A Virtual Queue Based Scheme To Support Real-Time Renegotiated VBR Video Streaming

Mei Han 1 and Yao Liang 2

1 Department of Electrical and Computer Engineering, Virginia Tech, Virginia, USA
2 Department of Computer and Information Science
Indiana University Purdue University Indianapolis, Indiana, USA
Email: yliang@cs.iupui.edu

Abstract: Variable bit rate (VBR) video traffic poses a unique challenge on network resource allocation and management for future packet networks. RED-VBR, a renegotiated deterministic VBR scheme, is a well-known approach proposed to support delay-sensitive VBR video traffic. However, the original RED-VBR suffers some limitations, such as the difficulty of dimensioning of D-BIND traffic descriptors for real-time videos, and the relatively high computation complexity. In this paper, we present a novel approach, referred to as virtual-queue-based RED-VBR, to overcome those limitations. In addition, we propose a simple and effective heuristic method to predict VBR video streaming performance in packet networks. Our proposed schemes have been demonstrated through extensive simulations with real-world MPEG-4 VBR video traces.

Keywords: multimedia networks, dynamic bandwidth allocation, quality-of-service, VBR video streaming, MPEG4.

1. Introduction

Variable bit rate (VBR) video traffic will be a significant portion of traffic in future packet-switched networks. VBR videos typically exhibit burstiness over multiple time scales (e.g., [1]), posing a unique challenge to achieve high overall network utilization while still preserving the required quality of service (QoS). Dynamic bandwidth allocation attempts to adaptively allocate network resources to capture the burstiness of VBR video traffic, and hence could be particularly promising. In general, dynamic bandwidth allocation for real-time VBR video transmissions can be in the forms of so-called either renegotiated constant bit rate (R-CBR) or renegotiated variable bit rate (R-VBR). RED-VBR, a renegotiated deterministic variable bit rate scheme proposed by Zhang and Knight [2], is a well-known R-VBR service. RED-VBR is based on the deterministic bounding interval-length dependent (D-BIND) traffic model [3]. A traffic model, or traffic descriptor, describes the characteristics of traffic with the use of countable traffic parameters. D-BIND traffic model specifies the traffic by a set of bounding rates over different intervals (i.e., rate-interval pairs) to accurately capture the burstiness of VBR video traffic. While D-BIND can only take the burstiness of VBR video traffic on smaller time scales into consideration, bandwidth renegotiation, on the other hand, attempts to effectively address the burstiness of VBR video traffic on longer time scales by allowing individual video source to renegotiate its resource request and QoS dynamically with the network system on-the-fly, when a video source’s rate changes dramatically. In this framework, consecutive renegotiation points divide the video stream into segments or renegotiation intervals. RED-VBR uses the online segmentation algorithm to determine when to renegotiate and how much more or less resources to request for the admitted real-time video streams. Any renegotiation request for more resource is honored or rejected due to the availability of resource, although the renegotiation request for resource release is always...
granted. If the requested resource is honored, the resource reservation for the corresponding VBR video stream will be updated to reflect this new change.

RED-VBR, with the use of the D-BIND traffic model, is aimed to build a statistical service on top of a deterministic service. The deterministic service is provided on segment level. As each source has to transmit according to the reserved D-BIND bounding rates, a deterministic service, guaranteeing no packet dropped or delivered beyond its delay bound, is provided for each segment. The statistical service is built on session level because renegotiation requests for more bandwidth may be denied/blocked due to the limited network resources. Thus, a renegotiation blocking probability is introduced to describe the probability of renegotiation failure for RED-VBR. However, the original RED-VBR suffers some limitations. One outstanding difficulty is how to choose the appropriate D-BIND window size in the RED-VBR for real-time video sources. That is, how many rate-interval pairs should be used in the RED-VBR. This is an important issue in both theoretical and practical sense. If the number of D-BIND rate-interval pairs used is too small, then the burstiness of VBR video traffic on small time scales cannot be effectively taken care of by the D-BIND traffic descriptor, which results in less efficiency in bandwidth renegotiation process. On the other hand, too many D-BIND rate-interval pairs add unnecessary implementation complexity of RED-VBR. In this paper (also see [4]), we attempt to address this difficulty and present a novel extension to the original RED-VBR, referred to as virtual-queue-based RED-VBR. With this virtual-queue-based RED-VBR approach, the number of D-BIND rate-interval pairs can be significantly reduced but still achieving comparable or better performance results as the original RED-VBR. Furthermore, the selection of the number of D-BIND rate-interval pairs in our approach is no longer difficult and tricky but rather straightforward and simple. Numerous simulations with real-world MPEG-4 VBR video traces have demonstrated our approach.

Individual VBR videos can exhibit dramatic different nature of traffic dynamics and burstiness. It is such individual video’s inherent nature of its traffic dynamics that makes network admission control and QoS management more complicated and inefficient. The existing research work regarding QoS of VBR video streaming and network utilization, in admission control (e.g., [5], [6]) and dynamic bandwidth allocation (e.g., [7], [8]), mainly focuses on the behavior of aggregate traffic with respect to the statistical multiplexing gain (SMG) of the network. While this aggregate traffic perspective is of significant importance and sheds light on the understanding of VBR video transmissions, the other important perspective of VBR video traffic – the inherent characteristics of each individual video flows did not receive enough attention. When the number of aggregated VBR video flows is not overwhelming in the network, which is common in practice, the different inherent characteristics of individual videos exhibit some clear “performance-related” nature beyond those basic descriptors (i.e., average rate, peak rate, etc.). Indeed, it is desirable to devise a simple and effective method to assess the inherent performance-related nature of individual VBR videos’ streaming in a packet network. To this end, we provide our insight into this problem, and propose a simple heuristic method to easily predict individual VBR videos’ inherent performances. Our method helps us to answer some fundamental and practical questions such as, for example, how to characterize and predict the inherent “performance” nature of individual VBR videos regardless of its aggregate traffic context?

The paper is organized as follows. Section 2 briefly reviews D-BIND and RED-VBR. We present our new virtual queue based RED-VBR approach in Section 3, and our simple method of predicting VBR video streaming performance in Section 4, respectively. Section 5 provides the simulation results and analyses. Finally, conclusions are given in Section 6.

2. D-BIND and RED-VBR

2.1 D-BIND

Let $A[\tau, \tau + t]$ denote the cumulative number of bits generated by a source arriving in the interval ranging from $\tau$ to $\tau + t$, and the empirical envelop $B^*(t)$, the tightest time-invariant bound over the
interval, is:

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

(1)

The D-BIND traffic model [3] characterizes the arrival traffic with a set of rate-interval pairs \( R_T = \{(r_k, t_k) \mid k = 1, 2, ..., P\} \), where \( r_k = q_k / t_k \). Here, \( r_k \) is the bounding rate is the total number of bits over the interval \( t_k \). Therefore, the traffic constraint function \( B_{R_T} \) of the D-BIND traffic model is:

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

(2)

with the assumption of \( B_{R_T}(0) = 0 \) at \( t_0 = 0 \). Let us define one frame time to be a time unit. For a video sequence with \( M \) frames, the tightest D-BIND traffic constraint function \( B^*_{R_T} \) should be constructed by the consecutive \( P \) rate-interval pairs where \( P = M \). It is a piecewise-linear time-invariant curve. Figure 1 illustrates 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 \( B_{R_T}(t) \). The D-BIND bounding rate is actually the maximum rate allowed for the associated interval. This allows smoothing out some level of bursts on smaller time scales. The D-BIND traffic model always upper bounds the video segment in any selected interval length.

![Figure 1. Comparison of traffic constraint functions.](image)

### 2.2 RED-VBR

The framework of RED-VBR can be illustrated in Figure 2. Note that this framework can be generally applicable to other R-VBR schemes in the dynamic bandwidth allocation system as well. Three key functions are associated with individual stream control: 1) to extract the bit statistics of VBR encoded video stream from the corresponding source, 2) to calculate the traffic parameters according to the embedded traffic model, and 3) to conduct the segmentation algorithm, determining when to renegotiate and how much bandwidth to request. Meanwhile, the primary function of the link control is to perform resource allocation according to the available resources. In general, the link control module is invoked by any request from a stream control module.

In the RED-VBR, when a new source requests for admission into the network, the stream control module sends a connection request to the link control module. Depending on the current link capacity,
the bounding rates of the admitted ongoing connections and the requirement of bounding rates for the new connection, the link control module performs an admission control check to determine if this connection request can be admitted without degrading the QoS of other admitted connections. Once a source is admitted into the network, the stream control module begins to update the D-BIND bounding rates frame by frame using the currently loaded frames. At the same time, in response to a significant bit rate change, the segmentation algorithm periodically initiates a renegotiation request and immediately sends this request to the link control. The link control then conducts renegotiation control (similar to admission control) to determine either to accept or to reject the renegotiation request. If a renegotiation request is rejected, the source would continue transmitting using the previously reserved resources by applying traffic shaping. In fact, due to this dynamic renegotiation procedure, the unused bandwidth released by some existing connections can be efficiently utilized by other admitted demanding connections to achieve high statistical multiplexing gain.

The original RED-VBR considers offline and online scenarios and specifies two kinds of segmentation algorithms accordingly [2]. Offline video segmentation algorithm can be applied to pre-recorded video streams. The offline algorithm takes an advantage of knowledge of the full video trace history to calculate the optimal renegotiation schedule in advance. Online video segmentation algorithm is used for real-time live video streams. The online algorithm uses the previous trace history to estimate future resource requirements in order to generate renegotiation requests. We focus on the online real-time scenario. Assume that a D-BIND traffic descriptor of $P$ consecutive rate-interval pairs is chosen. In RED-VBR, once a video stream is admitted into the network, the network control begins to update the video stream’s D-BIND bounding rates frame by frame using the currently loaded frames. Whenever any $r_i$ of online measured bounding rate $(r_1, r_2, \ldots, r_P)_\text{measured}$ is greater than the corresponding rate in the reserved bounding rates $(r_1, r_2, \ldots, r_P)_\text{reserved}$, a renegotiation request for more resources is initiated. Two parameters $\alpha$ and $\beta$ are used in RED-VBR to calculate newly requested bounding rates. The renegotiation request includes a set of new bounding rates $(r_1, r_2, \ldots, r_P)_\text{new}$ calculated by multiplying the currently online measured rates with $\alpha (\alpha \geq 1)$. If there is sufficient network bandwidth resource available, the updated bounding rates $(r_1, r_2, \ldots, r_P)_\text{new}$ replace the previously reserved bounding rates of that session. Otherwise, the session has to proceed according to the previously reserved bounding rates $(r_1, r_2, \ldots, r_P)_\text{reserved}$.

Policing can be applied based on the reserved D-BIND traffic descriptors for individual video streams in the network.

In RED-VBR, network link utilization is computed as follows. Let $d$ denote delay bound. Assume that the maximum capacity $Q_{\text{max}}$ 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:

$$
d = \frac{1}{C} \max_{t = t_k} \left\{ \sum_{j=1}^{N} B_{R_i}(t) + Q(t) - Ct \right\}, \quad j = 1, 2, \ldots, N; \quad k = 1, 2, \ldots, P
$$

(3)

where $C$ is the capacity of the outgoing link, $N$ is the maximum number of connected video streams at delay bound $d$, $P$ is the dimension of D-BIND descriptor, and $Q(t)$ is network buffer occupancy at time $t$. Given a certain renegotiation rejection probability, link utilization $u(d)$, a function of delay bound $d$, can be defined as follows:

$$
u(d) = \frac{\sum_{j=1}^{N} R_i (1 - \text{drop} \_ \text{rate}(d))}{C},
$$

(5)

where $R_i$ is the $i^{th}$ video source’s ultimate average rate, and $\text{drop} \_ \text{rate}(d)$ is the overall average drop rate at delay bound $d$. 

Int. Trans. on Systems Science and Applications, Vol. 6, No. 1, May 2010, pp. 60-72
2.3 Overlap-Based RED-VBR

Aimed to simplify the D-BIND descriptor computation in the original RED-VBR, an approximation of RED-VBR (termed as overlap-based RED-VBR herein) was also proposed in [3]. In the overlap-based RED-VBR, the D-BIND traffic model of dimension \( M \) is replaced with an approximation D-BIND traffic model of dimension \( M \) by only using the first \( P \) D-BIND rate-interval pairs, where \( P < M \). The rest bounding rates for an interval length longer than \( P \) are derived by an overlapping approximation approach. That is, the overlap-based RED-VBR approach first builds the traffic constraint function \( B_{RT}[t_1,...,t_P] \) from \( P \) rate-interval pairs carried in a renegotiation request. Then, the traffic constraint function \( B_{RT}[t_{P+1},...,t_M] \) for an interval length longer than \( P \) is approximately derived by overlapping \( B_{RT}[t_1,...,t_P] \) several times, as demonstrated in Figure 2.

![Figure 2. An illustration of the overlapping approximation in building traffic constraint function, in which GOP size is 6 and \( P = 6 \). The bounding rates \( (r_{P+1},r_{P+2},...,r_M) \) for an interval length longer than \( P \) are computed from \( B_{RT}[t_{P+1},...,t_M] \).](image)

3. Virtual Queue Based Approach

3.1 D-BIND Dimensioning

As indicated in Section 1, a significant difficulty in RED-VBR implementation is the dimensioning of D-BIND traffic descriptor. To thoroughly investigate this problem, we conducted extensive empirical study on the RED-VBR with numerous MPEG-4 encoded VBR real-world videos [9], including both high quality encoding and low quality encoding. The video traces adopted in our simulation include several types, such as movies, cartoons, sports events and TV sequences. We attempted to empirically identify the appropriate number of D-BIND bounding rates in the RED-VBR for the MPEG-4 video traces. The method is to gradually increase the number of D-BIND bounding rates \( P \) until the D-BIND descriptor can correctly capture the burstiness of video traffic up to the scale of renegotiation intervals, and thus no (or negligible) video data drop occurs after traffic shaping even when bandwidth renegotiation requests are not satisfied. Such a smallest sufficient D-BIND dimension corresponding to the tightest D-BIND traffic constraint function, denoted as \( M_{\text{min}} \), is called the minimal D-BIND model dimension. Clearly, it is unnecessary to have \( P > M_{\text{min}} \) due to computation overhead. Table 1 lists the observed \( M_{\text{min}} \) of the several MPEG-4 video traces from our empirical
investigation. (Note that \( M_{\text{min}} \) of each video trace is also, in some degree, dependent on its multiplexed video traffic environment.) A key observation from Table 1 is that the appropriate D-BIND dimension \( M_{\text{min}} \) in the RED-VBR approach with given \((\alpha, \beta)\) is video trace dependent, and could vary significantly from one video to another video. For real-time videos, however, video traces are not known in advance, making choosing appropriate D-BIND model dimension in the original RED-VBR scheme difficult and tricky in practice. When a selected D-BIND dimension \( M < M_{\text{min}} \), the deterministic characteristics of the D-BIND traffic model would be violated and thus the traffic characteristics of video traces cannot be sufficiently captured; on the other hand, when \( M > M_{\text{min}} \), the computational complexity would be unnecessarily increased.

Table 1. Empirical observation of D-BIND \( M_{\text{min}} \) under the following experimental conditions: the delay bound being 0.2 second, \( \alpha = 1.1 \) and \( \beta = 0.9 \) in the RED-VBR, each admitted connection being policed according to the reserved bounding rates, and the simulation time being 40,000 frames.

<table>
<thead>
<tr>
<th>Quantization Quality</th>
<th>Video Trace</th>
<th>D-BIND dimension ( M_{\text{min}} ) (Frame)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Quality</td>
<td>Silence of the Lambs</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>Star Wars IV</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>Mr. Bean</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>First Contact</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>Soccer</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>ARD News</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>Lecture Room</td>
<td>24</td>
</tr>
<tr>
<td>Low Quality</td>
<td>Silence of the Lambs</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>Star Wars IV</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>Mr. Bean</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>Robin Hood</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>First Contact</td>
<td>36</td>
</tr>
</tbody>
</table>

3.2 Virtual-Queue-Based RED-VBR

To address the difficulty of D-BIND descriptor dimensioning, we propose a new virtual queue based RED-VBR scheme. Our scheme is to introduce a system wide virtual queue \( VQ \) (i.e., virtual network buffer) into RED-VBR. The capacity of virtual queue is less than the actual network buffer, that is:

\[
VQ = \lambda Q, \quad 0 < \lambda \leq 1. \tag{6}
\]

In our proposed virtual-queue-based RED-VBR, Equations 3 and 4 are then modified as follows for renegotiation and admission control conditions:

\[
d = \frac{1}{C} \max \left\{ 0, \sum_{j=1}^{N} B_{j}^{k}(t) + VQ(t) - Ct \right\}, \quad j = 1, 2, ..., N; \quad k = 1, 2, ..., P \tag{7}
\]

\[
VQ(0) = 0. \tag{8}
\]

By choosing and adjusting ratio \( \lambda \), a much smaller D-BIND window size than \( M_{\text{min}} \) can be used for each video stream in the virtual-queue-based RED-VBR to achieve comparable QoS and network utilization performance as the original RED-VBR, but with significant reduction of the computational complexity of the RED-VBR. Furthermore, the number of D-BIND bounding rates used in virtual-queue-based RED-VBR is no longer video trace dependent, but can be easily selected in advance as, for example, the size of one GOP (group of pictures) of each VBR video.

By extensive empirical study, we found that \( \lambda \) can be well estimated based on the following empirical formula:

\[
\lambda = \begin{cases} 
1, & 0 \leq d \leq d_{\text{max}} \\
-4d + 4d_{\text{max}} + 1, & d_{\text{max}} \leq d \leq 0.2, 
\end{cases} \tag{9}
\]

Int. Trans. on Systems Science and Applications, Vol. 6, No. 1, May 2010, pp. 60-72
where $d_{\text{max}}$ is the maximum delay bound when $\lambda$ begins to decrease from its maximum value of 1, and is dependent on the overall characteristics of the multiplexed video streams in some degree.

4. Predicting VBR Video Streaming Performance

In general, for any VBR video streaming under VBR network service model, the overall network performance, i.e., statistical multiplexing gain, will be affected not only by individual video traffic dynamics but also by its associated traffic context. To avoid this complexity, we consider a simplified situation of network SMG performance where a homogeneous traffic context of flows multiplexed from the same video is assumed. This simplified network performance consideration is referred to as homogeneous-SMG (H-SMG). We also introduce a concept of homogeneous utilization performance for a VBR video streaming subject to H-SMG network setting. This concept enables us to study the inherent nature of each individual video’s potential performance limit, without dealing with the complicated interactions among aggregated traffic from various heterogeneous video sources in a network environment.

Our insight is that an individual VBR video’s potential network utilization performance limit probably relies largely on the degree of the effectiveness of its traffic bursts being smoothed out within its GOPs (groups of pictures). If the bursts can be effectively smoothed out within fewer GOPs, relatively high potential network utilization may be expected given identical QoS constraints. To formulate our idea, we define a new metric, referred to as smoothing ratio (SR) of a VBR video, as follows:

$$SR = 1 - \frac{\varphi_{\text{GoP}}}{\varphi_{\text{Frame}}}$$

where $\varphi_{\text{GoP}} = \frac{\text{GoP}_{\text{Peak}}}{\text{GoP}_{\text{Mean}}}$ and $\varphi_{\text{Frame}} = \frac{\text{Frame}_{\text{Peak}}}{\text{Frame}_{\text{Mean}}}$. A larger value of $SR$ indicates less GoP bursts compared to the frame bursts. Let $S(C, Q)$ denote a set of VBR videos with video coding standard $C$ at quantization level (quality level) $Q$, and let $v_i$ denote a video stream $i$. We can then represent our idea by the following hypothesis.

**Hypothesis:** $\forall (i, j) v_i, v_j \in S(C, Q)$, if $SR(v_j) > SR(v_i)$, then $v_i$ would likely, in a statistical sense, produce higher network utilization performance than that of $v_j$ in the sense of H-SMG under any identical QoS constraints.

Based on the hypothesis, we propose the following simple heuristic method to estimate the homogeneous network utilization performance for any given VBR video stream $v_k \in S(C, Q)$.

**Method:**

1. For a given VBR video category $S(C, Q)$, select $v_i, v_j \in S(C, Q)$ which are known a priori, and $SR(v_i) > SR(v_j)$. Calculate network utilization $U(v_i)$ and $U(v_j)$ in the sense of H-SMG;
2. Based on the values of $SR(v_i)$, $SR(v_j)$, $U(v_i)$, and $U(v_j)$, predict the intrinsic performance nature of a new VBR video $v_k \in S(C, Q)$ as follows:
   - $SR(v_k) \leq SR(v_j) \Rightarrow U(v_k) \leq U(v_j)$;
   - $SR(v_k) \geq SR(v_j) \Rightarrow U(v_k) \geq U(v_j)$;
   - $SR(v_j) < SR(v_k) < SR(v_j) \Rightarrow U(v_j) < U(v_k) < U(v_j)$.

5. Simulation Results

5.1 Simulation Setup

To thoroughly evaluate the performance of our virtual queue based approach, we conducted simulations to compare the virtual queue based RED-VBR scheme against the original RED-VBR and overlap-based RED-VBR [3]. Several real-world MPEG-4 VBR video traces are used in our simulations [9]. Frame rate is 25 frames/s and the GOP 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. More details about the video traces can be found in [9]. The simulations have been carried out on 40,000 frames for each video trace (approximately 26 minutes). Multiple video streams are generated through random starting points of each individual video trace. Single and mixed video traces are used in our simulations respectively. The link capacity \( C = 45 \text{ Mb/s} \) for high quality category videos, and \( C = 11 \text{ Mb/s} \) for low quality category videos, respectively. For simplicity, admission control uses the same algorithm as the renegotiation control. Blocking probability of 1% for renegotiation is used. Also, policing (i.e., traffic shaping) is applied to each video stream. The rejected video streams for bandwidth renegotiation will continue being transmitted under their previously reserved D-BIND bounding rates. In our virtual-queue-based RED-VBR, \( P=12 \) (the GOP size), whereas in the RED-VBR and overlap-based RED-VBR, \( P=M_{\min} \). Parameters \( (\alpha=1.3, \beta=0.7) \) are used for all three schemes.

5.2 Results on Virtual-Queue-Based Scheme

The simulation results are shown in Figures 3-8 and Table 2. Compared to the original RED-VBR, the virtual-queue-based RED-VBR has achieved a comparable utilization performance, but lower average drop rate and larger renegotiation interval. A larger average renegotiation interval means less renegotiations during video transmission, which is quite desirable. On the other hand, the virtual-queue-based RED-VBR has outperformed overlap-based RED-VBR in terms of network utilization when delay bound increases (with the exception of low quality video Robin Hood), with comparable average drop rate and renegotiation interval.

We highlight that the D-BIND window size \( P \) in the virtual-queue-based RED-VBR (i.e., \( P=12 \)) is not only readily determined in advance based on the video GOP size, but also a small fraction (e.g.,

![Figure 3. Network performance of Star Wars IV (high quality).](image)
Figure 4. Network performance of *Mr. Bean* (high quality).

Figure 5. Network performance of *Silence of the Lambs* (high quality).
Figure 6. Network performance of *Silence of the Lambs* (low quality).

Figure 7. Network performance of *Robin Hood* (low quality).
Figure 8. Network performance of mixed traces (Star Wars IV, Mr. Bean, Silence of the Lambs, and Soccer) (high quality).

Table 2. Comparisons of average drop rate and renegotiation interval.

<table>
<thead>
<tr>
<th>Video Trace</th>
<th>Scheme</th>
<th>Avg. data drop rate</th>
<th>Avg. renegotiation interval (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>StarWars IV (high quality)</td>
<td>Virtual queue based RED-VBR</td>
<td>0.42%</td>
<td>2.00</td>
</tr>
<tr>
<td></td>
<td>RED-VBR</td>
<td>0.70%</td>
<td>1.67</td>
</tr>
<tr>
<td></td>
<td>Overlap-based RED-VBR</td>
<td>0.41%</td>
<td>1.98</td>
</tr>
<tr>
<td>Mr. Bean (high quality)</td>
<td>Virtual queue based RED-VBR</td>
<td>0.40%</td>
<td>2.04</td>
</tr>
<tr>
<td></td>
<td>RED-VBR</td>
<td>0.58%</td>
<td>1.74</td>
</tr>
<tr>
<td></td>
<td>Overlap-based RED-VBR</td>
<td>0.36%</td>
<td>2.02</td>
</tr>
<tr>
<td>Silence of the Lambs (high quality)</td>
<td>Virtual queue based RED-VBR</td>
<td>0.81%</td>
<td>1.85</td>
</tr>
<tr>
<td></td>
<td>RED-VBR</td>
<td>1.07%</td>
<td>1.59</td>
</tr>
<tr>
<td></td>
<td>Overlap-based RED-VBR</td>
<td>0.70%</td>
<td>1.90</td>
</tr>
<tr>
<td>Silence of the Lambs (low quality)</td>
<td>Virtual queue based RED-VBR</td>
<td>1.82%</td>
<td>1.48</td>
</tr>
<tr>
<td></td>
<td>RED-VBR</td>
<td>2.46%</td>
<td>1.33</td>
</tr>
<tr>
<td></td>
<td>Overlap-based RED-VBR</td>
<td>1.69%</td>
<td>1.51</td>
</tr>
<tr>
<td>Robin Hood (low quality)</td>
<td>Virtual queue based RED-VBR</td>
<td>0.44%</td>
<td>1.83</td>
</tr>
<tr>
<td></td>
<td>RED-VBR</td>
<td>0.93%</td>
<td>1.47</td>
</tr>
<tr>
<td></td>
<td>Overlap-based RED-VBR</td>
<td>0.46%</td>
<td>1.89</td>
</tr>
<tr>
<td>Mixed traces (Star Wars IV, Silence of the Lambs, Mr. Bean, and Soccer) (high quality)</td>
<td>Virtual queue based RED-VBR</td>
<td>0.53%</td>
<td>1.82</td>
</tr>
<tr>
<td></td>
<td>RED-VBR</td>
<td>0.72%</td>
<td>1.58</td>
</tr>
<tr>
<td></td>
<td>Overlap-based RED-VBR</td>
<td>0.48%</td>
<td>1.82</td>
</tr>
</tbody>
</table>
one-third or one-fourth in the above experiments) of the $P$ used in the original RED-VBR and overlap-based RED-VBR schemes, leading to a significant reduction of the computation complexity. We note that although overlap-based RED-VBR was initially aimed to simplify the D-BIND descriptor computation in the original RED-VBR, it still has significantly higher computation overhead than the virtual-queue-based RED-VBR, and still has the same difficulty of choosing appropriate D-BIND window size for each individual video.

5.3 Results on Predicting VBR Video Performance

Based on the above simulation results, we can easily test and validate our proposed heuristic method of predicting VBR video streaming performance. Table 3 lists the relevant frame and GoP statistics of the video traces obtained from [9], the calculated SR values, and the above network performance results at the delay of 0.2 second with all the three different schemes. We can see clearly that the achievable network utilizations of various video traces are correctly characterized and predicted by videos’ smoothing ratio $SR$ values. That is, in each quantization quality category, the video with larger SR value exhibits higher network performance in the sense of H-MSG. Two important observations are obtained. The first one is that in general, our method works better as the delay bound increases, which is in consistent with the underlying idea of the smoothing ratio. The other important observation is that our method is insensitive to which dynamic bandwidth allocation scheme (either the virtual queue based RED-VBR, or the original RED-VBR, or the overlap-based RED-VBR) used. This is significant since it indeed shows that our introduced concept of smoothing ratio $SR$ captures the inherent nature of individual video’s potential performance limit to a significant extent. Note that our method should be applied to the videos within the same category.

Table 3. Statistics and smoothing ratio SR values of VBR MPEG-4 videos.

<table>
<thead>
<tr>
<th>Quantization Quality</th>
<th>Video Trace</th>
<th>Frame Statistics Peak/Mean $\varphi_{Frame}$</th>
<th>GOP Statistics Peak/Mean $\varphi_{GoP}$</th>
<th>Smoothing Ratio SR $1 - \frac{\varphi_{GoP}}{\varphi_{Frame}}$</th>
<th>Network Utilization (at the delay of 0.2 sec.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Our scheme</td>
</tr>
<tr>
<td>High</td>
<td>StarWars IV</td>
<td>6.81</td>
<td>4.29</td>
<td>37.0%</td>
<td>57.7%</td>
</tr>
<tr>
<td></td>
<td>Mr. Bean</td>
<td>5.24</td>
<td>3.73</td>
<td>28.8%</td>
<td>55.6%</td>
</tr>
<tr>
<td></td>
<td>Silence of the Lambs</td>
<td>7.73</td>
<td>6.22</td>
<td>19.5%</td>
<td>32.3%</td>
</tr>
<tr>
<td>Low</td>
<td>Silence of the Lambs</td>
<td>21.39</td>
<td>10.48</td>
<td>51.0%</td>
<td>24.3%</td>
</tr>
<tr>
<td></td>
<td>Robin Hood</td>
<td>10.56</td>
<td>3.34</td>
<td>68.4%</td>
<td>36.7%</td>
</tr>
</tbody>
</table>

6. Conclusions

The main contributions of this work are (1) proposing a virtual queue based RED-VBR scheme to overcome the difficulty of D-BIND descriptor dimensioning in the original RED-VBR and overlap-based RED-VBR, as well as to significantly reduce the computation complexity of the previous RED-VBR schemes; (2) showing network performance improvement of our virtual-queue-based RED-VBR in general over the previous RED-VBR schemes; (3) introducing a new concept and metric –
smoothing ratio – to improve our fundamental understanding of individual VBR videos’ intrinsic nature regarding network utilization performance, and proposing a simple and effective method based on the smoothing ratio to predict the inherent network performance limit of any given VBR video in the sense of H-SMG. Rigorous simulations have been conducted with MPEG-4 real-world VBR video traffic traces.

The original RED-VBR implies largely that the D-BIND window size realizes a concrete separation of the smaller-time-scales of video traffic burstiness from the longer-time-scales for each individual video source. Our insight is that, due to the heterogeneity nature of video sources, it is usually difficult if not impossible to appropriately determine the dimensioning of D-BIND descriptors for individual real-time video sources. Therefore, the introduced virtual queue in our scheme provides a means of relaxation on D-BIND window size in an R-VBR service to attempt systematically dealing with heterogeneous real-time video sources. In light of this, the proposed virtual queue based approach can not only apply to RED-VBR as demonstrated in this work, but can also apply to other R-VBR schemes with D-BIND traffic model, such as the predictive dynamic bandwidth allocation scheme presented in [7]. Due to its simplicity and effectiveness, our proposed method of predicting individual VBR video’s network performance would be desirable in practical use, when applied to any VBR video streaming service over packet networks (e.g., CATV networks), to support better network resource management. The future work includes devising a simple online learning method to estimate \( d_{\text{max}} \) in Formula (9), and the integration of the performance predicting method with existing admission controls and dynamic bandwidth allocation schemes.

References