BPB: A novel approach for obtaining network path characteristics in non-cooperative environments

Sven Hessler
Institute of Computer Science, University of Innsbruck
6020 Innsbruck, Austria
Email: sven.hessler@uibk.ac.at

***Abstract*— Due to the growth of unresponsive UDP traffic in the Internet, it becomes increasingly important for ISPs to amply shape the traffic that leaves their network. Ideally, flows should be forced to be TCP-friendly; to this end, knowledge about certain end-to-end path characteristics is needed. We present a suitable mechanism (the Burst-PiggyBack (BPB) technique) that obtains the necessary information at a device that is located close to the sender without requiring any changes at the communicating peers or in routers. The method's hypothesis is that an injected probe packet at the end of a burst is treated similar to the burst.**

I. Introduction

The increasing amount of unresponsive UDP traffic in the Internet (e.g. the growing number of VoIP flows) gives rise to concerns about the potential danger of a future congestion collapse [1]. If this trend does not change, there will be increasing pressure for ISPs to install traffic shapers which limit the magnitude of UDP aggregates that leaves their network. Then, the question of determining the maximum per-flow rate that can be allowed arises.

Ideally, an end-to-end data flow should be *TCP-friendly*, i.e. not send more than a standard TCP under comparable conditions [2]. A TCP sender updates its rate based on packet loss information and an RTT estimate, both are determined via feedback from the receiver — thus, for a traffic shaper to calculate the theoretical rate of a standard TCP under similar conditions, these network properties must be determined.

Without detailed knowledge about the application in question, it may however not be feasible to carry out strictly passive measurements, as the receiver may not always generate enough feedback that can be monitored.

In this research abstract, we present a first step towards such a traffic shaper: a novel method for measuring the RTT and estimating loss within the network (close to the sender side) without help from the receiver. Roughly, our approach is to inject ping-like packets into the network. These packets cause a reaction from the receiver without knowing about our mechanism. While TCP decisions are based on per-packet feedback, having our mechanism generate a probe packet for each payload packet would obviously be too much overhead. The existence of schemes such as TFRC [3] and TEAR [4], which were shown by their authors to attain TCP-friendly behavior with only one feedback packet per RTT (for TFRC, this is specified in RFC 3448 [5]) leads us to believe that such a small amount of signaling may be enough. Our method could be integrated in these schemes as well as RTCP-based TCP-friendly adaptation mechanisms such as LDA+ [6].

The remainder of this paper is organized as follows: Section 2 presents a survey of related research regarding the measurement of network parameters. In Section 3 we toroughly describe the BPB technique. Finally, future directions of our research are outlined.

II. Related work

The measurement of network metrics of an Internet end-to-end path is an important research area. Knowledge of performance information, such as delay, loss, and bandwidth, can be used in several ways: ISPs can keep track of and adjust their links, media servers can adapt their sending rate [7], and overlay networks can be created more robustly [8].

Previous work on estimation of network properties has focused on measurement of link capacity [9], [10], loss [11], and discovery of network properties like queuing behavior of routers [12]. Most of that research is based on the packet-pair technique, i.e. sending two packets one after the other within a very short time whereby the difference of the received packets could be used to calculate the bottleneck bandwidth. The reliability of this method is underlined by the following reasoning: if a pair of packets is sent across a network, it is highly likely that the second packet will be received provided that the first packet reached the sink. This thesis was validated experimentally [13], and theoretically for a $M/M/1/K$ queue model [14]. Our approach is based on these results; it actively injects a probe packet right after a burst of data which virtually builds a pair of packets (or a series thereof).

In addition to estimating loss, our method measures the round-trip time (RTT), which is an indicator for network utilization, and can be used to calculate the sender rate as a reference to enforce TCP-friendliness. The most common tool to measure the RTT along a network path is *ping*, which calculates the RTT from the timestamps of an ICMP Echo-request and the related ICMP Echo-response. Besides this classical method, the SYN/ACK mechanism of the TCP handshake protocol is employed by *ping*-like utilities, such as Sting [11]. While in our approach both methods can be applied to calculate the RTT, it mainly uses a combination of them – it sends a TCP or a UDP packet as request, and receives an ICMP message as response.

III. The Burst-PiggyBack (BPB) approach

The classical way to measure the RTT is to send an ICMP Echo-request packet, receive the corresponding ICMP Echo-reply packet, and calculate the difference of the timestamps. It is likely that a single ICMP packet will be treated differently in comparison to other flows on the same path if their behavior is not related to each other. We propose to append probe packets that will force the receiver to generate a response packet to bursts of non-responsive flows, e.g. by sending TCP or UDP connection requests to a blocked port. The difference between the time sending the probe-packet and the reception of the generated response packet can be used to estimate the RTT of the unresponsive flow (Figure 1), and the monitored loss events can be used to estimate the loss of the data flow, respectively. Finally, our hypothesis is, that the transmission delay and loss probability of the probe packets are similar to, and particularly not lower than, that which a packet within a burst of data will experience. If this hypothesis is correct, the loss behavior and the round-trip time of the data stream can be derived from that of the probe packets.

Obviously, a few issues must be considered when using bursts of packets instead of the packet-pair technique. The method of using packet-pairs is based on the assumption that both packets receive the same treatment in the network due to their close arrival times. But the time difference of the two departure times (inter-packet-delay) at the source is important too. Due to statistical multiplexing, accuracy in measuring the current network behavior decreases as the inter-packet-delay at the source increases. Therefore, in order to infer from one probe packet per burst the RTT and loss probability of the packets in the preceding burst, the departure times between the first packet and the last packet of the burst (referenced as burst-time) has to be short – a fraction of the RTT. We chose

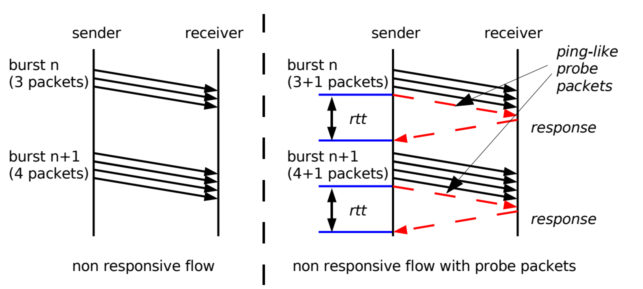

Fig. 1. Probe packets are injected after bursts (right) of a non-responsive flow (left)

to send TCP probe packets (TCP-Syn) that were addressed to a likely blocked port number instead of sending regular ICMP messages because manual pre-tests showed that some hosts which ignored ICMP echo packets responded with an ICMP error message to our TCP-Syn probes. Since we regard our work as a feasibility study, we decided that this method would suffice. Applications that use our approach, however, may use ICMP echo packers if applicable or should use more reliable ways to detect the appropriate probe packet type – for instance, a sender can try a series of different packet types (UDP, TCP, ICMP) until a response is received. If the recipient in question is a web server, a mechanism like Sting [11] can be used.

IV. Outlook on future research

The BPB-approach is based on the idea that the injected probe at the end of a burst will be treated similar to the burst itself. Despite the results obtained by simulations [15] show promising results, this assumption needs further investigation, particularly live Internet measurements. Such measurements will help us to estimate the impact and applicability of the BPB approach to real networks. Further, they will show the influence of the different answering behavior of the manifold types of networks and computers on the quality of the prediction of the BPB-technique. Last, but not least, the research on efficient burst detection techniques will be a major aspect of our future studies.

Acknowledgments

This work was partly funded by the Austrian Science Fund (FWF).

References

[1] S. Floyd and J. Kempf, "IAB Concerns Regarding Congestion Control for Voice Traffic in the Internet," RFC 3714, Mar. 2004.
[2] S. Floyd and K. Fall, "Promoting the use of end-to-end congestion control in the internet," *IEEE/ACM Transactions on Networks*, vol. 7, no. 4, pp. 458–472, 1999.
[3] S. Floyd, M. Handley, J. Padhye, and J. Widmer, "Equation-based congestion control for unicast applications," in *Proceedings of ACM SIGCOMM*, 2000.
[4] I. Rhee, V. Ozdemir, and Y. Yi, "Tear: Tcp emulation at receivers - flow control for multimedia streaming," NCSU Department of Computer Science, Tech. Rep., April 2000.
[5] M. Handley, S. Floyd, J. Padhye, and J. Widmer, "TCP Friendly Rate Control (TFRC): Protocol Specification," RFC 3448, Jan. 2003.
[6] D. Sisalem and A. Wolisz, "Lda+: A tcp-friendly adaptation scheme for multimedia communication," in *ICME*, 2000.
[7] X. Wang and H. Schulzrinne, "Comparison of adaptive internet multimedia applications," *IEICE Transactions on Communications*, vol. E82-B, no. 6, pp. 806–818, June 1999.
[8] I. Stoica, R. Morris, D. Karger, F. Kaashoek, and H. Balakrishnan, "Chord: A scalable content-addressable network," in *Proceedings of ACM SIGCOMM*, 2001.
[9] C. Dovrolis, P. Ramanathan, and D. Moore, "Packet dispersion techniques and capacity estimation," *IEEE/ACM Transactions on Networking*, December 2004.
[10] L.-J. Chen, T. Sun, D. Xu, M. Y. Sanadidi, and M. Gerla, "Access link capacity monitoring with tfrc probe," in *Proceedings of the 2nd Workshop on End-to-End Monitoring Techniques and Services (E2EMON)*, October 2004.
[11] S. Savage, "Sting: a tcp-based network measurement tool," in *USENIX Symposium on Internet Technologies and Systems*, October 1999, pp. 71–79.
[12] J. Liu and M. Crovella, "Using loss pairs to discover network properties," in *Proceedings of Internet Measurement Workshop*, 2001.
[13] V. Paxson, "End-to-end internet packet dynamics," *IEEE/ACM Transactions on Networking*, vol. 7, pp. 277–292, 1999.
[14] M. Coates and R. Nowak, "Network inference from passive unicast measurement," Rice University, Tech. Rep., 2000.
[15] S. Hessler, J.-A. Müller, and M. Welzl, "Bpb: A method for transparently obtaining network path characteristics close to the sender," in *Proceedings of IEEE 20th International Conference on Advanced Information Networking and Applications (AINA2006)*, Vienna, Austria, Apr 18-20 2006.