Video Conferencing Specific Considerations for RTP Congestion Control

Stephen Botzko and Mary Barnes

The lack of standards-based congestion control for videoconferencing applications frequently results in a poor experience. There are some proprietary congestion control schemes extant today. Even though some of these schemes use standard signaling, they often do not work well together. Some approaches are transmitter controlled. Systems using these approaches look at RTCP reports, and adapt their media flows accordingly. Other approaches are receiver controlled. These systems determine the lost packet rate directly from the RTP stream, and use flow control to reduce the media rate. Systems that use transmitter control generally will not work well when calling systems that use receiver control. In one direction the media flow is not controlled at all. In the other, the two systems compete for control.

Congestion control is particularly important for the videoconferencing applications, in part because traditional videoconferencing systems are severely impacted by packet loss. Modern video codecs use forward prediction, so any packet loss results in decoder divergence. This creates picture artifacts that persist until an I-frame (or IDR) is received. Since an I-frame is approximately 10 times the size of a typical P frame, frequent I-frame transmission results in a substantial reduction in video quality. As a result, effective congestion control can substantially improve overall video quality.

Layered video codecs such as SVC can help reduce these artifacts, since upper layers can be discarded without creating decoder divergence. However even when layered codecs are used, losses in the base layer have lasting effects on the video quality.

In videoconferencing, the required resolution and quality level often depends on the content being sent. A slide presentation requires that resolution be maintained, but frame rate can often be reduced. A view of a conference room usually requires a different tradeoff – it is often best to allow the resolution to drop in order to maintain motion handling.

Note that the use of multiple video streams is increasingly common. The combination of conference room and presentation video has been in use for several years, and multi-stream telepresence deployment is also growing. Many existing implementations jointly optimize the total media flow rate, rather than control individual RTP streams separately. For instance, presentation video might be prioritized above conference room video. In the case of telepresence, the application might send all conference room video streams at the same bandwidth, in order to keep the quality of video on the display wall consistent.

Of course if the available media bandwidth drops too low, then there is no value in sending the video at all – the quality or frame rate simply becomes too low to be useful for interactive communication. In such cases, multiple streams systems might reduce the number of flows. Single stream systems might turn off video altogether.

Since effective interactive communication also requires low-latency connection, the videoconferencing experience can be severely impaired by buffer bloat. Even If there is no packet loss presently on the connection, if the media flow rate saturates the connection surprisingly high latencies can result. We have seen one-way latencies in the 2-3 second range when ADSL home uplinks are saturated.

Many implementations today use proprietary congestion control algorithms. We describe one such algorithm, identifying gaps and limitations, in order to give some context to our position on application-specific considerations for video conferencing.

It is important to note that the goal of the congestion control algorithm, called Lost Packet Recovery (LPR), is application-specific - provide the best possible user experience given the limitations of the network connection. LPR is not intended to protect the network from collapse, or to allow multiple flows to share the network fairly.

LPR combines forward erasure correction (FEC) with an adaptive loss-based congestion control algorithm. When packet loss is detected, the amount of FEC repair packets needed to recover lost packets is estimated. The total media flow rate (video + FEC) is simultaneously reduced in order to hopefully eliminate the loss. This reduction is greater than the data rate lost on the connection path. In order to prevent FEC from making the congestion worse, we never allow the aggregated video+FEC media flow to exceed the negotiated call rate. The video and repair packets are combined into a single RTP stream at the transport layer.

If the FEC-protected video is free from packet loss for a period of time, then the congestion control algorithm periodically probes the connection path to estimate available bandwidth. If the probe indicates that the path can carry a higher media rate, then the bandwidth is speculatively increased. After a longer period of stable operation, the FEC protection is withdrawn.

We have found this combination of FEC and congestion avoidance to work quite well on most internet connections. However, it does have some limitations. Low latency is required for effective interactive communications. A fundamental limitation is that low-latency FEC cannot handle burst losses that exceed the latency bound.

In addition, maintaining a low latency bound also raises the bandwidth needed for repair packets. This can be partially overcome by reducing the RTP packet size, since increasing the packet rate makes FEC much more efficient. However, many consumer-grade routers are packet rate limited (not capable of running at wire-speed). If such a device is on the connection path, then increasing the packet rate can increase the loss rate - even if there is sufficient bandwidth.

We have also found that it is important to take the round trip time into account when developing a stable congestion control algorithm. If the feedback path is using a different transport (or is running at a different QOS level), then it is particularly important to measure the RTT on the actual forward RTP link to the return feedback path. A loopback message (RTP “ping” with a feedback path response) is one convenient way to make such a measurement. Monitoring round trip time also provides useful information on the amount of buffering used on the connection path.

Finally, voluntarily reducing the application bandwidth when congestion is detected does not ensure that the application will experience lower loss rates. In many cases, the competing flows simply increase their bandwidth – resulting in a video experience that is impaired by both packet loss and a lower media rate. Since the application goal is to provide the best overall user experience, it will turn off congestion control if it does not appear to help.

<u>Implications for Congestion Control for Interactive Real time Media</u>

(1) It would be useful to have interoperable congestion algorithms for videoconferencing applications.
(2) The videoconferencing application needs to have knowledge of the currently available bandwidth, so it can adjust its operating point to offer the best quality.
(3) The ideal congestion control algorithm would allow the sending application to dynamically prioritize media flows.
(4) The ideal congestion control algorithm would have knowledge of the minimum bandwidth required to support the application, and take that into account (in addition to "fairness").
(5) The ideal congestion control algorithm would also maintain the latency bound needed for interactive applications.
(6) It is useful for videoconferencing applications to be able to use FEC and RTP retransmission in conjunction with congestion control. It is reasonable for the repair packets to considered part of the media flow for congestion control purposes.
(7) In videoconferencing applications, all frame types are equally important. However, frames that are not used for prediction are less important than frames that are.
(8) Ideally the videoconferencing application would have an incentive to keep congestion control on. Perhaps maintaining a higher QOS for media flows that are controlled is one avenue.