Optimized Cross-Layer Design for Scalable Video Transmission Over the IEEE 802.11e Networks

Chuan Heng Foh, *Member, IEEE*, Yu Zhang, Zefeng Ni, Jianfei Cai, *Member, IEEE*, and King Ngi Ngan, *Fellow, IEEE*

***Abstract*—A cross-layer design for optimizing 3-D wavelet scalable video transmission over the IEEE 802.11e networks is proposed. A thorough study on the behavior of the IEEE 802.11e protocol is conducted. Based on our findings, all timescales rate control is developed featuring a unique property of soft capacity support for multimedia delivery. The design consists of a macro timescale and a micro timescale rate control schemes residing at the application layer and the network sublayer respectively. The macro rate control uses bandwidth estimation to achieve optimal bit allocation with minimum distortion. The micro rate control employs an adaptive mapping of packets from video classifications to appropriate network priorities which preemptively drops less important video packets to maximize the transmission protection to the important video packets. The performance is investigated by simulations highlighting advantages of our cross-layer design.**

***Index Terms*—Multimedia transmission, network performance analysis, scalable video, wireless LAN.**

I. Introduction

AS A RESULT of the high performance to price ratio, the IEEE 802.11 based wireless local area networks (WLANs) have been massively deployed in public and residential places for various wireless applications. Given the growing popularity of real-time services and multimedia based applications, it has recently become more critical to tailor IEEE 802.11 medium access control (MAC) protocol to meet the stringent requirements of such services. The IEEE 802.11 working group has developed a new standard known as the IEEE 802.11e [1] to provide the quality-of-service (QoS) support. The IEEE 802.11e defines a single coordination function, called the hybrid coordination function (HCF), which includes two medium access mechanisms: contention-based channel access and controlled channel access. In particular, the contention-based channel access is referred to as enhanced distributed channel access (EDCA), which extends the legacy distributed coordination function (DCF) [2] by providing the MAC layer with per-class service differentiation.

Accepted for publication
IEEE Transactions on Circuits and Systems for Video Technology 2007

Among various applications, video streaming is one of the most attractive applications for WLANs. However, due to the characteristics of wireless networks such as high error rate, limited bandwidth, time-varying channel conditions, limited battery power of wireless devices, and dynamic network users, wireless video streaming faces many challenges. From the application-layer coding point of view, wireless streaming requires video coding to be robust to channel impairments and adaptable to the network and diverse scenarios. The need for video adaptation becomes obvious. In general, video adaptation can be implemented in many ways such as bit stream switching [3] and transcoding [4]. Scalable video coding (SVC) [5] is the latest video coding technique designed for video adaptation. SVC provides great flexibility in video adaptation since it only needs to encode a video once and the resulted bit stream can be decoded at multiple reduced rates and resolutions.

Recently, we have seen extensive studies [6]–[9] on streaming scalable video over lossy channels. The common idea is to use unequal error/loss protection (UEP/ULP), i.e., giving the more important information more protection, to explore the fine granularity scalability provided by SVC. Such an ULP idea has been implemented differently in different network protocol layers. For example, in the application layer or the MAC layer, the ULP is often provided through using different forward error correction (FEC) codes [7] or different automatic repeat request (ARQ) strategies [8]. In the network layer, the DiffServ is often used to provide the ULP [9]. There are also quite a few physical layer approaches such as using OFDM to provide different physical channels with different priorities [10], unequally distributing transmission powers [11], and using different modulations. Most of these existing ULP approaches only consider a single end-to-end connection and focus on optimally distributing network resource among different priorities under the constraint of a fixed total network resource. However, from the entire network point of view, the resource distribution in one connection is not independent of other competing connections. In other words, the ULP adjustment at one user will affect other competing users. Thus, the ULP strategy for one user should aim at not only maximizing its own video quality but also minimizing the bad effect to other users.

In designing a cross-layer QoS support for video streaming over WLANs, it is necessary that the understanding of video streaming characteristics and network behavior are developed. Under a microscopic investigation, we evaluate the EDCA behavior, propose an appropriate video streaming technique, and introduce a practical cross-layer design to achieve optimized

video streaming over WLANs. We first recognize the importance of all timescale rate control and its direct influence to the video quality. Based on that, we propose a *macro* and a *micro* rate control schemes in terms of timescale at the application layer and network sublayer respectively.

At the application layer, the macro rate control minimizes the distortion of the video quality given the bandwidth constraint by optimal bit allocation in our considered SVC. At the network sublayer, the micro rate control performs further rate cut by packet drops when network experiences congestion before the application can react with the macro rate control. Through an adaptive QoS mapping, the micro rate control enforces ULP that preemptively sacrifices video packet with low importance to protect the transmission of those with high importance. This combination ensures optimization of video streaming over the IEEE 802.11e WLAN.

The rest of the paper is organized as follows. Section II provides overviews of the IEEE 802.11e EDCA and SVC operations. In Section III, a thorough study on the EDCA is presented, followed by the detailed description of our cross-layer design. Section IV presents the experiment results highlighting the benefits of our proposed design.

II. Background

A. Overview of IEEE 802.11e EDCA

In the IEEE 802.11e standard, the EDCA mechanism extends the DCF access mechanism to enhance the QoS support in the MAC-layer through introducing multiple access categories (ACs) to serve different types of traffic. In particular, a node implementing IEEE 802.11e MAC protocol provides four ACs that have independent transmission queues as shown in Fig. 1. Each AC, basically an enhanced variant of the DCF, contends for transmission opportunity (TXOP) using one set of the EDCA channel access parameters including the following.

- $\mathrm{CW}_{\min}$: Minimal Contention Window (CW) value for a given AC. Assigning a smaller value to $\mathrm{CW}_{\min}$ gives a higher TXOP. Each AC is given a particular $\mathrm{CW}_{\min}$.
- $\mathrm{CW}_{\max}$: Maximal CW value for a given AC. Similar to $\mathrm{CW}_{\min}$, $\mathrm{CW}_{\max}$ is also assigned on a per AC basis.
- AIFSN: Arbitration interframe space number. Each AC starts its backoff procedure after the channel is idle for a period according to AIFSN setting.
- $\mathrm{TXOPlimit}$: The limit of consecutive transmission. During a TXOP, a node is allowed to transmit multiple data frames but limited by $\mathrm{TXOPlimit}$.

Note that if the backoff counters of two or more ACs co-located in the same node elapse at the same time, a scheduler within the node treats the event as a *virtual collision*. The TXOP is given to the AC with the highest priority among the colliding ACs, and the other colliding ACs defer and try again later as if the collision occurred in the medium. Details of the IEEE 802.11e MAC protocol operation is described in [1]. In short, through differentiating the services among multi-class traffic, EDCA provides a certain level of QoS in WLANs for multimedia applications.

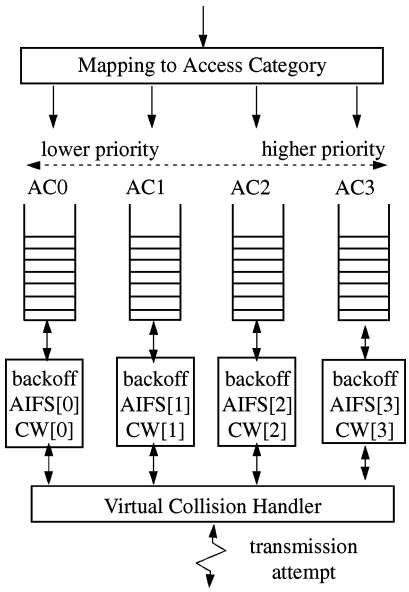

Fig. 1. Four access categories in an EDCA node.

B. Scalable Video Codec

Various SVC schemes have been proposed in the literature. Some of them are based on the traditional hybrid video coding scheme, such as MPEG-4 FGS. Many others utilize wavelet transform, either entirely based on 3-D wavelet or combining wavelet transform with motion compensated predictive coding [12], [13]. Up to now, the lifting scheme of the motion compensated temporal filtering (MCTF) proposed in [14] has become a popular approach for temporal decomposition since it can efficiently explore multiple-frame redundancies which is hardly achievable by conventional frame-to-frame (MPEG-4) or multiframe prediction (H.264) methods.

On the other hand, the development of scalable image coding is much more advanced than SVC. The latest image coding standard, JPEG2000 [15], provides a highly scalable (including component, quality, spatial and positional scalability) and also highly efficient (in terms of rate-distortion (R-D) performance) image codec. It is based on the concept of embedded block coding with optimal truncation (EBCOT) [16]. In this paper, in order to enable the easy adaptation of wireless video streaming, we develop a simple SVC scheme based on the integration of MCTF and JPEG2000. By exploring the unique features of MCTF and JPEG2000, the proposed codec not only achieves a competitive R-D performance but also has the property of high scalability.

Fig. 2 shows the structure of the developed encoder, and the decoding process simply follows an inverse procedure. The entire encoding process consists of three main components: MCTF, spatial coding and optimal bit truncation. In particular, each color component (YUV) of the original frames F_k is first filtered using MCTF with 5/3 wavelet. MCTF is applied iteratively to the set of lowpass bands in order to provide multiple frame rates in the final scalable bit stream. Through MCTF, we generate the motion vectors and many temporal bands (*T*-bands). Each *T*-band can be treated as an individual image. Then, JPEG2000 is used to encode these *T*-bands into multiple

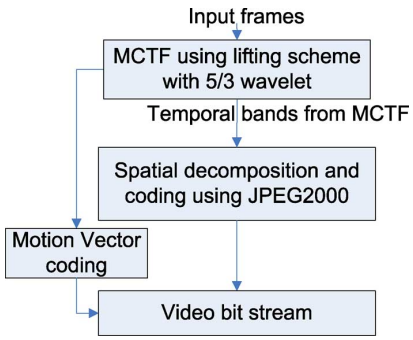

Fig. 2. Encoding structure.

quality layers, each of which has a R-D value. After removing those nonfeasible truncation points, optimal bit truncation is performed to reach the given target bitrate. The final video bit stream consists of the MV information generated by MCTF and the JPEG2000 bit stream for each *T*-band.

III. Cross-Layer Design

Employing a cross-layer based design for multimedia transmission over a network is a common approach to enhance the overall QoS support of the traffic delivery usually measured by video quality. Such a design introduces additional functions to build linkage between two protocol layers of different responsibilities with a single goal of achieving optimized multimedia delivery. In [17], the authors discuss various possible approaches to achieve cross-layer QoS design. Differentiating by the design emphasizes, the authors classify the cross-layer QoS design into five categories summarized as follows.

- **Top-down approach** relies on the higher layers to optimize their parameters and the parameters of their next lower layer in a top-down manner.
- **Bottom-up approach** lets the lower layers insulate the higher layers from losses and bandwidth variation.
- **Application-centric approach** uses either top-down or bottom-up approach to allow the application layer optimizing the lower layer parameters one at a time. However, due to the slower timescales in the application layer operation, this approach may not be optimal at all time.
- **MAC-centric approach** considers the passing of traffic information and requirements from the application layer to the MAC layer allowing the MAC layer to optimize the transmission. However, the approach is effective only when the application layer sets requirements that can be met by the MAC layer which is not a necessity in practice.
- **Integrated approach** mixes and matches the above approaches to provide a combined strategy for cross-layer QoS design.

Aiming to achieve a certain QoS support for multimedia delivery over the recently standardized EDCA operation in the IEEE 802.11e WLAN, it is important to understand the unique characteristics of EDCA. Attributed to its shared broadcast channel and contention based characteristics, we acknowledge the volatility of the IEEE 802.11e WLAN. Besides, congestion collapse, which is a common threat to contention based protocols, is likely to occur in the IEEE 802.11e WLAN under extreme load condition. We found that scalable video codec which provides high flexibility over rates and quality control serves as a good candidate for multimedia delivery over the IEEE 802.11e WLAN.

As the scalable video and EDCA are designed with no knowledge of each other, hence the marriage of the two requires additional entities to provide cross-layer cooperation. Our cross-layer design although adopts the complex integrated approach, the design complication is kept to minimum for operation practicability and implementation easiness. Targeting at achieving optimal video quality delivery using scalable video codec over EDCA, the design also features the so called *soft capacity* property, which adaptively sacrifices video quality of each existing user to accommodate new comers to the network.

Our cross-layer design uses rate control to optimize quality performance for multimedia delivery over EDCA. The application layer and network sublayer cooperate to decide the optimal transmission strategy for the scalable video stream. The common knowledge of the two layers is the available bandwidth which is the main factor dictating the employed transmission strategy at each layer. In brief, based on the detected network available bandwidth at the network sublayer, the application layer decides the encoding strategy and produces video streams that fit into the available bandwidth with the best possible video quality. However, the slower timescales at the application layer makes it difficult to respond to rapid bandwidth variation. To cope with this bandwidth variation at a smaller timescale, especially a sudden downside change in the available bandwidth, using the traffic information passed down from the application layer, the network sublayer reacts with a further transmission rate adjustment subjected to minimum distortion. This design of rate controls, where the application layer and the network sublayer perform *macro* and *micro* rate controls respectively, provides all timescale rate adaptation for multimedia delivery over EDCA.

Fig. 3 presents the block diagram of our design. The application layer consists of two components, namely optimal bit allocation and SVC packetization with relative priority index (RPI). The macro rate control is achieved by the coordination between the bandwidth estimation and optimal bit allocation components, where bandwidth estimation component measures the availability of the network capacity and indicates it to the optimal bit allocation component for distortion minimization. The micro rate control is achieved by the coordination between the SVC packetization with RPI and adaptive QoS mapping components, where each video packet is marked with an RPI indicating its importance, then the adaptive QoS mapping adaptively directs the packets to the appropriate EDCA prioritized queues for transmission so as to maximize the transmission of important video packets.

There is a number of cross-layer studies for multimedia transmission over QoS enabled networks [9], [18]–[21]. Many of these designs either employ application-centric or MAC-centric approach. In contradict, we propose an integrated approach where coordination crosses two layers of the application layer and the network sublayer. In the following subsections, we first study the characteristics of EDCA, leading to the proposal of the macro and the micro rate control schemes for our design. An

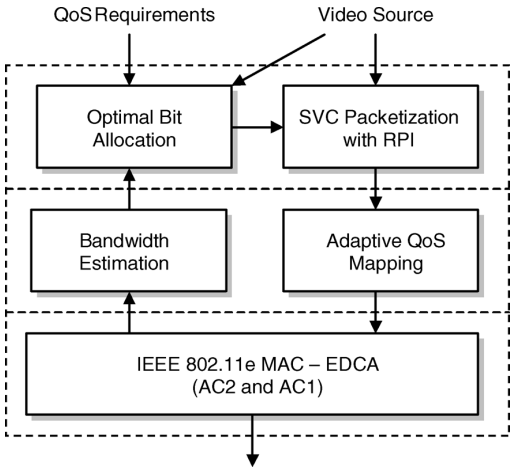

Fig. 3. Block diagram of the proposed cross-layer QoS design.

TABLE I
IEEE 802.11e MAC PROTOCOL SYSTEM PARAMETERS

Access category	AIFSN	CW_{min}	CW_{max}	Queue length	Maximum retry limit, r
AC3	2	7	15	25	8
AC2	2	15	31	25	8
AC1	3	31	1023	25	4
AC0	7	31	1023	25	4

available bandwidth estimation method is introduced, followed by the description of the optimal bit allocation and packetization with priority in 3-D wavelet SVC, and finally the design of the adaptive QoS mapping.

A. Study of the IEEE 802.11e EDCA Characteristics

Scalable video traffic consists of packets carrying video information of different importance indicated by RPI. These packet transmissions must receive according error and loss protections to exploit the scalable video benefit. Using the characteristics of the EDCA queues, we design a stream mapping strategy that adaptively maps packets of scalable video traffic onto two EDCA queues with ULP. We choose AC2 as one of the queue since its default purpose is to carry video traffic. AC1, which is used for best effort according to the IEEE 802.11e standard, is another chosen queue. It is used here to carry less important scalable video traffic. Packets mapped onto AC1 are prepared to make sacrifices in forms of packet drops under heavy network load conditions. These sacrifices make room for more important video packets to be transmitted successfully, and hence achieving higher QoS of video transmissions.

To understand the EDCA queue characteristics, we extend our earlier developed Markov Chain model [22] for the IEEE 802.11e EDCA performance analysis under saturation traffic condition to include the nonsaturation traffic condition. We consider the IEEE 802.11e system parameters listed in Table I. These constants follow the latest standard [1] whenever specified, otherwise, typical values applied.

According to our previous study in [22], due to the different AIFSN values standardized for AC2 and AC1, two Markov Chains are developed, referred as *Chain A* and *Chain B*, to model the backoff procedure of AC2 and AC1 respectively. The following list defines the variables in a particular chain associated with a certain AC.

- $W_{j,i}$ backoff window size at stage j of ACi, where $W_j = 2^{\min(j,m)}(\mathrm{CW}_{\min}+1)$ and $\mathrm{CW}_{\min}$ of the associated AC is specified in the IEEE 802.11e standard (see also Table I);
- m_i maximum backoff stage of ACi. That is, $W_m = \mathrm{CW}_{\max}+1$ where $\mathrm{CW}_{\max}$ of the associated AC is specified in the IEEE 802.11e standard (see also Table I);
- r_i maximum retry limit of ACi. We follow the value specified in [21] (see also Table I);
- u_i probability that after a successful transmission by a node using ACi, its queue remains empty after either an idle or a busy slot duration. This input variable provides unsaturation load adjustment of a node. Setting this probability to zero reduces the Markov Chain model to that of the saturation load ;.
- p_i probability that a transmission of ACi suffers a collision;
- q_1 probability that a transmission does not occur in a slot given that the previous slot is busy. This variable only applies in *Chain B*;
- q_2 probability that a transmission does not occur in a slot given that the previous slot is idle. This variable only applies in *Chain B*.

Figs. 4 and 5 show the developed Markov Chain for AC2 and AC1. The state $\{i,j\}$ in *Chain A* or the state $\{i,j,k\}$ in *Chain B* corresponds to the situation where a node currently carries value j in its backoff counter at the ith backoff stage, except for $j=-1$ which indicates that there is no packet awaiting for transmission in the node. Due to the different AIFSN settings in AC2 and AC1, an additional state variable k in *Chain B* is introduced to model the different actions that AC1 takes when the channel is detected to be busy and idle. Details of this explanation are given in [22]. We reuse the analytical approach and extend the model to include unsaturation traffic consideration. We achieve this by adding an extra state $\{0,-1\}$ or $\{0,-1,0\}$ into *Chain A* or *Chain B*, respectively. The new balance equations for the two chains are presented in the following.

Let $\alpha_{i,j}(p,u)$ be the stationary distribution of *Chain A* given that the collision and unsaturation probability for the associated AC are p and u, respectively. Owing to the chain regularities and imposing the stationary probability normalization, we have

$$\alpha_{0,0}(p,u) = \begin{cases} \left(\frac{\iota_r+\kappa}{\xi}+\frac{u}{1-u}\right)^{-1}, & r \le m \\ \left(\frac{\iota_m+\kappa+\nu}{\xi}+\frac{u}{1-u}\right)^{-1}, & r > m \end{cases} \quad (1)$$

where ξ, ι, κ, and ν are given by

$$\begin{aligned} \xi &= 2(1-2p)(1-p) \\ \iota_r &= W_0\left(1-(2p)^{r+1}\right)(1-p) \\ \iota_m &= W_0\left(1-(2p)^{m+1}\right)(1-p) \\ \kappa &= (1-2p)(1-p^{r+1}) \\ \nu &= W_0 2^m p^{m+1}(1-2p)(1-p^{r-m}). \end{aligned} \quad (2)$$

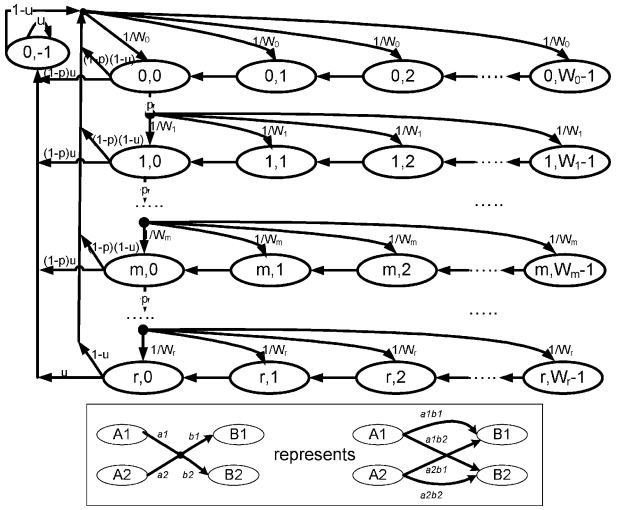

Fig. 4. Markov chain model for AC2 (chain A), where W_x, r, m, p and u refer to $W_{x,2}$, r_2, m_2, p_2 and u_2, respectively.

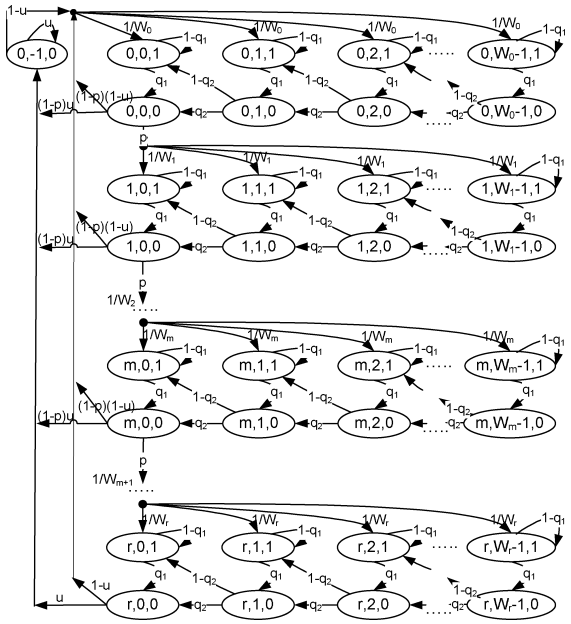

Fig. 5. Markov chain model for AC1 (Chain B), where W_x, r, m, p and u refer to $W_{x,1}$, r_1, m_1, p_1 and u_1, respectively.

Similarly for AC1, let $\beta_{i,j,k}(p,u,q_1,q_2)$ be the stationary distribution of *Chain B* given p, u, q_1 and q_2, we obtain

$$\beta_{0,0,0}(p,u,q_1,q_2) = \begin{cases} \left(\frac{(1+q_1-q_2)\iota_r+(1+q_1+q_2)\kappa}{q_1\xi}+\frac{u}{1-u}\right)^{-1}, & r\le m \\ \left(\frac{(1+q_1-q_2)(\iota_m+\nu)+(1+q_1+q_2)\kappa}{q_1\xi}+\frac{u}{1-u}\right)^{-1}, & r>m \end{cases} \quad (3)$$

where ξ, ι, κ, and ν are given by (2).

Using the developed Markov Chain results, in the following, we analyze the system performance. In this analysis, we restrict our study on only AC2 and AC1 for formulation clarity, although as in [22], a complete model for all four ACs of EDCA can be formulated. In our case, AC2 uses *Chain A* while AC1 uses *Chain B*. As in [22], the probabilities τ_2 (τ_1) that a node of AC2 (AC1) transmits in a slot is given by

$$\tau_i = \begin{cases} \frac{\left(1-p_i^{r_i+1}\right)\beta_{0,0,0}(p_i,u_i,q_1,q_2)}{(1-p_i)P_I}, & i=1 \\ \frac{\left(1-p_i^{r_i+1}\right)\alpha_{0,0}(p_i,u_i)}{1-p_i}, & i=2. \end{cases} \quad (4)$$

The probability τ_1 is conditioned on the probability that the previous slot is an idle slot as transmissions by AC1 nodes cannot occur after a busy slot. The probability P_I that a slot is idle is

$$P_I = q_1 P_B + q_2 P_I = \frac{q_1}{1+q_1-q_2} \quad (5)$$

where the probability P_B that a slot is busy given by $P_B = 1 - P_I$.

Let the number of nodes accessing the network of ACi be n_i where $i = 1, 2$, the collision probability p_i, considering also the virtual collision, can be written as

$$p_i = \begin{cases} 1-(1-\tau_i)^{n_i-1}\prod_{x\neq i}(1-\tau_x)^{n_x}, & i=1 \\ \left(1-\prod_{x\le i}(1-\tau_x)^{n_x-1}\prod_{x>i}(1-\tau_x)^{n_x}\right)P_B \\ +\left(1-\prod_{x\le i}(1-\tau_x)^{n_x-1}\prod_{x>i}(1-\tau_x)^{n_x}\right)P_I, & i=2. \end{cases} \quad (6)$$

Similar to [22], for AC1, q_1 and q_2 can be expressed by

$$\begin{aligned} q_1 &= (1-\tau_2)^{n_2} \\ q_2 &= \prod_{i=1}^{2}(1-\tau_i)^{n_i}. \end{aligned} \quad (7)$$

Equations (4)–(7) form a set of nonlinear equations that can be computed numerically where u_i and n_i are the inputs of the system. The input u_i describes the unsaturated probability of a node given a total of n_i statistical identical nodes of ACi in the network where $i = 1, 2$.

Three quantities of our main interest from the above analysis are P_I, P_C, and P_S describing the probabilities that a particular slot is idle, contains a collision, and carries a successful transmission, respectively. P_I is given in (5) and $P_C = 1 - P_I - P_S$. $P_S = P_{S1} + P_{S2}$ where P_{Si} is the probability that a slot carries a successful transmission from ACi. P_{Si} can be determined by

$$P_{Si} = \begin{cases} n_i\tau_i(1-\tau_i)^{n_i-1}\prod_{x\neq i}(1-\tau_x)^{n_x}, & i=1 \\ n_i\tau_i(1-\tau_i)^{n_i-1}\prod_{x\neq 1,i}(1-\tau_x)^{n_x}P_B \\ \quad +n_i\tau_i(1-\tau_i)^{n_i-1}\prod_{x\neq i}(1-\tau_x)^{n_x}P_I, & i=2. \end{cases} \quad (8)$$

Given the above results, two important performance measures, namely the throughput and the packet loss probability of ACi denoted by $U_i(u_2,n_2,u_1,n_1)$ and $L_i(u_2,n_2,u_1,n_1)$ can be computed by

$$\begin{aligned} U_i(u_2,n_2,u_1,n_1) &= \frac{P_{Si}E[P]}{P_I\sigma + P_S T_S + P_C T_C} \\ L_i(u_2,n_2,u_1,n_1) &= p_i^{r_i+1} \end{aligned} \quad (9)$$

where σ, $E[P]$, T_S, and T_C are the duration of an idle slot, the average duration of frame payload, successful transmission

TABLE II
IEEE 802.11e TIMING CONSTANTS FOR BASIC METHOD

Description	Value
Channel bit rate	11 Mbps
Idle slot duration, σ	20 μs
Successful transmission duration, T_S	$E[P] + 131.8182$ μs
Collision duration, T_S	$E[P] + 93.1818$ μs

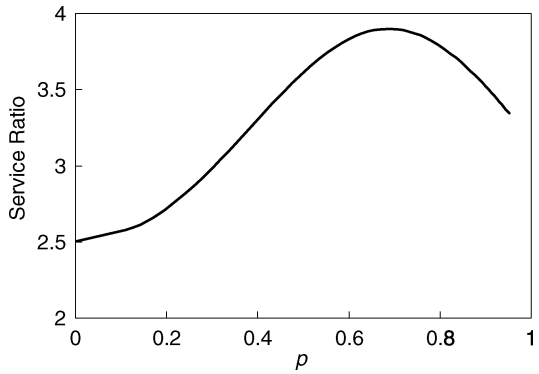

Fig. 6. Ratio of the throughput of AC2 to that of AC1 in an EDCA node under various network load conditions.

and collision, respectively. Table II summaries the duration constants according to the IEEE 802.11e standard. We consider only the basic access method due to our considered small video packet size in the practical situation. The detailed computation of the duration is available in [22]. With the above derivation of performance measures, we complete the description of our Markov Chain modeling for the IEEE 802.11e MAC protocol.

We first evaluate the service rates differentiation of the standardized IEEE 802.11e MAC protocol using the developed model. The service rates in different EDCA queues are mainly dictated by their contention window sizes and AIFSN settings. To analyze the service rate differentiation between AC2 and AC1, we compare the saturation throughput between AC2 and AC1 under various network load conditions. The saturation load condition is considered because it stresses the queue utilization and the channel usage, and threatens packet loss, which is of our main interest of study. The number of saturated nodes directly dictates the load of the network. Varying the number of saturated nodes gives a range of network load conditions. The network load condition can also be indicated by the collision probability, where with a low (high) number of saturated nodes, the collision probability that a packet suffers is low (high) indicating light (heavy) network load. Fig. 6 plots the service ratio of the throughput of AC2 to that of AC1 versus the collision probability of AC2. The collision probability of AC2 is a more accurate measure in this study because it counts only the external collision probability whereas that of AC1 includes also the internal collision which is termed the virtual collision in the IEEE 802.11e standard.

As shown in Fig. 6, the service ratio ranges from two to four under various load conditions. This means that AC2 is capable of transmitting two to four times faster than AC1 thus achieving service differentiation. This result provides a guideline for the design of mapping from the priority of scalable video traffic to the priority of the two considered EDCA queues. To optimize the queue usage in the system, it is necessary that packets in both queues be built or cleared up at a similar rate. Since our result as presented in Fig. 6 suggests that AC2 is capable to transmit more packets than AC1 in a given time interval, it is expected that more packets should be queued in AC2 than in AC1. Moreover, based on the EDCA queue metrics, AC2 and AC1 should serve high and low prioritized packets of the scalable video traffic, respectively.

However, maintaining the same utilization in the queues of AC2 and AC1 provides unbiased loss protection to the two queues since both queues are equally likely to drop packets under heavy load condition. To facilitate ULP in EDCA for the scalable video traffic, our mapping design considers a lower ratio of throughput of AC2 to that of AC1 than the service ratios presented in Fig. 6. Theoretically, any value below the presented curve may be adopted yielding different levels of inequality in loss protection. In order not to differ too far from the theoretical results so as to maintain certain optimality in queue utilization, we choose the service ratio two as our mapping strategy. In other words, with our design, AC2 is responsible for the top 2/3 scalable video traffic in terms of the priority, while AC1 is responsible for the remaining bottom 1/3 traffic. Since our mapping inserts lesser traffic to AC2 than the suggested optimum value, packet built-up rate in AC2 is slower, and hence packets queued in AC2 receive higher protection from overflow at the expense of the lower protection to AC1. It is thus expected that under heavy load conditions, overflow is likely to occur in AC1 before AC2. However, occurrence of overflow in AC2 is inevitable if the service ratio of the mapping remains unchanged. To tackle this problem, we further introduce an adaptiveness into our mapping design described in Section III-E to reduce the chances of overflow in AC2.

To further strengthen the ULP in EDCA, we follow the design in [21] that different maximum retry limit settings are applied to different EDCA queues. We demonstrate the effectiveness of this property in Fig. 7 by comparing the probabilities of packet loss as a result of maximum retry limit of the two queues under various network load conditions measured by the collision probability. As shown in the figure, especially in the range of small p values, the packet loss probability of AC2 remains closed to zero whereas that of AC1 becomes perceptible as early as $p = 0.2$. This comparison of packet loss rates between AC2 and AC1 also justify the effectiveness of the MAC-centric approach for the cross-layer design in [21].

Another important characteristics of the IEEE 802.11e MAC protocol is the congestion collapse phenomenon which occurs in all contention based protocols. Attempting to avoid collisions when their are likely to occur, the modern design in the IEEE 802.11e MAC protocol has eased but not eliminated congestion collapse from occurring. In the IEEE 802.11e MAC protocol operation, heavy load pushes up the chances of transmission collision on the channel. If load continues to increase, more transmission collisions will occur forcing the protocol to operate at a lower throughput level. To illustrate this phenomenon, in Fig. 8, we plot the channel capacity versus the number of nodes each transmitting 500 kbps traffic on AC2. The numerical computation for this study is described as follows where the throughput

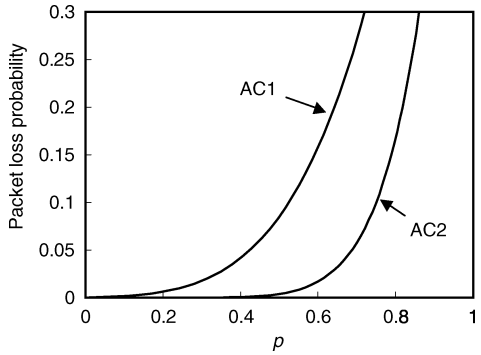

Fig. 7. Packet loss probability comparison between AC2 and AC1 in an EDCA node under various network load conditions.

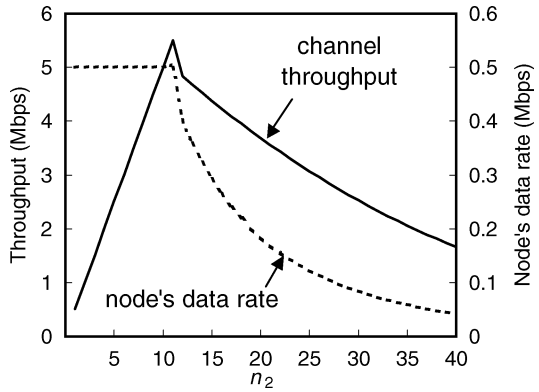

Fig. 8. Capacity of the IEEE 802.11e WLANs versus the number of nodes with each transmits 500 kbps.

for a network consisting n_2 nodes, $S(n_2)$ expressed in megabits per second, is given by

$$S(n_2) = \begin{cases} 0.5n_2, \exists u_2 \text{ such that } U_2(u_2, n_2, 0, 0) = 0.5n_2 \\ U_2(0, n_2, 0, 0), \quad \text{otherwise} \end{cases} \tag{10}$$

where u_2 is the probability in the range [0,1] that after its successful transmission, the queue of a node remains empty after either an idle or a busy slot duration. In Fig. 8, two curves are presented where the solid line describes overall channel throughput on the primary y axis while the dotted line draws throughput of each node on the secondary y axis. As can be seen from the figure, the current IEEE 802.11e with the considered system parameters can only support a maximum of 11 nodes transmitting 500 kbps each. Beyond this point, the offered load to the network exceeds that it can support, and hence all nodes enters the saturation state. Under the saturation load condition, further increase in the number of nodes pushes the protocol to operate at an even lower channel throughput. This clearly shows the IEEE 802.11e MAC protocol is not immune from congestion collapse.

Nevertheless, contention based protocols can be controlled to ease congestion collapse. Lam *et al.* presented one of the first classical papers in [23] discussing the stability of contention based protocols dated back in the mid-1970s. The IEEE 802.11e MAC protocol uses binary exponential backoff for this purpose, however, this backoff mechanism is aggressive and inaccurate in estimating the available bandwidth which causes high wastage due to collisions under the heavy load condition. Our aim is to achieve better congestion control through a micro and a macro rate control schemes so as to improve the stability of the IEEE 802.11e MAC protocol even under the heavy load condition.

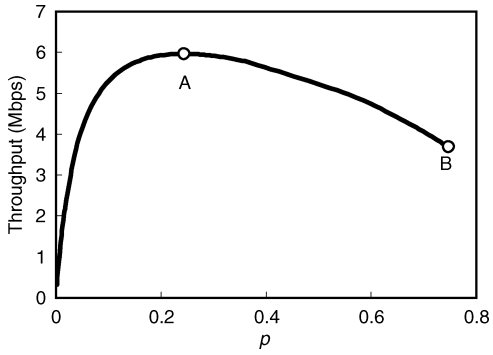

Fig. 9. Throughput versus collision probability for 20 active nodes.

It is indicated in Fig. 8 that the network enters saturation when the number of nodes increases higher than 11. In particular, for 20 nodes, the saturation throughput is around 3.6 Mbps (also indicated as point B in Fig. 9). Our next study varies u_2 and plots the throughput versus collision probability for 20 nodes on AC2, precisely, $U_2(u_2, 20, 0, 0)$ which is given in (9). Varying u_2 is equivalent to adjusting the source transmission rate of each node. The result is depicted in Fig. 9. As shown in the figure, the protocol may achieve as high as 6 Mbps (indicated as point A in Fig. 9). To achieve that, the collision probability must be maintained at a relatively low level, which is around 0.2 in this case. This operating point is achieved when each of the 20 nodes transmits 300 kbps. Comparing to the result given in Fig. 8, when each node attempts to transmit 500 kbps, the actual data rate for each node is only just below 200 kbps because each node suffers packet loss with a rate just over 300 kbps. Hence, controlling the source transmission rate helps reduce the collision probability leading to a higher operational throughput.

In summary, the above study has revealed several important characteristics of EDCA for cross-layer design, namely: 1) EDCA may facilitate ULP via a mapping scheme; 2) the ULP may be further enhanced by adjusting maximum retry limit settings for each EDCA queue as in [21]; and 3) despite the likelihood of congestion collapse, EDCA may operate at its highest achievable throughput by rate control in each source. Addressing to these findings, in our cross-layer design, we develop an adaptive mapping scheme and a rate control strategy in both the micro and the macro timescales accordingly.

B. Bandwidth Estimation in EDCA

Bandwidth estimation is an important component in our cross-layer design. The application layer and network sublayer depend on this estimation to achieve rate control and optimize QoS. The main role of the bandwidth estimation is to passively measure the network condition during a macro interval and provide estimation of available capacity that a node can access in the next macro interval. There are several methods proposed in the literature, from the earlier work of Cprobe [24], to some popular proposals of Spruce [25], Pathload [26], to some recent development of Probegap [27], wireless traffic probe [28] and

IdleGap [29]. While many of them deal with bandwidth estimation of a communication path in a wide area network (WAN), wireless traffic probe and IdleGap are specifically designed for the IEEE 802.11 DCF MAC protocol with different approaches.

Our design adopts IdleGap [29] to estimate the network bandwidth availability. However, the estimated available bandwidth will be shared among all nodes, thus the bandwidth accessible by each node is usually lower. To ensure that each node only utilizes its share of the available bandwidth, we introduce a simple method described as follows. We first notice that multimedia transmissions usually occurs in a form of continuous streams. The detection of a video packet from a node indicates the participation of multimedia transmissions. Hence, we propose that during a macro interval, not only a node estimates the network available bandwidth, $\tilde{R}$, using IdleGap, it also detects and counts the number of different nodes, $\tilde{N}$, that transmit AC2 packets. Using the two quantities, also including itself as a transmitting source, $\tilde{R}/(\tilde{N}+1)$ represents the estimated available bandwidth for a node in the next interval. If a node also transmitted $\tilde{T}$ video traffic during the current macro interval, then by the end of the current interval, the node estimates its available bandwidth, R, for the next interval simply by

$$R = \tilde{T} + \frac{\tilde{R}}{\tilde{N}+1}. \tag{11}$$

C. Optimal Bit Allocation in SVC

The problem of optimal bit truncation can be summarized as follows. Given the estimated bandwidth R from the network sublayer, how to optimally truncate each *T*-band JPEG2000 stream so that the overall distortion can be minimized. Given three types of scalabilities, there are many ways to allocate bits among *T*-bands by making trade off among frame rates, spatial resolution and signal-to-noise ratio (SNR) of individual frames. However, an optimal tradeoff among the three scalabilities is still an open question. In practice, this multiple scalability tradeoff problem is simplified due to the constraints of mobile devices. For example, spatial resolution is constrained by the screen size of mobile device and frame rate is limited by the processor speed. In this paper, we assume the spatial resolution is fully determined by the screen size of mobile devices, and we use a simple threshold-based approach to tradeoff among temporal scalability and SNR scalability. In particular, We mainly use the SNR scalability to match the estimated network bandwidth. Only at very low bit rate, when the individual frame quality becomes lower than a predefined peak SNR (PSNR) threshold, we reduce the frame rate to increase the quality of individual frames.

Considering each *T*-band JPEG2000 stream contains a number of quality layers whose corresponding rates and distortions are known, for simplicity we directly use these quality layers of JPEG2000 streams as the available truncation points although JPEG2000 provides even finer truncation points. In this way, we formulate the optimal bit truncation problem for the SNR scalability as

$$\min_{L_{B,C}} D_{\text{total}} = \sum_C \sum_B W_C \cdot G_B \cdot D_{L_{B,C},B,C} \tag{12}$$

subject to

$$R_{\text{total}} = \sum_C \sum_B R_{L_{B,C},B,C} < R_{\text{target}} - R_{MV} \tag{13}$$

where $L_{B,C}$ denote the quality layer L in the *T*-band B of the color component C, $R_{L_{B,C},B,C}$ and distortion $D_{L_{B,C},B,C}$ are the corresponding rate and distortion up to the quality layer $L_{B,C}$, R_{MV} is the bits consumed by motion vectors, and G_B is the energy gain associated for the *T*-band B and W_C is the weight ($W_C = 1$ by default) for each color component if we want to assign the YUV components with different priorities. G_B is determined by the filter coefficients and the level of MCTF that the *T*-band belongs to. For example, with four levels of MCTF and the common 5/3 filter, the energy gain is {5.06,2.43,1.62,1.08,0.72} for LLLL, LLLH, LLH, LH, and H band, respectively [30].

Similar to the optimal truncation used in JPEG2000 [16], this constrained minimization problem can be converted into an unconstrained problem by introducing a Lagrange multiplier λ, i.e.,

$$\min_{L_{B,C}} J = D_{\text{total}} + \lambda \cdot (R_{\text{total}} - R_{\text{target}}). \tag{14}$$

Note that not all the quality layers of each *T*-band are feasible truncation points because the R-D curve of the JPEG2000 stream is not strictly convex. Those unfeasible truncation points need to be removed so that the remaining truncation points will form a convex curve. In order to find the optimum λ for a given bit-rate R_{target}, we apply a binary searching method based on the rate and slope information, where the slope for a truncation point h of *T*-band B is calculated as

$$Slope_B^h = W_C \cdot G_B \cdot \frac{D_{h-1,B,C} - D_{h,B,C}}{R_{h,B,C} - R_{h-1,B,C}}. \tag{15}$$

We would like to point out that the resulted video bit stream by the above optimal bit allocation algorithm is actually a variable bit rate (VBR) bit stream with an average bit rate of R since the optimal bit allocation algorithm only minimizes the average distortion under the constraint of a total bit budget. A burst of packets may be generated at one time and passed to the MAC layer immediately for transmission. Since MAC layer is often implemented in firmware with very limited resources including buffer size, a burst of packets may cause unnecessary buffer overflow at the MAC layer. As such we apply the optimum bit allocation to each small time interval T_{interval} independently. In other words, the *T*-bands in each T_{interval} is allocated a total bandwidth of $R \cdot T_{\text{interval}}$. A leaky bucket is further applied to smooth out the bit rates allocated to each *T*-band. In this way, for an estimated network bandwidth R for a node, the application layer will generate a corresponding CBR-like bit stream to the bottom layer.

D. SVC Packetization With Relative Priority Index

For the packetized video transmission, *T*-band streams and MV data need to be assembled into individual network packets. In this paper, we simply put different color component into different packets and place motion vectors into the first packet of

their associated temporal high bands. The size of a packet is limited by a predefined maximum packet size P_{size}.

In order to provide adaptive QoS mapping in the network sublayer, it requires the application layer to provide the relative priorities of video packets. In this research, we apply the concept of relative priority index proposed in [9] to categorize different video packets. In particular, the authors in [9] focused on a specific type of RPI, relative loss index (RLI), and calculated the RLI by summarizing the normalized "video factors" (e.g., number of intra-coded MB, motion vector magnitude, etc) with appropriate weighting as

$$\text{RLI}_i = \sum_{n=1}^{N_{\text{VF}}} W^n \times \frac{\text{VF}_i^n}{m_i^n} \tag{16}$$

where N_{VF} stands for the number of video factors to be considered, W^n is the corresponding weight factors, VF_i^n is the magnitude of video factor VF activity n in the ith packet, and m_i^n is the sampling mean of VF^n up to the ith packet. The problem of such a general RLI model is that it can hardly approximate the actual loss impact in MSE. In [19], the authors extended their method to provide a more systematic RPI association by proposing a complicated MB-based corruption model. However, even with the new model, it is still not easy to estimate the loss impact of one packet to the MSE of the received video due to the drifting problem caused by the close-loop prediction in hybrid video codecs such as H.263.

It is relatively easy to approximate the packet loss impact for our developed SVC. This is because the open-loop structure of MCTF significantly reduces the effects of the drifting problem. In particular, the loss impact of a packet is defined as the corresponding distortion increase in reconstructed video in the case that the packet is lost while all other packets are correctly received. Mathematically, the loss impact of the ith packet in the T-band B of the color component C is calculated as

$$W_C \times G_B \times (D_{i-1,B,C} - D_{i,B,C}). \tag{17}$$

Moreover, in order to enable smooth adjustment of the QoS mapping in the network sublayer, we map the loss impact values obtained in (17) into integer RPI values (e.g., 0-63 with 8 bits representation), and also uniformly distribute the integer RPI values into different packets. The detailed procedure of generating RPI values is summarized as follows. First, we sort all the packets according to their calculated loss impact values in a decreasing order. Then, for each packet i, we identify its position Pos_i in the sorted list. Assuming the RPI values ranging from 0 to $M-1$, we define that the packets with smaller RPI value are more important. For a total number of N packets, we calculate RPI as

$$\text{RPI}_i = \lfloor Pos_i/N \times (M-1) \rfloor . \tag{18}$$

Fig. 10 shows an example of the loss impact results in MSE for different packets.[1] From the figure, we can see that some packets have extremely large loss impact. These packets are the first packets of each *T*-band stream, and their loss will cause the entire *T*-band undecodable.

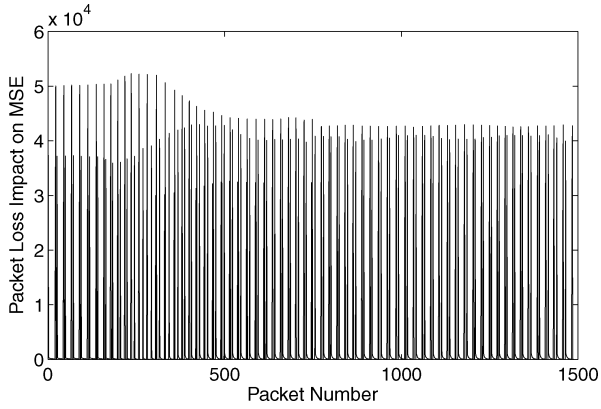

Fig. 10. Example of the loss impact results.

E. Adaptive QoS Mapping Algorithm

As another important task in the network sublayer, adaptive QoS mapping provides the essential ULP function to increase protection for important video traffic delivery at the expense of losing less important video traffic. The QoS mapping refers to the mapping from the priority of scalable video traffic to the priority of the two considered EDCA queues, AC2 and AC1. Mapping of cross-layer priorities allows the traffic to be treated differently for different levels of protection. Through the level of the inequality, the amount of sacrifices of the low priority traffic group in terms of packet drops can be induced, which in turn controls the total volume of traffic transmitted on the broadcast channel. With this approach, we achieve micro rate control in an indirect manner during a macro interval, and we ensure that dropped packets are most likely to occur at AC1.

Earlier our EDCA characteristics study reveals the service rate ratio of AC2 to AC1 and proposes the service ratio for the mapping to be set to two for ULP. As a result, 2/3 of the top prioritized scalable video traffic will be transmitted via AC2, and the remaining traffic will be transmitted via AC1. Packet drops are inevitable due to the variation in bandwidth usage of the IEEE 802.11e WLAN within a macro interval. If such event is not attended, under extreme conditions where the available bandwidth plunges dramatically, some level of packet drops may also occur at AC2 which may result in some significant drop in the video quality. It is therefore necessary to adjust the level of loss protection using packet drop event as an indicator by, for example, strengthening the loss protection at AC2 upon a detection of packet drops.

In the macro control, the application controls the source rate based on the bandwidth estimated from the network sublayer within an interval. With the advantage of knowing the total volume of generated traffic from the application within a macro interval, our adaptive QoS mapping attempts to maintain a two to one service ratio of AC2 to AC1 for the mapping based on the remaining traffic volume rather than the fixed total traffic volume. The mapping point hence floats and changes based

[1]We use the CIF Table-tennis sequence with two-level MCTF, a bit rate of 600 kbps, a packet size of 500 bytes and $T_{\text{interval}} = 32/30$ s.

TABLE III
DESCRIPTION OF ADAPTIVE QOS MAPPING ALGORITHM

Let $SRPI$ denote the set of RPI values.
Let $\hat{v_i}$ denote the remaining traffic volume of each RPI (in bits).
Let Λ denote the mapping pointer where packets with $RPI \leq \Lambda$ is inserted to AC2 queue, otherwise, they are inserted to A1 queue.

procedure ComputePointer($\hat{v_i}$)
  $\hat{v} := \Sigma_i \hat{v_i}$
  return $\max(i)$ such that $\Sigma_i \hat{v_i} \leq \frac{2}{3}\hat{v}$
end

procedure PacketDepartureEvent(i, s)
  // Packets departing the network may be due to either a successful transmission or a loss.
  // i is the RPI of the departing packet.
  // s is the size (in bits) of the departing packet.
  $\hat{v_i} := \hat{v_i} - s$
  Λ = ComputePointer($\hat{v_i}$)
end

procedure StartMacroIntervalEvent(c)
  // c is a vector containing the expected traffic volume of each RPI during a macro interval.
  for each i **in** $SRPI$ **do**
    $\hat{v_i} := c_i$
  end for
  Λ = ComputePointer($\hat{v_i}$)
end

on the situation of the network condition. Our proposed QoS mapping algorithm is detailed in Table III.

For the sake of understanding, we provide the illustration with a simple example to show the adaptiveness of the mapping adjustment. In Fig. 11, curve (i) describes the initial distribution of RPI at the start of a macro interval with a total volume of, say 500 kbits. After a certain period of time, some packets are successfully transmitted and some are dropped due to congestion. Since the ULP is in place, the packet drops should occur for those with higher RPI indicating lower importance. Curve (iii) shows the remaining traffic volume whereas curve (ii) depicts that without the packet drops for comparison. Since the distribution curve has lower tail, after that period, the new mapping pointer covering the 2/3 of the area measured from the left under curve (iii) shifted leftward to a lower RPI value according to our design. With this adjustment, lesser packets will be inserted to AC2, and its queue will experience overflow with an even lower probability, hence achieving higher loss protection. This adjustment of mapping pointer is further illustrated in an experiment presented in Section IV showing the adaptiveness of the QoS mapping.

IV. EXPERIMENTAL RESULTS

In this section, we present the performance of the 3-D wavelet SVC over the IEEE 802.11e WLANs. In our experiments, we consider the first 256 frames of the "Table-tennis" CIF sequence as our video source. The sequence is repeatedly transmitted for simulation time requirement. Two levels of MCTF decomposition is used during encoding of the video. The maximum video packet size is set to 500 bytes. The RPI ranges between 0 and 63. The macro interval is set to 32/30 s. We take the Y component PSNR as our video quality measurement.

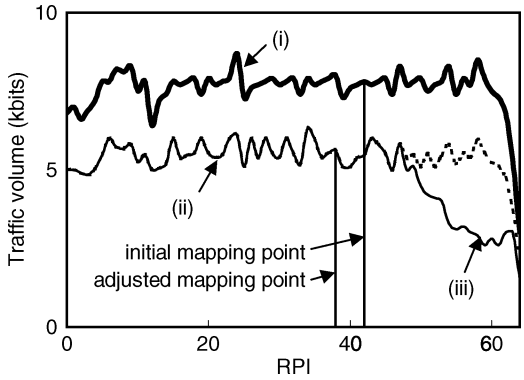

Fig. 11. Illustration of the proposed adaptive mapping scheme.

The experiments are conducted using ns2. The IEEE 802.11e settings follow the latest standard presented in Table I. The performance is also compared with DCF and EDCA. The former refers to the legacy DCF with setting following the IEEE 802.11b standard [2], and the latter refers to the EDCA using AC2 for video transmission with no cross-layer consideration. Our considered network is a single hop private infrastructure WLAN. This setup allows us to focus solely on the WLAN. Two scenarios are performed to show the behavior and the benefits of our cross-layer design.

A. Case 1: Soft Capacity Illustration

This experiment starts with a node, say node-0, transmitting a certain rate of video traffic. For every four macro interval duration, a new video node is added to the WLAN, until the number of nodes reaches 15 in the WLAN. In addition to the "Table-tennis" CIF sequence, here we also conduct experiment using "foreman" CIF sequence to test the sensitivity of our scheme on a different content.

For our proposed scheme, we consider that the macro rate control ranges from a minimum of 500 kbps to a maximum of 1 Mbps. A comparison is made with the performance results from that of DCF and EDCA. For PSNR performance comparison, the sent video PSNR is also measured as a benchmark. This PSNR measure represents the optimal PSNR.

Figs. 12 and 13 plot the PSNR performance measured at node-0 under DCF and EDCA, respectively. Two fixed source rates, which are 500 kbps and 1 Mbps, are used. As can be seen, DCF supports merely three (seven) nodes for the case of 1 Mbps (500 kbps) of video source rate before the PSNR plunges quickly due to network congestion. A simple calculation reveals that the maximum achievable network throughput is no more than 4 Mbps. Similar observation is made for EDCA as indicated in Fig. 13. The results also show the maximum achievable network throughput of no more than 4 Mbps. This result is consistent with our earlier analytical finding in Fig. 9 that when source rate is left uncontrolled, the network may be pushed to saturation where the protocol operates at a lower throughput level.

For our scheme, since the macro rate control is in place, each node varies its source rate according to the estimated available bandwidth from no less than 500 kbps to no more than 1 Mbps. A node always starts from the maximum specified rate

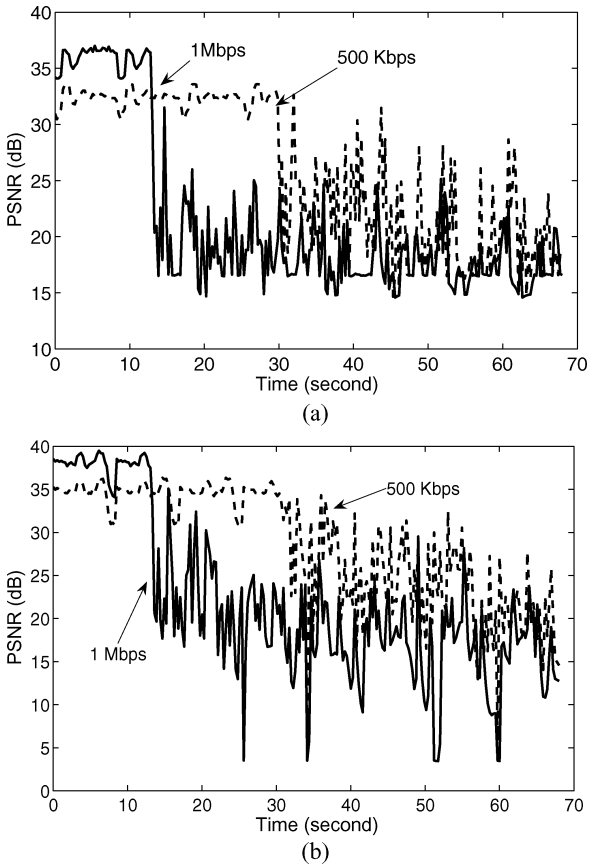

Fig. 12. PSNR performance of received video for DCF. (a) The table tennis clip. (b) the foreman clip.

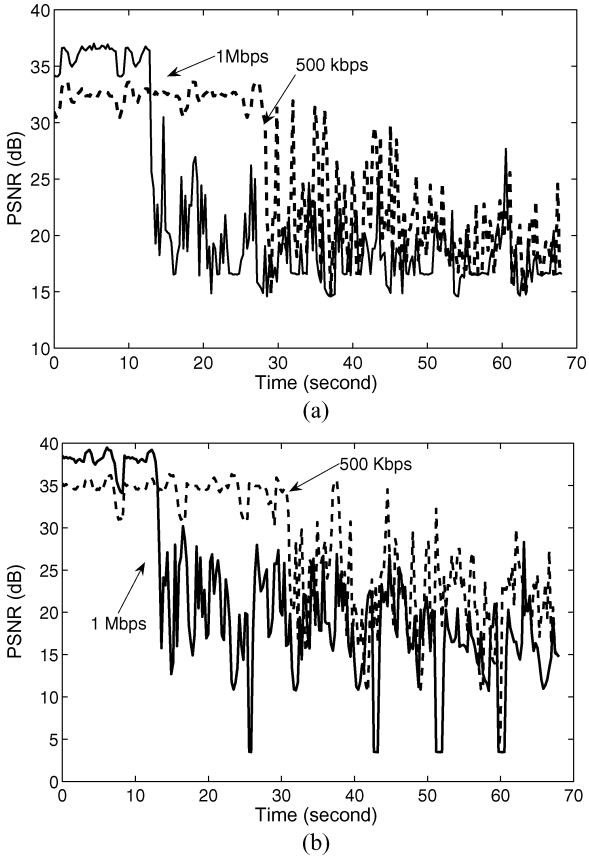

Fig. 13. PSNR performance of received video for EDCA. (a) The table tennis clip; (b) the foreman clip.

of 1 Mbps. The performance result is presented in Fig. 14. An immediate comparison among Figs. 12 and 13 show the performance advantage of our cross-layer design. The soft capacity property of our scheme is also clearly illustrated. As can be seen from Fig. 14, when the network consists of a small number of nodes, each node receives high quality video streams. As the simulation progresses, the number of nodes increases, all nodes adapt themselves to the changed network condition to accommodate more nodes by reducing the sending rate. Our earlier study of EDCA capacity presented in Fig. 8 indicates that EDCA may only support 11 nodes transmitting 500 kbps each, even though we specify our minimum macro rate to 500 kbps, the micro rate control preemptively drops video packets of low importance via AC1 so that the quality of the video streams maintained even at 15 nodes. In terms of the PSNR performance between the two different sequences, similar results are reported. We see that our scheme always provides better PSNR performance than the other two methods.

To further illustrate the performance advantages of our scheme, we compare the packet loss ratio (in percentage) in Figs. 15–17. As shown in the figures, DCF and EDCA share the similar behavior in terms of packet loss ratio. With no cross-layer consideration, important video packets may be dropped and hence quality of video streams cannot be maintained when packet loss occurs. On the other hand, the ULP in our adaptive QoS mapping protects video packets on AC2 by sacrificing packets on AC1. There is hardly any loss experienced at AC2. Fig. 18 depicts the adaptiveness of the mapping. The mapping pointer is reset at the beginning of every macro interval, but quickly adapts downwards to a lower value for a greater protection to AC2 when the packet loss occurs. This operation happens more frequently as the network approaches congestion.

We further plot the packet loss ratio from the application's point of view in Fig. 19. For clarity of presentation, instead of showing the packet loss ratio of all 64 RPIs, we partition them into three groups and compare the packet loss ratio among groups. As can be seen, groups of lower importance suffers high packet loss compared to that of higher importance.

Fig. 20 shows the action of the macro control in our cross-layer design. We see that as more nodes are added to the network intensifying the contention for bandwidth, the detected available bandwidth drops, and hence the source rate at node-0 slowly decreases from the initial maximum rate of 1 Mbps to the eventual minimum rate of 500 kbps.

B. Case 2: Competition With Best Effort Traffic

Since our scheme also uses AC1 for the transmission of video traffic, where AC1 is designed for best effort traffic, the study of performance in the presence of best effort is necessary to show the minimum influence of our scheme to the concurrent best effort traffic transmission. In this experiment, we consider four video sources, each with a fixed sending rate of 700 kbps. For

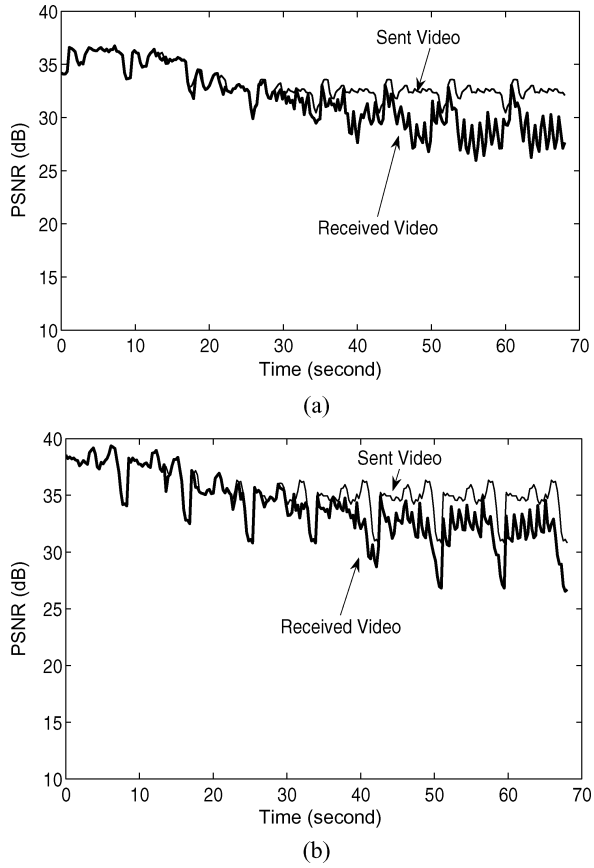

Fig. 14. PSNR performance of received video for our cross-layer design. (a) The table tennis clip; (b) the foreman clip.

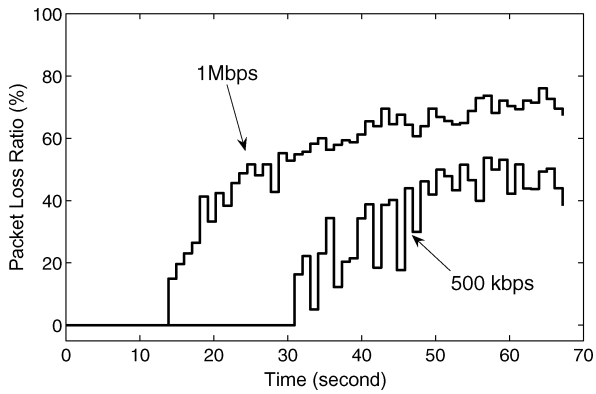

Fig. 15. Packet loss ratio for DCF.

our scheme, the minimum and maximum rates at the application layer are set to 700 kbps preventing a node from adjusting. This achieves consistency among all schemes for a fairer comparison.

Moreover, the simulation duration is set to 14 s. Best effort traffic in a form of CBR packets at rate 600 kbps is injected at each video node into the network at five seconds after the simulation has started. Fig. 21 shows the PSNR comparison among DCF, EDCA and our scheme, with the sent PSNR as the benchmark. As can be seen, while others suffer quality drop when best effort traffic is introduced into the network, the best effort traffic has little influence to our scheme. In fact, packet loss does occur in our scheme, however, due to the ULP, dropped packets are of

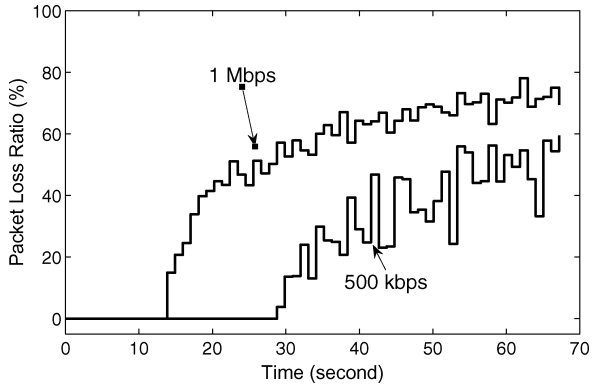

Fig. 16. Packet loss ratio for EDCA.

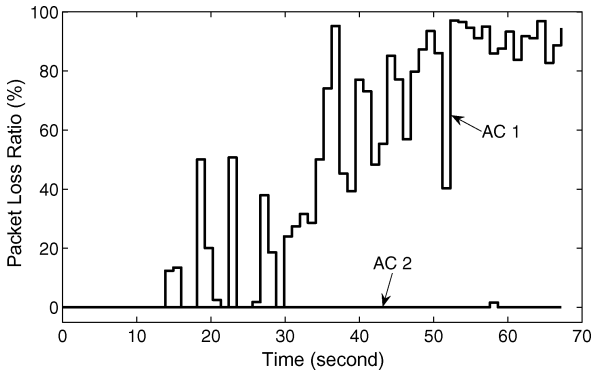

Fig. 17. Packet loss ratio for our cross-layer design.

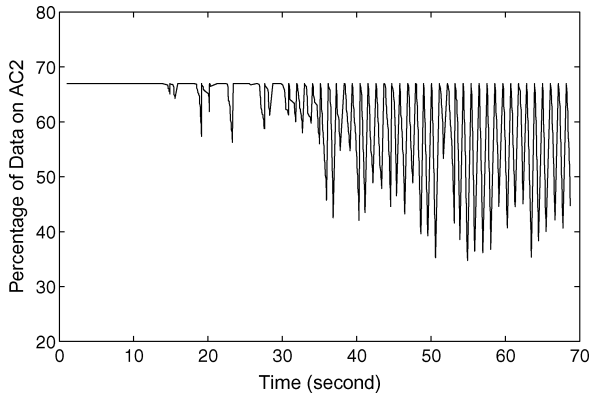

Fig. 18. Illustration of the QoS adaptive mapping in our cross-layer design.

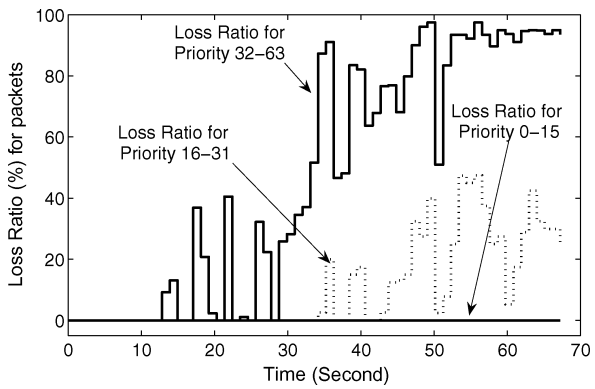

Fig. 19. Packet loss ratio of video packets observed at the video client.

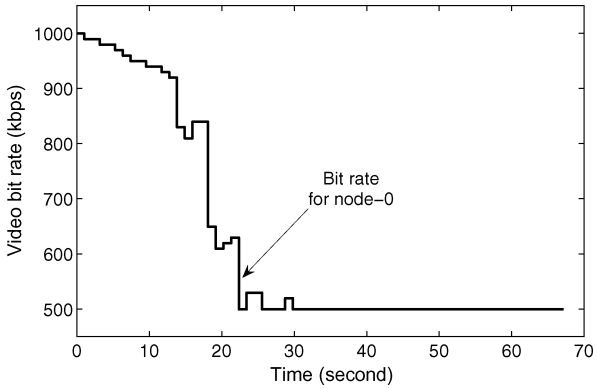

Fig. 20. Illustration of macro rate control in our cross-layer design.

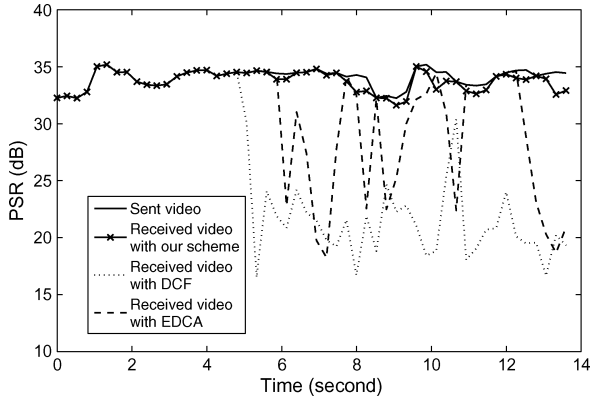

Fig. 21. Comparison of PSNR for various schemes for simultaneous transmission of video and best effort traffic.

low importance, and hence impact to the quality is kept minimum.

For the best effort traffic, from the simulation, we recorded that 394, 184, and 211 kbps of traffic per node was transmitted in DCF, EDCA and our scheme, respectively. Comparing the best effort traffic volume, DCF ranked the highest. This is because there is no differentiation between the video and best effort traffic in DCF, hence direct competition between the two traffic types occur with each receives its fair share. On the other hand, we notice that EDCA ranked the lowest. This is because EDCA parameter settings provide much higher TXOP to AC2 compared with that of AC1 which chokes traffic of AC1. Our scheme, although does not allow as much best effort traffic as that of DCF to be transmitted in the network, offers larger volume of best effort traffic to be transmitted than that of EDCA. This shows a higher friendliness of bandwidth sharing for our scheme compared to that of the IEEE 802.11e standard.

V. Conclusion

In this paper, we introduce a cross-layer design that achieves all timescales rate control for optimizing 3-D wavelet scalable video transmission over the recently standardized IEEE 802.11e networks. We studied thoroughly the characteristics of the IEEE 802.11e MAC protocol. Based on our findings on the analysis, we designed a rate control scheme that offers all timescales rate control. Four important components were developed, namely: 1) the network estimation component that estimates the available bandwidth; 2) the optimal bit allocation component that achieves minimum distortion given the bandwidth constraint obtained from the network estimation component; 3) the SVC packetization with RPI that marks the importance of video packets; and, finally, 4) the adaptive QoS mapping component that provides different levels of loss protection to different prioritized video packets. The interaction of the four components provide a macro timescale and a micro timescale rate control schemes.

Computer simulation was carried out to investigate the performance of our proposed design. Experimental results show that our scheme outperforms all other considered methods including DCF and the standardized EDCA with no cross-layer consideration. Moreover, the results also indicate the optimality of our design as it maintains a high received video PSNR that is closed to the measured sent video PSNR.

In addition, the soft capacity property of our design was illustrated showing that using the proposed rate control scheme, our design can accommodate more users by lowering the PSNR of each user. The macro rate control, ULP and adaptive QoS mapping were also demonstrated. Moreover, in the presence of best effort traffic, the experimental results further suggest that our design achieves better friendliness of bandwidth utilization than the current IEEE 802.11e standard.

References

[1] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 8: Medium Access Control (MAC) Quality of Service Enhancements*, IEEE Std 802.11e-2005, 2005.

[2] *IEEE Standard for Information Technology—Telecommunications and Information Exchange Between Systems—Specific Requirements Part 11: Wireless LAN MAC and PHY Specifications*, IEEE Std 802.11-1999, 1999.

[3] M. Karczewisz and R. Kurceren, "The SP- and SI- frames design for H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, pp. 637–644, Jul. 2003.

[4] A. Vetro, C. Christopoulos, and H. Sun, "Video transcoding architectures and techniques: An overview," *IEEE Signal Process. Mag.*, vol. 20, no. 2, pp. 18–29, Mar. 2003.

[5] J.-R. Ohm, "Advances in scalable video coding," *Proc. IEEE*, vol. 93, no. 1, pp. 42–56, Jan. 2005.

[6] P. A. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media," *Microsoft Research Technical Report*, Feb. 2001.

[7] X. K. Yang *et al.*, "A degressive error protection algorithm for MPEG-4 FGS video streaming," in *Proc. IEEE ICIP*, Rochester, NY, 2002, pp. 737–740.

[8] C. Liu and S. S. Chen, "Providing unequal reliability for transmitting layered videostreams over wireless networks by multi-ARQ schemes," in *Proc. IEEE ICIP*, 1999, vol. 3, pp. 100–104.

[9] J. Shin, J. W. Kim, and C. C. J. Kuo, "Quality-of-service mapping mechanism for packet video in differentiated services network," *IEEE Trans. Multimedia*, vol. 3, no. 2, pp. 219–231, Jun. 2001.

[10] H. Zhang, X.-G. Xia, Q. Zhang, and W. Zhu, "Precoded OFDM with adaptive vector channel allocation for scalable video transmission over frequency-selective fading channels," *IEEE Trans. Mobile Comput.*, vol. 1, no. 4, pp. 132–141, Apr. 2002.

[11] Q. Zhang, Z. Ji, W. Zhu, and Y.-Q. Zhang, "Power-minimized bit allocation for video communication over wireless channels," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 398–410, Jun. 2002.

[12] C. I. Podilchuk, N. Jayant, and N. Farvardin, "Three-dimensional subband coding of video," *IEEE Trans. Image Process.*, vol. 4, no. 2, pp. 125–139, Feb. 1995.

[13] J. R. Ohm, "Three-dimensional subband coding with motion compensation," *IEEE Trans. Image Process.*, vol. 3, no. 5, pp. 559–571, Sep. 1994.

[14] A. Secker and D. Taubman, "Motion-compensated highly scalable video compression using an adaptive 3-D wavelet transform based on lifting," in *Proc. IEEE ICIP*, 2001, vol. 2, pp. 1029–1032.

[15] D. S. Taubman and M. W. Marcellin, "JPEG2000: Standard for interactive imaging," *Proc. IEEE*, vol. 90, no. 8, pp. 1336–1357, Aug. 2002.
[16] D. Taubman, "High performance scalable image compression with EBCOT," *IEEE Trans. Image Process.*, vol. 9, no. 7, pp. 1158–1170, Jul. 2000.
[17] M. van der Schaar, S. Shankar, and N. , "Cross-layer wireless multimedia transmission: Challenges, principles, and new paradigms," *IEEE Wireless Commun.*, pp. 50–58, Aug. 2005.
[18] D. Wu, T. Hou, and Y.-Q. Zhang, "Transporting real-time video over the Internet: Challenges and approaches," *Proc. IEEE*, vol. 88, no. 12, pp. 1855–1875, Dec. 2000.
[19] J.-G. Kim, J. Shin, J. W. Kim, and C. C. J. Kuo, "Coordinated packetlevel protection with a corruption model for robust video transmission," in *Proc. VCIP*, 2001, pp. 410–421.
[20] W. Kumwilaisak, Y. T. Hou, Q. Zhang, W. Zhu, C.-C. J. Kuo, and Y.-Q. Zhang, "A cross-layer quality-of-service mapping architecture for video delivery in wireless network," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 10, pp. 1685–1698, Dec. 2003.
[21] A. Ksentini, M. Naimi, and A. Gueroui, "Toward an improvement of H.264 video transmission over IEEE 802.11e through a cross-layer architecture," *IEEE Commun. Mag.*, vol. 44, no. 1, pp. 107–114, Jan. 2006.
[22] J. W. Tantra, C. H. Foh, and A. B. Mnaouer, "Throughput and delay analysis of the IEEE 802.11e EDCA saturation," in *Proc. IEEE ICC*, 2005, vol. 5, pp. 3450–3454.
[23] S. S. Lam and L. Kleinrock, "Packet-switching in a multi-access broadcast channel: Dynamic control procedures," *IEEE Trans. Commun.*, vol. COM-23, no. 4, pp. 410–423, Sep. 1975.
[24] R. L. Carter and M. E. Crovella, "Dynamic server selection using bandwidth probing in wide-area networks," Comp. Sci. Dep., Boston Univ., MA, Tech. Rep. BU-CS-96-007, Mar. 1996.
[25] J. Strauss, D. Katabi, and F. Kaashoek, "A measurement study of available bandwidth estimation tools," in *Proc. ACM IMC*, 2003, pp. 39–44.
[26] M. Jain and C. Dovrolis, "Pathload: A measurement tool for end-to-end available bandwidth," presented at the Passive Active Meas., Mar. 2002.
[27] K. Lakshminarayanan, V. N. Padmanabhan, and J. Padhye, "Bandwidth estimation in broadband access networks," presented at the Proc. ACM IMC, 2004.
[28] M. Davis, "A wireless traffic probe for radio resource management and QoS provisioning in IEEE 802.11 WLANs," in *Proc. ACM Int. Symp. Modeling, Anal. Simul. Wireless Mobile Syst.*, 2004, pp. 234–243.
[29] H. K. Lee, V. Hall, K. H. Yum, K. I. Kim, and E. J. Kim, "Bandwidth estimation in wireless LANs for multimedia streaming services," in *Proc. IEEE ICME*, 2006, pp. 1181–1184.
[30] R. Xiong, J. Xu, F. Wu, S. Li, and Y. Q. Zhang, "Optimal subband rate allocation for spatial scalability in 3-D wavelet video coding with motion aligned temporal filtering," in *Proc. VCIP*, 2005, pp. 381–392.