

Network Working Group
Request for Comments: 4461
Category: Informational

S. Yasukawa, Ed.
NTT
April 2006

Signaling Requirements for Point-to-Multipoint
Traffic-Engineered MPLS Label Switched Paths (LSPs)

Status of This Memo

This memo provides information for the Internet community. It does not specify an Internet standard of any kind. Distribution of this memo is unlimited.

Copyright Notice

Copyright (C) The Internet Society (2006).

Abstract

This document presents a set of requirements for the establishment and maintenance of Point-to-Multipoint (P2MP) Traffic-Engineered (TE) Multiprotocol Label Switching (MPLS) Label Switched Paths (LSPs).

There is no intent to specify solution-specific details or application-specific requirements in this document.

The requirements presented in this document not only apply to packet-switched networks under the control of MPLS protocols, but also encompass the requirements of Layer Two Switching (L2SC), Time Division Multiplexing (TDM), lambda, and port switching networks managed by Generalized MPLS (GMPLS) protocols. Protocol solutions developed to meet the requirements set out in this document must attempt to be equally applicable to MPLS and GMPLS.

Table of Contents

1. Introduction	3
1.1. Non-Objectives	6
2. Definitions	6
2.1. Acronyms	6
2.2. Terminology	6
2.2.1. Terminology for Partial LSPs	8
2.3. Conventions	9
3. Problem Statement	9
3.1. Motivation	9
3.2. Requirements Overview	9
4. Detailed Requirements for P2MP TE Extensions	11
4.1. P2MP LSP	11
4.2. P2MP Explicit Routing	12
4.3. Explicit Path Loose Hops and Widely Scoped Abstract Nodes	13
4.4. P2MP TE LSP Establishment, Teardown, and Modification Mechanisms	14
4.5. Fragmentation	14
4.6. Failure Reporting and Error Recovery	15
4.7. Record Route of P2MP TE LSP	16
4.8. Call Admission Control (CAC) and QoS Control Mechanism of P2MP TE LSPs	17
4.9. Variation of LSP Parameters	17
4.10. Re-Optimization of P2MP TE LSPs	18
4.11. Merging of Tree Branches	18
4.12. Data Duplication	19
4.13. IPv4/IPv6 Support	20
4.14. P2MP MPLS Label	20
4.15. Advertisement of P2MP Capability	20
4.16. Multi-Access LANs	21
4.17. P2MP MPLS OAM	21
4.18. Scalability	21
4.18.1. Absolute Limits	22
4.19. Backwards Compatibility	24
4.20. GMPLS	24
4.21. P2MP Crankback Routing	25
5. Security Considerations	25
6. Acknowledgements	26
7. References	26
7.1. Normative References	26
7.2. Informative References	26

1. Introduction

Existing MPLS traffic engineering (MPLS-TE) allows for strict QoS guarantees, resource optimization, and fast failure recovery, but it is limited to point-to-point (P2P) LSPs. There is a desire to support point-to-multipoint (P2MP) services using traffic-engineered LSPs, and this clearly motivates enhancements of the base MPLS-TE tool box in order to support P2MP MPLS-TE LSPs.

A P2MP TE LSP is a TE LSP (per [RFC2702] and [RFC3031]) that has a single ingress LSR and one or more egress LSRs, and is unidirectional. P2MP services (that deliver data from a single source to one or more receivers) may be supported by any combination of P2P and P2MP LSPs depending on the degree of optimization required within the network, and such LSPs may be traffic-engineered again depending on the requirements of the network. Further, multipoint-to-multipoint (MP2MP) services (which deliver data from more than one source to one or more receivers) may be supported by a combination of P2P and P2MP LSPs.

[RFC2702] specifies requirements for traffic engineering over MPLS. In Section 2, it describes traffic engineering in some detail, and those definitions are equally applicable to traffic engineering in a point-to-multipoint service environment. They are not repeated here, but it is assumed that the reader is fully familiar with them.

Section 3.0 of [RFC2702] also explains how MPLS is particularly suited to traffic engineering; it presents the following eight reasons.

1. Explicit label switched paths that are not constrained by the destination-based forwarding paradigm can be easily created through manual administrative action or through automated action by the underlying protocols.
2. LSPs can potentially be maintained efficiently.
3. Traffic trunks can be instantiated and mapped onto LSPs.
4. A set of attributes can be associated with traffic trunks that modulate their behavioral characteristics.
5. A set of attributes can be associated with resources that constrain the placement of LSPs and traffic trunks across them.
6. MPLS allows for both traffic aggregation and disaggregation, whereas classical destination-only-based IP forwarding permits only aggregation.
7. It is relatively easy to integrate a "constraint-based routing" framework with MPLS.
8. A good implementation of MPLS can offer significantly lower overhead than competing alternatives for traffic engineering.

These points are equally applicable to point-to-multipoint traffic engineering. Points 1 and 7 are particularly important. Note that point 3 implies that the concept of a point-to-multipoint traffic trunk is defined and is supported by (or mapped onto) P2MP LSPs.

That is, the traffic flow for a point-to-multipoint LSP is not constrained to the path or paths that it would follow during multicast routing or shortest path destination-based routing, but it can be explicitly controlled through manual or automated action.

Further, the explicit paths that are used may be computed using algorithms based on a variety of constraints to produce all manner of tree shapes. For example, an explicit path may be cost-based [STEINER], shortest path, or QoS-based, or it may use some fair-cost QoS algorithm.

[RFC2702] also describes the functional capabilities required to fully support traffic engineering over MPLS in large networks.

This document presents a set of requirements for Point-to-Multipoint (P2MP) traffic engineering (TE) extensions to Multiprotocol Label Switching (MPLS). It specifies functional requirements for solutions to deliver P2MP TE LSPs.

Solutions that specify procedures for P2MP TE LSP setup MUST satisfy these requirements. There is no intent to specify solution-specific details or application-specific requirements in this document.

The requirements presented in this document apply equally to packet-switched networks under the control of MPLS protocols and to packet-switched, TDM, lambda, and port-switching networks managed by Generalized MPLS (GMPLS) protocols. Protocol solutions developed to meet the requirements set out in this document MUST attempt to be equally applicable to MPLS and GMPLS.

Existing MPLS TE mechanisms such as [RFC3209] do not support P2MP TE LSPs, so new mechanisms need to be developed. This SHOULD be achieved with maximum re-use of existing MPLS protocols.

Note that there is a separation between routing and signaling in MPLS TE. In particular, the path of the MPLS TE LSP is determined by performing a constraint-based computation (such as CSPF) on a traffic engineering database (TED). The contents of the TED may be collected through a variety of mechanisms.

This document focuses on requirements for establishing and maintaining P2MP MPLS TE LSPs through signaling protocols; routing protocols are out of scope. No assumptions are made about how the TED used as the basis for path computations for P2MP LSPs is formed.

This requirements document assumes the following conditions for P2MP MPLS TE LSP establishment and maintenance:

- o A P2MP TE LSP will be set up with TE constraints and will allow efficient packet or data replication at various branching points in the network. Although replication is a data plane issue, it is the responsibility of the control plane (acting in conjunction with the path computation component) to install LSPs in the network such that replication can be performed efficiently. Note that the notion of "efficient" replication is relative and may have different meanings depending on the objectives (see Section 4.2).
- o P2MP TE LSP setup mechanisms must include the ability to add/remove receivers to/from the P2MP service supported by an existing P2MP TE LSP.
- o Tunnel endpoints of P2MP TE LSP will be modified by adding/removing egress LSRs to/from an existing P2MP TE LSP. It is assumed that the rate of change of leaves of a P2MP LSP (that is, the rate at which new egress LSRs join, or old egress LSRs are pruned) is "not so high" because P2MP TE LSPs are assumed to be utilized for TE applications. This issue is discussed at greater length in Section 4.18.1.
- o A P2MP TE LSP may be protected by fast error recovery mechanisms to minimize disconnection of a P2MP service.
- o A set of attributes of the P2MP TE LSP (e.g., bandwidth, etc.) may be modified by some mechanism (e.g., make-before-break, etc.) to accommodate attribute changes to the P2MP service without impacting data traffic. These issues are discussed in Sections 4.6 and 4.10.

It is not a requirement that the ingress LSR must control the addition or removal of leaves from the P2MP tree.

It is this document's objective that a solution compliant to the requirements set out in this document MUST operate these P2MP TE capabilities in a scalable fashion.

1.1. Non-Objectives

For clarity, this section lists some items that are out of scope of this document.

It is assumed that some information elements describing the P2MP TE LSP are known to the ingress LSR prior to LSP establishment. For example, the ingress LSRs know the IP addresses that identify the egress LSRs of the P2MP TE LSP. The mechanisms by which the ingress LSR obtains this information is outside the scope of P2MP TE signaling and so is not included in this document. Other documents may complete the description of this function by providing automated, protocol-based ways of passing this information to the ingress LSR.

This document does not specify any requirements for the following functions.

- Non-TE LSPs (such as per-hop, routing-based LSPs).
- Discovery of egress leaves for a P2MP LSP.
- Hierarchical P2MP LSPs.
- OAM for P2MP LSPs.
- Inter-area and inter-AS P2MP TE LSPs.
- Applicability of P2MP MPLS TE LSPs to service scenarios.
- Specific application or application requirements.
- Algorithms for computing P2MP distribution trees.
- Multipoint-to-point LSPs.
- Multipoint-to-multipoint LSPs.
- Routing protocols.
- Construction of the traffic engineering database.
- Distribution of the information used to construct the traffic engineering database.

2. Definitions

2.1. Acronyms

P2P: Point-to-point

P2MP: Point-to-multipoint

2.2. Terminology

The reader is assumed to be familiar with the terminology in [RFC3031] and [RFC3209].

The following terms are defined for use in the context of P2MP TE LSPs only.

P2MP tree:

The ordered set of LSRs and TE links that comprise the path of a P2MP TE LSP from its ingress LSR to all of its egress LSRs.

ingress LSR:

The LSR that is responsible for initiating the signaling messages that set up the P2MP TE LSP.

egress LSR:

One of potentially many destinations of the P2MP TE LSP. Egress LSRs may also be referred to as leaf nodes or leaves.

bud LSR:

An LSR that is an egress LSR, but also has one or more directly connected downstream LSRs.

branch LSR:

An LSR that has more than one directly connected downstream LSR.

P2MP-ID (P2ID):

A unique identifier of a P2MP TE LSP, which is constant for the whole LSP regardless of the number of branches and/or leaves.

source:

The sender of traffic that is carried on a P2MP service supported by a P2MP LSP. The sender is not necessarily the ingress LSR of the P2MP LSP.

receiver:

A recipient of traffic carried on a P2MP service supported by a P2MP LSP. A receiver is not necessarily an egress LSR of the P2MP LSP. Zero, one, or more receivers may receive data through a given egress LSR.

2.2.1. Terminology for Partial LSPs

It is convenient to sub-divide P2MP trees for functional and representational reasons. A tree may be divided in two dimensions:

- A division may be made along the length of the tree. For example, the tree may be split into two components each running from the ingress LSR to a discrete set of egress LSRs. Upstream LSRs (for example, the ingress LSR) may be members of both components.
- A tree may be divided at a branch LSR (or any transit LSR) to produce a component of the tree that runs from the branch (or transit) LSR to all egress LSRs downstream of this point.

These two methods of splitting the P2MP tree can be combined, so it is useful to introduce some terminology to allow the partitioned trees to be clearly described.

Use the following designations:

Source (ingress) LSR - S
Leaf (egress) LSR - L
Branch LSR - B
Transit LSR - X (any single, arbitrary LSR that is not a source,
leaf or branch)
All - A
Partial (i.e., not all) - P

Define a new term:

Sub-LSP:

A segment of a P2MP TE LSP that runs from one of the LSP's LSRs to one or more of its other LSRs.

Using these new concepts, we can define any combination or split of the P2MP tree. For example:

S2L sub-LSP:

The path from the source to one specific leaf.

S2PL sub-LSP:

The path from the source to a set of leaves.

B2AL sub-LSP:

The path from a branch LSR to all downstream leaves.

X2X sub-LSP:

A component of the P2MP LSP that is a simple path that does not branch.

Note that the S2AL sub-LSP is equivalent to the P2MP LSP.

2.3. Conventions

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

3. Problem Statement

3.1. Motivation

As described in Section 1, traffic engineering and constraint-based routing (including Call Admission Control (CAC), explicit source routing, and bandwidth reservation) are required to enable efficient resource usage and strict QoS guarantees. Such mechanisms also make it possible to provide services across a congested network where conventional "shortest path first" forwarding paradigms would fail.

Existing MPLS TE mechanisms [RFC3209] and GMPLS TE mechanisms [RFC3473] only provide support for P2P TE LSPs. While it is possible to provide P2MP TE services using P2P TE LSPs, any such approach is potentially suboptimal since it may result in data replication at the ingress LSR, or in duplicate data traffic within the network.

Hence, to provide P2MP MPLS TE services in a fully efficient manner, it is necessary to specify specific requirements. These requirements can then be used when defining mechanisms for the use of existing protocols and/or extensions to existing protocols and/or new protocols.

3.2. Requirements Overview

This document states basic requirements for the setup of P2MP TE LSPs. The requirements apply to the signaling techniques only, and no assumptions are made about which routing protocols are run within the network, or about how the information that is used to construct the Traffic Engineering Database (TED) is distributed. These factors are out of the scope of this document.

A P2MP TE LSP path computation will take into account various constraints such as bandwidth, affinities, required level of protection and so on. The solution MUST allow for the computation of P2MP TE LSP paths that satisfy constraints, with the objective of

supporting various optimization criteria such as delays, bandwidth consumption in the network, or any other combinations. This is likely to require the presence of a TED, as well as the ability to signal the explicit path of an LSP.

A desired requirement is also to maximize the re-use of existing MPLS TE techniques and protocols where doing so does not adversely impact the function, simplicity, or scalability of the solution.

This document does not restrict the choice of signaling protocol used to set up a P2MP TE LSP, but note that [RFC3468] states

...the consensus reached by the Multiprotocol Label Switching (MPLS) Working Group within the IETF to focus its efforts on "Resource Reservation Protocol (RSVP)-TE: Extensions to RSVP for Label-Switched Paths (LSP) Tunnels" (RFC 3209) as the MPLS signalling protocol for traffic engineering applications...

The P2MP TE LSP setup mechanism MUST include the ability to add/remove egress LSRs to/from an existing P2MP TE LSP and MUST allow for the support of all the TE LSP management procedures already defined for P2P TE LSP. Further, when new TE LSP procedures are developed for P2P TE LSPs, equivalent or identical procedures SHOULD be developed for P2MP TE LSPs.

The computation of P2MP trees is implementation dependent and is beyond the scope of the solutions that are built with this document as a guideline.

Consider the following figure.

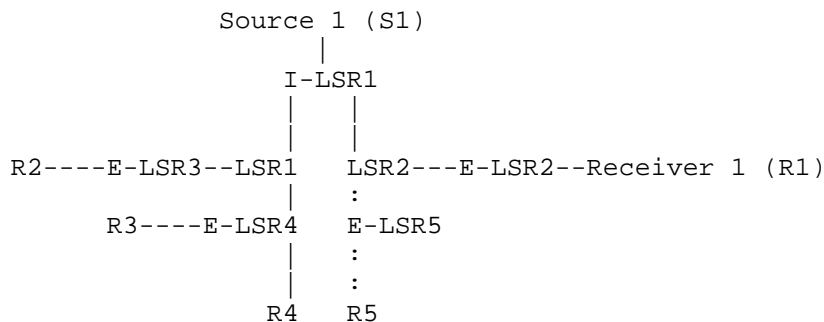


Figure 1

Figure 1 shows a single ingress LSR (I-LSR1), and four egress LSRs (E-LSR2, E-LSR3, E-LSR4, and E-LSR5). I-LSR1 is attached to a traffic source that is generating traffic for a P2MP application.

Receivers R1, R2, R3, and R4 are attached to E-LSR2, E-LSR3, and E-LSR4.

The following are the objectives of P2MP LSP establishment and use.

- a) A P2MP tree that satisfies various constraints is pre-determined, and details are supplied to I-LSR1.

Note that no assumption is made about whether the tree is provided to I-LSR1 or computed by I-LSR1. The solution SHOULD also allow for the support of a partial path by means of loose routing.

Typical constraints are bandwidth requirements, resource class affinities, fast rerouting, and preemption. There should not be any restriction on the possibility of supporting the set of constraints already defined for point-to-point TE LSPs. A new constraint may specify which LSRs should be used as branch LSRs for the P2MP LSP in order to take into account LSR capabilities or network constraints.

- b) A P2MP TE LSP is set up from I-LSR1 to E-LSR2, E-LSR3, and E-LSR4 using the tree information.
- c) In this case, the branch LSR1 should replicate incoming packets or data and send them to E-LSR3 and E-LSR4.
- d) If a new receiver (R5) expresses an interest in receiving traffic, a new tree is determined, and a B2L sub-LSP from LSR2 to E-LSR5 is grafted onto the P2MP TE LSP. LSR2 becomes a branch LSR.

4. Detailed Requirements for P2MP TE Extensions

4.1. P2MP LSP

The P2MP TE extensions MUST be applicable to the signaling of LSPs for different switching types. For example, it MUST be possible to signal a P2MP TE LSP in any switching medium, whether it is packet or non-packet based (including frame, cell, TDM, lambda, etc.).

As with P2P MPLS technology [RFC3031], traffic is classified with a FEC in this extension. All packets that belong to a particular FEC and that travel from a particular node MUST follow the same P2MP tree.

In order to scale to a large number of branches, P2MP TE LSPs SHOULD be identified by a unique identifier (the P2MP ID or P2ID) that is constant for the whole LSP regardless of the number of branches and/or leaves.

4.2. P2MP Explicit Routing

Various optimizations in P2MP tree formation need to be applied to meet various QoS requirements and operational constraints.

Some P2MP applications may request a bandwidth-guaranteed P2MP tree that satisfies end-to-end delay requirements. And some operators may want to set up a cost-minimum P2MP tree by specifying branch LSRs explicitly.

The P2MP TE solution therefore MUST provide a means of establishing arbitrary P2MP trees under the control of an external tree computation process, path configuration process, or dynamic tree computation process located on the ingress LSR. Figure 2 shows two typical examples.

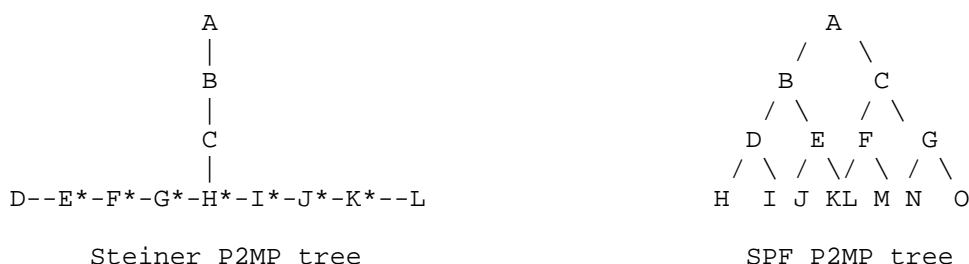


Figure 2: Examples of P2MP TE LSP topology

One example is the Steiner P2MP tree (cost-minimum P2MP tree) [STEINER]. This P2MP tree is suitable for constructing a cost-minimum P2MP tree so as to minimize the bandwidth consumption in the core. To realize this P2MP tree, several intermediate LSRs must be both MPLS data terminating LSRs and transit LSRs (LSRs E, F, G, H, I, J, and K in Figure 2). Therefore, the P2MP TE solution MUST support a mechanism that can set up this kind of bud LSR between an ingress LSR and egress LSRs. Note that this includes constrained Steiner trees that allow for the computation of a minimal cost trees with some other constraints such as a bounded delay between the source and every receiver.

Another example is a CSPF (Constraint Shortest Path First) P2MP tree. By some metric (which can be set upon any specific criteria like the delay, bandwidth, or a combination of those), one can calculate a shortest-path P2MP tree. This P2MP tree is suitable for carrying real-time traffic.

The solution **MUST** allow the operator to make use of any tree computation technique. In the former case, an efficient/optimal tree is defined as a minimal cost tree (Steiner tree), whereas in the later case, it is defined as the tree that provides shortest path between the source and any receiver.

To support explicit setup of any reasonable P2MP tree shape, a P2MP TE solution **MUST** support some form of explicit source-based control of the P2MP tree that can explicitly include particular LSRs as branch LSRs. This can be used by the ingress LSR to set up the P2MP TE LSP. For instance, a P2MP TE LSP can be represented simply as a whole tree or by its individual branches.

4.3. Explicit Path Loose Hops and Widely Scoped Abstract Nodes

A P2MP tree is completely specified if all the required branches and hops between a sender and leaf LSR are indicated.

A P2MP tree is partially specified if only a subset of intermediate branches and hops is indicated. This may be achieved using loose hops in the explicit path, or using widely scoped abstract nodes (that is, abstract nodes that are not simple [RFC3209]) such as IPv4 prefixes shorter than 32 bits, or AS numbers. A partially specified P2MP tree might be particularly useful in inter-area and inter-AS situations, although P2MP requirements for inter-area and inter-AS are beyond the scope of this document.

Protocol solutions **SHOULD** include a way to specify loose hops and widely scoped abstract nodes in the explicit source-based control of the P2MP tree as defined in the previous section. Where this support is provided, protocol solutions **MUST** allow downstream LSRs to apply further explicit control to the P2MP tree to resolve a partially specified tree into a (more) completely specified tree.

Protocol solutions **MUST** allow the P2MP tree to be completely specified at the ingress LSR where sufficient information exists to allow the full tree to be computed and where policies along the path (such as at domain boundaries) support full specification.

In all cases, the egress LSRs of the P2MP TE LSP must be fully specified either individually or through some collective identifier. Without this information, it is impossible to know where the TE LSP should be routed to.

In case of a tree being computed by some downstream LSRs (e.g., the case of hops specified as loose hops), the solution MUST provide protocol mechanisms for the ingress LSR of the P2MP TE LSP to learn the full P2MP tree. Note that this information may not always be obtainable owing to policy considerations, but where part of the path remains confidential, it MUST be reported through aggregation (for example, using an AS number).

4.4. P2MP TE LSP Establishment, Teardown, and Modification Mechanisms

The P2MP TE solution MUST support establishment, maintenance, and teardown of P2MP TE LSPs in a manner that is at least scalable in a linear way. This MUST include both the existence of very many LSPs at once, and the existence of very many destinations for a single P2MP LSP.

In addition to P2MP TE LSP establishment and teardown mechanisms, the solution SHOULD support a partial P2MP tree modification mechanism.

For the purpose of adding sub-P2MP TE LSPs to an existing P2MP TE LSP, the extensions SHOULD support a grafting mechanism. For the purpose of deleting a sub-P2MP TE LSPs from an existing P2MP TE LSP, the extensions SHOULD support a pruning mechanism.

It is RECOMMENDED that these grafting and pruning operations cause no additional processing in nodes that are not along the path to the grafting or pruning node, or that are downstream of the grafting or pruning node toward the grafted or pruned leaves. Moreover, both grafting and pruning operations MUST NOT disrupt traffic currently forwarded along the P2MP tree.

There is no assumption that the explicitly routed P2MP LSP remains on an optimal path after several grafts and prunes have occurred. In this context, scalable refers to the signaling process for the P2MP TE LSP. The TE nature of the LSP allows that re-optimization may take place from time to time to restore the optimality of the LSP.

4.5. Fragmentation

The P2MP TE solution MUST handle the situation where a single protocol message cannot contain all the information necessary to signal the establishment of the P2MP LSP. It MUST be possible to establish the LSP in these circumstances.

This situation may arise in either of the following circumstances.

- a. The ingress LSR cannot signal the whole tree in a single message.
- b. The information in a message expands to be too large (or is discovered to be too large) at some transit node. This may occur because of some increase in the information that needs to be signaled or because of a reduction in the size of signaling message that is supported.

The solution to these problems SHOULD NOT rely on IP fragmentation of protocol messages, and it is RECOMMENDED to rely on some protocol procedures specific to the signaling solution.

In the event that fragmented IP packets containing protocol messages are received, it is NOT RECOMMENDED that they are reassembled at the receiving LSR.

4.6. Failure Reporting and Error Recovery

Failure events may cause egress LSRs or sub-P2MP LSPs to become detached from the P2MP TE LSP. These events MUST be reported upstream as for a P2P LSP.

The solution SHOULD provide recovery techniques, such as protection and restoration, allowing recovery of any impacted sub-P2MP TE LSPs. In particular, a solution MUST provide fast protection mechanisms applicable to P2MP TE LSP similar to the solutions specified in [RFC4090] for P2P TE LSPs. Note also that no assumption is made about whether backup paths for P2MP TE LSPs should or should not be shared with P2P TE LSPs backup paths.

Note that the functions specified in [RFC4090] are currently specific to packet environments and do not apply to non-packet environments. Thus, while solutions MUST provide fast protection mechanisms similar to those specified in [RFC4090], this requirement is limited to the subset of the solution space that applies to packet-switched networks only.

Note that the requirements expressed in this document are general to all MPLS TE P2MP signaling, and any solution that meets them will therefore be general. Specific applications may have additional requirements or may want to relax some requirements stated in this document. This may lead to variations in the solution.

The solution SHOULD also support the ability to meet other network recovery requirements such as bandwidth protection and bounded propagation delay increase along the backup path during failure.

A P2MP TE solution MUST support the P2MP fast protection mechanism to handle P2MP applications sensitive to traffic disruption.

If the ingress LSR is informed of the failure of delivery to fewer than all the egress LSRs, this SHOULD NOT cause automatic teardown of the P2MP TE LSP. That is, while some egress LSRs remain connected to the P2MP tree, it SHOULD be a matter of local policy at the ingress LSR whether the P2MP LSP is retained.

When all egress LSRs downstream of a branch LSR have become disconnected from the P2MP tree, and some branch LSR is unable to restore connectivity to any of them by means of some recovery or protection mechanisms, the branch LSR MAY remove itself from the P2MP tree provided that it is not also an egress LSR (that is, a bud). Since the faults that severed the various downstream egress LSRs from the P2MP tree may be disparate, the branch LSR MUST report all such errors to its upstream neighbor. An upstream LSR or the ingress LSR can then decide to re-compute the path to those particular egress LSRs around the failure point.

Solutions MAY include the facility for transit LSRs and particularly branch LSRs to recompute sub-P2MP trees to restore them after failures. In the event of successful repair, error notifications SHOULD NOT be reported to upstream nodes, but the new paths are reported if route recording is in use. Crankback requirements are discussed in Section 4.21.

4.7. Record Route of P2MP TE LSP

Being able to identify the established topology of P2MP TE LSP is very important for various purposes such as management and operation of some local recovery mechanisms like Fast Reroute [RFC4090]. A network operator uses this information to manage P2MP TE LSPs.

Therefore, the P2MP TE solution MUST support a mechanism that can collect and update P2MP tree topology information after the P2MP LSP establishment and modification process.

It is RECOMMENDED that the information is collected in a data format that allows easy recognition of the P2MP tree topology.

The solution MUST support mechanisms for the recording of both outgoing interfaces and node-ids.

The solution MUST gracefully handle scaling issues concerned with the collection of P2MP tree information, including the case where the collected information is too large to be carried in a single protocol message.

4.8. Call Admission Control (CAC) and QoS Control Mechanism of P2MP TE LSPs

P2MP TE LSPs may share network resource with P2P TE LSPs. Therefore, it is important to use CAC and QoS in the same way as P2P TE LSPs for easy and scalable operation.

P2MP TE solutions MUST support both resource sharing and exclusive resource utilization to facilitate coexistence with other LSPs to the same destination(s).

P2MP TE solutions MUST be applicable to DiffServ-enabled networks that can provide consistent QoS control in P2MP LSP traffic.

Any solution SHOULD also satisfy the DS-TE requirements [RFC3564] and interoperate smoothly with current P2P DS-TE protocol specifications.

Note that this requirement document does not make any assumption on the type of bandwidth pool used for P2MP TE LSPs, which can either be shared with P2P TE LSP or be dedicated for P2MP use.

4.9. Variation of LSP Parameters

Certain parameters (such as priority and bandwidth) are associated with an LSP. The parameters are installed by the signaling exchanges associated with establishing and maintaining the LSP.

Any solution MUST NOT allow for variance of these parameters within a single P2MP LSP. That is:

- No attributes set and signaled by the ingress LSR of a P2MP LSP may be varied by downstream LSRs.
- There MUST be homogeneous QoS from the root to all leaves of a single P2MP LSP.

Changing the parameters for the whole tree MAY be supported, but the change MUST apply to the whole tree from ingress LSR to all egress LSRs.

4.10. Re-Optimization of P2MP TE LSPs

The detection of a more optimal path (for example, one with a lower overall cost) is an example of a situation where P2MP TE LSP re-routing may be required. While re-routing is in progress, an important requirement is to avoid double bandwidth reservation (over the common parts between the old and new LSP) thorough the use of resource sharing.

Make-before-break MUST be supported for a P2MP TE LSP to ensure that there is minimal traffic disruption when the P2MP TE LSP is re-routed.

Make-before-break that only applies to a sub-P2MP tree without impacting the data on all the other parts of the P2MP tree MUST be supported.

The solution SHOULD allow for make-before-break re-optimization of any subdivision of the P2MP LSP (S2PL sub-LSP, S2X sub-LSP, S2L sub-LSP, X2AL sub-LSP, B2PL sub-LSP, X2AL sub-LSP, or B2AL tree). Further, it SHOULD do so by minimizing the signaling impact on the rest of the P2MP LSP, and without affecting the ability of the management plane to manage the LSP.

The solution SHOULD also provide the ability for the ingress LSR to have strict control over the re-optimization process. The ingress LSR SHOULD be able to limit all re-optimization to be source-initiated.

Where sub-LSP re-optimization is allowed by the ingress LSR, such re-optimization MAY be initiated by a downstream LSR that is the root of the sub-LSP that is to be re-optimized. Sub-LSP re-optimization initiated by a downstream LSR MUST be carried out with the same regard to minimizing the impact on active traffic as was described above for other re-optimization.

4.11. Merging of Tree Branches

It is possible for a single transit LSR to receive multiple signaling messages for the same P2MP LSP but for different sets of destinations. These messages may be received from the same or different upstream nodes and may need to be passed on to the same or different downstream nodes.

This situation may arise as the result of the signaling solution definition or implementation options within the signaling solution. Further, it may happen during make-before-break re-optimization (Section 4.10).

It is even possible that it is necessary to construct distinct upstream branches in order to achieve the correct label choices in certain switching technologies managed by GMPLS (for example, photonic cross-connects where the selection of a particular lambda for the downstream branches is only available on different upstream switches).

The solution **MUST** support the case where multiple signaling messages for the same P2MP LSP are received at a single transit LSR and refer to the same upstream interface. In this case, the result of the protocol procedures **SHOULD** be a single data flow on the upstream interface.

The solution **SHOULD** support the case where multiple signaling messages for the same P2MP LSP are received at a single transit LSR and refer to different upstream interfaces, and where each signaling message results in the use of different downstream interfaces. This case represents data flows that cross at the LSR but that do not merge.

The solution **MAY** support the case where multiple signaling messages for the same P2MP LSP are received at a single transit LSR and refer to different upstream interfaces, and where the downstream interfaces are shared across the received signaling messages. This case represents the merging of data flows. A solution that supports this case **MUST** ensure that data is not replicated on the downstream interfaces.

An alternative to supporting this last case is for the signaling protocol to indicate an error such that the merge may be resolved by the upstream LSRs.

4.12. Data Duplication

Data duplication refers to the receipt by any recipient of duplicate instances of the data. In a packet environment, this means the receipt of duplicate packets. Although small-scale packet duplication (that is, a few packets over a relatively short period of time) should be a harmless (if inefficient) situation, certain existing and deployed applications will not tolerate packet duplication. Sustained packet duplication is, at best, a waste of network and processing resources and, at worst, may cause congestion and the inability to process the data correctly.

In a non-packet environment, data duplication means the duplication of some part of the signal that may lead to the replication of data or to the scrambling of data.

Data duplication may legitimately arise in various scenarios including re-optimization of active LSPs as described in the previous section, and protection of LSPs. Thus, it is impractical to regulate against data duplication in this document.

Instead, the solution:

- SHOULD limit to bounded transitory conditions the cases where network bandwidth is wasted by the existence of duplicate delivery paths.
- MUST limit the cases where duplicate data is delivered to an application to bounded transitory conditions.

4.13. IPv4/IPv6 Support

Any P2MP TE solution MUST support IPv4 and IPv6 addressing.

4.14. P2MP MPLS Label

A P2MP TE solution MUST allow the continued use of existing techniques to establish P2P LSPs (TE and otherwise) within the same network, and MUST allow the coexistence of P2P LSPs within the same network as P2MP TE LSPs.

A P2MP TE solution MUST be specified in such a way that it allows P2MP and P2P TE LSPs to be signaled on the same interface.

4.15. Advertisement of P2MP Capability

Several high-level requirements have been identified to determine the capabilities of LSRs within a P2MP network. The aim of such information is to facilitate the computation of P2MP trees using TE constraints within a network that contains LSRs that do not all have the same capability levels with respect to P2MP signaling and data forwarding.

These capabilities include, but are not limited to:

- The ability of an LSR to support branching.
- The ability of an LSR to act as an egress LSR and a branch LSR for the same LSP.
- The ability of an LSR to support P2MP MPLS-TE signaling.

4.16. Multi-Access LANs

P2MP MPLS TE may be used to traverse network segments that are provided by multi-access media such as Ethernet. In these cases, it is also possible that the entry point to the network segment is a branch LSR of the P2MP LSP.

Two options clearly exist:

- the branch LSR replicates the data and transmits multiple copies onto the segment.
- the branch LSR sends a single copy of the data to the segment and relies on the exit points to determine whether to receive and forward the data.

The first option has a significant data plane scaling issue since all replicated data must be sent through the same port and carried on the same segment. Thus, a solution SHOULD provide a mechanism for a branch LSR to send a single copy of the data onto a multi-access network to reach multiple (adjacent) downstream nodes. The second option may have control plane scaling issues.

4.17. P2MP MPLS OAM

The MPLS and GMPLS MIB modules MUST be enhanced to provide P2MP TE LSP management in line with whatever signaling solutions are developed.

In order to facilitate correct management, P2MP TE LSPs MUST have unique identifiers, since otherwise it is impossible to determine which LSP is being managed.

Further discussions of OAM are out of scope for this document. See [P2MP-OAM] for more details.

4.18. Scalability

Scalability is a key requirement in P2MP MPLS systems. Solutions MUST be designed to scale well with an increase in the number of any of the following:

- the number of recipients
- the number of egress LSRs
- the number of branch LSRs
- the number of branches

Both scalability of control plane operation (setup, maintenance, modification, and teardown) MUST be considered.

Key considerations MUST include:

- the amount of refresh processing associated with maintaining a P2MP TE LSP.
- the amount of protocol state that must be maintained by ingress and transit LSRs along a P2MP tree.
- the number of protocol messages required to set up or tear down a P2MP LSP as a function of the number of egress LSRs.
- the number of protocol messages required to repair a P2MP LSP after failure or to perform make-before-break.
- the amount of protocol information transmitted to manage a P2MP TE LSP (i.e., the message size).
- the amount of additional data distributed in potential routing extensions.
- the amount of additional control plane processing required in the network to detect whether an add/delete of a new branch is required, and in particular, the amount of processing in steady state when no add/delete is requested
- the amount of control plane processing required by the ingress, transit, and egress LSRs to add/delete a branch LSP to/from an existing P2MP LSP.

It is expected that the applicability of each solution will be evaluated with regards to the aforementioned scalability criteria.

4.18.1. Absolute Limits

In order to achieve the best solution for the problem space, it is helpful to clarify the boundaries for P2MP TE LSPs.

- Number of egress LSRs.

A scaling bound is placed on the solution mechanism such that a P2MP TE LSP MUST reduce to similar scaling properties as a P2P LSP when the number of egress LSRs reduces to one. That is, establishing a P2MP TE LSP to a single egress LSR should cost approximately as much as establishing a P2P LSP.

It is important to classify the issues of scaling within the context of traffic engineering. It is anticipated that the initial deployments of P2MP TE LSPs will be limited to a maximum of around a hundred egress LSRs, but that within five years deployments may increase this to several hundred, and that future deployments may require significantly larger numbers.

An acceptable upper bound for a solution, therefore, is one that scales linearly with the number of egress LSRs. It is expected that solutions will scale better than linearly.

Solutions that scale worse than linearly (that is, exponentially or polynomially) are not acceptable whatever the number of egress LSRs they could support.

- Number of branch LSRs.

Solutions MUST support all possibilities from one extreme of a single branch LSR that forks to all leaves on a separate branch, to the greatest number of branch LSRs which is $(n-1)$ for n egress LSRs. Assumptions MUST NOT be made in the solution regarding which topology is more common, and the solution MUST be designed to ensure scalability in all topologies.

- Dynamics of P2MP tree.

Recall that the mechanisms for determining which egress LSRs should be added to an LSP and for adding and removing egress LSRs from that group are out of the scope of this document. Nevertheless, it is useful to understand the expected rates of arrival and departure of egress LSRs, since this can impact the selection of solution techniques.

Again, this document is limited to traffic engineering, and in this model the rate of change of LSP egress LSRs may be expected to be lower than the rate of change of recipients in an IP multicast group.

Although the absolute number of egress LSRs coming and going is the important element for determining the scalability of a solution, note that a percentage may be a more comprehensible measure, but that this is not as significant for LSPs with a small number of recipients.

A working figure for an established P2MP TE LSP is less than 10% churn per day; that is, a relatively slow rate of churn.

We could say that a P2MP LSP would be shared by multiple multicast groups, so the dynamics of the P2MP LSP would be relatively small.

Solutions MUST optimize for such relatively low rates of change and are not required to optimize for significantly higher rates of change.

- Rate of change within the network.

It is also important to understand the scaling with regard to changes within the network. That is, one of the features of a P2MP TE LSP is that it can be robust or protected against network

failures, and it can be re-optimized to take advantage of newly available network resources.

It is more important that a solution be optimized for scaling with respect to recovery and re-optimization of the LSP than for change in the egress LSRs, because P2MP is used as a TE tool.

The solution MUST follow this distinction and optimize accordingly.

4.19. Backwards Compatibility

It SHOULD be an aim of any P2MP solution to offer as much backward compatibility as possible. An ideal that is probably impossible to achieve would be to offer P2MP services across legacy MPLS networks without any change to any LSR in the network.

If this ideal cannot be achieved, the aim SHOULD be to use legacy nodes as both transit non-branch LSRs and egress LSRs.

It is a further requirement for the solution that any LSR that implements the solution SHALL NOT be prohibited by that act from supporting P2P TE LSPs using existing signaling mechanisms. That is, unless doing so is administratively prohibited, P2P TE LSPs MUST be supported through a P2MP network.

Also, it is a requirement that P2MP TE LSPs MUST be able to coexist with IP unicast and IP multicast networks.

4.20. GMPLS

The requirement for P2MP services for non-packet switch interfaces is similar to that for Packet-Switch Capable (PSC) interfaces. Therefore, it is a requirement that reasonable attempts must be made to make all the features/mechanisms (and protocol extensions) that will be defined to provide MPLS P2MP TE LSPs equally applicable to P2MP PSC and non-PSC TE-LSPs. If the requirements of non-PSC networks over-complicate the PSC solution a decision may be taken to separate the solutions.

Solutions for MPLS P2MP TE-LSPs, when applied to GMPLS P2MP PSC or non-PSC TE-LSPs, MUST be compatible with the other features of GMPLS including:

- control and data plane separation;
- full support of numbered and unnumbered TE links;
- use of the arbitrary labels and labels for specific technologies, as well as negotiation of labels, where necessary, to support limited label processing and swapping capabilities;

- the ability to apply external control to the labels selected on each hop of the LSP, and to control the next hop label/port/interface for data after it reaches the egress LSR;
- support for graceful and alarm-free enablement and termination of LSPs;
- full support for protection including link-level protection, end-to-end protection, and segment protection;
- the ability to teardown an LSP from a downstream LSR, in particular, from the egress LSR;
- handling of Graceful Deletion procedures; and
- support for failure and restart or reconnection of the control plane without any disruption of the data plane.

In addition, since non-PSC TE-LSPs may have to be processed in environments where the "P2MP capability" could be limited, specific constraints may also apply during the P2MP TE Path computation. Being technology specific, these constraints are outside the scope of this document. However, technology-independent constraints (i.e., constraints that are applicable independently of the LSP class) SHOULD be allowed during P2MP TE LSP message processing. It has to be emphasized that path computation and management techniques shall be as close as possible to those being used for PSC P2P TE LSPs and P2MP TE LSPs.

4.21. P2MP Crankback Routing

P2MP solutions SHOULD support crankback requirements as defined in [CRANKBACK]. In particular, they SHOULD provide sufficient information to a branch LSR from downstream LSRs to allow the branch LSR to re-route a sub-LSP around any failures or problems in the network.

5. Security Considerations

This requirements document does not define any protocol extensions and does not, therefore, make any changes to any security models.

It is a requirement that any P2MP solution developed to meet some or all of the requirements expressed in this document MUST include mechanisms to enable the secure establishment and management of P2MP MPLS-TE LSPs. This includes, but is not limited to:

- mechanisms to ensure that the ingress LSR of a P2MP LSP is identified;
- mechanisms to ensure that communicating signaling entities can verify each other's identities;
- mechanisms to ensure that control plane messages are protected against spoofing and tampering;

- mechanisms to ensure that unauthorized leaves or branches are not added to the P2MP LSP; and
- mechanisms to protect signaling messages from snooping.

Note that P2MP signaling mechanisms built on P2P RSVP-TE signaling are likely to inherit all the security techniques and problems associated with RSVP-TE. These problems may be exacerbated in P2MP situations where security relationships may need to be maintained between an ingress LSR and multiple egress LSRs. Such issues are similar to security issues for IP multicast.

It is a requirement that documents offering solutions for P2MP LSPs MUST have detailed security sections.

6. Acknowledgements

The authors would like to thank George Swallow, Ichiro Inoue, Dean Cheng, Lou Berger, and Eric Rosen for their review and suggestions.

Thanks to Loa Andersson for his help resolving the final issues in this document and to Harald Alvestrand for a thorough GenArt review.

7. References

7.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC2702] Awduche, D., Malcolm, J., Agogbua, J., O'Dell, M., and J. McManus, "Requirements for Traffic Engineering Over MPLS", RFC 2702, September 1999.
- [RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, January 2001.
- [RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.

7.2. Informative References

- [RFC3468] Andersson, L. and G. Swallow, "The Multiprotocol Label Switching (MPLS) Working Group decision on MPLS signaling protocols", RFC 3468, February 2003.

- [RFC3473] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.
- [RFC3564] Le Faucheur, F. and W. Lai, "Requirements for Support of Differentiated Services-aware MPLS Traffic Engineering", RFC 3564, July 2003.
- [RFC4090] Pan, P., Swallow, G., and A. Atlas, "Fast Reroute Extensions to RSVP-TE for LSP Tunnels", RFC 4090, May 2005.
- [STEINER] H. Salama, et al., "Evaluation of Multicast Routing Algorithm for Real-Time Communication on High-Speed Networks," IEEE Journal on Selected Area in Communications, pp.332-345, 1997.
- [CRANKBACK] A. Farrel, A. Satyanarayana, A. Iwata, N. Fujita, G. Ash, S. Marshall, "Crankback Signaling Extensions for MPLS Signaling", Work in Progress, May 2005.
- [P2MP-OAM] S. Yasukawa, A. Farrel, D. King, and T. Nadeau, "OAM Requirements for Point-to-Multipoint MPLS Networks", Work in Progress, February 2006.

Editor's Address

Seisho Yasukawa
NTT Corporation
9-11, Midori-Cho 3-Chome
Musashino-Shi, Tokyo 180-8585,
Japan

Phone: +81 422 59 4769
EMail: yasukawa.seisho@lab.ntt.co.jp

Authors' Addresses

Dimitri Papadimitriou
Alcatel
Francis Wellensplein 1,
B-2018 Antwerpen,
Belgium

Phone : +32 3 240 8491
EMail: dimitri.papadimitriou@alcatel.be

JP Vasseur
Cisco Systems, Inc.
300 Beaver Brook Road

Boxborough, MA 01719,
USA

EMail: jpv@cisco.com

Yuji Kamite
NTT Communications Corporation
Tokyo Opera City Tower
3-20-2 Nishi Shinjuku, Shinjuku-ku,
Tokyo 163-1421,
Japan

EMail: y.kamite@ntt.com

Rahul Aggarwal
Juniper Networks
1194 North Mathilda Ave.
Sunnyvale, CA 94089

EMail: rahul@juniper.net

Alan Kullberg
Motorola Computer Group
120 Turnpike Rd.
Southborough, MA 01772
EMail: alan.kullberg@motorola.com

Adrian Farrel
Old Dog Consulting

Phone: +44 (0) 1978 860944
EMail: adrian@olddog.co.uk

Markus Jork
Quarry Technologies
8 New England Executive Park
Burlington, MA 01803

EMail: mjork@quarrytech.com

Andrew G. Malis
Tellabs
2730 Orchard Parkway
San Jose, CA 95134

Phone: +1 408 383 7223
EMail: andy.malis@tellabs.com

Jean-Louis Le Roux
France Telecom
2, avenue Pierre-Marzin
22307 Lannion Cedex
France

EMail: jeanlouis.leroux@francetelecom.com

Full Copyright Statement

Copyright (C) The Internet Society (2006).

This document is subject to the rights, licenses and restrictions contained in BCP 78, and except as set forth therein, the authors retain all their rights.

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Intellectual Property

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at <http://www.ietf.org/ipr>.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

Acknowledgement

Funding for the RFC Editor function is provided by the IETF Administrative Support Activity (IASA).

