Rate-based Pacing for Optical Packet Switched Networks with Very Small Optical RAM

Onur Alparslan, Shin'ichi Arakawa, Masayuki Murata
Graduate School of Information Science and Technology,
Osaka University, 1-5, Yamadaoka, Suita, 560-0871,
Osaka, Japan
{a-onur, arakawa, murata}@ist.osaka-u.ac.jp

Abstract— We show that by applying rate-based pacing at the edge nodes, very small optical RAM buffers can be enough for high utilization and low packet drop ratio inside core optical packet-switched (OPS) networks.

I. INTRODUCTION

Buffering optical packets in the network is one of the difficulties of Optical packet-switched (OPS) networks when compared with electronic packet-switched (EPS). In EPS networks, contention of packets is resolved by storing the contended packets in an electronic random access memory (RAM). Electronic RAM allows sending out the packets with O(1) reading operation when the output port is free. However, converting optical packets to electrical domain in order to use electronic RAM is not a feasible solution because of the processing limitations of EPS. Processing and switching must be done in the optical domain for high-speed operation.

Many researchers consider FDL buffers to resolve contentions in optical networks, because optical RAM is infeasible or has immature technology. However, optical RAM is under research, for example Takahashi *et. al.* [1] and NICT project (phase II) [2], but the problem is that optical RAM is not expected to have a large capacity, soon. Purpose of this paper is minimizing the optical RAM buffer requirements of OPS networks.

According to a rule-of-thumb, an output link of a router needs a buffer sized at $B = RTT * BW$, where *RTT* is the average round trip time of flows and *BW* is the bandwidth of output link, in order to achieve high utilization with TCP flows. This buffer requirement is too high for high speed OPS routers with very small amount of buffering capacity. Appenzeller *et al.* [3] showed that a buffer sized at $B = RTT * BW / \sqrt{n}$, where *n* is the number of TCP flows passing through the link, is enough for achieving high utilization. However, this buffer requirement is still high for high speed OPS routers with very small amount of buffering capacity. Recently, Enachescu *et al.* [4] proposed that *O(log W)* buffers are sufficient where *W* is the maximum congestion window size of flows when TCP flows are paced and the link is under-utilized. Ref. [4] proposes pacing by using Paced TCP or using access links much slower than OPS core links. Replacing TCP senders of clients with paced versions can be hard. Also using slow access links is not a preferred solution when there are applications that require high-bandwidth on the network. Therefore, it may be better to design a general architecture for OPS network that can achieve high utilization in a small buffered OPS network independent of the number of TCP or UDP flows, and does not require limiting the speed of access links, and does not require replacing sender or receiver TCP, UDP agents of computers using the network. Applying pacing to the input traffic at the edge nodes of an OPS network can be a good choice for achieving these goals. Also, even if TCP pacing is applied at the clients, the aggregated traffic arriving to the OPS network may end up behaving bursty due to a misbehaving router etc. Therefore, pacing at the edge of OPS network can be more effective on minimizing burstiness of traffic entering the OPS domain.

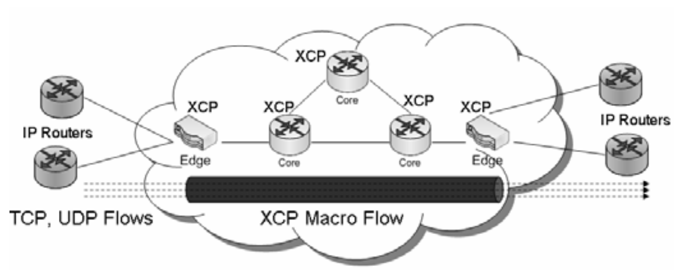

Figure 1. Architecture

In [5], we introduced an all-optical OPS network architecture that can achieve high utilization and low packet drop ratio by using FDL-based small buffering. In this paper, we show the buffer requirements and packet drop ratio by using simulations, when the same architecture is applied to optical RAM based buffered OPS networks.

II. ARCHITECTURE

In this architecture, we consider an OPS domain where packets enter and exit the OPS domain through edge nodes. We propose using a XCP-based [6] intra-domain congestion control protocol, similar to TeXCP [7], designed for WDM OPS networks for achieving high utilization and low packet drop ratio with small buffers. We show that XCP can be used for controlling and limiting the utilization level of each wavelength. Wavelength capacity must be explicitly given to XCP algorithm. If we give a false virtual capacity value less than actual wavelength capacity, XCP algorithm converges to the given virtual capacity and causes underutilization. Giving a target wavelength utilization less than actual wavelength capacity in XCP control algorithm can prevent queue buildups and allow operating at a utilization level that can provide a low packet drop ratio for QoS and traffic engineering considerations. If there is traffic between an edge source-destination node pair, a rate-based XCP macro flow is created, and incoming TCP and UDP packets of this edge pair are assigned the XCP macro flow as shown in Fig. 1. The edge nodes apply leaky-bucket pacing to the macro flows by using the rate information provided by XCP for minimizing the burstiness. XCP feedbacks of OPS edge nodes are carried in separate probe packets that XCP sender agents send only once in every control period. We are separating the control channel and data channels. Probe packets are carried on a separate single control wavelength that is slow enough for carrying only probe packets. Low transmission rate of control wavelength allows applying electronic conversion for updating the probe feedback and buffering the probe packets in electronic RAM in case of a contention. Variable sized IP packets enter OPS network without any assembling.

III. EVALUATION

NSFNET network is simulated by using a simulator developed on ns-2 [8]. There are a total of 28 nodes (14 core nodes+14 edge nodes) and 35 links (21 core links+14 edge links). All links (including edge and core links) apply optical packet switching. Edge nodes apply electronic buffering, but core routers use only optical RAM for buffering. All switches employ output buffering and cut-through bit-synchronized switching and buffering. It is assumed that there is a backlogged traffic at edge buffers, so each edge node sends traffic to all other edge nodes at the maximum possible rate controlled by XCP. The capacity of the data wavelength is set to 1Gbps. XCP's α, β, and λ parameters are selected as 0.2, 0.056 and 0.1, respectively. XCP control period of core routers and probe packet sending interval of edge routers is 50ms. 60% of the packets are 40Bytes and 40% of the packets are 1500Bytes. MSS is 1500Bytes. Fig. 3 shows the ratio of the packets dropped in NSFNET topology vs. the optical RAM size limit of output links of core routers by using our architecture and setting the target wavelength utilization of XCP to 30% and 90%. It can be seen that low packet drop ratio can be achieved with very small optical RAM buffering. If we want very low packet drop ratio like 10^{-6}, only around 4MSS (6Kbytes) and 7MSS (10.5Kbytes) optical RAM per link may be enough for 30% and 90% utilization, respectively. On the other hand, around 0.1% packet drop ratio may be enough for internet traffic. In this case, only around 2MSS (3Kbytes) and 4MSS (6Kbytes) optical RAM per link may be enough for 30% and 90% utilization in NSFNET, respectively.

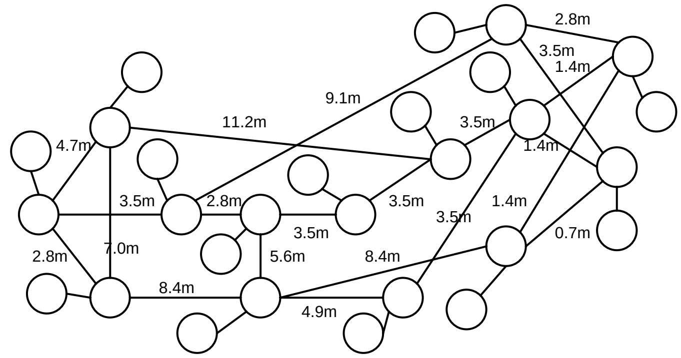

Figure 2. NSFNET topology

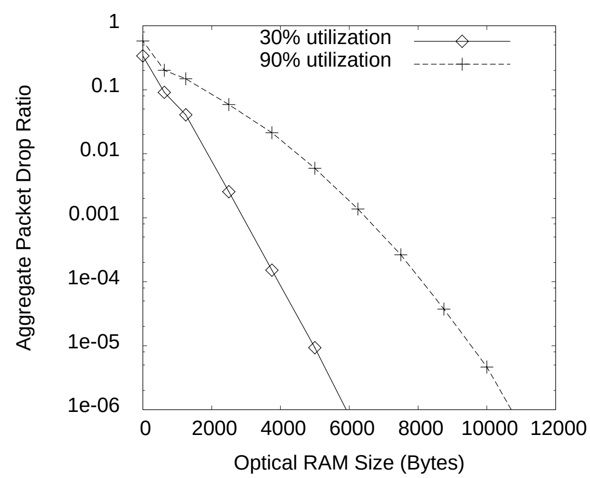

Figure 3. Aggregate packet drop ratio as a function of optical RAM size

IV. CONCLUSIONS

In this paper, we evaluated the packet drop ratio of a mesh OPS network using small optical RAM buffering with XCP based intra-domain congestion control and edge pacing. Preliminary simulation results in this paper show that only a few packet buffers per output may be enough for low packet drop ratio and high utilization. As a future work, we will evaluate the performance and buffer requirements of other switch architectures like input buffering, CIOQ. Also more realistic input traffic will be applied for showing the performance of the architecture with TCP flows and electronic buffer size requirements of edge nodes for pacing.

V. ACKNOWLEDGEMENT

This work is supported by the National Institute of Information and Communications Technology (NICT). The work of O. Alparslan was supported by Ministry of Education, Culture, Sports, Science and Technology, Japan.

REFERENCES

[1] R. Takahashi *et. al*., "Photonic random access memory for 40-Gb/s 16-b burst optical packets," *IEEE Photonics Technology Letters*, vol. 16, pp. 1185–1187, Apr. 2004.

[2] T. Aoyama, “New Generation Network(NWGN) Beyond NGN in Japan,” http://akari-project.nict.go.jp/document/INFOCOM2007.pdf.

[3] G. Appenzeller, N. McKeown, J. Sommers, and P. Barford, “Recent results on sizing router buffers,” in *Proceedings of the Network Systems Design Conference*, Oct. 2004.

[4] M. Enachescu, Y. Ganjali, A. Goel, N. McKeown, and T. Roughgarden, “Part III: Routers with very small buffers,” *ACM/SIGCOMM Computer Communication Review*, vol. 35, pp. 83–90, Jul. 2005.

[5] O. Alparslan, S. Arakawa, and M. Murata, “Optical rate-based paced XCP for small buffered optical packet switching networks,” in *Proceedings of PFLDnet*, pp. 117–124, Feb. 2006.

[6] D. Katabi, M. Handley, and C. Rohrs, “Internet congestion control for future high bandwidth-delay product environments,” in *Proceedings of ACM SIGCOMM*, Aug. 2002.

[7] S. Kandula, D. Katabi, B. Davie, and A. Charny, “Walking the tightrope: Responsive yet stable traffic engineering,” in *Proceedings of ACM SIGCOMM 2005*, Aug. 2005.

[8] S. McCanne and S. Floyd, “ns network simulator,” Web page:http://www.isi.edu/nsnam/ns/, Jul. 2002.