An Optical Burst Reordering Model for a Time-based Burst Assembly Scheme

Sebastian Gunreben
Universität Stuttgart
Institute of Communication Networks and Computer Engineering
Stuttgart, Germany
Email: gunreben@ikr.uni-stuttgart.de

***Abstract*—In optical burst switching networks, contention resolution schemes as well as contention avoidance schemes reduce the burst loss probability. These schemes delay the burst delivery and may change the burst arrival sequence. In this paper we present an analytic burst reordering model and derive analytically the impact of a time-based burst assembly scheme on the burst reordering pattern. Our results hold for a general network delay distribution. We apply the IETF WG IPPM reordering metrics and calculate explicitly three reordering metrics assuming a general burst delay distribution: the reordering degree, the extent metric for buffer dimensioning and the TCP relevant metric for TCP throughput estimation. We show that our analytic model represents a worst case reordering scenario, which enables studies on the upper layer protocol performance in OBS networks without excessive multi-layer simulations.**

***Index Terms*—burst reordering, time-based assembly**

I. Introduction

Optical burst switching (OBS, [1]) is a promising new network technology for core and metro networks based on wavelength division multiplex. It shows equal resource efficiency as optical packet switching, while it additionally mitigates the technical limitations of an all optical network.

At the OBS network edge, the OBS assembly unit aggregates incoming packets based on their destination address and optionally their service class. At the end of the assembly process, the assembly unit forwards the burst to the optical transmission unit heading to the destination node. The literature proposes various assembly schemes like time- or size-based assembly or a combination of both. Each of this schemes shows a different traffic characteristic of the departing bursts. Vega and Laevens present a survey of the traffic characteristics of these assembly strategies in [2] and [3].

In OBS networks, contention occurs on intermediate nodes if two or more bursts request the same wavelength at the same time. Given this situation, original OBS discards all but one successful burst. These burst losses degrade the transport service and stimulate the research on contention resolution [4], [5] and contention avoidance schemes, e. g., multipath routing [6] for reducing burst losses.

Two widely applied contention resolution schemes are buffering using fiber delay lines (FDL) and deflecting routing. Both contention resolution schemes and the multipath routing scheme delay individual bursts compared to the primarily planned path. As a result, the burst arrival order may change at the destination. Since each data burst is an aggregate of multiple packets, out-of-sequence burst delivery also implies a special out-of-sequence packet delivery, which may affect transport protocols and application protocols.

Transport protocols provide an unreliable or a reliable connection service to applications. Out-of-sequence packet delivery of the same flow affects the performance of these protocols [7]. For instance, the real time transmission protocol (RTP, [8]) providing an unreliable transport services for video and audio services suffers from packet reordering as studied in [7]. Consequently, it provides mechanisms to regain the original packet sequence, e. g., by a de-jitter buffer, or alternatively, discard out-of-sequence packets and degrade the service.

The transmission control protocol (TCP, [9]) is the most important representative of transport protocols for a reliable connection service in the Internet. The basic TCP congestion control algorithm [9] suffers from missing or out-of-sequence packets. The TCP receiver responds on incoming segments with the next expected segment sequence number. If the next expected segment does not arrive due to packet loss or delay, subsequent segment arrivals cause the receiver to respond with the missing segment sequence number. Every reception of the same response refers to a duplicate acknowledgment (dup-ack) at the sender. The sender maintains a dup-ack counter. Exceeding the dup-ack threshold triggers the fast retransmit algorithm. The sender resends the missing segment and halves its congestion window. Consequently, the TCP throughput decreases. This encouraged many studies to analyze the protocol performance of TCP in respect to loss and reordering properties of OBS networks. TCP extensions, e. g., TCP DSACK [10], [11], try to restore the original congestion window size if they detect reordering. These implementations are in an early stage and currently not widely deployed [12].

The literature extensively studies in [13]–[18] the impact of burst losses on TCP. These studies investigate an integrated TCP over OBS scenario by simulations or formal methods without analyzing the characteristics of the intermediate layers. Therefore, it is hard to identify the direct quantitative relationship between OBS network characteristics and TCP throughput.

The literature rarely studies the impact of burst reordering on TCP and other upper layer protocols. In [19], Perelló et al. quantify by simulation the impact of contention resolution schemes on optical burst reordering and estimate the TCP performance. They measure the amount of optical burst reordering in the same order of magnitude as the burst loss probability. These results emphasize the necessity for a detailed investigation on optical burst reordering. Schlosser et al. analyze in [20] the impact of burst deflection on the TCP performance by intensive simulations. They apply an integrative TCP over OBS network model including only a single alternative path. This gives a first insight in the performance of TCP in this scenario but does not allow any generalization in a network wide analysis with an arbitrary delay distribution between source and destination node.

A more basic problem is the definition of an out-of-sequence packet and the characterization of its out-of-sequence pattern. The literature proposes several different out-of-sequence metrics. Piratla et al. propose in [21] the reorder density to measure the amount and displacement of a packet. They compare their approach to the standardized metrics of the IETF in [22]. Also in the field of optical burst reordering Callegati et al. propose a simple measure for out-of-sequence bursts in [16]. Both metrics lack a standardized approach. Due to their individual nature and for a comprehensible study we consider the standardized metrics of the IETF [23].

In our previous work [24], [25] we proposed a first model to investigate the burst reordering phenomena analytically. We assumed the burst traffic characteristic of a time-based assembly scheme with one additional delay link. For this model, we derived analytically the reordering metrics and estimated the impact on TCP.

In here, we extent this burst reordering model to an arbitrary number of additional delay links. This extension models any network delay distribution between any source/destination pair in an OBS network. We apply the IETF WG IPPM reordering metrics [23] to classify and to quantify the amount of burst reordering. We provide the first time the exact analysis of these reordering metrics for our reordering model in a general delay environment. We show that our model approximates the expected burst reordering characteristics of a time-based burst assembly environment. The knowledge of the burst reordering characteristics together with the information on the number of packets per burst enables transport protocol or application protocol performance studies.

Each burst consists of a certain number of packets – in most cases – of different flows. We focus on one specific flow and assume exactly one packet of this flow in every burst of this source/destination node pair. Then, any burst out-of-sequence pattern equals the packet out-of-sequence pattern. In our previous work [19], [25] we show, that this assumption serves either as a worst case regarding reordering or provides the initial condition for further metrics.

Given this, our model enables an estimation of the transport protocol performance of e. g. TCP and UDP without extensive network simulations. The model quantifies analytically the buffer size and the retransmission ratio. Single layer studies on the OBS layer are sufficient, a simulation of the whole network stack ranging from the transport protocol layer down to the OBS network layer is not necessary.

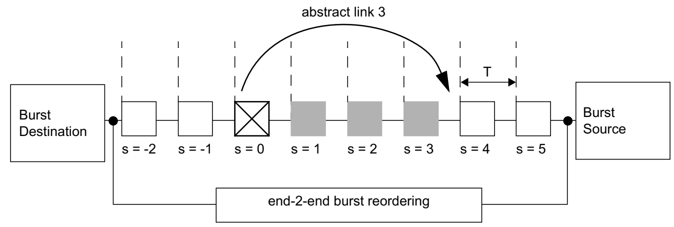

Fig. 1. Scenario with constant burst interdeparture time

We structure our paper in the following way: In section II we introduce the IETF reordering metrics. Section III introduces our generic reordering model for the time-based assembly scheme and derives its analytic solution. In section IV we show numerical results on the reordering metrics and point out the worst case properties of our model. In section V we summarize our findings.

II. Reordering Metrics

This section reviews the IP packet reordering definition and metrics of the IETF WG IPPM [23]. These metrics also hold for generic packet-switched networks like OBS networks.

Reordering definition: The source node assigns each packet a *sequence number*. The sequence numbers increase monotonically. At the destination node a three tuple (i, $s[i]$, $s'[i]$) characterizes each packet arrival. Index i numbers the arriving packet order at the destination. $s[i]$ denotes its sequence number and $s'[i]$ denotes its expected sequence number. The previously received packet determines the value of $s'[i]$. We distinguish two cases:

1) $s[i] < s'[i]$ packet i arrives *out-of-sequence* and $s'[i+1] = s'[i]$.
2) $s[i] \geq s'[i]$ packet i arrives *in order* and $s'[i+1] = s[i]+1$.

Literally, a packet arrives out-of-sequence, if there is one packet with a larger sequence number arriving prior to it. The first packet arrives in order by definition.

Reordering ratio: The reordering ratio reflects the ratio of the number of out-of-sequence packets to the total number of received packets. It equals the probability of an out-of-sequence packet arrival.

Reordering extent: The reordering extent metric quantifies the buffer size needed to restore packet order at the destination. It equals the number of packet arrivals between its nominal in order position and its actual arriving position. Formally, the extent e_i for an out-of-sequence packet i is $e_i = i - \min_{j<i}\{j : s[j] > s[i]\}$.

n_r**-Reordering metric:** This TCP-relevant metric quantifies the violation of the TCP dup-ack threshold. An n_r-reordered packet triggers n_r dup-acks. If there is a set of n_r packets directly preceding packet i and $s[i]$ is smaller than the sequence number of each of these packets, then these packets trigger n_r dup-acks. Formally, packet i is n_r-reordered if $s[j] > s[i]\ \forall j \in \{o : i - n_r \leq o < i, \text{ and } o \in \mathbb{N}\}$.

III. Burst reordering analysis

This section is twofold. First, we introduce our burst reordering model for a time-based assembly scheme and second, we provide the closed form solution of the reordering metrics. For this investigation, we assume a lossless burst network.

A. Burst reordering model

Our model considers the burst traffic sent from an OBS source node to an OBS destination node, c. f. figure 1. These bursts may follow different paths from source to destination node, due to contention resolution and avoidance schemes. Consequently, in an e2e consideration we observe a certain burst delay distribution.

We model each alternative path from source to destination node by one abstract link l. In general we assume m, $m \in \mathbb{N}^+$ parallel abstract links l_1 to l_m. Besides this, l_0 represents the primarily planned shortest path with no extra delay. m is finite as the network itself limits the number of alternative paths.

On each of this abstract links, the burst receives an additional delay, reflecting the time in an FDL or on a deflection path. This delay is, in general, predictable. For this reason we discretize this additional delay by the basic delay unit $\Delta \in \mathbb{R}^+$. Each abstract link represents an integer multiple delay of Δ.

We define a 3-tuple $(k, p_k, k\,\Delta)$ characterizing each abstract link l_k: the link number k, $0 \le k \le m$; the probability p_k, $0 \le p_k \le 1$ to follow l_k and the delay $k\,\Delta$ as an integer multiple of the basic delay unit. Note that the probabilities p_k are independent of each other, as in OBS bursts are switched independently of each other. Further the law of total probability holds: $\sum_{k=0}^{m} p_k = 1$.

Figure 2 depicts the general reordering scenario for one selected burst, i. e., the *test burst*, but our following considerations also hold for any other burst. The arrow line indicates the relative change of the position in the burst series *at the destination* if the burst follows an abstract link. We distinguish three kinds of bursts:

1) the test burst for which we evaluate the reordering metrics. Without loss of generality, its sequence number s is $s = 0$.
2) bursts departing later but arriving earlier than the test burst because of the delay of the test burst (gray).
3) bursts departing and arriving earlier than the test burst and bursts departing and arriving later than the delayed test burst (white).

B. Time-based assembly scheme

In this section we derive the three reordering metrics of section II for the burst departure traffic of a time-based assembly scheme. We approximate the burst interdeparture time by the timeout value T of the time-based assembly scheme. Consequently, we assume a constant interdeparture time of the bursts. This is in general possible, if the mean packet interarrival time is significantly smaller than the timeout parameter of the assembly [2].

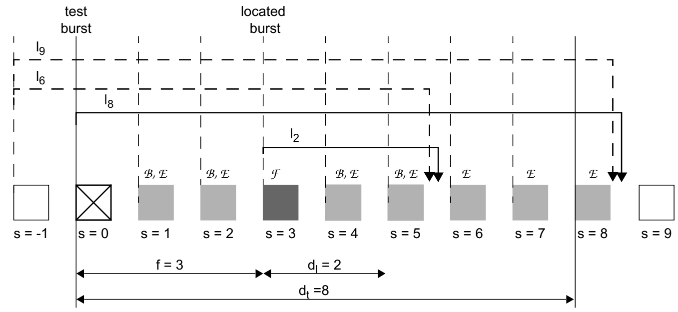

Fig. 2. Burst reordering model

We express the basic delay unit Δ by the interdeparture time T as delays smaller than the interdeparture time would not cause any reordering, i. e. $\Delta = T$. We identify the bursts by their sequence number s. As the burst delay is proportional to the constant interarrival time Δ, we abbreviate a delay of $d\,\Delta$ by d bursts. Next, we calculate the reordering metrics.

1) Burst reordering probability: We first concentrate on the burst reordering probability of the test burst. The test burst arrives out-of-sequence at the destination if there is at least one burst arrival with $s > 0$ prior to the test burst. The reordering probability is a joint probability of (a) the test burst receives a delay and (b) there is at least one burst arrival with larger sequence number than zero before the test burst. For condition (a) the test burst receives a delay of d_t with probability p_{d_t}. Then d_t candidate bursts may accomplish condition (b).

We derive the probability of (b) by its complement. The random variable of the delay of the test burst is D_t. Then the probability that the candidate burst j, $0 < j \le d_t$ does not accomplish condition (b) is $\mathrm{P}(B = 0 \,|\, D_t = d_t \,|\, J = j) = \sum_{k=d_t-j+1}^{m} p_k$. B denotes the arrival of burst J before the test burst. The joint probability that none of the candidate bursts accomplish condition (b) at the same time is $\mathrm{P}(B = 0 \,|\, D_t = d_t) = \prod_{j=1}^{d_t} \mathrm{P}(B = 0 \,|\, D_t = d_t \,|\, J = j)$. The complementary probability of $\mathrm{P}(B = 0 \,|\, D_t = d_t)$ accomplishes condition (b). The burst reordering probability results in $P = \sum_{d_t=1}^{m} p_{d_t}\ (1 - \mathrm{P}(B = 0 \,|\, D_t = d_t))$.

2) Reordering extent: In this section, we calculate the probability density function (PDF) of the burst reordering extent. We first depict the overall scenario. Second, we classify the bursts according to their delay characteristic. Third, we introduce random events for these bursts, which will lead to the burst extent PDF.

The extent equals the number of burst arrivals between the located burst and the test burst. According to the definition in section II, the located burst has the smallest sequence number greater than the test burst arriving prior to the test burst. The burst with $s = f$, $0 < f \le d_t$ is the located burst. It receives a delay of d_l, $0 \le d_l < d_t + f$. The bursts with $0 < s < f$ and the bursts with $f + 1 < s \le f + d_l$ in case of a delayed located burst, arrive earlier than the located burst.

Figure 2 depicts this scenario for a test burst delay of $d_t = 8$ and the located burst $f = 3$ with a delay of $d_l = 2$. Note that we extended the actual delay in the figure to highlight the burst arrival order.

$$P(E=e)=\sum_{d_t=1}^{m}\sum_{f=1}^{d_t}\sum_{d_l=0}^{(d_t-f-1)^+} p_{d_t}\, p_{d_l}\;\mathrm{P}(\mathfrak{F}\,|\,F=f\,|\,D_l=d_l)\; p_{e-1}(d_t,f,d_l) \tag{1}$$

$$P(N_r\geq n_r)=\sum_{d_t=1}^{m}\sum_{f=1}^{d_t}\sum_{d_l=0}^{(d_t-f-1)^+} p_{d_t}\, p_{d_l}\; p_{s>0}(n_r-1,d_t,f,d_l)\; p_{s<0}(0,d_t,f,d_l) \tag{2}$$

The bursts arriving prior to the test burst may each arrive at a different position to the test burst. The following three random events (re) structure each of these different positions and highlight their necessary conditions. We first define these random events and second assign each burst to one or more of these events. Afterwards, we calculate the probabilities of these random events.

re $\mathfrak{F}$ applies to the located burst only

re $\mathfrak{E}$ applies to all bursts arriving later than the located burst and prior to the test burst and thus contributes to the extent, and

re $\mathfrak{B}$ applies to bursts, which have to arrive later than the located burst due to the condition of the located burst.

Next, we classify the bursts according to these random events. For illustration purpose, we also assigned the bursts in figure 2 to the corresponding random events:

$s<0$**:** Bursts departing earlier than the test burst ($s<0$) may arrive later than the located burst and thus contribute to the extent. Event $\mathfrak{E}$ applies.

$0<s<f$**:** These bursts may contribute to the extent but overall they arrive later than the located burst f, due to the condition of the located burst. Events $\mathfrak{E}$ and $\mathfrak{B}$ apply.

$f<s\leq f+d_l$**:** If the located burst f is delayed, too, events $\mathfrak{E}$ and $\mathfrak{B}$ apply because of the same reasons as above.

$f+d_l<s\leq d_t$**:** Bursts which originally arrive later than the located burst but prior to the test burst contribute to the extent. Event $\mathfrak{E}$ applies.

Next, we calculate the probability of each of these random events. The different combinations of the bursts applying these random events lead to the extent metric.

a) Random event $\mathfrak{E}$: As previously mentioned, bursts with $s\leq d_t$ contribute to the burst extent metric if they arrive later than the located burst but prior to the test burst. The probability of a burst with sequence number s to arrive later than the located burst but prior to the test burst depends on its location S, the delay of the test burst D_t and the located burst F and its possible delay D_l.

We denote the probability for a burst with sequence number s arriving later than the located burst but prior to the test burst as $\mathrm{P}(B=1\,|\,S=s\,|\,D_t=d_t\,|\,F=f\,|\,D_l=d_l)$. Herein the random variable B represents the burst arrival ($B=1$) prior to the test burst and after the located burst, otherwise $B=0$. Due to space limitations, we abbreviate this probability by $p_{1s}(d_t,f,d_l)$. Its definition in (3) includes sums of probabilities for all possible values and combinations of d_l and f. These sums indicate the probability to follow different abstract links and finally to arrive later than the located burst but prior to the test burst.

$$p_{1s}(d_t,f,d_l) = \begin{cases} \sum_{\kappa=f-s}^{d_t-s} p_\kappa, & \text{if } s<0 \,\wedge\, d_l=0;\\ \sum_{\kappa=f-s}^{d_t-s-1} p_\kappa, & \text{if } 0<s<f \,\wedge\, d_l=0;\\ p_0+\sum_{\kappa=1}^{d_t-s-1} p_\kappa, & \text{if } f<s\leq d_t \,\wedge\, d_l=0;\\ \sum_{\kappa=f+d_l+1-s}^{d_t-s} p_\kappa, & \text{if } s<0 \,\wedge\, d_l\neq 0;\\ \sum_{\kappa=f+d_l+1-s}^{d_t-s-1} p_\kappa, & \text{if } 0<s\leq f+d_l \,\wedge\, d_l\neq 0;\\ p_0+\sum_{\kappa=1}^{d_t-s-1} p_\kappa, & \text{if } f+d_l<s\leq d_t \,\wedge\, d_l\neq 0;\\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

b) Random event $\mathfrak{B}$: Random event $\mathfrak{B}$ applies for bursts with $s>0$, which originally arrive prior to the located burst. These bursts must not arrive prior to the located burst as a necessary condition of the located burst. We apply the law of total probability and calculate the probability of event $\mathfrak{B}$ by its complementary. $\mathrm{P}(\bar{\mathfrak{B}})$ denotes the probability of a burst arrival for a specific burst prior to the located burst. This probability depends on the original location S of the burst and the located burst F and its delay D_l. $Q(B=1\,|\,S=s\,|\,F=f\,|\,D_l=d_l)$ denotes this probability. The random variable B indicates the burst arrival prior to the located burst. We abbreviate this probability by $q_{1s}(f,d_l)$. In (4) we derive the probabilities $q_{1s}(f,d_l)$ for all bursts applying random event $\mathfrak{B}$.

$$q_{1s}(f,d_l) = \begin{cases} 1-\sum_{\kappa=0}^{f-s-1} p_\kappa, & \text{if } 1<s<f \,\wedge\, d_l=0;\\ 1-\sum_{\kappa=0}^{f+d_l-s} p_\kappa, & \text{if } 1<s\leq f+d_l \,\wedge\, d_l\neq 0;\\ 1, & \text{otherwise.} \end{cases} \tag{4}$$

c) Conditional random events $\mathfrak{B}$ and $\mathfrak{E}$: Event $\mathfrak{B}$ is a necessary condition for the bursts with a smaller sequence number number than the located burst. These bursts also apply event $\mathfrak{E}$ (c.f. figure 2). This results in a conditional probability for these bursts contributing to the extent. Event $\mathfrak{B}$ conditions the probability that these bursts contribute to the extent by event $\mathfrak{E}$. We calculate this conditional probability $\mathrm{P}_{\mathfrak{B}}(\mathfrak{E})=\frac{\mathrm{P}(\mathfrak{B}\cap\mathfrak{E})}{\mathrm{P}(\mathfrak{B})}=\frac{\mathrm{P}(\mathfrak{E})}{\mathrm{P}(\mathfrak{B})}=\frac{\mathrm{P}(\mathfrak{E})}{1-\mathrm{P}(\bar{\mathfrak{B}})}$. As event $\mathfrak{E}$ includes random event $\mathfrak{B}$ as well, $\mathrm{P}(\mathfrak{B}\cap\mathfrak{E})=\mathrm{P}(\mathfrak{E})$ holds. With the previous expressions (3) and (4) we calculate the conditional probability:

$$p^*_{1s}(d_t,f,d_l)=\frac{p_{1s}(d_t,f,d_l)}{q_{1s}(f,d_l)} \tag{5}$$

d) Random event $\mathfrak{F}$: Each of the bursts with $0<s\leq d_t$ may serve as the located burst. The sequence number of the located burst is f. The located burst receives a delay of d_l with probability p_{d_l}. The necessary condition for the located burst is

the arrival of bursts with $0 < s < f$ later than the located burst f. This necessary probability depends on the sequence number F and the delay D_l of the located burst: $\mathrm{P}(\mathfrak{F} \,|\, F = f \,|\, D_l = d_l) = \prod_{s=1}^{d_l+f} q_{1s}(f, d_l)$. This joint probability requires a later arrival of bursts, which departed earlier than the located burst.

With the probabilities of these random events, we calculate the burst reordering extent.

e) Reordering extent: For a given scenario with a test burst delay of D_t, a located burst F and its delay D_l, we denote the probability of E burst arrivals between the located burst and the test burst by $P(E = e \,|\, D_t = d_t \,|\, F = f \,|\, D_l = d_l) = p_e(d_t, f, d_l)$. We derive this probability using the probability generator function (GF) on (3) and (5). The GF of $p_{1s}(d_t, f, d_l)$ includes two states:

$$G_{s,d_t,f,d_l}(z) = \sum_{i=0}^{1} p_{is}(d_t, f, d_l)\, z^i \tag{6}$$

$$= \begin{cases} p_{0s}(d_t, f, d_l) + p^*_{1s}(d_t, f, d_l)\, z & \text{if } 0 < s \leq f + d_l \\ p_{0s}(d_t, f, d_l) + p_{1s}(d_t, f, d_l)\, z & \text{otherwise} \end{cases}$$

The burst arrivals prior to the test burst and after the located burst are independent of each other. The sum of burst arrivals forming the extent is a joint probability experiment. The sum of random variables is the product of their GFs. The GF of the joint experiment is

$$G_{d_t,f,d_l}(z) = \prod_{s=-\infty}^{d_t} G_{s,d_t,f,d_l}(z) \tag{7}$$

The probability distribution function $p_e(d_t, f, d_l)$ corresponds to the derivation of the GF of the joint experiment:

$$p_e(d_t, f, d_l) = \frac{1}{e!} \frac{\partial^e}{\partial z^e} G_{d_t,f,d_l}(z) \bigg|_{z=0} \tag{8}$$

The reordering extent PDF considers every combination of the test burst delay D_t, the location of the located burst F and its delay D_l. Together they form a triple sum (1), where the located burst accounts to the overall extent.

3) n_r-Reordering metric: In this section we derive the complementary cumulative distribution function (CCDF) of the n_r-reordering metric.

The test burst arrives n_r-reordered at the destination if there are at least n_r subsequent burst arrivals with $s > 0$ prior to the test burst. This definition requires two conditions: (a) the test burst receives an extra delay. (b) the sequence of n_r burst arrivals with $s > 0$ at the destination excludes any arrival of bursts with sequence number $s < 0$. The first burst of this sequence is the burst with sequence number f, the located burst. The located burst f receives a delay of d_l.

We denote the probability of $n_r - 1$ burst arrivals between the located burst and the test burst $\mathrm{P}_{s>0}(B = n_r - 1 \,|\, D_t = d_t \,|\, F = f \,|\, D_l = d_l) = p_{s>0}(n_r - 1, d_t, f, d_l)$. Note that only bursts with $s > 0$ contribute to the extent. We denote the probability of no burst arrivals with $s < 0$ between the located burst and the test burst $\mathrm{P}_{s<0}(B = 0 \,|\, D_t = d_t \,|\, F = f \,|\, D_l = d_l) = p_{s<0}(0, d_t, f, d_l)$

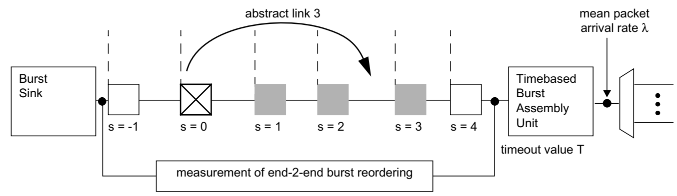

Fig. 3. Scenario with time-based burst assembly

The probability that a burst with sequence number s arrives later than the located burst but prior to the test burst depends on its location S and the delay of the test burst D_t and the located burst F and its delay D_l. These individual probabilities have already been derived in (3). We calculate both probabilities applying (6), (7) and (8).

$$p_{s<0}(0, d_t, f, d_l) = G_{s<0,d_t,f,d_l}(0) \tag{9}$$

$$G_{s<0,d_t,f,d_l}(z) = \prod_{s=-\infty}^{-1} G_{s,d_t,f,d_l}(z) \tag{10}$$

$$p_{s>0}(n_r - 1, d_t, f, d_l) = \frac{\partial^{n_r-1}}{\partial z^{n_r-1}} \frac{G_{s>0,d_t,f,d_l}(z)}{(n_r - 1)!} \bigg|_{z=0} \tag{11}$$

$$G_{s>0,d_t,f,d_l}(z) = \prod_{s=1}^{d_t} G_{s,d_t,f,d_l}(z) \tag{12}$$

With these results, a triple sum forms the CCDF of the n_r-reordering metric in (2).

IV. Impact of interdeparture time approximation

Our model with a constant interdeparture time is an approximation for the expected values of a scenario with a time-based assembly scheme. Any approximation and modelling introduces an error. In our case it influences the traffic characteristics of the burst departure process, which also influences the burst out-of-sequence pattern. In this section we highlight the impact of the approximation error on the burst out-of-sequence pattern and show that our model with a constant interdeparture time reflects the worst case scenario for a time-based assembly scheme.

A. Worst case considerations

Vega and Götz derive in [2] the interdeparture time distribution of a time-based assembly strategy with Poisson packet arrivals. Figure 3 depicts this scenario with two connected nodes and one representative abstract link of our reordering model.

The mean packet arrival rate is λ, while the timeout value of the assembly unit is T. Then, the burst interdeparture time distribution results in $f_z(t) = f_i(t-T) = \lambda \exp\left(-\lambda\,(t - T)\right)$. This corresponds to a shifted negative exponential distribution. The average burst interdeparture time becomes $\mathsf{E}\left[\hat{T}\right] = T + 1/\lambda$. In our model, we approximated the burst interdeparture time by the constant interdeparture time of $\Delta = T < \hat{T}$, which is strict smaller than the expected interdeparture time.

We further assumed the abstract links to delay bursts with a certain probability by integer multiples of Δ. We argued this spacing by the possibility to enable any reordering. As the real burst interdeparture time is larger than the original delay Δ, two subsequent bursts arrive in order, even if the first burst follows an abstract link with Δ delay. Thus the first abstract link l_1 does not enable any reordering. The same considerations apply for abstract links beyond l_1. Abstract link l_2 may delay a burst by one subsequent burst while l_3 may delay a burst by two subsequent burst and so on. Consequently, the largest displacement of abstract link l_m becomes impossible. The delay distribution changes and the probability density function of the extent and the n_r-reordering metrics shifts to the left.

Consequently, the reordering metrics of the expected scenario change. Again, we consider the three metrics. The burst reordering probability decreases as the probability for reordering decreases, i. e., there arrive in total less bursts of of sequence. We expect that the burst reordering probability of the constant interdeparture time scenario serves as an upper limit. The burst extent metric as well as the n_r-reordering metric also change. As the largest values are impossible now and the overall possibility for any reordering decreases, we expect that the CCDF of these metrics of the approximation scenario also serves an upper limit.

We simulated the setup of figure 3 to validate our expectations. Numerical results of selected delay distributions illustrate this property in the next section.

B. Numerical results

In this section we first show some illustrative results to visualize the burst reordering metrics of our model with a constant interdeparture time. Second, we compare these results to the values obtained by simulation.

We parametrize our reordering model with the probability of delay p, which corresponds to the complementary probability to follow l_0, the number of abstract links m and the delay distribution among the m abstract links. For the simulation in the second step we also choose some reasonable values for T and λ. We distinguish three different delay distributions:

- geometric distribution:
  $p_i = q\,(1-q)^{i-1}$ with $q = 1-(1-p)^{1/m}$,
- linear distribution: $p_i = 2\,i\,p/(m^2+m)$,
- complementary linear distribution:
  $p_i = 2\,(m-i+1)\,p/(m^2+m)$.

i gives the index of the abstract link with $1 \leq i \leq m$. The geometrically distributed delay may correspond to FDLs along a path. The linearly distributed delay may correspond to a deflection scenario, where long paths are likely, while the complementary linear distribution may correspond to a scenario where long paths are unlikely.

In the figures 4, 5 and 6 we depict numerical results of the analytic model of figure 2. In figure 4 we depict the burst reordering probability depending on the delay probability p. We depict the probability for $m = 5$ and $m = 15$ for our three delay distributions. The reordering probability increases with the number of abstract links. Further, the reordering probability becomes larger if abstract links with larger delays become more probable, e. g., in case of linear delay distributions.

In figure 5 we illustrate the burst extent PDF for the time-based assembly scheme with $p = 0.1$ and $m = 5$ for our three delay distributions. The three options show different behavior. The complement linear distribution is decreasing as smaller extent values are more likely than larger ones. The linear distribution is bell shaped as its maximum is moved towards larger extent values. The geometric distribution starts between both distributions and decreases only slightly until its knee.

In figure 6 we plotted the CCDF of the n_r-reordering metric for $m = 5$ and $m = 15$ for our three delay distributions and assume $p = 0.1$. If we assume a TCP scenario with a dup-ack threshold of three as proposed in [9], then the number of fast retransmit triggers varies between one and seven percent of the arriving packets depending on m and the delay distribution. In the scenario with $m = 15$, the probability of exceeding even higher thresholds is not negligible. OBS networks matching this scenario are critical to TCP.

We now compare our analytic results with the simulations results of the time based assembly unit of figure 3. We simulated this set-up with our event-driven simulation library [26]. For reasonable results we obtained the statistical values from ten batches each including at least one million burst arrivals. As the absolute values do not matter, we considered the following relative values. The product of the assembly timeout value and the mean packet arrival rate $T\,\lambda$ serves as a basic measure. A large product value leads to a mean burst interdeparture time relatively close to a constant value, while a small product value increases it significantly. In all scenarios we further assume a branch probability of 10%, i. e. $p = 0.1$. The number of abstract links represents with 5 and 15 a small and a large network, respectively.

Figures 7, 8 and 9 illustrate our results for this parameter set. Figure 7 depicts the impact of the reordering ratio for the different packet arrival rates. We show the reordering ratio of the product $T\,\lambda$. The figure depicts all three delay distributions for $m = 5$ and $m = 15$ abstract links. We omitted the calculated values for the reordering ratio as it is just below 0.10 (c.f. figure 4), but always above the simulated values. We observe with a decreasing $T\,\lambda$ a decreasing reordering ratio. If $T\,\lambda$ increases, the simulation values approach the calculated ones. The reordering ratio in the simulation scenario is always lower than the calculated value. The analytic model serves as an upper bound for the reordering ratio as expected.

In figure 8 we depict the CCDF of the burst reordering extent with $m = 15$ abstract links for $T\,\lambda = 1$. For each delay distribution we show two lines. The solid line shows the analytic result. The dashed line shows the simulation results of the simulation model of above. We observe the dashed lines always below the corresponding solid ones. Consequently, the CCDF of the burst reordering extent of our analytic model serves as an upper bound as expected. We omitted the results for $m = 5$ abstract links, as they show the same behaviour. For the complement linear delay distribution we show the simulation results for $T\,\lambda = 1000$, too. The large product value

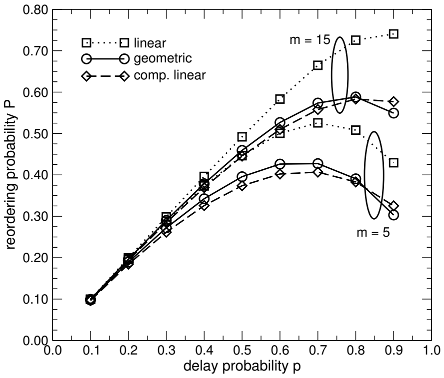

Fig. 4. Analytic model: reordering probability

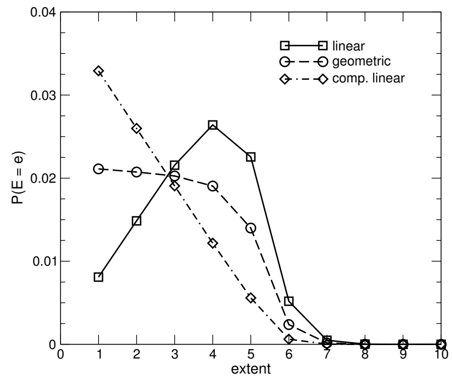

Fig. 5. Analytic model: extent PDF

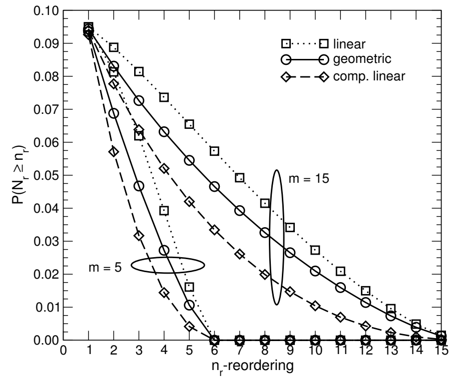

Fig. 6. Analytic model: n_r-reordering CCDF

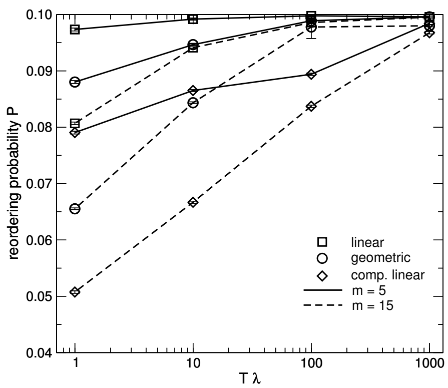

Fig. 7. Simulation model: reordering probability

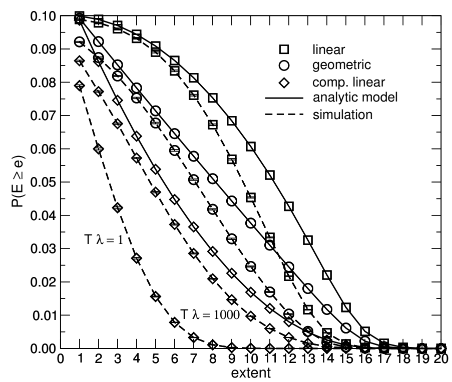

Fig. 8. Extent CCDF, $m = 15$, $T\,\lambda = 1$

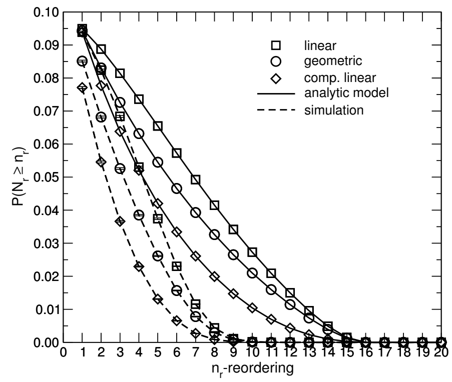

Fig. 9. n_r-reordering CCDF, $m = 15$, $T\,\lambda = 10$

indicates that the expected interdeparture value approaches the approximated interdeparture value. The dashed line highlights this explicitly.

Figure 9 completes this discussion with the comparison of the CCDF of the n_r-reordering metric for $T\,\lambda = 1$. We restrict ourselves to the results of the 15 abstract link scenario as they are also valid for the scenario with $m = 5$ abstract links. We show each of the three delay distributions in two lines. The solid line shows the results of the analytic model and the dashed line shows the corresponding simulation result. We observe the results on the constant interdeparture time serve as an upper bound for the expected n_r-reordering characteristic.

The analytic model with a constant interdeparture time serves for all three metrics with three different delay distributions as an upper limit. Worst case considerations are done more easily with the analytic model than with extensive simulations.

V. Conclusion

We proposed and analyzed an analytic burst reordering model for a time-based assembly scheme. Our model serves as an approximation for the burst reordering characteristic as we assumed a constant interdeparture time. For this model, we derived the three most important IETF reordering metrics for this approximation: the reordering ratio, the reordering extent and the TCP relevant metric. These metrics allow the dimensioning of the required OBS buffer capacity to resolve reordering and give an estimation on the expected dup-acks of the TCP protocol. Our results on the reordering metrics hold for a general OBS network delay distribution.

Besides these formal methods, we also simulated the time-based assembly scheme and measured the reordering metrics. The comparison of the measured values to the values obtained by our analytic model showed the worst case property of our analytic model. Our model with a constant interdeparture time serves as an upper bound for the reordering metrics. We reasoned this by the reduced reordering ratio and the shift of the delay distribution. Our ongoing work will proof this observation analytically.

As our analytic model serves as a worst case scenario, it simplifies any transport layer protocol investigations in respect to optical burst reordering. OBS network simulations may derive the network delay distribution by OBS layer simulations only, without simulating additional layers. This provides a clear layer separation and speeds up simulation time. The observed burst delay distribution serves as an input parameter for our reordering model. The model gives the expected burst reordering metrics as an upper bound analytically. Applying earlier results on the dependencies of burst and packet reordering, the packet reordering metrics may serve as an input parameter for analytic transport protocol models.

This structured layer per layer analysis on reordering provides an estimation on the expected transport protocol performances without excessive multi network-layer simulations.

Acknowledgment

The work described in this paper was carried out with the support of the BONE-project ("Building the Future Optical Network in Europe"), a Network of Excellence by the European Commission through the 7th ICT-Framework Programme.

References

[1] C. Qiao and M. Yoo, "Optical burst switching (OBS)—a new paradigm for an optical Internet," *Journal of High Speed Networks*, vol. 8, no. 1, pp. 69–84, January 1999.

[2] M. de Vega Rodrigo and J. Goetz, "An analytical study of optical burst switching aggregation strategies," in *Proceedings of the Third International Workshop on Optical Burst Switching (WOBS)*, San Jose, October 2004.

[3] K. Laevens, "Traffic characteristics inside optical burst switched networks," in *Proceedings of the Optical Networking and Communications Conference (OptiComm)*, 2002.

[4] C. M. Gauger, M. Köhn, and J. Scharf, "Performance of contention resolution strategies in OBS network scenarios," in *Proceedings of the 9th Optoelectronics and Communications Conference/3rd International Conference on the Optical Internet (OECC/COIN)*, Yokohama/Japan, July 2004.

[5] ——, "Comparison of contention resolution strategies in OBS network scenarios," in *Proceedings of the 6th International Conference on Transparent Optical Networks (ICTON)*, vol. 1, 2004, pp. 18–21.

[6] C. G. Argos, O. G. de Dios, and J. Aracil, "Adaptive Multi-Path Routing for OBS Networks," in *Proceedings of the 9th International Conference on Transparent Optical Networks (ICTON)*, vol. 3, no. 1-5, July 2007, pp. 299–302.

[7] L. Gharai, C. Perkins, and T. Lehman, "Packet reordering, high speed networks and transport protocol performance," *Computer Communications and Networks, 2004. ICCCN 2004. Proceedings. 13th International Conference on*, pp. 73–78, 11-13 Oct. 2004.

[8] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications," IETF, RFC 3550, Jul. 2003.

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

[10] S. Floyd, J. Mahdavi, M. Mathis, and M. Podolsky, "An Extension to the Selective Acknowledgement (SACK) Option for TCP," IETF, RFC 2883, Jul. 2000.

[11] M. Zhang, B. Karp, S. Floyd, and L. Peterson, "RR-TCP: a reordering-robust TCP with DSACK," *Network Protocols, 2003. Proceedings. 11th IEEE International Conference on*, pp. 95–106, 4-7 Nov. 2003.

[12] A. Medina, M. Allman, and S. Floyd, "Measuring the evolution of transport protocols in the internet," *SIGCOMM Comput. Commun. Rev.*, vol. 35, no. 2, pp. 37–52, 2005.

[13] X. Yu, J. Li, X. Cao, Y. Chen, and C. Qiao, "Traffic statistics and performance evaluation in optical burst switched networks," *IEEE/OSA Journal of Lightwave Technology*, vol. 22, no. 12, pp. 2722–2738, 2004.

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

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

[16] F. Callegati, W. Cerroni, and C. Raffaelli, "Impact of Optical Packet Loss and Reordering on TCP Performance," in *IEEE Global Telecommunications Conference (GLOBECOM '06)*, Nov. 2006, pp. 1–5.

[17] J. He and S. H. G. Chan, "TCP and UDP performance for Internet over optical packet-switched networks," *Computer Networks*, vol. 45, no. 4, pp. 505–521, Jul. 2004.

[18] J. P. Gelpke, M. Schlosser, E. Patzak, and H. Buchta, "Assessment of tcp performance in obs networks with load dependent contention," in *Proceedings of the 9th International Conference on Transparent Optical Networks (ICTON)*, 2007.

[19] J. Perelló, S. Gunreben, and S. Spadaro, "A Quantitative Evaluation of Reordering in OBS Networks and its Impact on TCP Performance," in *Proceedings of the Optical Network Design and Modelling Conference (ONDM)*, Mar. 2008.

[20] M. Schlosser, E. Patzak, and P. Gelpke, "Impact of deflection routing on TCP performance in optical burst switching networks," in *Proceedings of the 7th International Conference on Transparent Optical Networks (ICTON)*, vol. 1, Barcelona, June 2005, pp. 220–223.

[21] N. Piratla, A. Jayasumana, and T. Banka, "On reorder density and its application to characterization of packet reordering," *Local Computer Networks, 2005. 30th Anniversary. The IEEE Conference on*, pp. 156–165, 15-17 Nov. 2005.

[22] N. M. Piratla and A. P. Jayasumana, "Metrics for packet reordering—a comparative analysis," *Int. J. Commun. Syst.*, vol. 21, no. 1, pp. 99–113, 2008.

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

[24] S. Gunreben and G. Hu, "A multi-layer analysis on reordering in optical burst switched networks," *IEEE Communications Letters*, vol. 11, no. 12, pp. 1013–1015, Dec. 2007.

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

[26] S. Bodamer, K. Dolzer, C. Gauger, and M. Necker, *IND Simulation Library 2.3 User Guide Part I: Introduction*, Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, September 2004.