Adaptive Transmission of Multi-Layered Video over Wireless Fading Channels

Jun Xu, Xuemin (Sherman) Shen, *Senior Member, IEEE,* Jon W. Mark, *Life Fellow, IEEE,* and Jun Cai, *Member, IEEE*

***Abstract*— In this paper, adaptive transmission of scalable multi-layered video with quality of service (QoS) assurance over wireless channels is studied. By properly formulating the channel fading as a finite state Markov channel (FSMC) model, three rate adaptation schemes, namely, assured-rate-allocation, neighbor-interleaving, and swing-loaded schemes, are proposed to exploit the inherent multiplexing gain. An analytical model for QoS performance evaluation of video transmission over time-varying erroneous channels is derived. The accuracy of the analytical model is validated by simulations. Analytical and simulation results demonstrate that the proposed rate adaptation schemes can effectively improve the channel utilization and the system throughput.**

***Index Terms*— Multi-layered video, channel adaptive rate allocation, QoS provisioning, differentiated QoS, Rayleigh fading, finite state Markov channel (FSMC), wireless resource allocation.**

I. Introduction

WIRELESS Internet video services are expecting to be extensively deployed with the increasing real-time Internet video applications and the fast evolution of wireless networks. Unlike voice and data services, video applications consume relatively large bandwidth and require heterogeneous quality of service (QoS) requirements in terms of delay, bit error rate (BER) and video quality. However, providing static QoS satisfaction for video applications over wireless Internet is difficult due to the scarce radio resource and the dynamics of the available system capacity [1]. Therefore, statistical QoS becomes necessary since otherwise the resource will be extremely underutilized. Recently, multi-layered scalable video encoding has been used at the application layer, which provides opportunity to fast adapt a video source to the time-variant system capacity [1], [2]. How to adaptively transmit layered video streams over wireless fading channels with satisfactory QoS and efficient wireless resource utilization becomes an open and challenging research issue.

For a multi-layered video source, each layer may have the same QoS requirements and all layers can be transmitted together. This is called *integrated QoS provisioning*. On the other hand, different layers may have different QoS requirements, and resource allocation should consider each layer separately with *differentiated QoS provisioning*. There are a number of papers in the literature reporting on delivering video streams or multimedia traffic over wireless time-varying channels [3]–[13]. In [3], a Markov model based analytical approach is proposed. The model is characterized by switched batch Bernoulli processes with finite buffer size and has memory space specified by the buffer size. A limitation is that the Markov space may be too large if the buffer size is large in real-time applications. Moreover, because of the variable capacity, the QoS parameter obtained from the finite buffer model cannot be directly related to the delay requirement. In [4], an analytical upper bound for delay outage is obtained. Since the upper bound is kind of loose, designing rate adaption strategies based on such upper bound may not be efficient. A rate allocation scheme for optimal joint source and channel coding and an adaptive resource allocation scheme for scalable video transmission are proposed in [5] and [6], respectively. However, both schemes are based on the criterion of minimum-distortion or minimum-power consumption. An analytical model for rate allocation and QoS performance analysis is proposed in [7], where the partitioning of the Markov channel model has to match the traffic pattern. In [8], adaptive video transmission is discussed for wideband code division multiple access (WCDMA) systems. However, rate adaptation for layered video over time-varying wireless channels is not discussed. The simulation study of QoS performance is the focus of [9]–[11], which, however, may not fully reveal the relation between QoS provisioning and channel adaptive rate allocation. Another common approach is an *effective capacity* (EC) scheme [12], [13]. The main idea of the EC scheme is to impose a constraint on the source rate based on the QoS requirements and the time-varying channel capacity. The QoS requirements can be guaranteed if the source rate does not exceed the imposed constraint. However, the EC scheme is conservative in the sense that it overlooks the potential multiplexing gain and therefore may underutilize the system resource.

In this paper, adaptive transmission of scalable multi-layered video with QoS assurance over wireless channels is studied. The system employs adaptive channel coding in which the code rate adapts to the channel dynamics. With adaptive channel coding, the channel fading process is partitioned into a finite number of states at the code switching points, resulting in a Markov service rate process. We then propose three rate

Manuscript received October 21, 2005; revised April 13, 2006; accepted September 30, 2006. The associate editor coordinating the review of this paper and approving it for publication was S. Aissa. This work was supported by research grants from Bell University Laboratories (BUL) under the sponsorship of Bell Canada, and the Natural Science and Engineering Research Council of Canada under Strategic Project grant no. STPGP 257682.

J. Xu, X. Shen, and J. W. Mark are with the Department of Electrical and Computer Engineering, University of Waterloo, Ontario, Canada (e-mail: {j5xu, xshen, jwmark}@bbcr.uwaterloo.ca).

J. Cai is with the Department of Electrical and Computer Engineering, University of Manitoba, Manitoba, Canada (e-mail: jcai@ee.umanitoba.ca).

Digital Object Identifier 10.1109/TWC.2007.05838.

adaptation schemes, namely, assured-rate-allocation, neighbor-interleaving, and swing-loaded, for a single layer transmission by exploiting the inherent multiplexing gain. A novel analytical model is developed and a closed form packet loss rate due to delay and channel error is derived. After that, joint QoS provisioning among different layers, namely, differentiated and integrated QoS provisioning, is further discussed.

The major contribution of this paper is two-fold: i) an analytical model is derived for QoS performance analysis; ii) adaptive rate allocation schemes are proposed for multi-layered video transmission over wireless channels with QoS assurance, which can effectively improve the channel utilization and video throughput.

The remainder of the paper is organized as follows. Section II presents the system model which captures the features of layered video transmission, adaptive channel coding, and the propagation channel model. In Section III, channel partition and efficient rate adaptation schemes are proposed, and an analytical model is derived for QoS performance analysis. Section IV discusses differentiated and integrated QoS provisioning. Numerical results are given in Section V, followed by conclusions in Section VI.

II. System Model

With MPEG-4 codec, a video source can be encoded into a base layer (BL) which contains the most important information, and an enhanced layer (EL) which provides additional information for better video quality [2]. The EL can be further split into several sub-layers. The layered video can be structured to adaptively fit the variable capacity due to fluctuations resulting from channel fading, mobility, etc. In a layered video, different layers may have heterogeneous QoS requirements depending on the importance of the information they are involved. In general, the BL has a more stringent QoS requirement than the ELs since the correctness of the ELs is based on the correct reception of the BL. In this paper, a higher priority and more stringent QoS requirements are set to the BL so that the dropping of the ELs due to error in the BL is not detrimental.

Fig. 1 shows the system model of adaptive multi-layered video transmission over a downlink wireless fading channel. The full-layered video stream, which is received by the base station (BS) from the source node via the core network, is adaptively transmitted to a mobile station (MS) through the wireless medium. In the BS, the packets of BL and ELs are stored in two separate transmit buffers before transmission[1]. Both BL and ELs maintain their own buffers and manage their own link layer retransmissions. The erroneous packets will be selectively retransmitted by assuming that the acknowledgement can be received in one frame time such that retransmission can be carried out in the next frame. The packets successfully received within the delay bound are then decoded at the receiver. Rate adaptation primarily satisfies the transmission of the BL while the ELs are adaptively transmitted subject to the availability of the residual capacity.

[1]Usually the base and enhanced layers are not independent. We assume that the dependency can be recovered at the receiver buffer.

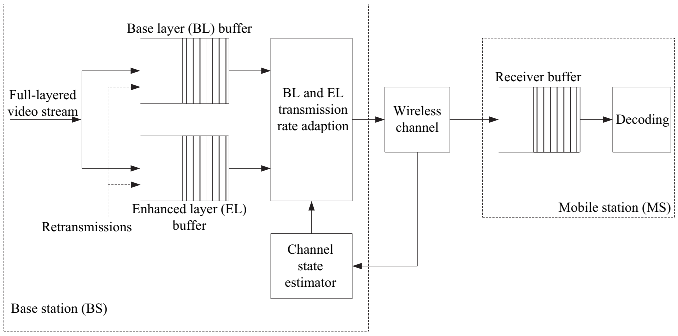

Fig. 1. Adaptive multi-layered video transmission over the wireless medium.

Adaptive forward error correction (FEC) is applied. The number of error correction bits is adaptively chosen to match the channel BER so that the throughput is maximized [16], [17]. In this paper, without loss of generality, an (n, k) BCH linear block code with sufficient interleaving is considered, where n is the block length, k is the number of information bits, and k/n is the *code rate*. When BCH is strictly used for error correction on a channel with BER p, the probability of decoding error is upper bounded by [19]

$$P_E \leq \sum_{i=t+1}^{n} \binom{n}{i} p^i (1-p)^{n-i} \tag{1}$$

where t is the random-error-correction capacity of the code. Let each link layer packet consist of n_1 code blocks. The link layer packet error rate becomes

$$PER_{link} = 1-(1-P_E)^{n_1} = 1-(\sum_{i=0}^{t} \binom{n}{i} p^i (1-p)^{n-i})^{n_1}, \tag{2}$$

and the information throughput can be calculated as

$$U = \frac{k}{n}(1-P_E)^{n_1}. \tag{3}$$

Fig. 2 shows the throughput of an adaptive FEC with $n = 255$, $n_1 = 20$ and different k values. It can be seen that as the channel BER increases, the throughput decreases for fixed (n, k). However, if k can be switched at certain BER values, such as the crossover points shown in the figure, the overall throughput can be significantly improved. Details on how to choose such switching points will be provided in the next section.

The wireless Rayleigh fading channel can be represented as a finite state Markov channel (FSMC) model [14], [15]. In FSMC, the received signal-to-noise ratio (SNR) is partitioned into N states. Transitions can exist between adjacent states or spread over the whole state space depending on the partition scheme used and the mobile speed. The steady-state probability of the ith state is $\pi_i = \int_{\gamma_i}^{\gamma_{i+1}} p(\gamma)d\gamma$, $i = 1, ..., N$, where γ_i is the SNR threshold for the ith state and $p(\gamma)$ denotes the probability density function of SNR. If the transitions happen only between adjacent states, the transition probabilities can be approximated as [14]

$$p_{i,i+1} \approx \frac{A(\gamma_{i+1})T_m}{\pi_i}, \quad i = 1, ..., N-1 \tag{4}$$

$$p_{i,i-1} \approx \frac{A(\gamma_i)T_m}{\pi_i}, \quad i = 2, ..., N \tag{5}$$

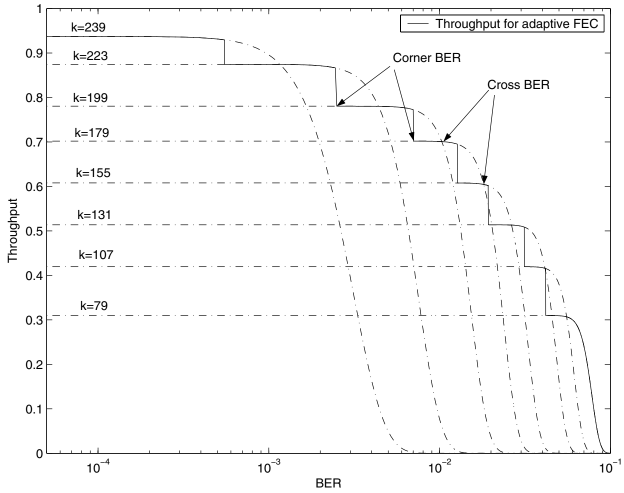

Fig. 2. Throughput for adaptive FEC using BCH (n,k) code, n=255.

where T_m is the frame length and $A(\gamma_i)$ is the level crossing rate, which can be calculated in one direction as

$$A(\gamma) = \sqrt{\frac{2\pi\gamma}{\gamma_0}} f_m \exp(-\frac{\gamma}{\gamma_0}). \quad (6)$$

In (6), f_m denotes the maximum Doppler frequency and γ_0 is the average SNR.

III. Adaptive Rate Allocation

In this section, adaptive rate allocation schemes are developed for a multi-layered video transmission with efficient channel utilization and QoS assurance. For a multi-layered video source, the total assigned bandwidth, referred to as *system bandwidth*, for both information and error correction bits is assumed fixed. Since the number of error correction bits may be increased or decreased subject to channel fluctuations, the bandwidth allocated for information bits varies accordingly. Assume channel information is available at the BS and the code rate is switched on a frame-by-frame basis. In what follows, we first present two channel partition methods for FSMC generation and reveal the characteristics of the resultant Markov service rate process (MSRP). We then discuss the channel efficiency and QoS assurance for BL transmission, and provide a QoS performance analysis model which can be generalized for BL, ELs and integrated BL/EL transmissions.

A. Partition of the FSMC

The partition of an FSMC is quite flexible and subject to specific applications. The partitioning criteria can be based on equal steady-state probability or equal residence time duration of each state [14], [15]. In this paper, we consider a new partition criterion that maximizes the average throughput. Without loss of generality, we consider an adaptive BCH block coding with nearly equal capacity interval, i.e., $\frac{k_{i+1}-k_i}{n} \simeq \frac{k_i-k_{i-1}}{n}$, where k_i/n is the code rate in the ith state.

As shown in Fig. 2, switching the code rate at clearly defined crossover points produces a best overall throughput [16]. The values of BER corresponding to crossover points are referred to as *cross-BERs*. However, for real-time video traffic with a stringent delay bound, only a limited number of retransmissions are allowed. Since the PER near cross-BERs may be relatively large, the increasing number of retransmissions may degrade the throughput and the QoS satisfaction of the real-time video. In order to combat such drawback of the cross-BERs, alternative switching points are *corner-BERs* (points), which can be found by setting a threshold ε_{th} on the maximum PER of a certain code rate such that $PER_{link} \leq \varepsilon_{th}$. Corner-BER is named because the switching is at the corner point where the throughput slope drops sharply for a given code rate. Obviously, cross-BER can provide larger throughput with loose QoS while corner-BER is more suitable for stringent QoS at the cost of throughput because of the cut-off effect.

In general, it is convenient to use SNR to represent channel status because it is easier to measure than BER. For a Gaussian channel and a given modulation scheme, such as BPSK, the relationship between SNR, γ, and the corresponding BER is given by

$$\gamma = 0.5 * [F^{-1}(1 - BER)]^2 \quad (7)$$

where F^{-1} is the inverse cumulative function of Gaussian distribution. For a given corner-BER or cross-BER, the resultant SNR can be set as the threshold to segment the Markov channel process. Since the fading channel can be modeled by an FSMC and the channel partitioning is affiliated with adaptive coding, the service rate is also a Markov process with the same characteristics as the FSMC, referred to as *Markov service rate process*.

A larger number of partitioned channel states is better for accurate modelling of the FSMC and providing sufficient granularity of the underlying Markov service rate process. However, if the number of states is too large, the residence duration on each state could become too short for practically estimating the channel conditions. To determine a suitable number of states in FSMC, a minimum average residence duration, τ_0, is set on each state such that $\tau_i \geq \tau_0$, where τ_i is the average residence duration on state i, $i = 1, ..., N$, and is a function of level crossing [14]:

$$\begin{aligned}\tau_i &= \frac{\pi_i}{A(\gamma_i) + A(\gamma_{i+1})} \\ &= \frac{1}{f_m} \cdot \frac{\pi_i}{\sqrt{\frac{2\pi\gamma_i}{\gamma_0}}\exp(-\frac{\gamma_i}{\gamma_0}) + \sqrt{\frac{2\pi\gamma_{i+1}}{\gamma_0}}\exp(-\frac{\gamma_{i+1}}{\gamma_0})}. \quad (8)\end{aligned}$$

B. Efficient base layer transmission

For differentiated QoS provisioning, a priority based scheme should be used for bandwidth allocation to each layer. In other words, the most important BL should be given the highest priority, and the remaining resource is then allocated to less important layers. Consider an L-layered video source. The instantaneous bandwidth allocation to each layer should satisfy

$$W_1(t) + W_2(t) + \cdots + W_L(t) = W \quad (9)$$

where $W_j(t)$, $j = 1, ..., L$, is the time-varying bandwidth allocation for layer j with QoS assurance, and W is the total bandwidth available. Therefore, for an N-state FSMC model with a code rate $\{k_i/n\}$ at state i, the effective state dependent service rate for layer j becomes

$$r_{j,i}(t) = k_i \cdot W_j(t)/n, \quad i = 1, 2, \ldots, N. \quad (10)$$

Without loss of generality, we consider two different QoSs for BL and ELs, and $W_1(t)$ and $W_2(t)$ are the bandwidth allocations, respectively.

For BL transmission, let u_{BL}, $v_{BL} = \{v_{BL,i}\}$, and $r_{BL} = \{r_{BL,i}\}$, $i = 1, ...N$, represent the source rate, state dependent retransmission rate and service rate, respectively. If u_{BL} is constant and the service rate at state i satisfies

$$r_{BL,i} \geq u_{BL}[1 + \delta_i] \geq u_{BL} + v_{BL,i} \quad (11)$$

where $\delta_i > 0$ is a state dependent parameter which will be calculated later, the packet dropping rate due to buffer overflow may be close to zero (there is no buffer accumulation in each state). If channel coding provides sufficiently small PER, the total packet loss rate can become negligibly small. We call the approach *assured-rate-allocation*, where the bandwidth allocation for both information and error correction bits of the BL at state i is

$$W_{1,i} = (n/k_i) \cdot u_{BL} \cdot [1 + \delta_i]. \quad (12)$$

Since wireless resource is scarce and the delay tolerance allows video packets to be buffered for a while before transmission, the channel multiplexing gain (besides adaptive coding) should be investigated. In other words, the packets can be buffered at some states (which might incur higher error rates), and then rate-compensated at other states (which might incur lower error rates). Specifically, consider two neighboring states i and $i+1$, where state $i+1$ has a better channel condition. Let $W'_{1,i} = (n/k_i) \cdot u_{BL} \cdot [1 - \alpha]$ and $W'_{1,i+1} = (n/k_{i+1}) \cdot u_{BL} \cdot [1 + \alpha]$, $0 \leq \alpha < 1$, $k_i < k_{i+1}$, i.e., we hold some packets at state i and compensate for their transmissions at state $i+1$. Assume there is no error or retransmission. If $\pi_i = \pi_{i+1}$, the inter-state gain is

$$\begin{aligned}(\frac{n}{k_i} + \frac{n}{k_{i+1}}) \cdot \pi_i \cdot u_{BL} - (W'_{1,i} + W'_{1,i+1})\pi_i \\ = (1/k_i - 1/k_{i+1}) \cdot n \cdot u_{BL} \cdot \alpha\pi_i.\end{aligned}$$

In general, we introduce α and β, the two tunable parameters with $0 \leq \alpha, \beta < 1$. If $(-\alpha)$ is associated with a state, the state is *overloaded* and the buffer occupancy increases. Otherwise, if $(+\beta)$ is associated with a state, the state is *underloaded* and the buffer tends to empty. Applying this to all the states, we can obtain a space with interleaved overload and underload states. Two interleaving methods are considered. One is that the overload and underload states are strictly interleaved with the coefficient vector $\theta = [..., 1-\alpha, 1+\beta, 1-\alpha, 1+\beta, ...]$. We call it *neighbor-interleaving*; the other is that the overload or underload states are clustered to one swing with $\theta = [..., 1-\alpha, 1-\alpha, 1+\beta, 1+\beta, ...]$, which is called *swing-loaded*. In either way, the bandwidth allocation at state i can be represented in a general form as

$$W'_{1,i} = (n/k_i) \cdot u_{BL} \cdot \theta_i \quad (13)$$

where θ_i is the ith element of θ. The tunable parameters, α and β, are determined based on the packet loss probability provided in Section III-C. Given θ, the bandwidth utilization improvement over the *assured-rate-allocation* scheme can be obtained by

$$\xi_{BL} = \frac{\sum_{i \in I}(W_{1,i} \cdot \pi_i) - \sum_{i \in I}(W'_{1,i} \cdot \pi_i)}{\sum_{i \in I}(W_{1,i} \cdot \pi_i)} \quad (14)$$

where I is the set of states that involve in state interleaving.

C. QoS performance analysis

In this subsection, the packet loss probability and the resource utilization efficiency of the proposed adaptive rate allocation schemes are analyzed.

Consider an N-state Markov service rate process. By taking into account the possible retransmissions, the approximate effective arrival rates (including new transmission and retransmissions) are $\{u_i + v_i\}$ and $u_1 + v_1 \leq u_2 + v_2 \leq \cdots \leq u_N + v_N$, where u_i and v_i denote the source and the average retransmission rates, respectively. Let the service rates be r_i, $i = 1, ..., N$. For the channel with a state dependent average packet error rate ϵ_i, an upper bound of the mean retransmission rate can be calculated as

$$v_i \leq \frac{(1-\epsilon_i)}{1-2\epsilon_i}\left(\sum_{j \in \{j: u_j < r_j\}} \epsilon_j p_{ji} u_j + \sum_{j \in \{j: u_j > r_j\}} \epsilon_j p_{ji} r_j\right), \quad i, j = 1, ..., N \quad (15)$$

where p_{ij} is the state transition probability of the Markov service rate process from state i to state j. If $u_i = u_{BL}, u_i < r_i$, and $\epsilon_i = \epsilon$, (11) can be rewritten as

$$r_{BL,i} \geq u_{BL}[1 + \frac{(1-\epsilon)\epsilon}{1-2\epsilon}\sum_{j=1}^{N} p_{ji}]. \quad (16)$$

Obviously, $\delta_i = 1 + \frac{(1-\epsilon)\epsilon}{1-2\epsilon}\sum_{j=1}^{N} p_{ji}$, $i, j = 1, ..., N$, for *assured-rate-allocation*.

Let $\mathbf{P} = \{p_{ij}\}$ be the state transition matrix of the Markov service rate process. Since the underlying Markov service rate process has the same characteristics as the FSMC, $\mathbf{P}$ can be calculated based on (4) and (5) when the mobility is low. Since the transition mainly occurs between the neighboring states, the transition matrix $\mathbf{P}$ is dense along the principal diagonal. When the mobility is relatively fast, the transition is not limited to neighboring states and the transition matrix $\mathbf{P}$ becomes defused around the principal diagonal and spreads over all the states. In either case, slow or fast mobility, $\mathbf{P}$ can be obtained by numerical methods, e.g., by monitoring a certain number of SNR sequences over time. The stationary probability $\mathbf{\Omega} = (\pi_1, ..., \pi_N)$ is independent of mobility and can be obtained by

$$\pi_i = \exp(-\frac{\gamma_i}{\gamma_0}) - \exp(-\frac{\gamma_{i+1}}{\gamma_0}), \quad i = 1, \ldots, N \quad (17)$$

The generator matrix $\mathbf{M}_N$ of the Markov service rate process can then be obtained by $\mathbf{M}_N = \mathbf{I} - \mathbf{P}$, where $\mathbf{I}$ is a $N \times N$ identity matrix.

We apply the fluid-flow approach [18] to analyze the queuing behavior of the transmit buffer by solving the following linear vector differential equation

$$\frac{d\mathbf{F}_N(x)}{dx} = \mathbf{F}_N(x)\mathbf{M}_N\mathbf{D}_N^{-1} \quad (18)$$

where $\mathbf{F}_N(x) = [F_1(x), F_2(x), \cdots, F_N(x)]$, $F_i(x) = \Pr[X \leq x, S = i]$, X and S are random variables denoting the buffer occupancy and the state of the Markov service rate process, respectively, and $\mathbf{D}_N = diag.[u_i + v_i - r_i]$, $1 \leq i \leq N$.

The solution of (18) can be readily obtained as

$$F_i(x) = \pi_i + \sum_{j \in \{z_j < 0\}} a_j \Phi_{ji} e^{z_j x} = \pi_i - \sum_{j \in \{z_j < 0\}} w_{ji} e^{z_j x} \quad (19)$$

where $i = 1,\ldots,N$, $(z_j, \mathbf{\Phi}_j = [\Phi_{j1}, \Phi_{j2}, \cdots, \Phi_{j2N}])$ is the (eigenvalue, eigenvector) pair satisfying the eigenvalue equation $z_j \mathbf{\Phi}_j \mathbf{D}_N = \mathbf{\Phi}_j \mathbf{M}_N$, $1 \le j \le N$, a_j's are the coefficients, and $w_{ji} \ge 0$ is the weight to be determined.

Since the real-time traffic has stringent delay requirement, the packet will be dropped if its transmission delay exceeds a threshold. Given the delay bound is D frames, a set of virtual buffer bounds $\{x_i(D),\ i = 1, 2, \ldots, N\}$ can be determined for each state of the Markov process, which is defined as the virtual buffer length corresponding to the delay bound D. Let $\mathcal{Z}_F = \{i \in N | u_i + v_i > r_i\}$ and $\mathcal{Z}_E = \{i \in N | u_i + v_i < r_i\}$ denote the set of overload and underload states, respectively. In the underload states, the buffer occupancy is low so that $\Pr[X > x_i(D), S = i | i \in \mathcal{Z}_E] = \pi_i - F_i(x_i(D) | i \in \mathcal{Z}_E) = 0$, or $\sum_{j \in \{z_j < 0\}} w_{ji} e^{z_j x_i(D)} = 0$. As a result, for all the underload states, $w_{ji} = 0$. In the overload states, $\Pr[X > x_i(D), S = i | i \in \mathcal{Z}_F] = \pi_i - F_i(x_i(D) \mid i \in \mathcal{Z}_F) = \sum_{j \in \{z_j < 0\}} w_{ji} e^{z_j x_i(D)}$. The probability of buffer overflowing a virtual bound in an overload state, $G_i(x)$, equals the weighted summation of exponentials with negative eigenvalues associated with overload states, i.e.,

$$G_i(x) = \sum_{j \in \{z_j < 0\}} w_{ji} e^{z_j x}. \tag{20}$$

Asymptotically, $G_i(x)$ can be governed by a few significant eigenvalues (the negative eigenvalues with small magnitude). The *significant* eigenvalues include the *dominant* (the largest negative) eigenvalue and several next largest negative eigenvalues, and are denoted by $\{z_j^*,\ j = 1, \ldots, J\}$, where J is the number of significant eigenvalues.

If all weights are equal, i.e., $w_{ji} = w_i$, $j = 1, \ldots, J$, then w_i of an overload state i can be obtained by solving

$$G_i(0^+) = \sum_{j=1}^{J} w_i e^{z_j^* \cdot 0^+} = \frac{(u_i + v_i - r_i)\pi_i^T}{\Omega u^T}, \tag{21}$$

i.e., $w_i = (u_i + v_i - r_i) \cdot \pi_i^T / [J \cdot \Omega u^T]$, where $u = \{u_i,\ i = 1, 2, \ldots, N\}$.

Considering that $r_i(t = 0)$ is the service rate at time $t = 0$, $x_i(D) = \sum_{j=0}^{D-1} r_i(t = j)$ is a random variable depending on the service rates of the next D frames. The expected value of $x_i(D)$ can be obtained as

$$\begin{aligned} \overline{x}_i(D) &= E\left[\sum_{j=0}^{D-1} r_i(t = j)\right] \\ &\doteq r_i(0)\left(1 + \sum_{j=1}^{N} p_{ij}^{(1)} + \sum_{j=1}^{N} p_{ij}^{(2)} + \cdots + \sum_{j=1}^{N} p_{ij}^{(D-1)}\right) \end{aligned} \tag{22}$$

where $p_{ij}^{(m)}$ is the (i,j)th element of the transition probability matrix $\mathbf{P}^{(m)} = \mathbf{P}^m$ after m evolutions, $m = 1, \ldots, D-1$. Denote t_d the time from the moment of a packet arrived at the transmit buffer to the moment when the packet is successfully received at the receiver. We have

$$\begin{aligned} &\Pr[t_d > D, S = i] = G_i[X > x_i(D)] \\ &\doteq \sum_{j=1}^{J} w_i e^{z_j^* \overline{x}_i(D)} \\ &= \sum_{j=1}^{J} w_i \cdot \exp[z_j \cdot r_i(0)(1 + \sum_{j=1}^{N} p_{ij}^{(1)} + \cdots + \sum_{j=1}^{N} p_{ij}^{(D-1)})]. \end{aligned} \tag{23}$$

The total packet dropping probability $P_d(D)$ due to delay exceeding the delay bound is given by

$$P_d(D) = \Pr[t_d > D] = \sum_{i \in \mathcal{Z}_F} \sum_{j=1}^{J} w_i e^{z_j^* \overline{x}_i(D)}. \tag{24}$$

If an erroneous packet is within its delay bound, it can be continuously retransmitted until it is correctly received or exceeds the delay bound. Given that the delay bound is D frames, the round trip acknowledgement can be received within one frame, and the average channel error rate is ϵ, the total packet loss probability P_L due to delay and channel error can be obtained as

$$P_L(D) \doteq \sum_{i=0}^{D-1} \epsilon^i \cdot P_d(D - i). \tag{25}$$

Equation (25) is a conservative evaluation of the packet loss probability when ϵ is small. The channel utilization (efficiency) is the ratio of the average successfully received throughput and the average channel capacity, and is given as

$$\eta = \frac{\Omega u^T (1 - P_L(T_d))}{\Omega r^T} \tag{26}$$

where $r = \{r_i,\ i = 1, 2, \ldots, N\}$ and Ω is the stationary probability vector of the Markov service rate process.

IV. Joint QoS Provisioning

Since the BL carries the most important information and has more stringent QoS requirement, the highest priority is given to BL in resource allocation. By further considering the granularity of multi-layered ELs and the relatively loose QoS requirement, we investigate the joint resource allocation and QoS provisioning for both BL and EL layers.

A. Differentiated QoS provisioning

Given the BL rate allocation, i.e., $W_1(t)$, based on either the assured or the state interleaving (swing) rate allocation scheme, the residual bandwidth, $W_2(t) = W - W_1(t)$, which is also time-varying, can be used for EL transmission. Because $W_1(t)$ for BL is designed based on the partition at corner-BERs, the residual bandwidth should be calculated based on the same channel partition method.

Since the packet loss requirement of EL may be less stringent than that of BL, channel partition for EL can either inherit BL by using corner-BERs or apply cross-BERs to obtain higher throughput. If the corner-BER partition is used, the service rate for EL is

$$r_{EL,i} = (k_i/n)(W - W_{1,i}). \tag{27}$$

The EL adaptation is then based on identical channel characteristics and coding rates as BL. This scheme is referred to as *EL-regular* adaptation. On the other hand, to fully utilize the coding gain, cross-BERs can be used for channel partition. Since EL bandwidth is inherited from BL partition, when cross-BERs are applied, in each state, there will be two code rates. To illustrate this, let Bth_i and Bth_{i+1} be two adjacent SNR thresholds for BL partition, and $Eth_{i'}$ be the SNR corresponding to the cross-BER in between. Then, the code rates for intervals $[Bth_i, Eth_{i'})$ and $[Eth_{i'}, Bth_{i+1})$ are k_i/n and k_{i+1}/n, respectively. As a result, there should be more states for channel fading modelling. This partition is called *EL-extended*, because the state space (the number of states) is extended. Since less conservative coding is applied in some states in *EL-extended*, the *effective throughput rate*, which is defined as the average rate of EL minus the packet loss rate (per frame), can be higher for EL-extended than that for EL-regular. The state space and characteristics of the resultant Markov service rate process for EL-extended can be obtained accordingly from the corresponding Markov channel partitions.

Let $\mathbf{M}_{BL}$ and $\mathbf{M}_{EL}$ be the generator matrices for the Markov rate processes of BL and EL, respectively. Let $u_{EL} = \{u_{EL,i}\}, v_{EL} = \{v_{EL,i}\}$ and $r_{EL} = \{r_{EL,i}\}$ denote the state dependent EL rate vectors of source, retransmission, and service, respectively, for $i = 1, ..., N_{EL}$, where N_{EL} is the state space size of the EL rate process. For multi-layered video sources, it can be further assumed that each EL sub-layer has the same constant source rate c_{EL}. Then the source rate of EL at state i is $u_{EL,i} = l_i \cdot c_{EL}$ and the total source rate is $u_i = u_{BL} + l_i \cdot c_{EL}$, where l_i is the number of EL sub-layers in state i. Note that $(u_{EL,i})$'s may or will not necessarily be bounded by the corresponding service rates. It is possible, or even desirable from the resource management point of view, to allow the coexistence of underload and overload states, to achieve the *inter-state multiplexing gain*. Therefore, given higher priority of BL bandwidth allocation as described in Section III, EL rate allocation is to determine the l_i's such that the residual bandwidth is highly utilized with satisfactory QoS of all EL sub-layers.

To analyze the queuing behaviors of both BL and EL buffers, linear differential equations in vector form can be respectively set for BL and EL queues as

$$\frac{d\mathbf{F}_{BL}(x)}{dx} = \mathbf{F}_{BL}(x)\mathbf{M}_{BL}\mathbf{D}_{BL}^{-1} \tag{28}$$

$$\frac{d\mathbf{F}_{EL}(y)}{dy} = \mathbf{F}_{EL}(y)\mathbf{M}_{EL}\mathbf{D}_{EL}^{-1}. \tag{29}$$

In (28), $\mathbf{F}_{BL}(x) = [F_{BL,1}(x), F_{BL,2}(x), \cdots, F_{BL,N_{BL}}(x)]$, $F_{BL,i}(x) = \Pr[X \le x, S_{BL} = i]$, X and S_{BL} are random variables denoting the BL buffer occupancy and the state of BL partition, respectively, $\mathbf{D}_{BL} = diag.[u_{BL,i} + v_{BL,i} - r_{BL,i}]$, $1 \le i \le N_{BL}$, and N_{BL} is the number of states for BL partition. Similarly, in (29), $\mathbf{F}_{EL}(y) = [F_{EL,1}(y), F_{EL,2}(y), \cdots, F_{EL,N_{EL}}(y)]$, $F_{EL,i}(y) = \Pr[Y \le y, S_{EL} = i]$, Y and S_{EL} are random variables denoting the EL buffer occupancy and the state of EL partition, respectively, $\mathbf{D}_{EL} = diag.[u_{EL,i} + v_{EL,i} - r_{EL,i}]$, $1 \le i \le N_{EL}$, and N_{EL} is the number of states for EL partition. The packet loss rate for BL or EL can be obtained by solving the respective differential equation (28) or (29), as shown in Section III-C. Specifically, given the BL QoS requirement and a constant u_{BL}, the service rate r_{BL} of (28) can be obtained, so does $W_1(t)$. Since the total assigned bandwidth W is fixed, the available state dependent residual bandwidth $W_2(t) = W - W_1(t)$ and r_{EL} can be calculated. Furthermore, EL adaptation u_{EL} of (29) can be calculated for better channel efficiency providing satisfactory QoS requirement of ELs.

In the EL-regular adaptation scheme, EL partition is identical to BL partition so that $\mathbf{M}_{EL}$=$\mathbf{M}_{BL}$. For the EL-extended scheme, since corner-SNRs are used for BL channel partition and cross-SNRs are used for EL channel partition, $\mathbf{M}_{EL}$ differs from $\mathbf{M}_{BL}$ and has to be calculated separately. In the EL-extended scheme, the PER in the SNR interval $[Bth_i, Eth_{i'}), i \in N_{BL}$, is bounded by the corresponding corner-PER of BL. To calculate the average PER in the interval $[Eth_{i'}, Bth_{i+1})$ for EL-extended partition, we assume BPSK modulation with $BER = Q(\sqrt{2SNR})$, where $Q(.)$ is the Q-function of normal distribution. The mean PER in this interval is

$$\epsilon_i' = \frac{\int_{Eth_{i'}}^{Bth_{i+1}} \frac{1}{\gamma_0}\exp\{-\frac{\gamma}{\gamma_0}\} PER_{link}(Q(\sqrt{2\gamma}))d\gamma}{\int_{Eth_{i'}}^{Bth_{i+1}} \frac{1}{\gamma_0}\exp\{-\frac{\gamma}{\gamma_0}\}d\gamma} \tag{30}$$

where $PER_{link}(\cdot)$ is given by (2). Given the average PER of a state, the retransmission rate v_{BL} and v_{EL} can be obtained.

B. Integrated QoS provisioning

For practical implementation, BL and ELs may have the same QoS requirements and are transmitted together. In this case, the system model in Fig. 1 can be revised by combining two transmission buffers into one single buffer. The source rate becomes $u_i = u_{BL} + l_i \cdot c_{EL}$, $i = 1, ...N$. Let $u = \{u_i\}, v = \{v_i\}$ and $r = \{r_i\}$ be the state dependent rate vectors of the source, retransmission and service, respectively. The channel partition method can be either based on corner-BERs or cross-BERs. The objective of rate adaptation is to maximize channel efficiency, $\eta = \frac{\Omega u^T(1-P_L(T_d))}{\Omega r^T}$, in terms of $l = \{l_i\}$ with QoS satisfaction.

We propose a heuristic method to determine the source rate parameters l_i as follows.

1) Set initial values of l_i's so that all states are underloaded, i.e., $u_i + v_i \le r_i$, $i = 1, ..., N$. The two end states, where the corresponding SNRs are very small and very large, should be conservatively underloaded, because the corresponding stationary probabilities may be relatively large and their average state residence durations may be longer. These states are difficult to compensate if they are overloaded.
2) Sort the l_i's in ascending order and increase l_i by 1 in order. Every time one l_i changes, the following conditions are checked: (a) traffic intensity $\rho = \frac{\Omega(u+v)^T}{\Omega r^T} \le 1 - \delta$ and (b) $P_L(D) \le \varepsilon$, where δ and ε are predefined thresholds. The first condition guarantees the stability of the system, and the second one provides the QoS assurance. Increment of l_i may cause some states to become overloaded and deteriorate the packet loss rate.

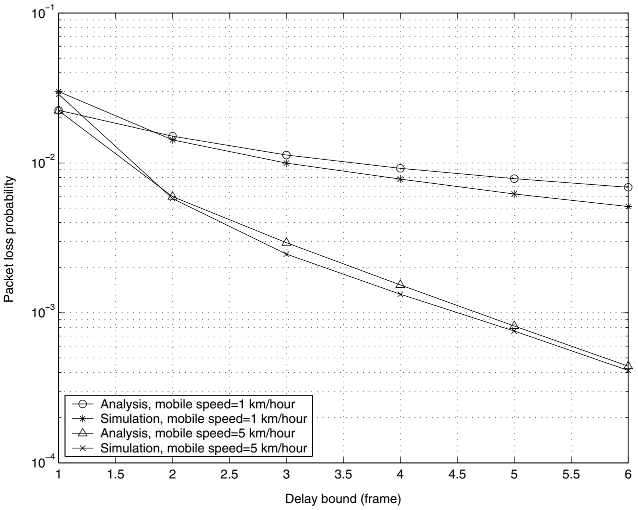

Fig. 3. Packet loss probability vs delay bound, integrated BL/EL, PER=0.01.

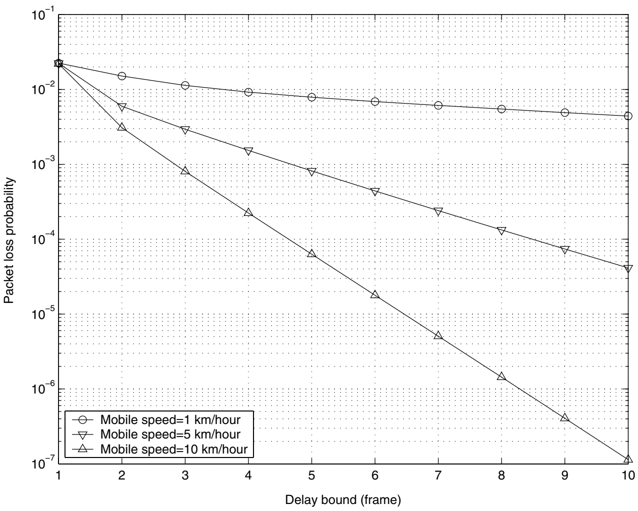

Fig. 4. Packet loss probability vs delay bound for various mobilities.

If condition (a) or (b) can not be satisfied, leave the current l_i intact and find the next available state for consideration. The iteration stops when all l_i's have been considered.

V. Numerical Results and Discussion

In this section, numerical results are presented to evaluate the proposed adaptive rate allocation schemes for real-time multi-layered video transmissions. Both suitability and scalability issues are discussed.

A. Simulation parameters

Consider a wireless system with Rayleigh fading channel. The radio time frame is set to 20 ms. The carrier frequency is 900M Hz. Mobile speed is from 0.5 to 10 km/hour (for speed larger than 10 km/hour, the average duration of intermediate states may be very short and few states should be chosen). The modulation scheme is BPSK. The fading condition is fixed in each time frame. The available (assigned) bandwidth and the service (or source) rates are measured in the number of link

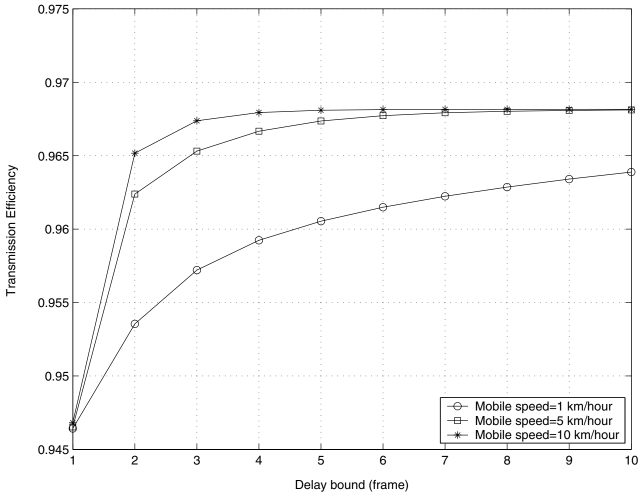

Fig. 5. Transmission efficiency vs delay bound for various mobilities.

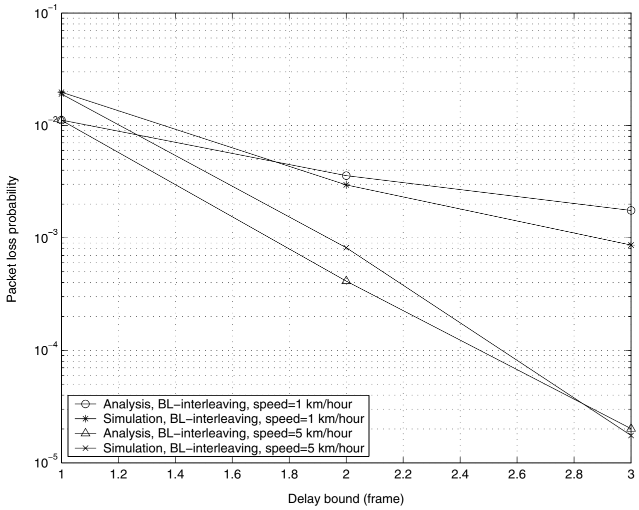

Fig. 6. Packet loss probability vs delay bound, BL adaptation by neighbor-interleaving.

layer packets per frame, where each packet contains $n_1 = 20$ block code words. Let the assigned bandwidth $W = 70$ (packets/frame), BL source rate $u_{BL} = 20$ (packets/frame) and each EL sub-layer source rate $c_{EL} = 10$ (packets/frame). EL sub-layers can be incorporated until the assigned bandwidth is reached. The BCH block code has a block size $n = 256$ and information length $k \in \{239\ 223\ 199\ 179\ 155\ 131\ 107\ 79\}$. The Markov fading process is partitioned by corner-SNRs (the corresponding SNRs to corner-BERs). The parameters of FSMC transition matrix can be evaluated by monitoring the SNR for a certain amount of time. In our simulation, transition matrices for mobile speed of 0.5km/hour and 5km/hour are used as shown on the top of the next page. However, the stationary probability, Ω=[.2922 .0384 .0617 .0495 .0656 .1006 .1117 .2803], is independent of mobile speed.

B. Integrated and differentiated QoS provisioning

We first consider the integrated transmission of BL and ELs with the same QoS, and then present the differentiated QoS

$$\mathbf{P}(0.5\text{km/hour}) = \begin{pmatrix} .9888 & .0112 & 0 & 0 & 0 & 0 & 0 & 0 \\ .0755 & .8382 & .0842 & 0 & 0 & 0 & 0 & 0 \\ 0 & .0492 & .9009 & .0500 & 0 & 0 & 0 & 0 \\ 0 & 0 & .0594 & .8802 & .0604 & 0 & 0 & 0 \\ 0 & 0 & 0 & .0442 & .9103 & .0456 & 0 & 0 \\ 0 & 0 & 0 & 0 & .0285 & .9457 & .0258 & 0 \\ 0 & 0 & 0 & 0 & 0 & .0233 & .9576 & .0191 \\ 0 & 0 & 0 & 0 & 0 & 0 & .0079 & .9921 \end{pmatrix}$$

$$\mathbf{P}(5.0\text{km/hour}) = \begin{pmatrix} .9043 & .0470 & .0392 & .0084 & .0011 & 0 & 0 & 0 \\ .3915 & .2298 & .2619 & .0924 & .0205 & .0039 & 0 & 0 \\ .1739 & .1724 & .3106 & .1949 & .1227 & .0256 & 0 & 0 \\ .0462 & .0633 & .2543 & .2724 & .2513 & .1085 & .0040 & 0 \\ .0072 & .0185 & .1345 & .1965 & .3156 & .2995 & .0282 & 0 \\ 0 & .0026 & .0132 & .0551 & .1970 & .4836 & .2378 & .0106 \\ 0 & 0 & 0 & .0009 & .0219 & .2092 & .5812 & .1867 \\ 0 & 0 & 0 & 0 & 0 & .0020 & .0735 & .9245 \end{pmatrix}.$$

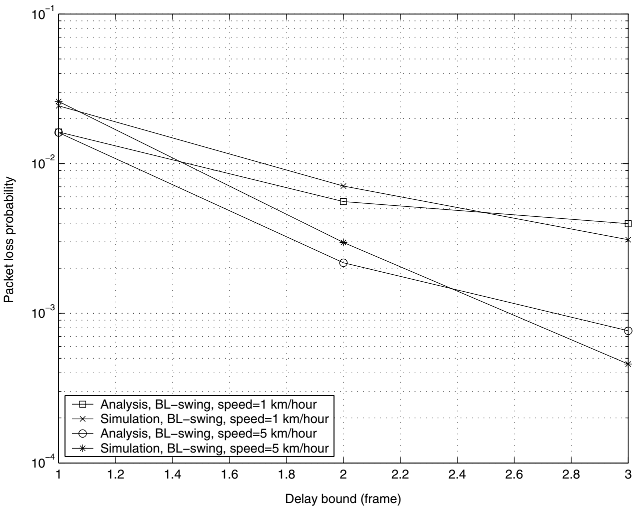

Fig. 7. Packet loss probability vs delay bound, BL adaptation by swing-loaded.

TABLE I
SIGNIFICANT EIGENVALUES VS MOBILE SPEED (SPEED UNIT: KM/HOUR)

speed=0.5	speed=1	speed=3	speed=5	speed=10
-.0012	-.0022	-.0073	-.0148	-.0301
-.0091	-.018	-.0572	-.1027	-.1696
-.0615	-.1236	-.3732	-.5364	-.6377
-.1835	-.3357	-.883	-.9567	-1.0889

provisioning.

In integrated QoS provisioning, due to the corner-BER based partition, the average PER of each state is bounded. Given $u_{BL} = 20$, $u_{EL} = 10$, and $W = 70$, the state dependent service rate $r = [22\ 29\ 36\ 43\ 49\ 55\ 61\ 65]$. In each state, resource is first allocated for BL transmission; one or more ELs are then selectively sent using the residual resource based on the channel state and QoS requirement. For QoS evaluation, the number of significant eigenvalues J is set to 2. Using matched sending rate [20 30 40 40 50 60 60 60], when the channel PER is bounded by 0.01, the packet loss probability and channel efficiency vs various mobility are shown in Figs. 3, 4, and 5. It can be seen that the analytical and simulation results agree reasonably well; as mobile speed increases, the packet loss rate decreases and the channel utilization increases. This can be explained by Table I, in which the first two rows of eigenvalues are used in calculating the packet loss rate. It is observed that the magnitude of eigenvalues become larger when the mobile speed increases, and this leads to a smaller packet loss rate. This can also be explained as follows. When the source rate is larger than the service rates in overload states, the packets are cumulated in the buffer. If the mobile speed or the fading is slow, the channel condition is likely to stay in the current state and packet loss occurs when the duration in an overload state exceeds the delay bound. However, if the mobile speed is relatively high, the fading is fast and the channel condition is likely to move to other states. If the channel status moves from an overload state to an underload state within the delay bound, the queued packets can be absorbed in the underload state and the inter-state multiplexing gain can be achieved. To demonstrate this, we tabulate the average time duration for each state with respect to mobile speed, as shown in Table II. It is observed that as the mobile speed increases, the average time duration in each state decreases.

In the case of differential QoS provisioning, both *neighbor-interleaving* and *swing-loaded* schemes are evaluated analytically and by simulation. The tunable parameters α and β are adjusted to balance the efficiency improvement and QoS. Figs. 6 and 7 show the performance of the *neighbor-interleaving* adaption with $\alpha = 0.05$, $\beta = 0.05$ and the *swing-loaded* adaptation with $\alpha = 0.1$, $\beta = 0.05$, respectively. It can be observed that the analytical result is conservatively reasonable compared with the simulation results. This is because only two significant eigenvalues are used for the calculation of the analytical results. For the BL-swing scheme, the queued packets in the overload states will encounter a series of underload states because the effect of fading diminishes. On

TABLE II

AVERAGE TIME DURATION VS MOBILE SPEED (SPEED UNIT: KM/HOUR, TIME UNIT: FRAME)

state	1	2	3	4	5	6	7	8
speed=1	50.191	3.267	5.192	4.1623	5.6159	9.1229	11.542	63.528
speed=3	16.73	1.089	1.7307	1.3874	1.872	3.041	3.8474	21.176
speed=5	10.038	0.65339	1.0384	0.83246	1.1232	1.8246	2.3084	12.706
speed=8	6.2738	0.40837	0.649	0.52029	0.70199	1.1404	1.4428	7.9411
speed=10	5.0191	0.3267	0.5192	0.41623	0.56159	0.91229	1.1542	6.3528

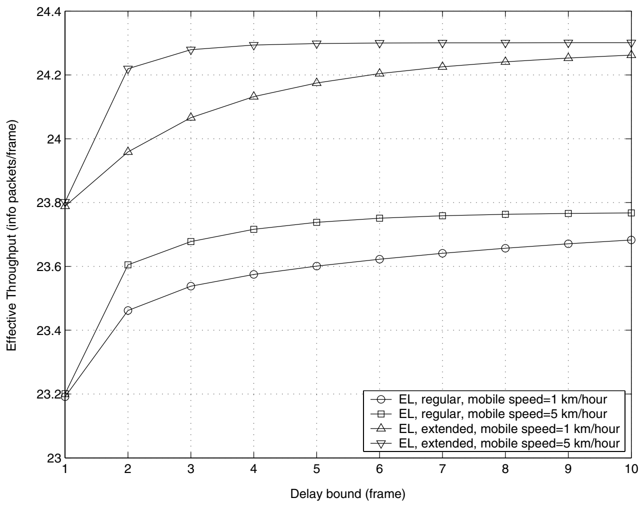

Fig. 8. Effective throughput rate for regular and extended EL schemes.

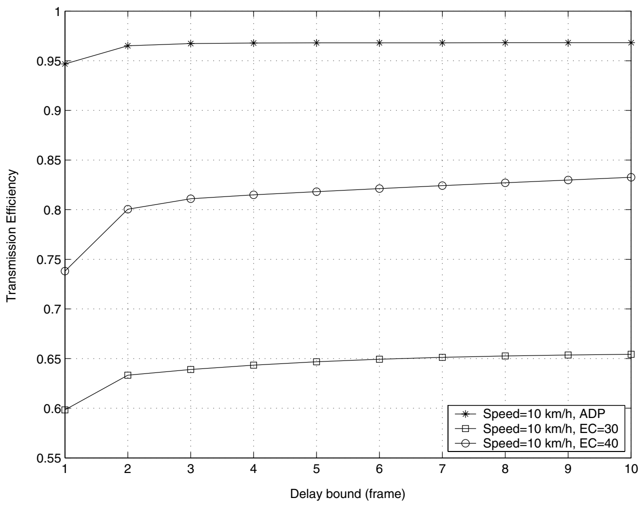

Fig. 10. Comparison of transmission efficiency for Rate-adaptation and EC schemes.

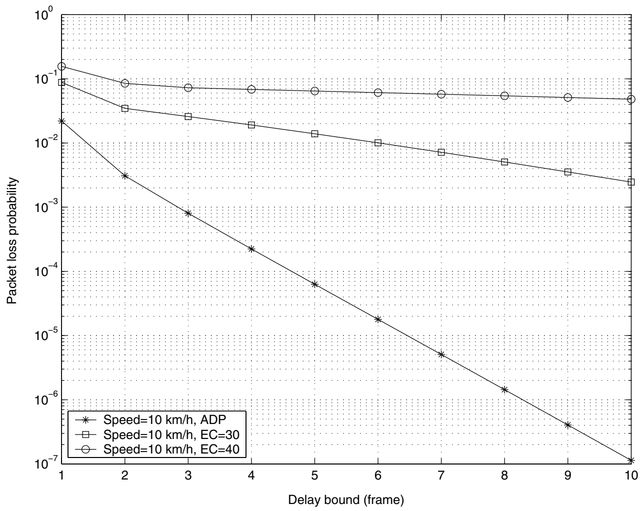

Fig. 9. Comparison of packet loss probabilities for Rate-adaptation and EC schemes.

the other hand, when the effect of fading increases, more packets will be queued since overload states tend to occur in batch. As long as the state transition (fading rate) is fast, the queued packets can be quickly released when channel conditions become good. For EL adaptation, the performance of both EL-regular and EL-extended schemes are compared. For the same packet loss rate, it can be seen in Fig. 8 that the EL-extended scheme can achieve better effective throughput rate than the EL-regular scheme.

We further compare the proposed rate adaptation schemes with the commonly used EC scheme [12], [13]. For the sake of generality, we consider the integrated transmission. The average service rate is 44.16 (calculated by $\Omega \cdot r^T$). Using the adaptive scheme in Section IV-B, the average source rate is 43 packets/frame. Constraints of 30 and 40 (packets/frame) are imposed on the source for the EC scheme. The comparisons are shown in Figs. 9 and 10, respectively. Fig. 9 shows that the packet loss rate of EC is worse when the constraint is relaxed. Even when the constraint is set as low as 30 compared with the average source rate of 43 in the adaptive scheme, the packet loss probability is much worse for EC than that for the adaptive scheme. In addition, as shown in Fig. 10, for the channel efficiency, the adaptive scheme has superior performance over EC.

VI. CONCLUSIONS

In this paper, adaptive rate allocation for scalable multi-layered video transmission over time-varying wireless channels has been investigated in terms of channel efficiency and QoS provisioning. Rate adaptation schemes for integrated and differentiated BL/EL real-time video transmission have been proposed. A novel analytical approach based on a Markov fluid-flow model has been presented to evaluate the QoS performance in terms of packet loss rate due to delay and channel error. Simulation results verify the accuracy of the proposed analytical model and demonstrate the efficiency of channel utilization.

REFERENCES

[1] Q. Zhang, W. Zhu, and Y.-Q. Zhang, "End-to-end QoS for video delivery over wireless Internet," *Proc. of IEEE*, vol. 193, pp. 123–134, Jan. 2005.

[2] H. Radha, Y. Chen, K. Parthasarathy, and R. Cohen, "Scalable Internet video using MPEG-4," *ELSEVIER Signal Processing: Image Commun.*, vol. 15, pp. 95–126, 1999.

[3] L. Galluccio, F. Licandro, G. Morabito, and G. Schembra, "An analytical framework for the design of intelligent algorithms for adaptive-rate MPEG video encoding in next-generation time-varying wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 23, pp. 369–384, Feb. 2005.

[4] T. Stockhammer, H. Jenkac, and G. Kuhn, "Streaming video over variable bit-rate wireless channels," *IEEE Trans. Multimedia*, vol. 6, pp. 268–277, Apr. 2004.

[5] M. Bystrom and T. Stockhammer, "Dependent source and channel rate allocation for video transmission," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 258–268, Jan. 2004.

[6] Q. Zhang, W. Zhu, and Y.-Q. Zhang, "Channel-adaptive resource allocation for scalable video transmission over 3G wireless network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, pp. 1049–1063, Aug. 2004.

[7] M. Hassan, M. M. Krunz, and I. Matta, "Markov-based channel characterization for tractable performance analysis in wireless packet networks," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 821–831, May 2004.

[8] J. Xu, X. Shen, Jon W. Mark, and J. Cai, "Efficient real-time video transmission in OVSF-CDMA system," in *Proc. IEEE WCNC05*, pp. 1329-1334, New Orleans, USA, Mar. 13-17, 2005.

[9] T. Ahmed, A. Mehaoua, R. Boutaba, and Y. Iraqi, "Adaptive packet video streaming over IP networks: a cross-layer approach," *IEEE J. Sel. Areas Commun.*, vol. 23, pp. 385–401, Feb. 2005.

[10] J. Liu, B. Li, Y. T. Hou, and I. Chlamtac, "On optimal layering and bandwidth allocation for multisession video broadcasting," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 656–667, Mar. 2004.

[11] Y. Pan, M. Lee, J. B. Kim, and T. Suda, "An end-to-end multipath smooth handoff scheme for stream media," *IEEE J. Sel. Areas Commun.*, vol. 22, pp. 653–663, May 2004.

[12] W. Kumwilaisakand, Y. T. Hou, Q. Zhang, W. Zhu, C. J. Kuo, and Y. Q. Zhang, "A cross-layer quality-of-service mapping architecture for video delivery in wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 21, pp. 1685–1698, Dec. 2003.

[13] D. Wu and R Negi, "Effective capacity: a wireless link model for support of quality of service," *IEEE Trans. Wireless Commun.*, vol. 2, pp. 630 –643, July 2003.

[14] Q. Zhang and S. Kassam, "Finie-state Markov model for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, pp. 1688–1692, Nov. 1999.

[15] H. S. Wang and N. Moayeri, "Finite-state Markov channel–a useful model for radio communication channels," *IEEE Trans. Commun.*, vol. 44, pp. 163–171, Feb. 1995.

[16] R. H. Deng and M. L. Lin, "A type I hybrid ARQ system with adaptive code rates," *IEEE Trans. Commun.*, vol. 18, pp. 733–737, Feb./Mar./Apr. 1995.

[17] B. Vucetic, "An adaptive coding scheme for time-varying channels," *IEEE Trans. Commun.*, vol. 39, pp. 653–663, May 1991.

[18] M. Schwartz, *Broadband Integrated Networks*. Prentice Hall, 1996.

[19] S. Lin and D. J. Costello, Jr., *Error Control Coding, Second Edition*. Prentice-Hall, 2004.

Jun Xu received the B.Eng. from Tsinghua University, China, MASc (2002) in Systems Design Engineering and Ph.D. (2006) in Electrical Engineering from University of Waterloo, Ontario, Canada. His current research interests include radio resource management for multimedia in wireless communication systems, mobility management, wireless internetworking, and the application of artificial intelligence in communication networks.

Xuemin (Sherman) Shen (M'97-SM'02) received the B.Sc. (1982) degree from Dalian Maritime University (China) and the M.Sc. (1987) and Ph.D. degrees (1990) from Rutgers University, New Jersey (USA), all in electrical engineering. Dr. Shen is with the Department of Electrical and Computer Engineering, University of Waterloo, Canada, where he is a Professor and the Associate Chair for Graduate Studies. Dr. Shen's research focuses on mobility and resource management in interconnected wireless/wireline networks, UWB wireless communications systems, wireless security, and ad hoc and sensor networks. He is a coauthor of two books, and has published more than 200 papers and book chapters in wireless communications and networks, control and filtering.

Dr. Shen serves as the Technical Program Committee Chair for IEEE Globecom'07, IEEE WCNC'07 Network Symposium, Qshine'05, IEEE Broadnet'05, WirelessCom'05, IFIP Networking'05, ISPAN'04, IEEE Globecom'03 Symposium on Next Generation Networks and Internet. He also serves as Associate Editor for *IEEE Transactions on Wireless Communications*, *IEEE Transactions on Vehicular Technology*, *ACM Wireless Networks*, *Computer Networks*, *Wireless Communications and Mobile Computing* (Wiley), etc., and the Guest Editor for *IEEE Journal on Selected Areas in Communications*, *IEEE Wireless Communications*, and *IEEE Communications Magazine*. Dr. Shen received the Premier's Research Excellence Award (PREA) from the Province of Ontario, Canada for demonstrated excellence of scientific and academic contributions in 2003, and the Outstanding Performance Award from the University of Waterloo, for outstanding contribution in teaching, scholarship and service in 2002. Dr. Shen is a registered Professional Engineer of Ontario, Canada.

Jon W. Mark (M'62-SM'80-F'88-LF'03) received the Ph.D. degree in electrical engineering from McMaster University, Canada in 1970. Upon graduation, he joined the Department of Electrical Engineering (now Electrical and Computer Engineering) at the University of Waterloo, became a full Professor in 1978, and served as Department Chairman from July 1984 to June 1990. In 1996, he established the Centre for Wireless Communications (CWC) at the University of Waterloo and has since been serving as the founding Director.

Dr. Mark was on sabbatical leave at the IBM Thomas Watson Research Center, Yorktown Heights, NY, as a Visiting Research Scientist (1976-77); at AT&T Bell Laboratories, Murray Hill, NJ, as a Resident Consultant (1982-83); at the Laboratoire MASI, Université Pierre et Marie Curie, Paris, France, as an Invited Professor (1990-91); and at the Department of Electrical Engineering, National University of Singapore, as a Visiting Professor (1994-95).

His current research interests are in wireless communications and wireless/wireline interworking, particularly in the areas of resource management, mobility management, and end-to-end information delivery with QoS provisioning.

A co-author of the text **Wireless Communications and Networking**, Prentice-Hall, 2003, Dr. Mark is a Life Fellow of the IEEE and has served as a member of a number of editorial boards, including *IEEE Transactions on Communications*, *ACM/Baltzer Wireless Networks*, *Telecommunication Systems*, etc. He was a member of the Inter-Society Steering Committee of the *IEEE/ACM Transactions on Networking* from 1992-2003, and a member of the IEEE COMSOC Awards Committee during the period 1995-1998.

Jun Cai (M'04) received the B.Sc. (1996) and the M.Sc. (1999) degrees from Xi'an Jiaotong University (China) and Ph.D. degree (2004) from University of Waterloo, Ontario (Canada), all in electrical engineering. From June 2004 to April 2006, he was with McMaster University as NSERC Postdoctoral Fellow. Since July 2006, he has been with the Department of Electrical and Computer Engineering, University of Manitoba, Canada, where he is an Assistant Professor. His current research interests include multimedia communication systems, mobility and resource management in 3G beyond wireless communication networks, and ad hoc and mesh networks. He is currently a holder of NSERC Associated Industrial Research Chair.