Investigation into Layer 3 Multicast Virtual Private Network Schemes

Muneer Ibrahim Bazama

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies in partial fulfillment of the requirements for the degree of

Master of Electrical and Computer Engineering

Under the auspices of Ottawa-Carleton Institute for School of Electrical Engineering and Computer Science

uOttawa

University of Ottawa
Ottawa, Ontario, Canada
February 2012

Abstract

The need of multicast applications such as Internet Protocol Television (IPTV) and dependent financial services require more scalable and reliable MVPN infrastructures. This diversity and breadth of services pose a challenge for operators to create an infrastructure that supports Layer 2 (ATM/Frame relay/Ethernet/PPP) and Layer 3(IPv4/IPv6) Virtual Private Networks. The difficulty is particularly true for virtual services that require complex control and data plane operations. Another challenge is to support emerging multicast applications incrementally on top of the existing Layer 3 VPN infrastructure without adding operational complexity.

In this thesis, we investigate and analyze several implementation methods of Multicast Virtual Private Network (MVPN) schemes by carrying out tests in a research testbed environment. These schemes are intended for offering multicast services over layer 3 VPN. However, some of these technologies can be tuned to offer multicast services over layer 2 VPN as well. We also provide tools and tactics on how to implement and evaluate the scalability and performance of two MPVN schemes in IP/MPLS core networks such as Rosen scheme and NG MVPN.

Acknowledgements

All praise is due to Allah, the almighty, the most merciful, the most beneficent, who has granted us knowledge and bestowed upon us good health to accomplish this work. I would like to thank my family, especially my beloved parents, may Allah bless them and admit them to Paradise,my wife Noha and my daughter Tesneem. I dedicate this for all the research time I spent away from them. Without their patience and support, I would not have been able to accomplish what I have.

I would like to extend my thanks and gratitude to my supervisor, Professor Hussein Mouftah of the School of Information Technology and Engineering at the University of Ottawa, for his advice, enduring patience, constant support and encouragement. I would also like to thank my colleagues Mr. Abdulbaset Hassan and Mr. Khaled Maamoon of the Optical Network Research Lab (ONRL) at the University of Ottawa for their productive debates and helpful insights.

Lastly, and certainly not least, I would like to express my gratitude to the staff and professors at the School of Electrical Engineering and Computer Sciences at the University of Ottawa for providing me with the means by which I was able to complete my degree.

Table of Contents

List of Figures

List of Tables

List of Acronyms

A-D	Auto-Discovery
ANT	Acreo's National Testbed
ATM	Asynchronous Transfer Mode
BGP	Border Gateway Protocol
C-RP	Customer-Rendezvous Point
CE	Customer Edge
FRR	Fast Re-Route
GE	Gigabit Ethernet
GRE	Generic Routing Encapsulation
iBGP	Internal Border Gateway Protocol
IETF	Internet Engineering Task Force
IGMP	Internet Group Multicast Protocol
IGP	Interior Gateway Protocol
IPSec	Internet Protocol Security
IPTV	Internet Protocol Television
IPV4	IP Version 4
IPV6	IP Version 6
IS-IS	Intermediate System - Intermediate System
ISP	Internet Service Provider

LAN	Local Area Network
LDP	Label Distribution Protocol
LER	Label Edge Router
LPE	Logical Provider Edge
LSA	Link State Advertisement
LSP	Label Switched Path
LSR	Label Switched Router
MAC	Media Access Control
MDT	Multicast Distribution Tree
mLDP	Multicast Label Distribution Protocol
MPLS	Multiprotocol Label Switching
MPLS-FRR	Multiprotocol Label Switching-Fast Reroute
MVPN	Multicast Virtual Private Network
MVRF	Multicast Virtual Routing and Forwarding
NG MVPN	Next Generation Multicast Virtual Private Network
NIC	Network Interface Card
NTP	Network Time Protocol
ONRL	Optical Network Research Lab
OSPF	Open Shortest Path First
P-RP	Provider-Rendezvous Point

P2MP	Point-to-Multipoint
P2P	Point-to-Point
PE	Provider Edge
PIM	Protocol Independent Multicast
PIM-DM	Protocol Independent Multicast-Dense Mode
PIM-SM	Protocol Independent Multicast-Spare Mode
PIM-SSM	Protocol Independent Multicast-Specific Source Mode
PMSI	Provider Multicast Service Interfaces
QoS	Quality of Service
RD	Route Distinguisher
RP	Rendezvous Point
RPF	Reverse Path Forwarding
RPT	Rendezvous Point Tree
RR	Route Reflector
RSVP	Resource Reservation Protocol
RSVP-TE	Resource Reservation Protocol-Traffic Engineering
RT	Route Target
SNMP	Simple Network Management Protocol
SONET	Synchronous Optical Network
SP	Service Provider

SPT	Shortest Path Tree
SUT	System Under Test
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
VLAN	Virtual Local Area Network
VM	Virtual Multicast
VPLS	Virtual Private LAN Service
VPN	Virtual Private Network
VR	Virtual Router
VRF	Virtual Routing and Forwarding

Chapter 1

Introduction

1.1 Background

The volume of multicast traffic has been growing primarily based on the emergence of video-based applications. Also there are a growing number of Layer 3(Internet Protocol) Virtual Private Network (VPN) customers that require to deliver Internet Protocol (IP) multicast service over the service provider network.

Similarly, multicast applications, such as Internet Protocol television (IPTV), Content distribution and Audio/ Video conferencing are quickly gaining popularity as is the number of networks with multiple, media-rich services merging over a shared MPLS infrastructure. As such, the demand for delivering multicast service across a BGP-MPLS infrastructure in a scalable and reliable way is also increasing [JUN09].

A Multicast Virtual Private Network (MVPN) is a technology that takes advantage of service provider networks to allow IP multicast traffic to securely traverse from one local VPN site to another geographically remote VPN site belonging to a same or different corporate.

Historically during the past years, the MVPN traffic has been carried on the top of a BGP-MPLS network using a virtual LAN model described in [ROS10]. This mode is based on dedicated virtual PE routers (Virtual Routing and Forwarding VRF concept of BGP/MPLS VPNs) and the customer traffic is tunnelled through the provider core by using Generic Routing Encapsulation (GRE) tunnels between the PEs.

Service providers adopted this approach in order to provide MVPN to their customers. As a result, they were faced with control and data plane scaling issues of an overlay model and the

maintenance of two routing/forwarding mechanisms: one for VPN unicast and one for VPN multicast service [MIN08].
For that reason, the IETF Layer 3 VPN working group published an IETF draft (2547bis-mcast) that outlines the new architecture for NG MVPNs, as well as an accompanying draft (2547bis-mcast-bgp) that proposes BGP as control plane for MVPNs and RSVP-TE P2MP as data plane for forwarding customer traffic through the provider core [JUN10].

In this thesis, we provide methods and tools that we have deployed to analyze the scalability of Rosen and NG MVPN schemes. In this work, we study the scalability problems in the control plane of MVPN schemes such as Protocol Independent Multicast (PIM) and Border Gateway Protocol (BGP) then we evaluate the two schemes through lab experimentation. For this purpose a multi-vendor testbed consisting of Juniper and Avaya equipment is established.

The following sections specify the motivations and objectives that are the key themes of this thesis. More detailed and specific solutions along with related works are provided in Chapters both 2 and 3.

1.2 Motivations

As multicast applications, such as IPTV and multimedia collaboration, gain popularity, and as the number of networks with different service needs merge over a shared MPLS infrastructure, the demand for a scalable, reliable MVPN service is rapidly increasing [JUN09].

The Rosen scheme as described in [ROS10] is solely based on virtual router architecture. With this virtual router model, control and data plane scaling issues arise, and MVPN providers must maintain two different routing and forwarding mechanisms for VPN unicast and multicast services.

The NG MVPN drafts as proposed by IETF introduced a BGP-based control plane, which is used for distributing all necessary routing information to enable VPN multicast service. The use of

BGP for distributing C-multicast routes results in the control traffic exchange being out-of-band from the data plane [JUN10b].

Up unit now, there are many service providers are still deploying the original scheme (Rosen Scheme) to distribute their multicast contents and even offer IPTV services to their customers despite the well-known scalability issues of that scheme. This is due either to interoperability problems (multi-vender deployment) or to device capability. Our motivation in this work is to study and evaluate the scalability problems that affecting each scheme, to identify which scheme that is the best candidate for being deployed in core networks and to set of mandatory steps, which can be used to provide a base for interoperable solution.

1.3 Objectives

The main objective of this thesis is to develop and implement methods and tools that investigate and analyze the scalability problems of MVPN schemes over an MPLS core network in very control environment. This is done by conducting research experiments in the Optical Network Research Lab (ONRL) at the University of Ottawa. These methods and tools allow us to study different technologies that have been proposed to solve scalability problems to delay sensitive multicast applications, such as voice and video traversing through service provider networks.

Such tools and methods help understanding the challenges of designing and implementing MVPN solutions and allowing service providers and telecom operators to choose the best interoperable solution for their networks. This is achieved by using network devices from different vendors to guarantee the interoperability option. Also, we demonstrate the advantages and disadvantages of each scheme in order to show the most scalable scheme. This is done by carry out several scenarios of both schemes on our testbed.

These scalability tests are performed in our work by increasing the number of Provider Edge (PE) routers in the network as well as the number of MVPN sites per each PE router. For generating multicast traffic, we have used a multicast tool (Multicast Hammer Tool) developed

by Nortel Networks to send and receive multicast packets. We have repeated our tests a number of times in order to ensure the accuracy of our results.

1.4 Thesis Contributions

The contributions that we have provided in this thesis can be summarized as follows:

- Novel Investigation into implementation methods of different Multicast Layer 3 Virtual Private Network technologies.
- Analyze the two most deployed Multicast VPN schemes; Rosen and NG MVPN in control environment in order to assess their scalability issues.
- Provide interoperability steps to ensure end-to-end functionality by building a testbed that consists of routers (a combination of real and emulated routers) and switches from different vendors.
- Demonstrate detailed analysis of the data obtained from both cases in which are presented to draw some conclusions with regard to the effectiveness and applicability of each scheme.
- Present key ideas and network design tips for a scalable multicast environment over service provider networks.

1.5 Outlines

The remainder of this thesis is organized as follows; Chapter 2 presents a review of the relevant background work related to Multicast mechanisms. Chapter 3 discusses multicast VPN techniques applicable in IP/MPLS network and shows the differences of and the application for each technique in a scientific perspective. It studies and evaluates new MVPN schemes that can reduce the operational and deployment in MPLS core networks. Chapter 4 presents a description of the several experiments that are conducted over the test networks as well as an analysis of the obtained results. Finally, Chapter 5 presents a conclusion for this thesis and provides some ideas for future research. Appendices A, B, and C show router configuration samples as well as some

verified working command that are used in our tests. Also, Appendix D shows confidence interval calculations for all our experimental results that shown in chapter 4.

Chapter 2

State of the Art of Multicast VPN Technologies

Multicast Virtual Private Network (MVPN) is a technology that provides multicast service over existing unicast VPN infrastructure (Layer 3 BGP-MPLS). Generally, multicast traffic is transmitted between private networks by encapsulating the original multicast packets within MPLS packets then carries it by means of multicast protocol. The need for an approach to deploy multicast services, which can use the same technology as used for deploying Layer 3 VPN for unicast services, leads scientists and researchers to conduct several experiments to develop a number of schemes that can reduce the operational and deployment risk.

In this chapter, we investigate a number of publications related to MVPN solutions. We also examine IP multicast experiments that have been performed in different multicast environments. We start with the most recent publications to demonstrate different mechanisms and structures that provide multicast solutions to service providers and enterprise networks.

2.1 Scalable IPTV Delivery to Home via VPN

In [QU10], ShuaiQu and Jonas Lindqvist discuss one of the most significant subjects related to Internet Protocol Television (IPTV) distribution, which is scalability. The authors demonstrate a solution to distribute IPTV by the use of VPN technology to remote end-users over the public IP network. This solution allows end-users over a wider geographical area to obtain IPTV service with less operational cost, enhanced security and better control.

Figure 2.1 shows the IPTV VPN network setup, which is on Acreo's National Testbed (ANT). The ANT is physically constructed on fiber optic infrastructure of the local municipality network in Hudiksvall, Sweden. The scheme presented in [QU10] is built on pure IP network to implement and distribute IPTV services to remote end-users.

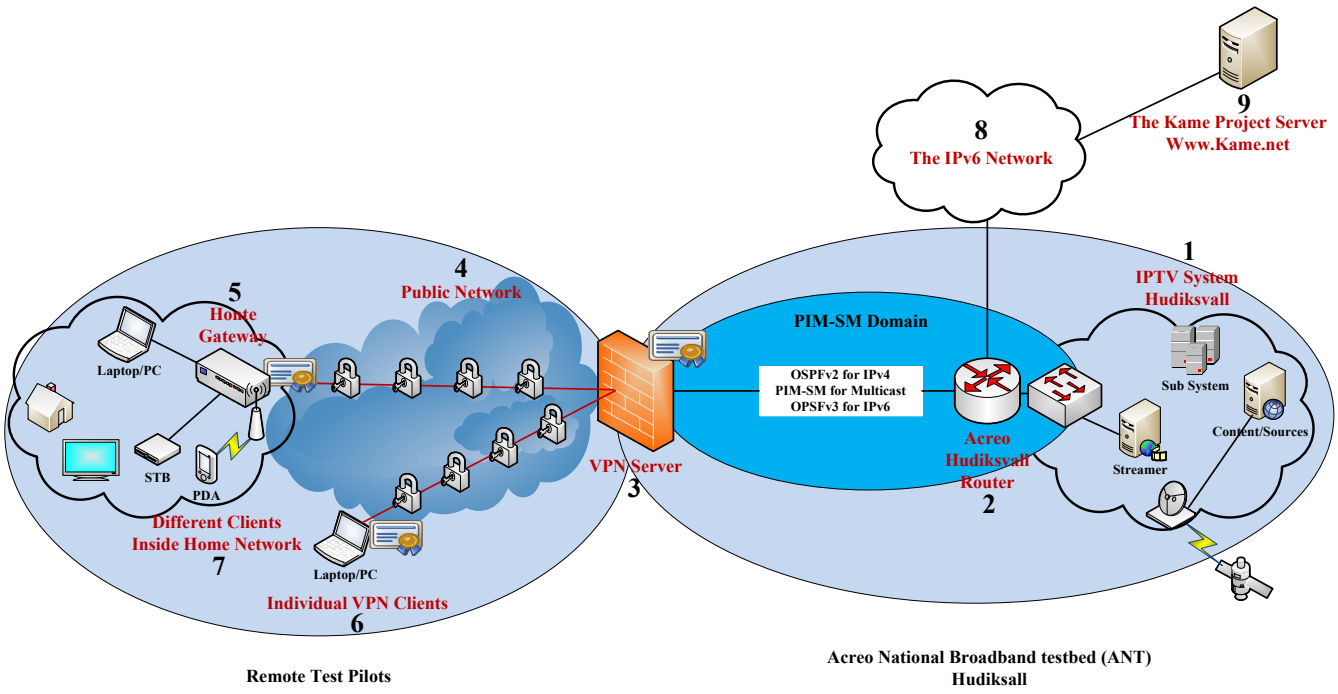


Figure 2.1: IPTV VPN Network Layout [QU10]

Additionally, the authors present measurement values for VPN connectivity such as network delay, network connectivity and capacities loss. There are six different options presented in the table above; option 1 is for IP network bandwidth test without VPN connections and option 5 is for VPN bandwidth test with data compression enabled.

The test results from the proposed scheme indicate that the qualities of IPTV service via VPN are adequate. However, we can see that there is a network capacity reduction of VPN due to network management traffic overhead. To conclude, there are several drawbacks of the proposed scheme:

Firstly, all the test cases in [QU10] have been performed and implemented on a pure IP network (not using MPLS technology). The MVPN MPLS solution, compared to other types of VPN such as (Internet Protocol security) IPsec VPN or ATM, is proved to be more cost efficient and can provide very high scalable solutions to IPTV service. In our experiments, we implemented both Protocol Independent Multicast (PIM) based and BGP based solution over MPLS network.

Secondly, the proposed scheme uses Protocol Independent Multicast-Sparse Mode (PIM-SM) as a multicast protocol, which has its own disadvantages such as scalability and performance issues. See chapter three for more details.

Finally, the approach in [QU10] does not consider any protection mechanisms in its structure (this is discussed in Chapter 4); its proposed solution only relies on Open Short Path First (OSPF) and PIM protocols to converge in case of failure.

2.2 Design and Analysis of Multicasting Experiment Based on PIM-DM

In [FEI09], Hong Fei and Bai Yu build a network experimental testbed based on Protocol Independent Multicast- Dense Mode (PIM-DM) to deliver multicast services. In their experiments, they have discovered that using PIM-DM as multicast protocol makes efficient use of bandwidth.

The PIM-DM assumes that when a source starts sending multicast data, all downstream devices are interested in receiving multicast traffic. Accordingly, the upstream PIM-DM router floods multicast datagram to all areas of the network. PIM-DM uses Reverse Path Forwarding (RPF) to prevent looping of multicast datagram while flooding. If some areas of the network do not have group members, then PIM-DM router will prune off the forwarding branch by instantiating prune state [ADA05].

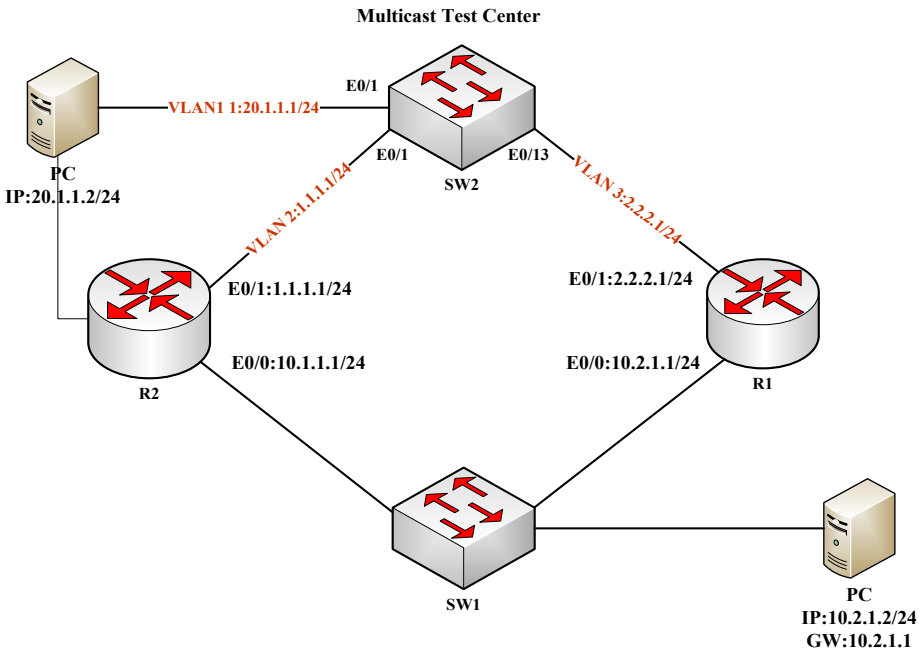


Figure 2.2: Network Setup based on PIM-DM [FEI09]

Figure 2.2 shows the network topology that has been used to provide multicast environment based on PIM-DM. the structure as is shown above, consists of two network switches and two routers with multicast enabled.

Based on several experiments, the authors in [FEI09] have discovered a number of interesting situations. They have found that PIM-SM makes flooding-pruned-flooding state change without an end receiver. Furthermore, they have noticed that the graft state is quite fast to establish for data receiver when it is initiated.

Yet, PIM-DM still may be a bit difficult to implement especially in a very large network since it has been originally designed for multicast LAN application. PIM-DM is an efficient protocol when the network is densely populated (most receivers are interested in multicast data). However, it does not scale well across larger domains in which most receivers are not interested in the data [MET10]. In our experiments, we have used PIM-SM as part of [Rosen Scheme] scheme for providing multicast services since it is the most deployed protocol by Internet Service Providers (ISP). We also show that even PIM-SM does not scale well when deployed in modern core networks.

2.3 Fast Rerouting for IP Multicast in Managed IPTV Networks

In [LUE09], Ralf Luebben et al propose and evaluate an efficient method for fast rerouting of IP multicast traffic during link failures in managed IPTV networks. The authors have developed an algorithm for tweaking IP link weights so that both the unicast and multicast routing paths between any two routers are failure disjoint. This has allowed unicast IP encapsulation for undelivered multicast packets during link failures. They also have showed that the proposed method can be realized with minor modification to the current PIM-SM [LUE09].

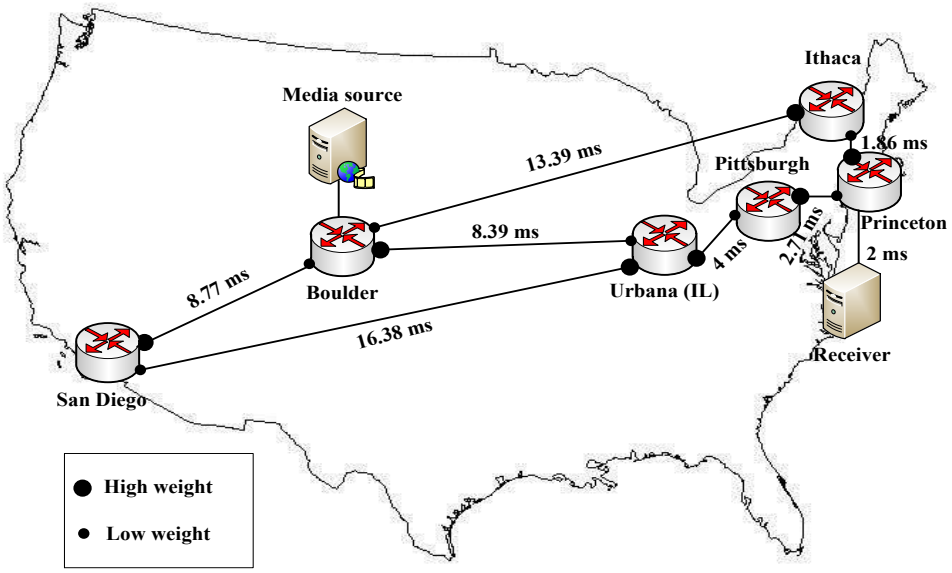
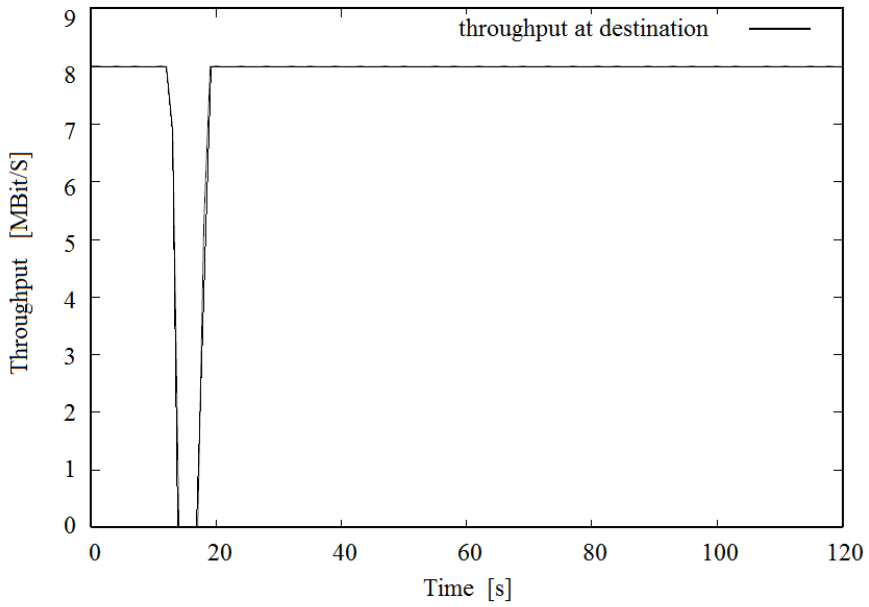


Figure 2.3: Backbone Topology [LUE09]

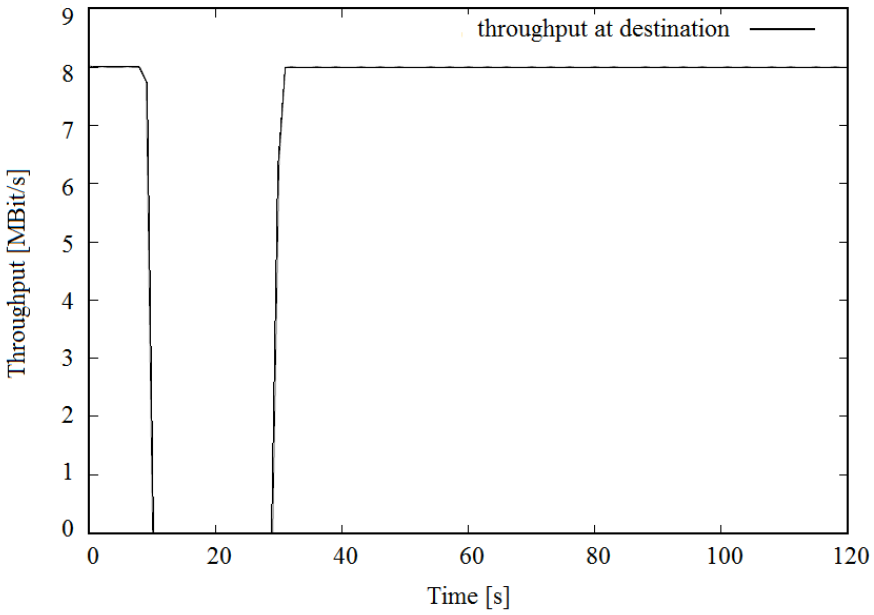
To evaluate the correctness of the proposed scheme, the authors have implemented a simplified version of the National Science Foundation Network (NSFNET) as emulation lab setup. The network topology in [LUE09] consists of six multicast routers, one media source and one receiver. The link weights have been configured as shown in Figure 2.3, and the bandwidth is set to 50 Mbps for each link.

In Figure 2.4, the diagram shows the throughput of a User Datagram Protocol (UDP) multicast stream without IP encapsulation restoration. It also shows that there has been improvement of

failure detection in conventional IP network by adjusting the *Hello* and *Dead* intervals in the Open Short Path First (OSPF) protocol.



(a) hello interval 1 s, dead interval 4 s



(b) hello interval 5 s, dead interval 20 s

Figure 2.4: Original PIM/OSPF throughput of an 8 Mbps UDP multicast stream [LUE09]

To evaluate the performance of the proposed scheme, the authors in [LUE09] have performed a number of tests to compare the PIM/OSPF throughput with and without the IP encapsulation restoration approach in case of a link failure.

To summarize, the proposed approach involves specific modifications to the existing PIM structure in order to perform an IP encapsulation process. It also requires deploying an algorithm to change the link weights to allow for disjoint backup path selection in case of a link failure. These modifications are huge disadvantages for the proposed scheme since modifying a protocol itself causes interoperability and scalability problems. Also, changing *hello* and *dead* intervals is a very risky operation as it may affect the failure detection operation in the OSPF Protocol.

In our case, we have chosen to deploy Resource Reservation Protocol-Traffic Engineering (RSVP-TE) as multicast data forwarding mechanism to generate Point-to-multipoint Label Switch Paths (P2MP LSPs).Our choice is based on the fact that P2MP LSPs have a recovery range from tens of milliseconds to fifty of milliseconds in case of failure. Also, we have kept the *Hello* and *Dead* intervals at their default values (10 and 40 seconds respectively) to guarantee the interior routing protocol stability.

2.4 Multicast Domain on E-Government MPLS VPN

In [FEN08], the authors present a general solution for IP multicast by means of MPLS VPN with multicast domains (MD) scheme. The main purpose behind this work is to allow IP multicasting of E-Government to be constructed with low cost, and high operational efficiency, as well as increase the network maintainability and controllability.

In Figure 2.5, the approach in [FEN08] shows the network setup that has been used to perform MD network testing, where PE1 has been works as a bureau core router, PE2 as the city convergence switching and PE3 as the county core switching.

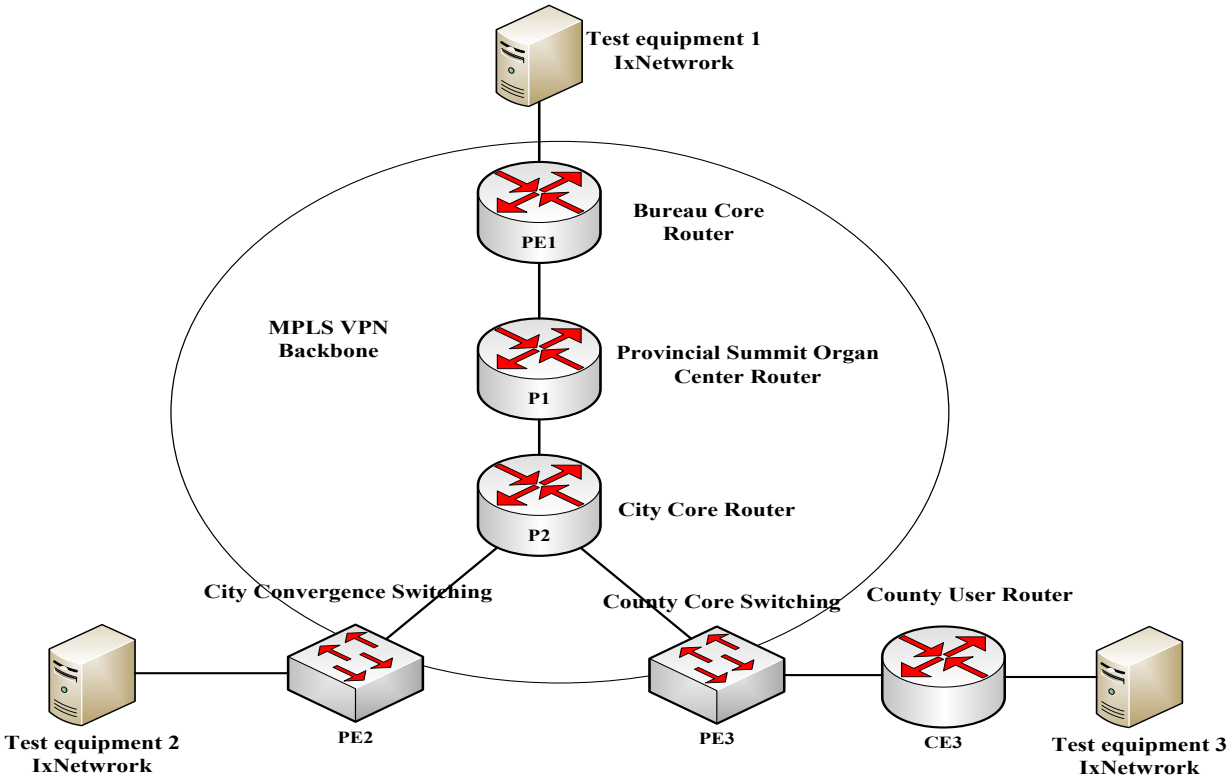


Figure 2.5: Test Network Scheme [FEN08]

Correspondingly, in Figure 2.6, the authors have presented results that demonstrate the bandwidth capacity and stability of multicast in MPLS VPN network MD system as well as the maximum loss rate. The frame receive rates are all above 99.99%, the maximum loss rates are below 7 frames/s and the total send rates are around 145350 frames/s.

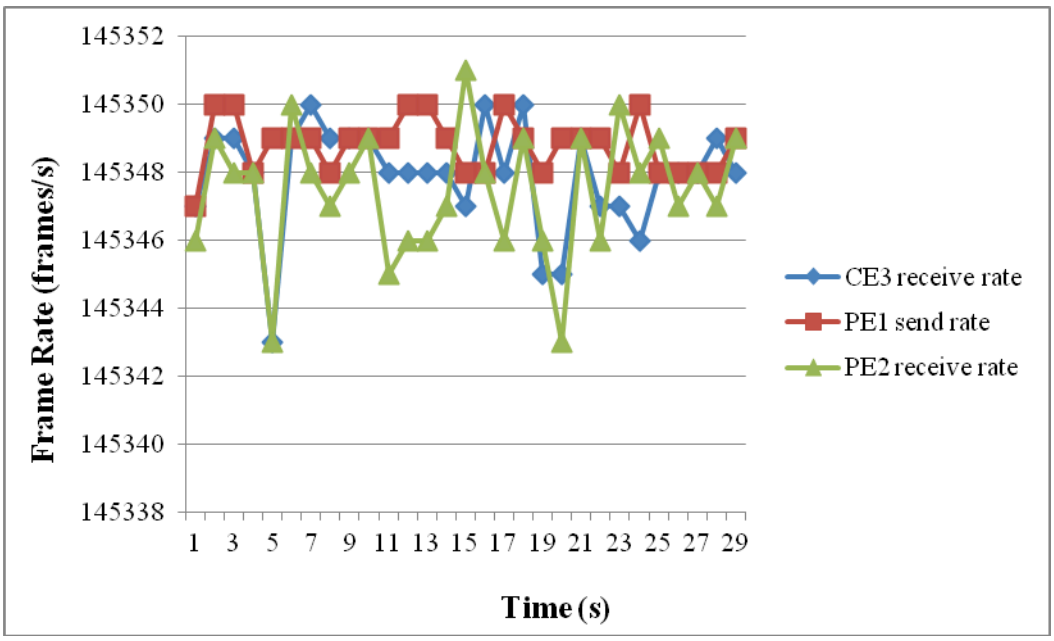


Figure 2.6: Bandwidth capacity and stability of MVPN [FEN08]

One of the major attractions of using MD solution is that the Provider (P) routers do not need hardware or software upgrade to enable new multicast features to support Multicast VPNs. Only native multicast is required in the core network to support multicast domains. Therefore, the operational risk is minimized in the service provider network when deploying MDs. However, doing so, places heavy load on P routers since they must participate in the MD and maintain forward state for each Multicast Distribution Tree (MDT) created.

Additionally, deploying MD requires the Customer domain and Service Provider (SP) domain to run PIM instances independently, which consumes memory and requires extensive CPU cycles on a router. In Chapter 4, we show the benefit of using Next Generation-Multicast Virtual Private Networks NG-MVPN (BGP based) approach since it does not put any burden on P routers as well as requiring less overhead in the control plane.

2.5 Delivering Multicast Services over MPLS Infrastructure

The authors in [ALA07] present a performance analysis of conveying multicast services across MPLS infrastructure networks. The multicast flows are carried via Layer 2 VPN Virtual Private LAN Services (VPLS) technology as transport media. Furthermore, the authors investigate the

reliability of multicasting over VPLS and compare it with other multicast protocols such as PIM in terms of jitter and delay.

In Figure 2.7, two multicast testbed setups have been formed; one based on PIM-SSM protocol and the other based on VPLS technology. The primary interest of these setups is measuring the performance of multicast traffic throughput, delay, jitter, and packet loss in the VPLS based and PIM-SSM based (Specific Source Multicast) network. For that reason several multicast experiments to evaluate the VPLS multicast performance against PIM have been conducted.

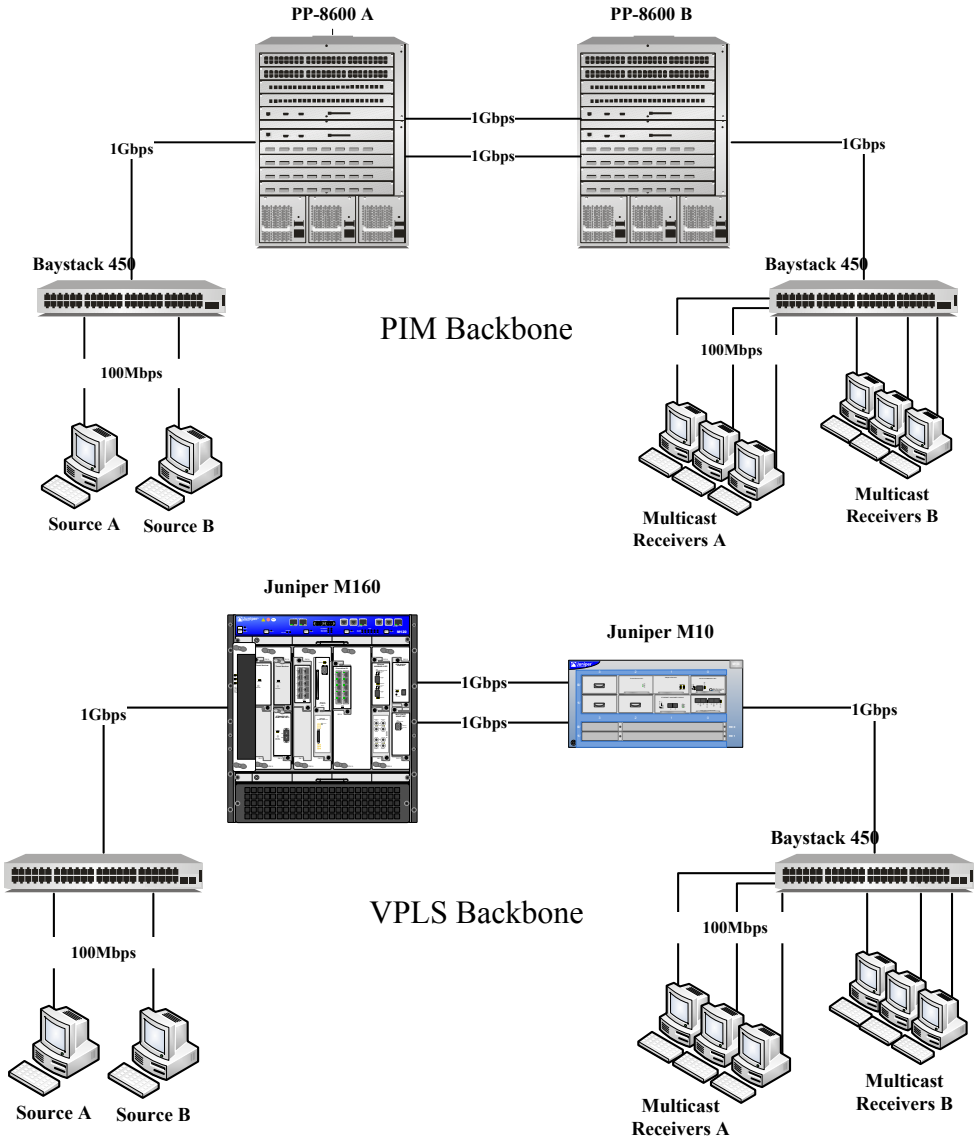


Figure 2.7: ONRL Testbed Setup [ALA07]

Figure 2.8 illustrates the percentage of the bandwidth wasted for the number of concurrent multicast sessions. As we can see, the implementation of PIM based approach requires more bandwidth to handle multicast traffic than the VPLS based approach. The bandwidth wasted is simply due to packets received in errors or collided.

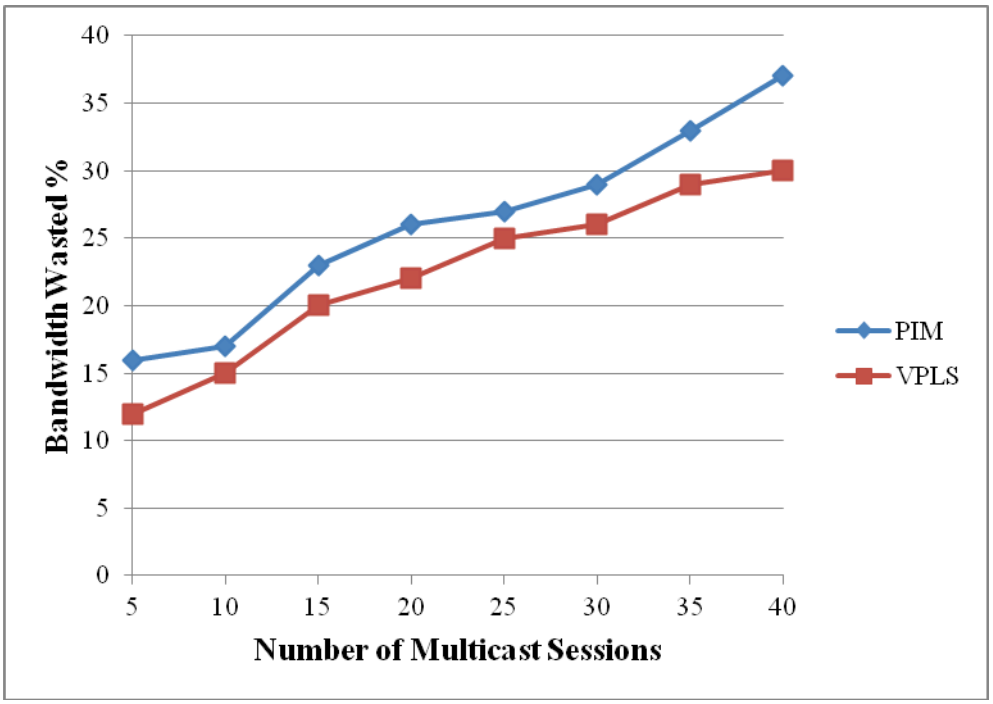


Figure 2.8: Bandwidth waste vs. No. of concurrent multicast sessions [ALA07]

Additionally, the authors mention that the restoration process takes approximately two to three seconds to switch to the backup link in case of PIM based approach. While in case of VPLS, it takes less than 50 ms, this being due to the performance of Fast Reroute mechanism in MPLS networks.

We conclude that PIM based approach does not promise sub-second convergence time due to its dependence on integrated routing protocols. Also, it provides limited control over the traffic paths, as many services depending on multicast routing tend to follow a similar path (the shortest path as computed by the IGP) [JUN09]. However, the VPLS approach as well has its own limitations. For instance, since VPLS implementations use ingress replication, it naturally does not offer bandwidth efficiency for multicast traffic [JUN09]. Therefore, in order to provide true

multicast services through VPLS, it must use Point to Multipoint Label Switch Paths (P2MP LSPs) with BGP VPLS, which allows replication in the network only where it is required.

2.6 A Scalable and Efficient Client Emulation Method for IP Multicast Performance Evaluation

In this work [WAT07], the authors propose a cost effective approach (client emulation method) for evaluating IP multicast forwarding performance of network devices by using an IEEE 802.1Q Tagged Virtual Local Area Network (VLAN) technology. This is achieved by combining a layer 2 (L2) switch with an Ethernet interface of a single PC. Furthermore, this paper describes the benefit of the proposed scheme for evaluating layer 2 and 3 devices and its ability for monitoring specific traffic destined for arbitrary clients.

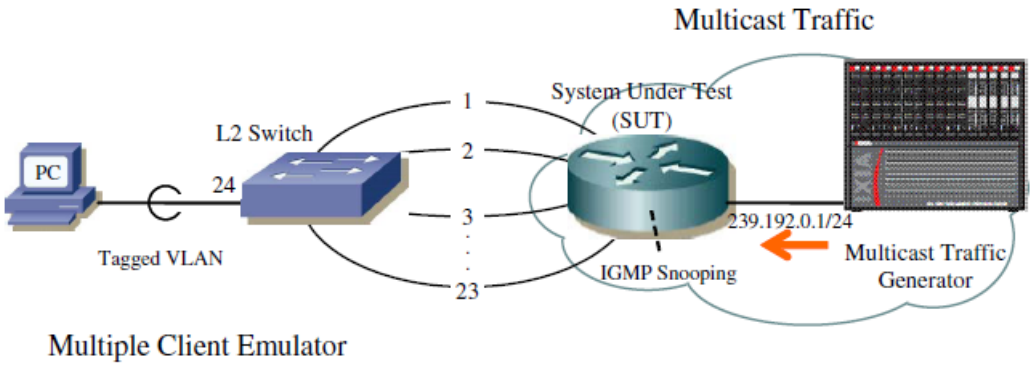


Figure 2.9: Client Emulation Conformance Test [WAT07]

The approach in this paper is generally based on the idea of tagged VLAN technology. The interface between PC and L2 switch shown in Figure 2.9 is a tagged (trunk) interface, while the interfaces between the L 2 switch and the System Under Test (SUT), which acts as Customer Edge (CE) device, are the untagged interfaces.

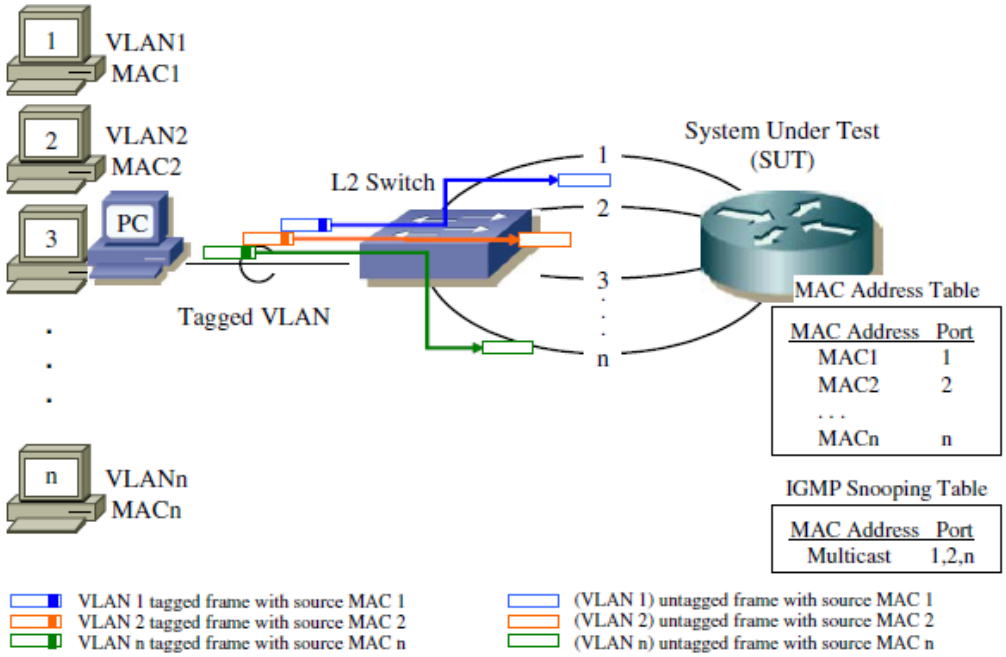


Figure 2.10: Forwarding of each Tagged VLAN Frame [WAT07]

Figure 2.10 shows how each Ethernet frame transmitted from one of the VLAN logical interfaces of the PC is given a corresponding tag to identify its network. For each transferred frame, the L2 switch examines the tag to determine the corresponding forwarding interface. The tag is then removed and transmitted through the corresponding interface of the layer 2 switch. Since the output frame is now a regular Ethernet frame, the SUT processes the frame as unique IGMP query [WAT07].

In our conclusion, we see the purpose of this paper is to introduce a method that diminishes the cost and eliminates the need of large number of clients in multicast testing environment. However, we find that the proposed approach has several drawbacks:

1) It affects network scalability since it introduces a single point of failure as well as introduces complexity design at customer side when deployed.

2) It is a costly design since it requires the SUT device (a router) to have as many interfaces as the number of clients (the cost of an Ethernet port on a router is greater than the cost of a network card on a PC).
3) The number of clients will be limited due to the fact that every network card is designed to have a limited buffer and can handle a certain amount of traffic. It also introduces the idea of MAC addresses collision.

In our test case, we have used a number of PCs equipped with four Gigabit Ethernet Network Interface Cards (NICs) acting as multicast receivers in order to distribute the multicast traffic on multiple PC and not creating a single point of failure. For a large deployment, it is possible to have multiple GE Quad ports NICs (e.g Intel PRO/1000 GT Quad Port Server Adapter) installed on a server to create a large number of clients without affecting the network scalability.

2.7 VMScope – A Virtual Multicast VPN Performance Monitor

Breslau et al. [BRE06] discuss the MVPN monitoring problem by introducing a virtual MVPN performance monitor (VMScope) designed to enable SP to monitor and troubleshoot their MVPN service. The VMScope uses tunnelling technology to establish virtual connections to routers in the provider network and uses these tunnels to enable remote monitoring of customer.

The main purpose of VMScope is to enable the SP to measure customer performance across the MVPN and provides a critical performance monitoring and troubleshooting tool. This is done, by maintaining a virtual P2P GRE tunnel with each PE router that it will monitor, as shown in Figure 2.11.

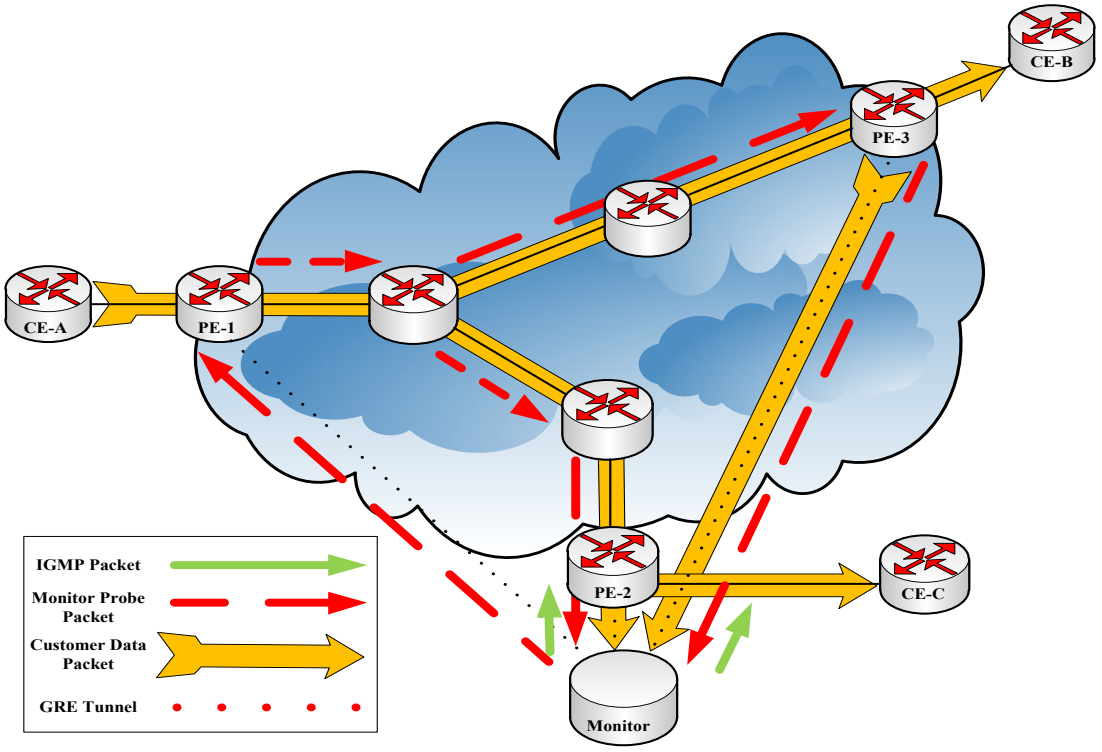


Figure 2.11: VMScope Operation [BRE06]

To conclude, the demonstrated results from testing the proposed structure are promising. However, there are some drawbacks; First, to prove the system's scalability, VMScope needs to be tested for a larger topology (two PE routers are not enough). Second, establishing point-to-point (P2P) GRE with each PE needs to be taken into consideration since this will put a lot of burden on both PEs and the monitoring tool.

In Chapter 4, we demonstrate our results, which have been obtained directly from routers log to get more accurate and realistic results. Moreover, network venders nowadays offer management tools that are built-in platform to monitor.

2.8 A Solution for IP Multicast VPNs based on Virtual Routers

The authors of [CHU06] propose a solution for IP multicast VPN based on the use of Virtual Routers (VR). The main idea is that all multicast traffic from the same VR is forced to share the same MDT. This is done by deploying Source Specific Tree/Shared Tree with the VR that is acting as source or RP. The authors then detail this solution according to process in local

customer sites, method of establishing multicast distribution trees in SP networks and forwarding flow of multicast data packets. The analysis shows that the mechanism can effectively reduce the router multicast states on SP backbone network, and ensures the bandwidth resource utilization.

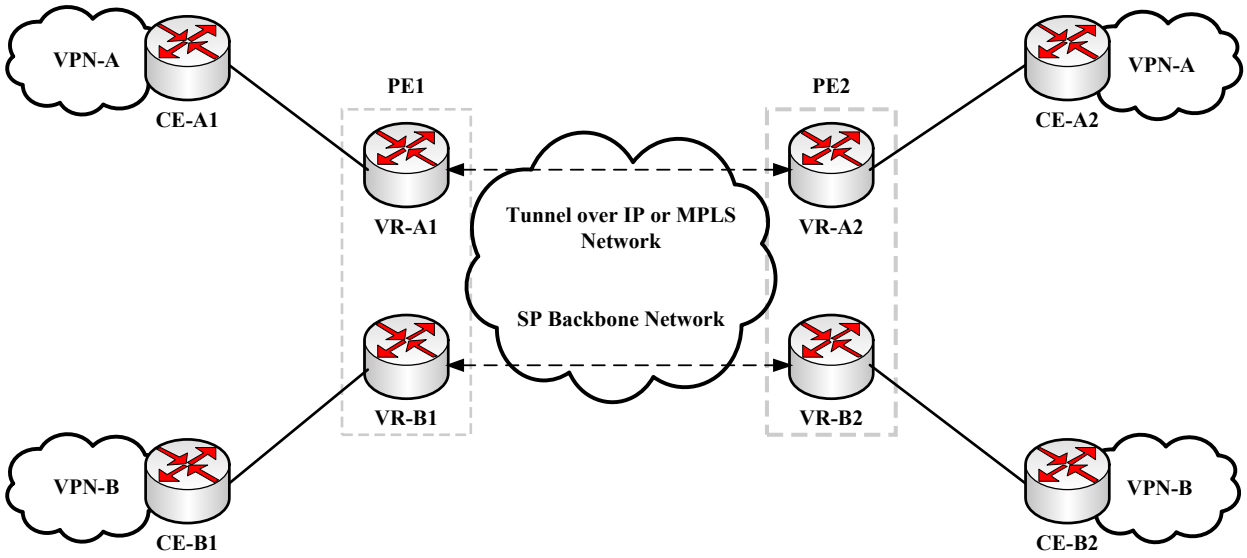


Figure 2.12: VR to VR connectivity through IP or MPLS tunnels [CHU06]

The VR-VPN model in Figure 2.12 allows customer sites to contact the SR backbone via the connection between Customer Edge (CE) device and VR. Furthermore, the CE device can connect to VR via any access link, and then forward all multicast traffic to the local VR. The VRs that belong to the same VPN domain must discover VPN membership and distribute reachability information. VRs only maintain route state for the VPN they belong to and they can be configured on one Provider Edge (PE) device. Furthermore, the authors in [CHU06] discuss several aspects of performance to evaluate the scheme such as scalability, resource optimization and security.

The authors argue that in order to improve the scalability, an aggregated tree must be deployed on a VR that acts as a Rendezvous Point (RP) in SP network. This is due to the fact that the larger number of multicast trees in the core, the less scalable network becomes. Additionally, the tree maintenance can be much less frequent process than in traditional multicast. So it does not only reduce payload on SP network effectively, but it also improves scalability.

One of the shortcomings of this scheme is implementing the Virtual Router model rather than deploying the aggregated routing model for VPN services. Another shortcoming by introducing different mechanisms in both control and data plans. Therefore, the proposed scheme does not only introduce a second set of mechanisms for multicast services, but also it affects the scalability and flexibility of the unicast VPN solution.

In contrast, we have deployed the Multicast BGP/MPLS-based approach, which reuses Layer 3VPN (L3VPN) unicast mechanisms with extensions for multicast. This approach retains as much as possible the scalability and flexibility of the unicast without adding additional mechanisms to the L3VPN solution.

2.9 Delivering Reliable Real-Time Multicast Services over Virtual Private LAN Service

Biradar et al. [BIR06] investigate the performances and reliability of multicasting by using VPLS technology as transport media. This is achieved by deploying multicast services such as delivery of high-quality video and audio conferencing on a VPLS network testbed.

The main goal in [BIR06] is to present the benefit of using VPLS technology as transport media for multicast services. The authors investigate the performance over IP/MPLS optical Ethernet public network to provide high reliability for transmitting IP Multicast applications. This is done by providing redundancy between end-points through MPLS protection and restoration techniques such as MPLS-Fast Reroute (MPLS-FRR).

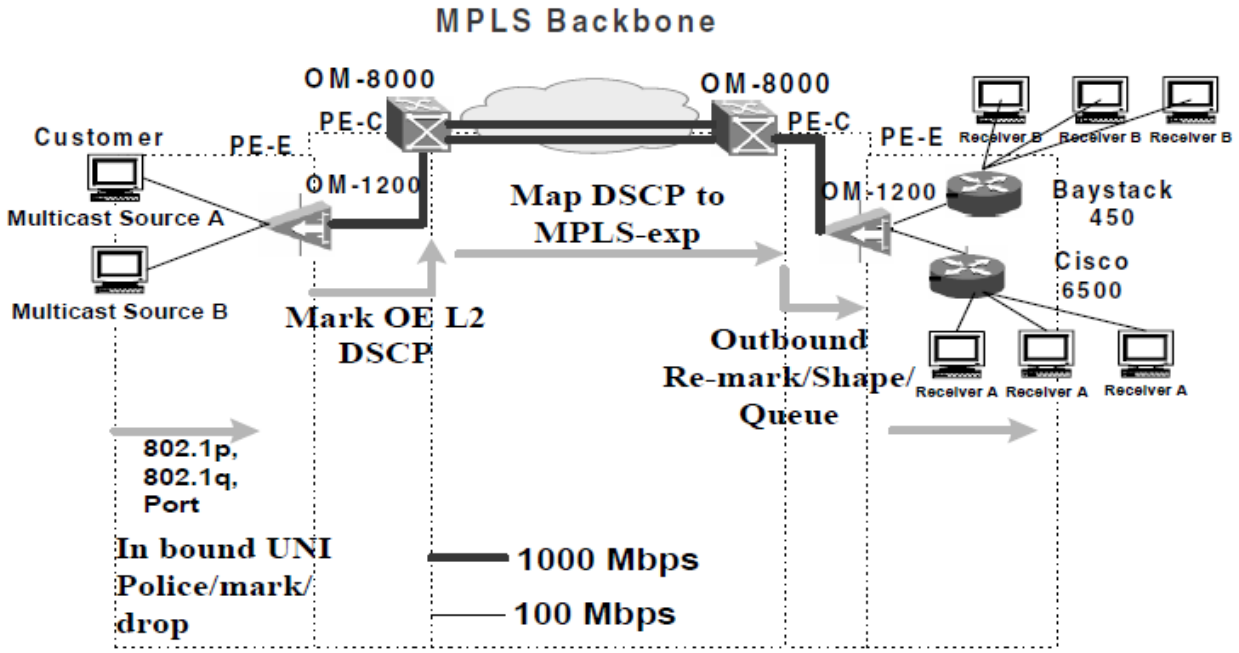


Figure 2.13: ONRL VPLS Testbed [BIR06]

Figure 2.13 demonstrates the relationship between PE-E and PE-C devices in the Logical Provider Edge (LPE) theoretical model, which has been originally proposed by Nortel Networks [draft-knight-l2vpn-lpe-ad-00.txt] and implemented according to the IETF model.

The PE-E acts as gateway between provider and customers, while the PE-C represents the provider edge router in the core. Furthermore, signalling between the PE-E and the PE-C controls the Auto-Discovery (A-D) process; this allows devices belonging to the same VPLS site to automatically discover each other across the MPLS network.

In Figure 2.14, it shows that the multicast packet loss gradually increases with the increase in the multicast rate for a given traffic load. This behaviour is due to a router/switch limited packet buffering or queuing capacity [BIR06] as well as due to the use of Label Distribution Protocol (LDP) instead of Resource Reservation Protocol- Traffic Engineering RSVP-TE with Fast Reroute mechanism.

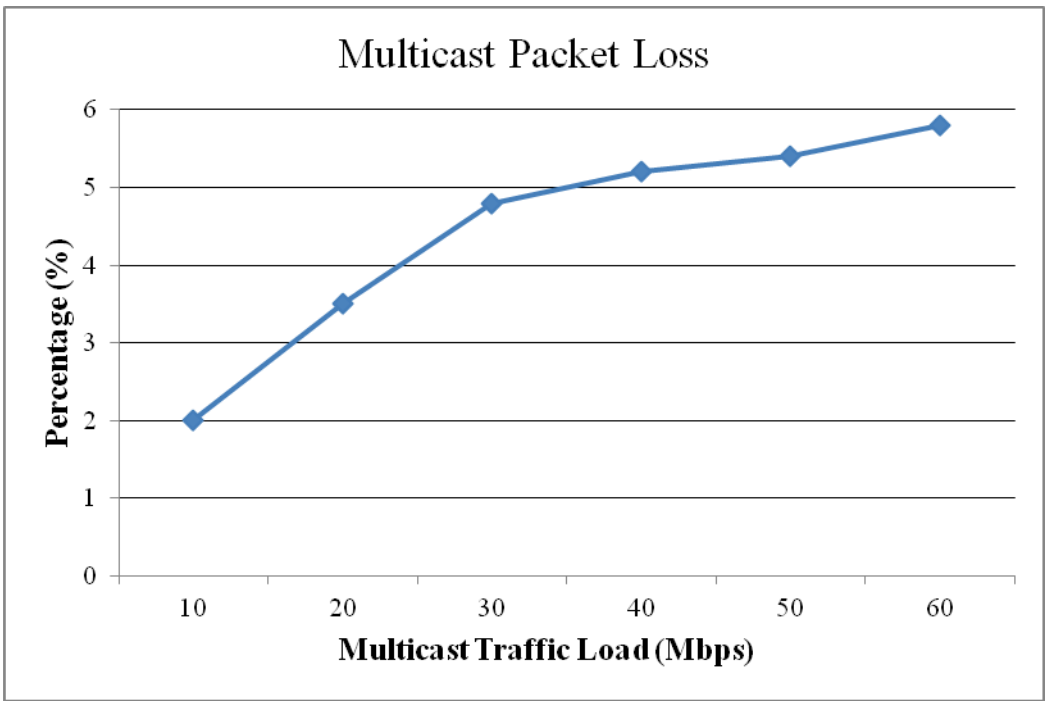


Figure 2.14: Multicast Packet Loss with various traffic loads [BIR06]

In our conclusion, the proposed solution in [BIR06] has a few drawbacks:

1) The solution uses VPLS implementation based on LPE theoretical model, which does not consider deploying any protection mechanisms such as MPLS-FRR. Instead it relays on IGP to converge in case of link or node failure.

2) As we mentioned earlier that VPLS implementations in general use ingress replication methods, which naturally do not offer any bandwidth efficiency for multicast traffic [JUN09]. Therefore, in order to provide true multicast services through VPLS, it must make use of P2MP LSPs with BGP VPLS, which allows replication on the network only where it is required.

2.10 Tunneling Multicast Traffic through Non-Multicast Aware Networks and Encryption Devices

The authors in [HIG01] propose a scheme that would allow the multicast traffic flows through several encryption devices and traverse a non-multicast environment. This is achieved through the use of GRE tunnels. The authors develop a solution that allowed dynamic routing of

multicast traffic throughout the network. In addition, this scheme is proposed as a solution to run multicast applications across networks that do not support multicast traffic.

Figure 2.15 presents the multicast tunnelling architecture that was adopted as a solution to satisfy the requirements of running multicast traffic over non-multicast environment. The M mark in the diagram below represents the gateway routers that provide site-to-site access via Wide Area Network (WAN). The proposed scheme in [HIG01] uses GRE tunnelling technology to hide the multicast addresses from encryption devices and non-multicast routers. Further, it associates Enhanced Internet Gateway Routing Protocol (EIGRP) with each GRE tunnel interface on a (M) router.

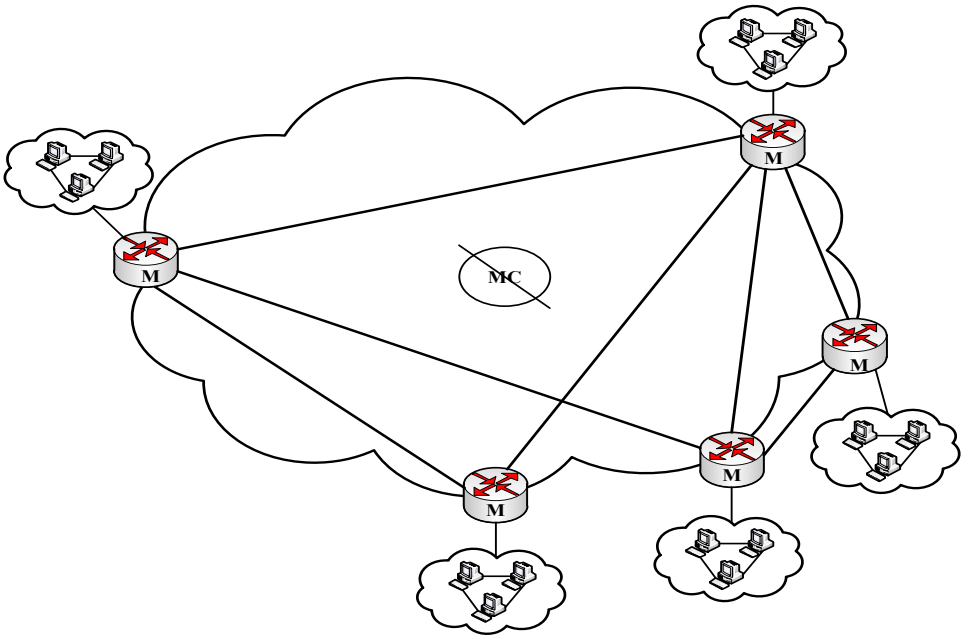


Figure 2.15: Multicast Tunnelling Architecture [HIG01]

To conclude, we can see that this proposed architecture is only valid for special cases such as the one presented in [HIG01] and it cannot be adopted as an enterprise solution for the following reasons:

First, P2P GRE tunnelling technology is designed specifically to carry unicast traffic, thus deploying it for multicast case would cause extensive overhead in the core since each router has to establish P2P GRE tunnels with all other routers in the network.

Second, as the authors in [HIG01] state that the encryption devices cannot see the true source and destination IP addresses, it is not possible to control which nodes are allowed to communicate with each other.

Last, if a router has tunnels to n other routers, then the configuration commands must be repeated manually n times on that router. For example, in a fully meshed network of n LANs, the configuration commands would be repeated n (n-1) times.

Chapter 3

Multicasting in Layer 3 VPNs

In this Chapter we demonstrate different MVPN schemes that we have employed on our experiments such as Rosen Scheme [ROS10] and NG MVPN [AGG09]. We start off by showing the operation mode of each scheme, and then we discuss the scalability issues of each technique and continue with presenting the advantages and disadvantages for each scheme when they are being deployed in MPLS core networks. Additionally, we present the recovery technique that each scheme uses when a failure occurs.

3.1 Introduction

Due to the rapid growth of multicast applications, such as IPTV and media-rich services merging over a shared MPLS infrastructure, MVPN technology has become an essential tool for providing multicast service over an existing Layer 3 VPN or as part of a transport infrastructure.

The existing Layer 3 BGP/MPLS IP VPN service, which was described in [ROS06], provides a VPN implementation with an optimal unicast routing through the SP backbone [ROS10]. Unfortunately, it does not support multicast VPNs; it, along with its IPv6 companion standard [CLE06], only supports IP unicast between VPN customer sites [ALC10]. Thus, to provide a scalable MVPN solution, two requirements must be fulfilled:

1. The need of a mechanism to carry multicast routing information, which acts as control plane. This information is flowed from the provider edge (PE) routers connected to the sites that contain the receivers to the PEs connected to the sites containing the sources [JUN10b]. This path allows the sources to know which receiver sites are interested in receiving multicast traffic.

2. A mechanism to carry multicast traffic, which acts as data plane. This information is carried from the PE routers connected to the sites that contain the sources to the PEs connected to the sites that contain the receivers [JUN10a]. This path enables the flow of multicast traffic from the sources to the receivers.

In the next section, we discuss the deployment of two MVPN schemes in order to provide multicast services over MPLS networks.

3.2 The ROSEN Scheme (PIM/GRE MVPN)

The Rosen scheme is a mechanism that provides multicast services over the existing unicast BGP/MPLS IP VPN model. This solution introduces the idea of Multicast Domains "MD". Where a PE is attached to a particular multicast-enabled VPN is said to belong to the corresponding MD.

This scheme uses either PIM-SM or PIM-SSM for exchanging control plane information, as well as setting up multicast forwarding state on the Layer 3 VPN infrastructure. The Customer domain multicast (C-multicast) protocol information, typically Customer PIM (C-PIM) join/prune messages, received from local Customer Edge (CE) routers is propagated to other PEs using these PE-PE PIM sessions across the VPN-specific virtual network [JUN10].

For each MD, there is a default Multicast Distribution Tree (MDT) through the backbone, which connects all PEs that belonging to that MD. Any given PE may be participating in as many Multicast Domains as there are VPNs attached to that PE, and each MD has its own MDT [ROS10]. The MDTs generally are multipoint GRE tunnels, which created by running different instance of PIM protocol on the provider side (P-PIM).

The GRE header used for tunnelling VPN multicast data and control traffic is a multicast group address assigned to that VPN by the provider. This multicast group address gives the GRE the multipoint property as these tunnels are, in fact, VPN multicast distribution trees (MDTs) [JUN09b]. The Default MDTs are constructed automatically as the PEs in the domain come up

and they do not relay on the existence of multicast traffic in the MD. If the provider uses Data MDTs, then these must be unique for each customer [CIS02].

3.2.1 Operation Mode of Rosen scheme

The Rosen scheme operates in two modes; PIM-SM and PIM-SSM modes. Each mode has its own advantages and disadvantages. It is said that the former mode is widely deployed by SPs because of its flexibility and performance. However, the PIM-SSM is gaining popularity because of the need of streaming applications such as IPTV and the ease of deployment.

3.2.1.1 PIM-SM MVPN

The PIM-SM mode uses shared trees and rendezvous point (RP) for auto-discovery of the PE routers. The PE that is the source of the multicast group encapsulates multicast data packets into a PIM register message and sends them by means of unicast to the RP router. The RP then builds a shortest-path tree (SPT) toward the source PE. The remote PE that acts as a receiver for the MDT multicast group sends (*, G) join messages toward the RP and joins the distribution tree for that group [JUN09]. CE routers do not have a PIM adjacency across the provider network with remote CE routers, but rather have an adjacency with their local routers and the PE router [CIS02].

The Figure 3.1 displays the traffic forwarding in PIM-SM mode, where a multicast source in site 1 of VPN A sends a multicast traffic to receivers (that are interested in receiving multicast traffic) in sites 2 and 3 of the same VPN. The traffic is delivered by means of a Default MDT (multipoint GRE tunnel) from PE1 to PE2 and PE3 and from there to the appropriate CEs, which are in this case CE2 and CE3.

On the other hand, the Data-MDT is used only to distribute multicast data that exceeds a certain configured threshold of Bandwidth (BW) to only PEs that have joined unique multicast groups. The word “Data” is appended as these multicast groups designed to be used for customers require a higher amount of bandwidth to deliver their data [CIS04]. In Figure 3.1, the diagram

shows a Data-MDT has been configured only to serve a customer connected to C3, which requires higher
BW.

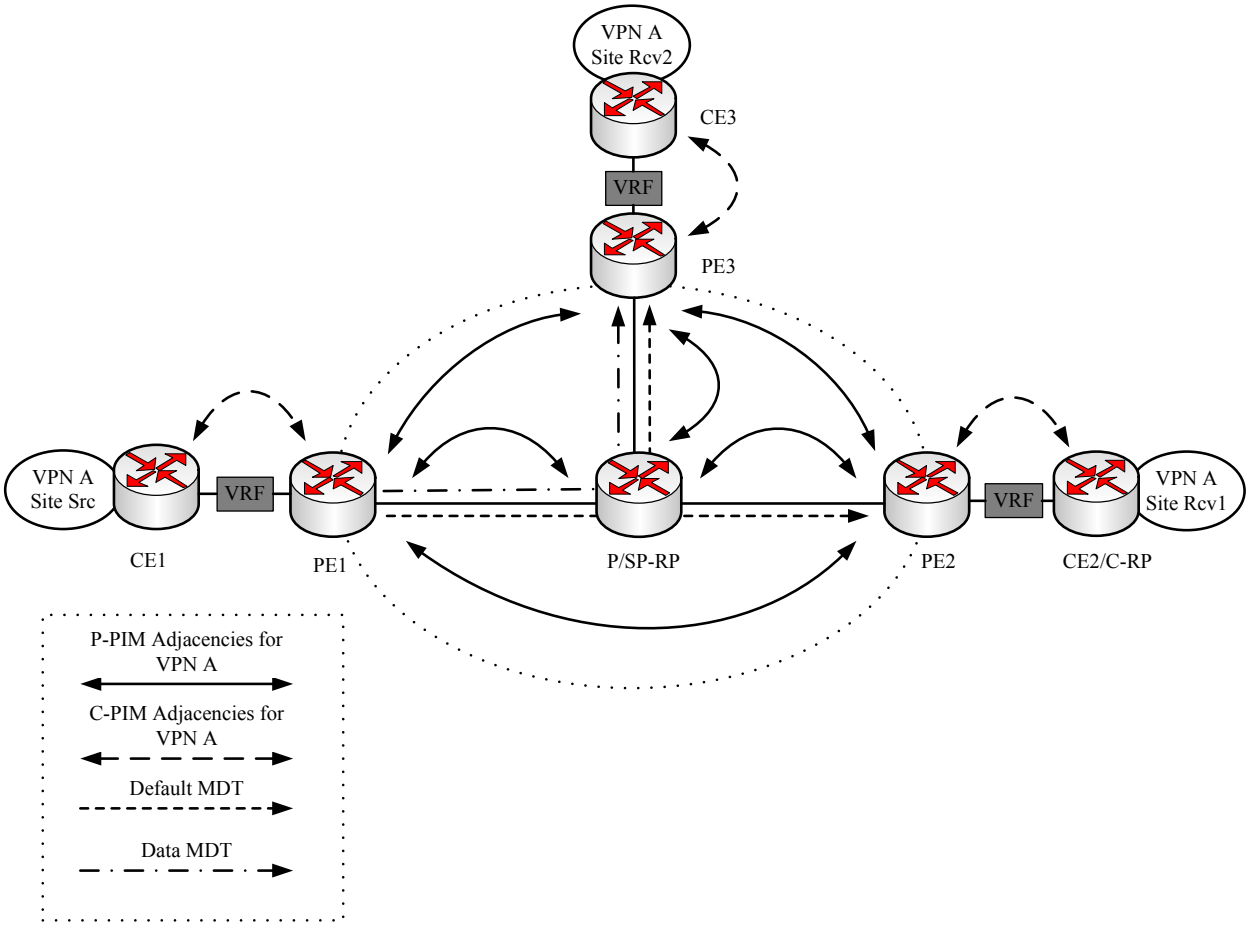


Figure 3.1: PIM-SM/GRE MVPN Operation Mode

3.2.1.2 PIM-SSM MVPN

In the contrary, the PIM -SSM mode uses BGP signalling for auto-discovery of the PE routers. Each PE sends an MDT-Subsequent Address Family Identifier (MDT-SAFI) BGP Network Layer Reachability Information (NLRI) advertisement as described in [NAL06]. The advertisement contains the following:

- The Route Distinguisher (RD)

- Unicast address of the PE router to which the source site is attached (usually the loopback)
- Multicast group address
- Route Target extended community attribute (RT)

Each remote PE router imports the MDT-SAFI advertisements from each of the other PE routers if the route target matches. Each PE router then joins the (S,G) tree rooted at each of the other PE routers. After a PE router discovers the other PE routers, the source and group are bound to the VRF through the multicast tunnel de-encapsulation interface [GAG10].

To propagate the PIM information between the PE routers and onwards to other VPN sites, PIM adjacencies are set up between the PEs that have sites in the same VPN (per-VPN Routing and Forwarding VRF basis). These per-VRF PIM adjacencies ensure that the necessary multicast trees can be set up within each customer VPN [MIN08].

A unique source address for the multicast packet in the provider network is required. This source address is recommended to be used as the source for the iBGP, as this address is used for the Reverse Path Forwarding (RPF) check at remote PE [CIS02].

The Figure 3.2 shows the traffic forwarding in PIM-SSM mode. A multicast source in site 1 of VPN A is sending traffic to receivers in sites 2 and 3 of the VPN A. The traffic must be delivered from PE1 to PE2 and PE3 and from there to the appropriate CEs, which are CE3 and CE4. In the same way, the source in Site 1 of VPN B is sending traffic to receivers in Sites 2 and 3 of VPN B.

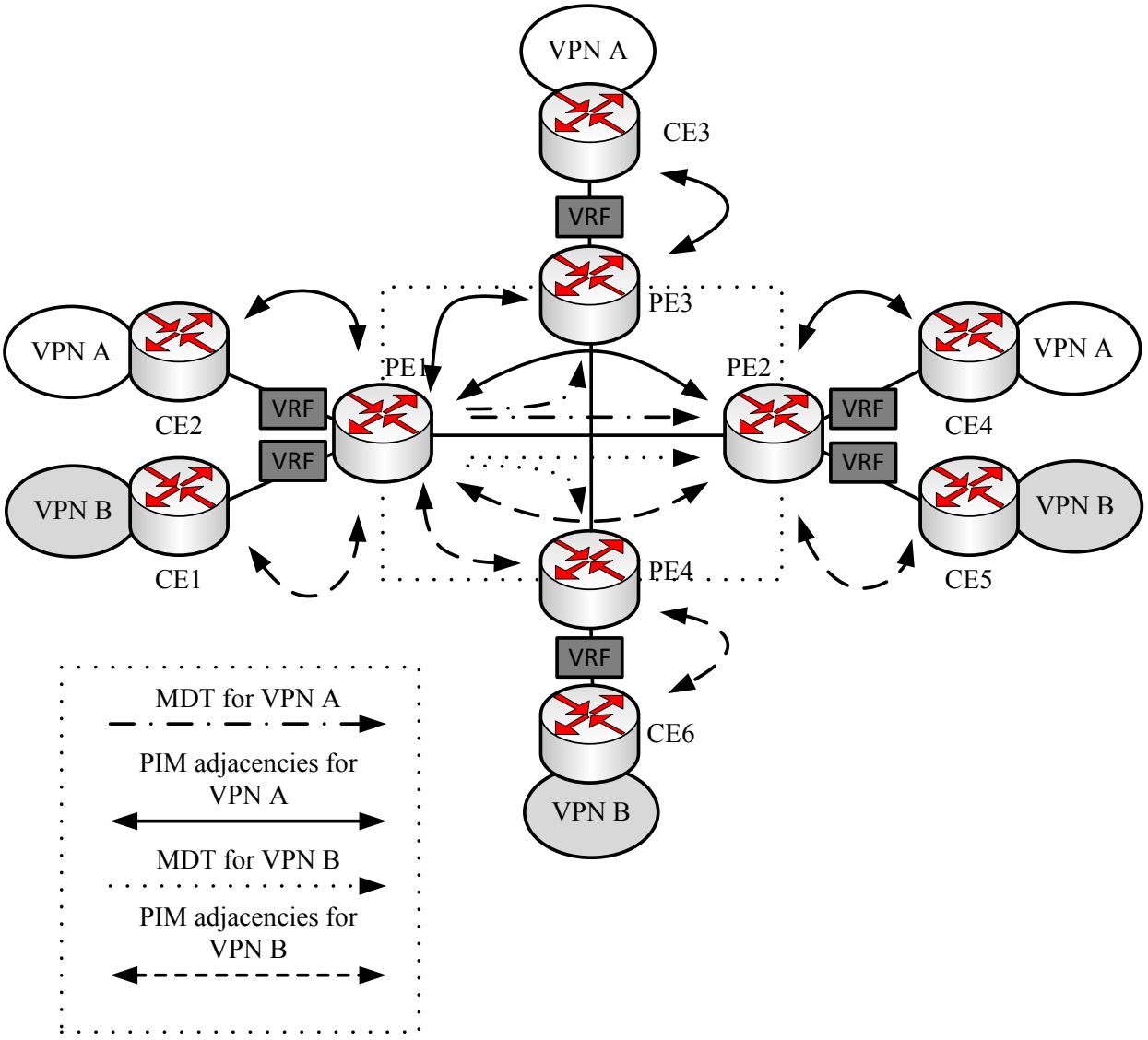


Figure 3.2: PIM-SSM/GRE MVPN Operation Mode

3.2.2 Network Recovery in Rosen Scheme

The Rosen scheme does not discuss network recovery procedure in case of link/node failure in [ROS10]. Instead, it relies solely on the interaction between Open Short Path First (OSPF) and PIM protocols to recover from a failure.

In general, most OSPF routers are able to notice that a link/node is down within a small number of seconds. This information then is spread via an updated router-Link State Advertisement

(LSA) to the rest of the OSPF routers. In many cases, the Router Dead Interval of OSPF can be used as a last resort to detect a link failure.

As soon as each router receives the new router-LSA, it recalculates its shortest path through Dijkstra's algorithm to destination. PIM can learn the topology change directly through a "notify" message from OSPF or indirectly by periodically polling the OSPF routing table. PIM needs to determine its RPF for each source in the source-group pair (S, G) or RP. If a new RPF has been discovered, PIM sends a Join message on the new RPF interface to form a new multicast tree [WAN00].

3.2.3 Advantages and disadvantages of Rosen Scheme

PIM is still considered the most popular multicast protocol for providing multicast services regardless of its scalability issues. In this section we show some of advantages and disadvantages of Rosen scheme.

3.2.3.1 Advantages of PIM

- Already deployed and proven for many SP and enterprise networks.
- PIM has a capability of building trees in the Provider core networks.
- It is capable of distribute C-multicast routes between the PEs.
- Works well without any changes.
- PIM supports multicast Label Distribution Protocol (mLDP) and offers smooth migration from GRE to MPLS encapsulation (not discussed in Draft Rosen).

3.2.3.2 Drawbacks of Rosen scheme

The most important issue with this scheme is that the PIM sessions are between the VRFs. Thus, for a given MPVN, a PE maintains a PIM session with every other PE that has membership in that MVPN. This complexity poses a significant scaling challenge. For example, with 1000

MVPNs per PE and 100 sites per MVPN, there would be 100,000 PIM neighbours per PE, which results in 3300 PIM hellos/second[MIN08].

- Large number of PIM adjacencies: Each PE has to maintain PIM adjacencies with all other PEs for which it has at least one MVPN in common per MVPN.
- Lack of support for aggregation: No ability to aggregate multiple MVPNs into a single inter-PE tunnel.
- Challenges in Inter-AS/Inter-provider deployment: 1) in control plane, because the requirement for direct peering between PEs, PEs in different ASs to have (direct) PIM routing peering.

  2) In data plane, because forces all providers to use the same tunnelling technology - GRE.
- Potentially large number of trees in the core:
- PIM-SM: for each VPN, there is a single multicast tree rooted at a rendezvous point (RP). If there are n VPNs in the network, then there are n multicast trees within the SP part of the network.
- PIM-SSM: for each VPN, there is a multicast tree rooted at each PE. Hence if there are n VPNs in the network present on each of m PEs, there are (m × n) multicast trees within the SP part of the network.
- Per-VPN state is maintained on the P routers because there is at least one MDT per VPN, the amount of state maintained by the P router is equal to at least the numbers of VPNs that have multicast support.
- Manual configuration of the MDT group address: The tunnel multicast destination address tells the receiving PEs which VPN the packet belongs to. For this purpose, the mapping between the multicast destination address and the VPN must be manually configured on each PE [MIN08].

3.3 NG MVPN (BGP/MPLS MVPN)

One of the major improvements of NG MVPN is to use BGP rather than PIM for carrying Customer information across (Join/Prune massages that PEs receive from the CE routers) the provider network by encoding it as a special customer multicast (C-multicast) route using BGP Multiprotocol Extensions [BAT07].

3.3.1 Distribution of C-multicast information in NG MVPN

Customer multicast (C-multicast) routing information exchange refers to the distribution of customer PIM (C-PIM) join/prune messages received from local CE routers to other PEs (towards the VPN multicast source) [JUN10a].This is accomplished by defining a new BGP NLRI (Network Layer Reachability Information), the MCAST-VPN NLRI. When a PE receives a PIM join message from an attached CE, indicating the presence of a receiver for a particular multicast group and source in the customer site, the PE converts it to a BGP update containing the multicast group and source. As a result, the information received by the local PE in the PIM join is advertised to the remote PEs as a C-multicast route carried in BGP. At the remote PE, the information carried in the C-multicast route is converted back into a PIM join message that is propagated using PIM to the remote CE [MIN08].

The Figure 3.3 shows the operation of distributing multicast routing information between PEs in NG MVPN environment. In this example the R1 and R2, which are multicast receivers located within RCV1 and RCV2 sites, interested in receiving multicast traffic from S, which in turn distribute Internet Group Management Protocol IGMP messages for a particular (S, G) to their Designated Routers (CE2 and CE3). The CE2 and CE3 then send PIM join messages for that (S, G) toward S to the PEs that is servicing them (PE2 and PE3).

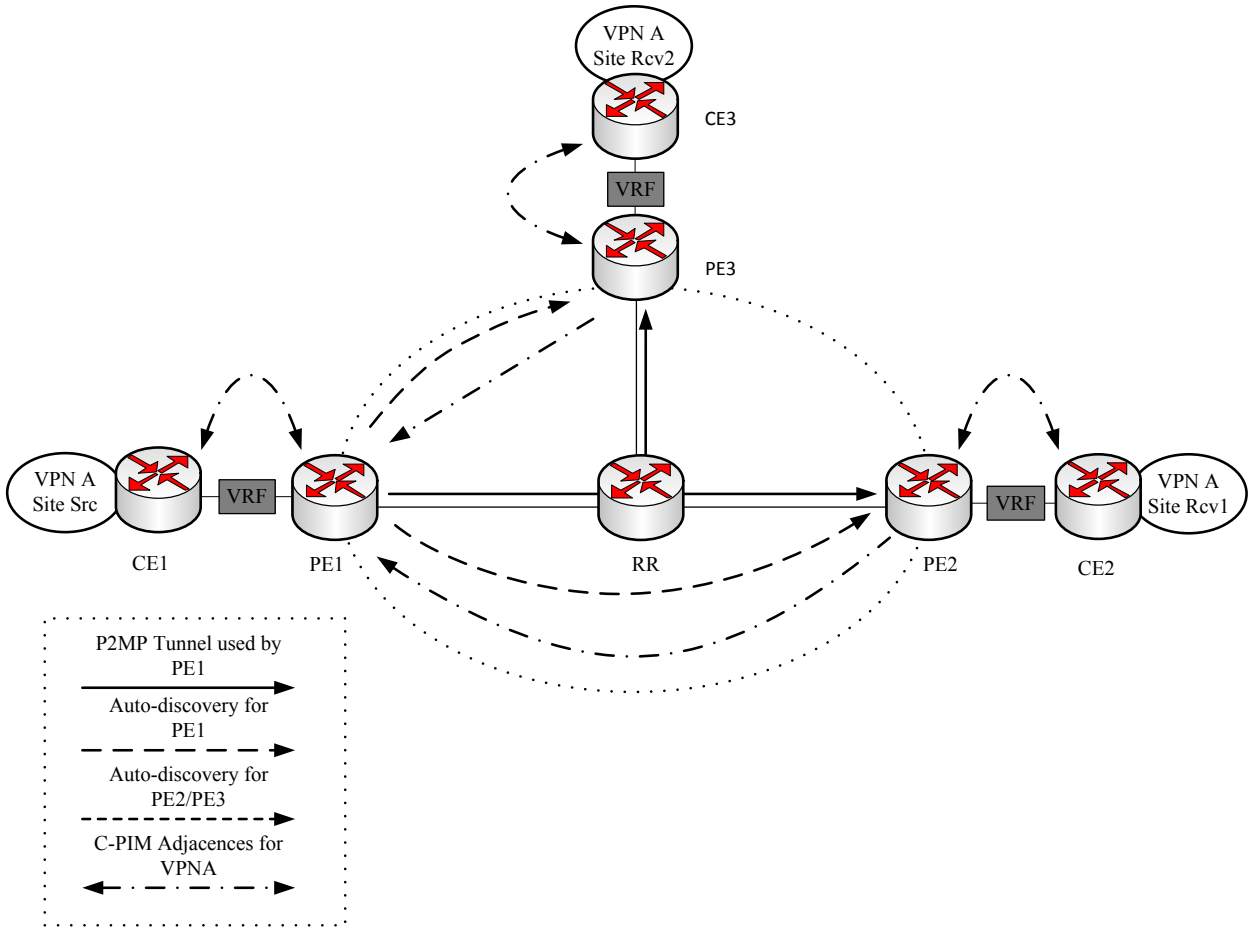


Figure 3.3: Distribution of C-multicast routing information in NG MVPN

When PE2 receives a PIM join from CE2, it uses the information carried in this PIM join to construct and originate a C-multicast route as follows:

- PE2 finds in the VRF associated with VPN A the unicast VPN-IPv4 route to S and extracts from this route the RD and the VRF Route Import extended community.
- PE2 builds a BGP C-multicast route that carries the source and group (S, G) information from the PIM join message received from the CE, the RD from the VPN-IPv4 route for S and an RT constructed from the VRF Route Import of the VPN-IPv4 route. Note that PE2 does not attach to the C-multicast route the 'regular' unicast RT associated with VPN A (the RT used by VPN-IPv4 routes).

- PE2 sends the C-multicast advertisement to its BGP peers, which in this case is the Rout Reflector (RR) as shown in the figure above.

The role of RR is to receive the C-multicast advertisement coming from PE2 and PE3 and aggregate and propagate them to PE1. Upon the receipt of the C-multicast route advertisement, several actions must be performed on PE1:

- PE1 accepts the C-multicast route into the VRF for VPN A because the C-multicast Import RT and matches the RT attached to the route. Any other PE receiving the route ignores it as it does not match its C-multicast Import RT, and the 'regular' unicast RT for VPN A is not attached to the route either.

- As a result of accepting the C-multicast route advertisement, PE1 creates (S, G) state in the VRF and propagates the (S, G) join towards CE1 using PIM instance running on PE1.

When CE1 delivers the massages to Source (Scr) site, S starts distributing multicast traffic towards the R1 and R2. The next section shows how multicast traffic is handled in NG MVPN.

3.3.2 Carrying of Multicast Traffic in NG MVPN

The NG MVPN scheme allows for a wide range of tunnelling technologies in the provider network. However, the most effective technology is the MPLS P2MP LSPs (RSVP-TE) as tunnels for transporting MVPN traffic between the PEs servicing the source and the PEs servicing the receivers [MIN08].

Many service providers with equipment supporting the NG MVPN service will choose to use RSVP-TE P2MP LSPs rather than multipoint GRE tunnels for MVPN environment because of the traffic engineering and fast resiliency advantages [ALC10].

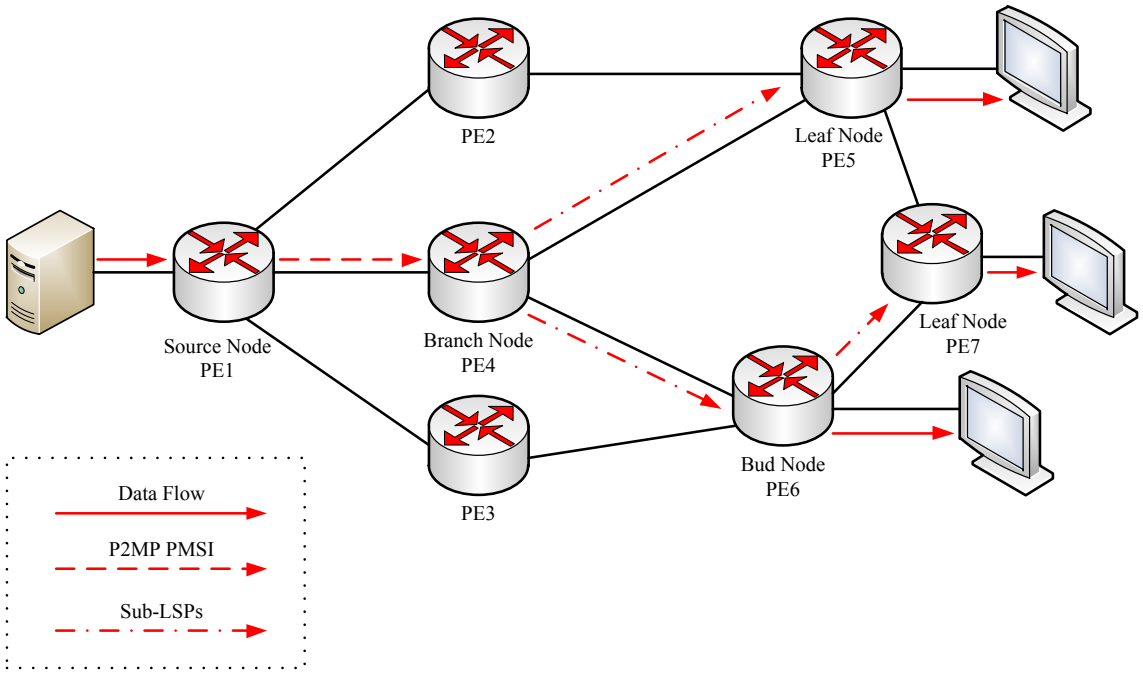


Figure 3.4: Composition of P2MP LSP

An RSVP-TE P2MP LSP is a unidirectional label switched path that takes in a packet at a root/source node and forwards a copy of the labelled packet to each of one or more leaf/destination nodes, with packet replication occurring at the source node or at the Label Switch Routers (LSRs) called branch nodes. An ingress node (PE) with an MVPN service initiates the set-up of an RSVP-TE P2MP LSP by signalling (sending PATH messages for) a set of source-to-leaf (S2L) sub-LSPs that collectively comprise the P2MP tree [ALA10]. The figure 3.4 demonstrates the role that each LSR plays when setting up a P2MP LSP.

- The source node (PE1) acts like a normal P2P LSP ingress (although it may also be a branch). There is always a single source node.
- Each leaf node (PE5 and PE7) behaves like a normal P2P LSP egress.
- Branch nodes (PE4) replicate data flows and forward them to multiple next-hops PEs.

Also, there are other nodes such as Bud nodes (PE6), a combination of a leaf and branch/transit nodes. They replicate data traffic and both pass it up to local endpoints and forward it to next-

hop PEs. Additionally, a transit node is any node, which forwards traffic (just like a normal P2P LSP transit node).

NG MVPN introduces two new types of Provider Multicast Service Interfaces (PMSI), which is an abstraction introduced in MVPN that refers to a "service" in the service provider's core network that can take a packet from one PE, belonging to one particular MVPN, and deliver a copy to some or all of the other PEs supporting that MVPN [ALC10]:

- **Inclusive I-PMSI:** A single multicast distribution tree in the backbone carrying all the multicast traffic from a specified set of one or more MVPNs. An inclusive tree carrying the traffic of more than one MVPN is an aggregate inclusive tree. All the PEs that attach to MVPN receiver sites using the tree belong to that inclusive tree.

- **Selective S-PMSI:** A single multicast distribution tree in the backbone carrying traffic for a specified set of one or more multicast groups. When multicast groups belonging to more than one MVPN are on the tree, it is called an aggregate selection tree [JUN08].

For an I-PMSI or S-PMSI of an MVPN service, the leaf nodes are discovered by the BGP auto-discovery process (I-PMSI A-D routes and Leaf A-D routes received by the source PE) and the S2L paths to these leaf nodes are calculated dynamically by the Constrained Shortest Path Firs (CSPF) algorithm, using the constraints (bandwidth, admin-groups) defined in an LSP template. An LSP template can be specific to one PMSI or used by many PMSIs. To protect P2MP LSP traffic from link failures, Fast Reroute (FRR) using P2P bypass LSPs may be used to recover from such failures in less than 50 ms [AGG07].

3.3.2.1 P2MP Link/Node Protection (MPLS Fast Reroute)

A key attraction of P2MP LSPs signalled using RSVP-TE is that MPLS fast reroute can be used for traffic protection, giving low failover times (< 50 ms) [HAS11]. In contrast, in normal IP multicast, the failover mechanisms are relatively slow (on the order of seconds), which is unacceptable for applications such as real-time video [MIN08].

3.3.3 Advantages of BGP

- Uniform control plane
    - Leverage of capabilities
        - Constrained distribution
        - Security
        - Summarization
        - Inter-AS, etc.

- MVPN membership is very similar to L3VPN membership
    - Matching on RT extended communities.
    - MVPN Membership is discovered automatically
    - Using Auto-Discovery (A-D) BGP routes
    - Different types of routes at different stages of discovery.
    - MVPN Data Plane Auto-Discovery
    - Same A-D routes discover and bind services to PMSI tunnels.
    - Does not setup the tunnel; rather than initiates the setup.

3.3.4 Advantages of NG MVPM Scheme

- Supports multiple transport options
    - Different MVPNs can use different tunnelling technologies (P2MP MPLS or PIM-SM GRE).

- Scalable and Extensible Signalling
    - BGP, same model as unicast Layer 3 VPN and supports auto-discovery of routes.
    - Each PE uses existing IBGP sessions, which may only require sessions with the route reflectors.

- Reducing data plane overhead

- It is possible to aggregate multiple (S,G) of a given MVPN into a single selective tunnel.
- Proven and flexible Inter-AS Operational model
- Seamlessly works with all three options (A, B and C as defined in RFC 4364) available for inter-AS unicast.
- Segmented inter-AS trees that allow each AS to independently run a different tunnelling technology.

- Lifting P-Tunnel Mesh Requirement
  - Allow providers to support MVPN customers who want to restrict.
- Multicast sources to a subset of its sites
  - Provides the SP with the flexibility to build pricing models for a MVPN service based on number of sources/receiver sites in a MVPN.
- Seamless support for Extranet and Hub and Spoke Topologies similar to Unicast VPN

3.4 Comparison of Rosen Scheme and NG MVPN

An important but often overlooked aspect of the comparison is the impact on the SP operations. The Rosen approach requires the deployment of PIM protocol in the SP network in order to support MVPN traffic. In contrast, the NG MVPN approach reuses the same protocols used by the unicast VPN solution, with extensions [MIN08]. In this section, Table 3.1 shows the comparison of Rosen and NG MVPN schemes in terms of transport option, signalling protocol, VPN traffic aggregation, inter-AS operation and provider tunnel mesh requirement.

	ROSEN Scheme	NG MVPN
Transport	Uses PIM-SM GRE tunnelling technology	Different MVPNs can use different tunnelling technologies such as P2MP MPLS or PIM-SM GRE
Reservation Mechanism	No resource reservations mechanism available	Resource reservation using MPLS RSVP-TE
Signalling	PIM	BGP, same model as unicast L3VPN (supports auto-discovery of routes)
PE-PE Signalling Sessions	Each PE needs a separate PIM adjacency with each remote PE per-VRF	Each PE uses existing iBGP sessions, which may only require sessions with the route reflectors
VPN Traffic Aggregation	No ability to aggregate multiple MVPNs into a single inter-PE tunnel	It is possible to aggregate multiple (S,G) of a given MVPN into a single selective tunnel and aggregate multiple P2MP LSPs using P2MP LSP hierarchy
Inter-AS Operations	Inter-AS operations B and C (RFC 4364) requires PEs in different Ass to have direct PIM routing peering	NG MVPN seamlessly works with all three options A, B and C (RFC 4364). It also has the concept of segmented inter-AS trees that allows each AS to independently run a different tunnelling technology
Provider Tunnel (P-tunnel) mesh requirement	Required between PEs, which forces providers to sell an MVPN service where every customer site can be a source and a receiver	P-tunnel mesh requirement is removed. Providing the flexibility to build pricing models for an MVPN service based on sites connected to wither sources or receivers or both
Recovery Mechanism	Relies solely on IGP to recover, relatively slow (on the order of seconds)	Uses MPLS fast reroute (FRR), low failover time (less than 50 ms)

Table 3.1: Comparison of Rosen scheme and NG MVPN

3.5 Summary

Multicast distribution techniques reduce the cost of delivering multi-site, multipoint video and content applications across WAN. It allows carriers to overcome the challenges of delivering high-bandwidth, mission-critical multicast applications in WAN environments. The MVPN service implementation delivers required multicast VPN service attributes such as Privacy, Multi-dimensional Scale, Geographic Reach, Inter-Company Delivery, High Availability, Traffic Engineering, End-to-End Bandwidth Efficiency, Application Assurance and Operational Consistency for next-generation multicast video and content application delivery [ALC10].

Although the original PIM-based solution described in [ROS10] offer to carry IP multicast customer traffic over a shared provider infrastructure, this approach departed from the L3VPN unicast model and suffered from scaling limitations both in terms of the maximum number of MVPNs it could support and in terms of the efficiency with which it could carry traffic through the provider network. In contrast, the BGP/MPLS-based approach reuses the unicast L3VPN unicast mechanisms with extensions as necessary, thus retaining as much as possible the flexibility and scalability of unicast [MIN08]. However, this scheme requires an extension to the existing Multiprotocol BGP along with advanced operational and troubleshooting knowledge to be deployed in service provider networks. In the next chapter, we carry out a number of tests using a multivendor testbed to better understand the operational challenges as well as the scalability issues of each scheme.

Chapter 4

Performance Evaluation and Analysis

This chapter demonstrates various number of methods and tools in which we have used to perform very advanced level of practical experiments. Our tests reflect scalability measures of different network techniques to carry out multicast traffic over IP/MPLS networks. All our tests have been carried out on the ONRL at the University of Ottawa. The testbed that we have used in our experiments consist of a mixture of real and emulated routers, along with high advanced network switches.

In this work, we have performed several experimental tests in order to evaluate the scalability of two well-known multicast schemes in Layer 3 VPN model; Rosen Scheme (PIM-SM/GRE) and NG MVPN (BGP/RSVP-TE P2MP) in IP /MPLS core networks. In our experiments, we have successively generated as many as five experimental tests to evaluated the scalability of the aforementioned schemes in terms of the overhead (the number of control messages that should be generated through the core) and number of sessions/adjacencies required between Provider Edge (PE) routers

The rest of this chapter is organized as follows: In Section 4.1, we describe the specification of the hardware that has been used to run our test scenarios. In Section 4.2, we define testbed setup. In Section 4.3, we define the methodology behind our experiments. In Section 4.4, we demonstrate the scalability of two multicast schemes through measuring the amount of the overhead generated by each protocol and the number of sessions/adjacencies for each protocol. In Section 4.3, we evaluate the scalability and analyze the test results for the Rosen and NG MVPN schemes.

4.1 Hardware Description

The hardware that we have used in our experiments involves multiple network vendors such as Avaya, Juniper as is shown in Figure 4.1. In addition to number of PCs equipped with a Gigabit Ethernet (GE) Network Interface Cards (NIC). This section illustrates detailed specification of each device in our network.

- **Juniper Routers**
  - All Juniper routers (M10/M160 series) are running JUNOS version 10.2 Internet Software, which supports variety of routing and switching protocols such as OSPF, ISIS, BGPv4, PIM, IGMP, MPLS-TE, RSVP-TE, LDP and Layer2/3 VPN implementation.
  - All Juniper routers are connected using very high speed links such as Packet over SONET (POS) OC-48, GE and Asynchronous Transfer Mode (ATM) OC-3 modules with capacity of 2.45 Gbps, 1 Gbps and 155 Mbps respectively.
  - Each router is uniquely assigned a loopback interface address along with a unique private IPv4 address for each interface.

- **Avaya Ethernet Router Switch 8600 series**
  - Both Avaya 8600s are running Software release version 3.51, which supports OSPF, BGPv4, PIM, IGMP and Layer 3 VPNs.
  - They are equipped with four GE ports.
  - They are designed to provide high-density connectivity for mid-size and large Each router is uniquely assigned a loopback interface address along with a unique private IPv4 address for each interface.

- **Avaya BayStack Ethernet Switches**
  - The Avaya Ethernet Switches are running BayStack Operating system Switching Software (BoSS) version 3, which offers highest level of security with featuring including Secure Shell SSH and Simple Network Management Protocol (SNMPv3).
  - 24 x Network – Ethernet 100Base-TX - RJ-45 to allow the end computers to be accessible through the network.
  - They include two built-in GE ports to connect to the MPLS domain.
  - They are designed to provide high-density desktop connectivity for mid-size and large enterprise customers wiring closets.
  - The two BayStack switches are connected to the MPLS domain using 100Base-T Ethernet links.

- **PC Host Computers**
  - Intel Core 2 Duo CPU E8200, 2.66 GHz processors, 2 GB RAM.
  - Windows XP SP2-based and are equipped with Gigabit Ethernet (GE) cards.
  - Nortel Multicast tool is installed to generate multicast traffic.
  - All PC hosts are synchronized using Windows NTP server to maintain a fixed clock with all devices on the network.

- **Linux-based Router Emulator (LRE)**
  - Intel Core 2 Duo CPU E8200, 2.66 GHz processors, 2 GB RAM.
  - Linux distribution; Ubuntu 10.04 is installed and each PC is equipped with at least three Gigabit Ethernet (GE) cards.

- Running a modified version of JUNOS (OLIVE) in order to turn a PC into one or two emulated routers.

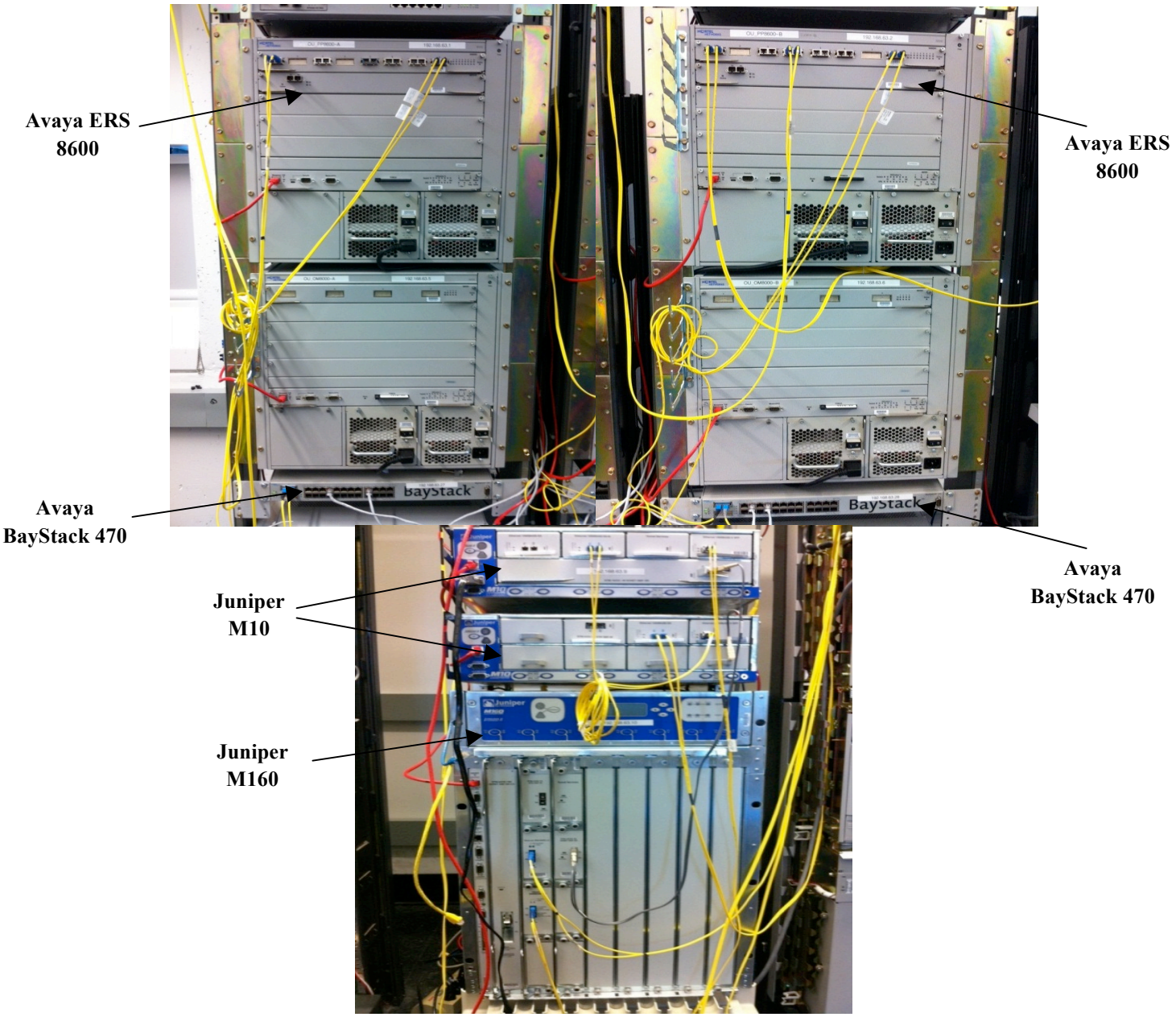


Figure 4.1: Hardware Setup

4.2 Testbed Setup

This section shows the testbed setup that we have built to perform our experiments. The testbed that is shown in Figure 4.2, consists of Juniper and Avaya equipment along with a number of Linux-based Router Emulators (LREs), they are configured as follows:

- Two M10 routers along with two LRE devices are configured to present the Provider Edge (PE) routers in the network. Furthermore, M160 router is configured as Provider (P) to handle and redirect multicast traffic as well as reduces the number of sessions between PEs in the Core.

- Two ERS 8600s along with two LRE devices and two 470 switches are configured to present the Customer Edge (CE) routers in our network. The 470 switches are designed to connect CE routers to customer VPN sites.

- Multiple PCs equipped with a multicast generator tool called as Multicast Hammer (MC Hammer) tool from Nortel and are placed in each of the customer's remote sites.

We have deployed the MC Hammer tool in order to generate multicast traffic from a single server to multiple clients in order to measure the number of control messages as well as the number of adjacencies/sessions required in the core network. It is important to mention that in order to measure the number of control messages for both schemes, we have completely relied on the router logs since they are more precise than any other tools.

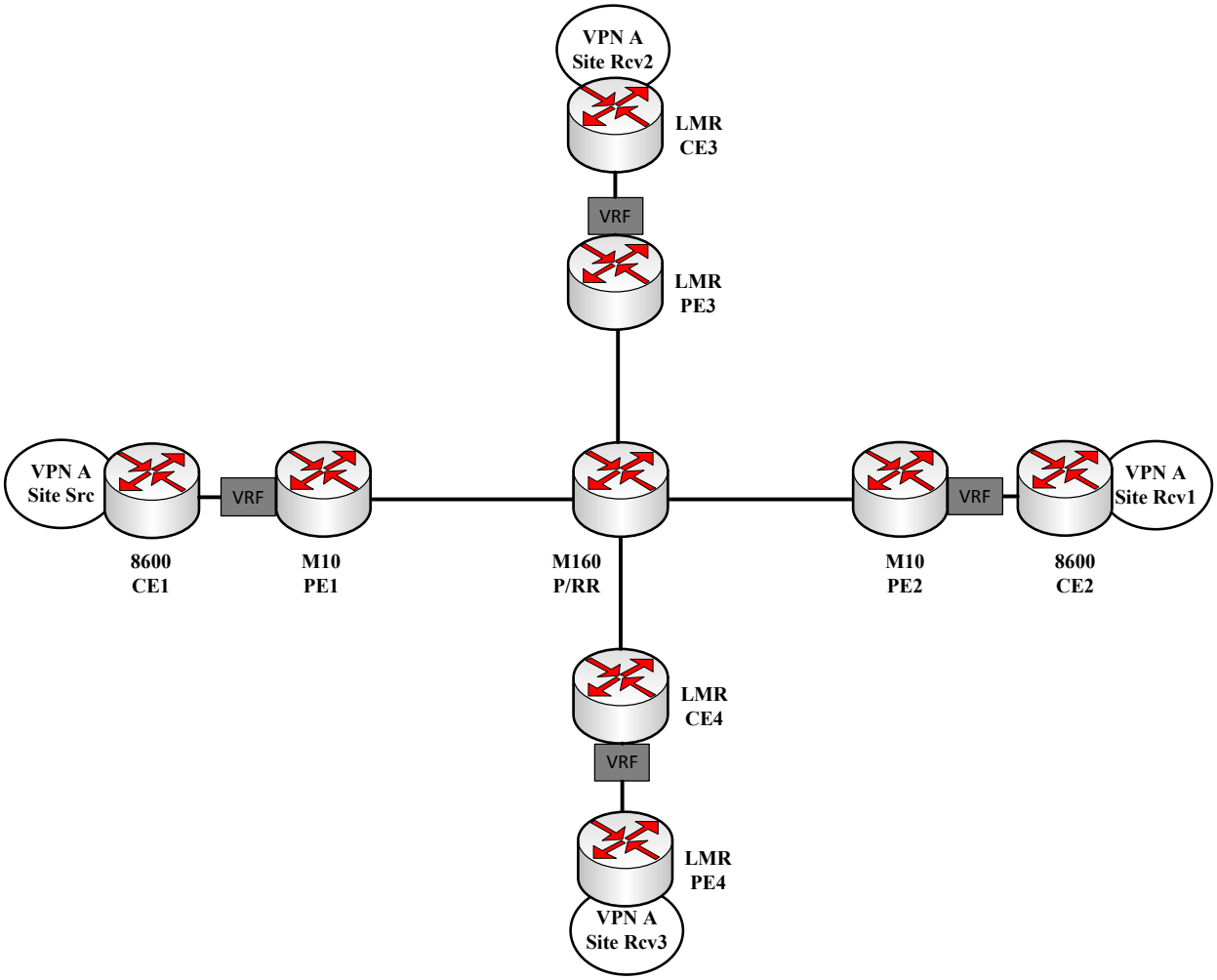


Figure 4.2: General Testbed Setup for MVPN

4.3 Experiment Methodology

In this section, we present the main practice behind our experiments. As we have mentioned in the previous section, we have deployed a MC Hammer tool to generate Multicast (UDP) traffic that is originated from a single host (sender) to multiple clients (receivers) in order to measure the scalability and resiliency of Rosen scheme and NG MVPN in the core network. Our evaluation is done through measuring the number of control messages as well as the number of adjacencies/sessions in the control plane.

In the first two scenarios of our experimentation, we measure the scalability of Rosen scheme by observing the exchange of the control plane information, as well as multicast forwarding states on the Layer 3 VPN infrastructure. Similarly, we perform additional three test scenarios to evaluate the scalability of NG MVPN scheme, which introduces a BGP-based control plane that distributes all necessary routing information to enable multicast service over the existing Layer 3 VPN infrastructure.

4.4 Scalability Scenarios

Scalability is one of the core requirements, if not the most required one, for multicast VPN environment. Thus, in order to discuss scalability problems for a certain scheme, it is very significant to consider number of parameters that affect the scalability of that scheme. As a result, we have defined some parameters such as the amount of overhead in the control plane as well as the number of established sessions/adjacencies in order to measure the control plane (BGP and PIM) scalability of each scheme.

In the subsequent sections, we demonstrate in details the scalability of the control plane in Draft Rosen and NG MVPN schemes in terms of the number of control plane messages as well as the number of established sessions/adjacencies.

4.4.1 Rosen scheme Scenarios

In the Rosen scheme scenarios, we show the number of control messages generated at the provider side along with the number of PIM adjacencies required between PEs when deploying PIM-SM as the control Plane of MVPN.

In general, we define the control messages as *PIM Join/Prune messages* only at the Provider side since *PIM Join/Prune messages* at the customer side are treated in the same fashion by both schemes. Additionally, the *Register* and *Register-stop messages* that generated by the PIM routers also have been considered in our experiments. What's more, the *PIM Hello messages* are

not counted since these are not messages that trigger specific action in a typical scenario. They are generated only to maintain the adjacency status between PIM routers.

4.4.1.1 Scenario 1 (No. of PEs vs. No. of multicast messages)

In a unicast Layer 3 VPN environment, all VPN states are contained within the PE routers. With MVPNs however, two types of PIM adjacencies are established: one between the CE and PE routers through a Multicast VPN routing and forwarding (MVRF) routing instance known as Customer-PIM (C-PIM) instance, the second between the main PE routers and their service Provider (P) core neighbours known as Provider-PIM (P-PIM) instance.

For MVPN to work correctly there must be two types of RPs: The VPN Customer PR (VPN C-RP) is an RP that resides within a VPN that connects the segments of a customer network. The Provider-RP (P-RP) resides within the service provider network itself [JUN04].

In this scenario, we have setup a MVPN testbed environment by deploying the Rosen scheme in PIM-SM in order to test the scalability of the control plane of the approach in respect to the number PE routers. As shown in Figure 4.3, our network topology contains nine routers: four routers as PEs, one router as RP and four routers as CEs and they are configured as follows:

- At the source side, a multicast source (Src1) with configured an IP address 10.10.11.2/24 is connected to an Ethernet switch (Baystack 470), in which is connected over a GE link to a CE (CE1) router. The CE1 router is configured with a loopback address 10.10.10.55/32.
- At the receiving side, one multicast receiver (Rcv1) configured with an IP address 10.10.12.2/24 is connected to an additional Ethernet switch Baystack 470, which is in turn connected through a GE interface to the CE2 router. The CE2 is assigned a loopback interface 10.10.10.60/32.
- Another two multicast receivers (Rcv2 and Rcv3) configured with IP addresses 10.10.13.2/24 and 10.10.13.3/24 are directly connected over GE links to the CE3 and

CE4 routers. The CE3 and CE4 routers are configured with loopback addresses 10.10.10.65/32 and 10.10.10.70/32 respectively.

- The Provider-RP configured with a loopback address 10.10.10.2/32 is connected to all PE routers using various high-speed links such as ATM OC-3, GE and SONET OC-48 to form Hub-and-Spoke (Star) topology.
- The PE1 router is configured with a primary loopback address 10.10.10.1/32 for iBGP session and configured as VPN Customer-RP with a second loop back address 10.10.47.101/32 in order to associate the local VRF (VPN A) with the CE1 router.
- In a similar way, the PE2, PE3 and PE4routers are also configured with primary loopback addresses 10.10.10.3/32, 10.10.10.4/32 and 10.10.10.5/32 for iBGP sessions while the second loopback addresses 10.10.47.102/32, 10.10.47.103/32 and 10.10.47.104/32 are configured to associate the local VRF (VPN A) with the CE2, CE3 and CE4 routers.

Auto-Discovery Phase

When a C-PIM instance is configured on a CE router, a PIM Hello message is generated and forwarded to a VRF table assigned (VPN A) for that customer on the local PE router. A GRE header then is added to that *PIM Hello message* with fields containing the multicast group address of 224.1.1.1/32 and the loopback address of the PE router.

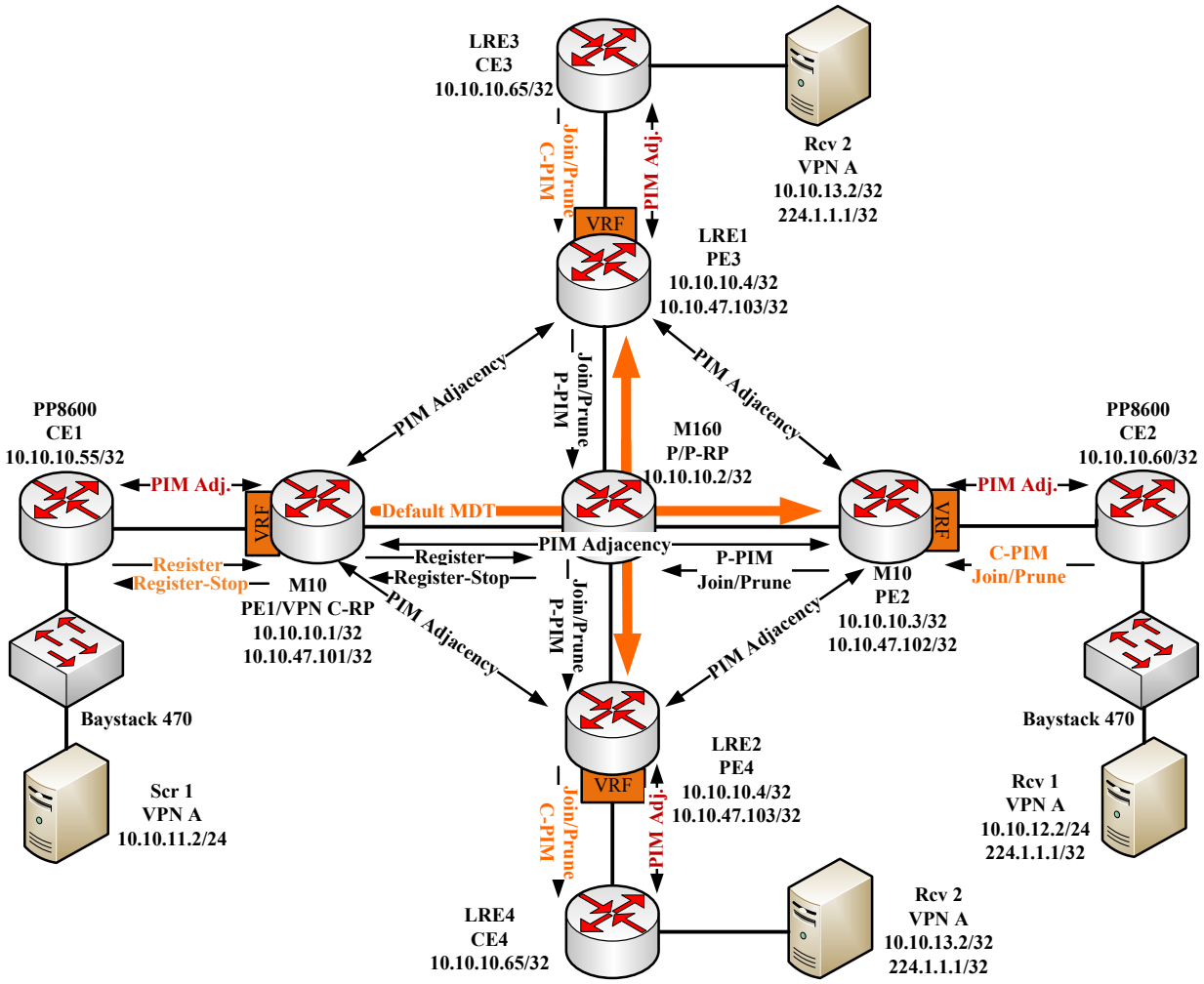


Figure 4.3: Testbed Setup for Scenario 1

A PIM register header is added to that PIM Hello message and forwards it through the core to the P-RP. The header of the *Register message* contains the destination address of the P-RP 10.10.10.2/32 and the PE1's loopback address 10.10.10.1/32 as the source address.

Upon receiving the message, the P-RP de-encapsulates the header of the *Register message* and sends the remaining GRE encapsulated *PIM Hello message* to all of the PE routers.

The P-PIM instance on each PE router handles the GRE encapsulated packet. Since the multicast group address is contained in the header of the *PIM Hello message*, each PE router de-encapsulates the GRE header attached to the packet and sends the remaining to reach the desired

multicast group address of 224.1.1.1 within the MVRF (VPN A), in which is forwarded to the local CE router.

PIM Join/Prune Phase

Once the receivers (Rcv1, Rcv2 and Rcv3) are interested in receiving multicast traffic, each sends a request message (*IGMP membership Report message*) to their local CE router. The purpose of this request is to inform the local CE router that the receiver wants to join a specific multicast group address 224.1.1.1/32, which is included in that message. This is achieved in our experiment by configuring the MC Hammer tool at Rcv1, Rcv2 and Rcv3 to generate *IGMPv2 membership Report message*.

In response, three *PIM Join messages* then generated and forwarded from CE2, CE3 and CE4 routers to the local PE2, PE3 and PE4 routers respectively. Next, the received *PIM Join messages* are further sent throughout the Multicast Tunnel (MT) after a GRE header added to each *PIM Join message*. The GRE header contains a VPN group ID and the PE's loopback address. The GRE encapsulated *Join message* is sent to the PE1 routers by means of the P-RP. When the PE1 router receives the packet, it strips off the GRE header and finds that the packet is destined to the VPN C-RP. Since the PE1 router is also the VPN C-RP, it installs (*, G) (*, 224.1.1.1) state for the multicast group address that it has received the Join message for. In this case, the multicast group address is 224.1.1.1/32.

Should one of the receivers desire to leave the multicast group 224.1.1.1/32, it sends an *IGMP Leave message* to the local CE router, which in turn sends a *PIM Prune message* destined to the VPN C-RP. Once the VPN C-RP receives the *PIM Prune message*, it removes the state that was previously installed for that multicast group.

This procedure is similar to the join procedure that was mentioned earlier in this section. In fact, a PE router running a PIM instance is capable of aggregating a number of *Join* and *Prune messages* into a single packet and sends it to the P-RP.

Data Forwarding Phase

Once the multicast source, which is directly connected to the CE1 router, has data to send, it starts forwarding traffic to multicast group address 224.1.1.1/32. This is achieved by deploying the MGT at the source node and it begins generating native multicast data. The designated router (DR), which is also the CE1 router, encapsulates these data into a *PIM Register message* and forwards it to the VPN C-RP through unicast routing over the Layer 3 VPN.

The VPN C-RP de-encapsulates the received data packet and installs (S, G) (10.10.11.2, 224.1.1.1) state for the multicast group 224.1.1.1/32 and the source address 10.10.11.2/24.

If there are no PIM Join messages received for that multicast group at the RP, *a Register-stop message* is sent through unicast routing over the Layer 3 VPN to stop forwarding *Register messages* to the RP.

PE routers must maintain additional state when the C-multicast routing protocol is PIM-SM. This requirement is because of the existing of P-RP, the receivers first join the shared tree rooted at C-RP (called C-RP Tree or C-RPT). However, as the VPN multicast sources become active, receivers learn the identity of the sources and join the tree rooted at the source (called customer shortest-path tree or C-SPT). The receivers then send a prune message to C-RP to stop the traffic coming through the shared tree for the group that they joined to via C-SPT. The switch from C-RPT to C-SPT is a complicated process requiring additional state.

4.4.1.2 Scenario 2 (No. of MVPN Sites vs. No. of Multicast messages)

In this Scenario, we have once more setup a MVPN testbed environment by deploying Rosen scheme in PIM-SM. However, this time we altered our network topology in order to test the scalability of the control plane of the approach in respect to the number of MVPN sites. As shown in the Figure 4.4, our network topology in this scenario consists of seven routers: two PE routers, a RP and four CE virtual routers. Each CE Virtual router is dedicated two NIC cards and a loopback interface (two OLIVE instances per a LRE):

- At the source side, two multicast sources (Src A and B) with configured IP addresses 10.10.11.2/24 and 10.10.13.2/24 are connected to CE1 and CE3 virtual routers, which in turn are assigned loopback addresses 10.10.10.55/32 and 10.10.10.60/32 respectively.
    - Additionally, another two multicast source (Src C and D) with configured IP address 10.10.15.2/24 and 10.10.17.2/24 are directly connected over GE links to the CE5 and CE7 virtual routers. These virtual routers CE5 and CE7 are assigned loopback addresses 10.10.10.65/32 and 10.10.10.70/32 respectively.
- At the receiver side, two receivers (Rcv A and B) with configured IP addresses 10.10.12.2/24 and 10.10.14.2/24 are connected to CE2 and CE4 routers, which are assigned loopback addresses 10.10.10.4/32 and 10.10.10.5/32 respectively.
    - Another set of receivers (Rcv C and D) configured with IP addresses 10.10.16.2/24 and 10.10.18.2/24 are connected to CE7 and CE8 routers, which are also, configured with loopback address 10.10.10.6/32 and 10.10.10.8/32.
- The Provider-RP configured with a loopback address 10.10.10.2/32 is connected to PE1 and PE2 routers using two different high-speed interfaces; ATM OC-3 and SONET OC-48.
- The PE1 router is configured with a primary loopback address 10.10.10.1/32 for iBGP session while it is configured as VPN Customer-RP with a second loop back address 10.10.47.101/32 to connect the local MVRFs (VPN A, B, C, and D) to the CE routers.
- In similar fashion, the PE2 router is also configured with a primary loopback address 10.10.10.3/32 for an iBGP session while a second loopback address 10.10.47.102/32 is configured to connect the local MVRFs (VPN A, B, C and D) to the CE routers.

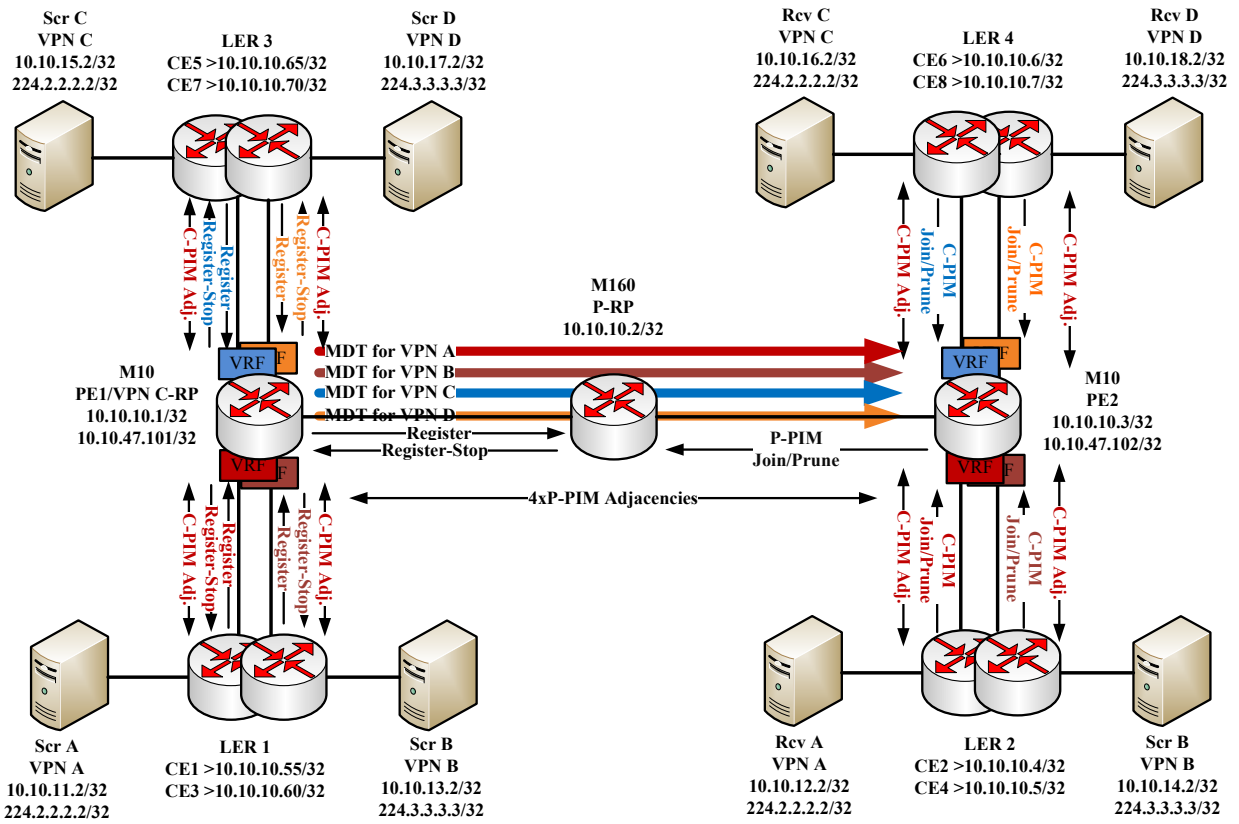


Figure 4.4:Testbed Setup for Scenario 2

In this scenario, the Auto-discovery Phase, the PIM Join/Prune Phase and the Data Forwarding Phase follow the same pattern as the previous one except that four C-PIM instances are now configured on both PE1 and PE2 routers for VPN A,B,C and D.

This means that PE1 will be receiving four separate *PIM Register messages* from CE1, CE3, CE5 and CE7 routers. Additionally, the PE1 router will also be generating four separate *PIM Register-stop messages* (one for each group).

Similarly, The PE2 router will be receiving four separate sets of *Join/Prune messages* from CE2, CE4, CE6 and CE8 routers.

4.4.2 NG MVPN Scenarios

In the NG MVPN scenarios, we show the number of control messages generated at the provider side in addition to the number of established sessions between PEs when BGP is deployed as the control Plane of MVPN. The main tasks of the control plane include MVPN auto-discovery, distribution of P-tunnel information, and PE-PE C-multicast route exchange.

A PE router that joins a BGP-based NG MVPN network is required to send a BGP update message that contains a MCAST-VPN NLRI. The value of the variable field depends on the route type; seven types of NG MVPN BGP routes are defined in the 2547 bis-mcast-bgp draft. The first five route types are called Auto-discovery (AD) MVPN routes, which is referred to as non-C-multicast MVPN routes. Type 6 and Type 7 routes are called C-multicast MVPN routes [JUN10].

In the following experiments, we only consider some types of NG MVPN BGP routes generated by the PE routers such as *Source active AD route* (SA AD route), *Shared tree join route* (C-multicast mvpn route) and *Source tree join route* (C-multicast mvpn route). As in the previous scenarios, the *C-PIM Join/Prune messages* generated on PE-CE links are not counted since they are treated in the same fashion by both schemes.

Additionally, other BGP messages such as *Open*, *Keepalive* and *Notification* messages are not counted since these are not messages that trigger specific action in our scenarios. They are only generated for establishing and maintaining session status between BGP Peers. We suppose that a message processed by a BGP Route Reflector (RR) to go from a receiver connected PE to the source connected PE.

4.4.2.1 Scenario 3 (No. of PEs with RR vs. No. of control messages)

In this scenario, we have setup a MVPN testbed environment by deploying the NG MVPN approach in order to test the scalability of the control plane of this approach in respect to the number PE routers. The network topology shown in Figure 4.5 is similar to the one in the first

scenario, except that P router is now assuming the role of RR in order to reduce the number of BGP peering sessions.

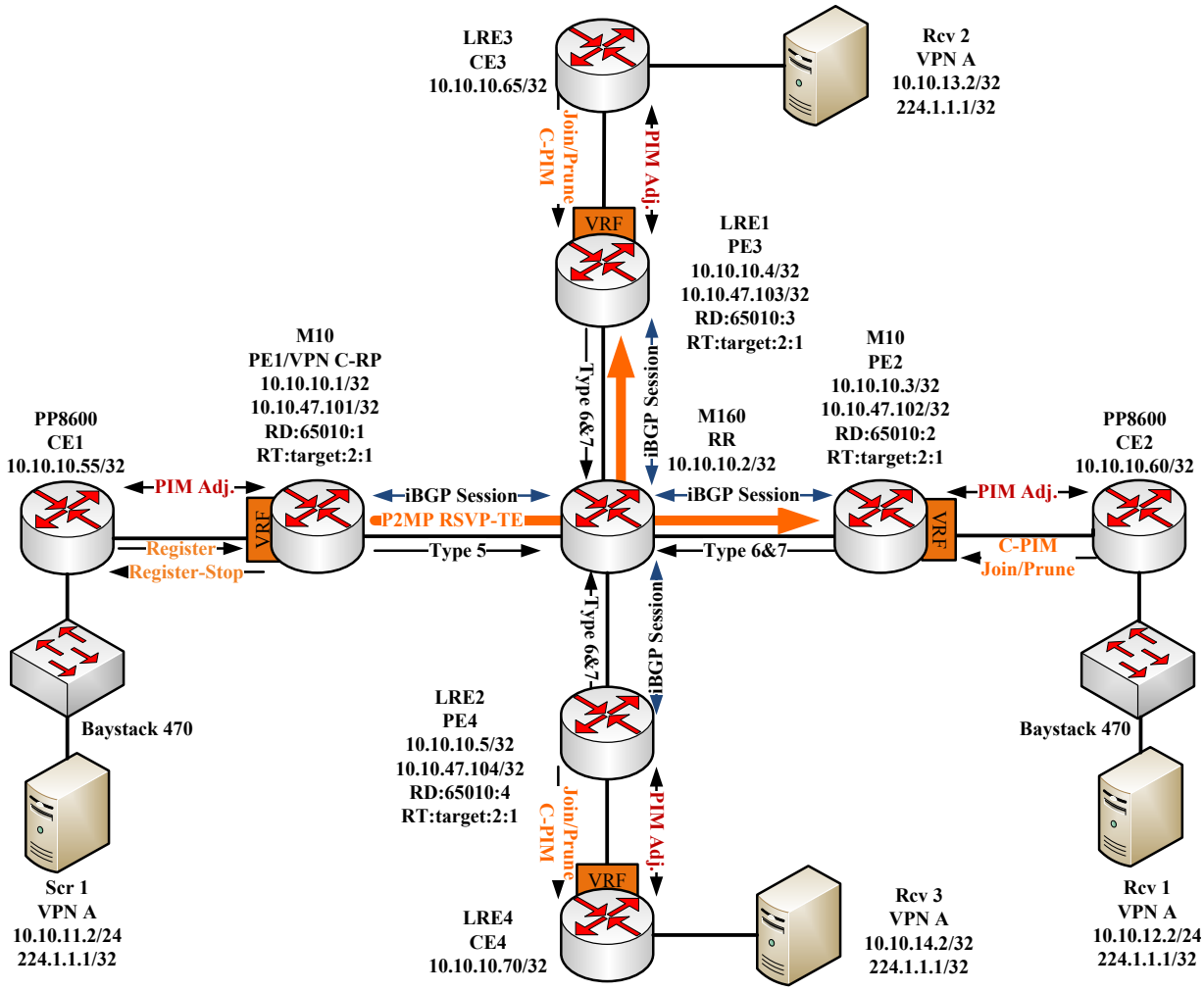


Figure 4.5: Testbed Setup for Scenario 3

Originating a Type 1 Auto-Discovery Route

Once BGP is configured, all PE routers start establishing BGP peering sessions with the RR. The use of RR is very essential in core networks because it relaxes the full mesh requirement when deploying iBGP within a local AS.

Next, all PE routers begins originating a local Auto-Discovery (AD) route known as *Intra-AS I-PMSI AD route* (Type 1) to each other. This route is used to find other PE routers that have sites of the same MVPN (VPN A) connected to them for instance.

The *Intra-AS I-PMSI AD* route format consists of three portions; *route type*: *RD configured of the local site*: *Loopback address of the originator*. In our experiment, the PE1 router originates the following *intra-AS AD route* 1:65010:1:10.10.10.1. In a similar manner, the PE2 and PE3 routers originate the subsequent *intra-AS AD routes* 1:65010:2:10.10.10.3 and 1:65010:3:10.10.10.4 respectively. Each PE router then advertises the AD route to the RR after attaching its RT to it. This guarantees that BGP active routes will be correctly imported to or exported from the appropriate MVRF (VPNA) configured on a PE router.

Additionally, the PE1 router also attaches a PMSI attribute **PMSI: Flags 0:RSVP-TE:label[0:0:0]:Session_13[10.10.10.1:0:32423:10.10.10.1]** to the AD route based, and then it establishes a provider tunnel (P2MP Inclusive tunnel) using RSVP-TE with all other PEs in the network. The PE2 and PE3 routers join the P2MP RSVP-TE through the Tunnel Identifier in the PMSI attribute.

Originating a Type 6 C-multicast Route

As soon as Rcv1, Rcv2 and Rc3 receivers become interested in multicast traffic, each sends a request message (*IGMP membership Report message*) to CE routers in order to join the multicast group address 224.1.1.1/32. This is achieved in our experiment by configuring the MGT at the receiving site (Rcv1/Rcv2) to generate *IGMPv2 membership Report message*.

As a result, the CE2, CE3 and CE4 routers will generate and advertise a *PIM Join message* (*, 224.1.1.1) destined to PE2, PE3 and PE4 routers respectively. The local PE routers in turn do a route lookup in their VRF tables (VPN A) for VPN C-RP configured for that group, which is the PE1 router in our experiment. Consequently, the PE routers construct a BGP C-multicast route known as *Shared tree join route* (Type 6) and advertise it to RR 6:65010:1:65010:32:10.10.47.101:32:224.1.1.1 after attaching RT and RD to it. Upon receiving

the BGP C-multicast routes, the RR forwards the received routes to the PE1 router.

The VPN C-RP (PE1) in turn installs those routes in its local VRF table and creates a (*, 224.1.1.1) state on the router. The (*, G) state indicates that the multicast source is unknown while in the same time ensures that the RP Tree is established between CE routers and VPN C-RP.

Originating a Type 5 Source Active Auto-Discovery Route

As the multicast source Scr1 becomes active, the PE1 (VPN C-RP) router starts receiving a PIM Register messages from CE1 router. Consequently, the PE1 router originates an SA A-D route and advertises it to PE2, PE3 and PE4 routers: 5:65010:1:32:10.10.11.2:32:224.1.1.1.
As soon as the source within the site connected to PE1 starts sending multicast data, the PIM Designated Router connected to the source originates C-PIM Register message and sends it to the C-RP, which is PE1. As a result, PE1 router originates a type 5 SA route upon receiving the first data packet or the C-PIM register messages.

Originating a Type6 and 7 C-multicast Route

Each PE router (except PE1) will originate and advertise two different types of C-multicast routes to the PE1 router and to each other via the RR:

- One Type 6 route 6:65010:1:65010:32:10.10.47.101:32:224.1.1.1 as a response to *PIM Join messages* from CE routers.
- Another Type 7 route 7:65010:1:65010:32:10.10.11.2:32:224.1.1.1 in response to the Type 5 routes received from PE1 router.

Only PE1 router accepts these routes because of the unique RT the route that carries while the other PE routers will discard the route they received from each other due to non-matching RT values.

The PE1 then compares the RT of C-multicast route to its RT extended community (whose value

is set to VRF Route Import community). If there is a match, the C-multicast route is accepted and (S, G) is passed to C-multicast protocol on PE1/VPNA to be processed. Finally, the PE1 creates state in C-PIM database and propagates (10.10.11.2, 224.1.1.1) to CE1 towards the source Scr 1.

4.4.2.2 Scenario 4 (No. of MVPN Sites with RR vs. No. of control messages)

As the previous scenario (4.4.2.1), we have setup a MVPN testbed environment by deploying the NG MVPN. However, this time we have altered our testbed design in order to evaluate scalability of the control plane of the NG MVPN approach in respect to the number of MVPN sites. The network topology shown in the Figure 4.6is similar to the one in scenario 2 except that P router now is assuming the role of RR.

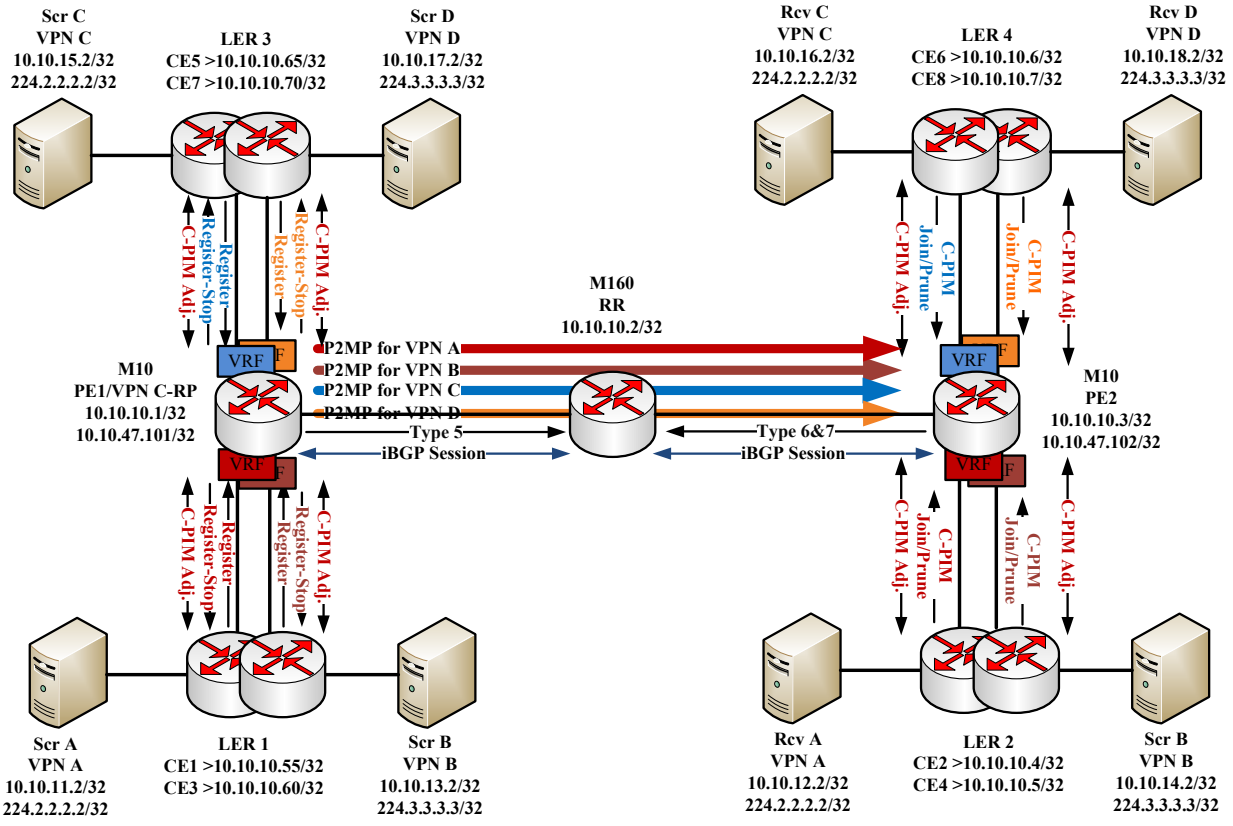


Figure 4.6: Testbed Setup for Scenario 4

In this scenario, the flow of the control messages (Type 5, 6 and 7 routes) is increased since there are four MVPN instances (VPN A, B, C and D) configured on each PE routers.

This means that VPN C-RP will be receiving four separate *PIM Register messages* from CE routers (CE1, CE3, CE5 and CE7). As a result, the PE1 router will generate and advertise a separate set of control messages to PE2 router; one for each configured MVPN site:

- One Type 1 routes for each VPN site (A, B, C and D) for PE auto-discovery; 1:65010:1:10.10.10.1, 1:65010:11:10.10.10.1, 1:65010:22:10.10.10.1 and 1:65010:33:10.10.10.1 respectively.
- One Type 5 routes for each VPN (A, B, C and D)as a response to *PIM Register messages* from CE routers; 5:65010:1:32:10.10.11.2:32:224.1.1.1, 5:65010:11:32:10.10.12.2:32:224.2.2.2, 5:65010:22:32:10.10.12.2:32:224.3.3.3 and 5:65010:33:32:10.10.12.2:32:224.1.1.4 respectively,

Additionally, the PE2 router will be receiving four separate sets of *Join/Prune messages* from CE routers. In response, the PE2 router will generate and advertise two distinct set of control messages to PE1 router; one for each configured MVPN site (VPN A, B, C and D):

- One Type 1 route (per each VPN site A, B, C and D) for PE auto-discovery;1:65010:2:10.10.10.3, 1:65010:12:10.10.10, 1:65010:13:10.10.10 and 1:65010:14:10.10.10 respectively.
- One Type 6 route for each configured VPN site as a response to *PIM Join messages* from CE routers; 6:65010:1:65010:32:10.10.47.101:32:224.1.1.1, 6:65010:11:65010:32:10.10.47.101:32:224.2.2.2, 6:65010:22:65010:32:10.10.47.101:32:224.2.2.3 and 6:65010:33:65010:32:10.10.47.101:32:224.2.2.4.
- One Type 7 route for each site VPN A, B, C and D in response to the Type 5 routes received from PE1 router;

7:65010:1:65010:32:10.10.11.2:32:224.1.1.1,7:65010:11:65010:32:10.10.12.2:32:224.2.2.2, 7:65010:22:65010:32:10.10.12.2:32:224.3.3.3 and 7:65010:33:65010:32:10.10.12.2:32:224.4.4.4.

4.4.2.3 Scenario 5 (No. of PEs without RR vs. No. of control messages)

In this scenario, we have setup a MVPN testbed environment similar to the one in Scenario 1 however this time; we have excluded the RR from the network topology in order to test the scalability of the control plane of the NG MVPN approach in respect to the number iBGP sessions as a worst case scenario. As shown in Figure 4.7, the network topology is similar to the one in the scenario 3 except that there is no RR. This results in large number of BGP peering sessions among PE routers.

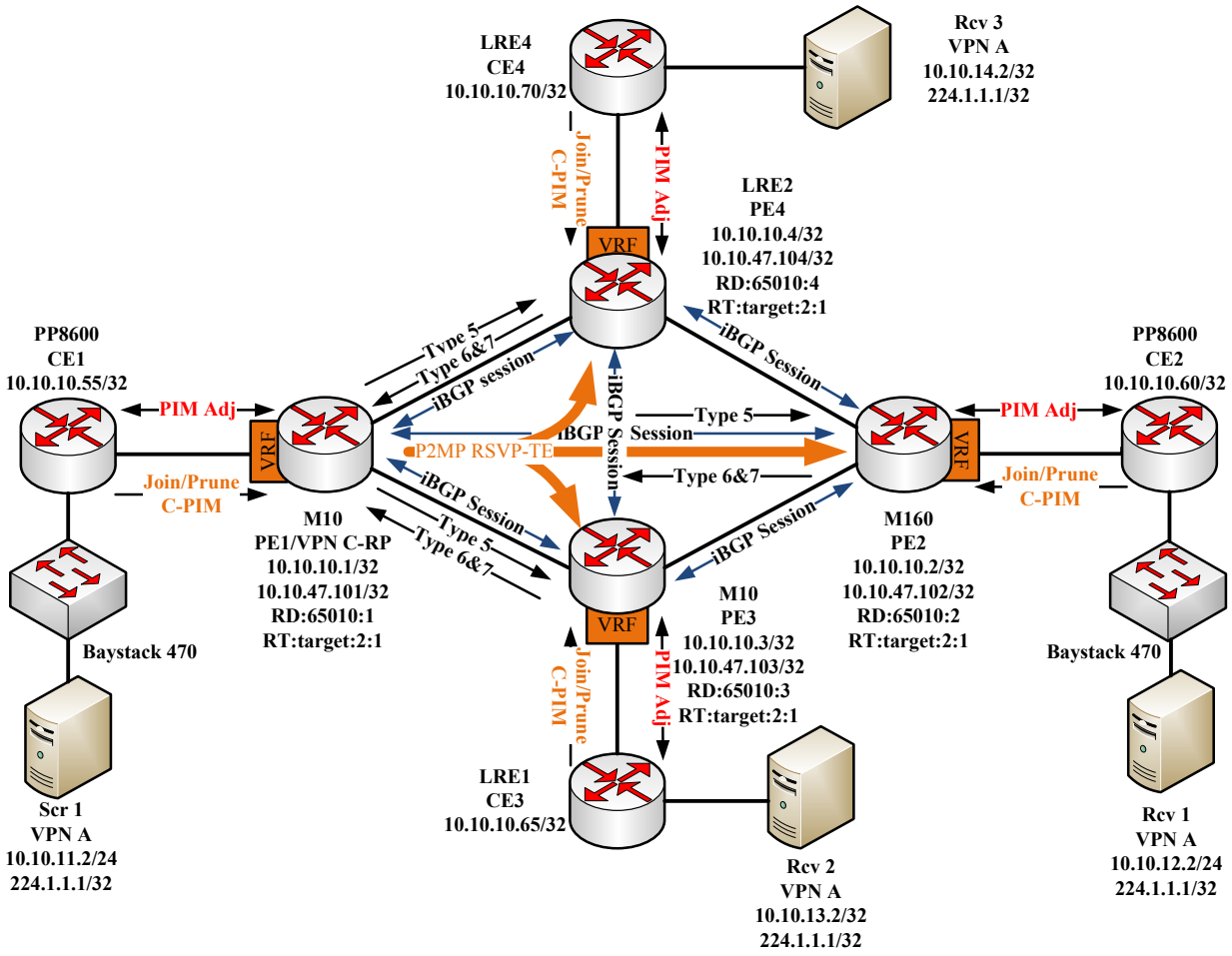


Figure 4.7:Testbed Setup for Scenario 5

In this scenario, the flow of the BGP control messages differs from the last previous scenarios since the RR is not included in our topology, which led to increase in the number of iBGP sessions because of the full mesh requirement by iBGP while the number of control messages is noticeably decreased when compared to Scenario 3.

Upon completion of the auto-discovery phase, the PE1 router generates and advertises three separate type 5 routes to PE2, PE3 and PE4 routers; one for each configured MVPN site (VPN A). Additionally, the PE2, PE3 and PE4 routers generate and advertise to PE1 router in response to *PIM Join messages* generated from CE routers as well as type 5 routes advertised from PE1 router.

4.5 Experimental Results

In this section, we show collected data that are obtained from our experiments in section 4.4 in order to help evaluate the scalability of the control plane in Rosen and NG MVPN schemes. These results are in form of number of control messages along with number of adjacencies/session in both schemes.

4.5.1 Rosen Scheme Results

In this section, we investigate the scalability of Rosen scheme's control plane by showing the results obtained from Scenarios 1 and 2. As we have mentioned in Section 4.2.1 and 4.2.2, each PE router is required to generate a number of specific PIM messages such as *PIM Register*, *PIM Register-stop*, *PIM Join* and *PIM Prune* according to its role. These messages are categorized as either trigger or periodic. The trigger PIM messages are sent during the initial operation of the scheme in order to allow the control plane to forward the multicast traffic from the source to the destination while the periodic PIM messages are primarily meant for maintaining the PIM state in the PE routers. We categorize these messages in order to better evaluating the scalability of

The Figure 4.8 shows the accumulated number of PIM control messages in respect to the number of PE routers in the initial operation. We have noticed that as we increase the number of PE routers in Scenario 1. This increment is accompanied with growing in the number of PIM adjacencies along with the number of BGP sessions. We have noticed that the behaviour of the Rosen scheme introduces scalability problems when there is a large deployment of PE routers due to the fact that the number of generated PIM messages is in function of number of PE routers.

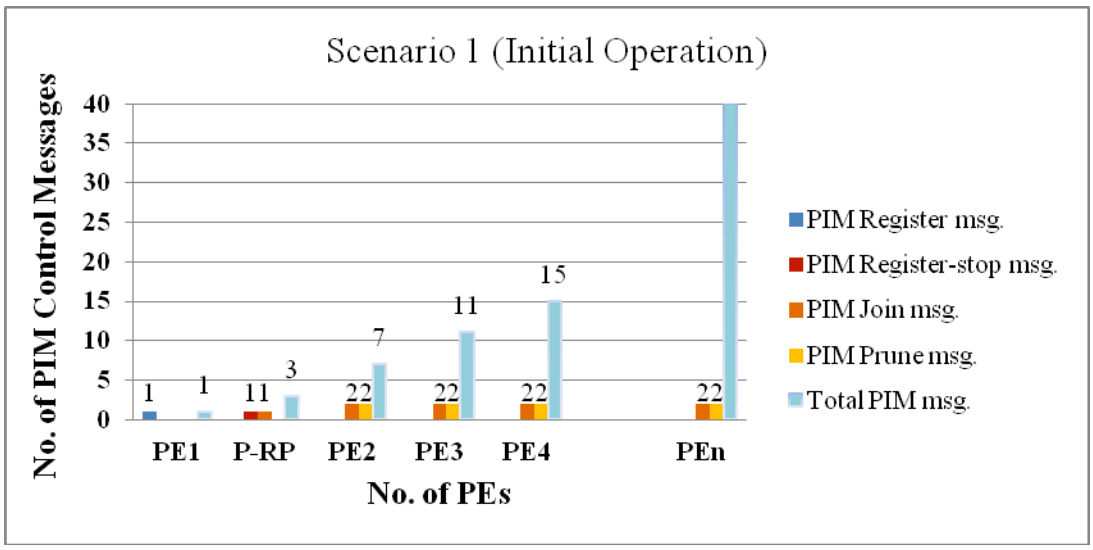


Figure 4.8: No. of PIM Control Messages verse No. of PE1 in Rosen scheme initial Operation

Correspondingly in Figure 4.9, we demonstrate the total number of PIM control messages in the steady operation of the scheme. This increasing is due to the fact that PIM is a soft state protocol. It sends PIM messages periodically (default interval is 60s) after establishing adjacencies between PE routers in order to maintain the PIM state. We have noticed that this increasing is correlated to the number of PE routers.

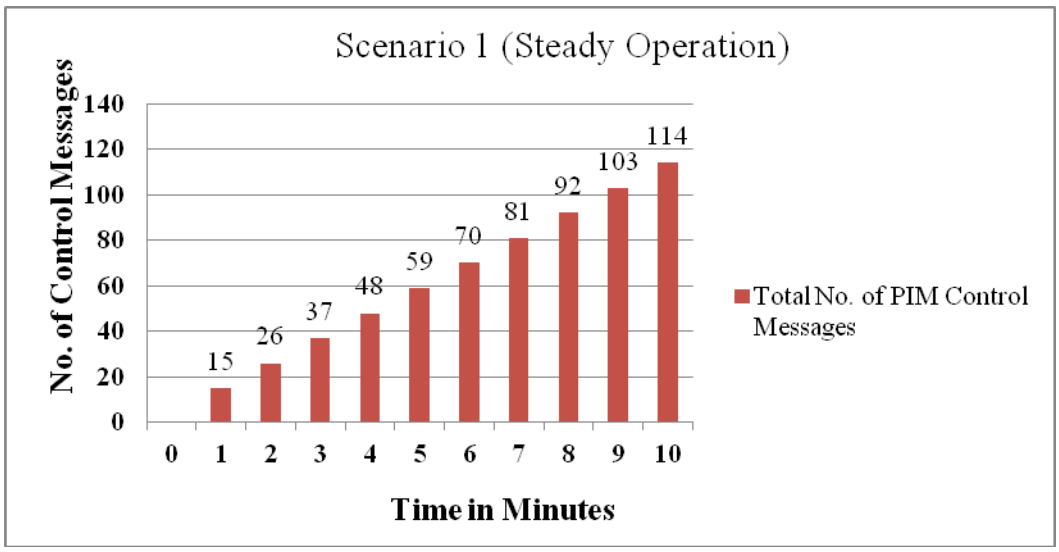


Figure 4.9: Total No. of PIM Control Messages in Rosen scheme Steady Operation

PIM messages for VPN A	PE1	P-RP	PE2	PE3	PE4	Total
P-PIM Adjacencies/ BGP Sessions	1 PE1-PE2 1 PE1-PE3 1 PE1-PE4		1 PE2-PE1 1 PE2-PE3 1 PE2-PE4	1 PE3-PE1 1 PE3-PE2 1 PE3-PE4	1 PE3-PE1 1 PE3-PE2 1 PE3-PE4	n(n-1) = 12
PIM Register Message	1 To P-RP					1
PIM Stop-Register Message		1 To P-RP				1
PIM Join Message		1 To PE1	1 To P-RP 1 To PE1	1 To P-RP 1 To PE1	1 To P-RP 1 To PE1	7
PIM Prune Message			1 To P-RP 1 To PE1	1 To P-RP 1 To PE1	1 To P-RP 1 To PE1	6
Time-based Operation	PIM sends periodic refresh messages every 60s					15

Table 4.1: PIM Control Messages for Scenario 1 in Rosen Scheme

Additionally, by examining the contents in Table 4.1, there is another factor that affects the scalability, which is the number of PIM adjacencies established between per-VRFs as show in Figure 4.10. In our network setup in Scenario 1, a total of 6 PIM adjacencies must be maintained per a single VRF (VPN A) as well as 6 BGP sessions per a PE router based on the following equation; (n (n–1)/2), where n presents the number of PE routers. This means that if we have100 PE routers in our network, this will require maintenance of 4950 PIM adjacencies as well as 4950 iBGP peering sessions leads to n(n-1) adjacencies/sessions in total.

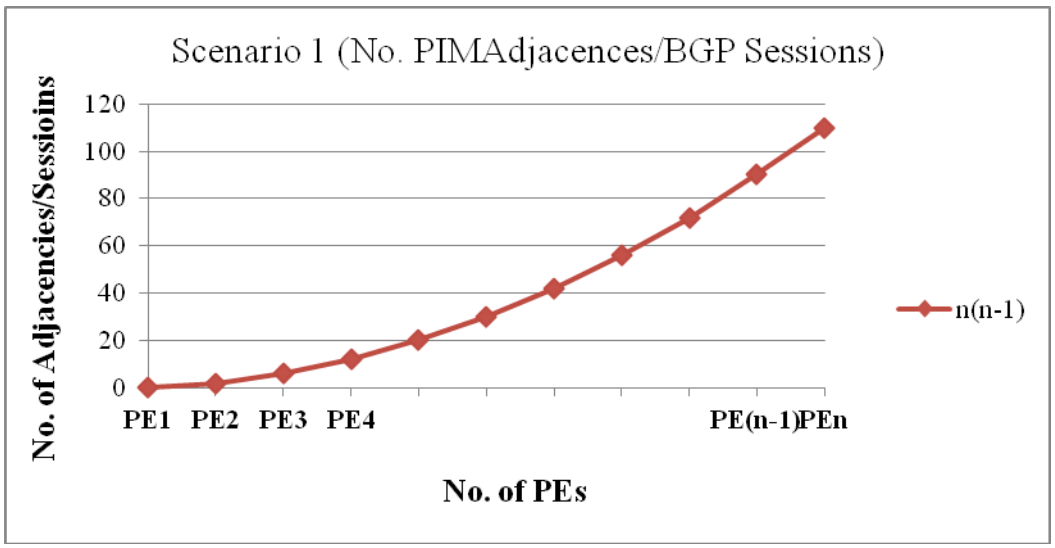


Figure 4.10: No. of Adjacencies/Sessions vs. No. of PEs in Rosen Scheme

Similarly, in Figure 4.11, we also show the relation between certain numbers of PIM control messages with the number of MVPN sites in the initial operation. We have set the number of PE routers to two, while we gradually increase the number of MVPN sites. We have noticed that as we increase the number of MVPN sites in Scenario 2, the number of PIM messages is increased linearly. This increment is also accompanied with huge growing in the number of PIM adjacencies along with the number of BGP sessions since we have four VRFs per each PE routers. This results in 4 PIM adjacencies (one per-VRF) Between PE1 and PE2 routers.

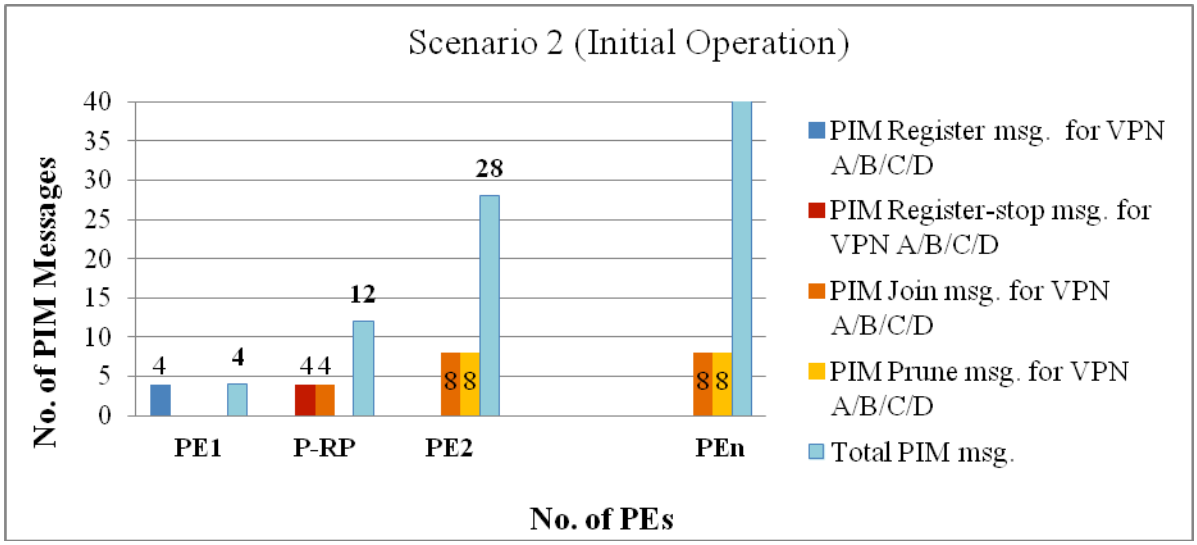


Figure 4.11: No. of PIM Control Messages verse No. of MVPN Sites in Rosen scheme initial Operation

Similarly in Figure 4.12, we demonstrate the total number of PIM control messages in the steady operation of the scheme. We have noticed that this increasing is related extensively to the number of MVPN sites. However, if we increase the number of MVPN along with the number of PE routers, we get greater impact on the scalability of the scheme. In our experiment, we have restricted the increasing in number of MVPN to four (VPN A/B/C/D) due to the limited number of devices.

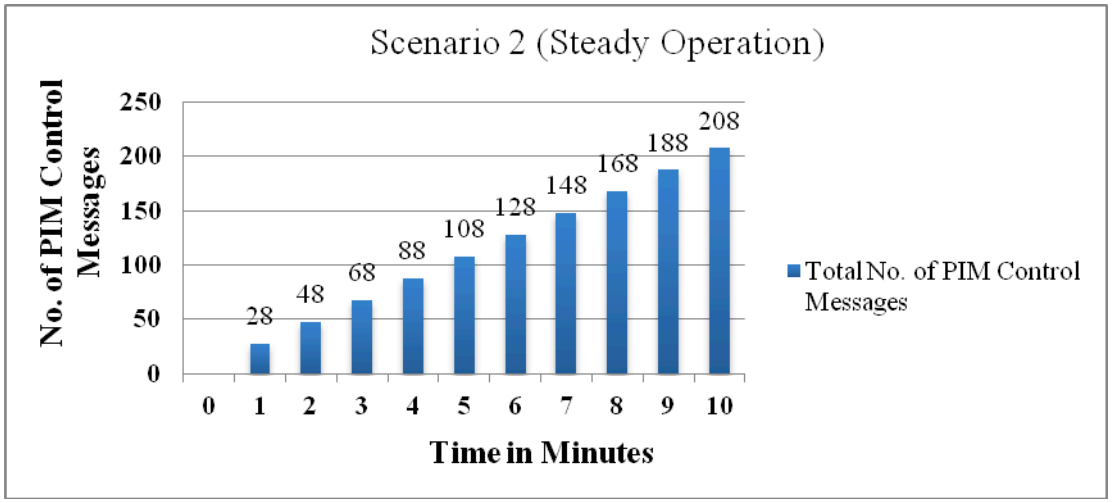


Figure 4.12: Total No. of PIM Control Messages verse in Rosen scheme Steady Operation

We have also noticed that the increasing in number of PIM messages in Scenario 2 is greater if compared to the results in Scenario 1. This indicates that the growing in number of MVPN sites has greater impact on scalability in Rosen scheme. This increasing can also be seen by observing the outcome from Table 4.2. Additionally, the total number of PIM Adjacencies is increased to 4 folds if compared to the results in scenario 1 due to the fact that PIM Adjacencies are established per-VRF basis as shown in Figure 4.13.

PIM messages for VPN A/B/C/D	PE1	P-RP	PE2	Total
P-PIM/iBGP Adjacencies/Sessions	1 PE1–PE2		1 PE2-PE1	2n(n-1) = 4
PIM Register Message	4 To P-RP			4
PIM Stop-Register Message		4 To PE1		4
PIM Join Message		4 To PE1	4 To P-RP 4 To PE1	12
PIM Prune Message			4 To P-RP 4 To PE1	8
Time-based Operation	PIM sends periodic refresh messages every 60s			28

Table 4.2: PIM Control Messages for Scenario 2 in Rosen Scheme

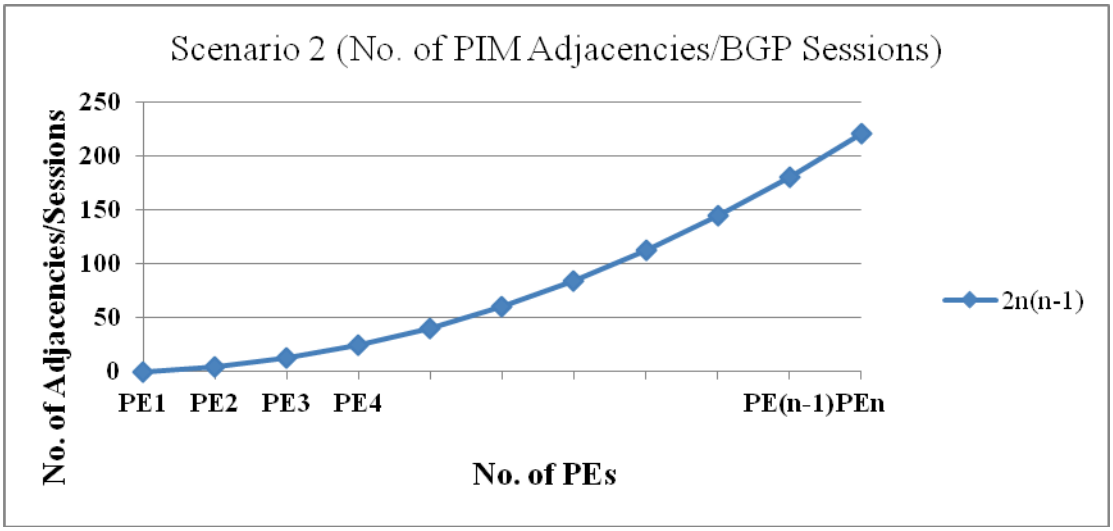


Figure 4.13: No. of Adjacencies/Sessions vs. No. of MVPNs/PEs in Rosen Scheme

4.5.2 NG MVPN Results

In this section, we investigate the scalability of NG MVPN's control plane based on the results obtained from Scenarios 3, 4 and 5. As we have seen in Section 4, each PE router is required to generate and advertise a number of specific BGP routes such as Type1, 3, 5, 6 and 7 MVPN

routes according to its role. These routes are stored in the MVRF tables and it is never refreshed due to the fact that BGP by default does not refresh its stored routes in routing table instead it uses BGP withdraw messages to remove the stale routes for its routing table. Therefore, there is no need of periodic MVPN control messages in BGP compared to what we have previously seen in PIM steady operation.

The Figure 4.14 shows the number of certain BGP messages in respect to the number of PE routers in the initial operation of NG MVPN scheme. Based on our experiments, we have noticed that as we increase the number of PE routers, the number of BGP messages is relatively increased.

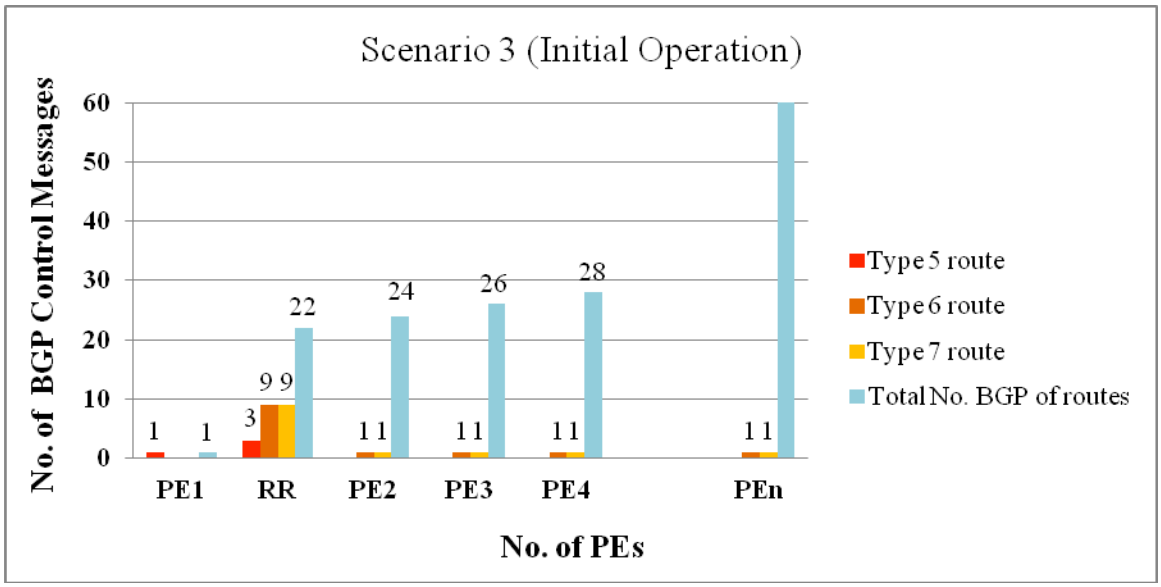


Figure 4.14: No. of BGP Control Messages verse No. of PEs in NG MVPN Initial Operation

It is worth to mention that we have not seen any changes to the number of BGP messages during the steady operation. This is not surprising because BGP by default does not refresh routes once they are installed in the routing table. BGP only installs or withdraw routes by means of its update messages.

Routes BGP for VPN A	PE1	RR	PE2	PE3	PE4	Total
iBGP Sessions	1 PE1-RR	To All PEs	1 PE2-RR	1 PE3-RR	1 PE4-RR	(n-1) = 3
Type 5Route	1 To RR	3 To All Clients				4
Type 6Route		9 To All Clients	1 To PE1	1 To PE1	1 To PE1	12
Type 7Route		9 To All Clients	1 To PE1	1 To PE1	1 To PE1	12
Time Base Operation	BGP routes never refreshed					28

Table 4.3: BGP Control Messages for Scenario 3 in NG MVPN Scheme

We also have noticed increasing in number of BGP sessions in scenario 3. However, this growth does not pose a great impact on the scalability as we have seen in scenario 1 and 2 due to the existing of RR. This is clearly observed from Figure 4.15. Furthermore, the data from Table 4.3 shows that deploying RRs is a good practice when there are a large number of PE routers in the network.

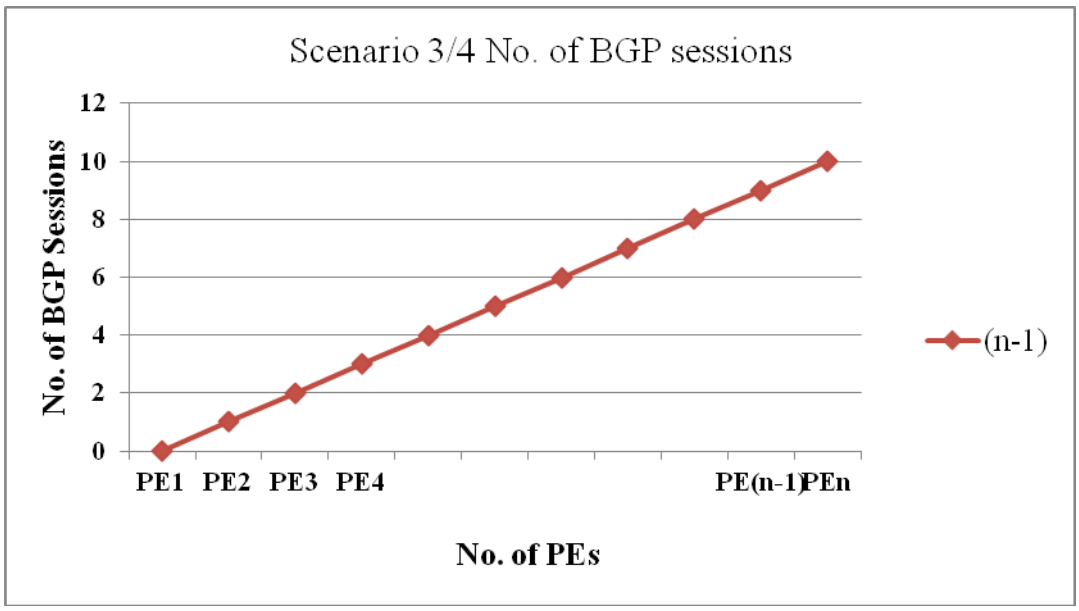


Figure 4.15: No. of BGP sessions VS. No. of PEs in NG MVPN Scheme

Correspondingly, in Figure 4.16, we show the relation between certain numbers of BGP control messages with the number of MVPN sites. We have restricted the number of PE routers to two due to the limited number of devices, while we gradually increase the number of MVPN sites to four (VPN A/B/C/D). We have noticed that as we increase the number of MVPN sites in Scenario 4, the number of BGP messages is increased. This increment is also accompanied with insignificant growth in the number of BGP sessions due to the deployment of RR as shown in Figure 4.15.

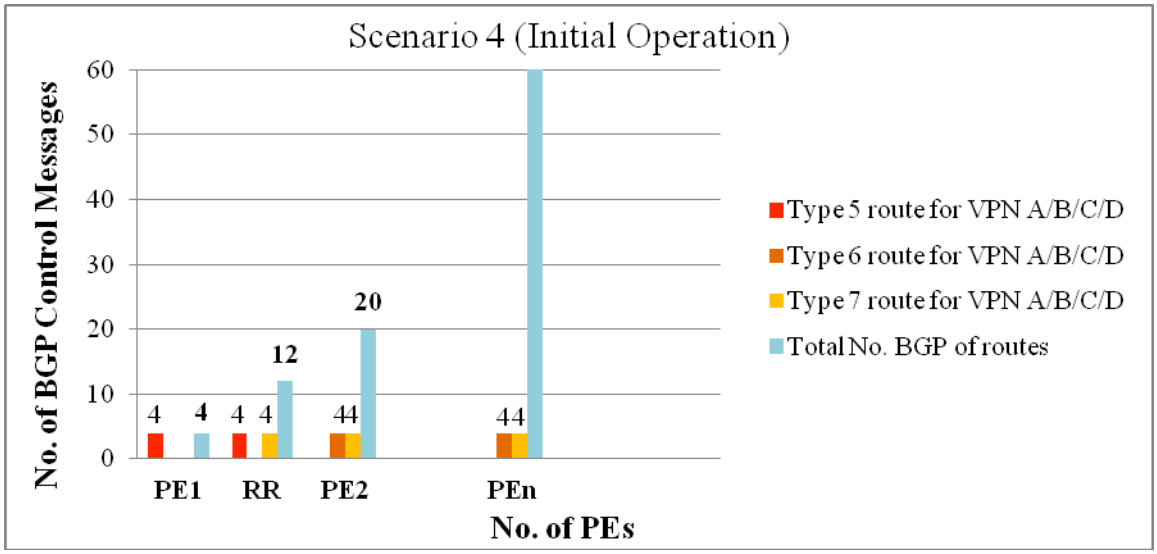


Figure 4.16: No. of BGP Control Messages verse No. of PEs in NG MVPN Initial Operation

Also, we have noticed that the increasing in number of BGP messages in Scenario 4 is less than the ones in Scenario 3. This indicates that the growing in number of MVPN sites has less impact on scalability in NG MVPN scheme. This also has been clarified by the results in Table 4.4.

BGP Routes for VPN A/B/C/D	PE1	RR	PE2	Total
iBGP Sessions	1 PE1-RR	To All PEs	1 PE2-RR	(n-1) = 1
Type 5Route	4 To RR	4 To All Clients		8
Type 6Route		4 To All Clients	4 To PE1	8
Type 7Route		4 To All Clients	4 To PE1	8
Time Base Operation	BGP routes never refreshed			24

Table 4.4: BGP Control Messages for Scenario 4 in NG MVPN Scheme

Once more in this scenario, we have not witnessed any changes to the number of BGP messages during the steady operation of the scheme. Moreover, from the results in Tables 4.3 and 4.4, we can see that the majority of BGP control messages are due to re-advertising behaviour of RR. This is of course trade off between number of messages and number of sessions.

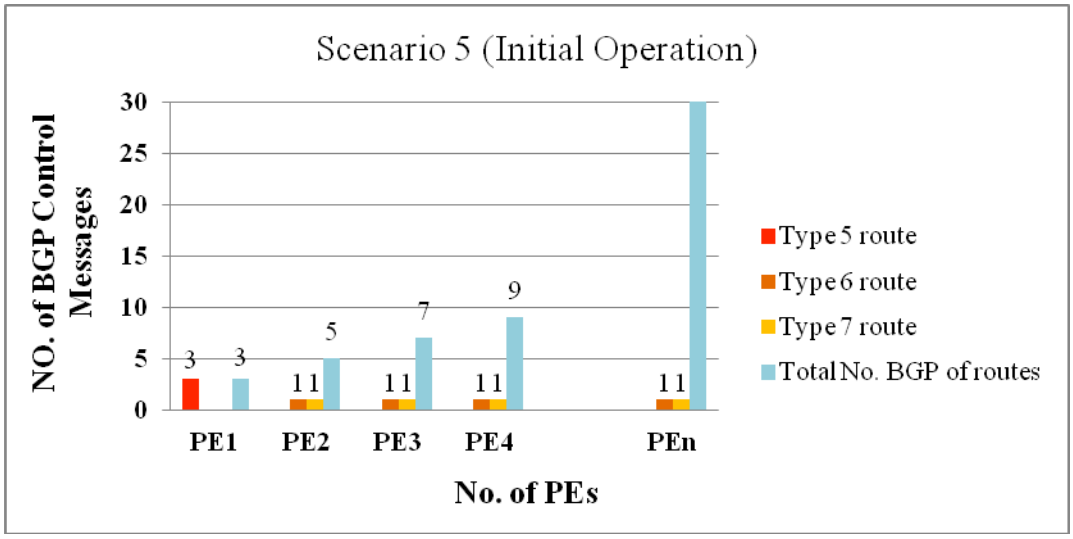


Figure 4.17: No. of BGP Control Messages verse No. of PEs without RR in NG MVPN Initial Operation

In Figure 4.17, we demonstrate the numbers of BGP control messages in respect to the number of PE routers. This time however we have excluded the RR from our calculation. The main reason behind this exclusion is to help understanding the impact of RR deployment on network scalability. We have noticed that without the existing of RR, the number of BGP messages is decreased compared to the results in Table 4.3 to reach total of nine messages when limiting the number of PE routers to four. As we have previously mentioned that RR re-advertisements make up most of BGP messages. However, we have noticed that there is growth in number of iBGP peering sessions because of the full-mesh connectivity requirement by the protocol implementation (n (n–1)/2), where n is the number of PE routers.

Also, we have noticed that there are no changes to the number of BGP messages during the steady state operation due to the same above-mentioned reasons. However, it is worth to mention that the increasing number of iBGP peering sessions is very significant and it can pose a great impact on the scalability of the network. These results are shown in Table 4.5.

BGP Routes	PE1	PE2	PE3	PE 4	Total
iBGP Sessions	1 PE1–PE2 1 PE1–PE3 1 PE1–PE4	1 PE2–PE1 1 PE2–PE3 1 PE2–PE4	1 PE3–PE1 1 PE3–PE2 1 PE3–PE4	1 PE4–PE1 1 PE4–PE2 1 PE1–PE3	$(n^2-n)/2 = 6$
Type 5Route	1 To PE2 1 To PE3 1 To PE4				3
Type 6Route		1 To PE1	1 To PE1	1 To PE1	3
Type 7Route		1 To PE1	1 To PE1	1 To PE1	3
Time-Base Operation	BGP routes never refreshed				9

Table 4.5: BGP Control Messages for Scenario 5 in NG MVPN Scheme

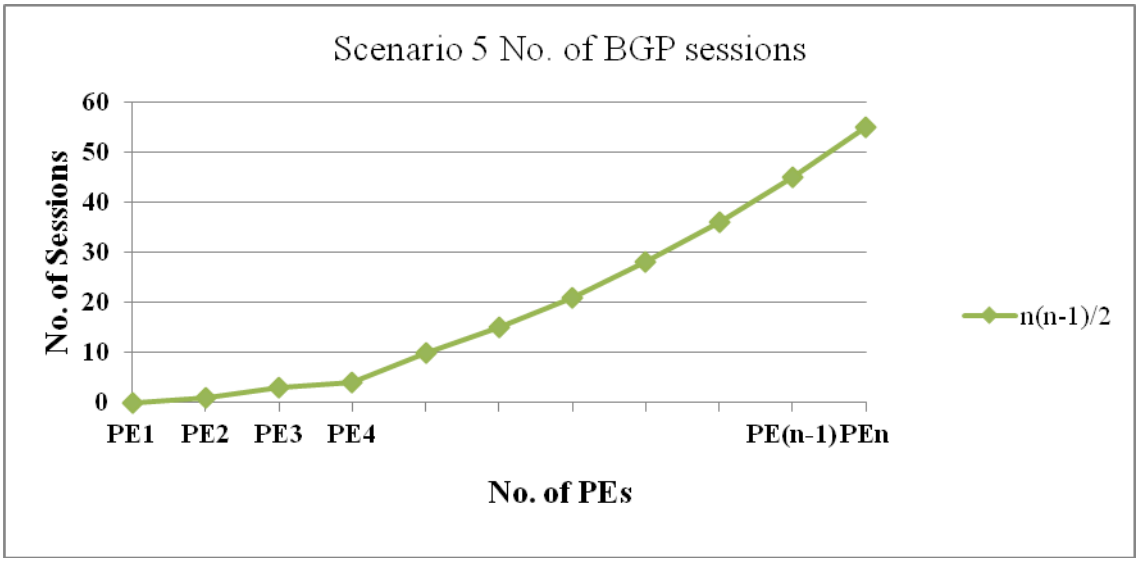


Figure 4.18: No. of BGP Session VS. No. of PEs without RR

4.6 Comparison

In this section, we have performed a set of scalability test comparisons between Rosen scheme and NG MVPN based on results obtained from our experiments. The reason behind this comparison is to show which scheme is more scalable than the other in terms of total number of control messages generated along with number of adjacencies/sessions required.

In Figure 4.18, we present the total number of control messages from Scenarios 1, 3 and 5 since they lies under the same test category. By looking at the graph below, we can see that Rosen scheme in Scenario 3 generates more control messages than the ones generated by NG MVPN scheme in Scenarios 3 and 5. Additionally, Rosen scheme needs to maintain a large number of adjacencies as well as a large number of iBGP peering sessions as the number of PE routers increases. This is due to the fact that MPLS/BGP unicast solution must be implemented prior to the deployment of Rosen scheme.

The graph also shows that the number of BGP messages in Scenario 5 is fewer than the ones in Scenario 3. This is due to that the RR re-advertises all routes received from one PE router to other PE routers, which leads to more BGP messages in the network. This is of course trade of between deployment of RR and number of iBGP sessions. However, as the number of PE routers grows higher, the deployment of RR becomes necessity in order to aid the scalability of the network.

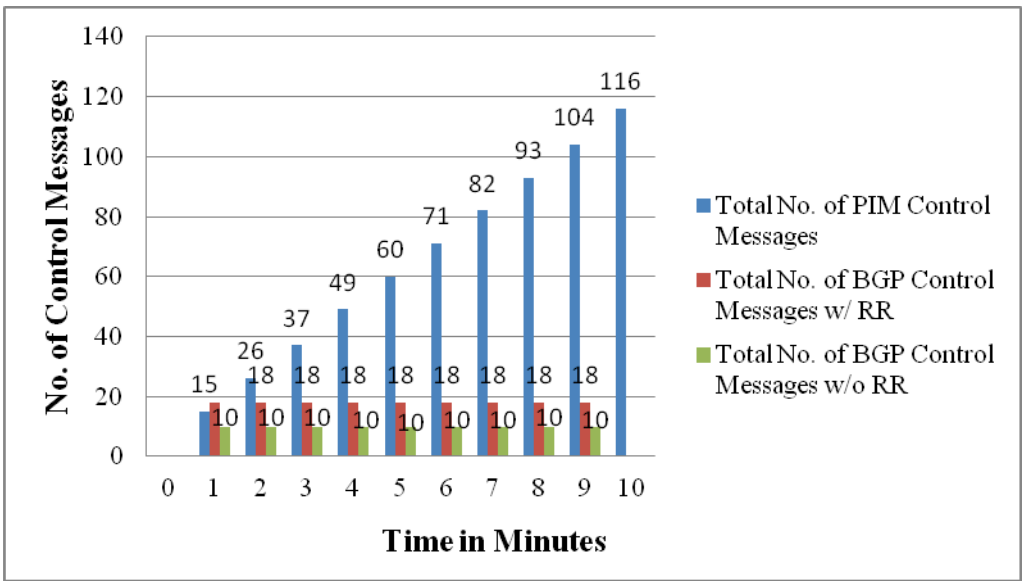


Figure 4.19: Comparison in No. of Control Messages between Scenario 1, 3 and 5

In Figure 4.19, we demonstrate the total number of control messages by comparing results from Scenarios 2 and 4 since they fall under the same test category. By looking at the graph below, we can see that Rosen scheme in Scenario 2 generates more control messages when compared with the results in Scenario 4. We have noticed that total number of control messages in Scenario 4 is much less what we have anticipated.

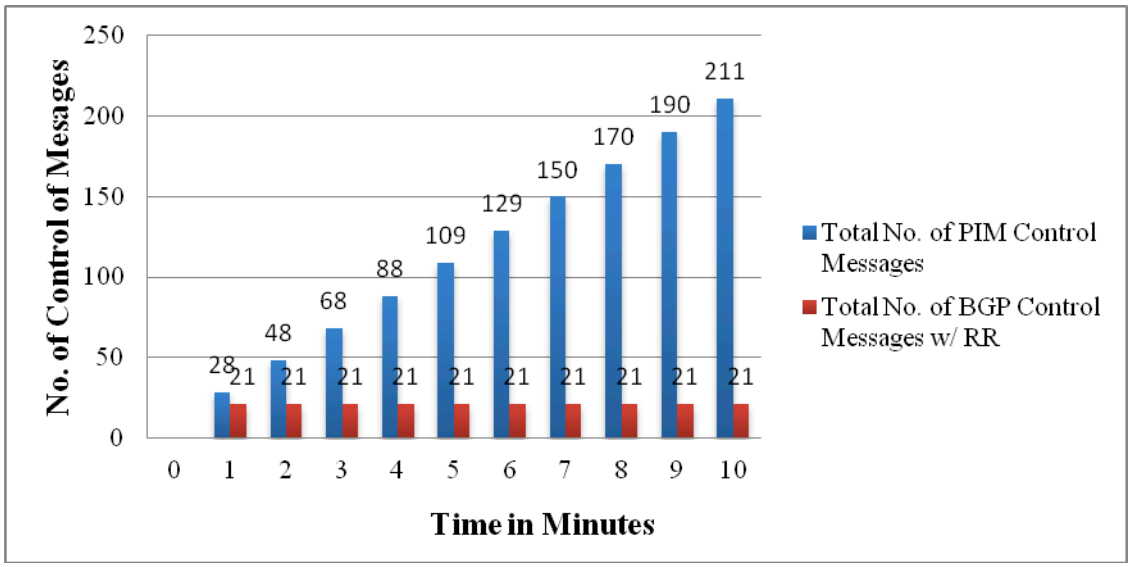


Figure 4.20: Comparison in No. of Control Messages between Scenario 4 and 5

4.7 Summary

In this chapter, we have performed several experimental tests in order to evaluate the scalability of Rosen scheme (PIM-SM/GRE) and NG MVPN (BGP/RSVP-TE P2MP) in IP /MPLS core networks. In our experiments, we have created several experiments as many as five test cases to evaluate the scalability in terms of number of control messages and number of sessions/adjacencies required between PE routers.

To summarize, the Rosen scheme departs from the 2547 model in that it implements the Virtual Router model rather than the Aggregated Routing model for VPN support and uses different mechanisms in both the control and the data planes. By doing so, it not only introduces a second set of mechanisms for multicast but also loses a lot of the scalability and flexibility of the unicast L3VPN solution.

Additionally, based on the results shown in the previous sections, we conclude that the NG MVPN scheme is more scalable than the Rosen scheme. Our conclusion came after we assist the number of control messages that each scheme requires to generate or advertise, taking into account the total number of adjacencies/sessions required in each scheme. Additionally, all our results shown in previous sections are mostly complied with the results in [MVPN-BGP].

Chapter 5

Concluding Remarks and Future Work

5.1 Concluding Remarks

The vast deployment of multicast applications, such as IPTV and media-rich services merging over the MPLS infrastructure has created requirements for efficient and reliable multicast transport, with guaranteed Quality of Service, through service provider core networks [MET09].

One solution is to deploy Rosen scheme as described in the IETF draft [ROS10]. This scheme is an overlay to BGP/MPLS IP VPNs because customer multicast state is signalled differently than customer unicast routes, and customer multicast packets are encapsulated differently than customer unicast packets. More specifically, the C-multicast signalling specified by Rosen scheme is PIM (as opposed to BGP for unicast routes), and the C-multicast traffic encapsulation is GRE (as opposed to MPLS for unicast traffic) [ALC10]. However, the addition of multicast brings an entire new set of protocols and procedures for the service provider.

Amore recently solution, which proposes to use BGP as the control plane for MVPN as described in IETF drafts [AGG09] and [MOR10]. This solution generalizes the key concept of the Rosen scheme in order to introduce new options for the C-multicast signalling and multicast traffic encapsulation.

In this work, we have provided methods and tools that study and evaluate the scalability and performance of multicast VPN schemes in IP/MPLS core networks. Our tests have been carried out on a testbed environment that consists of multi-vendor network devices to ensure the interoperability. Additionally, in this thesis, we have demonstrated pros and cons of each scheme in order to identify which scheme that is the best candidate to be deployed in core networks. Also, we have performed a set of scalability test comparisons between the Rosen scheme and the

NG MVPN based on results obtained from each experiment. The main reason behind this comparison is to show which scheme is more scalable than the other in terms of total number of control messages generated along with number of adjacencies/sessions required.

Our results show that the scalability of NG MVPN control plane is far less complex when compared to Rosen scheme's due to the fact that the Rosen scheme generates more control messages than the ones generated by NG MVPN's. Also, the number of peering sessions maintained in the Rosen scheme is almost twice the number of peering sessions in NG MVPN scheme (in case of RR deployment in the network). Therefore, the deployment of NG MVPN scheme over the Rosen scheme could reduce network complexity, scaling problems, and potential failure of core nodes by reducing the number of protocols (and hence the state/memory and CPU load) executing on core nodes.

5.2 Future Research

While all the tests we have performed in this thesis are comprehensive, there is still much work to be done in the area of MVPN. For instance, in our first two scenarios in Section 4.4.1, we have deployed Rosen scheme in PIM-SM environment. However, future studies might attempt to test the scalability of Rosen in PIM-SSM as described in [ROS10]. This requires PE auto discovery process across the multicast domain, which is done by the use of BGP Address Family named Multicast Data Tree Subsequent Address Family Identifier (MDT SAFI) defined in [NAL06]

Another possibility would be to use a different data plane, such as mLDP LSPs. This would allow PIM to take advantage of MPLS LSP instead of relying on purely IP tunnels. mLDP can construct LSPs without relying on or interacting with any multicast tree construction protocol. Therefore, migrating to or deploying mLDP for MVPN in the core can result in a common data plane for both unicast and multicast streams [CIS10].

Also in this work, we have mainly discussed issues concerning the control plane in MVPN schemes. However, more studies can be carried out on the data plane regarding performance and

resiliency issues. One of these issues is the efficiency of using S-PMSI tunnels over DATA-MDT in core networks and how they can affect the performance. Another issue would be to compare between P2MP RSVP-TE and mLDP in terms of network resiliency.

Additionally, these tests can further be carried out to include Multicast in Layer 2 VPNs. This topic is gaining attention from SP and enterprise businesses. For instance, VPLS is considered to be valuable solution to provide many-to-many Layer 2 VPN services. However, there are many challenges to provide multicast services through VPLS technology. One of these challenges is the requirement of scalable control plane such as BGP, another challenges concerning the need of reliable data plane such as P2MP RSVP-TE or mLDP LSPs.

Reference

[ADA07] D. Adami, S. Giordano, F. Mustacchio and M. Pagano, “Design and development of a DS-TE experimental testbed with Point-to-Multipoint LSP support”, 3rd Conference on Next Generation Internet Networks (EuroNGI 2007), Trondheim, Norway, pp. 14-20, May 2007.

[AGG07]R. Aggarwal, D. Papadimitriou and S. Yasukawa, “Extensions to Resource Reservation Protocol - Traffic Engineering (RSVP-TE) for Point-to-Multipoint TE Label Switched Paths (LSPs)”, RFC 4875, May 2007.

[AHS08] M. A.Chishti, S. Qureshi and A.H. Mir, “Implementation of Source Specific Multicast (SSM) in a test bed over Internet Protocol version 6 (IPv6)”, 2nd International Conference on Internet Multimedia Services Architecture and Applications (IMSAA 2008), Bangalore, India, pp.1-5, December 2008.

[ALA07] B. Alawieh and H. T. Mouftah, “Delivering Multicast Services over MPLS Infrastructure”, 12th IEEE Symposium on Computers and Communications (ISCC 2007), Aveiro, Brazil, pp. 509-514, June 2007.

[ALC10] Alcatel-Lucent, “Next-generation Layer 3 Multicast VPN (MVPN) Services”,CPG2896100217, Application Note, June 2010.

[ASH07] G. R. Ash, “Traffic Engineering and QoS Optimization of Integrated Voice & Data Networks”, ISBN 13: 978-0-12-370625-6, Morgan Kaufmann, October 2007.

[BAT07] T. Bates, R. Chandra, D. Katz and Y. Rekhter, “Multiprotocol Extensions for BGP4”, RCF 4760, January 2007.

[BEI02]I. VanBeijnum, “BGP”, O’Reilly, ISBN-13: 978-0596002541, California, USA, September 2002.

[BIR06]S. Biradar, B. Alawieh and H.T.Mouftah, “Delivering reliable real-time multicast services over virtual private LAN service”, 2nd International Conference on Testbeds and Research Infrastructures for the Development of Networks and Communities (TRIDENTCOM 2006), Barcelona, Spain, pp. 552-557, July 2006.

[BLA05] T. Blaga, V. Dobrota, K. Steenhaut, I. Trestian and G. Lazar, “Steps towards native IPv6 multicast: CastGate router with PIM-SM support”, the 14th IEEE Workshop on Local andMetropolitan Area Networks (LANMAN 2005), Crete, Greece, pp. 5-11, Sep2005.

[BRE06] L. Breslau, C. Chasey, N. Duffield, B. Fenner, Y. Maoz and S.Sen, “VMScope: a virtual multicast VPN performance monitor”, Proceedings of the 2006 SIGCOMM workshop on Internet network management(INM ’06), New York, USA, pp. 59-64,March2006.

[CIS02] Cisco Systems, “Multicast Virtual Private Networks”, Technology white paper, February2002.

[CIS06] Cisco Systems, “Multicast VPN Design Guide”, Design Guide, May 2006.

[CIS10] Cisco Systems, “Core of Multicast VPNs: Rationale for Using mLDP in the MVPN Core”, C11-598929-00, White paper, April 2010.

[CLE06]J. De Clercq, D. Ooms, M. Carugi, F. Le Faucheur, “BGP-MPLS IP Virtual Private Network (VPN) Extension for IPv6 VPN”, RFC 4659, September 2006.

[GOR02] W. J. Goralski, “Routing Policy and Protocols for Multivendor IP Networks”, Wiley, ISBN: 978-0-471-21592-9, September 2002.

[HAS10] A. Hassan, M. Bazama, T. Saad, and H. T. Mouftah, "Investigation of Fast Reroute Mechanisms in an Optical TestbedEnvironment", 2010High-Capacity Optical Networks and Enabling Technologies (HONET), Cairo, Egypt, pp. 247-251, February 2011.

[HIG01]L. Higgins, G. McDowell and B. VonTobel, "Tunneling multicast traffic through non-multicast-aware networks and encryption devices", Military Communications Conference on communications for Network-Centric Operations MILCOM 2001, pp. 296-300, August 2001.

[HON09] F. Hong and Y. Bai, "Design and analysis of multicasting experiment based on PIM-DM", Conference on Computer Science and Information Technology, Beijing, China, pp. 9-13, August2009.

[JUN08] Juniper Networks, "Deploying Next-Generation Multicast VPN", E. Gagala, white paper, May 2008.

[JUN09] Juniper Networks, "Best Practices for Video Transit on an MPLS Backbone",2000106-003-EN, White Paper, April 2009.

[JUN10] Juniper Networks, "Understanding Junos OS Next-Generation Multicast VPNs", 2000320-002-EN, White paper, January 2010.

[JUN10a] Juniper Networks, "NG MVPN BGP Route Types And Encodings",3500142-002-EN,Application note, May 2010.

[JUN10b] Juniper Networks, "Optimizing Media-Rich Content Delivery with Point-to-Multipoint Label Switched Paths", 2000274-002-EN, White Paper, August 2010.

[KUM08] K. Kumakit, I. Nakagawa, K. Nagami, T. Ogishit and S.Anot, "Design and evaluation of a P2MP MPLS-based hierarchical service management system", IEEE Symposium on Computers and Communications ISCC 2008, Marrakech, Morocco, pp. 794-799, Jul 2008.

[LAV03] T. Lavian, P. Wang, R. Durairaj, D. Hoang and F. Travostino, "Edge device multi-unicasting for video streaming", Conference on Telecommunications ICT 2003, pp. 1441-1447, Mar 2003.

[LIU08] F-H. Liu, J-M. Xu and H-B. Qin, "Multicast Domain on E-Government MPLS VPN", 11th IEEE International Conference on ICCT 2008, Hangzhou, China, pp. 355-358, June 2008.

[LUE09] R. Luebben, G. Li, D. Wang, R. Doverspike and X. Fu, "Fast Rerouting for IP Multicast in Managed IPTV Networks", IWQoS, Charleston, USA, pp. 1-5, Jul 2009.

[MAT06] A. Matrawy, W. Yi, I. Lambadaris and C-H. Lung, "MPLS-based multicast shared trees", Proceedings of the 4th Annual Communication Networks and Services Research Conference CNSR 2006, Moncton, Canada, pp. 234-242, May 2006.

[MAT07] I. Martinez-Yelmo, D. Larrabeiti, I. Soto and P. Pacyna, "Multicast traffic aggregation in MPLS-based VPN networks", IEEE Communications Magazine, pp. 78-85, October 2007.

[MET06]C.Metz, "Multiprotocol label switching and IP. Part 2. Multicast virtual private networks", IEEE Internet Computing Magazine, pp. 76-81, February 2006.

[MET09] Metaswitch Networks, "RSVP-TE Point-to-Multipoint", NicNeate,White paper, March 2009.

[MIN08]I. Minei and J. Lucek, "MPLS-enabled application: emerging developments and new technologies 2nd Edition", Wiley, ISBN: 978-0-470-98644-8, July 2008.

[MOY98] J. Moy, "OSPF Version 2", RFC 2328, April 1998.

[MVPN-BGP] T. Morin, B. Niven-Jenkins, R. Zhang, N. Leymann and N. Bitar,"Mandatory Features in a Layer 3 Multicast BGP/MPLS VPN Solution", draft-ietf-l3vpn-mvpn-

considerations-06, February 2010.

[NAD10] T.D. Nadeau, A.K. Koushik, and R. Cetin, “Multiprotocol Label Switching (MPLS) Traffic Engineering Management Information Base for Fast ReRoute”, Internet Draft, April 2010.

[NAL06] G.Nalawade, A.Sreekantiah, “Multicast Data Tree Subsequent Address Family Identifier”, Internet Draft, Draft-nalawade-idr-mdt-safi-03.txt, April 2006.

[OVU08] OVUM, "The benefits of P2MP label switched paths for MPLS networks”, white paper, November 2008.

[PAL05] F.Palmieri,” Evaluating Evolutionary IP-Based Transport Services on a Dark Fiber Large-Scale Network Testbed”, Lecture Notes in Computer Science ICN 2005, ISBN: 978-3-540-25339-6, November2005.

[PAN05] P. Pan, G. Swallow and A. Atlas, “Fast ReRoute Extensions to RSVP-TE for LSP Tunnels”, RFC 4090, May 2005.

[QU10] S. Qu and J.Lindqvist, “Scalable IPTV Delivery to Home via VPN”, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2010, ISBN: 978-3-642-12630-7, May 2010.

[**QUINN01]** B. Quinn and K. Almeroth, “IP Multicast Applications: Challenges and Solutions”, RFC 3170, September 2001.

[RAH08] M. A. Rahman, A. H. Kabir, K. A. M. Lutfullah, M. Z. Hassan and M. R. Amin, “Performance Analysis and the Study of the behaviour of MPLS Protocols”, Proceedings of the International Conference on Computer and Communication Engineering, Kuala Lumpur, Malaysia, pp. 226 – 229, May 2008

[ROS06] E. C. Rosen and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, February 2006.

[AGG09] R.Aggarwal and E. C. Rosen, ”BGP Encoding and Procedures for Multicast in MPLS/BGP IP VPNs”, Internet Draft, draft-ietf-l3vpn-2547bis-mcast-bgp-08.txt, October 2009.

[ROS10] E. C. Rosen, Y. Cai and I. Wijnands, “Cisco Systems’ Solution for Multicast in BGP/MPLS IP VPNs”, RFC6037, October 2010.

[ROS11] E. C. Rosen, Y. Cai, I.Wijnands, “MVPN: S-PMSI Join Extensions for mLDP-created Tunnels”, Internet Draft, November 2011.

[SAA05] T. Saad, B. Alawieh, and H. T.Mouftah, “Inter-VLAN VPNs over a High Performance Optical Testbed”**,** First International Conference on Testbeds and Research Infrastructures for the Development of Networks and Communities Tridentcom 2005, pp. 221-229, February 2005.

[SOR03] J. M. Soricelli, J. L. Hammond, G. D. Pildush, T. E. Van Meter, and T. M. Warble, “Juniper Networks Certified Internet Associate”,Sybex, ISBN: 0-7821-4071-8, June 2003.

[WAT07] M. Watari, A. Tagami, T. Hasegawa and S. Ano, “A Scalable and Efficient Client Emulation Method for IP Multicast Performance Evaluation”, 21st International Conference on Advanced Information Networking and Applications Workshops AINAW '07, Niagara Falls, Canada, pp. 142-147, May 2007.

[WEN07] C-C. Wen, C-S. Wu and K-J. Chen, “Centralized Control and Management Architecture Design for PIM-SM Based IP/MPLS Multicast Networks”, IEEE Global Telecommunications Conference GLOBECOM '07, Washington-DC, USA, pp. 443-447, November 2007.

[YU00]X. W. Yu, C. Schulzrinne, H. Stirpe and P. W. Wu, “IP Multicast Fault Recovery in PIM

over OSPF", Proceedings 2000 International Conference on Network Protocols, Osaka, Japan, pp. 116-125, November 2000.

[ZHO06] C-YZhou, H-K Zhang and Y. Liu, "A Solution for IP Multicast VPNs based on Virtual Routers", 8th International Conference on Signal Processing 2006, Beijing, China, pp. 255-258, April2006.

Appendix A

Router Configuration

PE1Router Configuration for Draft Rosen

```
ottawalab@M10> show configuration
## Last commit: 2010-80-21 01:20:42 EDT by ottawalab
version 10.2R1.8;
system {
host-name M10;
time-zone America/Montreal;
root-authentication {
encrypted-password "$1$BZ71x/9p$6LfXhfThg.LB9Zv7Bmu0f0"; ## SECRET-DATA
}
login {
userottawalab {
uid 2000;
classsuperuser;
authentication {
encrypted-password "$1$ZtqeCKR0$b9ENZLHV6R3DDoHu2rcpd1"; ## SECRET-DATA
}
services {
telnet;
}
syslog {
user * {
any emergency;
}
file messages {
```

```
any notice;
authorization info;
ntp {
peer 10.10.10.1;
interfaces {
ge-0/0/0 {
description connected_to_CE1_8600-A;
unit 0 {
familyinet {
address 10.0.67.14/30;
}
ge-0/2/0 {
description "connected_to_P_M10#2";
unit 0 {
familyinet {
address 10.0.78.5/30;
        }
familympls;
      }
fxp0 {
unit 0 {
familyinet {
address 192.168.63.9/24;
}
lo0 {
unit 0 {
familyinet {
address 10.10.10.1/32 {
primary;
      }
unit 1 {
familyinet {
address 10.10.47.101/32;
}
routing-options {
rib-groups {
vpna-mcast-rib {
export-rib vpn-a.inet.2;
import-rib vpn-a.inet.2;
}
autonomous-system 0.65010;
}
protocols {
rsvp {
traceoptions {
file rsvp-log;
flag all;
      }
interface ge-0/0/0.0;
interface ge-0/3/0.0;
```

```
interface ge-0/2/0.0;
}
mpls {
label-switched-path to-PE2 {
to 10.10.10.2;
        }
interface ge-0/0/0.0;
interface lo0.0;
interface ge-0/3/0.0;
interface ge-0/2/0.0;
    }
bgp {
traceoptions {
filebgp-log;
flag all;
        }
group group-mvpn {
type internal;
local-address 10.10.10.1;
familyinet-vpn {
unicast;
            }
neighbor 10.10.10.2;
neighbor 10.10.10.3;
neighbor 10.10.10.4;
        }
    }
ospf {
traffic-engineering {
shortcuts;
        }
area 0.0.0.0 {
interface lo0.0;
interface ge-0/2/0.0;
        }
    }
pim {
traceoptions {
filepim-log;
flag all;
        }
rp {
static {
address 10.10.10.2;
            }
        }
interface lo0.0 {
mode sparse;
version 2;
        }
interface ge-0/2/0.0 {
mode sparse;
```

```
version 2;
}
policy-options {
policy-statement bgp-to-ospf {
from protocol bgp;
then accept;
   }
}
routing-instances {
vpn-a {
instance-type vrf;
interface ge-0/0/0.0;
interface lo0.1;
route-distinguisher 65010:1;
vrf-target target:2:1;
vrf-table-label;
protocols {
ospf {
exportbgp-to-ospf;
area 0.0.0.0 {
interface all;
            }
         }
pim {
traceoptions {
filevpn-pim-log;
flag all;
            }
vpn-group-address 224.1.1.1;
rib-group inetvpna-mcast-rib;
rp {
local {
address 10.10.47.101;
               }
            }
interface lo0.1 {
mode sparse;

version 2;
            }
interface ge-0/0/0.0 {
mode sparse;
version 2;
            }
}
```

PE1 Router Configuration for NG MVPN

```
ttawalab@M10> show configuration
## Last commit: 2010-10-16 16:55:02 EDT by ottawalab
version 10.2R1.8;
system {
```

```
host-name M10;
time-zone America/Montreal;
ports {
console type vt100;
    }
root-authentication {
encrypted-password "$1$jg4a6y9h$v/ct5QUNqGhXfAnghUp7T0"; ## SECRET-DATA
    }
login {
userottawalab {
uid 2000;
class super-user;
authentication {
encrypted-password "$1$l.iv9D85$zIUJbkQFIYwWQYBIMb6Wz1"; ## SECR
ET-DATA
services {
ssh;
telnet;
    }
syslog {
user * {
any emergency;
        }
file messages {
any notice;
authorization info;
        }
file interactive-commands {
interactive-commands any;
        }
    }
ntp {
peer 10.10.10.2;
interfaces {
ge-0/0/0 {
description connected_to_CE1_8600-A;
unit 0 {
familyinet {
address 10.0.67.14/30;
    }
ge-0/3/0 {
description connected_to_CE2_8600-B;
unit 0 {
familyinet {
address 10.1.67.14/30;
            }
    }
so-1/0/0 {
unit 0 {
description connected_to_P_M160;
familyinet {
address 10.0.67.14/30;
```

```
        }
familympls;
      }
   }
fxp0 {
unit 0 {
familyinet {
address 192.168.63.9/24;
   }
lo0 {
unit 0 {
familyinet {
address 10.10.10.1/32 {
primary;
            }
      }
unit 1 {
familyinet {
address 10.10.47.101/32;
         }
      }
unit 2 {
familyinet {
address 10.10.48.101/32;
}
routing-options {
autonomous-system 0.65010;
}
protocols {
rsvp {
traceoptions {
file rsvp-log;
flag all;
      }
interface ge-0/0/0.0;
interface so-1/0/0.0;
interface ge-0/3/0.0;
   }
mpls {
label-switched-path to-PE2 {
to 10.10.10.3;
      }
label-switched-path p2mp-template-mvpn-1 {
traceoptions {
file p2mp-log;
flag all;
         }
template;
link-protection;
p2mp;
      }
label-switched-path p2mp-template-mvpn-2 {
```

```
template;
link-protection;
p2mp;
        }
label-switched-path to-PE3 {
to 10.10.10.4;
        }
interface ge-0/0/0.0;
interface lo0.0;
interface so-1/0/0.0;
interface ge-0/3/0.0;
    }
bgp {
traceoptions {
filebgp-log;
flag all;
        }
group group-mvpn {
type internal;
local-address 10.10.10.1;
familyinet-vpn {
unicast;
            }
familyinet-mvpn {
signaling;
            }
neighbor 10.10.10.2;
neighbor 10.10.10.3;
neighbor 10.10.10.4;
        }
    }
ospf {
traffic-engineering {
shortcuts;
        }
area 0.0.0.0 {
interface lo0.0;
interface so-1/0/0.0;
        }
}
policy-options {
policy-statement bgp-to-ospf {
from protocol bgp;
then accept;
    }
}
routing-instances {
vpn-a {
instance-type vrf;
interface ge-0/0/0.0;
interface lo0.1;
route-distinguisher 65010:1;
```

```
provider-tunnel {
rsvp-te {
label-switched-path-template {
p2mp-template-mvpn-1;
      }
vrf-target target:2:1;
vrf-table-label;
protocols {
ospf {
exportbgp-to-ospf;
area 0.0.0.0 {
interface all;
            }
         }
pim {
traceoptions {
filevpn-pim-log;
flag all;
            }
rp {
local {
address 10.10.47.101;
               }
            }
interface lo0.1 {
mode sparse;
version 2;
            }
interface ge-0/0/0.0 {
mode sparse;
version 2;
            }
         }
mvpn {
traceoptions {
filemvpn-log;
flag all;
            }
   }
vpn-b {
instance-type vrf;
interface ge-0/3/0.0;
interface lo0.2;
route-distinguisher 65010:11;
provider-tunnel {
rsvp-te {
label-switched-path-template {
p2mp-template-mvpn-2;
            }
      }
vrf-target target:2:2;
vrf-table-label;
```

```
protocols {
ospf {
exportbgp-to-ospf;
area 0.0.0.0 {
interface all;
                }
            }
pim {
rp {
local {
address 10.10.48.101;
                    }
                }
interface ge-0/3/0.0 {
mode sparse;
version 2;
                }
interface lo0.2 {
mode sparse;
version 2;
                }
            }
mvpn;
}
```

Appendix B

Rosen Scheme Work Verification

Show PIM-SM Hello, Join/Prune, Register and Register-Stop Messages

ottawalab@PE1> show log file pim-log

I Aug 21 19:00:02.519228 PIM ge-0/2/0.0 SENT 10.0.78.5 -> 224.0.0.13 V2 **JoinPrune** to 10.0.78.6 holdtime 210 groups 1 sum 0x87d6 len 34
Aug 21 19:00:02.519335 group 224.1.1.1 joins 1 prunes 0
Aug 21 19:00:02.519379 join list:
Aug 21 19:00:02.519438 source 10.10.10.2 flags sparse

Aug 21 19:00:02.549861 task_timer_reset: reset PIM.master_IF_Xmit

Aug 21 19:00:11.082510 PIM ge-0/2/0.0 RECV 10.0.78.6 -> 224.0.0.13 V2 **JoinPrune** to 10.0.78.5 holdtime 210 groups 1 sum 0x517f len 50
Aug 21 19:00:11.082712 group 224.1.1.1 joins 2 prunes 1
Aug 21 19:00:11.082756 join list:
Aug 21 19:00:11.082821 source 10.10.10.1 flags sparse,rptree,wildcard
Aug 21 19:00:11.082879 source 10.10.10.1 flags sparse
Aug 21 19:00:11.082924 prune list:
Aug 21 19:00:11.082980 source 10.10.10.2 flags sparse,rptree

Aug 21 19:00:11.083943 task_timer_uset: timer PIM.master_RxJoin<Touched> set to offset 3:30 at 19:03:41

Aug 21 19:00:11.908770 PIM ge-0/2/0.0 SENT 10.0.78.5 -> 224.0.0.13 V2 **Hello** hold 105 T-bit LAN prune 500 ms override 2000 mspri 1 genid 419d7ede sum 0x951d len 34

Aug 21 19:00:11.938352 task_timer_reset: reset PIM.master_IF_Xmit

Aug 21 19:00:29.491560 PIM ge-0/2/0.0 RECV 10.0.78.6 -> 224.0.0.13 V2 **Hello** hold 105 T-bit LAN prune 500 ms override 2000 mspri 1 genid 1faed122 sum 0x64c8 len 34

Aug 21 19:00:31.343710 PIM ge-0/2/0.0 RECV 10.10.10.2 -> 10.10.10.1 V2 **Register** Flags: 0x40000000 Border: 0 Null: 1 Source 10.10.10.2 Group 224.1.1.1 sum 0xa9f0 len 28

Aug 21 19:00:31.344502 PIM SENT 10.10.10.1 -> 10.10.10.2 V2 **RegisterStopSource** 10.10.10.2 **Group** 224.1.1.1 sum 0xe6d0 len 18

Aug 21 19:00:41.431785 task_timer_uset: timer PIM.master_IF_Xmit<Touched> set t
o interval 0.030000 jitter 10 at 19:00:39
Aug 21 19:00:41.461023 PIM ge-0/2/0.0 SENT 10.0.78.5 -> 224.0.0.13 V2 **Hello** hold
105 T-bit LAN prune 500 ms override 2000 mspri 1 genid 419d7ede sum 0x951d len
34

Aug 21 19:00:59.481710 PIM ge-0/2/0.0 RECV 10.0.78.6 -> 224.0.0.13 V2 **Hello** hold
105 T-bit LAN prune 500 ms override 2000 mspri 1 genid 1faed122 sum 0x64c8 len
34

Aug 21 19:01:02.488390 task_timer_uset: timer PIM.master_IF_Xmit<Touched> set t
o interval 0.030000 jitter 10 at 19:00:59

Aug 21 19:01:02.518650 PIM ge-0/2/0.0 SENT 10.0.78.5 -> 224.0.0.13 V2 **JoinPrune** to 10.0.78.6 holdtime 210 groups 1 sum 0x87d6 len 34
Aug 21 19:01:02.518762 group 224.1.1.1 joins 1 prunes 0
Aug 21 19:01:02.518807 join list:
Aug 21 19:01:02.518866 source 10.10.10.2 flags sparse

Aug 21 19:01:09.710192 task_timer_uset: timer PIM.master_IF_Xmit<Touched> set t
o interval 0.030000 jitter 10 at 19:01:09

Aug 21 19:01:09.740490 PIM ge-0/2/0.0 SENT 10.0.78.5 -> 224.0.0.13 V2 **Hello** hold 105 T-bit LAN prune 500 ms override 2000 mspri 1 genid 419d7ede sum 0x951d len
34

Aug 21 19:01:09.769116 task_timer_reset: reset PIM.master_IF_Xmit

Aug 21 19:01:11.082982 PIM ge-0/2/0.0 RECV 10.0.78.6 -> 224.0.0.13 V2 **JoinPrune** to 10.0.78.5 holdtime 210 groups 1 sum 0x517f len 50
Aug 21 19:01:11.083202 group 224.1.1.1 joins 2 prunes 1
Aug 21 19:01:11.083247 join list:
Aug 21 19:01:11.083310 source 10.10.10.1 flags sparse,rptree,wildcard
Aug 21 19:01:11.083368 source 10.10.10.1 flags sparse
Aug 21 19:01:11.083414 prune list:
Aug 21 19:01:11.083471 source 10.10.10.2 flags sparse,rptree

Aug 21 19:01:27.282696 PIM ge-0/2/0.0 RECV 10.0.78.6 -> 224.0.0.13 V2 **Hello** hold
105 T-bit LAN prune 500 ms override 2000 mspri 1 genid 1faed122 sum 0x64c8 len
34

Aug 21 19:01:27.283008 task_timer_uset: timer PIM.master_Nbr<Touched> set to of
fset 1:45 at 19:03:12

Aug 21 19:01:31.348340 PIM ge-0/2/0.0 RECV 10.10.10.2 -> 10.10.10.1 V2 **Register** Flags: 0x40000000 Border: 0 Null: 1 Source 10.10.10.2 Group 224.1.1.1 sum 0xa9f0 len 28

Aug 21 19:01:31.349217 PIM SENT 10.10.10.1 -> 10.10.10.2 V2 **RegisterStopSource**

10.10.10.2 **Group** 224.1.1.1 sum 0xe6d0 len 18

Aug 21 19:01:37.873994 PIM ge-0/2/0.0 SENT 10.0.78.5 -> 224.0.0.13 V2 **Hello** hold
105 T-bit LAN prune 500 ms override 2000 mspri 1 genid 419d7ede sum 0x951d len
34

Aug 21 19:01:37.904628 task_timer_reset: reset PIM.master_IF_Xmit

Aug 21 19:01:57.230734 PIM ge-0/2/0.0 RECV 10.0.78.6 -> 224.0.0.13 V2 **Hello** hold
105 T-bit LAN prune 500 ms override 2000 mspri 1 genid 1faed122 sum 0x64c8 len
34

Aug 21 19:02:02.488798 task_timer_uset: timer PIM.master_IF_Xmit<Touched> set t
o interval 0.030000 jitter 10 at 19:01:59

Aug 21 19:02:02.518057 PIM ge-0/2/0.0 SENT 10.0.78.5 -> 224.0.0.13 V2 **JoinPrune**
to 10.0.78.6 holdtime 210 groups 1 sum 0x87d6 len 34
Aug 21 19:02:02.518167 group 224.1.1.1 joins 1 prunes 0
Aug 21 19:02:02.518211 join list:
Aug 21 19:02:02.518271 source 10.10.10.2 flags sparse
Aug 21 19:02:02.545696 task_timer_reset: reset PIM.master_IF_Xmit

Aug 21 19:02:07.274042 task_timer_uset: timer PIM.master_IF_Xmit<Touched> set t
o interval 0.030000 jitter 10 at 19:02:06

Aug 21 19:02:07.304394 PIM ge-0/2/0.0 SENT 10.0.78.5 -> 224.0.0.13 V2 **Hello** hold
105 T-bit LAN prune 500 ms override 2000 mspri 1 genid 419d7ede sum 0x951d len
34

Aug 21 19:02:07.332922 task_timer_reset: reset PIM.master_IF_Xmit

Aug 21 19:02:11.082173 PIM ge-0/2/0.0 RECV 10.0.78.6 -> 224.0.0.13 V2 **JoinPrune**
to 10.0.78.5 holdtime 210 groups 1 sum 0x517f len 50
Aug 21 19:02:11.082438 group 224.1.1.1 joins 2 prunes 1
Aug 21 19:02:11.082590 join list:
Aug 21 19:02:11.082657 source 10.10.10.1 flags sparse,rptree,wildcard
Aug 21 19:02:11.082717 source 10.10.10.1 flags sparse
Aug 21 19:02:11.082763 prune list:
Aug 21 19:02:11.082821 source 10.10.10.2 flags sparse,rptree
Aug 21 19:02:11.083187 task_timer_uset: timer PIM.master_RxJoin<Touched> set to
offset 3:30 at 19:05:4

Appendix C

NG MVPN Work Verification

Originating a Type 1 AD Route

ottawalab@PE1> show route table vpn-a.mvpn.0 detail

vpn-a.mvpn.0: 2 destinations, 2 routes (2 active, 0 holddown, 0 hidden)

1:65010:1:10.10.10.1/240 (1 entry, 1 announced)

*MVPN Preference: 70

Next hop type: Indirect

Next-hop reference count: 3

Protocol next hop: 10.10.10.1

Indirect next hop: 0 -

State: <Active Int Ext>

Age: 1d 0:58:25 Metric2: 1

Task: mvpn global task

Announcement bits (3): 0-PIM.vpn-a 1-mvpn global task 2-BGP RT Background

AS path: I

Receiving a Type 1 AD Route

ottawalab@PE2> show route table l3vpn_50001.mvpn.0 detail
vpn-a.mvpn.0: 2 destinations, 2 routes (2 active, 0 holddown, 0 hidden)

1:65010:1:10.10.10.1/240 (1 entry, 1 announced)

*BGP Preference: 170/-101

PMSI: Flags 0:RSVP-TE:label[0:0:0]:Session_13[10.10.10.1:0:32423:10.10.10.1]

Next hop type: Indirect

Next-hop reference count: 2

Source: 10.10.10.1

Protocol next hop: 10.10.10.1

Indirect next hop: 2 no-forward

State: <Secondary Active Int Ext>

Local AS: 65010 Peer AS: 65010

Age: 22:28:53 Metric2: 2

Task: BGP_65010.10.10.10.1+179

Announcement bits (2): 0-PIM.vpn-a 1-mvpn global task

AS path: I

Communities: target:2:1

Import Accepted

Localpref: 100

Router ID: 10.10.10.1

Primary Routing Table bgp.mvpn.0

Originating a Type 5 Route

ottawalab@PE1> show route table vpn-a.mvpn.0 detail | find 5:

5:65010:1:32:10.10.11.2:32:224.1.1.1/240 (1 entry, 1 announced)

*PIM Preference: 105

Next hop type: Multicast (IPv4)

Next-hop reference count: 3

State: <Active Int>

Age: 2:20:17

Task: PIM.vpn-a

Announcement bits (3): 0-PIM.vpn-a 1-mvpn global task 2-BGP RT Background

AS path: I

```
ottawalab@PE1> show route advertising-protocol bgp 10.10.10.3 table vpn-a detail

* 5:65010:1:32:10.10.11.2:32:224.1.1.1/240 (1 entry, 1 announced)

 BGP group group-mvpn type Internal

    Route Distinguisher: 65010:1

Nexthop: Self

    Flags: Nexthop Change

Localpref: 100

    AS path: [65010] I

    Communities: target:2:1


ottawalab@PE1> show route table bgp.mvpn.0 detail

bgp.mvpn.0: 2 destinations, 2 routes (2 active, 0 holddown, 0 hidden)

5:65010:1:32:10.10.11.2:32:224.1.1.1/240 (1 entry, 0 announced)

       *BGP    Preference: 170/-101

            Next hop type: Indirect

            Next-hop reference count: 4

            Source: 10.10.10.1

            Protocol next hop: 10.10.10.1

            Indirect next hop: 2 no-forward

            State: <Active Int Ext>

            Local AS: 65010 Peer AS: 65010

            Age: 2:37:29    Metric2: 2

            Task: BGP_65010.10.10.10.1+179

            AS path: I

            Communities: target:2:1

            Import Accepted

Localpref: 100

            Router ID: 10.10.10.1

            Secondary Tables: vpn-a.mvpn.0
```

```
ottawalab@PE1> show route table vpn-a.mvpn.0 detail | find 5:
5:65010:1:32:10.10.11.2:32:224.1.1.1/240 (1 entry, 1 announced)
      *BGP    Preference: 170/-101
            Next hop type: Indirect
            Next-hop reference count: 4
            Source: 10.10.10.1
            Protocol next hop: 10.10.10.1
            Indirect next hop: 2 no-forward
            State: <Secondary Active Int Ext>
            Local AS: 65010 Peer AS: 65010
            Age: 2:50:11    Metric2: 2
            Task: BGP_65010.10.10.10.1+179
            Announcement bits (2): 0-PIM.vpn-a 1-mvpn global task
            AS path: I
            Communities: target:2:1
            Import Accepted
Localpref: 100
            Router ID: 10.10.10.1
            Primary Routing Table bgp.mvpn.0
```

Originating a Type 6 and 7 Route

```
ottawalab@PE2> show route table vpn-a.mvpn.0 detaill
6:65010:1:65010:32:10.10.47.101:32:224.1.1.1/240 (1 entry, 1 announced)
      *PIM    Preference: 105
            Next hop type: Multicast (IPv4)
            Next-hop reference count: 2
            State: <Active Int>
            Age: 22:50
            Task: PIM.vpn-a
```

```
        Announcement bits (2): 0-PIM.vpn-a 1-mvpn global task

        AS path: I

        Communities: no-advertise target:10.10.10.1:
7:65010:1:65010:32:10.10.11.2:32:224.1.1.1/240 (1 entry, 1 announced)
    *MVPN   Preference: 70

        Next hop type: Multicast (IPv4)

        Next-hop reference count: 2

        State: <Active Int Ext>

        Age: 22:50      Metric2: 1

        Task: mvpn global task

        Announcement bits (3): 0-PIM.vpn-a 1-mvpn global task 2-BGP RT Background

        AS path: I

        Communities: target:10.10.10.1:5
```

Appendix D

Confidence Interval Calculations

All experimental results that shown in the previous sections are measured by taking the mean of a series of *n* tests, each of long enough time to ensure uncorrelated results. All tests that have been presented in this thesis are identical and independent from each other. The *n* independent results are represented by X_1, X_2, $X_{3, \dots,}$ X_{n-1},X_n.

$$\text{The Mean } \bar{X} = \frac{1}{n}\sum_{i=1}^{n} Xi \qquad \text{(D.1)}$$

However, the mean of independent experimental tests $\bar{X}$ provide us with a single numerical value for the estimate of the expected value E|X| =μ. Therefore, in order to distinguish how good the estimate provided by $\bar{X}$ for the experimental results, it is necessary to compute the Variance $Var(X)$.

$$Var(X) = \frac{1}{n-1}\sum_{i}^{n}(Xi - \bar{X}) \qquad \text{(D.2)}$$

Small $Var(X)$ value indicates that the results are closely clustered around $\bar{X}$, and we can be confident that $\bar{X}$ is close to the E|X|. However, if $Var(X)$ is large, then the results are broadly distributed around $\bar{X}$, and we can't be confident that $\bar{X}$ is close to the E|X|. Alternatively, we can specify the interval of values that is highly likely to contain the true value of the parameter. This is done by specifying some high probability assumingly 1-α and then finds the interval [L(X), U (B)] such that the probability:

$$Pr[\, L(X) \leq \mu \geq L(X)] = 1 - \alpha \qquad \text{(D.3)}$$

The interval contains the true value of the parameters with probability $1-\alpha$. Such an interval is $1-\alpha \times 100\%$ confidence interval.

Using the standard deviation σ and *t* distributed table, the lower and upper limits of the 95% confidence interval can be calculated as follow:

$$\text{Lower limit } L(X) = \bar{X} - \frac{\sigma t\,[\frac{\alpha}{2}, n-1]}{\sqrt{n}} \qquad \text{(B.4)}$$

$$\text{Upper Limit } U(X) = \bar{X} + \frac{\sigma t\,[\frac{\alpha}{2}, n-1]}{\sqrt{n}} \qquad \text{(B.5)}$$

Where:

$$\alpha = 0.05$$

n= number of tests

$\bar{X}$= sample average

$$\sigma = \text{sample standard deviation} = \sqrt{V_x^2} = \sqrt{\frac{1}{n-1}\sum_i^n (\mathrm{X}i - \overline{\mathrm{X}})}$$

The confidence interval means that 95% of the experiments results fall within the interval. Throughout this thesis, the confidence interval is computed based on three independent tests. From the table of *t* distribution, the $t_{[\alpha/2,\sigma]}$ is found to be 2.447. It is observed that more than 95% of the results are within the calculated confidence interval for each experiment. See Tables D.1, D.2, D.3, D.4 and D.5 for confidence interval calculation. The tables show the number of control plane messages for both Draft Rosen and NG MVPN schemes in Chapter 4 when experiments are carried out on the network topology in Figures 4.12, 4.22, and 4.25 with the increasing of some parameters. The number of control massages for all tests are shown along with the calculated Mean $\overline{X}$, the upper and lower values of interval *U(X)- L(X)*. The confidence intervals are also shown on the figures.

No. Of PIM Massages	Experiment tests averages			$\bar{X}$	σ	U(X)	L(X)	Interval
	X_1	X_2	X_3					
T = 1	15	16	15	15.3333	0.57735	15.98665	14.68001	0.653321
T = 5	48	49	49	48.6667	0.57735	49.31999	48.01335	0.653321
T = 10	104	108	105	105.667	2.08167	108.0223	103.3111	2.355584
T = 15	165	164	162	163.667	1.52753	165.3952	161.9381	1.728526
T = 20	221	223	220	221.333	1.52753	223.0619	219.6048	1.728526

Table D.1: Confidence Interval Calculations in Scenario 1

No. Of PIM Massages	Experiment tests averages			$\bar{X}$	σ	U(X)	L(X)	Interval
	X_1	X_2	X_3					
T = 1	28	28	28	28	0	28	28	N/A
T = 5	88	90	88	88.66667	1.1547	88	88	1.306643
T = 10	194	192	190	192	2	194.2632	189.7368	2.263171
T = 15	295	292	291	292.6667	2.08167	295.0223	290.3111	2.355584
T = 20	395	394	391	393.3333	2.355584	395.6889	390.9777	2.355584

Table D.2: Confidence Interval Calculations in Scenario 2

No. of BGP messages	Experiment tests averages			$\bar{X}$	σ	U(X)	L(X)	Interval
	X_1	X_2	X_3					
T=1-20	18	19	18	18.33333	18.33333	18.9867	17.68001	0.653321

Table D.3: Confidence Interval Calculations in Scenario 3

No. of BGP messages	Experiment tests averages			$\bar{X}$	σ	U(X)	L(X)	Interval
	X_1	X_2	X_3					
T=1-20	20	23	21	21.33333	1.52753	23.06186	19.6048	1.728526

Table D.4: Confidence Interval Calculations in Scenario 4

No. of BGP messages	Experiment tests averages			$\bar{X}$	σ	U(X)	L(X)	Interval
	X_1	X_2	X_3					
T=1-20	9	9	11	9.666667	1.154701	10.97331	8.360024	1.306643

Table D.5: Confidence Interval Calculations in Scenario 5