Performance Evaluation 68 (2011) 237–255

Contents lists available at ScienceDirect

Performance Evaluation

journal homepage: www.elsevier.com/locate/peva

An optical burst reordering model for time-based and random selection assembly strategies☆

Sebastian Gunreben*

Universität Stuttgart, Stuttgart, Germany

ARTICLE INFO

Article history:
Received 4 March 2008
Received in revised form 17 June 2009
Accepted 16 November 2010
Available online 26 November 2010

Keywords:
TCP over OBS
Burst reordering
Time-based assembly
Random selection

ABSTRACT

Contention resolution schemes in optical burst switched networks (OBS) as well as contention avoidance schemes delay burst delivery and change the burst arrival sequence. The burst arrival sequence usually changes the packet arrival sequence and degrades the upper layer protocols performance, e.g., the throughput of the transmission control protocol (TCP).

In this paper, we present and analyze a detailed burst reordering model for two widely applied burst assembly strategies: time-based and random selection. We apply the IETF reordering metrics and calculate explicitly three reordering metrics: the reordering ratio, the reordering extent metric and the TCP relevant metric. These metrics allow estimating the degree of reordering in a certain network scenario. They estimate the buffer space at the destination to resolve reordering and quantify the number of duplicate acknowledgements relevant for investigations on the transmission control protocol.

We show that our model reflects the burst/packet reordering pattern of simulated OBS networks very well. Applying our model in a network emulation scenario, enables investigations on real protocol implementations in network emulation environments. It therefore serves as a substitute for extensive TCP over OBS network simulations with a focus on burst reordering.

1. Introduction

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

At the edge of an OBS network, the OBS assembly unit aggregates incoming packets based on their destination address and optionally their service class. Packets of the same aggregate belong to the same burst. At the end of the assembly process, the assembly unit forwards the burst to the optical transmission unit heading to the destination node on the path with the least delay. The assembly strategy, i.e., the criteria when to finish the assembly process, may be time-based, size-based or a combination of both. Vega and Goetz as well as Laevens survey the statistical properties of these assembly strategies in [2,3].

Prior to the actual burst transmission, a control packet heading for the same destination configures the intermediate nodes along the path to the destination. An offset time between control packet and burst realizes the necessary

☆ 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 and the Integrated Project NOBEL2 both funded by the European Commission through the 7th ICT-Framework Programme.

* Corresponding address: Institute of Communication Networks and Computer Engineering, Pfaffenwaldring 47, 70569 Stuttgart, Germany. Tel.: +49 71168567968; fax: +49 71168557968.
E-mail address: gunreben@ikr.uni-stuttgart.de.

0166-5316/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.peva.2010.11.006

configuration time. Contention occurs if two or more bursts request the same wavelength at the same time. The duration of contention is in the order of the burst transmission time. In this case, the original OBS discards all but the successful burst.

Contention resolution schemes (cf. survey in [4]) reduce the burst loss probability [5,6]. Besides wavelength conversion, literature proposes two additional contention resolution schemes: local buffering using fibre delay lines (FDL) and deflecting routing, i.e., forwarding the burst on an alternative path to the destination. Additionally to contention resolution schemes, contention avoidance may also reduce the burst loss probability, by e.g., multipath routing [7]. It avoids contention by load balancing on alternative paths.

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 pattern. Especially, if this affects packets of the same flow, it influences higher-layer protocol performance.

Out-of-sequence packets especially affect transport protocols [8]. They provide a reliable or unreliable but always an in-order connection service to applications. The transmission control protocol (TCP, [9]) and the newly developed stream control transmission protocol (SCTP, [10]) are the most important representatives of transport protocols for a reliable connection service. In addition, real-time protocols for unreliable transport services, e.g., the real-time transmission protocol (RTP, [11]) for video and audio services, also suffer from packet reordering [8]. They have to provide mechanisms, e.g., a de-jitter buffer, to regain original packet sequence or they discard out-of-sequence packets and degrade the service.

TCP is the dominant transport layer protocol on the Internet. The basic TCP congestion control algorithm [9] suffers from missing or out-of-sequence packets. The TCP receiver responds to 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. The sender refers to every reception of the same response as a duplicate acknowledgment (dup-ack). Thereby, 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, i.e., the amount of data sent as a whole. Consequently, the TCP throughput decreases. Thus, it is important to analyze the protocol performance of TCP in respect to loss and reordering properties of OBS networks. Modern TCP implementations, e.g., TCP SACK [12], try to detect and clear reordering in a smooth way to limit the TCP throughput degradation. These implementations are in an early stage and currently not widely deployed in operating systems.

The impact of burst losses on TCP has been studied extensively in literature [13–16]. These studies investigate an integrated TCP-over-OBS scenario by simulations or analytics. Their approach requires multi-scale simulations (burst and packet time scale) as well as a model for both entities (TCP and OBS). Especially the large number of implementation variations of TCP and OBS make it hard to give general results on the relationship between OBS network characteristics and TCP throughput. Additionally, each model introduces some error within the simulation or formal description, which may change the results with respect to real experiences. Consequently, it is more advisable — if possible — to use real protocol implementations of TCP and to model/simulate the unknown OBS network.

The impact of burst reordering on TCP and other upper layer protocols has not been studied in literature in such detail. Callegati et al. introduce in [13,17] a burst reordering framework for a WDM network, but their reordering definition misses the exact link between the reordering characteristics and the related TCP mechanisms. In [18], Perelló et al. quantify by simulation the impact of contention resolution schemes on optical burst reordering and estimated 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 [19] the impact of burst deflection by intensive simulations. They apply an integrative TCP-over-OBS network model including only a single alternative path. Thus, the results are not representative for a network-wide analysis with a different delay distribution between source and destination node.

An alternative approach of multi-layer investigations is to separate the OBS layer from the transport protocol layer. In case of our reordering studies, we choose the following approach. The first step derives the properties of an OBS network by OBS network simulations. Thereby, the simulations only consider the burst level. The second step abstracts the burst layer properties, in our case the reordering characteristics. The reordering characteristics may serve as parameters for a modified network emulation software [20] or [21] and enable investigations on real protocol implementations. A network emulation showing a certain reordering has been successfully implemented based on [21]. We skip the details on the implementation as they are out of the scope of this paper.

In our previous work [22–24], we proposed a first model to investigate the burst reordering phenomena analytically. We assumed a time-based assembly strategy and a discretized general burst delay characteristic to reason about burst reordering. For this model, we estimated the reordering metrics and focused on the TCP relevant metric to give an estimate on the TCP throughput.

In this paper, we provide a detailed analysis on the reordering characteristic for two widely applied assembly strategies: a *time-based* and a *random selection* assembly strategy [2]. We approximate the time-based assembly scheme by a constant burst inter-departure time. This is in general possible as the timeout value is magnitudes larger than the mean packet inter-arrival time [2]. On every arriving packet, the random selection assembly scheme decides with a certain probability if the burst is ready to send or not. This results in a geometric distribution of the number of packets per burst and in a Poisson burst departure process. The random selection assembly scheme provides the technical background for the studies on OBS network performance and network dimensioning of [25–27]. These studies apply for the network dimensioning

process the Erlang-B formula, which assumes a Poisson arrival process and may lead to wrong results in case of other traffic characteristics. Additionally, investigations on burst loss performance also assume a Poisson burst departure process [2].

This paper provides the exact analysis on burst reordering for these two assembly schemes in a general network delay environment. Thus, our results hold for any network delay distribution for each source/destination pair in an OBS network. We apply the IETF WG IPPM reordering metrics [28] to quantify the amount of burst reordering. We show that our model emulates measured reordering metrics of an OBS simulation with contention resolution schemes. For the emulation of the burst reordering pattern in a network emulation, we formulate an optimization problem to carry out the parameters for our reordering model. The solution provides the parameters for a network emulation environment showing the same packet reordering than the measured burst reordering of an OBS network simulation. This setup allows quantifying TCP performance in a real world scenario using original protocol stacks rather than using extensive multi-scale simulations with inaccurate models.

In this paper, we concentrate on burst reordering and assume exactly one packet of a considered packet flow in every burst of an end-to-end connection. Then the burst reordering characteristic equals the packet reordering characteristic. In our previous work [18], we prove this as a worst case scenario. We also show in [23], which corresponds to the findings of Schlosser et al. in [19], that more than one packet of the same flow per burst decreases the negative impact of burst reordering.

In Section 2, we introduce the IETF reordering metrics. Section 3 introduces our generic reordering model. Section 4 provides our analytic results on both assembly strategies. We compare our analytic results with simulation results of an OBS network and show the applicability of our model in Section 5. We summarize our work in Section 6.

2. Reordering metrics

This section reviews the IP packet reordering definition and metrics of the IETF WG IPPM [28]. These metrics also hold for generic packet-switched networks like OBS networks.

Reordering definition:

The definition of reordering requires the source node assigning each packet a *sequence number*. The sequence number increases monotonically. At the destination node a three tuple $(i, s[i], s'[i])$ characterizes each packet arrival. Index i indicates the arrival 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-order.
$s'[i]$ remains unchanged, i.e., $s'[i+1] = s'[i]$.
(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 ratio of reordered packets to the total amount of received packets refers to the reordering ratio. It equals the probability of an out-of-sequence arrival at the destination.

Reordering extent:

This metric estimates the buffer size needed to restore the 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 a reordered packet at arrival position i at the destination is

$$e_i = i - \min_{j<i} \{j : s[j] > s[i]\}. \tag{1}$$

Therein, the nominal in-order position is characterized by the smallest j, where the corresponding sequence number $s[j]$ is larger than the sequence number of the packet at position i. We name this special burst j as a *located burst* as it indicates the first packet out of a sequence of packets, which has a larger sequence number than the test burst.

n_r- *reordering metric*:

This TCP-relevant metric estimates the number of TCP dup-acks. It defines that a 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 each of these packets triggers one dup-ack. Formally, packet i is n_r-reordered if $s[j] > s[i]\ \forall j \in \{o : i - n_r \leq o < i, \text{ and } o \in \mathbb{N}\}$.

3. Burst reordering model

In this section, we introduce our burst reordering model and analyze the reordering characteristics of a single packet flow belonging to one application, e.g., packets of one TCP connection.

Fig. 1 shows two packet flows between two host systems at the network edge (R1a/b and R2a/b). We focus on one packet flow (R2a–R2b), but our considerations hold for the other packet flow, too. The packets of our flow enter the burst switched network at an edge node. The ingress assembly unit assembles them together with other packets. The network nodes switch

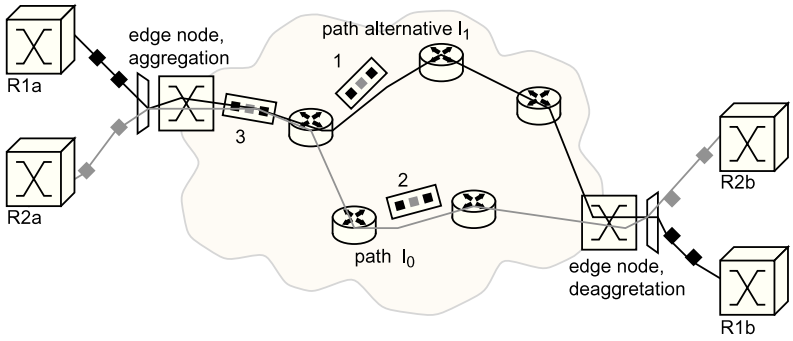

Fig. 1. OBS reordering scenario.

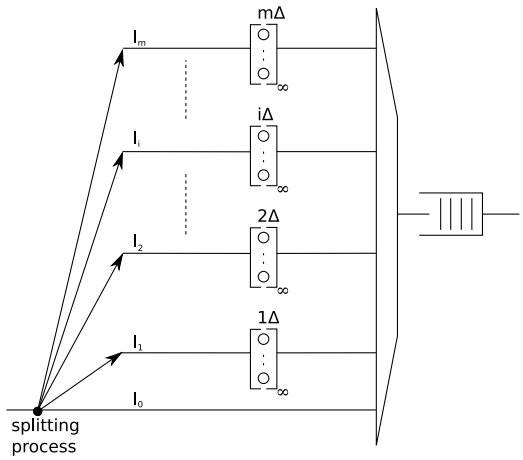

Fig. 2. Generic queueing model.

bursts to the next node on the primarily planned shortest path (l_0) to the destination edge node. In case of congestion, it uses a different path (l_1). After disassembling the burst at the destination edge node, the packets leave the egress edge node. We assume a lossless burst network and focus on the burst reordering phenomena.

Fig. 2 depicts the corresponding queuing model. We model every alternative path from source to destination node by one abstract link. Besides this, l_0 represents the primarily planned shortest path with no extra delay. In general, we assume $m,\ m \in \mathbb{N}$ parallel abstract links l_1 to l_m from source to destination node. Thereby, m is finite as the number of alternative paths in a network is limited.

The additional delay a burst receives in an FDL or on a deflection path is, in general, predictable. For this reason and for simplicity in calculation we discretize the additional delay by the basic delay unit $\Delta \in \mathbb{R}^+$. Each abstract link represents an integer multiple delay of Δ from source to destination. A set of an infinite number of servers realize the delay per abstract link.

Each burst follows independently one of these abstract links randomly. This reflects the probability to follow an FDL or a deflection path. In the figure, this corresponds to the initial splitting process. This assumption is reasonable as the time of congestion lies in the order of the burst transmission time, while the burst inter-arrival time is usually larger. Consequently, subsequent bursts find a different resource occupation state at the node and independence of burst switching can be assumed.

Summarizing, a 3-tuple $(k, p_k, k\ \Delta),\ 0 \leq p_k,\ 0 \leq k \leq m$ characterizes each abstract link l_k: k is the link number; p_k is the probability to follow l_k and $k\ \Delta$ is the burst delay as an integer multiple of the basic delay unit Δ. Further the law of total probability holds: $\sum_{k=0}^{m} p_k = 1$.

Fig. 3 depicts the general reordering scenario for one selected burst, i.e., a *test burst*. For clarity, it shows the possible delays of the test burst only. These 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 $s = 0$. The sequence number is also an identifier of each burst.
(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).

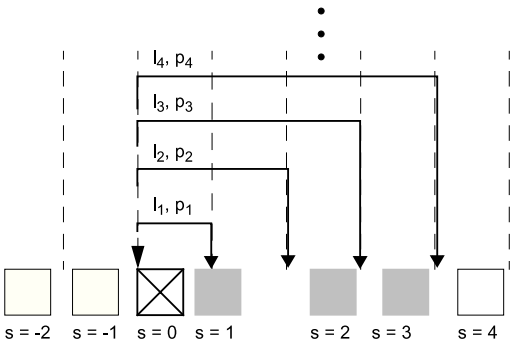

Fig. 3. Generic reordering model.

4. Burst reordering analysis

In this section, we derive the burst reorder metrics of Section 2 for two different burst departure processes. First, we consider a time-based assembly scheme with an approximate constant inter-arrival time. Second, we consider a random selection assembly scheme with a burst departure process showing Poisson characteristics.

4.1. Time-based assembly strategy

The burst departure process of a time-based assembly strategy shows a constant inter-departure time Δ if the packet arrival rate is sufficiently large [2]. We express the network delay by integer multiples of the inter-departure time as delays smaller than the inter-departure time would not cause any reordering. We identify the bursts by their sequence number s. As the burst delay is proportional to the constant inter-arrival time Δ, we abbreviate a delay of d Δ time units by d bursts. The next three sections calculate the reordering metrics of Section 2.

4.1.1. Burst reordering probability

According to Section 2, the test burst arrives out-of-sequence if the following condition holds: At the destination there is at least one burst arrival with sequence number $s > 0$ prior to the test burst. Consequently, the reordering probability is a joint probability that the test burst follows an abstract link $i > 0$ (a) and that there is at least one burst arrival with larger sequence number than zero before the test burst (b). Condition (a) assumes an arbitrary burst delay of d_t. In this situation, there are d_t candidate bursts, which may accomplish condition (b).

We derive the probability of (b) by the complementary probability, that none of the d_t bursts arrives earlier than the test burst. The random variable of the test burst delay is D_t. The random variable of a burst arrival before the test burst is B. Then the probability that the candidate burst j, $0 < j \leq d_t$ does not accomplish condition (b) is $\mathrm{P}(B = 0 \mid D_t = d_t, J = j) = \sum_{k=d_t-j+1}^{m} p_k$. The sum of probabilities represents all possible abstract links, which lead to a later arrival than the test burst. This sum considers the probabilities of the abstract link delays as well as the location of the burst j.

The joint probability that none of the candidate bursts accomplishes condition (b) at the same time is $\mathrm{P}(B = 0 \mid D_t = d_t) = \prod_{j=1}^{d_t} \mathrm{P}(B = 0 \mid D_t = d_t, J = j)$. The product consists of a joint probability all bursts for a later arrival than the test burst. If $\mathrm{P}(B = 0 \mid D_t = d_t)$ does not accomplish condition (b) (as all d_t bursts arrive later than the test burst), then due to the law of total probability the complementary probability does. The reordering probability results in

$$P = \sum_{d_t=1}^{m} p_{d_t} \left(1 - \mathrm{P}(B = 0 \mid D_t = d_t)\right) = \sum_{d_t=1}^{m} p_{d_t} \left(1 - \prod_{j=1}^{d_t} \sum_{k=d_t-j+1}^{m} p_k\right). \tag{2}$$

The outer sum considers all possible abstract links of the test burst with the corresponding probability. Within the brackets, the complementary distribution considers the joint probability of no burst arrival before the test burst.

The reordering probability equals the reordering ratio, which denotes the ratio of packets arriving out-of-sequence.

4.1.2. Reordering extent metric

In this section, we calculate the probability density function (pdf) of the reordering extent. The extent equals the number of burst arrivals between the located burst and the test burst. According to the definition in Section 2, the located burst has the smallest sequence number greater than the test burst arriving prior to the test burst.

According to this definition, the sequence number of the test burst is $s = f$ where $0 < f \leq d_t$. The located burst may also follow an abstract link of length d_l, where the delay obeys the following inequality as the located burst always arrives before the test burst: $0 \leq d_l < d_t + f$. The bursts with sequence number $0 < s < f$ and the bursts with sequence number $f + 1 < s \leq f + d_l$ in the case of a delayed located burst, arrive earlier than the located burst.

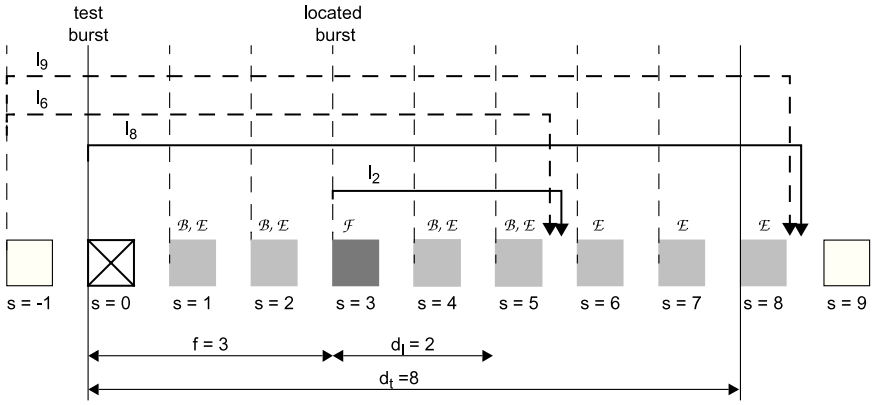

Fig. 4. Time-based assembly strategy.

Fig. 4 depicts this scenario for a test burst delay of $d_t = 8$ and the located burst $f = 3$, which is delayed by $d_l = 2$. Note, that we extend the actual delay in the figure to denote the order of the burst arrivals in case two bursts arrive after the same burst. If $f = 3$ becomes the located burst, it has to satisfy the condition of a located burst according to Section 2. In this example, the bursts with sequence numbers 1 and 2 as well as 4 and 5 need to arrive later then burst 3. The first two bursts because of their smaller sequence number and the second two bursts because of the condition of the smallest burst index with a larger sequence number than the test burst.

For evaluation, we need to distinguish these different cases. Therefore, we define three different random events. The figure also shows these random events:

- random event $\mathfrak{F}$ applies to the located burst only,
- random event $\mathfrak{E}$ applies to bursts arriving later than the located burst and prior to the test burst and thus define the extent,
- random event $\mathfrak{B}$ applies to bursts, which have to arrive later than the located burst due to the necessary condition of the located burst.

According to these random events we classify the bursts with respect to their sequence number (cf. Fig. 4).

$s < 0$	bursts with sequence number smaller than 0 may arrive later than the located burst and thus contribute to the extent. Event $\mathfrak{E}$ applies.
$0 < s < f$	for bursts with sequence number smaller than f but larger than zero, both events $\mathfrak{E}$ and $\mathfrak{B}$ apply. 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.
$f < s \leq f + d_l$	if the located burst f is delayed, too, the events $\mathfrak{E}$ and $\mathfrak{B}$ apply for the bursts between the located burst and the test burst.
$f + d_l < s \leq d_t$	bursts which originally arrive later than the located burst but prior to the test burst contribute to the extent. Event $\mathfrak{E}$ applies.

The next sections concentrate on the derivation of the probabilities of these events.

4.1.2.1. Random event $\mathfrak{E}$. Bursts with sequence number $s \leq d_t$ contribute to the extent if they arrive later than the located burst and prior to the test burst.

The probability of a burst with sequence number s arriving later than the located burst and prior to the test burst depends on three different properties:

- its location S,
- the delay of the test burst D_t,
- 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 $\mathrm{P}(B = 1 \mid S = s, D_t = d_t, F = f, D_l = d_l)$. 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 $p_{1s}(d_t, f, d_l)$. It evaluates in (3) the sums of probabilities, where each sum represents the probability to arrive later than the

located burst but prior to the test burst.

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

For instance, in Fig. 4, we consider the burst with sequence number $s = -1$. The probability that this burst arrives after the located burst $f = 3$, which follows l_{d_l} ($d_l = 2$) is $p_7 + p_8$. If burst $s = -1$ follows link l_6, it arrives prior to the located burst. If burst $s = -1$ follows link l_9, it arrives later than the test burst.

4.1.2.2. Random event $\mathfrak{B}$. Random event $\mathfrak{B}$ applies to 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 $\mathrm{P}(\mathfrak{B}) = 1 - \mathrm{P}(\bar{\mathfrak{B}})$.

Thereby, $\mathrm{P}(\bar{\mathfrak{B}})$ denotes the probability of a burst arrival for a specific burst prior to the located burst. This probability depends on the origin location S of the burst and the located burst F and its delay D_l. $Q(B = 1 \mid S = s, F = f, D_l = d_l)$ 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, which apply random event $\mathfrak{B}$.

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

This again results in sums of probabilities indicating a non-arrival before the test burst.

4.1.2.3. Conditional random events $\mathfrak{B}$ *and* $\mathfrak{E}$. Event $\mathfrak{B}$ is a necessary condition for the bursts with a smaller sequence number than the located burst. These bursts also apply event $\mathfrak{E}$ as depicted in Fig. 4. This results in a conditional probability for these bursts contributing to the extent. The probability that these bursts contribute to the extent (event $\mathfrak{E}$) is conditioned by event $\mathfrak{B}$. We calculate this conditional probability:

$$\mathrm{P}(\mathfrak{E}|\mathfrak{B}) = \frac{\mathrm{P}(\mathfrak{B}, \mathfrak{E})}{\mathrm{P}(\mathfrak{B})} = \frac{\mathrm{P}(\mathfrak{E})}{\mathrm{P}(\mathfrak{B})} = \frac{\mathrm{P}(\mathfrak{E})}{1 - \mathrm{P}(\bar{\mathfrak{B}})}. \tag{5}$$

The joint probability $\mathrm{P}(\mathfrak{B}, \mathfrak{E})$ is equal to the probability $\mathrm{P}(\mathfrak{E})$ as event $\mathfrak{E}$ includes random event $\mathfrak{B}$ as well. With the previous expressions $p_{1s}(d_t, f, d_l)$ from (3) and $q_{1s}(f, d_l)$ from (4) we get the conditional probability:

$$p^*_{1s}(d_t, f, d_l) = \frac{p_{1s}(d_t, f, d_l)}{q_{1s}(f, d_l)}. \tag{6}$$

4.1.2.4. Random event $\mathfrak{F}$. Each of the bursts with sequence number s in $0 < s \le d_t$ 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 sequence number $0 < s < f$ later than the located burst f. This necessary probability depends on the position/sequence number F and the delay D_l of the located burst. The necessary

condition for the located burst is:

$$\begin{aligned}\mathrm{P}(\mathfrak{F} \mid F=f, D_l=d_l) &= \prod_{s=1}^{d_l+f} Q(B=1|S=s, F=f, D_l=d_l)\\ &= \prod_{s=1}^{d_l+f} q_{1s}(f, d_l). \end{aligned} \tag{7}$$

This joint probability requires a later arrival of bursts, which depart earlier than the located burst. It considers all bursts with sequence numbers between 1 and the sequence number of the located burst plus its delay.

4.1.2.5. Reordering extent. With the above probability distributions for the different random events, we calculate the reordering extent distribution.

We denote $\mathrm{P}(E=e \mid D_t=d_t, F=f, D_l=d_l)$ the probability of E burst arrivals between the located burst and the test burst. Therein, the conditions are the delay of the test burst D_t, a located burst at position F with a delay D_l.

The next step considers the probability of every potential burst to contribute to the extent. The burst arrivals prior to the test burst and after the located burst are independent of each other. The composite of the number of burst arrivals forming the extent is a joint probability experiment. The discrete convolution of all bursts leads to the estimated probability of above. For the calculation of the convolution, we apply the probability generating function (GF).

We denote $\mathrm{P}(B=1|S=s, D_t=d_t, F=f, D_l=d_l)$ the probability that burst s contributes to the extent. The probability generating function for this distribution becomes:

$$\begin{aligned} G_{s,d_t,f,d_l}(z) &= \sum_{i=0}^{1} p_{is}(d_t, f, d_l)\, z^i\\ &= \begin{cases} p_{0s}(d_t, f, d_l) + p^*_{1s}(d_t, f, d_l)z, & \text{if } 0 < s \le f + d_l\\ p_{0s}(d_t, f, d_l) + p_{1s}(d_t, f, d_l)z, & \text{otherwise.}\end{cases}\end{aligned} \tag{8}$$

The GF of the distribution of burst arrivals after the located burst and prior the test burst is determined by the product of the GFs of all bursts arriving prior to the test burst.

$$G_{d_t,f,d_l}(z) = \prod_{s=f+d_l-m}^{d_t} G_{s,d_t,f,d_l}(z). \tag{9}$$

We derive the corresponding probability distribution function $\mathrm{P}(E=e \mid D_t=d_t, F=f, D_l=d_l)$ by the derivation of the GF of the joint experiment:

$$\mathrm{P}(E=e \mid D_t=d_t, F=f, D_l=d_l) = \frac{1}{e!}\frac{\partial^e}{\partial z^e} G_{d_t,f,d_l}(z)\bigg|_{z=0}. \tag{10}$$

The computational effort of the product in (9) and the derivation in (10) is relaxed by two less expensive steps. In (10) the eth derivation gives us the eth coefficient of the polynomial $G_{d_t,f,d_l}(z)$. We calculate this coefficient in (9) by applying the Cauchy product.

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:

$$\mathrm{P}(E=e) = \sum_{d_t=1}^{m}\sum_{f=1}^{d_t}\sum_{d_l=0}^{(d_t-f-1)^+} p_{d_t} p_{d_l} \mathrm{P}(\mathfrak{F} \mid F=f, D_l=d_l)\mathrm{P}(E=e-1 \mid D_t=d_t, F=f, D_l=d_l). \tag{11}$$

The outer sum represents the possible delay D_t of the test burst. The middle sum represents the position of the located burst F. The inner sum represents the delay D_l of the located burst. The three sums enclose a product of four factors. The first factor denotes the delay probability of the test burst. The second factor denotes the delay probability of the located burst. The third factor represents the conditional probability of the located burst (7). The last factor quantifies the probability of $e-1$ burst arrivals between the located burst and the test burst (10). The located burst itself accounts to the overall extent e.

4.1.3. 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.

From this definition we derive two conditions: (a) the test burst receives an extra delay and (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 located burst with sequence number f. 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 $\mathrm{P}_{s>0}(B=n_r-1 \mid D_t=d_t, F=f,$

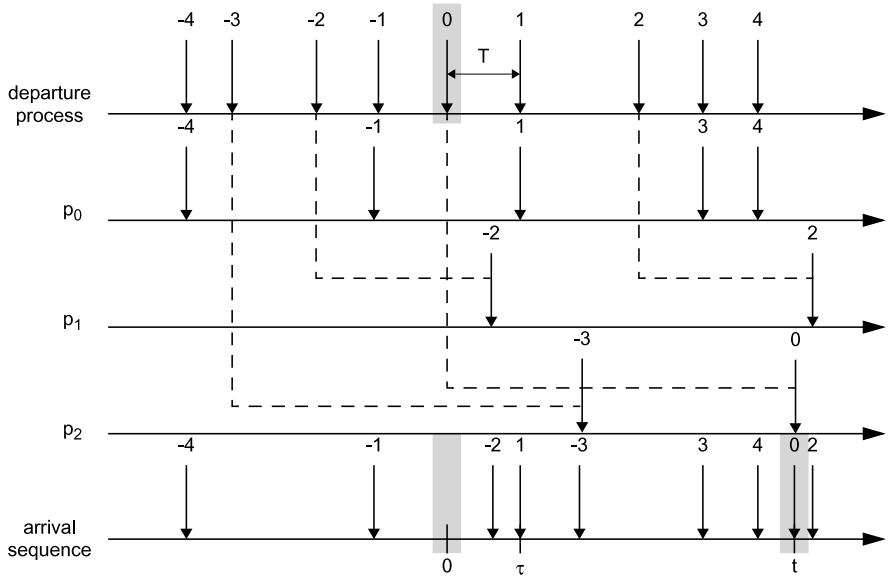

Fig. 5. Random selection assembly, Poisson splitting process for $m = 2$.

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

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, the delay of the test burst D_t, the located burst F and its delay D_l. (3) gives the individual probability. Applying (8)–(10), we derive both probabilities $\mathrm{P}_{s<0}$ and $\mathrm{P}_{s>0}$.

$$\mathrm{P}_{s<0}(B = 0 \mid D_t = d_t, F = f, D_l = d_l) = G_{s<0,d_t,f,d_l}(0) \tag{12}$$

$$G_{s<0,d_t,f,d_l}(z) = \prod_{s=f+d_l-m}^{-1} G_{s,d_t,f,d_l}(z) \tag{13}$$

$$\mathrm{P}_{s>0}(B = n_r - 1 \mid D_t = d_t, F = f, D_l = d_l) = \frac{\partial^{n_r-1}}{\partial z^{n_r-1}} \frac{G_{s>0,d_t,f,d_l}(z)}{(n_r-1)!}\bigg|_{z=0} \tag{14}$$

$$G_{s>0,d_t,f,d_l}(z) = \prod_{s=1}^{d_t} G_{s,d_t,f,d_l}(z). \tag{15}$$

(12) gives the probability of no burst arrival between the located burst and the test burst for bursts with sequence numbers $s < 0$ and (14) gives the probability of exactly $n_r - 1$ burst arrivals with a larger sequence number than the test burst between itself and the located burst.

Putting both results together, leads to the ccdf of the n_r-reordering metric of (16). The structure is similar to (11) except of the dependence on the sequence number of the burst arrivals.

$$\begin{aligned} P(N_r \geq n_r) = \sum_{d_t=1}^{m} \sum_{f=1}^{d_t} \sum_{d_l=0}^{(d_t-f-1)^+} & p_{d_t} p_{d_l} \mathrm{P}_{s>0}(B = n_r - 1 \mid D_t = d_t, F = f, D_l = d_l) \\ & \times \mathrm{P}_{s<0}(B = 0 \mid D_t = d_t, F = f, D_l = d_l). \end{aligned} \tag{16}$$

4.2. Random selection assembly strategy

In this section, we derive the reordering metrics for the random selection assembly strategy resulting in a Poisson departure process. The next paragraphs determine some general properties of the reordering model with focus on the Poisson process. Then, the following sections derive explicitly the reordering metrics. We apply the same reordering model than for constant inter-arrival times of the previous section.

In Fig. 5, we depict the point process of the burst departure process. On the first time axis, we show the origin burst departure process with the test burst $s = 0$. The random variable of the departure time is T. The probability distribution function of this inter-departure time is $f(t) = \lambda \exp(-\lambda t)$, with mean rate $\lambda = 1/\mathsf{E}[T]$.

On the second time axis, we show the burst arrival process of bursts following the abstract link l_0. Each of the remaining time axis depict the point process of bursts following abstract links l_1 and l_2 respectively. The superposition of abstract links $l_0 \cdots l_2$ leads to the experienced arrival order at the destination in the bottom time axis.

Joining an abstract link l_j with probability p_j forms a random splitting process. As a result, the burst inter-arrival times T_j on link l_j are also negative exponentially distributed [29] and form a Poisson process. Consequently, the inter-departure time distribution for link l_j becomes $f_j(t) = \lambda_j \exp\left(-\lambda_j t\right)$, where $\lambda_j = p_j\lambda$.

On the time axis of the arrival process, we define two points in time (cf. Fig. 5). The first one, the origin, corresponds to the arrival of the test burst $s = 0$ at the destination. The second one is an arbitrary point in time t, $t > 0$. The sequence numbers of bursts departing before the test burst are $s < 0$, while the sequence number of bursts departing after the test burst are $s > 0$.

We divide the interval $[0, t]$ on the arrival axis as well as on the abstract links l_j into two intervals $[0, \tau)$ and $[\tau, t]$. We calculate the probability distribution function of the random number X_j of burst arrivals on link l_j in the time interval $[\tau, t]$.

Bursts originally departing in the interval $[\tau - j\,\Delta, t - j\Delta]$ arrive in the interval $[\tau, t]$ if they follow link l_j. The pdf of the number of the number of burst arrivals in $[\tau, t]$ follows a Poisson distribution:

$$p_{x_j,\lambda_j}^{[\tau,t]} = \frac{(\lambda_j(t-\tau))^{x_j}}{x_j!} \exp\left(-\lambda_j(t-\tau)\right). \tag{17}$$

At the destination, the bursts on the individual abstract links are superposed together. We denote $\mathbb{J}$ as the set of abstract links superposing the bursts at the destination. The superposition of burst arrivals within the interval $[\tau, t]$ forms a compound experiment, where the random variable of burst arrivals is a sum of random variables $X = \sum_{j\in\mathbb{J}} X_j$.

The probability distribution function of X is the result of the discrete convolution of the probability density functions of all X_j. For the convolution, we apply the probability generating function on X_j. The GF of $p_{x_j,\lambda_j}^{[\tau,t]}$ and for compound arrival are:

$$G_j^{[\tau,t]}(z) = \sum_{i=0}^{\infty} p_{i,\lambda_j}^{[\tau,t]} z^i \qquad G^{[\tau,t]}(z) = \prod_{j\in\mathbb{J}} G_j^{[\tau,t]}(z). \tag{18}$$

The derivation of the GF identifies the probability density function $p_x^{[\tau,t]}$ of the random variable X of the compound experiment:

$$p_x^{[\tau,t]} = \frac{1}{x!}\frac{\partial^x}{\partial z^x} G^{[\tau,t]}(z)\bigg|_{z=0}. \tag{19}$$

The derivation of the reordering metrics, especially, the n_r-reordering metric requires the distinction of the bursts according to their sequence number.

In the observed interval $[\tau, t]$, bursts may arrive with sequence number $s > 0$ and $s < 0$ dependent on the abstract link l_j and the size of the interval $[\tau - j\Delta, t - j\Delta]$. We distinguish three cases:

(1) If $0 < \tau - j\Delta \wedge 0 < t - j\Delta$ bursts with sequence number $s < 0$ do not arrive in the interval $[\tau, t]$.
(2) If $0 > \tau - j\Delta \wedge 0 < t - j\Delta$ bursts with sequence number $s < 0$ as well as with $s > 0$ arrive in the interval $[\tau, t]$.
(3) If $0 > \tau - j\Delta \wedge 0 > t - j\Delta$ bursts with sequence number $s > 0$ do not arrive in the interval $[\tau, t]$.

For bursts with $s < 0$, we assume L_j bursts, for bursts with $s > 0$, we assume K_j bursts in $[\tau, t]$. The pdfs of L_j and K_j for the previous three cases are

$$p_{L_j,\lambda_j,s<0}^{[\tau,t]} = \begin{cases} p_{L_j,\lambda_j}^{[\tau,t]}, & \text{if } \tau - j\Delta < 0 \wedge t - j\Delta < 0 \\ p_{L_j,\lambda_j}^{[\tau-j\Delta,0]}, & \text{if } \tau - j\Delta < 0 \wedge t - j\Delta > 0 \\ 0, & \text{otherwise} \end{cases} \tag{20}$$

$$p_{K_j,\lambda_j,s>0}^{[\tau,t]} = \begin{cases} p_{K_j,\lambda_j}^{[\tau,t]}, & \text{if } \tau - j\Delta > 0 \wedge t - j\Delta > 0 \\ p_{K_j,\lambda_j}^{[0,t-j\Delta]}, & \text{if } \tau - j\Delta < 0 \wedge t - j\Delta > 0 \\ 0, & \text{otherwise.} \end{cases} \tag{21}$$

The different alternatives take into account the abstract link number and the position of τ and t. As a result, they modify the interval and apply (17).

Applying the method of the pdf for the compound event of multiple branches of set $\mathbb{J}$ for $s > 0$ and $s < 0$, we get $p_{L,s<0}^{[\tau,t]}$ and $p_{K,s>0}^{[\tau,t]}$, the pdf of the number of burst arrivals within $[\tau, t]$ for bursts with $s < 0$ and $s > 0$.

$$p_{L,s<0}^{[\tau,t]} = \frac{1}{L!}\frac{\partial^L}{\partial z^L} G_{s<0}^{[\tau,t]}(z)\bigg|_{z=0} = \frac{1}{L!}\frac{\partial^L}{\partial z^L}\left\{\prod_{j\in\mathbb{J}}\left(\sum_{i=0}^{\infty} p_{i,\lambda_j,s<0}^{[\tau,t]} z^i\right)\right\}\bigg|_{z=0} \tag{22}$$

$$p_{K,s>0}^{[\tau,t]} = \frac{1}{K!}\frac{\partial^K}{\partial z^K} G_{s>0}^{[\tau,t]}(z)\bigg|_{z=0} = \frac{1}{K!}\frac{\partial^K}{\partial z^K}\left\{\prod_{j\in\mathbb{J}}\left(\sum_{i=0}^{\infty} p_{i,\lambda_j,s>0}^{[\tau,t]} z^i\right)\right\}\bigg|_{z=0}. \tag{23}$$

In the next sections, we apply the previous results to determine the introduced reordering metrics.

4.2.1. Burst reordering probability

The probability of an out-of-sequence arrival is the compound probability to follow link l_{d_t} (with probability p_{d_t}) and the probability of at least one burst arrival in the interval $[0, d_t\Delta]$. We derive the latter probability by the complementary probability of no burst arrival and apply (23):

$$P = \sum_{d_t=1}^{m} p_{d_t} \left(1 - p_{0,s>0}^{[0,d_t\Delta]}\right). \tag{24}$$

In (23) the set $\mathbb{J}$ of considered links l_j is $\mathbb{J} = \{o : 0 \le o < d_t\}$. Bursts with $s > 0$ on links with a larger delay than $d_t\Delta$ do not arrive within $[0, t]$. As we requested the 0th derivative of (23) we simplify (24):

$$\begin{aligned} P &= \sum_{d_t=1}^{m} p_{d_t} \left(1 - \prod_{j=0}^{d_t-1} p_{0,\lambda_j}^{[0,(d_t-j)\Delta]}\right) = \sum_{d_t=1}^{m} p_{d_t} \left(1 - \prod_{j=0}^{d_t-1} \exp\left(-\lambda_j\Delta(d_t-j)\right)\right) \\ &= \sum_{d_t=1}^{m} p_{d_t} \left(1 - \exp\left(-\lambda\Delta \sum_{j=0}^{d_t-1} p_j(d_t-j)\right)\right). \end{aligned} \tag{25}$$

4.2.2. Reordering extent metric

In this section, we calculate the pdf of the reordering extent metric. According to the extent metric definition, we calculate the reordering extent in three steps: (a) identify the located burst with the smallest j as defined in (1). (b) Count the number of burst arrivals between the located burst and the test burst.

The delay of the test burst is a necessary condition of (a). Without loss of generality, we assume the test burst following link l_{d_t} receiving a delay of $d_t\Delta$. The located burst follows abstract link l_{d_l}, $0 \le d_l < d_t$, and arrives at $\tau = d_l\Delta + t$. t is the time between the departure of the located burst and the departure of the test burst. Consequently, the co-domain of τ is $t + d_l\Delta \le \tau \le d_t\Delta$. The probability of a burst arrival at τ is a compound probability of a burst departing at t and the probability to follow link l_{d_l}: $\mathrm{P}(B = 1 \mid D_l = d_l, t \le T < t + \mathrm{d}t)$, with B indicating a burst arrival.

The second part of this joint probability results in the instantaneous termination rate $\lambda = 1/\mathsf{E}[T]$.

$$\mathrm{P}(B = 1 \mid D_l = d_l, t \le T < t + \mathrm{d}t) = p_{d_l} \frac{\mathrm{P}(t \le T \le t + \mathrm{d}t)}{\mathrm{d}t}\mathrm{d}t \tag{26}$$

$$\lim_{\mathrm{d}t\to 0} \frac{\mathrm{P}(t \le T \le t + \mathrm{d}t)}{\mathrm{d}t} = \lim_{\mathrm{d}t\to 0} \frac{1 - \mathrm{e}^{-\lambda\,\mathrm{d}t}}{\mathrm{d}t} = \lambda. \tag{27}$$

The necessary condition of (a) restricts any burst arrival in $[0, \tau]$ with sequence numbers smaller than the located burst, i.e., bursts departed in $(0, t)$. These bursts must not follow abstract link l_j with $j < t/\Delta + d_l$ as they would arrive earlier to the located burst. The probability of this restriction again is a compound probability applying (23) with no burst arrival on the abstract links l_j with $j \in \{o : 0 \le o \le \hat{j} = \lfloor t/\Delta + d_l \rfloor\}$.

$$\mathrm{P}_{s>0}(B = 0 \mid 0 < T < \tau) = p_{0,s>0}^{[0,\tau]} = \prod_{j=0}^{\hat{j}} p_{0,\lambda_j,s>0}^{[0,\tau-j\Delta]}. \tag{28}$$

In the interval $[\tau, d_t\Delta]$, an arbitrary number of burst arrivals form the extent value. This is a compound probability based on the probability distribution function $p_{x_j,\lambda_j}^{[\tau,t]}$ (cf. (17)) of each abstract link j.

We apply (18) with $\mathbb{J} = \{o : 0 \le o \le m\}$ for the required number of $e - 1$ arrivals in the considered interval. The located burst completes the extent value to e:

$$P(B = e - 1 \mid \tau < T \le d_t\Delta) = p_{e-1}^{[\tau,t]} = \frac{1}{(e-1)!} \frac{\partial^{e-1}}{\partial z^{e-1}} G^{[\tau,d_t\Delta]}(z)\bigg|_{z=0}. \tag{29}$$

The compound of these necessary three conditions leads to the probability distribution function of the extent. We consider each possible delay of the test burst and of the located burst. We consider the located burst at every possible point in time t. We apply (26), (28), (29) and receive the probability distribution function of the burst extent metric.

$$\begin{aligned} \mathrm{P}(E = e) = \sum_{d_t=1}^{m} p_{d_t} \sum_{d_l=0}^{d_t-1} \int_0^{(d_t-d_l)\Delta} & \mathrm{P}(B = 0 \mid 0 < T < \tau) \\ & \times \mathrm{P}(B = 1 \mid t \le T < t + \mathrm{d}t)\mathrm{P}(B = e - 1 \mid \tau < T < d_t\Delta). \end{aligned} \tag{30}$$

Within the integral, the first factor denotes the probability of the condition of the located burst with no burst arrival before it. The second factor gives the probability of the located burst arrival exactly at time t, while the last factor determines the probability for the required burst arrivals in the interval between located burst and test burst. With help of (29) and (28),

we simplify the extent probability to:

$$\mathrm{P}(E = e) = \sum_{d_t=1}^{m} p_{d_t} \sum_{d_l=0}^{d_t-1} p_{d_l} \lambda \int_0^{(d_t-d_l)\Delta} p_{0,s>0}^{[0,\tau]} p_{e-1}^{[\tau,t]} \mathrm{d}t. \tag{31}$$

4.2.3. n_r-reordering metric

The n_r-reordering metric requires after the located burst consecutive arrivals of bursts with sequence numbers larger than the sequence number of the test burst. Again, we assume the located burst at τ. We also assume a test burst delay of $d_t\Delta$.

The ccdf of the n_r-reordering metric is a joint probability of two conditions: (a) arrival of the located burst at τ, (b) arrival of $n_r - 1$ bursts with appropriate sequence number in the interval $[\tau, d_t\Delta]$. Condition (b) forbids any burst arrival in $[\tau, d_t\Delta]$ with sequence number $s < 0$.

From the previous section, the probability of (a) is $\mathrm{P}(B = 1 \mid t \le T \le t + \mathrm{d}t)$. We derive the probability of (b) by the probability GF of the compound probability of the number of arrivals. We apply (23) for the bursts with $s > 0$ and (22) for the bursts with $s < 0$:

$$\mathrm{P}_{s>0}(B = n_r - 1 \mid \tau < T \le d_t\Delta) = p_{n_r-1,s>0}^{[\tau,d_t\Delta]} \tag{32}$$

$$\mathrm{P}_{s<0}(B = 0 \mid \tau < T \le d_t\Delta) = p_{0,s<0}^{[\tau,d_t\Delta]}. \tag{33}$$

Similar to (30), the overall ccdf of the n_r-reordering metric becomes:

$$\begin{aligned}\mathrm{P}(N_r \ge n_r) = \sum_{d_t=1}^{m} p_{d_t} \sum_{d_l=0}^{d_t-1} \int_0^{d_t\Delta} & \mathrm{P}_{s<0}(B = 0 \mid \tau < T \le d_t\Delta) \\ & \times \mathrm{P}(B = 1 \mid t \le T \le t + \mathrm{d}t)\mathrm{P}_{s>0}(B = n_r - 1 \mid \tau < T \le d_t\Delta).\end{aligned} \tag{34}$$

Applying (29) and (28) leads to the more compact version:

$$\mathrm{P}(N_r \ge n_r) = \sum_{d_t=1}^{m} p_{d_t} \sum_{d_l=0}^{d_t-1} \lambda p_{d_l} \int_0^{d_t\Delta} p_{n_r-1,s>0}^{[\tau,d_t\Delta]} p_{0,s<0}^{[\tau,d_t\Delta]} \mathrm{d}t. \tag{35}$$

5. Results

In this section, we first illustrate the reordering metrics of selected parameterizations of the model and second show its applicability on OBS network simulations. Thereby, we determine the OBS reordering characteristic for the parameterization of our model.

5.1. Numerical results

We show some illustrative results to visualize the burst reordering metrics. We parameterize our model by

- 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 comprehensive studies, we distinguished three different delay distributions:

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

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 Fig. 6, we depict the amount of out-of-sequence bursts in relation to the delay probability p. We depict the probability for both assembly strategies. $m = 5$ and $m = 15$ abstract links distribute the delay geometrically. The amount of reordering is higher in case of more abstract links, although the probability of delay is the same. The chance for an out-of-sequence arrival is higher, if there are more alternative paths to follow. We recognize a bell shaped curve, which starts at zero but did not reach zero at the other end due to the different delays on the abstract links.

In Fig. 7, 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 behaviour. The linear distribution is straight decreasing as smaller extent values are more likely than larger ones. The complement linear distribution is bell-shaped as its maximum is moved towards larger extent values. The geometric distribution start between both distributions and decreases only slightly until its knee at $e = 5$.

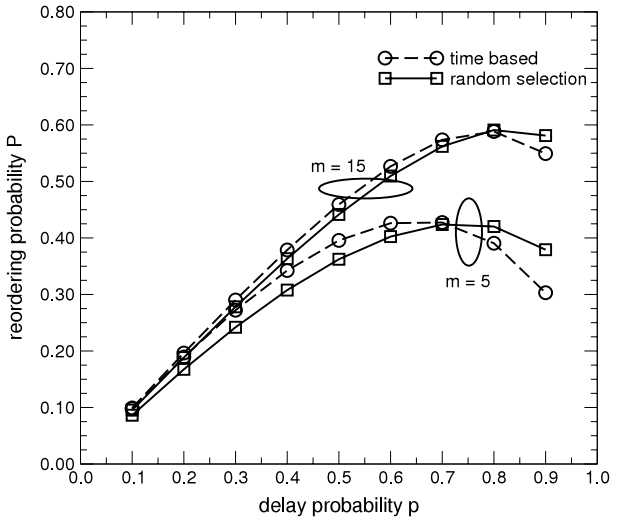

Fig. 6. Reordering probability $m = 5$, $m = 15$, geometric delay distribution.

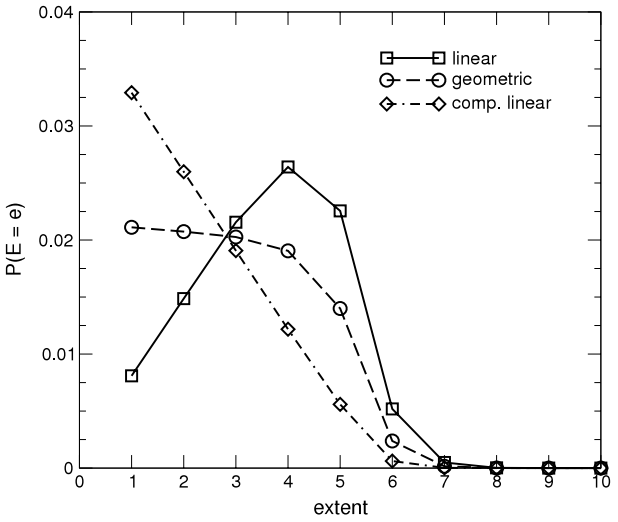

Fig. 7. Extent pdf, $m = 5$, $p = 0.1$, time-based assembly.

In Fig. 8, we illustrate the burst extent pdf for the random selection based assembly scheme with equivalent parameters. The earlier observations also apply in this scenario, but in a less extreme way. The three graphs are close together and show a smaller extent probability for small e. In case of $e > 5$ the random selection assembly strategy shows a larger extent value than the time-based assembly strategy.

In Figs. 9 and 10, we plotted the ccdf of the n_r-reordering metric for both assembly strategies. In case of a linear distributed delay we expect the largest amount of n_r-reordering. Is the delay complementarily linear distributed we expect the smallest amount of n_r-reordering. For small values of n_r the ccdf of the time-based assembly strategy in all cases is larger than for a random selection assembly. For larger n_r this property swaps.

5.2. Application of the reordering model

This section shows the applicability of our reordering model on simulations on optical burst switched networks regarding the burst reordering phenomena. In our earlier work [18], we quantified burst reordering in an OBS network simulation. The paper presents the average values on burst reordering extent and on the burst n_r-reordering metric and pointed out metrics required for TCP throughput estimation. Here, we present the reordering extent distribution and the n_r-reordering distribution and show that our model abstracts the end-to-end reordering properties well.

We consider OBS network simulations, which provide the reordering metrics of certain end-to-end connections for a certain burst traffic. The experienced out-of-sequence pattern is the result of an unknown network delay distribution.

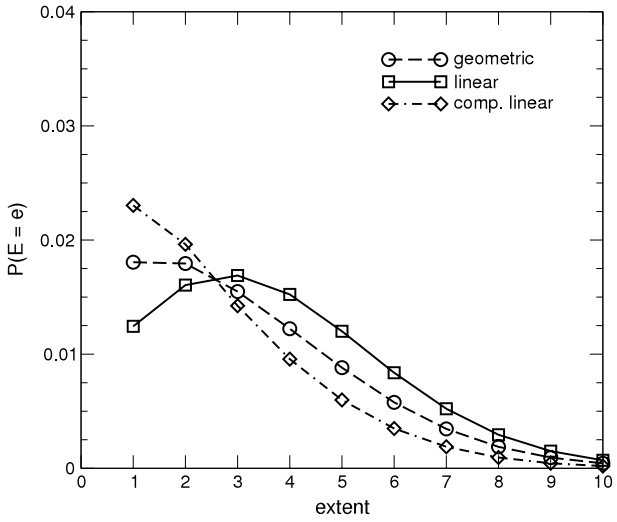

Fig. 8. Extent pdf, $m = 5$, $p = 0.1$, random selection assembly.

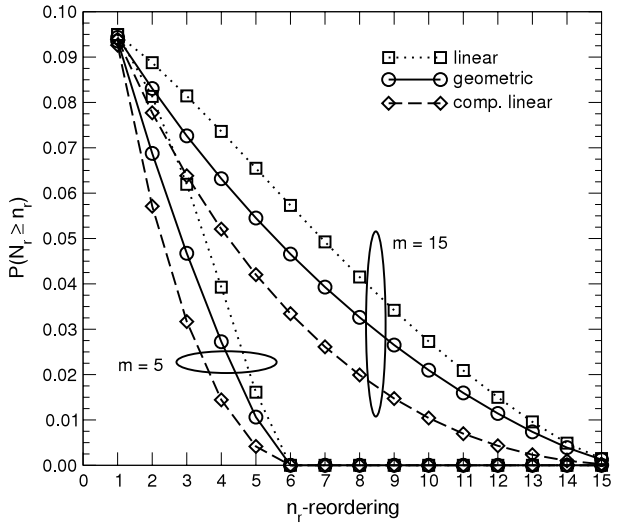

Fig. 9. n_r-reordering ccdf, $m = 5$, $p = 0.1$, time-based assembly.

The reordering model is able to emulate this out-of-sequence pattern but requires a network delay distribution. The aim of the next sections is to provide the methodology to obtain the parameters for the network delay distribution on a given out-of-sequence pattern.

In general, we apply the reordering model and the related reordering functions of the model showing constant inter-arrival times (cf. Section 4.1) independent of the original burst arrival characteristic. The reason lies in the out-of-sequence pattern for constant inter-arrival times. In [30], we proved that the reordering metric reach its maximum for constant inter-arrival times. Consequently, the model for constant inter-arrival times serves as an upper limit of the experienced out-of-sequence pattern.

5.2.1. Parameter estimation of the reordering model

The probability distribution function of the reordering extent and the n_r-reordering metric is discrete, non-linear and in an open form (cf. (11), (16), (31) and (35)). Additionally, these functions map an $m + 1$-dimensional input vector (network delay distribution) onto an n dimensional output vector (reordering metric) and shows highly non-linear operations. Abstracting these functions by f leads to the following generic equation:

$$\vec{y} = f(\vec{x}), \quad \text{where } \vec{x} \text{ is a probability vector.} \tag{36}$$

Therein, $\vec{y}$ represents either the pdf of the reordering extent or the ccdf of the n_r-reordering metric according to the model. $\vec{x}$ represents in either case the network delay distribution. If the vector degrades to a scalar, the equation and the function realize the reordering probability.

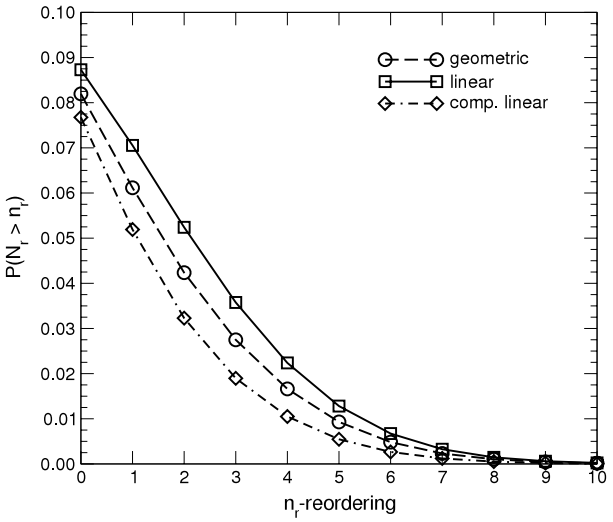

Fig. 10. n_r-reordering ccdf, $m = 5$, $p = 0.1$, random selection assembly.

A correct representation of the experienced reordering in an OBS network requires the configuration of the network delay parameter $\vec{x}$. This leads to the inversion of the reordering functions $\vec{x} = f^{-1}(\vec{y'})$, which is analytically very hard.

A pragmatic approach is to reformulate this problem to a constraint based optimization problem using the original functions for the reordering metrics. If $\vec{y'}$ denotes the estimated reordering vector, then (37) depicts the non-linear optimization problem with one single constraint:

$$\begin{aligned} &\min_{\vec{x}} \|\vec{y'} - f(\vec{x})\| \quad \text{where } \|\vec{a}\| = \sqrt{\sum_i a_i^2} \\ &2 = \sum x_i + \sum |x_i|. \end{aligned} \tag{37}$$

The original multi-dimensional function of (37) becomes a scalar function by applying the norm on the resulting vector. This maps an n-dimensional vector onto a 1-dimension scalar, which realizes an additional non-linearity. The second equation formulates the constraint $\vec{x}$ being a probability vector.

For these kinds of problems solvers exist. The Optimization Toolbox of Matlab [31] provides one of these solvers. The name of the applied solver is `fmincon`, which tries to find a minimum of a constrained nonlinear multivariable function. The solver applies the active set algorithm, which the following literature describes in depth: [32–35]. We apply this Matlab solver to determine the network delay distribution (abstract links) to experience the same reordering pattern than in an OBS end-to-end connection. We skip the details of the applied algorithm as they are out of the scope of this paper.

5.2.2. Simulation parameter

This section provides the background on the simulation environment, which provides the measurement data of the OBS burst reorder pattern. We use the OBS network simulation of [36], which bases on the event-driven simulation library SimLib [37].

The network model represents the 16-node COST 266 reference network (cf. Fig. 11) with equidistant nodes and link delays of 1 ms. The traffic matrix is population based and offers 9.9 Tbps to the network. The network has been dimensioned for the total of 9.9 Tbps and equivalent blocking probabilities on all links. We also consider different load scenarios represented by the parameter α. It reflects an over-provisioning factor, where $\alpha \geq 1$. For load variations, instead of decreasing the network traffic, we increase the network resources as described in [36]. In case of contention, 32 FDLs per node may avoid a burst loss. The burst departure process follows a Poisson process representing the random selection assembly strategy. Here, we consider the contention resolution scheme including wavelength conversion and FDL (ConvFDL) in the given order as described in [18].

We selected five arbitrary node pairs and showed the compliance of our model with the simulation results for an FDL scenario. The presented results are examples only, our model holds for any other node pairs, too. The solver of the previous section found the parameters of the model.

The Figs. 12–14 show the reordering extent metric for selected end-to-end connections of the reference network. The solver estimated the network delay distribution with the abstract link probabilities. In each of the figures, the results from the solver are compared to the simulation results. The figures show that the results from the solver match the simulation results very well in the core of the distribution. At the border on the left and to the right of each distribution the results differ slightly. In Fig. 12, these difference only occur at the far right of the distribution. There, the large confidence

Fig. 11. COST 266 reference network with 16 nodes.

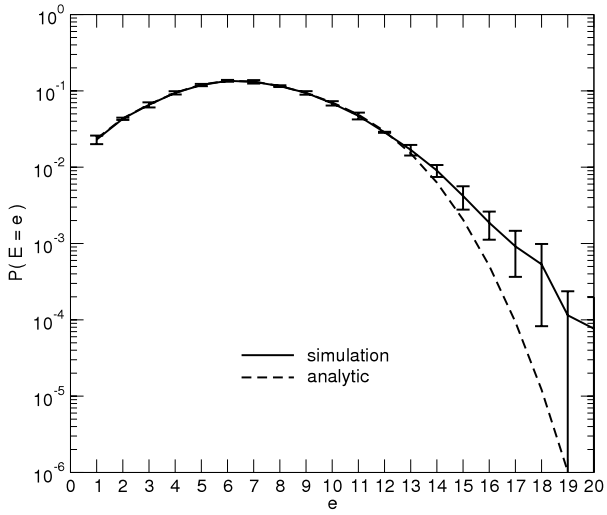

Fig. 12. Lyon–Milan, $\alpha = 1.0$, reordering extent.

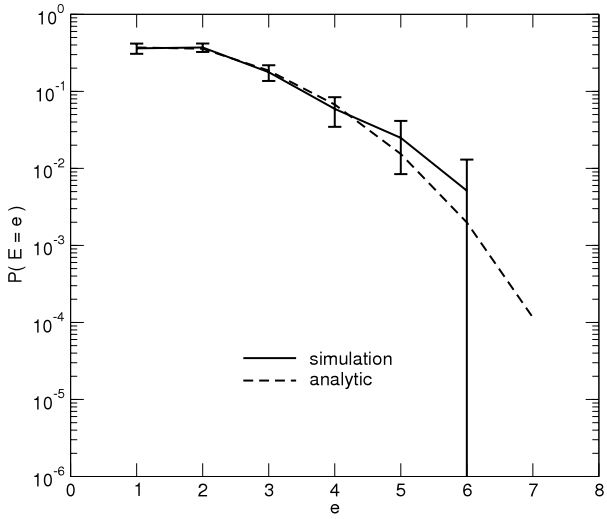

Fig. 13. Zagreb–Vienna, $\alpha = 1.2$, reordering extent.

intervals indicate, that these extent values occur rarely. For further analysis of any protocols in an network emulation environment this may have only a small impact. Figs. 12 and 13 show the extent pdf in relative values (the extent pdf of all reordered bursts), while Fig. 14 shows the distribution in absolute values for all bursts.

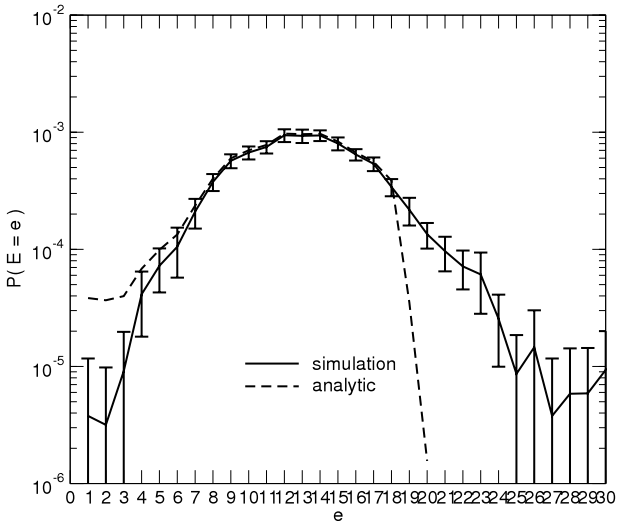

Fig. 14. Paris–Rome, $\alpha = 1.35$, reordering extent.

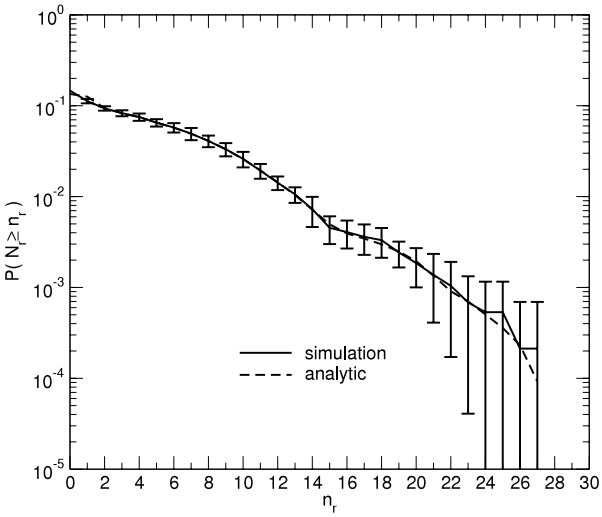

Fig. 15. Amsterdam–Lyon, $\alpha = 1.35$, n_r-reordering.

Figs. 15 and 16 show the results for the ccdf of the n_r-reordering metric of two end-to-end connections. The observations are similar to those given above. Analytics and simulations match quite well. For large values at the far end of the distribution, the values differ but lie within the confidence intervals of the simulation. The figure depicts the ccdf of the absolute values, showing also the reordering probability at the point $n_r = 0$ of the x-axis.

6. Conclusion

We proposed and analyzed a burst reordering model for two commonly applied burst assembly schemes. The time-based assembly strategies with a constant burst inter-departure time and the random selection assembly strategy with a negative exponentially distributed inter-departure time. We derived the three most important IETF reordering metrics. These metrics allow the dimensioning of the required OBS buffer capacity to resolve reordering and allow an estimation on the expected TCP throughput performance. We derived these metrics for an optical burst switching scenario and made no assumptions on the network delay distribution.

Our model enables a structured analysis on optical burst reordering. Investigations applying our reordering model cover a broader and deeper scope of burst reordering than an integrated network simulation is able to provide. A properly configured model substitutes optical network simulations on burst reordering. We showed its applicability exemplarily on selected links of a representative OBS network scenario. We found that our model reflects the network characteristics on burst reordering very well. The burst extent metric and the n_r-reordering metric both fit the simulation results.

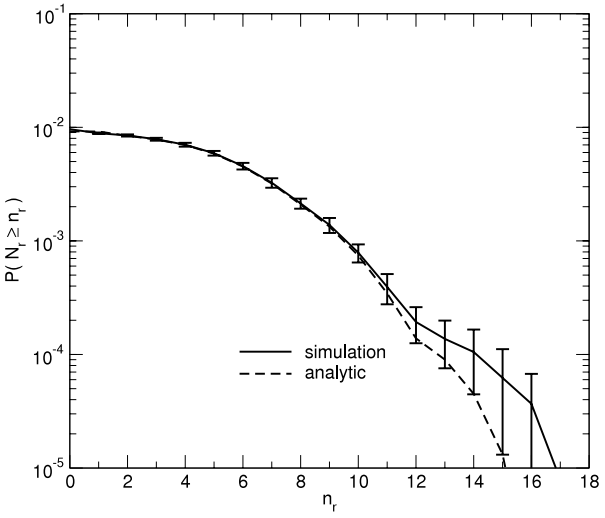

Fig. 16. Paris–Rome, $\alpha = 1.50$, n_r-reordering.

These results enable the parameterization of a network emulation environment creating the same reordering pattern as experienced in OBS network simulations. This network emulation setup is able to investigate real protocol implementations in presences of reordering, without modelling complex protocols. The parameters of the network emulation only require single layer OBS studies, which reduces the time for simulation and evaluation.

Acknowledgements

The author would like to thank Joachim and Michael Scharf as well as Guoqiang Hu for their valuable discussions and stimulating ideas. Additionally, anonymous reviewers also provided valuable feedback, the author would like to thank them for this.

References

[1] C. Qiao, M. Yoo, Optical burst switching (OBS)—a new paradigm for an optical Internet, Journal of High Speed Networks 8 (1) (1999) 69–84.
[2] M. de Vega Rodrigo, 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, 2004.
[3] K. Laevens, Traffic characteristics inside optical burst switched networks, in: Proceedings of the Optical Networking and Communications Conference, OptiComm, 2002.
[4] A. Rahbar, O. Yang, Contention avoidance and resolution schemes in bufferless all-optical packet-switched networks: a survey, IEEE Communications Surveys & Tutorials 10 (4) (2008) 94–107.
[5] C.M. Gauger, M. Köhn, 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, 2004.
[6] C.M. Gauger, M. Köhn, J. Scharf, 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.
[7] C.G. Argos, O.G. de Dios, J. Aracil, Adaptive multi-path routing for OBS networks, in: Proceedings of the 9th International Conference on Transparent Optical Networks, ICTON, vol. 3, 2007, pp. 299–302.
[8] L. Gharai, C. Perkins, T. Lehman, Packet reordering, high speed networks and transport protocol performance, in: Proceedings, 13th International Conference on Computer Communications and Networks, 2004, ICCCN 2004, 2004, pp. 73–78.
[9] M. Allman, V. Paxson, W. Stevens, TCP congestion control, RFC 2581, IETF, April 1999.
[10] R. Stewart, Stream control transmission protocol, RFC 4960, IETF, September 2007.
[11] H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson, RTP: a transport protocol for real-time applications, RFC 3550, IETF, July 2003.
[12] M. Mathis, J. Mahdavi, S. Floyd, A. Romanow, TCP selective acknowledgment options, RFC 2018, IETF, October 1996.
[13] F. Callegati, W. Cerroni, C. Raffaelli, Impact of optical packet loss and reordering on TCP performance, in: IEEE Global Telecommunications Conference, GLOBECOM'06, 2006, pp. 1–5.
[14] X. Yu, J. Li, X. Cao, Y. Chen, C. Qiao, Traffic statistics and performance evaluation in optical burst switched networks, IEEE/OSA Journal of Lightwave Technology 22 (12) (2004) 2722–2738.
[15] S. Gowda, R. Shenai, K. Sivalingam, 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.
[16] A. Detti, M. Listanti, Impact of segments aggregation on TCP Reno flows in optical burst switching networks, in: Proc. IEEE INFOCOM, 2002.
[17] F. Callegati, G. Muretto, C. Raffaelli, P. Zaffoni, W. Cerroni, A framework for performance evaluation of OPS congestion resolution, in: Proceedings of the IFIP Working Conference on Optical Network Design and Modelling, ONDM, 2005, pp. 242–249.
[18] J. Perelló, S. Gunreben, S. Spadaro, A quantitative evaluation of reordering in OBS networks and its impact on TCP performance, in: Proceedings of the IFIP Working Conference on Optical Network Design and Modelling, ONDM, 2008.
[19] M. Schlosser, E. Patzak, 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, Barcelona, vol. 1, 2005, pp. 220–223.
[20] M. Carson, D. Santay, NIST net: a Linux-based network emulation tool, SIGCOMM—Computer Communication Review 33 (3) (2003) 111–126.
[21] The Linux Foundation, Net:Em, The Linux Foundation, May 2009. URL: http://www.linuxfoundation.org/en/Net:Em.
[22] S. Gunreben, G. Hu, A multi-layer analysis on reordering in optical burst switched networks, IEEE Communications Letters 11 (12) (2007) 1013–1015.

[23] 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.
[24] S. Gunreben, An optical burst reordering model for a time-based burst assembly scheme, in: Proceedings of the International Workshop on Optical Burst/Packet Switching 2008, WOBS 2008, 2008.
[25] K. Dolzer, C.M. Gauger, J. Späth, S. Bodamer, Evaluation of reservation mechanisms for optical burst switching, AEÜ International Journal of Electronics and Communications 55 (1) (2001) 18–26.
[26] K. Dolzer, C.M. Gauger, On burst assembly in optical burst switching networks—a performance evaluation of just-enough-time, in: Proceedings of the 17th International Teletraffic Congress, ITC 17, Salvador, Brazil, 2001, pp. 149–160.
[27] H. Chaskar, S. Verma, R. Ravikanth, A framework to support IP over WDM using optical burst switching, in: Proceedings of the Optical Networks Workshop, Richardson, TX, 2000.
[28] A. Morton, L. Ciavattone, G. Ramachandran, S. Shalunov, J. Perser, Packet reordering metrics, RFC 4737, IETF, November 2006.
[29] P. Kühn, Approximate analysis of general queuing networks by decomposition, IEEE Transactions on Communications 27 (1) (1979) 113–126.
[30] S. Gunreben, O.G. de Dios, Why deterministic traffic shows the highest reordering ratio, in: Proceedings of the International Workshop on Optical Burst/Packet Switching, WOBS 2009, Madrid, 2009.
[31] MathWorks, Inc., Optimization Toolbox—User's Guide, 3rd ed., 2004.
[32] M.J.D. Powell, The Convergence of Variable Metric Methods For Nonlinearly Constrained Optimization Calculations, Academic Press, 1978.
[33] M.J.D. Powell, A Fast Algorithm for Nonlinearly Constrained Optimization Calculations, Vol. 630, Springer Verlag, 1978.
[34] S.P. Han, A globally convergent method for nonlinear programming, Journal of Optimization Theory and Applications 22 (1977) 297.
[35] P.E. Gill, Practical Optimization, Academic Press, 1981.
[36] C. Gauger, Novel network architecture for optical burst transport, Ph.D. Thesis, University of Stuttgart, Stuttgart, 2006.
[37] S. Bodamer, K. Dolzer, C. Gauger, M. Barisch, M. Necker, IKR simulation library 2.6 user guide, Institut fü Kommunikationsnetze und Rechnersysteme, Universität Stuttgart, June 2004.

Sebastian Gunreben received his Dipl.-Ing. degree in Mechatronics in 2004 from the University of Stuttgart, Germany. Since then he has been with the Institute of Communication Networks and Computer Engineering (IKR) at the University of Stuttgart. His work is dedicated to traffic engineering tasks in IP-over-WDM networks and control plane aspects of multi-layer and multi-domain networks. Currently, he is preparing his Ph.D. on the formal description of out-of-sequence packet arrivals in networks implementing packet assembly and multipath routing.