Implementing Congestion Avoidance for Network Transmission of Scalable Video Coding

Yuliang Li, Dritan Kaleshi, and Alistair Munro
Centre for Communications Research,
University of Bristol, Bristol, BS8 1UB, UK
E-mail: Yuliang.Li@bristol.ac.uk, Tel: +44-117-3315059

Abstract: In this paper we propose a congestion avoidance mechanism for the delivery of scalable video streams over IP networks. Different packets from different video layers are treated with different dropping priorities and are dropped by routers before the network experiences congestion conditions. This mechanism can largely protect higher priority traffic whilst providing bounded video quality guarantee for the video transmission. We have implemented this mechanism on a real-life test-bed that allows us to investigating the behaviour of scalable video packets transmission for various congestion levels and for different topologies with bottleneck points in the network. The experimental results confirm that dropping lower priority scalable video packets can be a feasible and effective way to address the heterogeneity of network resources.

1. Introduction

The use of scalable video for network transmission has been the focus of significant research efforts [1]. Scalable video coding (or layered video coding) divides the video bit stream into one base layer and one or more enhancement layers. Base layer has the highest priority, because the video cannot be reconstructed at the receiver if any of the base layer packets is lost. The higher the enhancement layers the lower priority they have. Enhancement layers can only be decoded in conjunction with the base layer and any lower enhancement layer(s), thus providing progressively increased video quality. This gives the possibility for receivers to adapt their reconstructed video quality to the number of video layers that they can receive depending on their bandwidth availability. The feasibility for scalable video coding transmission has been widely validated by using *N*etwork *S*imulator (NS-2) [2] [3]. In this paper we present our work for investigating real-life performance of transmitting scalable video coding in a network test-bed with congestion avoidance mechanism. The congestion avoidance can be enabled or disabled in the test-bed. When the network is congested, the test-bed without congestion avoidance drops any packets without considering their priority; the congestion avoidance test-bed monitors the bottleneck bandwidth utilisation and drops the lower priority scalable video packets before congestion happens. Our goal was to produce an experimental network system, which demonstrates the feasibility of applying congestion avoidance mechanism in routers to provide congestion avoidance and video quality guarantee for scalable video transmission. We have based our implementation on the CLICK modular router (version 1.4.3) [4], which is a flexible and modular framework for the rapid implementation of new services on software routers. The paper is organized as follows: Section 2 shows the test-bed system and describes the design and implementation details. Section 3 provides the experimental evaluation results. We conclude in section 4.

2. Test-bed Architecture and Implementation

2.1 Architecture

Figure 1 shows the topology of our scalable video coding network transmission test-bed. A scalable video packet receiver is connected with a scalable video packet generator via two servers. A background traffic generator sends standard UDP packets to the video receiver to create congestion. CLICK has been installed on the two Linux servers, acting as software routers (router-1 and router-2). In order to simplify the system and focus on the scalable video congestion avoidance, we initially have used static unicast routing tables on these two routers.

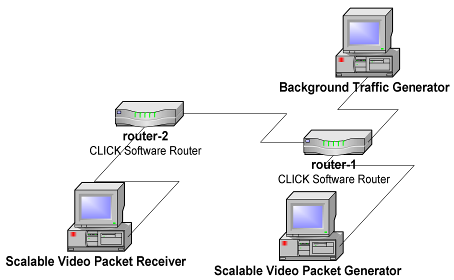


Figure 1: Test-bed Topology

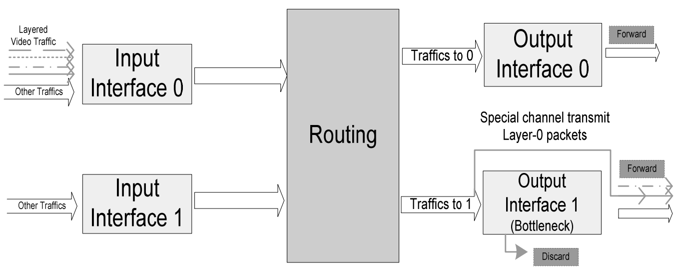
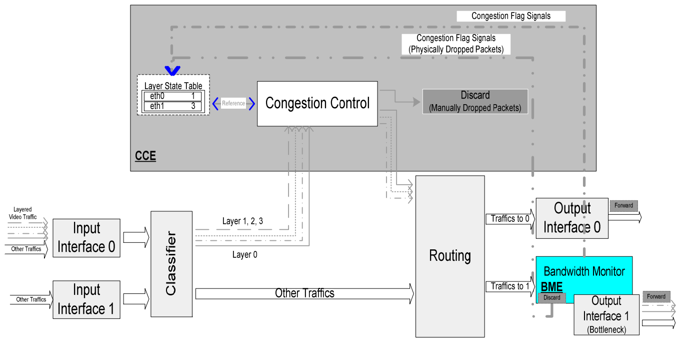


Figure 2: Normal Border Router Architecture

The scalable video generator generates the scalable video packets into network (reading them from a pre-encoded video testing file). The video receiver receives all the scalable video packets and extracts their video data payload into a file for off-line decoding and reconstruction. We define router-2 as the border router and its outgoing interface to the video receiver is the only place where congestion may happen. We set this bottleneck link bandwidth to 90KBps. We also consider the standard UDP packets having highest priority than the scalable video packets (i.e. under congestion conditions they will always go through). In order to thoroughly investigate the advantages for scalable video coding network transmission, we have set up two experimental cases by using the same network topology from Figure 1. We call them normal video transmission test-bed (normal test-bed) and congestion avoidance video transmission test-bed (congestion avoidance test-bed). The only difference between them is that they have different architectures for the border router, shown in Figure 2 and Figure 3 respectively. In Figure 2, the border router simply routes input packets to the proper output interface according to the destination IP address. When the output traffic rate exceeds the outgoing link bandwidth the router starts to drop any packets randomly without regarding to their priorities. We do not consider the influence of the Layer-0 packets in our system because (a) according to the scalable video coding technology even one lost Layer-0 packet can stop the decoding for whole video frame group, and (b) the number of Layer-0 packets is fairly small in our testing video file. Hence we use a special channel to guarantee that all the Layer-0 packets can be received at the video receiver.



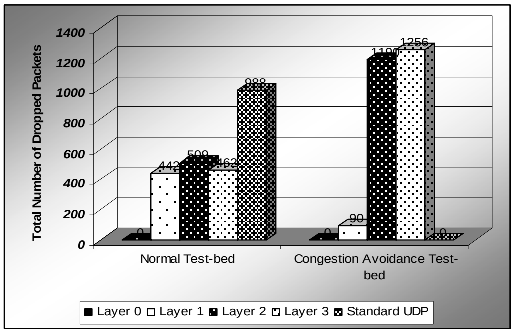


Figure 3: Congestion Avoidance Border Router Architecture   **Figure 4: Dropped Packets Comparison**

The use of a protected virtual channel to guarantee the Layer-0 packets transmission is not realistic; if congestion can be avoided then it is safe to assume that a minimum amount of link capacity will be available so that Layer-0 packets will never be physically dropped. A mechanism to avoid congestion, if well designed, can guarantee that no packets will be physically dropped, which means that the special channel used in the normal test-bed is not necessary for congestion avoidance test-bed. The internal architecture for congestion avoidance border router is shown in Figure 3. The two new elements that we have developed to support scalable video congestion avoidance are the *B*andwidth *M*onitor *E*lement (BME) and the *C*ongestion *C*ontrol *E*lement (CCE). A classifier has been used to separate the scalable video packets from the background UDP packets. All the video packets have been sent to the CCE. According to its layer state table and packet layer information, CCE may discard lower priority packets progressively. BME is used to monitor the bandwidth utilisation at the outgoing interface. It assesses the congestion level and generates different congestion flag signals to update the layer state table in the CCE. It is still possible for the outgoing rate to exceed the bottleneck link capacity, in which case packets are randomly dropped without considering their priorities. We call these randomly dropped packets *physically dropped packets* versus the packets that are discarded by the CCE, which are called *manually dropped packets*. The BME treats these physically dropped packets as an indication of congestion condition, and generates a higher degree congestion flag signals to notify the CCE that it needs to discard more packets in order to avoid any more physically dropped packets.

2.2 Congestion Avoidance Mechanism

The BME is used to monitor the outgoing interface bandwidth utilisation. The *E*xponentially *W*eighted *M*oving *A*verage (EWMA) algorithm has been used in the element to calculate the average bandwidth utilisation. According to the utilisation, the BME generates congestion flags that signal the congestion state to the CCE. It has three levels of congestion flag signals (*CC_C*, *CC_B*, and *CC_A*), which represent the different congestion degrees. When the bottleneck bandwidth utilisation is between 92% to 95%, BME assumes that there is a high probability that the bottleneck link is going to be congested. It generates *CC_C* flag signal to update the layer state table in the CCE; in this case the CCE will drop some of the lowest priority packets to try and avoid the possible congestion. When the bandwidth utilisation is between 96% to 98%, the BME predicts that the outgoing link is going to be congested very quickly. It generates *CC_B* signal to inform the CCE that it should increase the rate of packet drop. If the link utilisation goes over 98% the *CC_A* signal is sent. Under this condition the CCE can start dropping some higher priority packets.

The CCE has two input ports, which receive congestion flag signals and scalable video packets respectively. Upon receiving congestion flag signals, the CCE updates its layer state table. Upon receiving scalable video packets, the CCE decided to forward or drop the packet based on the layer state table and the packet layer information. The dropping algorithm is as follows. If the *CC_C* flag received, CCE will assign its layer state table to 1 for its correspondent outgoing interface, and it starts to drop the next 80 Layer-3 packets which are to be forwarded to that outgoing interface. If the *CC_B* flag is received, the layer state table will update to 2, and the next 60 Layer-3 and 40 Layer-2 packets will be dropped, with the layer state table reverting back to state 1 and continuing to drop Layer-3 packets. Under serious congestion conditions the *CC_A* flag will update the layer state table to 3 and the router will drop 60 Layer-3, 50 Layer-2, and 5 Layer-1 packets, then go to the state 2. This dropping policy can be shown to protect very well higher priority packets, and drop slowly and progressively the less important enhancement scalable video packets. From our experimental results it can be seen that no packets have been physically dropped; all the dropped packets are from the CCE, which is exactly our initial goal for congestion avoidance.

3. Experimental Evaluation

Our scalable video testing file includes 135 Layer-0 packets, 1277 Layer-1 packets, 1462 Layer-2 packets, and 1398 Layer-3 packets. The scalable video packet size varies between 32 to 800 bytes, with the average packet size 225-byte. The whole file has 144 frames, arranged for decoding into 9 frame groups, with 16 frames per group. Every frame group contains almost the same number of video packets; the average packet number for each layer in a frame group is 15, 142, 162, and 155 respectively. Scalable video generator generates the scalable video packets at the constant rate of 150pps. The total transmission rate for scalable video is about 33.75KBps (150pps * 225byte). The background traffic consists of 214-byte standard UDP packets. Using the same network topology from Figure 1, two scenarios have been set with the different background traffic transmission rates to investigate the network behaviour.

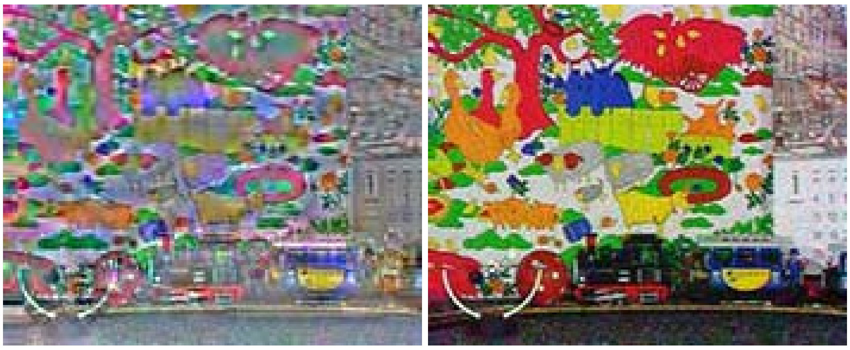


Figure 5: Video Quality Comparison

In the first scenario, the transmission rate of the background traffic is set to 85.6KBps (400pps). The total bandwidth required for transmission the video traffic and the background traffic is 119.35KBps, which is significantly greater than the 90KBps bottleneck bandwidth capacity, and the network is constantly at the congestion situation. For the normal video transmission test-bed, when congestion happens, all the packets (i.e except Layer-0 packets) will be randomly dropped at the congested outgoing interface without considering their priorities. From our experimental results (Figure 4), 442 Layer-1 packets, 509 Layer-2 packets, 462 Layer-3 packets, and 988 standard background UDP packets have been dropped. According to the scalable video coding theory, the higher enhancement scalable video packets can only progressively improve the video quality basing on the previous lower layers. As 442 Layer-1 packets have been discarded during the congestion period, most of the received Layer-2 and Layer-3 packets become useless for the video decoder. Note also that 988 standard UDP packets has been dropped, which represents a huge quality impact for the background traffic. On the contrary, congestion avoidance test-bed provides significantly (and visible) better solution to protect the higher priority packets by mainly dropping the enhanced (lower priority) scalable video packets before congestion happens. From Figure 4, 1190 Layer-2 and 1256 Layer-3 packets have been manually dropped by the CCE. Although 90 Layer-1 packets have also been dropped, note that there has no single packet loss for the most important Layer-0 and standard UDP packets (no packets have been physically dropped). Figure 5 shows the scalable video quality comparison for the received video frame (frame 125). As more Layer-1 packets have been lost from the normal test-bed, even though congestion avoidance test-bed drops almost the twice number of Layer-2 and Layer-3 packets, the video quality from normal video transmission test-bed (Figure 5-left) is visibly much worse than the congestion avoidance test-bed (Figure 5-right).

In the second scenario we are interested to investigate the dynamics of the congestion avoidance. For this we increase the background packet sending rate steadily in every 6 transmission seconds (250pps, 300pps, 350pps, 400pps and stay there until the end of the experiment). The correspondent total transmission rates for both video traffic and background traffic are 87.25KBps, 97.95KBps, 108.65KBps and 119.35KBps. The border router link capacity is still kept at 90KBps. With the increase of background traffic transmission rate, the border router outgoing link state changes from non-congested situation to fully congested situation. Figure 6 and Figure 7 show the experimental results for the number of scalable video packets which have been received in every frame group (the X axis indicates frame group number; the Y axis shows the received packet number for each layer in every frame group).

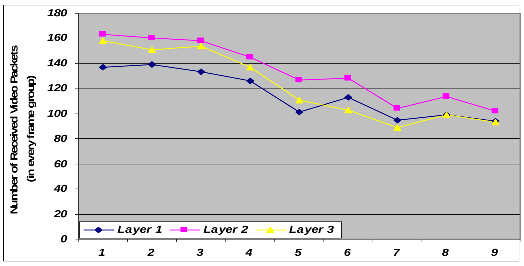


Figure 6: Normal Test-bed Evaluation

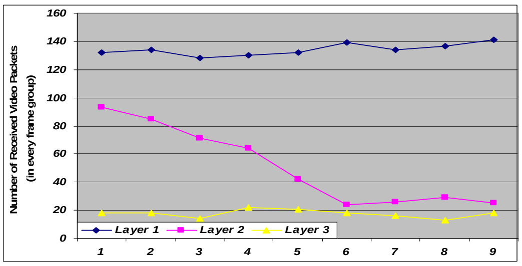


Figure 7: Congestion Avoidance Test-bed Evaluation

Figure 6 shows the experimental results for normal test-bed. It can be seen that at the beginning of the transmission, when no congestion exists, all the packets have been successfully received by the receiver. With the increase of background traffic, all received video packets decreases steadily. After the frame group 6, the number of received packets becomes stable (stationary state, as the background traffic stabilises at 400pps until the end of transmission). Figure 7 shows the congestion avoidance test-bed results. According to our congestion avoidance mechanism and the initial total traffic transmission rate from this scenario, the lower degree congestion flag signal will be sent at the very beginning to avoid the congestion, which is indicated by the packet drops for both Layer-2 and Layer-3 even at the first frame group transmission. About 90% of the lowest priority Layer-3 packets have been dropped, the percentage that is maintained until the end of the transmission. At the beginning, about half Layer-2 packets have been dropped. With the increase of background traffic transmission rate, more Layer-2 packets are dropped to release bandwidth for higher priority packets through. When the background transmission rate reaches 400pps (frame group 6 on the x-axis), the packet drop percentage becomes stable. Note through that, similarly with the first scenario, no Layer-0 and standard UDP packets (highest priority) have been dropped (i.e. no physical drops are caused by network congestion). On the contrary, normal video transmission test-bed still need a special channel to guarantee the Layer-0 packets transmission and the number for dropped background UDP packets is 678. The video quality for the normal video transmission test-bed is better than the congestion avoidance test-bed during the first two frame groups transmission; because during this period all packets have been received. Beyond that point, the quality becomes progressively worse as more Layer-1 packets are dropped. The video quality for the congestion avoidance test-bed is maintained at a good level and relative stable, independently of the variations in the congestion conditions in the network during the whole transmission period.

4. Conclusions

Recent advances in computing technology, video coding/compression technology, and high-speed networks have made it possible to provide multimedia services over the Internet. Scalable video coding is one of the best video compression technologies supporting the video transmission over heterogeneous networks with varying traffic conditions and possible congestion points. This paper describes the implementation of a network congestion avoidance mechanism that can be combined well with scalable video transmission over IP networks on a real-life network test-bed. The experimental results demonstrate that dropping lower priority scalable video packets can be a feasible and effective way to address the network heterogeneity and congestion conditions, whilst protecting critical (higher priority) traffic and providing guaranteed reconstructed video quality to the receiver.

References

[1] A. Matrawy and I. Lambadaris, "A Survey of Congestion Control Schemes for Multicast Video Applications", IEEE Communications Surveys & Tutorials 5(2), 2003.
[2] S. McCanne, V. Jacobson, and M. Vetterli, "Receiver-driven Layered Multicast", In Proceedings of ACM SIGCOMM, August 1996.
[3] K. Nakauchi, H. Morikawa, and T. Aoyama, "A Network-Supported Approach to Layered Multicast", In Proceedings of IEEE ICC 2001, June 2001.
[4] The Click Modular Router project: http://pdos.csail.mit.edu/click/#pubs