

Master's Thesis

Title

Cross-Layer Traffic Engineering in IP over WDM Networks

Supervisor

Professor Masayuki Murata

Author

Yuki Koizumi

February 15th, 2006

Graduate School of Information Science and Technology

Osaka University

Abstract

One approach for accommodating IP traffic on the wavelength division multiplexing (WDM) network is to construct a logical topology by establishing a set of lightpaths between nodes. The lightpath carries IP traffic and does not require any electronic packets processing at intermediate nodes, which reduces the load of packet processing at those nodes. However, if IP and WDM networks have their own routing functions, the lightpaths configured in the WDM network may not be fully utilized by IP unless the route of IP packets can be explicitly determined. Therefore, the integration of routing mechanisms is necessary to provide efficient resource utilization and flexible adaptation against traffic changes. In this thesis, we propose an integrated routing mechanism for IP over WDM networks. The key idea is to prepare a set of *virtual-links* based on the resource utilization of the WDM network, and then calculate the minimum cost route on the IP network. Our simulation results show that the throughput of our method is about 70 % higher than those of optimally designed logical topologies for a given traffic demand when traffic patterns change. We also develop a simulation method based on the fluid flow model, that is a flow-level simulation method, for simulating high-bandwidth and high-speed networks, such as IP over WDM networks. We verify the results of the flow-level simulation method by comparing with those of a conventional simulation method (packet-level simulation method). The flow-level simulation method consumes about 500 times less time than the packet-level simulation method.

Keywords

Wavelength Division Multiplexing, IP over WDM, Traffic Engineering, Routing, Integrated Routing Algorithm

Contents

1	Introduction	7
2	Traffic engineering in IP over WDM networks	13
2.1	Models of IP over WDM networks	13
2.1.1	Overlay model	13
2.1.2	Peer model	13
2.1.3	Augmented model	14
2.2	Approaches for traffic engineering	14
2.2.1	Routing in the overlay model	15
2.2.2	Routing in the peer model	16
3	Integrated routing for cross-layer traffic engineering	18
3.1	Network model	18
3.2	Concept of virtual-links	19
3.3	Integrated routing algorithm	20
3.3.1	Topology database	20
3.3.2	Route selection	21
3.4	Cost assignment to virtual-links	21
4	Flow-level simulation method for packets over WDM networks	23
4.1	Difficulties in packet-level simulation	23
4.2	Method for flow-level simulation	23
4.2.1	Fluid flow model	23
4.2.2	Application of the fluid flow model to network simulations	24
4.3	Verification of the flow-level simulation method	25
4.3.1	Simulation model	25
4.3.2	Performance evaluations	26
5	Effects of cross-layer traffic engineering in IP over WDM networks	28
5.1	Simulation model	28
5.2	Static logical topology design methods	28

5.3	Performance evaluations	31
5.3.1	Average end-to-end delay	31
5.3.2	Load of nodes	31
5.3.3	Network throughput	33
6	Conclusion	39
	Acknowledgements	40
	References	42

List of Figures

1	IP over point-to-point WDM architecture.	8
2	IP over reconfigurable WDM architecture.	10
3	An example of a logical topology viewed from IP's point of view.	11
4	Overlay model.	14
5	Peer model.	15
6	An example that the IP routing does not select a link provided by the WDM network.	16
7	Node architecture.	19
8	An example of a network with virtual-links.	20
9	Congestion resulting from increasing a link.	22
10	Calculation of the number of packets at a node.	25
11	The average end-to-end delay of flow-level simulation and packet-level simulation.	27
12	The number of packets in a node flow-level simulation and packet-level simulation.	27
13	NSFNET topology.	29
14	European Optical Network (EON) topology.	29
15	Average end-to-end delay without traffic change: NSFNET topology, the case of realistic traffic demand matrix.	32
16	Average end-to-end delay with traffic change: NSFNET topology, the case of realistic traffic demand matrix.	33
17	Average end-to-end delay without traffic change: NSFNET topology, the case of random traffic demand matrix.	34
18	Average end-to-end delay with traffic change: NSFNET topology, the case of random traffic demand matrix.	35
19	Average end-to-end delay without traffic change: EON topology, the case of random traffic demand matrix.	35
20	Average end-to-end delay with traffic change: EON topology, the case of random traffic demand matrix.	36
21	Load of nodes without traffic change: NSFNET topology, total traffic volume is 42 Gbps for all method.	36

22 Load of nodes with traffic change: NSFNET topology, total traffic volume is 42
Gbps for all method. 37

23 Amount of traffic volume of the integrated routing method: ingress, egress, and
transit traffic. 37

24 Network throughput. 38

List of Tables

1	Traffic demand matrix for NSFNET topology.	30
---	----------------------------------------------------	----

1 Introduction

Growth of Internet traffic still continues resulting from the increase of popularity of the Internet and from the advent of new network applications, such as P2P file sharing, voice communications, and grid computing. Internet backbone networks require more capacity for accommodating the current and future traffic demand in a cost-effective way. To cope with traffic demand in the current Internet, optical transmission technologies, such as Synchronous Optical Network (SONET) or Synchronous Digital Hierarchy (SDH) are deployed in many Internet service providers. In these optical technologies, however, electronic processing is still required to transmit packets over optical channels, which eventually will become a bottleneck as the traffic demand increases.

Wavelength Division Multiplexing (WDM) technology, which carries multiple wavelength channels on a single fiber, is expected to accommodate the traffic volume for the current and future Internet. By using WDM technology, a much lower cost solution to meet traffic demands can be achieved. Since the majority of Internet traffic is IP, it is an important task to develop network architectures which transport IP traffic on WDM networks. There are several possible approaches to transport IP traffic on the WDM network.

One is to use the Asynchronous Transfer Mode (ATM) technology on SONET/SDH, where SONET is built on the WDM network, which is a so called IP over ATM over SONET/SDH over WDM network. WDM is employed as a physical layer, which provides parallel transmission technology. Another approach is to use SONET/SDH, where SONET is built on the WDM network, which is a so called IP over SONET/SDH over WDM network. In these network architectures, WDM technology is employed as a physical layer which provides concurrent transmission via wavelengths. These network architectures require, however, processing for several protocol stacks, and thus increase protocol overheads and management complexity. Another and more promising solution is an IP over WDM network architecture, where IP packets are directly carried over the WDM network¹.

There are several alternatives for IP over WDM networks: IP over point-to-point WDM, IP over reconfigurable WDM, and IP over switched WDM [1]. The IP over point-to-point WDM architecture is shown in Figure 1. Under the IP over point-to-point WDM architecture, the WDM

¹More exactly, we need a data link layer protocol such as Point to Point Protocol (PPP) or High Level Data Link Control (HDLC) between IP and WDM protocol stack.

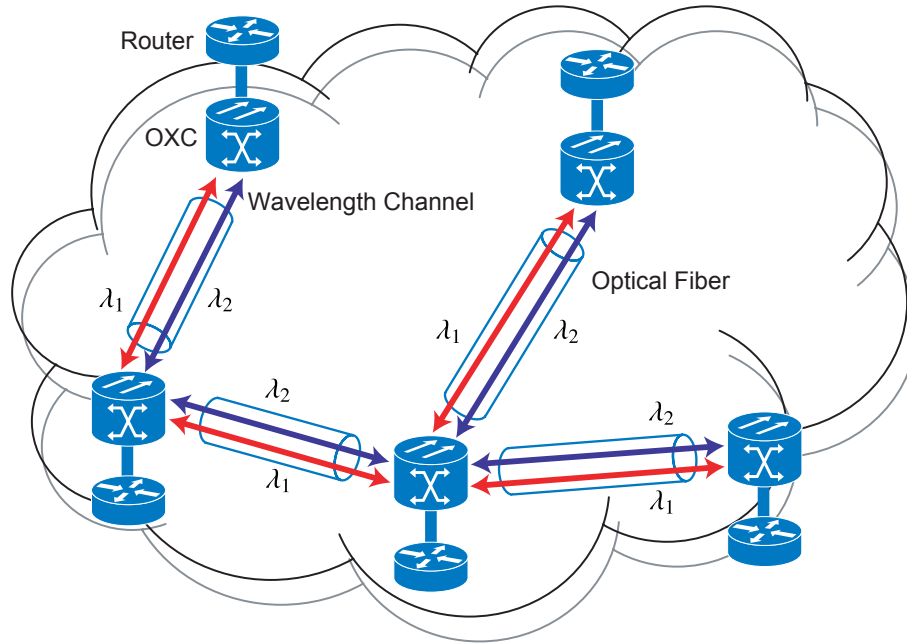


Figure 1: IP over point-to-point WDM architecture.

technology is used on optical fiber links. That is, each wavelength on the fiber is treated as a physical link between conventional IP routers, and therefore multiple links of wavelengths are provided between IP routers. In this case, the bandwidth of optical fiber links is actually increased by the WDM technology; however, it is insufficient to resolve the network bottleneck against the explosion of traffic demands since the bottleneck of the network is shifted to electronic routers.

The IP over reconfigurable WDM architecture, which is also called the Optical Circuit Switching (OCS) network architecture, is the promising candidate in the Internet backbone networks. In this architecture, wavelength channels (also known as lightpaths) are configured in the WDM layer (Figure 2), and IP packets are carried by lightpaths. Each intermediate node of lightpaths has optical cross-connects (OXCs) that bind an input wavelength channel to a specified output wavelength channel. IP packets on lightpaths are forwarded with no electronic packet processing at the IP routers in the intermediate nodes. In this way, lightpaths greatly reduce the load of IP routers. Hence, the IP over reconfigurable WDM network can accommodate more traffic volume than the IP over point-to-point WDM network since the IP over reconfigurable WDM network relieves the bottleneck, that is the processing capacity of IP routers. In this thesis, we investigate the IP over reconfigurable WDM architecture, and hereafter we refer to IP over reconfigurable WDM

networks by IP over WDM networks.

A set of lightpaths forms a logical topology for the IP routers. We show a simple example in Figure 2. In this figure, several lightpaths are placed by using a wavelength on links. The lightpath l_2 is configured using the wavelength λ_2 from the node N_1 to N_3 through node N_2 . Since l_2 is switched at a wavelength granularity, IP packets on l_2 are not aware of the IP router of node N_2 . Therefore, the IP layer recognizes the lightpath l_2 as a link which is directly connected from node N_1 to node N_3 . Hence, when the lightpaths are placed as shown in Figure 2, the logical topology constructed for the IP layer as shown in Figure 3 is different from physical topology.

In order to fully utilize the bandwidth provided by WDM networks, it is important to develop traffic engineering approaches. Without traffic engineering, some links can be heavily congested, while other links are less utilized. Traffic engineering objectives can be achieved by developing an efficient routing method. That is, to select appropriate routes for balancing the load of nodes or links. There are two types of logical topology design methods aiming at achieving traffic engineering objectives. One approach is the static logical topology design method, which statically configures a logical topology based on the traffic statistics (e.g., traffic demand matrices) measured for a certain period. Given the traffic demand, optimal logical topology can be obtained by Mixed Integer Linear Problem (MILP) solutions, and some heuristic design methods configure near optimal topologies [2,3]. However, the static design method cannot flexibly adapt the logical topology for changes of the traffic pattern, because the topology was configured for the traffic statistics before changing the traffic pattern. Another approach is the dynamic logical topology design method, which dynamically constructs logical topologies based on the current network status.

In IP over WDM networks, there are three different inter-networking models: peer, augmented, and overlay model [4]. One of the most important differences among these models is what type of information is shared between IP layer and optical layer. In the peer model, the topology and other network information (e.g., routing information and link state information) are shared between both layers, and then the unified routing mechanism controls the whole network (i.e., IP and WDM). On the other hand, in the overlay model, no information is shared between the two layers, and each layer is controlled independently by its own routing mechanism. The augmented model is a hybrid of the peer model and the overlay model. That is, the augmented model shares some part of information, such as reachability information, between the two layers according to a certain agreement, and the two layers are managed independently as like the overlay model.

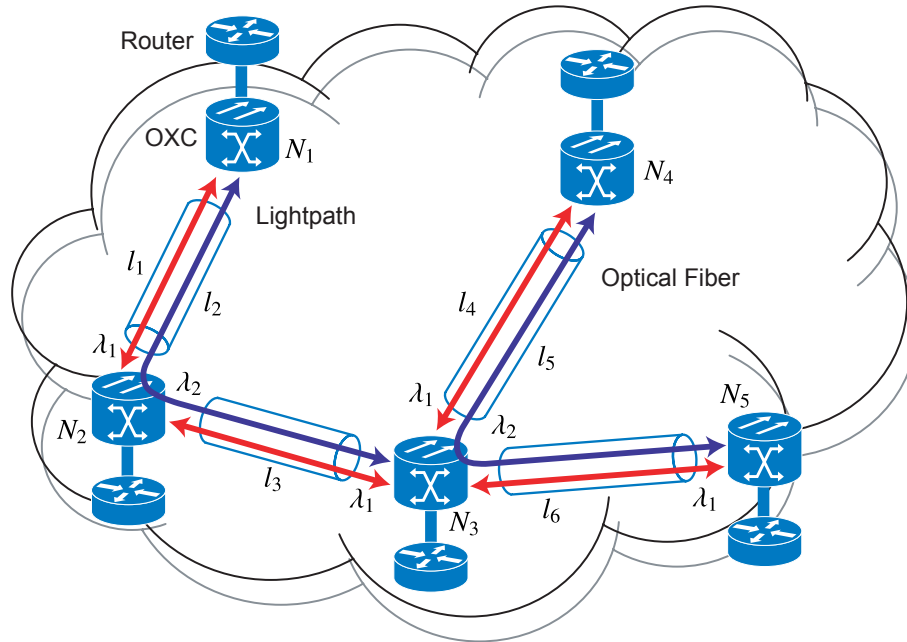


Figure 2: IP over reconfigurable WDM architecture.

For the overlay model, a logical topology is first constructed by setting lightpaths between nodes in the WDM network, and then existing routing protocols for IP packets are performed on that logical topology (see reference [5] and references there in). In this case, the IP protocols need not change. However, since both IP and WDM networks have their own routing functions, the lightpaths configured in the WDM network may not be fully utilized by IP. That is, the routes of IP packets are decided by IP routing mechanism, and IP does not necessarily use lightpaths provided by the WDM network. Therefore, the integration of routing mechanisms is necessary such that efficient resource utilization and flexible adaptation against traffic changes are provided.

The integrated routing method for the peer model is proposed in references [6–9]. In these papers, Multi Protocol Label Switching (MPLS) technology is used to specify the route of IP packets (MPLS-based IP over WDM networks). That is, the route of IP packets can be explicitly determined by MPLS's Label Switched Path (LSP). We will call such a network as IP/MPLS over WDM networks. These works develop integrated routing algorithms for calculating the route of both LSPs and lightpaths. The IP/MPLS over WDM network has a unified routing entity that collects all the topology information and link state information of both IP/MPLS and WDM networks. When traffic flows arrive, these routing methods calculate routes of the traffic flow where

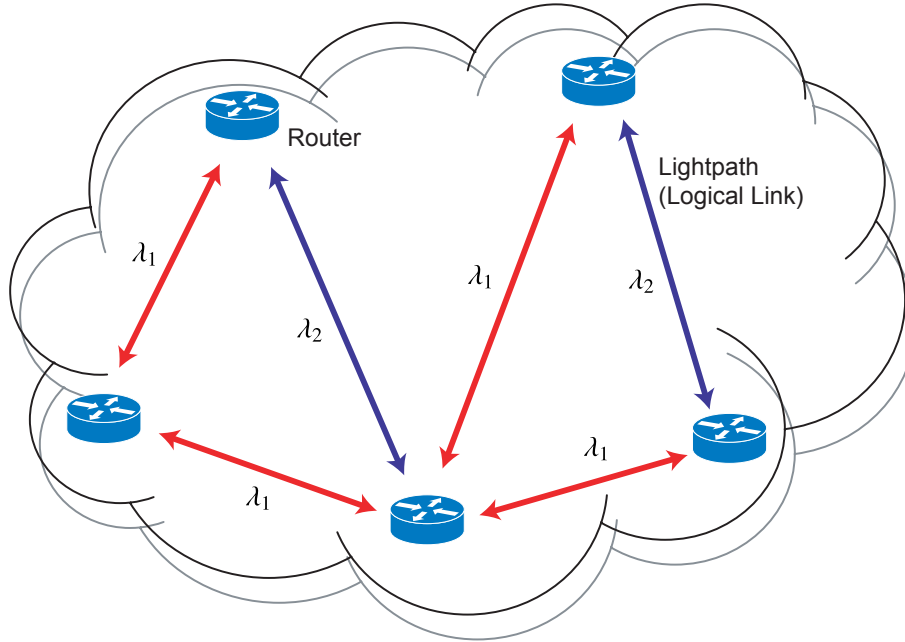


Figure 3: An example of a logical topology viewed from IP's point of view.

the required bandwidth is explicitly specified according to their performance objectives. When there is no appropriate route, LSP requests are blocked.

In the peer model, efficient route control can be performed since all network information is available. However, collecting the information requires the advertisement of the information of both IP layer and WDM layer. This leads to excessive update overhead, and therefore lacks scalability. On the other hand, in the overlay model, only the information of each layer needs to be advertised, and thus scales well comparing to the peer model. However, since information of network status is limited, it is difficult to perform efficient routing as we will show in Section 5 by comparing with the algorithm in [2].

In this thesis, we integrate IP routing and wavelength routing for IP over WDM network, and propose the routing method where IP packets are surely forwarded on the lightpath configured by the wavelength routing. For this purpose, we introduce a concept of *virtual-links* that are configured between IP routers. The virtual-link is a logical link which is not configured as a lightpath, but can be activated as a lightpath by reserving the wavelength resource according to the request. In the IP network, our routing method first calculates routes on the topology including these virtual-links. If a virtual-link is selected as a route of IP packets, a lightpath corresponding

to the virtual link is established. By this way, we can calculate the route of IP packets and the route of lightpaths simultaneously. In this thesis, we give the cost function to virtual-links aiming at reducing the load of nodes, and the results of computer simulations show the effectiveness of our proposed method in terms of the average end-to-end delay and throughput.

Unlike the previous works that investigate blocking probability of LSP requests, we intend to reveal the packet-level behavior in the IP directly over WDM networks. However, by the WDM technology, the traffic that can be accommodated in the network increases, which makes computer simulations difficult since there is a huge number of IP packets on WDM networks. In this thesis, we develop a simulation technique based on the fluid flow model that greatly reduces the number of packets that has to be processed during computer simulations. We compare the results of flow-level simulation based on the fluid flow model with the results of the packet-level simulation, and reveal that end-to-end packet delay shows almost the same.

The rest of this thesis is organized as follows. In Section 2, we describe how IP routing methods perform in IP over WDM networks, and introduce other traffic engineering approaches. In Section 3, we describe our integrated routing algorithm. Before evaluating the performance of our method, we show the flow-level simulation method that enables a computer simulation to finish within a reasonable time. We next evaluate the performance of our routing method in Section 5. Finally, we conclude this thesis in Section 6.

2 Traffic engineering in IP over WDM networks

As mentioned above, it is important to develop traffic engineering approaches in order to fully utilize bandwidth provided by WDM networks. Traffic engineering objectives can be achieved by developing efficient routing methods. In this section, we first describe three models of IP over WDM networks with focusing on information sharing between IP and WDM networks, and then routing methods for IP over WDM networks.

2.1 Models of IP over WDM networks

The IP over WDM network has control plane and data plane. The control plane manages the route information and link state information, and has a responsible for setup/tear down lightpaths. There are three models of information sharing between IP and WDM networks dependent on what kind of information and how much information is shared. We describe each model in the following sections. In the data plane, a WDM network is used as the transport network like other layer-2 networks, such as Ethernet and ATM.

2.1.1 Overlay model

Figure 4 illustrates an example network of the overlay model. In the overlay model, no network information is shared between an IP layer and a WDM layer. A WDM network and an IP network have their own control planes. The routing protocols, topology information, and signaling protocols in the IP network are independent of those of the WDM network. Under this network model, the IP layer and the WDM layer work as client-server model. The IP network sends lightpath setup/tear-down requests via User Network Interfaces (UNI) to the server, that is, the WDM layer, and the WDM layer configures or tears down lightpaths in response to these requests.

2.1.2 Peer model

In the peer model, the topology and other network information (e.g., routing information and link state information) are shared between both layers, and then the unified routing mechanism controls the whole network. In Figure 5, we illustrate an example of the peer model IP over WDM network.

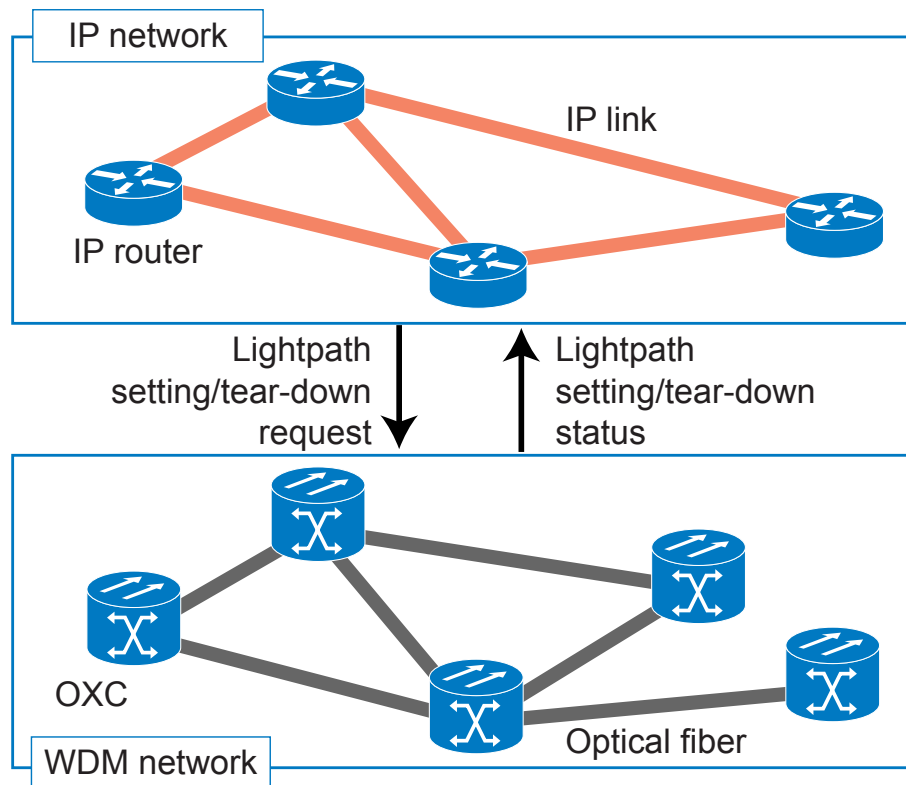


Figure 4: Overlay model.

2.1.3 Augmented model

The augmented model is the hybrid of the peer model and the overlay model. That is, the augmented model shares some part of information, such as reachability information, between two layers according to a certain agreement, and two layers are managed dependently as like the overlay model.

2.2 Approaches for traffic engineering

There are two kinds of performance objectives for traffic engineering. One is to maximize the traffic volume that can accommodate (traffic-oriented objective), and another is to minimize resources that are used to accommodate given traffic demand (resource-oriented objective). There are many routing approaches which focus on the traffic-oriented objective, and we also focus on that in this thesis. In the following section, we describe routing approaches for the overlay model and the peer model of IP over WDM networks.

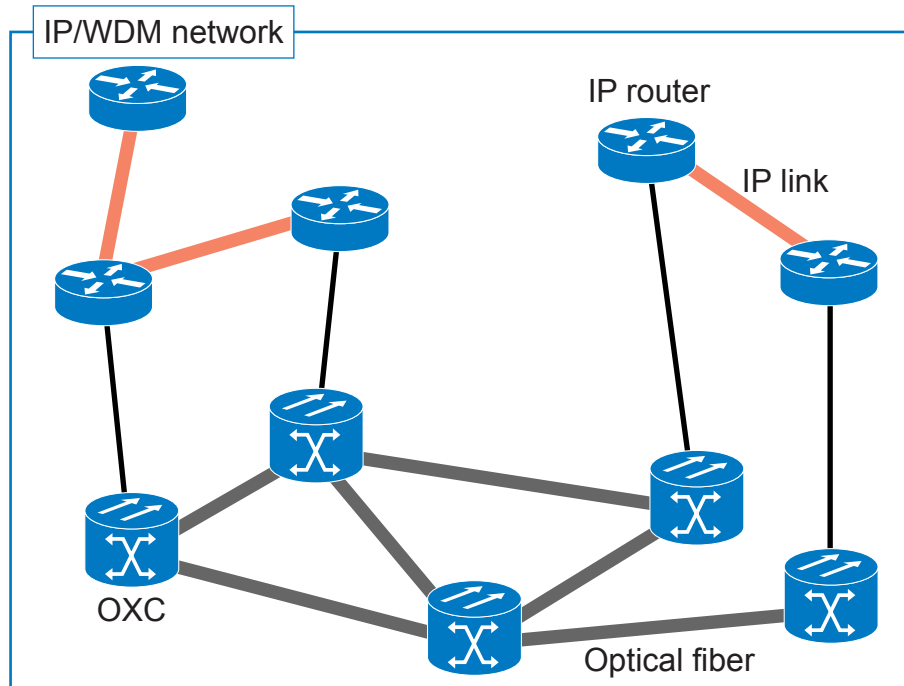


Figure 5: Peer model.

2.2.1 Routing in the overlay model

In the overlay model, a logical topology is first constructed by putting lightpaths between nodes in the WDM network, and then conventional IP routing protocols such as Open Shortest Path First (OSPF) and Intermediate System - Intermediate System (IS-IS) works on that logical topology. In this case, these routing protocols are performed without any extension for WDM networks. However, the IP routing mechanism and the WDM routing mechanism have been designed independently, therefore, IP does not necessarily select links provided by WDM networks. One typical example for such a problem is minimum delay routing in IP networks. In Figure 6, we illustrate a simple example of this problem. In this example, we assume that each optical fiber has two wavelengths, and the propagation delay of each fiber is 1 unit time. The delay of IP routers including processing delay and queuing delay is also 1 unit time, and that of OXC is 0. In this figure, six lightpaths are configured in the WDM network, and therefore a logical topology has six logical links. One long hop lightpath l_5 is configured between node N_2 and N_4 by using wavelength λ_2 because λ_2 on the optical fiber between node N_2 and N_3 has been already used for l_2 . There are

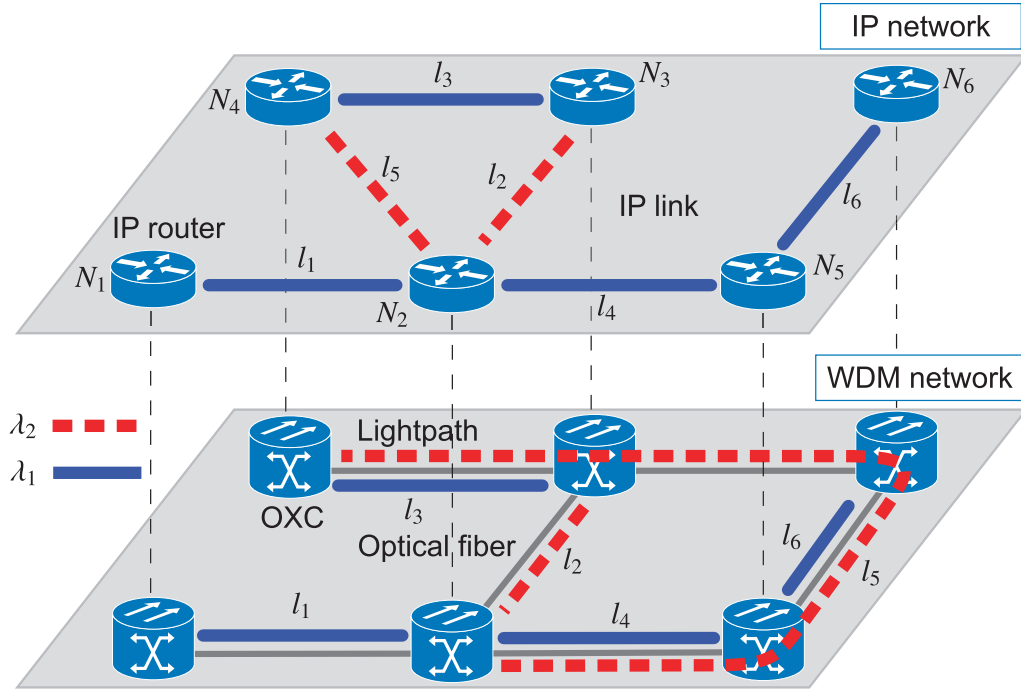


Figure 6: An example that the IP routing does not select a link provided by the WDM network.

two possible routes from node N_2 to N_4 , R_1 and R_2 . R_1 is one hop route $N_2 \rightarrow N_4$ using l_5 and R_2 is two hop route $N_2 \rightarrow N_3 \rightarrow N_4$ using l_2 and l_3 . The end-to-end delay of R_1 is 4 while that of R_2 is 3. Therefore, if IP use end-to-end delay as its routing metric, R_1 is never used as the route from N_2 to N_4 . This causes inefficient wavelength resource utilization. Therefore, we need an integrated routing mechanism aiming at efficient resource utilization. Some integrated routing method are proposed. In these works, they mainly assumed the peer model, that is, they used the information of both IP networks and WDM networks, such as bandwidth availability information of the IP network and wavelength availability information of WDM networks. In the following section, we present related works on the integrated routing.

2.2.2 Routing in the peer model

The integrated routing method for the peer model is proposed in references [6, 7]. In these papers, IP/MPLS over WDM network where the route of IP packets can be explicitly determined by MPLS's LSP is investigated. Reference [6], MIRA (Minimum Interference Routing Algorithm) is proposed as an integrated routing algorithm for calculating route of LSPs and the lightpaths.

The IP/MPLS network has a unified routing entity that collects all of the topology information and link state information of both IP/MPLS and WDM networks. For incrementally arriving traffic flows where required bandwidth is explicitly specified, MIRA calculates routes of the traffic flow. MIRA pays attention to residual bandwidth that will be used by future LSP requests, and therefore tries to maximize the residual bandwidth. When there is no appropriate route, LSP requests are blocked. Reference [10], the integrated routing method which takes inaccurate link state information is developed. This method use inaccuracy of link state and hop count as link cost metrics of LSPs. Reference [7] proposed an integrated routing method for the peer model of IP over WDM networks based on Generalized MPLS (GMPLS) framework. In this work, they assigned cost values to links aiming at reusing existing lightpaths for new LSP setting requests as much as possible. These works have assumed that IP traffic is mapped onto a bandwidth guaranteed LSP requests. Their main objective has been to improve blocking probability of LSP setting requests. Few works focus on the performance metrics of IP traffic (e.g., end-to-end delay or throughput). It is important to develop routing mechanisms aiming at improving the performance of IP traffic and to reveal the behavior of IP traffic. These works have assumed the peer model network. As mentioned above, however, the peer model lacks the scalability. Furthermore, there are few works on the integrated routing in the overlay model networks. In this thesis, therefore, we propose a new integrated routing method for the overlay model of IP over WDM networks aiming at maximizing traffic volume that networks can accommodate. In the next section, we describe our routing method.

3 Integrated routing for cross-layer traffic engineering

In this section, we introduce our network model and a concept of *virtual-links*, which is the key idea of our integrated routing method. Then, we elaborate on our routing algorithm and discuss our choice of cost metrics.

3.1 Network model

We consider the network that consists of optical fiber links having W wavelengths, and nodes that consist of both IP routers with WDM interfaces and OXCs. In Figure 7, we illustrate the node architecture used in this thesis. The upper part of this figure shows an IP router and the lower part shows an OXC. The IP router is connected to the OXC via a WDM interface which converts optical (electronic) signal to electronic (optical) signal. The control plane in the WDM layer manages the topology database with wavelength availability information. The IP router has a route manager which has three main blocks: the routing controller, the topology database, and OSPF-TE. In the route manager, the routing controller calculates routes based on our integrated routing algorithm on the logical topology specified in the topology database. This topology database is composed of only the information of IP layer, such as IP link, IP router status, and IP network connectivity since we assume the overlay internetworking architecture. OSPF-TE advertises and collects IP link state information. Note that, since we assume the overlay model, our method does not require any extension of existing IP routing protocols. The route manager in the IP layer sends setup or tear-down lightpaths requests to the control plane in the WDM layer according to the result of route calculation. When the control plane in the WDM layer receives these requests, it sets up new lightpaths and tears down existing lightpaths, and return results of those requests to IP route manager.

In our network architecture, one lightpath is statically set up between all adjacent nodes using one wavelength resource to ensure the end-to-end reachability. We refer to this lightpath as *persistent-lightpath*. The other wavelength resources are used for non-persistent lightpaths that are dynamically set up according to traffic changes.

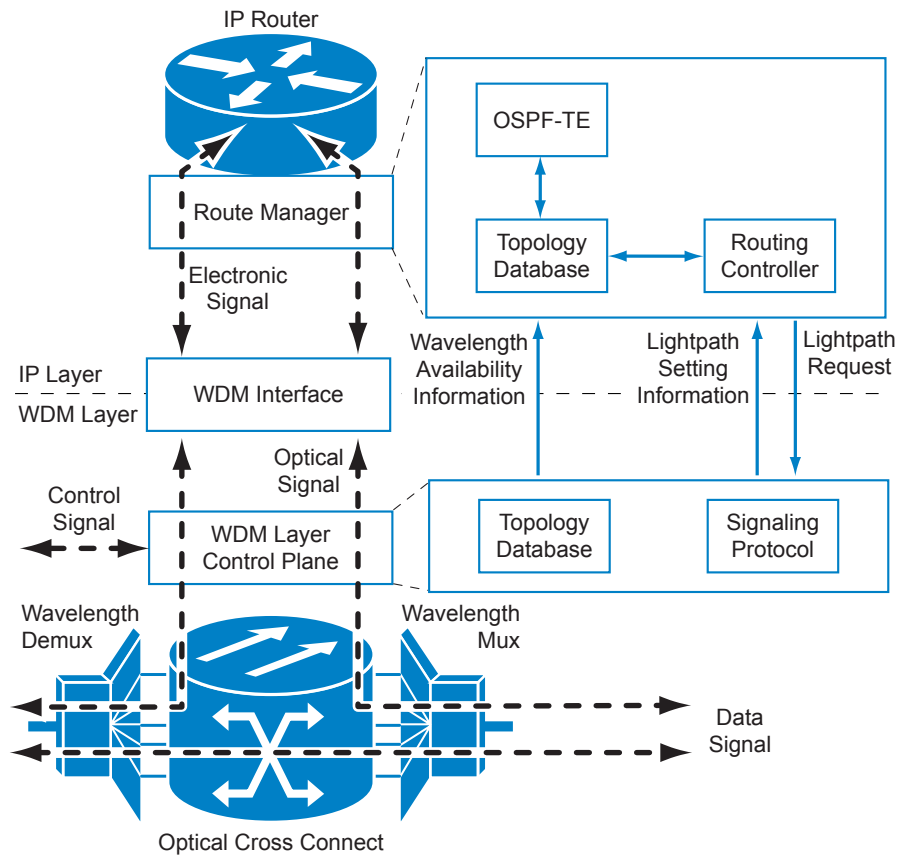


Figure 7: Node architecture.

3.2 Concept of virtual-links

To integrate IP routing and wavelength routing, we propose a concept of *virtual-links*. A virtual-link is a virtual link which has not yet been configured as a lightpath, but can be activated by reserving wavelength resources if necessary. Virtual-links are placed in a logical topology of the IP network when IP route manager calculates routes. An example of a network with virtual-links is illustrated in Figure 8. In this figure, persistent-lightpaths are configured between all adjacent nodes, and one non-persistent lightpath is configured from router R_1 to router R_6 . Three virtual-links (that includes from router R_1 to routers R_2 , R_4 , and R_5) are illustrated as dashed arrows in the figure. Each link and virtual link has a cost value that is used by the IP routing mechanism. IP routing method calculates routes from this logical topology including virtual-links and existing lightpaths. Note that in our proposal method, the cost values of virtual links give much impact on

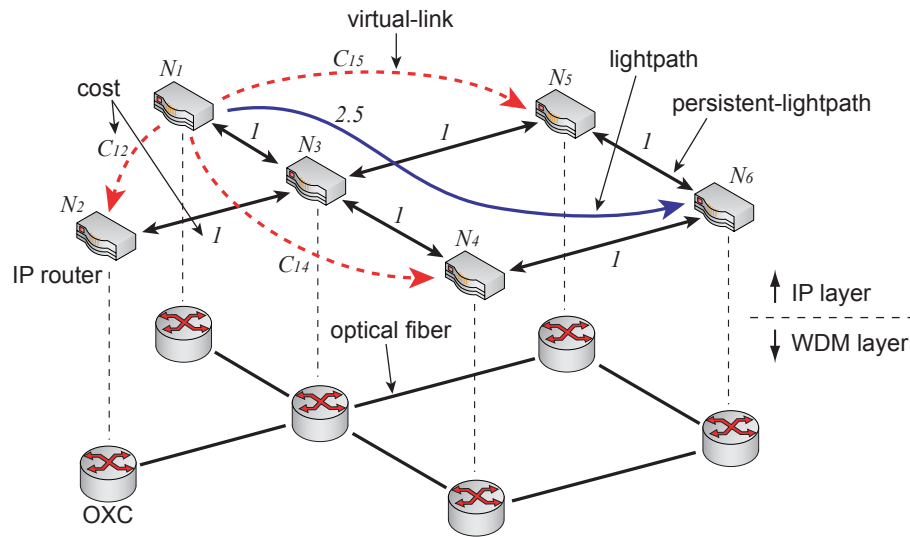


Figure 8: An example of a network with virtual-links.

the network performance. We discuss about the cost functions for the virtual-links in Section 3.2.

3.3 Integrated routing algorithm

3.3.1 Topology database

OSPF-TE in the IP router advertises the link state information, and each IP router collects all the link state information. Then, the IP router configures following logical topology, which includes the virtual-links and its cost.

Step 1: Set virtual-links from the source node to all destination nodes if there is no lightpath configured.

Step 2: Assign link cost values to each virtual-link.

Step 3: Update link cost values of existing non-persistent lightpaths.

Step 4: Set link cost values of persistent-lightpaths to 1.

Without loss of generality, we set the cost value of a persistent-lightpath to 1.0, and change the cost values of virtual-links.

3.3.2 Route selection

Based on the topology database obtained from the above procedures, the routing algorithm calculates routes of IP packets, and if virtual-link is selected as the part of IP routes, corresponding lightpath is dynamically configured on WDM networks.

Step 1: Calculate the minimum cost routes on the logical topology including virtual-links and existing lightpaths via the IP routing algorithm.

Step 2: If the resulting routes contain one or more virtual-links, send lightpath set up requests to the WDM network.

Step 3: Send tear down lightpath messages if there are existing lightpaths which are not used as routes of IP anymore.

The routing in IP layer and WDM layer is integrated by searching for the minimum cost routes from a logical topology where virtual-links are included. In this case, IP traffic is surely forwarded on the virtual-links (i.e., lightpaths) since lightpaths are selected by IP routing.

3.4 Cost assignment to virtual-links

In our approach, a fundamental problem is to select the cost function of virtual-links. The main objective of the cost function of virtual-links is to minimize the end-to-end delay. In order to minimize this, reducing the load of IP routers is important. However, setting lightpaths (virtual-links) may cause an unnecessary increase to the load of nodes. This is because more traffic may come to destination nodes of virtual links. We show a simple example in Figure 9. In Figure 9(a), two flows, f_1 and f_2 , are forwarded through route $N_1 \rightarrow N_3 \rightarrow N_5$ and $N_2 \rightarrow N_4 \rightarrow N_6 \rightarrow N_5$, respectively. Figure 9(b) illustrates the network after one lightpath is configured from node N_2 to N_3 . Since the lightpath is newly set, the route of flow f_2 is changed to $N_2 \rightarrow N_3 \rightarrow N_5$. In this case, traffic of f_2 that was passed through node N_4 and node N_6 is then passed through the node N_3 . Consequently, the load of N_3 increases. We, therefore, use the load of the destination node of the virtual-link as a parameter of our cost function of virtual-links to prevent from concentrating traffic on the destination node of the lightpath. Note that, IP router status can be obtained by using Simple Network Management Protocol (SNMP).

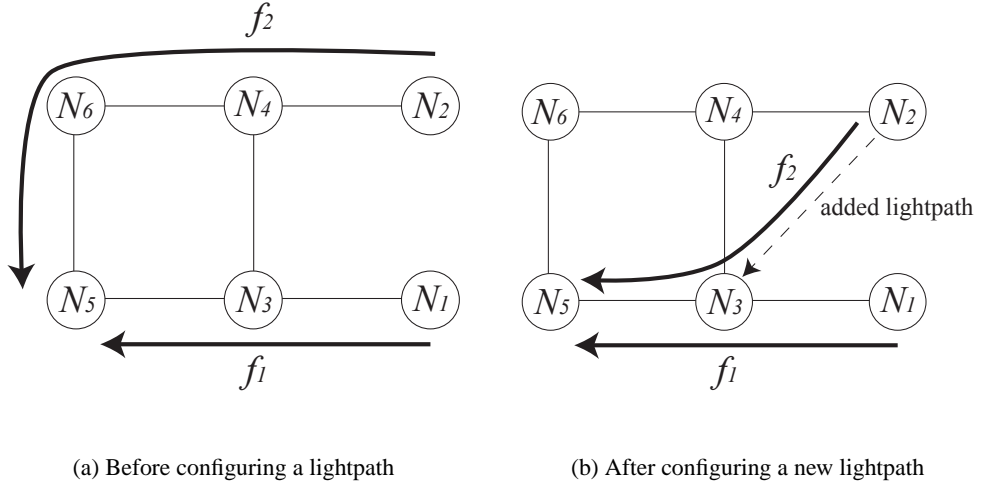


Figure 9: Congestion resulting from increasing a link.

The cost function of virtual-link from nodes i to j , C_{ij} , is given by,

$$C_{ij} = v_j^2 + \beta, \quad (1)$$

where v_j is the load of node j , and β is offset. As noted above, our primal purpose is to reduce the end-to-end delay, and therefore balances the load of nodes by using the cost function. We therefore use the convex cost function against the load of node v_j , which is expected to prevent the increase of the delay of node j . We set β to 0.5 to prevent IP routing method selecting unnecessarily long hop route. More exactly, to set the offset value to 0.5 makes IP routing mechanism always use a persistent-lightpath if there exists, since the cost value for a persistent-lightpath is set to 1.

The number of lightpaths N_{ij} between a node pair ij is decided as follows:

$$N_{ij} = \lceil b_{ij} \rceil,$$

where b_{ij} is the traffic volume that goes to node j through node i . Here, we assume traffic volume that goes to node j can be measured at node i . This equation means that at least N_{ij} lightpaths are configured in order to satisfy current traffic volume.

4 Flow-level simulation method for packets over WDM networks

In this section, we show a simulation method based on the fluid flow model, which is called flow-level simulation method. Note that we call a conventional discrete event simulation that process every packet as packet-level simulation. The flow-level simulation method enables computer simulations finish within a reasonable time. First, we describe why do flow-level simulation methods need, and then show the flow-level simulation method based on the fluid flow approximation. Finally, we compare the results obtained by flow-level simulation with those by packet-level simulation in order to verify flow-level simulation method.

4.1 Difficulties in packet-level simulation

A computer simulation is a commonly used method to understand and evaluate the performance of communication networks. Discrete event network simulators (such as OPNET [11] or NS-2 [12]) use events in order to simulate the movement of packets in the network. Although this provides accurate insight of the network state or performance, this also requires computers to process a lot of events to simulate networks. With the growth of the Internet , the scale of networks that have to simulate becomes larger. Moreover, new network technologies, such as WDM technology, provide a lot of high bandwidth links, and thus the traffic volume that networks can accommodate has been increasing significantly. Therefore, the huge amount of time and computer resources are consumed in order to simulate these networks.

In this thesis, we evaluate our integrated routing mechanism by the fluid flow approximation that enables much smaller number of packets to be processed in the computer simulations.

4.2 Method for flow-level simulation

4.2.1 Fluid flow model

First, we describe the fluid flow approximation model. In the simulation methods based on the fluid flow model, data traffic is modeled as not sequence of packets but a continuous fluid. Network simulators based on the fluid flow approximation abstract the model in order to focus on only changes in arrival rates of traffic flows. This can lead to significant decrease of computation time, although the information about the individual packets is lost. There are many works about

simulation methods based on the fluid flow model [13, 14]. These researches mainly develop methods for TCP-based networks, that is, they focus on the transient state of TCP flows. In our approach, we pay much attention to state of IP traffic and whole network. In the following section, we show our simulation method.

4.2.2 Application of the fluid flow model to network simulations

We define following terminologies to explain the flow-level simulation.

t_n : n -th time when arrival or departure of flows occur.

$L(t_n)$: Number of packets in the node at time t_n .

$a(t_n)$: Packet arrival rate to the node at time t_n .

a_i : Packet arrival rate in each flow i .

μ : Service rate of the node.

$W(t_n)$: Delay in the node at time t_n .

Here, we assume that the arrival process of packets in each flow follows Poisson with the average rate a_i , and the average rate is not changed.

Figure 10 illustrates the outline of the computational method of the packet arrival rate at the node. The figure illustrates three flows (flow₁, flow₂, flow₃) that arrive at the node. The flow₁, flow₂, and flow₃ arrive (depart) at time t_1 , t_2 , and t_3 (t_6 , t_4 , and t_5), respectively. Each flow has a parameter, a_i which is an arrival rate of packets in the flow _{i} . In other word, the interval of packet arrival for flow _{i} is $1/a_i$. The packet arrival rate for this node (denote $a(t_n)$) is the sum of the packet arrival rate a_i of each flow, and thus $a(t_n)$ changes according to the arrival/departure process of flows, as shown in the upper part of Figure 10.

The increase of the number of packets in the node between time t_{n-1} and t_n can be represented by $\{a(t_{n-1}) - \mu\} \times (t_n - t_{n-1})$. Here, $a(t_{n-1}) \times (t_n - t_{n-1})$ means the number of packets which arrived at the node between time t_{n-1} and t_n , and $\mu \times (t_n - t_{n-1})$ means the number of packets which departed from the node. Note that, if $\{a(t_{n-1}) - \mu\} \times (t_n - t_{n-1}) < 0$, the number of packets in the node decreases. Using the number of packets at time t_{n-1} (i.e., $L(t_{n-1})$), the total number of packets in the node at the time t_n can be expressed by,

$$L(t_n) = \{a(t_{n-1}) - \mu\} \times (t_n - t_{n-1}) + L(t_{n-1}).$$

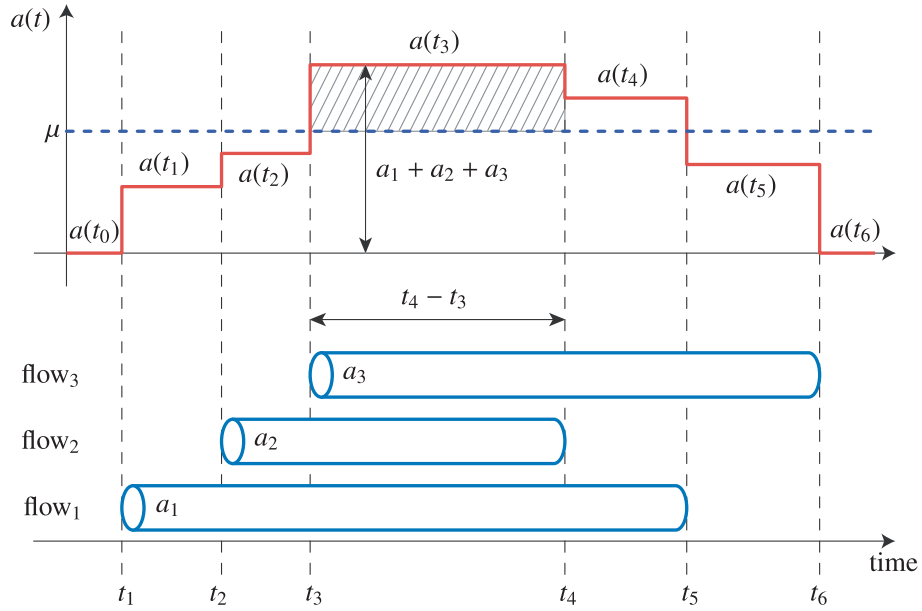


Figure 10: Calculation of the number of packets at a node.

Here, $L(t_n)$ is non-negative number, and therefore $L(t_n) = 0$ if $L(t_n) < 0$. Then, the delay experienced at a node can be calculated as $W(t_n) = L(t_n)/a(t_n)$ by applying Little's theorem.

In the flow-level simulation, only head and tail packets of flows are processed to update parameters, $a(t_n)$, $L(t_n)$, and $W(t_n)$, while in the packet-level simulation every packet in the flow have to be processed. That is, packet-level simulation updates $a(t_n)$, $L(t_n)$, and $W(t_n)$ whenever the packet arrives at the node.

4.3 Verification of the flow-level simulation method

To show the validity of the fluid flow model approximation, we compare the results of flow-level simulation with those of the packet-level simulation.

4.3.1 Simulation model

We compare two simulation methods under the same environment. In our validation, we use NSFNET topology, which has 14 nodes and 21 edges. The number of wavelength of each link is eight, and wavelength converters are not used in this simulation. The bandwidth of one wavelength is assumed to be 10 Gbps. The processing capacity of the IP router is set to 1 Gbps to perform

the packet-level simulation. The flow length is exponentially distributed with a mean value of 12 Mbyte and flows between each node pair ij arrive according to Poisson process with the average rate $\gamma \times d_{ij}$, where $D = \{d_{ij}\}$ is traffic demand matrix and γ is a scale factor of the traffic demand matrix. We use the traffic demand matrix shown in the reference [3]. Moreover, we use the same random number for both the flow-level and packet-level simulation.

4.3.2 Performance evaluations

Figure 11 shows the average end-to-end delay. Figure 11 shows the average end-to-end delay dependent on the total amount of traffic by changing the scale factor γ . From this figure, we observe that the average end-to-end delay obtained by flow-level simulation shows the similar tendency to the delay by packet-level simulation. Although the difference appears at a high load (greater than 3.5 Gbps) because of the fluid flow approximation, the saturation point is almost the same. In Figure 12, we show the queue behavior at a node. The vertical axis shows the number of packets in the node at the saturation point (3.5 Gbps), and the horizontal axis shows the simulation time. From this figure, we observe that the flow-level simulation method achieves high accuracy. The transient state, that is, the changes of router status obtained by flow-level simulation method also follows the values obtained by packet-level simulation method.

Flow-level simulation greatly reduces the simulation time. This is because in the flow-level simulation, only two packets for each flow are processed while every packet is processed in the packet-level simulation. Although, in this validation, we chose a small number of packets (1000 packets) as the average number of packets in one flow in order to finish packet-level simulation programs in a reasonable time, the number of packets that the packet-level network simulator must process is 500 times greater than that flow-level simulator must. Actual simulation time of flow-level simulation for one parameter setting is a few seconds while that of packet-level simulation is more than ten minutes. Moreover, memory requirement of flow-level simulation method is also much less than that of packet-level simulation. Thus, the computer simulation based on the fluid flow model is effective method for evaluating performance of high-bandwidth networks although the information of each packet is lost.

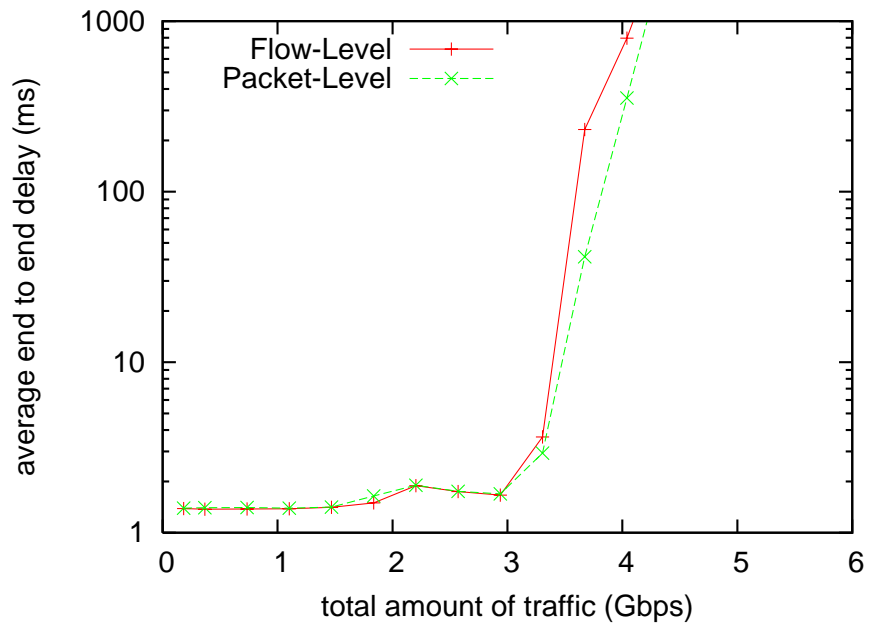


Figure 11: The average end-to-end delay of flow-level simulation and packet-level simulation.

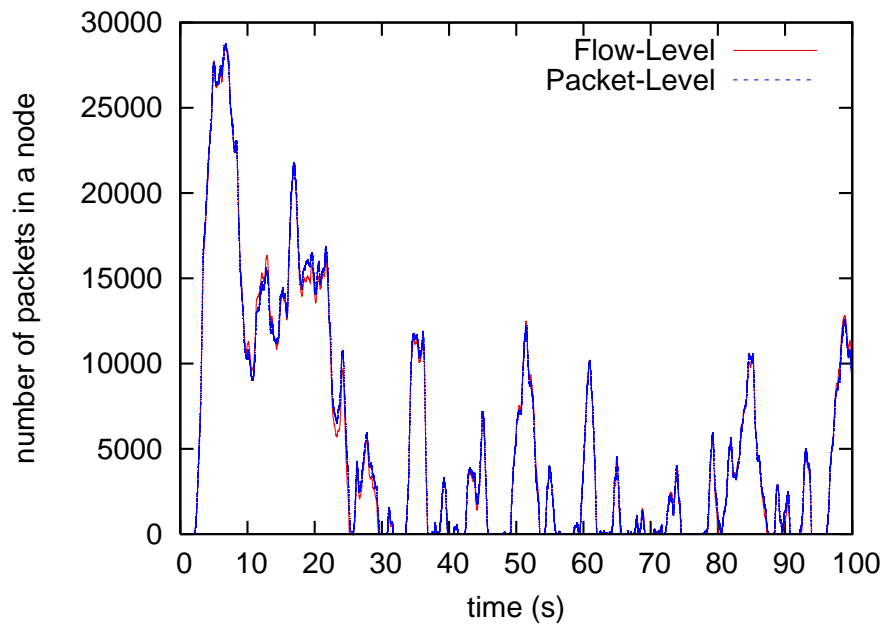


Figure 12: The number of packets in a node flow-level simulation and packet-level simulation.

5 Effects of cross-layer traffic engineering in IP over WDM networks

In this section, we evaluate the performance of the proposed algorithm through computer simulations with flow-level simulation method.

5.1 Simulation model

We evaluate the performance of the proposed routing algorithm through flow-level simulation method in Section 4 on the NSFNET (Figure 13) that has 14 nodes and 21 bidirectional links, and the European Optical Network (EON) (Figure 14) that has 19 nodes and 39 bidirectional links. A bidirectional link has an unidirectional optical fiber that multiplexes 8 wavelengths in each direction, and the transmission capacity of each wavelength is set to 10 Gbps. The processing capacity of IP router is set to 10 Gbps, and flow length is exponentially distributed with a mean value of 75 Mbyte (the size of packets is 1500 byte, and the average length of flows is 50000 packets). Under these parameters settings, the bottleneck of the network is processing capacity of routers. IP traffic flows between each node pair ij arrive according to Poisson process with the average rate $\gamma \times d_{ij}$, where $D = \{d_{ij}\}$ is traffic demand matrix and γ is scale factor of the traffic demand matrix. We use the traffic demand matrix shown in the reference [3]. In Table 1, we show this traffic demand matrix. For EON topology, we randomly generate a traffic matrix according to exponential distribution.

5.2 Static logical topology design methods

For comparison purpose, we use two logical topologies designed by SHLDA [2] (Shortest-Hop Logical topology Design Algorithm) and MLDA (Minimum delay Logical topology Design Algorithm) [3]. These methods generate logical topology based on given traffic demand matrix in order to achieve their performance objectives. MLDA use the performance metric d_{ij} for node-pair ij , where d_{ij} is the traffic demand from node i to node j , in order to construct a logical topology. First of all, it places one lightpath between two nodes if a fiber directly connecting these nodes exists. Then, it constructs a logical topology by configuring lightpaths in a descending order of traffic demand d_{ij} . Finally, if non-utilized wavelengths still exist, lightpaths are placed randomly by using those wavelengths. MLDA focuses on maximizing throughput of IP traffic by setting lightpaths in a descending order of traffic demand.

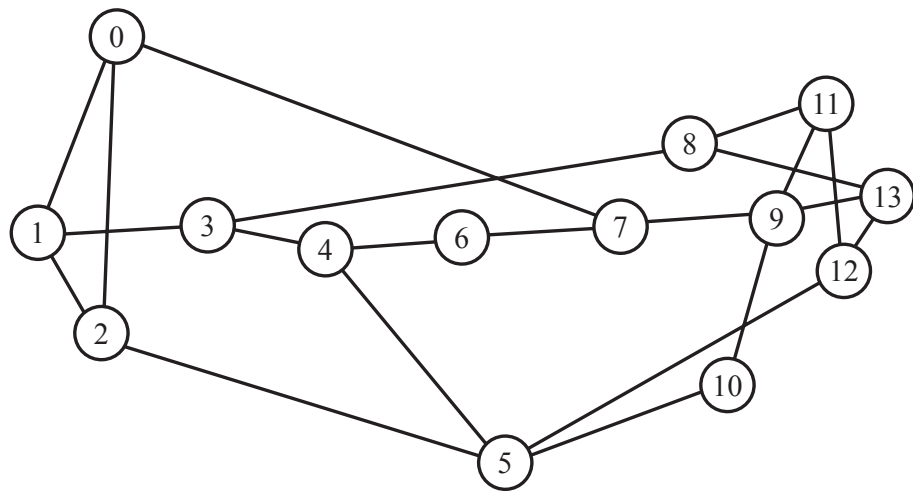


Figure 13: NSFNET topology.

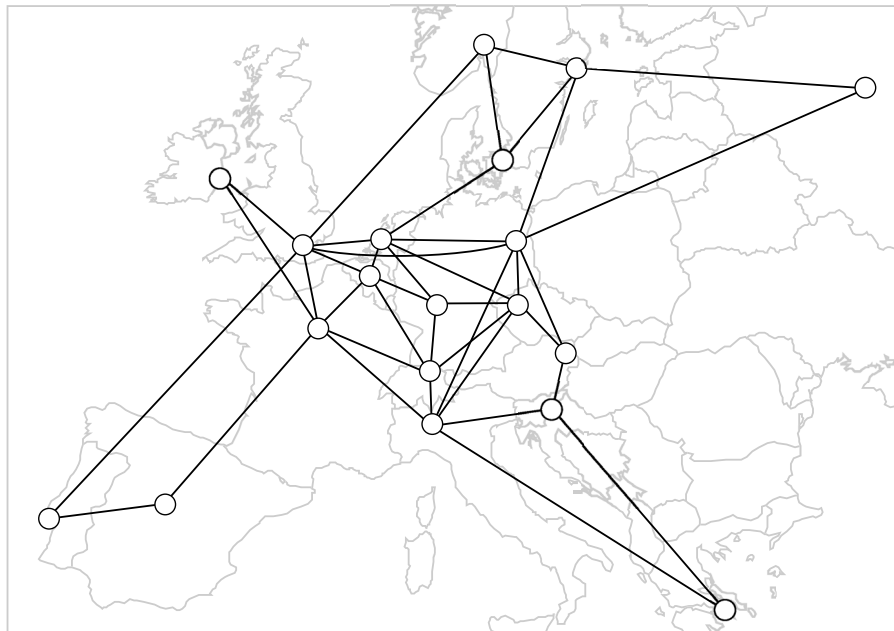


Figure 14: European Optical Network (EON) topology.

Table 1: Traffic demand matrix for NSFNET topology.

0	33.029	32.103	26.008	0.525	0.383	82.633	31.992	31.174	0.568	0.358	0.544	0.651	0.16
0.546	0	0.984	0.902	0.866	0.84	0.013	62.464	0.475	0.001	0.342	0.925	0.656	0.501
35.377	0.459	0	0.732	0.272	0.413	28.242	0.648	0.909	0.991	56.15	23.617	1.584	0.935
0.739	0.225	0.296	0	0.896	0.344	0.012	84.644	0.293	0.208	0.755	0.106	0.902	0.715
0.482	96.806	0.672	51.204	0	0.451	0.972	0.814	0.225	0.694	0.504	0.704	0.431	0.333
0.456	0.707	0.626	0.152	0.109	0	0.804	0.476	0.429	0.853	0.28	0.322	90.503	0.212
0.042	0.067	0.683	0.862	0.197	0.831	0	0.585	67.649	56.138	0.896	0.858	73.721	0.582
0.616	0.64	0.096	97.431	0.308	0.441	0.299	0	0.161	0.49	0.321	0.638	82.231	0.376
0.768	0.323	0.676	0.359	0.019	50.127	12.129	0.65	0	0.483	45.223	58.164	0.894	0.613
0.037	0.318	0.367	2.981	0.976	0.629	0.525	0.293	0.641	0	33.922	0.228	0.995	71.905
12.609	0.479	0.146	0.174	0.181	0.072	23.08	0.671	0.634	0.759	0	0.725	0.592	0.445
0.887	0.004	1.614	0.471	0.12	0.263	0.585	0.086	0.157	95.633	42.828	0	0.527	0.021
9.019	0.569	0.936	0.975	81.779	0.573	0.738	0.41	0.49	0.948	0.154	0.145	0	0.436
20.442	0.515	0.719	0.089	39.269	49.984	0.72	0.863	0.858	0.49	0.106	0.765	0.059	0

On the contrary, SHLDA takes the nature of route selection used in the IP routing protocol into account. More specifically, it is natural that a shorter path would be selected by the IP routing protocol for forwarding packets. Note here that a short path means that the number of lightpaths between two nodes is small. Therefore, SHLDA reduces hop counts of lightpaths (i.e., the number of lightpaths that the packet traverses). SHLDA uses the performance metric F_{ij} , which is determined by $F_{ij} = d_{ij} \times h_{ij}$, where h_{ij} is the hop count of the minimum hop route for node pair ij on the physical topology. Here, the hop count of the lightpath refers to the number of physical links (i.e., optical fibers) that the lightpath traverses. SHLDA configures lightpaths in a descending order of F_{ij} , and other steps are the same as MLDA. Since MLDA and SHLDA can use priori knowledge about future requests (i.e., traffic demand matrices) they generate nearly optimal topologies for the given traffic demand matrix. We compare our proposal with these near-optimal topologies in order to show efficiency of our method. In our simulation experiments, we use minimum hop routing on the logical topology generated by MLDA/SHLDA for the IP packet routing.

5.3 Performance evaluations

We evaluate the performance of our integrated routing method by three performance metrics. First, we evaluate the average end-to-end delay in Section 5.3.1, and then we evaluate the load of nodes in Section 5.3.2. Finally, in Section 5.3.3, we investigate the network throughput.

5.3.1 Average end-to-end delay

Figure 15 and Figure 16 shows the average end-to-end delay on NSFNET. The horizontal axis shows the total traffic volume in the network. Figure 15 shows the result when the traffic matrix in [3] is used. In this figure, the average end-to-end delay by our routing method and by logical topologies obtained from MLDA/SHLDA shows almost the same performance although the logical topologies by MLDA/SHDLA are designed for the given traffic demand. In Figure 16, we show the result when traffic change occurs. Here, we change the traffic matrix by randomly re-generating it every 200 seconds while keeping the sum of the matrix to be a constant value (i.e., x -axis value in the figure). As expected, changing traffic patterns degrades the performance of SHLDA and MLDA drastically. For our algorithm case, however, we observe little increase in the average end-to-end delay, and our proposed method shows almost the same performance as Figure 15 even when traffic demand matrix changes. This means that our integrated routing mechanism efficiently utilizes wavelength resources.

We next evaluate the proposal method on NSFNET topology using the random traffic matrix. Figure 17 and Figure 18 shows the average end-to-end delay on NSFNET topology dependent on the total amount of traffic. From this figure, we observe that our routing method outperforms conventional logical topology design methods. Moreover, the average end-to-end delay by our method keeps constant against the traffic change, while the delay of SHLDA or MLDA much depends on the traffic demand matrices. Figure 19 and Figure 20 depicts the results of our routing mechanism by applying to the EON topology, which shows the similar tendency that we have observed on the NSFNET topology.

5.3.2 Load of nodes

To see the efficiency of our method more clearly, we show the average load of each node in Figure 21 and Figure 22. Here, load of a node means that traffic volume processed in the node divided

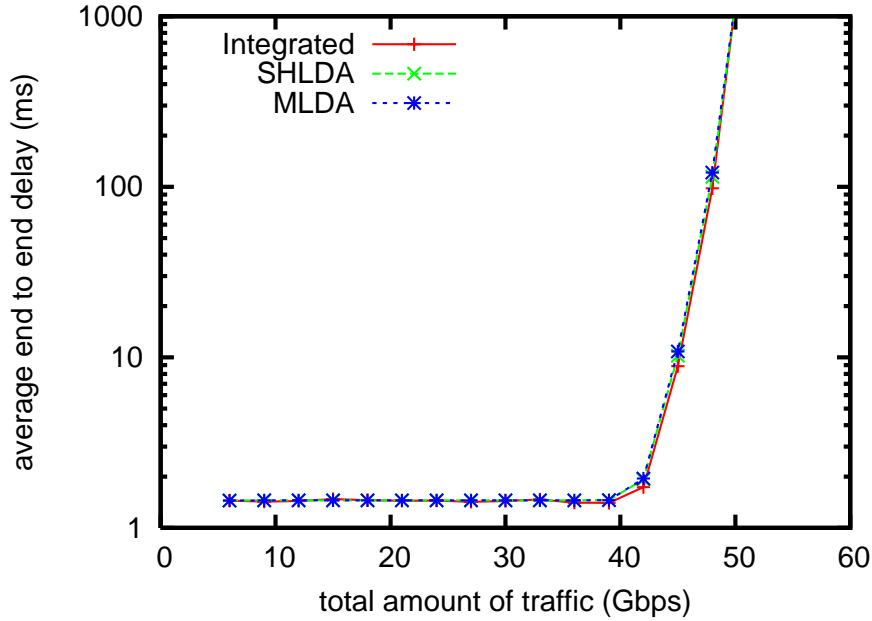


Figure 15: Average end-to-end delay without traffic change: NSFNET topology, the case of realistic traffic demand matrix.

by processing capacity of the node. In this simulation experiment, we use NSFNET topology, and traffic demand matrix in reference [3]. In obtaining this Figure, we set the total traffic demand when the average end-to-end delay begins to increase by referring Figure 15 and Figure 16. We set the total traffic volume in the network to 42 Gbps for all method in the case of both the static and dynamic traffic pattern. Figure 22 shows that our algorithm balances the load of nodes around 0.6 when slight congestion occurs in the network. On the contrary, in SHLDA and MLDA, the node 12 is being saturated while other nodes remain under-utilized. However, Figure 21 shows some nodes are high-loaded in the cases of both the proposed method and SHLDA/MLDA. To investigate the reason for this, we classify traffic volume processed in IP routers into three types; ingress traffic, egress traffic, and transit traffic. Ingress traffic is traffic that arrives at the node from outside of the network, egress traffic is traffic that leaves from network at the node. The transit traffic is traffic that is forwarded from the ingress line card to the fabric and then to the egress card. Figure 23 shows the amount of traffic volume for each traffic. For obtaining this result, we evaluate our method through NSFNET topology without traffic change. It is revealed that the dominant of the traffic volume processed at nodes is the ingress and the egress traffic, which must

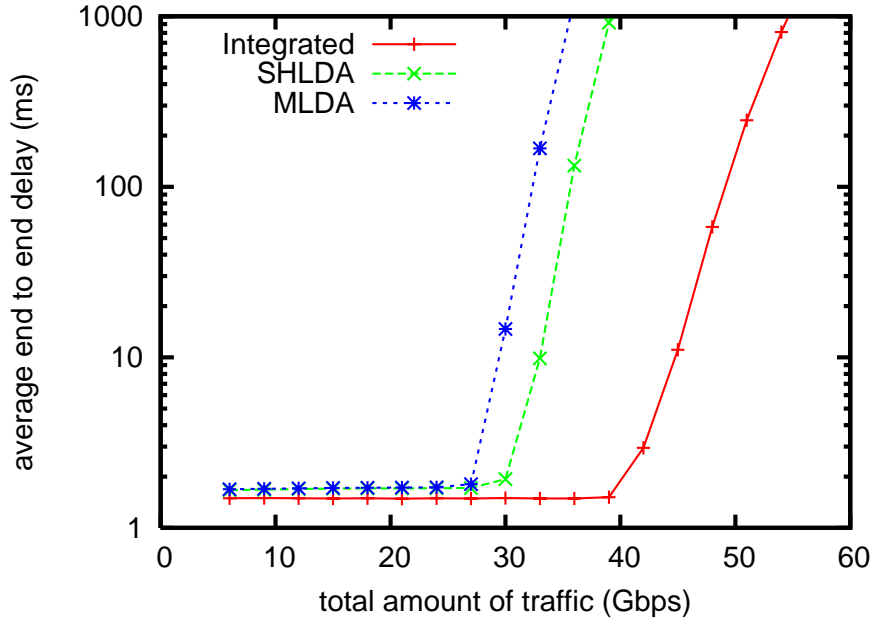


Figure 16: Average end-to-end delay with traffic change: NSFNET topology, the case of realistic traffic demand matrix.

be necessarily processed at nodes. Since the ingress traffic and the egress traffic depend on the traffic demand matrix, hereafter we focus on the behavior of the transit traffic. In Figure 23, the transit traffic processed at the nodes which have much ingress and egress traffic (e.g., node 6, 7, and 12) is nearly 0 Gbps. The transit traffic of low-loaded nodes is much less than that of high-loaded nodes. This means that, in our method, IP traffic is detoured via low-loaded nodes, and therefore, load balancing can be achieved. This leads to reduce the load of nodes.

5.3.3 Network throughput

We next show the network throughput achieved by the cross-layer traffic engineering. Here, we define the network throughput as maximum of traffic volume that the network can accommodate with keeping the average end-to-end delay less than 10 ms. In this simulation, we also use NSFNET topology. In Figure 24, we show the throughput when we use traffic demand matrix in Table 1. The throughput of our method is a little higher than those of SHLDA or MLDA when traffic does not change. Changing traffic patterns degrades the throughput of SHLDA/MLDA dramatically. When traffic patterns change, the logical topology generated by SHLDA can accommodate only

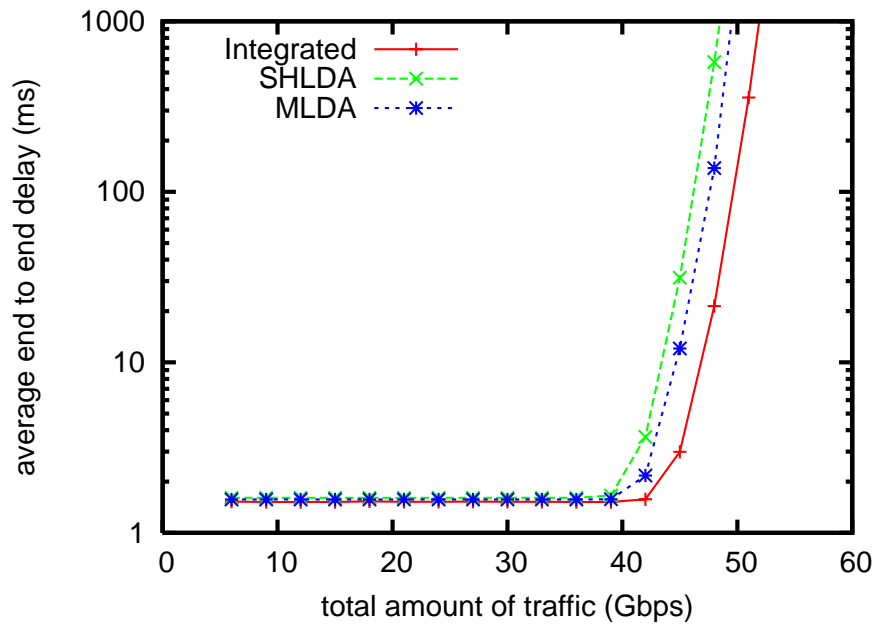


Figure 17: Average end-to-end delay without traffic change: NSFNET topology, the case of random traffic demand matrix.

56 % of traffic compared with the result without traffic change. The logical topology generated by MLDA can accommodate 62 % of traffic. However, changing traffic patterns does not affect the throughput of our method. The throughput of our method with traffic change is about only 10 % lower than that without traffic change. When traffic patterns change, our method can accommodate 79 % more traffic than SHLDA when traffic patterns change, and 54 % more traffic than MLDA.

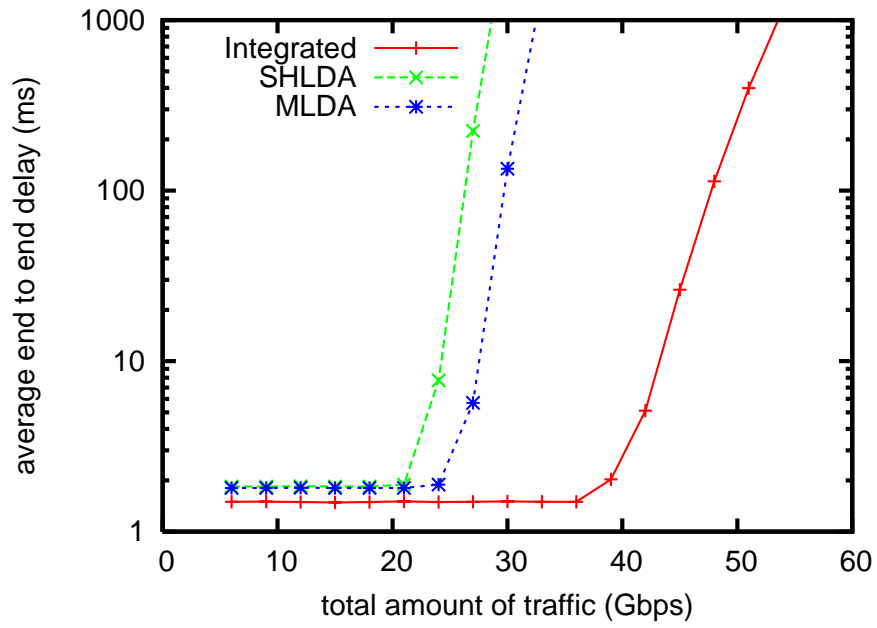


Figure 18: Average end-to-end delay with traffic change: NSFNET topology, the case of random traffic demand matrix.

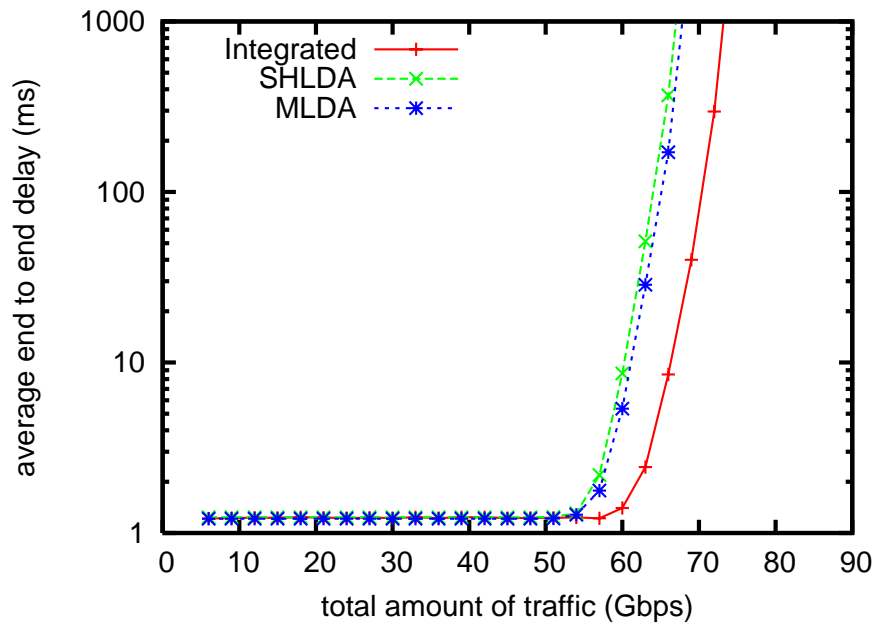


Figure 19: Average end-to-end delay without traffic change: EON topology, the case of random traffic demand matrix.

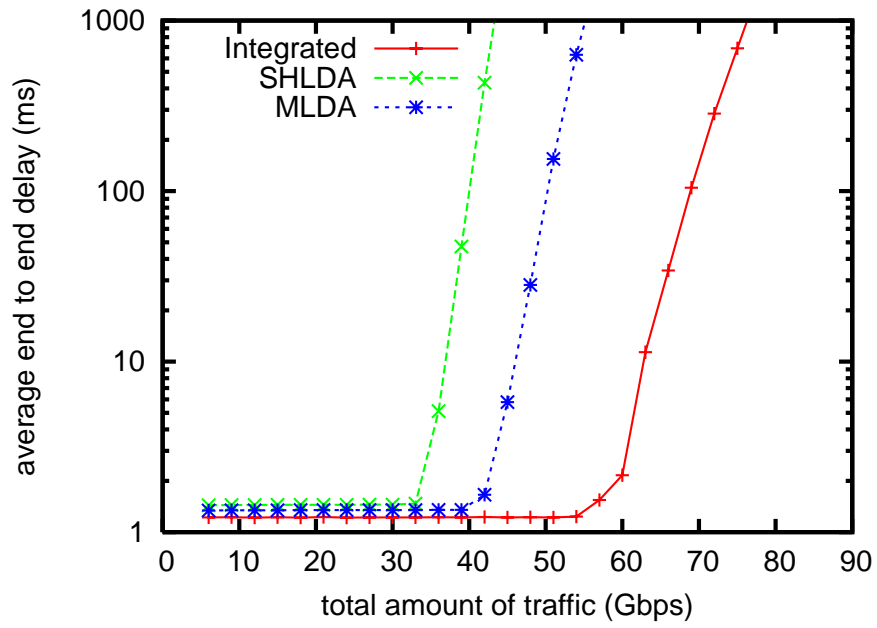


Figure 20: Average end-to-end delay with traffic change: EON topology, the case of random traffic demand matrix.

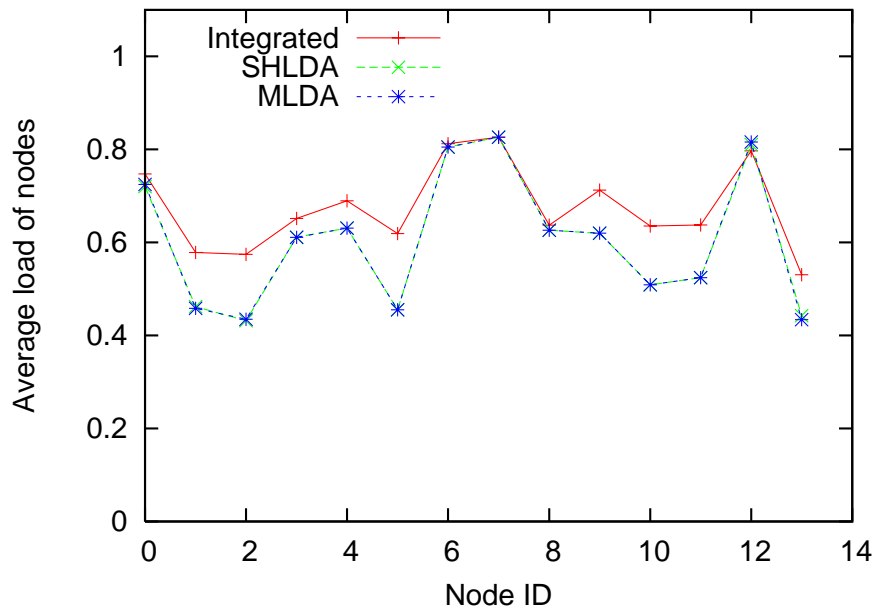


Figure 21: Load of nodes without traffic change: NSFNET topology, total traffic volume is 42 Gbps for all method.

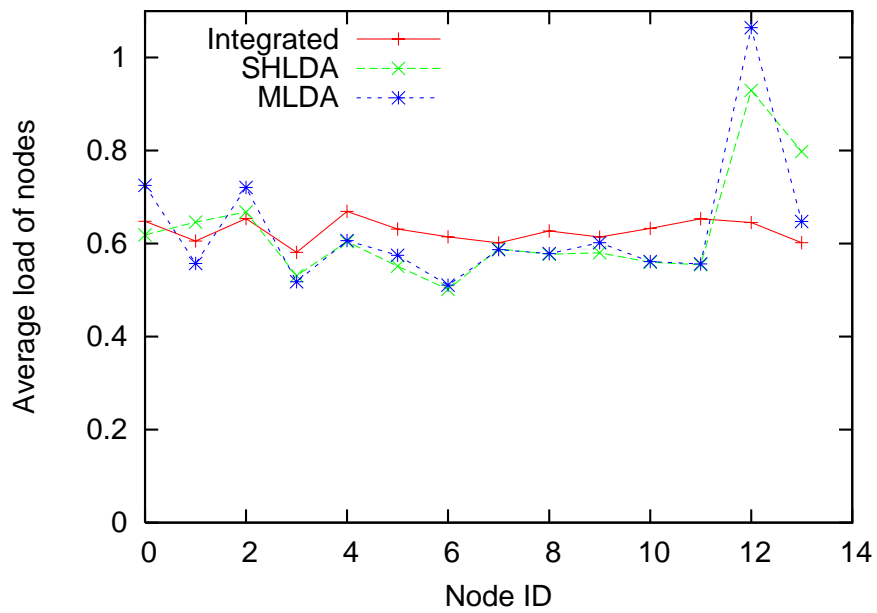


Figure 22: Load of nodes with traffic change: NSFNET topology, total traffic volume is 42 Gbps for all method.

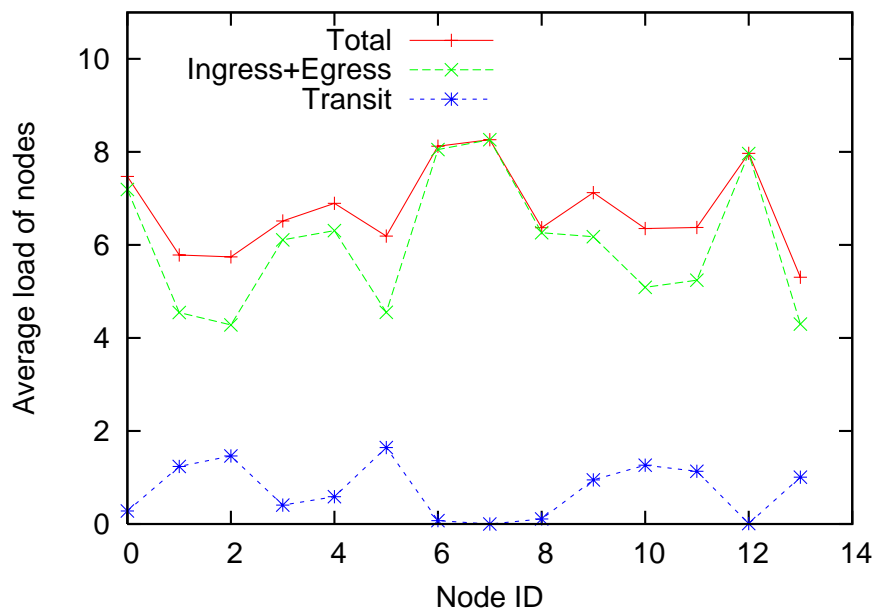


Figure 23: Amount of traffic volume of the integrated routing method: ingress, egress, and transit traffic.

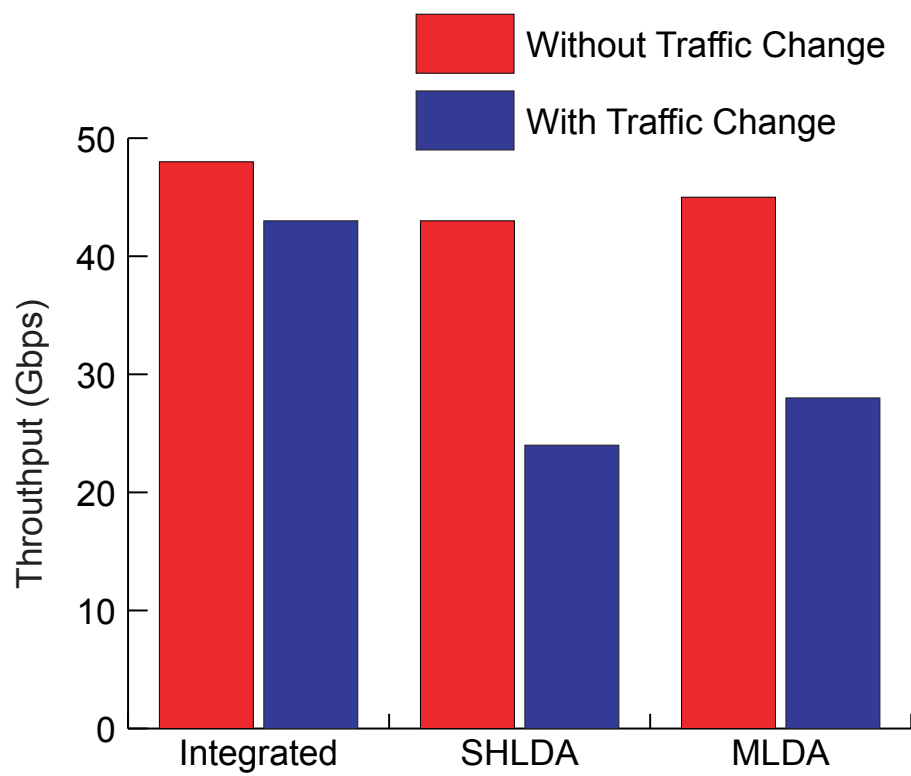


Figure 24: Network throughput.

6 Conclusion

In this thesis, we proposed a new integrated routing algorithm that used the virtual-link concept for selecting routes of IP packets, and found the necessity of dynamic lightpath configuration in IP over WDM networks. The main objective of our algorithm was to reduce the load of IP routers, and to maximize traffic volume that could be accommodated in the network. Simulation results showed that our method balanced the load of IP routers although static logical topology design methods could not. Therefore, the throughput of our method was 54 % – 79 % higher than those of static logical topology design methods when traffic patterns change. We also developed flow-level simulation method that uses the fluid flow model in order to reduce the costs of computer simulation. We verified the flow-level simulation by comparing conventional packet-level simulation, and the results of both the flow-level and packet-level simulation showed the same tendency for the performance metrics of end-to-end delay and the number of packets in nodes. The results also showed that flow-level simulation consumed about 500 times less time than packet-level, and also reduced the required memory, and thus can be applied to a large scale and high-bandwidth networks.

Acknowledgements

I would like to appreciate to my supervisor, Professor Masayuki Murata of Osaka University, for his pointed feedbacks, valuable comments, and encouragement.

All of my work would not be achieved without the support of Research Assistant Shin'ichi Arakawa. I sincerely appreciate for him to improve my work and to support preparations of this thesis.

I am most grateful to Professors Koso Murakami, Makoto Imase, Teruo Higashino, Hirotaka Nakano, and Tetsuji Satoh of Osaka University, for their appropriate guidance and invaluable firsthand advice.

Thank you to NiCT Hiroaki Harai for his advice which clarified my work. I am particularly thankful for the continued input of Associate Professor Naoki Wakamiya, Associate Professor Go Hasegawa, and Research Associate Masahiro Sasabe.

Finally, I owe a lot to my friends, colleagues and the Advanced Network Architecture Research Group of Osaka University. Our conversations and work together have greatly influenced this thesis.

References

- [1] K. H. Liu, *IP over WDM*. <http://www.wiley.com/>: John Wiley & Sons, Inc, Nov. 2002. ISBN: 0-470-84417-5.
- [2] J. Katou, S. Arakawa, and M. Murata, "A design method for logical topologies with stable packet routing in IP over WDM networks," *IEICE Transactions on Communications*, vol. E86-B, pp. 2350–2357, Aug. 2003.
- [3] R. Ramaswami and K. N. Sivarajan, "Design of logical topologies for wavelength-routed optical networks," *IEEE Journal on Selected Areas in Communications*, vol. 14, pp. 840–851, June 1996.
- [4] S. Koo, G. Sahin, and S. Subramaniam, "Dynamic LSP provisioning in overlay, augmented, and peer architectures for IP/MPLS over WDM networks," in *Proceedings of IEEE INFOCOM*, pp. 514–523, Mar. 2004.
- [5] R. Dutta and G. N. Rouskas, "A survey of virtual topology design algorithms for wavelength routed optical networks," *Optical Network Magazine*, vol. 1, pp. 73–89, Jan. 2000.
- [6] M. Kodialam and T. V. Lakshman, "Integrated dynamic IP and wavelength routing in IP over WDM networks," in *Proceedings of IEEE INFOCOM*, pp. 358–366, Apr. 2001.
- [7] J. Comellas, R. Martinez, J. Prat, V. Sales, and G. Junyent, "Integrated IP/WDM routing in GMPLS-based optical networks," *IEEE Network Magazine*, vol. 17, pp. 22–27, Mar./Apr. 2003.
- [8] D. Zhemin and M. Hamdi, "A simple integrated provisioning and protection scheme in GMPLS-based optical networks," in *Proceedings of ONDM*, pp. 295–301, Feb. 2005.
- [9] T. Ye, Q. Zeng, Y. Su, L. Leng, W. Wei, Z. Zhang, W. Guo, and Y. Jin, "On-line integrated routing in dynamic multifiber IP/WDM networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, pp. 1681–1691, Nov. 2004.
- [10] J. Li, G. Mohan, E. C. Tien, and K. C. Chua, "Dynamic routing with inaccurate link state information in integrated IP over WDM networks," *Computer Networks*, vol. 46, pp. 829–851, Dec. 2004.

- [11] “OPNET.” <http://www.opnet.com/>.
- [12] “The network simulator ns-2.” <http://www.isi.edu/nsnam/ns/>.
- [13] C. Kiddle, R. Simmonds, C. L. Williamson, and B. Unger, “Hybrid packet/fluid flow network simulation,” in *Proceedings of the 17th Workshop on Parallel and Distributed Simulation*, pp. 143–152, June 2003.
- [14] F. Baccelli and D. Hong, “Flow level simulation of large IP networks,” in *Proceedings of IEEE INFOCOM*, pp. 1911–1921, Mar. 2003.