NAIST-IS-MT0751144

Master's Thesis

Studies on Fairness Improvement and Effective Wavelength Utilization in OBS Mesh Networks

Badr Oulad Nassar

February 05, 2009

Department of Information Systems
Graduate School of Information Science
Nara Institute of Science and Technology

A Master's Thesis
submitted to Graduate School of Information Science,
Nara Institute of Science and Technology
in partial fulfillment of the requirements for the degree of
MASTER of ENGINEERING

Badr Oulad Nassar

Thesis Committee:

Professor Kenji Sugimoto	(Supervisor)
Professor Kotaro Minato	(Co-supervisor)
Associate Professor Kentaro Hirata	(Co-supervisor)
Associate Professor Yuichi Kaji	
Assistant Professor Takuji Tachibana	(Co-supervisor)

Studies on Fairness Improvement and Effective Wavelength Utilization in OBS Mesh Networks*

Badr Oulad Nassar

Abstract

In optical burst switching (OBS) networks, an optical burst is generated at the OBS ingress node, from multiple IP packets coming from access networks. After the optical burst is generated, its corresponding control packet reserves all the wavelengths until the destination node and then, the optical burst is transmitted along the reserved wavelengths, all optically, hence providing an ultra fast transmission speed. In OBS networks, the number of wavelength reservations is equal to the number of hops, and hence, bursts whose number of hops is large tend to be lost more frequently than those with a small number of hops. Such an unfairness is one of the most important issues in OBS networks, and many methods have been proposed in the literature to resolve this matter.

Many of the above mentioned methods improve fairness but their drawback is a significant increase in the overall burst loss probability. However, the recently proposed hop-based burst-cluster transmission not only improves fairness but also decreases the overall burst loss probability significantly. In this method, multiple bursts are assembled simultaneously and sorted from the smallest number of hops to the largest one. With this method, the burst loss probability for a large number of hops decreases, improving fairness. The performance of this method has been evaluated in OBS ring networks, and the effectiveness of this method has been shown. However, in mesh networks, the overall burst loss probability increases by using the method because the transmission of burst-cluster from an ingress node is often synchronized with that from other ingress node. Furthermore, the amount

*Master's Thesis, Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT0751144, February 05, 2009.

of traffic on each link is not necessarily the same, and hence the performance of hop-based burst-cluster transmission degrades.

In this thesis, in order to resolve the synchronization issue, we propose the utilization of random scheduling for hop-based burst-cluster transmission. In the random scheduling, the maximum waiting time is determined in advance from the maximum acceptable delay of an IP packet and the buffer size of the ingress node. Every time a burst-cluster is generated, a scheduler derives the actual waiting time at random based on the obtained maximum waiting time. When a timer becomes the actual waiting time, the burst cluster is transmitted from the scheduler.

Moreover, in order to adapt hop-based burst-cluster transmission when the traffic load is dynamic, we propose dynamic burst ordering. In this method, each ingress node calculates the burst loss probability for each number of hops using ACK and NACK messages. Based on the calculated probabilities, the ingress node changes the order of bursts within a burst-cluster dynamically. It is expected that this method can improve local fairness for each ingress node.

As far as the numerical examples are concerned, the performances of both methods are evaluated with Monte Carlo simulation. For random scheduling, we evaluate by its effectiveness for NSFNET. Numerical examples show that the random scheduling not only decreases the overall burst loss probability but also improve the fairness significantly. For dynamic burst ordering, we evaluate its performance in tandem networks by simulation. Numerical examples show that the proposed method is effective for improving local fairness for each ingress node regardless of the amount of traffic on each link. In addition, it is shown that the proposed method improves global fairness.

Keywords:

Optical burst switching, Fairness, Burst-cluster transmission, Synchronization, Random scheduling, Dynamic burst ordering.

OBS メッシュ網における公平性改善と効率波長予約使用に関する研究*

Badr Oulad Nassar

内容梗概

光バースト交換（Optical Burst Switching: OBS) では、アクセス網からエッジノードに到着した複数の IP パケットから光バーストを生成し，制御パケットが送信された後，生成した光バーストの送信が開始される．制御パケットは各ノードで光バーストを送信するための波長をあらかじめ予約し，光バーストは予約した波長を用いて光信号のまま目的ノードまで送られる．これにより，光バーストは電気信号に変換されることなく高速伝送が実現される．光バースト交換網では，各ノードで波長予約が必要となることから，伝送ホップ数が多い光バーストは伝送ホップ数の少ない光バーストよりも棄却されやすくなる．このような伝送ホップ数に関する不公平性は光バースト交換網で解決すべき重要な問題の一つであり，これまでに多くの公平性改善方式が提案されている．

しかしながら，これらの方式を用いると全体のバースト棄却率が増加してしまい，波長を有効利用することができない．その一方で，近年提案された伝送ホップ数型バーストクラスタ伝送は，公平性改善だけでなく波長有効利用も可能になる方式として注目されている．この伝送ホップ数型バーストクラスタ伝送では，出力リンクが同じで伝送ホップ数が異なる複数のバーストから，一つのバーストクラスタを生成し，バーストクラスタ単位で伝送を行う．ここで，バーストクラスタ内には各光バーストが伝送ホップ数順に配置され，ホップ数が最も少ない光バーストから送信される．この方式では，後方に配置された光バーストの波長予約が成功しやすくなるため，伝送ホップ数の大きいバーストの棄却率が低下し，結果として公平性改善が期待される．本方式の性能は光バーストリング網で評価されており，公平性改善と波長の有効利用の実現が示されている．しかしながら

---

*奈良先端科学技術大学院大学 情報科学研究科 情報システム学専攻 修士論文, NAIST-IS-MT0751144, 2009 年 2 月 05 日.

光バーストメッシュ網で利用する場合，異なるノードから送信された複数のバーストクラスタが，特定のノードにほぼ同時に繰り返し到着する傾向にあり，この同期現象により光バースト棄却率が増加してしまう．さらに，各リンクのトラヒック量が同じではないために，公平性改善の効果も減少してしまう．

そこで本論文では，同期現象による性能低下を防ぐために，ランダムスケジューリング方式を提案し，さらに各リンクのトラヒック量が異なる場合でも公平性が改善されるように動的バースト配置法を提案する．

ランダムスケジューリング方式では，バーストクラスタ伝送前にランダムな待ち時間を導入することで同期の解消を目指す．本方式では，導入される待ち時間の最大値をIPパケットの最大許容遅延，もしくは各ノードのバッファ容量から導出し，アプリケーションへの影響を最小限に抑えることができる．一方，動的バースト配置法では，バースト伝送の成否を示すACK/NACKメッセージを用いて各ホップ数のバースト棄却率を計算し，各送信ノードは計算結果を基にバーストクラスタ内のバースト順を動的に変化させる．具体的には，棄却率が低いホップ数のバーストを前方に配置し，棄却率が高いホップ数のバーストは後方に配置する．後方に配置される光バーストの伝送が成功しやすいため，公平性の改善が期待される．

両提案方式をOBSメッシュ網で用いた場合の性能を，モンテカルロシミュレーションで評価する．数値例から，ランダムスケジューリングの使用によりバーストクラスタの同期現象が解消され，全体の光バースト棄却率が減少することを示す．また，本方式は伝送遅延の制約があるアプリケーションでも利用でき，さらに公平性改善も実現されることを述べる．一方，動的バースト配置法に関しては，各ホップ数のバースト棄却率に与える影響を評価し，各送信ノードに対する局所的な公平性を改善できることを示す．また，光バースト棄却率の増加を抑えながら大局的な公平性も改善できることを確認する．

キーワード

光バースト交換, 公平性, バーストクラスタ伝送, 同期, ランダムスケジューリング, 動的バースト配置法

Contents

List of Figures

List of Tables

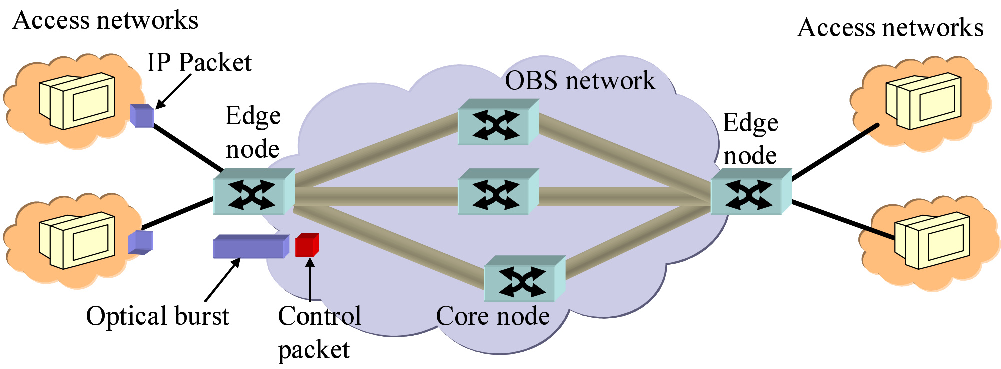


Figure 1. The OBS network.

1. Introduction

Recently, wavelength division multiplexing (WDM) networks that multiplex thousands of wavelengths into a single optical fiber are attractive as an integral part of the infrastructure of next-generation networks [1]. WDM Networks can readily support from hundreds of gigabits per second (Gbps) to several terabits per second (Tbps) on a single optical fiber [2, 3, 4, 5, 6, 7, 8].

To support bursty optical networks traffic on the Internet efficiently, optical burst switching (OBS), which is a WDM transmission scheme, was proposed [9]. In OBS, incoming traffic (IP packets) from clients at the access networks is aggregated at the ingress of the OBS network. As shown in Fig. 1, the transmission data unit is called optical burst, which is assembled from multiples IP packets that have the same OBS egress node. Each burst has its own control packet reserves the necessary wavelengths before burst transmission. Note that the control packet is processed electrically, in order to select the route and reserve a wavelength at each node, whereas the optical burst is transmitted all optically without O/E/O conversion [2]. Then, at the OBS egress node, the optical burst is disassembled into the original IP packets, and are sent to their respective access networks. Finally the wavelengths are released.

In OBS networks, bursts with different number of hops have different burst loss probabilities, resulting in the fairness issue as explained in Fig. 2. In this figure,

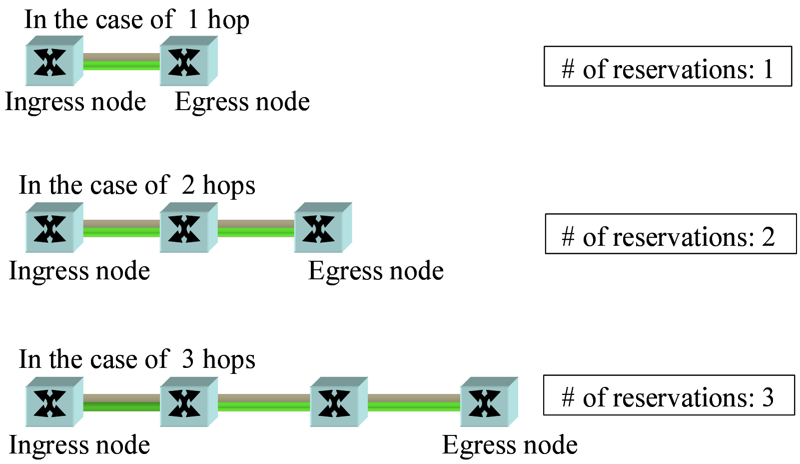


Figure 2. The Fairness issue.

we are considering three transmission examples: one-hop, two-hop and three-hop transmissions. Here, the control packet has to reserve a wavelength at all nodes, therefore in the first case, two wavelength reservations are required. In the second and third case, the control packet has to make two and three wavelength reservations, respectively. If there is no available wavelength at a node, the wavelength reservation fails and the burst is lost at that node. Therefore, the burst loss probability becomes large (small) as the number of hops increases (decreases). In order to solve fairness in terms of the number of hops, several methods have been proposed for the immediate reservation [10].

In balanced just-in-time (BJIT) [11], as a burst traverses many hops, this method gradually increases the number of wavelengths available for the burst. Therefore, the burst loss probability for a large (small) number of hops decreases (increases), improving fairness. However, by using this method, the overall burst loss probability increases significantly. Prioritized random-early discard (PRED) [11] is based on proactive burst dropping with a discarding probability that decreases as the number of hops increases. This method discards bursts based on probabilistic parameters at the ingress node, improving fairness. However, it is difficult to determine the optimal parameters in advance.

In addition, preemption mechanism has been utilized for improving fairness [12, 13, 14, 15]. In these methods, a burst whose number of hops is large can preempt another burst whose number of hops is small. This method can improve fairness more simply and more efficiently than BJIT and PRED, however, the overall burst loss probability increases slightly.

In [16], hop-based burst-cluster transmission has been proposed. In this method, a burst-cluster is generated from multiple bursts and bursts are sorted from lowest number of hops to highest within the burst-cluster. By using this method, bursts with large number of hops have many chances for wavelength reservation, and hence the loss probability of bursts whose number of hops is large (small) becomes small (large). The performance of hop-based burst-cluster transmission has been evaluated in a unidirectional ring network. Numerical examples have shown that this method can not only improve fairness but also decreases the overall burst loss probability.

However, when attempting to implement this method in other topologies such as mesh networks, we encountered two problems. First, the transmission of burst-clusters from an ingress node is often synchronized with that from other ingress nodes because the burst-cluster generation time is almost constant. Therefore, hop-based burst-cluster transmission increases the overall burst loss probability due to the synchronization. Second, in the hop-based burst-cluster transmission, the loss probability of a burst for each number of hops is affected by the amount of traffic on its last-hop link. For instance, if the traffic load on the last-hop link is high, a burst is more likely to be lost. This is because each control packet often performs the wavelength reservation process only at its last-hop link. As a result, local fairness for each ingress node is never improved by using the hop-based burst-cluster transmission.

In this thesis, in order to resolve the synchronization, we propose the utilization of random scheduling in hop-based burst-cluster transmission. In addition, we propose dynamic burst ordering for hop-based burst-cluster transmission, so that local fairness can be improved for all ingress nodes and global fairness can also be improved.

In random scheduling, the maximum waiting time is derived based on the acceptable delay of IP packets and the buffer size of the ingress node. Every time a burst-cluster is generated, a scheduler determines the actual waiting time at random based on the derived maximum waiting time. After a timer becomes the actual waiting time, the burst-cluster is transmitted from the ingress node. It is expected that the transmission interval is not constant and the synchronization can be resolved.

On the other hand, in dynamic burst ordering, each ingress node calculates the burst loss probability for each number of hops from the number of ACK and NACK messages. Based on the calculated loss probabilities, the ingress node changes the order of bursts within a burst-cluster dynamically. In this method the burst having the smallest loss probability is placed at the first position within the burst-cluster and the one with the highest loss probability is placed at the last position. This is because a burst in the rear part of a burst-cluster can reserve a wavelength with a higher probability. Therefore, it is expected that burst loss probabilities are almost the same regardless of the number of hops for each ingress node.

As for the performance evaluation, we use Monte Carlo simulation for both proposed methods. we evaluate the performance of hop-based burst-cluster transmission with random scheduling by simulation for NSFNET and investigate its effectiveness. On the other hand, dynamic burst ordering is evaluated in a tandem network.

The rest of this thesis is organized as follows. Section 2 summarizes the hop-based burst cluster transmission as the previous work and Sect. 3 describes our proposed random scheduling. The proposed dynamic burst ordering is presented in Sect. 4. Numerical examples are discussed in Sect. 5 and finally, conclusions are presented in Sect. 6.

2. Hop-Based Burst-Cluster Transmission

2.1 Overview

Burst-cluster transmission has been proposed to provide service differentiation in terms of the burst loss probability [17]. This method has been extended to resolve unfairness as hop-based burst-cluster transmission [16]. In this section, we explain how the hop-based burst-cluster transmission is utilized and how the fairness can be improved with this method.

We consider an OBS network with N nodes, and we focus on one ingress node which has $N-1$ egress nodes and L links ($L \leq N-1$). Here, the ith link is denoted as $l(i)$ and $N(i)$ denotes the number of egress nodes where bursts are transmitted with $l(i)$.

The ingress node has L burstifiers that correspond to the output links, and the burstifier for $l(i)$ has $N(i)$ queues which correspond to the egress nodes. Moreover, this node is composed of a scheduler and an OBS switch. When an IP packet arrives at the ingress node, it is stored into a queue according to its egress node.

Let T denote the timer value, and B_{total} the total size of IP packets stored in all queues of a burstifier. In addition, let $T_{\max}$ and $BC_{\max}$ denote the maximum burst assembly time and the maximum size of a burst-cluster, respectively. A burst-cluster is generated according to the following algorithm.

Step 1: T and B_{total} are set to 0.

Step 2: The timer starts when the first IP packet arrives at the burstifier. Furthermore, B_{total} is updated accordingly.

Step 3: As long as $T < T_{\max}$ and $B_{total} < BC_{\max}$, a new arriving IP packet is stored into its appropriate queue.

Step 4: When $T = T_{\max}$ or $B_{total} \geq BC_{\max}$, $N(i)$ bursts are assembled from $N(i)$ queues simultaneously.

Step 5: A burst-cluster is generated from the assembled bursts, where bursts are sorted from lowest number of hops to highest.

Figure 3 shows an OBS network with eight nodes, and we focus on the leftmost node, which is denoted as ingress node. This node has one link and seven egress nodes. At the ingress node, a burst-cluster is generated according to the above

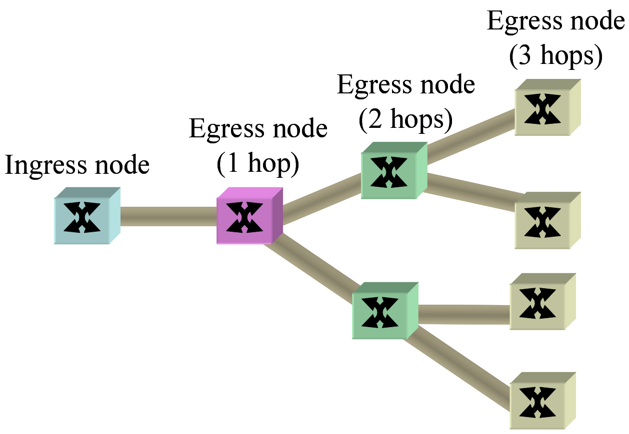


Figure 3. Example of OBS network.

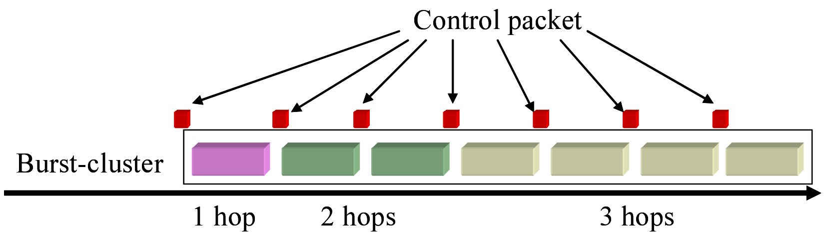


Figure 4. A burst-cluster in the case of Fig. 1.

algorithm (see Fig. 4). As shown in Fig. 3, there is only one egress node that is one-hop away from the ingress node, and therefore, the number of bursts for one hop is one within the burst-cluster. Similarly, there are two (four) egress nodes that are two-hops (three-hops) away from the ingress node, and hence the number of bursts for two (three) hops is two (four). When there are multiple egress nodes at two or more hops, bursts for each hop are ordered at random within the burst-cluster. Note that each burst has its own control packet.

Just after the burst-cluster is generated, it is forwarded to a scheduler. Then, the burst-cluster is transmitted to egress nodes along with the control packets from the scheduler. Here, the fundamental difference between the hop-based burst-cluster transmission and the original immediate reservation is that the control packet of the preceding burst reserves the wavelength not only for its own

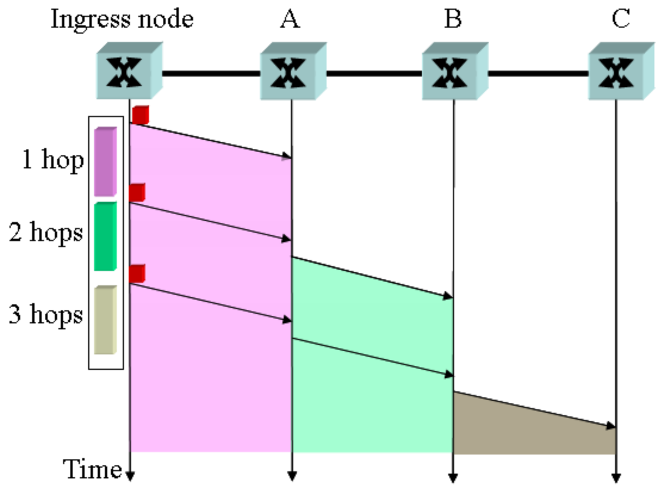


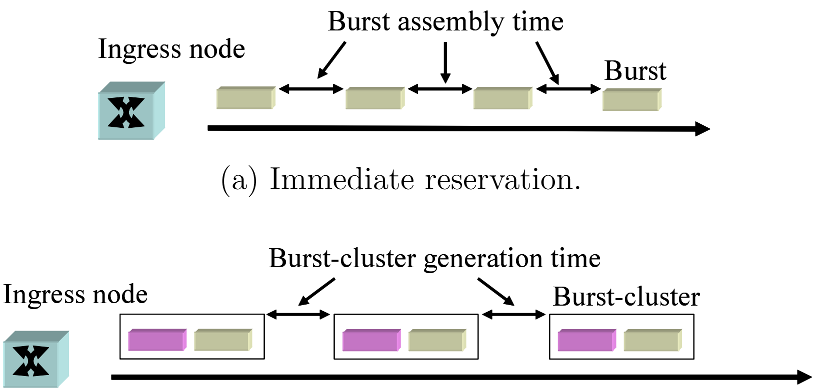
Figure 5. Hop-based burst-cluster transmission.

burst but also for the remaining bursts within the burst-cluster.

Figure 5 shows an example of hop-based burst-cluster transmission. In this figure, the ingress node has three egress nodes and a burst-cluster contains three bursts for the egress nodes. As shown in this figure, each control packet reserves a wavelength on only one link, if all wavelength reservation do not fail. Consequently, the number of wavelength reservations for each burst is almost the same regardless of the number of hops, improving fairness. Moreover, this method can decrease the redundant wavelength reservation which is required in the original immediate reservation. Hence, the hop-based burst-cluster transmission can also decrease the overall burst loss probability.

2.2 The Synchronization Issue

Figures 6(a) and (b) show examples of burst transmission from an ingress node for the traditional immediate reservation and the hop-based burst-cluster transmission, respectively. In these figures, bursts and burst-clusters are assembled and generated according to the timer/threshold-based algorithm which is effective for OBS networks [18]. As long as the amount of traffic does not change frequently, the interval between successive bursts and the interval between successive burst-



(a) Immediate reservation.

(b) Hop-based burst-cluster transmission.

Figure 6. The case of timer/threshold-based algorithm is used.

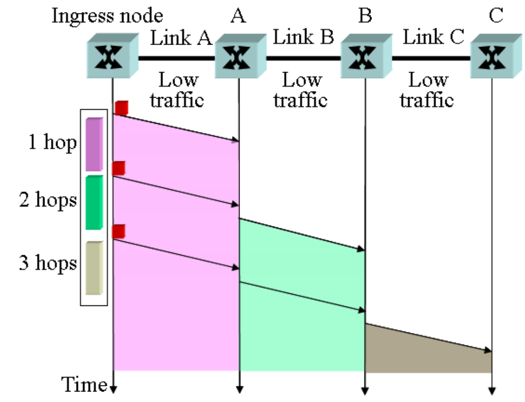
clusters become almost constant.

In addition, comparing Fig. 6(a) and Fig. 6(b), the size of a burst-cluster is much larger than that of an optical burst. Therefore, because of the repetitive constant time interval mentioned above, control packets coming from different source nodes arrive at the same intermediate node at almost the same time and hence, many control packets cannot reserve a wavelength at the node because there are no available wavelengths, leading to a massive burst loss. Consequently, in this situation, synchronization degrades the performance of the hop-based burst-cluster transmission significantly. It is indispensable to resolve this issue for the hop-based burst-cluster transmission.

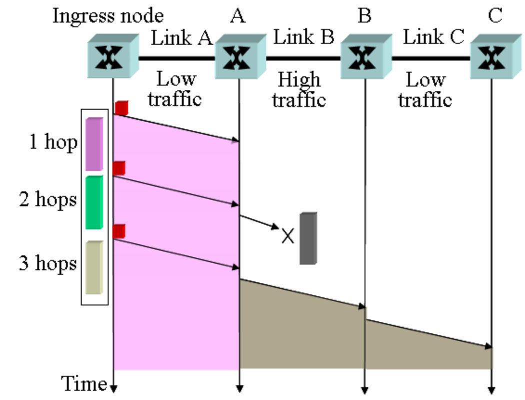
2.3 Impact of Traffic on the Last-Hop Link

As shown in subsection 2.1, in the hop-based burst-cluster transmission, it is expected that the number of wavelength reservations is only one regardless of the number of hops. This denotes that the wavelength reservation for each burst is performed only at its last-hop link.

Figure 7 shows a tandem network where the number of nodes is four. In this network, link A, link B and link C are the last-hop link of the one-hop burst, the two-hop burst, and the three-hop burst, respectively. In Fig. 7(a), the amount of traffic on every link is low, but in Fig. 7(b), only the amount of traffic on link B



(a) A case where the amount of traffic is low on each link.



(b) A case where the amount of traffic is high only on link B.

Figure 7. Hop-based burst-cluster transmission.

is high.

On link A, the control packet for the one-hop burst can reserve a wavelength easily due to low traffic, and the transmission of the one-hop burst succeeds. However, on link B, it is difficult for the control packet of the two-hop burst to reserve a wavelength due to congestion, and the two-hop burst is likely to be lost. Nevertheless, the control packet of the three-hop burst can reserve a wavelength on link B if the congestion has been resolved. Moreover, the control packet for the three-hop burst can reserve a wavelength on link C easily due to low traffic. As a result, the loss probability of the three-hop burst tends to be smaller than that of the two-hop burst.

From the above, having a situation where a three-hop burst has a smaller loss probability than a two-hop burst is undesirable. Therefore, hop-based burst-cluster transmission can not always improve the local fairness for all ingress nodes, when the traffic load changes dynamically.

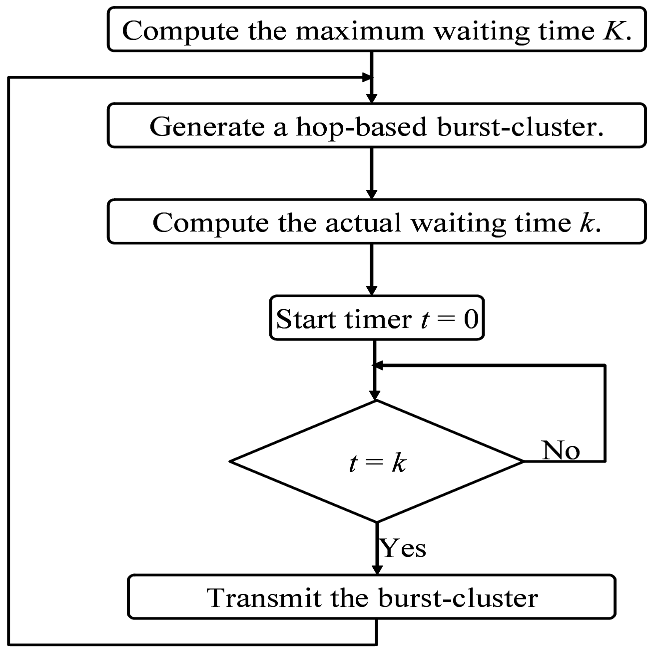


Figure 8. Random scheduling algorithm.

3. Random Scheduling

3.1 Random Scheduling Algorithm

In order to resolve the synchronization issue, we propose random scheduling for hop-based burst-cluster transmission. The random scheduling is used in the scheduler of each ingress node.

Figure 8 shows the random scheduling algorithm. As shown in this figure, first the random scheduler computes the maximum waiting time K. Every time a burst-cluster is generated, the actual waiting time k is determined at random between $[0,\ K]$ at the scheduler, and then a timer T starts. The burst-cluster and the corresponding control packet is transmitted from the scheduler after T becomes k.

Figure 9 shows the impact of random scheduling. As shown in this figure, the transmission interval between two successive burst-clusters is not constant.

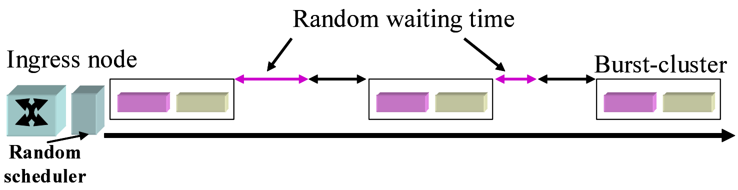


Figure 9. Transmission interval for the hop-based burst-cluster transmission with random scheduling.

Hence, the random scheduling is effective for resolving the synchronization in spite of its simplicity.

3.2 Derivation of the Maximum Waiting Time K for Delay-Sensitive Traffic

It is clear that the transmission delay of each burst becomes large by using random scheduling. For delay-sensitive traffic, a large transmission delay is critical in addition to the burst loss probability. The maximum waiting time K has to be derived considering the acceptable delay of delay-sensitive traffic.

Let D_{accept} denote the maximum acceptable delay of an IP packet inside OBS networks. Note that this delay does not include the transmission delay in access networks. For example, D_{accept} is from 50 to 100 ms for VoIP traffic [19]. H is the maximum number of hops, C is the transmission speed of a wavelength, and δ is the processing time of a control packet. In addition, T_S is the optical switching time and $P_{\max}$ is the maximum propagation delay of a burst within a burst-cluster. Please see Fig. 10 about these parameters.

When $T_{\max}$ is the maximum assembly time and $BC_{\max}$ is the maximum size of a burst-cluster, the maximum waiting time K is derived from the following inequality (see Fig. 10).

$$K \leq D_{accept} - \left\{ T_{\max} + (H+1) \times (\delta + T_S) + P_{\max} + \frac{BC_{\max}}{C} \right\}. \tag{1}$$

Here H and $P_{\max}$ are different for each ingress node and each output link. Therefore, from (1), K is different for each burst-cluster. However, the implementation is easy if the same K is used for all burst-clusters.

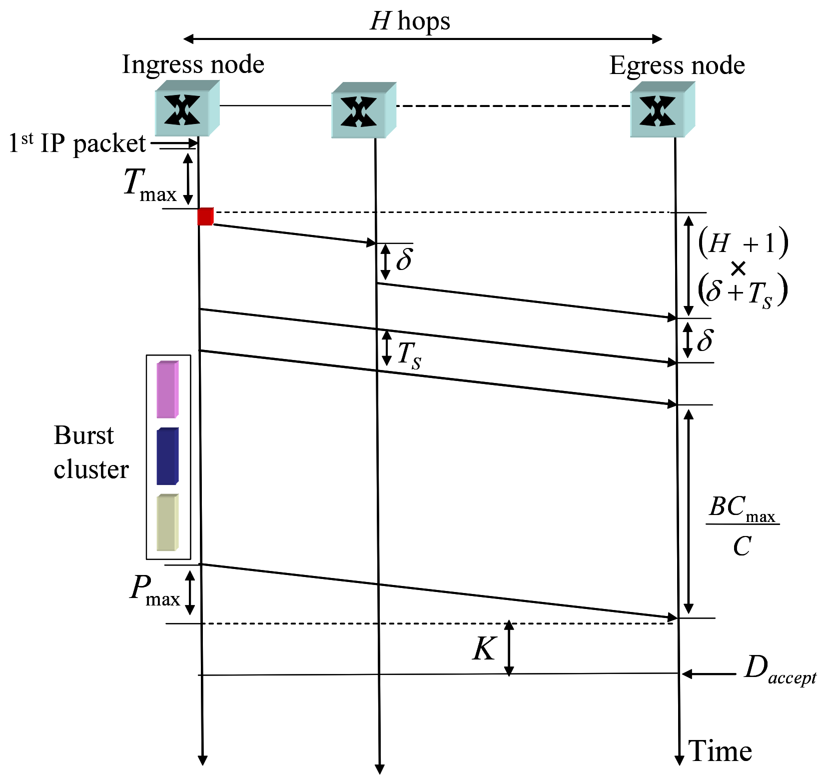


Figure 10. Derivation of K for delay-sensitive traffic.

3.3 Derivation of K for Delay-Tolerant Traffic

In our proposed method, a hop-based burst-cluster has to be stored for a duration k in an ingress node. If the maximum waiting time K is large, the ingress node may not store the generated burst-cluster in its buffer if there is not enough memory. Therefore, for delay-tolerant traffic, K should be determined from the buffer size.

Now, let $B_{tolerant}$ denote the maximum buffer size which is used to store the generated burst-clusters. In addition, let $T_{\min}$ denote the minimal burst-cluster generation time. In this case $B_{tolerant}$ has to be equal to or less than $BC_{\max} + \lceil K/T_{\min} \rceil \times BC_{\max}$, as shown in Fig. 11. As a result, K is derived from the following inequality.

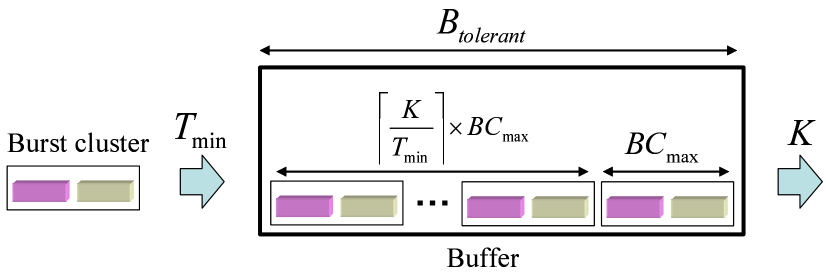


Figure 11. Derivation of K for delay-tolerant traffic.

$$K \leq \left\lfloor \frac{B_{tolerant}}{BC_{\max}} - 1 \right\rfloor \times T_{\min}. \tag{2}$$

In the burst-cluster generation algorithm shown in subsection 2.1, $T_{\min}$ is not defined. Therefore, $T_{\min}$ should be set from the maximum arrival rate of IP packets $\lambda_{\max}$. Otherwise, $T_{\min}$ may be set to a certain value, for example $0.5 \times T_{\max}$, for simplicity.

3.4 Derivation of K for Multiple Types of Traffic

In general, delay-sensitive traffic and delay-tolerant traffic are transmitted in OBS networks together. In this case, we have two kinds of derivations of the maximum waiting time K depending on the node structure. One is a case where an ingress node has two different buffers for delay-sensitive traffic and delay tolerant traffic, and the other is a case where an ingress node has only one buffer.

In the former case, a burst-cluster is generated from each buffer and the two kinds of traffic are never included in the same burst-cluster. Therefore, for a delay-sensitive burst-cluster, K is derived from (1), but for a delay-tolerant burst-cluster, K is derived from (2).

In the latter case, two kinds of traffic are included in the same burst-cluster. Hence the maximum waiting time K should be derived from the following equation.

$$K = \min\{K \text{ of } (1), K \text{ of } (2)\}. \tag{3}$$

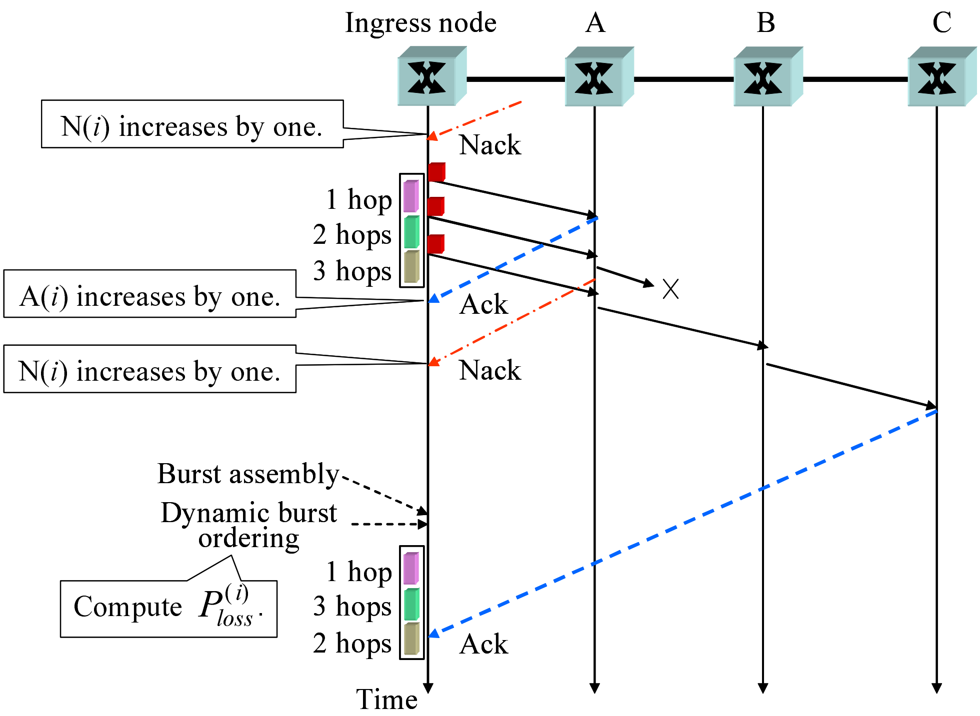


Figure 12. Dynamic burst ordering with ACK and NACK messages.

4. Dynamic Burst Ordering

In order to improve local fairness for each ingress node and improve global fairness significantly, we propose dynamic burst ordering for the hop-based burst-cluster transmission.

In the conventional hop-based burst-cluster transmission, a burst-cluster is generated from multiple bursts, and bursts are always arranged within the burst-cluster in order from the smallest number of hops to the largest one. On the other hand, in our proposed method, the order of bursts within a burst-cluster is changed dynamically.

4.1 Derivation of the Burst Loss Probability for Each Hop

The proposed method utilizes an ACK (NACK) message which is received by an ingress node when the burst transmission succeeds (fails). Let $A(i)$ denote the number of received ACK messages for i-hop and $N(i)$ denote the number of received NACK messages for i-hop. $A(i)$ and $N(i)$ increase by one when the ingress node receives ACK and NACK messages, respectively (see Fig. 12). Then, the burst loss probability for i-hop, $P_{loss}^{(i)}$, is calculated as follows.

$$P_{loss}^{(i)} = \frac{N(i)}{A(i) + N(i)} \ , \tag{4}$$

where the initial value of $P_{loss}^{(i)}$ is equal to zero for every i.

4.2 Dynamic Ordering Based on Burst Loss Probability

When multiple bursts are assembled simultaneously, the ingress node determines the order of bursts in a burst-cluster based on $P_{loss}^{(i)}$. When $P_{loss}^{(i)}$ is smaller than $P_{loss}^{(j)}$, the burst for i hop is arranged ahead of that for j hop. When $P_{loss}^{(i)}$ is equal to $P_{loss}^{(j)}$, the burst with a smaller number of hops is arranged ahead of that with a larger number of hops.

Figure 13 shows how dynamic burst ordering is performed when the maximum number of hops is three. If $P_{loss}^{(1)} \leq P_{loss}^{(2)} \leq P_{loss}^{(3)}$ is satisfied, the order of the three bursts is the same as in the conventional method (see Fig. 13 (a)). Remind that the transmission of a burst in the front part of the burst-cluster succeeds with higher probability than that in the rear part of the burst-cluster.

If the burst loss probabilities satisfy $P_{loss}^{(1)} \leq P_{loss}^{(3)} < P_{loss}^{(2)}$, the order of bursts changes as shown in Fig. 13(b). Moreover, in the case of $P_{loss}^{(3)} < P_{loss}^{(2)} < P_{loss}^{(1)}$, the order of the bursts changes as shown in Fig. 13(c).

The generated burst-cluster is transmitted along with multiple control packets, as is the case with the conventional method. Figure 14 shows how a burst-cluster is forwarded from an ingress node when dynamic burst ordering is used. In this figure, N denotes the number of egress nodes and H is the maximum number of hops for the burst-cluster. S_m and R_m $(1 \leq m \leq N)$ denote the $SETUP$ and the $RELEASE$ messages for the mth burst in the burst-cluster.

(a) Case 1: $P_{loss}^{(1)} \leq P_{loss}^{(2)} \leq P_{loss}^{(3)}$.

(b) Case 2: $P_{loss}^{(1)} \leq P_{loss}^{(3)} < P_{loss}^{(2)}$.

(c) Case 3: $P_{loss}^{(3)} < P_{loss}^{(2)} < P_{loss}^{(1)}$.

Figure 13. Dynamic burst ordering based on the calculated burst loss probabilities.

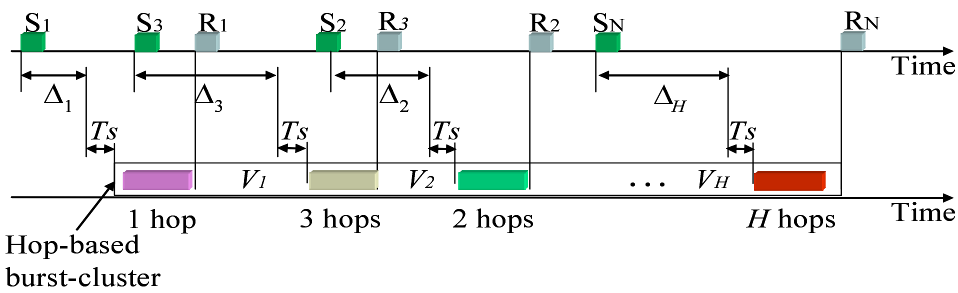
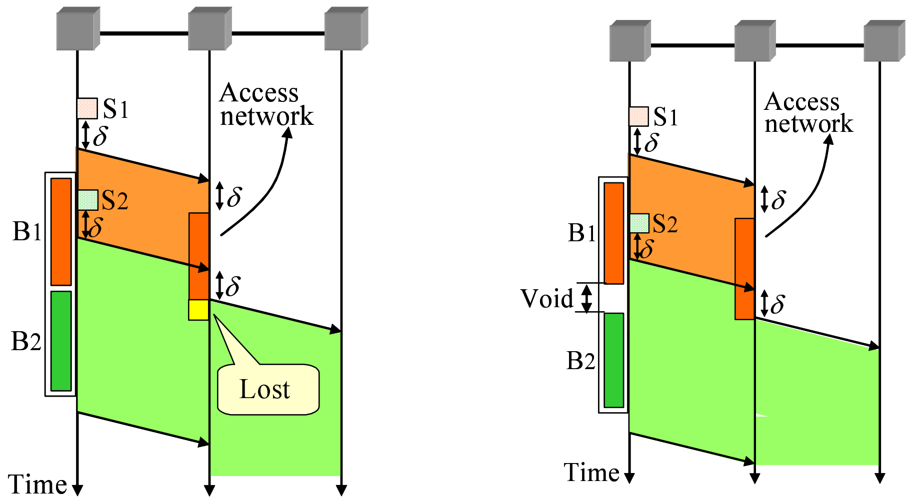


Figure 14. Hop-based burst-cluster transmission with dynamic burst ordering.

4.3 Derivation of the Void Size

In Fig. 14, hop-based burst-cluster transmission requires a void between two consecutive bursts although it increases the wavelength reservation time. Here, Fig. 15(a) shows a case where there is no void. As shown in this figure, when two bursts are forwarded to different output links, a preceding burst is preempted by the next one. This preemption occurs even if the number of hops of the next node is smaller than that of the preceding burst. In order to avoid such an undesirable preemption, a void is used, as shown in Fig. 15(b).

The size of each void can be determined from the number of hops of the next burst, the processing time of a control packet δ, the switching time T_s, and at which transmission hop the two bursts are switched to different output links. In



(a) Without void between two bursts. (b) With void between two bursts.

Figure 15. The impact of void on burst preemption.

order to decrease the redundant wavelength reservation, the accurate size of each void is required. On the other hand, in the proposed method, the order of bursts changes dynamically. In addition, the processing time of a control packet δ is very small. Therefore, for the simple implementation, we set the size of a void to $(i-1)\times\delta+T_s$ when the number of hops for the next node is i. This is denoted in Fig. 14 with V_1, V_2 through V_H. For instance, V_1 is the void necessary between the first burst (1 transmission hop) and the second burst (3 transmission hops), therefore $V_1=(3-1)\times\delta+T_s=2\delta+T_s$.

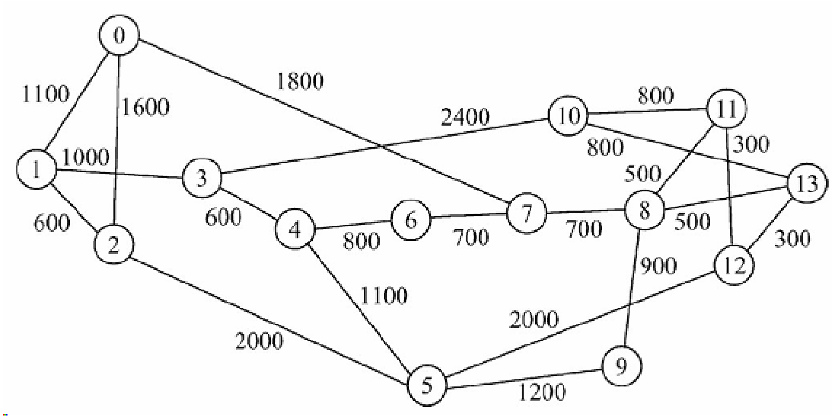


Figure 16. NSFNET topology.

5. Numerical Examples

5.1 Performance Evaluation of Random Scheduling

In this section, we evaluate the effectiveness of random scheduling for NSFNET by simulation (see Fig. 16). In NSFNET, the number of nodes is 14 and the number of links is 21. The number of wavelengths at each link is eight and the transmission speed of a wavelength is 10 Gbps. Please note that in NSFNET, the maximum number of hops between any ingress node and egress node is three. The distance between adjacent nodes is from 300 km to 2,800 km, and a static route between ingress and egress nodes is chosen according to the minimum hop routing. In addition, the processing time of a control packet is equal to 1.0 ms and the optical switching time is 0.1 ms.

We assume that delay-sensitive IP packets arrive to NSFNET according to the Poisson process with rate λ [packets/μs]. Ingress and egress nodes of an arriving IP packet are selected at random. The size of an IP packet is fixed equal to 1,250 bytes. We assume that the maximum acceptable delay of every IP packet is 55 ms.

In this network, we evaluate the performance of the burst-cluster transmission with random scheduling. In the proposed method, we use three pairs of $T_{\max}$

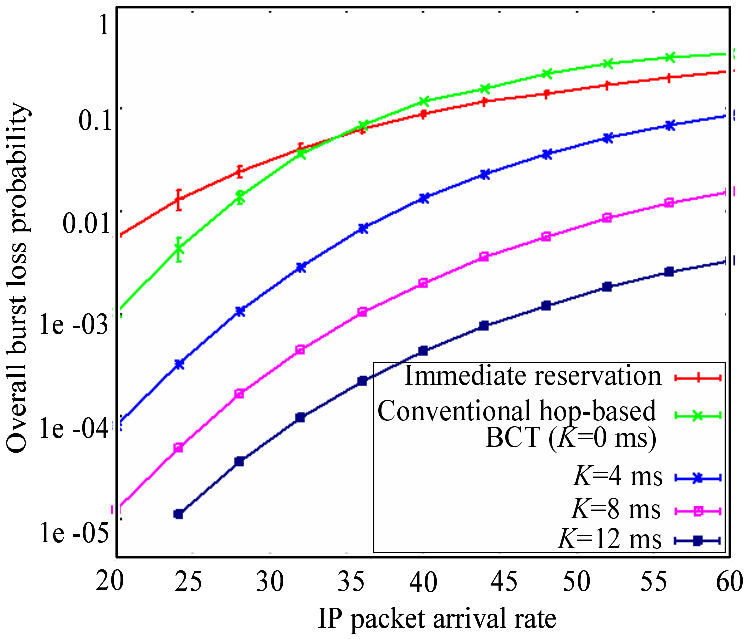


Figure 17. Burst loss probability vs. arrival rate.

and $BC_{\max}$ such as $T_{\max} = 5.0$ ms and $BC_{\max} = 50$ Mbits, $T_{\max} = 8.0$ ms and $BC_{\max} = 20$ Mbits, or $T_{\max} = 1.0$ ms and $BC_{\max} = 90$ Mbits. Here, the propagation time is between node 1 and node 13 is maximum (28 ms) in this network. Therefore, from (1), for nodes 1 and 13, K is equal to or less than $55 - \{5.0 + 4 \times (1 + 0.1) + 28 + 50/10\} = 55 - \{8.0 + 4 \times (1 + 0.1) + 28 + 20/10\} = 55 - \{1.0 + 4 \times (1 + 0.1) + 28 + 90/10\} = 12.6$ ms. In the following, we assume that every node uses the same maximum waiting time $K \leq 12$ ms, due to the simple implementation except for subsection 4.4. In addition, we evaluate the performance of the conventional hop-based burst-cluster transmission and that of the original immediate reservation.

5.1.1 Effect on the Overall Loss Probability

In this subsection, we investigate the impact of random scheduling on the overall burst loss probability. $T_{\max}$ and $BC_{\max}$ are set to 5.0 ms and 50 Mbits respectively. Figure 17 shows the overall burst loss probability against the arrival rate λ. In

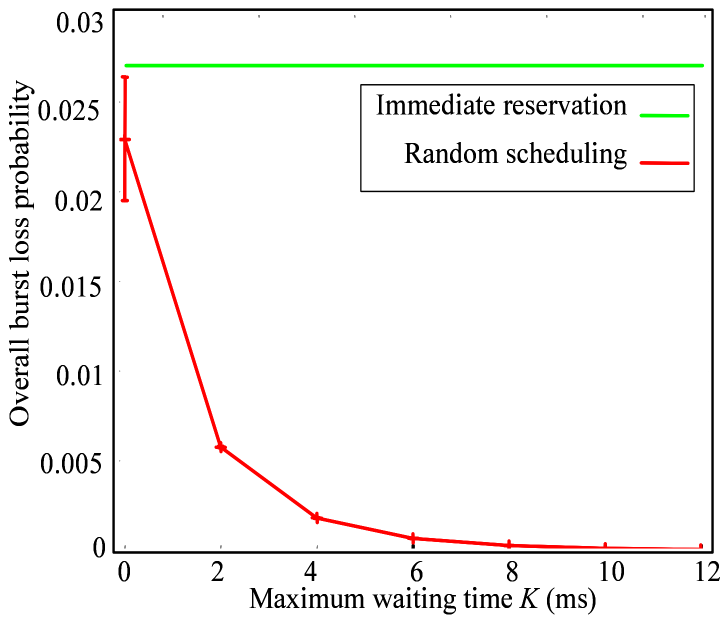


Figure 18. Burst loss probability vs. K when $\lambda = 30$.

this figure, a result for $K = 0$ denotes a result for the conventional hop-based burst-cluster transmission. From this figure, we find that the overall burst loss probabilities for all cases become large as the arrival rate increases. In addition, we find that when λ is larger than 36, the conventional hop-based burst-cluster transmission has a larger burst loss probability than the immediate reservation. This denotes that the impact of synchronization on the hop-based burst-cluster transmission is larger than that on the original immediate reservation when the arrival rate is large.

As for random scheduling, Fig. 17 shows the overall burst loss probabilities for three cases of K. Note that all cases satisfy inequality (1). From these results, we find that the overall burst loss probability is the smallest when K is 12 ms regardless of the arrival rate.

In addition, Figs. 18 and 19 show how the overall burst loss probability changes as K becomes large in cases of $\lambda = 30$ and $\lambda = 50$, respectively. Note that the overall burst loss probability for the immediate reservation is constant against K

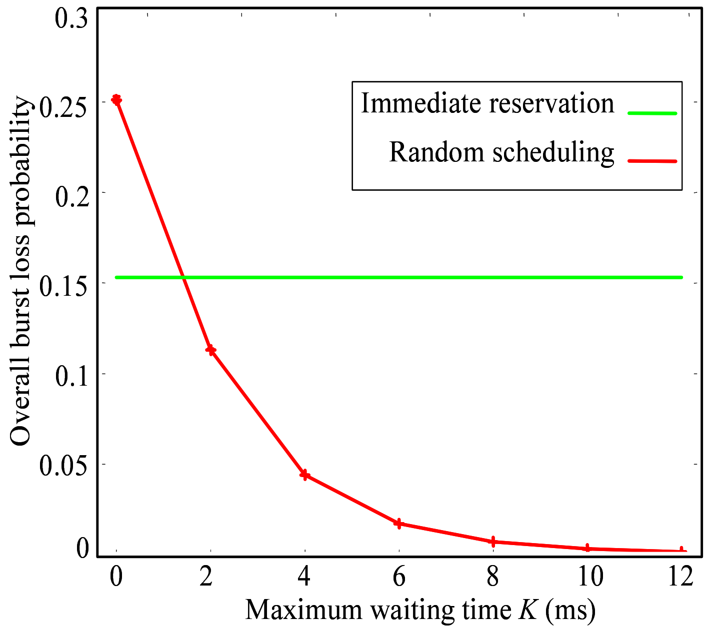


Figure 19. Burst loss probability vs. K when $\lambda = 50$.

because this method does not use random scheduling. From these two figures, we find that the overall burst loss probability becomes small as K becomes large. These results show that random scheduling is effective for decreasing the overall burst loss probability by resolving the synchronization issue.

5.1.2 Effect on the Maximum Transmission Delay

In this subsection, we investigate how random scheduling affects the maximum transmission delay. Figure 20 shows the maximum burst transmission delay in the case where $T_{\max}$ is 1.0 ms and $BC_{\max}$ is 90 Mbits, and Fig. 21 shows the maximum burst transmission delay in the case where $T_{\max}$ is 8.0 ms and $BC_{\max}$ is 20 Mbits.

From Fig. 20, we find that the maximum transmission delay increases as the arrival rate becomes large. In this figure, most burst-clusters are generated based on timer because $T_{\max}$ is small. Therefore, the size of the burst-cluster increases

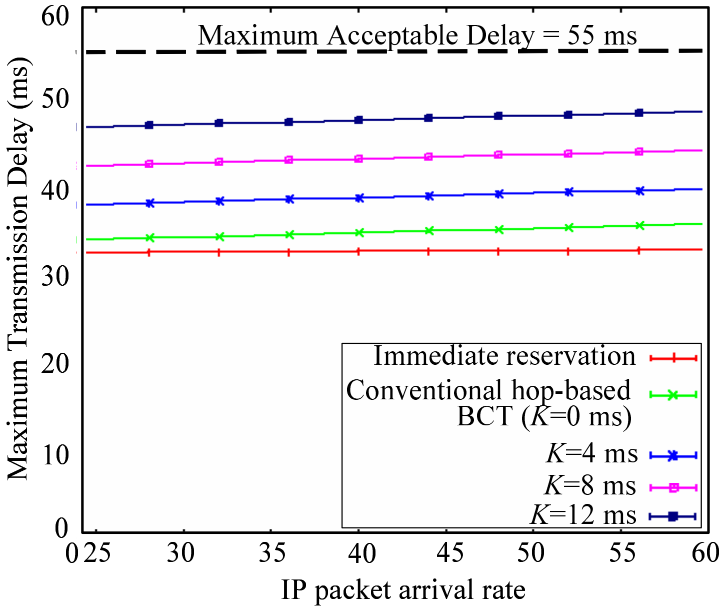


Figure 20. Maximum transmission delay vs. arrival rate when $T_{\max} = 1.0$ ms and $BC_{\max} = 90$ Mbits.

as the arrival rate becomes large. As a result, it takes much time for the rearmost IP packet within a burst to arrive at its egress node.

On the other hand, from Fig. 21, we find that the maximum transmission delay decreases as the arrival rate becomes large. In this case, most burst-clusters are generated when the size of the burst-cluster becomes $BC_{\max}$ due to a small $BC_{\max}$. Therefore, the generation time of a burst-cluster becomes small as the arrival rate increases. Consequently, the rearmost IP packet within a burst can arrive at its egress earlier.

In both cases, the maximum transmission delay increases as K becomes large. This is because a large K results in a large waiting time for each burst-cluster. Nevertheless, the maximum transmission delays for all cases satisfy the maximum acceptable transmission delay by deriving K from (1). Hence, random scheduling is available for delay-sensitive traffic.

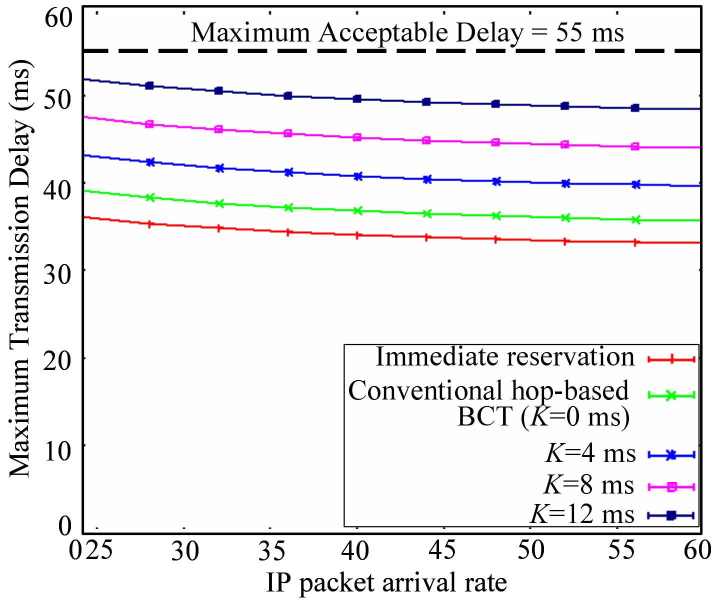


Figure 21. Maximum transmission delay vs. arrival rate when $T_{\max} = 8.0$ ms and $BC_{\max} = 20$ Mbits.

5.1.3 Effect on Fairness Improvement

In addition, we investigate how random scheduling affects fairness improvement. Figure 22 shows the loss probability for the i-hop bursts $P^{(i)}_{loss}$ in the case of $\lambda = 22$ for Fig. 17. From this figure, we find that the discrepancies among $P^{(i)}_{loss}$'s are significantly small in the proposed method. It is expected that random scheduling can also improve fairness. In the following, in order to evaluate fairness in terms of the number of hops, we use the fairness index [20] which is defined as follows.

$$Fairness\ Index = \frac{(\sum_{i=1}^{H} P^{(i)}_{loss})^2}{H \times \sum_{i=1}^{H} {P^{(i)}_{loss}}^2}\,, \tag{5}$$

where H is the maximum number of hops, i.e., H is equal to three for NSFNET.

Figure 23 shows the fairness index against the overall burst loss probability. When the fairness index is close (not close) to one, this index denotes that fairness is improved (not improved). From this figure, we find that the fairness index approaches to one as K becomes large. Therefore, random scheduling is also effective for improving fairness.

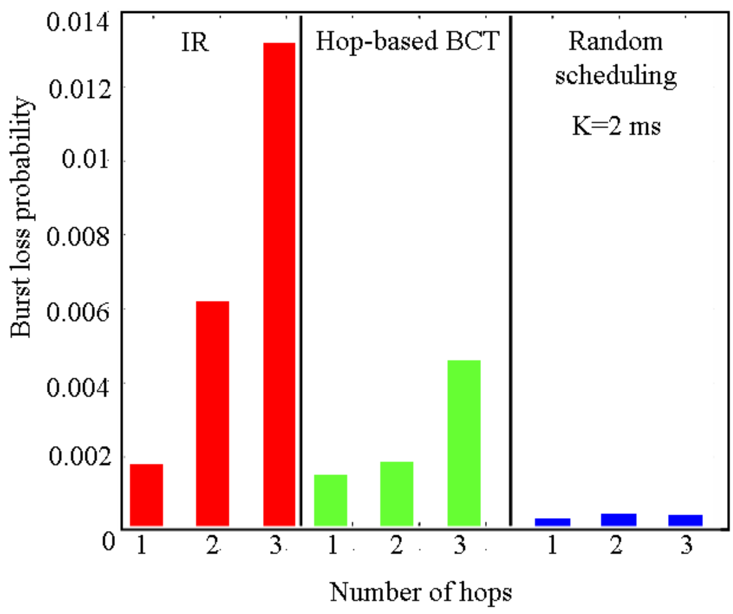


Figure 22. Burst loss probability for each number of hops in the case of $\lambda = 22$.

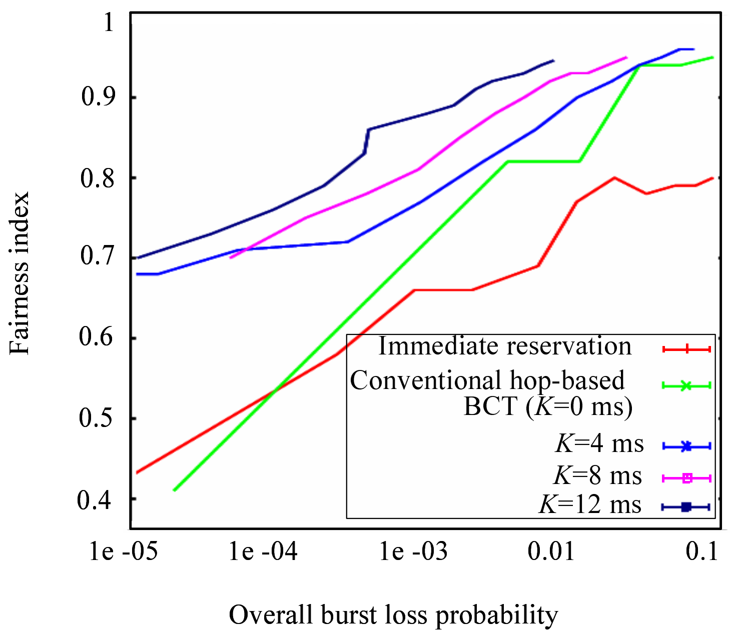


Figure 23. Fairness index vs. overall burst loss probability.

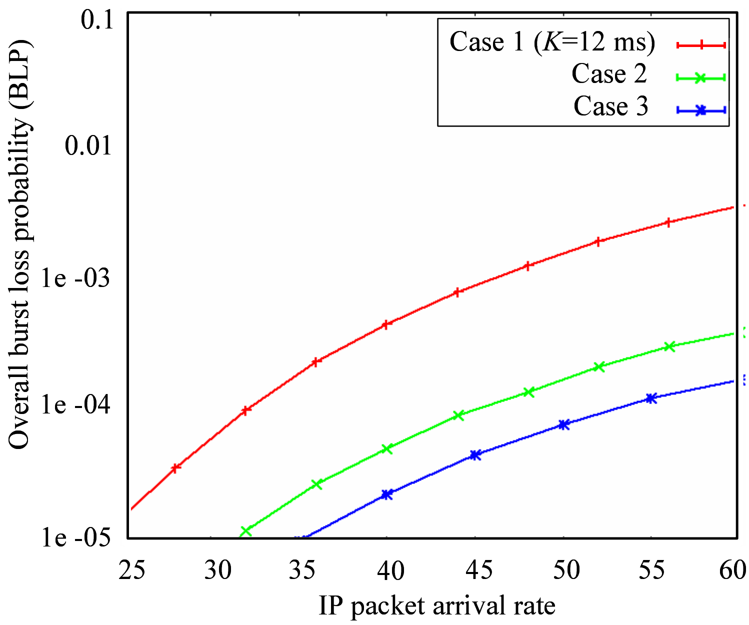


Figure 24. Burst loss probability vs. arrival rate.

5.1.4 Impact of the Maximum Waiting Time K

In the previous subsections, every node uses the same maximum waiting time K for each output link based on $P_{\max} = 28$ ms, which is the maximum propagation delay of the whole network (case 1). In fact, the maximum propagation delay for each ingress node is smaller than 28 ms, and hence each ingress node can use a larger K according to (1) (case 2). In addition, each node can use a different K for each of its output links (case 3).

In this subsection, we investigate how the performance of the proposed method changes for the above cases. Figure 24 shows the overall burst loss probability against the arrival rate λ. Here, all parameters are the same as in Fig. 17. In Fig. 24, for case 1 K is equal to 12 ms.

From Fig. 24, we can find that the overall burst loss probability for case 1 is the largest and that for case 3 is the smallest regardless of the arrival rate. This is because in case 3, many nodes has a local K that is higher than 12 ms.

Figure 25 shows the fairness index in the same situation. From this figure, the

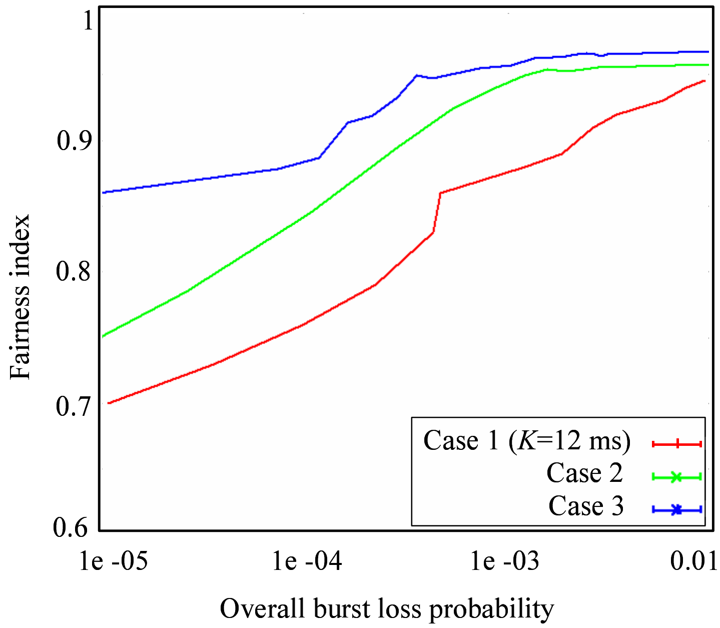


Figure 25. Fairness index vs. burst loss probability.

fairness indexes of case 2 and case 3 are larger than that of case 1. In addition, the fairness index of case 3 is the largest. As a result, the overall burst loss probability decreases and the fairness index is improved significantly when each node uses a different value of K for each output link.

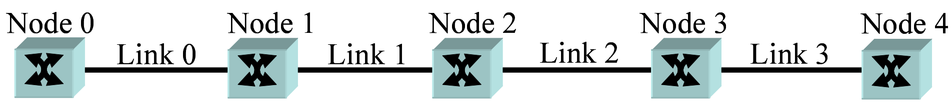


Figure 26. A tandem network with five nodes.

5.2 Performance Evaluation of Dynamic Burst Ordering

In this subsection, we evaluate the performance of our proposed dynamic burst ordering in a tandem network with five nodes by simulation. Figure 26 shows a tandem network, where each node and each link are numbered. The number of wavelengths on each link is eight and the transmission speed of a wavelength is 10 Gbps. The length of each link is 200 km. In addition, the processing time of a control packet is equal to 1.0 μs and the optical switching time is 1.0 μs.

We assume that IP packets arrive at the tandem network according to the Poisson process with rate 200 [packets/μs]. Ingress and egress nodes of an arriving IP packet are selected at random. The size of an IP packet is fixed equal to 1,250 bytes.

From the arriving IP packets, a burst-cluster is generated according to the timer/threshold based assembly algorithm, where the timeout value is 10 ms and the maximum burst-cluster size is 60 Mbits. The order of bursts is determined by using the dynamic burst ordering, and the generated burst-cluster is transmitted from the ingress node. We assume that the time interval between consecutive burst-cluster transmissions at the ingress node is exponentially distributed with rate λ [clusters/ms].

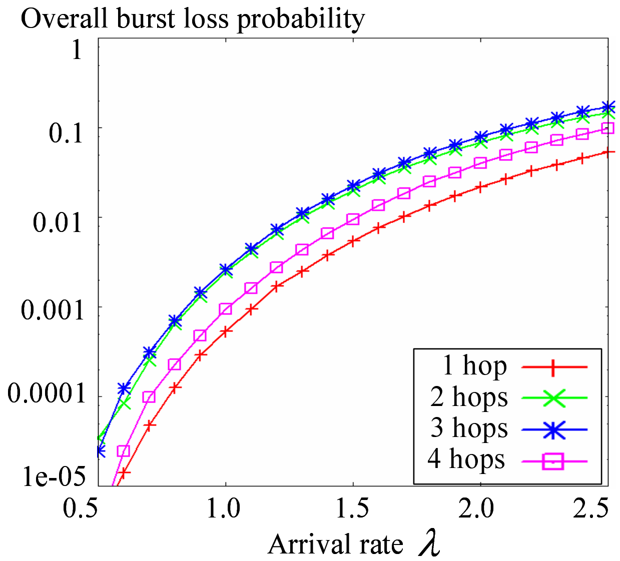
For the performance comparison, we also evaluate the performance of the conventional hop-based burst-cluster transmission where dynamic burst ordering is not used. In addition, we evaluate the performance of the original immediate reservation. For this method, we set the maximum burst size to 20 Mbits so that the burst sizes for the three methods are almost the same.

5.2.1 Impact on Local Fairness

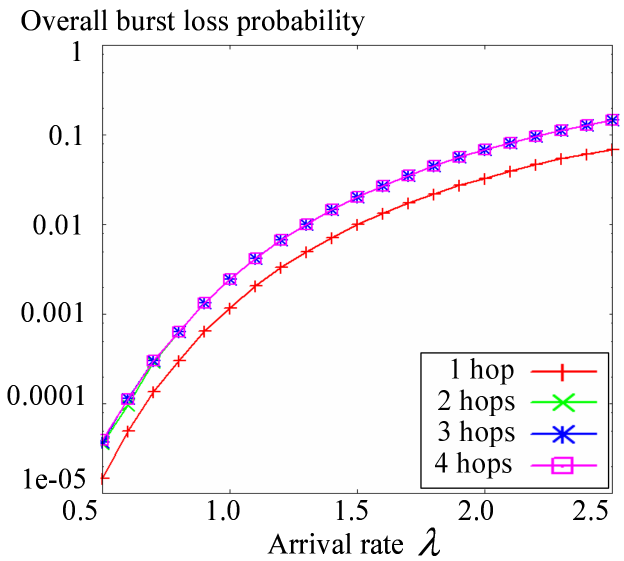
First, we investigate the impact of dynamic burst ordering on local fairness. Figure 27 shows the burst loss probability of each number of hops for the burst-cluster transmission from node 0 with output link 0. From Fig. 27(a), for the conventional method, we find that the burst loss probabilities of one hop and four hops are smaller than those of two hops and three hops. This is because the amount of traffic on link 0 and 3 are small, which are the last-hop links for one-hop burst and four-hop burst, respectively. Therefore the conventional burst-cluster transmission can not improve local fairness anymore. However, as shown in Fig. 27(b), the proposed method can improve the local fairness so that the burst loss probabilities of two hops and three hops never exceed that of four hops.

Figure 28 shows the burst loss probability of each number of hops from node 3 with output link 2. Note that there is no burst transmission of four hops. From Figs. 28(a) and (b), we find that the conventional method can not improve the local fairness but the proposed method can provide almost the same burst loss probability for each number of hops.

Figure 29 shows the fairness index [20] against the overall burst loss probability for three pairs of ingress node and output link. Here, when the fairness index is close (not close) to one, this index denotes that fairness is improved (not improved). From this figure, we find for all three cases, that the fairness index of dynamic burst ordering is much closer to one compared to that of the conventional hop-based burst-cluster transmission. Therefore, dynamic burst ordering is effective for improving fairness.

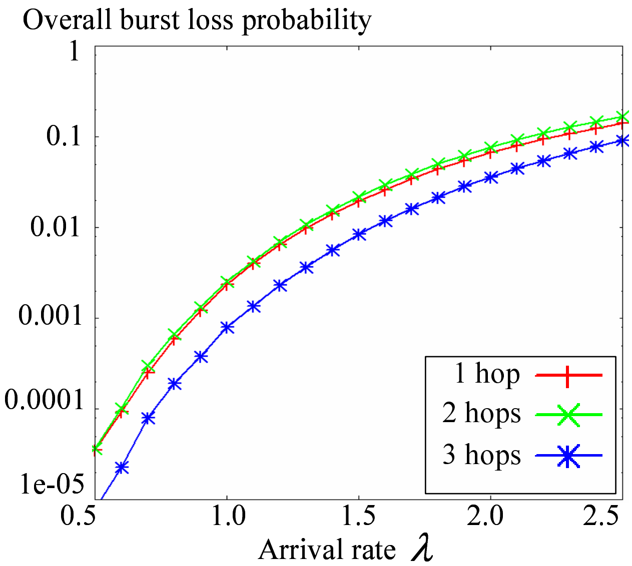


(a) Hop-based burst-cluster transmission.

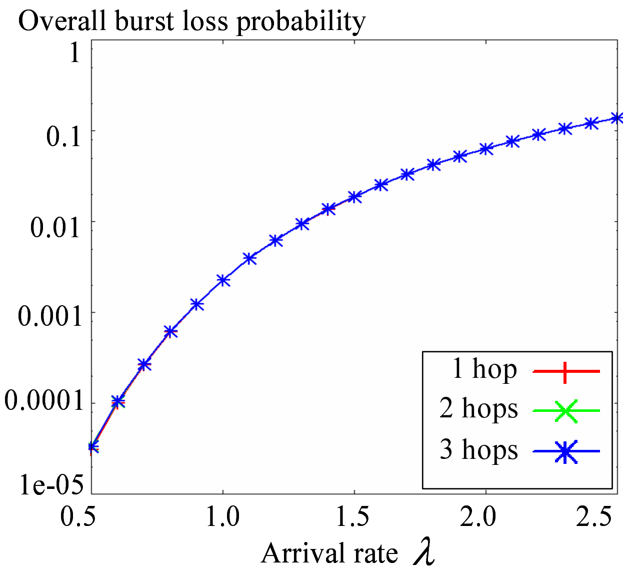


(b) Dynamic burst ordering.

Figure 27. Burst loss probability vs. arrival rate (node 0 and link 0).



(a) Hop-based burst-cluster transmission.



(b) Dynamic burst ordering.

Figure 28. Burst loss probability vs. arrival rate (node 3 and link 2).

5.2.2 Impact on Global Fairness

Next, we investigate the impact of dynamic burst ordering on global fairness. Figure 30 shows the fairness index against the overall burst loss probability in the tandem network. For the performance comparison, a result of the original immediate reservation is also shown. From this figure, we find that the conventional hop-based burst-cluster transmission can improve fairness further than the original immediate reservation, as expected. Besides, by using the proposed dynamic burst ordering, fairness can be improved significantly.

Figures 31(a) and (b) show the burst loss probability of each number of hops, in this network, for the conventional method and the proposed method, respectively. From these figures, we find that the discrepancies among the burst loss probabilities for the proposed method are smaller than those for the conventional method.

5.2.3 Impact on the Burst Loss Probability

Finally, we investigate how the burst loss probability changes by using the proposed method. Figure 32 shows the overall burst loss probabilities for the proposed method, the conventional hop-based burst-cluster transmission, and the original immediate reservation. From this figure, we find that the overall burst loss probability for the proposed method is larger than that for the conventional hop-based burst-cluster transmission. This is because burst loss probabilities which are small by using the conventional hop-based burst-cluster transmission increase in order to improve local fairness. Nevertheless, the proposed method can provide a smaller overall burst loss probability than the original immediate reservation.

Table 1 shows the largest burst loss probabilities and its number of hops for some pairs of ingress node and output link. From this table, it is shown that with the proposed method, each ingress node can decrease the largest burst loss probability among the ones for all egress nodes further than the conventional hop-based burst-cluster transmission. These results denote that all bursts can use wavelengths more fairly.

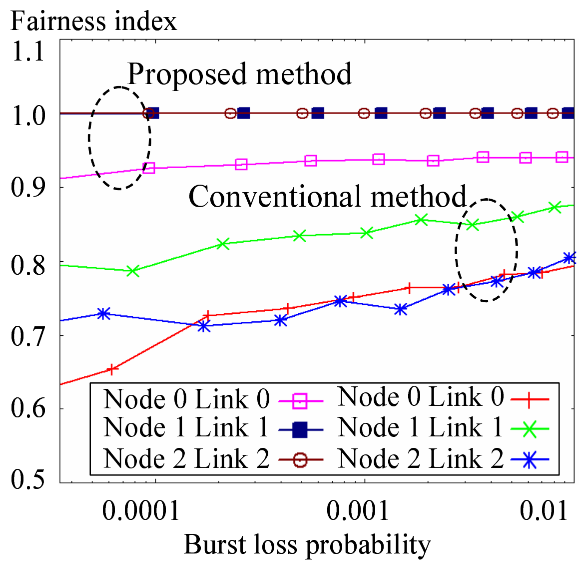


Figure 29. Fairness index vs. overall burst loss probability for each pair of ingress node and output link.

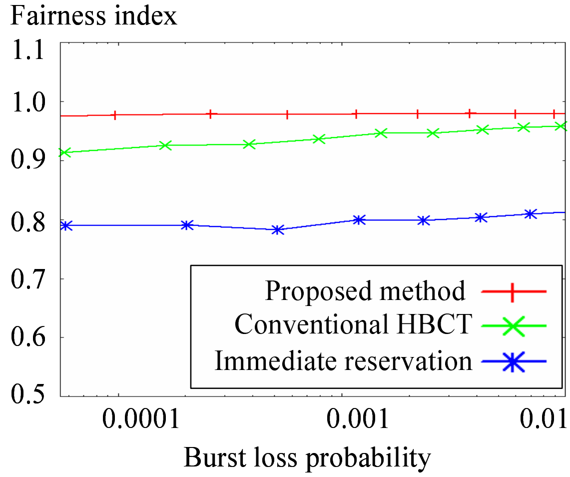
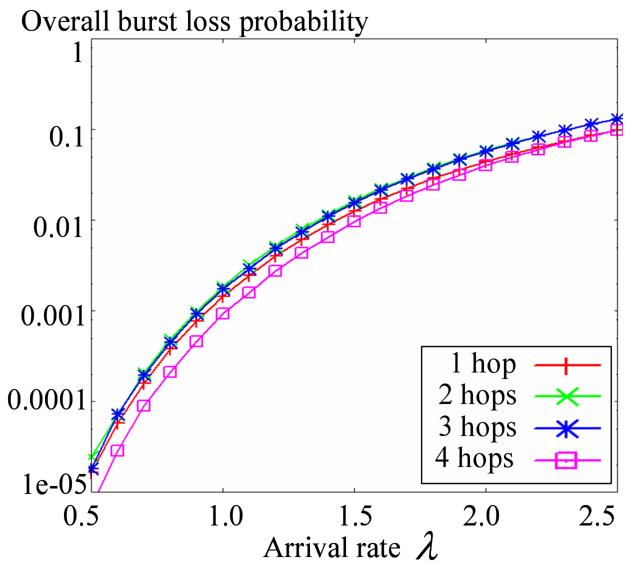
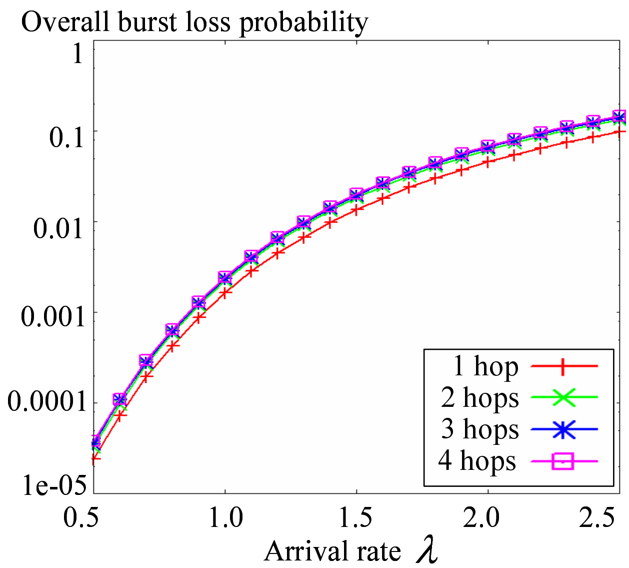


Figure 30. Fairness index in the tandem network vs. overall burst loss probability.



(a) Hop-based burst-cluster transmission.



(b) Dynamic burst ordering.

Figure 31. Burst loss probability in the tandem network vs. arrival rate.

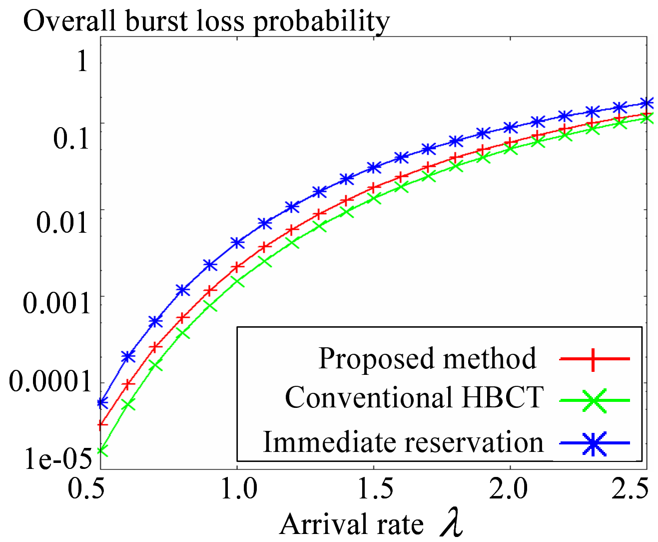


Figure 32. Overall burst loss probability vs. arrival rate.

Table 1. The largest burst loss probability for each pair of ingress node and output link when λ=1.0.

(Ingress node, Output link)	Conventional method		Proposed method	
	Largest loss prob.	Number of hops	Largest loss prob.	Number of Hops
(Node 0, Link 0)	2.64e-03	3 hop	2.45e-03	4 hop
(Node 1, Link 1)	2.44e-03	2 hop	2.27e-03	3 hop
(Node 2, Link 1)	2.37e-03	1 hop	1.94e-03	2 hop
(Node 2, Link 2)	2.39e-03	1 hop	1.93e-03	2 hop
(Node 3, Link 2)	2.53e-03	2 hop	2.28e-03	3 hop
(Node 4, Link 3)	2.72e-03	3 hop	2.50e-03	4 hop

6. Conclusions

In this thesis, we proposed the utilization of random scheduling and dynamic burst ordering for the hop-based burst-cluster transmission in order to resolve the synchronization issue and the dynamic traffic load issue, respectively.

Random scheduling can be used for delay-sensitive traffic, delay-tolerant traffic, and multiple types of traffic, by considering the acceptable transmission delay and the buffer size. We evaluated the performance of random scheduling by simulation for NSFNET. From the numerical examples, we found that random scheduling is effective for decreasing the overall burst loss probability. We also found that the proposed method is available for delay-sensitive traffic although the transmission delay becomes large. In addition, random scheduling can improve fairness more efficiently than the conventional hop-based burst-cluster transmission. If the maximum waiting time is determined for each ingress node, the overall burst loss probability and the fairness are improved significantly.

Dynamic burst ordering improves the local fairness for each node regardless of the amount of traffic on each link. We evaluated by simulation the performance of dynamic burst ordering for tandem networks. From the numerical examples, we found that the burst loss probabilities become almost the same for each ingress node by using the proposed method. In addition, the fairness index for the proposed method is much close to one, and we found that the proposed method can improve global fairness in the tandem network significantly. As a future work, it would be interesting to investigate the impact of the proposed method on the overall burst loss probability as well as the performance of the proposed method in other network topologies such as ARPA2. Moreover, this method will be extended so that exponential moving average is used to estimate burst loss probabilities.

Acknowledgements

This is definitely not the work of a single person. I am very much grateful to Professor Kenji Sugimoto of Nara Institute of Science and Technology, for his helpful comments and continuous encouragement. I am very thankful to Professor Kotaro Minato and Associate Professor Yuichi Kaji for their kind advice and comment. I appreciate very much Associate Professor Kentaro Hirata for his kind input and comments during laboratory meetings. I would like, particularly, to express my utmost gratitude and warmest regards to Assistant Professor Takuji Tachibana for his continuous support, valuable advice and guidance. He taught me a lot about research. Moreover, I am thankful to Assistant Professor Kiminao Kogiso for his suggestions. I also would like to thank Mrs. Yoko Hashimoto for helping me through the various administrative procedures concerning conferences. Finally, I would like to thank all members of the Systems Science Laboratory for their kind assistance and making my stay at this lab a memorable one.

References

[1] M. Yoo, C. Qiao, and S. Dixit, "QoS Performance of Optical Burst Switching in IP-Over-WDM Networks," *IEEE J. Select. Areas Comm.,* vol. 18, no. 10, pp. 2062-2071, Oct. 2000.

[2] M. Murata, K. Kitayama, and H. Miyahara, "IP Over a-Thousand-Wavelength Division Multiplexing: Is It Useful and Possible for Resolving the Network Bottlenecks?," *Optical Networks Magazine,* vol. 99, no. 674, pp. 55-60, Nov. 2000.

[3] P. Bonenfant and A. R. Moral, "Optical Data Networking," *IEEE Comm. Magazine,* vol. 38, no. 3, pp. 63-70, March 2000.

[4] I. Chlamtac, V. Elek, A. Fumagalli, and C. Szabo, "Scalable WDM Access Network Architecture Based on Photonic Slot Routing," *IEEE/ACM Trans. Networking,* vol. 7, no. 1, pp. 1-9, Feb. 1999.

[5] N. Ghani, S. Dixit, and T. S. Wang, "On IP-over-WDM Integration," *IEEE Comm. Magazine,* vol. 38, no. 3, pp. 72-84, March 2000.

[6] M. W. McKinnon, H. G. Perros, and G. N. Rouskas, "Performance Analysis of Broadcast WDM Networks Under IP Traffic," *Performance Evaluation,* vol. 36-37, pp. 333-358, Aug. 1999.

[7] E. Modiano, "Random Algorithms for Scheduling Multicast Traffic in WDM Broadcast-and-Select Networks," *IEEE/ACM Trans. Networking,* vol. 7, no. 3, pp. 425-434, June 1999.

[8] K. Struyve, N. Wauters, P. Falcao, P. Arijis, D. Colle, P. Demeester, and P. Lagasse, "Application, Design, and Evolution of WDM in GTS's Pan-European Transport Network," *IEEE Comm. Magazine,* vol. 38, no. 3, pp. 114-121, March 2000.

[9] C. Qiao and M. Woo, "Optical Burst Switching (OBS) - a New Paradigm for an Optical Internet," *Journal of High Speed Networks,* vol. 8, no. 1, pp. 69-84, March 1999.

[10] Y. Wei and R. I. McFarland, "Just-In-Time Signaling for WDM Optical Burst Switching Networks," *Journal of Lightwave Technology,* vol. 18, no. 12, pp. 2019-2037, Dec. 2000.

[11] B. Zhou, M. A. Bassiouni, and G. Li, "Improving Fairness in Optical Burst Switching Networks," *Journal of Optical Networking*, vol. 3, no. 4, pp. 214-228, April 2004.

[12] M. Ueda, T. Tachibana, and S. Kasahara, "Last-Hop Preemptive Scheme Based on the Number of Hops for Optical Burst Switching Networks," *The Journal of Optical Networking*, vol. 4, no. 10, pp. 648-660, April 2005.

[13] M. Ueda, T. Tachibana, and S. Kasahara, "Intermediate-Hop Preemption to Improve Fairness in Optical Burst Switching Networks," *IEICE Transactions on Communications*, Vol. E91-B, no. 3, pp. 1290-1298, Oct. 2005.

[14] T. Orawiwattanakul and Y. Ji, "Preemption Scheme for Improving Source Level Fairness in Optical Burst Switching," *Proc. of 4th International Conference on Innovations in Information Technology,* pp. 710-714, Nov. 2007.

[15] B. Zhou and M. A. Bassiouni, "Concurrent Enhancement of Network Throughput and Fairness in Optical Burst Switching Environments," *Photonic Network Communications,* vol. 14, no. 2, pp. 199-207, July 2007.

[16] T. Tachibana, S. Kasahara, and H. Harai, "Burst-Cluster Transmission Based on the Number of Hops for Improving Fairness in Optical Burst Switching Networks," in *Proc. IEEE Globecom 2006*, Dec. 2006.

[17] T. Tachibana and S. Kasahara, "Burst-Cluster Transmission: Service Differentiation Mechanism for Immediate Reservation in Optical Burst Switching Networks," *IEEE Commun. Mag.*, vol. 44, no. 5, pp. 46-55, May 2006.

[18] C. Yuan, Z. Li, and A. Xu, "A Novel Burst Assembly Algorithm for OBS Networks Based on Data-Length Time-Lag Product," *Proc. of Asia-Pacific Conference*, pp. 319-323, Oct. 2005.

[19] "Transmission Impairments Due to Speech Processing," ITU-T Recommendation G. 113, May 2002.

[20] R. Jain, D. Chiu, and W. Hawe, "A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer systems," *Tech. Rep. DEC-TR-301, Digital Equipment Corporation,* July 1984.