MEASUREMENT STUDY OF PACKET LOSS VERSUS DELAY IN CONGESTION DETECTION FOR VIDEO STREAMING

M. Paredes Farrera, M. Fleury, M. Ghanbari, K. Guild

University of Essex, Colchester C04 3SQ, United Kingdom

ABSTRACT

A measurement study is conducted of video streaming across a testbed with routers typical of those found at bottlenecks on the wired Internet. During 'bursty' traffic packet loss is not always fairly distributed between background flows and a video stream. The paper shows that packet loss indications may be unreliable whereas packet-by-packet delay, an alternative metric, has the ability to closely track queue length, responding to available bandwidth in a timely manner.

Index Terms— video streaming, congestion detection, traffic measurement

1. INTRODUCTION

Video streaming over an internet is accomplished by determining the available bandwidth and adapting the video rate to that bandwidth. Rate adaptation is achieved either by changing the quantization parameter at a live-video encoder or at a transcoder for pre-stored video or by varying the rate by a scalable video technique. A video stream is often transported through UDP but controlled at the application layer with a suitable congestion controller such as TCP Friendly Rate Control (TFRC) [1]. In TFRC, every round-trip time, packet loss feedback helps determine the output rate and other TCP emulators [2] also incorporate packet loss.

Unfortunately, packet loss primarily occurring across the tight link in a network path may *not* be a reliable indicator of congestion. Under conditions of 'bursty' traffic at the time of deciding on a change in video rate, packet losses might impact less on cross-traffic at the tight link and more on the video stream under control. The consequence could be that the video stream's rate would be erroneously quenched and the received video quality would deteriorate. On the other hand, because every packet can provide a meaningful signature or footprint of the queue size in a congested link, it is possible to observe and measure the level of congestion through packet-by-packet delay. This paper reinforces the view that delay-based congestion avoidance for video streaming under UDP transport [3] may be preferable, as it is less affected by differential packet loss.

Though delay-based control is relatively common for TCP-variants, for example [4], it is a rarity for video streaming. For one-way delay measurement, synchronization of clocks may become an issue for high-speed links in excess of 1 Gbit/s, when in [4] a GPS-based clock was employed. The alternative is a relative delay measure such as Inter-Packet Delay Variation [5].

Though live experiments are not as easy to arrange as simulations and are dependent on available equipment, they indicate practical effects that may not be detected in simulations. The experiments are performed on a simple video streaming network testbed with Cisco routers through measurement of H.263+ encoded video. Congestion control depends on the ability to detection congestion at a tight link, the link with the smallest instantaneously available bandwidth on a network path at any one moment in time. Hotspots at tight links occur at the network edges, at intersections between networks and at access points. Therefore, the testbed mimics typical conditions at such a tight link.

2. METHODOLOGY

Consider a simple model of a queue with two main components the queue and the server, as illustrated in Fig. 1 for a FIFO (drop tail) queuing discipline. While RED routers have some advantages for congestion control, their deployment apparently is not widespread, while Adaptive Queue Management is not always available for simpler routers.

Q_{delay} can be calculated by the division of Ql_V and λ, provided SI units are used. As in most commercial routers the queue size is returned as the number of packets, if the mean packet length, Pl_{mean}, of the arriving video stream is known then

$$Q_{delay} = \frac{Ql_V\,(pkt) \times Pl_{mean} \times 8}{\lambda\,(bit\ /\ s)} \qquad (1)$$

Eqn. (1) captures the relation between the queue model and delay leading to congestion. In order to understand how arriving traffic influences the delay in live streaming and

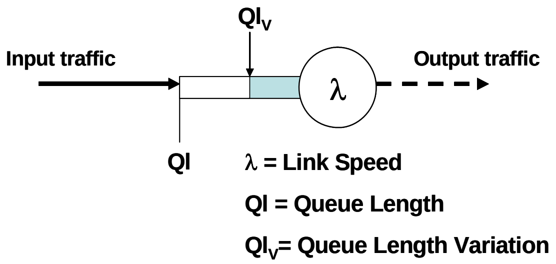

FIG. 1. QUEUE MODEL FOR A FIFO (DROP TAIL) QUEUING POLICY.

how the traffic behavior can be measured indirectly, a three-level congestion classification is introduced:

1. *No congestion*: Input traffic rate < λ
2. *Medium congestion*: Input traffic > λ and *Qlv* < *Ql.*
3. *Severe congestion*: Input traffic > λ and *Qlv* > *Ql* and, hence, packet loss.

The testbed in Fig. 2 has a 2 Mbit/s capacity serial link joining two Cisco routers, with otherwise 100 Mbit/s fast Ethernet links connecting the components. As cross-traffic arrives the available bandwidth varies, thus emulating the behavior of a tight link.

Network planners commonly recommend [6] to clients a T1 or E1 link with bandwidth of respectively 1.544 or 2.048 Mbit/s between a LAN and the border gateway (or a satellite link though with greater latency). Cisco 2600 series router [7], with operating system IOS C2600-I-M, version 12.2(13a), release fc2 were used. Cost is also a significant consideration in selection of such a router, and, hence, the same router is found at the LAN edge in network plans [8]. Transport was by UDP/IP while Point-to-Point Protocol (PPP) at the data link layer bridged the serial connection.

A variable bit-rate (VBR) H.263+ encoded video sequence represented the test video stream. Every Common Intermediate Format (CIF) frame was split into 18 macro-block row-wise slices (using H.263+'s slice mode rather than GOBs) for error resilience and then transported by allocating two slices per packet. In [9] it was established that 2-slices per packet is more efficient in these scenarios than a recommended 1-slice per packet scheme. The inter-packet gap (IPG) between sending packets was set to a constant value. The frame rate was 30 frame/s, resulting in a mean 270 packet/s. Table I summarizes the test video characteristics, which are for an "Interview" recording. This recording is a 'head and shoulders' video sequence that results in suitable data for the desired packetization lengths without causing packet fragmentation.

Cross-traffic patterns with a Normal or Pareto probability density function (pdf) were injected into the link at mean rates of 1.5, 1.8 and 1.9 Mbit/s. To determine the correlation of delay with the congestion level, these rates are intended to be representative of the three-level congestion classification. The Pareto pdf commonly represents aggregated traffic.

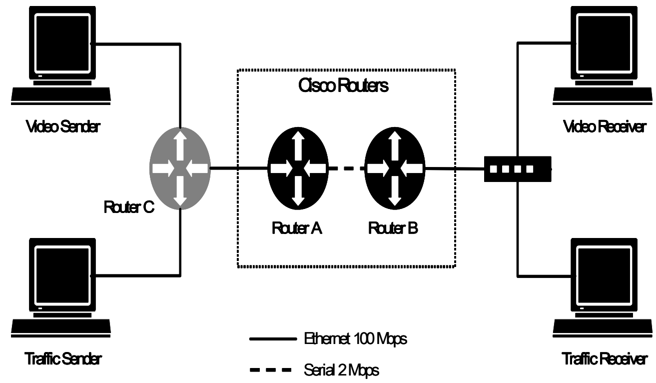

FIG.2 NETWORK TEST-BED WITH A SINGLE BOTTLENECK LINK.

Average encoded bit-rate	187 kb/s
Frame size (CIF)	352×288
Frame rate	30 f/s
Video duration	60 s
Intra refresh period	10 f

TABLE I. "INTERVIEW" ENCODED VIDEO STREAM CHARACTERISTICS

Ql	**Cross-traffic Characteristics**					
	Norm 1.5	**Norm 1.8**	**Norm 1.9**	**Par 1.5**	**Par 1.8**	**Par 1.9**
200	0	2.30	4.67	2.58	6.95	7.26
400	0	2.16	4.87	2.68	6.99	7.39

Ql is in packets, Norm = Normal pdf, Par = Pareto pdf and the cross-traffic rates are in Mbit/s.

TABLE II. PACKET LOSS FOR A TWO-SLICE PER PACKET CONSTANT IPG INPUT VIVDEO STREAM

Delay measurement was accomplished through timestamps in an RTP-like header, with clock synchronization through a Network Time Protocol stratum 2 server, distributed out-of-band over a 100 Mbit/s link. For the duration of each test and the accuracy required, this was felt to be sufficient. Propagation delay was neglected as it was too small to be relevant.

3. RESULTS

The occurrence of packet loss represents the maximum congestion level at which buffer overflow occurs at the router. Table II represents a summary of the packet loss measured as a result of video streaming in the scenarios of Section 2. An assumption might be that by increasing the cross-traffic rate the packet loss of the video stream will increase. This is the case when we observe the packet loss in the row with *Ql* = 200 packets. Because a Normal distribution is less bursty than the Pareto pdf the packet loss is less. The same tendency is observed with *Ql* = 400 packets, with the packet loss increasing as the available bandwidth from the video source's point-of-view decreases.

However, if packet loss detection is intended for the purposes of controlling congestion this may result in difficulties when the video steam transmitted represents a significant fraction of the total link traffic. In these experiments, the video stream represented around 10% of the total.

It might be thought that increasing the queue length at the input router would reduce packet loss. However, observe in Table II that this is not the case for Norm (1.9), Par (1.5), Par (1.8) and Par (1.9). The increase of the *Ql* from 200 to 400 packets did not reduce the packet loss in the video stream but increased it. This phenomenon was observed during our experiments when the background traffic rate was higher than that of the video stream. The background traffic was able to acquire a greater number of empty packet buffer slots in the router queue, while the video stream was relatively neglected.

When the background traffic is 'bursty', as occurs with the Pareto traffic, even with lower bitrates of 1.5 Mbit/s, video stream packet loss occurs. In this situation, the network link is not congested but exhibits the appearance of congestion to the video stream. Of course, a congestion rate controller based on packet losses would react as if congestion was occurring, whereas in fact the packet loss is due to the instantaneous pattern of cross-traffic packet arrivals and does not indicate the long term level of cross-traffic intensity. These effects were also observed for different values of *Ql* and Table II is simply a sample of these results.

By way of comparison, delay tracking was applied to five of the six cross-traffic scenarios recorded in Table II (with 'Norm 1.5' omitted as no congestion occurred). In the following Figures, in order to present the results, the left y-axis represents queueing delay, while the right y-axis represents the total bandwidth occupied on the serial link. The x-axis represents time during the 60 s video stream. In Figs. 3 and 4, the measured packet-by-packet packet delay of the video stream gradually rises to the maximum permitted by the given queue length. After the build up, the delay follows the bandwidth changes of the combined video- and cross-traffic. The time to reach the maximum delay varies based on the level of background traffic and the queue length: with cross-traffic of 1.8 Mbit/s, delay saturation is reached after 15 s and 20 s for respectively *Ql* = 200 and *Ql* = 400 packets. For a cross-traffic input of 1.9 Mbit/s, these times are approximately halved. Once the maximum is reached, delay closely tracks any fluctuations in the total presented input traffic.

Thus, delay measurement is a good means of knowing the level of congestion through observation of the rise time until the delay becomes stable. However, this observation can only be made if the background traffic is relatively well-behaved, for example in the unlikely event that it is Normally distributed. Of course, when delay has reached its maximum, packet loss has already begun to occur, which for encoded video, is unwelcome because of error propagation across the fragile compressed video stream. However, the rate of rise is still of interest and intelligent deductions can be made with a suitable congestion controller.

Cross-traffic is more likely to be 'bursty' within an internet, and consequently the results with Pareto background traffic are of interest. Figs. 5, 6 and 7 show the delay behavior with different levels of congestion. The delay response tracks all the peaks and lows present when the 2 Mbit/s threshold is crossed. However, there is no obvious build up in delay, as occurred when the background traffic was less bursty. Because of the background packet burst pattern, the maximum delay is rarely reached. The delay can also be much lower than that when the background traffic is well behaved.

In some cases, as in Fig. 5, the total presented bandwidth drops below the link capacity but the congestion is still tracked. However, in these graphs the bandwidth sampling period was every 1 s (imposed by the Cisco interface), while

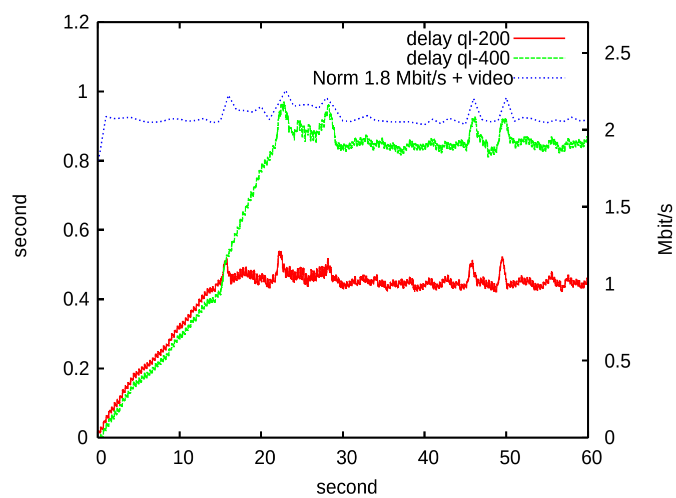

FIG. 3. DELAY AND BANDWIDTH WITH NORMAL PDF 1.8 MBIT/S CROSS-TRAFFIC

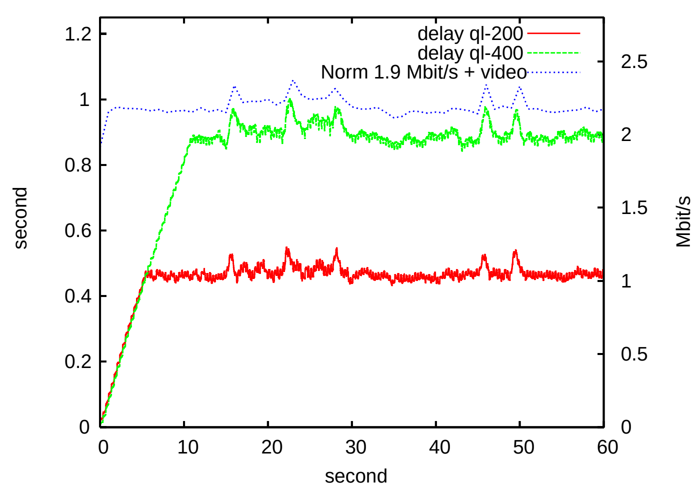

FIG. 4. DELAY AND BANDWIDTH WITH NORMAL PDF 1.9 MBIT/S CROSS-TRAFFIC

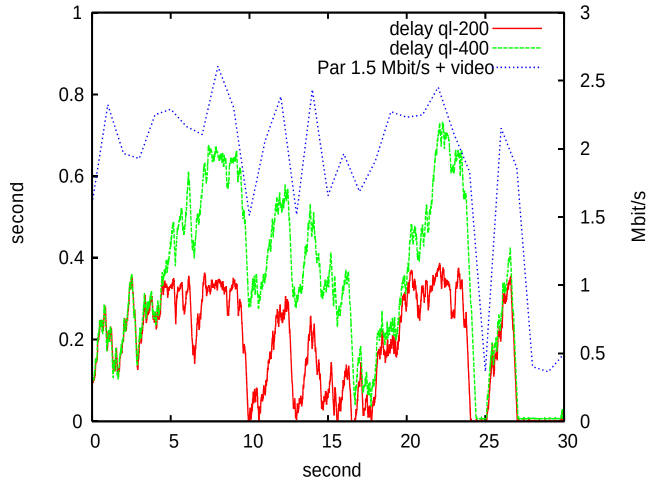

FIG. 5. DELAY AND BANDWIDTH WITH PARETO PDF 1.5 MBIT/S CROSS-TRAFFIC

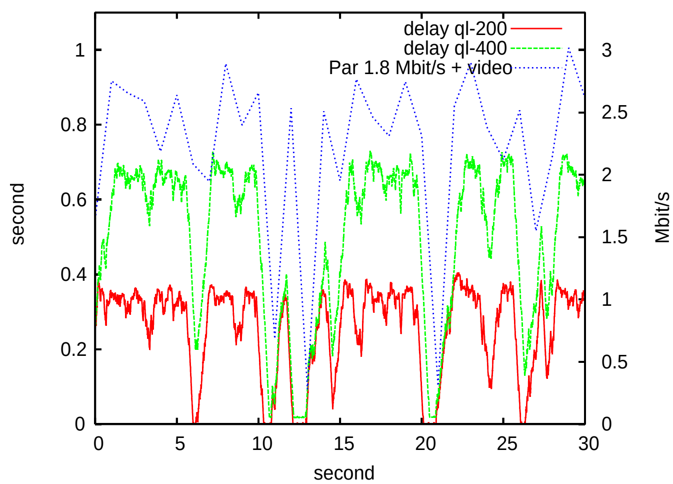

FIG. 6. DELAY AND BANDWIDTH WITH PARETO PDF 1.8 MBIT/S CROSS-TRAFFIC.

the packet-by-packet delay was sampled every time a packet was delivered, which, for the video stream, was around 0.004 s. Thus, the packet-by-packet delay actually has given a fine trace of what was occurring inside the router queue before the link. Therefore, even in these circumstances, delay serves as a reliable and responsive indicator of congestion.

5. CONCLUSIONS

From the measurement study, it is apparent that packet loss is not a precise indicator of congestion because packet loss can occur for many factors such as background traffic type (broadly 'bursty' or uniform) across the tight link on a streaming path. With suitable instrumentation it is possible to detect packet-by-packet delay, which this study indicates is a more reliable way of detecting the congestion level. Delay can indicate the queue lengths within a critical router,

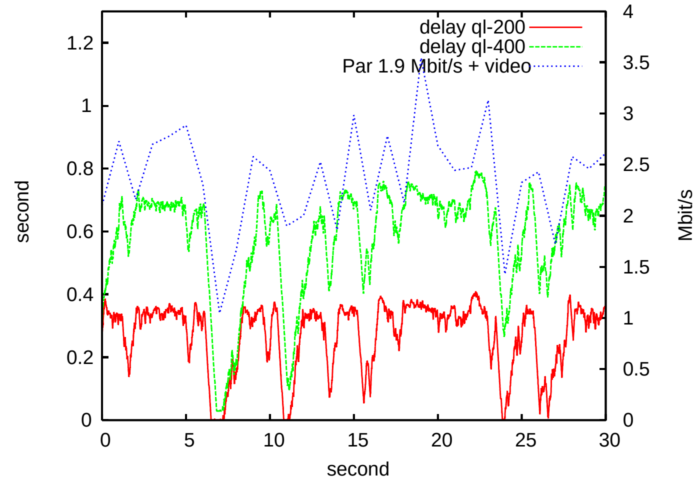

FIG. 7. DELAY AND BANDWIDTH WITH PARETO PDF 1.9 MBIT/S CROSS-TRAFFIC

though any congestion control algorithm must also try to reduce video quality fluctuations to avoid unsettling variations to the viewer. Further insight may be gained by varying the router queue discipline and explicitly passing multiple flows across the tight link.

Acknowledgment

This work was supported by the EPSRC, UK under grant nos. EP/C538692/1 and HIPNET EP/E002382/1.

References

[1] M. Handley, S. Floyd, J. Padyhe, and J. Widmer, "TCP friendly rate control (TFRC): Protocol specification", 2003, IETF RFC 3448.

[2] J. Widmer, R. Denda, and M. Mauve, "A survey on TCP-friendly congestion control", *IEEE Network*, vol. 15, no. 3, pp. 28–37, 2001.

[3] E. Jammeh, M. Fleury, and M. Ghanbari, "Delay-based congestion avoidance for video communication with fuzzy logic control", *16th Int. Packet Video Workshop*, Nov. 2007.

[4] M.C. Weigle, K. Jeffay, and F.D. Smith, "Delay-based early congestion detection and adaptation in TCP: impact on web performance", *Computer Comms.*, vol. 28, no. 8, pp. 837–850, May 2005.

[5] C. Demichelis and P. Chimento, "IP Packet Delay Variation Metric for IPPM", Draft-IETF-IPPM-IPDV-08.txt, 2001 (expired May 2002)

[6] S. Kieffer, W. Spicer, A. Schmidt, and S. Lyszyk, "Planning a data center", Technical report, Network System Architects, Inc., Denver, CO, 2003.

[7] G. Sackett, *Cisco Router Handbook*, 2nd edition, McGraw-Hill, New York, 2000.

[8] Cisco Systems, Inc., "LAN design guide for the midmarket", San Jose, CA, 2000.

[9] M. Paredes Farrera, M. Fleury, M. Mbise, and M. Ghanbari, "Best packetization scheme for H.263 internet video communication", *Electronic Letters*, vol. 40, no. 23, pp. 1476-1478, 2004.