Assured Horizon – A new Combined Framework for Burst Assembly and Reservation in Optical Burst Switched Networks

Klaus Dolzer

University of Stuttgart, Institute of Communication Networks and Computer Engineering,
IND, Pfaffenwaldring 47, D-70569 Stuttgart, Germany
Tel: +49 711 685 7969, Fax: +49 711 685 7983, E-Mail: dolzer@ind.uni-stuttgart.de

Abstract. While traffic management is believed to be best implemented in the IP layer – e. g. by GMPLS – an evolving questing in the context of IP-over-photonics is whether the optical layer can provide service differentiation as service to the IP layer. As a positive answer to that question, a new framework for optical burst switched networks is proposed that comprises a new burst assembly mechanism, a new reservation mechanism as well as the communication between them. In order to keep the core very simple, the complexity is moved to the ingress of the network where burst header processing is performed in a distributed way. Performance evaluations confirm that without relying on queueing in the core, assured Horizon efficiently yields service differentiation.

1 Introduction

IP-over-photonics is a promising approach for next-generation Internet. It allows to continue the success story of IP while efficiently making use of the huge bandwidth provided by the optical layer. While traffic management is believed to be best implemented in the IP layer – e. g. by GMPLS – an evolving question is whether the optical layer can provide service differentiation as service to the IP layer. However, before the optical layer is capable to provide such a service, the challenge remains to make the optical layer – which currently usually employs static, circuit switched pipes – more dynamic [15].

The currently standardized and already widely accepted GMPLS framework [2] is a first step in this direction. It is a circuit switched approach mainly for traffic management (e. g. resource control, (constraint-based) routing), but also some traffic engineering functionality is contained (e. g. classification at the ingress, source routing). However, dynamic characteristics of IP traffic allow only a lower utilization of statically allocated wavelengths (WLs). Furthermore, a large number of WLs may be required in larger networks to mash edge nodes. A next evolution step towards a packet-switched optical network is the reduction of the allocated granularity from WLs to bursts. Hereby, a burst is a ‘large packet’ containing many IP packets. Dependent on the length of such a burst, the granularity can range from packet switching to circuit switching. In this context, optical burst switching (OBS) [14], [17], [19] has been suggested as optical network architecture supporting one service class. OBS also includes ‘medium access control’ functionality to WLs whereas the overall control remains with GMPLS and thus in the IP layer.

The remainder of this paper is organized as follows. Section 2 gives a brief introduction to OBS and provides a classification of burst assembly mechanisms and OBS-QoS mechanisms. Section 3 introduces *assured Horizon*, a new framework comprising a new burst assembly mechanism, a new reservation mechanism as well as the communication between them. Section 4 provides a performance evaluation of *assured Horizon* with respect to resulting burst characteristics as well as achievable service differentiation.

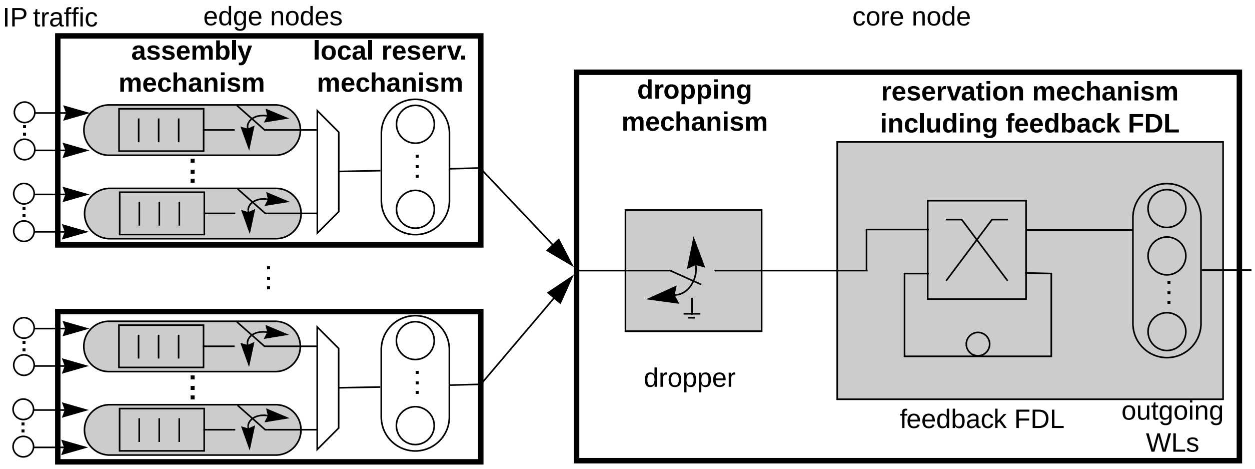

Figure 1: Block diagram of an OBS scenario including edge nodes and a core node

2 Optical Burst Switching (OBS)

2.1 Functionality

The main characteristics of OBS are the hybrid approach of out of band signalling and electronic processing of header information while data stays in the optical domain all the time. Further characteristics of OBS are one-pass reservation, variable length bursts, and no mandatory need for buffers, see e. g. [1], [13], [14], [17], [19].

Fig. 1 depicts a block diagram of an OBS scenario. Burst assembly is carried out at the network edge. Hereby, depending on the assembly policy, several assembly queues per destination may exist within one edge node. A local reservation mechanism controls access of bursts to outgoing WLs. In the core, the reservation mechanism of every core node may be supported by a burst dropper to enforce service differentiation. Fiber delay lines (FDLs) are not mandatory, but can be applied for further contention resolution [11].

An augmented view on OBS including the above GMPLS layer is discussed in [4] and also called labelled optical burst switching [13] or burst switching in virtual circuit mode [17]. Hereby, bursts are classified to forwarding equivalent classes, FECs, at the ingress. (Constraint-based) routing is carried out by GMPLS resulting in an allocation of a label for every FEC which fixes a path between ingress and egress. However, only fibers are determined whereas WLs are allocated dynamically for every burst by OBS.

At the moment, OBS is still at its definition phase which is indicated by a strongly increasing number of publications on new reservation mechanisms, Fig. 2 and Fig. 3, assembly mechanisms [12], [20], [4] and prototypes [10], [3]. In order to clarify their context, a classification of burst assembly as well as reservation mechanisms is presented.

2.2 Classification of burst assembly mechanisms

All burst assembly mechanisms available in literature are basically time-based in order to maintain a maximum waiting time in the assembly buffer. However, some slightly different flavours are available for mechanisms based on one-pass reservation. In [20] and [4] a purely time-based solution is suggested. However, [4] introduces an offset setting scheme that only allows bursts to leave the ingress node with a leaky bucket shaped interarrival time in order to obtain traffic smoothing. In [12], if the content of the assembly buffer is smaller than a threshold at timeout, the assembled burst is padded. It is shown that by doing so the traffic characteristics are significantly improved.

2.3 Classification of OBS-QoS mechanisms

When enhancing OBS to support service differentiation, three major challenges are faced. (i) There is only limited time for burst header processing in the core, (ii) there are

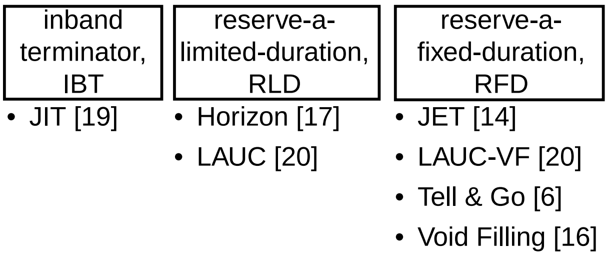

Figure 2: Classification of reservation mechanisms

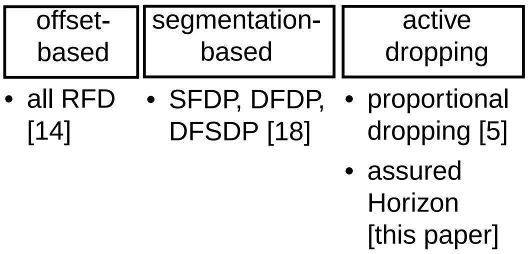

Figure 3: Classification of OBS-QoS mechanisms

no buffers beyond FDLs available for scheduling and (iii) one-pass reservation lacks to send any feedback about the state of the core to the edges. While the first challenge is compensated by electronically processing the burst header and delaying data by a constant offset, the two latter post an outstanding problem. The challenge is to find an algorithm to schedule bursts to outgoing WLs. Hereby, without relying on buffers, isolation between FECs or service classes has to be achieved. Furthermore, (iii) requests for an answer how to control possible overload that can significantly degrade the QoS.

In order to find an appropriate solution, different OBS-QoS mechanisms can be found in literature. All mechanisms classified in Fig. 3 apply one-pass reservation. Thus, the approaches in [7] and Tell & Wait [6] are not considered in this paper. Before discussing Fig. 3, the classification of one-class reservation mechanisms originally published in [9] is repeated in Fig. 2 as many OBS-QoS mechanisms directly rely on them.

Offset-based schemes like e. g. suggested in [14] for JET [14] require an RFD mechanism as different burst loss probabilities are obtained by offsets of different duration. The highest priority class has the largest offset and thus is able to reserve resources prior to all other classes. This yields a lower burst loss probability than that of lower priority classes which tend to fill gaps (voids) of higher priority bursts. However, [8] shows that this approach may be disadvantageous if the burst length of lower priority bursts exceed the one of higher priority bursts or if their distribution function has a greater variance.

Mechanisms classified as *segmentation-based* [18] require bursts to consist of several independent segments. In case of contention, some segments of a lower priority burst are either dropped or deflected whereas the remaining part of the burst can still be delivered to the egress. Thus, less bytes are lost compared to a solution with the granularity of whole bursts. However, this comes at the cost of increased complexity for burst assembly, burst scheduling as well as Byte ordering at the egress.

Finally, mechanisms based on *active dropping* implement a burst dropper in front of each core node. Dependent on a dropping policy – e.g. relative burst dropping probability [5] – some burst headers and their corresponding bursts are dropped and thus cannot compete for outgoing WLs. In order to perform dropping reasonably, the burst length has to be known prior to dropping. Accordingly, RLD as well as RFD mechanisms are suitable for active dropping based OBS-QoS mechanisms. This class of OBS-QoS mechanisms is promising as it allows to control burst arrivals based on sophisticated policies.

3 Assured Horizon framework

As none of the OBS-QoS mechanism available in literature is able to meet all three challenges introduced in Section 2.3 at the same time, a new framework called *assured Horizon* is proposed that comprises a new burst assembly mechanism, a new burst reservation mechanism as well as the communication between them. The main building blocks of this framework are (i) a coarse-grained (or static) bandwidth reservation for every FEC between ingress and egress, (ii) policing of that bandwidth reservation by the burst assemblers at the edges and (iii) – in order to allow for multiplexing gain – central enforcement of the policing at the core by a dropper in front of each core node.

3.1 Coarse-grained (or static) bandwidth reservation

The basic idea of assured Horizon is the introduction of a coarse-grained (or static) bandwidth reservation envelope r_i for every FEC i between ingress and egress. This corresponds to a ’weight’ of a ’weighted scheduler’ in the electrical domain. In case r_i is changed dynamically, a signalling protocol adapts r_i to the respective mean bitrate m_i of a FEC. The definition of such a protocol is beyond the scope of this paper and is left for further work. An allocation factor $f_i = r_i/m_i$ is defined that determines how much greater (smaller) the allocated bitrate envelope is compared to the mean rate of a FEC. f_i can e. g. be determined by the concept of effective bandwidth. Comparable to a weighted scheduler, a FEC i is given a higher priority by increasing f_i. An additional advantage of such a reservation envelope is that it allows for a coarse-grained (or static) *burst admission control* per FEC as every burst switch in the core admits or rejects new reservation requests based on its overall reserved bandwidth. Thus, a burst switch can control the amount of (admitted) traffic in order to avoid overload situations.

3.2 Burst assembly mechanism

The new burst assembly mechanism has an assembly queue and a timer for every FEC. It observes r_i by marking bursts as *compliant* (C) or *non-compliant* (NC), dependent on a burst conforming to r_i. The greater r_i, the greater the share of C bursts. In order to indicate whether a burst is C or NC, a new field called *burst drop priority* (BDP) is introduced. The name follows an idea in ATM where a bit called ‘cell loss priority’ indicates a cell exceeding its reservation. In the core, all evaluations to support service differentiation are solely based on that new BDP field. In a basic version, only C and NC bursts are distinguished. Nonetheless, an option is to extend this field in order to be able to differentiate between classes of NC bursts. An additional distinction between C bursts makes no sense, as non of them should be lost in the core, see Section 3.4. In the following, a time-based algorithm is presented that applies the basic marking scheme:

1. Upon arrival of an IP packet, the packet is classified to a FEC and forwarded to the respective queue and the respective timer is set to t_i (if it is not already set).
2. When a timer of a FEC expires, the assembly unit assembles a burst of maximum length which is still compliant to r_i. Hereby exponential averaging is applied in order to keep track of the already sent traffic volume. This burst is sent marked with C in the BDP. If r_i is zero, no burst is sent in this step.

3a. If the accumulated length of the IP packets remaining in the assembly queue exceeds a threshold q_i, they are all sent in a second burst marked NC. In order to further control the offered load, the amount of Bytes sent as NC may be bounded per FEC.

3b. Else, the non-compliant IP packets remain in the assembly buffer, the timer is set to t_i and arriving IP packets are added until the next expiration of the timer.

In the suggested mechanism, the timeout interval t_i allows to roughly adjust the resulting mean burst length $l_i = m_i \cdot t_i$ for every FEC. The threshold q_i compromises between the proportion of NC bursts and the waiting time in the assembly buffer.

3.3 Reservation mechanism of WLs at the network ingress

A local reservation mechanism controls the access to WLs at an edge node. Especially if the number of FECs is greater than the number of outgoing wavelengths w, collisions have to be avoided. As bursts are still waiting in an electronic queue, scheduling algorithms from the electronic domain can be applied. Thus e. g. first come first serve (FCFS) between FECs of the same priority and static priority between FECs of different priorities can be applied. Assuming reasonable dimensioning of outgoing optical bandwidth, it can be expected that no burst is lost at the ingress.

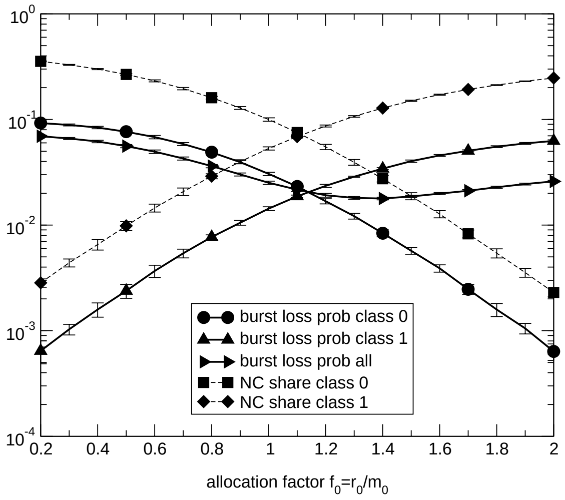

Figure 4: Impact of reservation envelope r_i

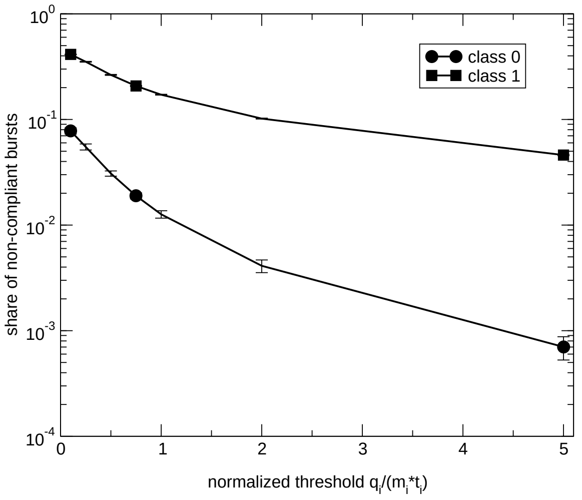

Figure 5: Impact of threshold q_i

3.4 Reservation mechanism of WLs at the network core supported by active dropping

The reservation mechanism at every core node is chosen – for simplicity – to Horizon [17]. However, any other mechanism classified in Section 2.3 as RLD or RFD can be applied. In any case, the reservation mechanism is only supposed to perform complete sharing of outgoing WLs. For central enforcement of the policing carried out by the burst assemblers, the reservation mechanism is supported by a burst dropper in front of every core node. Hereby, dropping is only based on the BDP field in the burst header and hence the core is stateless with respect to FECs.

A burst dropper has two states. A *regular state* where no burst is dropped and a *congestion state* c where all NC bursts are dropped. In case of differentiation between NC bursts (option indicated in Section 3.2), c is extended to sub-states c_k in which all bursts with a drop priority p smaller or equal to k are dropped. While this option allows for additional differentiation, it introduces more complexity.

State changes of the burst dropper are triggered from the reservation mechanism dependent on the number of currently allocated WLs w_a. Let w_c denote the number of allocated WLs where the congestion state starts, then $w_a \geq w_c$ indicates that the dropper is in congestion state. Hence, neither the dropper nor the reservation mechanism is required to perform any calculations to determine if a burst is dropped.

Dimensioning of w_c depends on the objective carried traffic and is a trade-off between overall burst losses and isolation between FECs. The aim is that at the objective carried traffic, only a negligible number of C bursts cannot find an outgoing WL and thus have to be discarded from the reservation mechanism. By doing so, this OBS-QoS mechanism realizes isolation between FECs, as most bursts that are marked as compliant can find an outgoing WL. Accordingly, this traffic is guaranteed a negligible burst loss probability. If there is only one global congestion state, there is no isolation between NC burst which all experience the same service. Mulitiplexing gain is achieved by (dynamic) partial sharing of WLs between C and NC bursts. The dynamic results from dedicating the last $w - w_c + 1$ not allocated WLs to C bursts.

4 Performance evaluation

Performance evaluations are carried out by simulations [21] in a scenario of a national backbone of the size of Germany. Comparable to Fig. 1, one core node with 8 WLs ($w_c = 6$) and a feedback FDL (delay of 0.2 ms, maximum 3 circulations) at 2.5 Gbps receives bursts from 10 edge nodes with 2 service classes each. Restriction to 8 WLs

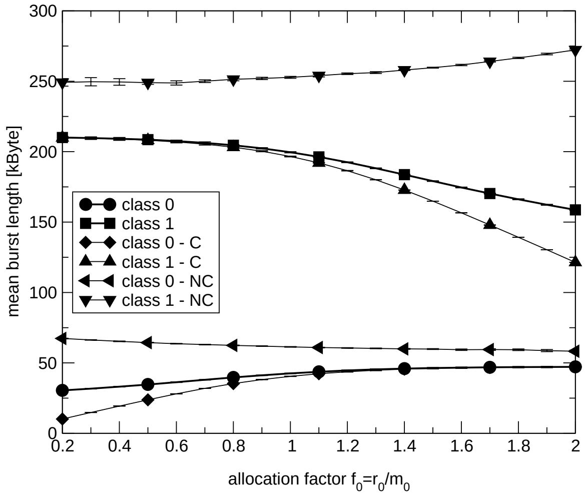

Figure 6: Mean burst length

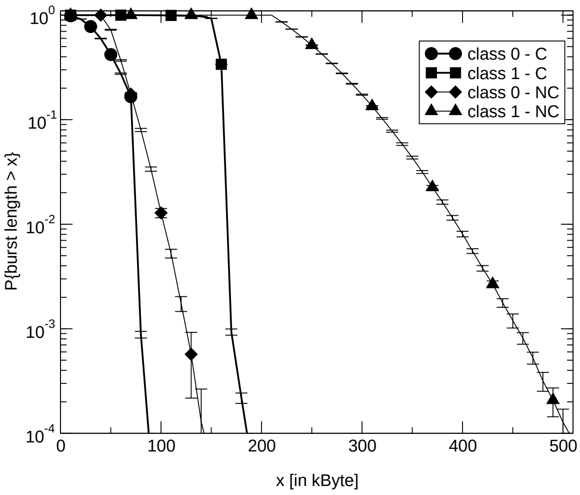

Figure 7: Burst length distribution

allows to perform simulations with sufficient accuracy. As shown in [9], a greater number of wavelengths significantly lowers burst losses. Header processing is assumed to be compensated by a dedicated FDL at every core node in order to avoid different offsets.

Bursts of high priority class 0 (share 0.3) and low priority class 1 are assembled from self-similar IP traffic which is segmented from files with negative exponentially distributed interarrival time and Pareto distributed size ($\alpha = 1.6$, mean = 10 kByte). The overall reserved rate $r = r_0 + r_1 = m$ and r_i are constant in a simulation. Thus, if r_0 is increased, r_1 is decreased according to $r_1 = (1 - 0.3 \cdot f_0) \cdot m_1 / 0.7$. Class-based results summarize statistics from FECs with identical parameters. In order to avoid synchronisation effects, a jitter of $\pm 5\%$ is added to $t_0 = 1\text{ms}$ and $t_1 = 2\text{ms}$. In Section 4.2, only the burst loss probability is discussed as end-to-end delay including queueing for burst assembly only exceeds 2ms for class 0 and 5ms for class 1 with a probability of 10^{-5}.

4.1 Impact of burst assembly on burst characteristics

In Fig. 4, the NC share of both classes is depicted against f_0 for an offered load of 0.6 per WL. Increasing f_j results in more bursts of service class j to be within the reservation envelope and thus marked as C. Comparable to the increase of a weight of a weighted scheduler in the electronic domain, this is the basis for this framework to service differentiation. $f_0 = 1.6$ is chosen for all further simulations as it allows for a differentiation of about one order of magnitude. Fig. 5 shows the decrease of the NC share against q_j. For greater q_j, the differentiation gets greater as the traffic volume of class 0 is smaller. $q_j = 1$ (also applied in Fig. 4) is chosen for all further simulations in order to compromise between proportion of NC bursts and edge delay.

Fig. 6 depicts the resulting mean burst length against f_0. The choice of q_j results in NC bursts which are in average longer than C bursts. According to the design of the assembly mechanism, smaller f_j results in reduced mean burst length. The separated statistics for C and NC bursts show that the mean C burst length is additionally decreased while the traffic volume is send in longer NC bursts. This behaviour is also confirmed by the burst length distributions depicted in Fig. 7. Whereas the burst length distribution of C bursts drops quickly, the heavy tail of the IP file size distribution is captured by NC bursts. This is especially advantageous as NC burst are only accepted if the carried traffic is low and thus the undesirable impact of long lasting congestions caused by very long NC burst is moderated.

4.2 Impact of framework parameters on burst loss probability

In Fig. 4, also the burst loss probabilities $P_{\text{Loss, j}}$ are depicted against f_0. According to

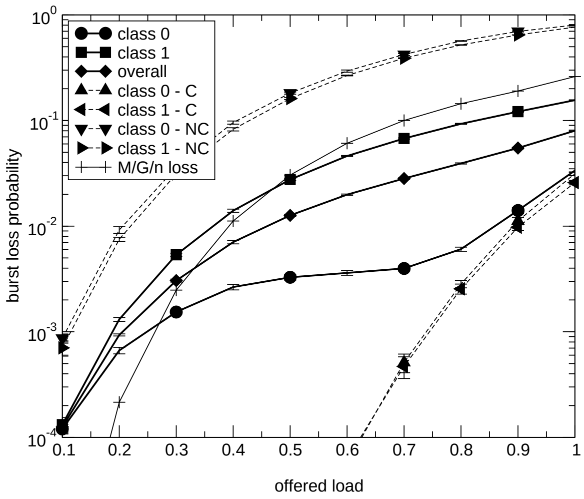

Figure 8: Impact of proportionally increases load

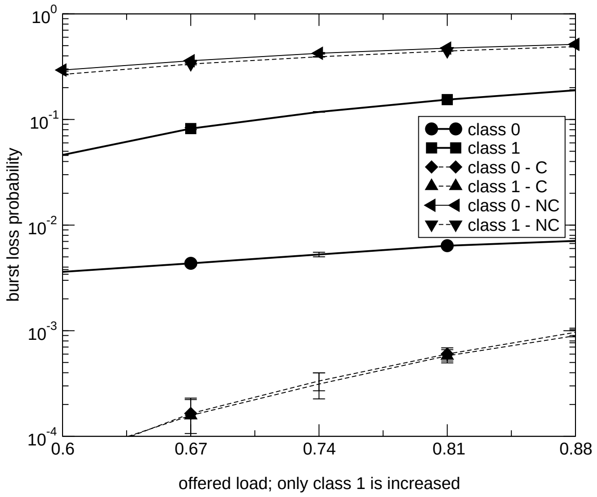

Figure 9: Impact of increased class 1 load

the design of the dropping mechanism, $P_{\text{Loss, j}}$ directly follows the progression of the share of NC bursts of class j. The overall burst loss probability $P_{\text{loss, all}}$ hereby only slight increases with increasing f_0. Their progression is typical for a weighted scheduler whose weights are varied. Hereby, the limit $f_j \to \infty$ corresponds to all burst of service class j being marked as C whereas all bursts of the other class are marked as NC. This is analogue to an electronic priority scheduler. Not shown due to space limitations, $P_{\text{Loss, j}}$ does not depend on the actual length of a burst and is the same for all NC bursts and C bursts, respectively. Concluding the discussion of Fig. 4, $P_{\text{Loss, j}}$ can be directly engineered for a FEC by the marking scheme at the edge of the network and its magnitude is determined by f_j. Thus a low burst loss probability of a high priority class does not automatically lead to a high burst loss probability of a low priority class.

In Fig. 8, $P_{\text{Loss, j}}$ is depicted against the offered load (per WL) in a scenario where the offered traffic as well as r_j are increased proportionally for all FECs. It can be seen that differentiation is achieved over the whole range of load with $P_{\text{Loss, 0}}$ staying reasonably constant around the objective offered load of 0.6. In order to avoid the strong increase of $P_{\text{Loss, 0}}$ for higher load, the overall admitted load at a core node should be bounded. Fig. 8 can be the basis for the dimensioning a such a boundary. For higher offered load, the increase of $P_{\text{Loss, all}}$ and $P_{\text{Loss, NC}}$ roughly follows the progression of the burst loss probability in an M/G/8 loss system which also indicates that the framework works reasonably.

Finally, Fig. 9 depicts a scenario where the absolute offered traffic of class 0 as well as r_j are kept constant and only the offered load of class 1 is increased starting from an overall offered load of 0.6. It can be seen that class 0 is protected from class 1 as $P_{\text{Loss, 0}}$ only moderately increases.

5 Conclusions

In this paper, a new combined framework for burst assembly and reservation called assured Horizon is introduced which comprises a new burst assembly mechanism, a new reservation mechanism as well as the communication between them. It addresses the three major three challenges for realization of OBS-QoS mechanisms which are not meet at the same time by published mechanisms, namely (i) limited time for header processing, (ii) no flexible buffer concept beyond FDLs and (iii) lack of feedback of the one-pass reservation. The basic idea is the introduction of a coarse-grained (or static) bandwidth reservation for every FEC between ingress and egress which also allows to control the load offered to a core node. The efficiency of that framework is achieved by carrying out header processing/policing by the assembly mechanism at the edges of the network in a distributed way. In order to allow for multiplexing gain, the policing is centrally enforced by a dropper in from of each core node.

Performance evaluations confirm that service differentiation is achieved whose order of magnitude can be adjusted within a broad range comparable to a electronic scheduler. Without relying on queueing, this framework provides protection between FECs.

Acknowledgement

The author thanks the ICIR group at International Computer Science Institute (ICSI) for their support during his sabbatical as well as his colleagues in Stuttgart – Stefan Bodamer and Christoph Gauger – for many inspiring discussions via e-mail.

This work was partly funded within the Transinet project by the German Bundesministerium für Bildung und Forschung under contract No. 01AK020C (http://www.transinet.de).

References

[1] Amstutz, S. R.: Burst switching - an update. IEEE Commun. Mag., Sept. 1989, pp. 50-57.

[2] Ashwood-Smith, P. et al.: Generalized Multi-Protocol Label Switching (GMPLS) Architecture. IETF draft, draft-ietf-ccamp-gmpls-architecture-02.txt, March 2002, work in progress.

[3] Baldine, I.; Rouskas, G.; Perros, H.; Stevenson, D.: JumpStart: A just-in-time signalling architecture for WDM burst-switched networks. IEEE Commun. Mag., Feb. 2002, pp. 82-89.

[4] Chaskar, H. M.; Verma, S.; Ravikanth, R.: A framework to support IP over WDM using optical burst switching. Proceedings of the Optical Networks Workshop, Richardson, Texas, 2000.

[5] Chen, Y.; Hamdi, M.; Tsang, D.: Proportional QoS over OBS networks. Proc. of the IEEE Global Telecommunications Conference (Globecom’ 01), San Antonio, Nov 2001.

[6] Detti, A.: Listanti, M.: Application of tell & go and tell & wait reservation strategies in an optical burst switching network: a performance comparison. Proceedings of the 8th IEEE International Conference on Telecommunications (ICT 2001), Bucharest, June 2001.

[7] Dueser, M.; Bayvel, P.: Bandwidth utilization and wavelength re-use in WDM optical burst-switched packet networks. Proceeding of 5th IFIP Working Conference on Optical Design and Modelling (ONDM 2001), vol. 1, Vienna, Feb. 2001.

[8] Dolzer, K.; Gauger, C.: On burst assembly in optical burst switching networks - a performance evaluation of Just-Enough-Time. Proc. 17th International Teletraffic Congress (ITC 17), Salvador da Bahia, Brazil, Dec. 2001, pp. 149-160.

[9] Dolzer, K.; Gauger, C.; Späth, J.; Bodamer, S.: Evaluation of reservation mechanisms for optical burst switching. AEÜ International Journal of Electronics and Communications, Vol. 55, No. 1, Jan. 2001.

[10] Eilenberger, G,: Optische Paketnetze – Alles optisch, oder?. Proc. 2. ITG Fachtagung Photonic Networks, Dresden, March 2001, pp. 109-114.

[11] Gauger, C.: Performance of converter pools for contention resolution in optical burst switching, Proceedings of OptiComm 2002, Boston, July 2002.

[12] Ge, A.; Callegati, F.: On optical burst switching and self-similar traffic, IEEE Commun. Letter 4, March 2000.

[13] Qiao, C.; Yoo, M.: Choices, features and issues in optical burst switching. Optical Networks Magazine, Vol. 1, No. 2, April 2000, pp. 36-44.

[14] Qiao, C.; Yoo, M.: Optical burst switching (OBS) - a new paradigm for an optical internet. Journal of High Speed Networks, No. 8, 1999, pp. 69-84.

[15] Spaeth, J.: Dynamic routing and resource allocation in WDM transport networks. Computer Networks, Vol. 32, May 2000, pp. 519-538.

[16] Tancevski, L; et. al.: Optical routing of asynchronous, variable length packets.” IEEE J. on Selected Areas in Communications, Vol. 18, No. 10, Oct. 2000, pp. 2084-2093.

[17] Turner, J. S.: Terabit burst switching. J. of High Speed Networks, No. 8, 1999, pp. 3-16.

[18] Vokkarane, V. M.; Jue, J. P.: ‘Prioritized Routing and Burst Segmentation for QoS in Optical Burst-Switched Networks, Optical Fiber Communication Conference 2002, Anaheim, CA, March 2002.

[19] Wei, J. Y.; Pastor, J. L.; Ramamurthy, R. S.; Tsai, Y.: Just-in-time optical burst switching for multi-wavelength networks. Proc. of Broadband (BC’99), 1999, pp. 339-352.

[20] Xiong, Y.; Vandenhoute, M.; Cankaya, H.: Control architecture in optical burst-switched WDM networks. IEEE J. on Selected Areas in Communications, Vol. 18, No. 10, Oct. 2000, pp. 1838-1851.

[21] http://www.ind.uni-stuttgart.de/INDSimLib/