

## Research Article

# Dynamic Bandwidth Allocation Based on Online Traffic Prediction for Real-Time MPEG-4 Video Streams

Yao Liang and Mei Han

*Department of Electrical and Computer Engineering, Advanced Research Institute,  
Virginia Polytechnic Institute and State University, Arlington, VA 22203, USA*

Received 12 August 2005; Revised 15 April 2006; Accepted 4 June 2006

Recommended by Ming-Ting Sun

The distinct characteristics of variable bit rate (VBR) video traffic and its quality of service (QoS) constraints have posed a unique challenge on network resource allocation and management for future integrated networks. Dynamic bandwidth allocation attempts to adaptively allocate resources to capture the burstiness of VBR video traffic, and therefore could potentially increase network utilization substantially while still satisfying the desired QoS requirements. We focus on prediction-based dynamic bandwidth allocation. In this context, the multiresolution learning neural-network-based traffic predictor is rigorously examined. A well-known-heuristic based approach RED-VBR scheme is used as a baseline for performance evaluation. Simulations using real-world MPEG-4 VBR video traces are conducted, and a comprehensive performance metrics is presented. In addition, a new concept of renegotiation control is introduced and a novel renegotiation control algorithm based on binary exponential backoff (BEB) is proposed to efficiently reduce renegotiation frequency.

Copyright © 2007 Hindawi Publishing Corporation. All rights reserved.

## 1. INTRODUCTION

Variable bit rate (VBR) video traffic, generated from diverse multimedia applications, is expected to be a significant portion of traffic in future integrated services networks. VBR videos require a sophisticated network resource allocation and management (NRAM) mechanism to support their strict delay and loss requirements. Due to the nature that VBR videos typically exhibit burstiness over multiple time scales [3, 4], it poses a unique challenge to the design of NRAM mechanism to achieve high overall network utilization, while still preserving the desired quality of service (QoS) requirements. The challenge is essential particularly for real-time VBR video transmission in which the traffic trace is unknown in advance. Clearly, any static allocating bandwidth (i.e., allocating exactly once for the entire video transmission session) for real-time VBR video streams would result in either inefficient utilization of bandwidth due to over allocation of resource or insufficient to support the required QoS due to the under allocation of resource. Dynamic bandwidth allocation attempts to adaptively allocate resources to capture the burstiness of VBR video traffic, and hence could be particularly promising for real-time VBR video transmissions [2, 5, 6]. One fundamental issue

in any dynamic bandwidth allocation mechanism is how to sense the traffic dynamics in advance in order to determine the resource needs. One approach is to devise some *heuristic* method to estimate the future traffic characteristics, such as RED-VBR scheme [2]. The other approach, called *predictive* or *prediction-based* dynamic bandwidth allocation, is to develop more sophisticated online network traffic predictor as a critical component in dynamic bandwidth allocation scheme to predict the future traffic [1, 5, 6]. Predictive dynamic bandwidth allocation introduces more complexity due to the additional complexity of online network traffic predictor, in the hope of being able to capture more accurately the dynamics of VBR video traffic and therefore to improve the performance. Yet, another technique to address efficient VBR video transmission is smoothing (e.g., [7, 8]).

Several schemes regarding predictive dynamic bandwidth allocation have been proposed. Chong et al. [5] presented a method of real-time video traffic prediction in the frequency domain to support VBR video transmission. Adas [9] studied adaptive linear regression traffic prediction for renegotiated CBR service. However, the work in both [5, 9] only considered single-step-ahead traffic prediction. In contrast, Liang [1] investigated multiresolution learning-based neural network (MRL-NN) online traffic predictor for long-term

traffic prediction up to hundreds of frames into the future, but the performance of MRL-NN traffic predictor was evaluated in terms of normalized mean square error (NMSE) not in a network setting. Wu et al. [6] investigated dynamic resource allocation through video content and traffic statistics, but under the assumption of a simplified service model/policy. Chiruvolu et al. [10] focused on allocation algorithm at ATM cell level, while Yoo et al. [11] focused on bandwidth allocation criteria rather than traffic prediction. Although a number of concrete predictive dynamic bandwidth allocation methods exist, little research has been conducted on predictive dynamic bandwidth allocation with a *realistic* service policy. The simplified service policy adopted in the previous studies (e.g., [6, 12]) assumes that any video source will be blocked whenever its bandwidth renegotiation is rejected, which is impractical. This motivates our work. A long-term VBR video traffic predictor is certainly desirable in such a study, because single-frame-ahead network traffic predictors (e.g., [5, 9, 10, 12]) would result in too frequent bandwidth renegotiations and thus the heavy reallocation control overhead which might eventually eliminate all merits that could be obtained from online traffic prediction. To this end, MRL-NN traffic predictor developed in [1] is employed and examined. Unlike the existing work [1, 5, 6, 9–11] which was performed using MPEG-1 or MPEG-2 video traces, we investigate real-world MPEG-4 video traces which have not been studied in dynamic bandwidth allocation. MPEG-4 video traffic might provide unique challenges to online traffic prediction and predictive dynamic bandwidth allocation. As noted in [13], the autocorrelation functions for MPEG-4 traces have more significant long-range dependencies (LRDs) than those for the similar MPEG-1 traces, and signals with significant LRD are typically more difficult to predict. To carefully investigate the strength and limitation of predictive dynamic bandwidth allocation approach, RED-VBR scheme [2], a well-known heuristic dynamic bandwidth allocation, is used as a baseline for the performance evaluation. By applying *realistic* service policy (described in Section 5.2), we are able to present our simulation results using more comprehensive performance metrics of utilization/delay, drop rate and renegotiation internal, as opposed to utilization/delay, and renegotiation internal used in the previous studies [6, 12]. This way, we are able to provide new insight which has not been available yet.

Another unique contribution of this work is to introduce a new concept and method of *renegotiation control* in predictive dynamic bandwidth allocation architecture. Renegotiation control is about to handle the reissue of renegotiation request whenever the previous renegotiation request fails due to lack of resource. A novel renegotiation control algorithm is proposed based on binary exponential backoff (BEB). Our simulation shows that the proposed renegotiation control algorithm can significantly reduce the renegotiation frequency.

The rest of the paper is organized as follows. Section 2 describes the general architecture of predictive dynamic bandwidth allocation. Section 3 briefly reviews MRL-NN traffic predictor [1]. In Section 4, renegotiation control algorithm based on BEB is proposed. Section 5 presents our empirical

study with careful analysis, and provides new insight. Conclusions and future work are given in Section 6.

## 2. PREDICTIVE DYNAMIC BANDWIDTH ALLOCATION

A general architecture of predictive dynamic bandwidth allocation can be illustrated as in Figure 1, in which the component of renegotiation control is newly proposed in this paper and will be discussed in Section 4. The other components shown in Figure 1 are briefly described as follows.

### (1) Extraction of traffic statistics/video content of VBR video streams

This is preprocessing for the online traffic prediction. Different types of preprocessing are required for the different approaches of online traffic prediction. For example, with video traffic-based predictor, some simple traffic statistics will be calculated as the input for the predictor, while with video content-based predictor, some sophisticated video content analysis technique will be performed to extract content features as the input for the predictor. Therefore, the complexity of this component is largely dependent upon the approach used for the online traffic prediction.

### (2) Online traffic predictor

This is a critical component of any predictive dynamic bandwidth allocation schemes. In principle, three types of online traffic predictors are currently available: traffic-based approach (e.g., [1, 5, 9]), video-content-based approach [12], and a combination of the above two [6].

### (3) VBR traffic descriptor computation

The traffic descriptor computation is based on the traffic model to be used for VBR video streams, such as deterministic bounding interval-length dependent (D-BIND) model [14], leaky bucket model [15] and so forth, and traffic descriptor attempts to describe and capture the characteristics of VBR video streams as accurately as possible.

### (4) Segmentation and future resource determination

Segmentation is to determine when individual VBR streams should renegotiate bandwidth resource with the system in order to preserve their required QoS. In other words, segmentation algorithm is about how to choose the renegotiation points along the VBR video transmission session, by which renegotiation intervals are generated. Two basic methods for segmentation are the fixed intervals and adaptive intervals, and the latter can be further divided into two categories: traffic-based and content-based. More details can be found in [6]. Future bandwidth determination is to determine how much bandwidth resource should be reserved for each renegotiation request.

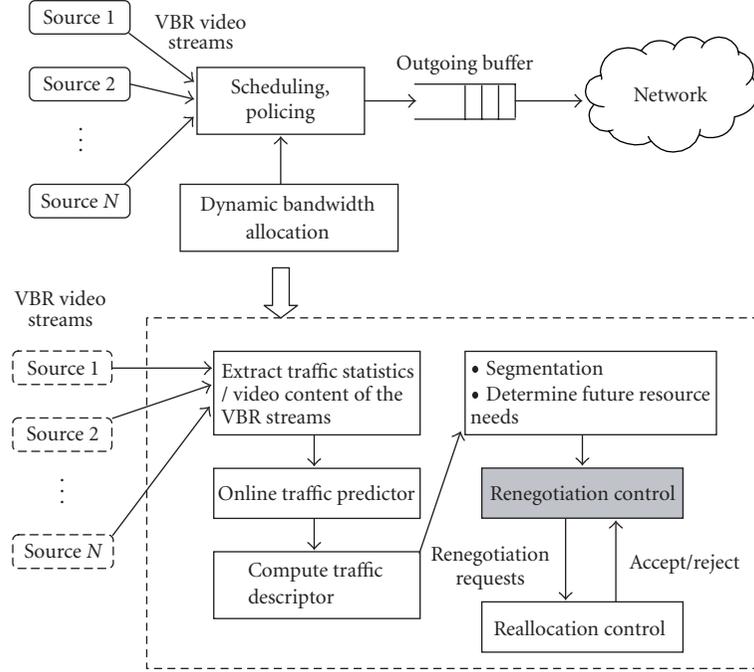


FIGURE 1: Dynamic bandwidth allocation architecture.

### (5) Reallocation control

This is to determine, based on the traffic model/descriptor used, whether the requested resource is honored or rejected due to the availability of resource. Connection admission control can also be consolidated into this component. Once the requested resource is honored, the resource reservation for the corresponding VBR video streams should be updated to reflect this new change. It also signals the scheduling and/or policing to be updated accordingly. The overall updated resource reservation will be used in the connection admission control later to determine if a new connection can be admitted.

## 3. ONLINE NETWORK TRAFFIC PREDICTOR

Aimed at maximizing the merit of prediction-based dynamic resource allocation, online VBR video traffic prediction has attracted a lot of attention. Most of the existing work (e.g., [5, 9, 10, 12]) only considered one-frame-ahead (i.e., single-step-ahead) traffic prediction due to the difficulty of highly burst, nonlinear, and self-similar nature of the VBR video traffic. In contrast, MRL-NN traffic predictor is for long-term VBR video traffic prediction [1], and is a three-layer feedforward neural network (NN) using multiresolution learning paradigm [16]. The inputs of the traffic predictor are the recent and current frame sizes and the output of the predictor is the predicted next frame size in the future. With iterated multistep prediction method, the predictor can predict as many frames into the future as needed. The multiresolution learning paradigm attempts to exploit the correlation structures in video stream traffic at

multiple resolution levels by first decomposing the original training traffic and approximating the traffic at different levels of detail, and then training the neural network traffic predictor from the coarsest resolution level to the finest resolution level during the learning process. Multiresolution analysis in wavelet theory is employed as a mathematical tool to decompose and approximate the original training traffic data for the MRL-NN traffic predictor. The MRL-NN traffic predictor is adopted in this predictive dynamic bandwidth allocation study.

Our online MRL-NN traffic predictor has structure 24-5-1, indicating a dimension of 24 for input layer, 5 hidden neurons with typical sigmoid activation function in the hidden layer, and one linear output neuron in the output layer. The choice of 24 input dimensions is to cover one or a few full size of group of picture (GOP) to readily capture traffic dynamics in a GOP and some dynamics between GOPs in video traffic. The chosen 24-5-1 structure is mainly empirical. With the multiresolution learning paradigm, three-resolution learning process is employed to construct the online traffic predictor, as an ordered sequence of learning/training activities  $A_1(r_1) \rightarrow A_2(r_2) \rightarrow A_3(r_3)$ , where  $A_1(r_1)$  represents a training activity for network traffic predictor using traffic data representation  $r_1$  at the coarsest resolution level, while  $A_3(r_3)$  represents a training activity at the finest resolution level of the original traffic data  $r_3$ . The first training activity  $A_1(r_1)$  starts with randomly initialized connection weights, and each subsequent learning activity starts with the connection weights resulting from the previous learning activity (see [1] for more details). In our study, 512 frames of each video trace will be initially used to train the online traffic predictor, and then the online predictor will be able to predict

the future  $M$  consecutive video frames iteratively. Here  $M$  is the look-ahead window size for the predicted future frames. Toward the end of each current window, online traffic prediction is activated for every admitted VBR video stream to produce the traffic prediction for the next  $M$  frames. In this study, we use fixed look-ahead prediction window size  $M$  (e.g.,  $M = 48$ ). Note that, as the admitted video streams proceed, the accuracy of the online traffic predictor will be gradually degraded. To overcome this, a periodical online updating of traffic predictor is necessary to maintain prediction accuracy. In our simulations, the traffic predictor is updated (e.g., slightly retrained) after every 10 800 frames (i.e., 7.2 minutes). It is an advantage of the MRL-NN traffic predictor that only the finest resolution learning is conducted for updating [1], as the fast dynamics of video traffic is more associated with the finest resolution. Thus, the traffic predictor's updating is much faster than its initial training.

#### 4. RENEGOTIATION CONTROL

When a renegotiation request is issued for any video transmission session, it will be checked immediately by reallocation control to verify if there is sufficient resource to meet the request. If the renegotiated resource can be accommodated, the renegotiation request is accepted, and otherwise is rejected. Currently, whenever a renegotiation request fails due to lack of resource, a dynamic bandwidth allocation scheme is simply to reissue the renegotiation request *consecutively* at the next frame time (e.g., RED-VBR). Given a frame rate  $f$  (frame/sec.), any two consecutive renegotiation requests are only at one frame time  $1/f$  apart. Although this approach tries to explore the statistical multiplexing gain (SMG) at fine resolution level, it could unnecessarily introduce a lot of renegotiation and control overhead due to potential consecutive renegotiation failures when the system lacks resource. In view of this, we introduce a new concept and method, referred to as renegotiation control, into dynamic bandwidth allocation architecture (as illustrated in Figure 1). The idea of the proposed renegotiation control is to avoid consecutive renegotiation requests after the previous renegotiation failure, by exponentially slowing down the process of reissuing renegotiations when the system is undergoing resource insufficiency. In this regard, the current aggressive (frame-by-frame) practice of re-renegotiations can be viewed as a default renegotiation control method. Inspired by the success of binary exponential backoff (BEB) algorithm in Ethernet and IEEE 802.11 distributed coordination function (DCF) for wireless local area networks, we believe that BEB can also naturally fit in the renegotiation control situation identified above. Thus, a BEB-based renegotiation control algorithm is proposed as in Algorithm 1, and is evaluated through simulations in Section 5.

#### 5. EMPIRICAL STUDY

Before we present our empirical study, we will briefly discuss the traffic descriptor used, the formalization of network link utilization computation, and the RED-VBR scheme which is

used as baseline for performance evaluation in our empirical study. (See [2, 14] for more details.)

#### 5.1. Traffic descriptor, link utilization, and RED-VBR

##### 5.1.1. Traffic descriptor

VBR video traffic descriptors describe the characteristics of VBR video traffic in integrated services networks with the use of countable parameters. Among available VBR video traffic descriptors, D-BIND-model-based traffic descriptor [14] is aimed to capture the traffic burstiness over different time scales. The D-BIND model has been widely studied and demonstrated in heuristic dynamic bandwidth allocation, such as RED-VBR scheme [2]. Therefore, D-BIND descriptor is adopted in our study. Let  $A[\tau, \tau + t]$  denote the cumulative bits generated by a source arriving in an interval ranging from  $\tau$  to  $\tau + 1$ , the so-called empirical envelop  $B^*(t)$ , the tightest time-invariant bound over the interval, is

$$B^*(t) = \sup_{\tau > 0} A[\tau, \tau + t] \quad \forall t > 0. \quad (1)$$

The D-BIND traffic model [14] characterizes the arrival traffic with a set of rate-interval pairs  $R_T = \{(r_k, t_k) \mid k = 1, 2, \dots, P\}$ , where  $r_k = q_k/t_k$ , the bounding rate over any interval of length  $t_k$ , and  $q_k$  is the maximum total number of bits over any interval of length  $t_k$ . Here  $P$  is the dimension of D-BIND traffic descriptor. Therefore, the traffic constraint function  $B_{R_T}$  of the D-BIND traffic model is

$$B_{R_T} = r_k t_k + \frac{r_k t_k - r_{k-1} t_{k-1}}{t_k - t_{k-1}} (t - t_k), \quad t_{k-1} \leq t \leq t_k, \quad (2)$$

with the assumption of  $B_{R_T}(0) = 0$  at  $t_0 = 0$ . For a video sequence with  $M$  frames, the tightest D-BIND traffic constraint function  $B_{R_T}^*$  can be constructed by the consecutive  $M$  rate-interval pairs (i.e.,  $P = M$ ). The relationship among the cumulative arrival function  $A(0, t)$ , the empirical envelope  $B^*(t)$ , the tightest D-BIND traffic constraint function  $B_{R_T}^*(t)$ , and the general D-BIND traffic constraint function  $B_{R_T}(t)$  is illustrated in Figure 2 (see [14] for more details).

##### 5.1.2. Network link utilization

Let  $d$  denote delay bound. Assume that the maximum capacity  $Q_{\max}$  of network buffer is sufficiently large for all possible delay bounds in the study. For a first-come-first-serve (FCFS) queuing policy, the upper bound on delay for all connections is computed as

$$d = \frac{1}{C} \max_{t \in t_k} \left\{ 0, \left\{ \sum_{j=1}^N B_{R_T}^j(t) + QL(t) - Ct \right\} \right\}, \quad (3)$$

$$j = 1, 2, \dots, N, \quad QL(0) = 0,$$

where  $C$  is the capacity of the outgoing link,  $N$  is the maximum number of connected video streams at delay bound  $d$ , and  $QL(t)$  is network buffer occupancy at time  $t$ . Given a certain blocking/rejection probability, link utilization  $u(d)$ ,

<b>Notation:</b>	$F$ :	the number of consecutive renegotiation rejections;
	$Delay$ :	the total time (number of frames) to be delayed for the next renegotiation;
	$M$ :	the look-ahead window size of predicted frames;
	$(f_1, f_2, \dots, f_M)$ :	$M$ consecutive frame sizes;
	$flag$ :	FAILURE/SUCCESS, the indication of renegotiation failure/success;
	$w$ :	the number of frames being waited before issuing the next renegotiation request.
<b>Initialization:</b>	$F \leftarrow 0; w \leftarrow 0; Delay \leftarrow 0;$	
<b>Renegotiation_Control:</b>	<b>if</b> ( $w \geq Delay$ ) <b>then</b> <b>for</b> $j \leftarrow 1$ <b>to</b> <i>number-of-admitted-streams</i> <b>do</b> Predict the next $M$ frames $(f_1, f_2, \dots, f_M)$ for stream $j$ ; Update stream $j$ 's traffic descriptor based on the $(f_1, f_2, \dots, f_M)$ ; <b>end for</b> $flag \leftarrow Renegotiation\_Request()$ ; <b>if</b> ( $flag = SUCCESS$ ) <b>then</b> $F \leftarrow 0; w \leftarrow 0; Delay \leftarrow 0;$ <b>else</b> $F \leftarrow F + 1; w \leftarrow 0;$ $Delay \leftarrow$ Randomly choose an integer from $[0, 2^F - 1]$ ; $Delay \leftarrow \min(Delay, M)$ ; <b>end if</b> <b>end if</b>	
<b>Video_Transmissions:</b>	<b>while</b> (outgoing buffer $\neq \emptyset$ ) Transmit the video frame from the head of the outgoing buffer; <b>if</b> ( $F > 0$ ) <b>then</b> $w \leftarrow w + 1;$ <b>end if</b> <b>end while</b>	

ALGORITHM 1: A BEB-based renegotiation control algorithm.

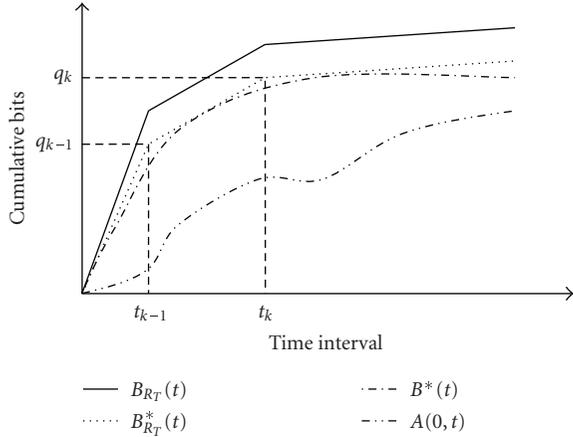


FIGURE 2: Illustration of traffic constraint functions.

a function of delay bound  $d$ , can be defined as follows:

$$u(d) = \frac{\sum_{i=1}^N \bar{R}_i (1 - \text{drop rate}(d))}{C}, \quad (4)$$

where  $\bar{R}_i$  is the  $i$ th video sources ultimate average rate, and  $\text{drop rate}(d)$  is the overall average drop rate at delay bound  $d$ .

### 5.1.3. RED-VBR scheme

RED-VBR [2] is used as a baseline to evaluate the performance of predictive dynamic bandwidth allocation. RED-VBR is a renegotiated VBR service and uses D-BIND traffic descriptor. Two parameters  $\alpha$  and  $\beta$  are used in RED-VBR to determine resource renegotiation. When D-BIND traffic descriptor indicates that the currently measured bandwidth exceeds the reserved resource, a renegotiation takes place, and it will reserve a new amount of bandwidth by multiplying the currently measured value by a factor  $\alpha$  ( $\alpha > 1$ ). When D-BIND traffic descriptor indicates that the currently measured bandwidth is consistently less than the reserved resource by a factor  $\beta$  ( $\beta < 1$ ), the currently reserved resource level will be lowered to release some resource, and any release renegotiation will be always successful. The choice of  $(\alpha, \beta)$  then impacts the tradeoff between average of renegotiation frequency and the overall network link utilization [2]. The larger  $\alpha$  and smaller  $\beta$  will result in smaller renegotiation frequency but at the expense of decreased network utilization. Typical values of  $(\alpha, \beta)$  are, for example, (1.1, 0.9) and (1.3, 0.7).

## 5.2. Service policy

Focused on the renegotiated VBR (R-VBR), our empirical study is conducted using trace-driven simulation. While the

simulation method seems similar to that in [6, 12], one *essential* distinction is the different service policies used. It was assumed in [6, 12], for simplicity, that once a renegotiation request for more bandwidth from a video source is rejected, that video source will be completely blocked (i.e., stopped) until its resource request is finally satisfied later. This service policy is, however, impractical. In contrast, we assume that when a request of a video source for more bandwidth is rejected, that video source will not be blocked but (shaped instead) will continuously transmit with its previously allocated bandwidth. Clearly, the service policy we adopted is more realistic. It also allows us to observe and study video data drop rate under each dynamic bandwidth allocation scheme in addition to link utilization and renegotiation interval, which has not been reported in literature yet. When D-BIND traffic descriptor is used, R-VBR service could be deterministic at *segment* level (a segment is an interval between any two consecutive renegotiation points) but would be stochastic at connection *session* level due to the possibility of bandwidth renegotiation failures. Once a renegotiation request fails, bit drop becomes not only possible but also important with the realistic service policy adopted in our study, but would not occur with the simplified service policy assumed in [6, 12]. This insight explains why the drop rate was not considered in the previous studies (e.g., [6, 12]). As a result of adopting the realistic service policy, we are able to present a comprehensive network performance metrics as utilization/delay, drop rate, and renegotiation interval in our empirical study. This approach would enable us to systematically investigate and better understand the fundamental aspects of prediction-based dynamic resource allocation. Note that as admission control is not a focus of this work, admission is simply considered as a usual bandwidth renegotiation, with which if the requested bandwidth resource cannot be satisfied the video source is declined, otherwise the source is admitted. More sophisticated admission control will be a future work.

### 5.3. MPEG-4 video traces and simulation setup

MPEG-4 provides very efficient video coding covering a wide range from low bit rates for wireless communications to high bit rates for high-quality levels. In our work, two real-world MPEG-4<sup>1</sup> VBR video traces are used: movies *Mr. Bean* (high quality) and *Robin Hood* (low quality) [17]. The resolution of display is 176 pels (width)  $\times$  144 pels (height). Frame rate is 25 frames/s. The GOP of the video traces is IBBPBBPBBPBB. For high-quality encoding, the quantization parameters for I frames, P frames, and B frames were fixed at 4, while for low-quality encoding, the quantization parameters were fixed at 10 for I frames, 14 for P frames, and 18 for B frames. Mean bit rates are 0.58 Mb/s for *Mr. Bean* (high quality) and 0.19 Mb/s for *Robin Hood* (low quality), respectively. More details about the video traces can be found in [17]. The simulation has been carried out on 40 000 frames of each video trace (approximately 26 minutes). For the experiments for each video

trace, multiple video sources are generated through random starting points of the video trace. The link capacity  $C = 45$  Mb/s for high-quality *Mr. Bean*, and  $C = 11$  Mb/s for low-quality *Robin Hood*, respectively. For our prediction-based R-VBR, we have the following.

- (i) A traffic predictor is applied to each VBR video stream.
- (ii) For simplicity, a fixed size  $M$  of look-ahead prediction window is adopted, with which a basic fixed renegotiation interval of  $M$  frames is used for any admitted video stream for renegotiation request. That is, for each admitted video stream, its online traffic predictor periodically predicts next  $M$  frames at the end of each current prediction window. However, whenever the admission of a new video source into the network takes place, the online prediction will be made immediately or based on the renegotiation control algorithm. The bounding rates of D-BIND traffic descriptor are updated accordingly based on new prediction of the next look-ahead  $M$  frames. If any updated traffic descriptor indicates a demand of the incoming traffic of a video source exceeding or falling below its currently allocated bandwidth resource, a renegotiation request is issued. Note that renegotiation for bandwidth release will be always satisfied. In our prediction-based R-VBR scheme, if any admitted video stream's bounding rates are updated, the bounding rates of D-BIND descriptors for the other admitted video streams will be updated simultaneously also to ensure that the bandwidth reallocation for all video streams is based on the most updated predictions.
- (iii) The prediction window size  $M$  is chosen as 48 frames for *Robin Hood* and as 36 frames for *Mr. Bean*, which equals to their D-BIND window size used in RED-VBR scheme, respectively, in our study.
- (iv) Two renegotiation methods are tested with prediction-based R-VBR scheme. One is the commonly used default method which keeps renegotiating *consecutively* at the next frame time (i.e., frame-by-frame) whenever the previous renegotiation request has failed [2]. The other method is based on the renegotiation control algorithm proposed in Section 4.

### 5.4. Results and analysis

We first conduct our simulation using the same setup of blocking probability of 1% for our prediction-based R-VBR scheme and the benchmark scheme RED-VBR [2] as the way used in [6, 12]. The simulation results are shown in Figures 3 and 4, for low-quality *Robin Hood* and high-quality *Mr. Bean*, respectively, where RED-VBR scheme is simulated with two sets of parameters  $\alpha$  and  $\beta$ . Moreover, both schemes use the common default method of renegotiation control. As we see, for both video traces, the network utilization of prediction-based R-VBR is consistently and significantly higher than that of RED-VBR for any given delay bound. The average renegotiation interval of the prediction-based R-VBR is 1.28 second and 1.11 second for *Robin Hood* and *Mr. Bean*, respectively, which is in between that of RED-VBR

<sup>1</sup> MPEG-4 Part2.

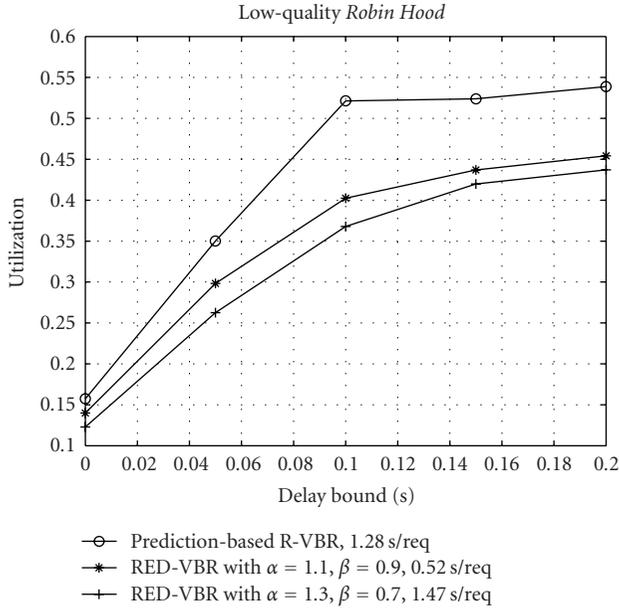


FIGURE 3: Comparison between online prediction-based R-VBR scheme and RED-VBR scheme for *Robin Hood*.

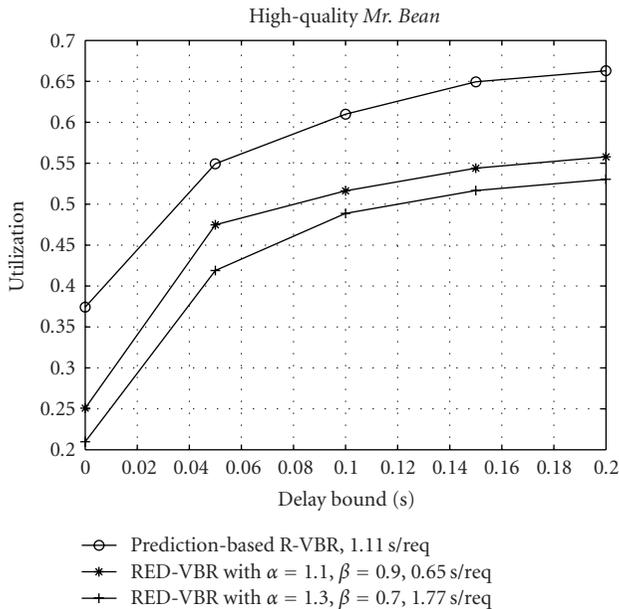


FIGURE 4: Comparison between online prediction-based R-VBR scheme and RED-VBR scheme for *Mr. Bean*.

( $\alpha = 1.3, \beta = 0.7$ ), that is, 1.47 second and 1.77 second, respectively, and that of RED-VBR ( $\alpha = 1.1, \beta = 0.9$ ) that is, 0.52 second and 0.65 second, respectively. This clearly indicates that the performance gain of the prediction-based R-VBR is beyond the adjustment of parameters  $\alpha$  and  $\beta$  of RED-VBR.

Table 1 shows the average drop rates for both schemes. While higher, the average drop rate of the prediction-based

R-VBR is still below 1%, which should well satisfy a wide range of real-time video applications. RED-VBR achieves a lower bit drop rate at the expense of significant lower network utilization. RED-VBR may be used to provide the video service with a very low (close to zero) drop rate. Prediction-based R-VBR can achieve around 20% utilization improvement with an average drop rate less than about 1%. Table 1 also lists the number of video streams accommodated into the network during our simulation. Clearly, the link utilization gains are due to the fact that the prediction-based R-VBR can accommodate significantly more simultaneous VBR video streams given the same link capacity. This indicates that prediction-based R-VBR can exploit SMG significantly better and utilize link capacity more efficiently.

It would be interesting to see the performance of the two approaches when they both accommodate the same number of video streams. To this end, we further conduct simulations with a different setup where the blocking probability for RED-VBR is increased until it accommodates the same number of video streams as those for the prediction-based R-VBR (i.e., both should have similar link utilization). Table 2 lists the performance results for both video traces. It is interesting to see that the average renegotiation interval for RED-VBR is dramatically reduced and much smaller than that of our prediction-based R-VBR (i.e., only about 1/3 for *Robin Hood*, and less than 1/2 for *Mr. Bean*, resp.).

To illustrate the limitation of single-step-ahead (i.e., single-frame-ahead) traffic prediction approach in dynamic bandwidth allocation, we further conduct simulations to compare the predictive dynamic bandwidth allocation schemes based on single-step-ahead predictor (SSP) with those based on the MRL-NN traffic predictor and RED-VBR. Our approach is to create a *perfect* SSP with zero prediction error in the simulations. Thus, a dynamic bandwidth allocation scheme based on the perfect SSP yields the *performance upper bound* for any possible SSPs. Movie *Mr. Bean* (high quality) is used in this experiment. The simulation setup is the same as that described in Section 5.3. Note that the prediction window size for *Mr. Bean* is 36, indicating that the employed MRL-NN traffic predictor is a 36-frame-ahead predictor. The simulation results are listed in Table 3. While the dynamic bandwidth allocation based on the perfect SSP can yield slightly better utilization performance compared to RED-VBR, its renegotiation interval is dramatically reduced to 0.04 second. Notice that the lapse of 0.04 second is one frame time given the frame rate of the video trace being 25 fps. This is not surprising because with single-frame-ahead traffic prediction, next frame has to be predicted for every frame. Accordingly, the dynamic bandwidth allocation has to be renegotiated for every frame time.

The second part of our empirical study focuses on examining the proposed renegotiation control algorithm applied to the prediction-based R-VBR scheme employing MRL-NN predictor through simulation. The link utilization results are shown in Figures 5 and 6, while the results of renegotiation frequency and bit drop rate are listed in Table 4. As can be expected, the proposed renegotiation control algorithm achieves slightly lower network link utilization compared to

TABLE 1: Comparison of bit drop rate and the number of video streams accommodated.

Video trace	Delay bound (s)	Average bit drop rate			Number of video streams accommodated		
		RED-VBR ( $\alpha = 1.1,$ $\beta = 0.9$ )	RED-VBR ( $\alpha = 1.3,$ $\beta = 0.7$ )	Prediction- based R-VBR	RED-VBR ( $\alpha = 1.1,$ $\beta = 0.9$ )	RED-VBR ( $\alpha = 1.3,$ $\beta = 0.7$ )	Prediction- based R-VBR
<i>Robin Hood</i>	0	0.01%	0	0.09%	8	7	9
	0.05	0	0	0.13%	17	15	20
	0.1	0.16%	0.07%	0.93%	23	21	30
	0.15	0.27%	0.20%	0.82%	25	24	30
	0.2	0.32%	0.25%	0.91%	26	25	31
<i>Mr. Bean</i>	0	0	0	0.63%	18	15	27
	0.05	0.05%	0	0.26%	34	30	38
	0.1	0.14%	0.02%	0.68%	37	35	44
	0.15	0.17%	0.06%	0.97%	39	37	47
	0.2	0.16%	0.08%	0.96%	40	38	48

TABLE 2: Performance comparisons of the prediction-based R-VBR versus RED-VBR with comparable network utilization.

Video trace	Delay bound (s)	RED-VBR ( $\alpha = 1.1, \beta = 0.9$ )			Prediction-based R-VBR without BEB		
		Average drop rate (%)	Utilization (%)	Average renegotiation interval (s)	Average drop rate (%)	Utilization (%)	Average renegotiation interval (s)
<i>Robin Hood</i>	0.05	0.13	35	0.326	0.13	35	1.073
	0.10	0.93	52	0.323	0.93	52	1.279
	0.15	0.82	52	0.431	0.82	52	1.349
	0.20	0.91	54	0.439	0.91	54	1.279
<i>Mr. Bean</i>	0.05	0.23	53	0.536	0.27	53	1.257
	0.10	0.61	61	0.467	0.68	61	1.107
	0.15	0.89	65	0.446	0.97	65	1.071
	0.20	0.90	66	0.450	0.97	66	1.080

that commonly used default method. This is because that the BEB-based renegotiation control algorithm will not intend to explore as much fine resolution SMG as the aggressive default method which keeps renegotiating *consecutively* frame-by-frame until success. Nevertheless, its utilization is still consistently and significantly better than that of RED-VBR ( $\alpha = 1.1, \beta = 0.9$ ). A unique gain of the proposed renegotiation control algorithm is to greatly reduce further the renegotiation frequency as shown in Table 4, and in addition, the lower average bit drop rate is also achieved. The dramatically reduced renegotiation frequency means the significant reduction of the overhead of traffic prediction, renegotiation computation and bandwidth reallocation control, compensated to the slightly decreased link utilization. Therefore, the proposed renegotiation control algorithm is favorable for better overall performance.

## 6. CONCLUSIONS

The key contributions of our work are (1) noting that a realistic service policy in trace-driven simulation for dynamic

resource allocation study is critical in revealing and understanding fundamental aspects of predictive dynamic bandwidth allocation approach; (2) showing significant improvement in performance of our predictive dynamic bandwidth allocation approach based on long-term online traffic prediction over RED-VBR and the predictive dynamic bandwidth allocation based on single-step-ahead traffic prediction; and (3) introducing the new concept of renegotiation control, and proposing and examining a BEB-based renegotiation control method. Rigorous empirical study has been conducted with MPEG-4 real-world VBR video traffic traces. The MRL-NN long-term traffic predictor [1] is examined. By adopting a realistic service policy, we present comprehensive performance metrics including bit drop rate for predictive dynamic bandwidth allocation simulation study. Our work suggests that the simplified service policy used in the previous studies (e.g., [6, 12]) might result in overestimating the benefits of prediction-based R-VBR as compared to heuristic-based R-VBR such as RED-VBR. The proposed BEB-based renegotiation control algorithm can effectively reduce potentially unnecessary connective

TABLE 3: R-VBR performance comparisons among perfect single-step-ahead prediction (SSP) based scheme, multistep-ahead prediction MRL-NN based scheme, and RED-VBR scheme.

Performance metrics	Delay bound (s)	Schemes		
		R-VBR with perfect SSP	R-VBR with MRL-NN predictor	RED-VBR ( $\alpha = 1.1, \beta = 0.9$ )
Utilization	0	27.9%	37.4%	25.1%
	0.05	48.8%	52.9%	47.5%
	0.1	53.0%	61.0%	51.6%
	0.15	55.7%	64.9%	54.4%
	0.2	57.1%	66.3%	55.8%
Drop rate	0	0.02%	0.63%	0.01%
	0.05	0.10%	0.27%	0.05%
	0.1	0.19%	0.68%	0.14%
	0.15	0.22%	0.97%	0.17%
	0.2	0.22%	0.97%	0.16%
Renegotiation interval (s)	0	0.04	1.03	0.66
	0.05	0.04	1.26	0.65
	0.1	0.04	1.11	0.66
	0.15	0.04	1.07	0.64
	0.2	0.04	1.08	0.64

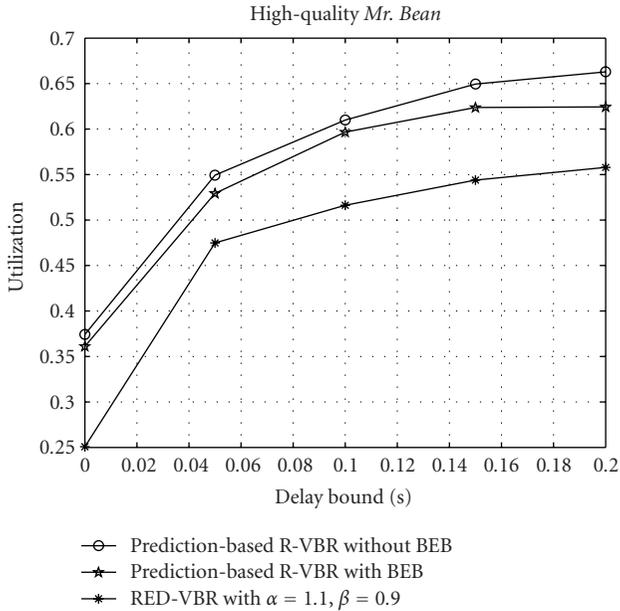


FIGURE 5: Effect of the BEB-based renegotiation control algorithm for *Mr. Bean*. (The performance of RED-VBR is also plotted to illustrate the minimal impact of the link utilization with BEB-based approach.)

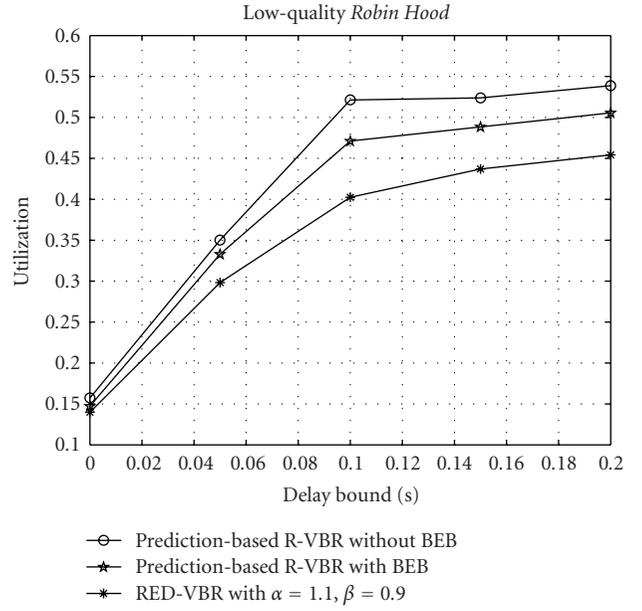


FIGURE 6: Effect of the BEB-based renegotiation control algorithm for *Robin Hood*. (The performance of RED-VBR is also plotted to illustrate the minimal impact of the link utilization with BEB-based approach.)

renegotiation failures, which is important for reducing renegotiation overhead. Although this study is conducted using R-VBR as a research vehicle, the insights and methodology

can also be applied to other types of dynamic bandwidth allocation schemes, such as renegotiated CBR (R-CBR). Future work includes investigating and quantitatively evaluating

TABLE 4: Effect of BEB-based renegotiation control algorithm on renegotiation interval and bit drop rate.

Video trace	Using BEB-based renegotiation control	Average drop rate	Average renegotiation interval (s)
<i>Robin Hood</i>	No	0.58%	1.18
	Yes	0.37%	1.81
<i>Mr. Bean</i>	No	0.70%	1.11
	Yes	0.52%	1.37

prediction-based R-CBR, as well as implementing real experimental system to conduct subjective comparison.

## REFERENCES

- [1] Y. Liang, "Real-time VBR video traffic prediction for dynamic bandwidth allocation," *IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews*, vol. 34, no. 1, pp. 32–47, 2004, (special issue on technologies that promote computational intelligence, openness and programmability in networks and Internet services—part III).
- [2] H. Zhang and E. W. Knightly, "RED-VBR: a renegotiation-based approach to support delay-sensitive VBR video," *Multimedia Systems*, vol. 5, no. 3, pp. 164–176, 1997.
- [3] M. W. Garrett and W. Willinger, "Analysis, modeling and generation of self-similar VBR video traffic," in *Proceedings of the ACM SIGCOMM Conference on Communications Architectures, Protocols and Applications*, pp. 269–280, London, UK, August–September 1994.
- [4] E. P. Rathgeb, "Modeling and performance comparison of policing mechanisms for ATM networks," *IEEE Journal on Selected Areas in Communications*, vol. 9, no. 3, pp. 325–334, 1991.
- [5] S. Chong, S. Li, and J. Ghosh, "Predictive dynamic bandwidth allocation for efficient transport of real-time VBR video over ATM," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 1, pp. 12–23, 1995.
- [6] M. Wu, R. A. Joyce, H.-S. Wong, L. Guan, and S.-Y. Kung, "Dynamic resource allocation via video content and short-term traffic statistics," *IEEE Transactions on Multimedia*, vol. 3, no. 2, pp. 186–199, 2001.
- [7] J.-Y. Le Boudec and O. Verscheure, "Optimal smoothing for guaranteed service," *IEEE/ACM Transactions on Networking*, vol. 8, no. 6, pp. 689–696, 2000.
- [8] J. Rexford, S. Sen, J. Dey, et al., "Online smoothing of live, variable-bit-rate video," in *Proceedings of the IEEE International Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV '97)*, pp. 235–243, St. Louis, Mo, USA, May 1997.
- [9] A. M. Adas, "Using adaptive linear prediction to support real-time VBR video under RCBR network service model," *IEEE/ACM Transactions on Networking*, vol. 6, no. 5, pp. 635–644, 1998.
- [10] G. Chiruvolu, R. Sankar, and N. Ranganathan, "VBR video traffic management using a predictor-based architecture," *Computer Communications*, vol. 23, no. 1, pp. 62–70, 2000.
- [11] S.-J. Yoo, K.-S. Kwak, and M. Kim, "Predictive and measurement-based dynamic resource management and QoS control for videos," *Computer Communications*, vol. 26, no. 14, pp. 1651–1661, 2003.
- [12] P. Bocheck and S.-F. Chang, "Content-based VBR traffic modeling and its application to dynamic network resource allocation," Res. Rep. 48c-98-20, Columbia University, New York, NY, USA, 1998.
- [13] A. Bhattacharya, A. G. Parlos, and A. F. Atiya, "Prediction of MPEG-coded video source traffic using recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 51, no. 8, pp. 2177–2190, 2003.
- [14] E. W. Knightly and H. Zhang, "D-BIND: an accurate traffic model for providing QoS guarantees to VBR traffic," *IEEE/ACM Transactions on Networking*, vol. 5, no. 2, pp. 219–231, 1997.
- [15] R. L. Cruz, "A calculus for network delay—I: network elements in isolation," *IEEE Transactions on Information Theory*, vol. 37, no. 1, pp. 114–131, 1991.
- [16] Y. Liang and E. W. Page, "Multiresolution learning paradigm and signal prediction," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2858–2864, 1997.
- [17] F. H. P. Fitzek and M. Reisslein, "MPEG-4 and H.263 video traces for network performance evaluation," *IEEE Network*, vol. 15, no. 6, pp. 40–54, 2001, (MPEG-4 traces, <http://trace.eas.asu.edu/TRACE/trace.html>).

**Yao Liang** received his B.S. degree in computer engineering and M.S. degree in computer science from Xi'an Jiaotong University, Xi'an, China. He received his Ph.D. degree in computer science from Clemson University, Clemson, SC, in 1997. He is currently an Assistant Professor in the Bradley Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University. His research interests include machine learning, neural networks, data mining, adaptive network control, multimedia networking, sensor networks and data management systems, and distributed and complex systems. Prior to joining Virginia Tech, he was a Technical Staff Member in Alcatel USA, Raleigh, NC, from 1997 to 2001. He is a Senior Member of the IEEE, and was a Member of the IEEE Signal Processing Society Machine Learning Technical Committee from 2003 to 2005. He has published numerous papers in refereed journals and international conferences proceedings.



**Mei Han** did receive the B.S. degree in electrical and electronic engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 1995, and the M.S. degree in electrical and computer engineering from Virginia Polytechnic Institute and State University (Virginia Tech), Alexandria, Virginia, in 2005. She was a Marketing Engineer from 1995 to 1999 and a Design Engineer from 1999 to 2001 at the Alcatel Shanghai Bell Co., Ltd. Her research interests include video communications and multimedia processing for wireless communications.

