Optimising Video Layers to the Available Bandwidth Based for Packet Networks

A Thakur[1], B Lines[2], P Reynolds[3], M. Knahl[4] and S. M. Furnell[5]

[1,2,4,5] Network Research Group, University of Plymouth, Plymouth, United Kingdom
[3] Orange SA, United Kingdom
email: aruna@jack.see.plymouth.ac.uk

Abstract

Layering of video stream in the Internet has been propounded as a promising way to satisfying receivers in heterogeneous environments. In this approach sender divides video stream into layers and sends them to receivers. The receivers accept and combine the layers based on their own capabilities. There are a number of layering techniques and this paper describes briefly the various video layering techniques and discusses the factors that influence the distribution of information among different layers and the decision of number of layers.

Keywords

Layered video coding, Codec.

1. Introduction

The demand of video and audio applications is being rapidly driven by the increasing usage of the Internet. This is due to the development of various multimedia applications involving live or on-demand transmissions. Among the various applications involving real time audio-video transmission is video-conferencing. Over the years, video conferencing over the Internet has become popular for personal chatting, distance education, teleconferencing, etc. because it offers a cheap and easy method of interaction among people located in diverse geographical locations.

In spite of the recent advances in video conferencing over the Internet, resulting in the availability of a number of commercially available tools, its usage has not become very popular due to a number of psychological and technical reasons. One of the technical challenges is to provide a service to users connected with various access bandwidths and terminal device processing capabilities so as to ensure intra-session fairness, by which all the users receive data at a rate appropriate to their own capabilities regardless of the capacities of other users. This basic requirement poses major problem because of the intrinsic heterogeneity and large scale of the Internet.

Over the years a number of approaches have been researched in order to solve the challenge of satisfying users having diverse capabilities. One of the approaches is to partition the video into base and enhancement layers. The base layer contains the basic information of the image that can be accepted by all terminals, and the enhancement layers provide further refinement to the basic quality of the image. At the receiving end, these layers can be combined, depending on the bandwidth limitation and the processing capabilities of the receivers to develop the program. The quality of the presented image depends upon the number of layers combined.

A number of protocols have been proposed in order to send video in a layered format. These protocols can be roughly divided into two main categories: *sender-initiated* and *receiver-initiated*. In the sender-initiated approach the sender multicasts video streams whose quality is adjusted based on the feedback information from the receivers. In the receiver-initiated approach the sender sends the signals at constant rate and in layers. Each layer uses a separate IP-multicast group address. The task of the receiver is to dynamically judge the number of layers it can receive and subscribe to one or more layers based on its capability. The layers are then joined to form the complete video. The best known receiver-initiated solution that has been proposed is the Receiver-driven Layered Multicast (RLM) (McCanne et al., 1996). Since, both the schemes have their respective advantages and disadvantages there are certain hybrid approaches that combine the good points of both.

A number of protocols have been propounded for transmitting video in layers. Most of these protocols try to solve the problems associated with the packet network. The question that still needs to be answered is how the video should be partitioned. In this paper various layering techniques have been described.

2. Basis for Layering

The layering can be done inside the codec or as a post-processing filter at the system level. In the former case the layering is done by the codec itself and is dependent upon the type of codec. While some of the codecs support layering others do not. Layered transmission is not supported by the H.261 codec, but it can be achieved by either modifying the codec or by adding a filter after codec. While H.263 does not support layering, the extended version of (H.263+ or H.263 Version 2) supports up to 15 layers. MPEG-2 supports layered representation but does not work efficiently at low bit rates because it relies on intra-frame updates to re-synchronize the decoder in the presence of errors or packet loss. In the latter case, a simple technique is used in which a filter parses through the output stream and reads the header. Depending upon the type of layering required the packets are directed to the appropriate multicast group. At the decoder, a multiplexer sequences video data from the different multicast groups and sends a stream to a decoder.

The suitable way for partitioning the video stream should be judged based on the satisfaction of the users and the requirements of the network. To judge user satisfaction, tests need to be carried out on the perceived quality of video (Thakur et al., 2001, 2004). A layering technique would be considered as network friendly if it satisfies the following criteria:

(a) The coding is simple enough to allow minimal real time processing by the sender and decoding is of low-complexity and require low-memory so that relatively unsophisticated devices can be used.
(b) The scheme is robust enough to handle wide differences in the heterogeneous environment and also handle unpredictable and dynamic network bandwidth changes in the network conditions.
(c) It is resilient enough to handle some packet losses during the transmission.

(d) It is able to handle both unicast and multicast applications.
(e) As far as possible, each layer is mutually independent. If that is not possible then the higher layer is dependent only on the lower layers.

Taking into consideration these points a good layering technique should consider the following three aspect of layering:

(a) The distribution of information among different layers. Type of layering plays an important role in the perceptual quality of video, load on network, load on sender and load on receiver (Jun-ichi et al., 1999).

(b) The number of layers the video needs to be divided into. Original work on layering (Ghanbari, 1989) proposed the two layered system comprising a base layer and an enhancement layer. However, end users of the application, who are not interested in the intricacies of the network, would like to have an "infinite" number of layers so as to get complete satisfaction based on the perceptual quality of video. The developer of the system has to find a "middle ground" between these two positions taking into consideration the fact that with an increase in the number of layers:

- It is possible to satisfy greater heterogeneity in the network but the range of bandwidth covered does not increase linearly with the number of layers.
- The perceptual quality normally increases with increase in the number of layers, but the quality does not increase linearly, it is even possible that the quality may fall with an increase in number of layers because of increased complexity and inefficiency. Coding complexity is the additional calculations that must be performed by the codec in order to partition the video into layers. There is a limit on the maximum number of layers to be sent, as beyond a certain number of layers perceptually there is hardly any improvement in the quality. This optimum number depends upon the type of codec and the partitioning method. Coding efficiency is defined as the ratio of the number of bits used by the non-layered codec to the number of bits used by the layered codec to achieve the same quality of picture.
- The number of layers into which the transmitted video can be either static or dynamic. In the receiver-initiated schemes the number of layers is static while in the sender-initiated schemes it is dynamic. For example, in the RLM, the authors considered a static 4 layer system; on the other hand, in HALM (Hybrid Adaptation Layered Multicast) (Liu et al, 2002) the sender dynamically optimises the number of layers. While the performance improves with an increase in the number of layers, 3 to 5 layers was found to be sufficient.

(c) The distribution of available bandwidth among the layers. The base layer normally contains the most important information. Ideally, the major share of the available bandwidth should be allocated to the base layer. On the other hand, if a large part of the bandwidth is given to one layer, the user does not have a strong reason to subscribe to enhancement layers.

3. Techniques for the Division of Video into Layers

Layering of video data for transmission can be achieved by any of the following four layering techniques currently available. The performance of these layering techniques depends upon the codec used. In the coding community, scalable coding is frequently used to refer to layered coding. The scalability can be achieved by scaling the frame speed (temporal scalability), frame size (spatial scalability) and frame quality (quality or SNR scalability) or a combination of these.

3.1. Temporal Layering

In this type of layering the Intra (I), Predictive (P) and Bi-directional (B) video frames are sent on different layers. For example, in developing Layered Video Multicast with Retransmission (LVMR) (Paul, S. et al, 1998) the simulation tests were carried out for transmitting video in four layers using MPEG-2, where I frames made the base layer, P frames made the first enhancement layer, and B_1, B_3, B_5 and B_7 made the second enhancement layer whilst B_2, B_4, B_6 and B_8 made the third enhancement layer. I-frames can be independently decoded, while P-frames require I-frames, and B-frames generally require both I and P-frames for decoding. During congestion, the order of preference for the dropping frames should in general be first B, then P and finally the I-frames. This is not always true since experiments have shown that dropping some particular B-frames may result in poorer performance than the dropping of some P-frames. Hence, there is a preferential order of frame drops that should be should be determined for achieving the best result.

Figure 1. An example of a frame from video sequence, showing the effect of drop in layers on the quality of video in temporal layering (Thakur, 2004).

Figure 2. An example of a frame from video sequence, showing the effect of drop in layers on the quality of video in spatial layering.

Temporal layering is easiest to implement and introduces almost no overhead because it can be done at the post-codec level by reading the header field to identify the frame type. The problem with this layering technique is that a reduction in layer numbers lead to severe degradation in the quality. If there is a loss of packets in the lower layers then some packets in the upper layer become unusable because the pictures at higher layer are based on the pictures in the lower layer. Figure 1 shows the effect of frame drops on the quality of video. In the system, the sender sends the video stream to the codec Intra-H261. The encoded stream is fed to a de-multiplexer, where the header of the RTP packets is read. On the basis of Macroblock address the de-multiplexer partitions the stream and sends it to different IP multicast addresses. At the receiver end, the multiplexer combines all the accepted layers and sends them to the decoder. Figure 1 shows the severe degradation of video quality due to "blockiness".

3.2. Spatial Layering

Spatial layering can be achieved by various ways including scaling the DCT coefficients, pyramid coding and spatial sub-band coding. Scaling of DCT coefficients can be done by various methods like:

- Layered quantization, in which 8x8 block of each image are transformed into frequency domain and the DCT coefficient magnitudes are partitioned into different layers with more significant ones sent to lower layers;
- Spectral separation, in which the video is split between the number of layers based on the spatial frequencies. Since the lower frequencies are better seen by humans as compared to the higher ones, they are sent in the lower layer while the higher frequencies are sent in higher layers.
- Spatial scaling, in which the spatial resolution of the image is increased from the lower to higher layers.

In pyramid coding the encoder first down-samples the image, compresses it using the required encoder and transmits it in the base layer. When the image is decompressed and up-sampled a much coarser copy of the original is obtained. The difference between the original and the up-sampled image is sent in the enhancement layer.

Spatial layering provides the best performance, especially when network conditions are poor, but has the highest implementation complexity. Figure 2 shows the effect of layer drops in the quality of video in spatial layering. Perceptual quality increases almost linearly with the increase in number of layers, and efficiency decreases almost linearly with the number of layers.

3.3. Data Partitioning (DP) layering

DP layering divides video into two layers, this technique is performed by allocating specific data bytes in the bit stream (i.e motion vector information and Discrete Cosine Transform (DCT) coefficients) into different layers. In MPEG-2 DP is done with the help of Priority Break Points (PBPs) which are placed in each slice header, which specifies the grouping of DCT coefficients into the layers. The layering method using these PBPs, the vectors and coefficients with higher importance can be directed to the base layer and the rest to the enhancement layers. It has been seen for MPEG-2 that the quality of picture increases through 4 layers, beyond this, there is not much increase (Jun-ichi et al., 1999).

After Temporal, DP layering is easiest to implement because it requires only a post-processor filter. The efficiency of this method is also high since only headers (namely, sequence, group of pictures, picture and slice headers) are included in all layers.

The major disadvantage of the DP layering is that even a small number of packet drops in the base layer degrades the quality significantly. It has been seen that the quality falls almost linearly with the number of packets lost in the enhancement layers and the enhancement layers are dependent upon the base layer. Therefore, the base layer should be protected at the expense of the enhancement layers.

3.4. Signal to Noise Ratio (SNR) Layering

SNR layering is done at the codec level by encoding the video using a quantizer scale in order to generate a base layer. The enhancement layer is later generated by encoding the difference between the original video and the base layer using a quantizer scale. For more than two layers, the same process is applied recursively using additional quantizer scale parameters. For SNR layering, the quality degrades linearly with the rate of packet loss in enhancement layer and drops rapidly when the data loss occurs in the base layer.

In this type of layering the efficiency is rather low because the DCT coefficients are divided among the layers, and information for every DCT coefficient is included in all the layers. The efficiency falls rapidly with an increase in the number of layers, so it is not practical to have more than 2 layers (Jun-ichi et al., 1999).

3.5. Combination layering

All the different types of layered coding schemes described above perform well according to their restricted premise, but the quality of video can be improved further by combing or exploiting the feature of multiple layering techniques. For example, this can be achieved by a trade-off between temporal and spatial resolution. Thus in the scenes involving high motion the stress is on the increased number of frames at the expense of resolution, hence more coarse frames should be sent on the lower layers and the resolution refinement should be sent on the higher layers. On the other hand in the case of scenes involving little motion the few frames having high resolution should be sent in lower layers and more frames should be sent in higher layers.

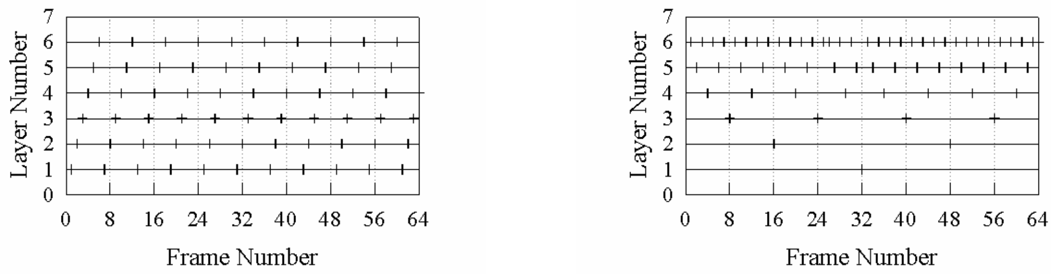
In another scheme the layering is first carried out using the spatio-temporal method. It is further separated into layers based on colour and resolution information.

4. Allocation of Bandwidth for Different Layers

Distribution of available bandwidth among different layers is an important factor determining the quality of video. It is important to provide maximum information to the base layer so that even if the enhancement layer is lost, the packets can be recreated. On the other hand, providing a major share of the bandwidth to the base layer negates the layering philosophy.

The available bandwidth can be allocated to different layers by a number of methods. It can be static or dynamically optimised. In static allocation two commonly used schemes are:

- Uniform allocation: in which the rates of all enhancement layers are equal. (Figure 3).
- Exponential allocation – in which the cumulative layer rates are exponentially spaced by a constant factor. For example, Receiver-driven Layered Multicast (RLM) divides the information so that more information is sent to higher numbered layers, N_L (No. of frames in a layer L) $\propto 2^{L-1}$ (Figure 3).



Uniform allocation **Exponential allocation**

Figure 3. Temporal scaling of video in static layer rate allocation (Thakur A 2004).

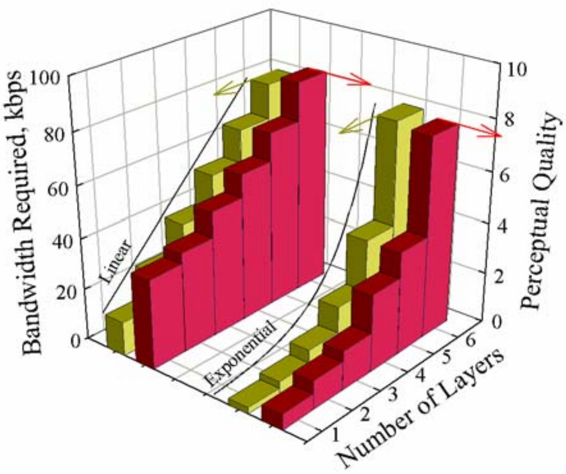
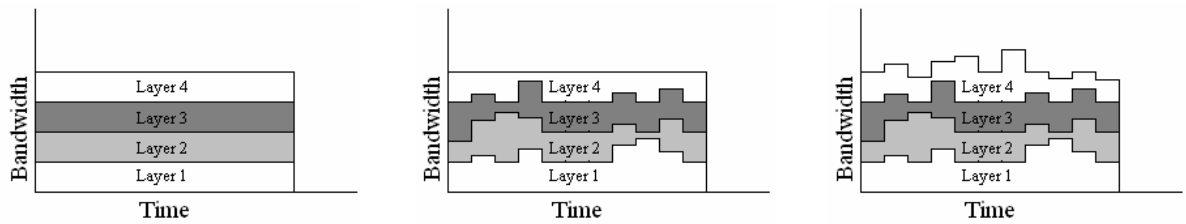


Figure 4. Effect of number of layers on the perceptual quality of video and bandwidth requirement.

As to be expected the bandwidth requirement and the perceptual quality of the video depends upon the allocation of bandwidth to different layers. Figure 4 shows the bandwidth allocation and perceptual quality in the case of Intra-H261 based temporal layering. In the case of uniform allocation the perceptual quality improves uniformly and in the case of exponential allocation the perceptual quality improves exponentially with the addition of layers.

In the sender-initiated layering, the number of layers and the capacity of each layer can vary dynamically so as to take maximum advantage of the available bandwidths. Figure 5 schematically shows the dynamic allocation of bandwidth to different layers. For example, in Source Adaptive Multi-Layered Multicast (SAMM) (Vickers et al., 1999) algorithms both the number of video layers being generated and the encoding and transmission rates of each video layer dynamically varies depending upon the feedback from the receivers.

This paper has described various layering techniques, their advantages and disadvantages. While a significant amount of work has been done in developing various layering protocols that are network friendly, relatively little work has been done in developing general principles that describe how the layering itself should be done.



Constant Total Constant Distribution | **Constant Total Variable Distribution** | **Variable Total Variable Distribution**

Figure 5. Allocation of bandwidth for different layers.

5. Future Work

It is proposed that in future the following work shall be done:

- <u>Study of parameters affecting the number of layers:</u> Detailed investigation shall be carried out to find out the effect of the number of layers on including the load on server, load on the receivers, scalability, efficiency of the codec, variations in network parameters and perceptual quality. Based on this work the optimum number of layers required to satisfy a reasonable number of receivers should be found.
- <u>Study of parameters affecting the distribution of frames among layers:</u> Research shall be carried out to study the effect of various ways of distribution of frames among different layers on the parameters like the load on the server, load on receivers, scalability, efficiency of codec, variation in network parameters and perceptual quality.

References

Ghanbari. M, (1989) “Two-layer coding of video signals for VBR networks”, *IEEE Journal on Selected Areas in Communications*, 7(5): 771-781.

Jun-ichi, K. Tobagi, F. A. Miguel, J. J, P. Emstad, P (1999), “Perceived quality and bandwidth characterization of layered MPEG-2 video encoding”, *Proc. SPIE International Symposium Voice, Video and Data Communications. Boston, M*

Liu, J. Li, B and Zhang, Y.-Q, (2002) “A Hybrid Adaptation Protocol for TCP or TCP – Friendly Layered Multicast and Its Optimal Rate Allocation”, *Proc. INFOCOM, New York*

McCanne, S. Jacobson, V. and Vetterli, M. (1996) “*Receiver-Driven Layered Multicast”. Proc. of ACM SIGCOMM, pp. 117-130.*

Paul, S. Li. X. and Ammar. M (1998) “Layered Video Multicast with Retransmission (LVMR): Evaluation of Hierarchical Rate Control”. *Proc of IEEE Infocom.*

Thakur, A. and Motycková, L. C. (2001), “Encoding of Video in Layers Based on Perceptual Quality of Video” Proceedings *of MIV'01, IEEE, Malta,, pp. 162-167.*

Thakur, A (2004), “Feed-Forward Controller for Layered Video Coding”, *MPhil thesis, Plymouth University, UK.*

Vickers, B. J. Albuquerque, C and Suda, T, (1999) “Source adaptive multi-layered multicast algorithms for real-time video distribution”, *Technical Report ICS-TR 99-45, University of California, Irvine.*