EVALUATION OF SCALABILITY AND BANDWIDTH EFFICIENCY OF MULTIPOINT TO MULTIPOINT HIERARCHY FOR FAST RECOVERY IN MPLS NETWORKS

Mohamad Chaitou\textsuperscript{1} and Hussein Charara\textsuperscript{1}

\textsuperscript{1}Lebanese University, Faculty of Science, Lebanon

ABSTRACT

Multi-Point to Multi-Point Traffic Engineering (MP2MP-TE) leads to an important scalability in Multi Protocol Label Switching-Traffic Engineering (MPLS-TE) networks. This paper emphasizes on the support of Fast-reroute (FRR) in MPLS-TE networks by using MP2MP bypass TE-tunnels. Hence, one MP2MP bypass TE-tunnel can be used to protect several primary TE-tunnels. During failure, the primary TE-tunnel is encapsulated into the MP2MP bypass TE-tunnel which calls for defining a new type of MPLS hierarchy, i.e. the multipoint to multipoint hierarchy. In this paper we present a simulation study that evaluates several fast rerouting scenarios depending on the number of leaves of a MP2MP bypass TE-tunnel and on the number of primary TE-tunnels that can be encapsulated into one MP2MP bypass TE-tunnel. In particular, the scalability/bandwidth efficiency tradeoff between these schemes is analyzed and valuable comparisons with the existing approaches are presented.

KEYWORDS

Simulations; MPLS; traffic engineering; fast reroute; multicast; multipoint-to-multipoint

1. INTRODUCTION

MP2MP MPLS-TE tunnels (Multi-Point to Multi-Point, Multi Protocol Label Switching-Traffic Engineering tunnels), which have been defined in [1], lead to an important scalability in MPLS-TE networks. This is due to the reduction in the number of TE-tunnels (a.k.a TE-LSP: TE-Label Switched Path) needed to maintain multipoint connectivity between edge routers.

Reducing the number of TE-LSPs is of prime importance as it improves, in the control plane, the scalability of the RSVP-TE (Resource reSerVation Protocol) signaling protocol by decreasing the memory and CPU consumed on a router.

Indeed, the scalability in the control plane is defined by the number of states required in order to establish the TE-LSPs. Each TE-LSP needs at least one state in each node it traverses. A state denotes the data information that must be stored at a node in order to maintain a TE-LSP. These states are of soft nature, that is, they should have to be refreshed at regular time intervals basis which creates a heavy burden on the routers CPUs. In fact, control plane traffic is processed by the CPU and not in hardware.

In addition, reducing the number of TE-LSPs reduces the length of MPLS tables in the data plane.
As an important application of MP2MP-TE tunnels is the support of Fast Reroute (FRR) procedures in order to protect point-to-point and point-to-multipoint TE-LSPs because they are particular cases of MP2MP TE-LSPs.

FRR is of prime importance for Multimedia applications such as VoIP, IPTV or video conferencing which have strong resiliency requirements, with a target of sub-50ms recovery upon link and/or node failure. However, a scalability problem arises when the traffic is of multicast nature such as IPTV (point to multipoint) or video conferencing (multipoint to multipoint).

FRR has been firstly defined in order to protect point to point (P2P) TE-LSPs. A primary P2P TE-LSP is protected by mean of a P2P TE-LSP, called bypass tunnel that encapsulates a backup P2P TE-LSP [2]. This is denoted as point to point MPLS hierarchy. That is, a bypass TE-LSP connecting the upstream node of the protected element, called Point of Local Repair (PLR), to the downstream node of the protected element, called Merge Point (MP), is used to encapsulate the backup P2P TE-LSP during failure.

To protect a primary point to multipoint (P2MP) TE-LSP, the use of point to point hierarchy is proposed in [3]. This method consists of establishing as many P2P bypass tunnels as the number of merge points downstream of the protected elements, which leads to a scalability problem. The use of point to multipoint MPLS hierarchy [4][5][6][7], however, helps in relieving the scalability problem since only one P2MP bypass tunnel is needed to protect an element.

Due to its bidirectional nature, the protection of an MP2MP primary TE-LSP by using point to point and/or point to multipoint hierarchy would cause a scalability problem since a full mesh of P2P or P2MP bypass TE-LSPs would have to be established between all nodes of the primary MP2MP TE-LSP adjacent to the protected element.

To overcome the above limitations, [7] proposed the use of multipoint to multipoint hierarchy i.e. the use of an MP2MP bypass TE-LSP to protect a link and/or a node of a primary MP2MP TE-LSP. Hence, one MP2MP bypass TE-LSP may encapsulate one or several MP2MP primary TE-LSPs, which is an important scalability improvement.

In addition, an MP2MP bypass tunnel can be used to encapsulate several primary P2P and/or P2MP TE-LSP which leads to an additional improvement of the scalability compared to the point to multipoint hierarchy.

Note that in [7] we focused on the signaling aspects, i.e. the control plane procedures, in order to support the multipoint to multipoint hierarchy. Also, we provided some architecture examples and presented a simple simulation study.

In this paper we present an extended simulation study in order to evaluate the performance of multipoint to multipoint hierarchy. Several scenarios are illustrated and evaluated. Indeed, in one of the scenarios we assume that the leaves of an MP2MP bypass TE-LSP cover all neighbors of a protected element while in another scenario we assume that only a subset of these neighbors, that is only the nodes of the primary MP2MP TE-LSP adjacent to the protected element, are covered.

Another important aspect reflected by the scenarios is the number of primary MP2MP TE-LSPs that can be encapsulated into one MP2MP bypass tunnel. This aspect is impacted by the selection of the node that constructs the MP2MP bypass tunnel, i.e. the node that calculates the path of the MP2MP bypass tunnel and that signals it in the control plane. We call this node “Upstream Protecting Node” (UPN). A detailed discussion about the determination of this node is given in
and it is beyond the scope of this paper. Moreover, several heuristics that can be used in order to calculate the path of the MP2MP bypass tunnel have been discussed in [1].

A scalability/bandwidth tradeoff exists between the different scenarios presented by this paper. We compare such scenarios with the use of P2P and P2MP bypass TE-LSPs by means of simulations.

Without loss of generality, the simulation model considers the protection of primary MP2MP TE-LSPs only. Indeed, primary P2P and P2MP TE-LSPs are particular cases of primary MP2MP TE-LSPs.

Our proposed model which incorporates traffic engineering in multi-point to multi-point environments is a novel contribution in order to support real time applications such as audio conferencing over IP networks [9].

This paper is organized as follows. Section 2 explains the scenarios that will be considered by the simulation model. In Section 3 we present several numerical examples to evaluate our propositions. Finally, Section 4 concludes the paper.

2. SIMULATION SCENARIOS AND ASSUMPTIONS

In this section we discuss the different scenarios and assumptions that will be evaluated by the simulation model.

We consider $l, l \geq 1$ primary MP2MP TE-LSPs to be protected as illustrated in Figure 1. These LSPs are named respectively: primary TE-LSP1, primary TE-LSP2 ... primary TE-LSP$l$. They cross a node named “P1”, which has $N_m, N_m \geq 2$ neighbors. We assume that these LSPs have the same characteristics (e.g. the same bandwidth requirements). Each of those primary LSPs crosses the same number of nodes $N_i, 2 \leq N_i \leq N_m$ among the neighbors of "P1" but not necessarily the same set of neighbors (for the sake of simplicity, Figure 1 shows that these primary LSPs cross the same set of neighbors, $N_1, N_2...N_l$. However, in the simulation model we consider the case where each primary TE-LSP traverses a number $N_i$ of nodes randomly uniformly distributed among the $N_m$ neighbors of "P1": $N_1, N_2...N_m$).
Figure 1 shows also a MP2MP bypass tunnel that is signaled such that it crosses a network connected to the neighbors of "P1" without using any link between "P1" and any of its neighbors. Hence, this bypass tunnel protects the traffic of all the primary TE-LSPs in the case of the failure of “P1” or any of the links N1—P1, N2—P1 … Nl—P1.

Note that the MP2MP bypass TE-LSP can be used to protect P2P (Nl=2), P2MP (Nl>=2 with only one source and several destinations) and MP2MP TE-LSPs (Nl>=2 with several sources and several destinations).

Another important issue is the node that builds or selects the MP2MP bypass tunnel, i.e. the node that calculates the path of the MP2MP bypass tunnel and that signals it in the control plane. A detailed discussion about the determination of this node is given in [7] and it is beyond the scope of this paper. However, it should be mentioned that this node is one of the neighbors of the node to be protected, e.g. node P1 in Figure 1, and it is determined independently for each primary TE-LSP.

For instance, how this node is determined for primary TE-LSP1 in Figure 1?

This node is one of the neighbors of P1 that are members of primary TE-LSP1 which are the nodes $N_1, N_2, ..., N_i$. We suppose that this node is randomly selected among these nodes.

Similarly, this node is determined independently for each primary TE-LSP.

In the following, this node is called Upstream Protecting Node (UPN).

In the simulation model we consider three cases for the MP2MP bypass tunnel. In addition, we take into account the cases of protecting the primary TE-LSPs by P2P and P2MP bypass tunnels respectively. Two performance criterions will be used in order to compare between these scenarios: the bandwidth wastage and the number of states caused by signaling the bypass TE-LSPs through a network that is used to bypass the primary TE-LSPs (see Figure 1). In order to simplify the simulation model and without loss of generality we can consider that this network is
reduced to one node "P2" as illustrated in Figure 2. That is, in the model we are interested by calculating the number of states in "P2". Note that in the general case, these states are spread among several nodes of the bypass network (Figure 1).

![Figure 2 The bypass network is replaced by one node](image)

**2.1. MP2MP bypass tunnel: Case 1**

In this case, we assume that regardless of the number of the primary TE-LSPs, only one MP2MP bypass tunnel exists. This has two consequences. First, the MP2MP bypass tunnel should have as leaves all the neighbors of the node to be protected, e.g. node P1 in Figure 3. In the case of Figure 3, the leaves are then the nodes $N_1, N_2, \ldots, N_m$. This is because it cannot be a priori predicted which nodes of the neighbors of P1 will be crossed by a primary TE-LSP (recall from the above discussion that each primary TE-LSP crosses a number $N_j$ of nodes randomly uniformly distributed among the $N_m$ neighbors of "P1": $N_1, N_2, \ldots, N_m$). Hence, we assume the worst case where the group of all the primary TE-LSPs crosses all the neighbors of P1.

Second, since the UPN is selected randomly and independently for each primary TE-LSP as described above, this means that it may exists several UPNs in the MP2MP bypass tunnel. For the above two reasons, Case 1 is henceforth called Full Covering (FC) multi-UPN mode.
It can be observed from Figure 3 that the bandwidth wastage occurs at nodes that do not belong to any of the primary TE-LSPs such as nodes $N_{i+1}, \ldots, N_m$.

Figure 4 below shows an example where $l = 3, N_m = 4, N_i = 3$. That is, there are three primary TE-LSPs ($l = 3$) and four neighbors for P1 ($N_m = 4$) and each primary TE-LSP crosses three neighbors of P1 ($N_i = 3$). Note that there is only MP2MP bypass tunnel and that node $N_m$ should drop the traffic of the three primary TE-LSPs.
2.2. MP2MP bypass tunnel: Case 2

In this case, the full covering assumption of Case 1 is adopted. However, we suppose that one MP2MP bypass tunnel encapsulates a subset of the primary LSPs crossing a subset of the neighbors of "P1": those that have the same UPN. In other words, for each subset of primary TE-LSPs that have the same UPN there will be a separate MP2MP bypass tunnel. Hence, Case 2 is called FC single-UPN mode.

![Diagram of MP2MP bypass tunnel](image)

Figure 5 An example of the full covering single-UPN scenario

Figure 5 shows an example where \( l = 3, N_m = 4, N_r = 3 \). Both primary TE-LSP 1 and primary TE-LSP2 in addition to primary TE-LSP3 pass through P1 and through the following subset of nodes among its neighbors: N1, N2, N3. We assume that the UPN of both primary TE-LSP1 and primary TE-LSP2 is node N1, while node N2 is the UPN of primary TE-LSP3. This implies that primary TE-LSP1 and primary TE-LSP2 will be encapsulated by the same MP2MP bypass tunnel while a second MP2MP bypass tunnel is required for primary TE-LSP3. Both bypass tunnel have the set of nodes N1, N2, N3 and N4, i.e. all the neighbors of P1, as leaves. Hence, node Nm will drop the traffic belonging to the three primary TE-LSPs

2.3. MP2MP bypass tunnel: Case 3

In this case, the full covering assumption of both Case 1 and Case 2 is no longer valid. However, we consider that one MP2MP bypass tunnel should encapsulate those primary TE-LSPs which cross exactly the same set of nodes among the neighbors of “P1”; in addition, these primary TE-LSPs should have the same UPN.
For instance, suppose that in Figure 6 below, we have 3 primary TE-LSPs and that node P1 has 4 neighbors N1, N2, N3 and N4. Suppose that: primary TE-LSP1 and primary TE-LSP2 cross P1, N1, N2 and N3 while primary TE-LSP3 crosses P1, N2, N3 and N4.

Suppose that the UPNs are: N1 for primary TE-LSP1, N2 for primary TE-LSP2 and N3 for primary TE-LSP3.

What should be the number of MP2MP bypass tunnels in this case?

The answer is three. Indeed, primary TE-LSP1 and primary TE-LSP2 cannot be encapsulated into the same MP2MP bypass tunnel although they have the same set of nodes crossing P1 (nodes N1, N2 and N3). This is because they have two different UPNs (N1 for primary TE-LSP1 and N2 for primary TE-LSP2). In addition, since primary TE-LSP3 crosses a different set of nodes among the neighbors of P1 (nodes N2, N3 and N4), it should be encapsulated into a separate MP2MP bypass tunnel.

Case 3 is henceforth called Exact Covering (EC) scenario.

The advantage of the EC mode is that the bandwidth wastage observed in the full covering modes (Case 1 and Case 2 above) is avoided. However, it incurs a control plane overhead burden which is discussed in [7].

**2.4. P2P bypass tunnels**

If the protection by P2P bypass TE-LSPs is desired then we consider a full mesh of bypass TE-LSPs between the neighbors of "P1" crossed by a primary MP2MP TE-LSP. If each neighbor of "P1" is crossed by at least one primary LSP then we will have $N_m(N_m - 1)$ P2P bypass TE-LSPs. Figure 7 illustrates an example where $l = 1, N_m = 4, N_f = 3$. There are two P2P bypass tunnels started from N1 (those in red), two P2P bypass tunnels started from N2 (those in white) and two P2P bypass tunnels started from N3 (those in blue). To define the bandwidth loss observe...
that in the particular case of Figure 7 (i.e. there exists one primary LSP), the two bypass TE-LSPs started from "N1" use the link "N1—P2". By observing that these two bypass TE-LSPs transport the same information, one may conclude that the proportion of data duplication on link "N1—P2" is 1/2. In general, this proportion is equal to \( \frac{(N_i - 2)}{(N_i - 1)} \forall N_i \). The same proportion of duplication exists on links “N2 – P2” and “N3—P2”. We define the bandwidth loss as the data duplication proportion on such links. The bandwidth loss will occur on such a link if the neighbor which it connects to "P2" belongs to at least one primary LSP.

![Diagram of Figure 7 Protection using P2P bypass tunnels](image)

**Figure 7** Protection using P2P bypass tunnels

### 2.5. P2MP bypass tunnels

Similarly for P2P bypass TE-LSPs, we consider a full mesh of P2MP bypass TE-LSPs between the neighbors of "P1" crossed by a primary MP2MP TE-LSP. For instance, in order to protect "P1" in Figure 8, we will have a total of \( N_i = 3 \) P2MP bypass TE-LSPs. It can be observed that there is no bandwidth loss in this case.
3. NUMERICAL RESULTS

In this section we investigate some numerical results by using the simulation model in order to illustrate the advantages of the MP2MP bypass TE-LSPs scenarios. Results are obtained by a simulator written in C. We emphasize on scalability and bandwidth wastage.

The bandwidth wastage has been defined in Section 2 above.

The scalability denotes the number of TE-LSP states. Let us illustrate how to derive this parameter for the different scenarios presented in Section 2.

For each segment of a bypass TE-LSP crossing a node we must add a state. For instance in Figure 4, which depicts the MP2MP FC multi-UPN scenario, the number of states required at node “P2” is equal to four. In Figure 5 (the MP2MP FC single-UPN scenario) eight states are required, while in Figure 6 (the MP2MP EC scenario) nine states are necessary. In addition, in Figure 7 (the P2P scenario) twelve states are added while in Figure 8 (the P2MP scenario) nine states are needed.

As previously mentioned, we consider that for each primary LSP, the $N_t$ nodes that it traverses are randomly uniformly distributed. Next we derive the number of states for the five scenarios: P2P bypass TE-LSPs, P2MP bypass TE-LSPs, FC-multi UPN, FC-single UPN and EC. In addition we estimate the traffic duplication in the P2P bypass TE-LSPs case and the bandwidth loss in the MP2MP case (FC-multi and single UPN cases).
Figure 9 Number of states (z-axis) as function of $N_m$ and $l_N$ for $N_l = \lfloor 0.6 \times N_m \rfloor$ (i.e. the integer part of $0.6 \times N_m$).

In Figure 9 we plot the number of states as function of $N_m$ and the number of primary LSPs. $N_l$ is chosen as the integer part of $0.6 \times N_m$. Figure 9 can be used to know the difference between the MP2MP scenarios and the P2P/P2MP scenarios as function of the degree (i.e. $N_m$) of a core node. Such study may help operators to show which scenario is suitable for their networks depending on the connectivity of the network and the number of edge nodes which is represented by the number of LSPs (i.e. one should expect that the number of LSPs increases as the number of edge nodes increases). Figure 9 shows also that MP2MP scenarios behave better in terms of scalability than the P2P and P2MP bypass TE-LSPs scenarios. In particular, the MP2MP EC-bypass TE-LSPs scenario enhances the scalability with respect to P2P and P2MP scenarios without leading to bandwidth loss. One may remark that as the degree of "P1" increases as the MP2MP scenarios become more interesting in terms of scalability. This better scalability compared to the use of P2P and P2MP bypass TE-LSPs can be explained by observing that the number of states in P2P case is in the order of $N_l^2$ and in P2MP case is of the order of $lN_l^2$ while it is proportional to $N_l$ or $N_m$ for the three MP2MP bypass TE-LSP scenarios. The number of states for the case MP2MP FC-bypass TE-LSPs multi-UPN is simply $N_m$.

The bandwidth loss is given in Figure 10. Since the difference between $N_l$ and $N_m$ is not severe ($N_l$ is less than or equal to $0.6 \times N_m$) the bandwidth loss for the MP2MP scenarios is less than the bandwidth loss for the P2P case. Figure 10 shows that as $N_m$ and the number of LSP increase, the bandwidth loss for the P2P case becomes more and more unacceptable. Indeed, as $N_m$ increases, $N_l$ will also increase (because $N_l$ is the integer part of $0.6 \times N_m$) and hence the average number of nodes, neighbors of "P1", crossed by primary LSPs will increase and hence
Surely, if the difference between \( N_l \) and \( N_m \) is high we should expect a high bandwidth loss for the MP2MP FC-bypass TE-LSPs scenarios. This is shown in Figure 11 where we consider that \( N_l = 3 \) regardless of the value of \( N_m \). In this case the bandwidth loss for the MP2MP FC-bypass TE-LSPs scenarios will continue increasing as \( N_m \) increases.

However, Figure 11 shows that the bandwidth loss in the case of the P2P bypass TE-LSPs remains greater than that of the MP2MP scenarios. One may notice (Figure 11) that even with \( N_l \) constant and for a given number of primary LSPs, the bandwidth loss in the case of P2P bypass TE-LSPs increases as \( N_m \) increases. This is because as \( N_m \) increases, the probability that the primary LSPs cross more nodes, neighbors of "P1", increases.

---

**Figure 10** Bandwidth loss for \( N_l = \left\lfloor 0.6 \times N_m \right\rfloor \)
To conclude: the EC-bypass case, which does not lead to bandwidth loss, has shown good performance also in terms of scalability. The FC-bypass multi-UPN case is the best in regards of enhancing the scalability but it leads to bandwidth loss if $N_i < N_m$. However for many realistic cases we have shown that this bandwidth loss is less than that of P2P bypass TE-LSPs. The FC-bypass single-UPN has the same bandwidth loss as the FC-bypass multi-UPN but with more number of states. Finally, in almost all cases the P2MP bypass TE-LSPs case has the worst performance in terms of scalability because the number of states is of the order of $IN_i^2$. As one may have observed, the reduction in the number states is caused by a reduction in the number of bypass TE-LSPs, i.e. scalability in both control and data plane is enhanced by using MP2MP bypass TE-LSPs.

4. CONCLUSIONS

In this paper we investigated a simulation model in order to evaluate the performance of fast reroute using multipoint to multipoint hierarchy. We have evaluated several multipoint to multipoint hierarchy scenarios depending on the number of leaves of a MP2MP bypass TE-tunnel and on the number of primary MP2MP TE-tunnels that can be encapsulated into one MP2MP bypass TE-tunnel. We demonstrated that multipoint to multipoint hierarchy leads to improve networks’ scalability compared to existing P2P and P2MP fast reroute mechanisms. Also, we studied the impact of nodes degree and the number of primary TE-LSPs on the data plane bandwidth wastage. We concluded that in most cases the proposed MP2MP scenarios lead to a good tradeoff between the scalability and bandwidth wastage when compared to P2P or P2MP scenarios. In particular, the best performance is exhibited by the MP2MP exact covering scenario.
REFERENCES


Authors

Mohamad Chaitou received an M.S. degree in computer and communications engineering from the Lebanese University of Beirut, Lebanon in 2002, and an M.S. degree in networking from the University of Pierre et Marie Curie (Paris 6), France, in 2003. In 2006, He obtained a Ph.D. degree in computer science from Telecom SudParis (TSP), France. From 2007 to 2008 He worked at Orange labs as a R&D engineer and He contributed to the development of the MPLS traffic engineering technology through several patents and research papers. He also contributed to RFC 6006 at the IETF. From 2008 to 2010, He worked as a consultant engineer for the account of Bouygues Telecom where He was involved in the convergence of IP backbone networks. Since 2010, He is an assistant professor at the Lebanese university. He is working in several research fields including: traffic engineering of enterprise networks, multicast, optical networks, performance evaluation and probabilistic modeling, cloud computing and wireless sensor networks.

Hussein CHARARA received an M.S. degree in Computer and Communications Engineering from the Lebanese University, Lebanon in 2002, and an M.S. degree in Networking and Telecommunications from the Institut National Polytechnique (INP-ENSEEIHT), France, in 2003. In 2007, He obtained a Ph.D. degree in Network, Telecom, Systems and Architecture from INP - IRIT, France. From 2006 to 2009 He worked for AIRBUS and THALES AV avionics as a R&D Engineer and PM. He contributed to the implementation & development of the Real time embedded AFDX networks such as for A380, A400M and Soukhoi RRJ programs. From 2009 to 2010, He worked for the ARECS GmbH - München where He was involved in the GPS/Galileo receivers modeling and simulations. Since 2010, He is an assistant professor at the Lebanese university. He is working in several research fields including: Traffic engineering of real time network, Performance Evaluation and QoS, Avionics and wireless sensor networks.