4-4 Advanced Lightpath Establishment for Distributed Computing

TACHIBANA Takuji and XU Sugang

Currently, it is expected that distributed computing environment is developed in wide-area networks by using wavelength division multiplexing (WDM) and lightpath switching. In this paper, in order to develop the distributed computing environment over lightpath switching networks, we study two new lightpath establishment approaches. These two approaches enable lightpaths to be effectively established in wide-area WDM networks and enable optical ring to be dynamically developed for the data transmission in multiple points.

Keywords
Lightpath switching, Grid, RSVP-TE, Multiple domains, Optical ring

1 Introduction

For the past several years, construction of a wide-area distributed computing environment (or “grid”) has been underway over the Internet. However, the difficulty involved in securing acceptable transmission delay and bandwidth over the Internet poses a significant problem in constructing such an environment, which must be able to handle large volumes of data on a scale from terabytes to petabytes.

Accordingly, in[2]-[5] the current feasibility of constructing a distributed computing environment over a lightpath network that makes use of wavelength division multiplexing (WDM) and lightpath switching[1] is examined. The lightpath network establishes a path on wavelength for data transmission (the “lightpath”) between two nodes (the source and destination nodes) and then transmits the data along the lightpath. By establishing this lightpath such that the endhosts performing the distributed computing are directly connected, it is possible to guarantee bandwidths larger than 1 Gbps between the endhosts.

When such a distributed computing environment is spread out over a wide area, connecting endhosts may require the establishment of a lightpath that passes through multiple domains. However, the network configuration within a single domain is often complex, and each domain is normally managed in accordance with a specific policy. Therefore, one domain cannot be expected to share the wavelength-usage information within the domain with neighboring domains freely. Thus, when establishing a lightpath through multiple domains it is necessary to allow masking—to the full extent possible—of the wavelength-usage information of the individual domains. In order to overcome this problem, the authors have proposed a method of establishing an end-to-end lightpath based on rank accounting[6][7], which makes efficient use of this wavelength-usage information about a limited number of wavelengths. A performance evaluation based on simulations has verified the effectiveness of the proposed method in the establishment of a lightpath passing through multiple domains.

For distributed computing, networks must be able to offer not only point to point data transmissions, but also provide for multipoint data transmission. The authors have proposed a technology for the construction of a lightpath and a management method for multipoint data transmission to allow for both types of transmission[8][9]. This system enables the construction of a wide-area distributed computing environment with guaranteed transmission between endhosts by taking full advantage of technologies in optical communication, photonic networks, and distributed computing. The proposed method has been implemented and tested. Both the effectiveness of the method and of the implemented system[9] have been verified through experiments.

This paper describes the lightpath establishment technology used for a lightpath network and presents an efficient method for establishing lightpaths passing though multiple domains. In addition, we also present a method for constructing a wide-area distributed computing environment with a unidirectional optical ring, also using lightpaths.

2 Lightpath establishment using RSVP-TE

Lightpath establishment using Resource Reservation Protocol-Traffic Engineering (RSVP-TE) signaling protocol for Generalized Multi-Protocol Label Switching (GMPLS) is currently under consideration for use in WDM networks[10][11]. Figure 1 shows an example of lightpath establishment by RSVP-TE signaling. With RSVP-TE, a lightpath is established using the *Path* and *Resv* messages.

The example presented in Fig. 1 corresponds to a case in which source node i receives a request for lightpath establishment from the user to destination node j. Node i generates a *Path* message and embeds the information on the wavelengths available for the next link within the message. In Fig. 1, $\{\lambda_1, \lambda_2, \lambda_4, \lambda_5, \lambda_6, \lambda_8\}$ are specified. Source node i then sends the *Path* message to destination node j.

When an intermediate node receives the *Path* message, it deletes any wavelengths unavailable for the next link from the wavelength-usage information contained within the *Path* message and then sends the message to destination node j.

When the destination node j receives the *Path* message, it selects a wavelength for the lightpath from among those indicated as available within the wavelength information in the *Path* message, in this case $\{\lambda_1, \lambda_4, \lambda_6, \lambda_8\}$. If, for example, the selected wavelength λ_4 is available for use, then this wavelength is reserved for the establishment of the lightpath. Destination node j then generates a *Resv* message specifying the selected wavelength (λ_4) and sends the *Resv* message to source node i.

Upon receipt of the *Resv* message, an intermediate node checks the usage status of the specified wavelength λ_4 of the message. If this wavelength is available for use, the intermediate node reserves it and sends the *Resv* message to source node i. When wavelength λ_4 is already occupied by another lightpath, the attempt at lightpath establishment fails, and a *ResvErr* message is sent to source node i. The lightpath is successfully established when source node i receives the *Resv* message.

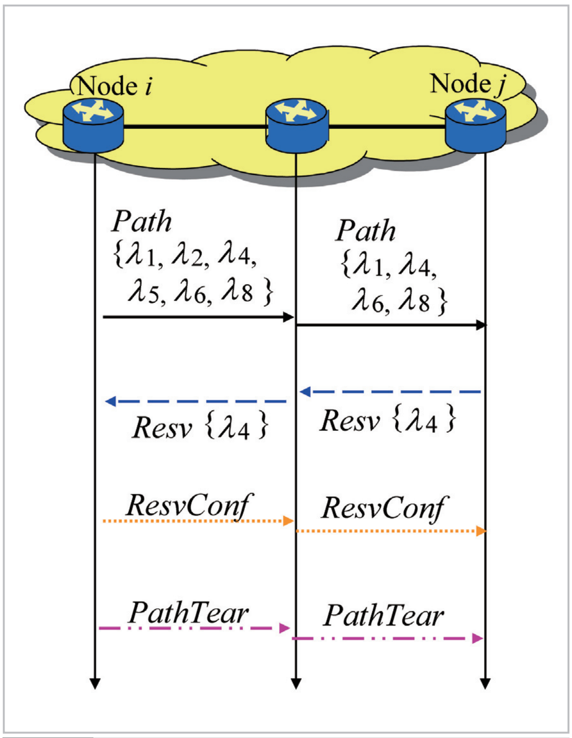

Fig. 1 Lightpath establishment using RSVP-TE signaling

3 Lightpath establishment in multi-domain WDM networks

Figure 2 shows a WDM network model consisting of multiple domains. When establishing lightpaths in a multi-domain WDM network using RSVP-TE, a given domain must provide its wavelength-usage information to other domains. However, each domain network is managed and operated in accordance with different policies; as a result we cannot assume that a given domain will share usage information about all wavelengths in the domain with other domains. Therefore, we have proposed a method of end-to-end lightpath establishment based on rank accounting, a method that permits the efficient establishment of lightpaths by sharing usage information on a limited number of wavelengths with other domains[6][7].

In end-to-end lightpath establishment based on rank accounting, the maximum number of wavelengths K which are shared with other domains is contracted among the domains. Conforming to this contract, each node provides to an adjacent node the wavelength-usage information for a maximum of K wavelengths. This contract allows each domain to disclose usage information for a limited number of wavelengths. Furthermore, each wavelength is allocated a wavelength rank, which indicates its order of priority in lightpath establishment, to provide for preferential selection among available wavelengths. The source node selects the contracted number of wavelengths according to wavelength rank, and makes efficient use of information for this limited number of wavelengths, thus facilitating the establishment of a lightpath passing through multiple domains.

Figure 2 shows the signaling procedure for end-to-end lightpath establishment based on rank accounting. In Fig. 2, the number of contracted wavelengths is set at $K = 3$. The source node executing the lightpath establishment selects wavelengths $\{\lambda_1, \lambda_2, \lambda_4\}$, a set comprising a number equal to or smaller than the contracted number, and uses the *Path* and *Resv* messages to establish the lightpath, as shown in Fig. 2.

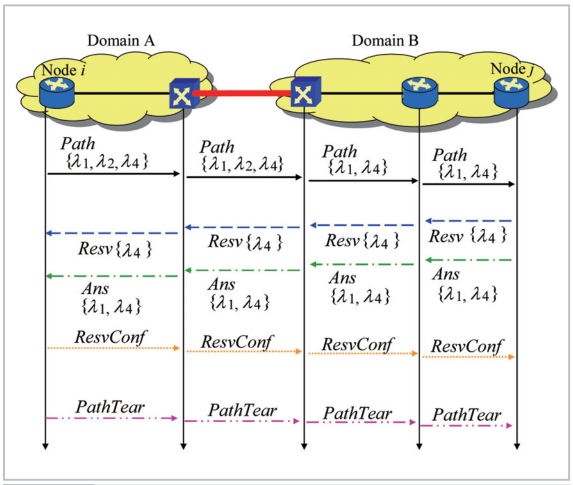

Fig.2 Lightpath establishment signaling in end-to-end lightpath establishment based on rank accounting

The destination node uses a control message called the *Ans* message to return the wavelength-usage information $\{\lambda_1, \lambda_4\}$ collected by the *Path* message to the source node. Upon receiving the *Ans* message, the source node updates the wavelength rank using formulas such as the exponential moving average (EMA) formula shown in Eq. (1).

$$r_{ij}^n(m) = \begin{cases} (1-\alpha)r_{ij}^n(m-1)+\alpha, \\ \quad \text{Wavelength } n \text{ is available for use,} \\ (1-\alpha)r_{ij}^n(m-1), \\ \quad \text{Wavelength } n \text{ is not available for use,} \\ r_{ij}^n(m-1), \\ \quad \text{Wavelength } n \text{ is not included} \\ \quad \text{in the } \textit{Path} \text{ message.} \end{cases} \tag{1}$$

Here, $r_{ij}^n(m)$ denotes the rank of wavelength n after establishment of the lightpath for the mth time from source node i to destination node j. The parameter α satisfies $0 \le \alpha \le 1$.

It can be seen from Eq. (1) that the rank of a wavelength available for the lightpath increases while one unavailable for use decreases. As such, preferential selection of wavelengths with higher ranks facilitates lightpath establishment across multiple domains.

Figure 3 presents the simulation results for performance evaluation of end-to-end lightpath establishment based on rank accounting in a WDM network consisting of two domains. Each domain consists of a 14-node NSFNET, and the number of wavelengths for each link is set at eight. The arrival of a request to this network for lightpath establishment passing through the two domains is assumed to take place in accordance with a Poisson process with a rate of 1.0. Furthermore, it is assumed that the request for lightpath establishment within each domain follows a Poisson arrival process with rate ξ. The holding time for each establishment of a lightpath follows an exponential distribution with a mean of 1 second. The processing time of the control message at each node is set at 1 millisecond.

Figure 3 shows the blocking probability in end-to-end lightpath establishment based on rank accounting for a number of contract wavelengths of K = 3. The figure also shows the blocking probability in lightpath establishment using the conventional RSVP-TE method. From Fig. 3, we can conclude that the blocking probability in end-to-end lightpath establishment based on rank accounting is consistently smaller than under the conventional method, regardless of the value of ξ. From this, we can further ascertain that a lightpath is efficiently established by using wavelength-usage information for a smaller number of wavelengths, specifically with the adoption of end-to-end lightpath establishment based on rank accounting.

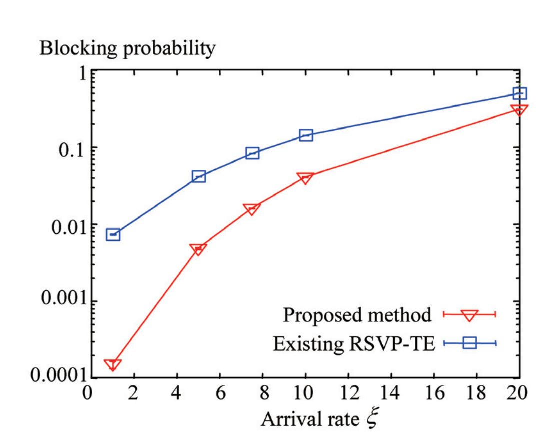

Fig.3 Simulation results

4 Construction of optical ring using lightpath for multipoint communication

Distributed computing requires the transfer of data between multiple hosts, demanding not only point to point communication but also multipoint communication. The authors have therefore proposed a method of constructing a unidirectional optical ring through dynamic lightpath establishment in order to realize multipoint data transmission on an optical network[8][9].

Figure 4 shows a multipoint distributed computing environment that employs a unidirectional optical ring. Since the optical ring does not demand the multicasting functions required for optical trees, the cost of construction can be reduced. The efficiency of use of wavelength resources is also expected to improve through the restriction of lightpath establishment to a single direction.

In the proposed method, the architecture of the optical ring is determined based on the following algorithm for constructing a unidirectional ring containing a number G of hosts $S = \{h_p, h_1, h_2, \ldots h_{G-1}\}$.

(1) Information on the host set $S = \{h_p, h_1, h_2, \ldots, h_{G-1}\}$ is initialized. The host set is divided into a host set L, for which a lightpath has already been set, and a host set U, for which a lightpath has not yet been set. Thus, the initialization process is equiva-

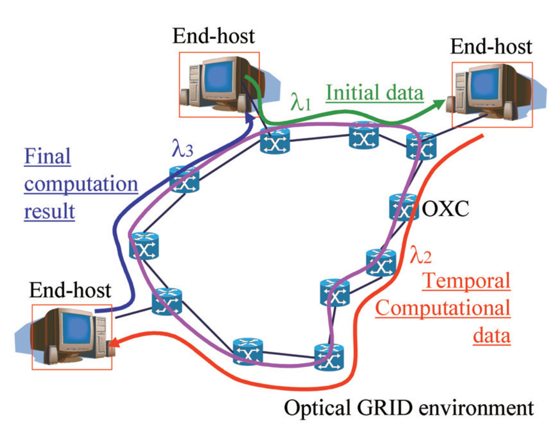

Fig.4 Multipoint distributed computing using optical ring

lent to setting $L = \{\}$ and $U = S = \{h_P, h_1, h_2, \ldots, h_{G-1}\}$.

(2) From set U, the parent host h_P is removed and added to set L. In other words, $L = \{h_P\}$ and $U = \{h_1, h_2, \ldots, h_{G-1}\}$. A host h with the minimum path cost (number of hops) from host h_P is selected from the hosts in set U and is appointed as the next host. A lightpath is then established from host h_P to host h.

(3) Host h is removed from set U and added to set L. For example, if $h = h_{G-1}$, then $L = \{h_P, h_{G-1}\}$ and $U = \{h_1, h_2, \ldots, h_{G-2}\}$. If U is not empty, the process of choosing the host with the minimum path cost from h and appointing it as the next node is performed, and a lightpath is established between host h to the next host selected. The process is repeated until $U = \{\}$.

(4) Finally, a lightpath is established from the last host added to L to the parent host h_P.

Note here that in the above algorithm, control signaling for lightpath establishment is required in steps (2) to (4). Figure 5 illustrates the signaling required in the establishment of the unidirectional optical ring. Figure 5 gives an example in which a ring network is created for three nodes a, b, and c. The lightpaths between nodes a-b, b-c, and c-a are established according to the RSVP-TE signaling described in Section **2**. Information on host sets L and U are transmitted using GRID_Group_Msg, and confirmation of the completion of lightpath establishment is performed using a GRID_Confirm_message.

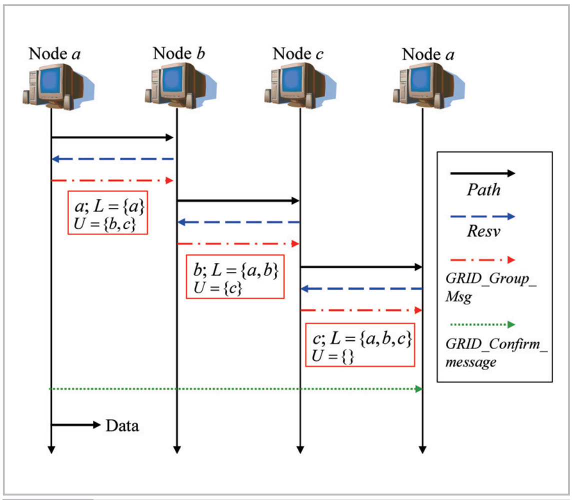

Fig.5 *Optical ring establishment signaling*

Figure 6 presents the results of an implementation experiment for optical ring establishment signaling[9]. The experimental network is composed of three nodes (a, b, c), and the data transmission links between the nodes are represented by dashed lines. The number of multiplexed wavelengths in each link is set at two, and the control plane is developed on a LAN. The results of this experiment show that a distributed computing environment using an optical ring may be dynamically constructed through adoption of the system implemented here.

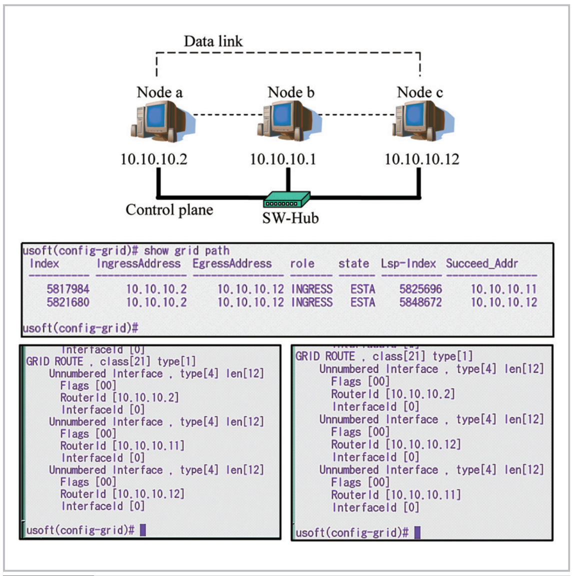

Fig.6 *Results of implementation experiment*

5 Conclusions

This paper introduced two advanced technologies for lightpath establishment when constructing a distributed computing environment using lightpath switching in a WDM network. We showed that a lightpath passing through multiple domains could be efficiently established under a method of end-to-end lightpath establishment based on rank accounting. We verified that with respect to optical ring construction technology using lightpaths, an optical ring for multipoint data transmission could be dynamically constructed using control signaling derived from RSVP-TE. The adoption of these two methods is expected to enable the dynamic construction of a large-scale distributed computing environment using lightpath switching.

Acknowledgements

We are grateful to Dr. Hiroaki Harai, chief researcher of the Ultrafast Photonic Network Group at NICT, for his discussions regarding the contents of this paper.

References

1 I. Chlamtac, A. Ganz, and G. Karmi, "Lightpath communications: An approach to high bandwidth optical WAN's", IEEE Transactions on Communications, Vol. 40, pp. 1171–1182, July 1992.

2 "OptIPuter", http://www.optiputer.net/.

3 "CA*Net 4", http://www.canarie.ca/.

4 "National Lambda Rail", http://www.nlr.net/.

5 H. Nakamoto, K. Baba, and M. Murata, "Proposal of a Shared Memory Access Method for Lambda Computing Environment", in Proc. Op-NeTec 2004, Oct. 2004.

6 T. Tachibana and H. Harai, "End-to-End Lightpath Establishment Based on Rank Accounting without Wavelength Conversion for Multi-Domain WDM Networks", in Proc. 31st European Conference and Exhibition on Optical Communication Conference (ECOC2005), No.We4.P.087, pp.675-676, Sept. 2005.

7 T. Tachibana and H. Harai, "Lightpath Establishment based on Rank Accounting in Multi-Domain WDM Networks", to appear in IEICE Transaction on Communications.

8 H. Harai and M. Murata, "Establishing Lightpaths of an Optical Ring for Distributed Computing Environment", in Proc. IEEE GridNets2005 (Second International Workshop on Networks for Grid Applications), pp. 488-495, 2005.

9 S. Xu and H. Harai, "Optical Ring Services in GMPLS Based Mesh Networks: An Implementation of Optical GRID", in Proc., Optical Fiber Communication Conference & Exposition and the National Fiber Optical Engineers Conference (OFC/NFOEC 2006), No. JThB74, Mar. 2006.

10 L. Berger, "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", IETF RFC 3471, Jan. 2003.

11 L. Berger, "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", IETF RFC 3473, Jan. 2003.

***TACHIBANA Takuji**, Dr. Eng.*

Nara Institute of Science and Technology (former: Expert Researcher,Ultrafast Photonic Network Group, Information and Network Systems Department)

Photonic Network Architecture, Lightpath Network

***Xu Sugang**, Ph.D.*

Limited Term Reseaercher, Network Architecture Group, New Generation Network Research Center (former: Ultrafast Photonic Network Group, Information and Network Systems Department)

Ultra-fast Network Control and Design