Energy Efficient H.263 Video Transmission in Power Saving Wireless LAN Infrastructure

Ahmad M. Kholaif, Terence D. Todd, *Member, IEEE*, Polychronis Koutsakis, *Member, IEEE*, and Aggelos Lazaris, *Member, IEEE*

***Abstract*—Wireless local area networks (WLANs) are now being used in places where access point (AP) power saving would be very desirable. Although this is not currently possible, modifications to the IEEE 802.11 protocol have recently been proposed which would permit this functionality. A difficult problem that arises is to maintain good client power saving when AP power saving is introduced. This is especially true when carrying multiple real-time traffic flows such as H.263 video, where the payload size varies randomly over short time periods. In this paper, we first present protocol modifications for power saving quality-of-service (QoS) enabled access points (PSQAP). We then focus on the transmission of bursty real-time H.263 video over a PSQAP. A variety of mechanisms are proposed for this purpose, which result in various tradeoffs between station and PSQAP power saving performance. Analytical models and simulation experiments are used to assess the performance of the video scheduling algorithms. The best overall performance is obtained using a mechanism that combines a novel variant of the Power Save Multi-Poll (PSMP) protocol with a scheduling mechanism based on a discrete autoregressive video prediction model.**

***Index Terms*—H.263, media access control, power saving, video transmission, wireless LAN.**

I. Introduction

WIRELESS local area network (WLAN) infrastructure is sometimes operated using battery power [1]. In these types of networks, access point (AP) power consumption is often significantly higher than that which would be possible if power saving procedures were available for the AP. Unfortunately, this is not currently possible, since IEEE 802.11 APs are required to be continuously active, even when there is no network activity [2].

End station power saving has always been strongly supported in the IEEE 802.11 standard.[1] The standard defines two states for an end station, i.e., the *Awake* and the *Doze* state. The Awake state occurs when the station radio interface is active and is capable of sensing radio channel activity and transmitting or receiving packets. The Doze state is associated with power saving mode (PSM), where parts of the radio circuitry are powered off, and a station is unable to perform any of the above functions. The IEEE 802.11 standards have introduced a variety of procedures which allow stations to transition between these two states while supporting various types of end applications [3]–[5].

In addition to the power saving features defined in IEEE 802.11, power saving in general has received a lot of attention in recent years [6]–[9]. In [10] and [11], the authors investigate application-specific protocols for reducing the power consumption of network interfaces in streaming media applications. Reference [12] addresses the issue of the interaction between energy-saving protocols and TCP performance, and presents a bounded slowdown (BSD) protocol that guarantees bounded delay while conserving energy. This idea is extended in [13], where a smart power save mode is proposed to guarantee delay performance. Reference [14] uses different IEEE 802.11-style beacon periods for different client nodes, which improves the response time performance for certain web applications. Anand *et al.* [15] implemented a self-tuning power management (STPM) module that adapts its power configuration using application layer hints. The study in [16] analyzed the impact of background traffic on the power consumption of mobile stations (MSs). This paper proposed a PSM protocol with scheduled frame delivery procedures and differentiated multicasting.

Although there has been a lot of work that deals with power saving at IEEE 802.11 client stations, very little has appeared which considers power saving on the access point. Protocols for this purpose have only recently been proposed in [17] and [18]. Reference [17] first introduced IEEE 802.11-based power saving access points (PSAPs) for use in solar powered WLAN mesh applications. Three different frame design arrangements were introduced for adapting the proposed AP sleep schedules. The second reference extends the IEEE 802.11 network allocation vector (NAV) mechanism using a dynamic scheme for updating the sleep and awake periods based on a network allocation map (NAM). The NAM is carried in the AP beacons and provides information indicating the times when the AP is in a power saving mode [18]. Unlike the protocols proposed in this paper, both of these references only consider best effort traffic with no quality-of-service (QoS) guarantees. For this reason there are not suitable for supporting real-time traffic.

Manuscript received February 27, 2009; revised October 06, 2009. First published November 24, 2009; current version published January 20, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Zhihai (Henry) He.

A. M. Kholaif is with the Research In Motion, Inc. Waterloo, ON N2L 3W8, Canada (e-mail: akholaif@rim.com).

T. D. Todd is with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4K1, Canada (e-mail: todd@mcmaster.ca).

P. Koutsakis and A. Lazaris are with the Department of Electronic and Computer Engineering, Technical University of Crete, Chania, Greece (e-mail: polk@telecom.tuc.gr; alazaris@telecom.tuc.gr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2009.2037380

[1]Power saving in the IEEE 802.11 standard includes power save mode (PSM) in IEEE 802.11b [3], automatic power save delivery (APSD) in IEEE 802.11e [4], and power save multi-poll (PSMP) in IEEE 802.11n [5].

IEEE 802.11 also includes support for real-time traffic. The IEEE 802.11e standard defines a hybrid coordination function (HCF) which achieves traffic priority differentiation so that real-time quality of service can occur [4]. The basic mechanism defines the times when a station is awarded transmission opportunities (TXOPs) and is therefore free to transmit packets. The basic procedures defined in IEEE 802.11e have been further developed in a number of studies. Ansel *et al.* [19], for example, presented an approach for efficient scheduling in a QoS enabled AP (QAP) based on measured queue sizes for each traffic stream. In [20], the mean TXOP is used for allocating channel access time and a token bucket algorithm allows nodes to vary their TXOPs over time. In [21], an *effective* TXOP is defined as that which can ensure that the packet loss probability is less than a preselected threshold. The effective TXOP duration of a newly arrived flow is first derived, and then standard IEEE 802.11e procedures are applied for packet access. The work in [22] uses a fixed TXOP value but offers multiple stream polls per service interval. Polling decisions are made by the QAP based on queue length and timing information.

In this paper, we first introduce MAC protocol extensions for an IEEE 802.11e-based power saving QoS enabled AP (PSQAP). The paper then focuses on the transmission of real-time video using the H.263 standard. Mechanisms are proposed which result in various tradeoffs between station and PSQAP power saving performance. Analytical models are proposed and simulation experiments are used to assess the performance of the video scheduling algorithms. Our results show that the power consumption at the PSQAPs can be significantly reduced while satisfying the QoS requirements for real-time traffic streams and while maintaining good client power saving performance. The best overall results are obtained using a mechanism that combines a novel variant of the IEEE 802.11n Power Save Multi-Poll (PSMP) protocol [5] with a video scheduling mechanism based on a discrete autoregressive model. The model is used to predict the aggregate bandwidth required for multiplexed videoconferencing streams.

The rest of this paper is organized as follows. Section II introduces the MAC protocol framework for PSQAPs. In Section III we consider the transmission of multiplexed real-time H.263 video streams over the PSQAP. Section III introduces a discrete autoregressive model which is used to predict the aggregate dynamic behavior of this type of videoconference traffic, and which is then integrated into the operation of the proposed protocols. A protocol based on modifications to the IEEE 802.11 PSMP mechanism is proposed which incorporates video prediction to reduce power consumption both at the PSQAP and at the client nodes. In Section IV we demonstrate the performance of the proposed mechanism through extensive simulations and comparisons with existing mechanisms. This is followed in Section V by new analytical models which compute PSQAP power consumption and can be used to validate the integrity of the simulation results. Finally, Section VI gives the conclusions of the paper.

II. Power Saving QoS Enabled AP (PSQAP)

In this section we give a brief overview of the basic service interval activation mechanism for PSQAPs [23]. The proposed protocol extends the concept of a NAM introduced in [18] for best effort traffic.

A power saving AP includes a NAM field in its periodic beacon broadcasts to coordinate traffic delivery and power saving at both end stations and at the AP. The NAM provides a simple layout of the APs current superframe, which consists of time intervals separated by boundaries that are marked as either fixed or moveable [18]. Further details of this mechanism can be found in [23], which includes some protocols based on a service interval activation mechanism. In Section III-B this mechanism is used in the design of a much improved protocol which combines a modified version of IEEE 802.11n PSMP and H.263 video prediction.

When real-time flows are present, a PSQAP schedules its awakening/sleeping pattern in a manner that satisfies the delay and packet loss requirements for the admitted flows. The proposed protocol employs a mechanism for dynamic activation and deactivation of scheduled service intervals [23]. The superframe is subdivided into a pool of N equal service intervals (SIs). The value of N determines the service interval length (t_{SI}) and the minimum delay bound for real-time streams. In Fig. 1(a), for example, a 100 ms superframe with eight SIs is shown. In this case the minimum delay bound that can be met by the PSQAP is 12.5 ms. Each SI consists of two sub-intervals, i.e., an active contention-based activity sub-interval (using IEEE 802.11 EDCA access) and an AP power saving sleep sub-interval. The boundary between these two subintervals is moveable and is dynamically updated. Initially, the AP advertises a beacon NAM that defines a single active SI which we refer to as the "Initial Access Period" (IAP), as shown in Fig. 1(a). The PSQAP may dynamically activate additional SIs in order to accommodate newly accepted traffic streams (TSs) or to accommodate increased levels of best-effort traffic. Note that in the diagrams of Fig. 1, un-activated SIs are shown as the cross-hatched timeline areas. During these periods the AP assumes a power saving sleep state, and as noted above, the AP also enters a sleep state during the sleep sub-interval of activated service intervals.

A QoS station (QSTA) may contend for channel access during the IAP to request the admission of a real-time traffic stream (TS). If the new TS is accepted, the PSQAP may activate additional SIs and adjust their activity sub-intervals in order to meet the QoS requirements for the admitted TS. Once a TS is accepted, the PSQAP instructs the corresponding QSTA to contend for channel access using IEEE 802.11e US-APSD [4] in a specific set of active SIs that meet the requirements of the TS.

A simple example is given to explain how the SI activation procedure works. In Fig. 1(a), the PSQAP is initialized and SI_1 is active with an EDCA activity interval set to 5 ms, leaving the remainder of the SI for AP power saving. In Fig. 1(b), a new best-effort station, QSTA_1, transmits during SI_1. The PSQAP extends the admissible part of SI_1 and advertises this in subsequent beacon NAMs. In Fig. 1(c), QSTA_2 has a new real-time TS (with a 50 ms delay bound) and the PSQAP adjusts the length of the access period of SI_1, activates SI_5, and advertises this in the NAM.

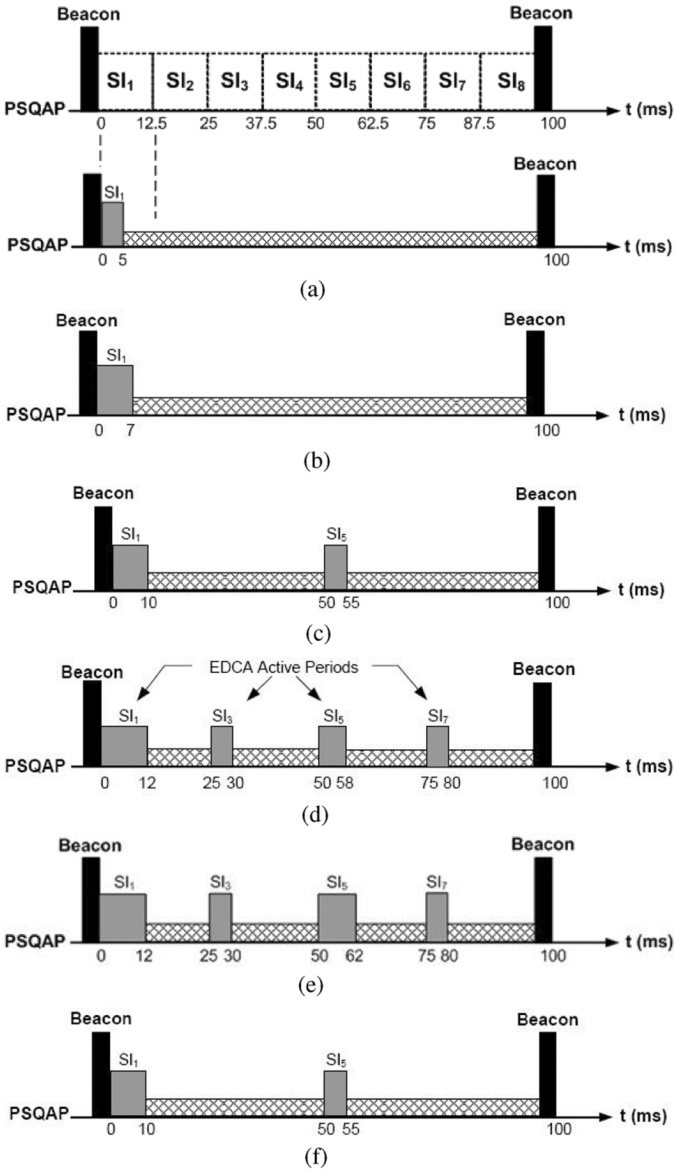

Fig. 1. Dynamic timeline update using dynamic service interval activation. (a) Single active SI: Initial access period. (b) Single SI with an extended access period. (c) Two active SIs with a 50 ms delay bound. (d) Four SIs with a 25 ms delay bound. (e) Four SIs with an extended access period in SI_5. (f) Deactivation of two SIs.

Fig. 1(d) shows $QSTA_3$ with a real-time TS which has a 25 ms delay bound. To accommodate this, the PSQAP activates both SI_3 and SI_7. Later, a best-effort station, $QSTA_4$, joins the IEEE 802.11 basic service set (BSS) [in Fig. 1(e)] and is assigned to transmit in SI_5. The right boundary of SI_5 is shifted to the right and the NAM is updated. After $QSTA_3$ terminates its TS [Fig. 1(f)], and since no other stations are assigned to SI_3 or SI_7, the PSQAP deactivates both service intervals and updates the NAM.

The proposed activation mechanism provides a flexible framework for attaining both station and AP power saving. An issue that arises is how to adjust the sleep behavior of the AP in a manner that achieves the best power saving at both the AP and at the client nodes. This problem is especially difficult when handling multiple real-time traffic flows whose bandwidth requirements are changing dynamically such as in H.263 videoconferencing applications. In the next section we show how some simple protocol extensions can be used to achieve this goal. In this type of traffic scenario, we find that the best mechanisms require the use of aggregate bandwidth prediction.

III. Energy Efficient H.263 Video Transmission for PSQAPs

H.263 is a widely used standard for videoconferencing applications and is used for compressing moving picture components at low bit rates [24], [25]. To achieve AP/client power saving, we use a discrete autoregressive (DAR) model in order to dynamically estimate the required aggregate video bandwidth. We first introduce the DAR model and then compare its performance with two other video traffic prediction approaches from the literature, using the proposed MAC protocol extensions. Our results, presented in Section III-A, motivated us to use the DAR(1) model in a more sophisticated scheme based on an improved version of the IEEE 802.11n PSMP protocol which is presented in Section III-B.

A. Videoconference Bandwidth Prediction

H.263 uses the idea of *PB* frames, i.e., two pictures being coded as a unit. With this coding option, the picture rate can be increased considerably without unduly increasing the transmission rate [26]. The work in [27] and [28] used five different long sequences of H.263 encoded videos (from [29]) with low or moderate motion, in order to derive a statistical model which fits well with the real data. The length of the videos varied from 45–60 min and the data for each trace consisted of a sequence of the number of packets per video frame. The work in [27] investigated the possibility of modeling the traces with a number of well-known distributions (gamma, lognormal, log-logistic, exponential, geometric, Weibull, Pearson V), and showed that the use of the gamma distribution (which has been shown in the past to be the best fit for MPEG-1, MPEG-2 and H.261 videos) is not a good choice for H.263 videoconference traces. The best fit among all studied distributions was achieved for all the traces with the use of the Pearson type V distribution. The superior quality of the Pearson V fit, in comparison to other distribution fits, was shown with the use of powerful goodness-of-fit tests like Q-Q plots [30], Kolmogorov-Smirnov (K-S) tests [30] and Kullback-Leibler (KL) tests [31]. For each one of these movies, the variable bit rate (VBR) coding version [in Quarter Common Intermediate Format (QCIF) resolution] was used.

Although the Pearson V was the best fit among all distributions, the degree of goodness-of-fit for the Pearson V varied for all traces. The reason that the Pearson V distribution cannot provide a perfect fit in any of the examined cases is that the high autocorrelation between successive video frames in a videoconference trace can never be perfectly captured by a distribution generating independent frame sizes according to a declared mean and standard deviation, and therefore none of the fitting attempts, as good as they might be, can achieve perfect accuracy. Still, very high accuracy can be achieved for *multiplexed* videoconference sources with the use

of a Discrete Autoregressive Model of order one [DAR(1) model]. A Discrete Autoregressive model of order p, denoted as $\mathrm{DAR}(p)$ [32], generates a stationary sequence of discrete random variables with an arbitrary probability distribution and with an autocorrelation structure similar to that of an autoregressive model. DAR(1) is a special case of a $\mathrm{DAR}(p)$ process and it is defined as follows: let V_n and Y_n be two sequences of independent random variables. The random variable V_n can take two values, 0 and 1, with probabilities $1-\rho$ and ρ, respectively. The random variable Y_n has a discrete state space S and $\Pr\{Y_n = i\} = \pi(i)$. The sequence of random variables X_n which is formed according to the linear model: $X_n = V_n X_{n-1} + (1 - V_n)Y_n$ is a DAR(1) process.

A DAR(1) process is a Markov chain with a discrete state space S and a transition matrix

$$\mathbf{P} = \rho\mathbf{I} + (1-\rho)\mathbf{Q} \tag{1}$$

where ρ is the autocorrelation coefficient, $\mathbf{I}$ is the identity matrix and $\mathbf{Q}$ is a matrix with $\mathbf{Q}_{ij} = \pi(j)$ for $i, j \in S$. Autocorrelations are usually plotted for a range, W, of lags. The autocorrelation can be calculated by the formula

$$\rho(W) = E[(X_i - \mu)(X_{i+w} - \mu)]/\sigma^2. \tag{2}$$

Here, μ is the mean and σ^2 is the variance of the frame size for a specific video trace. In [27] a DAR(1) model based on the Pearson V distribution was built and was shown through a series of statistical tests to provide high accuracy in modeling traffic from multiplexed H.263 videoconference sources. This model was also used in [28] in order to propose a new call admission control algorithm for videoconference traffic transmission over wireless cellular networks.

In the proposed protocols, the accurate predictions of the DAR(1) model are used to estimate the dynamic aggregate bandwidth requirements for multiplexed H.263 videoconference traffic served by the PSQAP. This is done by independently running the DAR(1) model for each trace and then adding the estimated bandwidths of all traces. In our study we use, without loss of generality, one of the five traces studied in [27] and [28], a video stream extracted and analyzed from a camera showing the events happening within an office ("Office Cam"). The trace has a mean of 90.3 Kbps, a peak of 1 Mbps, a standard deviation of 32.7 Kbps, an average packet size of 903 bytes and a maximum packet size of 5191 bytes. It needs to be emphasized that although we use one trace, we do not consider that each video starts from the beginning; for each new video trace arriving at the system a random start point in the "Office Cam" trace is chosen, therefore we are considering users who have the same mean, standard deviation and peak, but they do not transmit the same sequence. The estimate of the aggregate bandwidth is quite accurate even for a small number of sources (≥ 5), as shown in [27] and [28], and it is irrelevant if the sources have the same or different traffic characteristics; the DAR(1) modeling approach works equally well in both cases, as shown for example in [28] in a case where five video sources are present in the system, each one with different traffic characteristics. It is true that the DAR(1) model is not equally accurate for just one or two sources, as explained in [27] and [28]; still, the simplicity of the DAR(1) modeling approach, as well as the fact that the cases of multiplexed video users are of much more interest for call admission control purposes than the case of just one or two users in the system, makes it more than sufficient for use in a PSQAP.

The DAR(1) model uses the statistical data of multiplexed video traces to predict the channel resources required on a service interval basis. In other words, the capacity offered by the PSQAP to accommodate video traffic in the ith service interval, SI_i, is determined as $\mathrm{TB}_i = f(M(i))$, where $M(i)$ is the total number of admitted H.263 videoconference stations in SI_i. $f(M(i))$ is calculated as

$$\mathrm{TB}_i = \left[\mathrm{DAR}(1)_{M(i)} + M(i) \times \left(\frac{O}{R} + T_{\mathrm{overheads}}\right)\right] \mathrm{SF}(i-1) \tag{3}$$

where $\mathrm{DAR}(1)_{M(i)}$ is the predicted value from the DAR(1) model, of channel-time(s) required to transmit uplink packets for $M(i)$ video stations in SI_i. The aggregate bandwidth needed by a number of multiplexed video sources is estimated by independently running the DAR(1) model for each trace and then adding the estimated bandwidths of all traces. O is the per packet overhead, i.e., UDP/RTP/IP/MAC header overhead, R is the minimum physical transmission rate (bit/s) required to guarantee the QoS and $T_{\mathrm{overheads},i}$ denotes the IEEE 802.11e MAC overheads (e.g., SIFS, AIFS, ACK). $\mathrm{SF}(i-1)$ is the average *Surplus Factor* for the multiplexed video streams in the $(i-1)$th service interval and is used to compensate for the traffic loss caused by packet collisions or channel error. For the set of admitted video flows, the PSQAP updates $\mathrm{SF}(i+1)$ using a running weighted average

$$\mathrm{SF}(i+1) = \gamma \mathrm{SF}(i) + (1-\gamma)\frac{T_{\mathrm{Access}}(i)}{T_{\mathrm{Success}}(i)} \tag{4}$$

where $T_{\mathrm{Access}}(i)$ and $T_{\mathrm{Success}}(i)$ are the measured channel access times for the multiplexed video TSs regardless of success, and all successful transmission times as measured by the PSQAP, respectively. γ is a smoothing factor and is used to control the effect that fluctuations in $T_{\mathrm{Access},i}$ and $T_{\mathrm{Success},i}$ can cause to the values of the Surplus Factor. TB_i is the maximum allowable channel-time occupancy for videoconferencing traffic in a given SI_i.

Whenever a video packet is sent, successfully or not, the same amount of access time is subtracted from TB_i. A QSTA is not allowed to transmit any video packet once the corresponding TB_i is depleted. To decide upon the admission of a new videoconference flow, TS_j, that arrived in SI_i, the PSQAP first assumes that TS_j was actually admitted, and then it verifies that $\mathrm{TB'}_i < t_{\mathrm{CP}}$, otherwise TS_j is rejected. $\mathrm{TB'}_i$ is calculated as in (3) but with $M(i)$ replaced with $M(i)+1$ and t_{CP} is the admissible portion of t_{SI} allocated for contention access.

Using the above approach, we have used extensive simulation studies to compare the use of three different traffic prediction

TABLE I
SIMULATION PARAMETERS

Number of APs	1 (PSQAP)
Smoothing factor (f) / SF^{max}	0.65 / 2.0
R	$11Mbps$
$t_{BE} : t_{SI}$	1 : 5
Maximum Retry Limit	3
Maximum MPDU size	2304 bytes
UPD / RTP header	20 / 12 bytes
MAC / IP header	34 / 8 bytes
Beacon interval (B)	100 ms
SIFS / Slot time	$10/20$ μs
T_{ACK} / T_{PHY}	248 / 192μs
$\eta_{Tx}/\eta_{Rx}/\eta_D$	$750/500/8$ mW

mechanisms used to determine the time within a SI at which the PSQAP should terminate its (EDCA based) channel activity periods. Simulations were performed using a discrete event simulator written in the C programming language that includes detailed EDCA contention, the MAC protocol and the prediction-based scheduling algorithm presented above. The results presented are for a single AP including contention-based end station traffic. A trace file for a video stream extracted and analyzed from a camera showing the events happening within an office ("Office Cam") is used to generate video packets during the simulation runs. This trace included length and inter-arrival time information for the video packets. Randomization across multiple videoconference stations/sources is achieved by having each station start its videoconference session at a random packet index within the trace file. The results were obtained by averaging ten simulation runs, using a per run simulation time of 500 s, which was found to be sufficient to eliminate transient effects. In all of the results, the 95% confidence intervals are within 1% of the presented values. The various simulation parameters used in the experiments are summarized in Table I, and in all the comparisons presented, a superframe is assumed to have a single SI which (as discussed in Section II) consists of two sub-intervals, i.e., an EDCA activity sub-interval and an (AP power saving) sleep sub-interval. The SI length, t_{SI}, is 80 ms.

The first scheme uses the DAR(1) model presented above, and the other two are those from [33] and [34]. The work in [33], by Verbiest *et al.*, uses statistical parameters, i.e., average bit rate, the bit rate variance and the peak bit rate, to calculate the aggregate bandwidth demands (i.e., "equivalent bandwidth") for a given number of multiplexed VBR multimedia streams given a packet loss performance. Their analysis is based on the well-known fact that if a sufficient number of sources is multiplexed, the aggregate will approach a normal distribution with an average equal to the sum of the individual averages and a variance equal to the sum of the individual variances. The authors in [34] proposed methods to improve the number of admitted stations by creating multiple sub-flows from the global video flow, each with its own traffic specification. To simplify the design of the scheduler, their scheme shapes the traffic that is arriving at the MAC buffer through a twin leaky bucket that removes the burstiness of the video traffic. Based on a twin-leaky bucket analysis, the traffic specification (TSPEC) information of a video flow is used to calculate the effective bandwidth. In order to achieve an optimal scheduling policy, a low complexity linear-programming solution is proposed, which allocates the optimal transmission opportunity to each generated sub-flow in order to maximize utilization.

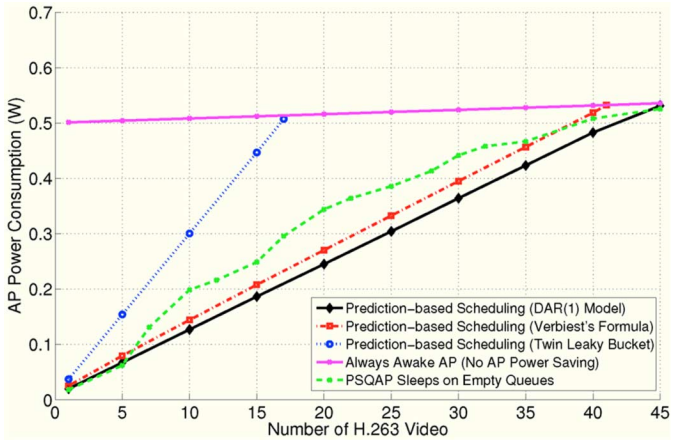

Fig. 2. AP power consumption.

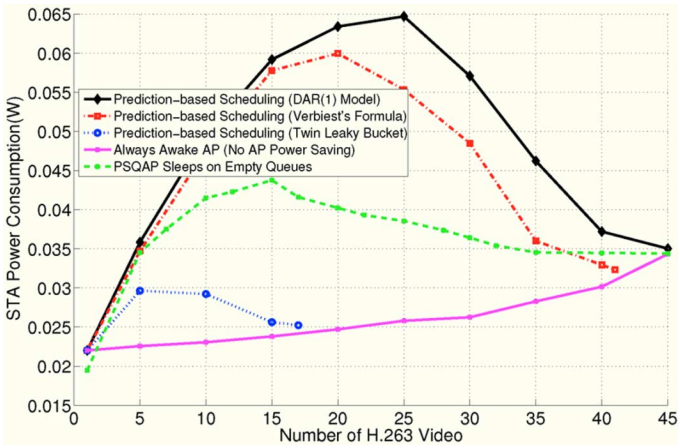

Fig. 3. Station power consumption.

The performance results for all the above-mentioned schemes are summarized in Figs. 2 and 3. We also include a mechanism, "PSQAP Sleeps on Empty Queues", where the PSQAP switches to sleep mode only when all QSTAs' queues are empty.

In Fig. 2, we can see that prediction-based video scheduling achieves better power saving at the AP compared to the "PSQAP Sleeps on Empty Queues" mechanism. It can be seen that the use of the DAR(1) prediction model results in better energy efficiency for the power-saving AP than the other two models. For the sake of comparison, we also show results for a conventional "always awake" AP. The effectiveness of the "Prediction-based (DAR Model)" scheme is even more evident compared to what the current IEEE 802.11 standard(s) can offer in terms of AP power saving. From Fig. 2, a PSQAP that uses DAR-prediction for bandwidth allocation is shown to need only 10%–50% of the total power consumed by an always awake AP to serve the same low/moderate video traffic load.

It can be seen from Fig. 3 that the DAR(1)-based scheduling algorithm exhibits the highest station power consumption values compared to the other schemes. Ironically, this is caused by the DAR(1) model accuracy. For this reason the PSQAP can significantly increase its power saving at the expense of enqueuing more video sources. The accurate aggregate bandwidth prediction with the use of the DAR(1) model leads to shorter

awake sub-intervals, which results in more contention among stations at the beginning of the SIs. Because of its superior prediction quality, the average MS power consumption values for the DAR(1) model are higher than those achieved by the algorithm using Verbiest's formula and are much higher than those of the Twin leaky bucket scheme from [34].

The above results illustrate the deficiencies in existing protocols which permit good power saving at both the stations and the AP. In order to prevent increases in power consumption at the station when AP power saving is used, we now introduce a more sophisticated design which more carefully controls the queuing and contention effects.

B. Improved IEEE 802.11n Unscheduled-PSMP

This section presents and evaluates the performance of a more sophisticated design that uses an improved version of the IEEE 802.11n US-PSMP power management mechanism [5] to further decrease station power consumption. In the standard US-PSMP, a PSMP-capable station awakens and sends a trigger frame using carrier sense multiple access with collision avoidance (CSMA/CA) to inform the AP that it has packets to send or to indicate its availability for receiving packets. Based on the queue information provided from PSMP-capable stations, the scheduler at the AP allocates uplink and downlink transmission times (i.e., UTT and DTT) for each station as part of a PSMP frame. A PSMP frame starts with a PSMP Action frame, which contains the time schedule to be followed by both the AP and stations during a PSMP frame [5].

In the proposed mechanism we exploit the fact that the arrivals of H.263 video packets are synchronous and hence a PSQAP can periodically assign PSMP uplink/downlink transmission opportunities (TXOPs) to stations without the need to receive trigger frames. We present and evaluate the performance of two different scheduling schemes that utilize our modified version of US-PSMP. The proposed mechanisms differ in terms of 1) the number of times a station needs to awaken and attempt channel access before it can successfully transmit a video packet, 2) the number of SIs per superframe, and 3) the mechanism by which a PSQAP determines the length of station TXOPs. For each case, the performance of both the PSQAP and stations is studied. We show that by using the accurate aggregate bandwidth prediction provided by the DAR(1) algorithm, and by minimizing the number of state transitions, a near-optimal power performance can be achieved for QSTAs and the power saving AP. The schemes are introduced as follows.

1) Prediction Assisted US-PSMP (Single Service Interval): Each PSMP frame starts with a PSMP Action frame that contains the schedule for TXOPs of individual QSTAs. Unlike IEEE 802.11n, transmissions within a PSMP period use CSMA/CA rather than just the timing assigned by the AP. This permits the timing to self-adjust and accommodate variable length transmission times. If a TXOP is insufficient to accommodate a packet transmission of a video station, the station will complete its transmission, which will delay the transmission of stations which have TXOPs assigned later in the PSMP frame. This will result in an increase in power consumption for those stations compared with the perfect scheduling case.

The order in which TXOPs are allocated to QSTAs in a PSMP frame is determined based on an earliest-deadline first (EDF) scheduling. In the first round of the algorithm, QSTAs are sorted in an ascending order based on their packet transmission deadlines. TXOPs are then assigned in a reverse order, i.e., videoconferencing QSTAs with early deadlines are assigned TXOPs late in a PSMP frame, while QSTAs with late deadlines are assigned TXOPs early in a PSMP frame. For some video stations, the assigned TXOPs can result in discarded video packets because of exceeded packet delay-bounds. For this reason, a second round of the algorithm is performed in which the satisfaction of delay bound for each QSTA is tested against its assigned TXOP. For a station QSTA_i (assumed to be located at position i in the sorted QSTAs list), if the assigned TXOP is found to violate the specified delay bound, then the algorithm searches for a potential candidate among all stations that follow QSTA_i in the list (i.e., QSTAs with later deadlines). Once such a candidate is found, both its TXOP and the QSTA_i's TXOP are swapped, and the test restarts from the beginning of the list and only stops when the produced TXOP assignment satisfies the delay bound for all video stations.

The proposed modifications to US-PSMP pose challenges on how to determine the best TXOP length, and three different mechanisms are evaluated. The first is based on worst case scheduling, where the maximum video packet size is used. In the second mechanism, prediction from the DAR(1) Model is used to obtain an accurate estimate of the total bandwidth that will be needed for all stations, and the TXOP length of a station is set to $1/N$ of that estimate. We show that this mechanism is the most energy-efficient approach and it achieves a near-optimal power performance for both the PSQAP and the stations. The third mechanism is the limiting case, where $\mathrm{TXOP} \rightarrow 0$, and in which videoconferencing QSTAs awaken at the beginning of an activity sub-interval to transmit their packets using conventional EDCA (US-APSD). Note that for the limiting case, the scheme works in an identical manner to the prediction-based with 1 EDCA SI scheme presented earlier.

2) US-PSMP With Multiple Service Intervals (No-Prediction): For comparison purposes we also include a non-predictive protocol where a PSQAP schedules four SIs per superframe, each having a PSMP frame. In this case, prediction is not needed for determining the length of TXOPs. However, we still rely on the PSQAP's knowledge of packet arrival times at individual stations. A video packet that arrives at the i_{th} SI, SI_i, will be queued at the corresponding station's buffer regardless of channel availability. In the following SI, SI_{i+1}, the QSTA is assigned a fixed length (usually short) TXOP. If the assigned TXOP is too short to accommodate the entire video packet, then the packet is fragmented. The PSQAP in SI_{i+2}, based on the station's queue length information, will schedule another TXOP that has the exact length required to send the remaining fragment. Since in this scheme it takes each video packet a maximum of three SIs (two SIs in addition to the SI at which the packet arrives) to obtain service, the choice of having four SIs per superframe ensures that video packets are served well within the 80 ms delay bound.

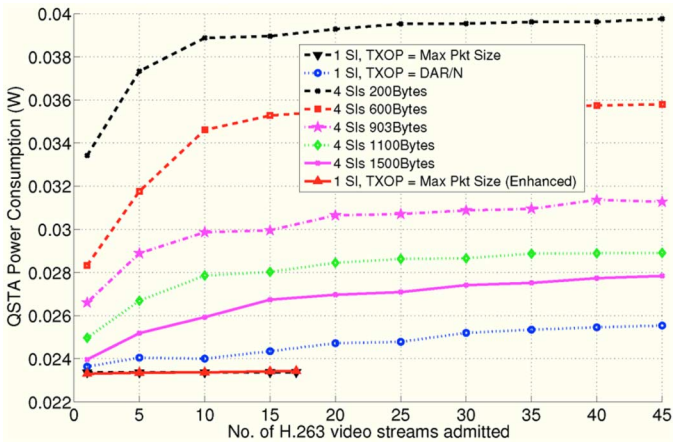

Fig. 4. Mean QSTA power consumption.

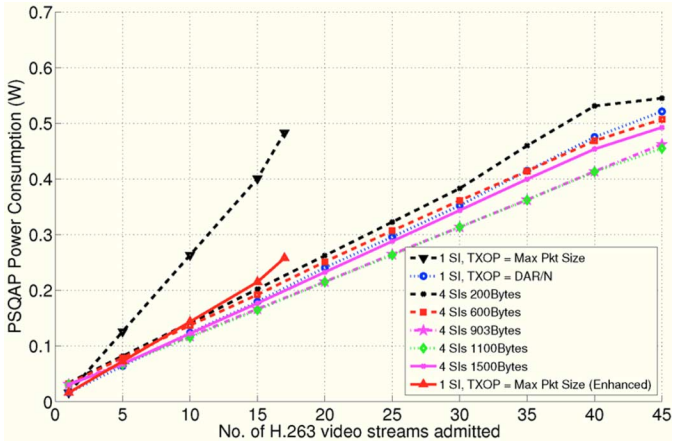

Fig. 5. Mean PSQAP power consumption.

IV. Performance Results

Figs. 4–6 show the results obtained from a simulation study of the two unscheduled PSMP-based designs and their variants introduced in Section III-B. The simulation assumes the same single AP network and contention-based end station traffic as in the simulation experiments at the end of Section III-A. We assume that a frequency plan is used so that the AP coverage areas are sufficiently isolated. The discrete event simulator was modified to include the two proposed unscheduled PSMP-based schemes. The curve labeled "1SI, TXOP = Max Pkt Size" corresponds to the use of the maximum video packet size to schedule TXOP intervals. As shown in Fig. 4, this mechanism guarantees optimal power performance for stations and thus it is a lower bound. This is true since a video packet transmission will always complete within the assigned TXOP and before the next station's TXOP begins, i.e., there is no queuing overhead. On the other hand, it can be seen from Fig. 5 that the same scheme is very energy inefficient from a PSQAP's perspective. This is also expected since for most of the video packets, the transmission completes well before the TXOP ends, leaving the PSQAP idle waiting and the channel unutilized until the beginning of the next TXOP. We also show results for an enhanced variant of this scheme, indicated by the curves labeled "1SI, TXOP = Max Pkt Size (Enhanced)". For this variant, a PSQAP also uses worst-case scheduling (i.e., TXOPs are scheduled based on the maximum video packet size). Once a station completes its packet transmission, the PSQAP makes a sleep decision based on the potential for power saving achieved by sleeping compared to the energy overheads required for sleeping the internal radio circuitry. As shown in Fig. 4, the station power performance is identical for both worst-case scheduling schemes. However, significant PSQAP power saving is achieved by the enhanced version. Worst-case scheduling schemes do not exploit the statistical multiplexing characteristics of variable-rate multimedia traffic such as H.263 videoconferencing streams. Therefore, a serious problem with these mechanisms is their inefficient usage of channel resources which is clearly reflected in the maximum number of H.263 video flows admitted (18) compared to 45 for the other curves.

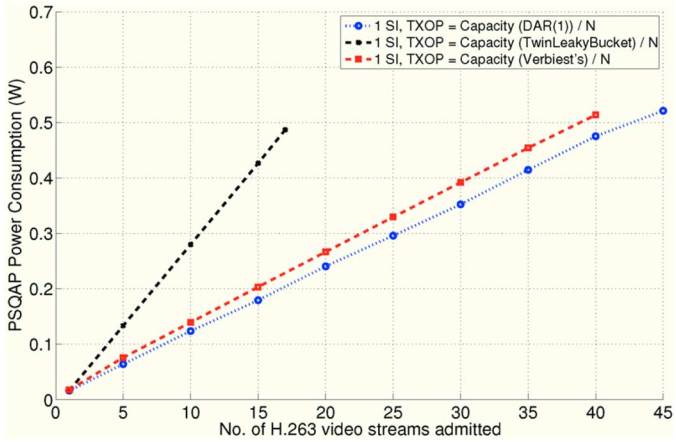

Fig. 6. Mean PSQAP power consumption results for prediction-assisted with one SI scheme using DAR(1) Model, Verbiest's Formula, and Twin-Leaky Bucket Analysis.

Fig. 4 also shows the power performance results for the single SI unscheduled PSMP-based scheduling scheme that uses the DAR(1) based algorithm. The results show the value of combining accurate capacity prediction with the energy-efficient MAC protocol. From Fig. 4, it can be seen that station power performance using this scheme closely approximates the lower bound on station power consumption. In numbers, the station power consumption under the prediction-assisted US-PSMP scheme (with one SI) is only 2 mW more than the lower bound, when the PSQAP is operating at the high end of the video traffic load range.

On the PSQAP side, the prediction-assisted scheme achieves very good power performance. The PSQAP consumes at most 15% more power than the lower value achieved by the curve labeled "4 SIs, 903 Bytes" which corresponds to the use of US-PSMP with the four SIs scheduling scheme. However, when comparing both mechanisms, Fig. 4 shows that the proposed algorithm saves about 22% of the station power. In general, and as shown in Fig. 4, the use of the prediction-assisted scheme with one SI achieves better station power performance compared to all other PSMP schemes that schedule four SIs/PSMP frames in a superframe. This interesting observation relates to the fact

that stations incur a significant power consumption when transitioning between awake and sleep states. Therefore, it is more energy-efficient to have each QSTA wake up only once per superframe to send its video backlog. For the same reason we can also explain the improved station power performance when using a larger TXOP size under the Multiple SIs scheme in Fig. 4. The larger the TXOP allocated for an MS, the more likely its packet transmissions will fit into this first TXOP and the MS can save the extra power that it would have consumed if it needed to awaken for a packet fragment transmission. The station power consumption reaches its lower bound when the TXOPs are scheduled based on the maximum video packet size. On the other hand, we can see that AP power consumption also decreases for larger TXOP sizes as shown in Fig. 5. In this case the AP will not need to schedule additional TXOPs. However, it can be seen that this trend stops when the TXOP exceeds 1100 bytes, and the reason is that for bigger TXOP values, the AP spends a significant portion of its awake time in an idle waiting state following the completion of relatively short packet transmissions.

Finally, Fig. 6 shows the results when using models from [33] and [34] to estimate the required capacity and determine the TXOP length. The superior prediction accuracy of the proposed framework using the DAR(1) algorithm achieves a lower power consumption on the PSQAP side without compromising the QoS requirements of H.263 multimedia streams.

V. Station Power Performance Analysis

This section provides analytical models for the power performance of a QSTA served by a PSQAP. The models and results presented also provide validation for the simulations which were used in Section IV. Our derivation of the average station power consumption is based on power calculations for a randomly chosen MS. In the remainder of this section, we refer to such a station as the "tagged" station.

A. QSTA Power Consumption Using Prediction-Assisted (Single EDCA SI)

In this protocol (from Section III-A), the PSQAP uses (3) to estimate the aggregate bandwidth requirements for multiplexed H.263 videoconference streams. Throughout the analysis, we use Ψ_s to indicate the group of video packets generated during the sleep period of a superframe and Ψ_a for video packets that are generated during the activity interval. The total number of admitted H.263 videoconferencing streams is denoted by N of which N_s denotes the number of QSTAs $\in \Psi_s$ and similarly, N_a denotes the number of QSTAs $\in \Psi_a$.

The station power consumption $\eta_{\rm MS}$ can be written as

$$\eta_{\rm MS} = \text{Superframes/sec} \times \zeta_{\rm SF} \tag{5}$$

in Watts, where $\zeta_{\rm SF}$ is the energy expenditure of a mobile station in a $T_{\rm SF}$ interval and is equal to

$$\zeta_{\rm SF} = \Pr\{\text{MS} \in \Psi_s\}\, \zeta_s + \Pr\{\text{MS} \in \Psi_a\}\, \zeta_a. \tag{6}$$

$\Pr\{MS \in \Psi_s\}$ and $\Pr\{MS \in \Psi_a\}$ are the probabilities that the tagged QSTA belongs to Ψ_s and Ψ_a, respectively, where $\Pr\{MS \in \Psi_s\} = N_s/N$ and $\Pr\{MS \in \Psi_a\} = N_a/N$; and $N_a + N_s = N$. ζ_s and ζ_a are the average energy (J) expended by a QSTA $\in \Psi_s$ and by a QSTA $\in \Psi_a$, respectively. The PSQAP is assumed to handle at most $N_{\max}$ concurrent H.263 videoconferencing sessions. To calculate N_a, we assume that the capacity of the PSQAP is discretized and normalized such that a new H.263 video stream consumes, on average, $1/N_{\max}$ of this capacity. When N simultaneous video flows are admitted, the PSQAP needs to roughly set $t_A = T_{\rm SF}N/N_{\max}$, where t_A is the activity sub-interval length which can be also written as $t_A = T_{\rm SF} - t_S$. From the simulation results presented in Fig. 2, $N_{\max}$ was found to be 45. Thus, N_a can be calculated as $(N \times t_A)/(T_{\rm SF}) = (N^2)/(N_{\max})$.

For a station that generates a packet during the sleep period, ζ_s can be expressed as

$$\zeta_s = \zeta_{{\rm Pkt},s} + P_c\, \zeta_{{\rm Pkt},c} + T_{Q,s}\, \eta_{Rx} + \zeta_{\rm Doze} + \zeta_{\rm sleep-wake} + \zeta_{\rm wake-sleep}. \tag{7}$$

Here, $\zeta_{{\rm Pkt},s}$ and $\zeta_{{\rm Pkt},c}$ are the energy expenditures (J) for an MS $\in \Psi_s$ while successfully transmitting a packet and during a packet collision, respectively. P_c is the packet collision probability for a station $\in \Psi_s$. η_{Rx} is the power (in Watts) consumed during receive/listen. $\zeta_{\rm Doze}$ is the energy expenditure (J) while the station is in the power saving state (normalized to a superframe interval) and is calculated as the product of the time in which the MS is sleeping and the power consumption during the doze mode ($\eta_{\rm Doze}$). $\zeta_{\rm sleep-wake}$ and $\zeta_{\rm awake-sleep}$ are the energies expended (J) during the doze-to-awake and awake-to-doze state transitions, respectively. To calculate P_c, recall that at the beginning of each activity period, all N_s stations that have packets in their queues awaken to contend for the channel. Also, an average of $N_s\, N_a/N$ new video packets ($\in \Psi_a$) are generated while N_s video packets are being served. This makes the average number of QSTAs contending for the channel at the beginning of an SI equal to $N_s(1 + N_a/N)$. For a tagged station that is randomly chosen, the packet collision probability can be written as

$$P_C = \frac{N_s\left(1 + \frac{N_a}{N}\right) - 1}{{\rm CW}_{\min}[\text{VIDEO}] - 1} \tag{8}$$

where ${\rm CW}_{\min}[\text{VIDEO}]$ is the initial contention window size for the VIDEO access category as defined for the IEEE 802.11e EDCA. Note that as an approximation, only first-order packet collisions are considered. The average queuing time $T_{Q,s}$ experienced by a tagged station ${\rm MS}_i \in \Psi_s$ can be calculated as

$$\begin{aligned} T_{Q,s} &= \frac{1}{N_s\left(1 + \frac{N_a}{N}\right)} \sum_{k=1}^{N_s\left(1+\frac{N_a}{N}\right)-1} k\,(T_{\rm Pkt} + T_{\rm DIFS}) \\ &= \frac{(T_{\rm Pkt} + T_{\rm DIFS})\left(N_s\left(1 + \frac{N_a}{N}\right) - 1\right)}{2} \end{aligned} \tag{9}$$

where $T_{\rm Pkt}$ is the packet transmission time(s) and it equals $(\text{Payload} + O)/(R) + (T_{\rm ACK} + T_{\rm SIFS} + T_{\rm DIFS})$. $\zeta_{{\rm Pkt},s}$ and

$\zeta_{\mathrm{Pkt},c}$ are calculated as

$$\begin{aligned}\zeta_{\mathrm{Pkt},s} = &\frac{\mathrm{Payload}+O}{R}\\ &\times \eta_{Tx} + (T_{\mathrm{ACK}} + T_{\mathrm{SIFS}} + T_{\mathrm{DIFS}}) \times \eta_{Rx}\\ &+ \eta_{Rx}\,(1-P_c)\,\mathrm{aSlotTime}\,\frac{\mathrm{CW_{min}[VIDEO]}-1}{2}\\ &+ \eta_{Rx}\,P_c\,\mathrm{aSlotTime}\,\frac{2\mathrm{CW_{min}[VIDEO]}-1}{2}\end{aligned} \tag{10}$$

and

$$\begin{aligned}\zeta_{\mathrm{Pkt},c} = &\frac{\mathrm{Payload}+O}{R} \times \eta_{Tx}\\ &+ (T_{\mathrm{SIFS}} + T_{\mathrm{DIFS}}) \times \eta_{Rx}\\ &+ \eta_{Rx}\,\mathrm{aSlotTime}\,\frac{\mathrm{CW_{min}[VIDEO]}-1}{2}.\end{aligned} \tag{11}$$

In this case, aSlotTime is the IEEE 802.11 MAC layer slot duration and η_{Tx} is the power (in Watts) consumed during transmit mode. Payload is the payload length of a video packet (in bits) and O is the per packet overhead, i.e., UDP/RTP/IP/MAC headers (in bits). $\mathrm{T_{ACK}}$, $\mathrm{T_{DIFS}}$ and $\mathrm{T_{SIFS}}$ are the channel time(s) for an acknowledgment frame, DIFS and SIFS, respectively.

Similarly, we can calculate ζ_a for a station that generates video packets during the awake period as

$$\begin{aligned}\zeta_a = \zeta_{\mathrm{Pkt},a} + T_{Q,a}\,\eta_{Rx} + P_{c,a}\,\zeta_{\mathrm{Pkt},c} + \zeta_{\mathrm{Doze}}\\ +2\,(\zeta_{\mathrm{sleep-wake}} + \zeta_{\mathrm{wake-sleep}}).\end{aligned} \tag{12}$$

Here, $\zeta_{\mathrm{Pkt},a}$ is the energy expenditure (J) for a $\mathrm{MS}_j \in \Psi_a$ while successfully transmitting a packet and is written as

$$\begin{aligned}\zeta_{\mathrm{Pkt},a} = &\frac{\mathrm{Payload}+O}{R} \times \eta_{Tx}\\ &+ (T_{\mathrm{ACK}} + T_{\mathrm{SIFS}} + T_{\mathrm{DIFS}}) \times \eta_{Rx}\\ &+ \eta_{Rx}\,\mathrm{aSlotTime}\,\frac{\mathrm{CW_{min}[VIDEO]}-1}{2}.\end{aligned} \tag{13}$$

$T_{Q,a}$ is the average queuing time experienced by a tagged station $\mathrm{MS}_j \in \Psi_a$ and it can be written as

$$\begin{aligned}T_{Q,a} = &\frac{N_s N_a}{N}\frac{T_{Q,s}}{N_a}\\ &+ \left(1-\frac{N_s}{N}\right)\left(\frac{N_a-1}{N_a}\right)(T_{\mathrm{Pkt}} + T_{\mathrm{DIFS}}).\end{aligned} \tag{14}$$

By substituting (6)–(14) into (5), the average station power consumption can be calculated. To verify the analysis presented above, the model is used to calculate an approximation to the average MS power consumption assuming the prediction-assisted EDCA (1 SI) scheduling scheme. The average packet size is assumed for video payload (i.e., $\mathrm{Payload} = 903$ Byte), $\mathrm{CW_{min}[VIDEO]} = 64$, $\mathrm{aSlotTime} = 20\mu$s and $O = 2704$ bit. $\zeta_{\mathrm{sleep-wake}}$ and $\zeta_{\mathrm{wake-sleep}}$ are modeled as 125 μJ and 250 μJ, respectively. The values for other parameters are listed in Table I.

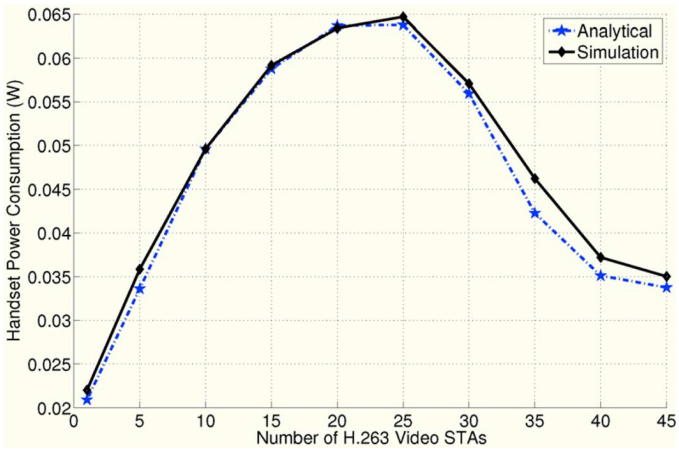

Fig. 7. Analytical versus simulation results for station power consumption under the prediction-assisted one SI (EDCA) scheduling scheme.

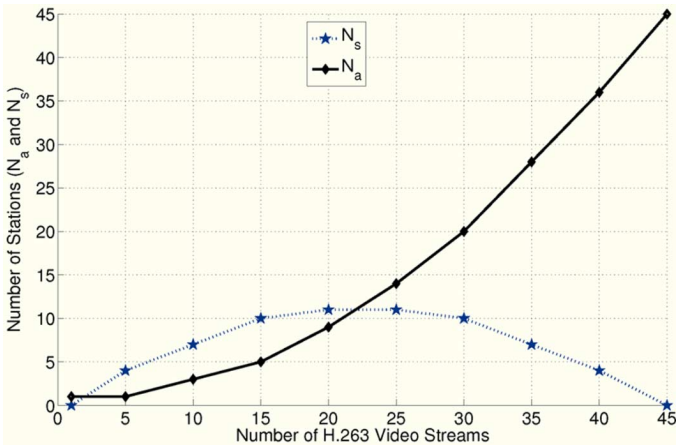

Fig. 8. Number of QSTAs in Ψ_a and Ψ_s versus total number of H.263 video stations admitted.

Fig. 7 shows the results obtained from the analytical model compared to simulation results previously shown in Fig. 3 for the curve labeled "Prediction-based (DAR Model)", shown in Fig. 3 to outperform the mechanisms from [33] and [34]. From Fig. 7, it can be seen that the analytical results closely match the results obtained through simulation. The effect of queuing and contention on station power performance is obvious in the range of 5–25 video streams. In this range and as shown in Fig. 8, most of the video packets are generated while the PSQAP is sleeping. This is true since for a small to moderate number of video flows, the ratio of the awake period to the total superframe interval is ≤ 0.5. Hence the quantity $\Pr\{\mathrm{MS} \in \Psi_s\}\,\zeta_s$ dominates the power consumption in (6) [and (5)]. However, as more video stations are admitted, N_s diminishes as a result of the PSQAP increasing the offered capacity in order to accommodate the increase in video traffic load. Consequently, the $\Pr\{\mathrm{MS} \in \Psi_a\}$ also increases and the term $\Pr\{\mathrm{MS} \in \Psi_a\}\,\zeta_a$ starts to dominate the station power consumption. From Fig. 8, we also see that N_s reaches zero when $N_{\max}$ video stations simultaneously exist at the PSQAP.

B. MS Power Consumption Under PSMP-Based With Multiple SIs Scheme

This mechanism was described in Section III-B. The power consumption derivation in this case is also based on (5). For this scheme, the energy expended by the station during a superframe interval, ζ_{SF}, can be defined as

$$\zeta_{\mathrm{SF}} = \zeta_{\mathrm{PSMP}} + \zeta_{\mathrm{Pkt(Payload<TXOP)}} + 2(\zeta_{\mathrm{awake-sleep}} + \zeta_{\mathrm{sleep-wake}}) + \zeta_{\mathrm{Doze}}. \quad (15)$$

In this case ζ_{PSMP} is the MS energy expenditure (J) while sampling a PSMP Action frame. $\zeta_{\mathrm{Pkt(Payload<TXOP)}}$ is the energy expenditure (J) during the transmission of an H.263 video packet of length Payload (in bits) and TXOP is the size (bits) of the fixed length PSMP TXOP that is first allocated to a newly arrived video packet in the PSMP-based with Multiple SI scheduling scheme. We can write

$$\zeta_{\mathrm{Pkt(Payload<TXOP)}} = \frac{\min(\mathrm{Payload}, \mathrm{TXOP}) + O}{R} \times \eta_{Tx} + (T_{\mathrm{ACK}} + T_{\mathrm{SIFS}}) \times \eta_{Rx}. \quad (16)$$

Equation (15) only considers the case where video packet payload is less than the first assigned PSMP-TXOP. However, if the assigned TXOP is too short to accommodate the entire video packet, the packet could be fragmented. In this case, the PSQAP uses the station's queue length information to schedule another PSMP TXOP (in the following SI) that has the exact length required to send the remaining packet fragment. The station will then need to awaken and sample the PSMP in the next SI to know its UTT. The station can then sleep to conserve energy and then awaken at its scheduled PSMP TXOP for packet transmission. To account for this latter case, (15) can be rewritten as

$$\zeta_{\mathrm{SF}} = 2\,\zeta_{\mathrm{PSMP}} + \zeta_{\mathrm{Pkt(Payload<TXOP)}} + \zeta_{\mathrm{Pkt(Payload>TXOP)}} + 4(\zeta_{\mathrm{awake-sleep}} + \zeta_{\mathrm{sleep-wake}}). \quad (17)$$

$\zeta_{\mathrm{Pkt(Payload>TXOP)}}$ is the energy expended by the station to transmit a packet fragment and is expressed as

$$\zeta_{\mathrm{Pkt(Payload>TXOP)}} = \frac{(\mathrm{Payload} - \mathrm{TXOP}) + O}{R} \times \eta_{Tx} + (T_{\mathrm{ACK}} + T_{\mathrm{SIFS}}) \times \eta_{Rx}. \quad (18)$$

The model was verified by calculating the station power consumption for five different lengths of the first PSMP TXOP including the two limiting cases (i.e., 200 and 5191 bytes) shown in Fig. 4 for the unscheduled PSMP-based scheme. Table II shows the results obtained from both the analytical model and simulations for an H.263 videoconferencing QSTA under the unscheduled PSMP-based with Multiple SIs Scheduling Scheme. Simulation results in Table II are the average of the power consumption values shown in Fig. 4 for the curves labeled "4 SIs 200 Bytes" and "1 SI, $\mathrm{TXOP} = \mathrm{Max}$ Pkt Size", respectively. Results from the table show that the analytic model closely approximates the results from the simulations. The excess in station power consumption when a small TXOP size is used can be explained in light of (17). Since a TXOP of size 200 bytes is likely insufficient to accommodate an H.263 video packet transmission (the minimum packet size $= 234$ Bytes), the MS still needs to awaken for a second PSMP-TXOP. As a result, the MS samples a second PSMP Action frame in order to know its UTT and then awakens again for packet transmission. The corresponding extra power components, ζ_{PSMP} and $2\,(\zeta_{\mathrm{awake-sleep}} + \zeta_{\mathrm{sleep-wake}})$, are shown in (17).

TABLE II
ANALYTICAL VERSUS SIMULATION RESULTS FOR STATION POWER CONSUMPTION (IN WATTS) UNDER THE PSMP-BASED WITH FOUR SIS SCHEDULING SCHEME

TXOP	Simulations	Analytical
200 Byte	0.0396	0.0408
600 Byte	0.0358	0.0368
903 Byte	0.0313	0.032
1100 Byte	0.029	0.0299
5191 Byte (Max.)	0.0233	0.0241

VI. CONCLUSION

In this paper we have presented a novel version of the IEEE 802.11e US-APSD access mechanism that includes access point power saving (PSQAP). This type of protocol is intended for use in battery powered wireless mesh networks where infrastructure power saving is used. We first considered a media access control protocol that implements a selective service interval activation/deactivation mechanism which uses network allocation maps for beacon advertisement. We then showed how bursty H.263 video can be efficiently transmitted over a PSQAP. An enhanced version of the IEEE 802.11n US-PSMP power management mechanism was used to provide a near optimal power performance for a PSQAP serving videoconferencing flows. The proposed mechanism uses a discrete autoregressive model of order one (DAR(1)) to accurately predict the dynamic aggregate bandwidth requirements for multiplexed H.263 videoconferencing streams. Results from a comparative study were presented, which evaluated several algorithms for video capacity prediction and resource allocation under the contention-based US-APSD channel access mechanism. Analytical models and simulation experiments were used to assess the performance of two video scheduling schemes that were based on a modified version of the IEEE 802.11n unscheduled PSMP power management.

Our simulation results show that bandwidth allocation using the proposed DAR(1)-based algorithm achieves the best power conservation performance at the PSQAP. The use of prediction-assisted scheduling combined with multiple SIs can roughly halve the station power consumption. However, the achieved improvements do not extend to high loading values and power consumption levels are at least double the lower-bound. A near optimal power performance for both the PSQAP and stations was achieved by using a modified version of the US-PSMP with one SI, where the length of the activity sub-interval is determined using DAR(1) prediction. In the case where stations with unknown videoconferencing content (i.e., with unknown traffic parameters) enter the network, we have shown in [28] for cellular networks that the system would have to search its pool

of "modes" (a mode corresponds to a set of traffic parameters) and find the mode which is most similar to the set of parameters declared by the unknown user. Then, the DAR(1) model of that "mode" is used for the station with the unknown content. We have simulated this case in our system in [28], focusing on the most "difficult" of the implementations examined, i.e., the handoff traffic arriving at various time intervals in the system originated from unknown modes, with which the system had to accommodate very quickly, in order to incorporate the new calls. The results presented in [28] showed that the difference between our mechanism's estimation and the actual videoconference bandwidth requirements was very small and averaged at only 0.8% over all the studied scenarios. For this reason we do not conduct this type of study in the present work, to avoid repetition. Recently [35], [36], it has been shown that our DAR(1) modeling approach for H.263 videoconference traffic can be applied, with slightly larger complexity, to MPEG-4 and H.264 videoconference streams, therefore our future work will include the necessary modifications (use of separate DAR(1) models for I, P and B frames of traces encoded with these standards) in order to show that our power saving mechanism works equally well for these streams.

The results which have been presented strongly depend on the number of video sources that a PSQAP is currently supporting. In our simulations we have assumed that good signal strength is provided throughout the wireless coverage area of the PSQAP so that capacity is maximized, as would typically be the case in an indoor network deployed for real-time QoS. Our conclusions therefore reflect the performance that one would expect for a system which is deployed for this purpose.

REFERENCES

[1] *SolarMESH*, McMaster Univ., Hamilton, ON, Canada, 2008. [Online]. Available: http://owl.mcmaster.ca/solarmesh/.

[2] A. Farbod and T. Todd, "Resource allocation and outage control for solar-powered WLAN Mesh networks," *IEEE Trans. Mobile Comput.*, vol. 6, no. 8, pp. 960–970, Aug. 2007.

[3] *ANSI/IEEE Std 802.11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications* , IEEE Standard 802.11, 1999, IEEE Press.

[4] *IEEE Standard 802.lie*, Media Access Control (MAC) Enhancements for Quality of Service (QoS), Nov. 2005, IEEE Press.

[5] *IEEE Standards Department, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Amendment 4: Enhancements for Higher Throughput* , IEEE P802.11n/D3.00, 2007, IEEE Press.

[6] J. P. Monks, V. Bharghavan, and W.-M. W. Hwu, "A power controlled multiple access protocol for wireless packet networks," in *Proc. INFOCOM 2001*, Apr. 2001, vol. 1, pp. 219–228.

[7] Y.-C. Tseng, C.-S. Hsub, and T.-Y. Hsieh, "Power-saving protocols for IEEE 802.11-based multi-hop ad hoc networks," *Comput. Netw., Elsevier*, vol. 43, no. 3, pp. 317–337, Oct. 2003.

[8] Y. Chen, N. Smavatkul, and S. Emeott, "Power management for VoIP over IEEE 802.11 WLAN," in *Proc. WCNC'2004*, Mar. 2004, vol. 3, pp. 1648–1653.

[9] K. M. Sivalingam, M. B. Srivastava, and P. Agrawal, "Performance comparison of battery power consumption in wireless multiple access protocols," *Wireless Netw.*, vol. 5, no. 6, pp. 445–460, 1999.

[10] S. Chandra and A. Vahdat, "Application-specific network management for energy-aware streaming of popular multimedia formats," in *Proc. General Track Annu. USENIX Annual Technical Conf. (ATEC)*, Berkeley, CA, 2002, pp. 329–342.

[11] S. Chandra, "Wireless network interface energy consumption: Implications for popular streaming formats," *Multimedia Syst.*, vol. 9, no. 2, pp. 185–201, 2003.

[12] R. Krashinsky and H. Balakrishnan, "Minimizing energy for wireless web access with bounded slowdown," *Wireless Netw.*, vol. 11, no. 1, pp. 135–148, 2005.

[13] D. Qiao and K. G. Shin, "Smart power-saving mode for IEEE 802.11 wireless LANs," in *Proc. 24th Annu. Joint Conf. IEEE Computer and Communications Societies (INFOCOM)*, Mar. 2005, vol. 3, pp. 1573–1583.

[14] S. Nath, Z. Anderson, and S. Seshan, "Choosing beacon periods to improve response times for wireless HTTP clients," in *Proc. 2nd Int. Workshop Mobility Management & Wireless Access Protocols (MobiWac)*, New York, 2004, pp. 43–50.

[15] M. Anand, E. B. Nightingale, and J. Flinn, "Self-tuning wireless network power management," *Wireless Netw.*, vol. 11, no. 4, pp. 451–469, 2005.

[16] Y. He, R. Yuan, X. Ma, J. Li, and C. Wang, "Scheduled PSM for minimizing energy in wireless LANs," in *Proc. IEEE Int. Conf. Network Protocols (ICNP)*, Oct. 2007, pp. 154–163.

[17] F. Zhang, T. D. Todd, D. Zhao, and V. Kezys, "Power saving access points for IEEE 802.11 wireless network infrastructure," *IEEE Trans. Mobile Comput.*, vol. 5, no. 2, pp. 144–156, Feb. 2006.

[18] Y. Li, T. D. Todd, and D. Zhao, "Access point power saving in solar/battery powered IEEE 802.11 ESS mesh networks," in *Proc. IEEE QSHINE*, 2005.

[19] P. Ansel, Q. Ni, and T. Turletti, "An efficient scheduling scheme for IEEE 802.11e," in *Proc. IEEE Workshop Modeling and Optimization in Mobile, Ad-Hoc and Wireless Networking*, Cambridge, U.K., Mar. 2004.

[20] A. Grilo, M. Macedo, and M. Nunes, "A scheduling algorithm for QoS support in IEEE 802.11E networks," *IEEE Wireless Commun. Mag.*, vol. 10, no. 3, pp. 36–43, Jun. 2003.

[21] W. F. Fan, D. Y. Gao, D. H. K. Tsang, and B. Bensaou, "Admission control for variable bit rate traffic in IEEE 802.11e WLANs," in *Proc. IEEE LANMAN Workshop*, Mill Valley, CA, 2004, pp. 61–66.

[22] N. Ramos, D. Panihrahi, and S. Dey, "Dynamic adaptation policies to improve quality of service of multimedia applications in WLAN networks," in *Proc. Int. Workshop Broadband Wireless Multimedia*, San Jose, CA, 2004.

[23] A. M. Kholaif, T. D. Todd, P. Koutsakis, and M. N. Smadi, "QoS-enabled power saving access points for IEEE 802.11e networks," in *Proc. IEEE WCNC*, Las Vegas, NV, Apr. 2008.

[24] S. X. Ng, J. Y. Chung, P. Cherriman, and L. Hanzo, "Burst-by-burst adaptive decision feedback equalized TCM, TTCM and BICM for H.263-assisted wireless video telephony," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 3, pp. 363–374, Mar. 2006.

[25] E. A. V. Navaro, J. R. Mas, J. F. Navajas, and C. P. Alcega, "Performance of a 3G-based mobile telemedicine system," in *Proc. IEEE Consumer Communications and Networking Conf. (CCNC)*, Las Vegas, NV, 2006, vol. 2, pp. 1023–1027.

[26] I. T. Union, ITU-T, Recommendation H.263, 1996.

[27] P. Koutsakis, "A new model for multiplexed VBR H.263 video-conference traffic," in *Proc. 49th IEEE GLOBECOM 2006*, San Francisco, CA, 2006.

[28] S. Chatziperis, P. Koutsakis, and M. Paterakis, "A new call admission control mechanism for multimedia traffic over next generation wireless cellular networks," *IEEE Trans. Mobile Comput.*, vol. 7, no. 1, pp. 95–112, Jan. 2008.

[29] F. H. P. Fitzek and M. Reisslein, "MPEG-4 and H.263 video traces for network performance evaluation," *IEEE Netw.*, vol. 15, no. 6, pp. 40–54, Nov./Dec. 2001.

[30] A. M. Law and W. D. Kelton, *Simulation Modeling & Analysis*, 2nd ed. New York: McGraw-Hill, 1991.

[31] K. P. Burnham and D. R. Anderson, *Model Selection and Multi-Model Inference*. New York: Springer-Verlag, 2002.

[32] P. A. Jacobs and P. A. W. Lewis, "Time series generated by mixtures," *J. Time Series Anal.*, vol. 4, no. 1, pp. 19–36, 1983.

[33] W. Verbiest, L. Pinnoo, and B. Voeten, "Impact of the ATM concept on video coding," *IEEE J. Select. Areas Commun.*, vol. 6, no. 9, pp. 1623–1632, Dec. 1988.

[34] M. van der Schaar, Y. Andreopoulos, and Z. Hu, "Optimized scalable video streaming over IEEE 802.11a/e HCCA wireless networks under delay constraints," *IEEE Trans. Mobile Comput.*, vol. 5, no. 6, pp. 755–768, Jun. 2006.

[35] A. Lazaris, P. Koutsakis, and M. Paterakis, "A new model for video traffic originating from multiplexed MPEG-4 video-conference streams," *Perform. Eval. J., Elsevier*, vol. 65, no. 1, pp. 51–70, 2008.

[36] A. Lazaris and P. Koutsakis, "Modeling video traffic from multiplexed H.264 videoconference streams," in *Proc. IEEE GLOBECOM 2008*, New Orleans, LA, Dec. 2008.

Ahmad M. Kholaif received the B.Sc. and M.Sc. degrees in computer engineering from Cairo University, Giza, Egypt, in 2000 and 2003, respectively, and the Ph.D. degree in electrical and computer engineering from McMaster University, Hamilton, ON, Canada, in 2008.

He is currently a WLAN Driver Software Developer with Research In Motion Ltd., Waterloo, ON, Canada, where he is actively involved in the investigation and development of low-power WLAN MAC implementations on the handheld side. In 2003, he instructed advanced networking courses at the National Telecommunication Institute in Cairo, Egypt. Prior to that, he worked as a Teaching Assistant in the Department of Computer Engineering at Cairo University and in the Electronics Engineering Department at the American University in Cairo. In 2000–2001, he worked as a Software Engineer in the Wireless Organization Group at Lucent Technologies in Cairo, where he investigated a number of research projects in the wireless networking area.

Dr. Kholaif formerly received the Ontario Graduate Scholarship, and he has a number of patent applications in the WLAN area where some of his patents had already been implemented on commercially available smart phones.

Terence D. Todd (M'76) received the B.A.Sc., M.A.Sc., and Ph.D. degrees in electrical engineering from the University of Waterloo, Waterloo, ON, Canada.

While at the University of Waterloo, he also spent three years as a Research Associate with the Computer Communications Networks Group (CCNG). During that time, he designed and lead the development of an early local area network testbed, called Welnet. He is currently a Professor of electrical and computer engineering at McMaster University, Hamilton, ON, Canada. He spent 1991 on research leave in the Distributed Systems Research Department at AT&T Bell Laboratories, Murray Hill, NJ. During that time, he worked on the characterization of one of the first ATM switches and developed techniques for mesh network media access control. He also spent 1998 on research leave at The Olivetti and Oracle Research Laboratory (ORL), Cambridge, U.K. While at ORL, he worked on the Piconet Project which was an early embedded wireless network testbed. His research interests include metropolitan/local area networks, wireless communications, and the performance analysis of computer communication networks and systems. At McMaster University, he has been the Principal Investigator on a number of research projects in the optical and wireless networking areas, and he currently directs a group working on wireless mesh networks, including those which operate from energy sustainable power sources such as solar power. His research includes both theoretical and practically oriented topics. For this reason, he has been heavily funded by the Canadian telecommunications industry, and his graduates typically find jobs in both academia and in leading companies in the field.

Prof. Todd is a past chairholder of the NSERC/RIM/CITO Chair on Wireless Networking. He is a past Editor of the IEEE/ACM TRANSACTIONS ON NETWORKING and is currently an editor for the IEEE TRANSACTIONS ON MOBILE COMPUTING. He is a Professional Engineer in the province of Ontario.

Polychronis Koutsakis (M'06) was born in Hania, Greece, in 1974. He received the five-year Diploma in electrical engineering from the University of Patras, Patras, Greece, in 1997 and the M.Sc. and Ph.D. degrees in electronic and computer engineering from the Technical University of Crete, Crete, Greece, in 1999 and 2002, respectively.

He was a Visiting Lecturer at the Electronic and Computer Engineering Department of the Technical University of Crete for three years (2003–2006). From July 2006 until December 2008, he was an Assistant Professor at the Electrical and Computer Engineering Department of McMaster University, Hamilton, ON, Canada, with his research being funded by the National Sciences and Engineering Research Council of Canada (NSERC). In January 2009, he joined the Electronic and Computer Engineering Department of the Technical University of Crete, as an Assistant Professor. His research interests focus on the design, modeling and performance evaluation of computer communication networks, and especially on the design and evaluation of multiple access schemes for multimedia integration over wireless networks, on call admission control and traffic policing schemes for both wireless and wired networks, on multiple access control protocols for mobile satellite networks, wireless sensor networks and powerline networks, and on traffic modeling. He has authored more than 80 peer-reviewed papers in the above-mentioned areas.

Dr. Koutsakis has served as a Guest Editor for an issue of the *ACM Mobile Computing and Communications Review*, as TPC co-chair for the 4th ACM WMUNEP 2008, as a project proposal reviewer for NSERC and for the Dutch Technology Foundation, and serves as Editor for the *Elsevier Computer Communications Journal*, the *International Journal on Advances in Networks and Services* and the *International Journal on Advances in Telecommunications*, as a TPC member annually for most major IEEE conferences in his research field and as a reviewer for almost all the major journal publications focused on his research field. He is a Voting Member of the IEEE Multimedia Communications Technical Committee (MMTC).

Aggelos Lazaris (M'07) was born in Athens, Greece, in 1982. He received the five-year Diploma and the M.Sc. degree in electronic and computer engineering from the Technical University of Crete, Crete, Greece, in 2006 and 2008, respectively. He is currently pursuing the Ph.D. degree at the Department of Computer Science and Engineering at the University of California, Riverside.

His research interests span the broad area of computer communication networks with emphases on traffic modeling, resource allocation for wireless networks, video encoding, and transmission. He has authored ten papers in the above-mentioned areas.