EDGE ROUTER MULTICASTING WITH MPLS TRAFFIC ENGINEERING

Baijian Yang
Department of Computer Science and Engineering
Michigan State University, East Lansing, MI
Email: yangbaij@msu.edu

Prasant Mohapatra
Department of Computer Science
University of California, Davis, CA
Email: prasant@cs.ucdavis.edu

ABSTRACT

Explicit routing in MPLS is utilized in traffic engineering to maximize the operational network performance and to provide Quality of Service (QoS). However, difficulties arise while integrating native IP multicasting with MPLS traffic engineering, such as point-to-multipoint or multipoint-to-multipoint LSPs layout design and traffic aggregation. In this paper, we have proposed an edge router multicasting (ERM) scheme by limiting branching point of multicast delivery tree to only the edges of MPLS domains. As a result, multicast LSP setups, multicast flow assignments, and multicast traffic aggregation are reduced to unicast problems. We have studied two types of ERM routing protocols in the paper. The first approach is based on modifications to the existing multicast protocols, while the second approach applies Steiner tree-based heuristic routing algorithm in the edge router multicasting environment. The simulation results demonstrate that the ERM scheme based on Steiner tree heuristic can provide near-optimal performance. The results also demonstrate that ERM provides a traffic engineering friendly approach without sacrificing the benefits of native IP multicasting.

1. INTRODUCTION

Several IP multicasting techniques have been proposed to support point-to-multipoint communications by sharing link resources at the network layer. The advantages of IP multicasting include reduction in network resource consumption and source link stress. Examples of applications that could benefit through multicasting include audio and video distribution, push applications, audio and video conferencing and in general, large amount of data transfer from a single to multiple locations [2]. Most of these applications usually have Quality of Service (QoS) requirements, which include bandwidth, bounded delay, and low loss rate. So the constraints of QoS provisioning should be also considered while supporting multicast communications.

Several techniques have been proposed by the IETF for QoS provisioning in the Internet. One of the approaches, Multiprotocol Label Switching (MPLS) is being considered for scalable QoS provisioning. In this paper, we focus on the support of multicasting in MPLS domains.

1.1. MPLS and MPLS Traffic Engineering

The fundamental idea of Multi-Protocol Label Switching (MPLS) [1] involves assigning short, fixed length labels to the packets at the ingress point of the network. In ATM environment, the label is encoded in the VCI/VPI field. In IP network, a 32-bit 'shim' header is inserted between the network layer header and the data link layer header. When packets are forwarded within an MPLS domain, the MPLS capable routers, termed as Label Switching Routers (LSRs), only examine the label rather than the IP header.

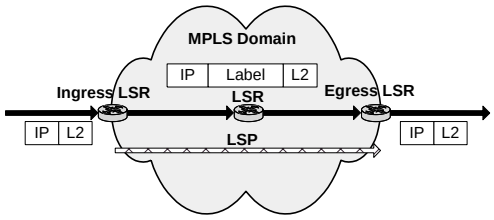


Fig. 1. MPLS Illustration.

As depicted in Fig.1, when a packet from a non-MPLS domain arrives at an MPLS domain, an MPLS header will be generated and inserted at the ingress LSR based on the IP header in the packet and local routing information. Within the MPLS domain, the LSR examines the incoming label, look up the forwarding table, and replaces it with an outgoing label. Thus, the packet is switched to the next LSR. Before a packet leaves the MPLS domain, the header will be removed. The path between the ingress LSR and egress LSR are called Label Switching Path (LSP), which can be set up using Label Distribution Protocol (LDP) or RSVP.

MPLS enriches the classical routing functionality by separating the forwarding components and path controlling components. It allows packets to be forwarded along a pre-configured LSP rather than the conventional shortest path, thus provides a means for traffic engineering (MPLS-TE) [3]. Adopting online or offline optimization algorithms, MPLS-TE can maximize operational network performance and balance traffic load. Moreover, working together with RSVP or DiffServ, MPLS-TE also provides a scalable QoS scheme. Typically, the procedures of MPLS-TE can be described as following:

- LSPs are pre-established between each ingress and egress node pair.

- Packets are classified into different Forwarding Equivalent Classes (FECs) when arriving at an ingress node.
- FECs are then grouped into traffic trunks, which are defined as routable objects placed inside of an LSP
- Finally, traffic trunks are mapped to LSPs which can satisfy their QoS requirements with optimized network performance.

Two primary problems of MPLS-TE are layout design and flow assignment. It would be efficient to run off-line algorithms if we have a priori knowledge about traffic demands and patterns. But such assumption is not valid in practice. Some online algorithms have been proposed to address LSPs layouts and flow assignments for unicast traffic [4]. However, to the best of our knowledge, no algorithms have been proposed yet for traffic engineering of multicast flows in MPLS domains.

1.2. Difficulties in Supporting IP Multicast in MPLS Domains

As MPLS is standardized by IETF and is expected to be implemented in the near future, it is inevitable to address the issues of supporting IP multicast in MPLS domain. Furthermore, the power of MPLS traffic engineering has the potential to provide QoS for IP multicast communications.

While MPLS offers great flexibility in packet forwarding, it does not enrich the functionality of native IP multicast routing. On the contrary, problems arise while mapping layer 3 multicast trees onto layer 2 LSPs. Thus a number of issues need to be addressed, such as flood and prune, source/shared trees, uni/bi-directional trees, and encapsulated multicast [5]. Specifically, while leveraging the power of MPLS traffic engineering to support QoS-aware multicasting, several difficulties arise, some of which are itemized as follows.

- *LSP design:* The multicast tree structure requires establishing point-to-multipoint LSPs or even multipoint-to-multipoint LSPs. In current MPLS architecture, only point-to-point LSP has been addressed. MPLS does not exclude other type of LSPs, but no mechanism has been standardized for this purpose. In fact, to the best of authors knowledge, only multipoint-to-point LSP has been studied so far [7], which is proposed to save label space. Moreover, dynamic multicast group membership indicates that multicast associated LSPs are volatile. The consequences are tremendous signaling overhead, and over-consumed labels. The design of efficient multicast-enabled LSPs layout is still an intriguing issue for researchers.
- *Traffic Aggregation.* In the context of MPLS, as mentioned in Section 1.1, traffic is aggregated and mapped to LSPs at the entrance of the network to achieve scalability. This feature will not be suitable for multicast traffic. To handle this situation, one needs to devise algorithms that can aggregate unicast flows with multicast flows as well as aggregate multiple multicast flows. Unfortunately, current studies on the aggregatability of multicast are limited to the forwarding state of each router rather an LSP consisting of a group of routers/switches in sequence.
- *Coexistence of Layer 2 and Layer 3 forwarding in core LSRs.* There are two cases where layer 2 incoming labels alone cannot determine the outgoing labels. The first case is due to the switch-over from a shared tree to a source based tree. In this situation, it might happen that certain on-tree routers are on both trees, and have both forwarding state (*,G) and (S,G) for the same destination address G. The other case occurs if labels are assigned inappropriately. Suppose a multicast flow is mapped to the same label as some unicast flows. Then at the branching node of the multicast tree, the label will be split. In both of the cases, it mandates such LSRs examine the layer 3 header as well as the layer 2 label. This requirement is at odds with the current MPLS standard, where it only demands edge LSRs be capable of layer 3 forwarding.

1.3. Solution and Paper Organization

To get around the difficulties mentioned above, and to facilitate multicasting in MPLS domains, we propose an *edge routers multicasting* (ERM) protocol. In the ERM technique, multicast trees are formed by branching only at the edge routers. Packets are routed through the branches using the MPLS tunnels established by the core routers. ERM facilitates multicast LSP set ups and the aggregation of multicast and unicast traffic. Simulation results on a variety of network topologies have been provided to demonstrate the feasibility and performance benefits of ERM.

The rest of the paper is organized as follows. The motivations for ERM is outlined in Section 2. The basic ERM protocol is described in Section 3 followed by the extended ERM2 protocol in Section 4. The performance results are discussed in Section 5. Section 6 describes the related work followed by the concluding remarks in Section 7.

2. MOTIVATION

We assume that in MPLS domain, multicast group members are directly attached to edge LSRs, and core LSRs are only connected with other LSRs. The proposed edge router multicasting scheme tries to construct a multicast tree whose branching points are only located at edge LSRs. As shown in Fig.2, edge LSRs ER1, ER2, ER3 and ER4 are active members of a multicast group. Fig.2(a) depicts the multicast tree produced by conventional IP multicast routing protocols. The branching nodes are core LSRs CR1, CR2 and CR3. In ERM, a multicast tree branches at edge LSRs ER1 and ER4, and is connected by pre-connected LSPs, namely LSP1, LSP2, and LSP3 respectively, as shown in Fig.2(b).

By limiting branching points only at the edges, conceptually, ERM converts a multicast flow into multiple quasi unicast

flows at the network layer. Compared to native IP multicasting, ERM scheme has distinct advantages that are itemized as follows.

1. *Simplifies LSP setup.* Since the diverging nodes of the tree are only located at edge LSRs, there is no need to create and maintain point-to-multipoint or multipoint-to-multipoint LSPs. Instead, a tree can be decomposed and mapped to multiple point-to-point LSPs.

2. *Makes multicast flows aggregatable.* Each branch of a multicast flow can be aggregated with other unicast flows which share the same ingress and egress LSRs. Thus the scalability of MPLS traffic engineering will not be compromised.

3. *Relaxes the requirements at core routers.* One of the reasons that IP multicast is not widely implemented is because of the fact that many core routers in the backbone are not multicast ready [2]. As the core routers are usually carrying out critical missions, they are unlikely to be upgraded off-line in the near future. Edge router multicasting approach can be designed in such a way that it poses little or no multicasting restrictions on core routers.

4. *Requires no encapsulation to setup multicast tunnels.* When a multicast router communicates with its multicast peers through non-multicast routers, a typical solution is manually-built tunnels by IP-in-IP encapsulation. That is, a whole IP header is inserted in the packet leaving from the upstream peer and then it is removed at the downstream peer. While in MPLS environment, LSPs can be directly used as multicast tunnels if multicast peers are edge routers.

3. EDGE ROUTER MULTICASTING (ERM) PROTOCOL

ERM consists of three fundamental components: edge router multicast routing, multicast LSPs mapping, and edge router multicast forwarding.

3.1. Edge Router Multicast Routing

We focus on intra-domain routing scheme since inter-domain routing protocols like MSDP/MBGP and BGMP allow each autonomous system to have its own multicast implementation.

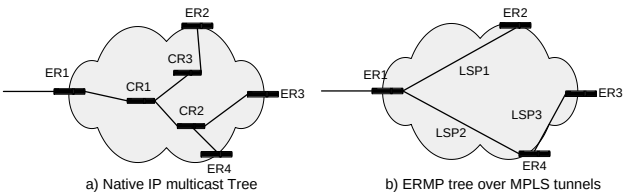


Fig. 2. ERMP Illustration.

For ERM, different multicast routing algorithms need to be developed to construct the ERM trees. We first present a simple solution by slightly modifying the existing IP multicast routing protocols. In the next section, a Steiner tree-based heuristic routing algorithms will be discussed.

For sparse mode IP multicasting like PIM-SM, and CBT, multicast trees are constructed by explicit join. To extend these routing schemes for edge router multicasting, the following two steps need to be adopted.

- Select edge routers as the core or Rendezvous Point (RP) of the tree.
- Allow a sub-tree to join only at the edge routers.

For dense mode protocols, such as DVMRP [**?**], multicast delivery trees are built by flood and prune approach. To support ERM, the process should be changed to 'flood and acknowledge'. Each edge router should inform its upstream peer explicitly whether it has any active members on its outgoing interfaces. Each edge router should keep the multicast state, and record its next downstream edge routers, if any. Reverse path forwarding algorithms can also be employed to limit the impact of flooding.

In both modes, core LSRs are involved in building multicast trees, but they do not need to maintain the multicast state. ERM routing tables at edge LSRs should record its downstream peers in addition to downstream outgoing interfaces. For example, in Fig.3, for multicast state (4.10.25.10, 234.62.37.6), the outgoing interface is 1 and 2, with downstream edge peer 63.42.7.91 and 63.1.3.85, respectively.

Source IP	Group IP	Outgoing Interfaces	Downstream Edge Peer IP
4.10.25.10	234.62.37.6	1	63.42.7.91
		2	63.1.3.85
18.2.36.88	242.11.7.8	2	63.15.9.28

Fig. 3. Multicast Routing Table at an Edge LSRs.

3.2. Multicast LSP Mapping

After the multicast routing process, each edge LSR has the knowledge about its downstream peers. A multicast flow can thus be mapped onto multiple LSPs based on downstream destination addresses of an edge LSR and QoS requirements of the flow as if there are multiple unicast flows destined to downstream peers. In Fig.3, a multicast flow from 4.10.25.10 to 234.62.37.6 will be mapped onto two unicast LSPs destined to 63.42.7.91 and 63.1.3.85, respectively.

3.3. Edge Router Multicast Forwarding

When multicast packets needs to be forwarded in the ERM protocol, edge LSRs need to duplicate packets based on their routing table, and assign the corresponding MPLS labels. Core LSRs do not have to duplicate any packets. The forwarding decisions can be made by simply examining the incoming labels.

In fact, core LSRs do not have to distinguish whether a label is associated with multicasting or not, because in ERM, they only have one outgoing interface for each incoming packet.

4. EXTENSION TO ERM ROUTING

The multicast routing approach described in the previous section is easy to implement and it requires only minor modifications in the current multicasting protocols. However, it still demands core routers participate in the multicast routing process. In MPLS-TE, we assume that network resource usage and availability is either available from centralized management nodes or from each edge LSRs. Thus an ERM-based Steiner heuristic tree can be constructed without the involvement of core LSRs, which leads to an extended version of the ERM protocol, termed as ERM2 in this paper.

4.1. Basic Characteristics

- *Source-based Tree.* EMR2 constructs a multicast tree per source. Source-based tree has an advantages over core-based tree in address allocation, since each source can freely pick any address and create a unique (S,G) state. Moreover, core-based tree are typically shared among the group members, which also requires the support of bi-directional trees. Bi-directional LSPs are still under investigation in the current MPLS architecture.

- *Explicit Join.* We avoid using the flood-and-prune approach for the following reasons. First, the density of a multicast group is likely to be sparse compared to the size of Internet. Explicit join would be more efficient in such scenario. Second, flood and prune are traffic driven, not control driven. When a multicast flow arrives at the edge of a network, it needs to set up a tree first, only after which the flow assignment algorithm can be executed to map the flow onto an LSP. This will increase the latency for the delivery of the very first packet.

- *Centralized Control.* We propose a dedicated node called "Multicast Manager" (MM) in ERM2. The role of MM is different with that of "core" or "RP" in CBT and PIM. MM is not designated to be the root of a delivery tree. Rather, it functions like a DNS server, and is responsible for group membership management in an MPLS domain. It keeps a record of current active on-tree edge routers and returns a list of candidates to a new receiver. The merits of centralized control includes easy implementation and simplified routing algorithms.

- *Protocol Independence.* In view of heterogeneous nature of internet, ERM2 is designed to be independent of unicast routing protocols. Thus it can be implemented on top of distance vector protocol as well as link state protocol.

4.2. ERM2 Illustration

The "Join" process in ERM2 can be illustrated by an example depicted in Fig.4, where edge router E1, E2, E4, and E7 are on-tree routers of multicast group G. Suppose edge router E5 wants to join group G. The routing procedure is enumerated as follows.

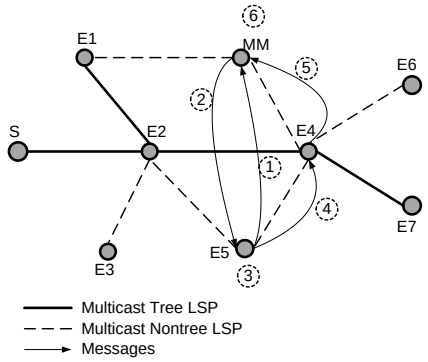


Fig. 4. ERMP2 Join Example

1. Edge router E5 sends a QUERY message to MM.
2. MM returns an ANSWER message with a list of candidates to E5. In this example, the candidates are S, E1, E2, E4, and E7.
3. Based on its own routing table, or resource availability, E5 picks the bestcandidate, say E4, as the join point.
4. E5 sends a JOIN message to E4 and E4 create an outgoing entry for state (S,G).
5. If successful, E4 inform MM that E5 is now an active on-tree edge router through an ADD message.
6. MM inserts E5 in the active member list.

An edge node will leave a multicast tree when two condition are met. First, it detects that there is no active member directly attached to it by Internet Group Management Protocol (IGMP) report. Second, it does not have any downstream peer. The leaving node will send a SUBTRACT message to the MM to update the member list, and a PRUNE message to its upstream peer.

5. PERFORMANCE ANALYSIS

Network topology and group density are two major factors which affect performance of multicast routing protocols. We chose three types of network topologies, Waxman1, Waxman2, and locality model, because they are considered close to real-life network topology [10]. A variety of flat random graphs have been proposed to model networks in aim to reflect realistic network topologies. All the variations randomly distribute vertices in a plane and add an edge between each pair of vertices with certain probabilitic parameters. We have chosen three commonly used random graph models in our study,

namely Waxman1, Waxman2, and locality. The edge distribution functions are summarized in Table 1

Model	Edge Probability
Waxman1	$\alpha e^{-d/(\beta L)}$
Waxman2	$\alpha e^{-rand(0,L)/(\beta L)}$
Locality	$\begin{cases} \alpha & if\ d < L \times radius \\ \beta & if\ d \geq L \times radius \end{cases}$

Table 1. Edge Probability of selected flat random graph models.

In Table 1, $0 < \alpha, \beta \leq 1$, d is the Euclidean distance between two vertices, and L is the maximum distance between any two vertices. Intutively, locality model has the richest short distnace connectivity in three models, while Waxman1 generate less long distance edges than Waxman2.

For each model, we use GIT network topology generator produced 1024 nodes flat network. Among them 300 out of 1024 nodes are randomly selected as the edge routers. The simulation results are collected and tabulated by recording performance metrics in different topologies by increasing group members from 5 to 300. For ERM2 routing, we assume each edge router picks the node with least cost path as the join point.

5.1. Relative Tree Cost

Relative tree cost is defined as the ratio of the tree cost over the sum of unicast path cost. Fig.5(a), Fig.5(b) and Fig.5(c) show tcomparison of relative tree costs. In the figures, OPT refers to optimal results produced by the Steiner Tree algorithm. For all the topologies, ERM yields worst relative tree cost, while ERM2 incurs less cost than DVMRP and even demonstrates near-optimal performance. These results prove that edge router multicast scheme may not necessary leads to very high tree cost. As a matter of fact, with careful design, it could be more efficient than least-cost unicast path tree built by protocols like DVMRP. Another interesting observation inferred from Fig.5 is that the widely accepted multicast protocols like DVMRP only save half of the link cost when all the edge nodes join a multicast tree.

5.2. Link Stress

Stressed links refer to those links that have multiple identical packets on the outgoing interface. The number of the identical packets is denoted as link stress. For native multicast protocols, link stress is always equals to one. For unicast, source node link stress equals to the total number of on tree node numbers in a domain. Combining tree cost results presented in Fig.5, the most important feature of multicast may be releaving link stress, rather than saving bandwidth. The ERM protocol could introduce stressed links. Results of link stress are plotted in Fig.6(a), Fig.6(b) and Fig.6(c). ERM and ERM2 both have average link stress between 2 to 3, and the ratio of stressed link are both less than 20%. However, the maximum link stress of ERM is much higher than ERM2. In the worst case, the maximum link stress is as high as nearly 40. Link stress performance can be easily improved by adding maximum link stress restrictions. The side effect of this restriction would produce worse results for other performance metrics like tree cost and relative delay.

6. RELATED WORK

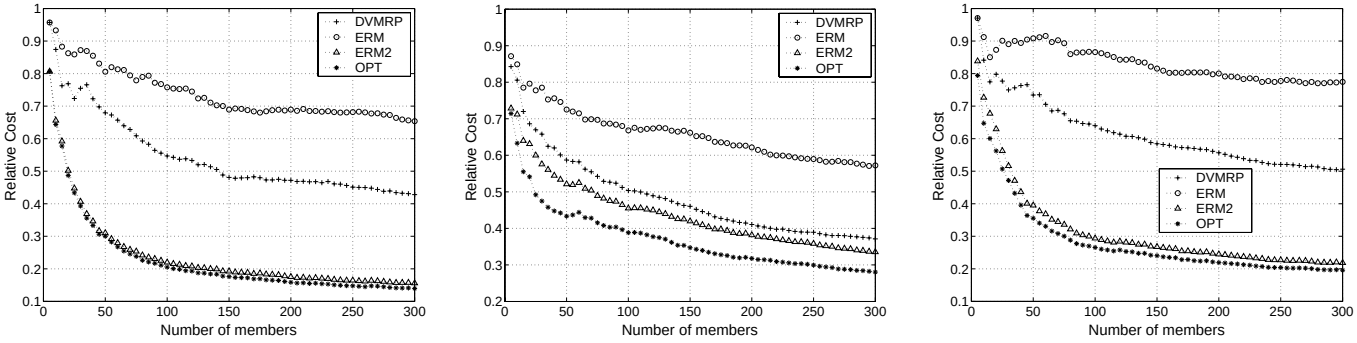
General issues of supporting native IP multicast in MPLS are identified and discussed in [5]. In addition to the concept of a hybrid of L2 and L3 forwarding, label distribution, and LSP setup trigger mode, the authors have proposed a framework for IP multicasting in MPLS domains. However, they did not address issues related to traffic engineering of multicasting or aggregating label assignment schemes in MPLS domains. The proposed ERM scheme eliminates most of the problems mentioned in [5] since supporting ERM in MPLS can be conceived as a label-switched approach for multiple simultaneous unicast flows. Problems of traffic aggregation and label assignment can thus be reduced to that of unicast flows.

An MPLS Multicast Tree (MMT) scheme was introduced in [6] to remove multicast forwarding state in non-branching nodes by dynamically setting up LSP tunnels between upstream branching nodes and downstream branching nodes. Like ERM, MMT can dramatically reduces forwarding states. However, MMT still needs to set up and update LSPs between edge LSRs and core LSRs (if some core LSRs are branching nodes of multicast trees). As a result, the core LSRs have to support the coexistence of L2/L3 forwarding schemes. Normally LSPs are built between edge LSRs. LSPs produced by MMT may not necessarily be able to aggregate with other unicast LSPs. However, in ERM, there would be no need to set up any LSPs between edge LSRs and core LSRs, which enables ERM to aggregate both multicast and unicast traffic. Another difference between MMT and ERM is that the multicast tree is centrally calculated in MMT, while basic ERM is fully distributed, and the extended ERM (ERM2) is partially distributed.

Some end-host based multicasting approaches, such as [8, 9], can also avoid problems described in Section 1.1. Instead of building a multicast tree on network layer, a shared tree/mesh is set up on the application layer among the active member hosts. While end-host multicasting offers an easy and general implementation of multipoint communication, it has limitations in scalability and QoS support due to complicated group management and the absence of network layer support. ERM is an alternative network layer multicasting which is designed in to provide QoS with MPLS traffic engineering.

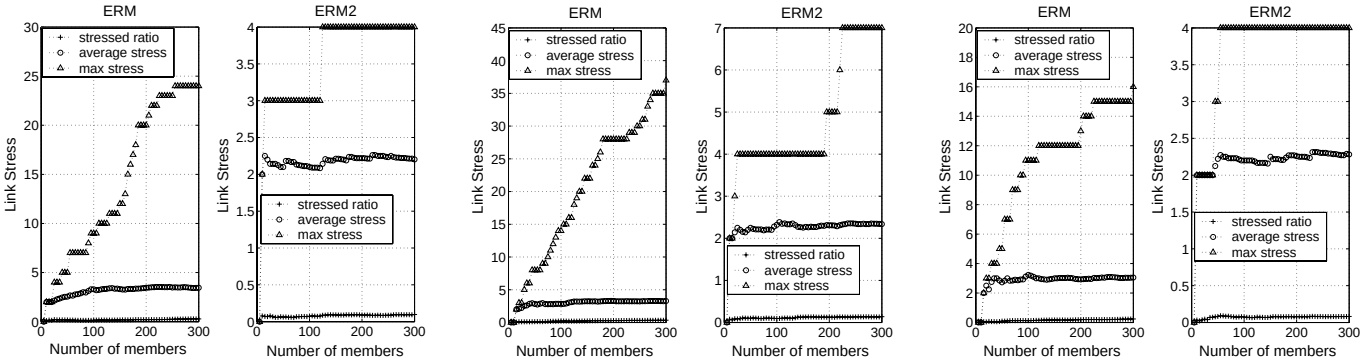
7. CONCLUSIONS

We have proposed an edge router multicasting approach in MPLS traffic engineering environment. ERM converts the design of point-to-multipoint LSP setup to a multiple point-to-point LSP problems, and make multicast traffic suitable for aggregation. In the ERM protocols, the multicast trees branch only at the edge routers and use the MPLS tunnels set up by the core routers. In addition, the proposed approach does not



(a) locality (b) waxman1 (c) waxman2

Fig. 5. Relative tree cost comparison.



(a) locality (b) waxman1 (c) waxman2

Fig. 6. Link Stress

loses the strength of native IP multicast. The implementation of the ERM protocol in incrementally deployable as it does not requires any changes in the core routers. Simulation results show that the proposed ERM2 has near optimized tree cost, low link stress, and incurs low delay.

8. REFERENCES

[1] E. Rosen, A. Viswanathan, and R.Callon, *Multiprotocol Label Switching Architecture*, RFC 3031, Jan. 2001.

[2] C. Diot, B. N. Levine, B. Lyles, H. Kassem and D. Balensiefen, "Deployment Issues for the IP Multicast Service and Architecture" *IEEE Network*, pp.78-88, Jan/Feb 2000.

[3] D. Awduche, J. Malcolm, J. Agogbua, M. O'Dell, and J. McManus, *Requirements for Traffic Engineering Over MPLS*, RFC 2702, Sep. 1999.

[4] P. Aukia, M. Kodialam, P. Koppol, T. Lashman, H. Sarin, and B. Suter "RATES: A server for MPLS Traffic Engineering" *IEEE Network Magazine*, pp.34-41, Mar./Apr. 2000.

[5] D. Ooms, B. Sales, W. Livens, A. Acharya, F. Griffoul, and F. Ansari, "Framework for IP Multicast in MPLS", work in progress, http://www.ietf.org/internet-drafts/draft-ietf-mpls-multicast-07.txt.

[6] A. Boudani and B. Cousin, "An Effective Solution for Multicast Scalability: The MPLS Multicast Tree (MMT)", work in progress, http://search.ietf.org/internet-drafts/draft-boudani-mpls-multicast-tree-00.txt.

[7] H. Saito, Y. Miyao, and M. Yoshida, "Traffic Engineering using Muliple Multipoint-to-Point LSPs", IEEE Infocom, 2000.

[8] P.Francis, "Yoid: Extending the Internet Architecture", Work in progress, Apr.2000, http://www.yallcast.com.

[9] Y. Chu, S.G. Rao, H. Zhang, "A Case For End System Multicast", *Proceedings of ACM SIGMETRICS*, pp.1-12, Jun.2000.

[10] K. Calvert, M. Doar and E. W. Zegura, "Modeling Internet Topology" *IEEE Communications*, Jun. 1997.