Wireless Networks 10, 133–146, 2004

MPEG-4 Video Transmission over Wireless Networks: A Link Level Performance Study

JI-AN ZHAO and BO LI *
Department of Computer Science, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China

CHI-WAH KOK
Department of Electrical and Electronic Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China

ISHFAQ AHMAD
Department of Computer Science and Engineering, University of Texas at Arlington, Arlington, TX 76019-0015, USA

Abstract. With the scalability and flexibility of the MPEG-4 and the emergence of the broadband wireless network, wireless multimedia services are foreseen to become deployed in the near future. Transporting MPEG-4 video over the broadband wireless network is expected to be an important component of many emerging multimedia applications. One of the critical issues for multimedia applications is to ensure that the quality-of-service (QoS) requirement to be maintained at an acceptable level. This is further challenged in that such a service guarantee must be achieved under unreliable and time-varying wireless channels. In this paper we study the link level performance of MPEG-4 video transmission over the uplink of an unreliable wireless channel. We introduce the discrete time batch Markovian arrival process (DBMAP) with two types of arrivals to model the MPEG-4 video source, which takes into account the inherent nature of the adaptiveness of the video traffic. We prove that in a hidden Markov modeled (HMM) wireless channel with probabilistic transmission, the service time for an arbitrary radio link control (RLC) burst follows phase type (PH-type) distribution. We show that the link level performance of a wireless video transmission system can be modeled by a DBMAP/PH/1 priority queue, and present computation algorithm and numerical results for the queueing model. Extensive simulations are carried out on the queueing behavior of the video transmission buffer, as well as on the packet level error behavior of the video data. The results demonstrate that video quality can be substantially improved by preserving the high priority video data during the transmission.

Keywords: DBMAP with marked transitions, HMM channel, PH-type distribution, DBMAP/PH/1 priority queue

1. Introduction

We have recently witnessed a phenomenal growth in the development and deployment of wireless services, evident from the proliferation of cellular data services and the emerging wireless multimedia applications. Various issues such as service priority, call admission control, retransmission scheme, flow control are now being examined for the next generation wireless cellular networks. With the scalability and flexibility of the MPEG-4 and the emergence of the broadband wireless network, wireless multimedia services are foreseen to become deployed in the near future [28]. Transporting MPEG-4 video over the broadband wireless network is expected to be an important component of many emerging multimedia applications. How to ensure the quality-of-service (QoS) of video based applications to be maintained at an acceptable level becomes a critical issue. This is further challenged by the inherent nature of wireless networks, in that this service guarantee must be achieved an under unreliable and time-varying channel condition.

There are two important issues that need to be dealt with for video transmission over the wireless network: (1) link unreliability, as wireless channels typically are much noisy and have both small-scale (multipath) and large-scale (shadowing) fading [25,27], which can cause considerable degradation of the video quality; (2) real-time video has stringent delay requirement, which means if the video packets arrive late, the video can not be played out smoothly [28]. MPEG-4 since its standardization is poised to become the enabling technology for multimedia applications [16]. With the flexibility and efficiency provided by a new form of visual data encoding called Visual Object (VO), it is expected that MPEG-4 will be capable of supporting both interactive content-based video services and conventional streaming video [29]. We are interested in system level performance modeling for MPEG-4 video transmission over the wireless networks. Specifically, in this paper we present a queueing model to study the video transmission over the uplink of an unreliable wireless channel. The objective is to examine queueing behavior at the sender side by taking into consideration the unique characteristics of the wireless channel and the QoS requirement of the video applications, namely the unreliable transmission and delay constrains. We introduce the discrete time batch Markovian arrival process (DBMAP) with two types of arrivals to model the MPEG-4 video source, which takes into account the inherent nature of the adaptiveness of the video

* Corresponding author.
E-mail: bli@cs.ust.hk

traffic; we prove that in a hidden Markov modeled (HMM) wireless channel with probabilistic transmission, the service time for an arbitrary radio link control (RLC) burst follows discrete time phase type (PH-type) distribution; we show that the link level performance of a wireless video transmission system can be modeled by a DBMAP/PH/1 priority queue, and present computation algorithm and numerical results for the queueing model. Extensive simulations are carried out on the queueing behavior of the video transmission buffer, as well as on the packet level error behavior of the video data. The results demonstrate that the video quality can be substantially improved by preserving the high priority video data during the transmission.

The rest of the paper is organized as follows. In section 2 we present the DBMAP based source model for MPEG-4 video traffic. In section 3 we introduce the HMM based probabilistic transmission model for the wireless channel, and derive the PH-type service model. In section 4 we introduce the matrix analytical solution to the DBMAP/PH/1 priority queue. Simulation and numerical results are reported in section 5. Section 6 concludes the paper with discussions on possible avenues for further research.

2. Video traffic model

It has been recognized that a compressed video stream often has a high peak-to-mean ratio. Further, video traffic is highly correlated in nature, and the correlation is present over large time scales. This distinguished feature is called the long range dependence (LRD) property [2]. The LRD feature has a negative impact on the network performance, which can lead to challenge in network traffic engineering. To model the LRD property of the video traffic, self-similar based models were developed in [5,12]. Studies in [10] showed that although the LRD property of video traffic may have a strong negative impact on the network performance, the effects are significant only if the LRD causes the busy period of the network channel to be long enough such that the long lag traffic can accumulate and come into play. Real-time VBR video often has stringent QoS requirements on delay and loss, therefore, sufficient network bandwidth must be allocated for the video traffic and the buffer size need to be limited. Thus the traffic intensity and the busy period can not be very high. In such situation, the short range dependence property of video data is more important to predict the network performance. Studies in [17] revealed that the dominant correlation factors for heterogeneous traffic can be captured by the low frequency band of the input spectral function in the frequency domain; and that correlated data traffic can be modeled by Markov modulated Poisson process. This means that for real-time VBR video, traditional Markov based models can still be valuable. In the literature there are other types of source models for video traffic, such as transform expand sample (TES) based and auto-regressive (AR) based processes. A survey of statistical video source models can be found in [13]. However, we are especially interested in Markov based video models, since in many conditions they can be accurate enough to predict the network performance [14], and the involved queueing analysis is solvable.

Due to the versatility of the Markovian arrival process (MAP) [20], it had been widely applied in modeling multimedia data traffic [4]. In recent time, MAP was used to model the aggregated Internet traffic [15]. MPEG-4 video is defined in the form of multiple video objects (VOs), which can be independently encoded, multiplexed and transmitted. This provides the potential for content-based interactivity, and also results in high efficient encoding, which makes it particularly appealing in running over low-bit rate wireless networks. Further, the contents of modern video like MPEG-4 are usually encoded in more than one layers, or in more than one types of video frames (e.g., I-frame, P-frame, or B-frame), with different significance in affecting the final decoded video quality. Hence, there is a need to model a single traffic source with multiple types of data arrivals. The MAP with marked transitions [9] can be extremely useful for this purpose, since the underlying Markov chain can be defined in a way that different type of state transitions correspond to different types of data arrivals. This naturally leads to the representation of MPEG-4 video traffic by using an MAP with multiple types of state transitions. In this paper, for simplicity, we consider video source with two traffic classes only. The model, however, is generally applicable to any hierarchically encoded video streams.

2.1. DBMAP with two types of arrivals

In this section we introduce the discrete time batch Markov arrival process (DBMAP) with marked transitions, which is a natural extension of the original DBMAP [4] and is called the marked DBMAP process hereafter. We consider an n-state DBMAP with *two* types of arrivals, from the class-1 and class-2 traffic, respectively. In real applications, the maximum batch size of a stochastic arrival process is usually limited. Let the maximum batch size for class-1 traffic arrival to be b_1, and the maximum batch size for class-2 traffic arrival to be b_2. The corresponding parameter matrices for the marked DBMAP are given by $\{D_{00}, D_{01}, \dots, D_{b_1 b_2}\}$, each $D_{i_1 i_2}$ is an $n \times n$ matrix. Suppose that at time t, $t \geqslant 0$, the underlying Markov chain of the arrival process is in state j, $1 \leqslant j \leqslant n$. Then at time epoch $t+1$, with conditional probability $D_{i_1 i_2}(j, j')$[1], where $0 \leqslant i_1 \leqslant b_1$ and $0 \leqslant i_2 \leqslant b_2$, the arrival process transits to state j', $1 \leqslant j' \leqslant n$, which is triggered by an arrival from the class-1 traffic with batch size of i_1, and an arrival from the class-2 traffic with batch size of i_2, simultaneously. Note that i_1 and i_2 might be 0, in which case the process transits without any traffic arrival.

Let the transition probability matrix of the underlying Markov chain for the arrival process to be D (also an $n \times n$

[1] We use the notation $D_{i_1 i_2}(j, j')$ to represent an element in the matrix $D_{i_1 i_2}$.

matrix), then we have $D = \sum_{i_1=0}^{b_1} \sum_{i_2=0}^{b_2} D_{i_1 i_2}$, i.e., an element $D(j, j')$ in the matrix D is given by

$$D(j, j') = \sum_{i_1=0}^{b_1} \sum_{i_2=0}^{b_2} D_{i_1 i_2}(j, j').$$

Now let us focus on the arrival process in a state j, it can have arbitrary number of arrivals from the two traffic classes, and results a transition to another state j'. This is given by the conditional probability $D_{i_1 i_2}(j, j')$, therefore we have

$$\sum_{i_1=1}^{b_1} \sum_{i_2=1}^{b_2} \sum_{j'=1}^{n} D_{i_1 i_2}(j, j') = 1.$$

Thus $D_{i_1 i_2}$ can be understood as the transition probability matrix for a simultaneous batch arrival from the two traffic classes, with batch size of i_1 and i_2, respectively. In other words, when the arrival process is in the state j, the next arrival is determined by this probability $D_{i_1 i_2}(j, j')$. More precisely, this should be the conditional probabilities for the arrivals from each of the two traffic classes, with batch size of i_1 and i_2, and with state transition from j to j'. There are total $n(1 + b_1)(1 + b_2)$ such probabilities. Notice that the above essentially implies $De = e$, in which e is an all 1 column vector[2].

We assume the arrival process is in stationary state and the initial probability vector is given by $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_n]$, which satisfies $\alpha e = 1$ and $\alpha = \alpha D$. Therefore, we can derive the average arrival rate for each traffic class, λ_1 and λ_2. The arrival rate of class-1 traffic λ_1 is given by

$$\lambda_1 = \alpha \left(\sum_{i_1=0}^{b_1} \sum_{i_2=0}^{b_2} i_1 D_{i_1 i_2} \right) e,$$

and the class-2 arrival rate λ_2 is given by

$$\lambda_2 = \alpha \left(\sum_{i_1=0}^{b_1} \sum_{i_2=0}^{b_2} i_2 D_{i_1 i_2} \right) e,$$

total traffic arrival rate is

$$\lambda = \lambda_1 + \lambda_2.$$

2.2. MPEG-4 video source model

The basic transmission unit for the next generation wireless networks is defined by a time slot. During each time slot, a specific size of data block, depending on channel coding method, can be transmitted. This unit of data block is referred as radio link control (RLC) burst[3]. High layer data such as video packets are usually first segmented into RLC bursts. These RLC bursts are then fed into the wireless transmission buffer waiting for transmission. As a result, the arrival of one video packet corresponds to a batch arrival of RLC data bursts.

We assume the video traffic can be modeled by a Markov modulated process with correlated batch arrivals. The underlying Markov chain of the arrival process can be constructed in a number of ways, for example, by scene segmentation technique [14], by frequency domain data analysis [18], or by considering the group of picture (GOP) pattern of the video sequence [3,19]. We consider video traffic with two classes of arrivals, e.g., class-1 can represent critical data like I-frame or base layer traffic, and class-2 can represent less-critical data like P-frame or enhancement layer traffic. Previous studies on video traffic have shown that the video frame size probability density function (pdf) can be fitted to log-normal distribution or Gamma distribution [8,11,24]. We assume that the class-1 video frame size pdf is given by $f_1(x)$, and the class-2 video frame size pdf is given by $f_2(y)$. Further, we assume that the transition probability matrix for the underlying Markov chain of the video arrival process is given by $D = D_0 + D_1 + D_2$, in which D_0 corresponds to state transitions of the Markov chain without data arrival, D_1 corresponds to state transitions with arrival of a class-1 video frame, and D_2 corresponds to state transitions with arrival of a class-2 video frame. Let b be the size of the RLC data burst. Then the video traffic can be modeled by a marked DBMAP process defined in section 2.1, with the following parameters:

$$\begin{aligned} D_{00} &= D_0, \\ D_{i_1 0} &= D_1 \int_{(i_1-1)b}^{i_1 b} f_1(x)\,\mathrm{d}x, \quad i_1 > 0, \\ D_{0 i_2} &= D_2 \int_{(i_2-1)b}^{i_2 b} f_2(y)\,\mathrm{d}y, \quad i_2 > 0. \end{aligned} \tag{1}$$

We next show how MPEG-4 video can be modeled as a marked DBMAP process. A MPEG-4 test sequence *Akiyo* is considered. The video test sequence is encoded with fixed quantizer equals 31, with frame rate at 30 fps. The GOP pattern repeats with 1 I-frame followed by a sequence of 14 P-frames. The picture size is 176 × 144 in pixel and the video source bit rate approximately equals to 16 kbps. The test sequence is featured with small picture size and slow motion, which is a common feature for video phone applications like wireless/mobile video. The frame size statistics of the overall video sequence follows a periodic pattern as shown in figure 1. Showing in figure 2 are the frame size distributions, which reveals that the I-frame size nearly follows normal distribution with the following pdf:

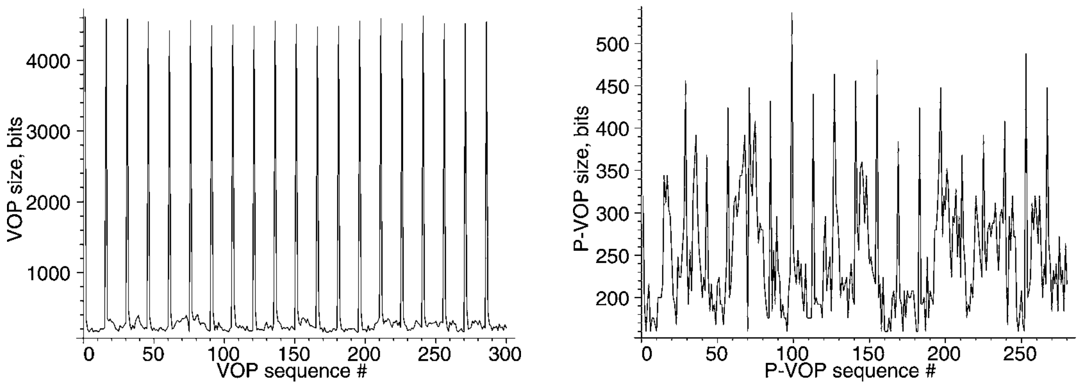
$$f_i(x) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-(x-\mu)^2/(2\sigma^2)}, \quad \sigma = 51.75,\ \mu = 4535.60,$$

with a mean value of nearly 4535 in terms of bits; and the P-frame size nearly follows a log-normal distribution with pdf

$$f_p(x) = \frac{1}{\sqrt{2\pi}\sigma x} \mathrm{e}^{-(\ln x-\mu)^2/(2\sigma^2)}, \\ \sigma = 0.2757,\ \mu = 5.4722,$$

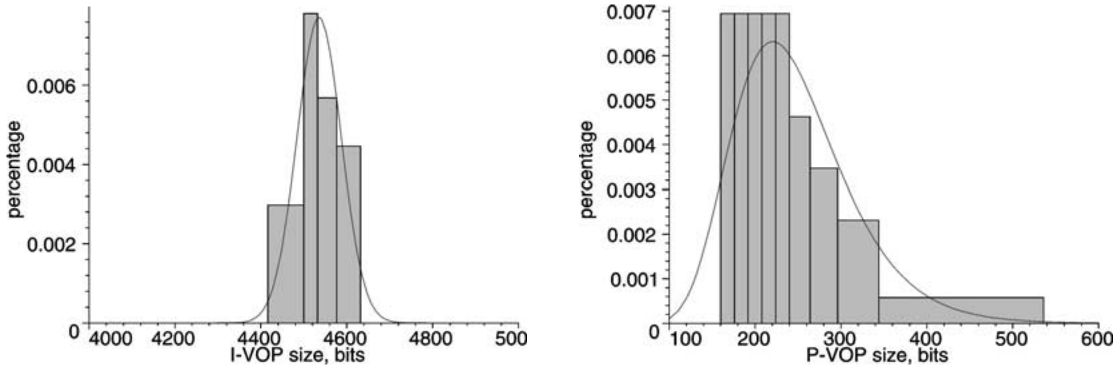
[2] We let e denote a column vector with all elements being 1's.

[3] In cdma2000 network, link layer data burst is referred as radio link protocol (RLP) burst.



(a) (b)

Figure 1. (a) All frame sequence; (b) P-frame sequence.



(a) (b)

Figure 2. (a) I-frame size distribution; (b) P-frame size distribution.

with mean of about 238 bits. Similar observation on frame size pdf had been reported in [8]. Since the maximum size of the video frames is less than 5 K bits, we assume the network layer packetizes each video frame into a single packet, and we ignore all the packet header.

Low bit rate video traffic and the periodicity property of the GOP pattern can be modeled by Markov modulated process [3]. In 3G wireless networks, the duration of a time slot is in scale of micro-seconds. If the time slot duration is 0.625 ms, a GOP can span as far as 853 time slots, given that the video frame rate is 30 fps. This requires nearly 1600 states in the Markov chains in order to precisely model the periodicity property. Such a large state space is computationally challenging. For simplicity, we consider a 2-state Markov chain to model the arrival process. Assume an I-frame is always followed by a sequence of P-frames; furthermore, the P-frame sequence is terminated by an I-frame with probability a; and once an I-frame appears, the process would repeat itself. Thus the video arrival process has 2 states, I-state (state 0) and P-state (state 1). In state 0, only P-frame can arrive; in state 1, a consecutive P-frame can arrive with probability $1 - a$, and a refreshing I-frame can arrive with probability a. Thus the transition probability matrix of the arrival process is

$$D = \begin{bmatrix} 0 & 1 \\ a & 1-a \end{bmatrix}.$$

Given the above transition probability matrix D, the video frame rate f, the time slot duration t, the I-frame and P-frame ratio r in the GOP pattern, the average channel data rate d, and the frame size distribution functions $F_i(x) = \int_0^x f_i(s)\,\mathrm{d}s$ and $F_p(x) = \int_0^x f_p(s)\,\mathrm{d}s$, $D_{i_1 i_2}$ can be derived as follows:

$$\begin{aligned} D_{i_1 0} &= ftr/(r+1)\big(F_i(i_1\,\mathrm{d}t) - F_i\big((i_1-1)\,\mathrm{d}t\big)\big)D, \\ D_{0 i_2} &= ft/(r+1)\big(F_p(i_2\,\mathrm{d}t) - F_p\big((i_2-1)\,\mathrm{d}t\big)\big)D. \end{aligned} \tag{2}$$

For example, if we assume the frame rate is 30 fps, time slot duration is 0.625 ms, $a = 0.1$, I-frame and P-frame ratio is $1 : 14$, the bit rate of the wireless channel is 28.8 kbps, then the parameters of video source are given by the positive matrices $D_{240,0}, D_{241,0}, \ldots, D_{256,0}$ for I-frame arrivals, and $D_{0,8}, D_{0,9}, \ldots, D_{0,32}$ for P-frame arrivals. As a result, all the matrices D_{i_1,i_2} could be readily determined, e.g.,

$$D_{246,0} = \begin{bmatrix} 0 & 0.0000280678 \\ 0.00000280678 & 0.000025261 \end{bmatrix},$$

$$D_{0,30} = \begin{bmatrix} 0 & 0.0000104835 \\ 0.00000104835 & 0.00000943518 \end{bmatrix}.$$

For the range other than $240 < i_1 < 256$ and $8 < i_2 < 32$, all the $D_{i_1 i_2}$'s are assumed to be zero matrices, this is because there is a very small probability for a video frame has packet size within that range.

We use the above example (2) to show that the marked DBMAP process can be used to model MPEG-4 video. Generally, in order to precisely model the video traffic, more states can be introduced to the underlying Markov chain of the arrival process, depending on how the Markov chain is constructed; and more accurate estimation of the frame size pdf can be applied by more elaborate statistical data analysis. However, as demonstrated by (1), the resulting model still belongs to the marked DBMAP process.

3. Link level service time distribution

In this section, we summarize the wireless channel characteristics and introduce the hidden Markov wireless channel model. To overcome the negative effect of link unreliability on the achieved video quality, we consider automatic repeat request (ARQ) protocol with overdue control for video data transmission, and prove that the link level service process follows discrete time PH-type distribution.

Wireless channel characterization, because of its importance on the system design, has been well studied in literature [23]. Its impact on the transmitted signal can be roughly divided into three categories: interference among neighboring cells, multipath fading, and path loss. Moreover, mobile terminals are usually in frequent movement. As a result, wireless channel exhibits correlated errors, that is, the impairment experienced by consecutive transmission data bursts are correlated. It has been shown that the error correlations have an important performance impact [6,32]. Taking the above factors into consideration, we propose to use the notion of a probabilistic transmission in that each transmission can result in a failure with a certain probability, and adopt the use of the hidden Markov modeled (HMM) channel. Consider a binary wireless channel that is time slotted and one date burst can be transmitted in each time slot. The HMM channel is defined by two transition probability matrices P_E, $E = 0, 1$. Suppose at time t, the channel is in state i, $1 \leqslant i \leqslant m$. The elements of P_E are defined such that at time $t + 1$, the channel transits from state i to state j, with a correct transmission probability of $P_0(i, j)$, and an error transmission probability of $P_1(i, j)$. P_0 and P_1 are called transition probability matrix for correct transmission and error transmission, respectively. The wireless channel evolves following the underlying Markov chain with transition probability matrix $P_0 + P_1$. Assume the channel process is stationary and the steady state probability vector is given by θ, then we have $\theta = \theta(P_0 + P_1)$ and $\theta e = 1$. The HMM channel is rather general and analytically tractable, which can be applied for error process at various levels. This had been extensively studied in [26].

There are a number of error control methods in wireless networks, such as forward error correction (FEC) and automatic repeat request (ARQ). In practical systems, FEC (using either convolutional or turbo coding) is usually implemented at the physical layer. Here, we only consider the ARQ at the link layer[4]. Given the one-hop nature of the wireless communication, we ignore the feedback delay for the acknowledgment such that retransmission is invoked immediately in the next time slot when error occurs. Since real-time video is considered, the delay of each video packet and the corresponding RLC bursts needs to be kept below certain threshold. Therefore, we assume an overdue time is set appropriately such that once the overdue time is exceeded, further (re)transmission attempt of the serving RLC bursts of the video packet will stop, the overdue RLC burst will be discarded, and the corresponding video packet is called erroneous.

We now prove that, by the above ARQ overdue control protocol, the service time for an arbitrary RLC burst in an m state HMM wireless channel follows discrete time PH-type distribution [21]. Assume the maximum tolerable transmission time for a RLC burst is d, $d > 1$, the service time or the channel occupation time till correct receipt or overdue can be derived as following. Define the state space for the service process as $\{(l, j), 1 < l < d, 1 < j < m\}$, in which a state (l, j) corresponds to the situation that the serving RLC burst is in lth (re)transmission and the channel is in state j. The channel itself evolves according to transition matrix $P = P_0 + P_1$. When the channel evolves according to the substochastic matrix P_0, the transmission is successful and the service process will terminate, this happens with transition probability vector of $P_0 e$; when the channel evolves according to the substochastic matrix P_1, the transmission is unsuccessful, thus l changes to $l + 1$ and j will change accordingly. Therefore, the transition probability matrix for the service process is given by

$$S = \begin{bmatrix} \mathbf{0} & P_1 & \mathbf{0} & \ldots & \mathbf{0} & P_0 e \\ \mathbf{0} & \mathbf{0} & P_1 & \ldots & \mathbf{0} & P_0 e \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \ddots & P_1 & P_0 e \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} & e \\ 0 & 0 & 0 & \ldots & 0 & 1 \end{bmatrix},$$

where P_0, P_1 are $m \times m$ size sub-stochastic matrices, $\mathbf{0}$ is zero matrix. S is $(dm + 1) \times (dm + 1)$ size transition matrix, and d is the channel occupation time. Let S' be the left-upper sub-matrix of S, with the last row and last column removed, resulting in a $dm \times dm$ size sub-stochastic matrix. It is obvious the service time follows a PH-type distribution with representation (β, S'), where $\beta = [1, 0, 0, \ldots]$. As a result, the service time for an arbitrary RLC burst in the HMM wireless channel with ARQ overdue control follows the discrete time PH-type distribution.

4. Queueing analysis for system level performance

Modern video encoding like MPEG-4 provides inherent adaptation flexibility for wireless video streaming [28]. With

[4] In real systems to be precise, ARQ protocol is implemented at the link access control (LAC) layer and radio link protocol (RLP) layer.

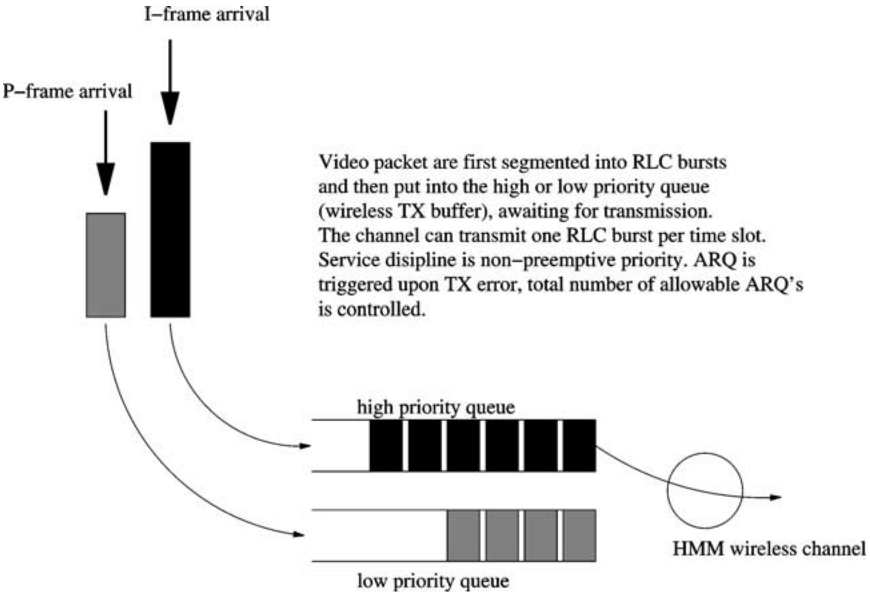

I–frame arrival
P–frame arrival
Video packet are first segmented into RLC bursts
and then put into the high or low priority queue
(wireless TX buffer), awaiting for transmission.
The channel can transmit one RLC burst per time slot.
Service disipline is non–preemptive priority. ARQ is
triggered upon TX error, total number of allowable ARQ's
is controlled.
high priority queue
low priority queue
HMM wireless channel

Figure 3. A diagram of DBMAP/PH/1 with priority queueing system.

adaptive encoding, the video contents are organized with different significance for video reconstruction. Thus when the channel condition temporarily becomes bad, or when the effective channel bandwidth temporarily becomes low, less important video data like P-frame bursts can be truncated in order to preserve sufficient bandwidth to transmit more important video data like I-frame bursts. Such adaptation can result better video quality than just randomly discarding all the video data since I-frame data play more important role in video decoding. For this purpose we consider priority based scheduling protocol for video data transmission, that is, video packets are first segmented into RLC bursts and then put into the link layer transmission buffer, the I-frame bursts are scheduled for transmission with higher priority than the P-frame bursts. We are interested in the system level performance modeling for video transmission over an HMM wireless channel. Since the video source can be modeled as the DBMAP with marked transitions, and the link layer video data transmission time follows PH-type distribution, the performance of the wireless video transmission system thus can be modeled by the DBMAP/PH/1 priority queue, as shown in figure 3.

In this section, we show that the DBMAP/PH/1 priority queue can be solved by using matrix analytical methods. We only consider the non-preemptive priority case. For the preemptive priority case, refer to [31] for more information; for the single arrival DMAP/PH/1 priority queue, refer to [1]. Consider a queueing system as follows:

1. The arrival process follows the marked DBMAP with 2 priority levels, defined by $\{D_{00}, D_{01}, D_{02}, \ldots, D_{b_1 b_2}\}$ as in section 2.1. All $D_{i_1 i_2}$ matrices are of dimension $n \times n$.
2. The service processes for both the high and low priority jobs follow discrete time PH-type distributions, with representation of (β_1, S_1) for the high priority jobs, and (β_2, S_2) for the low priority jobs, respectively. β_1 is row vector with m_1 elements, and S_1 is $m_1 \times m_1$ sub-stochastic matrix; similarly, β_2 is row vector with m_2 elements, and S_2 is $m_2 \times m_2$ sub-stochastic matrix. Here we assume $m_1 = m_2 = m$.
3. There is a single server in the system and the service discipline is non-preemptive priority in nature.

We define the system state space $\Delta = \{(n_1, n_2, i, s, p)\}$, in which a state (n_1, n_2, i, s, p) in Δ represents:

- The arrival process is in state s, $s = 1, 2, \ldots, n$;
- i is an index variable. $i = 1$ represents that a high priority job is being served; and $i = 2$ represents that a low priority job is being served;
- The service process is in phase p, depending on which type of job is in service i. When $i = 1$, $p = 1, 2, \ldots, m_1$; when $i = 2$, $p = 1, 2, \ldots, m_2$;
- There are n_1 high priority jobs in the system, including the one being served, if any;
- In the case of $n_1 = 0$ and $n_2 = 0$, state (n_1, n_2, i, s, p) can be reduced to $(0, 0, s)$, since the whole system is idle and there is no need to record the service phase. In this case the number of high and low priority jobs in the system are all zero's;

  In the case of $n_1 = 0$ and $n_2 \geqslant 1$, state (n_1, n_2, i, s, p) can be reduced to $(0, n_2, s, p)$, since only a low priority job can be in service. In this case n_2 denotes the number of low priority jobs in the system;

  In the case of $n_1 \geqslant 1$, the server may be busy with a high or low priority job, due to the non-preemptive service discipline. In this case, n_2 denotes the number of low priority jobs waiting in the queue. Thus if the server is busy serving a low priority job, the total number of low priority jobs in the system is $n_2 + 1$, otherwise the total number of low priority jobs in the system is n_2.

The DBMAP/PH/1 priority queueing system can be described by an M/G/1 type Markov chain [22] with the state space Δ. The transition probability matrix P of the related Markov chain is given by

$$P = \begin{bmatrix} B_{00} & B_{01} & B_{02} & \ldots & B_{0b_1} & & & \\ B_{10} & A_0 & A_1 & \ldots & A_{b_1-1} & A_{b_1} & & \\ & A_{-1} & A_0 & \ldots & A_{b_1-2} & A_{b_1-1} & A_{b_1} & \\ & & A_{-1} & \ldots & A_{b_1-3} & A_{b_1-2} & A_{b_1-1} & A_{b_1} \\ & & & \ddots & \ddots & \ddots & \ddots & \ddots \; \ddots \end{bmatrix}, \quad (3)$$

with the following elements[5]:

$$B_{00} = \begin{bmatrix} B_{00}^{00} & B_{00}^{01} & B_{00}^{02} & \cdots & B_{00}^{0b_2} & & & \\ B_{00}^{10} & B_{00}^{0} & B_{00}^{1} & \cdots & B_{00}^{b_2-1} & B_{00}^{b_2} & & \\ & B_{00}^{-1} & B_{00}^{0} & \cdots & B_{00}^{b_2-2} & B_{00}^{b_2-1} & B_{00}^{b_2} & \\ & & B_{00}^{-1} & \cdots & B_{00}^{b_2-3} & B_{00}^{b_2-2} & B_{00}^{b_2-1} & B_{00}^{b_2} \\ & & & \ddots & \ddots & \ddots & \ddots & \ddots \; \ddots \end{bmatrix},$$

where

$$\begin{aligned} B_{00}^{00} &= D_{00}, \\ B_{00}^{0i_2} &= D_{0i_2} \otimes \beta_2, \quad i_2 = 1, 2, \ldots, b_2 \\ &\quad (\otimes \text{ denotes the Kronecker product}), \\ B_{00}^{10} &= D_{00} \otimes S_2^0, \\ B_{00}^{i_2} &= D_{0i_2} \otimes S_2 + D_{0(i_2+1)} \otimes S_2^0 \beta_2, \\ &\quad i_2 = 0, 1, \ldots, b_2 - 1, \\ B_{00}^{b_2} &= D_{0b_2} \otimes S_2, \\ B_{00}^{-1} &= D_{00} \otimes S_2^0 \beta_2 \end{aligned}$$

and

$$B_{0i_1} = \begin{bmatrix} B_{0i_1}^{00} & B_{0i_1}^{01} & B_{0i_1}^{02} & \cdots & B_{0i_1}^{0b_2} & \\ B_{0i_1}^{-1} & B_{0i_1}^{0} & B_{0i_1}^{1} & \cdots & B_{0i_1}^{b_2-1} & \\ & B_{0i_1}^{-1} & B_{0i_1}^{0} & \cdots & B_{0i_1}^{b_2-2} & B_{01}^{b_2-1} \\ & & \ddots & \ddots & \ddots & \ddots \; \ddots \end{bmatrix},$$

$$i_1 = 1, 2, \ldots, b_1,$$

where

$$\begin{aligned} B_{0i_1}^{0i_2} &= [\, D_{i_1 i_2} \otimes \beta_1 \quad 0], \quad i_2 = 0, 1, \ldots, b_2, \\ B_{0i_1}^{-1} &= [\, D_{i_1 0} \otimes S_2^0 \beta_1 \quad D_{i_1 0} \otimes S_2 \,], \\ B_{0i_1}^{i_2} &= [\, D_{i_1(i_2+1)} \otimes S_2^0 \beta_1 \quad D_{i_1(i_2+1)} \otimes S_2 \,], \\ &\quad i_2 = 0, 1, \ldots, b_2 - 1 \end{aligned}$$

and

$$B_{10} = \begin{bmatrix} B_{10}^{00} & B_{10}^{1} & B_{10}^{2} & \cdots & B_{10}^{b_2} & \\ & B_{10}^{0} & B_{10}^{1} & \cdots & B_{10}^{b_2-1} & B_{10}^{b_2} \\ & & \ddots & \ddots & \ddots & \ddots \; \ddots \end{bmatrix},$$

where

$$B_{10}^{00} = \begin{bmatrix} D_{00} \otimes S_1^0 \\ 0 \end{bmatrix}$$

$$B_{10}^{i_2} = \begin{bmatrix} D_{0i_2} \otimes S_1^0 \beta_2 \\ 0 \end{bmatrix}, \quad i_2 = 0, 1, \ldots, b_2$$

and

$$A_{i_1} = \begin{bmatrix} A_{i_1}^0 & A_{i_1}^1 & A_{i_1}^2 & \cdots & A_{i_1}^{b_2} & \\ & A_{i_1}^0 & A_{i_1}^1 & \cdots & A_{i_1}^{b_2-1} & A_{i_1}^{b_2} \\ & & \ddots & \ddots & \ddots & \ddots \; \ddots \end{bmatrix},$$

$$i_1 = 0, 1, \ldots, b_1,$$

where

$$A_{i_1}^{i_2} = \begin{bmatrix} D_{(i_1+1)i_2} \otimes S_1^0 \beta_1 + D_{i_1 i_2} \otimes S_1 & 0 \\ D_{i_1 i_2} \otimes S_2^0 \beta_1 & D_{i_1 i_2} \otimes S_2 \end{bmatrix},$$

$$\text{for } i_1 = 0, 1, \ldots, b_1 - 1, \; i_2 = 0, 1, \ldots, b_2,$$

$$A_{b_1}^{i_2} = \begin{bmatrix} D_{b_1 i_2} \otimes S_1 & 0 \\ D_{b_1 i_2} \otimes S_2^0 \beta_1 & D_{b_1 i_2} \otimes S_2 \end{bmatrix},$$

$$i_2 = 0, 1, \ldots, b_2$$

and

$$A_{-1} = \begin{bmatrix} A_{-1}^0 & A_{-1}^1 & A_{-1}^2 & \cdots & A_{-1}^{b_2} & \\ & A_{-1}^0 & A_{-1}^1 & \cdots & A_{-1}^{b_2-1} & A_{-1}^{b_2} \\ & & \ddots & \ddots & \ddots & \ddots \; \ddots \end{bmatrix}$$

where

$$A_{-1}^{i_2} = \begin{bmatrix} D_{0i_2} \otimes S_1^0 \beta_1 & 0 \\ 0 & 0 \end{bmatrix}, \quad i_2 = 0, 1, \ldots, b_2.$$

A Markov chain is called GI/G/1 type if the transition probability matrix has the row-repeating property [7]; and is further called M/G/1 type if the transition probability matrix has upper-Hessenberg form. We note that Markov chain (3) is M/G/1 type. Standard M/G/1 type Markov chain can have infinite number of levels, but usually the phase number is finite. If the phase dimension is finite, the Markov chain can be solved by adopting matrix analytic methods [7,30]. The main challenge to solve the Markov chain (3) is the fact that the transition probability matrix has infinite number of levels and infinite number of phases, i.e., all the elements B_{0i_1} and A_{i_1} are of infinite size.

We next examine the R and G measures [30] R_k (for $k \geqslant 1$), $R_{0,j}$ (for $j > 0$), G_k (for $k \geqslant 1$), $G_{i,0}$ (for $i > 0$), Φ and Φ_0, of the M/G/1 type Markov chain (3). Obviously, for the G measures, only G_1 and $G_{1,0}$ have definition since the transition matrix is upper-Hessenberg form; for the R measures, all the R_k and $R_{0,k}$, for $k > b_1$, have no definition, since the transition matrix is banded with finite number of bands b_1. By the

[5] In this paper, we let the superscript $^{-1}$ denote an index, while let the bold superscript $^{\mathbf{-1}}$ denote the inverse of a matrix. We let $\otimes$ denote the Kronecker product.

results from [30] and the above facts, we have the following recursive formulas for R_k, $R_{0,k}$, and Φ_0:

$$\begin{aligned}R_k &= (A_k + R_{k+1}A_{-1})(I - A_0 - R_1A_{-1})^{-\mathbf{1}},\\ &\quad k = 1, 2, \ldots, b_1 - 1,\\ R_{b_1} &= A_{b_1}(I - A_0 - R_1A_{-1})^{-\mathbf{1}},\\ R_{0,k} &= (B_{0k} + R_{0,k+1}A_{-1})(I - A_0 - R_1A_{-1})^{-\mathbf{1}},\\ &\quad k = 1, 2, \ldots, b_1 - 1,\\ R_{0,b_1} &= B_{0b_1}(I - A_0 - R_1A_{-1})^{-\mathbf{1}},\\ \Phi_0 &= B_{00} + R_{0,1}B_{10}.\end{aligned} \tag{4}$$

Further, as shown in [31], R_k, $R_{0,k}$ and Φ_0 have row-repeating and triangular structures, which makes it possible to compute all the items in (4). Therefore, we can compute π_n, where $\pi_n = [\pi_{n0}, \pi_{n1}, \pi_{n2}, \ldots]$, $n = 0, 1, \ldots$, the stationary probability vector for the M/G/1 type Markov chain (3), in the following four steps:

1. In the first step, we compute R_k, $R_{0,k}$ and Φ_0 by the following iterative algorithm:
   (a) Set $R_k = A_k$, $R_{0,k} = B_{0k}$.
   (b) Update R_k, $R_{0,k}$ recursively according to (4), till to a sufficient large index i^* such that $R_k^{i^*}$, $R_{0,k}^{0i^*}$ and $R_{0,k}^{i^*}$ are small enough for truncation. Here $R_k^{i^*}$, $R_{0,k}^{0i^*}$ and $R_{0,k}^{i^*}$ are the elements of R_k and $R_{0,k}$.
   (c) Compute Φ_0 according to (4).
2. In the second step, we compute π_0 by making use of $\pi_0 = \pi_0\Phi_0$. Note that Φ_0 is an upper-Hessenberg form transition matrix with row-repeating property, which is also M/G/1 type. By standard matrix analytic method [7], $\pi_0 = \pi_0\Phi_0$ can be solved.
3. In the third step, we compute π_n in a similar way in [7]:
$$\pi_n = \begin{cases} \pi_0 R_{0,n} + \sum_{k=1}^{n-1} \pi_k R_{n-k}, & 1 \leqslant n \leqslant b_1,\\ \sum_{k=1}^{b_1} \pi_{n-k}R_k, & n > b_1.\end{cases}$$
4. Finally, we normalize π_{ij} and ensure
$$\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\pi_{ij}e = 1.$$

We define the following notations:

$P\{idle\}$: probability that the server is idle,
$P\{busy\}$: probability that the server is busy,
$P\{busy_h\}$: probability that the server is busy with a high priority job,
$P\{busy_l\}$: probability that the server is busy with a low priority job,
$P\{N_h = i\}$: probability that there are i high priority jobs in the system,
$P\{N_l = j\}$: probability that there are j low priority jobs in the system,
$P\{Q_h = i\}$: probability that there are i high priority jobs waiting in the queue,
$P\{Q_l = j\}$: probability that there are j low priority jobs waiting in the queue.

Based on π_{ij}, the following performance measures for the DBMAP/PH/1 priority queueing model can be derived:

$$\begin{aligned}P\{idle\} &= \pi_{00}e,\\ P\{busy\} &= 1 - \pi_{00}e,\\ P\{busy_h\} &= 1 - \sum_{j=0}^{\infty}\pi_{0j}e,\\ P\{busy_l\} &= \sum_{j=1}^{\infty}\pi_{0j}e,\\ P\{N_h = i\} &= \sum_{j=0}^{\infty}\pi_{ij}e, \quad i \geqslant 0,\\ P\{N_l = 0\} &= \pi_{00}e + \sum_{i=1}^{\infty}\pi_{i0}e_h,\\ P\{N_l = j\} &= \pi_{0j}e + \sum_{i=1}^{\infty}\pi_{ij}e_h + \sum_{i=1}^{\infty}\pi_{i(j-1)}e_l, \quad j \geqslant 1,\end{aligned}$$

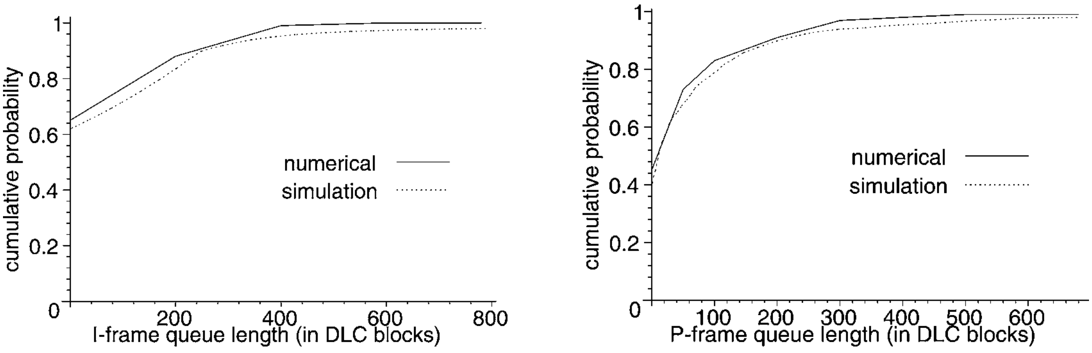
where e_h and e_l are column vectors of dimension $2nm$, $e_h = \langle e, e_0\rangle$, $e_l = \langle e_0, e\rangle$, and e_0 is an nm dimensional column vector with all members being 0's;

$$\begin{aligned}P\{Q_h = 0\} &= P\{N_h = 1\} + P\{N_h = 0\},\\ P\{Q_h = i\} &= P\{N_h = i + 1\}, \quad i \geqslant 1,\\ P\{Q_l = 0\} &= \pi_{00}e + \pi_{01}e + \sum_{i=1}^{\infty}\pi_{i0}e,\\ P\{Q_l = j\} &= \pi_{0(j+1)}e + \sum_{i=1}^{\infty}\pi_{ij}e, \quad j \geqslant 1.\end{aligned}$$

By $P\{Q_h = i\}$ and $P\{Q_l = j\}$, we essentially obtain the queue length distribution for both the high and the low priority queue.

5. Numerical and simulation results

In the first half of this section we present numerical results for the computation algorithm developed in section 4; in the second half of this section we present simulation results for the queueing behavior of the video transmission buffer. The parameters of the arrival process are derived from section 2.2 for the video sequence *Akiyo*. The settings for the PH-type service process are given by (β_1, S_1) and (β_2, S_2) for the high



(a) (b)

Figure 4. Comparison of the queue length distribution: numerical vs. simulation.

Table 1
Buffer overflow probability vs. the given buffer size. P_{oh}: overflow probability for the high priority buffer (I-frame queue). P_{ol}: overflow probability for the low priority buffer (P-frame queue).

	P_{oh}		P_{ol}	
Buffer size	Numerical	Simulation	Numerical	Simulation
50	–	0.42	–	0.32
100	–	0.29	0.18	0.22
200	0.12	0.14	0.09	0.10
300	–	0.08	0.03	0.04
400	0.03	0.035	0.02	0.03
500	–	0.03	0.016	0.02
600	0.015	0.018	0.01	0.011
700	–	0.012	0.01	0.01
780	0.01	0.011	–	–

and low priority queue, respectively, as follows:

$$S_1 = \begin{bmatrix} 0.0 & 0.05 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.05 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.05 \\ 0.0 & 0.0 & 0.0 & 0.0 \end{bmatrix},$$

$$S_2 = \begin{bmatrix} 0.0 & 0.05 & 0.0 \\ 0.0 & 0.0 & 0.05 \\ 0.0 & 0.0 & 0.0 \end{bmatrix},$$

$\beta_1 = [1.0, 0.0, 0.0, 0.0]$, $\beta_2 = [1.0, 0.0, 0.0]$. The above service setting corresponds to the case that the average RLC burst transmission error rate is 0.05, and the maximum allowable ARQ's for I-frame bursts and P-frame bursts are 4 and 3, respectively.

5.1. Numerical results for the computation algorithm

We apply the computation algorithm for the DBMAP/PH/1 priority queue and compare the numerical results with the simulation results. Simulation of the marked DBMAP arrival process is done in the following way. Let $D_{00}, D_{01}, \ldots, D_{b_1 b_2}$ be the parameter matrices for the arrival process. Let α be the initial vector, with $\alpha = \alpha D$ and $D = \sum_{i_1=0}^{b_1} \sum_{i_2=0}^{b_2} D_{i_1 i_2}$. Let $J(t)$ be the state of the underlying Markov chain immediately after time t, let $Y_1(t)$ and $Y_2(t)$ be the numbers of arrivals for the high and low priority queue, respectively, at the tth transition. We simulate the sequence of random variables $\{J(t), Y_1(t), Y_2(t)\}$ for t up to some value of *max_time_slots* with the following numerical method.

- Arrival process initialization.
  Generate the initial state $J(0)$ according to a multinomial variate with density α, set $t = 0$, $Y_1(t) = 0$, $Y_2(t) = 0$.
- Arrival process state machine evolution.
  while ($t <$ *max_time_slots*)
  {

  Choose the state $J(t+1)$ as a multinomial variate with density

  $$(D_{J(t),1}, D_{J(t),2}, \ldots, D_{J(t),n}).$$

  Set

  $$Y_1(t+1) = \left\lfloor \frac{i}{b_2+1} \right\rfloor, \quad Y_2(t+1) = i \bmod (b_2+1),$$

  where the index i is generated as a multinomial variate with the following density:

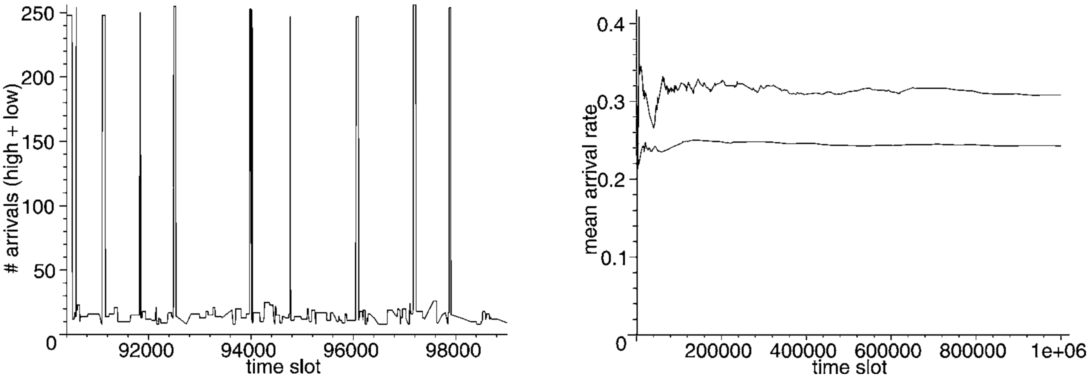
  $$\left(\frac{[D_{00}]_{J(t),J(t+1)}}{D_{J(t),J(t+1)}}, \frac{[D_{01}]_{J(t),J(t+1)}}{D_{J(t),J(t+1)}}, \ldots, \frac{[D_{b_1 b_2}]_{J(t),J(t+1)}}{D_{J(t),J(t+1)}} \right).$$

  Enqueue a batch of $Y_1(t+1)$ high priority bursts and a batch of $Y_2(t+1)$ low priority bursts into the system. Note that $Y_1(t+1)$ and $Y_2(t+1)$ might be 0.

  Set $t = t + 1$.

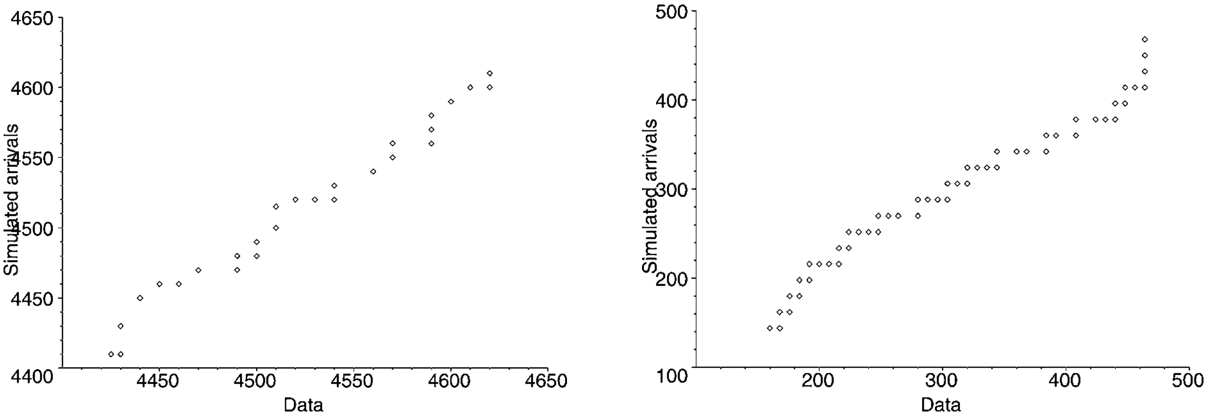
  }

As a result, the state machine for the arrival process is completely driven by the vector α, the stochastic matrix D, and the substochastic matrices $D_{00}, D_{01}, \ldots, D_{b_1 b_2}$. We calculate the cumulative queue length distribution function (cdf) for both the high and low priority queues using the algorithm developed in section 4. We also collect the queue length statistics from the simulation results. Shown in figure 4 is the comparison of the queue length cdf obtained from the numerical results and the simulation results. It can be observed that both the high and low priority queue, the queue length cdf curves are very close. From the queue length cdf curves, we can further obtain the buffer overflow probabilities, P_{oh} and P_{ol} for the high and low priority queue, respectively, if a maximum value of the buffer size is given. The results are shown in table 1.



(a) (b)

Figure 5. (a) Arrival patterns; (b) mean arrival rates.



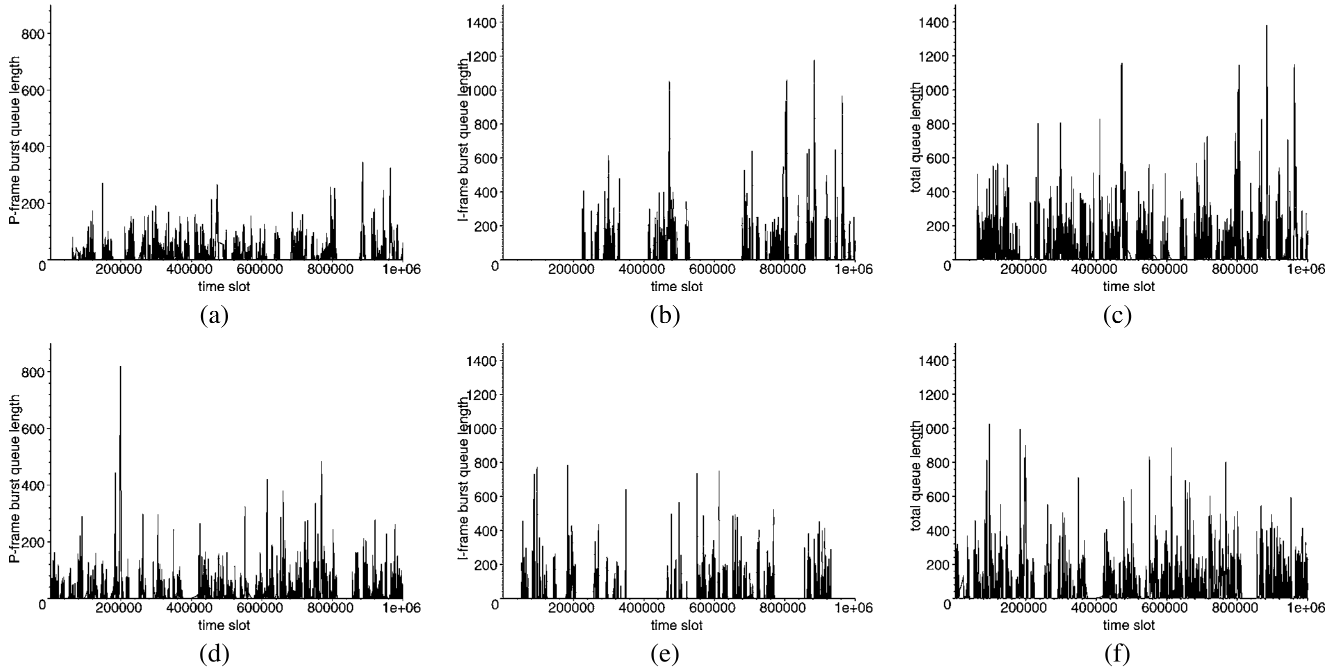
(a) (b)

Figure 6. QQ plots for trace data and simulated arrival data. (a) For I-frame arrivals; (b) for P-frame arrivals.

5.2. Simulation results for the video transmission

We simulate the wireless video transmission system, concentrating on the queueing behavior of the video transmission buffer. Figure 5(a) are the plots of the sample arrivals generated by the traffic model. The arrival bursts in the graph are clustered, which reveals that the data arrivals are bursty and highly correlated. Further, the simulated arrival process as shown in figure 5(a) has a pseudo-periodic feature. This demonstrates that the proposed arrival model can effectively grasp the correlation and periodic nature of the video traffic. In figure 5(b), the above curve is the mean arrival rate for the high priority queue, and the below curve is the mean arrival rate for the low priority queue. We observe that the mean arrival rate for the high and low priority queue are around 0.31 and 0.24, respectively, which are very close to the analytical derived arrival rate of 0.31438 and 0.2426. QQ plots for video trace data and simulated video arrival data are shown in figure 6(a) for I-frame data arrivals, and in figure 6(b) for P-frame data arrivals.

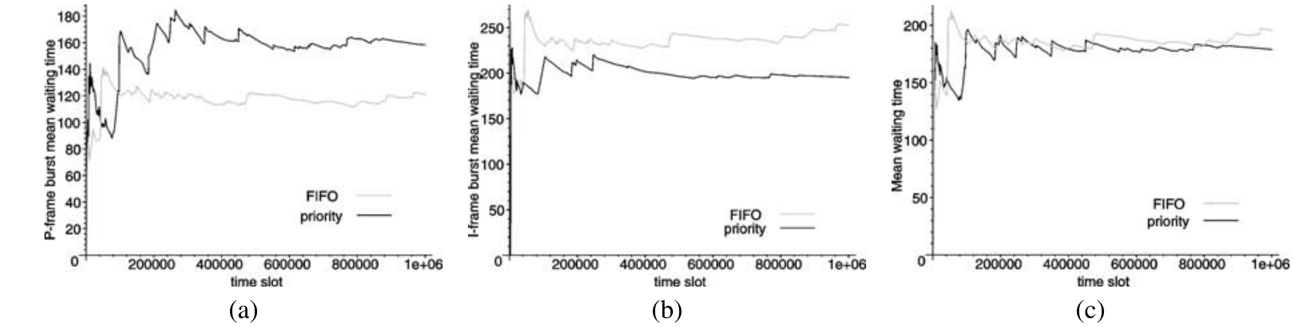
We also simulate the non-priority scheduling case, in which all the video data bursts are transmitted by first in first out (FIFO) order, and compare the queue length dynamics with the priority scheduling case. Figures 7(a)–(c) show the queue length fluctuation against time for the FIFO case. We find that both the I-frame bursts and P-frame bursts experience a high degree of queue length fluctuation. This is because the video traffic source has bursty and correlation nature, and the wireless channel is inherently instable. Further, we also find that I-frame bursts experience a much higher degree of dynamics. This is because the I-frame packet has a much larger size than the P-frame packet, with a ratio near 10 : 1. Thus an I-frame packet arrival will cause a significant increase of the queue length, which leads to greater fluctuation. Figures 7(d)–(f) show the queue length fluctuation against time for the priority scheduling case. The queue fluctuation for the low priority queue shown in figure 7(d) tends to be larger than that of the FIFO queueing case, as shown in figure 7(a). This demonstrates that, by priority scheduling, the low priority P-frame data bursts experienced a certain degree of service degradation. However, when examine the queue length dynamics for the I-frame bursts in figure 7(e), we find that it has much less fluctuation than the FIFO case, as in figure 7(b). Moreover, the total queue length dynamics as in figure 7(f) also has much less fluctuation than the FIFO case as in figure 7(c). This demonstrates that by priority scheduling, I-frame data get better service quality, and the overall service quality is also improved. Observed from figures 7(b) and (e), the maximum queue length for the I-frame bursts reaches as high as 1200 for the FIFO case, but it is



(a) (b) (c)

(d) (e) (f)

Figure 7. Comparison of the queue length dynamics. (a) P-frame bursts, FIFO case; (b) I-frame bursts, FIFO case; (c) total queue, FIFO case; (d) P-frame bursts, priority case; (e) I-frame bursts, priority case; (f) total queue, priority case.



(a) (b) (c)

Figure 8. Comparison of the mean waiting time. (a) For P-frame bursts; (b) for I-frame bursts; (c) for all RLC bursts.

only as high as 800 for the priority case. The maximum total queue length also decreases from near 1400 for the FIFO case to less than 1100 for the priority case, as shown in figures 7(c) and (f). The result reveals that the required video transmission buffer is reduced when priority scheduling is adopted.

In real-time video application, a play-back buffer is maintained in the client to overcome the non-uniform arrival rate of the video data, thus a smooth and constant frame rate video can be played-back in the client. The size of the play-back buffer is partly determined by the hold time, which in turn is determined by the waiting time and service time of the transmission queue in the sender. Real-time video application is delay sensitive, as a result, a low hold time should be maintained, which in turn require a low waiting and low service time. Shown in figures 8(a)–(c) are the comparison of the mean waiting time for the FIFO queueing and priority queueing cases. The mean waiting time for the low priority bursts is increased from 120 for the FIFO case to 160 for the priority case, as in figure 8(a). This again shows that the low priority P-frame bursts experienced a degree of service degradation. When observe from figure 8(b), we find that the mean waiting time for the high priority I-frame bursts decreased from 240 for FIFO queueing to 200 for priority queueing. Figure 8(c) also shows that the overall mean waiting time is decreased from 190 to 180 by taking priority queueing. Since non-preemptive priority scheduling is adopted, prioritizing the arrival bursts will not introduce extra delay in serving the bursts after classification, thus the mean service time should be the same as that for the FIFO queueing case. As a result, a lower hold time can be expected by adopting priority queueing. This demonstrates that priority based scheduling has better real-time support than the FIFO scheduling.

We now examine the error behavior for different combinations of the scheduling protocol and allowable ARQ number. We first examine the packet error rate when both the high and low priority RLC bursts have equal allowable ARQ number of 3, as shown in figure 9. From figure 9(b), it is evident that the packet level error rate for P-frames increases

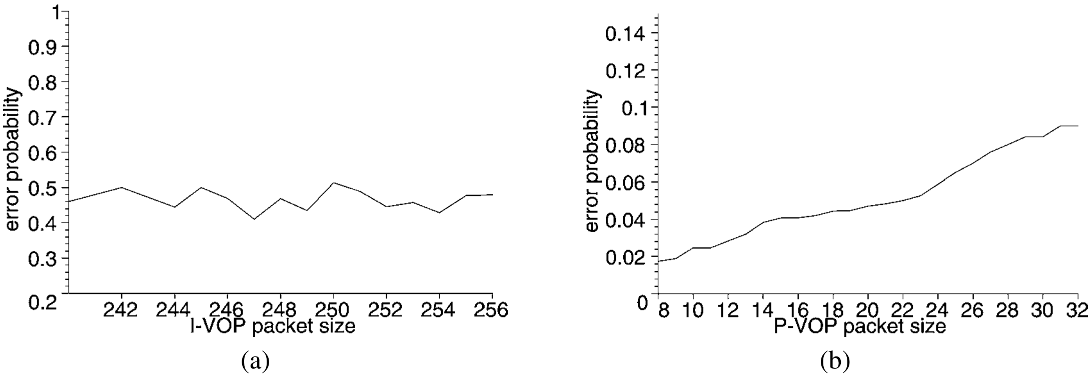


Figure 9. Packet error rate vs. packet size (both I-frame and P-frame ARQ = 3). (a) I-frame error rate; (b) P-frame error rate.

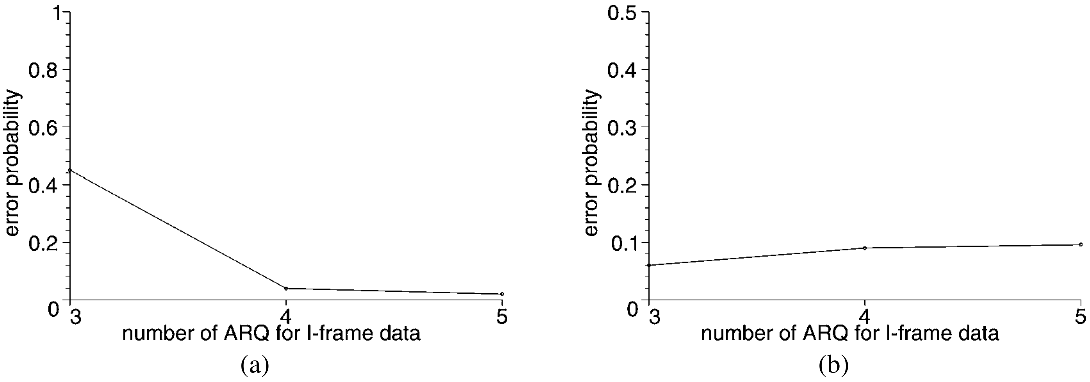


Figure 10. Average packet error rate (I-frame ARQ = 3, 4, 5; P-frame ARQ = 3 only). (a) I-frame error rate; (b) P-frame error rate.

Table 2
Frame error rate and PSNR measure for the erroneous frames.

	I-frame ARQ = 3 P-frame ARQ = 3	I-frame ARQ = 4 P-frame ARQ = 3
FIFO scheduling	I-frame error rate: 48.06% P-frame error rate: 5.30% erroneous I-frame PSNR: 21.24 erroneous P-frame PSNR: 19.46	I-frame error rate: 6.19% P-frame error rate: 7.64% erroneous I-frame PSNR: 23.58 erroneous P-frame PSNR: 18.30
Priority scheduling	I-frame error rate: 45.70% P-frame error rate: 6.09% erroneous I-frame PSNR: 22.18 erroneous P-frame PSNR: 18.56	I-frame error rate: 3.70% P-frame error rate: 8.54% erroneous I-frame PSNR: 24.40 erroneous P-frame PSNR: 17.88

with the packet size, i.e., the larger the packet size, the higher packet error rate it has. However, the packet level error rate for I-frames is more stable, as shown in figure 9(a). The reason is that I-frames have less variation in packet size; further, the packet size variation has smaller effect on the packet error rate, since basically an I-frame packet consists of much larger number of RLC bursts than a P-frame. From figure 9, we also observe that the average I-frame packet error rate is about 45.7%, which is significantly larger than the average P-frame packet error rate 6%. This is again caused by the fact that the average I-frame packet size are significantly larger than the average P-frame packet size. The I-frame video packet would require more RLC bursts to be transmitted, thus would be easier to get error under the same channel condition, when compared with the smaller P-frame packet. The high I-frame error rate is undesirable since I-frame packets are much more important for partial information decoding in the video player. Therefore, we keep the allowable P-frame ARQ number to be 3 only, and increase the allowable I-frame ARQ number to 4. The new simulation result is shown in figure 10. We find that the I-frame packet error rate drops to as low as 3.7% when ARQ = 4, as in figure 10(a); while the P-frame packet error rate does not change too much, as in figure 10(b). This demonstrates that we can differentiate the service quality for the I-frame and P-frame data by controlling the allowable ARQ number for each priority queue, such that the high priority I-frame data can achieve better QoS. When the allowable I-frame ARQ number is increased to 5, only minor improvement is achieved for I-frame error rate. As a result, we can conclude that only limited number of ARQ's is needed to sufficiently preserve the I-frame data.

Since the erroneous video frames may still be used to reconstruct the video pictures by the decoder, we calculate the peak signal-to-noise ratio (PSNR) measures for the erroneous video frames, and show the results in table 2. We find that by taking priority scheduling, and by allowing more ARQ transmission for I-frame bursts, we can reduce the I-frame packet error rate, and improve the PSNR value for the erroneous I-frames. The results demonstrate that by taking proper link level scheduling and ARQ control, less I-frame would get error; and for those erroneous I-frames, the degree of damage can also be noticeably less. The results offer hints for designing optimal link level scheduling and control protocol in order to facilitate faithful video playback in time-varying wireless channel condition.

6. Conclusions

In this paper we study the link level performance of video transmission over the uplink of an unreliable wireless channel, under the assumption that the link level errors are correlated and follow a hidden Markov model. The DBMAP process with marked transitions is proposed to model the video source. The RLC data burst transmission time is proved to follow probabilistic phase type distribution. We show that the link level performance of a wireless video transmission system can be modeled by a DBMAP/PH/1 priority queue, and present computation algorithm and numerical results for the queueing model. Extensive simulation results on the queue length and waiting time of the video transmission buffer has been presented. Packet level error rate and PSNR value for the erroneous video frames are reported. The results demonstrate that video quality can be substantially improved by taking proper link level scheduling and ARQ control. We are currently working on developing traffic model for scalable video encoded in multiple layers based on the marked DBMAP process. We are also interested in performance evaluation of different scheduling protocols for scalable video transmission over the wireless networks.

Acknowledgement

Bo Li's research was supported in part by grants from RGC under contracts HKUST6196/02E and HKUST6402/03E, a NSFC/RGC joint grant under contract N_HKUST605/02, and a grant from Microsoft Research under the contract MCCL02/03.EG01.

References

[1] A.S. Alfa, Matrix-geometric solution of discrete time MAP/PH/1 priority queue, Naval Research Logistics 45 (1998) 23–50.

[2] J. Beran, R. Sherman, M.S. Taqqu and W. Willinger, Long-range dependence in variable-bit-rate video traffic, IEEE Transactions on Communications 43 (1995) 1566–1579.

[3] N. Blefari-Melazzi, G. Russo and P. Talone, Traffic modelling of video sources codified at very low bit rate with the H.263 test model, in: *Proc. of IEEE ICC'97* (1997) pp. 528–533.

[4] C. Blondia, A discrete time batch Markovian arrival process as B-ISDN traffic model, Belgian Journal of Operations Research, Statistics and Computer Science 32(3/4) (1993).

[5] M.W. Garrett and W. Willinger, Analysis, modeling and generation of self-similar VBR video traffic, in: *Proc. of ACM SIGCOMM'94*, London, UK (August 1994) pp. 269–280.

[6] E.N. Gilbert, Capacity of a burst-noise channel, The Bell System Technical Journal 39 (September 1960) 1253–1265.

[7] W.K. Grassmann and D.A. Stanford, Matrix analytic methods, in: *Computational Probability* (Kluwer Academic, 1999) pp. 153–203.

[8] E. He, F. Du, X. Dong, L.M. Ni and H.D. Hughes, Video traffic modeling over wireless networks, in: *Proc. of IEEE ICC'2000* (2000) pp. 536–542.

[9] Q.-M. He and M.F. Neuts, Markov chains with marked transitions, Stochastic Processes and Their Applications 74(1) (1998) 37–52.

[10] D.P. Heyman and T.V. Lakshman, What are the implications of long-range dependence for VBR-video traffic engineering?, IEEE/ACM Transactions on Networking 4 (June 1996) 301–317.

[11] D.P. Heyman, A. Tabatabai and T.V. Lakshman, Statistical analysis and simulation study of video teleconference traffic in ATM networks, IEEE Transactions on Circuits and Systems for Video Technology 2(1) (1992) 49–59.

[12] C. Huang, M. Devetsikiotis, L. Lambadaris and A. Kaye, Modeling and simulation of self-similar variable bit rate compressed video: a unified approach, in: *Proc. of ACM SIGCOMM'95*, Cambridge, MA (August 1995).

[13] M. Izquierdo and D. Reeves, A survey of statistical source models for variable-bit-rate compressed video, Multimedia Systems 7(3) (1999) 199–213.

[14] P.R. Jelenkovic, A.A. Lazar and N. Semret, The effect of multiple time scales and subexponentiality of MPEG video streams on queueing behavior, IEEE Journal on Selected Areas in Communications 15(6) (1997) 1052–1071.

[15] A. Klemm, C. Lindemann and M. Lohmann, Traffic modeling of IP networks using the batch Markovian arrival process, in: *Proc. of 12th International Conference on Modeling Tools and Techniques for Computer and Communication System Performance Evaluation*, London, UK (April 2002).

[16] R. Koenen, MPEG-4 Overview – V.18 – Singapore Version, ISO/IEC JTC1/SC29/WG11 N4030 (March 2001).

[17] S.Q. Li and C.L. Hwang, Queue response to input correlation functions: continuous spectral analysis, IEEE/ACM Transactions on Networking 1(6) (1993) 678–692.

[18] S.Q. Li, S. Park and D. Arifler, SMAQ: a measurement-based tool for traffic modeling and queuing analysis, Part I: design methodologies and software architecture, IEEE Communications Magazine 36(8) (1998) 56–70.

[19] A. Lombardo, G. Morabito and G. Schembra, An accurate and treatable Markov model of MPEG-video traffic, in: *Proc. of IEEE INFOCOM'98* (1998) pp. 217–224.

[20] D.M. Lucantoni, K.S. Meier-Hellstern and M.F. Neuts, A single-server queue with server vacations and a class of non-renewal arrival processes, Advances in Applied Probability 22 (1990) 676–705.

[21] M.F. Neuts, *Matrix-Geometric Solutions in Stochastic Models – An Algorithmic Approach* (The Johns Hopkins University Press, 1981).

[22] M.F. Neuts, *Structured Stochastic Matrices of M/G/1 Type and Their Applications* (Marcel Dekker, New York, 1989).

[23] S. Rappaport, *Wireless Communications: Principles and Practice* (Prentice-Hall, Englewood Cliffs, NJ, 1996).

[24] O. Rose, Statistical properties of MPEG video traffic and their impact on traffic modeling in ATM systems, in: *Proc. of 20th Conference on Local Computer Networks* (1995) pp. 397–406.

[25] B. Sklar, Raleigh fading channels in mobile digital communication systems, Part I: characterization, IEEE Communications Magazine 35(7) (1997).

[26] W. Turin, *Digital Transmission Systems: Performance Analysis and Modeling* (McGrawHill, 1998).

[27] J. Vilasenor, Y.-Q. Zhang and J. Wen, Robust video coding algorithm and systems, Proceedings of IEEE 87(10) (1999).

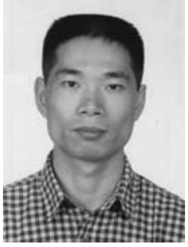
[28] D.-P. Wu, Y.T. Hou and Y.-Q. Zhang, Scalable video coding and transport over broadband wireless networks, Proceedings of IEEE 89(1) (2001).

[29] D.-P. Wu, Y.T. Hou, W. Zhu, H.J. Lee, Y.-Q. Zhang and H.J. Chao, On end-to-end architecture for transporting MPEG-4 video over the Internet, IEEE Transactions on Circuit Systems for Video Technology 10(6) (2000).

[30] Y.Q. Zhao, Censoring technique in studying block-structured Markov chains, in: *Proc. of the Third International Conference on Matrix Analytic Methods in Stochastic Models*, Leuven, Belgium (2000).

[31] J.-A. Zhao, B. Li, X.-R. Cao and I. Ahmad, Matrix analytic solution for DBMAP/PH/1 priority queues, Queueing Systems, Theory and Applications, submitted.

[32] M. Zorzi and R. Rao, On channel modeling for delay analysis of packet communications over wireless links, in: *Proc. of the 36th Annual Allerton Conference*, Allerton House, Monticello, IL (September 1998).



Ji-An Zhao received the B.E. degree in information science from the Beijing Information Technology Institute, Beijing, China, in 1992, and the M.E. degree in computer science from the Computing Center, Chinese Academy of Sciences, Beijing, China, in 1995. From 1995 to 1997 he was a Ph.D. student at the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. Since Spring 1998 he is a Ph.D. student at the Computer Science Department, the Hong Kong University of Science and Technology. His research interests include MPEG-4 video transmission, wireless and mobile networks, performance evaluation, queueing analysis and real-time computing. He is an IEEE student member.
E-mail: zhaojian@cs.ust.hk

Bo Li received the B.S. (summa cum laude) and M.S. degrees in computer science from Tsinghua University, Beijing, P.R. China, in 1987 and 1989, respectively, and the Ph.D. degree in computer engineering from University of Massachusetts at Amherst in 1993. Between 1994 and 1996, he worked on high performance routers and ATM switches in IBM Networking System Division, Research Triangle Park, North Carolina. Since then, he has been with the Computer Science Department, the Hong Kong University of Science and Technology. His current research interests include wireless mobile networking supporting multimedia, video multicast and all optical networks using WDM. He has published over 60 journal papers in the above areas. He has been an editor or a guest editor for 15 journals such as IEEE Transactions on Wireless Communications, IEEE Transactions on Vehicular Technology, IEEE Journal on Selected Areas in Communications, IEEE Communications Magazine, ACM Journal of Wireless Networks, ACM Mobile Computing and Communications, ACM Performance Evaluation, ACM Mobile Network and Applications etc. He currently serves as the Co-TPC Chair for IEEE Infocom'2004.
E-mail: bli@cs.ust.hk

Chi-Wah Kok received the B.Eng. and M.Phil. degrees from the City University of Hong Kong, and the M.Sc. and Ph.D. degrees from the University of Wisconsin, Madison, all in electrical and engineering, in 1990, 1993, 1996 and 1997, respectively. He joined the Department of Electrical and Electronic Engineering, Hong Kong University of Science and Technology since 1999, and is currently an assistant professor. Prior to that he was with several industrial corporations, and universities which include SONY US Research Laboratory, Stanford University and Lattice Semiconductor Corp. His research interests are in the area of signal processing, with applications in multimedia and communications.
E-mail: eekok@ust.hk

Ishfaq Ahmad received a B.Sc. degree in electrical engineering from the University of Engineering and Technology, Lahore, Pakistan, in 1985, and an M.S. degree in computer engineering and a Ph.D. degree in computer science from Syracuse University, New York, USA, in 1987 and 1992, respectively. His recent research focus has been on developing parallel programming tools, scheduling and mapping algorithms for scalable architectures, heterogeneous computing systems, distributed multimedia systems, video compression techniques, and web management. His work in the above areas is published in over 100 technical papers in refereed journals and conferences, with best paper awards at Supercomputing 90 (New York), Supercomputing 91 (Albuquerque), and 2001 International Conference on Parallel Processing (Spain). He is currently a Professor of Computer Science and Engineering in the CSE Department of the University of Texas at Arlington. He is an Associate Editor of Cluster Computing, Journal of Parallel and Distributed Computing, IEEE Transactions on Circuits and Systems for Video Technology, IEEE Concurrency, and IEEE Distributed Systems Online.
E-mail: iahmad@cse.uta.edu