Copyright Notice

© 2007 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

This material is presented to ensure timely dissemination of scholarly and technical work. Copyright and all rights therein are retained by authors or by other copyright holders. All persons copying this information are expected to adhere to the terms and constraints invoked by each author's copyright. In most cases, these works may not be reposted without the explicit permission of the copyright holder.

Institute of Communication Networks and Computer Engineering
University of Stuttgart
Pfaffenwaldring 47, D-70569 Stuttgart, Germany
Phone: ++49-711-685-68026, Fax: ++49-711-685-67983
Email: mail@ikr.uni-stuttgart.de, http://www.ikr.uni-stuttgart.de

A Multi-layer Analysis on Reordering in Optical Burst Switched Networks

***Abstract*—In this paper we introduce a generic burst reordering model and evaluate analytically reordering characteristic on both burst and packet layer in an OBS network. We apply the packet reordering definition and reordering metrics newly specified by IETF IPPM WG for precise characterization of burst and packet reordering. The results of the analysis can be used directly for reordering buffer dimensioning and estimation of TCP throughput.**

***Index Terms*—OBS, Multi-layer, TCP, Reordering**

I. Introduction

OBS contention resolution schemes of buffering and deflection routing introduce variable delay, that may cause disorder delivery of bursts. Since each data burst is aggregated from multiple packets, burst disorder delivery also implies special packet disorder pattern, which has an impact on the higher layer protocol performance.

TCP is the dominant transport layer protocol in the Internet. The basic TCP congestion control [1] suffers from the packet reordering as it may interpret out-of-sequence packets as lost. If the duplicate acknowledgment (dup-ack) threshold at the sender is exceeded, the fast retransmit algorithm is triggered. The missing packet is retransmitted and additionally the sender halves its congestion window, which decreases TCP throughput. Thus, it is important to analyze the TCP over OBS performance.

The impact of OBS on TCP, mostly with respect to the OBS burst loss property, has been studied extensively in literature. Most studies, e. g., [2], [3] investigated an integrated TCP over OBS scenario by simulations without looking into the characteristic of the intermediate layers. Therefore, it was hard to identify the direct quantitative relationship between OBS network parameters and TCP throughput.

Our work analyzes reordering phenomenon on both OBS burst layer and IP packet layer, aiming to quantify the reordering metrics closely associated with the network performance. To the best of our knowledge, this is the first multi-layer analysis for the burst/packet reordering in OBS networks. We characterize the burst/packet reordering caused by burst deflecting and burst buffering by applying the IETF reordering metrics. The results can be directly applied to derive the performance measures, typically for reordering buffer dimensioning and TCP throughput estimation. Burst losses are not considered in our model.

Section II introduces the IETF reordering metrics and their significance for the network performance. In section III we present our generic burst reordering model and describe the reordering scenario. The reordering characteristic of OBS bursts and encapsulated IP packets is derived in Section IV. The paper is summarized in section V.

II. Reordering Metrics

This section reviews the IP packet reordering definition and metrics of the IETF WG IPPM [4]. They hold for generic packet-switched networks like OBS networks.

Reordering Definition: At the source node each packet is assigned a unique, monotonically increasing sequence number. At the destination node the system state is characterized by a three tuple (i, $s[i]$, $s'[i]$) for each packet arrival. Index i numbers the packet arriving order at the destination. The sequence number of this packet is denoted by $s[i]$. $s'[i]$ denotes the expected sequence number for the packet i. Its value is determined from the previously received packet. An arriving packet i with sequence number $s[i]$ is reordered, if $s[i] < s'[i]$. In this case the value of $s'[i]$ remains unchanged. Otherwise, if $s[i] \geq s'[i]$ the packet is in order and $s'[i] = s[i] + 1$. The first received packet is regarded as in-order.

Reordering Ratio: It is the ratio of the number of reordered packets to the total number of received packets.

Reordering Extent: This metric quantifies the minimal buffer size needed to restore packet order at the destination. It equals to the distance of a reordered packet to its original in-order position in the sequence of packet arrivals. Formally, the extent e_i for a reordered packet i is $e_i = i - \min_{j<i}\{j : s[j] > s[i]\}$

TCP-relevant Metric: The TCP-relevant metric quantifies the violation of the TCP dup-ack threshold. An n_r-reordered packet triggers n_r dup-acks. Packet i is n_r-reordered if $s[j] > s[i]\ \forall j \in \{t : i - n_r \leq t < i\}$.

III. Burst reordering model

We assume that only the payload packets of one TCP flow are affected by reordering and the acknowledgment packets of the same flow arrive in order on the return path.

When a burst is delayed by a FDL or deflected onto an alternative route, the additional delay is generally predictable (i.e., the constant FDL length or the additional propagation delay of the alternative route). Therefore, a generic model can be build that consists of two nodes: source and destination node. They are interconnected by $1+m$ parallel abstract links. One link represents the case that the burst receives no extra delay due to the mentioned contention resolution schemes. On the contrary, the remaining m links represent the cases that an extra e2e delay of a certain value is introduced. Each burst sent from the source to the destination independently chooses one of the $1 + m$ links with a certain probability.

For a realistic network, m is large and the analysis is complex. Aiming to provide a first insight into the reordering characteristic on the burst/packet layers, in this paper a simplified but insightful scenario is observed. Only two links

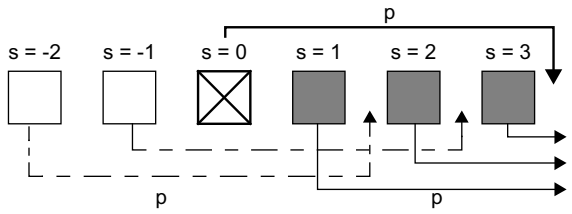

Fig. 1. Principle burst reordering for $e_B = 3$

A and B are considered, i. e. $m = 1$. While on Link A there is no additional delay, Link B adds a constant delay Δ to each burst. The probability that Link B is followed (i. e. deflected or buffered) is p. Otherwise, the burst follows Link A with probability $1 - p$.

Fig. 1 depicts the scenario that the bursts are sequentially sent from the source with constant inter-departure time T_B. In total we distinguish three kinds of bursts:

1) the test burst for which the reordering metrics is evaluated. Without loss of generality, its sequence number is set to $s = 0$ and is used as a reference number in the following presentation.
2) bursts with larger sequence numbers than 0 (gray).
3) bursts with smaller sequence number than 0 (white).

The arrow line indicates the relative change of the position in the burst series at the destination if the burst follows Link B. The distance of shift equals to $e_B = \lfloor \Delta / T_B \rfloor$ bursts.

It is assumed that each optical burst carries n_P packets of the same TCP flow. The probability that a packet becomes jth packet ($1 \le j \le n_P$) of the burst is uniformly distributed. The total number of packets per burst is also constant corresponding to a size-based burst assembly scheme.

IV. Reordering Analysis

In this section we calculate the reordering characteristic of bursts and packets at the destination. We consider the scenario in Fig. 1 to derive the reordering probability, probability of the reordering extent and the probability to be n_r-reordered.

A. Reordering probability

Let P denote the burst reordering probability and let P^* denote the packet reordering probability. The reordering probability is equivalent to the reordering ratio metric because each burst selects the link independently.

Burst reordering: According to the reordering definition, the test burst is reordered if the current expected sequence number s' is larger than 0. This occurs if and only if a burst with sequence number $s > 0$ arrives earlier than the test burst. In our scenario, this means that the test burst follows Link B and at least one burst whose sequence number $s : 1 \le s \le e_B$ follows Link A (Fig. 1, e. g., $e_B = 3$). This leads to $\mathrm{P} = p\,(1 - p^{e_B})$. From this the maximum reordering probability can be derived, which explains well the simulation results reported in [5].

Packet reordering: It is straightforward to see that a packet is reordered if and only if the burst containing the packet is reordered. Therefore, $\mathrm{P}^* = \mathrm{P}$.

B. Reordering extent probability

Let E and E^* denote the reordering extent of the test burst and an arbitrary packet in the burst, respectively.

Burst reordering extent: Note that the test burst receives a certain extent only if it follows Link B. There are two steps in determining the reordering extent of this test burst:

1) locate the burst with the smallest sequence number $s : 1 \le s \le e_B$ that follows Link A.
2) count the number of packet arrivals between this located burst (included) and the test burst at the destination.

Let I denote the sequence number of the burst found according to 1), where $1 \le I \le e_B$. According to 2), the reordering extent is at least 1 referring to the burst $s = I$. So, E can be expressed as $E = 1 + K + L$. K is the number of bursts having sequence number $s : I < s \le e_B$ (gray) and following Link A. So, $0 \le K \le e_B - I + 1$. L is the number of bursts having sequence number $s : I - e_B \le s < 0$ (white) and following Link B. So, $0 \le L \le e_B - I$.

The maximum of K and L leads to $E \le 1 + 2\,(e_B - I)$. Equivalently, $I \le \lfloor e_B + (1 - E)/2 \rfloor$ since I is of integer values. It means that I cannot exceed $\iota_{max} = \lfloor e_B + (1 - E)/2 \rfloor$ so that the value of E is still possible. By the law of total probability, the probability of the reordering extent E of the test burst can be calculated from (1), (2) and (3).

$$\mathrm{P}(E = e) = \sum_{\iota=1}^{\iota_{max}} \mathrm{P}(I = \iota)\,\mathrm{P}(E = e |\, I = \iota) \tag{1}$$

$$\mathrm{P}(I = \iota) = p^{\iota}\,(1 - p) \tag{2}$$

Note that $\mathrm{P}(I = \iota)$ means that the burst of $s = \iota$ follows Link A and the test burst as well as all bursts of $s : 1 \le s \le \iota - 1$ follow Link B, which leads to the geometric distribution of (2).

In the derivation of $\mathrm{P}(K = k |\, I = \iota)$ in (3), note that K and L are independent random variables but must fulfill $E = 1 + K + L$. That requires that K is at least $k_{min} = \max(0, E - 1 + I - e_B)$, while the maximum of K is $k_{max} = \min(E - 1, e_B - I)$. For any value of $k : k_{min} \le k \le k_{max}$, the k bursts can be an arbitrary subset of the bursts with $s : \iota < s \le e_B$, which leads to the binomial distribution.

Packet reordering extent: If a burst is reordered with an extent of E, the reordering extent of a packet of this burst can be calculated as $E^* = n_P\,E + V - 1$, where V denotes that this packet is the Vth packet of the observed flow in the burst and $1 \le V \le n_P$. It can be seen that with a given value of E^* above equation leads to a single solution of $E = \lfloor E^*/n_P \rfloor$ and $V = E^* - n_P\,\lfloor E^*/n_P \rfloor + 1$. Since E and V are independent of each other and V is uniformly distributed, it can be obtained:

$$\begin{aligned}\mathrm{P}(E^* = e) &= \mathrm{P}(E = \lfloor e/n_P \rfloor)\ \mathrm{P}(V = e - n_P\ \lfloor e/n_P \rfloor + 1) \\ &= 1/n_P\ \mathrm{P}(E = \lfloor e/n_P \rfloor)\end{aligned}$$

Combined with (1), this can be used to dimension the TCP receiver buffer size to reconstruct the packet order.

C. n_r-reordering probability

We use N_r and N_r^* to denote the random variable of the parameter n_r for the test burst and an arbitrary packet of the burst, respectively.

$$\begin{aligned} \mathrm{P}(E=e|\, I=\iota) &= \sum_{k=k_{min}}^{k_{max}} \mathrm{P}(K=k, L=e-k|\, I=\iota) = \sum_{k=k_{min}}^{k_{max}} \mathrm{P}(K=k|\, I=\iota)\ \mathrm{P}(L=e-k|\, I=\iota) \\ &= \sum_{k=k_{min}}^{k_{max}} \binom{e_B-\iota}{k} (1-p)^k\, p^{e_B-\iota-k} \binom{e_B-\iota}{e-k-1} p^{e-k-1}\, (1-p)^{e_B-\iota-(e+k+1)} \end{aligned} \tag{3}$$

Burst n_r-reorder: The complementary probability $\mathrm{P}(N_r \geq n_r)$ will be derived. According to the definition in Section II, a N_r-reordered burst has $N_r \geq n_r$ if and only if at least n_r bursts that arrive consecutively at the destination immediately before the test burst, have their sequence number $s > 0$. In the studied scenario, this is equivalent to the combination of two conditions. There exists a burst with sequence number $s = w : 1 \leq w \leq e_B - n_r + 1$ such that:

1) this burst follows Link A and there are exactly $n_r - 1$ bursts with sequence number $s : w < s \leq e_B$ that follow Link A. The test burst follows Link B.
2) the bursts with sequence number $s : w - e_B \leq s < 0$ follow Link A.

Here 1) assures that there are at least n_r bursts that arrive earlier than the test burst and have sequence number larger than 0. 2) prevents a burst with sequence number smaller than 0 arriving between these n_r bursts. The probability for 1) is denoted by $p_1(w)$ and can be calculated as:

$$p_1(w) = (1-p) \binom{e_B - w}{n_r - 1} (1-p)^{n_r-1}\, p^{e_B-w-(n_r-1)}\, p$$

The probability for 2) is denoted by $p_2(w)$ and can be calculated by $p_2(w) = (1-p)^{e_B-w}$. As 1) and 2) hold independently of each other, the complementary distribution for N_r can be obtained by:

$$\mathrm{P}(N_r \geq n_r) = \sum_{w=1}^{e_B-n_r+1} p_1(w)\, p_2(w) \tag{4}$$

From (4) the probability distribution of N_r can be easily obtained by recursive subtraction.

Packet n_r-reorder: If a burst is N_r-reordered and there are n_P packets per burst, then the first packet in the burst is N_r^*-reordered with $N_r^* = n_P\, N_r$. The other packets in the burst are not n_r-reordered because the preceding packet always has a smaller sequence number. In overall, the probability for a test packet to be n_r-reordered is:

$$\mathrm{P}(N_r^* = n_r) = \begin{cases} 1/n_P\ \mathrm{P}(N_r = z), & \text{if } n_r = z\, n_P,\ z \in \mathbb{N} \\ 0, & \text{otherwise;} \end{cases}$$

This shows that the probability of n_r-reordering is reduced by the number of packets per burst. The probability of a packet to invoke the fast retransmit algorithm with a dup-ack threshold of δ is the sum of probabilities for $N_r^* \geq \delta$.

$$\mathrm{P}(N_r^* \geq \delta) = \sum_{d=\delta}^{e_B\, n_P} \mathrm{P}(N_r^* = d) = \frac{1}{n_P}\ \mathrm{P}(N_r \geq \left\lceil \frac{\delta}{n_P} \right\rceil)$$

This can be used to estimate the pseudo packet loss probability, which is the key input parameter for analytic TCP performance models (overview in [6]). This equation can also guide the adjustment of the dup-ack threshold for throughput optimization.

V. Conclusion

In this paper we proposed and analyzed a generic model to grasp the reordering impact from buffering and deflecting of bursts in OBS networks in a multi-layer perspective. The reordering is characterized on both the burst and packet layer by applying the IETF reordering metrics. Our work provides the first analytical characterization of burst/packet reordering and paves the way to an integrated performance analysis for TCP over OBS.

References

[1] M. Allman, V. Paxson, and W. Stevens, "TCP Congestion Control ," IETF, RFC 2581, Apr. 1999.

[2] S. Gowda, R. Shenai, K. Sivalingam, and H. Cankaya, "Performance evaluation of TCP over optical burst-switched (OBS) WDM networks," in *Proc. of IEEE ICC*, vol. 2, 2003, pp. 1433–1437 vol.2.

[3] A. Detti and M. Listanti, "Impact of segments aggregation on TCP Reno flows in optical burst switching networks," in *Proc. of IEEE INFOCOM*, 2002.

[4] A. Morton, L. Ciavattone, G. Ramachandran, S. Shalunov, and J. Perser, "Packet Reordering Metrics," IETF, RFC 4737, Nov. 2006.

[5] S. Gunreben, "Multi-layer Analysis to Quantify the Impact of Optical Burst Reordering on TCP Performance," in *Proc. of the 9th International Conference on Transparent Optical Networks (ICTON)*, 2007.

[6] I. Khalifa and L. Trajkovic, "An overview and comparison of analytical TCP models," in *Proceedings of the 2004 International Symposium on Circuits and Systems, 2004. ISCAS '04.* , vol. 5, May 2004, pp. 469–472.