Multicasting in MPLS Networks

Kang Bin Wang

A Thesis

in

The Department

of

Electrical and Computer Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Applied Science at

Concordia University

Montreal, Quebec, Canada

June 2003

Your file Votre référence
ISBN: 0-612-83877-3
Our file Notre référence
ISBN: 0-612-83877-3

ABSTRACT

Multicasting in MPLS Networks

Kang Bin Wang

The explosive growth of the Internet has generated interests from researchers around the world. New applications of the Internet, such as video, audio and mission-critical applications, challenge researchers to develop more effective ways to deliver these applications to end users. Multicasting, QoS and MPLS emerge as new technologies and are being developed. The integration of MPLS, multicasting and/or QoS is still an open topic. RFC3353, released in August 2002, gives an overview of IP multicast in an MPLS environment. This thesis first presents RFC3353 in brief and then an extension to RFC for multicast aggregation and loop prevention in MPLS networks are discussed. FEC aggregation is proposed where IP header is not affected in the aggregation in MPLS networks. A mechanism to dynamically aggregate multicast traffic in MPLS network is presented in this thesis. Both source specific tree and shared tree are considered as network topologies. Multicast loops may exist in MPLS network during network transition. The multicast loop forming in PIM-SM, PIM-DM and DVMRP is discussed in MPLS networks. The Colored thread and the No thread mechanisms for multicast loop prevention are provided in this thesis. In addition to aggregation and loop prevention, DiffServ multicast traffic merging in MPLS networks is also considered in this thesis. Two multicast traffic merging modes, root-initiated and leaf-initiated, are discussed in such DiffServ-supported MPLS networks. Multicast routing protocols such as PIM-SM, PIM-DM and DVMRP are considered here.

Acknowledgements

I would like to take this opportunity to express my deepest respect and gratitude to my supervisor, Dr. Anjali Agarwal. She not only guided me in the choice of the topic and the progress of development of the thesis, but also gave me constant technical assistant and valuable suggestions.

Special thanks must be given to my wife, Dong Mei Han, and my lovely daughter, Shi Yuan Wang. Thanks for their support and understanding when I engaged in doing this thesis in the late night and on the weekend.

Thanks also to our parents, my parent and my wife's parent, for their constant support and encouragement during my studying.

To my whole family ……

Table of Contents

List of Figures

List of Abbreviations

AF:	Assured Forwarding
AME:	Aggregation Management Entity
AS:	Autonomous System
ATM:	Asynchronous Transfer Mode
BA:	Behavior Aggregate
BB:	Bandwidth Broker
BE:	Best Effort
BGP:	Border Gateway Protocol
CBT:	Core Based Trees
CLS:	Controlled-Load Service
COPS:	Common Open Policy Service
CoS:	Class of Services
CR-LDP:	Constraint-based Routed LDP
DE:	Default
DiffServ (DS):	Differentiated Services
DR:	Designated Router
DRUM:	DiffServ and RSVP/IntServ Use of MPLS
DSCP:	DiffServ Code Point
DSMCast:	DiffServ Multicast
DVMRP:	Distance Vector Multicast Routing Protocol
EF:	Expedited Forwarding

E-LSP:	EXP-Inferred-PSC
Exp:	Experiment
FEC:	Forwarding Equivalence Class
FEC-to-NHLFE:	Next Label Forwarding Entry
FF:	Fixed Filter
GS:	Guaranteed Service
IETF:	Internet Engineering Task Force
IGMP:	Internet Group Management Protocol
IntServ:	Integrated Services
IS-IS:	Intermediate System-to-Intermediate System
ISP:	Internet Service Provider
LARM:	Label Abort Request Message
LCB:	Loop Control Block
LDP:	Label Distribution Protocol
LE:	Limited Effort
LER:	Label Edge Router
L-LSP:	Label-Only-Inferred-PSC
LMM:	Label Mapping Message
LRlsM:	Label Release Message
LRM:	Label Request Message
LSP:	Label Switched Path
LSR:	Label Switched Router

MIDB:	Multicasting Information Database
MPLS:	Multi-Protocol Label Switching
MRP:	Multicast Routing Protocol
MRT:	Multicast Routing Table
NMS:	Network Management System
NRS:	Neglected Reservation Subtree
Ntf:	Notification Message
OSPF:	Open Shortest Path First
PHB:	Per Hop Behavior
PIM-DM:	Protocol Independent Multicast – Dense Mode
PIM-SM:	Protocol Independent Multicast – Sparse Mode
PSC:	PHB Scheduling Class
QoS:	Quality of Service
RP:	Rendezvous Point
RPF:	Reverse Path Forwarding
RPT:	Rendezvous Point Tree
RSVP:	Resource ReServation Protocol
RSVP-TE:	Resource ReServation Protocol – Traffic Engineering
SE:	Shared Explicit
SLA:	Service Level Agreement
SPT:	Shortest Path Tree
TCB:	Thread Control Block

TCS:	Traffic Conditions Specification
TERO:	Tree Explicit Route Object
TLV:	Type Length Value
ToS:	Type of Services
TRRO:	Tree Record Route Object
TTL:	Time To Live
VC:	Virtual Channel
WF:	Wildcard Filter

Chapter 1 Introduction

Quality of Service (QoS) has become an important issue in the Internet in recent years. In the beginning, the Internet was used to transmit only data and all the data were treated with the same service level, the Best Effort (BE), although IP protocol was designed to support different classes of service. Today data, video, audio and mission critical applications are merged together and sent over the same network. Networks are now required to treat these applications differently because users may like to pay more to an Internet Service Provider (ISP) for a better service or as another example a minimal level of performance must be guaranteed to mission critical applications. These are achieved by different QoS requirements that provide the necessary bandwidth resources, controlled latency and jitter required by interactive traffic and improve data loss characteristics of the network [1]. Some new applications like video conferencing, distance learning and network gaming, not only require QoS but also use multicasting technology. MPLS [2-5] (Multi-Protocol Label Switching), IntServ [6] (Integrated Services) and DiffServ or DS [7] (Differentiated Services) are widely used to provide QoS in IP networks..

This Chapter provides the background knowledge of MPLS, multicast and QoS. Motivation, objectives and organization of the thesis are also provided here.

1.1 Multiple Protocol Label Switching (MPLS)

MPLS [2-5] is developed by the Internet Engineering Task Force (IETF) working group. It provides fast Internet traffic delivery by using label switching. It integrates label swapping with network layer routing and provides an approach to let network administrators implement Traffic Engineering and other features. MPLS supports multiple protocols in both data link layer (L2) and network layer (L3).

1.1.1 MPLS shim header

In conventional IP forwarding, routers assign received packets into Forwarding Equivalence Class (FEC) based on the "longest match" for packets' destination address. This process is done by every hop until packets reach destination. In MPLS, a particular packet is assigned into a particular FEC just once as the packet enters the network. Generic MPLS adds a shim header (Figure 1.1) between the headers of data link header and IP header. In an MPLS network, ingress routers label the traffic based on FEC, and core routers check the label to forward the traffic to next node. At egress node the label is removed and the traffic is restored to original. This pushes the complexity to edge routers, leaving the core routers simply to implement fast forwarding.

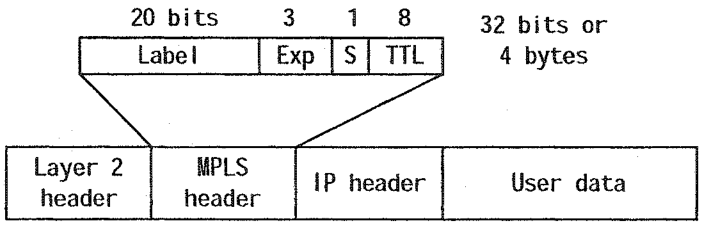


Figure 1.1 MPLS shim header

- The label field (20-bit) carries the actual value of the MPLS label.
- The Exp field (3-bit) is for experimental use, and mostly be used for CoS which can affect

the queuing and discard algorithms applied to the packet as it is transmitted through the network.

- The stack (S) field (1-bit) supports a hierarchical label stack.
- The TTL (time-to-live) field (8-bit) provides conventional IP TTL functionality.

Labels are bound to an FEC, and their assignment decisions are based on the forwarding criteria such as unicast destination routing, QoS, traffic engineering and/or multicast.

1.1.2 MPLS label distribution

In MPLS, traffic is forwarded along a Label Switched Path (LSP). MPLS labels need to be distributed along routers in LSP. MPLS does not specify specific label distribution mechanism. Labels can be piggybacked by existing signaling protocol or distributed by employing a separate protocol. Label Distribution Protocol (LDP) [8] is developed to separately distribute labels. By using LDP, Label Edge Routers (LERs) or Label Switched Routers (LSRs) map IP layer routing information into data-link layer to set up an LSP. Each LSP is associated with a specific FEC. Label distribution mode can be Downstream On Demand or Downstream Unsolicited. LSP can be set up with independent or ordered control. Figure 1.2 gives an example of LSP setup with downstream on demand, ordered control.

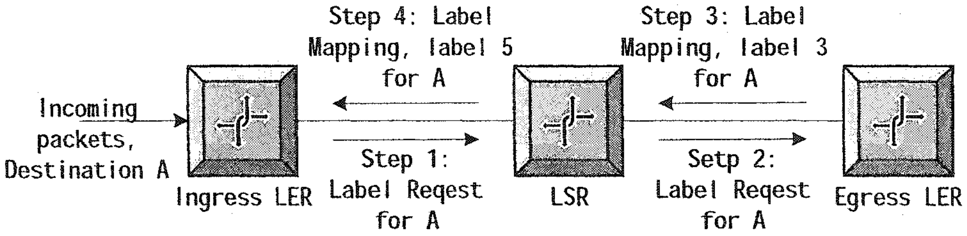
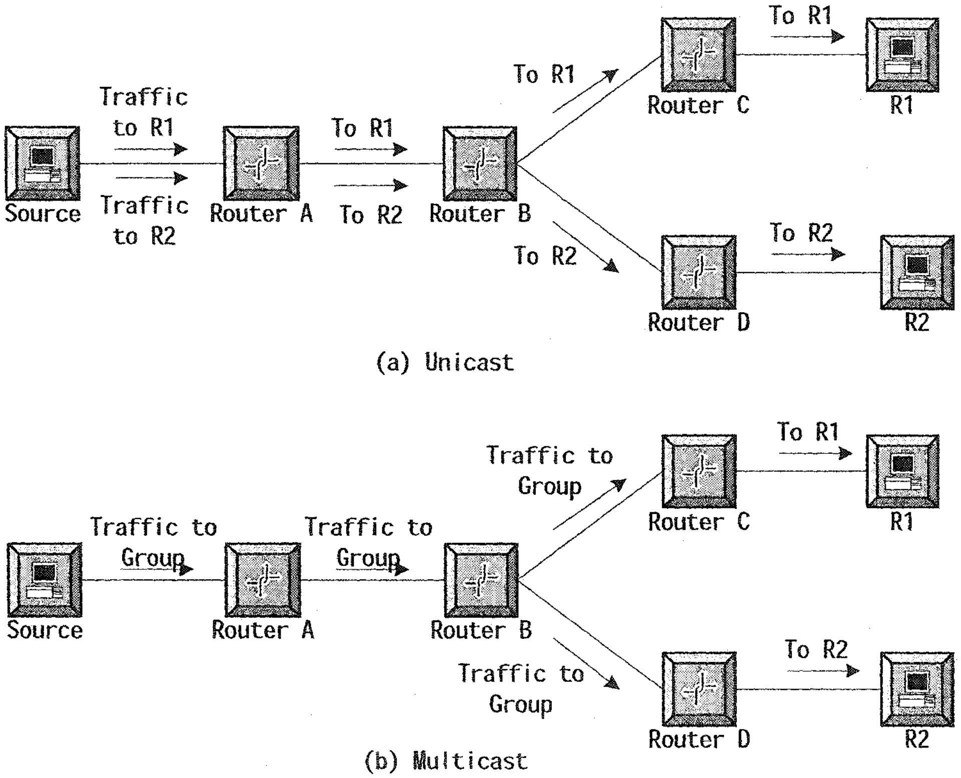

Step 4: Label Mapping, label 5 for A
Step 3: Label Mapping, label 3 for A
Incoming packets, Destination A
Ingress LER
Step 1: Label Reqest for A
LSR
Setp 2: Label Reqest for A
Egress LER

Figure 1.2 LSP setup

1.2 Multicasting

Multicasting [9,10] provides a way to save bandwidth compared to the use of several unicast channels in some applications by duplicating multicast flows only to the multicast tree branches. Multicasting technology sends traffic to one or group of receivers on the network at the same time and also reduces the server load. Unlike unicast routing protocol, which maintains a same shortest path for a certain <source, destination> pair, multicast routing protocol has to find one or more network topologies for a given group. In Figure 1.3 (a), traffic from source has to be sent twice along links between source and router B because of unicast <source, destination> pair. In Figure 1.3 (b), where traffic is sent to a group, only one copy of traffic is transmitted between source and router B. The topology can be either shared-based tree (Figure 1.4) where the source does not matter, e.g. PIM-SM [11] (Protocol Independent Multicast, Sparse Mode, where the tree is referred as Rendezvous Point Tree RPT); or source specified tree (Figure 1.3 (b)), e.g. DVMRP [12] (Distance Vector Multicast Routing Protocol), PIM-DM [13] (Protocol Independent Multicast, Dense Mode) and PIM-SM. Source tree is usually referred as Shortest Path Tree (SPT). DVMRP and PIM-DM use flood-and-prune to establish multicast tree, whereas PIM-SM joins or leaves group explicitly. Multicast makes the network complicated because of the dynamic group joining/leaving and uncertain destination address.



(a) Unicast

(b) Multicast

Figure 1.3 Comparison of unicast and multicast

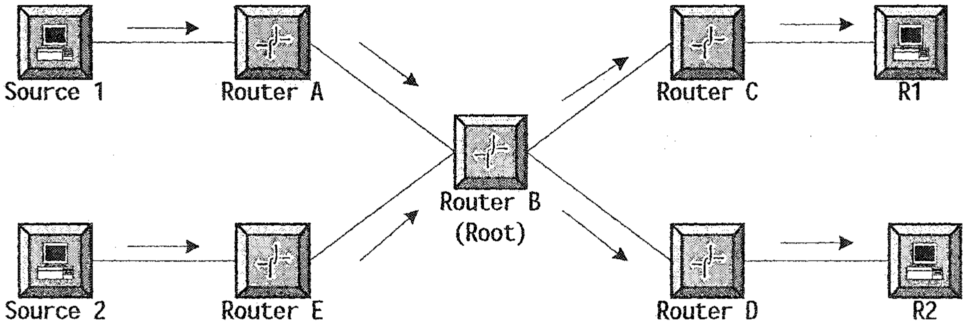


Figure 1.4 Shared tree of multicast

1.3 QoS

QoS gives a way to provide better services to selected traffic over various technologies. IntServ and DiffServ are two common ways to implement QoS.

1.3.1 IntServ

IntServ [6] was introduced to obtain QoS based on per-flow. Different applications can choose different controlled services to deliver their packets. Although IntServ gets more granularity, the routers have to maintain per-flow state and IntServ is generally not considered to be scalable. RSVP (Resource ReSerVation Protocol) [14, 15] is a signaling protocol used by IntServ to set up resources in the sender to receiver direction.

1.3.2 DiffServ

DiffServ [7] is more scalable than IntServ because the implementation of DiffServ is based on per-class. In DiffServ domain, edge routers classify the packets based on DSCP (DiffServ Code Point) and meter, mark, delay or drop traffic according to TCS (Traffic Conditions Specification). However, core routers need to only check the DSCP to obtain QoS parameters. DiffServ needs support of IP protocol because DSCP field substitutes the ToS (Type of Services) field in IPv4 or the CoS (Class of Services) field in IPv6.

1.4 Motivation

The explosive growth of the Internet has generated interest from researchers around the world. New applications of the Internet, such as video, audio and mission-critical applications, challenge researchers to develop more effective ways to deliver these applications to end users. Multicasting, QoS and MPLS emerge as new technologies and are being developed. Today's Internet looks like a much bigger picture than before. Figure 1.5 is an example of the Internet topology where core network and stub networks belong to different ASs (Autonomous Systems) to together make a bigger network. IP unicast routing protocols, such as IS-IS (Intermediate System-to-Intermediate System), OSPF (Open Shortest Path First), and multicast routing protocols (MRP), such as PIM-SM [11], DVMRP [12] and PIM-DM [13], are mainly employed in the AS.

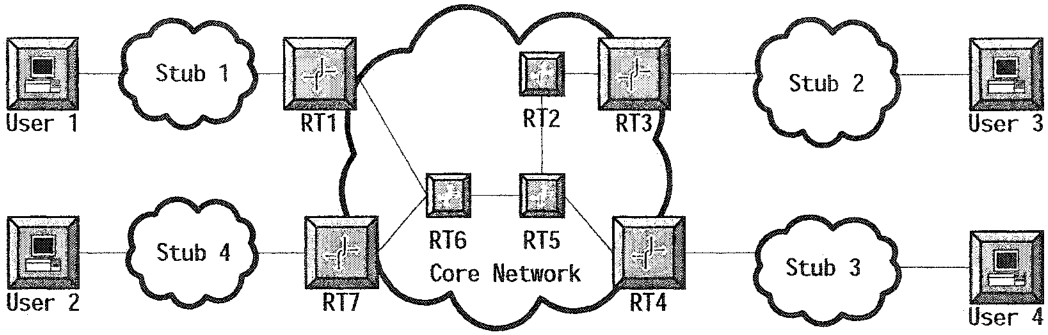


Figure 1.5 An example of Internet topology

New technologies provide better ways to transmit traffic. MPLS uses label switching to forward traffic. Other new features, such as traffic engineering, are also implemented by MPLS. Multicasting technology sends traffic to one or group of receivers on the network at the same time, and bandwidth is saved compared to unicasting. Multicast applications such as videoconference and distance learning are emerging in exponential growth. QoS gives a way to provide better services to selected traffic over various technologies.

Based on our knowledge, integration of MPLS, multicasting and/or QoS is still an open topic. We focus our interests on this new field.

1.5 Objectives

RFC3353 [16] gives an overview of IP multicast in an MPLS environment. The following issues arising when multicasting is supported in MPLS networks are addressed in the RFC:

- Aggregation
- Flood and Prune
- Source/Shared trees
- Co-existence of source and shared trees
- Uni/Bi-directional shared trees
- Encapsulated multicast data
- Loop-free-ness

Multicast routing protocols like PIM-DM, DVMRP use flood and prune mechanism to set up multicast tree. The tree is volatile due to receivers dynamically joining or leaving group. Traffic driven manner can be used to trigger LSP setup.

Some MRPs like PIM-SM [11], DVMRP [12] and PIM-DM [13] construct source specific trees whereas others like PIM-SM and Core Based Trees (CBT) set up shared trees. Shared trees consume fewer labels than source trees if LSPs are set up.

PIM-SM supports both source tree and shared tree. Co-existence of source tree and shared tree makes the LSP setup complicated if MPLS is to support PIM-SM. Four possible solutions are given in RFC3353 [16].

Shared tree in PIM-SM is uni-directional, but is bi-directional in CBT. Bi-directional tree is difficult to handle in MPLS network.

In PIM-SM, before sources join the group, multicast traffic is encapsulated in Register message to be forwarded to Rendezvous Point (RP). Then source tree is built between RP and Designated Router (DR) of the source. Non-member sources of bi-directional shared tree encapsulate the data towards the root node. This situation needs to be handled in MPLS network.

The above issues have been analyzed in RFC3353 [16]. But aggregation and loop-free-ness are left for future study. This thesis proposes solutions to these two problems. Dynamic multicast aggregation and multicast loop prevention in MPLS networks are presented. Multicast traffic merging with different QoS is also discussed.

1.6 Organization of the thesis

First Chapter presents background knowledge of MPLS, multicast and QoS, in addition to motivation and objectives. An overview of supporting multicast in MPLS networks is given in Chapter 2. Based on RFC3353 [16], this thesis gives solutions to the following topics. First, aggregation of traffic in MPLS networks is analyzed, and dynamic multicast aggregation is presented in Chapter 3. Then, loop prevention in MPLS networks is considered, and multicast loop prevention mechanisms are discussed in Chapter 4. Last, DiffServ multicast traffic merging in MPLS networks is provided in Chapter 5. Scenarios about the interworking between DiffServ and MPLS are given. Finally, conclusions are included in Chapter 6.

Chapter 2 Overview of Supporting Multicast in IP Networks

Multicasting applications such as video conferencing and distance learning require QoS. MPLS, IntServ and DiffServ can be used to deliver traffic with QoS. Agarwal and Wang [17] overview these different mechanisms to support multicast in IP networks. The interworking of these mechanisms within core and edge networks is also discussed.

2.1 IntServ and multicast support

IntServ [6] was developed with the need of real time applications, such as teleconferencing and remote seminars. It provides an Internet service model that includes best effort, real time and controlled link sharing and is inherently capable of supporting multicast. IntServ architecture [15] separates two function units, one for implementing or controlling QoS and another for signaling requests. There are two main QoS in IntServ, Guaranteed Service (GS) and Controlled-Load Service (CLS). Signaling is implemented by RSVP, however QoS control service is opaque to RSVP.

IntServ is application-based and provides per flow QoS. RSVP [14] cooperates with both unicast and multicast routing protocols. Resource reservation is implemented by RSVP Path and Resv messages. Since RSVP is receiver-oriented, receivers of the same multicast group in the SPT or RPT make resource reservation requests upstream towards the sender. These requests must be merged in the replication routers for scalability. RSVP provides three reservation styles: Fixed Filter (FF), Shared Explicit (SE) and Wildcard Filter (WF). Because RSVP routers have to maintain per flow soft state along the path, reservation overhead is logarithmic and IntServ is less scalable.

2.2 DiffServ and multicast support

Expedited Forwarding (EF) and Assured Forwarding (AF) are two main classes of QoS defined in DiffServ networks. Supporting multicast in DiffServ introduces some issues [7] that are not presented in unicast applications. Figure 2.1 shows multicast application in the DiffServ network. The number of receivers in multicast group may change frequently because users join/leave group dynamically. It cannot be predicted in advance how much resource is needed for a specific group. DiffServ network may connect to other peering DiffServ networks to provide end-to-end QoS. These different DiffServ networks may have different SLA (Service Level Agreement). DSCP value of DiffServ traffic is assigned in ingress router and a mapping between different SLAs may be needed. In order to reduce classification and traffic conditioning burden at the egress router, enforcement can be made at the ingress router of the first DiffServ domain. This can be achieved for unicast traffic because traffic is predictable. In case of multicast, traffic may be replicated at any place of DiffServ domain and go into different downstream peering DiffServ domains. It is difficult to get an agreement among all receivers. The possible solution given by Blake, et al. [7] is to isolate the unicast traffic from multicast traffic.

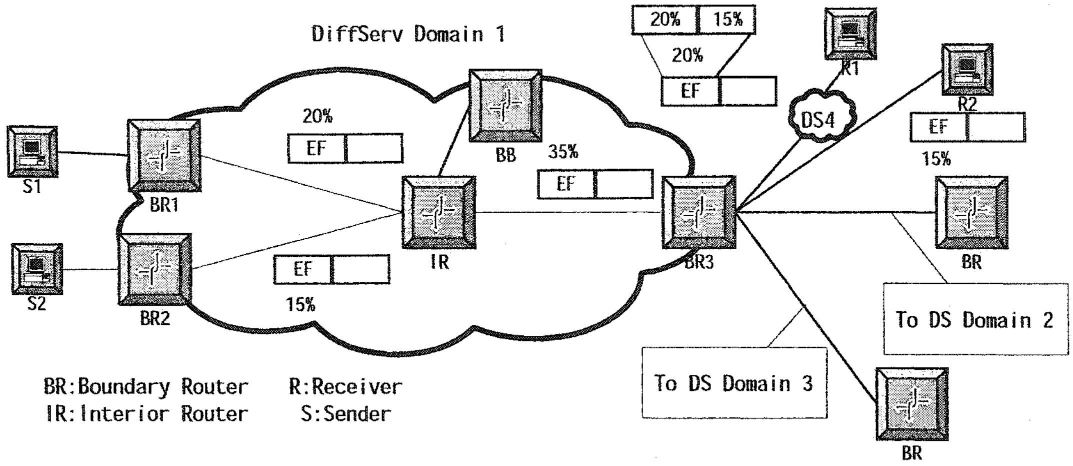


Figure 2.1 DiffServ and Multicast

Dynamic multicast group membership may cause other problems apart from resource reservation uncertainty. For example, consider two receivers and two groups in DiffServ Domain 1 of Figure 2.1 where S1 is the sender of group 1 and S2 is for group 2. R1 and R2 join group 1 and group 2 respectively. Group 1 reserves 20% of the bandwidth to provide EF, and group 2 gets 15% bandwidth for EF. In this case there are no problems and bandwidth is allocated satisfactorily. If R1 wants to join group 2, R1 should get default QoS, which is 15% bandwidth for EF. A 35% bandwidth is therefore needed for EF at the branch between BR3 and DS4. The boundary router BR3 has the traffic conditioning function, and DS4 uses additional 15% BW without permission from BR3. Therefore, BR3 will drop some packets to force the traffic to use admitted 20% BW. Without per-flow state in BR3, it will drop packets belonging to group 1 or group 2 randomly. Hence traffic of group 1 is adversely affected. This is a Neglected Reservation Subtree (NRS) problem given in [18, 19] and a solution is also provided there. Receivers can use BE or Limited Effort (LE), where LE is lower than BE and gets part of bandwidth from BE to join groups. When required resources are confirmed, A Bandwidth Broker (BB), which manages QoS resources in a given domain based on SLA, assigns a new DSCP, such that receivers can get better service than BE or LE.

There are different ways to support multicast in DiffServ network. Core routers can be multicast-capable, multicast-incapable or multicast traffic can be encapsulated to transfer to core routers. If core routers support multicast they have to maintain a multicast tree state per group. This makes core routers complex and not scalable as they have to check the state for every multicast tree. If core routers are unaware of multicast traffic, traffic is replicated at boundary routers. Striegel and Manimaran [20] have given a new architecture for multicast support in DiffServ domain. In this DiffServ Multicast (DSMCast) approach, core routers do not need to maintain multicast tree. The tree information is encapsulated in packet's header, and leaves core routers a little simpler and more scalable. However, encapsulation incurs additional costs. "Fat" header consumes additional bandwidth for each packet, and additional CPU cost is also incurred due to header processing. Under basic DSMCast model, the tree is built based solely on the network topology. Multicast traffic is transmitted like unicast apart from that core routers have to inspect the header to make replication when needed. The extension headers include identification field for DS core nodes, appropriate branching information, tunneling bit to bypass DS-non-capable nodes, and adaptive DS field to adapt to the heterogeneous DSCP requirements by the different receivers.

The request that receivers join /leave the group is forwarded by the egress router to the ingress router to be processed. The construction of the multicast tree is then done by the ingress router. Member's join /leave structure is discussed in [21].

Bless and Wehrle [22] add an extension of DS entry to multicast routing table to support heterogeneous DiffServ multicast group. This makes different branches in the same multicast tree get different QoS possible, but routers have to maintain a relatively larger routing table.

2.3 MPLS and multicast support

Unlike DiffServ and multicast, where the two are in the same network layer, MPLS and multicast are in the different layers. Ooms and Livens [23], and Ooms, et al. [16] have presented general issues in supporting IP multicast in MPLS networks.

MPLS routers include two separate components: control and forwarding. Control component uses standard routing protocols in L3 to exchange information with other routers to build and maintain a forwarding table. When packets arrive, the forwarding component (based on a label swapping forwarding algorithm) searches the forwarding table maintained by the control component to make a routing decision for each packet. Multicast traffic is delivered to all receivers of the same group along multicast distribution tree. Multicast routing protocols usually use Reverse Path Forwarding (RPF) or other incoming interface check to determine if the packet received belongs to a particular multicast group. Therefore in MPLS, multicast tree should be built on a per-interface basis by combining label value and incoming interface.

There are three ways to initiate label assignment: topology-driven, request-driven and traffic-driven. When MPLS is used to transmit unicast traffic, LSP is usually triggered by the network topology. In this case LSP already exists before traffic is transmitted. If topology-driven is applied to multicast, L3 tree needs to be mapped to L2 tree. MPLS-capable routers also have to maintain multicast tree. If multicast network topology changes frequently because of members dynamically joining /leaving, MPLS-capable routers spend a lot of resources to cope with these changes. Traffic-driven approach only sets up LSP to branches with traffic. It consumes fewer labels than topology-driven approach. This may take a longer setup time of LSP, but is better for the longer life span multicast group members. LSPs can also be set up by request-driven approach. For explicit multicast members joining/leaving protocols, such as PIM-SM and CBT, join/prune messages can be used to trigger LSP. The drawback is that multicast routing tree has to be constructed twice in L3 and in L2. RSVP Path and Resv messages can also be used to trigger LSP.

Label distribution can be achieved by dedicated protocols (extension may be needed to support multicast), e.g. LDP [8], or by piggybacking on routing protocols. For DVMRP and PIM-DM themselves there are no routing protocols that can be used for piggybacking. Join/prune messages of both PIM-SM and CBT, and RSVP are two good candidates for piggybacking.

Figure 2.2 presents some problems in an MPLS multicast network. Suppose R3 and R4 register to the same multicast group. R3 connects to MPLS network directly whereas R4 connects to an MPLS-incapable network. LER4 has to maintain two forwarding tables, one in L2 to LSR4 to reach R3 and another in L3 to non-MPLS network to reach R4. This is a mixed L2 and L3 forwarding [16] in a single router. Multicast traffic will go through different outgoing interfaces due to replication.

Consider the situation configured in Figure 2.2 where S1 and S2 are senders and R1 and R2 are receivers of the same group and LSR3 is the Rendezvous Point (RP). Assume initially multicast traffic is delivered along RPT. Multicast traffic coming from iif2 (Incoming Interface) is duplicated and forwarded to oif1 (Outgoing Interface) and oif2. Later suppose R1 joins SPT from S1. The multicast traffic from S1 is transferred to R2 by RPT and to R1 by SPT. Traffic coming in iif1 from S1 goes to oif1. Traffic from iif2 belonging to S1 only goes to oif2 and traffic belonging to S2 is forwarded to both oif1 and oif2. This is co-existence of source and shared trees in PIM-SM [16] and LER3 may have to maintain (*, G) and (S, G) tree state for the same group. This may cause transmitting duplicate packets to the same receiver during switchover from RPT to SPT. Ooms, et al. [16] propose different solutions to solve this problem, either by resorting to L3 routing or by using source specific label switching.

Setting up a source specific LSP is a solution to co-existence problem of SPT and RPT in PIM-SM. An example is given in [24]. Although ATM (Asynchronous Transfer Mode) is used in L2 in [24], the same methodology can be applied to general MPLS-capable L2 switches. Multiple labels are assigned in shared tree according to one VC (Virtual Channel) per source. Therefore, MRT (Multicast Routing Table) is needed to map shared tree to source-specific tree, and construct FEC-to-NHLFE (Next Hop Label Forwarding Entry). Label assignment is data-driven. Both Upstream Implicit and Downstream on Demand can be used. LDP is modified to support source-group tree structure. Labels reclamation is initiated from downstream LSR.

It is also possible to make core routers simple by pushing packets replication to LER [25]. Core routers do not have to maintain multicast routing state per group. Member's join/prune messages are forwarded to ingress router like unicast messages. Multicast traffic is duplicated there and transmitted to receivers transparent to core routers. PIM-SM multicast protocol has been discussed in [25]. In order to make the problem simpler, LER is assumed to be the candidate of RP. All LSPs are pre-established by using LDP [8] or CR-LDP [26] (Constraint-based Routed LDP) to communicate PIM messages between LERs as well as to transmit multicast data. The trade-off is between bandwidth inefficiency and multicast non-capable core routers.

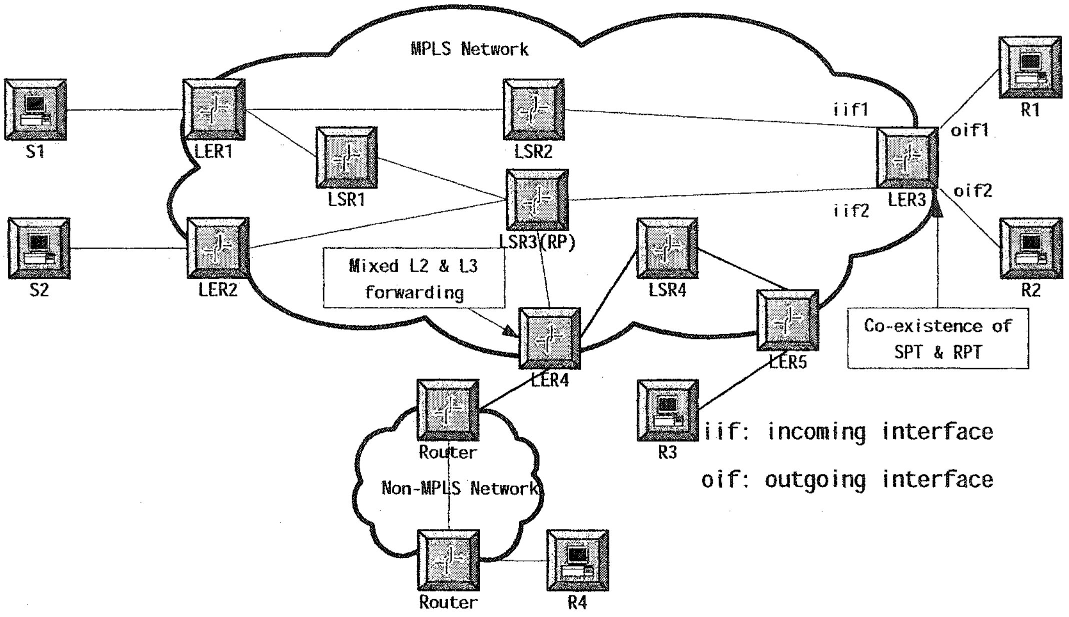


Figure 2.2 Supporting multicast in MPLS

Traffic triggered label distribution is also discussed in [27]. Data packet is first used to trigger label distribution. Graft and Prune messages of Dense Mode multicast routing protocols are then used to add new branches or remove branches from the tree.

Some networks may not support BGP (Border Gateway Protocol) when multicast traffic spans multiple MPLS networks belonging to different ASs (Autonomous System); PIM-SM join message cannot reach the source in other AS. Ooms, et al. [28] have provided a solution to create a pre-calculated multicast tree by extending MPLS signaling protocol other than multicast routing protocols. The tree is calculated based either on QoS requirement by the NMS (Network Management System) or by the initiator root or leaf of the tree in MPLS. Both root-initiated traffic engineered tree and leaf-initiated traffic engineered tree are mentioned in [28]. All tree information is embedded in signaling protocol, therefore, interior routers in MPLS do not need to support multicast. Ooms, et al. [28] have deployed CR-LDP [26] as signaling protocol and extensions for CR-LDP are discussed in detail. It should be noted that the tree constructed by this approach is immediately on L2 and applicable to applications of multicast that do not require dynamic tree.

2.4 Interworking between DiffServ and IntServ to support multicast

IntServ and DiffServ are complementary tools to provide end-to-end QoS. Su and Hwang [29] have discussed the multicast support to achieve end-to-end QoS with IntServ and DiffServ based networks. In Figure 2.3, the core network is DiffServ-capable whereas the two stub networks are IntServ-capable. Senders and receivers connect to stub networks. RSVP is used as signaling protocol in IntServ, and COPS (Common Open Policy Service) is employed in DiffServ by the BB. When multicast group is formed, sender of the group broadcasts RSVP Path message to all receivers by tunneling through the DiffServ domain. Receivers who want to get better service than BE reply back with RSVP Resv message to sender. When Resv arrives, DiffServ issues COPS request to BB that then interacts with all routers inside the domain to reserve resources for receivers. In order to support different reservation styles (SE, FF and WF) of RSVP in DiffServ domain, some modifications to COPS objects are needed. Su and Hwang [29] have shown the end-to-end signaling procedures for DVMRP and PIM-SM in both FF and SE.

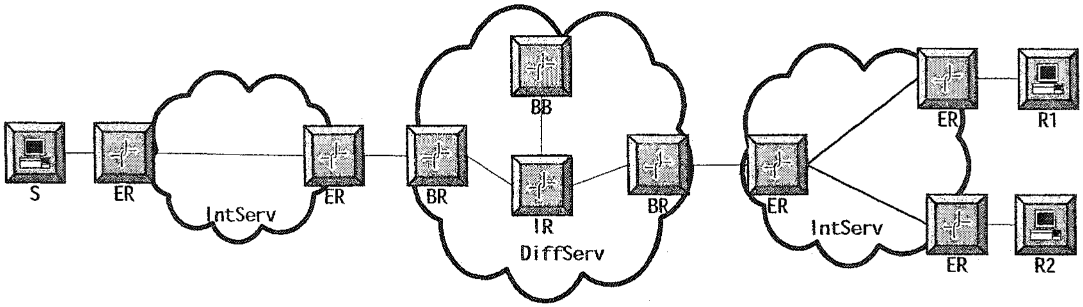


Figure 2.3 Interworking between DiffServ and IntServ to support multicast

IntServ supports GS and CLS whereas EF and AF are supported in DiffServ. Figure 2.4 shows the mapping of service level needed between IntServ and DiffServ.

CoS in IntServ	CoS in DiffServ	MPLS Service Class
GS	EF	Gold
CLS	AF	Silver/Bronze
	BE	

Figure 2.4 A way of mapping CoS among IntServ, DiffServ and MPLS

In DiffServ core routers are able to change DSCP in order to support different reservation styles. Boundary routers and PIM-SM RP have the same functionality, such as classification, metering.

2.5 Interworking between MPLS and IntServ to support multicast

RSVP-TE (Traffic Engineering) is a specification of extensions to RSVP for establishing explicit LSPs in MPLS networks. RSVP-TE for LSP tunnels in RFC3209 [30] is restricted to unicast LSPs. Researchers [31, 32] have specified extensions to RFC3209 to support multicast.

These proposals do not require or rely on traditional multicast routing protocol to set up the multicast tree. They assume that the tree calculation is done by other means, be it online or offline. The tree pre-calculated in this way must satisfy the TE parameters used in MPLS. Two kinds of trees are proposed, root (sender)-initiated tree originated from ingress LER (root) and leaf (receiver)-initiated tree originated from egress LER (leaf).

Cheng [31] has discussed root-initiated tree. Tree explicit route object (TERO), tree record route object (TRRO) and their respective sub-objects extending from RFC3209 are used to calculate multicast tree. RSVP Path message containing pre-calculated tree is forwarded to all leaf LERs, Resv message is transmitted back to root LER and merged from downstream LSRs to upstream LSR.

Chung [32] has supported both sender-initiated tree and receiver-initiated tree. Sender-initiated tree is used for new source of multicast group; receiver-initiated tree is used for dynamically joining/leaving receivers. The tree information is pre-gathered in MIDB (Multicasting Information Database) by using IGMP (Internet Group Management Protocol). For a new sender (Figure 2.5), root LSR uses this tree information to forward Path message to LSR-RP1, LSR-RP2, LSR1, LSR2 and LSR4, and confirmed by Resv message. After that, data traffic from sender is transmitted to receiver1 and receiver3. New receiver2 uses Join message to join the group, and Join message is forwarded to root LSR. This initiates the recalculation of the tree. After that, Root LSR sends Join message back to LSR-RP2 with a TREE object. Then LSR-RP2 sends Path message to LSR3 and gets back Resv message from LSR3. Receiver2 can now receive multicast data. LSR-RP is a LSR that has more than one branch participating in multicast group.

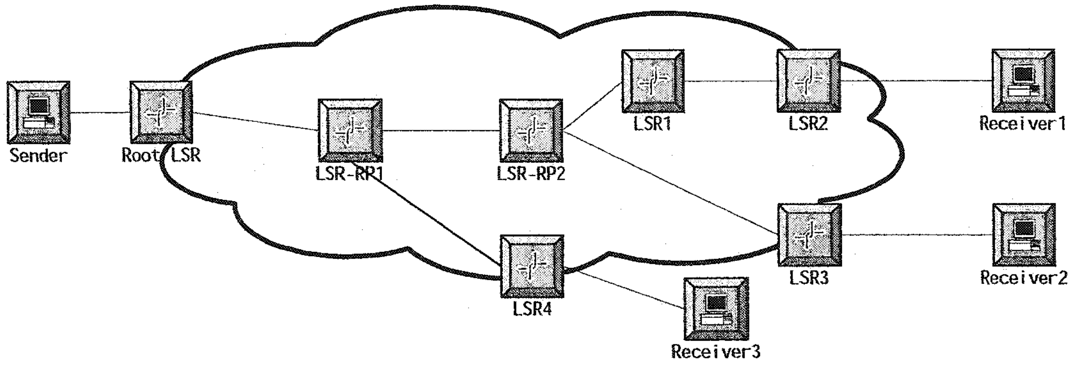


Figure 2.5 Using RSVP-TE in MPLS to support multicast

RSVP-TE provides a way to support multicast in MPLS. This approach is scalable because LSRs other than root LSR do not need to maintain MIDB. But periodical Path and Resv message contribute more traffic to network. Multicast supporting by RSVP-TE is restricted to source specific multicast group.

2.6 Interworking between MPLS and DiffServ

The basic difference between MPLS and DiffServ is that MPLS is in L2 and DiffServ is in IP Header. Clearly a mapping is needed between the two approaches. DSCP has 6 bits whereas MPLS has a 3-bit EXP field. If DSCP maps to EXP, some information will be lost. A mapping has been proposed by Rouhana and Horlait in [33] and is duplicated in Figure 2.4. They also propose DRUM (DiffServ and RSVP/IntServ Use of MPLS), an architecture that delivers end-to-end QoS for both DiffServ and IntServ networks, and the mapping is based on DSCP, CoS in IntServ and EXP. Andrikopoulos and Pavlou [34] have used ATM in L2. Part of VCI is used to map DSCP. Interior DS-compliant ATM-LSRs perform the appropriate traffic management function on ATM cells to interpret DSCP correctly. Some extensions are needed in MPLS and LDP to make them

DiffServ-capable.

Faucheur, et al. [35] have given detailed information about MPLS support of DiffServ. EXP-Inferred-PSC (PHB Scheduling Class) LSPs (E-LSP) and Label-Only-Inferred-PSC LSPs (L-LSP) are discussed specifically. If multiple Behavior Aggregates (BA) are mapped to a single LSP, the EXP field of MPLS shim header is used to determine the Per-Hop-Behavior (PHB). Up to eight BAs can be supported by one LSP for a given FEC. This is referred to as E-LSP. In L-LSP, separate LSPs are established to map different BAs. LSRs infer exclusively PSC from label value and EXP field conveys the Drop precedence. Label forwarding models for DiffServ LSRs, detailed operation of E-LSP and L-LSP are discussed in more detail in [35].

Label Distribution in DiffServ-MPLS may be piggybacked or may use specific protocol. RSVP with extension is piggybacked to establish LSPs supporting DiffServ in MPLS networks. CR-LDP [26] extended from LDP, describes LSPs with QoS requirements with some parameters.

Moh, et al. [36] have proposed an enhancement to E-LSP. It distributes traffic with the same class type to the same LSP. This is referred to as per-class TE or DS-TE. Labels are encoded to both FEC and service class (or class type) information. Three class types are defined: type 0 (BE), type 1 (EF) and type 2 (AF1 and AF2). Load is balanced by class type.

Jing, et al. [37] have presented an implementation algorithm to support DiffServ in MPLS-based ATM switches.

The published literature in interworking MPLS and DiffServ networks relate to unicast applications. Problems arise when multicast traffic spans multiple Internet clouds. In Figure 2.6, DS1, DS2, DS3 and MPLS network support multicasting. DS1, DS2 and DS3 may use different SLAs in each domain. The same issues related to supporting multicast in DiffServ and MPLS need to be considered here. The questions related to when the label is to be created and destroyed, how to distribute the label, and how to provide end-to-end QoS spanning multiple different networks need to be addressed. Multicast routing state may need to be maintained by all routers from end to end, or this function may be pushed to edge routers of each network and leave the core routers simple. For PIM-SM, it has to be decided where to locate the RP and how to cope with co-existence of SPT and RPT and problems mentioned in [16]. Extensions in LDP or CR-LDP are needed to support multicast and DiffServ in MPLS.

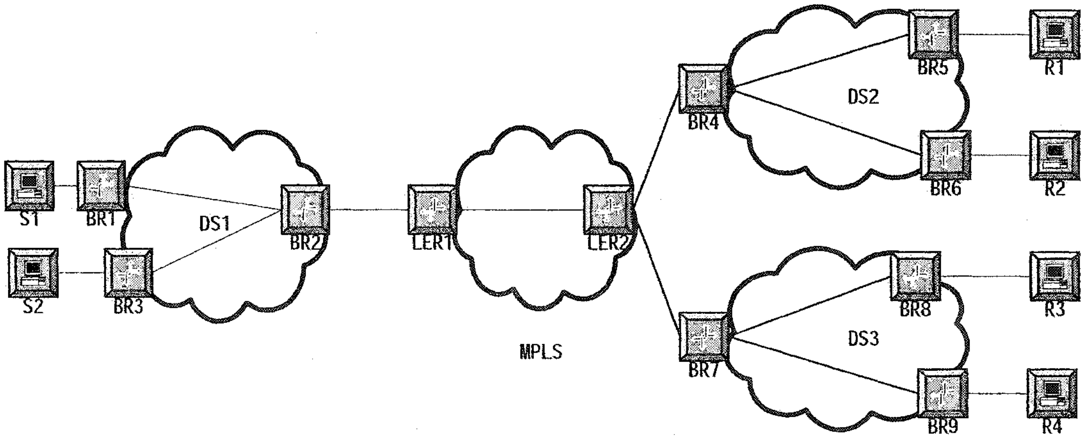


Figure 2.6 Interworking between DiffServ and MPLS to support multicast

2.7 Conclusions

IntServ, DiffServ and MPLS are technologies currently used in Internet to provide QoS. Multicast applications, such as videoconferencing and distance learning, are emerging in a fast step. In this Chapter, general issues on supporting QoS to multicast are discussed and different ways to implement IP multicast in IntServ, DiffServ or MPLS based networks are analyzed in detail.

IntServ is inherently capable of supporting multicast and provides per-flow QoS. Scalability is critical and must be considered when QoS is required for multicast applications. Per-class based DiffServ is more scalable than per-flow based IntServ. MPLS and DiffServ have similarities in some aspects. They both have "fat" edge routers and "slim" core routers. They are however basically different in architecture. MPLS runs over L2, whereas DiffServ needs support of IP protocol because DSCP field substitutes the ToS in IPv4 or the CoS in IPv6. When considering support multicast, the network architecture plays an important role and must be determined first. DiffServ, MPLS or both can be employed in core network, and multicast can be supported in entire network or only in edge routers. Each router needs to maintain a multicast tree if the entire network supports multicast. If multicast is supported only in edge routers of core network, multicast tree must be pre-calculated and multicast traffic is tunneled across core routers. In this case, core routers are more scalable. Some issues must be considered before supporting multicast in MPLS network. Multicast LSP setup trigger strategy and label distribution schemes should be addressed. Mixed L2 and L3 forwarding in multicast routing protocols, co-existence of source and shared trees in PIM-SM, handling LSP setup in PIM-SM-enabled MPLS network where one is source tree between ingress LER and RP and the other is shared tree between RP and egress LERs, dynamic receivers joining/leaving, as well as flooding and pruning are main issues when supporting multicast in MPLS network.

Subsequently, network configurations based on these different architectures and their interworking to provide end-to-end QoS are also discussed. The interworking among IntServ, DiffServ and MPLS is more complicated in multicast network than that in unicast network. Cooperation of different technologies in one network, end-to-end QoS provision and service type mappings are important points. IntServ and DiffServ are complementary tools to provide end-to-end QoS to multicast traffic. IntServ and DiffServ deal with the issues relevant to resource reservation and traffic policing. Here RSVP and COPS are used as signaling protocols to support QoS and multicast. RSVP-TE is used by MPLS to interwork with IntServ and is currently restricted to source specific multicast group. However, supporting DiffServ in MPLS is a scalable solution. E-LSP and L-LSP are used to support DiffServ in MPLS for unicast applications. Extensions to MPLS and label distribution protocols should be considered if MPLS is to support multicast and DiffServ. Implementation of end-to-end QoS spanning multiple different networks is still in research if interworking MPLS, DiffServ and multicast. When interworking the current different technologies such as IntServ, DiffServ, MPLS and multicast, mapping of service types is needed to deliver end-to-end QoS.

This overview provides information to researchers and service providers to get a general view of supporting QoS to IP multicast. What have been done, what technologies are available, and what needs to be researched and developed are clearly presented here.

2.8 The work of the thesis

Based on this overview, it is observed that extensive research work needs to be done when interworking IntServ, DiffServ, MPLS and multicast, and it is hoped that there is a reasonable potential research and development activity in this area. The work of this thesis is based on RFC3353 [16], which analyzes the issues on supporting multicast in MPLS networks and aggregation and loop-free-ness are left for future study. This thesis proposes solutions to these two problems. Chapter 3 focuses on aggregation of multicast traffic in MPLS networks and dynamic multicast aggregation mechanism is presented. Chapter 4 presents two mechanisms to implement multicast loop prevention in MPLS layer. Multicast traffic merging with different QoS is also discussed in Chapter 5.

Chapter 3 Aggregation in MPLS Networks

Aggregation means that traffic with the same characteristics can be delivered together on one route such that the size of routing tables is reduced. Aggregation can be applied to both unicast and multicast traffic. By sharing an aggregated tree with multiple multicast groups for multicast traffic, the routing states in IP routers are reduced largely. Multicast aggregation in MPLS networks saves labels consumed by multicast trees. This Chapter analyzes multicast aggregation in both IP and MPLS networks, and a new mechanism supporting dynamic multicast aggregation is presented.

3.1 Multicast aggregation in IP networks

Cui et al [38] presents a multicast aggregation scheme. The main idea is to force traffic from multiple multicast groups to share the same distribution tree, an aggregated tree. This would reduce the routing states in the routers because core routers only maintain an aggregated tree instead of multicast tree per group. This scheme is applied to IP layer routing protocol.

Figure 3.1 shows an example of a transit network. Multicast traffic from group G0, G1 and G2 enters the transit network from ingress router RA and exits by egress routers RD, RE, RF, RG and RH. RB and RC are core routers. By setting up an aggregated multicast tree starting from RA instead of one tree per each multicast group, fewer states are maintained by core routers RB and RC in the transit network. From Figure 3.1, we can see the tree of G0 and G1 is exactly the same, but tree of G2 has fewer branches than that of G0 or G1 because there is no active receiver of G2 in RD, RF and RH. The aggregated tree is a perfect match [38] for groups G0 and G1, and a leaky match [38] for group G2. By following the aggregated tree, traffic from G2 wastes the bandwidth from RB to RC and from RC to RD, RF and RH. This is the tradeoff for state reduction.

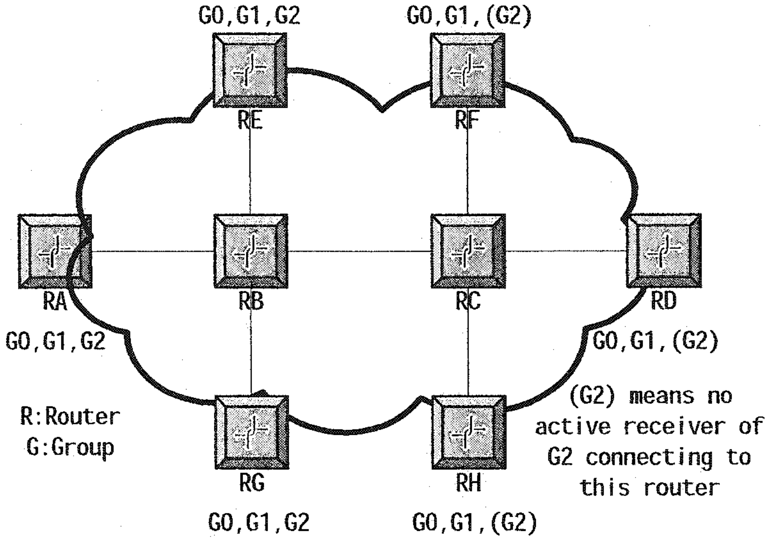


Figure 3.1 Multicast aggregation scheme

A logical management entity called tree manager is introduced in [38] to manage aggregated tree and to match multicast groups to aggregated tree. The aggregated tree may be pre-defined or established on demand. Some prerequisites must be satisfied, such as to keep the original group information in the data packets such that outgoing edge routers can recover this information and then know the forwarding of these packets further. In addition, the aggregated tree information must be stored in the forwarding packets in order to let core router forward these packets along the aggregated tree.

The aggregated multicast tree scheme can be applied to both source tree and shared tree. Figure 3.2 shows two source specific multicast trees have the same incoming edge router RA. RA is configured to act as aggregated tree manager, and aggregated tree originates from RA. In Figure

3.3, Source 1 and Source 2 are in group G1 and group G2 respectively. Two groups have the same rendezvous point (RB). So RP is configured to be the tree manager. It is a little complicated for bi-directional multicast tree in CBT. RB is the Root for G1 and G2 in Figure 3.4. Source 1 and Source 2 are in G1 and Source 3 in G2. All receivers join both G1 and G2. Receiver 1 gets multicast traffic from Source 3 by RF-RC-RD, and not from RB-RC-RD. Multicast traffic can be short cut to on-tree receivers. Multiple tree managers are needed to handle this situation; one tree manager for one source or one group.

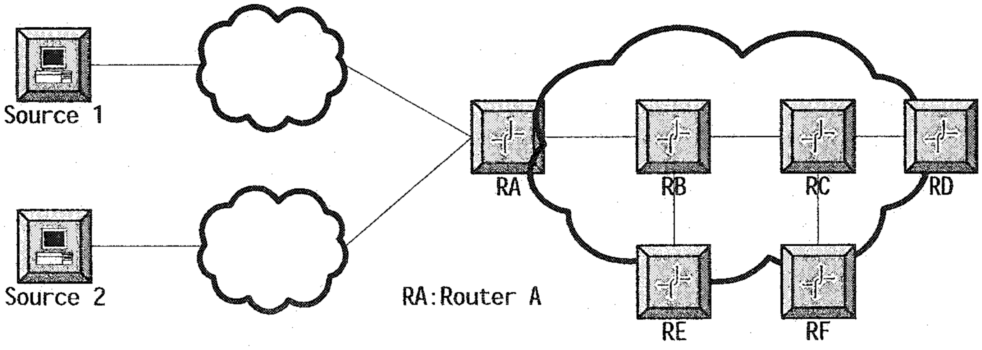


Figure 3.2 Source specific multicast aggregation

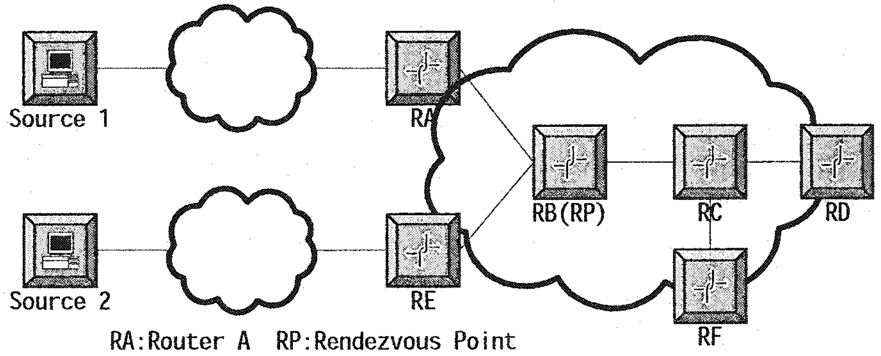


Figure 3.3 Uni-directional shared multicast aggregation

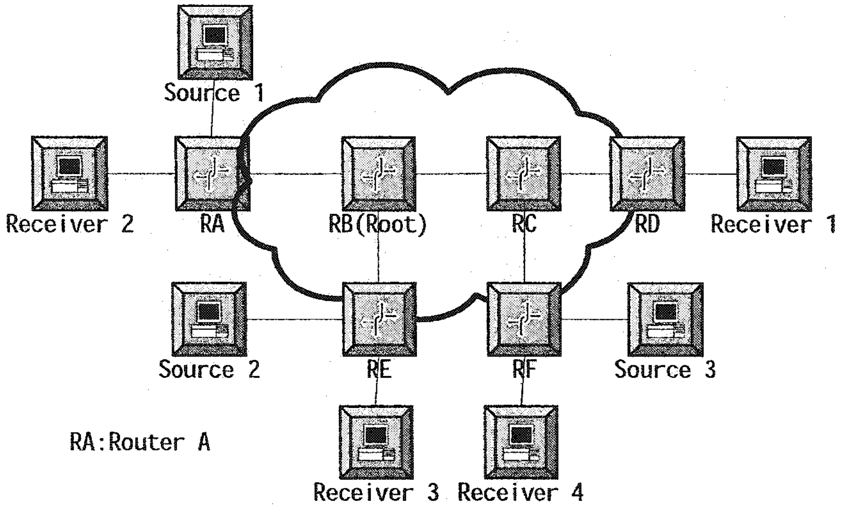


Figure 3.4 Bi-directional shared multicast aggregation

3.2 Aggregation in MPLS networks

3.2.1 Unicast aggregation in MPLS network

Aggregation [2] in MPLS network means FEC aggregation. It is a FEC union or a collection of FECs that have the same route inside the routing table. This FEC union itself is also a FEC. The component FECs in FEC union may have distinct address prefixes, but all traffic of FEC union follow the same route in MPLS network. So one label or one LSP for FEC union may be configured compared to one distinct outgoing label for each component FEC. Aggregation saves the number of labels as well as LSP setting up control traffic, and changes FEC granularity from fine to coarse. In unicast, FEC is usually configured according to destination address or destination address prefix which appears in routing table. Figure 3.5 shows an example of unicast FEC before and after aggregation. For example, all traffic to 191.1.1.x and 202.2.2.x go through from R_i to R_e

in MPLS domain. FEC1 is configured as 191.1.1.x, and FEC2 as 202.2.2.x. Two LSPs are established on link between R_i and R_e before aggregation, one for FEC1, and another for FEC2 as shown in Figure 3.5a. If FEC aggregation (see Figure 3.5b) is implemented in R_i, a new FEC, FEC3 (R_e ID) that is a FEC union of FEC1 and FEC2, is configured. Only one LSP is needed on link between R_i and R_e after aggregation. Aggregation saves the number of labels as well as the amount of LSP setting up control traffic.

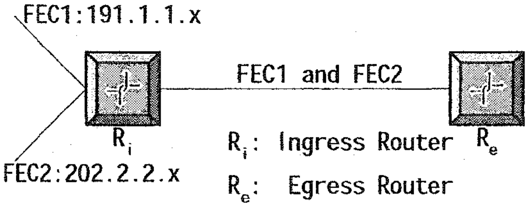


Figure 3.5a Unicast FEC before aggregation

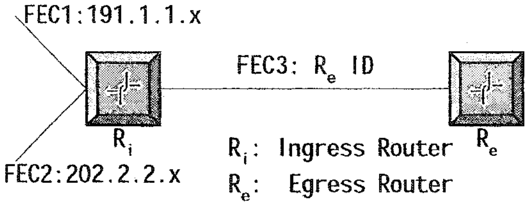


Figure 3.5b Unicast FEC after aggregation

3.2.2 Multicast aggregation in MPLS networks

Aggregation in unicast traffic is straightforward because the traffic knows where to go and how to get there. As receivers join or leave group dynamically in multicast, traffic is forwarded to group address and follows the multicast tree. Multicast FEC configuration is based either on sole group address in case of shared tree or on <source, group> pair in case of source specific tree.

There are different ways to implement FEC aggregation of multicast traffic in MPLS network. One way is to map aggregated IP layer multicast tree to MPLS network. Another is to apply explicit routing to multicast traffic in MPLS network. These two possible implementations are mentioned in [38]. An aggregated tree calculated by tree manager per incoming edge router in SPT or per Rendezvous Point (RP) in RPT is introduced. This tree is maintained in the IP layer. If MPLS tree is mapped directly from IP layer, the same solutions employed in IP layer can be employed in MPLS network. If MPLS multicast tree is set up by explicit routing, a centralized tree manager can be employed to pre-calculate the aggregated tree. These two solutions require pre-calculated multicast tree to implement aggregation in MPLS network. A mapping is also needed to map original multicast tree to new calculated aggregation tree. The original group information must be kept in the packets to let outgoing routers restore them and forward these packets to neighbor networks or receivers. A new aggregation mechanism, dynamic multicast FEC aggregation, is therefore required.

3.2.3 Dynamic multicast aggregation

In this Chapter, a new mechanism of dynamic multicast aggregation is proposed. In order to support multicast aggregation in MPLS network, all MPLS routers must support multicast routing protocol. These MPLS routers should dynamically and automatically aggregate the multicast traffic from different multicast groups. To support this, the MPLS routers should have the function of aggregation management, which can be achieved by introducing an Aggregation Management

Entity (AME) that checks the multicast traffic from different groups according to the outgoing interface. If different multicast group traffic goes through the same outgoing interface, AME issues a request to set up a LSP for a new FEC on this outgoing interface. The aggregate of group addresses are mapped as the new FEC. The downstream LSRs check all applicable multicast groups based on this new FEC to set up LSP. The LSP is set up on combined outgoing interfaces of these group addresses under this new FEC according to the multicast routing tables of these groups. LSP setup messages are needed when a new group comes. For example, traffic from two groups follows the same outgoing interface oif1 in a LSR, an FEC for each group needs to be configured before aggregation. If dynamic multicast FEC aggregation is employed, only one FEC representing all groups on this outgoing interface is needed. This dynamic multicast aggregation mechanism can be applied to both shared tree and source specific tree and are discussed below separately.

3.2.3.1 PIM-SM

In case of PIM-SM used as multicast routing protocol, routers have to maintain a shared tree to forward multicast traffic to receivers. Consider an example in Figure 3.6. There are two groups where Source 1 is the source of group G1 (224.1.2.1) and Source 2 is for group G2 (225.1.2.2). RP1 is the RP of G1 and RP2 is for G2. R1, R3, R4, R5 and R6 are receivers of G1 and R2 is the receiver of G2. Because R1 and R2 are in the same network, multicast traffic from RP1 and RP2 must follow the LSR2 and LSR3 to reach LER3. In LSR2, iif0 is the incoming interface of traffic from G1, and iif1 is for traffic from G2. Oif1 is the outgoing interface of both traffic of G1 and traffic of G2. Before multicast FEC aggregation, two FECs/LSPs are needed, an FEC/LSP1

(FEC1: 224.1.2.1) for G1 and an FEC/LSP2 (225.1.2.2) for G2. AME realizes that traffic of G1 and G2 goes through the same route from LSR2 to reach receivers, so a new FEC3 (224.1.2.1, 225.1.2.2) is configured. This mechanism is not scalable as the number of group increases greatly. New LSP setup follows down the multicast trees of both G1 and G2 starting from the aggregation point. A new LSP3 is established and the old LSPs are released between LSR2 and LER3 and LER4. Multicast traffic from both G1 and G2 must follow the new LSP to reach LER3 and LER4. This may cause an additional bandwidth requirement between LSR3 and LER4. LSP1 is needed between LSR2 and LER3 and LER4 because of the receivers R1, R3 and R4, but LSP2 is only established between LSR2 and LER3 because only R2 needs traffic from G2. New established LSP3 cannot identify which part of traffic comes from G1 and which part from G2. LSP3 forwards all traffic from G1 and G2 to LER3 and LER4. This wastes the bandwidth between LSR3 and LER4 and LER4 has to discard traffic from G2. We trade off bandwidth for number of label reduction. Bandwidth waste may become a benefit if R3 or R4 joins G2 later, R3 or R4 can then get traffic from G2 with no additional LSP setup because it is already in LER4. It should be noted that LSP between LSR2 and LER5 is not affected as no receiver of G2 connects to LER5.

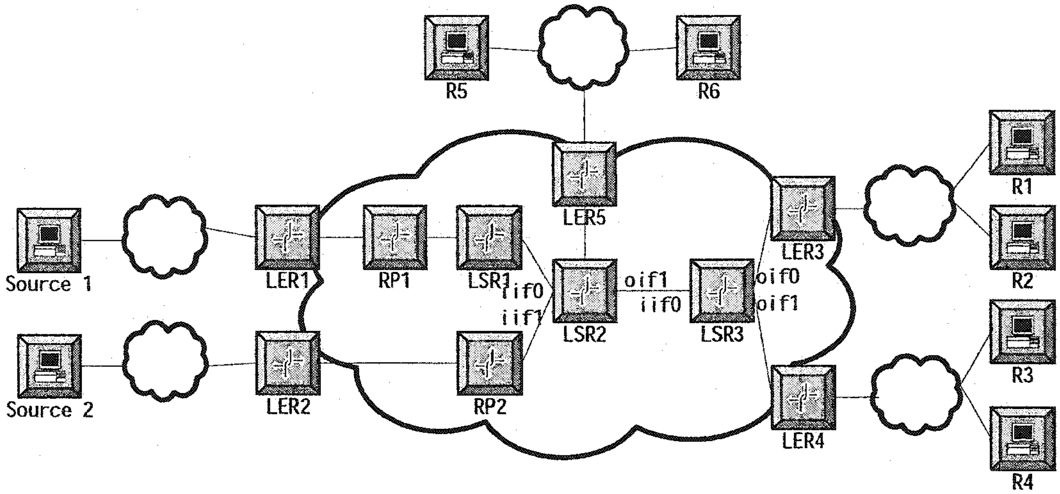


Figure 3.6 Dynamic multicast aggregation in PIM-SM

3.2.3.2 DVMRP or PIM-DM

If DVMRP or PIM-DM is configured, source specific tree is maintained by multicast-enabled MPLS routers in IP layer. The same aggregation procedure is followed as PIM-SM (see Figure 3.7). R1 to R6 belong to the similar groups as that in Figure 3.6. LER1 is the incoming ingress router in MPLS network for Source 1, and LER2 is for Source 2. Aggregation occurs in LSR2.

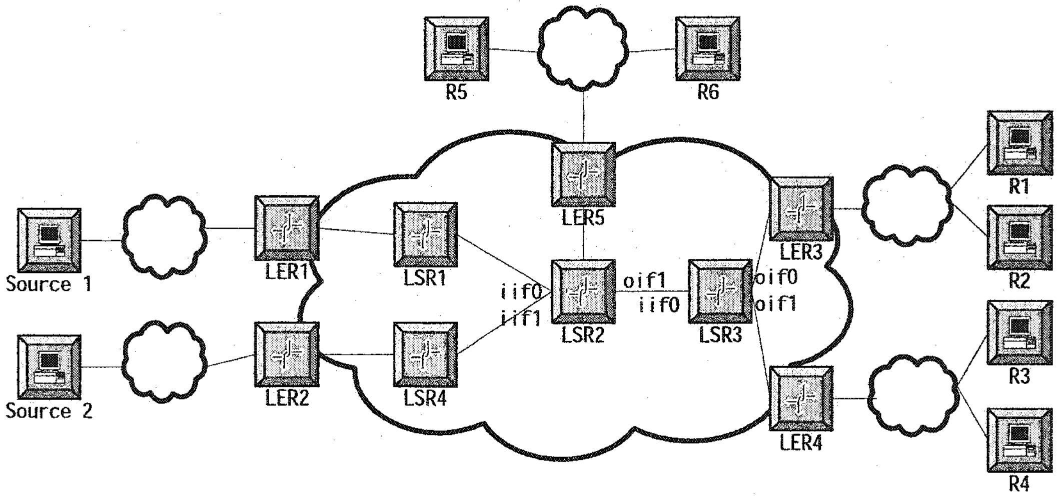


Figure 3.7 Dynamic multicast aggregation in DVMRP/PIM-DM

3.2.3.3 FEC before and after aggregation

Figure 3.8 and Figure 3.9 show a reduction in the number of FECs after aggregation. New FEC only relates to the group addresses and is independent of source address.

FEC before aggregation	FEC after aggregation
224.1.2.1	<224.1.2.1, 225.1.2.2>
225.1.2.2	

Figure 3.8 FEC based on group addresses in case of PIM-SM

FEC before aggregation	FEC after aggregation
Source 1, 224.1.2.1	<224.1.2.1, 225.1.2.2>
Source 2, 225.1.2.2	

Figure 3.9 FEC based on group addresses in case of DVMRP/PIM-DM

3.3 Multicast aggregation in MPLS with DiffServ

Multicast aggregation is also possible for MPLS networks supporting DiffServ. DiffServ provides a scalable service differentiation because the implementation of DiffServ is based on per-class. In DiffServ domain, edge routers classify the packets based on DSCP (DiffServ Code Point) and meter, mark, delay or drop traffic according to TCS (Traffic Conditions Specification). Core routers only need to check the DSCP to obtain QoS parameters. If DiffServ and multicast are both supported in MPLS network, multiple multicast tree for the same group will be maintained based on different DSCP. This means multiple LSPs for the same group in MPLS network. FECs can be configured to <S, G, DSCP> or <*, G, DSCP>. Multicast aggregation can be applied in this situation. If traffic from different multicast group has similar group addresses and the same DSCP, the traffic can be aggregated together to reduce the number of labels consumed. Figure 3.10 gives an example. We assume that PIM-SM is configured in MPLS network and DiffServ is also enabled. Source 1 and Source 3 belong to the same group G1 (224.1.2.1), and obtain AF11 and AF21 for their traffic respectively. Source 2 and Source 4 both subscribe to group G2 (225.1.2.2), and their traffic assumes AF11 and AF21 respectively. Both group G1 and G2 maintain two multicast trees, one for AF11 traffic and another for AF21 traffic. From LSR2, four LSPs are maintained. Aggregation can start from LSR2 like non-DiffServ supporting MPLS network. We aggregate LSPs based on DiffServ and group address prefix. After aggregation, only two LSPs are needed. One is for FEC (224.1.2.1, 225.1.2.2, AF11) and another LSP is for FEC (224.1.2.1, 225.1.2.2, AF21).

In case of source specific tree in PIM-DM or DVMRP in DiffServ enabled MPLS network, aggregation policy can be implemented as above example.

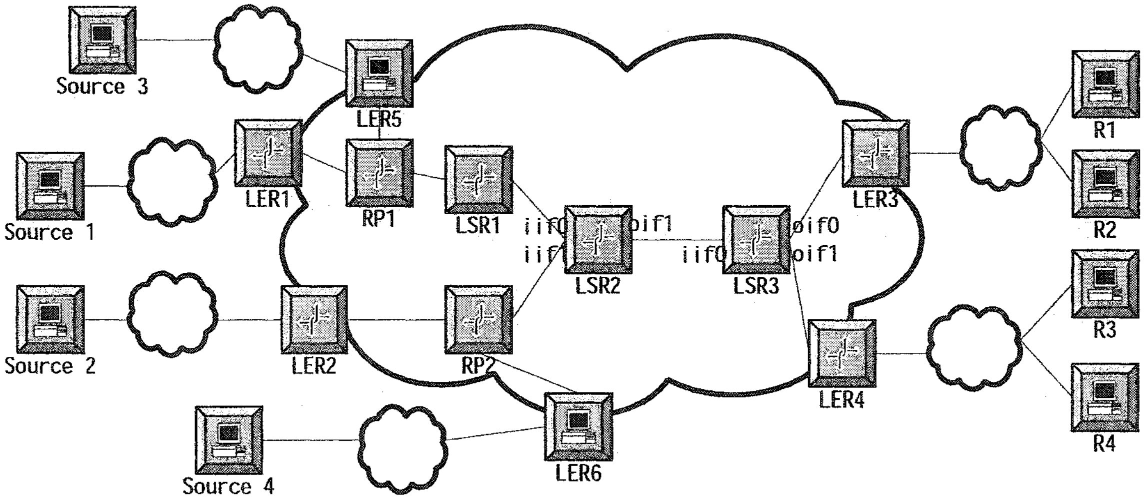


Figure 3.10 Multicast aggregation with QoS in MPLS network

3.4 Aggregation Management Entity (AME)

Every MPLS router has to maintain an Aggregation Management Entity (AME) to support dynamic multicast aggregation. AME keeps all aggregated FECs and their component information as the following.

FEC1

List of group addresses

…

FECn

List of Group addresses

Figure 3.11 gives the flow chart of AME.

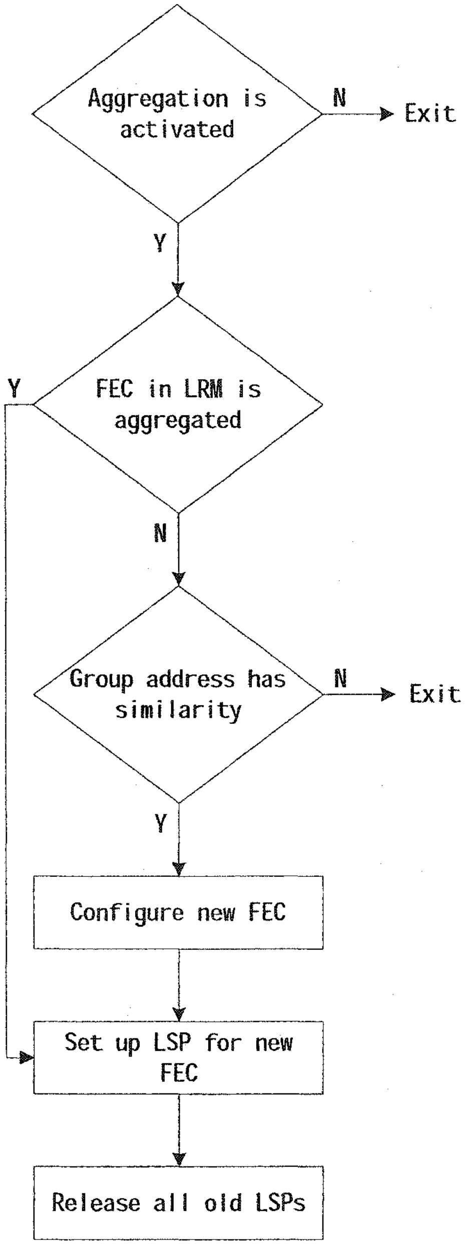
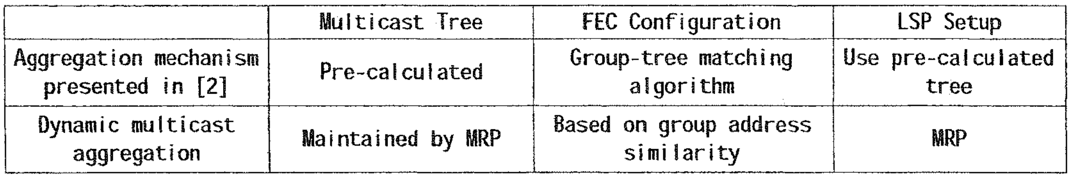


Figure 3.11 Flow chart of AME

3.5 Comparison

In the following Figure 3.12, we compare dynamic multicast aggregation with the aggregation mechanisms presented in [38].



	Multicast Tree	FEC Configuration	LSP Setup
Aggregation mechanism presented in [2]	Pre-calculated	Group-tree matching algorithm	Use pre-calculated tree
Dynamic multicast aggregation	Maintained by MRP	Based on group address similarity	MRP

Figure 3.12 Aggregation mechanisms comparison

3.6 Discussion on Multicast aggregation

By providing multicast aggregation in MPLS networks with or without QoS, number of labels/LSPs is reduced in participating MPLS routers, and in addition fewer LSP control messages are needed. However, as we have seen above, there is a trade-off between this kind of overhead reduction and bandwidth usage because after aggregation routers may deliver traffic to MPLS egress routers to which non-group-receivers connect.

Aggregating multicast traffic in MPLS network has some advantages compared with doing it in the IP layer. Before aggregation in IP layer, the original group information is required to be provided in order to restore this information when multicast traffic exits this network and goes to its neighbor network. This is not needed in MPLS aggregation. Here, the FEC granularity is modified from fine to coarse. Nothing is modified to traffic's IP layer header.

3.7 Conclusion

Multicast aggregation takes place among different multicast groups. Multicast aggregation in IP layer is different from that in MPLS network. Aggregating multicast traffic in IP layer is to reduce multicast states in IP routers. In MPLS, network aggregation is FEC aggregation where IP header is not affected. Multicast aggregation in MPLS network has advantages over that in IP layer. A mechanism to dynamically aggregate multicast traffic in MPLS network is proposed in this Chapter. There is a tradeoff between label reduction and bandwidth usage. Both source specific tree and shared tree are considered as network topologies. New FEC is introduced and the related advantages are also covered in the Chapter. The dynamic multicast aggregation proposed can also be implemented in DiffServ-enabled MPLS networks.

Chapter 4 Loop Prevention in MPLS Network

Multicast routing protocol uses either flood-and-prune in PIM-DM and DVMRP or explicit join/leave messages in PIM-SM to build multicast routing tree. During the transient phase, where the network topology is converging, if a flood-and-prune mechanism is used, there is a high probability that routing loops are formed because participating nodes or routers do not get full knowledge of changing topology of the network in DVMRP and PIM-DM. In PIM-SM, PIM Join message may also form loop during the transient state of network [39]. When network topology becomes stable, i.e. after converging, the network should be free of loops.

This Chapter gives a brief view of multicast loop prevention in IP networks. Multicast loop forming is then discussed, and finally two multicast loop prevention mechanisms in MPLS networks are proposed.

4.1 Multicast loop prevention in IP layer

DVMRP, which is similar in many ways to RIP with extension to support multicast, is tightly coupled to the underlying unicast routing algorithm. PIM has been shown to suffer from temporary loops resulting from the use of inconsistent unicast routing information [39].

Multicast routing protocols use Reverse Path Forwarding (RPF) to check if the multicast traffic arrives at the correct incoming interface. DVMRP maintains a separate multicast routing table and uses it to do RPF check, whereas PIM uses existing unicast routing table to perform RPF check. As the RPF check prevents multicast traffic coming from more than one incoming interface, it prevents forming multicast routing loops.

4.2 Multicast loop forming in MPLS network

Multicast loop may also exist in MPLS network. There are three ways to initiate multicast LSP setup: request driven, topology driven and traffic driven [16]. Request driven intercepts explicit multicast routing control message, and it cannot be used by flood-and-prune mechanism. Topology driven needs to map IP layer multicast routing table to MPLS layer. Traffic driven can be used by any existing multicast routing protocol, and LSP is set up upon traffic arrival. We discuss multicast loops in MPLS network based on the traffic driven case.

For flood-and-prune multicast routing protocols, when the first traffic comes to an MPLS edge router, LSP setup is initiated, and LSP setup accompanies first traffic to flood MPLS network to all egress LSRs. This means that multicast trees in IP layer and in MPLS layer are formed at the same time. In IP layer multicast routing protocols execute RPF to check if the multicast traffic comes from the correct incoming interface, then to make decision to discard or accept traffic. A prune message is sent out from check-failed incoming interfaces. However, there is no such mechanism in MPLS layer to do RPF check. Initial multicast flooding tree is maintained by LSRs, LSP may be set up to more than one incoming interface in a single LSR for the same multicast traffic. Therefore, there is a very high probability to form a loop in the MPLS layer in comparison with the IP layer. Multicast loop also exists in MPLS layer in multicast routing protocols with explicit control messages. The following sections will give an example in DVMRP, PIM-DM and PIM-SM respectively.

4.2.1 Multicast loop forming in DVMRP

DVMRP is a truly multicast routing protocol because it maintains its own multicast routing table and sends and receives multicast route update messages to update the table. When multicast traffic comes during the transient phase, while the network topology is converging, or during route updates, there is a possibility to form a loop both in the IP layer and in the MPLS layer.

Multicast traffic floods a truncated broadcast tree in DVMRP. This tree is formed and maintained by routing update messages. During the transient state, routers may not get full network topology change information. Consider the topology in Figure 4.1, A is the ingress LER for multicast traffic. Multicast traffic will go through interior LSRs and flood to egress LER J and K, then forward further to other networks or to receivers directly.

Immediately after a network topology change, network routers may not get updated route report and a multicast loop may form. Assume node F in Figure 4.1 (a) and (b) does not get updated route report from H that link H-J is not available when multicast traffic comes to F. Loop D-F-H-D is formed in Figure 4.1 (a) since the shortest path D-F-H-J is not available. Similarly, loop C-D-F-H-C is formed in Figure 4.1 (b) since its shortest path C-D-F-H-J is not available. RPF check in IP layer can be used to check correct incoming interface for multicast traffic and break the loop. But in MPLS layer nothing can be done if no new mechanism is employed. In (c), assume multicast traffic comes from the incoming interface of link E-F during transient. After network topology converges, traffic will come from the incoming interface of link D-F. It should be noted here that the same multicast traffic comes to F from two different incoming interfaces in MPLS layer since both the paths are active. There is no mechanism in MPLS to release the transient path. This causes duplicated transmission of the traffic.

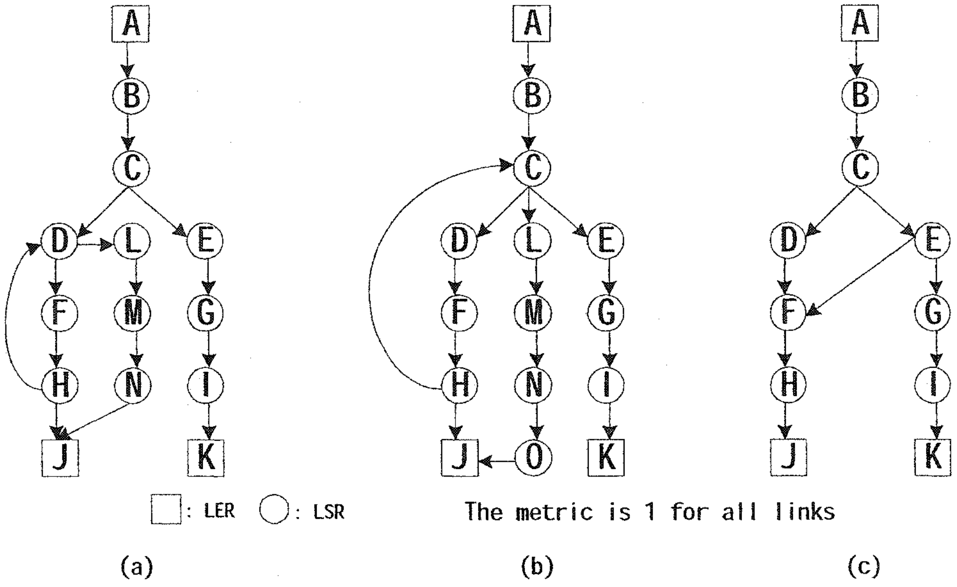


Figure 4.1 MPLS multicast loop in DVMRP

4.2.2 Multicast loop forming in PIM-DM

Unlike DVMRP, PIM-DM floods the entire network at the beginning of incoming multicast traffic to build broadcast tree toward all edge routers. Then PIM-DM uses existing unicast routing protocol to do RPF check, and prevent loops in IP layer. So multicast traffic duplication may happen even while not in the transient state. Compare Figure 4.1 (c) with Figure 4.2, traffic from link E-F and D-F cannot come at the same time because DVMRP maintains its own multicast routing table, but this happens if PIM-DM is implemented in the network. Then PIM-DM uses existing unicast routing table to do RPF check. In Figure 4.2, the MPLS multicast loop D-F-H-D is formed when traffic floods network if link H-J is not available. The multicast traffic loops and duplication may occur at the same time in Figure 4.2.

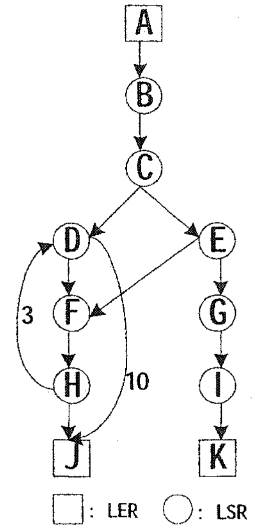


Figure 4.2 MPLS multicast loop in PIM-DM

4.2.3 Multicast loop forming in PIM-SM

PIM-SM employs explicit Join/Prune messages to form multicast tree. Underlying unicast routing protocol is used to find the reverse path to join or prune group. Even so, during transient state, temporary loop may still be formed [39]. This loop in IP layer may cause loop forming in MPLS layer. If no mechanism is implemented in MPLS layer, this loop will exist until LSP is released even though the loop in IP layer is broken.

Figure 4.3 gives an example. When receiver R joins group to A, the Join message is forwarded from A to B and then to C. It is further forwarded to D in order to reach RP by the shortest path. Because of link failure of link D-E, D has no other link but D-B, so a loop is formed for Join message in IP layer (see Figure 4.3 (a)). After loop is detected in IP layer, actions are taken by C to forward Join to E and finally to RP (see Figure 4.3 (b)). So LSP setup is initiated from RP to E, C, B and A. If at this time Join state is still kept in B and D for any reasons, a loop C-B-D-C is formed in MPLS layer (see Figure 4.3 (c)).

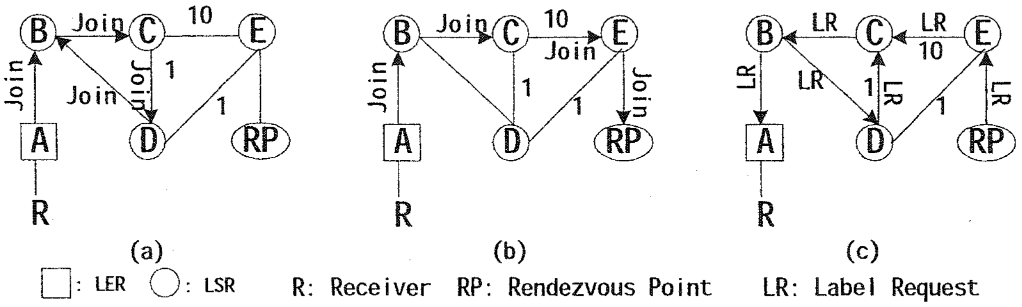


Figure 4.3 MPLS multicast loop in PIM-SM

4.3 Multicast loop prevention

As we have seen in Section 4.2, there are two types of multicast loops in MPLS layer. Multicast loops such as D-F-H-D in Figure 4.1 (a) and Figure 4.2, C-D-F-H-C in Figure 4.1 and C-B-D-C in Figure 4.3 are real loops. If nothing is done to prevent these loops, traffic entering the loop will be forwarded forever or until the traffic's timer is out. It can be seen in Figure 4.1 and 4-2 that if traffic comes from more than one incoming interface in MPLS layer, multiple copies of traffic will be forwarded that can preempt the network resources. This is not a real loop but has the similar effect as the real loop. This duplication should also be prevented. In this thesis, when we mention loops, other than specified clearly, it means either one of the two situations.

Node F in Figure 4.1 (c) and Figure 4.2 has two incoming interfaces. One of them should be marked as loop. There are two possible solutions to prevent loops in MPLS layer. The simplest way is to consider the multicast traffic incoming interfaces coming later than the first one marked as loops. In this case the link used to transmit multicast traffic in the MPLS layer may be different from that in the IP layer with no MPLS employed. For example, for node F link D-F is marked loop because it comes later than link E-F. In IP layer, however, link D-F passes RPF check correctly whereas link E-F fails. Another way is to access IP layer multicast tree regularly to get the correct traffic flow information, then consider the incoming interfaces not in the tree as loops, and set up LSP along the correct incoming interface. This keeps the consistency of multicast tree in IP and MPLS layer.

4.4 Multicast loop prevention - Colored thread mechanism

This thesis presents a simple mechanism to detect and prevent setting up multicast MPLS LSP that can form loops in MPLS network. This mechanism, based on "thread", is similar to that presented in [40,41] with changes made to adapt to multicast. The mechanism can be used with ordered downstream-on-demand label distribution.

4.4.1 Basic mechanism

When multicast LSP setup is initiated in a MPLS network, LSRs set up a separate thread that represents a different color for each of its outgoing interface and extend these threads downstream. This thread is accompanied by LSP path setup message. The thread messages are required at the beginning of LSP setup, and they are also needed when dynamic sources and receivers join/leave a group. If the downstream LSR has only one outgoing interface, it keeps the received color thread and further extends it downstream. If the downstream LSR has more than one outgoing interface, it assigns a new colored thread for each of its outgoing interfaces and forwards them downstream, and the colors of its outgoing interfaces are different from the color of its incoming interface.

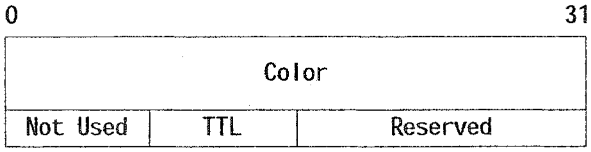
When the thread reaches egress LER, it means no loop is found. A special colored thread, which is transparent color, is rewound to the thread initiator, and label mapping is assigned along LSP that has this special color.

When a LSR receives a colored thread if it already has one colored thread in another one of its incoming interfaces, this means a loop is formed. A stalling flag in Thread Control Block (TCB) is signed. This thread is transmitted back to the thread initiator, and LSRs along the path are updated to sign stalling flags. No label mapping is assigned in looped LSP.

When a new branch of multicast tree or next hop change is found, only the nodes that are downstream of change are involved in a new thread distribution procedure.

4.4.2 Thread

A thread in Figure 4.4 is composed of Color, Not Used, TTL and Reserved part.



0		31
Color		
Not Used	TTL	Reserved

Figure 4.4 A thread

4.4.2.1 Color

A color is unique in network in both time and space. The router ID and an event handle in this LSR are used to form color. Special value all 0s indicating transparent color is reserved to indicate loop free.

4.4.2.2 Not Used

This field is not used here.

4.4.2.3 TTL

TTL is needed to guarantee this thread message will eventually time out if the loop is not prevented in some cases.

4.4.2.4 Reserved

This part is reserved for future use.

4.4.3 Thread primitive actions

4.4.3.1 Thread Extending

The nodes that initiate a new thread are thread initiator. The nodes that are candidates of thread initiator can be ingress LER, root or branch point of the tree. When a node receives a colored thread from upstream node, if no loop is found, the node extends the thread downstream. This process is thread extending.

Figure 4.5 gives examples of thread extending. In (a), Node A is ingress LER or root LSR. When multicast traffic comes to the node A, a thread Blue is created and forwarded to downstream LSR B. Link C-F in (b) is a new branch, node C is a thread initiator because it is a branch point. A new color Black is set for the new branch. Node B and C in (c) extend the thread color downstream and make no changes. In (d) link C-E is new. A new color Red is created to the new branch while color in old link C-D is changed from Blue to Green as every outgoing interface in a branch point is set a different color and these colors are different from the color of its incoming interface.

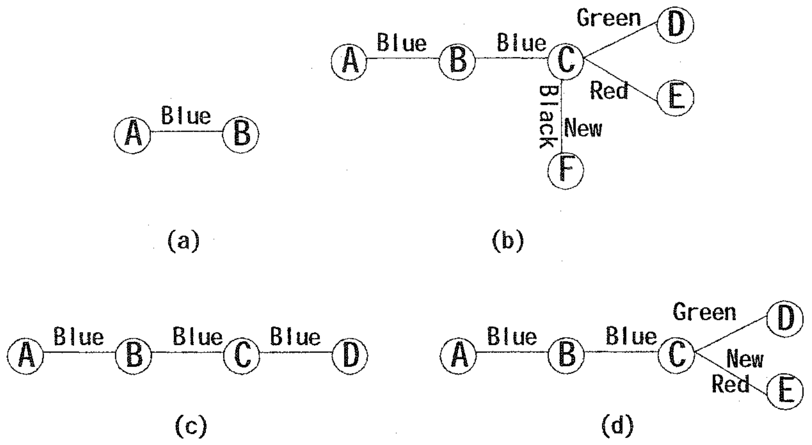


Figure 4.5 Examples of thread extending

4.4.3.2 Thread Withdrawing

In case of a broken next hop or a released LSP, thread withdrawing is sent downstream to withdraw the thread. For example, consider a link between LSR A and LSR B in Figure 4.6 (a). If the LSP setup is released, the thread from A to B must also be withdrawn. In Figure 4.6 (b) and (c), link C-D is withdrawn. Link C-E is changed to Blue from Red in (b) because only one colored link is left, but colored link C-E is not changed in (c) because its incoming link is transparent.

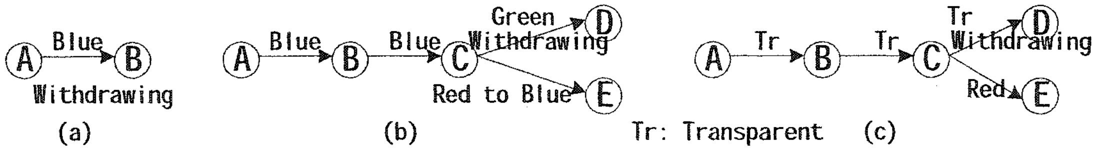


Figure 4.6 Examples of thread withdrawing

4.4.3.3 Thread Rewinding

When a thread reaches egress LER, it means this part of LSP is loop-free. So a transparent thread is sent back toward the thread initiator. If the upstream thread of the thread initiator is still colored, it continues to set the thread Transparent and send it back even if other downstream colored threads may still exist. This process proceeds until the root of the tree, the ingress LER or such an LSR that has its upstream thread already transparent.

In Figure 4.7, node D is egress LER. When colored thread Green reaches D, the thread is rewound, and the color is changed to Transparent. Although C has other colored threads, thread rewinding proceeds toward A by following the reverse LSP setup direction.

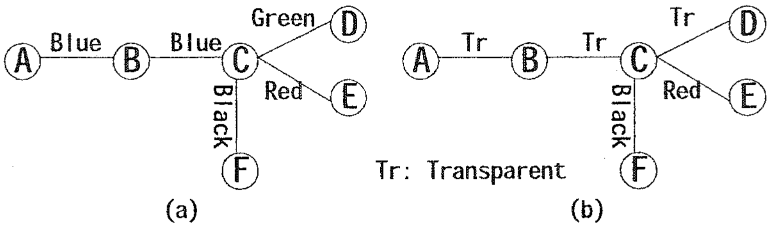


Figure 4.7 Examples of thread rewinding

4.4.3.4 Thread stalling

When a node gets more than one thread from its incoming interfaces for the same FEC, loops are found. The node keeps one thread and sets the stalling flag on all other threads. Thread stalling is reversed to the thread initiator. The nodes along the path set the stalling flag of Thread Control Block (TCB).

In Figure 4.8 (a), a loop is formed with pure Green thread whereas it is made of Green thread and Blue thread in (b). In both cases, Green thread is stalled. More than one incoming interface for the same multicast FEC is found in (c), and Red thread is stalled.

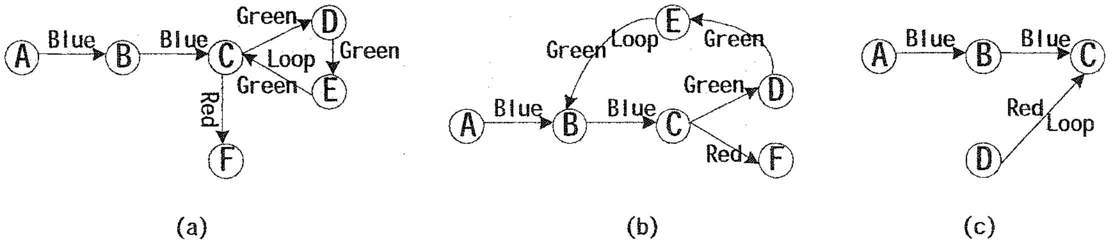


Figure 4.8 Examples of thread stalling

4.4.4 Thread Control Block

In order to support loop prevention, a Thread Control Block (TCB) is needed in participating nodes. TCB maintains the following:

FEC

Incoming interfaces

Each incoming interface has the following information:

Upstream LSR node

Color

Stalling flag

Label

Outgoing interfaces

Each outgoing interface has the following information:

Downstream LSR node

Color

Stalling flag

Label

4.4.5 Examples with Colored thread mechanism

Consider Figure 4.9 (a) where the same multicast tree is used as Figure 4.1 (a). When a colored thread comes to C, C sets up a new thread, Red, to D. D forwards Red to F and then to H. Because of failure of link H-J, H extends Red to D. Here a loop is found, and the Red thread is stalled on D, F and H. After link H-J is recovered, H withdraws Red from D, and tries to reach J. A label is issued since J is the egress LER. Red is changed to Transparent in D, F and H as shown in Figure 4.9 (b). In another case, D withdraws Red from F, then H and D, and tries to reach J along L, M and N. When label request message successfully reaches J, label mapping can go through this way in the reverse direction. Transparent is set on these routers as shown in Figure 4.9 (C).

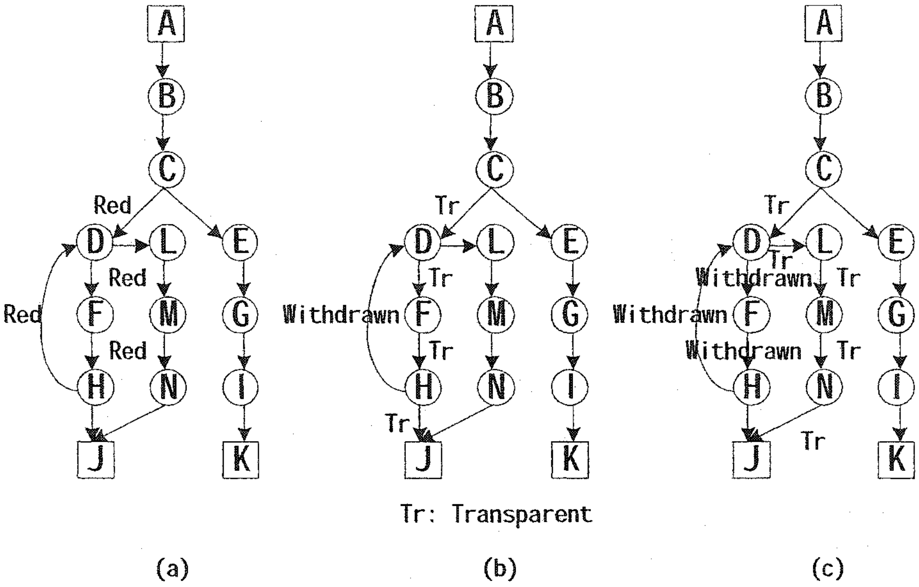


Figure 4.9 Loop prevention example 1

Consider Figure 4.10 (a) where the same multicast tree is used as Figure 4.2. When multicast traffic comes to D and E, D and E issue Red and Green thread respectively. Assuming F receives Red first, the Red is extended to F and H. It is also extended back to D because of failure of link H-J. A loop is formed, and Green thread is also marked as loop. After recovery of link H-J, H withdraws Red thread and tries to reach J by link H-J. Label is issued from J. Red thread is changed to Transparent, but Green thread is still stalled. F informs E to withdraw Green thread. See Figure 4.10 (b).

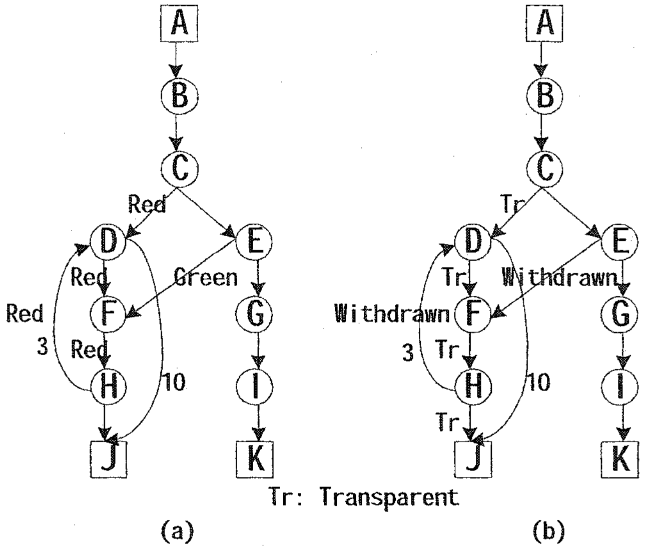


Figure 4.10 Loop prevention example 2

Figure 4.11 (a) provides loop prevention for the example in Figure 4.3 (c). Here B reaches A by Blue thread, and Blue and Red thread are changed to Transparent. The interfaces of routers still have Green thread as stalled. C informs D and D then informs B to withdraw Green thread. See Figure 4.11 (b).

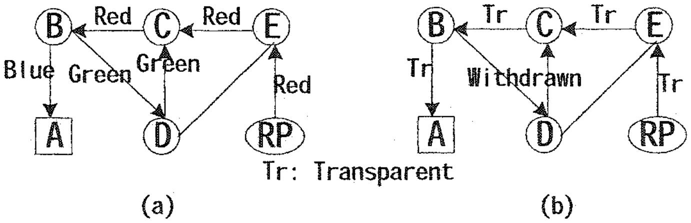


Figure 4.11 Loop prevention example 3

4.5 Multicast loop prevention - No thread mechanism

4.5.1 Basic idea

We now present a simple mechanism to detect and prevent setting up multicast MPLS LSP that can form loops in MPLS network. This mechanism uses the number of incoming interfaces of multicast traffic to detect the loop. If the number of incoming interfaces is greater than one, then based on some rules all branches except one are detected as loops and LSP setup along these paths is stalled. This is a simple but effective way to detect and prevent multicast loops in MPLS networks. The mechanism can be used with ordered downstream-on-demand label distribution. A Loop Control Block (LCB) is needed to implement this mechanism.

4.5.2 LSP setup

LSP setup message can be issued from ingress LER, root of the tree or a branch point that has a new branch. The setup message is forwarded to reach egress LER. Figure 4.12 gives examples of LSP setup. Link C-F is a new branch in Figure 4.12(c), node C therefore issues LSP setup message to F.

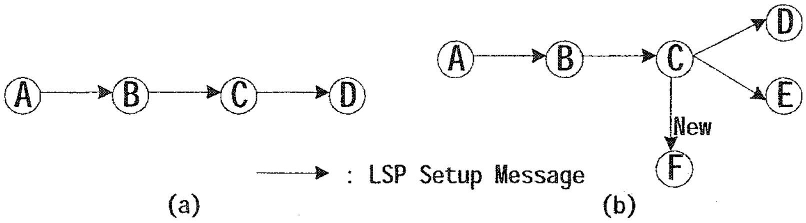


Figure 4.12 Examples of LSP setup

4.5.3 LSP setup stalling

A loop may be formed when a node gets more than one incoming interface for same FEC. This information is provided to all nodes until the nearest upstream branch point. LSP setup along the path is stalled to prevent a loop. Path setup in C-D, D-E, E-C of Figure 4.13 (a), C-D, D-E, E-B of Figure 4.13 (b) and D-C of Figure 4.13 (c) is stalled.

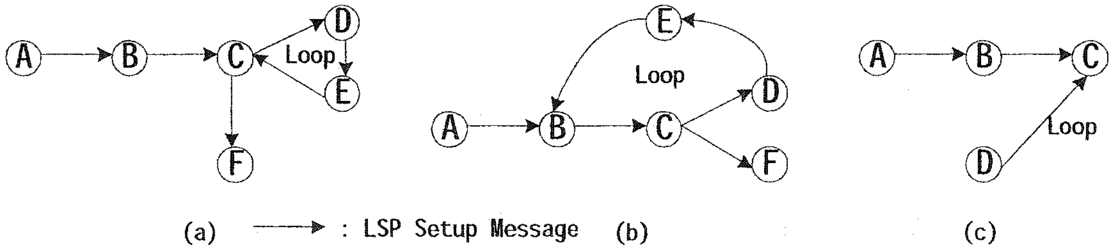


Figure 4.13 Examples of LSP setup stalling

4.5.4 Loop free LSP

When a LSP setup reaches egress LER, a loop free LSP is found. Label message can be sent in the reverse direction to LSP setup initiator. In Figure 4.14, node D is egress LER. A loop free LSP is set up between node A and D regardless of node C having other outstanding LSP setup.

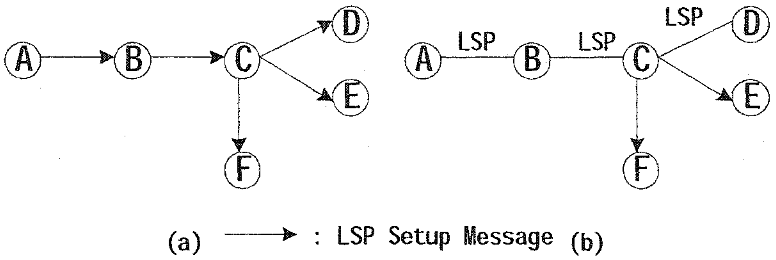


Figure 4.14 Example of loop freeness

4.5.5 Loop Control Block

In order to support loop prevention, a Loop Control Block (LCB) is needed in participating nodes. LCB maintains the following:

FEC

Incoming interfaces

Each incoming interface has the following information:

Upstream LSR node

Stalling flag

Label

Outgoing interfaces

Each outgoing interface has the following information:

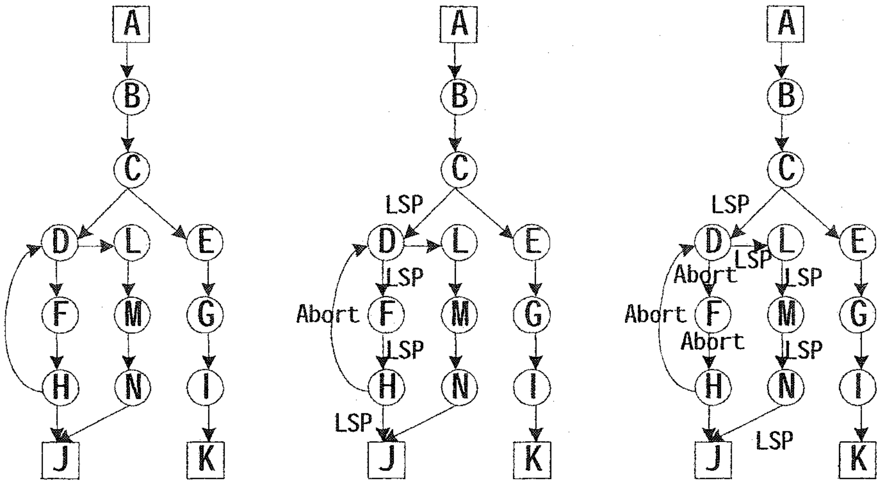
Downstream LSR node

Stalling flag

Label

4.5.6 Examples with No thread mechanism

Figure 4.15 provides loop prevention for the example in Figure 4.1 (a). When label request message comes to C, C requests label from D. This request is relayed to F and H. Due to the failure of link H-J, H requests label from D. A loop is formed, and label setup is stalled on D, F and H. After link H-J is recovered, H aborts label request to D, and tries to reach J. Label is issued since J is an egress LER,. LSP is set up in D, F and H as shown in Figure 4.15 (b). In another case, if D tries to reach J along L, M and N successfully, label mapping can go through this way. LSP is set up on these routers as shown in Figure 4.15 (c).



(a)

(b)

(c)

Figure 4.15 Loop prevention example

4.6 Other issues

In traffic-driven LSP setup, when the branches of the tree are pruned, LSRs in MPLS layer do not have knowledge about it. Multicast traffic is still forwarded to these nodes. Request-driven may be combined with traffic-driven to set up LSP. For example, if Prune message is received by a router, the route is removed in IP layer. And LSP setup corresponding to this route should also be released or aborted. This keeps the multicast traffic in MPLS layer follow the same route as that in IP layer if no MPLS is employed.

4.7 Conclusion

Multicast loops may be formed during the transient state of the network topology being changed. In IP layer, RPF is employed to prevent multicast loop forming. However, there is no such mechanism in MPLS layer to prevent multicast loops. In this Chapter, multicast loop forming in PIM-SM, PIM-DM and DVMRP based networks are discussed and specific situations are also given. Then two mechanisms are proposed to prevent multicast loops in MPLS layer. One of them is Colored thread mechanism, and another is No thread mechanism. The Colored thread proposed for multicast is an extension to the mechanism of unicast as given in [40,41]. The No thread is a new approach proposed here. These two mechanisms are analyzed in detail and examples are also provided. The mechanisms are used with ordered downstream-on-demand label distribution.

Chapter 5 Multicast Traffic Merging with QoS

If QoS is supported in MPLS networks, more than one LSPs may be needed for multicast traffic with different QoS for the same group. Number of LSPs is based on per group per QoS. In this case, one class of service can be chosen to the multicast traffic and merging can be employed to reduce the number of labels consumed by the multicast traffic. In this Chapter, two multicast traffic merging modes [42], root-initiated and leaf-initiated, are discussed in a DiffServ-supported MPLS network. Multicast routing protocols such as PIM-SM, PIM-DM and DVMRP are considered.

5.1 Problem statement

Figure 5.1 gives a network topology. In this network model, both MPLS network and stub DiffServ networks support multicast. Also DiffServ is supported in core MPLS network. Multicast traffic starts from sources and transverses multiple DiffServ networks and MPLS network to reach receivers. One multicast routing protocol is employed at one time in all networks here.

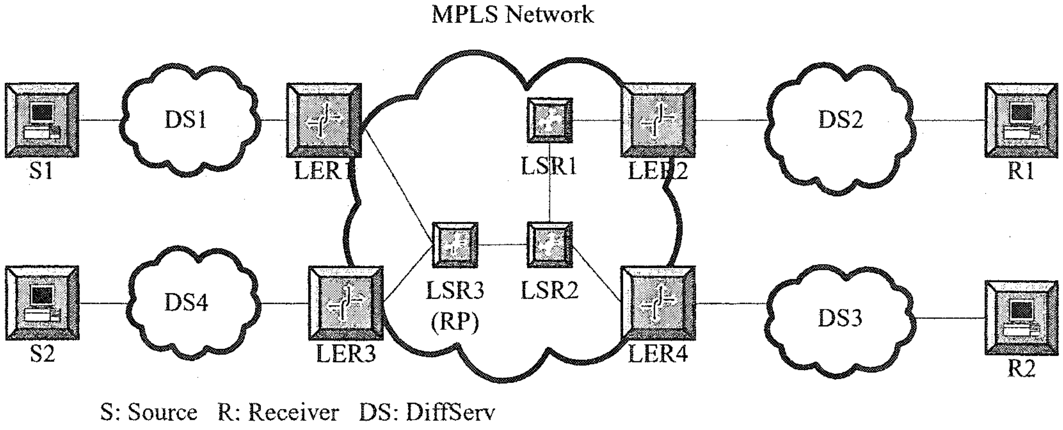


Figure 5.1 Network topology

In order to support end-to-end QoS to multicast traffic when the traffic spans multiple networks to reach receivers, interworking between MPLS and DiffServ networks needs to be discussed. The published literature in this field relates to only unicast applications. Problems arise when multicast traffic is considered. In Figure 5.1, DS1, DS2, DS3, DS4 and MPLS network support multicasting. DS1, DS2, DS3 and DS4 may use different SLAs in each domain. The issues related to supporting multicast in DiffServ-enabled MPLS network are considered here. If multicast traffic for the same group comes from different sources with different QoS, for example, in PIM-SM, the traffic merging may be needed to reduce the number of labels consumed. In this case root-initiated merging can be employed. If QoS from receiver side is considered, leaf-initiated merging can be implemented. In both cases one LSP is needed for the same group.

5.2 Merging mechanism

This Chapter presents two merging modes: root-initiated and leaf-initiated. In root-initiated merging mode, multicast traffic merging node is the root of the multicast tree or ingress node. In leaf-initiated merging mode, multicast traffic merging starts from the egress nodes towards the root of the tree or ingress node. For PIM-SM, all traffic from different sources meet at Rendezvous Point (RP), from where it is delivered to receivers. If these traffic flows have different DSCPs (Differentiated Services Codepoint), and no merging in RP is employed, separate LSPs per DSCP are needed from RP to egress LERs. If root-initiated merging is employed in RP, only one merged LSP is needed. If QoS is required from receiver side, leaf-initiated merging can be used to set up

LSP. Also one LSP is needed in this case. Root-initiated merging is mainly used in PIM-SM whereas leaf-initiated merging can be used in PIM-SM, PIM-DM and DVMRP. Root-initiated merging in PIM-DM and DVMRP is a simplified version. In PIM-SM, merging only takes place between RP and egress LERs in both cases of root-initiated and leaf-initiated merging. LSP setup between ingress LER and RP uses FEC of <source, group address> and is not affected by merging. If LSR fails to reserve required resources at any point, IP routing is used instead between the nearest upstream LSR and egress LER. The merging proposed here takes place among sources or receivers in the same multicast group. After merging, the number of labels or the number of LSPs is reduced, and the best QoS among sources or receivers is chosen. Moreover, more bandwidth with the best QoS may be required from all these traffic. The LSP setup messages are required at the beginning of LSP setup, and they are also needed when dynamic sources and receivers join/leave a group. The following scenarios specify the LSP setup process according to different multicast routing protocols (MRP) employed. We first consider PIM-SM in section 5.2.1 and then PIM-DM and DVMRP in section 5.2.2.

5.2.1 PIM-SM

Figure 5.2 provides a sample network topology considered for PIM-SM.

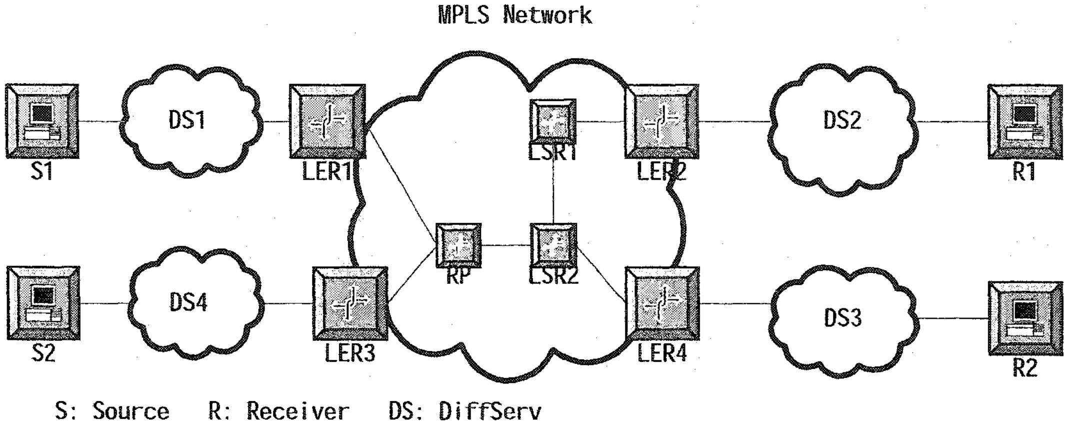


Figure 5.2 Network topology for PIM-SM

The following are the assumptions for scenario1, which considers root-initiated merging, and scenario2, which considers leaf-initiated merging:

- PIM-SM is used as multicast routing protocol
- There is one multicast group, Group1. R1 and R2 have already joined Group1.
- S1 gets AF21, Assured Forwarding of Differentiated Service, for the traffic of Group1 according to the SLA (Service-Level Agreement) between S1 and DS1 service provider.
- S2 gets AF31 for the traffic of Group1 according to the SLA between S2 and DS4 service provider.
- R1 gets AF31 for the traffic of Group1 according to the SLA between R1 and DS2 service provider.
- R2 gets AF41 for the traffic of Group1 according to the SLA between R2 and DS3 service provider.
- S2 joins Group1 later than S1.
- MPLS network supports two merging modes, root-initiated and leaf-initiated.

- CR-LDP is used as label distribution protocol.
- L-LSP is to be set up.
- LSP Trigger Strategy: traffic-driven
- Label Allocation and Distribution Scheme: downstream-on-demand
- Label Distribution Control Mode: ordered
- DiffServ Support: supported
- RP has the function of both ingress LER to map the traffic to FEC and to initiate LSP setup and egress LER.

In PIM-SM, sources of Group1, S1 and S2, register to the RP of Group1 by sending Register Message before multicasting traffic to RP via a source specific shortest-path tree (SPT). SPT (S, G) is established between source and RP after finishing source registration. RPT (*, G) is established between RP and receivers.

In the following, we consider two merging scenarios for PIM-SM. Special cases we discussed after scenario 1 and scenario 2 are the same other than the cases of receivers joining and leaving group.

5.2.1.1 Scenario1: root-initiated merging

The procedure to set up LSP where merging is root-initiated is given below:

- RP receives first multicast traffic encapsulated in PIM-SM source Register Message, and finds that the merging mode is root-initiated, that DSCP is encoded in multicast traffic and that DiffServ is supported in MPLS network, then RP starts to setup LSP. RP issues a Label Request Message (LRM) with traffic engineering parameters TLV (TE1) and DiffServ

TLV (AF21) requested by S1 to its downstream LSRs towards egress LERs. RP is to set up a LSP with FEC <*, Group1>.

- After SPT is set up, S1 sends its first multicast packet via SPT to RP, gets AF21 from DS1, and is forwarded to LER1.
- LER1 finds that this is the first multicast packet from S1 via SPT, that DiffServ is supported, that FEC is <S1, Group1>, that MRP is PIM-SM and that LER1 itself is not the RP of Group1. LER1 issues a LRM with TE1 and AF21 to be supported by the LSP to request label from downstream LSRs selected by PIM-SM. This message is propagated to RP. LER1 is to set up a LSP with FEC <S1, Group1>.
- When RP receives the LRM, it responds a Label Mapping Message (LMM) with reserved TE1 to its upstream LSR, i.e. LER1. So LSP between LER1 and RP is established.
- When the LRM initiated by RP arrives at LER2, LER2 releases a label for LSP with AF21 and returns it back with a LMM with reserved TE1 to upstream LSR1. LMM is propagated to LSR2, RP. So LSP between RP and LER2 is established.
- When the LRM initiated by RP arrives at LER4, LER4 releases a label for LSP with AF21 and returns it back with an LMM to upstream LSR2. LSR2 finds that there is no pending LRM from upstream LSR and that merging mode is root-initiated, so it does not forward this LMM upstream. LSP between LSR2 and LER4 is established.
- Multicast packets of Group1 are forwarded along LSP to LER2. The packets will go into DS2, DSCP is changed from AF21 to AF31 according to the SLA between R1 and DS2, then is forwarded to reach R1.
- Multicast packets of Group1 are forwarded along LSP to LER4. The packets will go into

DS3, DSCP is changed from AF21 to AF41 according to the SLA between R2 and DS3, then is forwarded to reach R2.

- Consider now that S2 sends traffic to Group1.
- When RP receives the traffic encapsulated in PIM-SM Register Message from Designated Router (DR) connecting directly to S2, it finds that there is a LSP that already exists for Group1 with better QoS (AF21) and that the merging mode is root-initiated. So RP decides to dynamically reserve more resources and takes AF21 for traffic of Group1. RP issues a LRM with AF21 and combined TE1 and TE2. This process is the same as the setup of new LSP. If succeeded, new LSP with new TE parameters is established.
- After S2 joins Group1, first multicast packet from S2 gets AF31 from DS4 and is forwarded to LER3. LER3 issues a LRM with TE2 and AF31 to downstream LSRs until RP. RP responds a LMM with reserved TE2 to its upstream LSP, i.e. LER3, so LSP with FEC <S2, Group1> between LER3 and RP is established.
- If S2 gets AF11 for its traffic, the LSP needs to be rebuilt because AF11 is better than AF21. The setup procedure is the same as the initial LSP setup. See the bottom part of Figure 5.3.
- If at any point, whether in setting up new LSP or modifying LSP, the specified resources are not available, a Notification Message (Ntf) is issued to its upstream LSR. The upstream LSR then aborts label request process by issuing a Label Abort Request Message (LARM) to downstream LSRs toward egress LER and forwards multicast traffic at Layer 3 to its downstream routers until egress LER (or RP). The original LSP is released starting from this LSR until egress LER in case of modifying LSP.

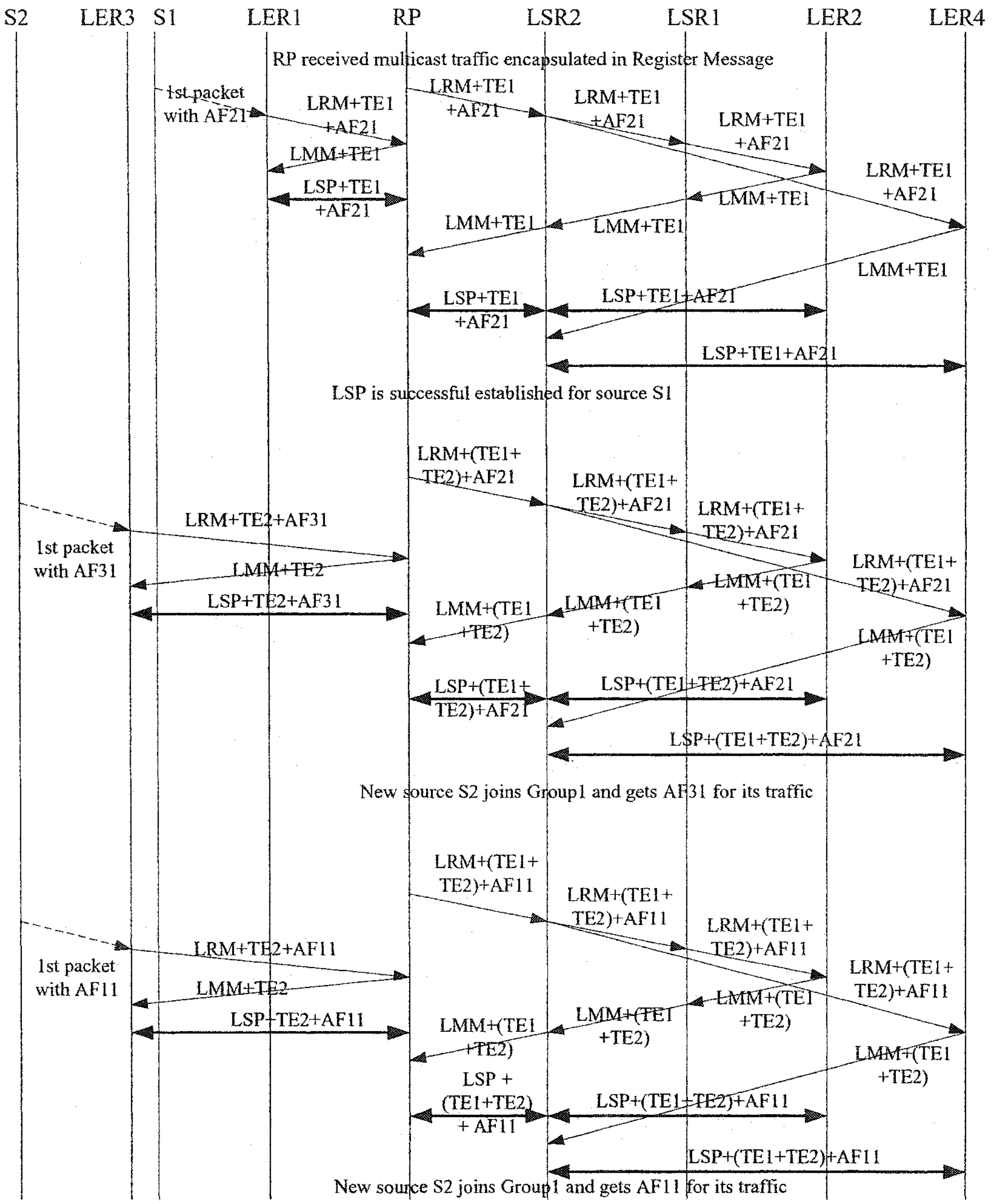


Figure 5.3 LSP setup procedure for scenario 1

Next, we consider different cases of receivers joining and leaving group and LSRs failing to reserve required resource.

Case 1: R2 leaves group

If receiver R2 leaves Group1, Prune message is sent to LER4 and LSR2. LSR2 then releases LSP between LSR2 and LER4. The corresponding sequence is shown in Figure 5.4.

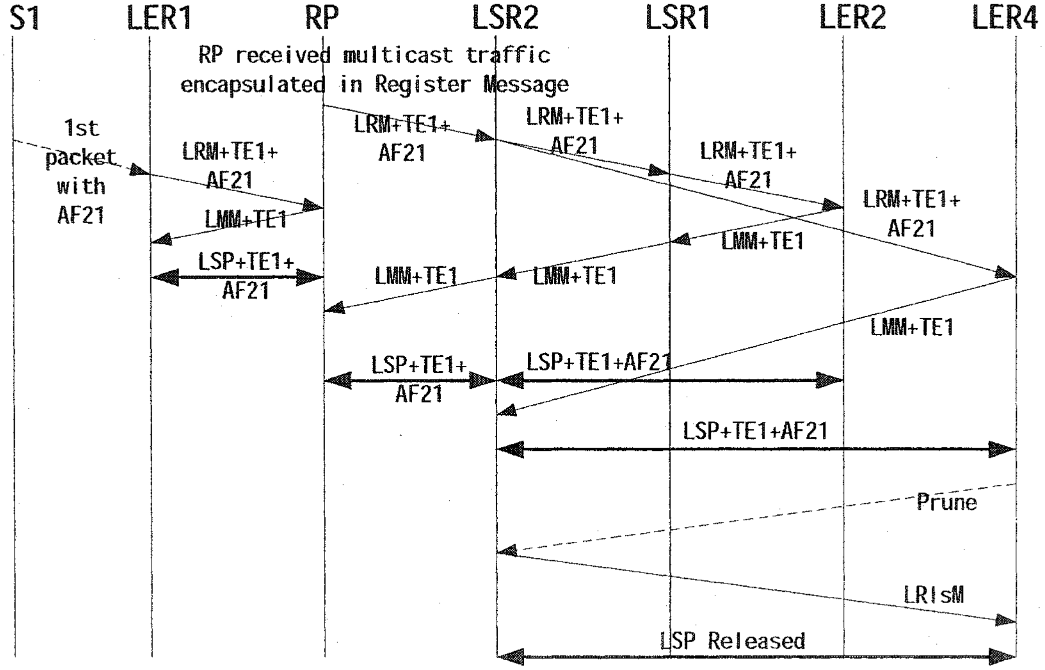


Figure 5.4 R2 leaves group

Case 2: LSR2 fails to reserve requested resources

If LSR2 fails to reserve requested resources, Notification Message is sent to RP. RP issues Label Abort Request Message to downstream LSRs until reaching LERs. Then traffic is forwarded in IP layer. The corresponding sequence is shown in Figure 5.5.

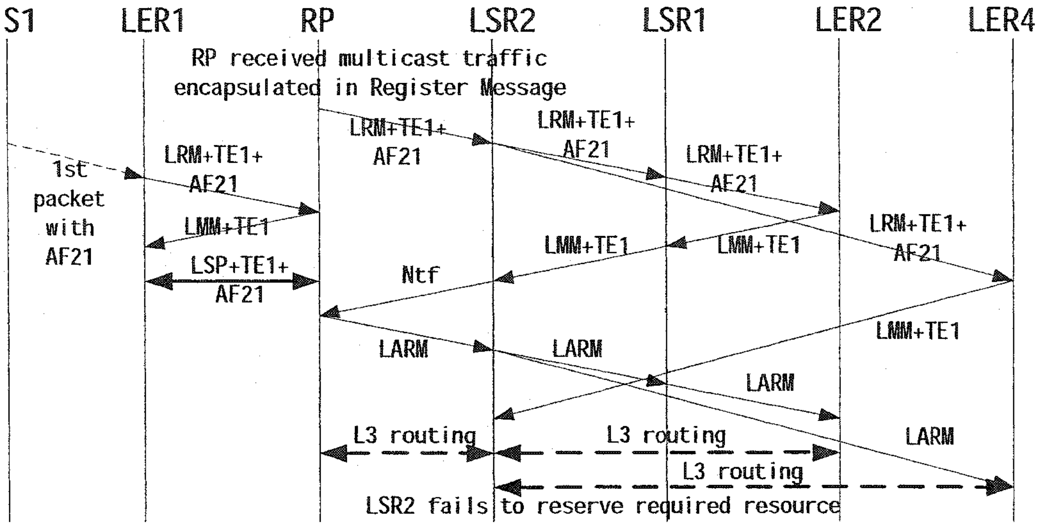


Figure 5.5 LSR2 fails to reserve requested resources

Case 3: LSR2 fails to dynamically reserve requested resources

If LSR2 fails to dynamically reserve requested resources from S1 and S2, Notification Message is sent to RP. RP issues Label Abort Request Message to downstream LSRs until reaching LERs. Then original LSP is released and traffic is forwarded in IP layer. The corresponding sequence is shown in Figure 5.6.

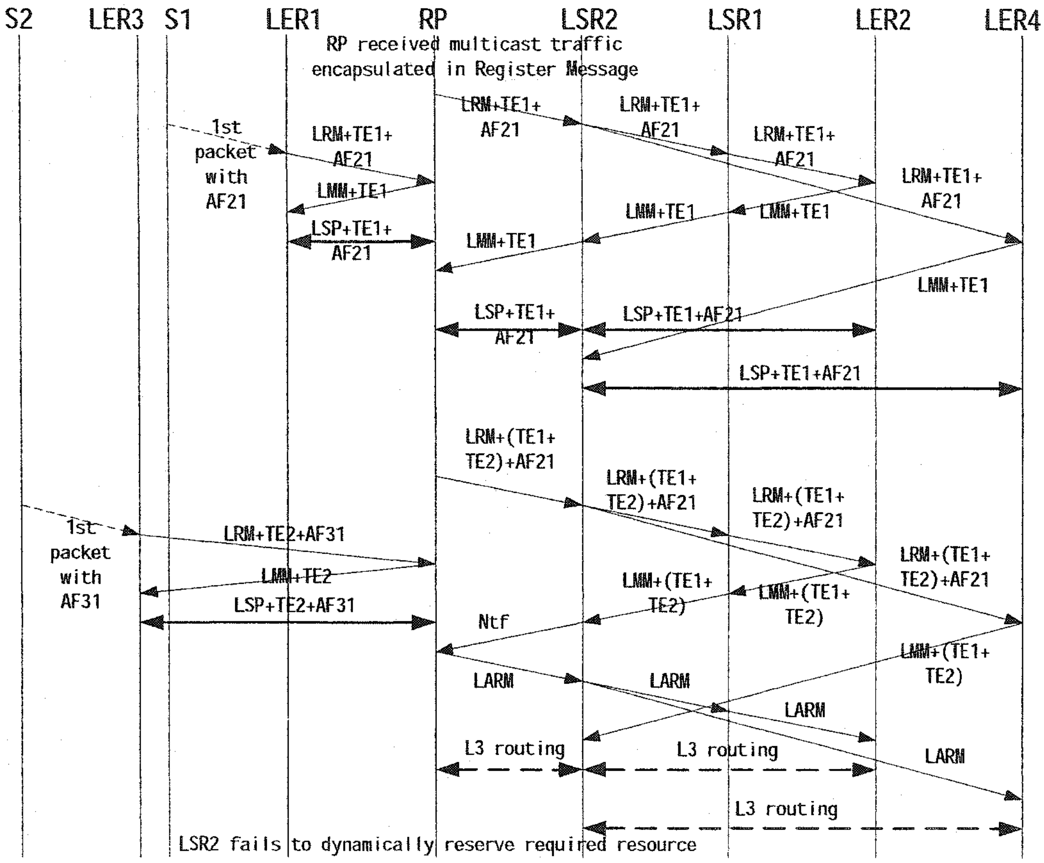


Figure 5.6 LSR2 fails to dynamically reserve requested resources

Case 4: RP fails to reserve requested resources

If RP2 fails to reserve requested resources for LSP with FEC <S1, Group1>, Notification Message is sent to LER1. LER1 issues Label Abort Request Message to downstream LSRs until reaching RP. Then traffic from S1 is forwarded to RP in IP layer. The corresponding sequence is shown in Figure 5.7.

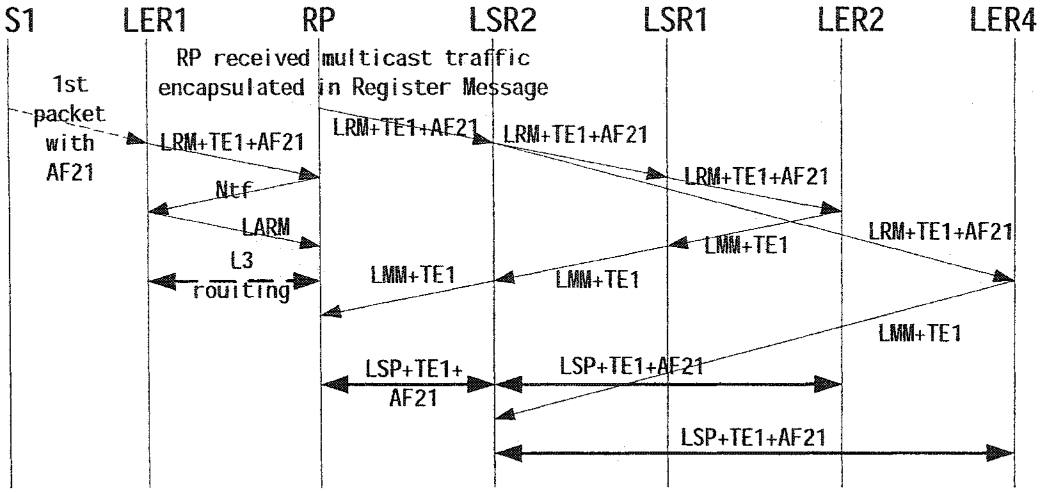


Figure 5.7 RP fails to reserve requested resources

Case 5: Receiver R3 joins group

Figure 5.8 shows a receiver R3 joins Group1.

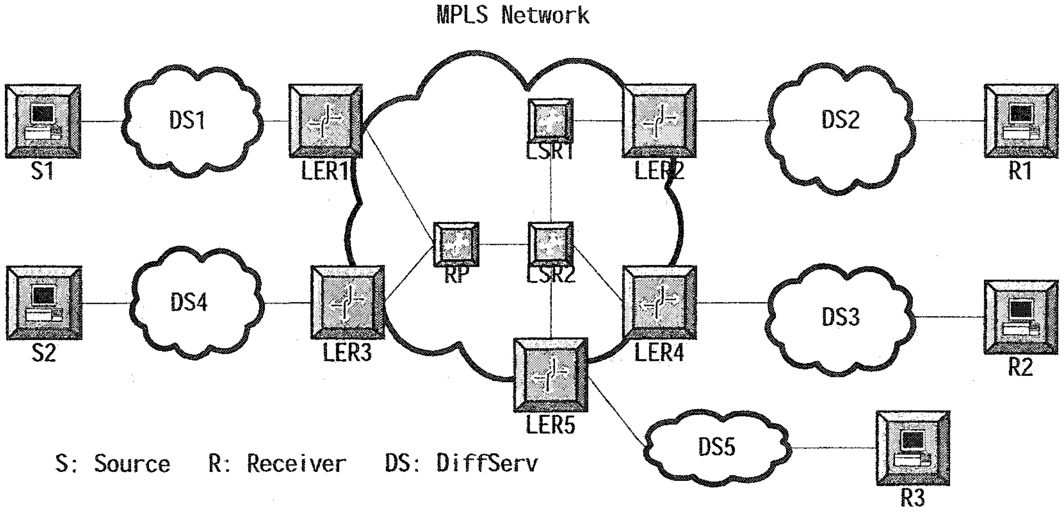


Figure 5.8 Network topology for PIM-SM with receiver R3

When a receiver R3 joins Group1, LSP between LSR2 and LER5 is established. The corresponding sequence is shown in Figure 5.9.

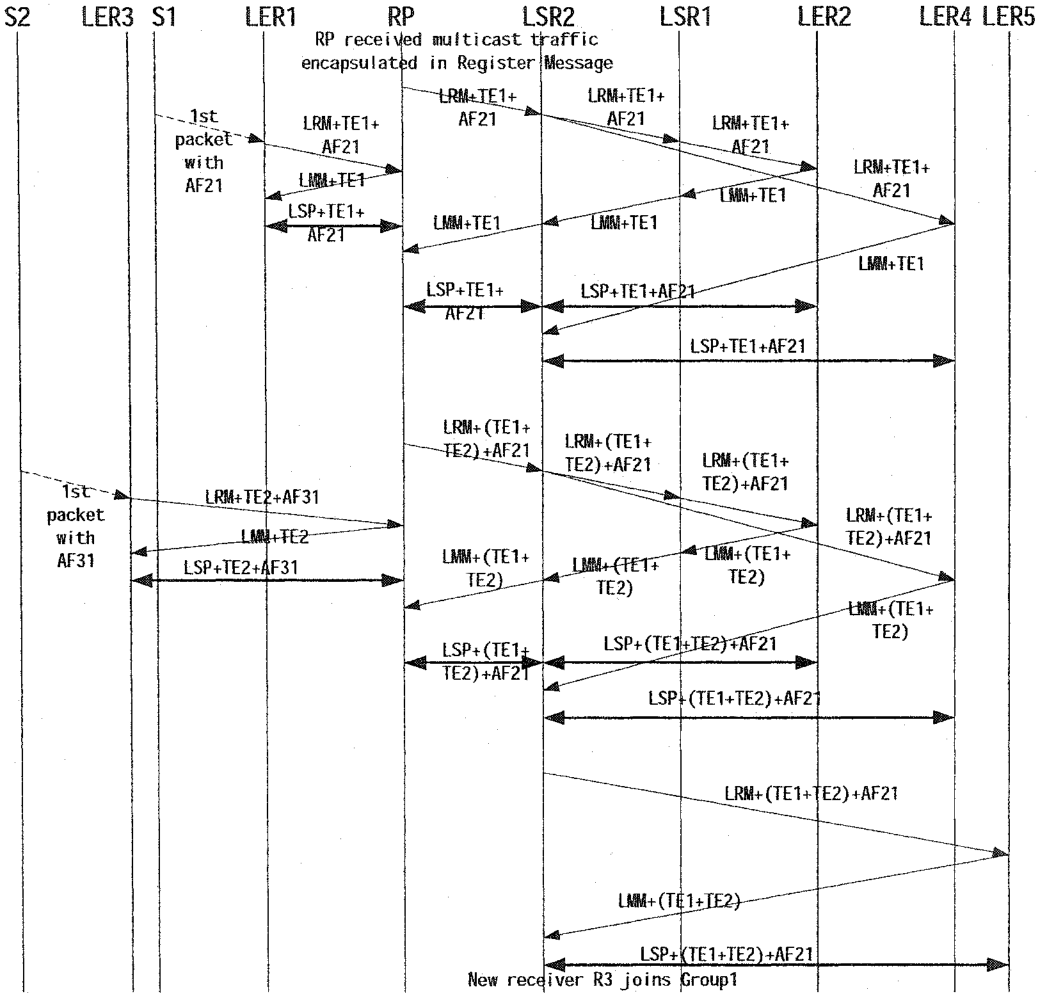


Figure 5.9 R3 joins Group 1

5.2.1.2 Scenario 2: leaf-initiated merging

The procedure to set up LSP where merging is leaf-initiated is given below:

- RP receives first multicast traffic encapsulated in PIM-SM source Register Message, and finds that the merging mode is leaf-initiated, that DSCP is encoded in multicast traffic and that DiffServ is supported in MPLS network. RP starts to setup LSP. RP issues a LRM to its downstream LSRs towards egress LERs. RP is to set up a LSP with FEC <*, Group1>.

- After SPT is set up, S1 sends its first multicast packet via SPT to RP, gets AF21 from DS1, and is forwarded to LER1.
- LER1 finds that this is the first multicast packet from S1 via SPT, that DiffServ is supported, that FEC is <S1, group1>, that MRP is PIM-SM and that LER1 itself is not RP of Group1. LER1 issues a LRM with TE1 and AF21 to request label from downstream LSRs selected by PIM-SM to set up LSP based on FEC of <S1, Group1>. This message is propagated to RP. LER1 is to set up a LSP with FEC <S1, Group1>.
- When RP receives the LRM, it responds a LMM with AF21 and reserved resource TE1 to LER1. So LSP between LER1 and RP is established.
- When the LRM initiated by RP arrives LER2, LER2 releases a label and returns it back to its upstream LSR1 with a LMM containing AF31 and traffic engineering parameters requested by receiver 1 (TEr1) that R1 should get, and the same process is propagated to RP.
- When the LRM initiated by RP arrives LER4, LER4 releases a label and returns it back to its upstream LSR2 with a LMM containing AF41 and TEr2 that R2 should get.
- After LSR2 receives the LMM from LER4, it finds that there is no pending LRM, that one LSP already exists for this group, and that merging mode is leaf-initiated. So it reserves resource with better QoS parameters chosen from combined TEr1 and TEr2 and chooses better QoS AF31 from (AF31, AF41). If successful, LSR2 sends a LMM with new TE parameters and new DS-TLV to its upstream LSR (RP).
- LSP between RP and LSR2 is modified. And LSP between LSR2 and LER4 is established.

- Consider now that S2 sends traffic to Group1.
- When RP receives the traffic encapsulated in PIM-SM Register Message from DR connecting directly to S2, it finds that there is a LSP that already exists for Group1 and that the merging mode is leaf-initiated. So RP does nothing but forwards this multicast traffic downstream existing LSP
- When first multicast packet from S2 gets AF31 from DS4 and is forwarded to LER3, LER3 issues a LRM to downstream LSRs. When RP receives the LRM, RP issues a LMM with TE2 and AF31 to its upstream LSR (LER3). If succeeded, new LSP with FEC <S2, Group1> is established.

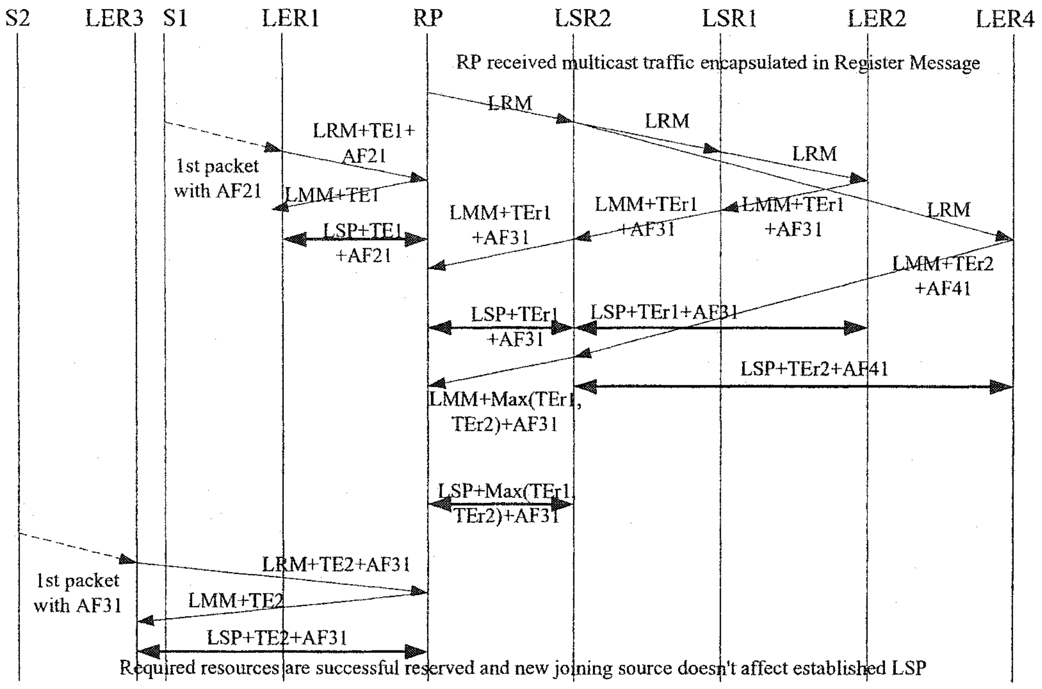


Figure 5.10 LSP setup procedure for scenario 2

Case 1: R2 leaves group

If receiver R2 leaves Group1, Prune message is sent to LER4 and LSR2. LSR2 then releases LSP between LSR2 and LER4. The corresponding sequence is shown in Figure 5.11.

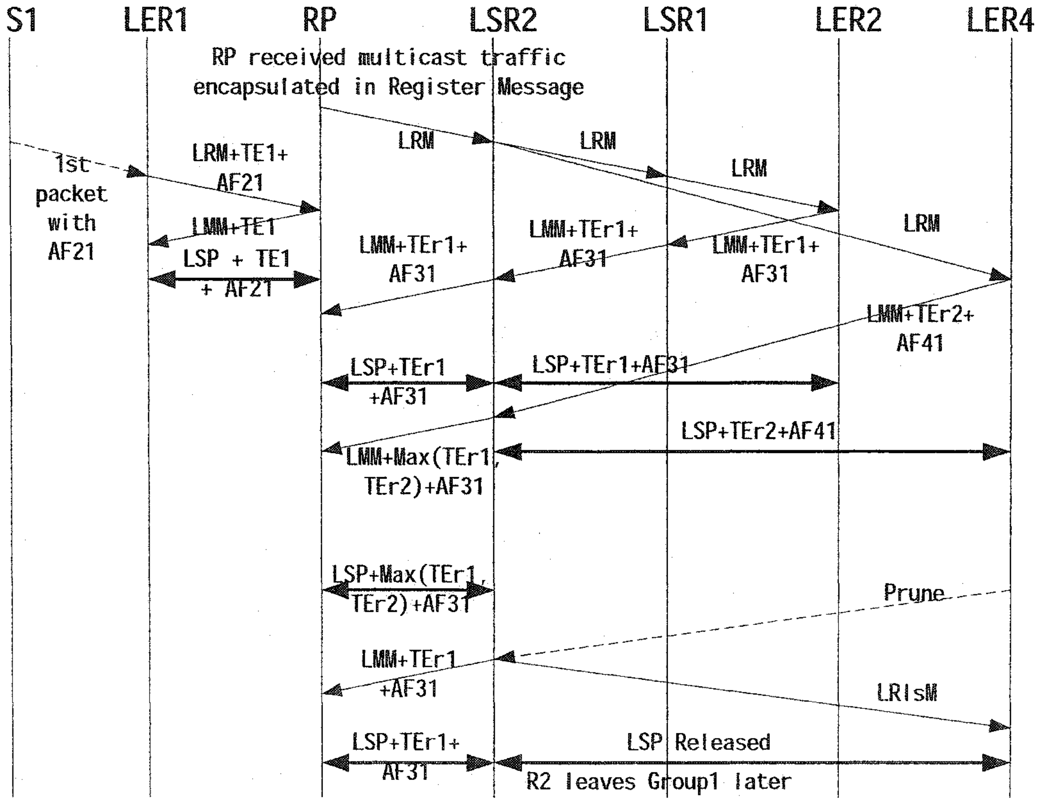


Figure 5.11 R2 leaves Group 1

Case 2: LSR1 fails to reserve requested resources

If LSR1 fails to reserve requested resources from R1, Notification Message is sent to LSR2. LSR2 issues Label Abort Request Message to downstream LSRs until reaching LER2. Then traffic is forwarded in IP layer between LSR2 and LER2. The corresponding sequence is shown in Figure 5.12.

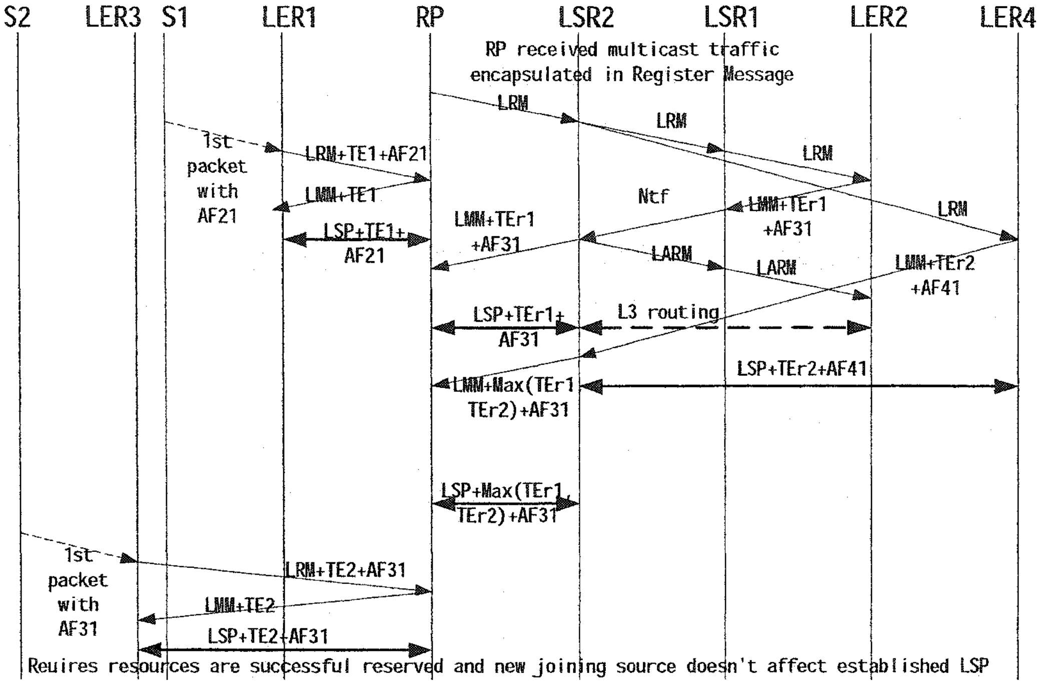


Figure 5.12 LSR1 fails to reserve requested resources

5.2.2 PIM-DM & DVMRP

Figure 5.2 provides a sample network topology considered for PIM-DM and DVMRP.

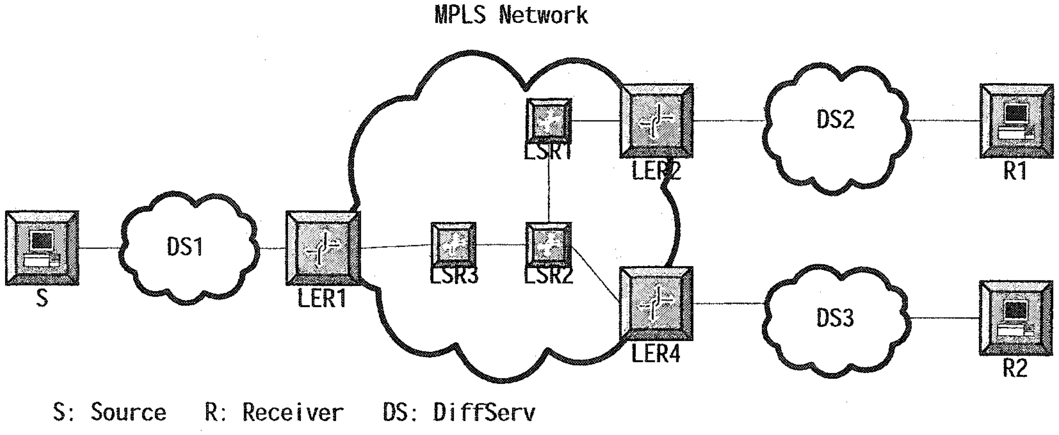


Figure 5.13 Network topology for PIM-DM and DVMRP

The following are the assumptions for scenario3 that considers root-initiated merging and scenario4 that considers leaf-initiated merging:

- PIM-DM or DVMRP is used as multicast routing protocol
- There is one multicast group, Group1.
- S gets AF21 for the traffic of Group1 according to the SLA between S and DS1 service provider
- R1 gets AF31 for the traffic of Group1 according to the SLA between R1 and DS2 service provider.
- R2 gets AF41 for the traffic of Group1 according to the SLA between R2 and DS3 service provider.
- MPLS network supports two merging modes, root-initiated and leaf-initiated.
- CR-LDP is used as label distribution protocol.
- L-LSP is to be set up.
- LSP Trigger Strategy: traffic-driven
- Label Allocation and Distribution Scheme: downstream-on-demand
- Label Distribution Control Mode: ordered
- DiffServ Support: supported

In the following, we consider two merging scenarios for PIM-DM and DVMRP. Special cases we discussed after scenario 3 and scenario 4 are the same other than the cases of receivers joining and leaving group.

5.2.2.1 Scenario3: root-initiated Merging

The procedure to set up LSP where merging is root-initiated is given below:

- S sends its first multicast packet to Group1, gets AF21 from DS1, and is forwarded to LER1.
- LER1 finds that this is the first multicast packet from S via SPT, that MRP is PIM-DM or DVMRP, that DiffServ is supported, that FEC is <S, Group1> and that the merging mode is root-initiated, LER1 issues a LRM with TE1 and AF21 to be supported by the LSP to request label from downsteam LSRs selected by MRP. This message is propagated to all egress LERs. LER1 is to set up a LSP with FEC <S, Group1>.
- When the LRM arrives at LER2, LER2 releases a label and returns it back with a LMM with reserved TE1 to upstream LSR1. The LMM is propagated to LSR2, LSR3 and LER1. So LSP between LER1 and LER2 is established.
- When the LRM arrives at LER4, LER4 releases a label and returns it back with a LMM with reserved TE1 to upstream LSR2. LSR2 finds that there is no pending LRM from upstream LSR and that merging mode is root-initiated, so it does not forward this LMM upstream. LSP between LSR2 and LER4 is established.
- Multicast packets of Group1 are forwarded along LSP to LER2. The packets will go into DS2, DSCP is changed from AF21 to AF31 according to the SLA between R1 and DS2, then is forwarded it to reach R1.
- Multicast packets of Group1 are forwarded along LSP to LER4. The packets will go into DS3, DSCP is changed from AF21 to AF41 according to the SLA between R2 and DS3,

then is forwarded it to reach R2.

- If LER4 fails to reserve required resources, an Ntf message is sent to LSR2. LSP setup is aborted between LSR2 and LER4, and traffic is routed in L3.
- Assume there is no receiver actively connecting to DS3, a Prune message is forwarded to LER4, and to LSR2. After receiving Prune message, LSR2 issues a Label Release Message (LRlsM) to downstream LER4. So LSP between LSR2 and LER4 is released.
- If a receiver joins Group1 by DS3, LSR2 will receive a Graft message. LSR2 re-issues a LRM with TE1 and AF21 to LER4 and gets response from LER4 by a LMM with reserved TE1, so LSP between LSR2 and LER4 is re-established.
- Or if multicast traffic floods the network again, LSR2 re-issues a LRM with TE1 and AF21 to LER4 and gets response from LER4 by a LMM, so LSP between LSR2 and LER4 is re-established.

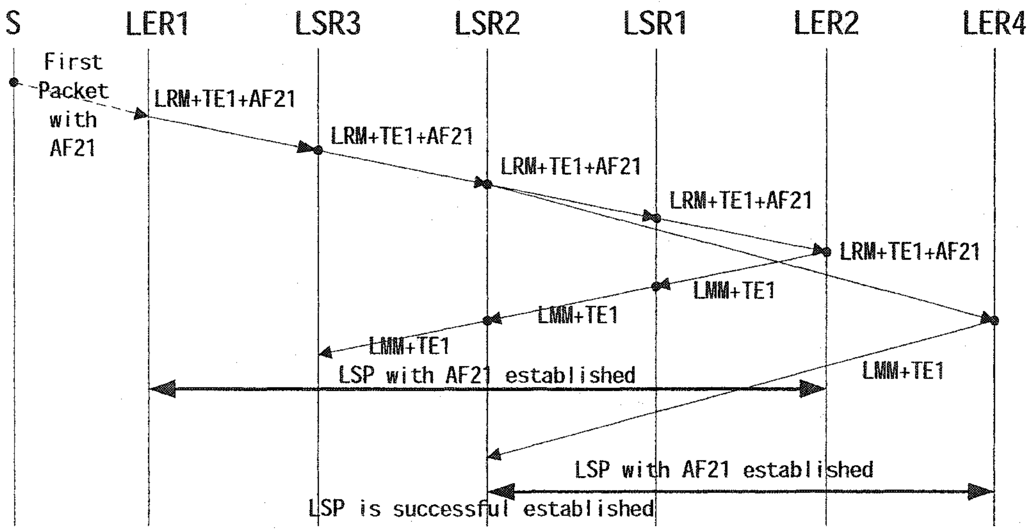


Figure 5.14 LSP setup procedure for scenario 3

Next, we consider different cases of receivers joining and leaving group and LSRs failing to reserve required resource.

Case 1: LER4 fails to reserve requested resources

If LER4 fails to reserve requested resources, Notification Message is sent to LSR2. LSR2 issues Label Abort Request Message to downstream LSRs until reaching LER4. Then traffic is forwarded in IP layer between LSR2 and LER4. The corresponding sequence is shown in Figure 5.15.

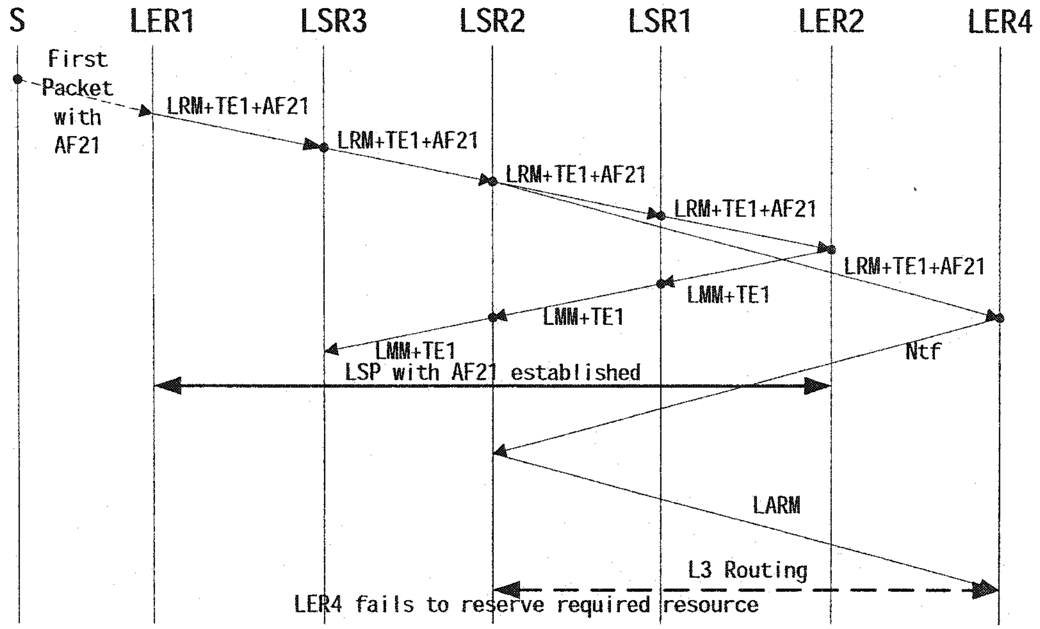


Figure 5.15 LER4 fails to reserve requested resources

Case 2: R2 leaves group, joins group again or flooding again

If receiver R2 leaves Group1, Prune message is sent to LER4 and LSR2. LSR2 then releases

LSP between LSR2 and LER4.

If R2 joins Group1 again, Graft message is sent to LER4 and LSR2. LSR2 then reestablishes LSP between LSR2 and LER4.

The corresponding sequence is shown in Figure 5.16.

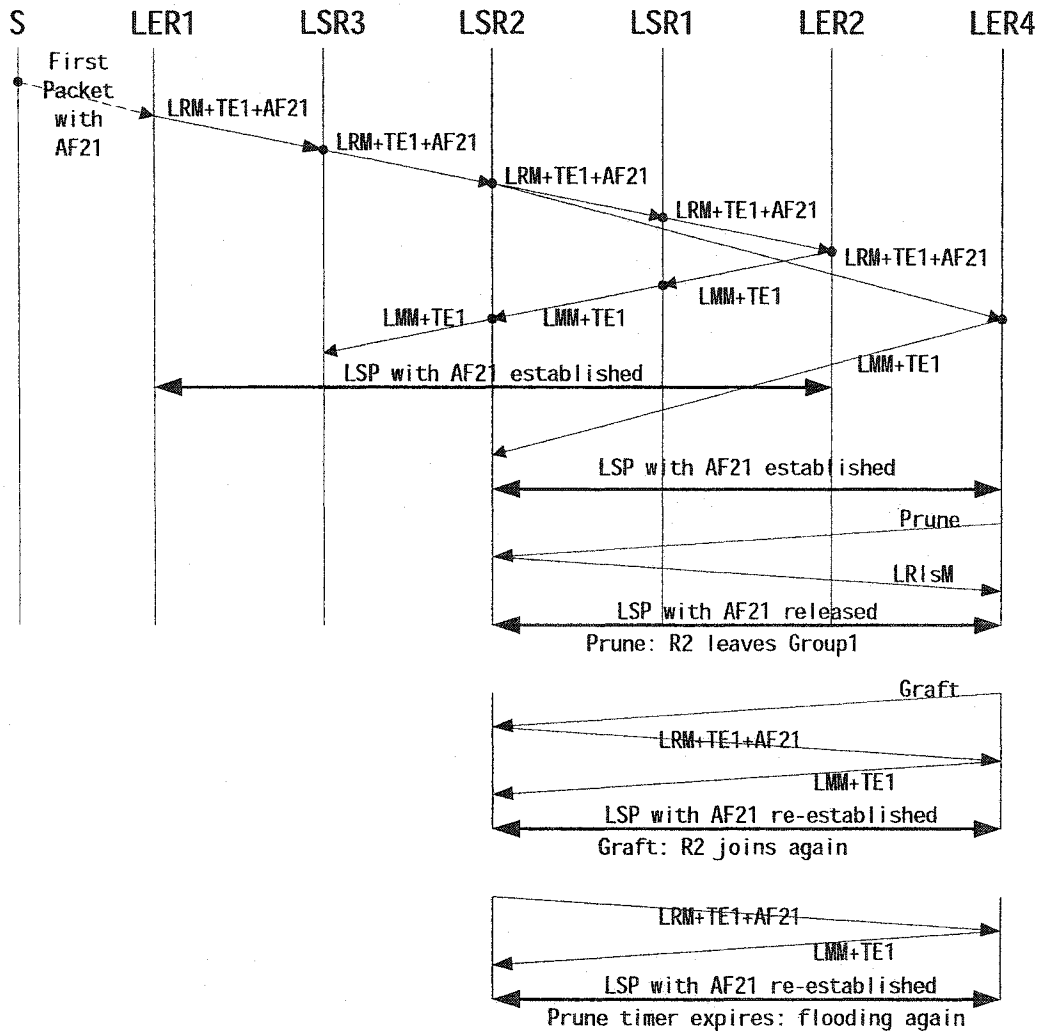


Figure 5.16 R4 leaves, rejoins Group 1

5.2.2.2 Scenario 4: leaf-initiated merging

The procedure to set up LSP where merging is leaf-initiated is given below:

- S sends its first multicast packet to Group1, gets AF21 from DS1, and is forwarded to LER1.
- LER1 finds that this is the first multicast packet from S, that MRP is PIM-DM or DVMRP and that the merging mode is leaf-initiated, LER1 issues a LRM to request label from downstream LSRs selected by MRP. This message is propagated to all egress LERs. LER1 is to set up a LSP with FEC <S, Group1> as PIM-DM or DVMRP is used.
- When the LRM arrives LER2, LER2 gives a label and returns it back to its upstream LSR1 with a LMM containing AF31 and TEr1 that R1 should get, and propagates toward LER1. LSP between LER1 and LER2 is established.
- When the LRM arrives LER4, LER4 releases a label and returns it back to its upstream LSR2 with a LMM containing AF41 and TEr2 that R2 should get.
- After LSR2 receives the LMM from LER4, it finds that there is no pending LRM, that one LSP already existed for this group, and that merging mode is leaf-initiated. So it reserves resource with better QoS parameters chosen from combined TEr1 and TEr2 and chooses better QoS AF31 from (AF31, AF41). If successful, LSR2 sends a LMM with new TE parameters and new DS-TLV to its upstream LSR.
- LSP between LER1 and LSR2 is modified, and LSP between LSR2 and LER4 is established.

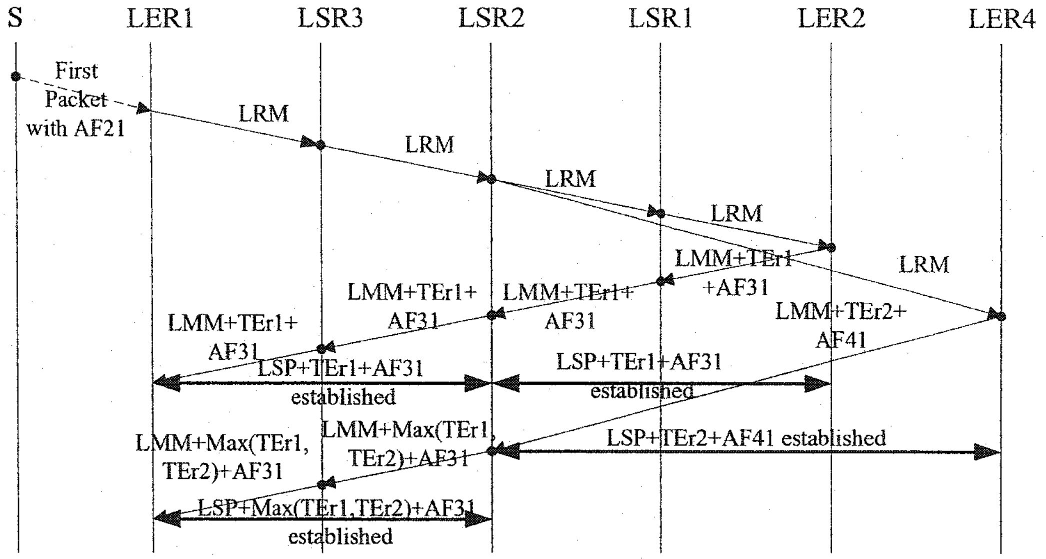


Figure 5.17 LSP setup procedure for scenario 4

Next, we consider a case of receivers leaving group.

Case 1: R2 leaves group

If receiver R2 leaves Group1, Prune message is sent to LER4 and LSR2. LSR2 then releases LSP between LSR2 and LER4. The corresponding sequence is shown in Figure 5.18.

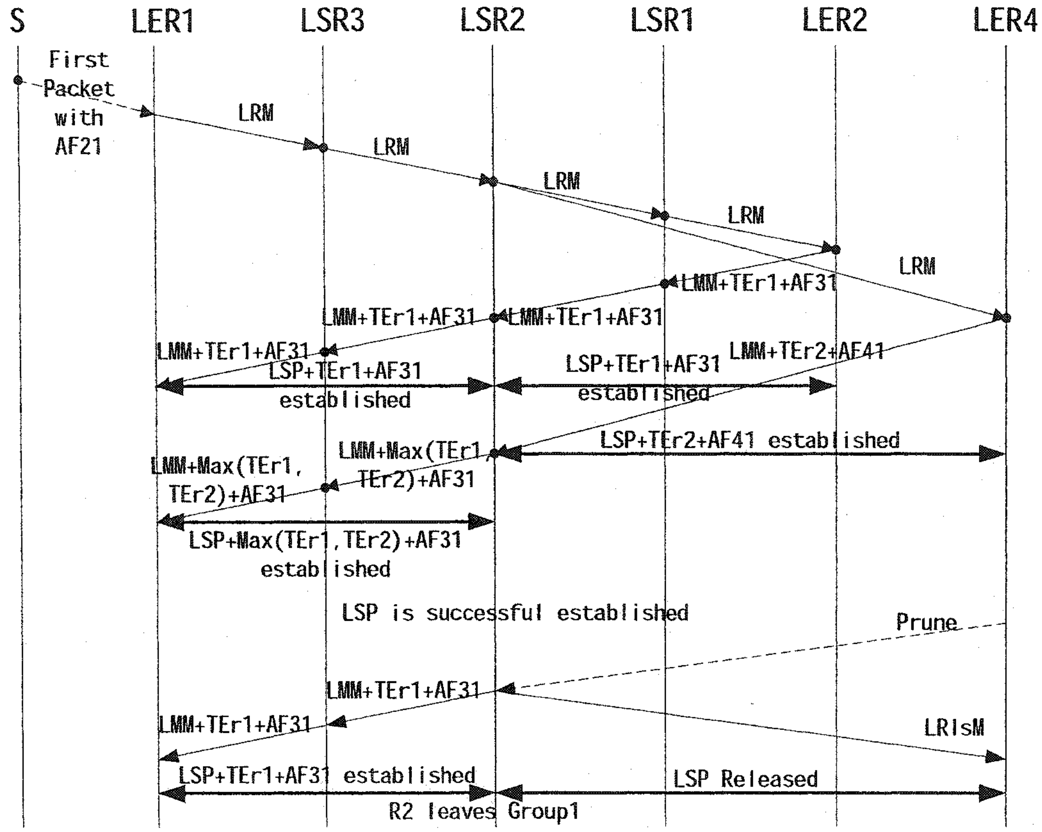


Figure 5.18 R2 leaves Group 1

5.3 LDP Messages

In the following, LSP setup messages are shown here.

Figure 5.19 shows the Label Request Message. Traffic TLV and DiffServ TLV are added as optional parameters to support traffic engineering and DiffServ respectively. They are only included in LRM in case of root-initiated merging.

0										1										2										3	
0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1
0	Label Request (0x0401)															Message Length															
Message ID																															
FEC TLV																															
LSPID TLV																															
Traffic TLV (Optional)																															
DiffServ TLV (Optional)																															

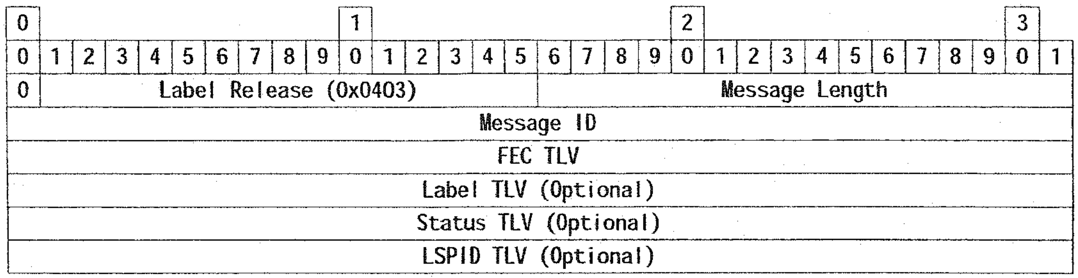
Figure 5.19 Label Request Message

Figure 5.20 shows the Label Mapping Message. LSPID TLV, Traffic TLV and DiffServ TLV are added as optional parameters to support traffic engineering and DiffServ. LSPID TLV and Traffic TLV can be included in both root-initiated merging and leaf-initiated merging. DiffServ TLV is only included in leaf-initiated merging.

0										1										2										3	
0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1
0	Label Mapping (0x0400)															Message Length															
Message ID																															
FEC TLV																															
Label TLV																															
Label Request Message ID TLV																															
LSPID TLV (Optional)																															
Traffic TLV (Optional)																															
DiffServ TLV (Optional)																															

Figure 5.20 Label Mapping Message

Figure 5.21 shows the Label Release Message. Label TLV is suggested in [35]. LSPID TLV is added here to make it clear which LSP is to be released.

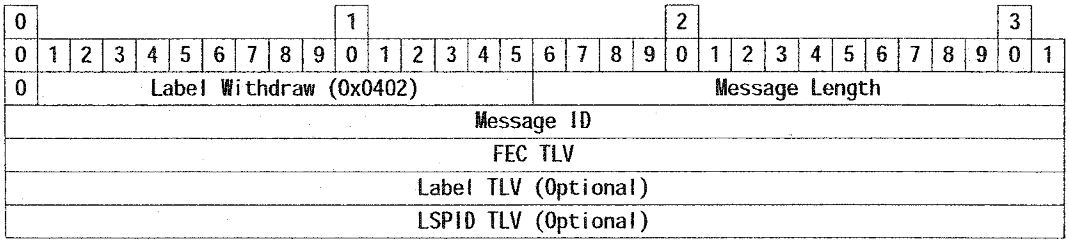


0	1	2	3
0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9	0 1

0	Label Release (0x0403)	Message Length
Message ID		
FEC TLV		
Label TLV (Optional)		
Status TLV (Optional)		
LSPID TLV (Optional)		

Figure 5.21 Label Release Message

Figure 5.22 shows the Label Withdraw Message. Label TLV is suggested in [35]. LSPID TLV is added here to make it clear which LSP is to be withdrawn.

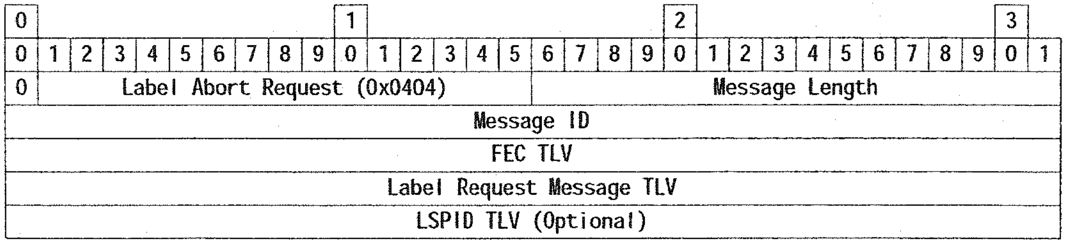


0	1	2	3
0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9	0 1

0	Label Withdraw (0x0402)	Message Length
Message ID		
FEC TLV		
Label TLV (Optional)		
LSPID TLV (Optional)		

Figure 5.22 Label Withdraw Message

Figure 5.23 shows the Label Abort Request Message. LSPID TLV is added here to make it clear which LSP is to be aborted.



0	1	2	3
0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9	0 1

0	Label Abort Request (0x0404)	Message Length
Message ID		
FEC TLV		
Label Request Message TLV		
LSPID TLV (Optional)		

Figure 5.23 Label Abort Request Message

Figure 5.24 shows the Notification Message. Traffic TLV and DiffServ TLV are added as optional parameters to support traffic engineering and DiffServ respectively. FEC TLV is added to indicate which FEC is related to this notification message.

<table>
<tr><td>0</td><td colspan="9"></td><td>1</td><td colspan="9"></td><td>2</td><td colspan="9"></td><td>3</td><td></td></tr>
<tr><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>0</td><td>1</td></tr>
<tr><td>0</td><td colspan="15">Notification (0x0001)</td><td colspan="16">Message Length</td></tr>
<tr><td colspan="32">Message ID</td></tr>
<tr><td colspan="32">Status TLV</td></tr>
<tr><td colspan="32">FEC TLV (Optional)</td></tr>
<tr><td colspan="32">DiffServ TLV (Optional)</td></tr>
<tr><td colspan="32">Traffic TLV (Optional)</td></tr>
</table>

Figure 5.24 Notification Message

Figure 5.25 and 5.26 show two new FECs. Figure 5.25 is for group address FEC and Figure 5.26 is for source and group address pair FEC.

<table>
<tr><td>0</td><td colspan="9"></td><td>1</td><td colspan="9"></td><td>2</td><td colspan="9"></td><td>3</td><td></td></tr>
<tr><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>0</td><td>1</td></tr>
<tr><td colspan="8">FEC Element Type</td><td colspan="16">Address Family</td><td colspan="8">Length</td></tr>
<tr><td colspan="32">Multicast Group Address</td></tr>
</table>

Figure 5.25 (*, G) FEC element

<table>
<tr><td>0</td><td colspan="9"></td><td>1</td><td colspan="9"></td><td>2</td><td colspan="9"></td><td>3</td><td></td></tr>
<tr><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>0</td><td>1</td></tr>
<tr><td colspan="8">FEC Element Type</td><td colspan="16">Address Family</td><td colspan="8">Length</td></tr>
<tr><td colspan="32">Source Address</td></tr>
<tr><td colspan="32">Multicast Group Address</td></tr>
</table>

Figure 5.26 (S, G) FEC Element

Figure 5.27 shows the FEC TLV format.

<table>
<tr><td>0</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>1</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>2</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>3</td><td></td></tr>
<tr><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>0</td><td>1</td></tr>
<tr><td>0</td><td>0</td><td colspan="14">FEC (0x0100)</td><td colspan="16">Length</td></tr>
<tr><td colspan="32">FEC Element</td></tr>
</table>

Figure 5.27 FEC TLV

Figure 5.28 shows the DiffServ TLV format [35].

<table>
<tr><td>0</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>1</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>2</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>3</td><td></td></tr>
<tr><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>0</td><td>1</td></tr>
<tr><td>U</td><td>F</td><td colspan="14">Type = PSC (0x901)</td><td colspan="16">Length</td></tr>
<tr><td>T</td><td colspan="15">Reserved</td><td colspan="16">PSC</td></tr>
</table>

Figure 5.28 DiffServ TLV

T: 1 bit

LSP Type. This is set to 1 for an L-LSP

Reserved: 15 bits

This field is reserved. It must be set to zero on transmission and must be ignored on receipt.

PSC: 16 bits

The PSC indicates that a PHB Scheduling Class to be supported by the LSP.

5.4 Rules for setting up LSP

5.4.1 PIM-SM as multicast routing protocol

RP can map FEC to multicast packets and issue LRM to downstream LSRs based on one of the following conditions (RP acts as ingress LER here):

- RP receives multicast traffic encapsulated in PIM-SM Register Message from Designated Router (DR) connecting directly to source for a multicast group and there is/are active receiver(s) for this group.
- A (*, G) Join Message of PIM-SM is received by RP and there is at least one active source for this group.

When RP receives a LRM from its upstream LSR for a multicast group, RP responds immediately with LMM to its upstream LSR to build LSP between ingress LER of this source and RP (RP acts as egress LER here).

FEC-TLV may be included in Notification Message.

Each LSR can issue Label Release Message (LRlsM) when a LSR receives a Prune Message of PIM-SM from its outgoing interface (oif) for this group, and this LSR still has other oif(s) which is/are active for this group after pruning this oif.

Each LSR must propagate LRlsM and LARM to its downstream LSRs until reaching egress LERs or RP.

In leaf-initiated merging mode, LSR may issue more than one of LMM to its upstream LSR if all of the following conditions are satisfied:

- This LSR receives LMM from its downstream LSR to its pending LRM and there is no pending LRM to its upstream LSR.
- The merging mode is leaf-initiated.
- The LSPID is the same as that of an existing LSPID.
- Traffic parameters or/and QoS need to be changed between this LSR and its upstream LSR.

This LMM will be forwarded to upstream LSRs until reaching RP.

In leaf-initiated merging mode, LRM does not include TE-TLV and DS-TLV whereas Notification Message and LMM must include TE-TLV and DS-TLV.

5.4.2 PIM-DM or DVMRP as multicast routing protocol:

Each LSR can issue Label Release Message (LRlsM) when a LSR receives a Prune Message of PIM-SM from its outgoing interface (oif) for this group, and this LSR still has other oif(s) which is/are active for this group after pruning this oif.

Each LSR must propagate LRlsM and LARM to its downstream LSRs until reaching egress LERs or RP.

FEC-TLV may be included in Notification Message.

In leaf-initiated merging mode, LSR may issue more than one of LMM to its upstream LSR if all of the following conditions are satisfied:

- This LSR receives LMM from its downstream LSR to its pending LRM and there is no pending LRM to its upstream LSR.
- The merging mode is leaf-initiated.

- The LSPID is the same as that of an existed LSPID.
- Traffic parameters or/and QoS need to be changed between this LSR and its upstream LSR.

This LMM will be forwarded to upstream LSRs until reaching ingress LER.

In leaf-initiated merging mode, LRM does not include TE-TLV and DS-TLV whereas Notification Message and LMM must include TE-TLV and DS-TLV.

5.5 Conclusion

In this Chapter, two multicast traffic merging modes [42], root-initiated and leaf-initiated, are proposed in a DiffServ-supported MPLS network. The merging proposed here takes place among sources or receivers in the same multicast group. Multicast routing protocols such as PIM-SM, PIM-DM and DVMRP are considered. Root-initiated merging is mainly used in PIM-SM whereas leaf-initiated merging can be used in PIM-SM, PIM-DM and DVMRP. Root-initiated merging in PIM-DM and DVMRP is a simplified version. Scenarios give detailed LSP setup procedures. By merging multicast traffic with different DSCPs, one LSP per multicast group is needed. Both the number of labels and control messages are reduced in participating MPLS routers. The best QoS among sources or receivers is chosen. Moreover, more bandwidth with the best QoS may be required from all these traffic.

Chapter 6 Conclusions

Internet applications such as video conferencing, distance learning and network gaming, not only require QoS but also use multicasting technology. With the integration of multicasting and MPLS, their performance can be further improved. Chapter 2 of this thesis overviews MPLS, multicasting, IntServ and DiffServ to support multicast in IP networks. The interworking of these technologies within core and edge networks is discussed.

RFC3353 gives an overview of IP multicast in an MPLS environment. This thesis analyzes multicast aggregation and loop prevention in MPLS networks mentioned in RFC3353. The possible solutions to these two problems are proposed. Dynamic multicast aggregation and multicast loop prevention in MPLS networks are presented. Multicast traffic merging with different QoS in MPLS is also discussed.

Multicast aggregation in MPLS networks is an open topic in RFC3353. The aggregation takes place among different multicast groups. Multicast aggregation in IP layer is different from that in MPLS network. Aggregating multicast traffic in IP layer is to reduce multicast states in IP routers. In MPLS, network aggregation is FEC aggregation where IP header is not affected. Multicast aggregation in MPLS network has advantages over that in IP layer. A mechanism to dynamically aggregate multicast traffic in MPLS network is proposed in this thesis. There is a tradeoff between label reduction and bandwidth usage. Both source specific tree and shared tree are considered as network topologies. New FEC is introduced and the related advantages are also covered in the Chapter. The dynamic multicast aggregation proposed can also be implemented in DiffServ-enabled MPLS networks.

Another topic left for future study in RFC3353 is multicast loop prevention in MPLS networks. Multicast loops may be formed during the transient state of the network topology being changed. In IP layer, RPF is employed to prevent multicast loop forming. However, there is no such mechanism in MPLS layer to prevent multicast loops. In this thesis, multicast loop forming in PIM-SM, PIM-DM and DVMRP based networks are discussed and specific situations are also given. Then two mechanisms are proposed to prevent multicast loops in MPLS layer. One of them is Colored thread mechanism, and another is No thread mechanism. The Colored thread proposed for multicast is an extension to the mechanism of unicast as given in [40,41]. The No thread is a new approach proposed here. These two mechanisms are analyzed in detail and examples are also provided. The mechanisms are used with ordered downstream-on-demand label distribution.

Multicast traffic merging is the last topic in this thesis. Two multicast traffic merging mechanisms, root-initiated and leaf-initiated, are proposed in a DiffServ-supported MPLS network. The merging mechanisms proposed here takes place among sources or receivers in the same multicast group. Multicast routing protocols such as PIM-SM, PIM-DM and DVMRP are considered. Root-initiated merging is mainly used in PIM-SM whereas leaf-initiated merging can be used in PIM-SM, PIM-DM and DVMRP. Root-initiated merging in PIM-DM and DVMRP is a simplified version. Scenarios give detailed LSP setup procedures. By implementing merging to multicast traffic with different DSCPs, one LSP per multicast group is needed. Both the number of labels and control messages are reduced in participating MPLS routers. The best QoS among sources or receivers is chosen. Moreover, more bandwidth with the best QoS may be required from all these traffic.

References

[1] Adriano Pezzuto, "What is Quality of Service", www.qosmagazine.net

[2] Eric C. Rosen, Arun Viswanathan, Ross Callon, "Multiprotcol Label Switching Architecture", RFC3301, January 2001

[3] Bruce Davie, Yakov Rekhter, "MPLS Technology and Applications", Morgan Kaufmann Publisher, ISBN: 1-55860-656-4

[4] F. Le Faucheur, "IETF Multiprotocol Label Switching (MPLS) Architecture", ICATM-98, 1998 1st IEEE International Conference on ATM, pp 6-15

[5] J. Lawrence, "Designing Multiprotocol Label Switching Networks", IEEE Communications Magazine, Volume: 39 Issue: 7, July 2001, pp 134-142

[6] R. Braden, D. Clark, S. Shenker, "Integrated Services in the Internet Architecture: an overview", RFC1633, June 1994

[7] S. Blake, et al, "An architecture for Differentiated Services", RFC2475, December. 1998

[8] Loa Andersson, et al, "LDP Specification", RFC3036, January 2001

[9] Beau Williamson, "Developing IP Multicast Networks Volume 1", Cisco Press, ISBN: 1-57870-077-9

[10] B. Baurens, et al, "Cooperative Environments for Distributed Systems Engineering", Chapter 6: Communications, Springer-Verlag, 2001, ISBN: 3-540-43083-0

[11] Bill Fenner, Mark Handley, Hugh Holbrook, Isidor Kouvelas, "Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)", draft-ietf-pim-sm-v2-new-07.txt, March 2, 2003

[12] D. Waitzman, C. Partridge, S. Deering, "Distance Vector Multicast Routing Protocol",

RFC1075, November 1988

[13] Andrew Adams, Jonathan Nicholas, William Siadak, "Protocol Independent Multicast - Dense Mode (PIM-DM): Protocol Specification (Revised)", draft-ietf-pim-dm-new-v2-03.txt, February 2003

[14] R. Braden, et al, "Resource ReSerVation Protocol (RSVP)-Version 1 Functional Specification", RFC2205, Sep. 1997

[15] J. Wroclawski, "The use of RSVP with IETF Integrated Services", RFC2210, Sep. 1997

[16] Dirk Ooms, et al, "Overview of IP Multicast in a MPLS Environment", RFC3353, August 2002

[17] Anjali Agarwal, KangBin Wang, Supporting Quality of Service in IP Multicast Networks, article in press in Journal of Computer Communications

[18] R. Bless, K. Wehrle, "Group communication in differentiated services networks", First IEEE/ACM International Symposium on Cluster Computing and the Grid, 2001, pp 618-625

[19] R. Bless, K. Wehrle, "IP Multicast in Differentiated Services Networks", draft-bless-diffserv-multicast-06.txt, February, 2003

[20] A. Striegel, G. Manimaran, "A scalable approach for DiffServ multicasting", ICC 2001, IEEE International Conference on Communications, Vol 8, pp 2327-2331

[21] A. Striegel, G. Manimaran, "A scalable protocol for member join/leave in DiffServ Multicast", Proc. of Local Computer Networks (LCN), Tampa, Florida, Nov. 2001

[22] R. Bless, K. Wehrle, "DS Multicast Router Extension", draft-bless-diffserv-mcast-routerext-00.txt, July, 2001

[23] D. Ooms, W. Livens, "IP multicast in MPLS networks", Proceedings of the IEEE Conference

on High Performance Switching and Routing, 2000, pp 301-305

[24] Arup Acharya, F. Griffoul, F. Ansari, "IP multicast support in MPLS" ATM Workshop, 1999. IEEE Proceedings, pp 211-218

[25] Jaihyung Cho; Min Young Chung, "A simple method for implementing PIM to ATM based MPLS networks", Ninth IEEE International Conference on Networks, 2001, pp 362-365

[26] B. Jamoussi, et al, "Constraint-Based LSP Setup using LDP", RFC3212, January 2002

[27] Zhongshan Zhang, Keping Long, Wendong Wang, Shiduan Cheng, "The new mechanism for MPLS supporting IP multicast", IEEE APCCAS 2000, pp 247-250

[28] D. Ooms, R. Hoebeke, P. Cheval, L. Wu, "MPLS Multicast Traffic Engineering", draft-ooms-mpls-multicast-te-01.txt, Feb. 2002

[29] Heng-Chi Su; Ren-Hung Hwang, "Multicast provision in a differentiated services network", 15th International Conference on Information Networking, 2001, pp 189-196

[30] D. Awduche, et al, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC3209, Dec. 2001

[31] Dean Cheng, "RSVP-TE: Extensions to RSVP for Multicast LSP Tunnels", draft-cheng-mpls-rsvp-multicast-er-00.txt, Oct. 2001

[32] Jong-Moon Chung, "RSVP-TE Extensions for MPLS Multicasting Services", draft-chung-mpls-rsvp-multicasting-00.txt, Feb. 2002

[33] N. Rouhana, E. Horlait, "Differentiated services and integrated services use of MPLS", ISCC 2000. Fifth IEEE Symposium on Computers and Communications, 2000, pp 194-199

[34] I. Andrikopoulos, G. Pavlou, "Supporting differentiated services in MPLS networks", IWQoS'99, Seventh International Workshop on Quality of Service, 1999, pp 207-215

[35] Francois Le Faucheur, et al, "Multi-Protocol Label Switching (MPLS) Support of

Differentiated Services", RFC3270, May 2002

[36] Melody Moh, Belle Wei, Jane Huijing Zhu, "Supporting differentiated services with per-class traffic engineering in MPLS", Tenth International Conference on Computer Communications and Networks, 2001 pp 354-360

[37] Zhigang Jing, Lemin Li, Hairong Sun, "Supporting differentiated services in MPLS-based ATM switches", APCC/OECC'99, Fifth Asia-Pacific Conference on Communications and Fourth Optoelectronics and Communications Conference, 1999, pp 91-93 vol.1

[38] Jun-Hong Cui, et al, "Aggregated multicast: A Scheme to Reduce Multicast States", draft-cui-multicast-aggregation-01.txt, September 2002

[39] M. Parsa, J.J. Garcia-Luna-Aceves, "Scalable Internet Multicast Routing", Proceeding of ICCCN'95, pages 162-166, 1995

[40] Yoshihiro Ohba, Yasuhiro Katsube, Eric Rosen, Paul Doolan, "MPLS Loop Prevention Mechanism", RFC3063, February 2001

[41] Yoshihiro Ohba, "Issues on Loop Prevention in MPLS networks", IEEE Communications Magazine, December 1999

[42] KangBin Wang, Anjali Agarwal, Multicast Traffic Merging in Diffserv-Supported MPLS Networks, CCECE'03, May 4-7, 2003, Montreal, Canada