On Design and Architecture of an IP over WDM Optical Network Control Plane

Chunsheng Xin^{1,2}, Ti-Shiang Wang¹, Yinghua Ye¹, Myungsik Yoo¹
Sudhir Dixit¹, and Chunming Qiao²

1. Nokia Research Center 5 Wayside Rd Burlington, MA01803 U.S.A

2. Department of Computer Science and Engineering, SUNY Buffalo, Buffalo, NY 14260, U.S.A

Key words: IP over Optical Network, Control Plane, WDM

Abstract: To demonstrate an IP-based routing and signalling approach in IP over WDM optical networks, we present the design and architecture of a WDM optical network control plane. With a modular design, the control plane effectively implements the key functionality such as routing, signalling, protection/restoration and quality of service support.

1. INTRODUCTION

The rapid growth of Internet and new digital services is driving the Internet service providers to provide huge bandwidth. With technical and economical feasibility, WDM optical networks are becoming an ideal infrastructure candidate in core networks and even access networks due to its potentially unlimited bandwidth. Although WDM optical networks are already in use to provide the point-to-point connection for a multiple-layer architecture, such as IP over ATM over SONET over WDM, to transport IP traffic, this approach has experienced high management cost and complexity for the service providers. Therefore, there is a consensus recently in the research community and industry to utilize a two-layer architecture, i.e. IP over WDM, to transport IP traffic directly over WDM optical networks. However, the rapid and effective bandwidth provisioning and protection/restoration remain the challenging issues for IP over WDM

The original version of this chapter was revised: The copyright line was incorrect. This has been corrected. The Erratum to this chapter is available at DOI: [10.1007/978-0-387-35491-0_28](https://doi.org/10.1007/978-0-387-35491-0_28)

A. Jukan (ed.), *Towards an Optical Internet*

© IFIP International Federation for Information Processing 2002

optical networks. Currently, there are many ongoing research efforts on the control requirements and mechanisms for IP over WDM optical networks, [1 – 9].

In this paper, we demonstrate how to use the existing IP protocols and mechanisms to introduce the intelligence into the optical crossconnect (OXC) in WDM optical networks by designing an IP-based optical control plane for the OXC. The control plane may be integrated into the same box of an OXC, or run on a separate IP/MPLS router, which is connected to the OXC. A dedicated and preconfigured wavelength, called *control channel*, between two neighboring OXCs, is terminated on both ends and used for the communication between the neighboring control planes (running on the neighboring OXCs). Each control message generated in a control plane and destined to a peering control plane goes through E-O conversion and is sent to the *control channel*. The control message goes through O-E conversion in the neighboring OXC and is received by the neighboring control plane. If this neighboring control plane is not the destination, the control message is relayed to the next hop control plane. This procedure continues until the control message arrives at the destination control plane.

This paper is outlined as follows. Section 2 introduces the general architecture of our proposed optical network control plane. A detailed description of a GUI component for the control plane is presented in section 2.1 and the function of resource management module is introduced in section 2.2. In section 2.3, we discuss the function of connection module. To illustrate the importance of survivability in optical network, section 2.4 presents the function of protection and restoration module in the control plane. Section 2.5 introduces the operation and function of the main module. Finally, in section 3, we conclude our current work and propose some future works.

2. GENERAL ARCHITECTURE

In this paper, we consider a mesh-type optical network without wavelength conversion capability. That is, the optical network that consists of edge and core optical nodes interconnected by fiber links is used to connect high-speed IP/MPLS client routers in IP networks, as shown in Fig. 1. Each optical node consists of an optical cross connect (OXC) and a control plane running upon the OXC, as shown in Fig. 2. The control plane is an IP-based controller that employs IP protocols to operate the underlying OXC. The border optical node can be connected by IP/MPLS routers (client routers). The peering control planes interact with each other to provide fast and effective provisioning (e.g. lightpath setup, teardown and abort), fault

monitoring, protection and restoration in WDM optical networks. For example, in order to set up a lightpath, the source control plane works with peering control planes to compute an end-to-end routing path and configure the switching fabric of OXCs along the path by a signalling protocol.

Figure 1. Generic Optical Network

Figure 2. Architecture of an Optical Node

In terms of routing and signalling across the IP and optical boundary (User Network Interface, or UNI), three network models (peer model, overlay model and integrated model) have been proposed for IP and optical internetworking. In the peer model, a single routing protocol instance runs

across both IP and optical domains, i.e. the OXC is a peer of a router. In the integrated model, separate routing protocol instances run in IP and optical domain respectively. However, partial routing information can be leaked between the routing protocol instances. For example, an inter-domain routing protocol (e.g. BGP) can be used to exchange the IP reachability information. In this scenario, the optical domain as a whole is a peer of the IP domain. In the overlay model, the routing and signalling in optical domain is also independent of the routing and signalling of IP domain, as in the integrated model. Furthermore, the separate routing protocol instance does not pass any routing information into each other. Therefore the border router (the router attached to optical domain) may have to register its address to optical domain and query the reachable external address (of another border router). After an initial virtual topology is formed over some border routers (their addresses are obtained through query), a routing protocol instance among the border routers can be run on this overlay topology.

In this paper, our focus is on designing an IP-based control plane for the optical domain. “IP-based” means that the control plane uses the IP protocols to control and operate the OXC in the optical network. Our objective is to demonstrate an approach to build an optical control plane using IP-based routing and signalling mechanism. As far as the internetworking with IP domain is concerned, the control plane does not differentiate the three UNI network models described above, but can be easily extended to support all of them. To demonstrate the generic functionality of the control plane without getting involved into the specifics of the UNI network models, we have designed a graphical user interface (GUI) component, called control plane GUI (CPGUI). First of all, CPGUI provides a user-friendly interface for the service provider to monitor, manage and operate the optical network. Second, CPGUI can emulate client routers and graphically demonstrate the key functionality of control plane, such as service provisioning and protection/restoration procedure. CPGUI can be deployed either as a centralized network management station or as a distributed management system on multiple nodes.

Fig. 3 shows the structure of a control plane and its interaction with CPGUI. Each control plane consists of four modules: main module (MM), connection module (CM), resource management module (RMM), and protection/restoration module (PRM). Socket connection is used to provide a communication infrastructure between the neighboring control planes, and between the control plane and CPGUI. The current implementation of socket connection is utilizing TCP stream socket.

Figure 3. Structure of a control plane and CPGUI

Figure 4. A network-wide view of control planes and CPGUI

CM is used for the lightpath signalling and maintenance (such as lightpath setup and teardown). TCP-based signalling is used by the control plane for its simplicity and flexibility. RMM is used for routing and wavelength assignment (RWA), topology and resource discovery, and service differentiation (QoS) support. PRM provides the functions of fault monitoring, and fast protection and restoration. The objective of the MM is to receive the incoming messages and works closely with other modules to process the requests. Fig. 4 shows a network-wide view of control planes

and CPGUI. Each control plane is associated with an individual optical node. In addition to the socket connections among neighboring control planes (representing the physical connectivity of the optical nodes), each control plane has its socket connection to a "node object" in CPGUI. Thus, through real-time message exchange, CPGUI can provide dynamic network views and issue client requests (for example, lightpath setup and teardown).

Figure 5. The Lightpath View and Event View of CPGUI

2.1 CPGUI

CPGUI is a tool to provide a graphical user interface to monitor, manage and operate the optical network through the information exchange with control planes. It can also be used to emulate client IP routers to request service provisioning, accept the network state information from the control plane and display the network state graphically.

In CPGUI, there are three network views: topology edit view, lightpath view and event view. In topology edit view, the network elements, such as node, link, and fiber can be created and edited to form the physical topology of the optical network. In lightpath view, CPGUI is able to perform the operations such as lightpath setup/teardown, lightpath physical path and status inquiry, lightpath protection path inquiry and its dynamic restoration

procedure, and network element state inquiry, such as OXC switching state, fiber state and wavelength state. In event view, an event history list is displayed. An event is generated when CPGUI receives a message from any control plane. The message can possibly be either an acknowledgement (ACK) of a previous request invoked by CPGUI to that control plane, or some fault notifications. Based on our design, an event is displayed in a single row that includes a series of attributes, such as event type, event generation time and source. Fig. 5 shows the lightpath view (at the top) and event view (at the bottom). In Fig. 5, there is a link failure between node 2 and node 6. The traffic from node 1 to node 10 whose primary path travels through the failure link is shifted to its protection path (node 1, node 5, node 9, node 10). Fig. 6 shows the switch state of node 5 in Fig 5. The switch is configured to switch <port 2, lambda 1> to <port 3, lambda 1> and <port 4, lambda 2> to <port 5, lambda 1> for lightpath {node 1, node 10} and lightpath {node 8, node 3}, where node 8 and node 2 are connected to port 2 and port 3 respectively, and node 1 and node 9 are connected to port 4 and port 5 respectively.

Figure 6. The Switch State of an OXC

2.2 Resource Management Module (RMM)

The main function of RMM is for RWA, topology and resource discovery, and QoS support. In the current implementation, the explicit source routing is used and the source control plane uses the shortest path first algorithm to compute an end-to-end routing path based on the network topology. The computed routing path is cached for the subsequent requests to save the per request routing computation overhead. The first-fit algorithm is used to assign the wavelength based on the resource availability along the computed routing path, which can be obtained from the resource database. The constraints based routing and other wavelength assignment will be added to enhance the resource management.

A topology connectivity matrix (TCM) is used to store the network topology. Fig. 7 shows a TCM of an optical network with N nodes, where each node is addressed by a “node ID” between 1 and N. Row vector *i* in the figure represents the link cost between node *i* and node 1 through N, where the last element is a sequence number extracted from the most recent topology update message (TUM) (to be discussed later). The link cost is generally relevant to the resource usage and the link cost “0” indicates the link is in fault or there is no link between two nodes. The active connectivity between a node and its neighbors is detected by the neighbor discovery mechanism (to be discussed later) and stored into a local connectivity vector (LCV), which is the same as a row vector in TCM with the last element excluded, “seq_num”. The link cost in LCV changes with the consideration of the link usage.

	1	-----	i	-----	N	Seq_num
1	11	-----	0	-----	8	101
⋮	⋮		⋮		⋮	⋮
row i	0	-----	6	-----	20	12
⋮	⋮		⋮		⋮	⋮
N	8	-----	11	-----	13	899

Figure 7. Topology Connectivity Matrix

Initially, when a control plane is started or rebooted, in order to build its own TCM, it will request its neighboring nodes for their TCMs. Then both this control plane and its neighboring control planes update their LCVs and broadcast their LCVs to the network. All other operational nodes in the network will accept those LCVs and update their TCMs. Within some period, the TCMs of all nodes will converge to a stable state that represents the current network topology.

To discover and maintain the active connectivity between the neighboring nodes, each control plane sends *keep-alive* messages periodically to *control channels* in all outgoing links. If a control plane cannot receive a *keep-alive* message from a neighboring control channel within some initially negotiated hold time, this link is considered as failed. The PRM will be invoked to process the fault. On the other hand, if a control plane receives a *keep-alive* from a failed link, the link is considered as restored. When the LCV is updated due to a link cost change, a link failure or restoration, it is broadcasted to the network using the same flooding mechanism as employed by OSPF.

To understand the flooding mechanism, without losing the generality, let us consider only one end of the failure link and assume this end is node A. Because of the link failure, the LCV of control plane A, named LCV_A, is updated. Then, a sequence number (SN_A) is acquired from a local sequence number generator. A topology update message, named TUM_A, which includes LCV_A, node A ID (its value is assumed to be 'k') and SN_A, is generated in RMM and sent on the *control channels* of all outgoing working links. When control plane B (CP_B) receives TUM_A, RMM of CP_B determines if this is the first time to receive TUM_A by comparing the SN_A in TUM_A with the "seq_num" in the kth row vector ("k" is node A ID) in its TCM. If SN_A in TUM_A is larger than the "seq_num" in the row vector, it means this is a new message and RMM of CP_B overwrites kth row vector of TCM with LCV_A and SN_A. Then RMM of CP_B sends TUM_A to CP_B's neighboring control planes. If the SN_A is smaller, it means the received TUM_A is an out-of-date message and CP_B deletes TUM_A.

If a control plane (denoted as CP_C) fails, every neighboring control plane of CP_C will not be able to receive the *keep-alive* message from CP_C. Then, every neighboring control plane broadcasts a TUM to the network. With all of these TUMs, any control plane can determine that CP_C failed and reset the row vector (representing the connectivity of node C to other nodes) in TCM.

RMM also manages resource availability exchange. At each control plane, the local resource availability is stored in a local resource table (LRT), as shown in Table 1. In an optical node, each port is attached with one single

fiber and the maximum number of ports is assumed as "M". The "node ID" in the row of "port k" represents the neighboring node ID connected by the fiber attached to "port k". The " λ_i -status" in the row of "port k" indicates the state of the i^{th} wavelength on the fiber attached to "port k", where $1 \leq i \leq W$. A wavelength can be in one of the following four states: "used and preemptable", "used and non-preemptable", "available", "fault". The "used and preemptable" state indicates this wavelength is being used by a low priority lightpath which can be pre-empted by a high priority lightpath.

Table 1. Local Resource Table

Port 1	node ID	λ_1 status
		...
		λ_w status
...	...	...
		...
		...
Port k	node ID	λ_1 status
		...
		λ_w status
...	...	...
		...
		...
Port M	node ID	λ_1 status
		...
		λ_w status

The global resource database is represented by a global resource table (GRT) which consists of a collection of LRTs of all nodes in the entire network. During the initialisation phase while a node starts or reboots, the establishment of initial GRT, the advertisement of LRTs and the GRTs convergence of the entire network are similar to the procedure in the topology discovery as described above. After the initialisation phase, a node is operational and ready to both accept LRTs advertised by other nodes and broadcast its own LRT when there is a change.

Another function of RMM is working with PRM and CM to support service differentiation. To demonstrate the QoS capability of the control plane, without loss of generality, we define three service types: high priority (mission critical), normal priority (protection sensitive), low priority (best-effort). According to the traffic characteristics and the end-to-end performance requirements, all client lightpath setup requests are classified into the three service types. The mission critical service needs a dedicated

protection path and the establishment of both primary and protection lightpaths may preempt the existing lightpath of best-effort type. The protection sensitive service requires a shared protection path, but the establishment of both primary and protection lightpaths cannot preempt any existing lightpath. On the other hand, once the primary and protection lightpath of this service type have been established, both of them cannot be preempted. A protection sensitive service request may be rejected even at the source node if either a primary path or a protection path cannot be found. The best-effort service only requires restoration function and may be preempted by the mission critical traffic. Upon receiving a mission critical lightpath setup request, RMM computes the primary and protection lightpaths, which may preempt some wavelength resource from the existing lightpaths. When CM of an intermediate control plane receives the setup request for a mission critical lightpath and the assigned wavelength for the mission critical lightpath is being utilized by a best-effort lightpath, a lightpath abort message is sent out for the preempted best-effort lightpath and the wavelength is allocated to the mission critical lightpath.

Table 2. Lightpath Table Entry

Lightpath ID			status	Protection restoration	Input port ID	Output port ID	Lambda ID
Source Node ID	Destination node ID	Lightpath sequence number					

2.3 Connection Module (CM)

The main function of CM is lightpath signalling and maintenance. At each node, a lightpath table (LT), maintained by CM, is used to manage all lightpaths (originating, passing-through and terminating) over the OXC. Table 2 shows the structure of an entry in LT, named lightpath table entry (LTE). In each LTE, “lightpath identifier” (lightpath ID) includes the source node ID, destination node ID and lightpath sequence number. The node ID is used to address the optical node. Although IP addressing mechanism can be used, the current implementation uses a simple addressing mechanism such that each node is assigned a unique ID (an integer) and other network elements, e.g. fiber and wavelength, are addressed using a selector (an ID, but not globally unique). Lightpath sequence number is used to identify the lightpaths originating at each node. Thus, a lightpath ID uniquely identifies a lightpath in the network domain. The “status” attribute in LTE indicates the state of a lightpath (precreate, reserved, active, or delete). The "Protection/Restoration" attribute indicates the QoS type (or service type) as described in RMM section. The “input/output port ID” represents the ID of the incoming/outgoing port of the lightpath in the OXC.

The “lambda ID” indicates which wavelength is occupied in the outgoing port specified by "output port ID".

Figure 8. Lightpath State Transfer Graph

Fig. 8 shows the life cycle of a lightpath at each control plane. We assume that the lightpath is in the "precreate" state initially. When a lightpath setup is requested, CM starts to process the request and reserve the wavelength for this lightpath. If the reservation succeeds, CM allocates an LTE in LT and sets all attributes in the LTE, with the “status” attribute set to “reserved”. When CM receives a setup ACK for this lightpath, the "status" attribute is set to “active” and the OXC switching fabric is configured for this lightpath based on the LTE information. Then the client traffic can be transported through this lightpath. If CM receives a teardown, abort, or timeout message for an existing lightpath (“reserved” or "active" state), the LTE of this lightpath in LT will be cleared and the reserved wavelength will be released. Conceptually, the lightpath is deleted.

To illustrate the end-to-end signalling procedure, we take the example of lightpath setup and consider the signalling of the working path. As a lightpath setup is requested from CPGUI, RMM of the source control plane utilizes the TCM and GRT to compute an end-to-end routing path and assign an available wavelength for the lightpath. RMM then sends the request with the computed working path to PRM which handles the protection and restoration. For the working (primary) path, RMM of the source control plane reserves the assigned wavelength by checking whether the status of this wavelength in the associated entry of LRT (specified by the output port) is available. If yes, the status for the assigned wavelength is set to "used". Then the CM of the source control plane allocates a LTE in LT for this lightpath and sets all attributes in the LTE, with the status set to “reserved”.

Afterwards, CM creates and sends a lightpath setup message, which includes the explicit routing path and the assigned wavelength computed by RMM, to the downstream control plane. Similarly, the RMM of the downstream control plane also reserves the assigned wavelength and updates LRT, and the CM allocates a LTE in LT, sets all attributes, with the status set to "reserved", and forwards the lightpath setup message to its own downstream control plane (specified by the explicit routing path). This procedure continues until the destination control plane receives the setup message and reserves the assigned wavelength. If this final reservation succeeds, a setup ACK is sent back to the source control plane along the routing path. The CMs of all intermediate control planes set the status attribute to "active" in the LTE of this lightpath and configure the OXC switching fabric based on the information of LTE. If the reservation fails at any control plane, e.g. the assigned wavelength is not available, a setup NAK is sent back to the source control plane. The intermediate control planes release the reserved wavelength and update the corresponding entry in the LT and LRT. The setup ACK or NAK will be sent to CPGUI by the CM of the source control plane.

2.4 Protection/Restoration Module (PRM)

PRM provides the function of protection/restoration path selection, fault detection and notification. Because of the similarity to compute protection path and restoration path, we take the protection path selection as an example to illustrate the merit of the PRM. Protection method includes fiber level protection (FLP) and channel level protection (CLP). In this paper, we consider to use channel level protection. CLP scheme will result in high network utilization. Because mesh network is rich in connectivity, the spare resources can be shared between multiple protection paths as long as they are not likely to simultaneously require it. We consider the end-to-end protection/restoration paths, and the disjointed-edge is selected as the restored path in PRM. This disjointed-edge path can provide fast recovery without diagnosing the fault location.

When PRM receives a protection/restoration request with the working path, it builds the Shared Risk Link Group (SRLG) of the working path. SRLG refers those resources which share the same physical ducts or fiber links as the working path. Then, PRM submits SRLG to RMM. RMM temporarily excludes SRLG from resource database (e.g. TCM and GRT) and computes the secondary lightpath as protection. Once the secondary lightpath is obtained, RMM will recover the SRLG in resource database, and send the secondary lightpath to PRM. According to the information from RMM, PRM will update its own protection/restoration routing table. In the

protection case, the secondary path is pre-calculated. Once the source control plane detects (or is notified of) a failure, it will look up its own protection routing table, and send a lightpath setup request to CM to set up the protection/restoration lightpath. If the secondary lightpath cannot be computed by RMM (e.g. due to lack of resource), the lightpath setup request will be rejected and a setup NAK is sent to the CPGUI by RMM.

Fault detection is responsible for monitoring network status and initiating protection/restoration process. It includes hardware detection to detect the low layer impairments, such as loss of signal (LOS) and higher layer detection by using the link probing mechanism. All control planes along the working path are responsible for detection. The rapid fault notification also greatly affects the protection /restoration time. While detecting the failure, the control plane will send out a failure notification by transmitting a failure indication signal (FIS) to its upstream nodes. This notification is relayed hop-by-hop to the upstream control plane until the notification reaches the source control plane. Once the source control plane receives the FIS, it will check if the reliability attribute of lightpath is protection or restoration. If the protection path is requested, the source node will look up the protection routing table and invoke the CM. If the restoration path is requested, it will involve rerouting process. For instance, the protection/restoration path selection process is invoked in PRM to calculate the rerouting path to recover the working traffic.

2.5 Main Module (MM)

The main function of MM consists of initializing the control plane, including all modules, waiting for incoming messages or requests from its neighboring control planes, CPGUI, and invoking other modules to process the message or request. When a control plane is started initially or rebooted (due to some failure), at first, the control plane establishes the *control channels* among neighboring control planes (the neighboring information is manually configured) and between a control plane and CPGUI (if CPGUI is running), and create all tables and data structures, such as LRT, LCV, LT, GRT, TCM, etc. Then the control plane queries the neighboring control planes to obtain the correct data in those tables if it is necessary. After the initialization, MM begins accepting requests from CPGUI or the neighboring control planes and invoking other modules to process the requests.

3. CONCLUSION AND FUTURE WORKS

In this paper, we have demonstrated the concept of control plane for IP over WDM optical networks. The proposed control plane is capable of several key functions, such as routing, signalling and protection/restoration. With the help of CPGUI, the proposed control plane can graphically demonstrate the service provisioning and protection/restoration. In the future, we will explore the more effective routing algorithm, which can provide the desirable traffic engineering capability to improve the resource utilization in the optical network, and the more effective protection/restoration mechanism. We will also evaluate and compare the performance of different approaches for IP over WDM integration.

4. ACRONYM

ACK: Acknowledge
CM: Connection Module
CLP: Channel Level Protection
CPGUI: Control Plane Graphical User Interface
FIS: Failure Indication Signal
FLP: Fiber Level Protection
GRT: Global Resource Table
LCV: Topology Vector
LOS: Loss of Signal
LRT: Local Resource Table
LT: Lightpath Table
LTE: Lightpath Table Entry
MM: Main Module
MPLS: Multi-protocol Lambda Switching
NAK: Negative Acknowledge
OSPF: Open Shortest Path First
OXC: Optical Crossconnect
PRM: Protection/Restoration Module
QoS: Quality of Service
RMM: Resource Management Module
RWA: Routing and Wavelength Assignment
SRLG: Shared Risk Link Group
TCM: Topology Connectivity Matrix
TUM: Topology Update Message

5. REFERENCE

- [1] N. Chandhok, etc. "IP over Optical Networks: A Summary of Issues", Internet draft, draft-osu-ipo-mpls-issues-01.txt, 2000
- [2] B. Rajagopalan, etc. "IP over Optical Networks: A Framework", Internet draft, draft-many-ip-optical-framework-01.txt, 2000
- [3] G. Wang, etc. "Extensions to OSPF/IS-IS for optical Routing", Internet draft, draft-wang-ospf-isis-lambda-te-routing-00.txt, 2000
- [4] D. Pendarakis, etc. "Routing Information Exchange in Optical Networks", Internet draft, draft-prs-optical-routing-00.ps, 2000
- [5] S. Chaudhuri, etc. "Control of Lightpaths in an Optical Network", Internet draft, draft-chaudhuri-ip-olxc-control-00.txt, 2000
- [6] D. Awduche, etc. "Multi-Protocol Lambda Switching: Combining MPLS Traffic Engineering Control With Optical Crossconnects", Internet draft, draft-awduche-mpls-te-optical-02.txt, 2000
- [7] D. Basak, etc. "Multi-protocol Lambda Switching: Issues in Combining MPLS Traffic Engineering Control With Optical Crossconnects", Internet draft, draft-basak-mpls-oxc-issues-01.txt, 2000
- [8] M. Krishnaswamy, etc. "MPLS control plane for Switched Optical Networks", Internet draft, draft-krishnaswamy-mpls-son-00.txt, 2000
- [9] Z. Tang, etc. "Extensions to CR-LDP for Path Establishment in Optical Networks", Internet draft, draft-tang-crlldp-optical-00.txt, 2000
- [10] Y. Ye, M. Ali, T.S. Wang, and S. Dixit, "Design hybrid protection scheme for survivable wavelength-routed optical transport networks", Proc. Of European Conference on Networks & Optical Communication, June 2000
- [11] Y. Rekhter, etc. "A Border Gateway Protocol 4 (BGP-4)", RFC 1771