Abstract: Receiver-driven Layered Multicast (RLM) for short is a rate-adaptation protocol framework proposed for multimedia streaming by McCanne et al (ACM SIGCOMM 1996, pp. 117–130). In this paper, we propose an adaptive content-aware scaling for video RLM. In our system, receiver who detected network congestion drops layer(s) according to the picture’s “motion” which can be estimated in real time. User studies show that this can remarkably improve receiver’s satisfaction at the same or reduced bandwidth.

1. Introduction

IP Multicast is considered the easiest solution to the scale problem for real-time multimedia multipoint communication. For adapting to network heterogeneity, many studies have been done so far, including an elegant approach of Receiver-driven Layered Multicast (RLM).

RLM is a rate-adaptation protocol framework which is proposed by McCanne, Jacobson and Vetterli [2]. The basic RLM can be described as follows.

- Sender prepares a media source which is encoded into a number of layers that can be incrementally combined to provide progressive refinement.
- Each layer of the media source is then transmitted on a separate multicast group.
- Each receiver runs the next control loop:
  - On congestion, drop a layer;
  - On spare capacity, add a layer.

Since the sender is not required to do extra work on congestion, it is easier to adapt to network heterogeneity with RLM than a sender-driven protocol. Further studies or variants of the RLM have been done by many successive researches, see e.g., [3], [5], [7].

Since the basic RLM considers the utilization of a media source which is encoded into a number of layers that can be incrementally combined to provide progressive refinement, the effectiveness of RLM depends on the “contents” of the media. Recent studies try to solve this problem by adding more flexible (i.e., adaptive) controls to improve receiver’s satisfaction without increasing network traffic, see e.g., [1]. Unfortunately, most of these studies are very complicated and they concentrate on controls for abstract and given satisfaction parameters with no actual user study being done.

In this paper, we give a simple, adaptive and content-aware scaling for video RLM and show that the proposed system is effective in improving receiver’s satisfaction at the same or reduced bandwidth.

2. Approach

We follow the RLM system [5]. Videos are encoded in MPEG-1 format with three types, I, P and B, pictures (for more about the MPEG format, see e.g., [4]). A natural layer division would have three layers; each consists of a different type of pictures. This gives us a temporal scaling. For more flexibility (quality scaling), each picture is further divided into low-frequency (i.e. main) part and high-frequency (auxiliary) part. This is done by dividing DCT coefficients for each picture.

Therefore we have six layers: IL, IH, PL, PH, BL and BH, where IL (respectively IH) is the low-frequency (high-frequency) part of I picture, and so on (thus I=IL+IH, P=PL+PH and B=BL+BH). Notice that a P(L or H) picture cannot be decoded without the referred I picture, and so cannot a B(L or H) picture without the referred I or P picture(s). Thus if there is no congestion, a receiver would add layers in the sequence

\[ IL \rightarrow I \rightarrow I+PL \rightarrow I+P \rightarrow I+P+BL \rightarrow I+P+B, \]

and on congestion, drop layers in the reverse manner.

Notice that layer combinations I+PL and I+P+BL contain pictures of low-frequency only (PL and BL respectively). As will be shown in the next section, we observe that they may not be preferred by users for a certain kind of videos which contains “low motion” and the most important factor to the receiver is not temporal continuance but picture resolution. That is, users may prefer I and I+P instead of I+PL and I+P+BL.

This suggests that a better congestion control should take account of the motion in the pictures. Our strategy is to drop layer(s) until one of modes IL, I, I+P reaches the low-motion condition. More precisely, we choose I (respectively I+P) instead of I+PL (I+P+BL) if the estimated (current) motion is less than a threshold S, even if the network capacity is spare.

3. Experiment and Analysis

We estimate the average motion in a picture by

\[ V = \frac{1}{N} \sum_{i=1}^{N} v_i, \quad (1) \]

where \( N \) is the number of macroblocks and \( v_i \) is the (absolute) value of the motion vector associated with macroblock \( i \). We note that, since motion vectors are stored in the transmitted MPEG data, it is not difficult to compute \( V \) in real time.

We compute \( V \) every one second. On congestion, the mean value of five recent \( V \)s is calculated as the

*Corresponding author.
current motion. Threshold $S$ is chosen to be 10. We have tested two kinds of video sources: Soccer with high motion ($V > S$ in most time) and OHP-Presentation with low motion ($V < S$). (Actually more video sources are tested but the above two seem to be the two most representative examples.)

Experiments have been taken between two PCs which are connected by two Cisco routers as shown in Figure 1. The network bandwidth is set to 1.2Mbps while the MPEG data is encoded at bit-rate 1.3Mbps (including an extra 128Kbps layer of audio).

![Figure 1. The experiment environment.](image)

Firstly, twenty students are asked to evaluate the qualities of different layer combinations. The mean values of the evaluations are shown in Figure 2. (For I+P+B, we have adjust the network bandwidth in the evaluation tests to fit for receiving data).

![Figure 2. Picture quality evaluations.](image)

Notice that, as the amount of motion decreases, low-frequency only pictures (I+PL and I+P+BL) tend to get worse grades. Actually, Figure 2 shows that the observers prefer I (respectively I+P) instead of I+PL (I+P+BL respectively) for low-motion videos. If in addition to pictures, the sound is also important to the receivers, however, further studies need to consider more carefully on the received layers to get the best trade-off between the sound and the pictures.

### 4. Conclusion and remark

In this paper, we have shown a practical, adaptive and content-aware scaling for RLM. User studies show that it can remarkably improve receiver’s satisfaction at reduced bandwidth than a normal RLM for a certain kind of video sources.

We note that, independent of our work, Tripathi and Claypool [6] also considered a user study based on the picture’s motion. They took a quite different, sender-driven approach, which we find is less practical because it requires to adjust the bit-rate of the video at the server side on the fly (thus it cannot perform efficiently for multiple receivers).

Finally we remark a future work. In order to get fair evaluations on the picture quality, our experiments have muted the sound (but receives the sound layer). It shows that the picture quality of I (I+P respectively) is better than I+PL (I+P+BL respectively) for low-motion videos. If in addition to pictures, the sound is also important to the receivers, however, further studies need to consider more carefully on the received layers to get the best trade-off between the sound and the pictures.

### References


