jamk.fi

MPLS Segment Routing in Junos

Teemu Heikkilä

Bachelor's Thesis
May 2021
School of Technology, Communication and Transport
Degree Programme in Information and Communications Technology

Jyväskylän ammattikorkeakoulu
JAMK University of Applied Sciences

jamk.fi

Kuvailulehti

Tekijä(t)	Julkaisun laji	Päivämäärä
Heikkilä Teemu	Opinnäytetyö, AMK	Toukokuu 2021
	Sivumäärä 51	Julkaisun kieli Englanti
		Verkkojulkaisulupa myönnetty: x

Työn nimi
MPLS Segmenttireititys Junoksella

Tutkinto-ohjelma
Tieto- ja viestintätekniikka, Insinööri (AMK)

Työn ohjaaja(t)
Jani Immonen, Karo Saharinen

Toimeksiantaja(t)
JYVSECTEC

Tiivistelmä

MPLS on miltei kaksi vuosikymmentä vanha teknologia. Vaikka se on ollut käytössä jo vuosia, niin verkkopalveluiden eksponentiaalinen kasvu on tekemässä siitä riittämättömän palvelemaan nykyistä infrastruktuuria. Jatkuvasti lisääntyvä verkkoon liitettävien laitteiden lukumäärä ja kasvavat kaistavaatimukset ovat tuoneet ilmi, että runkoverkon arkkitehtuuri on uusiutumisen tarpeessa. Vanha MPLS-teknologia ei ole riittävän skaalautumiskykyinen, puhuttamaan siitä, että sen ylläpitäminen on haastavaa. Potentiaalisena vastauksena näihin ongelmiin on Segmenttireititys. IEFT:n SPRING työpajassa kehittely Segmenttireititys on ehdokkaana korvaamaan vanhanaikaisen MPLS-verkon. Se on molempia, sekä skaalautuva että yksinkertainen. Tämän lisäksi Segmenttireititys on suunniteltu siten että sen tuotantoon tuominen olemassa olevaan infrastruktuuriin on mahdollisimman helppoa. Segmenttireitityksessä IGP on vastuussa leimojen jakamisessa, joten tarve useille eri protokollille poistuu.

Aloittaen kahden Segmenttireitityksen edeltäjän, MPLS:n ja lähdereitityksen lyhyestä historiasta ja toiminnasta, dokumentti esittää Segmenttireitityksen helposti ymmärrettävällä tavalla. Kun eri segmenttityypit ja niiden toimintaperiaate on esitetty, teknologian toiminta todennetaan laboratorioympäristössä. Ympäristö on rakennettu kokonaan Juniper Networksin Junos virtuaali-MX-reitittimillä. Koska Juniperin laitteet ovat tyypillisiä valintoja runkoverkkoihin, niin tämä antaa realistisen kuvan teknologin käytöstä tuotannossa.

Lopullisena tuloksena on sekä teoreettinen esitys, että käytännön toteutus Segmenttireititysarkkitehtuurista. Laboratoriototeutus todentaa myös Segmenttireitityksen hyötyjä todellisessa ympäristössä. Tämä tuo myös varsinaisen käyttäjän näkökulmaa Segmenttireititykseen, pelkkien teknologiatuottajien ja laitevalmistajien sijaan.

Avainsanat (asiasanat)
Segment Routing (SR) Multi Protocol Label Switching (MPLS) Traffic Engineering (TE) Junos Juniper Networks

Muut tiedot (Salassa pidettävät liitteet)

jamk.fi

Description

Author(s)	Type of publication	Date
Heikkilä Teemu	Bachelor's thesis	May 2021
		Language of publication: English
	Number of pages: 51	Permission for web publication: x

Title of publication
MPLS Segment Routing in Junos

Degree programme
Information and Communication Technology, Bachelor of Engineering

Supervisor(s)
Jani Immonen, Karo Saharinen

Assigned by
JYVSECTEC

Abstract

MPLS is almost two decades old technology. While it was sufficient in the past, the exponential growth for Internet related services have found it lacking. Due to rapidly increasing number of devices added to Internet and demand for ever greater bandwidth requirements, the existing network core architecture is simply no longer up to speed. MPLS is both cumbersome to upkeep and troubleshoot and has been found not to scale well enough. The potential answer to these challenges is Segment Routing. Developed by the IETF's SPRING Work Group, the Segment Routing is purposed to be successor for traditional MPLS networks. Being both scalable and simple, Segment Routing is easy to implement to existing networks. Since with it the label distribution is done via Interior Gateway Protocols it also eliminates need for multiple protocols all the while having seamless deployment.

Beginning from the brief history and explanation of the two technologies to which Segment Routing builds upon, the goal is to give introduction to Segment Routing in a way that is easy to understand and deploy. After explaining the concepts and different types of segments and the overall benefits of Segment Routing, a sample laboratory environment is configured. In lab, the concepts of Segment Routing are tested in practice and their configurations and results proven and documented. The lab consist solely Juniper Networks MX virtual routers, so configurations are seen from point of view of one of the industry leaders.

Result is a both theoretical explanation of Segment Routing architecture and practical immolation, in which the actual configurations and benefits of this technology are tested. Practical implantation also gives actual users point of view on Segment Routing, in contrast of having only the technology developers and device manufacturers input.

Keywords/tags (subjects)
Segment Routing (SR) Multi Protocol Label Switching (MPLS) Traffic Engineering (TE) Junos Juniper Networks

Miscellaneous (Confidential information)

Table of Contents

Figures

Tables

1 Technologies and abbreviations

BGP	Border Gateway Protocol
ECMP	Equal-Cost Multi-Path
FIB	Forwarding Information Base
IETF	Internet Engineering Task Force
IoT	Internet of Things
IS-IS	Intermediate System to Intermediate System
LAN	Local Area Network
LDP	Label Distribution Protocol
LSP	Label Switched Path
OSPF	Open Shortest Path First
PHP	Penultimate hop popping
MPLS	Multi Protocol Label Switching
RFC	Request for Comments
RSVP-TE	Resource Reservation Protocol – Traffic Engineering
SID	Segment Identifier
SR	Segment Routing
TI-LFA	Topology-Independent Loop-Free Alternate
VPN	Virtual Private Network
WAN	Wide Area Network

2 Preface

Use of Internet related services have kept evolving rapidly. During past decades number of Internet users have risen from millions to billions, and number of devices connected to the internet is growing exponentially (Litmanen, I. 2017). Ease of access and number of resources being allocated towards cloud mean that users and organizations using the Internet keep growing, while new IoT-devices need to be integrated to network. Internet service providers are facing the challenge of ever growing demand for service. Massive bandwidth requirements and number of devices need to be implemented to infrastructure are creating a difficult scenario for existing networks.

Yet at the same time however, the network core architectures have remained largely unchanged (Litmanen, I. 2017). Service Provider network core architectures are still relaying on technologies which are more than decade old, which would be considered to be ancient from IT-point of view. MPLS used in network core forwarding is nearly 20 year old, and while functioning, it too has started to show flaws. MPLS both cumbersome to troubleshoot and due to its many components and related protocols it is also poorly scalable.

Segment Routing is one the purposed solitons to upgrade existing core network architecture and provide sustainable network environment for ever growing demand. By modernizing the pre-existing source routing paradigm, Segment routing is both scalable and simple, as well as seamless to install in existing network implementations.

The purpose of this bachelor's thesis is to explain the Segment Routing architecture in a manner that is easy to understand and digest. Document provides explanation to concepts of Segment routing and their usage, as well as a lab implementation where some of the use cases of segment routing are shown in practice. Lab is implemented over Juniper Junos virtual routers, providing a view on Segment Routing from one of the industry leaders.

2.1 Assigner

This Bachelor's thesis was made for JYVSECTEC and in collaboration with their staff. The purpose was to crate and develop material for organization's upcoming new courses. These future courses consist several different fields and technologies of networking such as local area networks, access networks, network backbone, data center networks and automations of networks.

2.1.1 JyvSecTec

JYVSECTEC – Jyväskylä Security Technology is the leading independent cyber security research, development, and training center in Finland (JYVSECTEC. 2021). They operate as a part of JAMK University of Applied Science's Institute of Information Technology, and their main office is located at the university's facility. This collaboration grants them multidisciplinary experts from different fields of studies. JYVSECTEC specializes in cyber security, incident response, emerging new technologies and general IT technologies related but no limited to networks and servers. They were founded in 2011 and as a small company they currently employ around 11-50 employees, some of whom also teach or work under JAMK University of Applied Science's also.

JYVSECTEC offers both training and exercise for both students of JAMK University of Applied Science's and for private companies and institutions. These training and exercise sessions include things like cyber exercises that provide opportunities for organizations to demonstrate their capabilities and effectiveness how they people, processes, and technology to protect their information and assets. There are different types of these exercises such as Live, DFIR and Capture the Flag.

Besides training JYVSECTEC provides consulting services to their customers in different fields of information and cyber security. Consulting given can be related to matters such as reliability of their used software, devices, or systems. Consulting service can also be given in matter of development of used information security practices

and methods used by the client. All these are customized to meet requirements of each individual customer.

Third service carried out by JYVSECTEC is Research and Development, or R&D In short. JYVSECTEC performs different types of cyber security related research projects. These can be either separately funded projects or joint research that co-operated with given organization. Their partners are both research institutes and companies. With later the goal of research is often to improve or develop organization capabilities, services, or products.

For above mentioned exercises and their other research and training activities, JYCSECTEC deploys their unique Realistic Global Cyber Environment (RGCE). This is large and feature rich live cyber range that mirrors the actual real-world Internet to certain extend. To achieve this the RGCE combines both virtualization techniques, and physical devices and business specific systems. All these utilized in modernized ways. RGCE environment can also be tailored and adjusted to create customized practice environment for requesting organization's specific training, exercise, or research and development needs. The use of RGCE has many benefits. The main one being that training and exercise is no longer restricted to small laboratories or virtualized environments, but now these can be carried out in environment that represents real, live networks.

3 Segment Routing

3.1 Basic Concepts

Segment Routing is a new upcoming traffic-engineering technology which is under development by the IETF's SPRING Work Group. Segment Routing, also refenced as a SPRING, is basically a purposed substituent to a deprecating MPLS-technology. Segment Routing defines two forwarding plane encapsulation methods. These are Multi Protocol Label Switching (MPLS) and IPv6 with a Segment Routing Extension Header

(Farrel, A. Bonica, R. 2017). This thesis considers only the MPLS method. MPLS Segment routing, as one may already have guessed, re-uses concepts heavily from other two existing technologies, from the forementioned MPLS and Source Routing. Both of those are major largely used in today's networks. Therefore, it is quite helpful to have some knowledge of bot MPLS and Source Routing concepts

3.2 MPLS

Multi Protocol Label Switching (MPLS) is a widely used technique that uses labels to forward packets. The main benefit of these labels are that nodes in MPLS network do not need to make look ups to their routing tables during forwarding decisions. Instead, packet header is only analyzed one time when it enters MPLS network at the ingress router. During early times when routers were not exactly stuffed on resources, standard IP lookup was somewhat taxing process for them. Several steps are taken by the router when packet is received. Device must open the frame, compare its destination to routing table to find best prefix and perform ARP to determine which layer-2 address use as a source (David, P. 2020). MPLS was created to vastly speed up this process and save resources of a router. Multi Protocol Label Switching does not have its own datagram format. Instead, it operates between layer-2 and layer-3. MPLS labels are inserted between those two. See Figure 1 for reference.

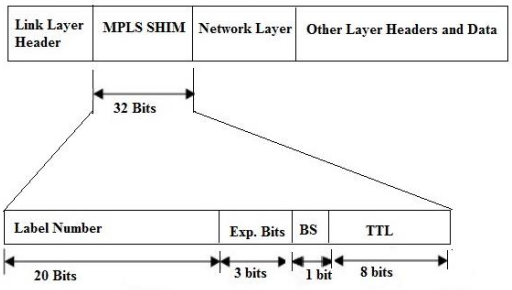


Figure 1. MPLS frame (ExamGuides. 2021)

MPLS relies mainly on Label Distribution Protocol (LDP). LDP is used to advertise label bindings, but alternative protocols are also available. LDP is responsible for creating

and advertising label mappings for network entries in a router's RIB, Routing Information database. LDP both crates and advertises label binding for RIB entries. These can be learned from all routing source protocols except for Border Gateway Protocol BGP, which has its own mechanism to advertise label mappings. It is important to know that labels are only locally significant. This meaning that each router has their own label binding for given prefix that are received from neighboring nodes (David, P. 2020). MPLS works by relaying three main process during label handling. These processes are shown in figure 1. Label switching Router, LSR looks at the topmost labels and performs operation suitable for packet.

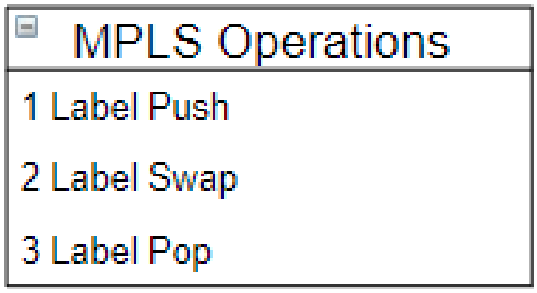


Figure 2. MPLS Operations

Label Push happens when IP packet arrives at LSR and it pushed a label on top of packet. LRS can also push label on top of existing label in case there is one already on top of stack (David, P. 2020).

Label Swap is performed when LSR forwards the packet to its next-hop node. Since labels are only locally significant, forwarding means removing incoming label and replacing it with next-hop label (David, P. 2020).

Label Pop means removing the label from packet or removing the label that is located at the top of a stack if packet has multiple ones (David, P. 2020).

MPLS is still very widely used in modern networks and heavily adapted by Service Providers. However, it is quite complicated to manage because it relies on other routing protocols for transferring control plane data. MPLS troubleshooting can also be difficult and cumbersome since MPLS labels are locally significant. This makes it harder to track down any problems. Traffic in MPLS network can be easily disrupted by failures in network because network convergence during re-signaling process is relatively slow.

3.3 Source Routing

Originally, the basic routing mode of IPV4 packets was based on destination routing paradigm. Only the destination address of a packet is used during route selection. There are no other influencers in path selection and properties of a sender do not matter (David, P. 2018). This is still a very common way of routing today, and a classic example of destination based routing is the route along the shortest path to destination. Figure 3 shows stander flow of packet when destination routing is applied.

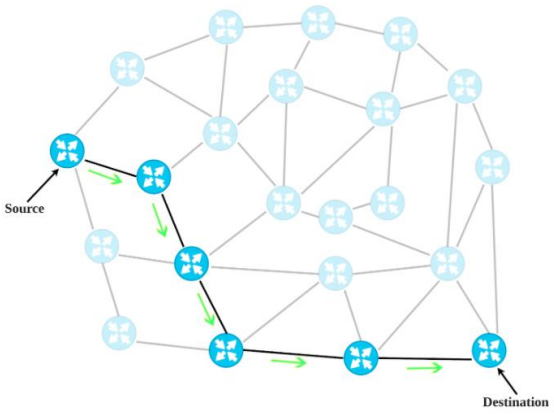


Figure 3. Destination Routing (David P. 2018)

IPv4 specification hover also came with two routing modes. These were Loose Source and Record Route (LSRR), and Strict Source and Record Route (SRRR). They would had allowed the source host more control about the path packet would take. Both would be specified in the options field (David, P. 2018). When looking IPv4 header in

figure 4, this is found after the destination address-field and comes last before the actual data.

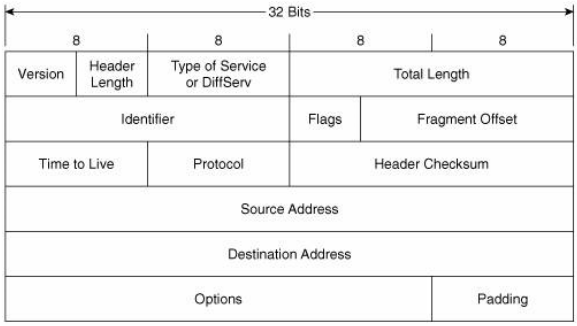


Figure 4. IPv4 Header (Kaur, R. 2009)

Use for these two was to allow sender to specify the path the packet would take. What hops packet need to traverse on its way to destination. This meant that source or sender could dictate that route selection, partially or totally (David, P. 2018). Both the two modes, LSRR and SSRR can be described in summary:

- Loose-mode specifies that packet should pass through the listed hops
- Strict-mode specifies the exact path that packet takes on a hop-by-hop basis

In Source Routing, sending node can dictate an arbitrary path that packet would take. This is entirely independent from typical short path selection of traditional destination routing. It is important to know that nodes on the path do not store state about the path of packets explicit paths. The explicit path is stored in the packet itself (David, P.). Figure 4 show source routing path on red, where packet takes alternative route instead of the shortest path.

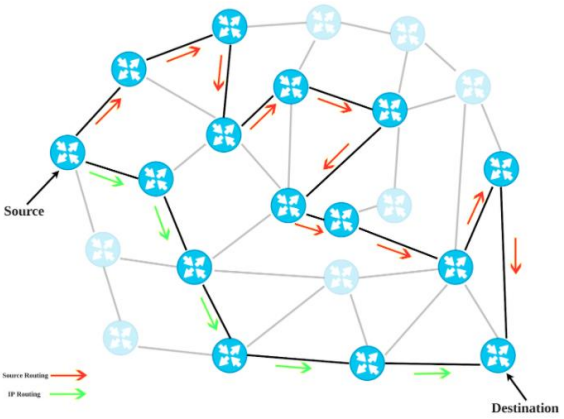


Figure 5. Source Routing (David P. 2018)

Source Routing was meant to provide a form of host-based Traffic Engineering. It would also have had helped extend the troubleshooting toolkit. Network administrators could discover and explore the network and track and pinpoint potential failure on a specific route towards the destination. But originally the Source Routing was proved to be a significant security concern. It opened too many possibilities for malicious individuals. For example, by using Source Routing the packets could have been directed to a private IPv4 networks which was not normally advertised by routing protocols. Attacker would only have to know which nodes to traverse until reaching the one that had the private network directly connected. Internet Engineering Task Force (IETF) published a document that recommend disabling the Source Routing capabilities of the devices. Modern devices are not source-routing-enabled out of the box and for the time being Source Routing became deprecated (David, P. 2018).

3.4 Segment Routing

From MPLS forwarding perspective, Segment routing builds top of the existing MPLS forwarding paradigm. It does not change how labeled packets are forwarded. Considering existing MPLS control-plane however, there two changes to well-known MPLS control-plane policies. First, for certain segment types, labels may have identical values on all routers inside the SR-domain. In MPLS the labels have local significance,

meaning they are unique for each node. Secondly, LDP is no longer needed. Instead, the label binding to each IGP segment are advertised by either OSPF or IS-IS.

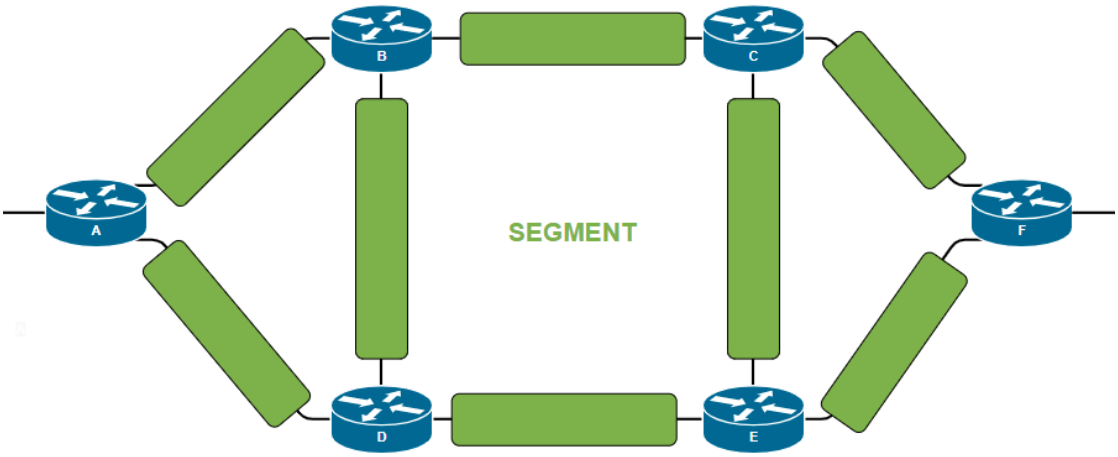


Figure 6. Segments

Figure 6 show basic fundamental idea of segments in network, each green block between routers is a segment between two or more nodes. The overall topology view of Segment Routing domain with different SIDs are show further below in figure 7. The most basic and fundamental concepts of Segment Routing are listed below.

Segment Routing Domain: A collection of network nodes that together participate in Segment Routing protocols. Node within this Segment Routing domain can execute either ingress, transit, or egress procedures. (Juniper Networks. 2021)

Segment Path: An ordered list of segments. This connects the SR ingress node to SR egress node. Most often this is the least-cost path from ingress to egress. However, it does not need to be. (Juniper Networks. 2021)

Segment Routing Segment: A routing forwarding instructions that have data packet to traverse a section of network topology. Segment Routing itself defines several different segment types, the two most common and used being adjacency and prefix

segments. An adjacency segment is simply a single-hop tunnel between two nodes that is strictly forwarded. In other words, it causes a packet to traverse a specified link associated with an interior gateway protocol (IGP) adjacency between two nodes, irrespective of the link cost. A prefix segment is a multihop tunnel that uses equal cost multihop-aware shortest path links to reach a prefix. (Juniper Networks. 2021)

4 Segment Routing Architecture

Architecture of Segment Routing consist of both SR data-plane and SR control-plane. These are the framework upon which Segment Routing is implemented upon.

Segment Routing data plane defines to encoded sequence of segments which are applied to packet. IT also defines forwarding semantics, meaning how each node would process the packet based on segment (Filsfils C, Nainar N, Pignataro C, Cardona J, Francois P).

Segment Routing control-plane on the other hand defines how segment identifiers (SID) are distributed and advertised among network nodes. It defines also how nodes would be instructed to apply given sequence of segments on a flow (Filsfils C, Nainar N, Pignataro C, Cardona J, Francois P).

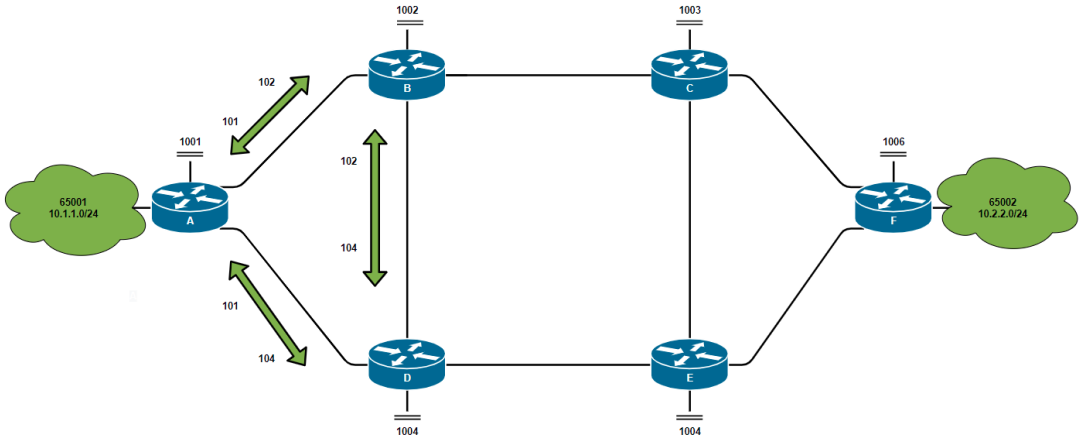


Figure 7. Segment Routing domain

4.1 Data Plane

4.1.1 Segments

In Segment Routing, there are two major segment classing which are defined. These are Global segment and Local Segments. Global segment is an ID value which has significance inside the entire Segment Routing domain. Each and every node inside said domain knows about this value. Therefore, each node also assigns same instruction in its label-FIB. Local Segments is an ID which holds only local value as its name suggest. Only node that advertises this SID will and can execute action associated to this SID. These values are locally configured and not allocated from SRGB.

4.1.2 Segment Routing Global Block

Segment Routing Global Block (SRGB) is range of labels that are reserved for Segment Routing in the label switching database. Each node located inside SR-domain is assigned a value from SRGB as a segment identifier (SID). Values have global significance (Nurminen, P. 2017). In other words, SRGB is the range of label values used in Segment Routing. It is strongly recommended to use same SRBG on all Segment Routing capable nodes within SR-domain. SRGB is a vendor specific range, meaning

that different vendors have different ranges. But since SRGB can be configured manually, it can have value range suited for network need or requirements.

4.1.3 Segment Identifiers

Each Node in Segment Routing domain has their own Segment Identifier (SID). Like their name suggest, Segment Identifiers are used identify different parts of SR network. Segment identifiers can be associated in to two major classes, which are Prefix-SID and Adjacency-SID. Prefix-SID's are global identities and are located in the FIB-table of each node in SR-domain. Adjacency-SID's on the other hand are local and therefore they are added to locally to only those nodes which advertise it (Nurminen, P. 2017). Interior gateway protocol (IGP) are used to distribute both types of segment prefixes. IGPs used and supported are either IS-IS or OSPF.

4.1.4 Prefix-SID

Prefix-SID is a global segment (unless explicitly advertised otherwise) and therefore it is unique within Segment Routing domain. Prefix-SID is associated with IP prefix to which it works as identifier. Its context includes the prefix, topology, and algorithm. It is possible to allocate multiple SIDs to same prefix, however the forementioned tuple (prefix, topology, and algorithm) must be unique. Prefix-Sid is configured manually from Segment Routing Global Block (SRGB) range of labels. Labels are then distributed by IGP, IS-IS or OSPF. Prefix Segment is very similar to a loose source routing hop (RFC8402). Figure 8 shows idea behind prefix-SID. Networks 10.1.1.0/24 and 10.2.2.0/24 have SID 65001 and 65002 assigned to them.

Prefix Segment includes path computation and uses it to steer packets along the shortest path to their destination. If there are multiple equally cost paths, then Prefix Segment also performs Equal Cost Multi Path (ECPM) -load balancing to steer traffic. Prefix-SID can be used to identify both single nodes and groups of nodes (Nurminen, P. 2017).

Prefix-SID has two subtypes associated to it. **Node-SID** and **Anycast-SID**

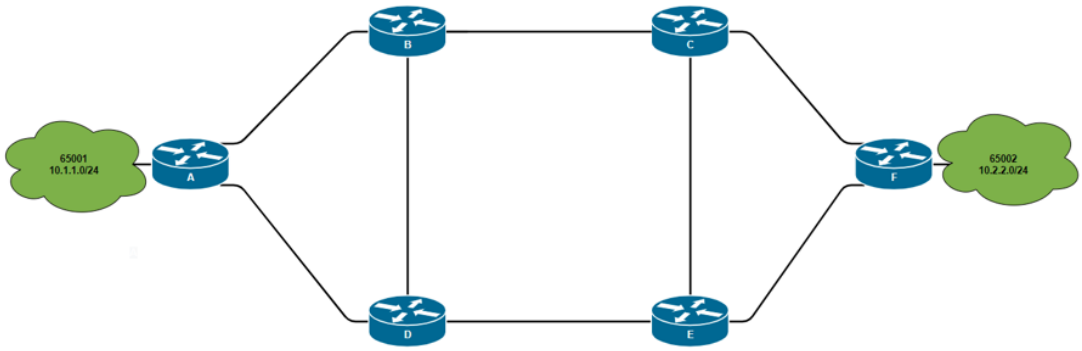


Figure 8. Prefix-SID

4.1.4.1 Node-SID

Node-SID is the first subtype of Prefix SID. It is used to reference a specific node. It can only be associated with IP-address of which is used to identify that specific node. This IP-address could be the loopback address of a router for example, something which is commonly used for these types of purposes. Node-SID that is associated to nodes must be unique within routing domain, and unique among routers located in there. It must not be associated with another routers in SR-domain (Nurminen, P. 2017). Node-SIDs is represented in figure 9, where device-A for example has Node-SID of 1001.

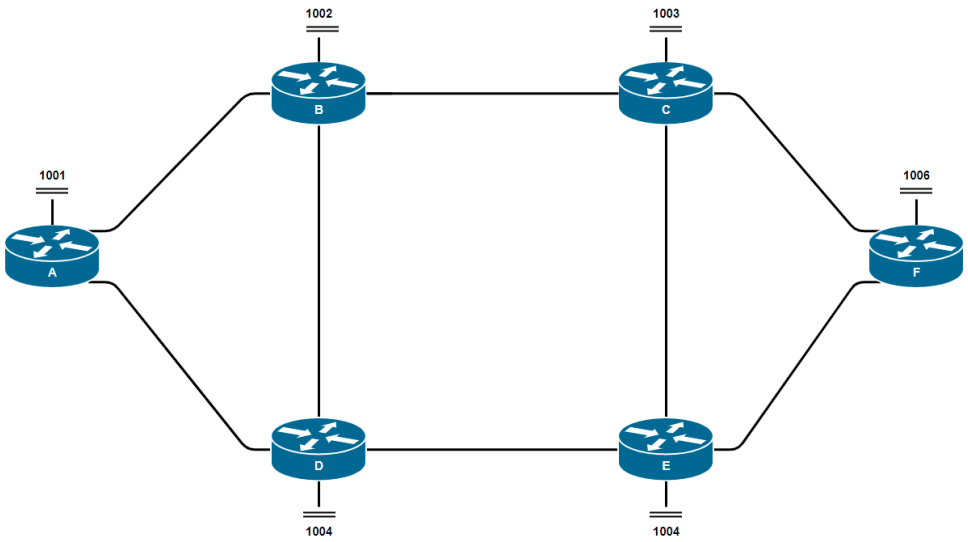


Figure 9. Node-SID

4.1.4.2 Anycast-SID

Second subtype of Prefix-SID is the Anycast-SID. Opposite to previous Node-SID, the Anycast-SID must not be used to refer a specific node. Anycast-SID is used to identify a group of nodes to whom a Prefix is associated with. Anycast group includes two or more nodes. Within an anycast group, all routers in an SR domain must advertise the same prefix with the same SID value. Anycast segment enforces EMCP to load balance the traffic (Nurminen, P. 2017).

4.1.5 Adjacency-SID

Adjacency-SID represent a specific adjacency. Adjacencies are formed by local node and remote node at the other end of adjacency . This adjacency can be an egress interface of a node towards a neighboring router. Adjacency-SID labels are again distributed by IGP such as IS-IS or OSPF. Adjacency segments are local segment. Because of this Adjacency-SIDs are locally unique relative to a specific router. This means that only that specific router adds adjacency identifier to its FIB-table (Nurminen, P. 2017). Adjacency-SIDs are used when traffic is steered to specific link, even though it may not be the shortest path to destination prefix. Adjacency Segment is very similar to a strict source routing hop. Adjacency-SID is shown in figure 10. Device-A has two adjacencies, 102 to B and 104 to D.

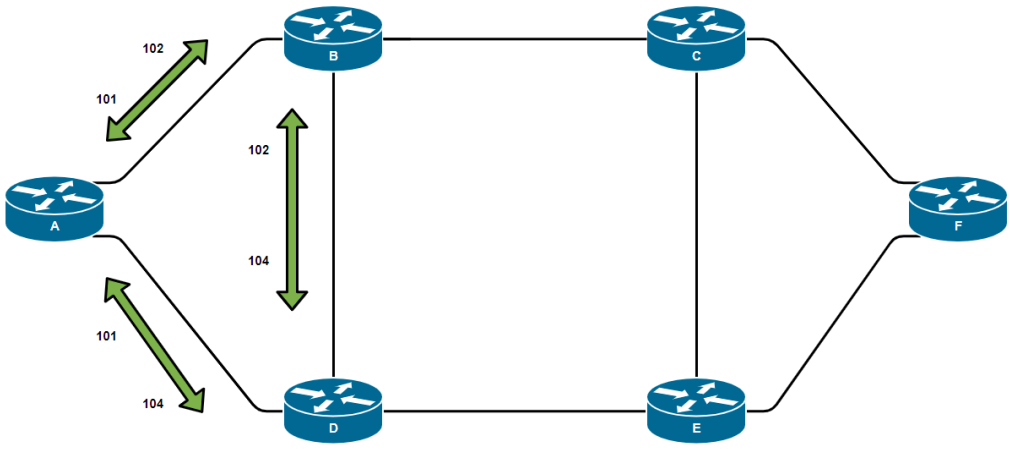


Figure 10. Adjacency-SID

4.1.6 BGP Prefix-SID

BGP Prefix-SID represents the shortest path to a specific BGP prefix. It is similar to earlier Prefix-SID (IGP) in that it too holds global significance in Segment Routing domain. However, BPG prefix-SID, as the name suggest is advertised by BGP-protocol, while normal prefix-SID is advertised by IGP such as IS-IS or OSPF. BGP Prefix Segment is ECMP aware, meaning that it can load balance traffic should there be multiple paths for traffic to reach its destination. BGP-SID labels are assigned as local labels by a network node to its neighbors. That local labels is then advertised as adjacency-SID to neighbor during standard BGP-link state updates (RFC8669. 2019). In figure 11, the prefix from WAN is reachable from domain via Prefix-SID of 65100.

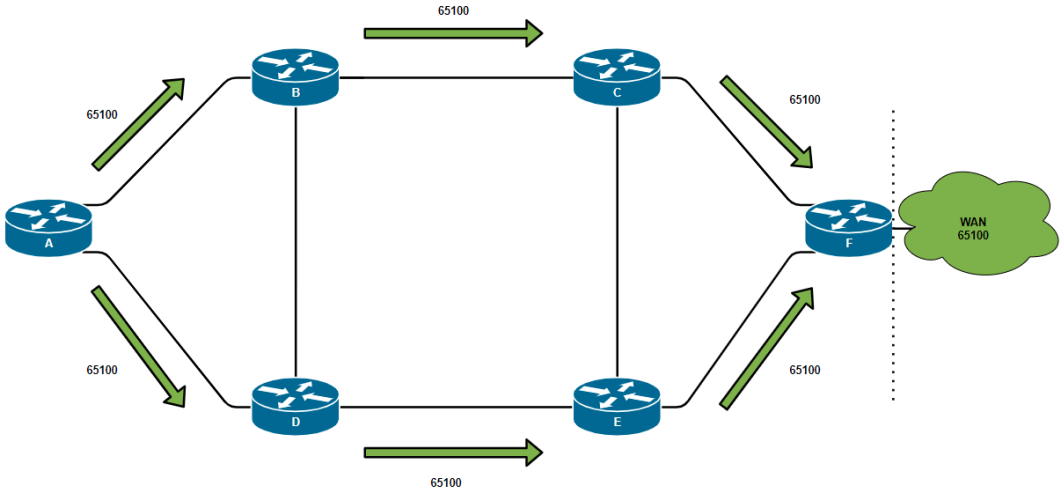


Figure 11. BGP Prefix-SID

4.1.7 Operations

Network nodes that are enabled with Segment Routing can support following operations: Push, Next, Continue. Due to Segment Routing similarities with Multi Protocol Label Switching, these operations can be derived from MPLS. Comparison between MPLS and SR operations are shown in figure 12.

MPLS Operations	SR Operations
1 Label Push	1 Push
2 Label Swap	2 Next
3 Label Pop	3 Continue

Figure 12. MPLS Operations compared to Segment Routing

Push operation is insertion the of segment at top of segment label stack. This can be directly compared to MPLS-Push operation, where receiving router pushes the new label on top of label stack. Segment is added top of SR-header and is then set as active segment (RFC8402. 2018).

Next is the operation after active segment is completed. This consist of inspecting of a next segment, which then becomes active. This is comparable to MPLS-Pop operation, where topmost label of a stack is removed. Next segment is marked as an active segment (RFC8402. 2018).

Continue operation means that active segment is not yet completed, and it remains as active. Node performs the forwarding action based on current active segment (RFC8402. 2018).

4.2 Control-Plane

4.2.1 Forwarding

In segment routing, each segment ends in a segment endpoint. Instead of putting single label on a packet, the SR ingress node puts multiple labels. Each of the labels represent a segment in the network. Figure 13 shows a simple of SR-domain, where each of the routers are SR-enabled. Router-A forward's packet to router-F. Ingress performs the PUSH-operation inserting the labels on arriving packet before forwarding it further to SR-domain Rest of the nodes along the path follow with NEXT-operation, removing the topmost label after each segment (Juniper Networks. 2020).

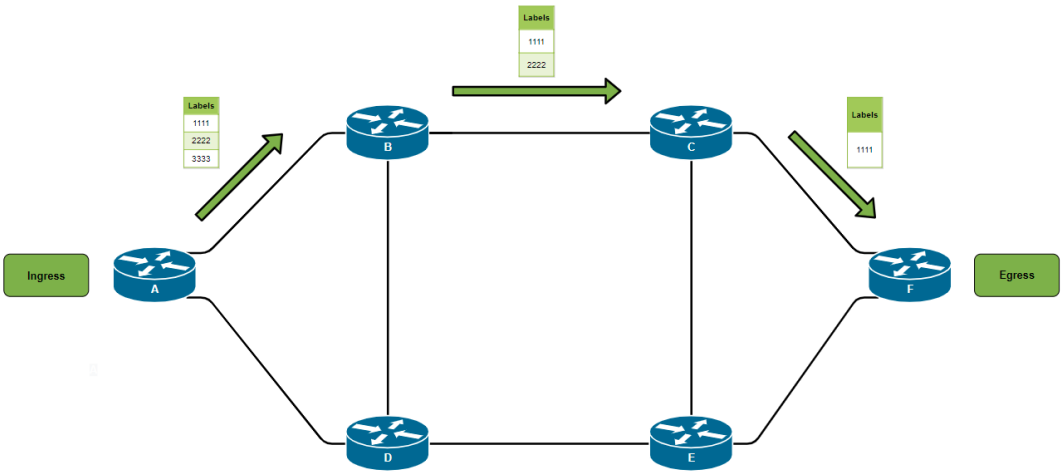


Figure 13. Segment Routing Overview

Router-A, the SR ingress node sends the packet to the first segment egress node, which is the router-B. Router-B looks at the outermost label and processes it. After which, this outermost label is removed. After this the SR node will forward the packet to the second egress node, Router-C. Second egress node takes a look at the current most outermost label and again processes it. Outermost label is then removed, and packet is sent to a third egress SR node. This process is repeated on every single egress node on a packets path until it finally arrives to its destination at the egress node. At final egress node packet has no labels left whatsoever.

4.2.2 Segment Advertisement

In Segment Routing, the control-plane defines how the SIDs are advertised and communicated among the devices in the network. However, they need to be advertised via routing protocol. Currently Segment Routing Supports couple of most common protocols. These are IS-IS and OSPF for IGP and BGP. Node-SID and Adjacency-SID are advertised via link-state IGP protocol. BPG prefix-SID are advertised by BPG protocol. Therefore, segments can be classified to both IGP -and BPG-segments (Filsfils C, Nainar N, Pignataro C, Cardona J, Francois P).

4.2.2.1 IGP

Interior Gateway Protocols are able the advertise Segments when protocols are included with extensions. Extensions are included amongst the standard IGP link-state

updates. Extensions are available for IS-IS and OSPF IGPs, them being currently two Interior Gateways protocols which support Segment Routing.

IS-IS includes collection of TLVs (Type/Length/Value) which are used in its standard link state advertisements (LSA). TVLs can be further be included with sub-TLV where required informed is encoded. With the use of these sub-TLVs, adding new extensions is relatively simple. Therefore, adding new extensions can be done either by defining a new TLV or extending already defined TLV by adding sub-TLVs to it. Segment Routing defines both Prefix-SID and Adjacency-SID extensions for IS-IS sub-TLV. By using these extensions IS-IS and advertise routers own SR-relate information to other nodes (RFC8667. 2019).

OSPF also has its own extensions to enable Segment Routing SID advertisements with other SR-capable nodes. To advertise its Segment Routing capabilities , OSPF add new Opaque LSA. This contains the SR-algorithm TLV and SRGB-TLV. SR-Algorithm TLV is a top-level TLV. Sub-TLVs are then added for both Prefix-SID and Adjacency-SID. Much like IS-IS the existing OSPF architecture can be extended to provide Segment Routing capabilities (RFC8665. 2019).

4.2.2.1.1 BGP

BGP-Prefix Segments are BPG segments and attached to BGP Prefixes. BGP Prefix-SID are advertised by optional and transitive BGP Prefix-SID attribute. This include in the standard BGP-update signal. BGP Prefix-SID is attached to normal IP-prefix and is encoded as a set of TLV (Type/Length/Value) elements. Following TLVs are defined, and they are only used when Segment Routing is applied to the MPLS data-plane:

- Label-Index TLV
- Originator SRGB TLV

The Label-Index TLV is not optional, and it has to be present in the Prefix-SID attribute attached to Labeled IPv4/IPv6 unicast prefix. It contains BGP Prefix-SIDs index value (RFC8669). Label-Index TLV is represented in the figure 14.

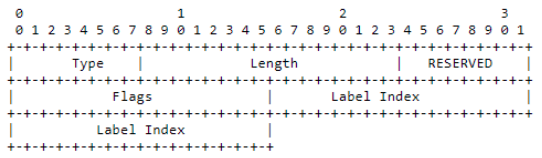


Figure 14. Label-Index TLV (RFC8669. 2019)

Originator SRGB TLV is optional. Its format is displayed in the figure 15. The Originator SRGB contains SRGB of the node which is originating the prefix. It describes the SRGB of that node where BGP Prefix Segment end. It may only appear on those refix-SID attributes which are attached o prefixes of SAFI 4 (labeled unicast) (RFC8669. 2019**)**.

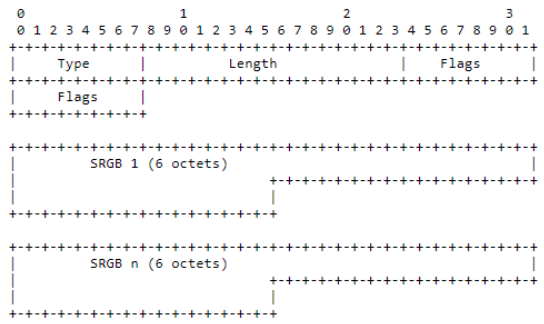


Figure 15. Originator SRGB TLV (RFC8669. 2019)

4.3 Benefits of Segment Routing

The primary benefits of Segment Routing are its abilities to both simplify the network and reduce the resource utilization. This makes it overall easier to manage and operate network running Segment Routing. Segment routing is essentially proposed as a replacement for LDP or RSVP-TE, the two protocols that handle label distribution in MPLS. With Segment Routing, the most immediate benefit is that both of the forementioned protocols can be eliminated from network. Instead, the label distribution is now handled by IGP (currently supported in ISIS and OSPF). Additional protocols be removed from MPLS, and everything can be run only with ISIS or OSFP (Gregory, T.

2016). This leads to simplified infrastructure that is both more manageable to implement and troubleshoot when necessary.

Some of the most typical and common reasons what makes Segment Routing beneficial are listed below in the next chapters.

4.3.1 Scalability and Simplicity

Due to its simplified behavior, the Segment Routing can offer a solution that scales way beyond current solutions used in traffic engineering. Segment Routing has little to no recourse impact for the network core nodes. It also has neither no per tunnel control nor data plane state in the core of the network. All this makes possible to create massive full-mesh end to end MPLS networks. With Segment Routing traffic engineered path only need to be programmed on the edge of domain. Per-flow state needs to be maintained only at ingress nodes which sit at domain edge. Path signaling is not required at all (Arista Networks. 2016).

Since in Segment Routing, the labels are distributed in the routing protocol itself (IGP or BGP) and that the core has no requirements to maintain per flow state makes segment routing simpler to operate the network at scale (Arista Networks. 2016).

4.3.2 Seamless Deployment

Because Segment Routing can be considered as an upgraded version of traditional MPLS network, implementing it to existing network is not only simple but almost seamless. Segment routing itself runs natively on MPLS at best only simple software upgrade will enable existing hardware infrastructure to run it. It can also run together with existing LDP network which makes migration to Segment Routing even less painless (Arista Networks. 2016).

4.3.3 Fewer Protocols

Transit nodes no longer need to maintain per path information. When transit nodes on a path of packet do not need to maintain this per path information, there is no

need for special protocols to distribute this information. Fewer protocols also make troubleshooting the network more simple and easier. The network can simply run one IGP such as IS-IS or OSPF. Directed LDP sessions between network core routers are also avoided with Segment Routing (Arista Networks. 2016).

While above is very much true, the Segment Routing can be made to interwork with existing and well know Label Distribution Protocols. For example, LDP and RSVP or RSVP-TE which have typically been used for this purpose. With LDP, only simple mapping of Segment Routing to LDP labels is needed to be configured. And with RSVP a binding segment identifier is configured. This makes overall inter-working Segment Routing domains and non-SR domains both simple and scalable. Other benefits of it is that it allows path migration for Internet Operators and Service Providers running legacy systems.

4.3.4 Faster Convergence

Network can converge much more quickly after change in network topology. If link goes down due to a breaking, the restoration process is faster. This meaning that alternative route is selected. Segment Routing provides network with resiliency through headend restoration and topology-independent loop-free alternate (TI-LFA) technology, which helps with path reliability during network outages (Juniper Networks. 2020).

In tradition traffic engineering, a single link failure may affect hundreds of even thousand paths. All these paths would need to re-signal about the change in topology at same time, causing large amount of unwanted network traffic where many, many nodes sending protocol and path changes. This will lead to a control plane congestion for example which itself is unwanted condition of a device (Abdelrahman, M. 2018).

In segment routing, a node directly upstream of a broken link will signal that a network link has failed. This information is relayed to path ingress nodes only. Ingress nodes can then send packet to a network with updated set of labels that makes packet traverse different path. One that circumvent the broken link. No longer is

there a need to re-signal every path at the same time. Just the need to inform ingress nodes about changes in the topology (Abdelrahman, M. 2018).

With overall faster network convergence and recovery, effect of microloops are also mitigated. Microloops are typically caused by unexpected and sudden changes in topology, simple because every node is not able to convergence simultaneously. Typical such events are link up or link down events. Large scale networks such Service Provider environment have wide variety different types of routers that are used. Different types of routers have different processing speed and mentioned link up/down events may cause short-term packets loss in traffic. With Segment Routing this is avoided as in the vent of link up/down event the traffic is forced to use the backup path using a temporary Segment list until other routers have converged correctly (Abdelrahman, M. 2018).

4.3.5 Centralized Traffic Engineering

In previous notes, it is assumed that the ingress node is the one doing the path selection process for the packets. While in some cases ingress node is sufficient enough to make the routing decision, which segments packet need to traverse to go from ingress to egress, it is not always the case. Different types of controllers or orchestration platforms can directly interact with segment routing and adjust traffic flow on the fly (Juniper Networks. 2020).

Segment routing may use a centralized controller that is the one doing the path computation. This controller can provide benefits beyond of path computing where issues such as simply link bandwidth reservation are involved. Controller can maintain a global view of entire network. It can receive information from every single individual network element about link congestion for example. With controller events like congestion could cause application to optimize the placement of segment routing traffic and will re-route the traffic to path with available bandwidth. Because of controller has a global view of the network, it may be able to make far better path computation decision than a single network element could make (Juniper Networks. 2020).

5 Lab Implementation

5.1 Research approach

This thesis takes more practical approach to research objectives. Main purpose is to implement and verify Segment Routing technology in a small scale laboratory environment. Technologies and protocols are pre-existing and developed by IETF SPRING group. In this lab they are applied to live environment using specific vendor methods and practices, that vendor being Juniper Networks Junos. Basis of configurations and their theoretical use-cases are retrieved from vendors' own manuals and guides. These are then applied to laboratory environment. Testing and verifications are done solely with devices themselves, using their command-line interface. This may be connectivity test after committing SR configurations. Verifications may also be done by seeing how SR-related configurations modify other protocols on device, such as routing tables. Devices in lab domain have mostly similar configurations, expect for addressing and Segment Identifiers. Segment Routing related special features are implemented to vMX1 unless they are required elsewhere also.

5.2 Topology explained

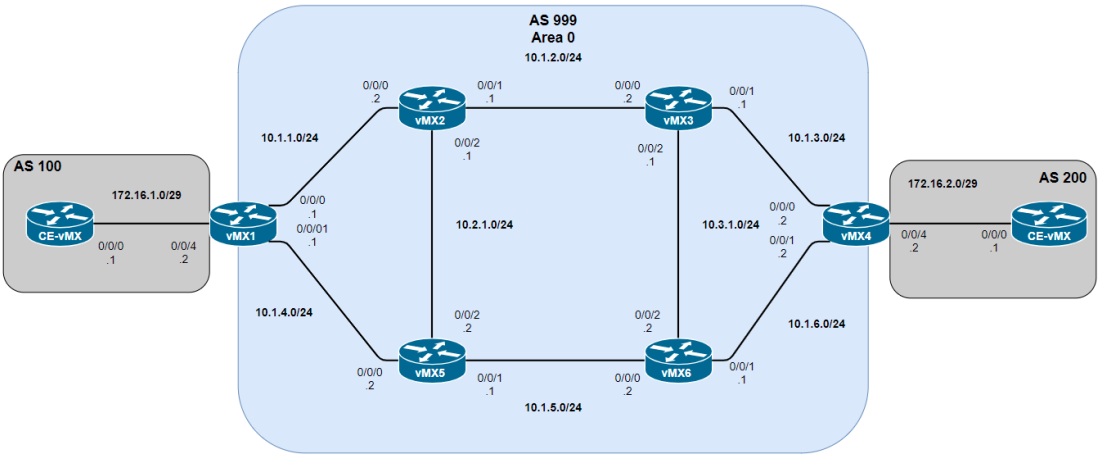


Figure 16. Laboratory Topology

Figure 16 shows the laboratory topology used in this thesis. Autonomous System, AS 999 is represents Service Provider MPLS core and includes Juniper vMX routers 1 to 6. vMX1 ja vMX6 are provider edge (PE) routers which connect to two imaginary customer sites on both sides of topology. Customer Edges are both AS100 and AS200, and they in turn have their own vMX routers, the CE-vMXs. Customer sites have their private networks from 172.16.0.0/16 and router loopback-addresses in the same range. CE routers have only basic configuration which simulates LAN-subnet and eBGP that allows them to connect to MPLS -core.

Autonomous System 999 in the middle which represent ISP-core network will have Segment Routing enabled. Devices will have minimal MPLS configuration that enables SR in domain. This means that MPLS is enabled on each core interface, however label distribution protocols such LDP are not configured. For IGP, Intermediate System to Intermediate System (IS-IS) is used. IS-IS is also enabled in every core interface, and area is set to 47. Routers in PE-core also have iBGP peering with each other. Customer sites have EBGP with PE-edger routers, and private address range is routed over MPLS L3 VPN, which routes prefix 172.16.0.0./16 from left-side-site to right-side -site over MPLS. Router from CE-routers are filtered to only previously mentioned prefix is advertised to PE-core.

Figure 17 has simply a ping test from HOST1-LAN to HOST2-LAN. In MPLS, the host traffic is forward from lan-to-lan via L3-vpn named HOST-LAN-VPN. Traceroute would too but due to nature of MPLS the hops in domain would time out the ICMP-echoes, even though they are passing the actual data.

```
student@vmx7> ping 172.16.255.2 detail
PING 172.16.255.2 (172.16.255.2): 56 data bytes
64 bytes from 172.16.255.2 via ge-0/0/0.0: icmp_seq=0 ttl=60 time=4.298 ms
64 bytes from 172.16.255.2 via ge-0/0/0.0: icmp_seq=1 ttl=60 time=4.359 ms
64 bytes from 172.16.255.2 via ge-0/0/0.0: icmp_seq=2 ttl=60 time=4.494 ms
^C
--- 172.16.255.2 ping statistics ---
3 packets transmitted, 3 packets received, 0% packet loss
round-trip min/avg/max/stddev = 4.298/4.384/4.494/0.082 ms
```

Figure 17. Connection from HOST1 to HOST2

Table 1 shows loopback address of each router in topology. These addresses are also used as router IDs in routing protocols and as a source of protocol advertisement origins. Network for each is /24.

See appendix 1 for clearer topology picture.

Table 1. Router Adressing

Router	Loopback IP
vMX1	100.0.0.101
vMX2	100.0.0.102
vMX3	100.0.0.103
vMX4	100.0.0.104
vMX5	100.0.0.105
vMX6	100.0.0.106
vMX7 / CE-vMX1	100.0.0.107
vMX8 / CE-vMX2	100.0.0.108

5.2.1 Device configurations

All core router, vmx1 to vmx6 have largely same identical base configuration, difference being at this point their addresses and router-ids. ISIS adjacencies are formed and iBGP peering within routers in core-domain is established. Every core interface has MPLS enabled, as this is a MUST for segment routing to work.

The very minimum to enable segment routing within Juniper routes is to set source packet routing under IGP protocol. See figure 18, there “source-packet-routing” is enabled under ISIS protocol. This triggers router for SR-operations and device will begin to allocate labels dynamically.

```
student@vmx1# show protocols isis
interface ge-0/0/0.0;
interface ge-0/0/1.0;
interface lo0.0 {
    passive;
}
source-packet-routing;
level 1 disable;
traffic-engineering {
    family inet {
        shortcuts;
    }
}
```

Figure 18. source-packet-routing

After setting has been configured and committed, dynamically allocated can be checked by viewing ISIS neighbor adjacencies. Seen in figure 19 is output of ISIS adjacencies to its neighbors. Link via Interface ge-0/0/0 connecting to vMX2 has been allocated adjacency-SID 788 and link ge-0/0/1 to vMX5 has been allocated adj-SID of 787. Worth of mentioning here is that only IPv4 labels are being allocated due to IPV6 addresses missing from device configurations (interfaces and ISIS for example).

```
student@vmx1# run show isis adjacency detail
vmx2
  Interface: ge-0/0/0.0, Level: 2, State: Up, Expires in 8 secs
  Priority: 64, Up/Down transitions: 1, Last transition: 06:25:54 ago
  Circuit type: 2, Speaks: IP, IPv6, MAC address: 0:50:56:88:b2:e9
  Topologies: Unicast
  Restart capable: Yes, Adjacency advertisement: Advertise
  LAN id: vmx2.02, IP addresses: 10.1.1.2
  Level 2 IPv4 Adj-SID: 788

vmx5
  Interface: ge-0/0/1.0, Level: 2, State: Up, Expires in 21 secs
  Priority: 64, Up/Down transitions: 1, Last transition: 06:23:32 ago
  Circuit type: 2, Speaks: IP, IPv6, MAC address: 0:50:56:88:35:30
  Topologies: Unicast
  Restart capable: Yes, Adjacency advertisement: Advertise
  LAN id: vmx1.02, IP addresses: 10.1.4.2
  Level 2 IPv4 Adj-SID: 787
```

Figure 19. ISIS adajency-SID labels

Segment Routing labels can also be verified from routers MPLS route table. Figure 20 is MPLS route table of vMX1, where above Adjacency-labels are also seen as the SR operations. Looking routers show that when vMX1 receives packet with top label of

787, label is popped and forward via ge-0/0/1 to vMX5. L-ISIS remarks seen in table are short for Labeled ISIS.

```
student@vmx1# run show route table mpls.0

mpls.0: 10 destinations, 10 routes (10 active, 0 holddown, 0 hidden)
+ = Active Route, - = Last Active, * = Both

0                  *[MPLS/0] 06:30:14, metric 1
                      to table inet.0
0(S=0)             *[MPLS/0] 06:30:14, metric 1
                      to table mpls.0
1                  *[MPLS/0] 06:30:14, metric 1
                      Receive
2                  *[MPLS/0] 06:30:14, metric 1
                      to table inet6.0
2(S=0)             *[MPLS/0] 06:30:14, metric 1
                      to table mpls.0
13                 *[MPLS/0] 06:30:14, metric 1
                      Receive
787                *[L-ISIS/14] 05:43:56, metric 0
                    >  to 10.1.4.2 via ge-0/0/1.0, Pop
787(S=0)           *[L-ISIS/14] 05:43:56, metric 0
                    >  to 10.1.4.2 via ge-0/0/1.0, Pop
788                *[L-ISIS/14] 05:43:56, metric 0
                    >  to 10.1.1.2 via ge-0/0/0.0, Pop
788(S=0)           *[L-ISIS/14] 05:43:56, metric 0
                    >  to 10.1.1.2 via ge-0/0/0.0, Pop
```

Figure 20. vMX1 MPLS Route

To enable shorted path forwarding in Segment routing MPLS, node segments need to be configured. While traditionally LPD can perform this, Node-SIDs have considerable advantages over, notably the source routing capabilities. Node-SIDs are associated with loopback of each MPLS vMX in domain. Two key pieces of information were re-quited to be configured to perform shortest path forwarding with SR. Node-SID and Segment Routing Global Block. Routing between different ranges is possible but hav-ing same label-range on domain has quite useful effect.

SRGB is set under protocols, ISIS. Staring label in this lab is 9000 and range of labels is 100. Meaning that laboratory domain as a whole has the SRGB of 9000 to 9999. SRGB has global significance and need to same within each router in domain. Node-SIDs, are the opposite. They will be unique with each router, vMX1 having the 101, vMX2

the 102 and so on. Overall label-value then to reach each is simply SRBG plus manually set Node-SID, vMX1 for example will be 9101.

After Segment routing related setting have been applied to node, they are visible under chosen IGP protocol. In lab, setting can be verified under ISIS. In figure 21 is screen capture on vMX1 ISIS overview. Segment Routing related info ins seen in the middle, under SPRING.

From ISIS overview, few useful things related to SR can be verified. Primarily that SR is used in conjunction with IGP, but also that SRBG has been successfully applied to router. Lastly the Node-SID is which is applied to device is listed there also.

```
student@vmx1# run show isis overview
Instance: master
  Router ID: 100.0.0.101
  Hostname: vmx1
  Sysid: 1000.0000.0101
  Areaid: 47
  Adjacency holddown: enabled
  Maximum Areas: 3
  LSP life time: 1200
  Attached bit evaluation: enabled
  SPF delay: 200 msec, SPF holddown: 5000 msec, SPF rapid runs: 3
  IPv4 is enabled, IPv6 is enabled, SPRING based MPLS is enabled
  Traffic engineering: enabled
  Traffic engineering v6: disabled
  Restart: Disabled
    Helper mode: Enabled
  Layer2-map: Disabled
  Source Packet Routing (SPRING): Enabled
    SRGB Config Range :
      SRGB Start-Label : 9000, SRGB Index-Range : 1000
    SRGB Block Allocation: Success
      SRGB Start Index : 9000, SRGB Size : 1000, Label-Range: [ 9000, 9999 ]
    Node Segments: Enabled
      Ipv4 Index : 101
    SRv6: Disabled
  Post Convergence Backup: Disabled
  Level 1
    Internal route preference: 15
    External route preference: 160
    Prefix export count: 0
    Wide metrics are enabled, Narrow metrics are enabled
    Source Packet Routing is enabled
  Level 2
    Internal route preference: 18
    External route preference: 165
    Prefix export count: 0
    Wide metrics are enabled, Narrow metrics are enabled
    Source Packet Routing is enabled
```

Figure 21. ISIS overwiev of vMX1 with SPRING

SRBG configuration also now enables inet.3 route table, or “label-table” . Inet.3 table form vMX1 in figure 22 shows labeled path to each MPLS enable node in domain, namely their loopback address. Inet.3 table is one more step verifying that MPLS is functioning properly in domain and since labels reflect those set in SRGB, then is safe to assume that MPLS is running SR-routing.

```
student@vmx1# run show route table inet.3

inet.3: 5 destinations, 5 routes (5 active, 0 holddown, 0 hidden)
+ = Active Route, - = Last Active, * = Both

100.0.0.102/32     *[L-ISIS/14] 2w0d 23:45:41, metric 10
                    >  to 10.1.1.2 via ge-0/0/0.0
100.0.0.103/32     *[L-ISIS/14] 2w0d 09:26:56, metric 20
                    >  to 10.1.1.2 via ge-0/0/0.0, Push 9103
100.0.0.104/32     *[L-ISIS/14] 2w0d 09:25:28, metric 30
                       to 10.1.1.2 via ge-0/0/0.0, Push 9104
                    >  to 10.1.4.2 via ge-0/0/1.0, Push 9104
100.0.0.105/32     *[L-ISIS/14] 2w0d 23:41:57, metric 10
                    >  to 10.1.4.2 via ge-0/0/1.0
100.0.0.106/32     *[L-ISIS/14] 2w0d 09:25:28, metric 20
                    >  to 10.1.4.2 via ge-0/0/1.0, Push 9106
```

Figure 22. vMX1 Inet.3 table

5.2.2 Traffic Engineering

In previous chapter Segment Routing was activated on devices. However, traffic still flows using standard shortest path route. To actually provide actual example of Source Routing, traffic will be diverted to flow throughout alternate path instead of typical shortest one. See the topology again in figure 23. RED arrows show the shortest path route which traffic is would normally take, and green arrows are representing the alternate route that we want to traffic to take by using source routing. VMX routers have their Node-SID which were set earlier marked next to them, and with SR traffic can be steered in MPLS with them

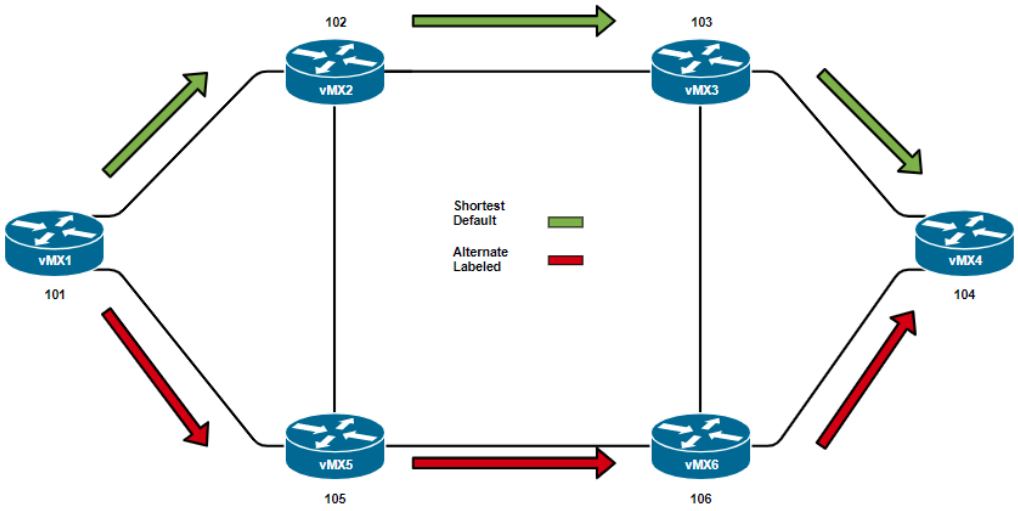


Figure 23. IGP and Labeled path

Alternate route will have traffic flowing the path 101 – 105 – 106 – 104, taking alternative route instead of a default IGP-shorts path. Looking at route table of vMX1 shows that currently route to Host2-LAN, 172.16.225.2, from ingress node (vMX1) is through ge-0/0/0 and ISIS-adjacency confirms that to be vMX2. Label 9104 is being pushed top.

```
student@vmx1# run show route 172.16.255.2

HOST-LAN-VPN.inet.0: 5 destinations, 6 routes (5 active, 0 holddown, 0 hidden)
+ = Active Route, - = Last Active, * = Both

172.16.255.2/32    *[BGP/170] 01:05:13, localpref 100, from 100.0.0.104
                      AS path: 200 I, validation-state: unverified
                    > to 10.1.1.2 via ge-0/0/0.0, Push 19, Push 9104(top)
                      to 10.1.4.2 via ge-0/0/1.0, Push 19, Push 9104(top)

student@vmx1# run show isis adjacency
Interface             System         L State       Hold (secs) SNPA
ge-0/0/0.0            vmx2           2  Up                8  0:50:56:88:b2:e9
ge-0/0/1.0            vmx5           2  Up               25  0:50:56:88:35:30
```

And when checking what vMX2 does with label 9104, it currently forwards it to vMX3. VMX3 in turn would send traffic to vMX4 which is the MPLS egress node in this case.

student@vmx2> show route label 9104

mpls.0: 20 destinations, 20 routes (20 active, 0 holddown, 0 hidden)
*+ = Active Route, - = Last Active, * = Both*

*9104 *[L-ISIS/14] 2w4d 12:12:12, metric 20*
> to 10.1.2.2 via ge-0/0/1.0, Swap 9104

To steer this towards vMX5 instead, a labeled-route is set using Node-SIDs is configured on ingress node, vMX1.This is achieved via hop-by-hop sequence configured at ingress router. In figure 24 is output from ingress router, vMX1. Labeled segment list is set with name SR-LABEL-PATH which is used in route to destination set with name ALTERNATE-TO-VMX4. Only a single hop is currently set to vMX5 via label 9105 that is also Node-SID of said router.

```
student@vmx1# show protocols source-packet-routing
segment-list SR-LABEL-PATH {
    hop1 label 9105;
}
source-routing-path ALTERNATIVE-TO-VMX4 {
    to 100.0.0.104;
    primary {
        SR-LABEL-PATH;
    }
}
```

Figure 24. Labeled static route

Checking route to 172.16.255.2 from vMX1 shows that packet is now indeed send to vMX5. It however does not any more labels to forward packet further, so at this point traffic between HOST-networks is interrupted. Addition hops are required in labeled path configuration on vMX1.

student@vmx1# run show route 172.16.255.2

HOST-LAN-VPN.inet.0: 5 destinations, 6 routes (5 active, 0 holddown, 0 hidden)

*+ = Active Route, - = Last Active, * = Both*

*172.16.255.2/32 *[BGP/170] 4d 02:59:28, localpref 100, from 100.0.0.104*
AS path: 200 I, validation-state: unverified
> to 10.1.4.2 via ge-0/0/1.0, Push 19

Configuring rest of the required hops to reach vMX4 (that has router to HOST2 network) resumes connectivity. Route table from vMX1 also has some new points of interest. Now label 9106 (vMX6) is pushed to top before 9104. VPN label 19 stays there since we are using VPN route between Host-LANs. Figure 25 has output from vMX1 that includes rest of the hops in labeled path as well simple ping test over VPN route.

student@vmx1# run show route 172.16.255.2

HOST-LAN-VPN.inet.0: 5 destinations, 6 routes (5 active, 0 holddown, 0 hidden)
*+ = Active Route, - = Last Active, * = Both*

*172.16.255.2/32 *[BGP/170] 4d 04:18:42, localpref 100, from 100.0.0.104*
AS path: 200 I, validation-state: unverified
> to 10.1.4.2 via ge-0/0/1.0, Push 19, Push 9104, Push 9106(top)

```
[edit protocols source-packet-routing segment-list SR-LABEL-PATH]
student@vmx1# show
hop1 label 9105;
hop2 label 9106;
hop3 label 9104;

[edit protocols source-packet-routing segment-list SR-LABEL-PATH]
student@vmx1# run ping 172.16.255.2 routing-instance HOST-LAN-VPN
PING 172.16.255.2 (172.16.255.2): 56 data bytes
64 bytes from 172.16.255.2: icmp_seq=0 ttl=59 time=4.404 ms
64 bytes from 172.16.255.2: icmp_seq=1 ttl=59 time=4.679 ms
64 bytes from 172.16.255.2: icmp_seq=2 ttl=59 time=5.433 ms
^C
--- 172.16.255.2 ping statistics ---
3 packets transmitted, 3 packets received, 0% packet loss
round-trip min/avg/max/stddev = 4.404/4.839/5.433/0.435 ms
```

Figure 25. Labeled path connectivity

Now due how MPSL works and how lab topology is set up, once packet reaches ether vMX3 or vMX6 the MPSL penultimate-hop-pooing is going to kick in. Since both after mentioned nodes are the second-to-last hop to network 172.16.0.0/16, the nodes are going to remove labels and forward packet to vMX4 if traffic is destined to that network. Something to keep in mind.

5.2.3 Traffic Protection

MPLS networks do have protection mechanisms included in their exiting protocols. LDP has the loop-free alternate (LFA) and RSVP-TE does have its own ways of recovering in the cases of loss of link, node, or fate-sharing resources. But neither of those

is without their own issues. LFA is dependent on topology to provide protection, and RSVP-TE's experience feature set cause additional resource overhead.

Segment Routings idea with traffic protection is to build backup path right from the beginning, instead of during network convergence event. Topology Independent Loop-Free Alternate, TI-LFA used by SR provides link failure, node failure, and fate-sharing protection in case of unexpected link or node issues within the Segment Routing domain. Backup path chosen by TI-LFA do follow post convergence path signaled by IGP when failure occurs in a network. However, this transition time is significantly lower in milliseconds. Link-protection via TI-LFA enables backup routes in case of link between nodes goes down. Node protection does more than that. In case a whole node goes down, the backup path is not computed over failed node. Fate-sharing options ensure that backup path does not share network links or other physical medium paths with original primary path.

Demonstrating this can be done via adjusting topology one more time. In figure 25, topology has been included with modified link metrics for ISIS. This is because TI-LFA in the post-convergence path, the cost of links from neighboring node is expected to increase by a certain amount. Green arrows represent the IGP shortest route. Red sign is would-be-broken link, from xMX1 ge-0/0/0 to vMX2 ge-0/0/0. Idea is to have backup route between vMX1 and vMX4 ready, so user traffic between host-LANs would not be interrupted.

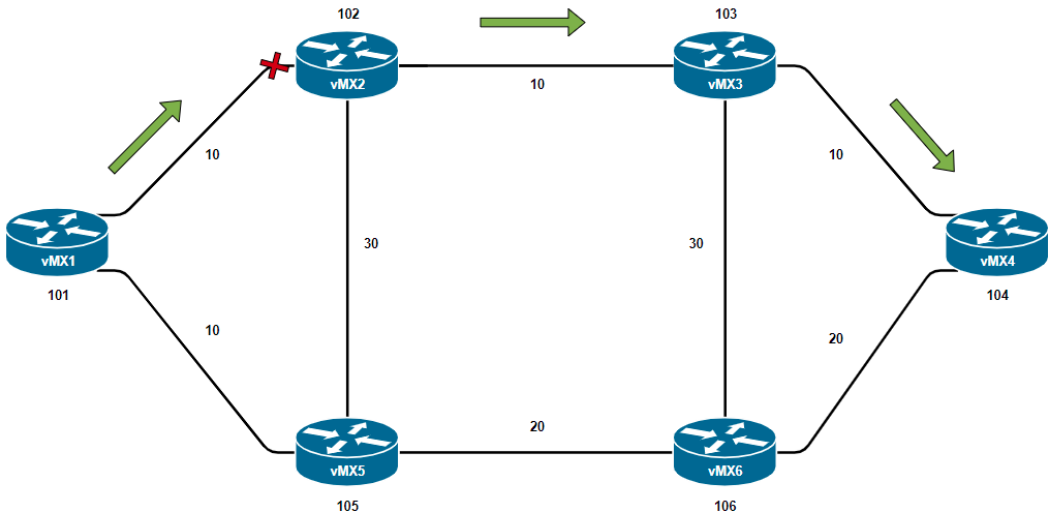


Figure 26. IGP link-cost topolgy

Inet.3 route table form vMX1 does show route to each node in domain. However, table is populated only with current active routes. Were there to be a node or link failure in network, table would be included with new routes after IGP convergence.

```
student@vmx1> show route table inet.3

inet.3: 5 destinations, 5 routes (5 active, 0 holddown, 0 hidden)
+ = Active Route, - = Last Active, * = Both

100.0.0.102/32     *[L-ISIS/14] 00:00:15, metric 10
                    > to 10.1.1.2 via ge-0/0/0.0
100.0.0.103/32     *[L-ISIS/14] 00:00:15, metric 20
                    > to 10.1.1.2 via ge-0/0/0.0, Push 9103
100.0.0.104/32     *[L-ISIS/14] 00:00:15, metric 30
                    > to 10.1.1.2 via ge-0/0/0.0, Push 9104
100.0.0.105/32     *[L-ISIS/14] 00:00:15, metric 10
                    > to 10.1.4.2 via ge-0/0/1.0
100.0.0.106/32     *[L-ISIS/14] 00:00:15, metric 30
                      to 10.1.1.2 via ge-0/0/0.0, Push 9106
                    > to 10.1.4.2 via ge-0/0/1.0, Push 9106
```

Rapid ping is run from vMX1 to vMX4 to maximize packets per seconds while interface ge-0/0/0 is manually disabled from vMX2. This is to simulated link failure. With IGP, single packet is lost during the process, visualized in figure 27.

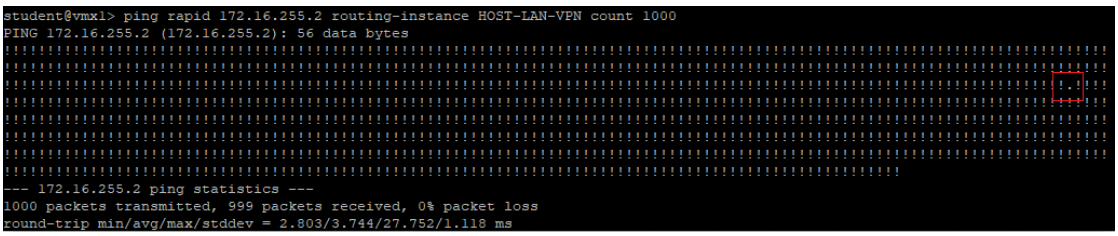


Figure 27. Rapid ping with IGP

TI-LFA is going to be enabled on vMX1. This would protect traffic running through router when primary link fails, instead traffic sis swiftly switched to backup-path. Settings are applied again under protocols. Verifying them is however somewhat difficult. When TI-LFA features are enabled, they do show up under IGP protocol interface overview, ISIS in this case as in figure 28.

```
student@vmx1# run show isis interface extensive | match "ge-|Protection"
ge-0/0/0.0
    Post convergence Protection:Enabled, Fate sharing: Yes, Srlg: No, Node cost: 2000
ge-0/0/1.0
    Post convergence Protection:Enabled, Fate sharing: No, Srlg: No, Node cost: 16777215
```

Figure 28. ISIS TFI-LFA interfaces

Another way of verifying that Link -and Node-protection is enabled is to look at inet.3 table on vMX1 after configurations have been committed, figure 29. It now has been populated with backup path's Alternate routes to each node in domain is seen on table, as well as their SIDs. Backup to vMX4 for example is through ge-0/0/1, label 9106 (vMX6) is being pushed to top.

```
student@vmx1# run show route table inet.3 100.0.0.104

inet.3: 5 destinations, 5 routes (5 active, 0 holddown, 0 hidden)
+ = Active Route, - = Last Active, * = Both

100.0.0.104/32     *[L-ISIS/14] 09:30:53, metric 30
                    >  to 10.1.1.2 via ge-0/0/0.0, Push 9104
                       to 10.1.4.2 via ge-0/0/1.0, Push 9104, Push 9106(top)
```

Figure 29. TI-LFA backup route

If rapid ping is run again now when TI_LFA is enabled, result are similar to when only IGP convergence was used. While TI-LFA does provides faster convergence, it is not instantaneous either.

--- 172.16.255.2 ping statistics ---
1000 packets transmitted, 999 packets received, 0% packet loss
round-trip min/avg/max/stddev = 2.707/3.566/42.779/1.647 ms

6 Results and Discussion

In a small scale environment such as this topology, some things are easier, and some are harder. Basic segment Routing configurations such as SRBG are easy to commit due to a small number of devices. Features such as TI-LFA are somewhat harder to demonstrate because hops between ingress and egress are minimal. Main feature of TI-LFA is also fast convergence and it is designed to minimalize traffic interruptions. In this lab, only a handful of routers need to be traversed and physical distance is non-existent, therefore demonstrating latency caused by network convergence is difficult to pull off. Plenty of Segment Routing features are also enabled by external controller, such as Juniper NorthStar. ECMP and latency adjustment are one such feature.

Setting up the laboratory environment for actual Segment Routing testing took some time and slight planning. While other configurations such, IPG, BGP and MPLS are not really a main focus, implementing them top topology needs some considerations for them to work properly. And they are still requited for SR to work. But once first device had functioning configurations, it's running-config could be used as a temple for others since mostly only addressing needed to be changes. Juniper Networks also has plenty of example configuration and explanations for older technologies and protocols on their webpages. These were easy enough to implement in lab topology.

The very basic setting for Segment Routing were also quite easy to set up once some basic information was gathered about them. This could be either because SR is made designed to be easy to implement in pre-existing networks topologies and on top of existing routing protocols. Or because they are the ones that have the most third party documentation made of. Once past SRGB however, the information becomes vastly more difficult to be run in to. Juniper Networks do have basic examples of these too, yet they feel either minimalistic or require good knowledge of other more obscure features. Acquiring knowledge from other sources felt almost impossible at this point also. This may be due to SR being new technology and not used in actual production networks in same volume as older technologies.

Lab network was a success, other than TI-LFA settings which left little bit inconclusive feeling. TI-LFA speeds the restoration after network convergence in milliseconds. Testing this via ping from device is not effective, and traffic generator would have been needed. Obviously, not all configurations were success on first commit. Problematic thing about his one was the fact that there were no troubleshooting documented online. No forums or blog post containing problem descriptions and troubleshooting taken by individuals and no troubleshooting steps by developers either. Segment routing is not perfect technology either, so current unavailability of information on troubleshooting can make things complicated.

Segment Routing is still relatively young technology. This means that finding information can be somewhat difficult. Market leaders like Juniper and Cisco do have quite large knowledgebase about theoretical parts of SR, and Juniper Networks do have their own sample configurations for some of the features. There is not much user cases or experience at this point though. That being said, published information about Segment Routing is vastly more now than it was few years ago. Overall, in this laboratory, the most arduous part was setting up the topology and basic connectivity settings on each device.

7 Conclusions

I've have had firsthand experience from live customer environment MPLS-networks. But while they are daunting and complex, they are still currently very much needed. Without them the whole backbone of Internet would be in peril. That being said I also very much agree that they are cumbersome and truly seem difficult to scale upward. And ´when something eventually breaks and requires troubleshooting, figuring out the ins and outs of traditional MPLS suddenly becomes not only time consuming, but also exceedingly difficult. Adding new networks or devices to network is complex process from both the party adding them and also for those parties through whom data may traverse.

With Segment Routing, we can indeed remove now redundant label protocols from network. This alone can be most helpful. But having global labels which are allotted from known and adjustable block is truly the thing which simplifies things regarding MPLS networks. Having known base for labels has also additional benefits as seen in lab testing during this thesis, such labeled paths. There are many more traffic engineering solutions available due them too, that are too many in numbers to be all include in this document. And since SR is still young technology, there is more to be expected in future.

I am quite happy with this thesis. Like mentioned earlier, SR has too much to be included in single document. But just with these configurations and reading thought theoretical insides of MPLS and SR gave most helpful insight to my future career, as well as renewed my skills with Juniper Networks devices. And in regards of Segment Routing, I do believe there is plenty more to come and be learned, but this was a basis for furthers learning. And continuous learning and adapting to new technologies is very much requirement IT-professionals.

I would like to express special gratitude to JYVSECTEC who provided me with actual live Juniper virtual MX routers, and to Jyväskylä University of Applied Sciences, JAMK who patiently aided me during this thesis.

References

Abous us. 2021 JYVSECTEC. Accessed January 2021. Retrieved from https://jyvsectec.fi/about/overview/

Abdelrahman, M. 2018. Segment Routing Benefits. LinkedIn Blog. Accessed January 2021. Retrieved from https://www.linkedin.com/pulse/segment-routing-benefits-mohamed-abdelrahman-ccie-sp-pmp/

Arista Networks. 2016. MPLS Segment Routing. Whitepaper. Accessed January 2021. Retrieved from arista.com/assets/data/pdf/Whitepapers/MPLSSegmentRouting_Whitepaper.pdf

David, P. 2020. MPLS History and building blocks. Cisco Learning Network. Accessed March 2021. Retrieved from https://learningnetwork.cisco.com/s/article/MPLS-History-and-building-blocks

David, P. 2018. Introduction to Segment Routing. Cisco Learning Network. Accessed March 2021. Retrieved from https://learningnetwork.cisco.com/s/blogs/a0D3i000002SKP6EAO/introduction-to-segment-routing

ExamGuides.2021. CNA ICND2 Exam Cram Notes: MPLS. Chapter 3. WAN Technologies. Accessed May 2021. Retrieved from https://www.examguides.com/CCNA-ICND2/ccna-icnd2-15.htm

Farrel, A. Bonica, R 2017. Segment Routing: Cutting Through the Hype and Finding the IETF's Innovative Nugget of Gold. IETF Journal. Accessed February 2021. Retrieved from https://www.ietfjournal.org/segment-routing-cutting-through-the-hype-and-finding-the-ietfs-innovative-nugget-of-gold/

Filsfils, C. Nainar, N. Pignataro, C. Cardona, J. Francois, P. The Segment Routing Architecture. Cisco Systems. Accessed February 2021. Retrieved from https://eprints.networks.imdea.org/1306/1/the_segment-routing_architecture_2015.pdf

Gregory, T. 2016. Segment Routing on JUNOS – The basics. 2016. CCIE Personal Tech Blog. Accessed January 2021. Retrieved from https://tgregory.org/2016/08/13/segment-routing-on-junos-the-basics/

Litmanen, I. 2017. Segment Routing. Thesis. Accessed April 2021. Retrieved from https://www.theseus.fi/bitstream/handle/10024/124568/Litmanen_Ilpo.pdf?sequence=1

Kaur, R. 2009. IPv4 Header. WordPress Blog. Accessed May 2021. Retrieved from https://advancedinternettechnologies.wordpress.com/ipv4-header/

Khare A. Barth, C. 2019. Day One: Inside Segment Routing, 10th edition. Juniper Networks Books.

Lucek, J. Szarkowicz, K. 2018. Day One: Configuring Segment Routing with Junos, 10th edition. Juniper Networks Books.

Nurminen P.2017. Centralized Segment Routing. Thesis. Accessed March 2021. Retrieved from
https://www.theseus.fi/bitstream/handle/10024/123779/Nurminen_Paul.pdf?sequence=1

RFC8402. 2018. Segment Routing Architecture. Section 2. Internet Engineering Task Force (IETF). Accessed February 2021. Retrieved from
https://tools.ietf.org/html/rfc8402#section-2

RFC8665. 2019. OSPF Extensions for Segment Routing. Internet Engineering Task Force (IETF). Accessed May 2021. Retrieved from
https://tools.ietf.org/html/rfc8665

RFC8667. 2019. IS-IS Extensions for Segment Routing. Internet Engineering Task Force (IETF). Accessed May 2021. Retrieved from
https://tools.ietf.org/html/rfc8667

RFC8669. 2019. Segment Routing Prefix Segment Identifier Extensions for BGP. Section 2. Internet Engineering Task Force (IETF). Accessed February 2021. Retrieved from
https://tools.ietf.org/html/rfc8669

Benefits of Segment Routing. 2020. Juniper Networks. Accessed January 2021. Retrieved from
https://www.youtube.com/watch?v=ogjAp_isBIg

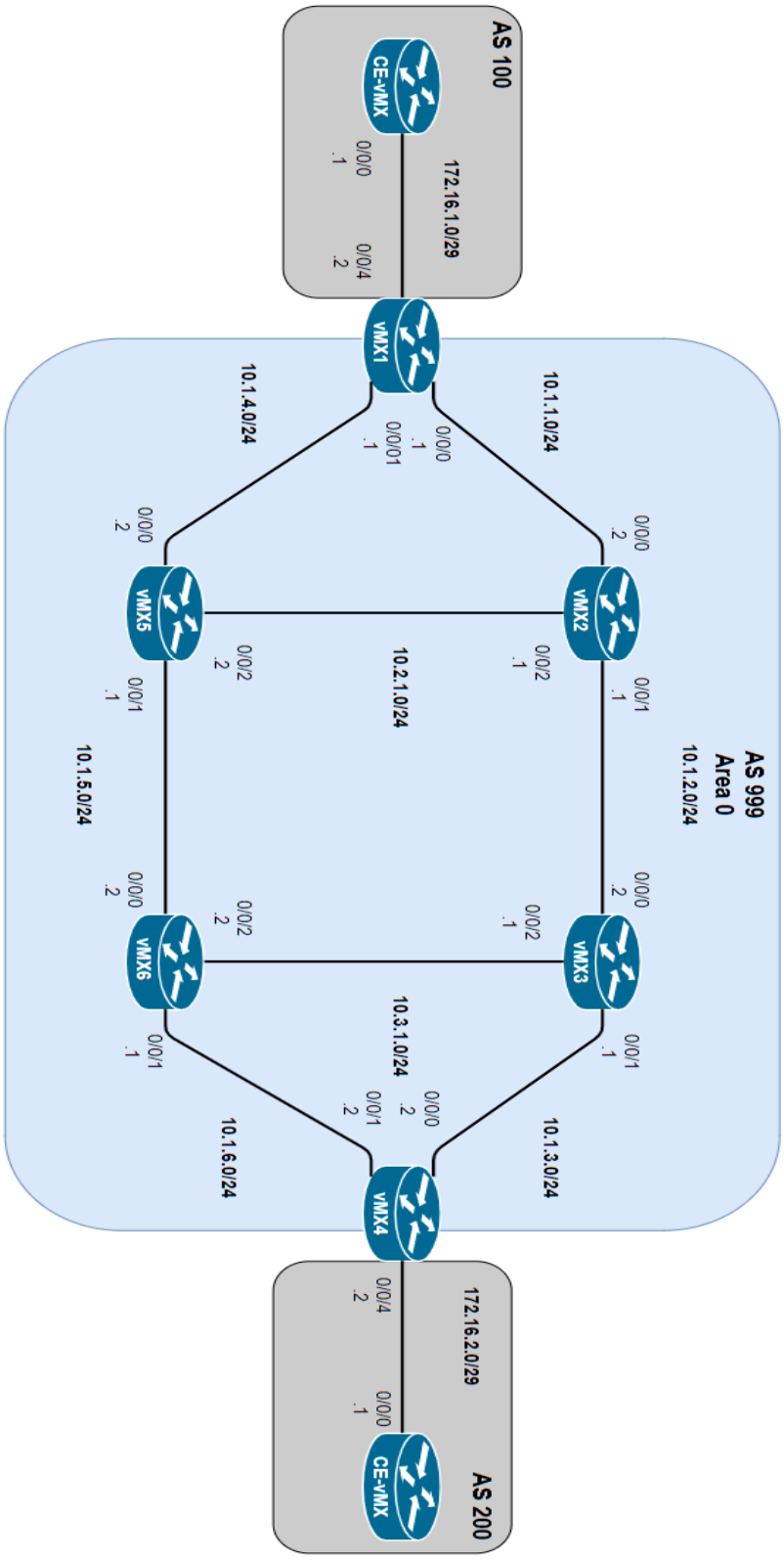
Segment Routing fundamentals. 2020. Juniper Networks. Accessed January 2021. Retrieved from
https://www.youtube.com/watch?v=wv2ZiS1-AB4

Segment Routing Overview. 2020. Juniper Networks. Accessed January 2021. Retrieved from
https://www.youtube.com/watch?v=HXs6keKpkK0

What is segment routing? Juniper Networks. Accessed March 2021.
Retrieved from
https://www.juniper.net/us/en/products-services/what-is/segment-routing/

Appendixes

Appendix 1. Lab topology



Appendix 2. vMX1 configuration

```
set version 21.1R1.11
set system services
set system time-zone Europe/Helsinki
set system syslog file messages any notice
set system syslog file messages authorization info
set system syslog file messages archive world-readable
set system syslog file interactive-commands interactive-commands any
set system processes dhcp-service traceoptions file dhcp_logfile
set system processes dhcp-service traceoptions file size 10m
set system processes dhcp-service traceoptions level all
set system processes dhcp-service traceoptions flag packet
set chassis fpc 0 lite-mode
set chassis network-services enhanced-ip
set interfaces ge-0/0/0 unit 0 description to-vmx2
set interfaces ge-0/0/0 unit 0 family inet address 10.1.1.1/24
set interfaces ge-0/0/0 unit 0 family iso
set interfaces ge-0/0/0 unit 0 family mpls
set interfaces ge-0/0/1 unit 0 description to-vmx5
set interfaces ge-0/0/1 unit 0 family inet address 10.1.4.1/24
set interfaces ge-0/0/1 unit 0 family iso
set interfaces ge-0/0/1 unit 0 family mpls
set interfaces ge-0/0/4 description to-HOST1-LAN
set interfaces ge-0/0/4 unit 0 description to-HOST-1
set interfaces ge-0/0/4 unit 0 family inet address 172.16.1.2/30
set interfaces ge-0/0/7 description Management
set interfaces ge-0/0/7 unit 0 family inet address 192.168.255.101/24
set interfaces lo0 unit 0 family inet address 100.0.0.101/32
set interfaces lo0 unit 0 family iso address 47.1000.0000.0101.00
set routing-instances HOST-LAN-VPN instance-type vrf
set routing-instances HOST-LAN-VPN protocols bgp group HOST1-LAN type external
set routing-instances HOST-LAN-VPN protocols bgp group HOST1-LAN peer-as 100
set routing-instances HOST-LAN-VPN protocols bgp group HOST1-LAN neighbor
172.16.1.1
set routing-instances HOST-LAN-VPN interface ge-0/0/4.0
set routing-instances HOST-LAN-VPN route-distinguisher 100.0.0.101:69
set routing-instances HOST-LAN-VPN vrf-target target:999:69
set protocols router-advertisement interface fxp0.0
set protocols bgp group ibgp type internal
set protocols bgp group ibgp local-address 100.0.0.101
set protocols bgp group ibgp family inet-vpn unicast
set protocols bgp group ibgp neighbor 100.0.0.104
set protocols bgp group ibgp neighbor 100.0.0.102
set protocols bgp group ibgp neighbor 100.0.0.103
set protocols bgp group ibgp neighbor 100.0.0.105
set protocols bgp group ibgp neighbor 100.0.0.106
```

```
set protocols isis interface ge-0/0/0.0 level 2 post-convergence-lfa node-protection
cost 2000
set protocols isis interface ge-0/0/0.0 level 2 post-convergence-lfa fate-sharing-pro-
tection
set protocols isis interface ge-0/0/1.0 level 2 post-convergence-lfa node-protection
set protocols isis interface lo0.0 passive
set protocols isis source-packet-routing srgb start-label 9000
set protocols isis source-packet-routing srgb index-range 1000
set protocols isis source-packet-routing node-segment ipv4-index 101
set protocols isis level 1 disable
set protocols isis backup-spf-options use-post-convergence-lfa
set protocols isis backup-spf-options use-source-packet-routing
set protocols mpls interface ge-0/0/0.0
set protocols mpls interface ge-0/0/1.0
set protocols source-packet-routing
set protocols lldp interface ge-0/0/0
set protocols lldp interface ge-0/0/1
set protocols lldp interface ge-0/0/2
set protocols lldp interface ge-0/0/4
set routing-options router-id 100.0.0.101
set routing-options autonomous-system 999
set routing-options fate-sharing group FS-GROUP use-for-post-convergence-lfa
set routing-options fate-sharing group FS-GROUP from 10.1.1.1 to 10.1.1.2
set routing-options fate-sharing group FS-GROUP from 10.2.1.2 to 10.2.1.1
```

Appendix 3. CE-vMX1 configuration

```
set version 21.1R1.11
set system services
set system time-zone Europe/Helsinki
set system syslog file messages any notice
set system syslog file messages authorization info
set system syslog file messages archive world-readable
set system syslog file interactive-commands interactive-commands any
set system processes dhcp-service traceoptions file dhcp_logfile
set system processes dhcp-service traceoptions file size 10m
set system processes dhcp-service traceoptions level all
set system processes dhcp-service traceoptions flag packet
set chassis fpc 0 lite-mode
set interfaces ge-0/0/0 description to-vmx1-mpls
set interfaces ge-0/0/0 unit 0 family inet address 172.16.1.1/30
set interfaces ge-0/0/7 description Management
set interfaces ge-0/0/7 unit 0 family inet address 192.168.255.107/24
set interfaces lo0 unit 0 family inet address 172.16.255.1/32
set policy-options policy-statement ADVERTISE-DIRECT term 1 from protocol direct
```

```
set policy-options policy-statement ADVERTISE-DIRECT term 1 from route-filter
172.16.0.0/16 orlonger
set policy-options policy-statement ADVERTISE-DIRECT term 1 then accept
set protocols router-advertisement interface fxp0.0
set protocols bgp group VMX1 type external
set protocols bgp group VMX1 export ADVERTISE-DIRECT
set protocols bgp group VMX1 peer-as 999
set protocols bgp group VMX1 neighbor 172.16.1.2
set protocols lldp interface ge-0/0/0
set routing-options router-id 172.16.255.1
set routing-options autonomous-system 100
```