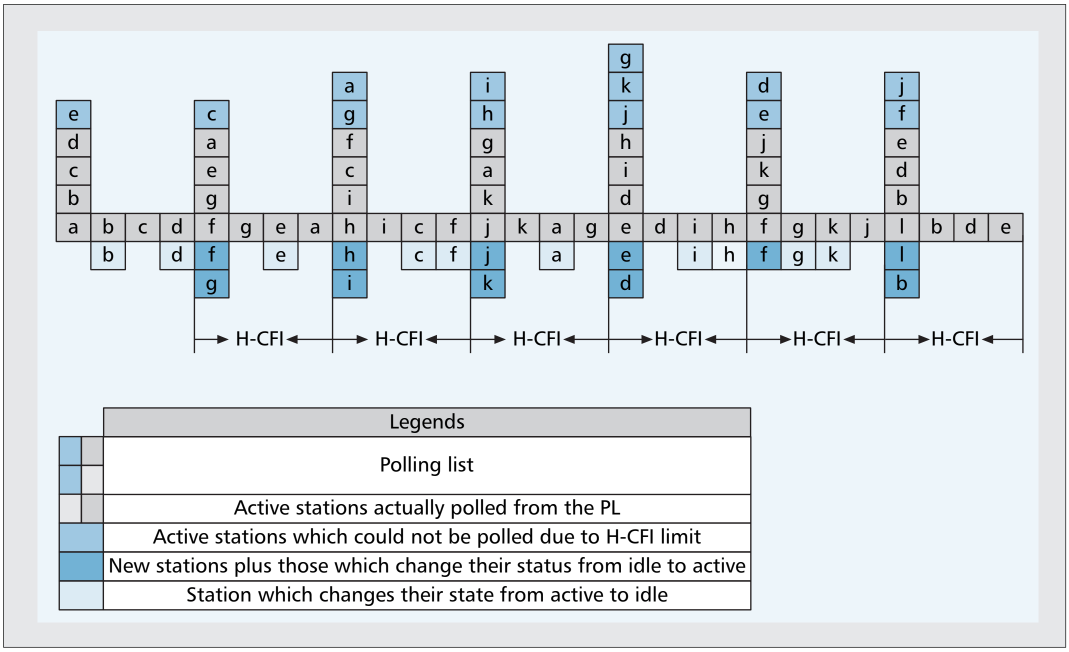


Figure 2. *Sequence of actions during contention-free activity detection interval of H-CFA.*

streams. This can be achieved through the negotiation of TSPECs as envisaged in [4]. To this end, we introduce a Traffic Stream Admission Control (TS-AC) algorithm that will keep the number of admitted voice users below some measured maximum count. The TS-AC function is incorporated in the AP that provides measurement-based CAC. Since the H-CFI must fulfill the delay bound requirements of all admitted TSs, its length is limited (to, say, *H_CFI_MAXlength*) according to the minimum negotiated delay bound of all the TSPECs and cannot go beyond that maximum limit. The TS-AC function keeps track of the unused bandwidth (*H_CFI_Unused*) during each H-CFI (i.e., *H_CFI_Unused* = *H_CFI_MAXlength* — *H_CFI_Timer*, where *H_CFI_Timer* is the actual bandwidth used during the current H-CFI). When a new call admission request comes, it differentiates the requested bandwidth from the measured unused



■ **Figure 3.** *Dynamic polling list management (D-PLM) and traffic stream admission control (TS-AC) algorithms.*

The proposed TS-AC algorithm maximizes the capacity by exploiting one characteristic of voice traffic that it can tolerate non-consecutive packet loss to some acceptable level. This means that all of the admitted traffic streams may not be served during one H-CFI.

bandwidth in the current H-CFI. If the differential is positive, it admits the call; otherwise, the call admission request is denied.

The proposed TS-AC algorithm maximizes capacity by exploiting one characteristic of voice traffic: it can tolerate nonconsecutive packet loss to some acceptable level. This means that all admitted TSs may not be served during one H-CFI. Therefore, in the worst scenario the maximum number of admissible TSs depends on the fact that no single user in the PL should remain unserved during two consecutive H-CFIs so that an unserved user does not undergo consecutive frame loss. Therefore, the AP polls the active users in the PL in a round-robin fashion but stops where the maximum limit of H-CFI comes. In the next H-CFI it will start polling from the user where it had left off in the last H-CFI. The TS-AC algorithm quantifies the increased capacity from the measured maximum number of TSs that will definitely be polled during the *H_CFI_MAXlength* period. This quantification mechanism is based on the maximum negotiated QoS level in terms of percentage of times an admitted voice user may undergo nonconsecutive packet loss. Therefore, according to this algorithm, the greater the maximum negotiated QoS level, the larger the capacity, and vice versa. The D-PLM and TS-AC functions are demonstrated in Fig. 3.

Let us represent the maximum number of TSs polled during the H-CFI by N_p, the maximum negotiated QoS level in terms of percentage of times an admitted voice user may undergo nonconsecutive packet loss during its entire voice session life by QoS_max, and the resulting increased number of TSs (i.e., the sum of the number of TSs that are polled during the H-CFI and cannot be polled due to the maximum limit of the H-CFI) by N_a. Then

$$QoS\_max = \left[\frac{N_a - N_p}{N_p}\right] \times 100 \Rightarrow N_a = round\_off\left(\frac{N_p}{1 - \frac{QoS\_max}{100}}\right). \tag{2}$$

Performance Evaluation

We compare the proposed H-CFA protocol to a few other similar protocols in the literature through simulations. We do not consider schemes that address fairness or priority QoS issues, or use (fully or partially) contention-based medium access approaches for silence suppression, because these do not conform to our objectives. We compare H-CFA with schemes that are also purely contention-free such as the adaptive polling MAC scheme [10], CSSR polling [9], and round-robin polling in the IEEE 802.11e MAC standard [4]. We also evaluate the proposed TS-AC algorithm through simulations.

Performance Measures and Simulation Parameters

We evaluate two very important performance measures. One is system capacity, the maximum number of voice users the AP can poll during each service interval or that can be satisfied according to the TSPEC parameters under the call admission control mechanism. The other is the wireless medium access waiting time delays to the first talk spurt frames when silence to talk spurt state change occurs.

For the TS-AC algorithm, we determine the number of satisfied voice users among the potential users served with their guaranteed delay

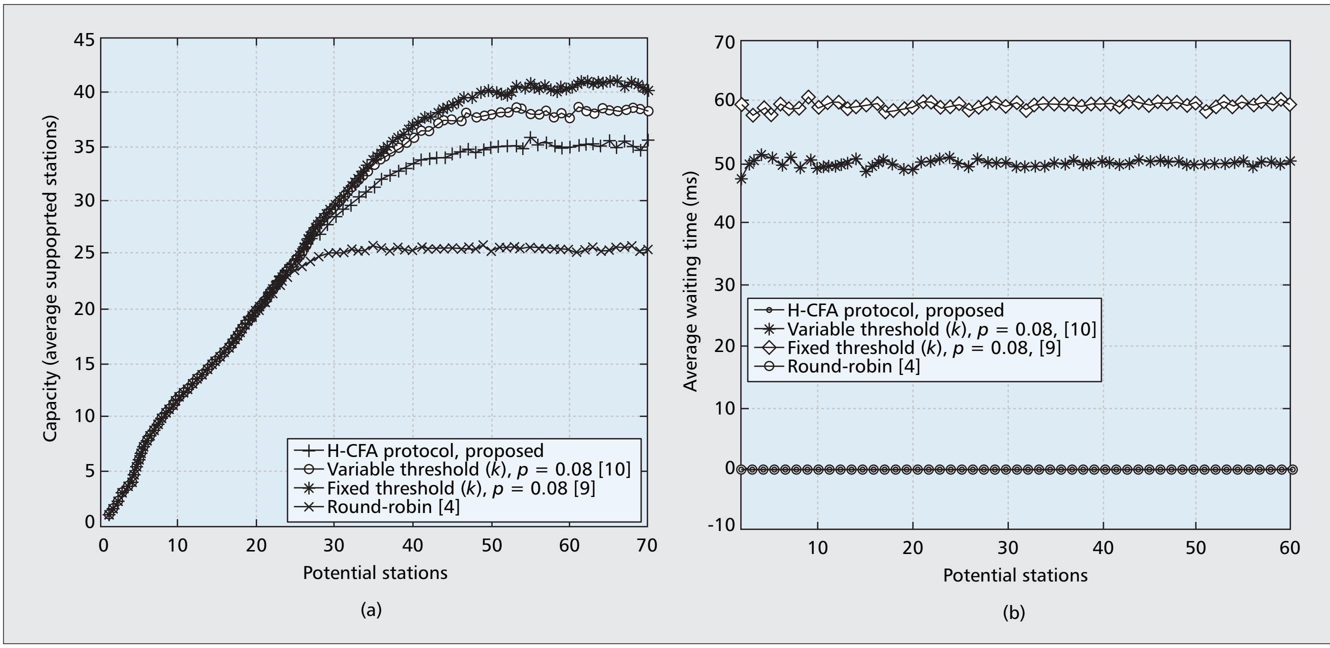


Figure 4. *a) Variations in WLAN capacities under different CF-MAC schemes; and b) waiting time delay experienced by the first talk-spurt.*

bound requirements who do not suffer a packet loss rate above the agreed QoS level.

We consider a two-state Markov model of uplink voice activity. One of the states is the idle state with mean duration of 1.35 s; the other is the active state with mean duration of 1 s (both idle and active periods have exponential distribution). We assume a packet generation rate of 64 kb/s. The transmission rate is assumed to be 2 Mb/s. For the schemes in [9, 10], we assume that the value of $p = 0.08$. The lengths of H-CFI, PIFS, and SIFS are 20 ms, 30 µs, and 10 µs, respectively. Also, the lengths of CF-poll, data frame, beacon frame, H-CF-poll, and BR frame (in bytes) are 20, 160, 88, 30, and 14, respectively. We assume that the channel is loss-free, and users are always in power-awake mode, or at least in soft power-sleep mode where they can always hear the broadcasts during H-CFIs.

Performance Results

H-CFA Protocol — The adaptive polling MAC scheme in [10] supports 38 users (with $p = 0.08$). This scheme saves bandwidth by not polling users for decreasing threshold time intervals during idle state. The CSSR polling scheme in [9] supports up to 40 users for the same parameter value. The reason behind this is that it uses fixed threshold time intervals for not polling the users during their silent state, and thus saves more bandwidth than the scheme in [10]. The round-robin scheme in [4] supports only 25 users due to the fact that it causes bandwidth wastage in polling the users all the time whether they are silent or talking, as seen in Fig. 4a.

The average waiting time delay to the first talk spurt packet in the adaptive polling MAC scheme in [10] is 50 ms. This is due to the decreasing threshold time intervals during the silent state when the frame(s) has to wait until the ongoing threshold time is over. In case of CSSR polling scheme in [9], waiting time delay is even larger (i.e., 60 ms). This is because the fixed threshold time intervals offer more time to arriving frames for waiting. The values of waiting times in these two schemes indicate that the first arriving frame has to wait for more than two consecutive H-CFs, which is not tolerable for voice users. These can be observed in Fig. 4b.

The proposed H-CFA protocol outclasses all the above discussed schemes. It can support 35 voice users by suppressing their silence periods. H-CFA does this job efficiently and independently without causing waiting time delays or consecutive frame losses to first talk spurt frames, as opposed to [9, 10] (Fig. 4a, b).

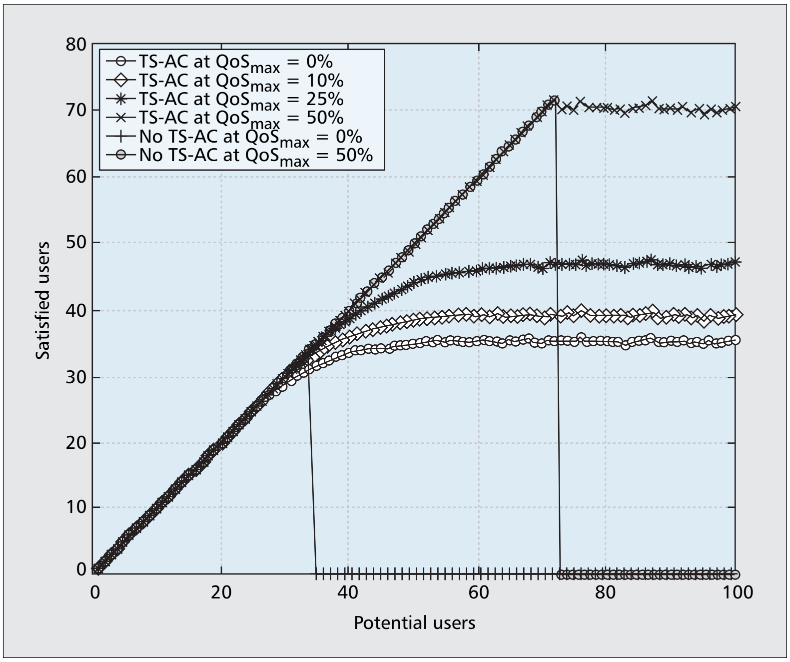
TS-AC Algorithm — We apply the TS-AC algorithm along with the H-CFA protocol to evaluate the maximum number of voice users that can be satisfied for different levels of maximum negotiated QoS. The results are shown in Fig. 5. At $QoS\_max = 0$ percent (i.e., each admitted user is polled in each H-CFI), the average number of satisfied users (N_a) is about 35. At $QoS\_max = 10$ percent, $N_a = 39$. At $QoS\_max = 25$ percent, $N_a = 47$. At $QoS\_max = 50$ percent, $N_a = 70$. Without application of the proposed CAC scheme at $QoS\_max = 0$ percent and $QoS\_max = 50$ percent, the results show that the number of satisfied users drops to zero after the respective maximum limit is reached.

Conclusion

The QoS-enhanced version of IEEE 802.11, IEEE 802.11e, has suggested solutions to the issues related to parametric QoS provisioning for real-time services in general. However, regarding real-time voice traffic, there are still open research issues. The WiFi standard does not suggest any mechanism for silence suppression in bursty voice calls, whereas VoIP technology does this in wired

IP networks. The hardest part of carrying out silence suppression in wireless LANs is for the central controller (the access point) to detect as soon as possible the arrival of first talk spurt frames at silent voice users. Also, due to the peculiar characteristics of the wireless medium, it is not easy for the central controller to predict the transmission time of a polled user, specially when there are large variances in PHY rates and frame sizes.

We have addressed two core QoS issues, efficient silence suppression and call admission control, related to provisioning of VoIP services in WiFi networks. We have discussed the challenges related to designing WiFi MAC architecture and provisioning QoS to voice services. The related works in the recent literature have been reviewed. In this connection we have proposed and analyzed a QoS-aware WiFi MAC protocol, the Hybrid Contention-Free Access (H-CFA) protocol, and a VoIP call admission control algorithm, the Traffic Stream Admission Control (TS-AC) algorithm. The H-CFA protocol provides substantial multiplexing capacity gain through silence suppression and purely contention-free medium access. TS-AC suggests efficient admission control for consistent delay-bound guarantees and further maximizes the capacity through exploiting the voice characteristic that it can tolerate some level of nonconsecutive packet loss. Performance analysis of H-CFA in comparison to other contention-free protocols including the IEEE 802.11 round-robin protocol shows far better performance of H-CFA regarding system capacity and waiting time delays to first talk spurt packets. H-CFA provides capacity gain up to 40 percent over the round-robin scheme used in the WiFi standard at no cost except some minor structural changes. Performance results of the TS-AC algorithm shows that it not only provides consistent delay satisfaction to the voice user, it also further increases the capacity gain up to 100 percent at the cost of different accepted QoS levels to voice users.



■ **Figure 5.** *Variations in the number of satisfied voice users at various QoS levels with and without TS-AC scheme.*

References

[1] ITU-T Rec. G.114, "General Characteristics of International Telephone Connections and International Telephone Circuits One-Way Transmission Time," Feb. 1996.

[2] J. Y. Yeh and C. Chen, "Support of Multimedia Services with the IEEE 802.11 MAC Protocol," *Proc. IEEE ICC '02*, 2002, pp. 600–04.

[3] R. S. Ranasinghe, L. L. H. Andrew, and D. Everitt, "Impact of Polling Strategy on Capacity of 802.11 based Wireless Multimedia LANs," *Proc. IEEE Int'l. Conf. Networks*, Brisbane, Australia, 1999.

[4] IEEE Std. 802.11e/D13.0, "Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Amendment: Medium Access Control (MAC) Quality of Service (QoS) Enhancements," Jan. 2005.

[5] J. Charzinski, "Activity Polling and Activity Contention in Media Access Control Protocols," *IEEE JSAC*, vol. 18, no. 9, 2000, pp. 1562–71.

[6] O. Sharon and E. Altman, "An Efficient Polling MAC for Wireless LANs," *IEEE/ACM Trans. Net.*, vol. 9, no. 4, Aug. 2001.

[7] P. Wang, H. Jiang, and W. Zhuang, "IEEE 802.11e Enhancement for Voice Service," *IEEE Wireless Commun.*, vol. 13, no. 1, Feb. 2006, pp. 30–35.

[8] R. Y. W. Lam *et al.*, "Polling-based Protocols for Packet Voice Transport over IEEE 802.11 Wireless Local Area Networks," *IEEE Wireless Commun.*, vol. 13, no. 1, Feb. 2006, pp. 22–29.

[9] E. Ziouva and T. Antonakopoulos, "Efficient Voice Communications Over IEEE 802.11 WLANs Using Improved PCF Procedures," *Proc. 3rd Int'l. Network Conf.*, University of Plymouth, U.K., 16–18 July, 2002.

[10] Y.- J. Kim and Y.-J. Suh, "Adaptive Polling MAC Schemes for IEEE 802.11 Wireless LANs Supporting Voice-Over-IP (VoIP) Services," *Wireless Commun. Mobile Comp.*, vol. 4, 2004, pp. 903–16.

[11] M. Veeraraghavan, N. Cocker, and T. Moors, "Support of Voice Services in IEEE 802.11 Wireless LANs," *Proc. IEEE INFOCOM '01*, vol. 1, 22–26 Apr. 2001, pp. 488–97.

[12] A. Bazzi *et al.*, "WLAN Call Admission Control Strategies for Voice Traffic over Integrated 3G/WLAN Networks," *Consumer Commun. and Networking Conf. '06*, vol. 2, Jan. 2006, pp.1234–38.

Biographies

IRSHAD A. QAIMKHANI (irshadkk@ee.umanitoba.ca), currently pursuing his M.Sc. degree in electrical engineering from the University of Manitoba, Canada, received his B.Eng. degree in electrical engineering from N.E.D. University of Engineering and Technology, Pakistan, in 1991. His research areas are cross-PHY-link-layer modeling, performance evaluation, and optimization in designing wireless communication networks, and efficient wireless medium access control protocols for capacity and real-time QoS enhancement and bandwidth adaptation. He has been extensively engaged in the industry as professional electrical engineer for several years.

EKRAM HOSSAIN [S'98, M'01, SM'06] (ekram@ee.umanitoba.ca) is currently an associate professor in the Department of Electrical and Computer Engineering at the University of Manitoba. He received his Ph.D. in electrical engineering from the University of Victoria, Canada, in 2000. His current research interests include design, analysis, and optimization of wireless communication networks and cognitive radio networks. He is a co-editor of *Cognitive Wireless Communication Networks* (Springer, 2007) and *Wireless Mesh Networks: Architectures and Protocols* (Springer, 2007), and a co-author of *An Introduction to Network Simulator NS2* (Springer, 2008). He serves as an editor for *IEEE Transactions on Mobile Computing*, *IEEE Transactions on Wireless Communications*, *IEEE Transactions on Vehicular Technology*, *IEEE Wireless Communications*, and several other international journals. He has served as a guest editor for special issues of *IEEE Communications Magazine* on Cross-Layer Protocol Engineering for Wireless Mobile Networks and Advances in Mobile Multimedia, and *IEEE Wireless Communications* on Radio Resource Management and Protocol Engineering for IEEE 802.16. He served as Technical Program Chair for CWNets '07 and WiNITS '07 held in conjunction with QShine '07, August 14–17, Vancouver, Canada. He served as Technical Protram Co-Chair for the Symposium on Next Generation Mobile Networks (NGMN '06) and NGMN '07 held in conjection with the ACM International Wireless Communications and Mobile Computing Conference '06 and '07.