FLOW OPTIMIZATION IN MPLS OVER DWDM

A dissertation submitted towards the partial fulfillment of requirements

for the

award of the degree

Of

Master of Engineering

(Electronics and Communication Engineering)

By

PRADEEP KUMAR

College Roll No. : 09/E&C/06

University Roll No. : 12281

Under the supervision and guidance of

Mr. Rajesh Rohilla

Asst. Professor (E&C)

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

DELHI COLLEGE OF ENGINEERING

UNIVERSITY OF DELHI

2008-09

CERTIFICATE

DELHI COLLEGE OF ENGINEERING

(Govt. of National Capital Territory of Delhi)

BAWANA ROAD, DELHI – 110042

Date: _____

This is to certify that the work contained in this dissertation entitled **"Flow Optimization In MPLS Over DWDM"** submitted by **Pradeep Kumar,** University Roll No- **12281** in the requirement for the partial fulfillment for the award of the degree of Master of Engineering in Electronics & Communication Engineering,

This is an account of his work carried out under our guidance and supervision in the academic year 2008-2009.

Mr. Rajesh Rohilla
Asst. Professor
Dept. of Electronics and Communication
Delhi College of Engineering, Delhi

ACKNOWLEDGEMENT

It is a great pleasure to have the opportunity to extend my heartiest felt gratitude to everybody who helped me throughout the course of this thesis.

It is distinct pleasure to express my deep sense of gratitude and indebtedness to my learned supervisor Mr. Rajesh Rohilla for their invaluable guidance, encouragement and patient reviews. I am very thankful to Prof. Asok Bhattacharyya H.O.D Electronics & Communication Engineering Department who allows me to do project under the Guidance of Mr. Rajesh Rohilla on Flow Optimisation in MPLS over DWDM. With their continuous inspiration only, it becomes possible to complete this dissertation and both of them kept on boosting me with time, to put an extra ounce of effort to realize this work.

I would also like to take this opportunity to present my sincere regards to all the faculty members of the Department for their support and encouragement.

Pradeep Kumar
M.E. (Electronics & Communication)
College Roll No. 09/E&C/06
University Roll No. 12281
Dept. of Electronics & Communication Engineering
Delhi College of Engineering, Delhi-110042

ABSTRACT

Multicasting is a useful service to support applications, like video/audio on-demand or teleconferencing, consume a large amount of network bandwidth because of, first the volume of the transmitted data and, second the larger number of application members. The shared aggregated tree approach in multilayer network enables optimization of the whole network in much more effective way comparing to the single-layer method, where each layer is optimized separately what cannot guarantee the global optimality of the solution.

In this thesis we have developed an algorithm for the evaluation of the performance of label aggregation mechanism. We are using a mechanism of group aggregation for multicast in MPLS networks to reduce related states during the multicast process and alleviate the usage problem of the limited label space. The essence of this mechanism is to let the multicast sessions that have the same multicast tree share the same label. But because not all of group/tree mapping are perfect match, we proposed a new approach of leaky tree matching to further reduce the label space and improve the state scalability.

Table of Contents

Table of Figures

Chapter 1

INTRODUCTION

1.1 OBJECTIVE

The objective of this dissertation is to study IP multicasting in MPLS networks and optimize the flow of data packet. The performance of multicast label aggregation algorithm can be evaluated on the basis of average number of label and bandwidth utilized metrics. It will also compare the performance of algorithms used for label aggregation

1.2 BACKGROUND

Many applications, like video/audio on-demand or teleconferencing, can consume a large amount of network bandwidth because of, first the volume of the transmitted data and, second the larger number of application members. Multicasting is a useful service to support such applications. Can be describe as follows:

- IP multicast is an efficient way of delivering data to a large group of receivers by sharing link bandwidth.
- It utilizes a tree structure, on which data packets are duplicated only on branch nodes and forwarding over each node is done once.

The challenges in designing optimized multicasting network therefore are to overcome these limitations.

- **IP multicast to a medium or large national network** -With many multicast groups, we will have state scalability problems which arise more with more different active multicast groups. This causes the forwarding table to grow more and more, indicating more memory requirement and slower forwarding processing.

- **QoS-** if QoS is applied to multicast, the problem becomes even worse, because for individual multicast group we need to maintain resources besides routes, e.g. bandwidth, delay.

Real-time multicast applications need mechanism to support QoS, but the most of architecture does not handle QoS efficiently, due to multicast routing state not scaling well.

Aggregate multicasting The idea of aggregated multicast is that, instead of constructing a tree for each individual multicast session in the backbone network, one can have multiple multicast sessions share a single tree to reduce multicast state and tree maintenance overhead at the network core.

The shared aggregated tree approach in multilayer network enables optimization of the whole network in much more effective way comparing to the single-layer method, where each layer is optimized separately what cannot guarantee the global optimality of the solution.

There are various options for distributing multicast traffic of different groups over a shared aggregated tree. MPLS (Multiprotocol Label Switching) is an efficient solution in which labels can identify different aggregated trees.

Multiprotocol Label Switching - MPLS is a switching technology placed at L3 OSI model, which fully improves performances of a core network. The main idea is to forward packets based on a short label, fixed length, instead on network address. Labels are assigned to packets when entering an MPLS domain. Inside an MPLS domain, forwarding decision is solely based on the labels. When leaving the MPLS domain, labels are removed and packets are forwarded in the conventional fashion.

According to many opinions coupling MPLS (Multiprotocol Label Switching) on the top of wavelength-routed WDM (Wavelength Division Multiplexing) is an interesting proposal for constructing of transport networks.

Modeling of multilayer networks is more complex comparing to single layer approach. In this thesis we assume that the network is already constructed - we do not consider facility capacity planning and topological design.

1.3THE APPROACH

In this thesis we have developed an algorithm for the evaluation of the performance of label aggregation mechanism. We are using a mechanism of group aggregation for multicast in MPLS networks to reduce related states during the multicast process and alleviate the usage problem of the limited label space. The essence of this mechanism is to let the multicast sessions that have the same multicast tree share the same label. But because not all of group/tree mapping are perfect match, we proposed a new approach of leaky tree matching to further reduce the label space and improve the state scalability.

But in this approach there may be some bandwidth wasted. In order to control load balance and avoid the wasting bandwidth, we are using the concept of the bandwidth threshold. In our work we are using DWDM network backbone at layer two to improve the bandwidth requirement and provide protection for multicast session.

We use the architecture of MPLS over optical network, taken through the overlay model, which is the most practical model. We developed an interactive simulation program on Matlab for evaluation and analysis the performance of different mechanism.

Chapter 2

WDM OPTICAL NETWORKS

Fiber optic communications have provided us with high-speed communications with enormous bandwidth potential. Although fibers can support very high data rates (nearly 50 terabits per second,), the associated electronic processing hardware will typically not be able to keep up with such speeds. Hence, electronic handling of data at network nodes basically limits the throughput of the network. Further, electronic processing is required because optical storage and processing technologies are not mature yet. Hence, data that must be stored or processed at an intermediate node has to be converted to its electronic form and stored in an electronic buffer memory. A routing decision is then made and the data is then queued at the output port, converted back into its optical form and transmitted towards its final destination.

Networks can be classified into three generations depending on the technology used at the physical level. The first generation networks used copper based technologies. Second generation networks used a combination of copper and optical technologies. Third generation networks are all optical networks. These networks are yet to become practical because of the challenges involved in routing and buffering in the optical domain without an intermediate conversion to the electronic domain.

Current networks use a mix of copper and optical based technologies. To improve the throughput of the network and to minimize transmission delay the network architecture must both reduce the number of times a message is processed by the intermediate nodes (and thus reduce the number of times an optical signal is converted back and forth between the electronic and optical domains) and must streamline the processing at each node. The irregular nature of most existing networks doesn't necessarily allow this. Hence complex routing tables are often used to make routing decisions less complicated and less time-

consuming. If the network could be connected in a regular uniform pattern, routing decisions could be significantly simplified thereby reducing the processing times at the intermediate nodes. But because of many real world constraints, a regular uniform pattern in building a network may not be feasible. Also economic reasons necessitate reuse of existing fiber connections in networks which restricts the physical topology options.

2.1 Optical Fiber Principles

Light has an information carrying capacity 10,000 times greater than the highest radio frequencies. Advantages of optical fibers over copper transmission lines include the ability to carry signals over long distances with low error rates, immunity to electrical interference, and security. The first fiber optic communication had been experimentally tested in the nineteenth century. However, it was in the second half of the twentieth century that the technology began to advance rapidly and began being used in practical networks. After the viability of transmitting light over fiber had been established, the next step in the development of fiber optics was to find a light source that would be sufficiently powerful and narrow. *Light emitting diodes* (LED) and *laser diodes* (LD) proved capable of meeting these requirements. Researchers in the mid 1960s proposed that optical fiber might be a suitable transmission medium for light. There was an obstacle, however, and that was the loss of signal strength (or attenuation) in the glass with which they were working. In 1970, Corning produced the first communication-grade *fibers*. With attenuation less than 20 decibels 2 per kilometre (dB/km), this purified glass fiber exceeded the threshold for making fiber optics a viable technology.

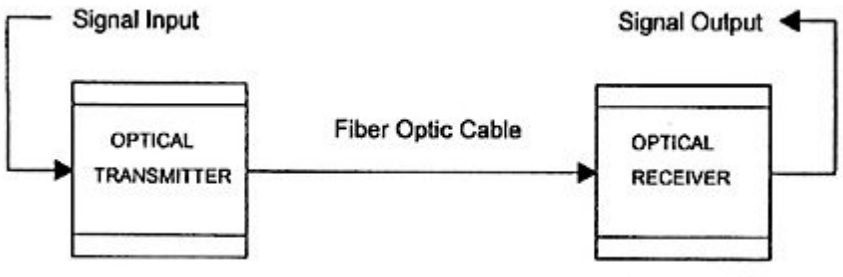

Figure 2.1 Optical Transmission System

A basic optical communication system shown in Figure 2.1 consists of an optical *transmitter*, an optical *receiver* and optical fiber as the communication medium. There are several other components that go into the system to make it practical, such as optical add/drop multiplexers, optical amplifiers, switches and wavelength converters.

An optical fiber is made of very thin glass rods composed of two parts: the inner portion of the rod or *core* and the surrounding layer or *cladding*. The core and cladding have different *indices* of refraction with the core having n_1 and the cladding n_2 ($n_1 > n_2$). Light injected into the core of a glass fiber will follow the physical path of that fiber due to the total internal reflection of the light between the core and the cladding. A plastic sheathing around the fiber provides the mechanical protection known as the jacket. It protects the core and cladding from shocks that might affect their optical or physical properties.
There are two general categories of optical fiber: *single-mode* and *multimode.*

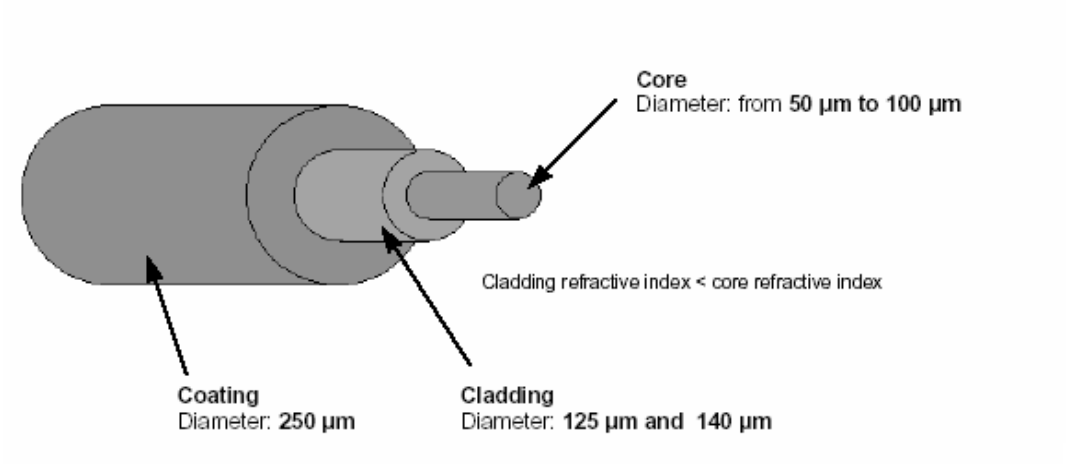

Figure 2.2 Multimode Fibre

Multimode fiber (Figure 2.2) was the first type of fiber to be commercialized. It has much larger core than single-mode fiber, allowing hundreds of modes of light to propagate through the fiber simultaneously. Additionally, the larger core diameter of multimode fiber facilitates the use of lower-cost optical transmitters (such as light emitting diodes or vertical cavity surface emitting lasers) and connectors. Single-mode fiber, on the other hand, has a much smaller core that

allows only one mode of light at a time to propagate through the core. While it might appear that multimode fibers have higher capacity, in fact the opposite is true. Single-mode fibers are designed to maintain *spatial* and *spectral integrity* of each optical signal over longer distances, allowing more information to be transmitted. Its tremendous information-carrying capacity and low intrinsic loss have made single-mode fiber the ideal transmission medium for a multitude of applications. Single-mode fiber is typically used for longer-distance and higher-bandwidth applications (see Figure 2.3).

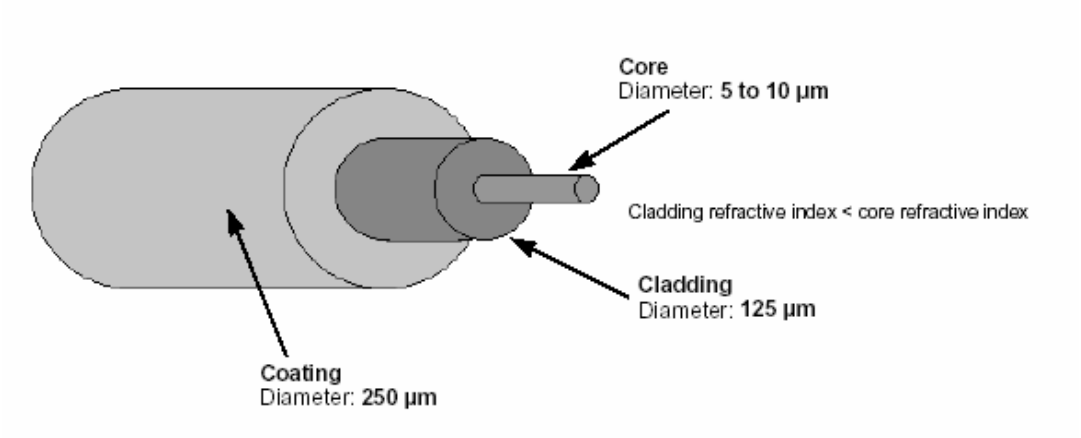

Figure 2.3 Single-mode fibre

2.2 Wavelength Division Multiplexing

WDM enables the utilization of a significant portion of the available fiber bandwidth by allowing many independent signals to be transmitted simultaneously on one fiber, with each signal being on a different wavelength). Routing and detection of these signals can be accomplished independently, with the wavelength determining the communication path by acting as the signature address of the origin, destination or routing. Components are therefore required that are wavelength selective, allowing for the transmission, recovery, or routing of specific wavelengths. A simple WDM system is shown in Figure 2.4.

After being transmitted through a high-bandwidth optical fiber, the combined optical signals must be de-multiplexed at the receiving end. One way to do that is

to distribute the total optical power to the output ports and then require that each receiver electively recovers only one wavelength using a tuneable optical filter. Each laser is modulated at a given speed and the total aggregate capacity being transmitted along the high-bandwidth fiber is the sum total of the bit rates of the individual lasers.

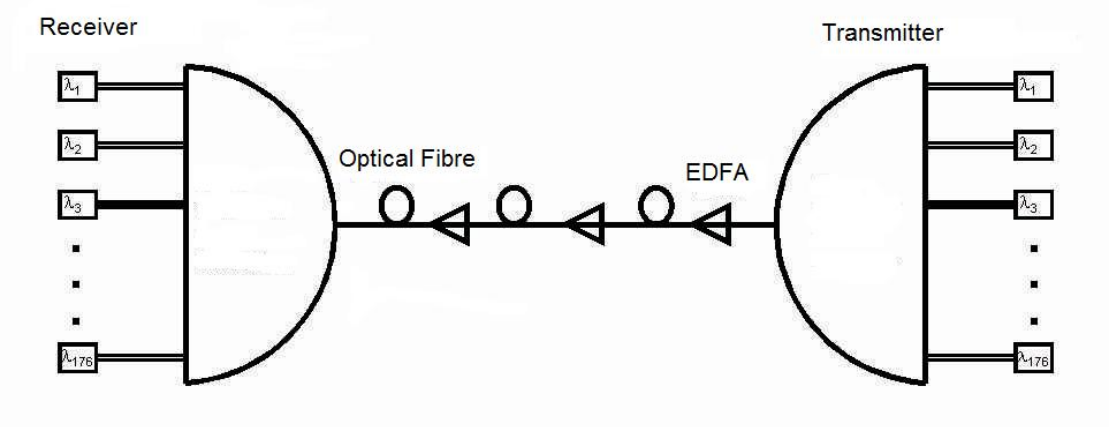

Figure 2.4 A Simple WDM System

2.3 Components of WDM System

2.3.1 Optical Amplifiers

In fiber optic communications systems, problems arise from the fact that no fiber material is perfectly transparent. The visible-light or infrared (IR) beams carried by a fiber are attenuated as they travel through the material. This necessitates the use of *repeaters* in spans of optical fiber longer than about 100 kilometers.

A conventional repeater puts a modulated optical signal through three stages: (1) optical-to-electronic conversion, (2) electronic signal amplification, and (3) electronic-to optical conversion. Repeaters of this type limit the bandwidth of the signals that can be transmitted in long spans of fiber optic cable. This is because, even if a laser beam can transmit several gigabits per second of data, the electronic circuits of a conventional repeater cannot.

The commercial development of WDM networks was made possible by the development of *optical amplifiers* known as *EDFA's* (*Erbium Doped Fiber*

Amplifiers) which provide a way to optically amplify all the wavelengths at the same time, regardless of their individual bit rates, modulation schemes or power levels. An EDFA amplifier is an optical repeater that amplifies a modulated laser beam directly, without opto-electronic and electro-optical conversion. The device uses a short length of optical fiber doped with the rare earth element erbium. When the signal-carrying laser beam pass through this fiber, external energy is applied, usually at infrared wavelengths. This so called pumping excites the atoms in the erbium-doped section of optical fiber increasing the intensity of the laser beams passing through. The beams emerging from the EDFA retain all of their original modulation characteristics, but are higher in energy than the input beams.

2.3.2 Add/Drop Multiplexer

In many WDM networks it is necessary to drop some traffic at intermediate points along the route between the end points. A wavelength add/drop multiplexer (WADM) is used for that purpose. A typical WADM is shown in Figure 2.5.

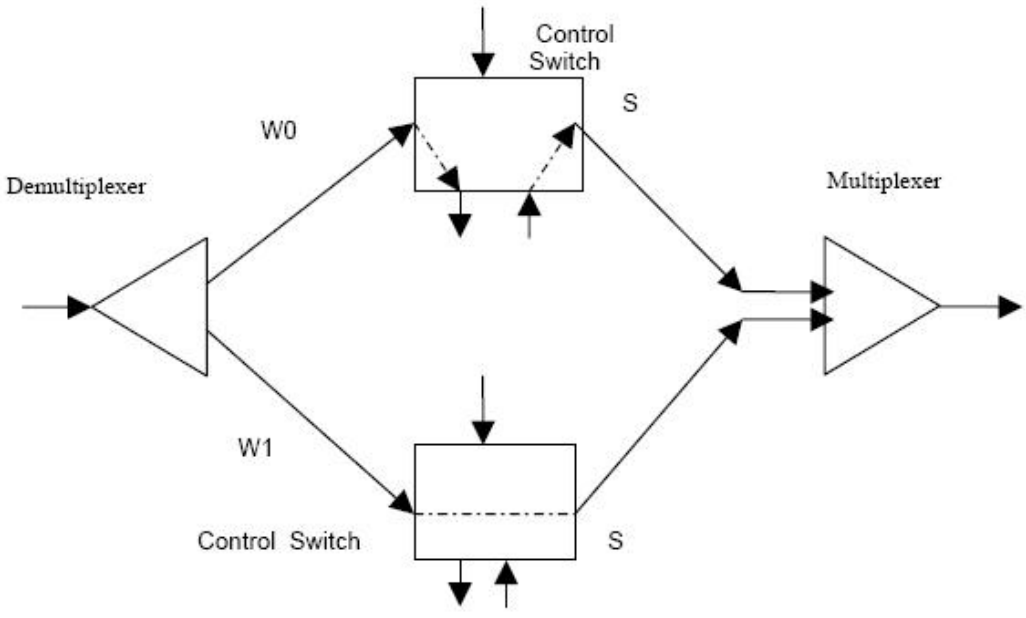

Figure 2.5 Wavelength Add/Drop Multiplexer

A WADM can be realized using 2x2 switches and a de-multiplexer. If the control switch is in the *bar state*, then the signal on the corresponding wavelength passes through the WADM. If the switch is in the *cross state*, then the signal on

the corresponding wavelength is dropped locally, and another signal of the same wavelength may be added.

2.3.3 Wavelength Cross-connect

Efficient use of fiber facilities at the optical level obviously becomes critical as service providers begin to move wavelengths around the world. In the optical domain, a network element is needed that can accept various wavelengths on input ports and route them to appropriate output ports in the network. *Routing* and *grooming* are key areas that must be addressed. This is the function of the OXC, as shown in Figure 2.6.

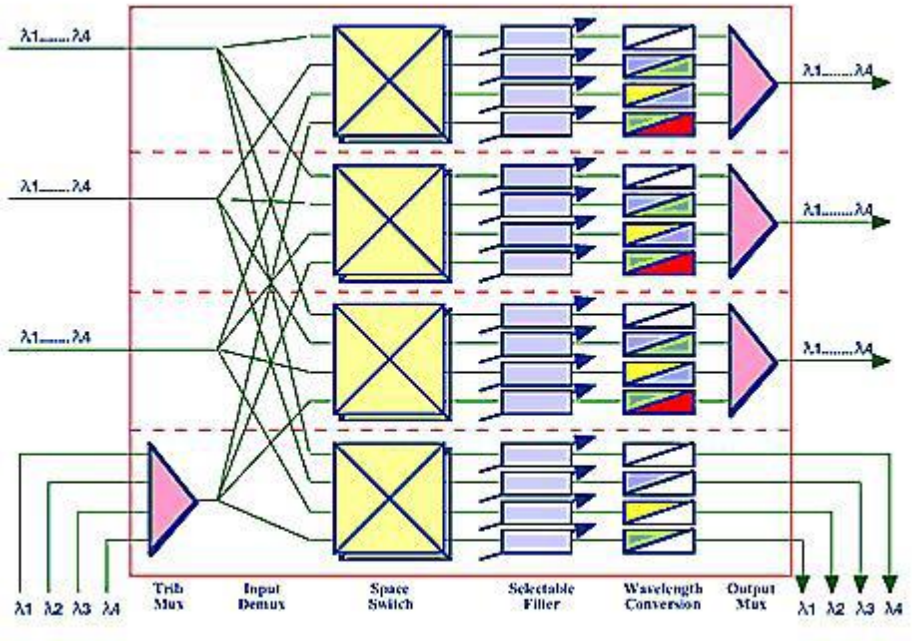

Figure 2.6 Wavelength Cross-connect Block Diagram

The function of this element is to provide (under network control), the ability to connect or switch any input wavelength channel from an input fiber (or port) to any one of the output fibers (or ports) in the optical domain. Digital cross-connect systems are deployed and provide the critical function of grooming traffic to fill output ports on the system efficiently. To accomplish this, the OXC needs three building blocks (See Figure 2.7):

Fiber switching - the ability to route all of the wavelengths on an incoming fiber to a different outgoing fiber.

Wavelength switching - the ability to switch specific wavelengths from an incoming fiber to multiple outgoing fibers.

Wavelength conversion - the ability to take incoming wavelengths and convert them (on the fly) to another optical frequency on the outgoing port. This may be necessary to achieve strictly non-blocking architectures when using wavelength switching.

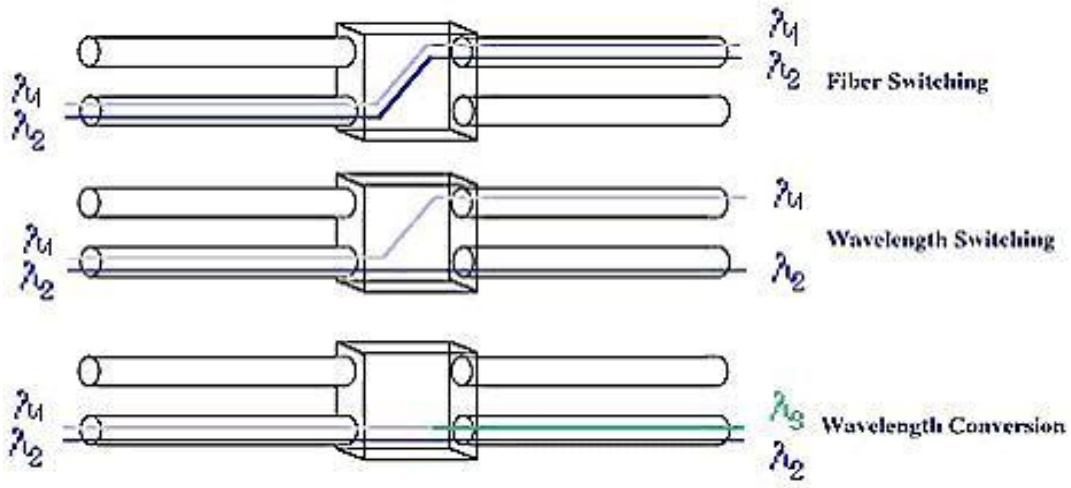

Figure 2.7 Optical Cross-Connects

2.4 WDM Optical Network Architectures

Three major classes of WDM optical network architectures are, broadcast and select networks, wavelength routed networks, and linear lightwave networks.

2.4.1 Broadcast and Select Networks

In this type of network (see Figure 2.8) all the nodes are connected to a central *star coupler*, which mediates all the communications among the nodes in the network. All the nodes in the network are equipped with fixed number of tuneable transmitters and receivers. To receive a particular wavelength all the destinations tune their receivers to that wavelength and start receiving the signal. All simultaneous transmissions occur at different wavelengths.

The role of the star coupler is to combine all these signals and then to broadcast the combined signal to all the nodes. Multiple communications can take place concurrently by appropriately tuning the receivers. These networks involve single hop transmission to the destination without any intermediate opto-electronic conversion. The problem with this type of network is that of collision which occurs when two or more nodes try to transmit simultaneously on the same wavelength. Also, power is not efficiently utilized.

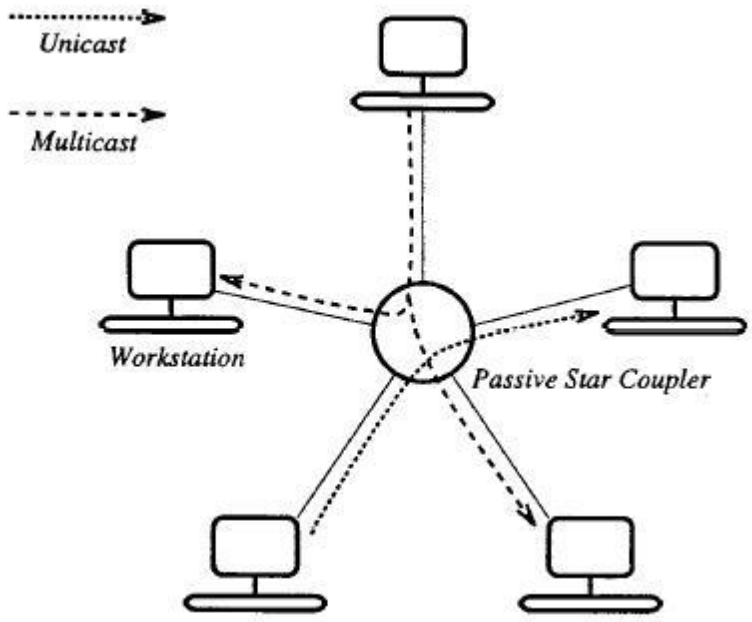

Figure 2.8 Broadcast and Select Network

2.4.2 Wavelength Routed Networks

A wavelength routed network consists of *Wavelength Cross Connects* interconnected by point-to-point fiber links in an arbitrary topology. Each end-user is connected to an active switch via a fiber link. Each node consists of transmitters and receivers, both of which may be wavelength tuneable. See Figure 2.9 for illustration.

The end-nodes tune their transmitters and receivers to the wavelength used for the *lightpath*. A lightpath is an all-optical communication channel between two

nodes in the network and may span more than one fiber link. The basic requirement in a wavelength routed optical network is that no two lightpaths traversing the same fiber link can use the same wavelength channel and that a lightpath uses the same wavelength across all links that it traverses. Selecting routes and wavelength is referred to as the routing and wavelength assignment (RWA) problem.

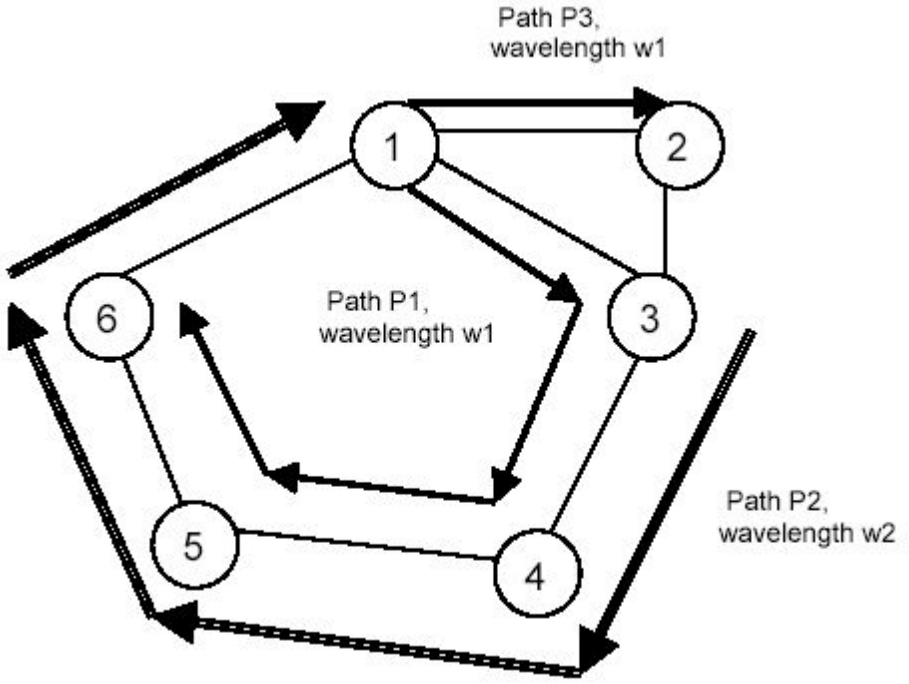

Figure 2.9 Wavelength Routed Network

2.4.3 Linear Lightwave Networks

These networks utilize the idea of partitioning the usable optical spectrum into wavelengths or *wavebands* as shown in Figure 2.10. These networks use two levels of partitioning and several such wavebands are multiplexed on a fiber. Several wavelengths are multiplexed onto a single waveband. So, unlike a wavelength routed network, linear lightwave network nodes de-multiplex, switch, and multiplex wavebands not wavelengths. Since the linear lightwave network doesn't distinguish between individual wavelengths within a waveband individual wavelengths are separated from each other at the receiving node.

These networks have the additional constraints of *inseparability* and *combining* signals from distinct sources. According to the inseparability constraint, channels belonging to the same waveband when combined on the same fiber cannot be separated within the network. Thus they travel together after the point where they were combined. The distinct source combining constraint states that on any fiber only signals from distinct sources are allowed to be combined.

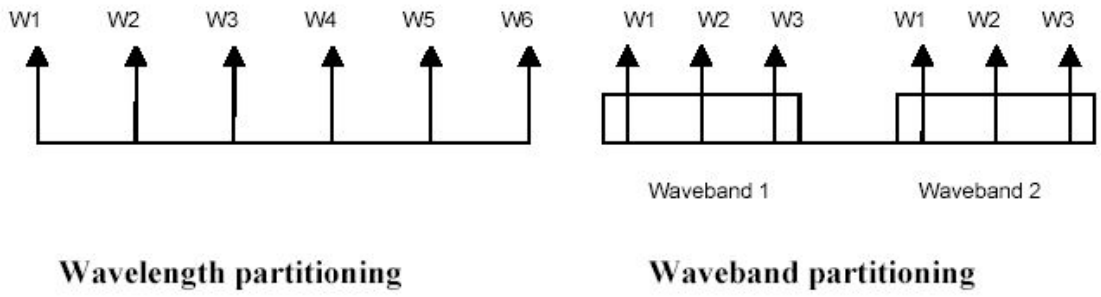

Figure 2.10 Wavelength and Waveband Partitioning

2.5 Future of WDM Optical Networks

Optical networks provide easy to manage high bandwidth services both for Internet exchanges and for local area networking applications. The introduction of WDM technology in optical fiber networks can be considered as a way of replacing the central switching functions to distributed network functions like optical add/drop multiplexers or repeaters. Thus, WDM supported Optical Internet and related services are expected to be a major driving force in future networking architectures. It is predicted that it will contribute towards substantially reducing the complexity and thus the cost of future Internet services. There is a great potential to eventually move to an all optical switching and all optical routing architectures as these technologies mature.

Chapter 3

MPLS

Multiprotocol Label Switching (MPLS), originating in IPv4, was initially proposed to improve forwarding speed. Its core technology can be extended to multiple network protocols, such as IPv6, Internet Packet Exchange (IPX), and Connectionless Network Protocol (CLNP). That is what the term multiprotocol means. MPLS integrates both Layer 2 fast switching and Layer 3 routing and forwarding, satisfying the networking requirements of various new applications.

3.1 MPLS AND ITS COMPONENTS

MPLS is a switching technology placed at L3 OSI model, which fully improves performances of a core network. The main idea is to forward packets based on a short label, fixed length, instead on network address. Labels are assigned to packets when entering an MPLS domain. Inside an MPLS domain, forwarding decision is solely based on the labels. When leaving the MPLS domain, labels are removed and packets are forwarded in the conventional fashion. Following are its basic components.

3.1.1 FEC

As a forwarding technology based on classification, MPLS groups packets to be forwarded in the same manner into a class called the forwarding equivalence class (FEC). That is, packets of the same FEC are handled in the same way.

The classification of FECs is very flexible. It can be based on any combination of source address, destination address, source port, destination port, protocol type and VPN. For example, in the traditional IP forwarding using longest match, all packets to the same destination belongs to the same FEC.

3.1.2 Label

A label is a short fixed length identifier for identifying a FEC. A FEC may correspond to multiple labels in scenarios where, for example, load sharing is required, while a label can only represent a single FEC

A label is carried in the header of a packet. It does not contain any topology information and is local significant. A label is four octets, or 32 bits, in length. Figure 3.1 illustrates its format.

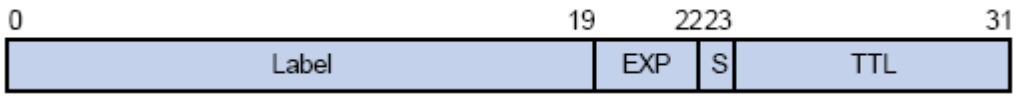

Figure 3.1 Format of a label

A label consists of four fields:

- Label: Label value of 20 bits. Used as the pointer for forwarding.
- Exp: For QoS, three bits in length.
- S: Flag for indicating whether the label is at the bottom of the label stack, one bit in length. 1 indicates that the label is at the bottom of the label stack. This field is very useful when there are multiple levels of MPLS labels.
- TTL: Time to live (TTL) for the label. Eight bits in length. This field has the same meaning as that for an IP packet.

Similar to the VPI/VCI in ATM and the DLCI in frame relay, an MPLS label functions as a connection identifier. If the link layer protocol has a label field like

VPI/VCI in ATM or DLCI in frame relay, the MPLS label is encapsulated in that field. Otherwise, it is inserted between the data link layer header and the network layer header as a shim. As such, an MPLS label can be supported by any link layer protocol. Figure 3.2 shows the place of a label in a packet.

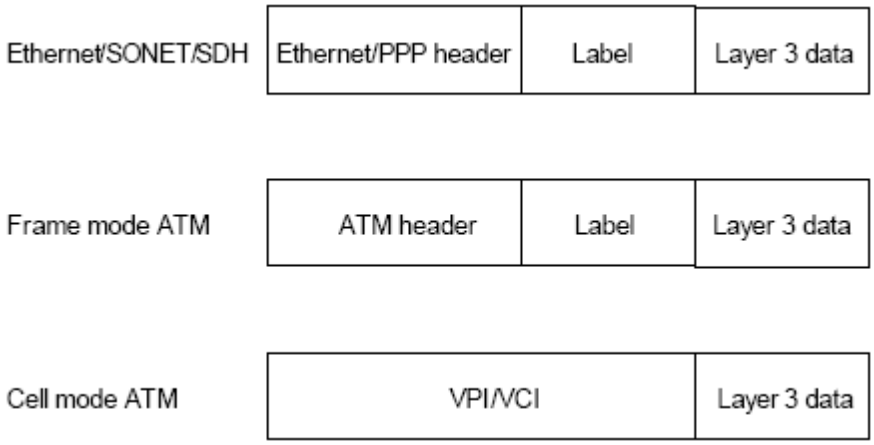

Figure 3.2 Place of a label in a packet

3.1.3. LSRs and LERs

LSR: A Label switching router (LSR) is a high speed router device in the core of an MPLS network that participates in the establishment of LSPs using label signalling protocol and high speed switching of the data traffic based on established paths.

LER: A Label Edge outer (LER) is a device that operates at the edge of the access network and MPLS network. LERs support multiple ports connected to dissimilar network (such as frame relay, ATM, and Ethernet) and forwards this traffic on to the MPLS network after establishing LSPs, using the label siganaling protocol at the egress and distributing the traffic back to the access networks at the egress.

The LER plays a very important role in the assignment of labels, as traffic enters or exits an MPLS network.

3.1.4. LSP

Label switched path (LSP) means the path along which a FEC travels through an MPLS network. Along an LSP, two neighboring LSRs are called upstream LSR and downstream LSR respectively. In Figure 3.3, R2 is the downstream LSR of R1, while R1 is the upstream LSR of R2.

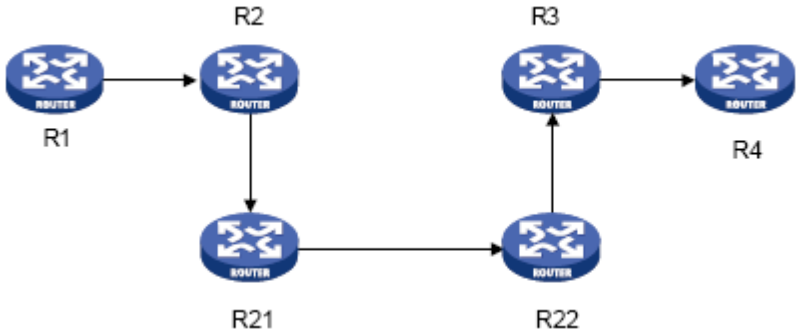

Figure 3.3 Diagram for an LSP

An LSP is a unidirectional path from the ingress of the MPLS network to the egress. It functions like a virtual circuit in ATM or frame relay. Each node of an LSP is an LSR.

3.1.5. LDP

Label Distribution Protocol (LDP) means the protocol used by MPLS for control. An LDP has the same functions as a signaling protocol on a traditional network. It classifies FECs, distributes labels, and establishes and maintains LSPs.

MPLS supports multiple label distribution protocols of either of the following two types:

- Those dedicated for label distribution, such as LDP and Constraint-based Routing using LDP (CR-LDP).
- The existing protocols that are extended to support label distribution, such as Border Gateway Protocol (BGP) and Resource Reservation Protocol (RSVP).

In addition, we can configure static LSPs.

3.1.6. LSP Tunneling

MPLS support LSP tunneling. An LSR of an LSP and its downstream LSR are not necessarily on a path provided by the routing protocol. That is, MPLS allows an LSP to be established between two LSRs that are not on a path established by the routing protocol. In this case, the two LSRs are respectively the start point and end point of the LSP, and the LSP is an LSP tunnel, which does not use the traditional network layer encapsulation tunneling technology. For example, the LSP <R2→R21→R22→R3> in Figure 3.3 is a tunnel between R2 and R3.

If the path that a tunnel traverses is exactly the hop-by-hop route established by the routing protocol, the tunnel is called a hop-by-hop routed tunnel. Otherwise, the tunnel is called an explicitly routed tunnel.

3.1.7. Multi-level label stack

MPLS allows a packet to carry a number of labels organized as a last-in first-out (LIFO) stack, which is called a label stack. A packet with a label stack can travel along more than one level of LSP tunnel. At the ingress and egress of each tunnel, these operations can be performed on the top of a stack: PUSH and POP.

MPLS has no limit to the depth of a label stack. For a label stack with a depth of m, the label at the bottom is of level 1, while the label at the top has a level of m. An unlabeled packet can be considered as a packet with an empty label stack, that is, a label stack whose depth is 0.

3.2 ARCHITECTURE OF MPLS

3.2.1 Structure of the MPLS network

As shown in Figure 3.4, the element of an MPLS network is LSR and LER. LSRs in the same routing or administrative domain form an MPLS domain. In an MPLS domain, LSRs residing at the domain border to connect with other networks are label edge routers (LERs), while those within the MPLS domain are core LSRs.

All core LSRs, which can be routers running MPLS or ATM-LSRs upgraded from ATM switches, use MPLS to communicate, while LERs interact with devices outside the domain that use traditional IP technologies.

Each packet entering an MPLS network is labeled on the ingress LER and then forwarded along an LSP to the egress LER. All the intermediate LSRs are called transit LSRs.

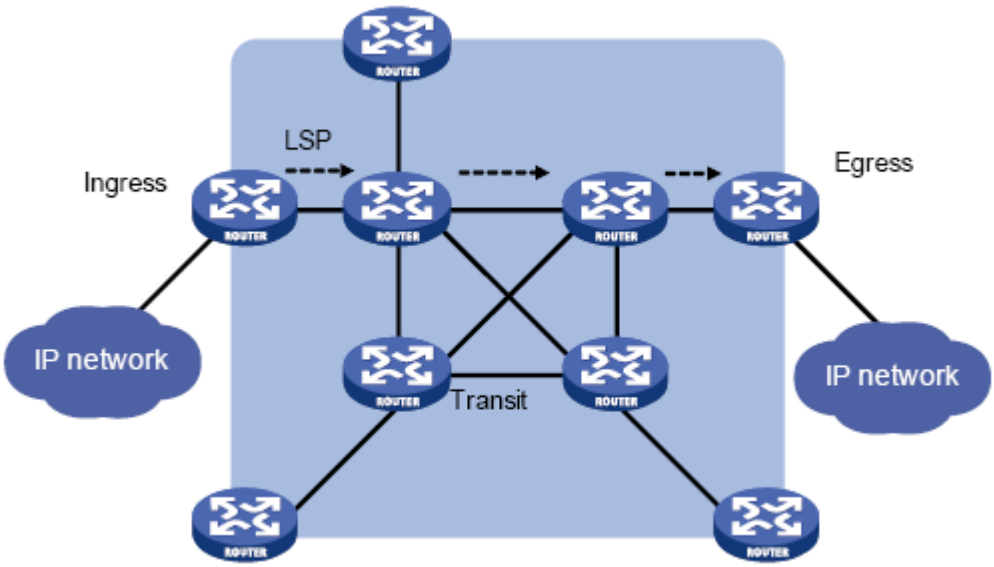

Figure 3.4 Structure of the MPLS network

The following step describes how MPLS operates:

1) First, the LDP protocol and the traditional routing protocol (such as OSPF and ISIS) work together on each LSR to establish the routing table and the label information base (LIB) for intended FECs.

2) Upon receiving a packet, the ingress LER completes the Layer 3 functions, determines the FEC to which the packet belongs, labels the packet, and forwards the labeled packet to the next hop along the LSP.

3) After receiving a packet, each transit LSR looks up its label forwarding table for the next hop according to the label of the packet and forwards the packet to the next hop. None of the transit LSRs performs Layer 3 processing.

When the egress LER receives the packet, it removes the label from the packet and performs IP forwarding.

Obviously, MPLS is not a service or application, but actually a tunneling technology and a routing and switching technology platform combining label switching with Layer 3 routing. This platform supports multiple upper layer protocols and services, as well as secure transmission of information to a certain degree.

As shown in Figure 3.5, an LSR consists of two components:

- Control plane: Implements label distribution and routing, establishes the LFIB, and builds and tears LSPs.
- Forwarding plane: Forwards packets according to the LFIB.

An LER forwards both labeled packets and IP packets on the forwarding plane and therefore uses both the LFIB and the FIB. An ordinary LSR only needs to forward labeled packets and therefore uses only the LFIB.

3.2.2 Structure of an LSR

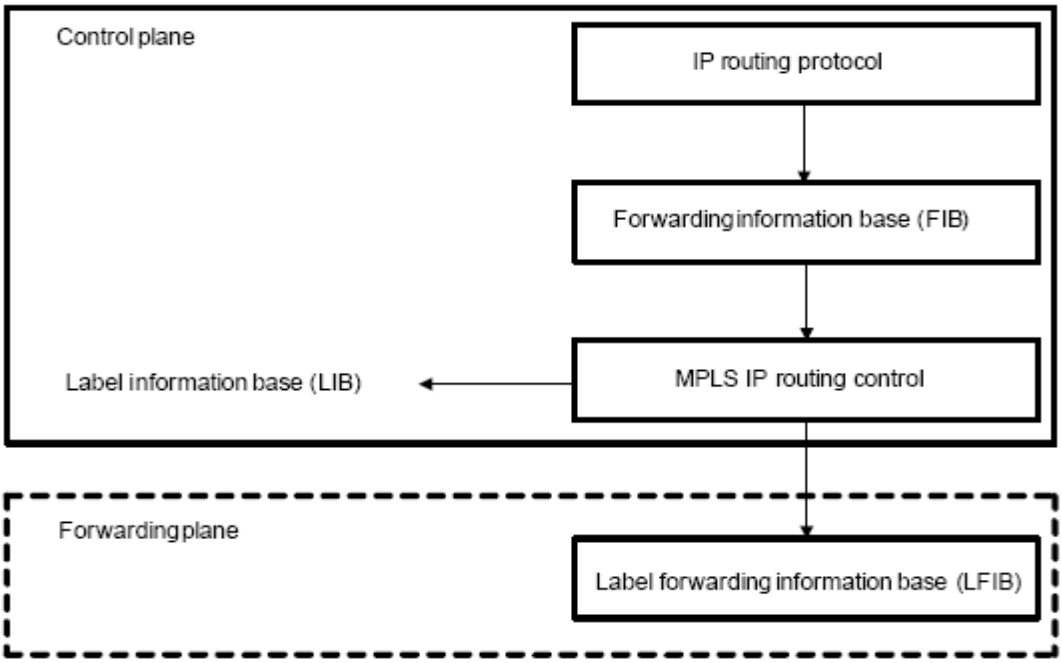

Figure 3.5 Structure of an LSR

3.3 Basic Operation of Label and LDP

3.3.1Label Advertisement and Management

In MPLS, the decision to assign a particular label to a particular FEC is made by the downstream LSR. The downstream LSR informs the upstream LSR of the assignment. That is, labels are advertised in the upstream direction.

3.3.1.1 Label advertisement mode

Two label advertisement modes are available:

- Downstream on demand (DoD): In this mode, a downstream LSR binds a label to a particular FEC and advertises the binding only when it receives a label request from its upstream LSR.
- Downstream unsolicited (DU): In this mode, a downstream LSR does not wait for any label request from an upstream LSR before binding a label to a particular FEC.

An upstream LSR and its downstream LSR must use the same label advertisement mode; otherwise, no LSP can be established normally. For more information, refer to LDP Label Distribution.

3.3.1.2 Label distribution control mode

There are two label distribution control modes:

- **Independent:** In this mode, an LSR can notify label binding messages upstream anytime. The drawback of this mode is that an LSR may have advertised to the upstream LSR the binding of a label to a particular FEC when it receives a binding from its downstream LSR.
- **Ordered:** In this mode, an LSR can send label binding messages about a FEC upstream only when it receives a specific label binding message from the next hop for a FEC or the LSR itself is the egress node of the FEC.

3.3.1.3 Label retention mode

Label retention mode dictates how to process a label to FEC binding that is received by an LSR but not useful at the moment. There are two label retention modes:

- **Liberal:** In this mode, an LSR keeps any received label to FEC binding regardless of whether the binding is from its next hop for the FEC or not.
- **Conservative:** In this mode, an LSR keeps only label to FEC bindings that are from its next hops for the FECs.

In liberal mode, an LSR can adapt to route changes quickly; while in conservative mode, there are less label to FEC bindings for an LSR to advertise and keep.

The conservative label retention mode is usually used together with the DoD mode on LSRs with limited label space.

3.3.1.4 Basic concepts for label switching

- Next hop label forwarding entry (NHLFE): Operation to be performed on the label, which can be Push or Swap.
- FEC to NHLFE map (FTN): Mapping of a FEC to an NHLFE at the ingress node.
- Incoming label map (ILM): Mapping of each incoming label to a set of NHLFEs. The operations performed for each incoming label includes Null and Pop.

3.3.1.5 Label switching process

Each packet is classified into a certain FEC at the ingress LER. Packets of the same FEC travel along the same path in the MPLS domain, that is, the same LSP. For each incoming packet, an LSR examines the label, uses the ILM to map the label to an NHLFE, replaces the old label with a new label, and then forwards the labeled packet to the next hop.

3.3.2 Fundamental Operation of LDP

LDP goes through four phases in operation: discovery, session establishment and maintenance, LSP establishment and maintenance, and session termination.

3.3.2.1 Discovery

In this phase, an LSR who wants to establish a session sends Hello messages to its neighbouring LSRs periodically, announcing its presence. This way, LSRs can automatically find their peers without manual configuration.

LDP provides two discovery mechanisms:

- Basic discovery mechanism

The basic discovery mechanism is used to discover local LDP peers, that is, LSRs directly connected at link layer, and to further establish local LDP sessions.

Using this mechanism, an LSR periodically sends LDP link Hellos as UDP packets out an interface to the multicast address known as “all routers on this subnet”. An LDP link Hello message carries information about the LDP identifier of a given interface and some other information. Receipt of an LDP link Hello message on an interface indicates that a potential LDP peer is connected to the interface at link layer.

- Extended discovery mechanism

The extended discovery mechanism is used to discover remote LDP peers, that is, LSRs not directly connected at link layer, and to further establish remote LDP sessions.

Using this mechanism, an LSR periodically sends LDP targeted Hellos as UDP packets to a given IP address.

An LDP targeted Hello message carries information about the LDP identifier of a given LSR and some other information. Receipt of an LDP targeted Hello message on an LSR indicates that a potential LDP peer is connected to the LSR at network layer.

At the end of the discovery phase, Hello adjacency is established between LSRs, and LDP is ready to initiate session establishment.

3.3.2.2 Session establishment and maintenance

In this phase, LSRs pass through two steps to establish sessions between them:

1) Establishing transport layer connections (that is, TCP connections) between them.

2) Initializing sessions and negotiating session parameters such as the LDP version, label distribution mode, timers, and label spaces.

After establishing sessions between them, LSRs send Hello messages and Keepalive messages to maintain those sessions.

3.3.2.3 LSP establishment and maintenance

Establishing an LSP is to bind FECs with labels and notify adjacent LSRs of the bindings. This is implemented by LDP. The following takes DoD mode as an example to illustrate the primary steps:

1) When the network topology changes and an LER finds in its routing table a new destination address that does not belong to any existing FEC, the LER creates a new FEC for the destination address and determine the route for the FEC to use. Then, the LER creates a label request message that contains the FEC requiring a label and sends the message to its downstream LSR.

2) Upon receiving the label request message, the downstream LSR records this request message, finds in its routing table the next hop for the FEC, and sends the label request message to its own downstream LSR.

3) When the label request message reaches the destination node or the egress of the MPLS network, if the node has any spare label, it validates the label request message and assigns a label to the FEC. Then, the node creates a label mapping message containing the assigned label and sends the message to its upstream LSR.

4) Upon receiving the label mapping message, an LSR checks the status of the corresponding label request message that is locally maintained. If it has information about the request message, the LSR assigns a label to the FEC, and adds an entry in its LFIB for the binding, and sends the label mapping message on to its upstream LSR.

5) When the ingress LER receives the label mapping message, it also adds an entry in its LFIB. Up to this point, the LSP is established, and packets of the FEC can be label switched along the LSP.

3.3.2.4 Session termination

LDP checks Hello messages to determine adjacency and checks Keepalive messages to determine the integrity of sessions.

LDP uses different timers for adjacency and session maintenance:

- **Hello timer:** LDP peers periodically send Hello messages to indicate that they intend to keep the Hello adjacency. If the timer expires but an LSR still does not receive any new Hello message from its peer, it removes the Hello adjacency.
- **Keepalive timer:** LDP peers keep LDP sessions by periodically sending Keepalive message over LDP session connections. If the timer expires but an LSR still does not receive any new Keepalive message, it closes the connection and terminates the LDP session.

Chapter 4

MPLS OVER WDM NETWORKS

4.1 NETWORK MODEL

The network model addressed in this thesis is a two layer model: MPLS over WDM. The lower layer – optical transport layer applying WDM – consists of nodes represented by optical cross-connects (OXCs) that perform wavelength routing operations and optical links - fibers. The upper layer – MPLS layer – includes nodes represented by MPLS routers, namely label switching routers. A set of lightpaths (wavelengths) provisioned by WDM layer forms a logical topology for the MPLS routers. i.e. lightpaths represent in MPLS layer.

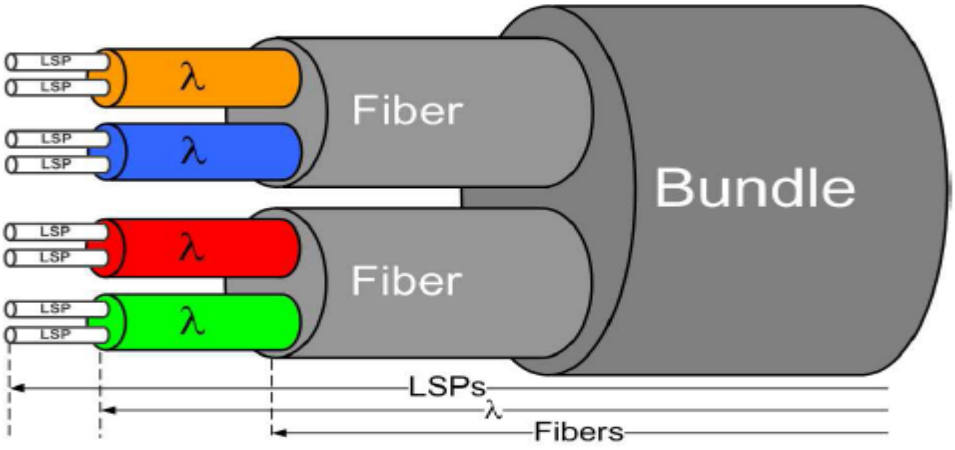

Figure 4.1. Model of optical link carrying MPLS traffic

Figure. 4.1 shows a model of a physical arc, which is a bundle of fibers. Each fiber can support a fixed number of wavelengths.

4.2 MPLS over WDM architecture

Based on the relationship of control planes in the two layers, the architecture of MPLS over optical network is generally classified into three inter connection

model namely the overlay model, the augmented model, and the peer model. In this thesis, we focus on the overlay model, which is the most practical model.

4.3 Overlay Model

In the overlay model, each LSR keeps only MPLS layer information, such as residual capacity on all of the existing logical links and the number of unused ports in the LSRs. The MPLS layer only receives a response of whether a requested lightpath can be set up or not, from the WDM layer.

Therefore, in the overlay model, a network has to decide whether it should use the existing logical links or open new lightpath(s) for a new arriving request. If it chooses to use the existing logical topology, then how to route the request over the existing logical topology has to be decided. If the network would open new lightpath(s), it has to decide the logical edge(s) (LSR pair(s)) on which to open the lightpaths, without any network information from the WDM layer.

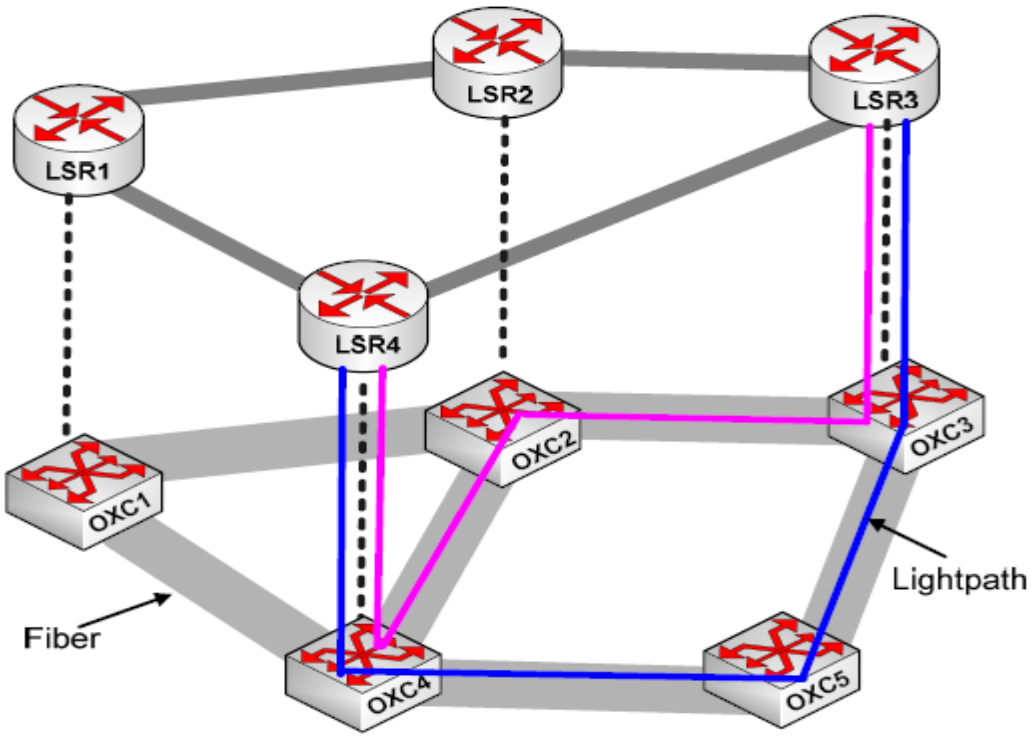

Figure. 4.2. MPLS over WDM architecture

In the overlay model for the establishment of new LSP either the MPLS layer or the optical layer can be selected. Figure. 4.2 shows a simple example to illustrate MPLS over WDM architecture. For the establishment of a new request of LSP the logical link between LSR3 and LSR4 either the existing virtual path LSR1-LSR2 can be used or any of the new two lightpaths (wavelengths) can be used. However, these two lightpaths are routed in two various paths in the WDM layer: OXC3-OXC5-OXC4 and OXC3-OXC2-OXC4.

4.4 Recovery Strategies

Many failure *recovery* mechanisms have been proposed for generic networks in the literature. The application of these mechanisms is not restricted to optical networks alone. Classifications have been done based on the nature of route computation, (i.e. centralized or distributed), by the layer where they take place (WDM, MPLS, IP…), by the type of protection (link-based or path-based), and by the computation timing (precomputed or real time).

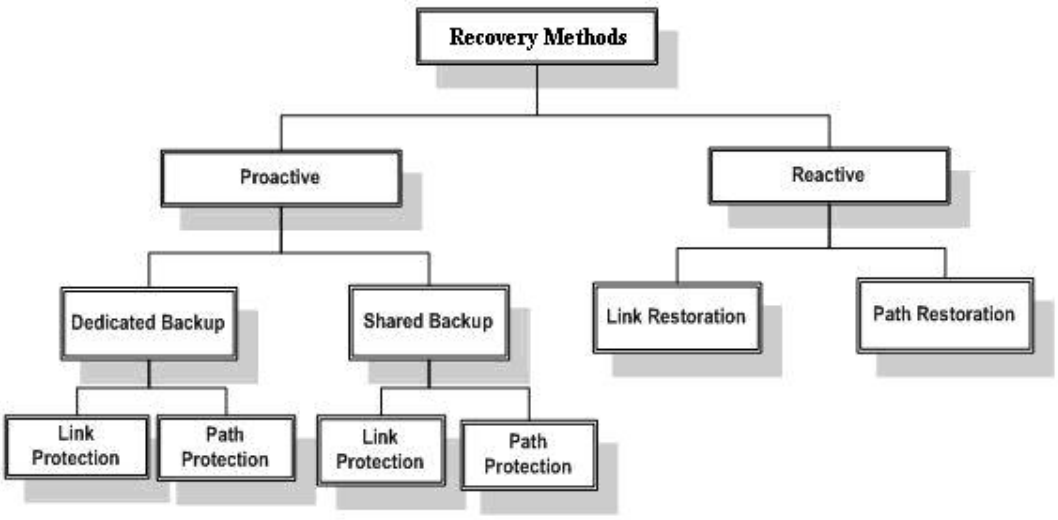

Figure 4.3 Classification of Restoration Methods

The two main categories for classifying the recovery schemes at the broadest level, are *proactive* and *reactive* techniques. The proactive or protection techniques allocate and reserve the backup resources in advance, thus, providing fast recovery on preplanned paths at the expense of an inefficient use of resources. The reactive or *restoration* methods can be classified as illustrated in Figure 4.3.

The restoration techniques make use of real time availability of resources. They provide a slower recovery but they do not reserve the resources for backup paths. The latency of restoration schemes will thus be higher than that of protective schemes.

In the reactive approach, when a failure occurs, a search for an alternate path is initiated. In the absence of failures or with a few failures, the overhead of the reactive approach is low. However, this approach may not be successful if there are no resources present at the time of actual recovery.

In case the recovery is computed in a distributed fashion contention may occur to win over the resources that are being needed to recover from some other failure simultaneously. This will result in several retries to recover. In the proactive approach, the backup paths are computed and the resources are reserved at the time of establishing the primary session. This method reserves resources, is faster, and always guarantees restoration.

The proactive or reactive schemes can be either *link based* or *path based*. When a component fails, *link based methods* select an alternate path between the end nodes of the failed link. This alternate path along with the intact part of the primary path is used for the recovery. This method is illustrated in Figure 4.4, which shows a primary lightpath, p1, and two backup lightpaths, b11 and b12, on a wavelength. When link 0−1 fails, backup path b11 is used.

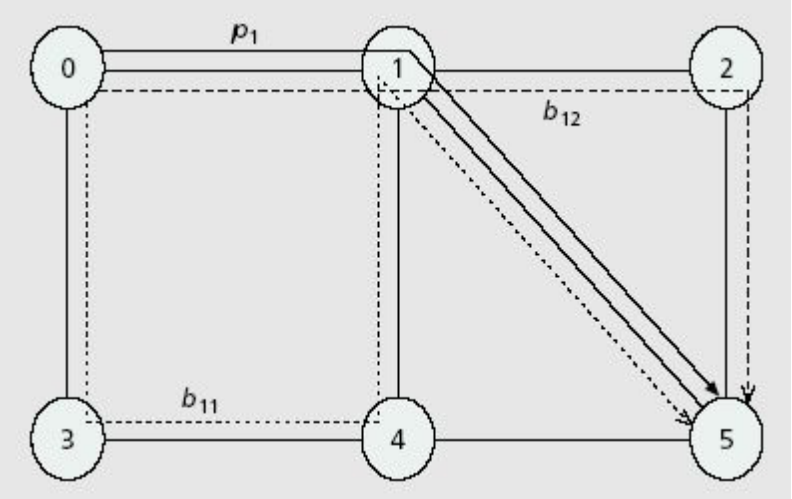

Figure 4.4 Link-Based Backup Path

When link 1−5 fails backup path b12 is used. It can be observed that b12 is routed around link 1−5 while retaining the working segment of p1. Note that the working segment of the primary lightpath is retained in the backup path.

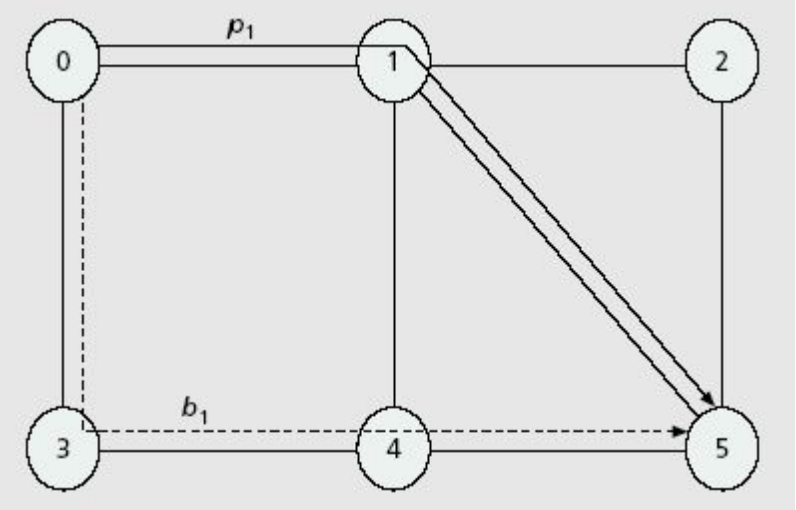

Figure 4.5 Path-Based Backup Path

In the case of *path based methods*, a backup path is computed between the end nodes of the failed primary lightpath. The backup path can use any wavelength independent of the one used by the corresponding primary lightpath. The path-based restoration method is illustrated in Figure 4.5. Figure 4.5 shows a primary lightpath, p1, and its backup lightpath, b1, on a given wavelength. Note that b1 is

established between the end nodes of p1, and the working segment of p1 is not utilized by b1. If none of the channels are shared between any two backup channels, then the method is referred to as *dedicated backup restoration*. This method ends up reserving a lot of resources and is not resource efficient. This method is illustrated in Figure 4.6. The Figure shows two primary lightpaths, p1 and p2, and their respective backup lightpaths b1 and b2. It can be observed that b1 and b2 do not share any wavelength channel.

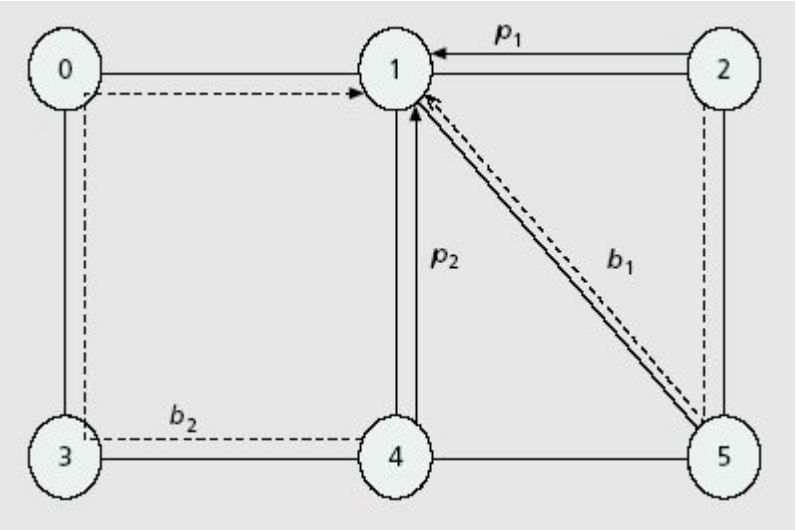

Figure 4.6 Dedicated Backup Path Reservation

If no two failures can occur at the same time then their backup channels can share channels. This is referred to as *shared backup restoration* and is illustrated in the Figure 4.7. The Figure shows two primary lightpaths p1 and p2 and their respective backup lightpaths b1 and b2 on a certain wavelength. As p1 and p2 are disjoint, they do not fail at the same time under the single link failure fault model. Therefore, b1 and b2 can share the wavelength on link 5−1. This shared channel will be used by b1 when link 2 − 1 fails and by b2 when link 5−1 fails. It is to be noted that "*segment*" based recovery is also possible. In this case backup is provided at the segment level rather than the link or path level, where a segment is a subpath.

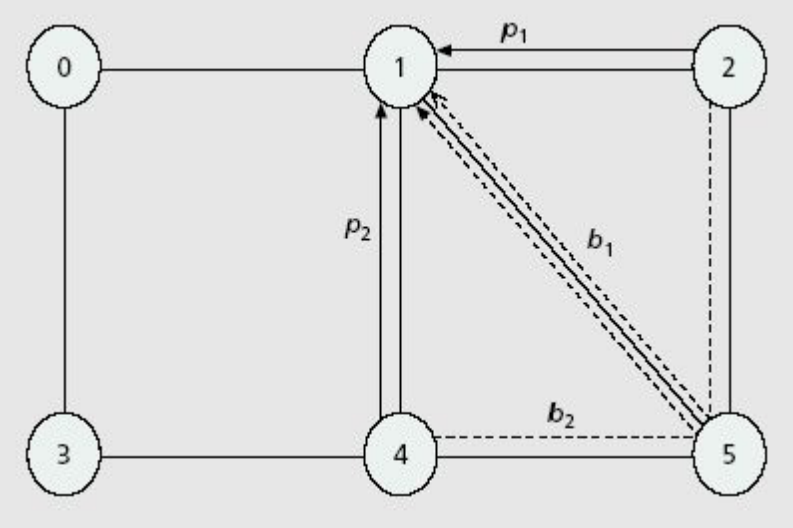

Figure 4.7 Shared Backup Path Restoration

4.5 Problem Definition

Increasing the efficiency of Internet resources utilization is very important. Multicasting in MPLS is a useful and effective operation for doing so. Multicasting in MPLS has problems in label distribution. In unicast, a destination IP address includes two parts, the subnet address and the host address. So, the destinations which have the same subnet address belong to the same subnet, which can be represented by a single record of the routing table.

In MPLS networks, for each routing table record, we map it to a FEC and distributes it a label, set up a LSP for it. Thus every subnet destinations only needs a label. But in multicast networks, destination address doesn't possess any information about the subnet of destination, so we cannot tell whether the destinations belong to a same subnet or not. And we cannot discern whether the different destinations belong to the same multicast group.

So for different FECs, carrying different forwarding information, we set up different LSPs and distribute different labels. This limits the number of multicast sessions that can be set up in MPLS networks. This problem becomes more apparent in large scale networks. And it also limits the scalability of multicast in MPLS networks. This problem is very common in the now existing MPLS multicast mechanisms.

Chapter 5

MULTICASTING IN MPLS NETWORKS

The key for MPLS multicast lays in label distribution mechanisms. In general, MPLS unicast is request driven. But multicast permits sources and receivers to join and leave a multicast session and the corresponding multicast tree dynamically. The multicast tree set up by multicast routing algorithm is also not permanent but time-limited and easily changed. So, the request driven model can bring cost for sending signals and is time-consuming if being directly transplanted into multicast networks. So, the approach for setting up LSP that works well in unicast networks is not appropriate for multicast networks.

To solve this problem, several solutions have been proposed. All those solutions use the dataflow-driven distribution. These dataflow-driven mechanisms can be classified into two sorts: first, upstream-node-based label distribution, i.e. the root driven and leaves driven distribution as proposed by IETF. And second, downstream-node-based label distribution. The optimizing work of this thesis is based on the latter.

5.1 Multicasting concept

The process of multicasting can be described as follows:

- When a LSR receives a packet, it consults the LIB according to the incoming port and incoming label, trying to find out the matching record, if success, then encapsulates the packet with the found outgoing label then forwards the packet through the found outgoing port.

- If LSR cannot find the corresponding record in LIB, it means that there is no label to match the Forwarding Equivalence Class (FEC) and so a new label should be created. It then starts the IP layer routing algorithm to find

should-be outgoing port and then forwards it. Meanwhile, LSR updates LIB with the record being comprised by FEC, incoming label, outgoing label, incoming port, and outgoing port.

- The downstream LSR receives the FEC, because it is a new one and no label has been bound to it. So it repeats the process done by the previous LSR.

- After the first IP routing and label binding, and by declaring the binding information through Label Distribution Protocol (LDP), the establishment of multicast LSP is completed. The following packets will be transmitted directly through the LSP.

An Example

Consider a MPLS network routing domain shown in Figure 5.1. LER and LSR are the routers of the MPLS backbone network. Domain A has a service provider, and domain B, C, and D and E are customer networks.

Suppose there are four multicast sessions in this network. The sources of sessions are all from the service provider (referred to as S) in domain A.

Session 1: Multicast address is G1. Group members include B1, B2 in domain B, C1, C4 in domain C and D1, D3 in domain D. Then its multicast tree is (A-LER A-LSR1- LSR2-LER B-B-LSR3-LSR4-LER C-C-LSR 3-LSR 5-LER D-D.

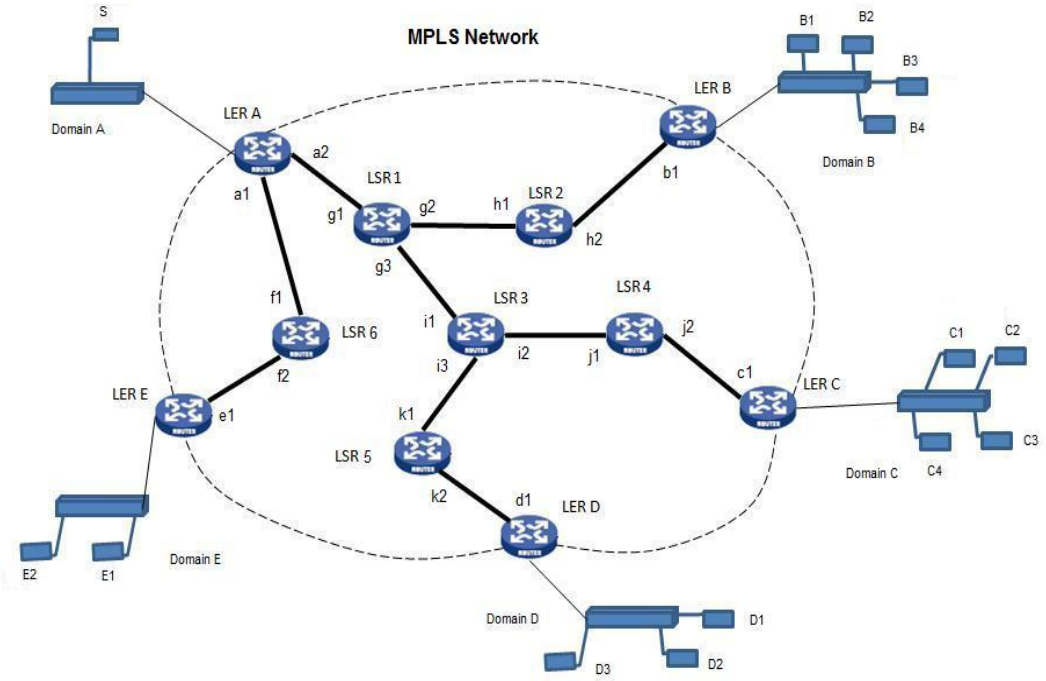

Figure 5.1 MPLS Network Model

Session 2: Multicast address is G2. Group members include B3, B4 in domain B, C1, C3 in domain C and D2 in domain D. Then its multicast tree is (A-LER A-LSR1- LSR2-LER B-B-LSR3-LSR4-LER C-C-LSR 3-LSR 5-LER D-D).

Session 3: Multicast address is G3. Its group members are C4 in domain C and E1 in domain E. Then its multicast tree is (A-LERA-LSR1-LSR3-LSR4-LER-C-LSR3-LSR6-LER E-E).

Session 4: Multicast address is G4. Its group members are C2, C4 in domain C and D1, D2 in domain D. Then its multicast tree is (A-LER A-LSR1-LSR3-LSR4-LER C-C-LSR 3-LSR 5-LER D-D).

The following Figure 5.2 shows the label switching table formed by a non-label-aggregation MPLS multicast mechanism.

LER A		
FEC	O/Int	O/Label
(S,G1)	a2	10
(S,G2)	a2	11
(S,G3)	a2	12
(S,G3)	a1	12
(S,G4)	a2	13

LSR 1			
I/Int	I/Label	O/Int	O/Label
g1	10	g2	14
g1	10	g3	14
g1	11	g2	15
g1	11	g3	15
g1	12	g3	16
g1	13	g3	17

LSR 2			
I/Int	I/Label	O/Int	O/Label
h2	14	g2	16
h2	15	g3	17

LSR 4			
I/Int	I/Label	O/Int	O/Label
j1	18	j2	22
j1	19	j2	23
j1	20	j2	24
j1	21	j2	25

LSR 6			
I/Int	I/Label	O/Int	O/Label
f1	12	k2	13

LSR 3			
I/Int	I/Label	O/Int	O/Label
i1	14	i2	18
i1	14	i3	18
i1	15	i2	19
i1	15	i3	19
i1	16	i2	20
i1	17	i2	21
i1	17	i3	21

LSR 5			
I/Int	I/Label	O/Int	O/Label
k1	18	k2	22
k1	19	k2	23
k1	21	k2	24

Figure 5.2. Label switching table without label aggregation

As illustrated in Figure.2, every multicast session means a different FEC: (S, G1), (S, G2), (S, G3), (S, G4). MPLS distributes a different label to each FEC. Thus LER1 needs four labels. But as showed in this example, session 1 and session 2 share one label. So the mechanism without label aggregation has a waste of labels. To solve this problem, we use a mechanism that adopts label aggregation and can save a larger number of labels.

5.2 Label aggregation concept

Aggregated multicast is targeted as an interdomain multicast provisioning mechanism in the transport networks. The idea of aggregated multicast is that, instead of constructing a tree for each individual multicast session in the backbone network, one can have multiple multicast sessions share a single tree to reduce multicast state and tree maintenance overhead at the network core.

5.2.1 Multicasting Mechanism with Label Aggregation

In the mechanism of label aggregation, every ingress LER must keep all the nodes of multicast trees of all multicast sessions. The multicast tree can be set up by the multicast routing algorithm. A mapping table which is to be established according to multicast trees is necessary for each LER. All the nodes of every multicast tree are included in that table.

Every table record consists of (Length, Multicast Tree nodes, Label Assigned, O/Int). O/Int represents the outgoing port for forwarding. Length equals to the number of nodes in a multicast tree. The following table shows the structure of the mapping table in our example.

Multicast Tree Table

Length	Multicast Tree Nodes	Label Assigned	O/Int
10	LER-A, LSR1, LSR2, LER-B, LSR3, LSR4, LER-C, LSR3, LSR 5, LER-D	10	a2
8	LER-A, LSR1, LSR3, LSR4, LER-C, LSR3, LSR6, LER-E	11	a1,a2
8	LER-A, LSR1, LSR3, LSR4, LER-C, LSR 3, LSR 5, LER-D	12	a2

5.2.2 Process of label aggregation

The process of label aggregation can be summarised as:

- When a packet arrives, first check its IP, and find the FEC it belongs to. According to the FEC, search the label switching table, if find the matching record, then encapsulate the packet with the matching label and forward it to next hop.
- If fail to find the matching record in table, then find the corresponding multicast tree of the FEC in the route table, and then search the tree node table, to find the matching record. If there is one record of the same tree

nodes and of the same order of tree nodes then encapsulate the packet with the label of the found record then forward it to next hop. And meanwhile update the label switching table with the new record (FEC, label, O/Int), and finish.

- If no matching record in the tree node table, distribute a new label for this packet, establish an LSP, the process of this step is the same as the mechanism without label aggregation. Create a label randomly and distribute to the packet, update the label switching table with record (FEC, Label, O/Int), and update the tree node table, then encapsulate the packet with the label, and forward it to next hop.

5.2.3 Algorithm: Label Aggregation

Definitions: LABEL_TABEL keeps the record of label assigned to the FEC. TREE_TABLE keeps the record of multicast tree running.

begin

Initialize with empty LABEL_TABLE and TREE_TABLE

while (Packet arrived)

Find a group address (Find FEC)

Search LABEL_TABLE

if (Match found)

Encapsulate the packet with label mapping the FEC

(go to packet transmit)

if(No match found)

Find multicast *routing table*

Search multicast TREE_TABLE

if (match found with same length and same nodes)

Encapsulate the packet with label mapping the entry

Update the LABEL_TABLE.

(go to packet transmit)

if(No match found)

Encapsulate the packet with new label.

Establish an LSP

Update the LABEL_TABLE and TREE_TABLE.

end if

end if

Transmit the packet with the applied label.

end while

end begin.

Example

Consider the MPLS network of Figure 5.1. The following Figure 5.3 shows the label switching table of each router of our example. We can see that the sessions of different FECs but of the same multicast tree have the same label. In the example, LER-A distribute the same label 10 to FEC: (S, G1) and FEC: (S G2). So a smaller number of labels are needed.

LER A		
FEC	O/Int	O/Label
(S,G1)	a2	10
(S,G2)	a2	10
(S,G3)	a2	11
(S,G3)	a1	11
(S,G4)	a2	12

LSR 1			
I/Int	I/Label	O/Int	O/Label
g1	10	g2	13
g1	10	g3	13
g1	11	g3	14
g1	12	g3	15

LSR 2			
I/Int	I/Label	O/Int	O/Label
h2	13	g2	14

LSR 4			
I/Int	I/Label	O/Int	O/Label
j1	16	j2	19
j1	17	j2	20
j1	18	j2	21

LSR 6			
I/Int	I/Label	O/Int	O/Label
f1	11	k2	12

LSR 3			
I/Int	I/Label	O/Int	O/Label
i1	13	i2	16
i1	13	i3	16
i1	14	i2	17
i1	15	i2	18
i1	15	i3	18

LSR 5			
I/Int	I/Label	O/Int	O/Label
k1	16	k2	19
k1	18	k2	20

Figure 5.3 Label switching table with label aggregation

5.3 Multicasting Mechanism with Improved Label Aggregation

In the mechanism of improved label aggregation, we are using the concept of *leaky match tree*. A match is called leaky if all the destination nodes in requested multicast group are found in multicast tree and the number of destination node is less than that of multicast destination nodes.

The disadvantage in using *leaky match tree* concept is that a certain amount of bandwidth would be waste to deliver data to nodes that are not involved for the group. Hence there is a trade-off between label aggregation and bandwidth usage.

We propose a concept of *utilisation threshold*. Utilisation threshold is a measure of bandwidth waste while using the leaky match for label aggregation. For the selection of leaky match tree the value of matching parameter (M_para) must be greater than the utilisation threshold.

Value of utilisation threshold and matching parameter can be calculated as follows:

5.3.1 Calculation of matching parameter (M_para)

If number of LERs involved in requested multicast group is p, and the number of LERs involved in selected multicast tree q, then

matching parameter (M_para) = p/q.

5.3.2 Calculation of utilization threshold (Uth)

If band width usage is greater than 30%,

Then $k = \infty$,

Otherwise $k = 1$.

Utilization threshold (Uth) = $k \times$Mth.

Where Mth is matching threshold and is the measure of perfectness of the matching of requested multicast in comparison to running multicast sessions.

5.3.3 Process of Improved Label Aggregation

The process of label aggregation with leaky match can be described as follows:

- When a packet arrives, first check its IP, and find the FEC it belongs to. According to the FEC, search the label switching table and tree node table, to find the matching record.
- If no matching record found in the tree node table, search for leaky match. If there is one record, which also satisfied the utilisation threshold condition, then encapsulate the packet with the label of the found record then forward it to next hop. And meanwhile update the label switching table with the new record (FEC, label, O/Int), and finish.
- If no matching record in the tree node table, distribute a new label for this packet, establish an LSP, the process of this step is the same as the mechanism without label aggregation.

5.3.4 Algorithm: Improved Label Aggregation

Definitions: LABEL_TABEL keeps the record of label assigned to the FEC. TREE_TABLE keeps the record of multicast tree running. NET_MAT is the network matrix with cost as the matrix elements. MC_GROUP is the set of destination nodes in the current multicast session.

begin

Initialize with empty LABEL_TABLE and TREE_TABLE

while (Packet arrived)

Find a group address (Find FEC)

Search LABEL_TABLE

if (Match found)

Encapsulate the packet with label mapping the FEC

(go to packet transmit)

if(No match found)

Find multicast *routing table*

Search multicast TREE_TABLE

if (match found with same length and same nodes)

Encapsulate the packet with label mapping the entry

Update the LABEL_TABLE.

(go to packet transmit)

if(No match found)

Compare the bandwidth threshold

Search the leaky multicast tree.

if (success and M_para>Uth)

Select the tree with largest value of M_para.

Encapsulate the packet with label mapping the entry. Update the LABEL_TABLE.

(go to packet transmit)

if(condition not matched)

Encapsulate the packet with new label.

Establish an LSP

Update the LABEL_TABLE and TREE_TABLE.

end if

end if

end if

Transmit the packet with the applied label.

end while

end begin

Consider the MPLS network of Figure 5.1. The following Figure 5.4 shows the label switching table of each router of our example. We can see that the sessions of different FECs but of the same multicast tree and leaky match tree have the same label. In the example, LER-A distribute the same label 10 to FEC: (S, G1), FEC: (S G2) and FEC(S, G4). So a smaller number of labels are needed.

LER A		
FEC	O/Int	O/Label
(S,G1)	a2	10
(S,G2)	a2	10
(S,G3)	a2	11
(S,G3)	a1	11
(S,G4)	a2	10

LSR 1			
I/Int	I/Label	O/Int	O/Label
g1	10	g2	12
g1	10	g3	12
g1	11	g3	13

LSR 2			
I/Int	I/Label	O/Int	O/Label
h2	12	g2	13

LSR 4			
I/Int	I/Label	O/Int	O/Label
j1	16	j2	18
j1	17	j2	19

LSR 6			
I/Int	I/Label	O/Int	O/Label
f1	11	k2	12

LSR 3			
I/Int	I/Label	O/Int	O/Label
i1	12	i2	16
i1	12	i3	16
i1	13	i2	17

LSR 5			
I/Int	I/Label	O/Int	O/Label
k1	16	k2	17

Figure 5.4 Label switching table with improved label aggregation

5.4 Mathematical analysis of average number of label

Suppose there is one MPLS network that has *n* egress LERs.

So the number of possible multicast tree $u = 2^n-1$.

The probability for each multicast tree = *1/u*.

If the number of multicast sessions is *s*, there are *min (u, s)* possible cases. In case k, the algorithm establishes $k(1 \le k \le \min(u, s))$ different multicast trees. So *k* labels are needed in this case. Suppose the probability for case k is *pr (k)*.

Where *pr (k)* is the probability of the following case: select *k* $(1 \le k \le \min(u, s))$ different trees from the total *u* trees. And distribute these *k* individual trees to *s* sessions; allow different sessions to have the same tree. But each of the *k* individual trees must appear at least once.

pr (k) can be calculated as follows:

Probable cases of distribute *u* trees to *s* multicast sessions, allow different sessions have the same tree, is u^s.

Probable cases of selection of *k* individual trees from *u* trees, = C_u^k.

Suppose the number of the probable cases to distribute the selected *k* trees to *s* sessions, different sessions may have the same tree but each of the *k* individual trees must be distributed to at least one session is x_k, then:

$$x_t = t^s - \sum_{i=1}^{t-1} C_t^i x_i \text{ for } (2 \le t \le k), x_1 = 1 \quad (1)$$

$$pr(k) = \frac{C_u^k}{u^s} \times x_k \quad (2)$$

So the average number of labels needed in our mechanism is:

$$label = \sum_{k=1}^{\min(u,s)} pr(k) \times k \quad (3)$$

As we know that the probability of case k to be occur, for lower value of k, in case of improved label aggregation is higher than that of label aggregation. Hence it is clear from the equation (3) that the average numbers of label required are less in case of improved label aggregation.

Chapter 6

Results and Discussions

This chapter evaluates the behaviour of each of the algorithms in previous Chapter. For that purpose we have implemented a simulator on MATLAB. Simulation are carried with n numbers of egress node and Mth (0.5<Mth<1)

6.1 Performance Metrics

The following parameters are utilized in analysing the performance of various schemes.

a. **Average number of Label**: This metrics gives fair idea about the scalability. It indicates the average number of label that will be used with increasing the number of multicast session.

b. **Average bandwidth used**: This metric gives an idea of usage of network. It indicates the average Bandwidth that will be used while increasing the number of multicast session.

The above parameters are measured versus the following variables for different

a. **Multicast session**: Multicast session request is increased from 1 to approximately double the number of total multicast tree possible. Membership size of multicast session is taken randomly.

b. **Threshold**: We will vary the threshold value from 0.5 to 1, while the number of multicast session is constant.

The parameter average number of label will be measured against Multicast session only. The parameter average bandwidth used will be measured against Multicast session, while for improved label aggregation scheme it also measured against the utilization threshold value. All performance parameter evaluated for different size of the MPLS network by considering different number of egress node.

Performance metric are evaluated as follows:

a. Set the value of utilization threshold in-between 0.5 and 1, for improved label aggregation and 1 for label aggregation.

b. Increase the number of multicast session.

6.2 Results

Figure 6.1 and 6.2 shows as the number of multicast session increases average number of label increases. Result shows that average number of label required in case label aggregation is lower than without label aggregation. Also there is a significant improvement in Label assignment as the number of request increases. While average number of label required in improved label aggregation is lower than the label aggregation mechanism.

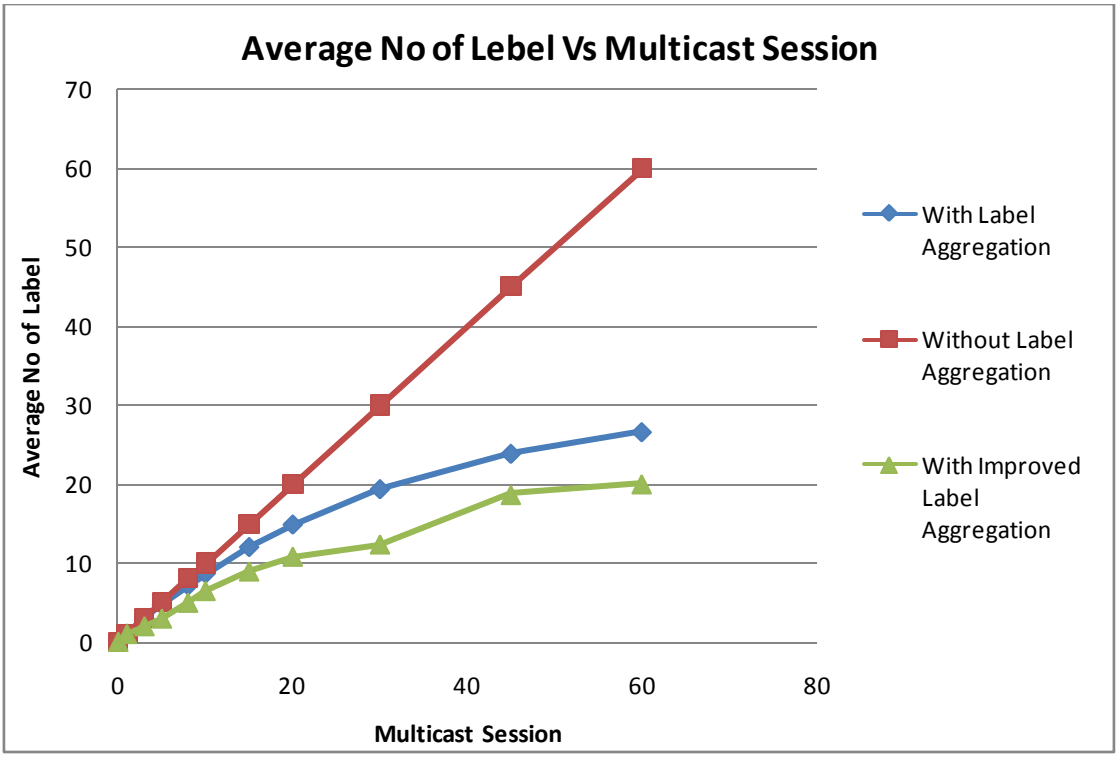

Figure 6.1 Average Number of Label Vs Multicast Session for n=5

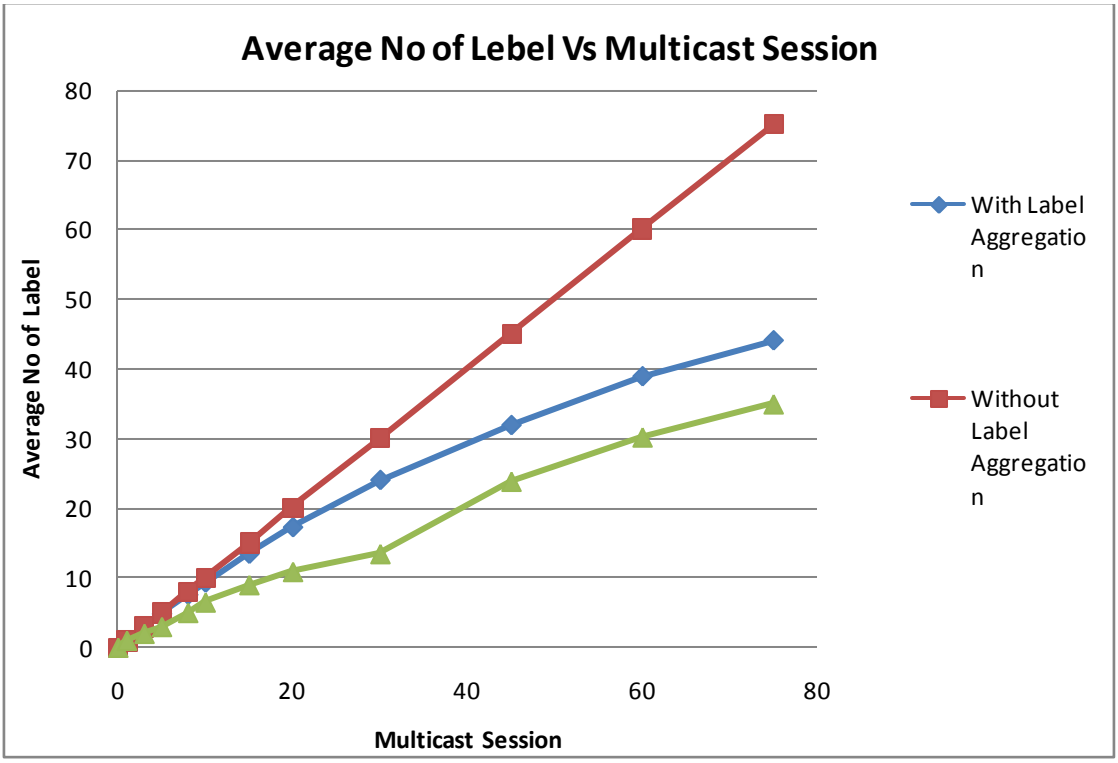

Figure 6.2 Average Number of Label Vs Multicast Session for n=7

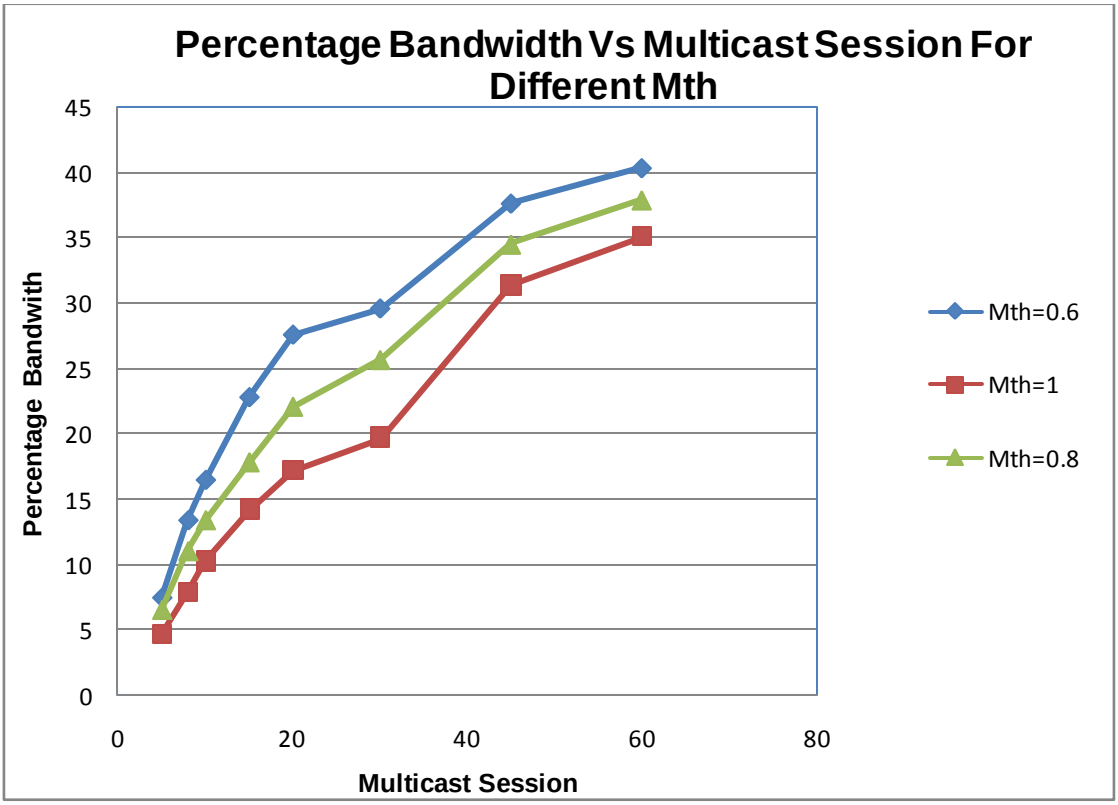

Figure 6.3 Percentage Bandwidth Vs Multicast Session For Different Mth (n=5)

Figure 6.3 and 6.4 shows that Percentage bandwidth used increases with decrease in the value of math threshold (Mth). Percentage bandwidth consumption is lower in case of simple label aggregation mechanism (Corresponds to Mth=1). Also the difference is higher for the multicast session request between 25%-60%. This effect is due to the bandwidth utilisation constraint imposed on Improved Label Aggregation mechanism.

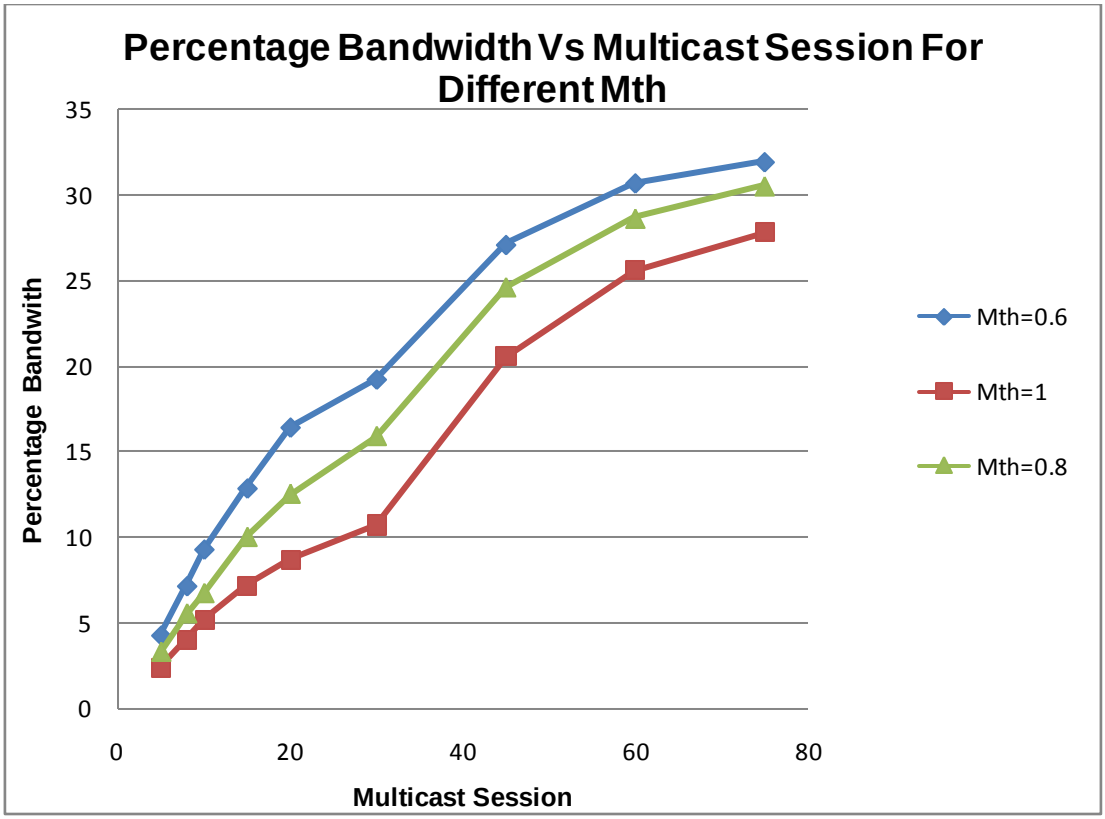

Figure 6.4 Percentage Bandwidth Vs Multicast Session For Different Mth (n=7)

Comparison of all the performance parameter shows that there is improvement in scalability with a slight increase in bandwidth consumption while using the improved label aggregation mechanism.

Chapter 7

CONCLUSION AND FUTURE WORKS

CONCLUSION

Label aggregation mechanism provides the scalability and hence enhances the overall performance of the network. For large scale network improved label aggregation mechanism gives the overall better performance in terms of scalability and hence optimises the flow across the MPLS network by reducing the average number of labels required. There is a trade-off between bandwidth consumption and average number of label in case of improved label aggregation mechanism.

FUTURE WORKS

Future research could aims at introducing congestion control mechanism and improving the load balance of the whole network. We have used bandwidth consumption as a variable for threshold calculation. Other network parameter e.g. link stress, delay etc. could be use for better support MPLS multicast and improve the scalability of multicast in MPLS networks.

REFERENCES

[1]. Huang Weili, Guo Hongyan, "*A mechanism of Label Aggregation for Multicast in MPLS Networks*", Computer science and information technology, 2008 ICCSIT apos.08, international conference, 29 Aug,2008, pp:504-508.

[2] E. Rosen, A. Viswanathan, and R. Callon, "*Multiprotocol label switching architecture*," *IETF RFC3031*, January 2001.

[3]. D. Thaler and M. Handley, "*On the Aggregatability of multicast forwarding state*", Proceedings of IEEE INFOCOM, March 2000.

[4]. N. K. Singhal and B. Mukherjee, "*Protecting Multicast Sessions in WDM Optical Mesh Networks*", Journal of Lightwave Technology, Vol. 21, No. 4, April 2003.

[5]. A. Matrawy, C.-H. Lung, and I. Lambadaris, "*A Framework for MPLS Path Setup in Uni-Directional Multicast Shared Trees*," In Proc. of SPIE Optics East, (in the Conference on Performance, QoS, and Control of Next-Generation Communication Networks), 2004.

[6]. Puype B., Groebbens A., De Maesschalck S., Colle D., Lievens I., Pickavet M., Demeester P., "*Benefits of GMPLS for multilayer recovery"*, IEEE Communications Magazine, Vol. 43, No. 7, 2005, pp. 51-59.

[7]. J-H Cui, L. Lao, M. Faloutsos, M. Gerla:, "*AQoSM: Scalable QoS Multicast Provisioning in Diff-Serv Networks*", GlobeCom 2005.

[8]. M. Bag-Mohammadi, S. Samadian-Barzoki, M.Nikoopour, N.Yazdani and N.Rezaee, "*A case for dense-mode multicast support in MPLS*," Computers and Communications, 2004. Proceedings. ISCC 2004. Ninth International Symposium, 28 June-1 July 2004.

[9]. A. Boudani and B. Cousin, "*A New Approach to Construct Multicast Trees in MPLS Networks*," Proc. Of Seventh International Symposiumon Computers and Communications, 2002.

[10]. G. Rouskas, and H. G. Perros, "*A Tutorial on Optical Networks*," Networking 2002 Tutorials, LNCS 2497, Pg. 155–193, 2002.

[11]. K. Lee, Kai-Yueng Siu, "An Algorithmic Framework for Protection Switching in WDM Networks".

[12]. Banerjee A., Drake J., Lang J. P., Turner B., Kompella K., Y. Rekhter Y, "*Generalized multiprotocol label switching: An overview of routing and management enhancements"*, IEEE Communication Magazine, No. 1 2001, pp. 144-150.

[13]. Y. Ye, C. Assi, S. Dixit, and M. A. Ali, "A simple dynamic integrated provisioning/protection scheme in IP over WDM networks," *IEEE Commun. Mag.*, vol. 39, no. 11, pp. 174–182, Nov. 2001.

[14]. P. Ashwood-Smith *et al.*, "Generalize multi-protocol label switching (GMPLS) architecture," RFC 3945, Oct. 2004.