

# Efficient Channel Utilization for Real-Time Video in OVSF-CDMA Systems with QoS Assurance

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

**Abstract**—In this paper, the utilization of real-time video service in the downlink of an orthogonal variable spreading factor code division multiple access (OVSF-CDMA) system is studied. By modeling the video traffic and wireless channel as a joint Markov modulated process, and properly partitioning the states of the Markov process, an adaptive rate allocation scheme is proposed for real-time video transmission with quality of service provisioning while achieving high channel utilization. The scheme is applicable for packet switching and frame-by-frame real-time video transmission, and incorporates both the physical layer and network layer characteristics. For QoS provisions, the closed form expressions of packet delay and loss probability are derived based on the Markov model. Analytical and simulation results demonstrate that the proposed scheme can significantly improve the channel utilization over the commonly used effective bandwidth approach.

**Index Terms**—Multimedia traffic, OVSF-CDMA, QoS assurance, radio resource allocation, real-time video, wireless networks.

## I. INTRODUCTION

THE third generation (3G) and beyond wireless networks are expected to provide multimedia services (including voice, video and data) with quality of service (QoS) provisioning. For a packet switching wireless network, QoS parameters include the transmission bit error rate (BER) at the physical layer and the packet loss probability (due to transmission error and network congestion) at the higher layers. Wideband code division multiple access (W-CDMA) has been adopted as the multiple access technology in 3G wireless networks. In W-CDMA networks, two major platforms are ordinarily used for multimedia transmission: multi-code CDMA (MC-CDMA) and orthogonal variable spreading factor CDMA (OVSF-CDMA)[1]. Although the former has been proved more suitable for multimedia applications, the latter is widely applied in commercial systems such as IMT2000 and cdma2000, due to the simple mobile terminals. OVSF-CDMA supports multimedia services by allocating orthogonal codes with variable lengths to users so that it can achieve multiple service rates in frame-by-frame based wireless communications.

There has been a number of research works reported in the literature that address traffic regulation and resource allocation

in wireless CDMA networks supporting heterogeneous services. Centered around the theme of QoS provisioning, these works focus on power control, transmission rate allocation, scheduling and capacity optimization, etc. Transmission of real-time video by the proper configuration of multiple codes is explored in [2]. A multiple access protocol for integration of multimedia traffic for UMTS/IMT-2000 based W-CDMA is proposed in [3], but only the performance of voice and data is evaluated. In [4], radio resource management strategies, including uplink/downlink scheduling and congestion control, are proposed for W-CDMA systems; however, no analytical formulation is presented. Recent works in this area include those reported in [5] and [6]. In [5], optimal power management is discussed and in [6], an effective capacity scheme is proposed based on a wireless link model. Moreover, as the major service supported by communication networks, real-time video transmission has also been extensively investigated in broadband wireline networks (e.g., ATM networks) in terms of QoS provisioning and efficient bandwidth allocation. The fundamental strategies of regulating video traffic in ATM networks, such as effective bandwidth scheme, are well discussed in [7]. In [8], a dynamic resource allocation approach is proposed based on the prediction of future traffic using the content and traffic information of short video segments. A fair bandwidth allocation policy is presented in [9]. This scheme achieves good QoS performance by maintaining a fair distribution of buffer length across the video streams. A diffusion process is used in [10] to model the statistical multiplexing of Markov modulated video traffic for bandwidth allocation and admission control. As these schemes are proposed for wireline broadband networks, they cannot be directly applied to wireless networks, especially in OVSF-CDMA networks, since (a) multiplexing video streams in CDMA coded channels is not so straightforward as in wireline networks; and (b) the variable service rate in OVSF-CDMA channels is essentially stepwise. Therefore, a method to efficiently transmit variable rate video over OVSF-CDMA channels is needed to enhance radio resource utilization.

Downlink service scheduling is a critical radio resource management function due to the perceived heavier traffic in the downlink than that in the uplink. The situation becomes even more critical in OVSF-CDMA systems because of the limitation of orthogonal code set [11], [12]. Non-proper rate allocation may result in inefficient channel utilization and higher load on code administration, and introduce extra interference if a high level coded rate is unnecessarily assigned.

Manuscript received January 21, 2004; revised December 13, 2004; accepted April 12, 2005. The associate editor coordinating the review of this paper and approving it for publication was R. Negi.

The authors are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1 (e-mail: {jxu, xshen, jwmark, jcai}@bbcr.uwaterloo.ca).

Digital Object Identifier 10.1109/TWC.2006.04031

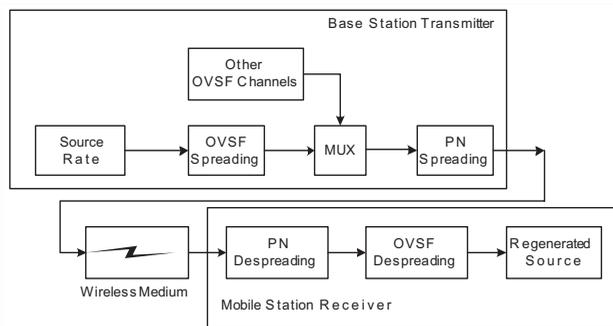


Fig. 1. Downlink transmission of the OVFS-CDMA system.

By taking into account the variable characteristics of the video traffic, it should be more preferable to transmit a variable video stream using adaptive rates rather than a peak or constant rate. Most of the research works on OVFS code administration and rate allocation have aimed either at data traffic [12], or at a mixture of delay sensitive (voice, video) and non-delay sensitive data [13], [14]. However, to the best of our knowledge, there is no analytical work on adaptive assignment of variable transmission rate to real-time video with guaranteed QoS in an OVFS-CDMA system.

We are concerned with the downlink channel utilization for real-time video service in an OVFS-CDMA system. The work presented in the sequel is motivated by the possibility of improving channel utilization using adaptive rate allocation. By modeling video traffic as a Markov modulated process, and properly segmenting the states of the Markov process, an adaptive rate allocation scheme is proposed to efficiently support real-time video transmission while guaranteeing the QoS requirements in terms of stringent packet delay and loss probability. The proposed scheme incorporates both the physical layer and network layer characteristics, and is applicable for packet switching networks and frame-by-frame video transmission. Closed form expressions of packet delay and loss probability are derived based on a two-dimensional Markov process. Analytical and simulation results show that the proposed scheme can significantly improve the channel utilization over the commonly used effective bandwidth approach.

The remainder of the paper is organized as follows. In Section II, the system, channel, and video traffic models of the OVFS-CDMA downlink are presented. In Section III, an adaptive rate allocation scheme is proposed for variable rate video traffic flows with the guaranteed QoS requirements. In Section IV, the QoS performance in terms of packet delay and packet loss probability is analyzed. Numerical results are given in Section V, followed by conclusions in Section VI.

## II. SYSTEM MODEL

Fig. 1 shows the transceiver structure of the downlink OVFS-CDMA system. At the base station (BS) transmitter, the source data is first spread by the OVFS code which is a unique orthogonal Walsh sequence. The choice of OVFS code is determined by the transmission rate of the source data. All traffic spread by the OVFS codes are multiplexed and are further spread using a pseudonoise (PN) code, which is used

to separate different BSs [1], to obtain the transmitted signal. At the mobile station (MS) receiver, the received signal is despread by the same PN sequence and the same OVFS code to recover the original source data.

### A. The Two-State Wireless Channel

The transmitted signal passes through the wireless channel which introduces path loss, shadowing, and multipath fading. In wideband CDMA systems, power control and FEC/ARQ (forward error correction/automatic repeat request) are ordinarily used to mitigate the channel impairments and to reduce the packet error rate (PER) (PER is a function of BER subject to coding). Since PER depends on the target SIR and the channel coding scheme [15], for a given coding scheme, power control can be considered as the primary way to compensate fading so that the residual average PER after power control becomes small. Depending on the effectiveness of power control, the errors can be randomly distributed or in bursty pattern [16]. To characterize the compensated wireless channel, a Gilbert model [17], [18], which is a two-state Markov chain, is used. In the Gilbert model, let state 0 denote a good channel condition with PER  $\epsilon_0$ , and state 1 denote a bad channel condition with PER  $\epsilon_1 (\gg \epsilon_0)$ . Given the stationary probabilities of the channel model  $(1 - q)$  and  $q$  for state 0 and state 1, respectively, the average packet error rate can be calculated as  $\epsilon = (1 - q)\epsilon_0 + q\epsilon_1$ . When  $\epsilon_0 \rightarrow 0$ , the bursty error rate  $\epsilon_1 \cong \epsilon/q$ . When  $q = 1$ , the channel becomes a random erroneous channel with average PER  $\epsilon$ .

### B. Downlink OVFS Channelization

OVFS codes, implemented by Walsh sequences, are used for channelization of W-CDMA downlink. The mutual orthogonality of Walsh sequences ensures that modulated signals for different MSs are orthogonal, so that the multiple access interference (MAI) among the users in the same cell is eliminated. The OVFS codes can be represented by a binary tree with each leaf (node) representing a unique code, and can support traffic flows with variable rates. The heterogeneous traffic streams with different QoS requirements are spread by the orthogonal codes with different lengths (i.e., different spreading gains) and multiplexed with transmissions by other users in the same cell.

Although the OVFS codes have perfect orthogonality, there are limitations in applications. First, the spreading factor can only be changed in a stepwise way, i.e., increasing or decreasing by the power of two. Second, the possible code blocking<sup>1</sup> makes it difficult to assign OVFS codes. Only mutually orthogonal codes can be used for simultaneous transmissions; otherwise, data streams spread by the same code will result in destructive interference. In a packet switching system, OVFS codes can be changed on a frame-by-frame basis so that adaptive rate transmission can be achieved to match different channel and traffic conditions.

<sup>1</sup>Although imaginary number theory is applied to overcome code blocking [21].

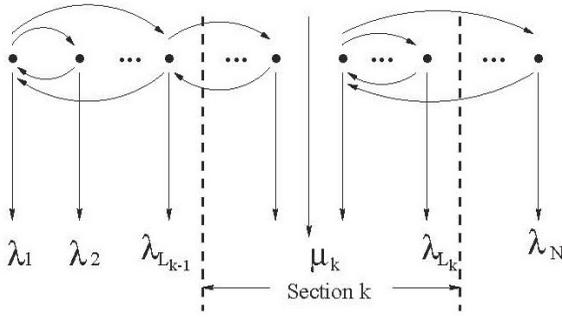


Fig. 2. Partition of the video source rate span according to the service rates.

### C. MMRP Video Traffic Model

A number of video traffic models, including autoregressive (AR) stochastic modeling and the Markov modulated process, have been proposed in the literature [7], [19]. Both models are believed to give rather accurate approximation. The AR model is a recursive Gaussian process which is suitable for generating simulated video sequences, while the Markov model is suitable for analytical purpose [7], [20]. In the Markov model, the rate variation of a video sequence is modulated by a continuous-time finite-state Markov process. Since the time is discretized into fixed length frames, this process is also referred as the Markov modulated rate process (MMRP) [20]. The variable traffic rate can be quantized into several levels, and each level is represented by a state of the MMRP. Transitions between states are governed by the underlying continuous-time Markov chain.

Let  $N$  denote the state number of the MMRP. The infinitesimal generator matrix  $\mathbf{M}$  for the general continuous-time Markov chain is an  $N \times N$  matrix given by  $\mathbf{M} = \{m_{ij}\}$ ,  $i, j = 1, \dots, N$ , where  $m_{ii} = -\sum_{j=1, j \neq i}^N m_{ij}$ . Let  $\boldsymbol{\pi} = [\pi_1, \pi_2, \dots, \pi_N]$  represent the stationary probability. It can be shown that  $\boldsymbol{\pi}\mathbf{M} = \mathbf{0}$ , and  $\boldsymbol{\pi}$  can be obtained by conditioning on  $\sum_{i=1}^N \pi_i = 1$ . To get the infinitesimal generator matrix  $\mathbf{M}$ , one can first monitor the video sequence for a certain amount of frames to get the transition probability matrix  $\mathbf{P}$  of the corresponding discrete Markov chain, given by  $\mathbf{P} = \{p_{ij}\}$ ,  $i, j = 1, \dots, N$ , where  $p_{ij}$  is the one-step transition probability from state  $i$  to  $j$ . According to [20], the matrix  $\mathbf{M}$  can be obtained by  $\mathbf{M} = \mathbf{I} - \mathbf{P}$ , where  $\mathbf{I}$  is a  $N \times N$  identity matrix.

## III. RATE ALLOCATION

In this section, an adaptive allocation of the transmission rate to each video source for efficient channel utilization is considered. Packet switching allows to switch OVSF codes among various rate levels on a frame by frame basis. If the variable source rate can be matched by different service rates with guaranteed QoS requirements, channel utilization can be improved.

### A. Partition on the Rate Span

Consider an  $N$ -state MMRP model for a video source with state dependent arrival rate  $\lambda_i$ ,  $i = 1, \dots, N$ . The source can be

buffered for transmission or retransmission. Let  $v_i$  be the mean retransmission rate at state  $i$ . For efficient resource utilization, the physical layer should be designed in such a way that the probability of retransmission is very low (e.g., 1%). Without loss of generality, consider the rate span with approximate effective rates (including new transmission and retransmission)  $\lambda_i + v_i$ ,  $i = 1, \dots, N$ , and  $\lambda_1 + v_1 \leq \lambda_2 + v_2 \leq \dots \leq \lambda_N + v_N$ . Furthermore, the rate span is covered by at least  $K$  layers of OVSF coded rates, i.e., letting the service rate  $\mu^{(1)}$  be the coded rate at a certain layer of the code tree, and  $\mu^{(2)} = 2^1 \cdot \mu^{(1)}$ ,  $\mu^{(K)} = 2^{K-1} \cdot \mu^{(1)}$ , then  $\lambda_1 < \mu^{(1)}, \mu^{(2)}, \dots, \mu^{(K-1)} < \lambda_N$  where  $\mu^{(K)}$  is allowed to be larger or smaller than  $\lambda_N$ . For rate allocation and performance analysis, the source rate span is partitioned into  $K$  sections, with each section having a unique service rate  $\mu^{(k)}$  ( $\mu_k$ ), as shown in Fig. 2. The effective mean arrival rates can be remarked and listed in ascending order:

$$\lambda_1 + v_1, \dots, \lambda_{L_1} + v_{L_1}; \dots, \lambda_{L_2} + v_{L_2}; \\ \dots, \lambda_{L_{K-1}} + v_{L_{K-1}}; \dots, \lambda_N + v_N$$

where  $L_j$ ,  $j = 1, \dots, K-1$ , are boundary states that separate the rate span into  $K$  sections. For notational convenience, we denote  $L_0$  for state 1 and  $L_K$  for state  $N$ . In the  $k$ th section,  $k = 1, \dots, K$ , the traffic flow with the corresponding arrival rates is serviced by the coded transmission rate  $\mu^{(k)}$ . This  $K$ -section process is referred to as a partitioned Markov modulated process. In what follows, we propose two strategies to partition the rate span, *server-wins-all* and *state-selective*.

*Server-wins-all* (where mean arrival rates in each section are upper bounded by the given service rate) — The rate partition is based on the following rule:  $\mu^{(k)} > \lambda_{L_k} + v_{L_k}$ ,  $k = 1, \dots, K-1$ . For section  $K$ ,  $\mu^{(K)} \geq \lambda_N$ , to be consistent with the criteria in determining the number of partitions. With a large buffer for packets waiting for transmission, the packet loss probability is very low in this case.

*State-selective* (where mean arrival rates in each section can be larger than the given service rate) — Consider the  $k$ th section with service rate  $\mu^{(k)}$ . The service rate of the  $k$ th section falls between two consecutive boundary states, i.e.,  $\lambda_{L_{k-1}} + v_{L_{k-1}} < \mu^{(k)} < \lambda_{L_k} + v_{L_k}$ . The buffer occupancy tends to increase when the states in the  $k$ th section have the arrival rates larger than  $\mu^{(k)}$  and these states are called *overload* states; otherwise, the buffer occupancy tends to decrease when the arrival rates of the states are less than  $\mu^{(k)}$  and these states are called *underload* states [7]. Hence, the  $k$ th section of the Markov chain is further partitioned into two sets,  $\mathcal{Z}_F^k \triangleq \{i : L_{k-1} < i \leq L_k | \lambda_i + v_i > \mu^{(k)}\}$  for overload states and  $\mathcal{Z}_E^k \triangleq \{i : L_{k-1} < i \leq L_k | \lambda_i + v_i < \mu^{(k)}\}$  for underload states, where  $k = 1, 2, \dots, K$ . Furthermore, the Markov chain with the  $K$  sections is constructed by interleaving *overload* and *underload* states. If '0' and '1' are used to represent overload and underload states, respectively, the Markov chain can be represented by an  $N$ -dimensional binary vector.

For comparison, in *server-wins-all* mode, since the transmission rate in each partition is set to be larger than all arrival rates in the same partition, better loss rate as well as delay performance can be achieved. However, in packet transmission, a small delay tolerance is allowed and the

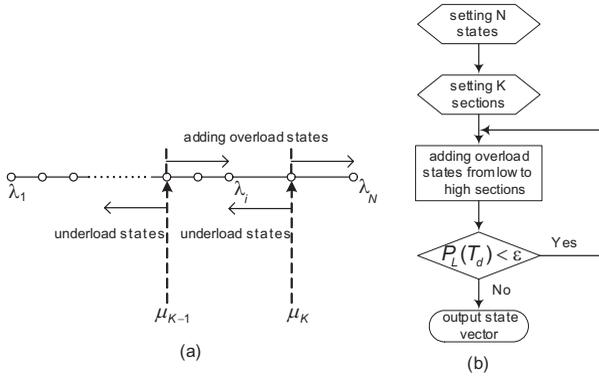


Fig. 3. Selecting overload states.

arrived packets can be queued when the instant arrival rate is larger than the service rate. In addition, in a multi-service-rate system, the delay tolerance provides an *inter-state multiplexing gain* in the sense that a packet queued in the overload states may be transmitted (rate-compensated) in the underload states. Therefore, the *state-selective* mode, where the service rate can be less than some arrival rates in the given section, could further improve the system utilization with satisfactory packet loss and delay requirements by a properly designed partitioning scheme.

### B. Adaptive Rate Allocation

In this section, an adaptive rate allocation scheme is proposed to properly set the overload and underload states so that the QoS is assured and the wireless resource can be highly utilized. Consider one video stream whose arrival rate is evenly quantized with  $N$  quantization levels. The quantization level  $N$  is determined by the rate levels of the OVFS system that covers the quantized arrival rate span with  $\mu^{(K)} \geq \lambda_N$ .

Initially, the states are partitioned such that in each section, all the arrival rates are less than the allocated service rate, i.e., all the states are underload. We then insert the overload states to each section except section  $K$ . When the state status changes, the boundary of the section also changes. The overload states are added from the low to the high sections, as shown in Fig. 3(a). When the overload state in the  $(K - 1)$ th section reaches  $\lambda_N$ , we reset this section as the highest, and update the number of sections from  $K$  to  $K - 1$ .

Each time an overload state is added, the state vector switches from '1' to '0' for that state with the section boundary changed, and the packet loss probability,  $P_L(T_d)$ , which will be derived in Section IV, in terms of delay bound  $T_d$  and channel error, is evaluated and compared with a predefined threshold  $\varepsilon$ , shown in Fig. 3(b). The process of adding overload states in each section continues until the threshold is exceeded. When adding overload states in a lower section, only the upper boundary of this section may be affected. It means that the number of required sections (rate levels) can be reduced. The output state vector is filled with interleaved 1's and 0's. Since the service rates cover quite a large span, two or three levels of OVFS codes should be sufficient to cover the arrival rates. Therefore, the iterations for setting selective states are very limited. In Section IV, further elaboration on

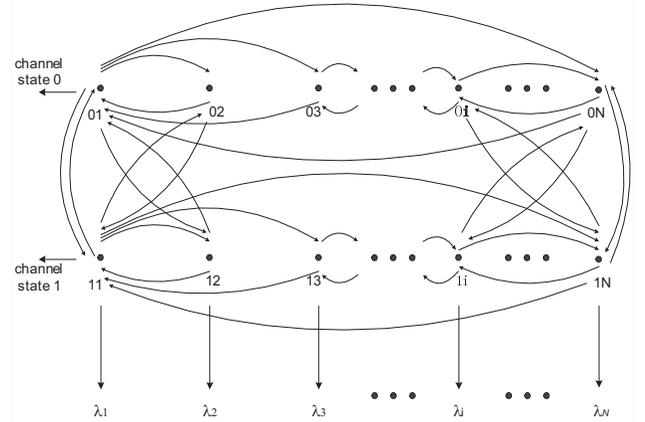


Fig. 4. MMRP model for two-state bursty channel.

the channel utilization and rate adaptation will be provided when the packet loss probability is derived.

### C. Traffic Rate Prediction

The rate allocation scheme is based on frame-by-frame transmission. If the service is of the wait-to-go mode, i.e., the arrived traffic in the current frame has to wait for service at least till the next frame, buffer occupancy can be measured and used for rate adaptation, but also introduce transmission delay. Moreover, due to the condition of code blocking, the required high service rate for video may randomly block other ongoing services. In order to achieve efficiency, code administration and QoS satisfaction, rate prediction is used for the semantics of frame-by-frame scheduling in OVFS-CDMA, since this can facilitate the implementation of the rate adaptation process.

It is observed in various experiments that there exists long or short term correlation in video sequences [7], [19]. This makes the video traffic more predictable as compared with data traffic. It is shown in [8] that accurately predicting video traffic rate benefits the QoS provisioning and bandwidth utilization. In this paper, one-step prediction is considered for simplicity, which is based on the available one-step transition probability matrix of the discrete Markov chain, and the predictor is referred to as *MMRP predictor*. Given the arrival rate  $R_n$  of the current frame, which can be quantized to a rate level of the Markov chain, say  $\lambda_h$ ,  $h = 1, \dots, N$ , the incoming mean rate  $R_{n+1}$  of the next frame can be predicted by

$$\begin{aligned}
 R_{n+1} &= \\
 &= \sum_{j=1}^N \Pr[\text{next state} = j | \text{current state} = h] \cdot \lambda_j \\
 &= \sum_{j=1}^N p_{hj} \lambda_j
 \end{aligned} \tag{1}$$

where  $p_{hj}$  is the transition probability as defined in Section II.

## IV. QOS PERFORMANCE ANALYSIS

In this section, the packet loss probability and the resource utilization efficiency of the proposed adaptive rate allocation scheme is analyzed based on the fluid model analysis [7].

In the transmitter, the arrived packets are buffered if they are not transmitted immediately. The erroneous packets will be selectively retransmitted by assuming that the acknowledgement can be received in one frame such that the retransmission can be carried out in the next frame. The retransmission packets are served with higher priority over the new arrival packets in each frame. Taking into account the possible retransmission, the effective arrive rate at state  $i$  becomes  $\lambda_i + v_i$ , where  $v_i$  denotes the average retransmission rate. For the randomly erroneous channel with small average error rate  $\epsilon$ , i.e., small retransmission rate  $v_i$ , an upper bound of the mean retransmission rate can be calculated from the  $N$ -state partitioned Markov model of Fig. 2 (see Appendix).

$$v_i \leq \frac{(1-\epsilon)\epsilon}{1-2\epsilon} \left( \sum_{j \in \{j: \lambda_j < \mu_j\}} p_{ji} \lambda_j + \sum_{j \in \{j: \lambda_j > \mu_j\}} p_{ji} \mu_j \right) \quad i, j = 1, \dots, N. \quad (2)$$

The wireless channel is represented by the two-state Gilbert model. The  $N$ -state Markov chain for the video traffic flow is expanded to a two-dimensional partitioned Markov chain with  $2N$  states, as shown in Fig. 4. In this figure, states (01) to (0N) are  $N$  arrival rates when the channel condition is good (channel state 0), while states (11) to (1N) are states when the channel condition is bad (channel state 1). Let states 1 to  $N$  represent the states (01) to (0N), and states  $N+1$  to  $2N$  represent the states (11) to (1N). Given the transition probabilities of the Gilbert model, the transition probability matrix of the underlying  $2N$ -state discrete Markov chain,  $\mathbf{P}_{2N}$ , which has dimension  $2N \times 2N$ , can be obtained as

$$\mathbf{P}_{2N} = \begin{bmatrix} \mathbf{P}(1-q) & \mathbf{P}q \\ \mathbf{P}(1-q) & \mathbf{P}q \end{bmatrix} \quad (3)$$

where  $\mathbf{P}$  is the  $N \times N$  transition matrix of source rates. The steady state probability vector  $\boldsymbol{\Omega} = (\Omega_1, \Omega_2, \dots, \Omega_{2N})$  can be calculated from  $\mathbf{P}_{2N}$ , or by  $\boldsymbol{\Omega} = [\boldsymbol{\pi}(1-q), \boldsymbol{\pi}q]$ .

The infinitesimal generator matrix,  $\mathbf{M}_{2N}$ , for the continuous-time  $2N$ -state Markov chain is a  $2N \times 2N$  matrix given by  $\mathbf{M}_{2N} = \mathbf{I} - \mathbf{P}_{2N}$ , where  $\mathbf{I}$  is a  $2N \times 2N$  identity matrix. Due to the random channel, there is also a retransmission rate associated with each of the  $2N$  states. Assuming that, in the erroneous states (states (11) to (1N)), the service rate is set sufficiently large to cover the arrival rates, i.e., the *server-wins-all* mode, a conservative approximation of the mean of the retransmission rates  $v_i$  in state  $i$  ( $i = 1, \dots, 2N$ ) can be calculated as

$$v_i \cong \epsilon_1 \sum_{j \in \{\text{all erroneous states}\}} p'_{ji} \lambda_j \quad (4)$$

where  $p'_{ji}$  is the  $j$ th element of matrix  $\mathbf{P}_{2N}$ ,  $i, j = 1, \dots, 2N$ . In the following,  $v_i$  will also be used as the approximate retransmission rate for *state selective* since in the bad channel, most states work under the *server-wins-all* mode.

The linear differential equation of the  $2N$ -state Markov model in vector form is given by

$$\frac{d\mathbf{F}_{2N}(x)}{dx} = \mathbf{F}_{2N}(x)\mathbf{M}_{2N}\mathbf{D}_{2N}^{-1} \quad (5)$$

where  $\mathbf{F}_{2N}(x) \triangleq [F_1(x), F_2(x), \dots, F_{2N}(x)]$ ,  $F_i(x) = \Pr[X \leq x, S = i]$ ,  $X$  and  $S$  are random variables

denoting the buffer occupancy and the MMRP state respectively, and  $\mathbf{D}_{2N} \triangleq \text{diag}[\lambda_i + v_i - \mu_i]$ ,  $1 \leq i \leq 2N$ . Consider different partition methods (service rate allocation schemes) for states 1 to  $N$  and for states  $N+1$  to  $2N$ . There can be a total of  $2N$  interleaved *underload* and *overload* states for the bursty channel, given as  $\mathcal{Z}_E = \{i \in 2N | \lambda_i + v_i < \mu_i\}$  for underload states, and  $\mathcal{Z}_F = \{i \in 2N | \lambda_i + v_i > \mu_i\}$  for overload states.

The solution of the linear differential equations given in (5) can be readily obtained by the sums of exponentials

$$\mathbf{F}_{2N}(x) = \sum_{j=1}^{2N} a_j \boldsymbol{\Phi}_j e^{z_j x} \quad (6)$$

where  $(z_j, \boldsymbol{\Phi}_j = [\Phi_{j1}, \Phi_{j2}, \dots, \Phi_{j2N}])$  is the (eigenvalue, eigenvector) pair satisfying the eigenvalue equation  $z_j \boldsymbol{\Phi}_j \mathbf{D}_{2N} = \boldsymbol{\Phi}_j \mathbf{M}_{2N}$ ,  $1 \leq j \leq 2N$ . The values of the coefficients  $a_j$ 's can be determined by the additional information such as initial values. From  $\boldsymbol{\Omega} \mathbf{M}_{2N} = \mathbf{0}$ ,  $(0, \boldsymbol{\Omega})$  is one of the eigen-pairs. Let  $\boldsymbol{\Phi}_1 = \boldsymbol{\Omega}$  and  $z_1 = 0$ , (6) becomes

$$\mathbf{F}_{2N}(x) = a_1 \boldsymbol{\Omega} + \sum_{j=2}^{2N} a_j \boldsymbol{\Phi}_j e^{z_j x}. \quad (7)$$

Define the *traffic intensity*  $\rho$  as

$$\begin{aligned} \rho &= \frac{\text{Average arrival rate} + \text{Average retransmission rate}}{\text{Average service rate}} \\ &= \frac{\boldsymbol{\Omega}(\boldsymbol{\lambda} + \mathbf{v})^T}{\boldsymbol{\Omega}\boldsymbol{\mu}^T} \end{aligned} \quad (8)$$

where  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_{2N})$ ,  $\mathbf{v} = (v_1, \dots, v_{2N})$  and  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_{2N})$ . Note that it is necessary that the traffic intensity should be evaluated whenever the service rate changes, such that  $\rho < 1 - \delta$ , where  $\delta$  is a reliability margin parameter. If  $\rho < 1$ , we say the system is *underloaded* (stable); otherwise, it is *overloaded*. When  $\rho < 1$ , for a sufficiently large buffer,  $F_i(x) \leq \Omega_i$ ; and when  $x \rightarrow +\infty$ ,  $F_i(x) \rightarrow \Omega_i$ . To solve (7), we can set  $a_1 = 1$ . Since all the positive eigenvalues are impractical and have to be ruled out, the associate coefficients  $a_j = 0$  for  $z_j > 0$ . Therefore, (7) can be rewritten as

$$\begin{aligned} F_i(x) &= \Omega_i + \sum_{j \in \{z_j < 0\}} a_j \Phi_{ji} e^{z_j x} \\ &= \Omega_i - \sum_{j \in \{z_j < 0\}} w_{ji} e^{z_j x}, \quad i = 1, \dots, 2N \end{aligned} \quad (9)$$

where  $w_{ji}$  ( $w_{ji} \geq 0$ ) is the weight to be determined.

Due to the stringent delay bound requirement of the real-time traffic, when the delay exceeds a threshold, the packet is dropped. Thus, we set virtual bounds  $\{x_i\}$  in the buffer associated with each state of the Markov chain. In the underload states, the buffer occupancy is low so that  $\Pr[X > x_i, S = i | i \in \mathcal{Z}_E] = 0$ . Since  $\Pr[X > x_i, S = i | i \in \mathcal{Z}_E] = \Omega_i - F_i(x_i | i \in \mathcal{Z}_E)$ , we get  $\sum_{j \in \{z_j < 0\}} w_{ji} e^{z_j x_i} = 0$ . As a result, for all the underload states,  $w_{ji} = 0$ . In the overload states,  $\Pr[X > x_i, S = i | i \in \mathcal{Z}_F] = \Omega_i - F_i(x_i | i \in \mathcal{Z}_F) = \sum_{j \in \{z_j < 0\}} w_{ji} e^{z_j x_i}$ . The probability of buffer overflow for an overload state,  $G_i(x)$ , is the weighted sum of exponential

expressions in terms of negative eigenvalues associated with overload states, i.e.,

$$G_i(x) \triangleq \sum_{j \in \{z_j < 0\}} w_{j_i} e^{z_j x}. \quad (10)$$

Although  $\rho < 1$  guarantees that the system is stable, during any epoch, the instantaneous traffic intensity can exceed 1 causing a temporary overload. For a sufficiently large buffer, the dropping probability in an overload state becomes asymptotically stable, i.e., when  $x \rightarrow \infty$ ,  $e^{-|z_j|x} \rightarrow 0$ . Therefore,  $G_i(x)$  can be governed by a few lower eigenvalues (the negative eigenvalues with small magnitude), referred to as *significant* eigenvalues<sup>2</sup>. These *significant* eigenvalues are selected to be the largest negative ones in descending order, noted as  $\{z_j^*, j = 1, \dots, J$ , and  $J$  is the number of significant eigenvalues. The dominant and the significant eigenvalue are the same when  $J = 1$ .

At the initial instant of a transmission frame, assuming delay is 1 frame, i.e., in time interval  $0^+ \leq t \leq 1^-$ , there is no buffer tolerance ( $x = 0$ ), then the difference between the arrival rate and service rate plus the number of erroneous packets per frame is the number of dropped packets. If the weights are set all equal, i.e.,  $w_{j_i} = w_i$ ,  $j = 1, \dots, J$ , then  $w_i$  of the random channel overload state  $i$  can be obtained by solving

$$\begin{aligned} G_i(0^+) &= \sum_{j=1}^J w_i e^{z_j^* 0^+} \\ &= \frac{[\max\{(\lambda_i - \mu_i), 0\} + \min\{\mu_i, \lambda_i\} \cdot \epsilon] \cdot \Omega_i}{\Omega \lambda^T}, \\ & \quad i \in \{\text{overload states}\} \end{aligned} \quad (11)$$

Thus,

$$w_i = \frac{[\max\{(\lambda_i - \mu_i), 0\} + \min\{\mu_i, \lambda_i\} \cdot \epsilon] \cdot \Omega_i}{J \cdot \Omega \lambda^T}.$$

For the two-state channel, the error rate in the two states can be averaged out, thus the weight of an overload state can be decided by  $w_i = (\lambda_i + v_i - \mu_i) \cdot \Omega_i / [J \cdot \Omega \lambda^T]$ . Asymptotically, given the delay bound  $T_d = D$ , the packet dropping happens in all the overload states  $\mathcal{Z}_F$  when the buffer occupancy exceeds the virtual buffer occupancy  $x(D)$ , which is defined as the virtual buffer length corresponding to the delay bound  $D$ . Considering that  $\mu_i(t)$  is the rate of the current time  $t$  at state  $i$ ,  $x_i(D) = \sum_{j=1}^D \mu_i(t+j)$  is a random variable depending on the service rates of the next  $D$  frames. Denote  $t_d$  the time from the moment of a packet arrival to the moment when the packet is successfully received. We have

$$\begin{aligned} \Pr[t_d > T_d, S = i] &= G_i[X > x_i(D)] \\ &= \sum_{j=1}^J w_i \left\{ \sum_{\{\cup(\mu_i)\}} e^{z_j x_i(D)} \right. \\ & \quad \cdot \Pr[\mu_i(t+1), \mu_i(t+2), \dots, \mu_i(t+D)] \left. \right\} \end{aligned} \quad (12)$$

where  $\{\cup(\mu_i)\}$  is the set of all possible combinations of the rate sequence  $[\mu_i(t+1), \mu_i(t+2), \dots, \mu_i(t+D)]$ . Consider the

<sup>2</sup>The *significant* eigenvalues include the *dominant* (the largest negative) eigenvalue and a few next largest negative eigenvalues.

facts that 1) the number of service rate levels (sections)  $K$  is small, e.g.,  $K = 2$ ; 2) when  $D$  is small, since there is strong correlation among the arrival rates, it can be assumed that the service rate does not change much during  $D$ ; and 3) when  $D$  is large, the summation  $x_i(D) = \sum_{j=1}^D \mu_i(t+j)$  is more likely dependent on  $D$ , given  $\mu_i$  varying in a small number of  $K$  levels. The approximate buffer bound  $x_i(T_d) \cong T_d \mu_i$ , and the total packet dropping probability  $P_d(T_d)$  due to delay is given by

$$\begin{aligned} P_d(T_d) &= \Pr[t_d > T_d] = \sum_{i \in \mathcal{Z}_F} G_i(T_d \mu_i) \\ &= \sum_{i \in \mathcal{Z}_F} \sum_{j=1}^J w_i e^{z_j^* T_d \mu_i}. \end{aligned} \quad (13)$$

The problem of finding the packet drop probability due to delay is solved by simply finding the significant eigenvalues. Our interest is to find the total packet loss probability due to delay and error. If an erroneous packet is within its delay bound, it can be continuously retransmitted until being correctly received or exceeding the delay bound. Given the delay bound  $T_d = D$  frames, and assuming that the round trip acknowledgement can be received within one frame, with the average channel error rate  $P(E) = \epsilon$ , the total packet loss probability  $P_L$  due to delay and channel error can be obtained by (14), shown at the top of the next page, where  $\Delta(\epsilon) = \epsilon P_d(D) + \dots + \epsilon^{D-1} P_d(2)$ . Equation (14) is the conservative evaluation of the packet loss probability when  $\epsilon$  is small and  $\Delta(\epsilon)$  is negligible.

Two performance metrics are defined to evaluate the proposed adaptive rate allocation scheme: the channel utilization  $\gamma$ , defined as the ratio of traffic throughput rate over the channel capacity (the average service rate), and the channel efficiency  $\eta$ , defined as the ratio of goodput rate (packets correctly received in unit time) over the channel capacity, given by  $\eta = [1 - P_L(T_d)] \cdot \Omega \lambda^T / [\Omega \mu^T]$ .

The service rate is usually fixed in an OVFS-CDMA system. The objective of the adaptive rate allocation is to properly partition the sections within the arrival rate span of the Markov chain and to select suitable states to be served by a certain level of the transmission rate, such that the packet loss probability and the delay requirement can be satisfied and high channel utilization can be achieved. To pursue an optimal rate allocation scheme in terms of maximal channel efficiency, we may look at the maximization of  $\eta$  in terms of rate vector  $\mu$ , subject to  $\rho = \frac{\Omega(\lambda+v)^T}{\Omega \mu^T} \leq 1 - \delta$  and  $P_L(T_d) \leq \epsilon$ , where  $\delta$  and  $\epsilon$  are predefined thresholds. The optimal solution may be found by direct search over rate vector  $\mu$  and manipulating the binary state vector. However, this issue is beyond the focus of this paper and can be a future investigation.

## V. NUMERICAL RESULTS AND DISCUSSION

In this section, numerical results of the proposed adaptive rate allocation scheme based on computer simulations and mathematical analysis are presented. The suitability and scalability of the proposed scheme will be discussed later.

$$\begin{aligned}
P_L(T_d) &= P(t_d > D) + P(E)P(D-1 < t_d < D) + \dots + [P(E)]^{D-1}P(1 < t_d < 2) \\
&= P_d(D) + P(E)[P_d(D-1) - P_d(D)] + \dots + [P(E)]^{D-1}[P_d(1) - P_d(2)] \\
&= P_d(D) + P(E)P_d(D-1) + \dots + [P(E)]^{D-1}P_d(1) - \Delta(\epsilon) \\
&\cong P_d(D) + P(E)P_d(D-1) + \dots + [P(E)]^{D-1}P_d(1)
\end{aligned} \tag{14}$$

TABLE I  
SIGNIFICANT EIGENVALUES

Average Channel PER	$z_1^*$	$z_2^*$
0.01	-0.0322	-0.0962
0.02	-0.0236	-0.0844
0.03	-0.0165	-0.0732
0.04	-0.0108	-0.0631
0.05	-0.0065	-0.0544
0.06	-0.0034	-0.0471

### A. System Parameters

In the simulation, the radio frame of CDMA downlink transmission is set to 20 ms (though it can be the multiple of 10 ms). Twenty video sequences are used in the simulation to characterize the statistics of video traffic. Each video flow has a mean of 0.52 bit/pixel (equivalent to the mean value of 20 packets/frame by proper conversion), standard deviation of 0.23 bit/pixel, and time constant<sup>3</sup> equal to 3.9. The video packet flow is generated using the autoregressive model described by  $\lambda(n) = \max\{0.925\lambda(n-1) + 0.088w(n), 0\}$ , where  $w(n)$  is a stationary Gaussian variable with mean  $E(w) = 0.4434$  and unit variance. The simulated sequence is then represented by the MMRP model. The observed rate span of the video sequence is  $[0, 50+)$  packets/frame, which is evenly quantized into eight rate levels (states), i.e.,  $\lambda = [3 \ 9 \ 15 \ 21 \ 27 \ 33 \ 39 \ 45]$ , named states 1 to 8 in ascending order.

The monitored discrete transition probability  $\mathbf{P}$  and the stationary probability  $\pi$  based on the simulated sequence of 1000 frames are given by the equation at the top of the next page, and  $\pi = [0.0501 \ 0.1107 \ 0.2206 \ 0.2725 \ 0.2032 \ 0.1034 \ 0.0333 \ 0.0062]$ . The two-state channel parameter is  $q = 0.3$ .

### B. Packet Loss Probability and Channel Efficiency

The transition probability matrix  $\mathbf{P}_{2N}$  and stationary probability  $\Omega$  can be readily obtained through (3). The infinitesimal generator matrices  $\mathbf{M}$  and  $\mathbf{M}_{2N}$  can then be calculated. Consider that the leaf rate is 20, and two layers of the codes can be used,  $R_1 = 20$  and  $R_2 = 40$ . Then the number of sections of each eight-state traffic rate span is two. An example of adaptive rate design is: for the quantized rate span  $\lambda$ , Rate Scheme 0 =  $[R_1 R_1 R_1 R_1 R_1 R_2 R_2 R_2]$  for the good channel states, and Rate Scheme 1 =  $[R_1 R_1 R_1 R_1 R_2 R_2 R_2 R_2]$  for the bad channel states, respectively. Specifically, the eight states of the Markov modulated process under the error free channel are divided into two sections: section 1 having states (1, 2, 3, 4, 5) with service rate  $R_1$ , and section 2 having

<sup>3</sup>Time constant represents the exponential behavior of the covariance of the video sequence, see [7].

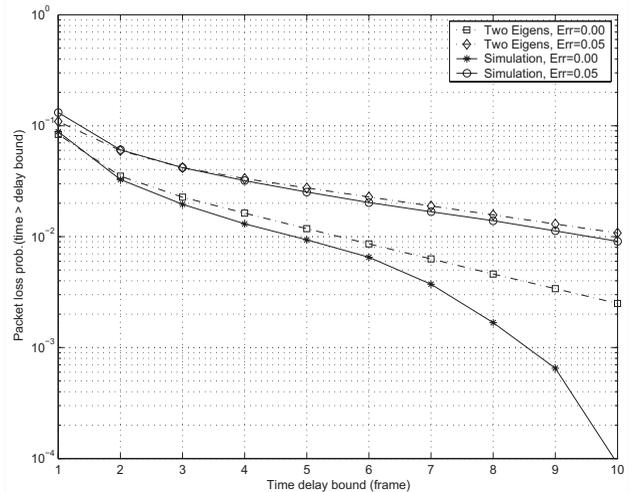


Fig. 5. Packet loss probability versus the delay bound for the random channel.

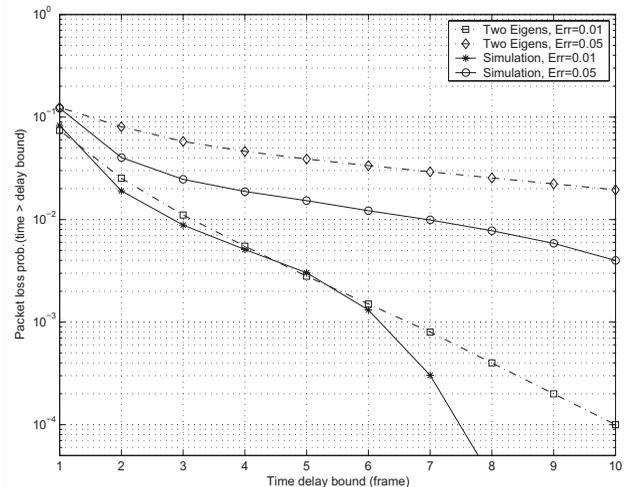


Fig. 6. Packet loss probability versus the delay bound for the two-state channel.

states (6, 7, 8) with service rate  $R_2$ . The state vector is  $[1 \ 1 \ 1 \ 0 \ 0 \ 1 \ 1 \ 0]$ .

Based on the Rate Schemes 0 and 1, the analytical results of packet loss probabilities are then obtained in terms of various delay requirements, channel types (error-free channel, random error channel, and Gilbert model) and channel errors. For (13), the number of significant eigenvalues is selected to be  $J = 2$ . Table I shows the variation of significant eigenvalues with respect to the change of channel error rate. It is observed that as the average channel PER increases, the significant eigenvalues become larger.

The packet loss probabilities for random and bursty channels with different average PERs are shown in Figs. 5 and 6,

$$\mathbf{P} = \begin{bmatrix}
 .6819 & .2946 & .0236 & 0 & 0 & 0 & 0 & 0 \\
 .1325 & .5542 & .2878 & .0155 & 0 & 0 & 0 & 0 \\
 .0057 & .1472 & .5596 & .2730 & .0144 & 0 & 0 & 0 \\
 0 & .0072 & .2180 & .5696 & .2008 & .0044 & 0 & 0 \\
 0 & .0007 & .0177 & .2641 & .5695 & .1437 & .0043 & 0 \\
 0 & 0 & 0 & .0163 & .2790 & .5777 & .1196 & .0075 \\
 0 & 0 & 0 & 0 & .0221 & .3879 & .4922 & .0978 \\
 0 & 0 & 0 & 0 & 0 & .0544 & .5922 & .3534
 \end{bmatrix}$$

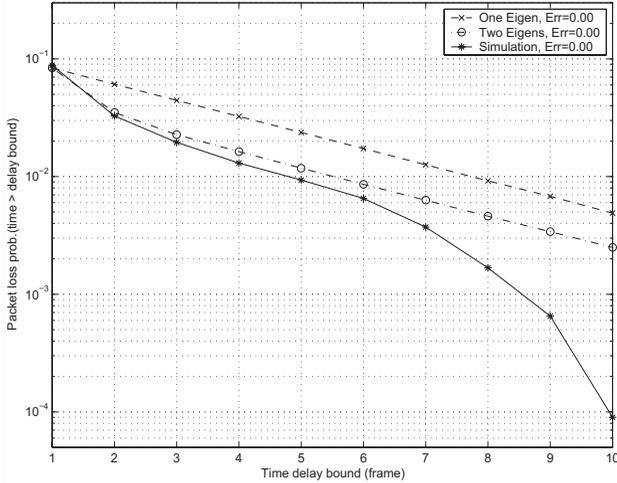


Fig. 7. Packet loss probabilities for different number of eigenvalues.

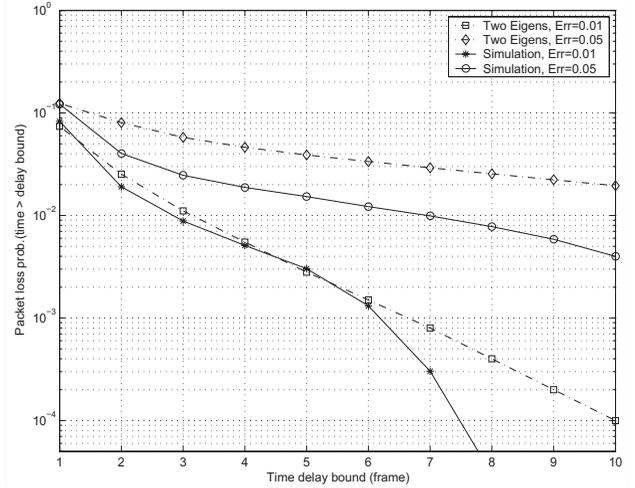


Fig. 9. Channel efficiency versus the delay bound for the two-state channel.

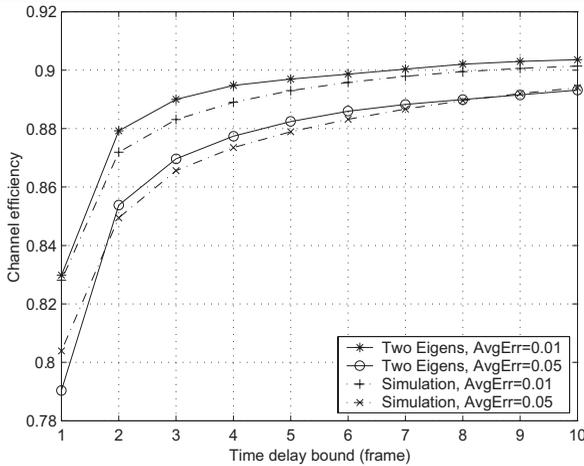


Fig. 8. Channel efficiency versus the delay bound for the random channel.

respectively. It is observed that the loss probability decreases as the delay bound increases, and the analytical results agree well with the simulation results for random channel and for bursty channel with PER of 0.01, respectively. As bursty PER is increased to 0.05, the analytical result is conservative, partially due to the conservative estimations of  $v_i$ 's and  $P_L$ . It is also observed that the analytical results are conservative compared with the simulation ones when the delay bound exceeds 6. For comparison, we also calculate the loss probability using only the dominant eigenvalue. The comparison between results obtained using the dominant eigenvalue, two significant eigenvalues and the simulation for an error free

(PER=0) channel are shown in Fig. 7. It can be seen that using the weighted sum of the significant eigenvalue exponential expressions can achieve better approximation of the packet loss probability than that using the dominant eigenvalue with the tradeoff of computational complexity.

Given the quantized arrival rate  $\lambda$  and the service rate allocation schemes, we can calculate the steady state channel efficiencies. For an error free channel, with state dependent service rate  $\mu = [20 \ 20 \ 20 \ 20 \ 20 \ 40 \ 40 \ 40]$ , if the loss probability is negligible, the channel efficiency can be calculated by  $\eta = \lambda \pi^T / [\mu \pi^T] = 0.9037$ . If the overload state is dealt with conservatively by setting  $\mu = [20 \ 20 \ 20 \ 20 \ 40 \ 40 \ 40 \ 40]$  which might be similar to the *server-wins-all* mode, the asymptotic efficiency tends to  $\eta = 0.7627$ . Considering the two-state channel, the service rates are designed as  $\mu_0 = [20 \ 20 \ 20 \ 20 \ 20 \ 40 \ 40 \ 40]$  for channel state 0, and  $\mu_1 = [20 \ 20 \ 20 \ 20 \ 40 \ 40 \ 40 \ 40]$  for channel state 1. By using the stationary probability  $\Omega$  and assuming that the loss probability is negligibly small, the channel efficiency is obtained by  $\eta = [\lambda; \lambda] \Omega^T / \{[\mu_0; \mu_1] \Omega^T\} = 0.8579$ . Figs. 8 and 9 demonstrate the transmission efficiency over the random and two-state bursty channels. It is observed that a less stringent delay requirement leads to a higher resource utilization efficiency, and the analytical results agree quite well with the simulation results.

For comparison of the proposed rate allocation scheme with some existing schemes, we assume no channel error and apply the *effective bandwidth* method [7], [22] for the same video

TABLE II  
COMPARISON OF CHANNEL UTILIZATION

Approaches	Dropping prob.	Channel utilization	
		regular (test) case	ideal case
effective bandwidth	less than $10^{-4}$	50.00%	58.82%
server-wins-all	less than $10^{-4}$	76.72%	–
state-selective (1)	$10^{-4}$	86.56%	–
state-selective (2)	$10^{-3}$	88.64%	–
state-selective (3)	$10^{-2}$	91.96%	–

sequence, given by

$$C_E = A \cdot M \cdot [(1 - \kappa)/2 + \sqrt{((1 - \kappa)/2)^2 + \kappa u}]$$

where  $M$  is the number of ON-OFF mini-sources representing the video sequence,  $A$  is the rate in the ON state of each mini-source,  $\kappa = \frac{\beta \cdot C_E \cdot T_d}{A(1-u) \cdot \ln(1/P_L)}$ ,  $u = \frac{\alpha}{\alpha + \beta}$ , and  $C_E T_d$  represents the size of queuing buffer for the video. By setting the design requirements  $T_d = 3$ ,  $P_L = 10^{-2}$  and  $M = 4$ , the effective bandwidth  $C_E = 34$  (packets/frame), given the effective transmission rate, we can obtain the packet dropping probability of  $10^{-4}$  in the simulation of the error free channel. Ideally, if  $C_E$  happens to be the leaf rate of the code tree, and taking into account  $E(\lambda) = \lambda \pi^T = 20$ , the channel efficiency (utilization) is  $\frac{20}{34} = 0.5882$ . If the leaf rate is 20 or 40, then the efficiency of the *effective bandwidth transmission rate* scheme ( $C_E = 40$ ) is given by  $\eta = \frac{20}{40} = 0.50$ . That is, the channel utilization can be significantly improved by the proposed rate adaption scheme.

Finally, we summarize the comparison results of the proposed approaches with the *effective bandwidth* scheme for an error free channel in Table II. In general the delay requirement for video is around 60-100 ms. With a frame length of 20 ms, the delay bound  $T_d$  in frames is reasonably set to  $T_d = 3$ . It can be seen that the proposed adaptive rate allocation scheme significantly improves the channel utilization of the OVSF-CDMA system compared with the effective bandwidth scheme.

## VI. CONCLUSIONS

In this paper, an adaptive rate allocation scheme for real-time video traffic in OVSF-CDMA system has been proposed. By modeling the video traffic and wireless channel as a joint Markov modulated process, an analytical approach based on segmentation of the rate span of the Markov process has been developed for adaptively allocating transmission rate and provisioning QoS in terms of delay and packet loss over the wireless channels. Simulation results show that the proposed approach is robust as long as the system is not overloaded, and is reasonably accurate. Compared with the commonly used effective bandwidth scheme, the proposed rate allocation scheme achieves superior QoS performance and significant improvement on the channel (code) utilization.

## ACKNOWLEDGEMENT

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 (NSERC) of Canada under the Strategic Project program. The authors wish to thank the anonymous reviewers for their helpful reviews and suggestions.

## APPENDIX

### DERIVATION OF EQUATION (A-2)

Let the delay bound be  $D$ , and the erroneous packet can be selectively retransmitted within one frame. Let  $t_d$  be the time from the moment of a packet arrival to the moment when the packet is successfully received. Denote the retransmission rate  $v_i$  at state  $i$ . Given the average PER  $P(E)$ , the new erroneous packets that need retransmission per frame are  $e_i = \lambda_i \cdot P(E)$ , for  $i \in \{i : \lambda_i < \mu_i\}$ , and  $e_i = (\mu_i - v_i) \cdot P(E)$ , for  $i \in \{i : \lambda_i > \mu_i\}$ , respectively. The remaining packets in the frame,  $Q_{i,remain}$ ,  $i \in N$  in the retransmission queue after one retransmission is

$$\begin{aligned} Q_{i,remain} &= v_i \cdot [P(E)P(t_d \leq D - 1) + P^2(E) \\ &\quad P(t_d \leq D - 2) + \dots + P^{D-1}(E)P(t_d \leq 1)] \\ &\leq v_i \cdot \frac{P(E)}{1 - P(E)} \end{aligned} \quad (A-1)$$

Given the new and remaining erroneous packets in frame  $n$ , the retransmission rate of frame  $n + 1$  in state  $i$  is shown at the top of the next page. Substituting  $\epsilon$  for  $P(E)$  and taking expectations of  $v_i$ 's on both sides of (A-3), we obtain (A-2).

## REFERENCES

- [1] C. Smith and D. Collins, *3G Wireless Networks*. New York: McGraw-Hill Telecom, 2002.
- [2] P.-R. Chang and C.-F. Lin, "Design of spread spectrum multicode CDMA transport architecture for multimedia services," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 99–111, Jan. 2000.
- [3] R. Fantacci and S. Nannicini, "Multiple access protocol for inter-gration of variable bit rate multimedia traffic in UMTS/IMT-2000 based on wideband CDMA," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 1441–1454, Aug. 2000.
- [4] O. Sallent, J. Perez-Romero, F. Casadevall, and R. Agusti, "An emulator framework for a new radio resource management for QoS guaranteed services in W-CDMA systems," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 1893–1904, Oct. 2001.
- [5] I. M. Kim and H. M. Kim, "An Optimum Power Management Scheme for Wireless Video Service in CDMA Systems," *IEEE Trans. Wireless Commun.*, vol. 2, pp. 81–91, Jan. 2003.
- [6] 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.
- [7] M. Schwartz, *Broadband Integrated Networks*. Upper Saddle River, NJ: Prentice Hall, 1996.
- [8] M. Wu, R. Joyce, H.-S. Wong, L. Guan, and S.-Y. Kung, "Dynamic resource allocation via video content and short-term traffic statistics," *IEEE Trans. Multimedia*, vol. 3, pp. 186–199, June 2001.
- [9] S. Biswas and R. Izmailov, "Design of a fair bandwidth allocation policy for VBR traffic in ATM networks," *IEEE/ACM Trans. Networking*, vol. 8, pp. 212–223, Apr. 2000.
- [10] Q. Ren and H. Kobayashi, "Diffusion approximation modeling for Markov modulated bursty traffic and its applications to bandwidth allocation in ATM networks," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 679–691, June 1998.
- [11] T. Minn and K.-Y. Siu, "Dynamic assignment of orthogonal variable-spreading factor codes in W-CDMA," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 1429–1440, Aug. 2000.
- [12] A. Kam, T. Minn, and K.-Y. Siu, "Supporting rate guarantee and fair access for bursty data traffic in W-CDMA," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 2121–2130, Nov. 2001.
- [13] T.-L. Sheu and J.-H. Hou, "A code assignment scheme for cdma wireless networks with integrated voice and data traffic," in *Proc. IEEE Int'l Conf. on Telecom*, Feb./Mar. 2003, vol. 1, pp. 774–780.

$$v_i(n+1) = \sum_{j=1}^N p_{ji} e_j(n) + Q_{i,remain}(n) \quad (A-2)$$

$$\begin{aligned} &= \left( \sum_{j \in \{j: \lambda_j < \mu_j\}} p_{ji} \lambda_j + \sum_{j \in \{j: \lambda_j > \mu_j\}} p_{ji} \mu_j \right) P(E) \\ &\quad - \sum_{j \in \{j: \lambda_j > \mu_j\}} p_{ji} v_j(n) \cdot P(E) + Q_{i,remain}(n) \\ &\leq \left( \sum_{j \in \{j: \lambda_j < \mu_j\}} p_{ji} \lambda_j + \sum_{j \in \{j: \lambda_j > \mu_j\}} p_{ji} \mu_j \right) P(E) + Q_{i,remain}(n) \\ &\leq \left( \sum_{j \in \{j: \lambda_j < \mu_j\}} p_{ji} \lambda_j + \sum_{j \in \{j: \lambda_j > \mu_j\}} p_{ji} \mu_j \right) P(E) + v_i(n) \cdot \frac{P(E)}{1 - P(E)} \end{aligned} \quad (A-3)$$

- [14] J. E. Fossa and N. Davis, "Dynamic code assignment improves channel utilization for bursty traffic in third-generation wireless networks," in *Proc. IEEE International Conference on Communications*, Apr./May 2002, vol. 5, pp. 3061–3065.
- [15] S. Lin and D. J. Costello, Jr., *Error Control Coding: Fundamentals and Applications*. Upper Saddle River, NJ: Prentice-Hall, 1983.
- [16] J. S. Lee and L. E. Miller, *CDMA Systems Engineering Handbook*. Boston: Artech House, 1998.
- [17] E. N. Gilbert, "Capacity of burst noise channels," *Bell Syst. Tech. J.*, vol. 39, pp. 1253–1256, 1960.
- [18] S. Kallel, "Analysis of memory and incremental redundancy ARQ schemes over a nonstationary channel," *IEEE Trans. Commun.*, vol. 40, pp. 1474–1480, Sep. 1992.
- [19] B. Maglaris, D. Anastassiou, P. Sen, G. Karlsson, and J. D. Robbins, "Performance models of statistical multiplexing in packet video communications," *IEEE Trans. Commun.*, vol. 36, pp. 834–844, July 1988.
- [20] P. Skelly, M. Schwartz, and S. Dixit, "A histogram-based model for video traffic behavior in an ATM multiplexer," *IEEE/ACM Trans. Networking*, vol. 1, pp. 446–459, Aug. 1993.
- [21] S. Lee, *Spread Spectrum CDMA: IS-95 and IS-2000 for RF Communications*. New York: McGraw-Hill, 2002.
- [22] A. I. Elwalid and D. Mitra, "Effective bandwidth of general Markovian traffic sources and admission control of high speed networks," *IEEE/ACM Trans. Networking*, vol. 1, pp. 329–343, June 1993.



**Jon W. Mark** (M'62-SM'80-F'88-LF'03) received the B.A.Sc. degree from the University of Toronto, Toronto, Ontario, Canada in 1962 and the M.Eng. and Ph.D. degrees from McMaster University, Hamilton, Ontario, Canada in 1968 and 1970, respectively, all in electrical engineering.

From 1962 to 1970, he was an Engineer and then Senior Engineer at Canadian Westinghouse Co. Ltd., Hamilton, Ontario, Canada. In September 1970, he joined the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario, Canada, where he is currently a Distinguished Professor Emeritus. He served as the Department Chairman during the period July 1984–June 1990. In 1996 he established the Centre for Wireless Communication (CWC) at the University of Waterloo and is currently serving as its founding Director.

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

He has previously worked in the areas of adaptive equalization, image and video coding, spread spectrum communications, computer communication networks, ATM switches design, and traffic management. His current research interests are in broadband and wireless communication, resource and mobility management, and cross domain interworking. He recently co-author the text entitled *Wireless Communications and Networking*, Prentice-Hall 2003.

A life Fellow of IEEE, Dr. Mark is the recipient of the 2000 Canadian Award for Telecommunications Research and the 2000 Award of Merit of the Education Foundation of the Federation of Chinese Canadian Professionals. He has served as an editor of IEEE TRANSACTIONS ON COMMUNICATIONS (1983–1990), a member of the Inter-Society Steering Committee of the IEEEACM Transactions on Networking (1992–2003), a member of the IEEE Communications Society Awards Committee (1995–1998), an editor of *Wireless Networks* (1993–2004), and an associate editor of *Telecommunication Systems* (1994–2004).



**Jun Cai** received the B. Eng. Degree (1996) in radio techniques and the M. Eng. Degree (1999) in communication and information systems from Xi'an Jiaotong University, China, and the Ph.D. degree (2004) in electrical engineering from University of Waterloo, Canada. He is currently doing research as Postdoctoral Fellow in electrical and computer engineering, University of Waterloo, Canada. His research interests include channel estimation, interference cancellation, and resource management in wireless communication systems.



**Jun Xu** received B. Eng. from Tsinghua University, China, in 1989, and MASc from University of Waterloo, Canada, in 2002. He is currently pursuing the PhD degree in electrical and computer engineering, University of Waterloo, Canada. His research interests include mobility, QoS performance analysis and resource management in wireless 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. From September 1990 to September 1993, he was first with the Howard University, Washington D.C., and then the University of Alberta, Edmonton (Canada). Since October 1993, he has been 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 Qshine'05, Co-Chair for IEEE Broadnet'05, WirelessCom'05, IFIP Networking'05, IS-PAN'04, I