A Delivery System for Scalable Video Streaming Using the Scalability Extension of H.264/AVC over DiffServ Networks

Lishui Chen, Guizhong Liu
School of Electronics and Information Engineering, Xi’an Jiaotong University, 710049, Xi’an, China
chenls@mail.xjtu.edu.cn liuzg@mail.xjtu.edu.cn

Abstract

With the rapid development of multimedia applications over Internet, the requirement of high bandwidth and stringent Quality of Services is more and more urgent. In this paper, we present a delivery system for streaming video using the scalability extension of H.264/AVC over DiffServ networks. It first maps the Network Abstraction Layer Unit of the SVC from different spatial-temporal layers and different quality layers using fine grain SNR scalability (FGS) to 16 types of source marks. Then it uses the Improved Two Rate Three Color Marker (ITRTCM) at the ingress of a DiffServ domain to map these source marks to three drop precedence according to current network conditions. The core routers of DiffServ networks delivery those packets according the predefined Per-Hop Behaviors (PHBs) associated with a Differentiated Services Code Points (DSCP). Experimental results show that the proposed delivery system can provide better guarantee of QoS to the streaming applications using SVC.

Index Terms: QoS, SVC, FGS, DSCP and ITRTCM.

1. Introduction

Video stream applications have become one of the dominant types of traffic over Internet in recent years, but current IP based Internet offers only “best effort” services and never provide any Quality of Services (QoS) guarantee to those applications. Therefore, two network architectures, Integrated Service (IntServ) [1] and Differentiated Service (DiffServ) [2], have been proposed by Internet Engineering Task Force (IETF) to provide end-to-end QoS. But DiffServ is more attractive to provide scalable network QoS for its simplicity and scalability.

In DiffServ networks, legacy packet markers like Two Rate Three Color Marker (TRTCM) [3] are markers placed at the edge router of a DS domain to mark packets to different Differentiated Services Code Points (DSCP) to ensure the packets with higher priority gain better services. But they are not suitable for guaranteeing QoS of video applications. In order to use the DiffServ architecture in a better way, several researchers have proposed media-aware application-level source marking schemes [4 5], which mark video packets according to their relative importance measured in specific ways, and then those packets are mapped to certain DSCPs according to their source marks by the edge router of a DS domain using the Enhanced Token Bucket Three Color Marker (ETBTCM) [4] and the Improved Two Rate Three Color Marker (ITRTCM) [5], respectively. But those source marking scheme are not suitable for layered streaming applications, because layered video compression standards like the scalability extension of H.264/AVC [6] has their own characteristics. A method for layered MPEG video transmission over IP DiffServ has been proposed in [7], but there they only simply map base layer packets to low drop precedence, enhancement layer packets to medium drop precedence and they do not consider the condition of the network.

In this paper we propose a novel source marking scheme to mark video packets of SVC streaming applications according to their relative importance and we use the scalability extension of H.264/AVC as illustration of this method. When those packets arrive at the edge router of a DS domain, the edge router maps those source marks to DSCPs using the edge router marker ITRTCM, which marks packets according to not only their source marks but also the current network condition.

The remainder of this paper is organized as follows. We first analyze the specific of the new SVC extension of H.264/AVC, introduce the source marking scheme in details in Section 2. The simulation experiment design and experiment results are presented in Section 3. Section 4 gives a brief conclusion of this paper.
2. Delivery system for scalable video streaming

2.1 Characteristics of scalability extension of H.264/AVC

The scalable extension of H.264/AVC [6] is related to MPEG-4 AVC [6], so it is also divided into two parts, the Video Coding Layer (VCL) and the Network Abstraction Layer (NAL) [8]. There are three main scalability aspects, i.e. temporal, spatial and quality scalability in VCL [8-9]. By pyramidal representation of the spatial levels the encoder achieves spatial scalability, where the generation of a lower resolution is from a high resolution layer signal by down-sampling. In each spatial layer, temporal scalability is enabled by employing either motion-compensated temporal filtering (MCTF) or hierarchical B-pictures. The residual signal resulting from intra prediction or motion compensated inter prediction is transform coded within each spatial layer. For quality scalability, a quality base layer residual provides minimum reconstruction quality at each spatial layer. This quality base layer can be encoded into an AVC compliant stream if no inter layer prediction is applied. And quality enhancement layers are additionally encoded within each spatial layer. This is the so-called fine grain quality (SNR) scalability (FGS) coding. In NAL, data of each specific spatial-temporal resolution and quality layer is encapsulated into one NALU.

2.2 The source marking scheme

Fig. 1 shows an example generate a serial of NALUs using hierarchical B-Picture techniques with two spatial layers and four temporal levels and it is from [10]. The pictures in layer 0 are down-sampling from the pictures in layer 1 by a factor two. When coding pictures in layer 1, the motion and texture information of layer 0 are scaled and up-sampled to predict the motion and texture information of layer 1. In each spatial layer, there are a quality base layer and a FGS layer. There are two NALUs for one picture and they contain the quality base layer data and the enhancement layer data, respectively. Note that these NALUs are serialized in encoding order, but not in picture display order.

From Fig. 1, we can gain the following three important facts. First, a higher spatial layer is predicted from a lower spatial layer, if the packets from the lower spatial layers lost in their way, the packets from the higher spatial layer can never be decoded, so the lower spatial layer packets are more important. Second, within a spatial layer, due to the simple fact that the temporal levels are generated by using hierarchical B-Picture techniques, so the previously coded pictures are more important than the later coded pictures in a GoP. Third, within each spatial-temporal resolution it is undoubted that the quality base layer packets are more important than enhancement layers’ packets and the packets from the lower enhancement layers are more important than the packets from the higher enhancement layers. Additionally, in this study we use only the quality base layer of the pictures in the lower layer for inter-layer prediction, so all the quality base layer packets are more important than all the enhancement layers packets within a GoP.

Let \( S_i \) be the spatial layer, \( Q_j \) be the quality layer and \( T_k \) be the temporal level of an NALU, then \( (S_i, Q_j, T_k) \) can represent the spatial-temporal resolution and the quality degree of the NALU. And \( (S_i, Q_j, T_k) \) can also reveal the relative importance of the NALU. Take the NALUs shown in Fig. 1 for example, there \( S_i = 0, 1, Q_j = 0, 1 \) (0 stands for quality base layer and 1 stands for the FGS layer) and \( T_k = 0, 1, 2, 3 \), according to the principle above there are totally 12 \( (2 \times 2 \times 3) \) different important grades. Let \( P(S_i, Q_j, T_k) \) (ranges from 0 to 11) be the priority of the NALU of \( (S_i, Q_j, T_k) \), 0 be the most important grade...
and 12 be the least important grade. Then the mapping is:

\[
\begin{align*}
P(0,0,0) &= 0 \\
P(0,0,1) &= 1 \\
&\quad \cdots \\
P(1,1,3) &= 11
\end{align*}
\]

Note that the total number of important grades gain from this mapping scheme above is related to the specific coding structure, that is, the size of the GoP, the number of spatial layers, the number of temporal levels and the number of enhancement layers. In order to use the edge router marking algorithm ITRTCM, which was designed for mapping 16 types of source marks to three, we should map these important grades to 16 source marks.

Let \( N \) be total number of important grades obtained from the mapping scheme above, \( i \) be the important grade of a NALU and \( \text{prior}[i] \) be the source marks of the NALU need to be computed, we can obtain the 16 source marks as follows.

\[
\text{prior}[i] = \left\lfloor \frac{16}{N^i} \right\rfloor + 1, \ i = (0 - N - 1)
\]

where \( \lfloor T \rfloor \) represents the greatest integer smaller than \( T \), and \( \text{prior}[i] \) ranges from 1 to 16.

3. Simulation and experimental results

3.1 Experiment Environment

We use the Network Simulation 2 (NS2) simulator [11] to complete our experiment. Fig. 2 shows the experiment simulation network topology which is a simple network with one DS domain. In the network there are an ingress router which is responsible for measuring and marking the traffic arrived at the DS network, a core router in which the PHBs are implemented and one bottleneck whose bandwidth is set according to the required different levels of traffic load. Video sender, in which the reference software JSVM-3 encoder [12] is implemented, is sending the video streaming to the Video receiver; There are two competing traffics, A CBR traffic with a rate of 550kb/s (\( R_1 \)) to CBR receiver and an exponential distribution on-off traffic with the mean packet size of 1000 bytes, the burst time interval 100 ms, the idle time interval 50 ms and the rate of 450kB/s (\( R_2 \)) to On-off traffic receiver respectively.

The source video sequence concludes 2260 frames from BUS, CITY, CREW, FOOTBALL, FOREMAN, MOBILE and SOCCER with a GoP size of 8. When the video bit stream with average rate 4Mb/s (\( R_3 \)) is being generated, there are three spatial layers, layer 0 has QCIF resolution and layer 1 has CIF resolution, and those two layers have three temporal levels with 3.75, 7.5 and 15 Hz respectively, and layer 2 has CIF resolution and four temporal levels with 3.75, 7.5, 15 and 30Hz respectively. Each spatial-temporal resolution has a quality base layer and two additional FGS layers. Then the total number of important grades of this coding configure is 30 according to (1) and we map these important grades to 16 source marks according to (2).

In order to test the performance of our video delivery scheme more efficiently, we can change the bandwidth of the bottleneck to obtain 5 different levels of payloads (90, 100, 110,120 and 130% of the bottleneck’s capacity). The total input load (\( \text{Total}_{\text{input}} \)) of the network in this simulation is 5Mb/s (\( R_1 + R_2 + R_3 \)) and the corresponding five levels’ bandwidths of bottleneck (\( BW \)) are respectively 5.56Mb/s (\( 5/0.9 \approx 5.56 \)), 5Mb/s, 4.55Mb/s, 4.17Mb/s, and 3.85Mb/s. For each input load level, three subscription levels (\( R_{af} \)) of DiffServ AF (40, 60 and 80% of the bottleneck’s capacity) are considered. Here \( R_{af} \) stands for the rate of “green” packets committed to the network and then the CIR parameter in the ingress marker should be set up according to (3).

\[
\text{CIR} = R_{af} \cdot BW \cdot R_{af} / \text{Total}_{\text{input}}, \ (R = R_1, R_2, R_3)
\]

The other parameters, CBS, PIR and PBS, are set up using the rules proposed in [20].

\[
\text{PIR} = \alpha \cdot \text{CIR}, \ \text{CBS} = \beta \cdot \text{CIR}, \ \text{PBS} = \beta \cdot \text{PIR}
\]

where \( \alpha \) and \( \beta \) are constants and \( \alpha = 2 \) and \( \beta = 1.5 \) are used in this study.

The core router in the DS domain implements the Weighted Random Early Detection (WRED) mechanism for the active queue management and using Assured Forwarding (AF) PHB [13]. The WRED parameters, namely the minimum threshold, the maximum threshold and the drop probability, are specified respectively as 10, 15 and 0.02 for the green packets, 7, 10 and 0.10 for the yellow packets and 3, 6, and 0.20 for the red packets in this simulation.

We compared our source marking scheme using ITRTCM as edge router marker with the legacy packets marker and the method LayerMapping proposed in [12].
3.2 Throughput

Fig. 3 shows the throughput of the three different marking schemes under payload 130% with three kinds of $R_w$ (AF in figures) and they can represent the other cases. From these subfigures the conclusion that the throughputs using source marking scheme proposed in this paper with ITRTCM is higher than those of other two methods.

3.3 Loss Rate of the Network

The loss rate is the ratio of number of lost packets within a $(S, Q, T_e)$ layer to the number of this layer packets sent at the server and those loss rates of with $R_w = 60\%$ and Payload = 130\% are listed in Table 1, where M1, M2 and M3 represent the mark algorithms our scheme, TRTCM and LayerMapping, S0, S1 and S2 represent the three spatial layers respectively and T0 to T4 represent the temporal levels in each spatial layer respectively. Although the LayerMapping method can protect the lowest spatial layer packets very well, packets losses of higher spatial layers are too high, and we can never gain any benefit from the SVC. It is very clear from Table 1 that our method can better protect the more important packets and TRTCM can never suffice this purpose, and generally, the loss rates of our source marking scheme with ITRTCM is gradually higher along with packets importance gradient lower.

Table 1. Loss rate of the network ($R_w = 60\%$ and Payload = 130\%)

<table>
<thead>
<tr>
<th>Quality base layer</th>
<th>Enhancement layer 1</th>
<th>Enhancement layer 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>M2</td>
<td>M3</td>
</tr>
<tr>
<td>T0</td>
<td>2.47</td>
<td>15.54</td>
</tr>
<tr>
<td>T1</td>
<td>4.19</td>
<td>6.38</td>
</tr>
<tr>
<td>T2</td>
<td>3.66</td>
<td>6.00</td>
</tr>
<tr>
<td>S1</td>
<td>3.88</td>
<td>12.36</td>
</tr>
<tr>
<td>T0</td>
<td>4.96</td>
<td>18.79</td>
</tr>
<tr>
<td>T1</td>
<td>8.31</td>
<td>24.60</td>
</tr>
<tr>
<td>T2</td>
<td>22.61</td>
<td>36.36</td>
</tr>
<tr>
<td>S2</td>
<td>T0</td>
<td>24.46</td>
</tr>
<tr>
<td>T1</td>
<td>26.37</td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>32.54</td>
<td></td>
</tr>
</tbody>
</table>

3.4 Total Drop Rate at the Client

The loss rates of the network can only reflect the relative different protections of different marking schemes when network congestion occurs. Based on the fact that if a quality base layer packet is lost, its corresponding enhancement layer packets will be useless, if a lower enhancement layer packet is lost, its corresponding higher enhancement layer packets will be useless, we remove the useless packets at the client to gain the drop rate which is the ratio of number of lost and useless packets within a $(S, Q, T_e)$ layer to the number of this layer packets sent at the server, and here we present the drop rates with $R_w = 60\%$ and Payload = 130\% in Table 2. The drop rate can reflect the end-to-end quality of different delivery schemes in a certain extent. From the table we can see that our proposed source marking is much better than other schemes. The conclusion that our proposed delivery system can provide better guarantee of QoS to streaming application using the scalability extension of H.264/AVC over DiffServ networks can be drawn.

Table 2. Drop rate at the client ($R_w = 60\%$ and Payload = 130\%)

<table>
<thead>
<tr>
<th>Quality base layer</th>
<th>Enhancement layer 1</th>
<th>Enhancement layer 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>M2</td>
<td>M3</td>
</tr>
<tr>
<td>T0</td>
<td>2.47</td>
<td>15.54</td>
</tr>
<tr>
<td>T1</td>
<td>3.19</td>
<td>6.38</td>
</tr>
<tr>
<td>T2</td>
<td>3.00</td>
<td>5.66</td>
</tr>
<tr>
<td>S1</td>
<td>3.88</td>
<td>12.36</td>
</tr>
<tr>
<td>T0</td>
<td>4.96</td>
<td>18.79</td>
</tr>
<tr>
<td>T1</td>
<td>9.35</td>
<td>24.60</td>
</tr>
<tr>
<td>T2</td>
<td>24.46</td>
<td>36.36</td>
</tr>
<tr>
<td>S2</td>
<td>T0</td>
<td>24.46</td>
</tr>
<tr>
<td>T1</td>
<td>26.37</td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>32.54</td>
<td></td>
</tr>
</tbody>
</table>
4. Conclusions

In this paper, by taking into consideration the relative importance of NALUs from different spatial-temporal layers and different quality layers using FGS, we proposed a source marking scheme for streaming applications using the scalability extension of H.264/AVC over DiffServ network. At the edge of the DiffServ network, the marking scheme ITRTCM was used to measure the rate of the streaming traffic and marked the packets according to their source marks and at the same considering the current network condition. By comparing the simulation results with those of TRTCM and LayerMapping, a simple conclusion can be drawn that the proposed source marking scheme can well reflect the relative importance of the scalability extension of H.264/AVC packets, and with the cooperation of the edge router marking algorithm ITRTCM we can gain a better end-to-end delivery quality.

5. References