



TAMPERE UNIVERSITY OF TECHNOLOGY

JARKKO VILKKI
PLANNING TOOLS FOR MPLS NETWORKS
MASTER OF SCIENCE THESIS

Examiner: professor Jarmo Harju
The examiner and the topic
approved in the Department Council
of Faculty of Automation,
Mechanical and Materials
Engineering 13th of January 2010

ABSTRACT

TAMPERE UNIVERSITY OF TECHNOLOGY

Master's Degree Programme in Automation Technology

VILKKI, JARKKO: Planning Tools for MPLS Networks

Master of Science Thesis, 55 pages

January, 2010

Major: Communication Networks and Protocols

Examiner: Professor Jarmo Harju

Keywords: MPLS, Network Planning

The networks utilizing MPLS functionality for routing their traffic are continuously growing larger and the functionality needed in these networks is coming more complex. For this reason the network planner would benefit from some kind of tools for the planning to be successful. The purpose of this thesis is to find out the key functionality, features and attributes of the planning tools used for MPLS enabled networks.

The thesis is divided to three parts. In the first part the technique behind the MPLS is revealed. This part goes through the key functionality of MPLS and protocols used in MPLS networks. The second part is about the network planning process and what it means in the case of MPLS enabled network. The third part introduces two different planning tools designed for MPLS planning.

The study indicates that the MPLS planning has many phases which are needed to be utilized in the planning tool for MPLS network. The functionality is also wide in MPLS networks and that reflects to the functionality needed in the planning tool. This study suggests that the tool planned must be scalable to networks of thousands of POPs and hundreds of thousand LSPs. The key functionality includes LSP design, TE optimization, importing and exporting of network topologies as whole and VPN planning in Layer 2 and Layer 3.

TIIVISTELMÄ

TAMPEREEN TEKNILLINEN YLIOPISTO

Automaatiotekniikan koulutusohjelma

VILKKI, JARKKO: Planning Tools for MPLS Networks

Diplomityö, 55 sivua

Tammikuu, 2010

Pääaine: Tietoliikenneverkot ja –protokollat

Tarkastaja: professori Jarmo Harju

Avainsanat: MPLS, Verkkosuunnittelu

Verkot, joissa MPLS-tekniikkaa (Multi Protocol Label Switching) käytetään pakettien reitittämiseen, kasvavat jatkuvasti yhä suuremmiksi ja toiminnallisuus, jota verkoissa tarvitaan, monipuolistuu koko ajan. Tämän syyn vuoksi verkon suunnittelija tarvitsee yhä parempia apuvälineitä, jotta suunnittelu olisi onnistunutta, optimaalista ja tuottaisi halutun tuloksen. Tämän diplomityön tarkoitus on selvittää tärkeimmät toiminnallisuudet ja ominaisuudet, joita MPLS-verkkojen suunnitteluun laadittu työkalu vaatii.

Diplomityö on jaettu kolmeen osaan. Ensimmäisessä osassa valotetaan MPLS-verkkojen käyttämää tekniikkaa. Tuossa osiossa käydään läpi tekniikat ja protokollat, joita MPLS-verkot käyttävät erinäisiin tehtäviin. Ensin käydään läpi yleisesti miksi MPLS-tekniikkaa ylipäättään tarvitaan ja miksi sitä käytetään verkkojen reitittämiseen. Tämän jälkeen tarkastellaan MPLS-protokollan otsikkokenttää ja sen osien käyttötarkoitukset selitetään. Sitten tarkastellaan MPLS-verkon rakennetta ja siihen kuuluvia laitteita. Seuraavaksi siirrytään osioon, joka selvittää kaikki yleisesti MPLS-polkujen rakentamiseen käytettävät protokollat ja miten ne eroavat toisistaan. Tämän jälkeen kerrotaan MPLS-vuonohjauksesta Differentiated Services-tekniikan avulla ja siitä miten se auttaa erilaisten liikenneluokkien erittelyssä MPLS-liikenteessä. Viimeinen kohta tässä osassa listaa erilaiset VPN-yhteydet, jotka ovat mahdollisia MPLS-tekniikkaa käytettäessä. Osio selventää näiden tekniikoiden eroavaisuudet ja mahdollisuudet, joita nämä MPLS-tekniikan avulla toteutettavat VPN-yhteydet suovat verrattuna aiempiin VPN-toteutuksiin.

Toinen osa tässä diplomityössä kertoo verkon suunnittelusta. Ensin käydään läpi verkon suunnittelua yleisellä tasolla. Tämä osa sisältää verkon suunnittelun eri vaiheet pääosittain: erilaiset ennustusmallit esitellään ja selvitetään mitoituksen ja vuonohjauksen rooli verkkosuunnittelussa. Näiden jälkeen siirrytään yleisestä verkonsuunnittelusta osioihin, joita käytetään MPLS-verkon suunnittelussa ja joiden yleisesti oletetaan tai halutaan löytyvän MPLS-verkkoihin tarkoitettusta suunnittelutyökalusta. Viimeinen kohta kertoo toiminnallisuus- ja skaalautuvuushaasteista, joihin MPLS:n on tekniikkana vastattava nykypäivänä.

Kolmannessa osiossa tarkastellaan kahta eri suunnittelutyökalua, jotka on laadittu MPLS-verkkojen suunnittelua varten: WANDL-yhtiön julkaisemaa IP/MPLSView:ta ja

Aria Networks Oy:n julkaisemaa iVNT:ta. Tässä osiossa käydään läpi näiden työkalujen toiminnallisuutta kertomalla erilaisista simulaatiomahdollisuuksista, joita kumpikin työkalu tarjoaa. Lisäksi kerrotaan mitä toimintoja ja protokollia näihin työkaluihin on mallinnettu, miten hyvin työkalut skaalautuvat kaupallisten MPLS-verkkojen tarpeisiin ja minkälaisista moduuleista työkalut on rakennettu.

Työn lopussa on pohdittu näiden kolmen osion perusteella, että mitkä ominaisuudet tulisi ottaa huomioon MPLS-verkon suunnittelutyökalua laadittaessa ja millä tavalla nämä ominaisuudet tulisi toteuttaa työkalussa. Näiden jälkeen on työhön vielä tehty loppuyhteenveto, joka kertoo työ tuloksista ja mahdollisista jatkokehitysmahdollisuuksista.

MPLS-verkon suunnittelu koostuu monesta eri vaiheesta, ja jokainen vaihe sisältää suuren määrän toiminnallisuusvaatimuksia. Nämä toiminnallisuusvaatimukset on mallinnettava MPLS-verkkojen suunnitteluun laaditussa työkalussa, jos halutaan että työkalu pystyy mallintamaan koko verkon suunnitteluprosessin alusta loppuun.

Tärkeimmät toiminnallisuudet, jotka MPLS-verkon suunnittelutyökalun tulee omata ovat simulointimahdollisuudet MPLS-poluille (LSP:t), MPLS-TE:lle, eri VPN-tyypeille ja DiffServ-liikenteelle, sillä nämä ovat tärkeimmät toiminnallisuudet MPLS-verkoissa tänä päivänä. Jos edellä mainittu toiminnallisuus on toteutettu ja mallinnettu suunnittelutyökalussa ja työkalu osaa optimoida liikennettä hyvin saadaan verkon pääoma- ja operaationaaliset kulut laskemaan. MPLS-verkon suunnittelutyökalua laadittaessa on myös tärkeää ottaa huomioon työkalun skaalautuvuusominaisuudet. Runkoverkot voivat koostua tänä päivänä tuhansista solmuista ja sadoista tuhansista liikennevirroista, joten suunnittelutyökalun tulisi omata toiminnallisuutta joka automatisoi joitain vaiheita verkon suunnittelussa, mikä mahdollistaa tämän kokoluokan verkkojen suunnittelun. Tällainen toiminnallisuus voisi esimerkiksi olla automatisoitu vuonohjaus ja verkkojen topologiakokonaisuuden vienti ja tuonti suunnittelutyökaluun ja siitä ulos.

Table of Contents

1.	INTRODUCTION.....	1
2.	MPLS.....	2
2.1.	WHY MPLS?	2
2.2.	MPLS HEADER.....	2
2.3.	LABEL ASSIGNING	3
2.4.	MPLS NETWORK STRUCTURE AND FECs	4
2.5.	LABEL DISTRIBUTION.....	5
2.5.1.	LDP.....	5
2.5.2.	CR-LDP	8
2.5.3.	RSVP-TE	10
2.5.4.	MPLS-BGP.....	12
2.6.	MPLS/DIFFSERV	13
2.6.1.	Benefits of using DiffServ over MPLS.....	13
2.6.2.	DiffServ mapping to MPLS.....	13
2.6.3.	Traffic Engineering in DiffServ	14
2.7.	MPLS VPNs.....	15
2.7.1.	BGP/MPLS VPN.....	16
2.7.2.	Layer 2 MPLS point-to-point VPNs.....	19
2.7.3.	Virtual Private LAN Service	21
2.7.4.	Hierarchical VPLS.....	23
3.	NETWORK PLANNING	24
3.1.	NETWORK PLANNING PROCESS.....	24
3.1.1.	Telecommunications forecast.....	25
3.1.2.	Dimensioning.....	27
3.1.3.	Traffic Engineering	27
3.2.	DIFFERENT PHASES OF MPLS NETWORK PLANNING.....	28
3.2.1.	Forecasting in the MPLS network	29
3.2.2.	SLA Specification.....	30
3.2.3.	Network discovery	30
3.2.4.	Simulation and Failure Case analysis.....	31
3.2.5.	Optimisation, TE and configuration.....	32
3.2.6.	Capacity requirements and network provisioning	32
3.2.7.	Network operations.....	33
3.3.	THE CHALLENGES OF THE MPLS NETWORKS	33
3.3.1.	Scaling challenges of the MPLS networks.....	35
3.3.2.	Functionality challenges of MPLS networks	36
4.	EXISTING MPLS PLANNING TOOLS	38
4.1.	IP/MPLSVIEW.....	38
4.2.	IVNT	43
5.	EVALUATION OF THE ASPECTS OF PLANNING TOOLS	47
5.1.	FUNCTIONALITY IMPLEMENTATION IN PLANNING TOOLS	47
5.1.1.	Forecasting in the MPLS planning tool	47
5.1.2.	Making the network discovery using the planning tool	48
5.1.3.	Simulation in planning tool.....	49
5.1.4.	Optimisation with TE and configuration work in the planning tool	49

5.1.5.	<i>Network Provisioning in a planning tool</i>	<i>50</i>
5.1.6.	<i>Network operations as a part of the planning tool</i>	<i>50</i>
6.	CONCLUSIONS.....	52
	REFERENCES	53

ABREVIATIONS AND NOTATION

ATM	<i>Asynchronous Transfer Mode</i> , cell-based switching technique that uses asynchronous time division multiplexing.
AS	<i>Autonomous System</i> , is a collection of connected Internet Protocol routing prefixes under the control of one or more network operators that presents a common, clearly defined routing policy to the Internet.
BC	<i>Bandwidth Constraint</i> , is a guarantee for some type of traffic to travel in link in a DiffServ aware network.
BGP	<i>Border Gateway Protocol</i> , is the core routing protocol of the Internet. It maintains a table of IP networks or prefixes which designate network reachability among autonomous systems.
CAPEX	<i>Capital expenditure</i> , are expenditures creating future benefits. A capital expenditure is incurred when a business spends money either to buy fixed assets or to add to the value of an existing fixed asset with a useful life that extends beyond the taxable year.
CBWFQ	<i>Class-Based Weighted Fair Queuing</i> , extends the standard WFQ functionality to provide support for user-defined traffic classes.
CE	<i>Customer Edge</i> , is the general way to call the customer's devices which are connected directly to network managed by the service provider.
CoS	<i>Class of Service</i> , is a 3 bit field within an Ethernet frame header when using 802.1Q tagging. The field specifies a priority value of between 0 and 7 inclusive that can be used by Quality of Service disciplines to differentiate traffic.
CSPF	<i>Constrained Path First</i> , is an extension of shortest path algorithms. The path computed using CSPF is a shortest path fulfilling a set of constraints.
CSV	<i>Comma-Separated Values</i> , is used for the digital storage of data structured in a table of lists form, where each associated item (member) in a group is in association with others also separated by the commas of its set. It is widely used to pass data between computers with different internal word sizes, data formatting needs, and so forth.
CT	<i>Class Type</i> , the set of Traffic Trunks crossing a link that is governed by a specific set of Bandwidth Constraints. CT is used for the purposes of link bandwidth allocation, constraint-based routing, and admission control. A given Traffic Trunk belongs to the same CT on all links.
CW	<i>Control Word</i> , is an addition to Layer 2 traffic to prevent missequencing of frames.

DiffServ	<i>Differentiated Services</i> , is a computer networking architecture that specifies a simple, scalable and coarse-grained mechanism for classifying, managing network traffic and providing Quality of Service guarantees on modern IP networks.
DSCP	<i>Differentiated Services Code Point</i> , is a field in the header for packet classification purposes of IP or MPLS packets using DiffServ technique.
ERO	<i>Explicit Routing Object</i> , is a way to distribute the explicit route information to nodes along LSP, enabling resource reservations to be made along explicitly routed paths in RSVP-TE.
FEC	<i>Forwarding Equivalence Class</i> , is a term used in Multiprotocol Label Switching to describe a set of packets with similar and / or identical characteristics which may be forwarded the same way.
FRR	<i>Fast ReRoute</i> , is a Multiprotocol Label Switching resiliency technology to provide fast traffic recovery upon link or router failures for mission critical services.
GoS	<i>Grade of Service</i> , is a probability used in telecommunications to express the quality of voice service.
IGP	<i>Interior Gateway Protocol</i> , is a routing protocol that is used within an autonomous system. RIP, OSPF and IS-IS are examples of IGP.
IP	<i>Internet Protocol</i> , protocol used for communicating data across a packet-switched internetwork using the Internet Protocol Suite, also referred to as TCP/IP.
IS-IS	<i>Intermediate System-to-Intermediate System</i> , is a link-state routing protocol.
LDP	<i>Label Distribution Protocol</i> , is a protocol in which two Label Switch Routers (LSR) exchange label mapping information.
LER	<i>Label Edge Router</i> , is a router that operates at the edge of an MPLS network.
LLR	<i>Least Load Routing</i> , a routing method which selects the least load links for establishing connections.
LSP	<i>Label Switched Path</i> , a path through an MPLS network, set up by a signaling protocol.
LSR	<i>Label Switching Router</i> , is a type of a router located in the middle of a Multiprotocol Label Switching network. It is responsible for switching the labels used to route packets.
MAM	<i>Maximum Allocation Model</i> , is a method of dividing link capacity between different CTs.
MDRR	<i>Modified Deficit Round Robin</i> , is a modified weighted round robin scheduling discipline.

MED	<i>Multi-Exit-Discriminator</i> , is a hint to external neighbors about the preferred path into an autonomous system that has multiple entry points.
MIB	<i>Management Information Base</i> , is a computing information repository used by Simple Network Management Protocol and other implementations.
MPLS	<i>Multi Protocol Label Switching</i> , is a mechanism in high-performance telecommunications networks which directs and carries data from one network node to the next.
MRE	<i>Mean Relative Error</i> , is mean absolute error divided by mean reference trajectory for the total transport distance.
NLRI	<i>Network Layer Reachability Information</i> , is a field used in BGP messaging.
OPEX	<i>Operational expenditure</i> , is an ongoing cost for running a product, business, or system.
OSI model	<i>Open Systems Interconnection model</i> , is a way of sub-dividing a Network Communications System into smaller parts (called layers).
OSPF	<i>Open Shortest Path First</i> , is a dynamic routing protocol for use in Internet Protocol networks.
P2MP	<i>Point-to-Multipoint</i> , refers to communication which is accomplished via a specific and distinct type of multipoint connection, providing multiple paths from a single location to multiple locations.
P2P	<i>Point-to-Point</i> , is a permanent link between two endpoints in a network.
PDU	<i>Protocol Data Unit</i> , a unit of data which is specified in a protocol of a given layer and which consists of protocol-control information and possibly user data of that layer.
PE	<i>Provider Edge</i> , is the general way to call the service provider's devices which are connected directly to customer networks.
PHB	<i>Per-Hop Behavior</i> , defines the policy and priority applied to a packet when traversing a hop (such as a router) in a DiffServ network.
POP	<i>Point of Presence</i> , is an artificial demarcation point or interface point between communications entities.
PSC	<i>PHB Scheduling Class</i> , A PHB group for which a common constraint is that ordering of packets must be preserved.
Q-in-Q	<i>Q-in-Q</i> is an Ethernet networking standard for Ethernet frame formats.

QoS	<i>Quality of Service</i> , is the ability to provide different priority to different applications, users, or data flows, or to guarantee a certain level of performance to a data flow.
RD	<i>Route Distinguisher</i> , is an address qualifier used only within a single internet service provider's Multi-Protocol Label Switching network.
RDM	<i>Russian Doll Model</i> , is a method of dividing link capacity between different CTs.
RT	<i>Route Target</i> , is way to link VPN traffic to a certain VRF.
SAFI	<i>Subsequent Address Family Indicator</i> , is a field used in BGP messaging for identification the set of Network Layer protocols.
SDH	<i>Synchronous Digital Hierarchy</i> , is a standardized multiplexing protocol that transfers multiple digital bit streams over optical fiber using lasers or light-emitting diodes.
SNMP	<i>Simple Network Management Protocol</i> , is an UDP-based network protocol. It is used mostly in network management systems to monitor network-attached devices for conditions that warrant administrative attention.
SONET	<i>Synchronous Optical Networking</i> , is a standardized multiplexing protocol that transfers multiple digital bit streams over optical fiber using lasers or light-emitting diodes.
SPF	<i>Shortest Path First</i> , is a graph search algorithm that solves the single-source shortest path problem for a graph with nonnegative edge path costs, producing a shortest path tree.
SRG	<i>Shared Risk Group</i> , look SRLG.
SRLG	<i>Shared Risk Link Group</i> , refer to situations where links in a network share a common fiber (or a common physical attribute). If one link fails, other links in the group may fail too. Links in the group have a shared risk.
TE	<i>Traffic Engineering</i> , is a common name given to measures which direct the traffic in the network to certain direction.
TED	<i>Traffic Engineering Database</i> , part of IGP's for computing the constraint-based shortest path between two end points.
TLV	<i>Type-Length-Value element</i> , optional information encoded inside of the PDU.
Traffic Trunk	An aggregation of traffic flows of the same class which are placed inside a Label Switched Path.
TTL	<i>Time-to-live</i> , is a limit on the period of time or number of iterations or transmissions in computer and computer network technology that a unit of data (e.g. a packet) can experience before it should be discarded.

VC	<i>Virtual Circuit</i> , a connection oriented communication service that is delivered by means of packet mode communication.
VE ID	<i>VPLS Edge Identifier</i> , is a way to indentify the VPLS edge devices.
VPLS	<i>Virtual Private LAN Service</i> , is a way to provide Ethernet based multipoint to multipoint communication over IP/MPLS networks.
VPN	<i>Virtual Private Network</i> , is a method to link two computers through an underlying local or wide-area network, while encapsulating the data and keeping it private.
VRF	<i>VPN Routing and Forwarding tables</i> , is a technology that allows multiple instances of a routing table to co-exist within the same router at the same time.
WFQ	<i>Weighted fair queuing</i> , is a data packet scheduling technique allowing different scheduling priorities to statistically multiplexed data flows.
XML	<i>eXtensible Markup Language</i> , is a set of rules for encoding documents electronically. It is also used in interchanging data over the Internet.

1. INTRODUCTION

More and more network operators are starting to use next generation network techniques in telecommunication networks. One of the techniques used is MPLS (Multi Protocol Label Switching). MPLS offers fast and reliable routing, compared to plain IP routing used. MPLS also offers a wide range of functionality and high scalability to meet the needs of growing amount of subscribers in the networks in general. Because the size of the networks has increased, the need for good and thorough planning has also increased. The network planners have more and more demanding task at hand. For this reason the need for design tools for MPLS networks has risen.

This thesis studies what kind of things a planning tool for MPLS networks would have to take into account and what functionality it should have for maximizing the help offered to the network planner. The idea for this study came from Tellabs Inc.

The thesis is divided to six chapters including the introduction. The second chapter consists mostly of the techniques used in MPLS. The explaining of the techniques and functionality is important for the reader to understand the wide range of functionality MPLS has to offer and it also gives a good hint of the scale of functionality that has to be taken into account when implementing a planning tool for MPLS. In the third chapter the principles used in network planning are introduced and the ways how they are used when designing a MPLS network are shown. In this chapter the focus is also on the challenges that today's MPLS network faces. In the fourth chapter some the planning tools at the market today are introduced and functionality and attributes of two of them are studied more thoroughly. The fifth chapter is about discussion of desired qualities for planning tools for MPLS networks based on all the earlier chapters. In the sixth chapter are the conclusions.

2. MPLS

In this chapter, the things why MPLS is used nowadays in networks for flow control instead of traditional IP-routing are discussed. After that the concentration is on techniques used in this protocol.

2.1. Why MPLS?

MPLS operates between OSI model layers (which I will call just layers from here on) 2 and 3. Where IP uses routing based on the hop by hop principle, where the header is analyzed at every router individually, in MPLS LSPs (Label Switched Paths) are used. The LSPs are analog to Virtual Circuits in ATM technology and used in similar way to make tunnels between distant nodes. This reduces the amount of calculation needed to be done at every router on packets way, which makes routing of a single packet much faster than in traditional IP-routing. Where in IP-routing the header of the packet needs to be analyzed in every router, in MPLS the header is only analyzed in the entrance and exit of the domain. In the core of the domain only the label is used for forwarding the packet.

Even though MPLS uses LSPs like VCs are used in ATM, MPLS can be integrated to any possible existing network infrastructure as Ethernet, Frame Relay or ATM. This is actually the greatest reason for implementing MPLS technology to already existing networks. Subscribers can be aggregated on an MPLS edge without a need to change any of the devices or access technology and they still get the benefits that MPLS offers. [1]

MPLS offers protection against data spoofing. Because the packet is only analyzed in the head and end of an MPLS tunnel, these are the only places where tempering of data can be done. This is much more secure than in traditional IP-network where data can be added to the packet at every router. [2]

2.2. MPLS Header

MPLS operates between layer 2 and 3 like already mentioned. The MPLS labels are attached to packets which are 32 bits long. The packet is divided to four different fields: label value field, EXP field, S field and TTL field.

The label value field is 20 bits long. MPLS packets are forwarded using this field. This value determines the forwarding equivalence class (FEC).

EXP field was originally reserved for an experimental use, but nowadays it is generally used for class of service implementation. LSRs (Label Switched Routers) and LERs (Label Edge Routers) use this field to decide where in the queue the packet should be placed. Only in couple of cases the label value field is used, instead of EXP field, to determine queuing behavior applied to the packet.

S-bit or bottom of the stack bit is a one bit field which is used when several MPLS headers are stacked on each other.

The Last field is time-to-live field (TTL). This field is the same kind of field which is implemented to IP headers. This field is used to avoid forwarding loops and is also used for path-tracing. The count on the TTL-field is decreased at every hop. If the count reaches zero the packet is dropped. [2]

2.3. Label assigning

Label assigning is done at every node of the MPLS domain. This means that the labels are of local significance only.

LSRs have three kinds of functions to use with labels: PUSH, POP and SWAP. PUSH function can be done for IP packet or MPLS packet. If PUSH is done for IP packet, the procedure is as follows:

1. New label is added.
2. TTL from IP is copied to label.
3. CoS (Class of Service) is assigned exploiting Diffserv.

If PUSH is done for MPLS packet, the procedure goes as follows:

1. Add one more label.
2. S-field is set to 1.
3. CoS is copied from penultimate label.
4. TTL is set to 225.

POP function can be done for one or multiple (stacked) labels. If POP is executed for one label only, it means that the packet is at the egress node of the domain and that the routing continues as IP-routing from that point.

1. Last label is removed.
2. TTL is copied to IP-header
3. Routing is continued with IP-packet

If POP is executed for stack of labels, only the upmost label is removed and routing continues as MPLS switching.

1. Remove the latest label.
2. Option of copying CoS from EXP and TTL-fields.
3. Continue MPLS switching.

The last function SWAP changes the label to another. This function is mainly used in the core part of the MPLS domain.

1. Change the upmost label value to the wanted.
2. S field and EXP fields are copied to the new label.
3. TTL is decreased with 1 and the value copied to the new label.

LSRs are allowed to execute multiple functions for incoming packets. For example with PUSH many labels can be created for single packet or new label can be added and the upmost label changed at the same time. This flexibility brings great amount of functionality to the MPLS.

2.4. MPLS Network Structure and FECs

MPLS network routers are divided to two different types of devices: the core routers and the edge devices. The placement of the device in the network decides the functions and tasks of the device. The edge devices are located at the edges of the MPLS domain and are called Label Edge Routers or Provider Edge (PE) routers. The core routers are known as Label Switching Routers or Provider (P) routers.

When a packet enters a network where MPLS is used, the FEC is decided. This is the only time when the decision is needed to make in a certain domain. This is also the only point where the header is analyzed. After that all the routing decisions are made by using labels. FEC's are represented between two MPLS routers with labels.

If Ru and Rd are two adjacent LSR's, they will make an agreement which label to use to describe certain FEC. This agreement binds a certain label value for Ru's "outgoing label" for certain FEC and certain label value for Rd's "incoming label". The label values are not usually the same between bindings somewhere else in the domain.

In the MPLS architecture, the label binding to a certain FEC is always done by the downstream LSR. With the downstream LSR, the LSR which is downstream with the respect of the binding is meant. [3]

2.5. Label Distribution

After LSR has made a binding of a certain FEC to a certain label, it has to distribute this binding information to all adjacent LSRs. Label distribution protocol also encompasses any MPLS capability learning conversations between two label distribution peers. MPLS architecture does not assume that there would be only one label distribution protocol at use. As this document was written there were two new distribution protocols in use [MPLS-LDP], [MPLS-CR-LDP] and also couple of the routing protocols designed for other use are piggybacked to MPLS [MPLS-BGP], [MPLS-RSVP-tunnels].

2.5.1. LDP

LDP is a set of functions and messages by which LSRs establish LSPs through a network by mapping network-layer routing information directly to data-link layer switched paths. LDP is based on the IGP structure. When LDP is used as signaling protocol, neighboring peers exchange labels and make label bindings independently, and the labels are distributed automatically. This means that in LDP, the traffic engineering cannot be done. Traffic engineering was introduced to LDP in CR-LDP.

LDP discovery is made using UDP messages. There are two different ways to discover other LSRs: basic discovery and extended discovery methods. In the basic discovery method “Hello’s” are sent periodically to the well-known LDP discovery port as multicast messages. These messages carry the LDP identifier for the label space which the LSR intends to use for the interface of the incoming message and possibly some additional information. The second method is so called extended discovery method. This discovery method is for a use of distant nodes. In this method, the “Hello’s” are sent as unicast messages to specific address. The address of the source LSR is included in the message. [4]

After the “Hello”- messages are changed, the LDP session is established. The session establishment is done in two phases: Transport connection establishment and Session initialization. In Transport connection establishment phase two LSRs determine the transport addresses to use in the TCP connection in the Session initialization phase. The LSR which gets the higher address will attempt to establish the LDP TCP connection to another. [4]

After the transport connection is made, the Session initialization phase is started. First thing in this phase is the exchange of LDP initialization messages. These messages include the information about the negotiation of parameters used for the session. The parameters negotiated include LDP protocol version, label distribution method, timer values, VPI/VCI ranges for label controlled ATM, DLCI ranges for label controlled Frame Relay, etc. Successful negotiation completes establishment of an LDP session between two LSRs for the advertisement of label spaces for these two. Active LSR always sends the initializing messages and passive LSR answers to these messages with

“keepalive” messages, which means that the parameters proposed are accepted by the passive node and the session is operational. If active LSR gets a Session Rejected/Parameters Error Notification message as a reply for the initializing message, this means that the parameters are not accepted by the passive node and the TCP connection is closed. The LDP sessions are controlled with “keepalive timers”. If “keepalive timer” has run out and LSR has not received any LDP PDUs, the session is terminated by closing the transport connection. [4]

LDP has two ways for the label distribution. The first one is called Downstream on Demand distribution. In this method LSR distributes FEC label bindings on explicit request from another LSR. The second one is called Downstream Unsolicited distribution. This method allows LSR to distribute FEC label bindings to LSRs which have not explicitly requested those. Both of these methods can be utilized at the same time in a MPLS domain. The principle of the Downstream on Demand is shown in Figure 2.1. Figure 2.2 presents the principle of the Downstream Unsolicited method.

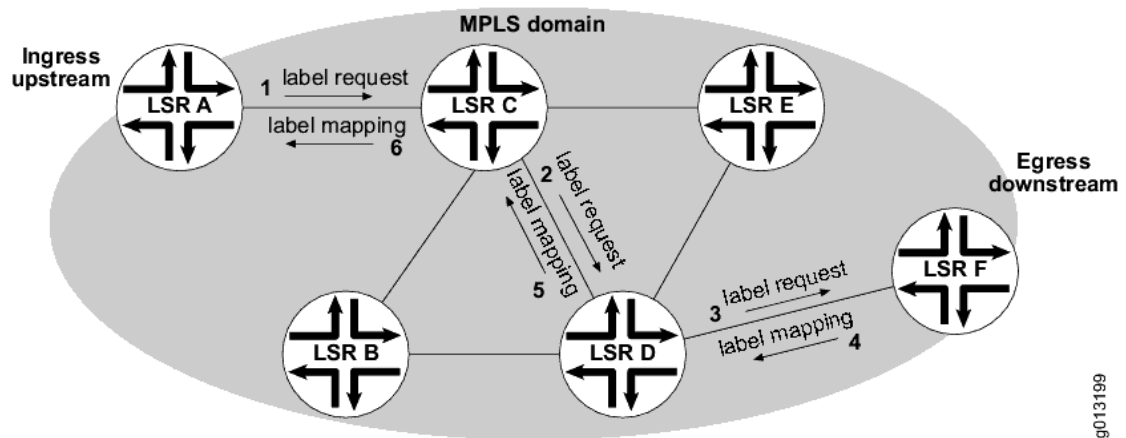


Figure 2.1 Downstream on Demand [5]

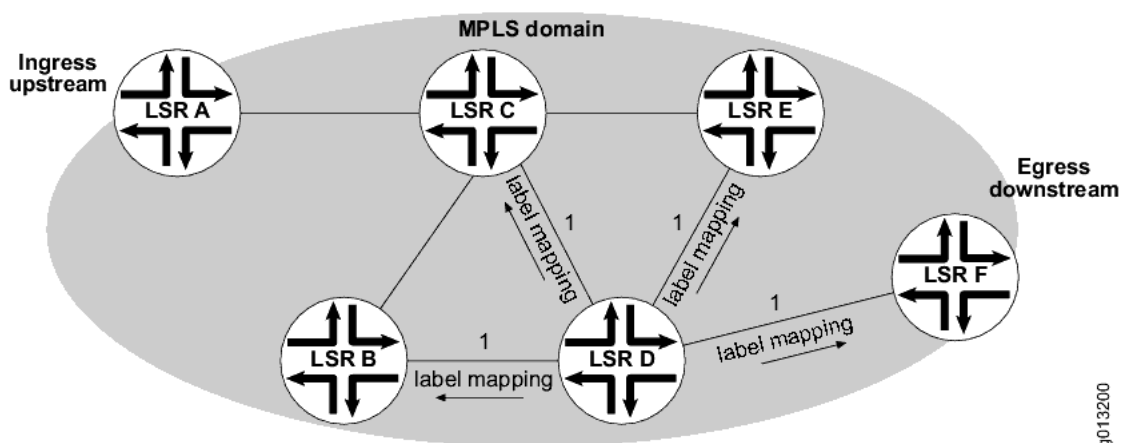


Figure 2.2 Downstream Unsolicited [5]

LDP has two different views for LSR how to deal with label bindings for a FEC learned from a neighbor that is not its next hop for the FEC. The first way is that LSR retain only those advertised label mappings which will be used to forward packets. This

method is called Conservative Label Retention Mode and its principle is illustrated in Figure 2.3.

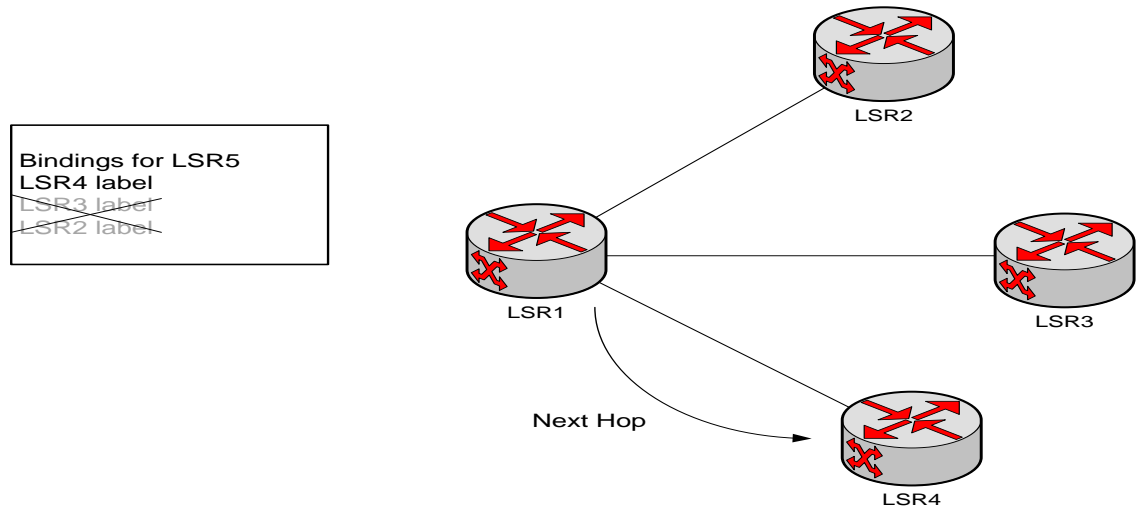


Figure 2.3 Example of Conservative Label Retention Mode [6]

The main advantage for the Conservative mode is that only the labels which are used for forwarding are allocated and maintained. This mode is this why highly used in devices where the label space is inherently limited as in ATM switches.

The second method which LDP uses is Liberal Label Retention mode. In this mode all the advertised label mappings for all routes may be received from all the LDP peers. Every label mapping received is retained regardless of whether the LSR which advertised it is the next hop for the advertised mapping. The main advantage for this method is that reaction to changes can be very fast, because labels already exist. Disadvantages are distribution and maintaining of unneeded label mappings. This mode's principle is shown in the Figure 2.4.

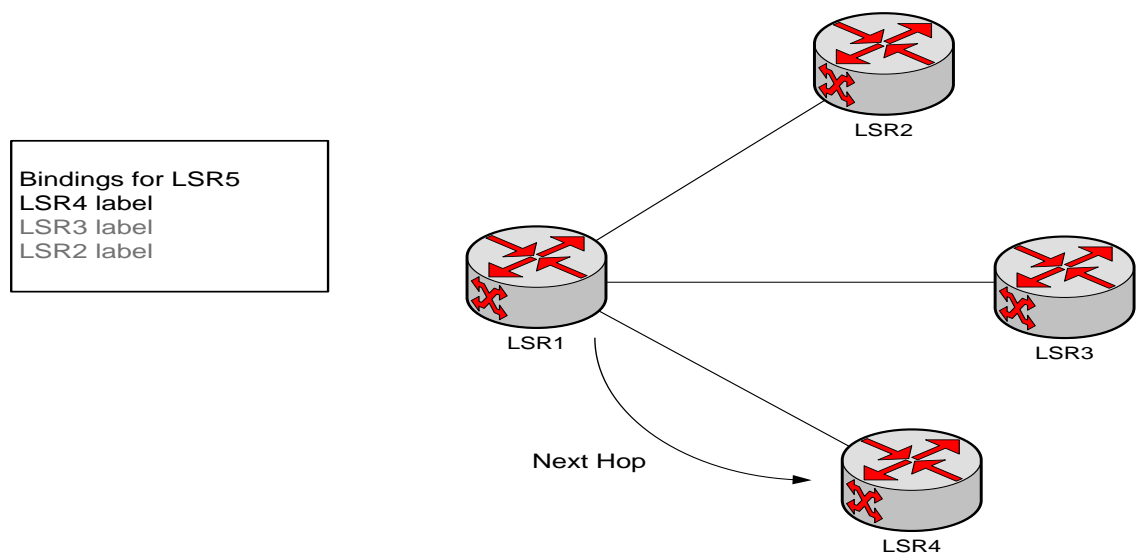


Figure 2.4 Example of Liberal Label Retention Mode [6]

In LDP there is a configurable option for loop detection. This option provides a mechanism for finding looping LSPs and preventing Label Request messages from looping in the presence of non-merge capable LSRs. Loop detection mechanism is utilized via Path Vector and Hop Count TLVs carried by Label Request and Label Mapping messages. If loop detection is wanted to be utilized in MPLS domain, then it should be turned on in all LSRs within that domain. Otherwise the loop detection will not operate properly and some loops will not be detected or some loops will be falsely detected.

2.5.2. CR-LDP

CR-LDP stands for Constrained Based Label Distribution Protocol and it is not actually an independent label distribution protocol but an extension for LDP. CR-LDP brings the possibility of constraint-based routing to MPLS networking which is not accessible with using LDP. With CR-LDP an LSP can be setup based on explicit route constrains, QoS constrains and other constrains. In other words, CR-LDP introduces CR-LSPs to MPLS routing.

CR-LDP works similar way as LDP but it adds TLV parameters to messages sent, which dictate the routes of the CR-LSPs. Some of these parameters are mandatory and some are optional. Some of the TLVs and their type-values are presented in Table 2.1.

TLV	Type
Explicit Route TLV	0x0800
Ipv4 Prefix ER-Hop TLV	0x0801
Ipv6 Prefix ER-Hop TLV	0x0802
Autonomous System Number ER-Hop TLV	0x0803
LSP-ID ER-Hop TLV	0x0804
Traffic Parameters TLV	0x0810
Preemption TLV	0x0820
LSPID TLV	0x0821
Resource Class TLV	0x0822
Route Pinning TLV	0x0823

Table 2.1 Some TLVs used in CR-LDP

Most important of these TLVs used in the MPLS-TE is Traffic Parameters TLV. This TLV is used to signal Traffic Parameter values. These parameters define the link properties which have to be fulfilled with in every link in the LSP. Traffic Parameters TLV is showed in Table 2.2.

U	F	Traf. Param. TLV	Length	
Flags		Frequency	Reserved	Weight
Peak Data Rate (PDR)				
Peak Burst Size (PBS)				
Committed Data Rate (CDR)				
Committed Burst Size (CBS)				
Excess Burst Size (EBS)				

Table 2.2 Traffic Parameters TLV

Frequency field tells at what granularity the Committed Data Rate (CDR) allocated to the CR-LSP is made available. Weight tells CR-LSP's relative share of the possible excess bandwidth above its committed rate. Peak Data Rate field (PDR) defines the maximum allowed rate at which data should be sent to the CR-LSP. Peak Burst Size (PBS) defines the biggest data burst size which can be sent to the CR-LSP. Committed Data Rate (CDR) tells the rate that the MPLS domain commits to be available to the CR-LSP, just like the Committed Burst Size (CBS) the burst size that MPLS domain commits to deliver to the CR-LSP at minimum. Excess Burst Size (EBS) can be used to measure the extent by which the traffic sent on a CR-LSP exceeds the committed rate. [7]

All of these parameters can be negotiated to lower values by any node of the CR-LSP with transforming the parameters in the Label Request messages, if those parameters are flagged with N-flag as negotiable. Once the label bindings are done, negotiation is not anymore possible. The situation is demonstrated in the Figure 2.5.

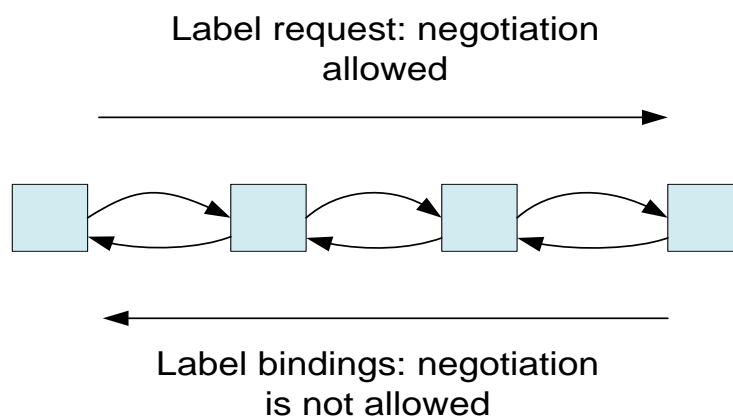


Figure 2.5 Principle of Traffic Parameter negotiation [6]

Prioritization of LSP's is done with optional TLV called Preemption TLV. With this TLV LSP SetupPriority and HoldingPriority can be changed. Both of these parameters can have values between [0,7], 7 being the lowest priority and 0 being the highest.

Because Preemption TLV is optional TLV the default value for both parameters is 4. The comparison is done between new LSP's SetupPriority and the HoldingPriority of a already existing LSP. The LSP which has the lower prioritization can be interrupted.

2.5.3. RSVP-TE

RSVP-TE stands for Resource Reservation Protocol – Traffic engineering. RSVP-TE was not originally meant to use in MPLS label distribution but with certain extension made into the protocol, it is nowadays utilized also in MPLS networking. Hosts and LSRs that support both RSVP and MPLS can associate labels with RSVP flows.

The basic functionality of the RSVP-TE is done with PATH and RESV messages. PATH message is sent by the upstream node. In this message there is a Label Request object inserted inside. This object indicates that a label binding for this path is requested and it also provides an indication of the network layer protocol that is to be carried over this path. The PATH message includes the chance of addition of EXPLICIT_ROUTE object to it. This object is added by the sender node. With this object the paths taken by the label-switched RSVP-MPLS flows can be pre-determined, independent of traditional IP-routing. For detecting loops the sender can also add a RECORD_ROUTE object to the message. Finally, the SESSION_ATTRIBUTE object can be added to the PATH message to aid in session identification and diagnostics. Some additional control information, such as setup and hold priorities, resource affinities and local protection, are included to this object. If a node is incapable of doing label bindings, a PathErr message is created and the information about this node is sent to the sender node. [8]

The RESV message is created by the egress node of the path. This message is sent backwards via the path created by the PATH message. Each node that receives the RESV message containing a Label object uses that Label for any outgoing traffic associated with this LSP tunnel. If the receiver is not the ingress node of this LSP, it allocates a new label and replaces the old label with this new label in the corresponding LABEL object in the RESV message which is then sent to the next LSR in the LSP upstream. Figure 2.6 shows the principle of RSVP-TE LSP founding procedure. [8]

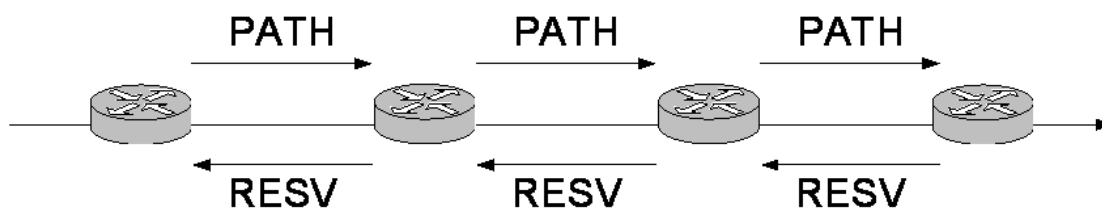


Figure 2.6 RSVP-TE Path finding [6]

In case of link failure, LSPs sometimes need to be torn down. With RSVP-TE two messages are sent in the case of a link break: PATHTEAR and RESVTEAR. PATHTEAR is sent to the ingress by the upstream node of the broken link. Message

includes router notification option and it clears all the soft states in the routers, which means that all the reservations made for this LSP are cleaned from the routers. RESVTEAR message is created by the downstream node of the broken link to the egress node of the LSP. This message also clears soft states of the routers on the route. Figure 1.7 shows the idea LSP tear down when RSVP-TE is used. [8]

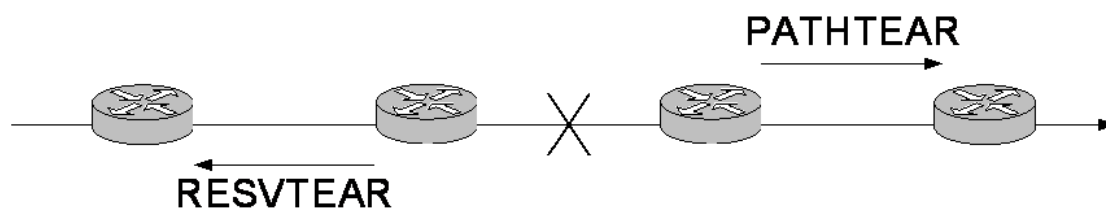


Figure 2.7 RSVP link tear down [6]

Error messages are created in a way as link failure messages. When error occurs in a RSVP-TE driven MPLS network, two messages are sent. PATHERR message is sent upstream towards ingress node and RESVERR towards egress node. Error messages can be generated because of numerous different error types for example “routing problem”, “unknown object class” or “routing error”. These messages are only of advisory type and they do not clear the states of the routers as they pass through them. In figure 1.8 the principle of the error messages is shown.

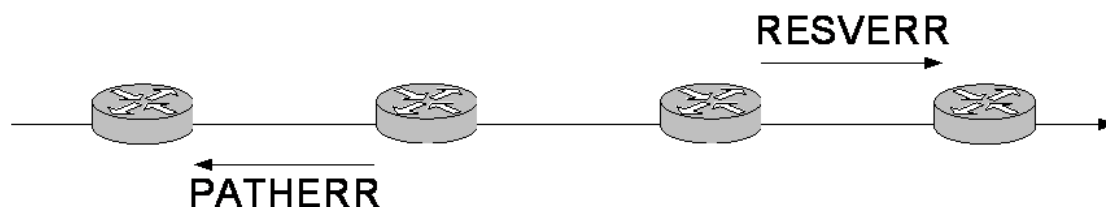


Figure 2.8 RSVP error messages [6]

The greatest gain in using the RSVP-TE in MPLS networks are its traffic engineering abilities. Traffic engineering is done with explicit routes which are created with EXPLICIT_ROUTE objects. This object allows the ingress node to determine a specific route for specific flow to follow. In an explicit route, the nodes can be determined to be “loose” or “strict”. “Loose” nodes have to be on the route but there can be transit nodes between “loose” nodes. The route between two “loose” nodes is calculated using Shortest path calculation. “Strict” nodes have to be directly connected.

Explicit routes can be set manually or automatically. If routes are set manually, it requires a great amount of knowledge about the topology of the network and could lead to mistakes and errors. If explicit route is decided to be set automatically, it is done with TED (Traffic Engineering Database) analysis and all the nodes are “strict”. TED is based on IGP with traffic engineering support. Routing protocols IS-IS and OSPF have been extended to support the use of TED. These protocols are used for distributing

information about the nodes and links and also administrative information. Ingress node uses CSPF with TED to determine explicit route.

RSVP-TE enables the use of administrative groups when creating EROs (Explicit Routing Objects). Each interface in LSR is assigned to a administrative group. These groups are normally divided to “colors” or types, like best-effort or voice, to distinguish between them. Group information is distributed with IGP among nodes. When creating an ERO, CSPF can be configured so that some certain groups can be included and certain can be excluded. Figure 1.9 shows an example of use of administrative groups in a part of a MPLS network.

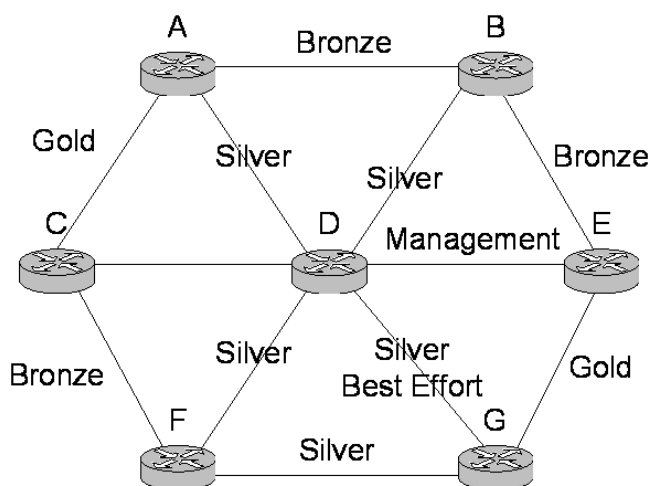


Figure 2.9 RSVP Administrative groups example [6]

2.5.4. MPLS-BGP

BGP (Boarder Gateway Protocol) is a signaling protocol which is designed to be used in traditional IP traffic between different ASs (Autonomous System). With some additions to the protocol, it can also be used in MPLS networks. The label mapping information for the route is piggybacked in the same BGP Update message that is used to distribute the route itself. [9]

There are two advantages for using BGP for carrying label information. The first one is that if two adjacent LSRs are also BGP peers, there is no need for use any other label distribution protocol. The second advantage concerns the interior LSRs, which serve only to carry traffic between the exterior LSRs. The exterior LSRs speak BGP between each others, but the interior router need not receive any routes from the BGP speakers. This is because the BGP speakers have distributed MPLS labels along with each route they distribute with BGP.[9]

Label distribution is piggybacked in the BGP update message by using the BGP-4 Multiprotocol Extensions attribute. The label is encoded into the NLRI (Network Layer

Reachability Information) field of the attribute and the SAFI (Subsequent Address Family Identifier) tells that the NLRI contains a label.

The greatest minus side of using BGP as a signaling protocol for a MPLS network is its lack of traffic engineering features. The creation of constraint routes is impossible with only BGP. The LSRs need RSVP with TE or CR-LDP for creating constraint routes.

2.6. MPLS/DiffServ

DiffServ (Differentiated Services) is an architecture which implements scalable differentiation in the Internet. DiffServ brings scalability to the network and makes it possible to implement QoS for aggregated traffic patterns. Because MPLS itself does not provide QoS for the traffic, DiffServ is also implemented nowadays in MPLS traffic.

2.6.1. Benefits of using DiffServ over MPLS

Using DiffServ over MPLS provides many benefits to the network. The two technologies can be said to fulfill each other in ways that cannot be exceeded with using either of them by itself. Some of the benefits are listed below [10]:

1. Traffic flows can be isolated by using LSPs, which complements the QoS provided by DiffServ.
2. The protection provided by MPLS architecture strengthens the QoS guarantees.
3. MPLS preemption can be used for prioritizing the traffic
4. TE of MPLS networks brings dynamics to the QoS provided by the DiffServ.

2.6.2. DiffServ mapping to MPLS

Per-Hop Behavior (PHB), which means what kind of priority and policy is applied to a packet when traversing a hop, is indicated in DiffServ with a 6 bit value called DSCP (Differentiated Services Code Point). This integer is a part of a IP-header using DiffServ. When importing DiffServ, DSCP had to be imported somehow to the MPLS header, because the LSRs do not read the IP-header at all. This was accomplished by using the EXP field of the MPLS header.

The problem for mapping DSCP to the EXP field of the MPLS header is that the EXP field is only 3 bits long when DSCP uses 6 bits. In [10] two different methods are proposed. The first one called EXP-Inferred-PSC LSP (E-LSP). This method supports DiffServ for 8 or less PHBs. In this method one EXP integer means one PHB. EXP field includes the PHB scheduling class (PSC) and the drop reference field. Mapping rules

are not strict, but are left for the networks administrator to be decided. Table 2.3 shows an example of E-LSP mapping.[10]

EXP Field		PHB
001	---->	AF11
010	---->	AF12
011	---->	AF13

Table 2.3 E-LSP PHB mapping example [6]

Because the EXP field is only 3 bits long and hence can display only 8 PHBs, the support for more PHBs has to be solved with some other mean than E-LSP. For this the Label-Inferred PSC LSP (L-LSP) was introduced. In this method DSCP can be mapped to the label of the MPLS packet itself. During the label establishment the PSC is explicitly signaled to the LSR with using CR-LDP or RSVP-TE, so that after that the PSC can be figured with just looking the label of the packet. EXP field can still be utilized in this method, but this time for solving the drop precedence of the packet. Table 2.4 shows an example of L-LSP mapping.[10]

EXP Field		PSC		PHB
000		DF	---->	DF
000		CSn	---->	CSn
001		AFn	---->	AFn1
010		AFn	---->	AFn2
011		AFn	---->	AFn3
000		EF	---->	EF

Table 2.4 L-LSP PHB mapping example [6]

2.6.3. Traffic Engineering in DiffServ

MPLS-TE itself lacks the functionality for traffic policy. MPLS treats every packet with the same priority. This brings the need for the solution described in [11]. [11] Brings the CT (Class Types) concept to MPLS networks utilizing DiffServ. CTs are usually used when bandwidth is needed to be reserved for certain types of traffic. This can be done because the type of the traffic (for example VoIP) or because of operator's other needs (for example bandwidth promises to clients). With DiffServ a total of eight CTs with different bandwidth reservation shares can be utilized to MPLS networks. LSPs supporting CT are called DiffServ-TE LSPs. In every DiffServ-TE LSP every node keeps individual queues for every CT. If DiffServ-TE is not supported by a node in signaling stage, an error message is sent back to the upstream node and the LSP will be created without a CT value. [11]

Bandwidth distribution and reservation is done in percents. Percent of available bandwidth for a certain CT is called a BC (Bandwidth Constraint). There are two BC models at the moment utilized. The first one is called MAM (Maximum Allocation Model). In this model the available bandwidth is divided between different CTs strictly. Figure 2.10 shows the principle of the MAM BC division.[11]

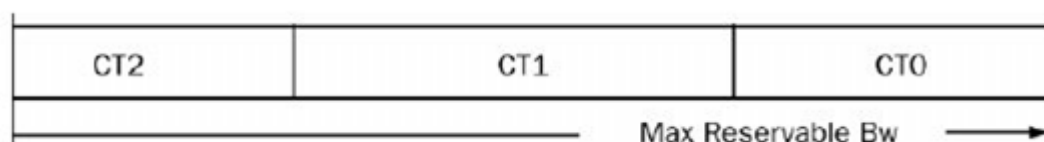


Figure 2.10 MAM principle [6]

In this model different CTs cannot use the bandwidth reserved to another CT but the plus side is that it isolates the CTs well.

The other model used for BC division between CTs is called RDM (Russian Doll Model). In this model the BCs are divided so that they can include more than one CT. Figure 2.11 demonstrates the principle of RDM BC division.

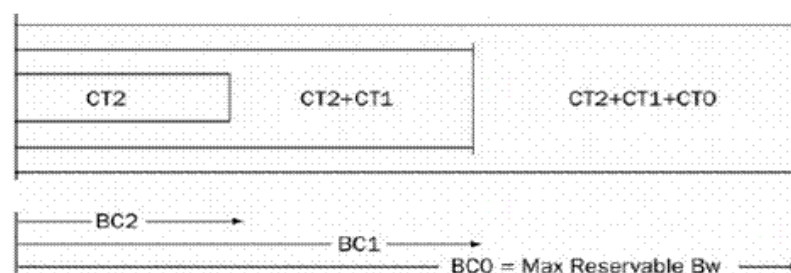


Figure 2.11 RDM principle [6]

The advantages of this method are that the available bandwidth can be better shared with the different CTs and that the network becomes more dynamic when utilizing this method. The biggest minus side is that the traffic is not as well isolated as with using MAM. [11]

2.7. MPLS VPNs

VPN (Virtual Private Network) is a technique which connects two or more client networks to each other. Usually between those networks is a network which is administered by an ISP. VPNs can be carried out in different layers of the OSI model. In MPLS network the VPNs are carried out in the data link layer (Layer 2) or in the network layer (Layer 3). Next the principles of the VPNs of both layers are explained, starting from the Layer 3 VPNs (BGP/MPLS) and moving then on to Layer 2 solutions.

2.7.1. BGP/MPLS VPN

One of the greatest advantages of the MPLS technique comes with the BGP/MPLS VPNs. When talking about MPLS networks, traffic engineering and fast reroute are usually thought as the greatest benefits which come with the MPLS technique. These things are of course major benefits of the technique but the main reason for the provider to deploy MPLS technique to its network is usually the Layer 3 VPNs.[2]

The main advantage of the BGP/MPLS VPNs is that the customer does not have to do any changes for its network to utilize the VPNs and have the same QoS and privacy guarantees with the connected dispersed sites as a private network. Customer can still use the private address space, does not require many configuration changes when adding a new site to its network and also the routing stays simple in the customers network. For the provider BGP/MPLS VPN allows to connect the customer's sites with fairly minimum cost and still have a support for a huge amount of customers/customer sites. [2]

BGP/MPLS is a so called PE-based (Provider Edge) VPN. In these VPNs the CE (Customer Edge) routers do not peer with other CE routers but only with directly attached PE routers. This way of making VPNs has multiple advantages compared to VPN which is so called overlay VPN (CE routers peer with another CE routers):

- Adding a new customer site to a VPN requires configuration to the new site's CE router and PE router only.
- The number of points of control in the network is not necessary increased when a new customer site is added.
- Single infrastructure is used for the service of all VPN customers
- There is no need for exact traffic matrix for provisioning the bandwidth between the customer sites. Instead only the amount of traffic flowing in and out of the site is needed because the provider's infrastructure is used to carry the traffic
- Increasing the bandwidth between sites requires only increasing the bandwidth between the CE and PE. No upgrading circuits or leasing lines is necessary.
- The routing is easier to CE.
- Different customer sites can have different routing protocols

The early PE-based solutions were all based on the IP-header. Even though they solved many problems they still had many downsides mainly because of the IP forwarding:

- The use of private address is precluded. Routers in the provider's network cannot distinguish the traffic of different VPNs if the destination IP addresses are allowed to be the same.
- The default routes cannot be used because the differentiation between default routes of different customers cannot be done.

- The scalability of the solution is limited. If the provider's routers must maintain the state for all the VPN routes, the maximum amount of VPN routes is limited to the forwarding table size of the core routers.

The solution for these problems arrived with BGP/MPLS. BGP/MPLS uses MPLS tunneling to overcome the problems discussed above. BGP/MPLS VPN utilizes the labeling to make tunnels between two PE routers.

With BGP/MPLS VPN solution the main goals are to isolate the traffic between the different VPNs and connect the customer sites with keeping the possibility to use private addresses spaces in each site in the manner that the addresses can overlap with each other. In BGP/MPLS VPNs the PE routers keep different routing tables for each VPN route. These tables are called VRFs (VPN routing and forwarding tables). VRFs is the main tool for achieve the goals set for the BGP/MPLS VPN.

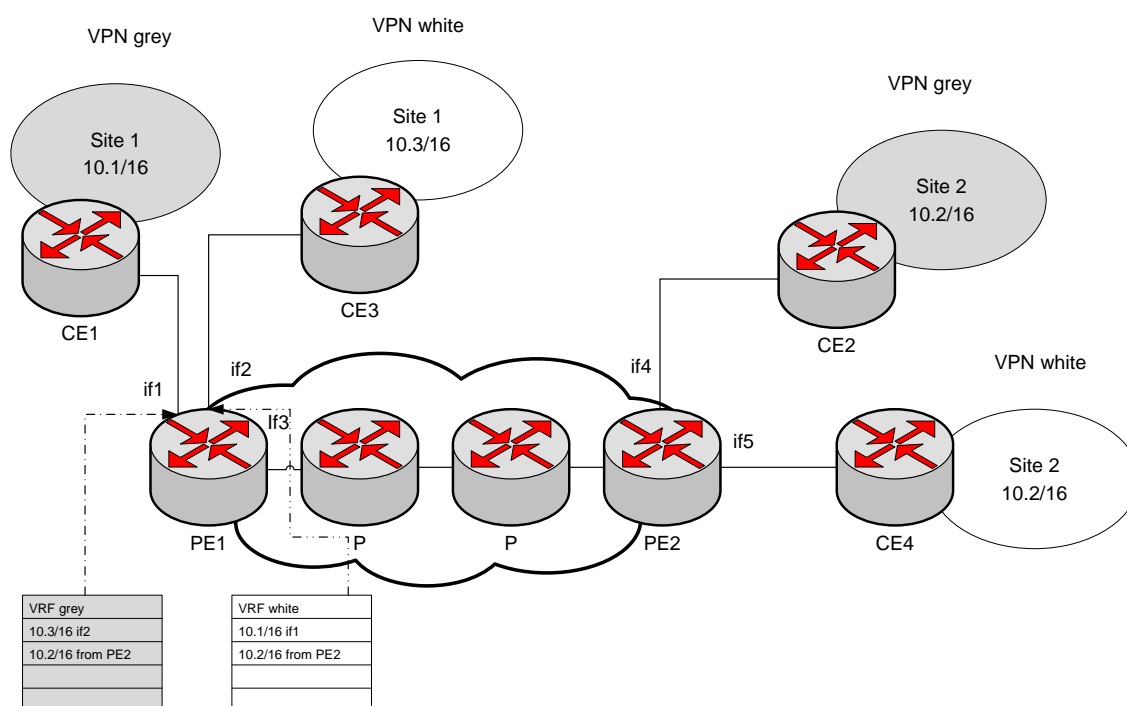


Figure 2.12 Network with two customers

In a Figure 2.12 the main principle of VRFs is shown. In this imaginary network there are two VPNs in use: grey and white. In each VRF, the first row shows the name of the VRF. The lines below that, tell the network which is tied to the VRF and the source where the information from that was learned. The VRFs are tied down to a certain interface (the interfaces can be also logical as VCI/VPI or VLAN) for a PE1 to know which table to use when an IP packet comes from customers site. This is much more scalable solution than to use individual PEs for each customer sites. [2]

To preserve private, and thus overlapping, address spaces when carrying VPNs MPLS/BGP solution needs to moderate the IP addresses used. BGP speakers can only install one route to given address prefix. This why the IP address is modified with a

prefix called the RD (Route Distinguisher). When RD is attached to an IP address, the new address is referred as the VPN-IPv4 address. Before advertising a customer VPN route in BGP, the PE routers attach to it the appropriate RD for the VPN site, transforming it into a VPN-IP route. When receiving a VPN-IP route, the PE converts the route back to plain IP by removing the RD. The RD itself is an 8-byte quantity consisting of three fields: a two-byte type field, an administrator field and an assigned number field.

Routing information distribution of VPNs is needed to be constrained to minimize the routing traffic. The requirement is broader than simply separating the routing information per VPN, for two reasons: customers may require arbitrary and complex connectivity models between their sites and support for overlapping VPNs means that the same route must be present in several VPN routing tables. For BGP/MPLS VPNs, this is done using BGP extended communities. Extended communities are 48 bit long and it consist two parts. The first part is 16 bits long and comes from the AS number allocated to the service provider. The second part is 32 bits long and is decided by the service provider. The second part is generally called the RT (Route Target). The RT is what accomplishes the constrained route distribution between PEs and ends up defining the connectivity available between the VPN sites. If the routes advertised do not match the RT they are discarded. But the routers can ask them again with BGP's route-refresh capability if the need arises.

VPN routes are distributed as VPN-IP prefixes between the PEs using BGP. The next hop of such route is the address of the advertising PE. The traffic must be tunneled between PEs, for two reasons: the P routers have no information on the VPN routes and the BGP information is for VPN-IP addresses, which are not routable. MPLS provides a great way to create these tunnels: associate a label (the VPN label) with a VPN route. When the packet is forwarded at the ingress PE, the VPN traffic is labeled with the VPN label and sent to the egress PE. Based on the VPN label, the egress PE can demultiplex the traffic to the correct VPN. With BGP the distribution of the label is done along with the VPN route information. MPLS makes sure that the P routers do not have to maintain separate state for each one of the PE-PE VPN tunnels. This is done with label stacking, which allows the creation of a hierarchy of tunnels. The VPN tunnel label is put in the bottom of the stack and the PE-PE tunnel label is put on top of VPN tunnel label. Forwarding is done always with the top label only, so the P routers need not maintain any state regarding the VPN tunnels. The forwarding at egress PE can be done in two fashion: an MPLS lookup on the VPN label to determine the appropriate VRF, followed by an IP lookup in that VRF or an MPLS lookup based on the VPN label, in which case the label provides an outgoing interface. Most vendors support the both types of lookup and let the user decide between the two through configuration.[2]

BGP/MPLS solution offers lots of benefits for both, the provider and the customer, sides. BGP/MPLS VPNs allow the customer to offload routing between the sites to the provider and enable the service provider to offer better services to its customers, such as firewall and authentication. For provider the BGP/MPLS VPN approach allows to

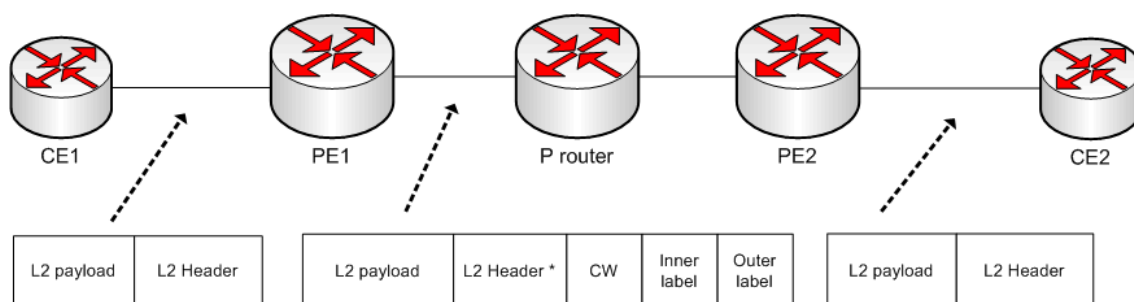
leverage the infrastructure to service multiple VPN customers, rather than managing a virtual backbone for each customer. PE-PE MPLS tunnels are used to carry the traffic between different customer sites. With this, the complexity is kept at the PE routers and the service is very scalable. Tunneling using MPLS also enables good hierarchy of routing, traffic identification of different VPNs at the egress PE and straightforward and low-cost protection against packet spoofing.[2]

2.7.2. Layer 2 MPLS point-to-point VPNs

Connecting customer sites with VPNs is also possible on Layer 2 of the OSI model (e.g. ATM, Ethernet and Frame Relay). Differences between Layer 2 and Layer 2 VPNs are as follows:

1. In the Layer 2 case, no routing interaction is done between CE and PE.
2. In the Layer 2 case, the customer can run any Layer 3 protocol between sites. The SP network is simply transporting Layer 2 frames and is unaware of the used Layer 3 protocol.
3. Multiple (logical) interfaces between each CE and the corresponding PE are required in the Layer 2 case, one per remote CE that CE needs to connect to. In the Layer 3 case, one connection between each CE and the corresponding PE is sufficient as the PE is responsible for routing the traffic towards the appropriate egress CE.

There are two main approaches with Layer 2 VPN solutions: one involving LDP signaling and other based on the BGP signaling. In the forwarding plane the two approaches are the same, in terms of how Layer 2 frames are encapsulated for transport across the MPLS network. However these two approaches differ a great amount on the control plane.



* Which parts of the layer 2 header are transported over the MPLS core depends on the layer 2 protocol

Figure 2.13 Forwarding plane operation of L2 transport over MPLS

A single point-to-point Layer 2 connection provided over an MPLS network is generally called a pseudowire, to convey the principle that as far as possible the MPLS network should be invisible to the end customer, in such way that the two CEs

interconnected by the pseudowire appear to be directly connected back to back. Figure 2.13 shows a cross-section through an imaginary MPLS network. Packets are pseudowired between CE1 and CE2. The following forwarding operations are carried out by PE1, when the Layer 2 frame arrives there:

1. Parts of the Layer 2 frame which are not needed to be transported are removed from the frame.
2. In some cases, a four-byte Control Word (CW) is prepended to the Layer 2 frame. The Control Word can include a sequence number so that the egress PE can detect mis-sequencing of packets. The Control Word may also contain flags corresponding to control bits within the header of the native Layer 2 frame. This allows the value of those control bits to be conveyed across the core of the network to the egress PE without the need to transport the entire native Layer 2 header.
3. PE1 looks the value of the MPLS inner label that PE2 expects for the frame and prepends an MPLS header having that label value.
4. PE1 determines how to reach PE2. As with L3 VPN, the network operator has a choice of tunneling technologies in the core, including LDP and RSVP-signaled LSPs and Generic Routing Encapsulation and IPSec tunnels. If an LDP or RSVP-signaled LSP is used, PE1 determines the MPLS label value required to reach PE2 and stacks an MPLS header containing that label value on top of the inner MPLS header. In networks where MPLS transport is not used between PEs, PE1 determines the appropriate tunnel to reach PE2.
5. When PE2 receives the packet, it examines the label value of the inner MPLS header before popping it. From that, it determines that the underlying L2 frame must be sent on certain interface to CE2 (for example VLAN XX). If the Control Word is present, PE2 may check the sequence number and take appropriate action if the packet is out of sequence. The processing of the sequence number by the egress PE is optional. Actions taken by receiving PE when it receives packet out of sequence are to drop the packet or reorder the packets into the correct sequence. PE2 then regenerates the L2 frame, which may involve determining the values of control bits in the frame header by referencing the corresponding flags in the Control Word.

Control plane operations for Layer 2 VPNs differ between the used distribution protocol (LDP and BGP), but share some common characteristics:

1. When forwarding traffic from a local CE through a remote PE to a remote CE the protocol offers a way to know the value of the VPN label (inner label) that the remote PE expects.
2. They offer a means for signaling characteristics of the pseudowire, such as media type and MTU.

3. Pseudowire formed is bidirectional. Hence, if there is a problem with transport in one direction, forwarding is not allowed to occur in the opposite direction.
4. They offer the means for a PE to indicate to remote PEs that there is a problem with connectivity, for example if the link to a CE goes down.

The differences in between the protocols are seen in the pseudowire creation phase: in LDP, the information of which remote PE(s) the PE needs to build the pseudowire is not automatically solved, but in the BGP scheme there are auto build discovery properties available.[2]

2.7.3. Virtual Private LAN Service

Virtual Private LAN Service (VPLS) is a way to offer Ethernet like multipoint-to-multipoint like connection between customer's sites. In other words, it enables the service provider's network to appear as an LAN to the end-user. The reason for utilization of this service is that compared to the BGP/MPLS mechanism customer does not need to configure any routing protocol or static routes to run between CE and PE. In comparison to L2 VPN, the customer does not need to build an overlay network with point-to-point connections provisioned by the service provider which needs great amount of expertise of that field. VPLS offers easy-to-use service for the customers to interconnect their equipments as they were in the same LAN. Also a mixture of different access media can be used within the same VPLS service instance, to cater for different types of sites that the customer might have.[2]

In VPLS the forwarding decisions are made with the MAC (Media Access Control) address of the destination. Forwarding decision is made in the ingress PE of the provider's network. For this reason, the PE routers are fully meshed with pseudowires. This is so that a PE receiving a frame from another PE can identify which VPLS the frame belongs to, on the basis of the pseudowire label. The tunneling between the PEs is typically done with LDP- or RSVP-signaled LSPs, but also GRE or IPSec tunnels can be used. Each PE maintains separate forwarding tables for each VPLS. For example, if a PE is connected to customer X's site and to customer Y's site, it keeps two separate forwarding tables, one for each customer. [2]

How the forwarding of the VPLS traffic is done, depends on the frame format, in other words are the frames unicast type or are they broadcast or multicast frames. If the frames are unicast type, the forwarding method depends on the ingress PE's forwarding table of wanted VPLS. If the ingress PE does not know the location of the destination's MAC address, it floods the packet to all ports, excluding the arrival port of the packet. Because all of the PEs are fully meshed with pseudowires, all of the PEs which have a member of the wanted VPLS attached receive the flooded packet and this also prevents the forming of the forwarding loops. Because the packet is forwarded with certain pseudowires, the PEs now know which VPLS forwarding table to look for the MAC address (certain pseudowire, certain VPLS forwarding table). If the PE does not

find the location of the wanted MAC address in the forwarding table, it floods the frame to its local ports facing the customer and belonging to the wanted VPLS. Also when the frames arrive to each PE, they add the sender's MAC address and the next hop which they get from the arrival port of the packet. When the ingress PE sends the frame to another PE it pushes two labels into the frame: the MPLS transport tunnel label (the outer label) and the pseudowire label of the VPLS (inner label). The outer one directs the packet to correct LSP and the inner one informs the VPLS used. At the time of the writing the multicast and broadcast of the frames in the VPLS work in similar manner (no RFC done about this). In both cases the ingress PE floods frames towards all sites of the VPLS, in other words floods the frame to all ports in certain VPLS. When other PEs receive the packet, they flood it to all ports facing the CEs which belong to that certain VPLS. [2]

Talking about the control plane mechanism, there are two different aspects to take in notice. The first one is how does the PE knows what other PEs there are connected in the same VPLS, in other words how the discovery happens. The other one is how the full mesh of pseudowires set up is done between the PEs, in other words the signaling aspect. At the moment there are two different schemes for doing the both tasks: LDP and BGP scheme. [2]

The LDP based scheme is pretty similar to the LDP scheme used in the point-to-point Layer 2 connections. In order to signal the full mesh of pseudowires required, a full mesh of targeted LDP sessions are required between the PEs, or at least each pair of PEs that have VPLS in common. This is if H-VPLS (Hierarchical VPLS) is not used. The only thing that is different compared to the scheme used in the point-to-point scheme is the usage of the VC ID field of the frame. In the point-to-point connections the VC ID field is used for identify a particular pseudowire. The value of the VC ID field is configured in this scheme to be the same for a certain VPLS instance on all PEs. This way the VC ID allows a PE to identify which VPLS instance the LDP message refers to. LDP also lacks the in-built auto discovery mechanism, so the LDP sessions must be manually configured to each PE router. This session is used for communicating the value of the inner label that must be used for each pseudowire. Other option is to use some external auto discovery mechanism. Some mechanism, like Radius or BGP are proposed for this use, but at the moment LDP itself still lacks an auto discovery mechanism.

The BGP signaling and auto discovery scheme is also pretty similar in VPLS as it is in Layer 2 VPN and in BGP/MPLS VPN. It fundamentally has the following components:

- A means for the PE to know which PEs are members of a certain VPLS, in other words it has an in-built auto discovery.
- A means for a PE to know the pseudowire label expected by a given remote PE for a given VPLS. This is known as signaling.

The functionality is similar to the BGP scheme used in L2 and L3 VPN. On each PE an RD and an RT are configured for each VPLS. The RT is same for a particular VPLS across all PEs and used to identify the VPLS to which the incoming BGP message is related. RD is as in L2 and L3 VPN schemes. [2]

On each PE, for each VPLS, an identifier is configured. This identifier is called the VE ID (VPLS Edge Identifier). Each PE involved in particular VPLS must be configured with a different VE ID. BGP is used to advertise the VE ID to other PEs in the network. This along with the other information passed on in the NLRI, provides the means to calculate the value of the pseudowire label required to reach the advertising PE. [2]

2.7.4. Hierarchical VPLS

Hierarchical VPLS (H-VPLS) is a scheme added to LDP based Layer VPN to solve the problem caused by the thing, that a full mesh of LDP sessions is required between PE routers. This causes in all but the very smallest of deployments burdensome administrative overhead. H-VPLS gives some help to this problem.[2]

In H-VPLS the access device is treated as a PE and pseudowires are provisioned between it and every other, as a basic VPLS. Another option is to utilize pseudowires or Q-in-Q logical interfaces between the access device and the selected PE. The VPLS core pseudowires (hub) are augmented with access pseudowires (spoke) to form a two tier hierarchy. [12]

3. NETWORK PLANNING

In this chapter, the concentration is on the planning process used in network designing generally, on the phases used in the MPLS network planning and in the challenges MPLS has to face as a network technique today.

3.1. Network planning process

Before launching a new telecommunication network, it is very important to plan different aspects of the new network as well as one can. The network planning consists of five different layers, which all have to be properly done to have a properly functioning network. The five different layers are listed below:

- Business Planning
- Long Term and Medium Term Network Planning
- Short Term Network Planning
- IT Asset Sourcing
- Operations and Maintenance

To network planning to be successful, the first step to be done is the acquisition of external information. This information includes the followed things:

- Forecasts of how the new network or service will operate
- The economic information (costs and tariffs)
- The technical detail of the network's capabilities

Before the network planning can start, the people responsible must have made the choices of the used protocols and transmission technologies.

After these initial steps have been made, the network planning process consists of three different phases. The phases are listed below:

- Topological Design
- Network-Synthesis
- Network Realization

Topological Design involves the actual placing of the wanted components and how to connect them to form a network capable of transmitting traffic a predetermined

amount as low cost as possible. In this phase the Graph theory is widely used to determine the costs of switching and transmissions to find out the optimum placing for the switches and concentrators. Network-Synthesis phase involves determining the size of the components used, subject to performance criteria such as the Grade of the Service. In this phase the typically used method is called the nonlinear optimization. It involves determining the required Grade of Service, cost of transmission and some other parameters. Information gathered from the previous phase is used to calculate a routing plan and the size of the components used. Network Realization stage determines how to meet the capacity requirements and ensures the reliability within the network. In this phase the commonly used method is called the multicommodity flow optimization which involves determining the information relating to the demand, costs and reliability. This information is then used for calculation of an actual physical circuit plan.[13]

3.1.1. Telecommunications forecast

Forecasting of the planned network's traffic intensity and traffic load is important for its planning and design phase. There are two different methods mainly used to form a forecasting model of the network. The first one is to take a similar network already utilized and make the forecasts based on the measurement data got from this network. However, this "comparison network" is rarely available and because of this the method of telecommunications forecast is used to estimate the expected traffic intensity. Telecommunications forecast method includes the following steps:

- Definition of the problem
- Data acquisition/preparation
- Choice of the forecasting method
- Forecast analysis
- Documentation and analysis of the result

The first phase is typical for any forecast process. It basically means that the planner must be sure of the problem he is trying to find the solution. This phase may sound unnecessary or not needed but this is actually as important as the rest of the phases. The target of the forecast must be defined as specifically as possible to get the forecast model to simulate the future network's traffic properties precise enough.

In the data acquisition phase the gathered data is needed to be prepared for the forecast process. If the data contains errors, the forecast will also. This process is commonly known as "scrubbing" of the data. Scrubbing the data concentrates on removal of certain data points, which are called the "outliers". The outliers are data that lie outside the normal pattern. These data points are caused by anomalous events

and are unlikely to recur. Removing these data points improves the “data integrity” and so increases the accuracy of the forecast.

For the forecast itself there are several methods in use. The list below lists some of the methods:

- Judgment based methods
 - Delphi method
 - Extrapolation
- Survey Methods
- Time Series methods
 - Exponential smoothing
 - Cyclical and seasonal trends
 - Statistical models
- Analogous methods
- Causal models

The judgment based models are based on the knowledge and experience of the experts of the area that is being conducted. In the Delphi method a series of questions is directed to experts of the area where the forecast is wanted. The forecaster makes a summary out of the answers and sends it back to the experts to be reanalyzed. If the experts are not satisfied with the forecast, they can revise their opinions. This method is considered to be highly unreliable way of making a forecast. Extrapolation is a usual method used in forecasting generally. This method relies in the information acquired from the subject wanted to be forecasted. The data acquired from the previous events is plotted and if some kind of pattern is found, the pattern is tried to be extended into the future moments also, and so the forecast is formed. In extrapolation there are different extrapolation rules (for example S-shaped logistic curve, Gompertz curve or the Catastrophic Curve) one can follow to make the forecast. The rule to be followed is decided by the maker of the forecast.

The survey methods are based on the opinions of the customers. In order for the survey to be successful, first the target group has to be recognized. Once this is done, a sample must be chosen. The sample is a subset of target group and has to be chosen so that it reflects everyone in the wanted target group. The survey is formed by series of questions for the sample group and the answers are recorded. The answers got from the survey must be analyzed with statistical and analytical methods. The results of the analysis should be checked after this with some other forecast method to get a better probability for the survey forecast to be true and accurate.

Time series method uses data samples measured periodically. With these samples a model of the subject is made and it can be extrapolated to the future. Time series method can be used for many different purposes depending on which set of assumptions is used in the forecast. Exponential smoothing is based on the moving average of the data and can be used for example in sales for sales mean forecasting. Cyclical and

seasonal trends method searches from the data acquired some patterns or trends which have repeated in cyclic or seasonal periods. The makers of the forecast use the current data to adjust the pattern to better fit in this period's data. With the values got from this are then processed to forecast for the remainder of the current season or the cycle. Statistical model is based on the statistical relationship of different variables. Models are based on the current data, which extrapolated to the future. Erlang B and C formulae are examples of statistical model.

Analogous methods concentrate on finding similarities between foreign events and events which are being studied. The foreign events usually have more data samples than in the studied events. Analogous methods are divided to two groups, which are qualitative models and quantative models. In qualitative method the comparison between the foreign events is made in way that it forms a symbolical similarity. In quantative method the quantity is compared and forecasted.

Causal models are the most complex of the forecast methods used, but also usually the most accurate. In this method the goal is to create a all-inclusive model of the true situation which takes all the parameters and variables in notice. With the calculation power of the modern day computers this method has become more and more "competitive" solution to be used in forecasting. Especially in telecommunications forecasting this method is highly used.

3.1.2. Dimensioning

The main purpose for this phase of the network planning is to determine the minimum capacity requirements that allow the wanted GoS level to be achieved. For this reason the dimensioning is made with the maximum traffic thought to be generated in the network for a certain moment. This moment is generally called the peak hour. [13]

Dimensioning process has many components under it. Dimensioning consist of determining the network topology, making a routing plan and traffic matrix and making the GoS requirements. The information received from these sub-phases is then used for to determine the maximum capacity of the switches and the maximum number of channels required between the switches. [13]

A thumb rule for dimensioning is that the planner must ensure that the traffic load should never reach the load of 100 percent. With this rule in mind, the planner must continuously maintain and upgrade the resources to meet the changing requirements of the network. This is done with ongoing measurements of the networks traffic and with some forecasting method. [13]

3.1.3. Traffic Engineering

If the network engineering consists of adding the components like links, routers and switches to the network, the traffic engineering concentrates on controlling the path taken by traffic through the network. There are many reasons why the network operators

want to influence the path taken by the traffic in their network. The most popular reason for TE today is improving utilization of network resources. The goal for the TE is to avoid a situation where the parts of the network are congested while others are underutilized. Other reasons for using traffic engineering include ensuring that the path has certain wanted characteristics, ensuring that transmission resources are available along a particular path, and determining which traffic gets priority at time of resource crunch. [2]

3.2. Different phases of MPLS network planning

MPLS networks, like all the different types of networks, have many different planning phases. In this chapter those planning phases and what they mean in the MPLS scheme are discussed.

MPLS network planning can be thought of a sum of many fields of planning. Figure 3.1 shows different fields of MPLS network planning. All the phases are divided to smaller parts. Traffic forecasting is created on the base of traffic demand matrix and service subscription. The Network Discovery phase is a sum of physical and logical topology and the QoS model is also decided in this phase. Physical topology means the devices and links of the network designed and logical topology is the arrangement of the devices in the network and how they communicate with each other. SLA specification is based on the SLA requirements, which are mainly agreed in cooperation with the customer. After these phases are planned, the network planner can do his/her first simulation run, to do analysis how the network is behaving. After the first simulation network planner can go on with optimization of the network, as well as with traffic engineering which can balance and optimize the traffic delivered in the network. This phase helps also the next phase which contains discovering the capacity requirements and provisioning the network elements to services of different customers. The last phase is network operation which basically means the supervision, monitoring and maintaining of the network designed. After these phases, the first model of the network is planned.

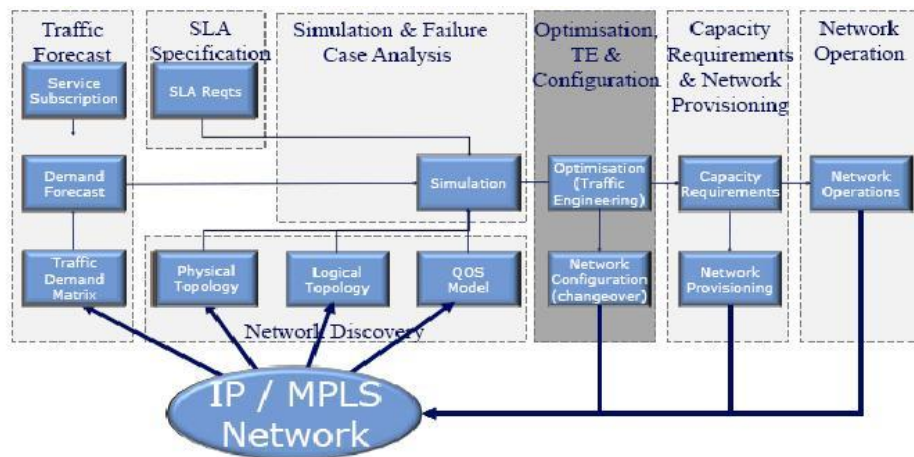


Figure 3.1 Phases of a Network planning [14]

3.2.1. Forecasting in the MPLS network

Like mentioned before the forecasting phase of the network planning is done with traffic demand matrix (which is generally called a traffic matrix) and pricing of the service subscription.

The decision of choosing the traffic matrix used to model the traffic of the MPLS network is not an easy or straightforward task. The first thing that makes it hard is the wide selection of different model types. The second thing is that the MPLS networks can be very different. The differences come in a form of size (quantity of nodes and links and subscribers served), shape (the topology model used), functionality, and traffic types and so on. Based on the study made on the traffic matrix estimation on a large MPLS/IP backbone [15], the traffic matrix estimate model which models the traffic best was very situation dependent. In the study, traffic information was gathered from two different backbone MPLS networks, one located in American continent and another in Europe. The traffic matrices for the networks were generated using a simulation of the network. In this case the simulation was done with designing tool MATE done by Cariden and the data gathered was then converted to a routing matrix. The result performance evaluation of different traffic matrices was based on MRE (Mean Relative Error). The best traffic matrix modelling method was different in the both of the networks, the entropy method giving the best MRE in the American network and the Bayes method giving the best MRE in the European network. [14]

The service subscription phase defines the service level which can be given to a certain customer. The network has a limited amount of resources which can be divided among the customers. Service subscriptions include many parameters which dictate the services which can be offered to customers.

3.2.2. SLA Specification

SLA (Service Level Agreement) is defined when the contract between the customer and the service provider is signed. SLAs in general can be very different, but in networking it usually is based on the same attributes. In MPLS case the SLAs require broader range of service level metrics than more traditional technologies. The SLAs can vary by service provider within a region and region-to-region. Even if the SLAs can vary, they usually include some key metrics which can be found in the most of the SLAs. The metrics are site-to-site packet delivery, site-to-site round trip delay, jitter, site availability and maximum time to repair.

The next example is from AT&T's SLA applicable to the Dedicated Internet Access family provided by AT&T Internet Services and/or AT&T Long Distance, and it shows one example how the metrics can be defined in MPLS based network. [16]

- A. Network availability: This defines the percentage of time the IP/MPLS Backbone Network shall be available to deliver traffic to/from other AT&T POP locations on the IP/MPLS Backbone measured over a calendar month.
- B. Network Latency: This defines the roundtrip POP-to-POP latency on the IP/MPLS Backbone Network.
- C. Network Packet Loss: This defines the percentage of monthly average packet loss between AT&T POPs on the IP/MPLS Backbone.
- D. Off-Net Performance (AT&T -KB40): This defines the performance level needed to have from outside the AT&T maintained network.
- E. Internet Service Availability: This defines the percentage Dedicated Internet Access Service shall be available.
- F. On-Line Information: This defines where the all SLA computations, methodologies, and credit requests are available for customer to survey.
- G. Scheduled Network Maintenance: This defines how and when the maintenance of the network equipment software, network equipment hardware, or Network capacity is carried out.
- H. Emergency Network Maintenance: This refers to efforts to correct network conditions that are likely to cause a material Service outage and that require immediate action.
- I. Force Majeure: This frees AT&T and the customer from the responsibility in the case of acts or occurrences beyond their reasonable control

3.2.3. Network discovery

Network discovery phase is divided to physical topology discovery, logical topology discovery and QoS discovery. Physical and logical topology is often added in the same phase and with using the same tool of addition. The network topology can be very versatile in networks which utilise MPLS technique because of the different nature of

routers utilised and the different network layers used. Also the CE devices can be very different, so the SE devices need to be very functional.

Network topology discovery process is usually made one of two different methods. In the first method, the discovery is done by adding one equipment or logical link at the time. First the POPs are added, and after that, the links connecting the POPs are inserted to the topology. After that the logical topology is generated. This one by one addition is generally done via GUI in most of the planning tools at the moment. When the devices are added one by one, the entire configuration has to be inserted manually to the devices or logical links in the network. This configuration consists of device interfaces, device types, inventory details, protocols and technologies running on the device (if not preconfigured to the device model), virtual connections, route objectives, and so on[17].

The other way to do the network discovery process is to import the wanted network topology as whole with using some kind of importing feature of modelling/planning or management tool for MPLS networks. The tool gets the topology information from some other management tool or from some file which has the topology information saved in it in format which is can be accessed with the modelling/planning tool. This is the mainly used method in current real life MPLS networks, mainly due its scalability. Forming the network topology in a case of wide network with thousands or more POPs is not possible with adding one element at time, or in a reasonable time period anyway. The topology and configuration info of MPLS network, as networks in general, is stored in database of some kind. Usually networks nowadays use so called automated discovery, where the info of the POP is automatically send to the database, which is called MIB (Management Information Base) in general, when the state of the element is changed or new element is added to the network using SNMP (Simple Network Management Protocol) messages or some other same type of messaging. When the information is stored in one place, it can be very easily accessed by the administrators of the network or other people who need the data for some purpose, for example network planners. This data can be usually sent to the planning/management tool via different messaging schemes like XML or CSV [17].

The last part of network discovery is the QoS modelling. The QoS functionality in the MPLS is made with using DiffServ or IntServ at the moment. The Intserv is usually not considered to real world MPLS networks due its poor scalability features. So the DiffServ is at the moment the only decision to be reckoned with for commercial MPLS networks.

3.2.4. Simulation and Failure Case analysis

The simulation part is the main focus of all planning tools on the market at the moment. In the simulation part the attributes of the existing/planned network can be observed, the effects of possible future network topology and attribute/functionality changes can be simulated or different failure scenarios can be run to see how the network survives from them.

Simulation of MPLS networks needs little bit more functionality than for example IP-traffic simulation. Because MPLS has lots of different functions and it can run on many link layer techniques, it needs also very wide simulation choices. The simulation has to simulate VPNs of different types, QoS as well as basic MPLS functions such as label switching, LDP, CR-LDP, RSVP-TE and many more not listed functions.

3.2.5. Optimisation, TE and configuration

After the wanted simulation cases have been run the network simulated is ready to be optimised. The optimisation can be done for numerous things: the optimization can be structural, traffic engineering, financial (which it usually is in some manner), metrics changing (configuration) et cetera. Probably in MPLS networks case the most important and at the same time most money saving part of the optimization phase is the traffic engineering. Because this can be done with adding no new capital to the network and it can actually gain capital by adding new customers due to the released bandwidth after the optimization. Traffic engineering can also help with the congestion problems with redirecting some traffic of congested links or nodes.

After the optimisation is done for the simulated network on the results got from the different simulation cases, the results are now ready to be imported to the real network. This is done by configuring some of the parameters or devices in the network. This to be efficient, the wanted configuration changes have to be distributed to the devices in such manner that it does not need a lot of manual labour. There are two reasons for this. The first one is that the error probability increases in the configuration if it is done manually. The second one is that the changes made to the configuration can be in large scale. To do this manually can be very time consuming task and it is can also be very expensive to hire many experts to do this kind of optimization configuration. Many planning tools offer the functionality to export the configuration done in the simulator to the real network. This is done using different type of messaging traffic.

3.2.6. Capacity requirements and network provisioning

After the TE is done for the network based on the original/pre-estimated capacity requirements, it is time to take a look how these requirements can be decreased. TE part of the planning usually optimises the traffic in a way that the capacity requirements decrease in some parts of the network because of the traffic redirected.

In MPLS scheme the network provisioning can be done quite easily and with simplified methods. The provisioning between different services can be done almost entirely with VPNs. Because of different class types and end-to-end routing done by LSPs provisioning, the network manager/planner can utilise for example BGP to do the LSP provisioning task from start to finish. Only elements which need more specific configuration are the ingress and egress routers of the LSPs. This configuration can also be automated by utilising some planning or management tool. The tool utilised can offer

the functionality to export the configuration by messaging the devices needed by some messaging protocol and do the configuration as 'remote'.

3.2.7. Network operations

The last part of planning's iteration round is the network operations phase. The network operations mainly consist of management duties for the network. These management duties are not done because something has changed in the topology or in the functionality of the network, but because the gained functionality is wanted to be kept in the same level all the time. From this phase the planning tool gets additional information for the model so that it can be tuned to model the network, its functionality and the traffic flows in the network. In this phase also all of the maintenance duties are scheduled for the elements. For this part the service providers or the network vendors usually utilise some kind of centralised operations tool or device. These elements have varying functionality, but they usually have some basic functionality in common. These tools usually at least include the capturing and storage of flow information feature. In the next list the key features of one of these tools, Lancop's StealthWatch, are listed [18]:

- Provides an integrated, real-time overview of network usage, network performance and host integrity.
- Utilizes existing NetFlow™ and sFlow® routers and switches as well as native captured flow information to provide a comprehensive and cost-effect approach to network security and optimization.
- Reduces the time necessary to diagnose and separate security and network events from each other, and then to generate a response
- Supports segmented, high speed internal networks and fully meshed network environments.
- Documents historical information and trend analysis that accelerates network performance capacity planning and advanced resource management

3.3. The challenges of the MPLS networks

Before talking about the challenges of MPLS planning tool, some things about the challenges of MPLS networks nowadays in general have to be mentioned, because the needs of the MPLS networks are increased from the beginning of this millennium. The major challenges, that MPLS networks had and are still having, are the scalability and functional requirements with today's growing amount of subscribers and services and which are needed.

The next example lightens the situation little bit more. British Telecom has offered BGP/MPLS based VPN services from the year 2001. Table 3.1 shows the growth in the platform and features in a period of 2001-2007.

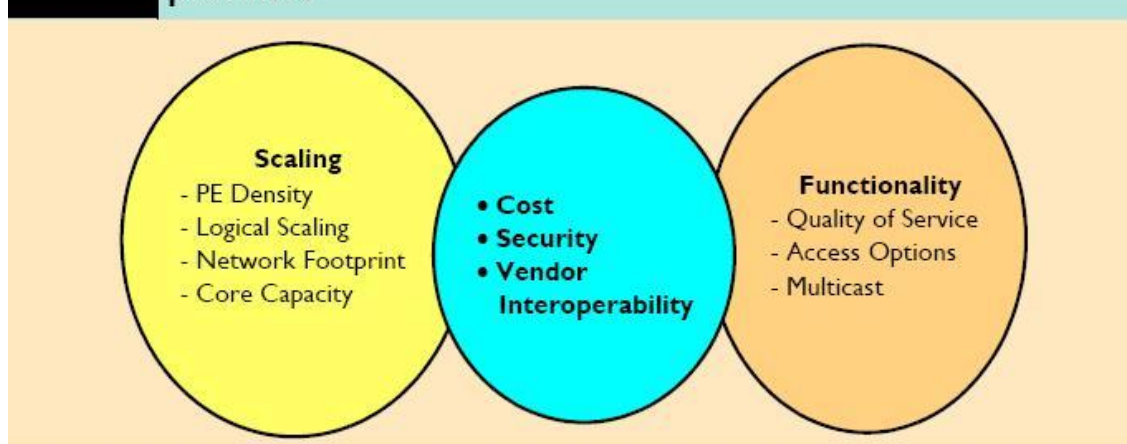
Table 1 BT's BGP/MPLS VPN platform growth and feature development in UK over 6 years

'Growth' Metric	Circa 2001	Circa 2007
Number of PE routers	<50	>1000
Maximum WAN link speed	OC-3/STM-1 (155 Mbit/s)	OC-48/STM-16 (2.44 Gbit/s)
Maximum customer access speed	E3/T3 (34 Mbit/s, 45 Mbit/s)	Gigabit Ethernet (1 Gbit/s)
Number of PoP locations	<10	~100
Number of distinct customer VPNs supported	Tens	Thousands
Number of customer ports	<1000	Tens of thousands
Number of customer routes	Thousands	Hundreds of thousands
Service Feature	Circa 2001	Circa 2007
Quality-of-Service and traffic categorisation	Best-effort only, supporting data traffic	6 DiffServ Code Point (DSCP) service classes, supporting voice, video, premium data and best-effort data
Access connectivity options	Leased line only	Leased line, layer 2 (ATM, Frame Relay), Ethernet, xDSL
Customer routing	Static only	Static, default or BGP
Access resilience options	Fixed line only	Fixed line, dial, xDSL as back-up

Table 3.1 Growth in MPLS platform of British Telecom [19]

As can be seen the scalability of the MPLS technique is not the only aspect which has been put to the test. With the more delay sensitive traffic types like VoIP, streaming video and multimedia teleconferencing the QoS and traffic categorisation needs have emerged. At the same time the customer networks have continued to grow larger and more complex, the techniques the customer uses in a transport have also increased as well as the routing methods used [19].

For the growth of the networks to be sustainable, the service providers need to keep the ability to support the VPN services in scalable manner and continue to develop an array of service features. The scaling and functionality aspects are always needed to be planned together with cost, security and vendor interoperability to make the developments more sustainable. The challenges in engineering for MPLS service providers and their relationship are shown in the Figure 3.2.

Figure 2 Main engineering challenges for BGP/MPLS VPN service providers**Figure 3.2 Engineering challenges for MPLS service providers [19]**

3.3.1. Scaling challenges of the MPLS networks

There are four main challenges which underpin the scalability of the MPLS platform and those challenges are PE density, logical scaling, network footprint and core capacity as they are shown in Figure 3.2.

One of the critical aspects of the physical network scalability is the ability to support a rapidly growing number of customer accesses without having to deploy an unbearable amount of PE access routers. Costs per port can be driven down and the scaling can be strengthened with a deployment of dense aggregation PE devices. The aggregation of PE devices reduces the accommodation requirements such as racking, power and ventilation.

It is not only the cost-efficiency aspect which encourages keeping the number of PE routers in check by densely aggregating large numbers of customer access circuits. Maintaining logical scalability is also very important. In a BGP/MPLS VPN network, logical state is required to maintain connectivity and reachability between PEs in the form of LSPs and BGP peering. LSPs are created in the control plane and once they are established they facilitate MPLS packet forwarding in the data plane. BGP peerings are formed in the control plane via TCP sessions which in turn facilitate the exchange of customer specific VPN routing information stored in PE routers. The accumulation of logical state has a direct relationship with the number of PE routers in the network used to terminate customer access connections. [19]

Ability to support a sizeable network footprint providing wide ranging geographical coverage is also a critical aspect of physical scaling. Building points of presence (POPs) and the WAN connectivity between them can become a significant cost, however many providers of VPN services target obvious geographical locations like big cities and financial centres. For a single service provider it is usually not possible to invest in network infrastructure in all countries due to a prohibitive build cost. Service providers need to make commercial agreement with other service providers to form a strategic partnership and reach the wanted geographical location. Creating these multi-AS backbones forms always more risk because the whole backbone network is not under a control/management of the service provider which has the contract with the customer. The physical scaling challenge of extending a network footprint is a good example of situation where the combined constraints of costs and security, provide a key influence over the preferred expansion strategy. [19]

The last element of the scalability is related with increasing challenge of handling large volumes of customers in the core network. The fundamental aim of effective core capacity management is to deploy the necessary core WAN links to support the traffic requirements and meet the necessary performance criteria such as packet throughput, delay and jitter. This must be done in way that it will be cost efficient. Expensive WAN link capacity upgrades are consequently avoided or deferred wherever possible. [19]

Deferral of WAN links upgrades can be achieved by optimising the flow of traffic in the core network such that the load of traffic is spread as evenly as possible around

available WAN links in the network topology. MPLS-TE is the way to balance the load made by the traffic in the network. The tunnels made with MPLS TE allows, for example, high cost links to be avoided and lower cost links to be used in preference, thus ensuring that traffic is spread around the network in an optimal fashion.[19]

3.3.2. Functionality challenges of MPLS networks

In the early versions of MPLS networks there was no possibility to apply QoS as a means of differentiating the classification and treatment of IP traffic entering the network. Nowadays the DiffServ technology is imported also to MPLS networks. At the moment there are two different techniques which can classify different types of traffic to eight (E-LSP) or 64 (L-LSP). At the moment service providers do not usually need more than eight types of traffic classes, so for the most service providers the E-LSPs are adequate enough to prioritise different types of traffic. But however, the CTs must be planned still with a great care. It is very important to make the prioritization so, that the most urgent and important type of traffic will be treated with the ‘privileges’ wanted.

MPLS VPN customers encompass a wide and diverse range in terms of number of access circuits, geographic coverage and access bandwidth. In order to keep their own costs down, the customer of IP VPN services continually seeks the most cost efficient way as part of their overall service delivery. In the beginning of MPLS networks, the principal access technology supported was leased line, such as E1/T1. More recently techniques like Ethernet, xDSL and wireless access technologies have got more popular choice for the ‘last mile’ access technology. [19]

The unique characteristics of a wide range of access types are important for the service provider, because the technical requirements for the PE devices come much wider with that. In other words, the functionality for the PE device which delivers traffic to an Ethernet based customer network will not probably be optimal counterpart for the xDSL based customer network. The service provider must meet the challenges offered by the different access technologies. The PE devices must be aligned in terms of the service features offered by the VPN service provided (for example QoS and multicast functionalities), irrespective of the access technology used.

IP multicast involves efficient transmission of a single datagram to a selected number of receiving hosts with suitable packet replication and forwarding in the core network. This results in a single multicast data stream to replace multiple point-to-point data flows. Multicasting has many uses in today’s core networking as in networking in general. For example IT companies can use multicasting to distribute regular software updates to company employees. Multicasting works also well when company’s CEOs want to hold ‘IPTV’ broadcasts as for example to distribute info for certain parts in the organization. Multicasting is also used in the finance sector, for the distribution of real market data between financial service providers and clients. [19]

When planning an MPLS network to accomplish multicast or P2MP functionality the planner has a very challenging task at hand, mainly due to the all pervading nature

of the required engineering developments. Multicasting impacts on most of the aspects of the network including QoS, traffic engineering and management, billing, dimensioning, network diagnostics and performance reporting.

4. EXISTING MPLS PLANNING TOOLS

At the moment there are some MPLS planning tools existing to aid the planning and managing a network. Most of these tools are of corporal nature, but also some free-to-use applications can be found from this field. In this chapter the concentration is on those existing tools. The functionality of two different MPLS planning tools is introduced: IP/MPLSView by WANL, Inc. and iVNT by Aria Networks Ltd.

There are also other tools for MPLS planning, but those are not examined in this thesis due to scarce information available. Some other tools, which are designed for MPLS planning or which have an MPLS module in them, are listed in the list below:

- MATE made by Cariden
- SP Guru made by OPNET
- OnePlan made by VPI Systems

4.1. IP/MPLSView

IP/MPLSView is a tool made by WANDL, Inc. (Wide Area Network Design Laboratory) for multi-layer Traffic Management and Traffic Engineering. IP/MPLSView offers much functionality for network design and planning and also for network engineering and operations in MPLS networks. Figure 4.1 shows different modules in MPLS network planning supported by the IP/MPLSView.

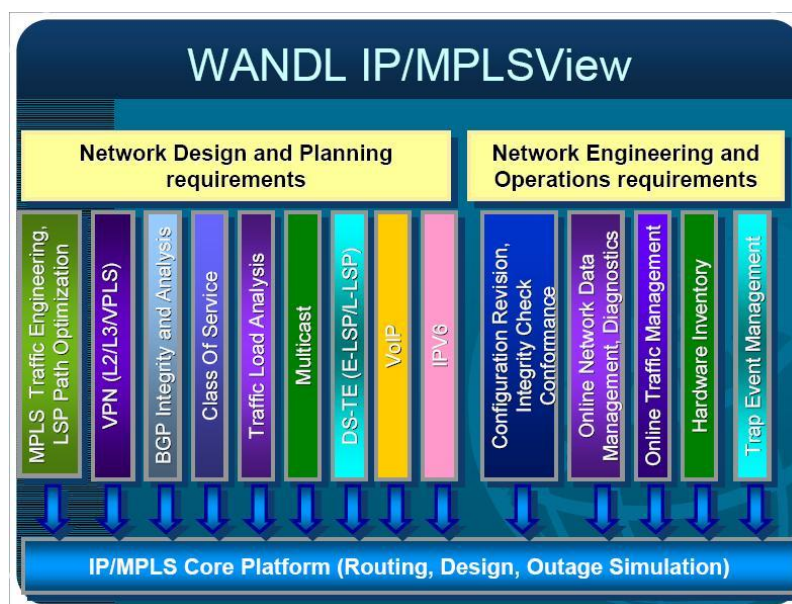


Figure 4.1 IP/MPLSView functionality structure [20].

Figure 4.1 shows that the IP/MPLSView's functionality is divided to two categories: to network design and planning supporting modules and to network engineering and operations supporting modules. Both of these aspects are important when talking about MPLS network planning/designing. Good and thorough planning is a key aspect for the future network engineering and maintenance, but it can also be said that function works in both ways: good network engineering and maintenance gives good prerequisites for future addition or change planning in the network.

The minimum hardware/system requirements for the IP/MPLSView to run are of course dependent on the size of the simulated network and the detail the simulation is wanted to run, but Wandl has released recommended system configuration where the IP/MPLS should run without major difficulties. These system configurations are shown in the figure 4.2.

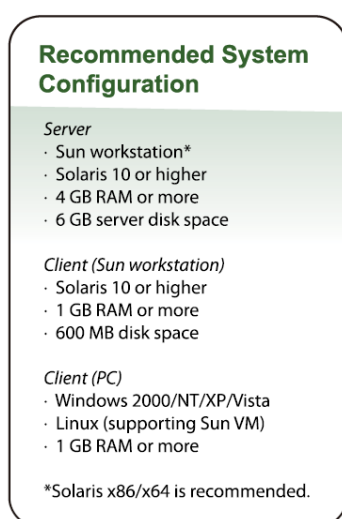


Figure 4.2 Recommended System Configuration for IP/MPLSView [21]

Next the different functionalities of the IP/MPLSView tool are discussed. The concentration is on the major functions that IP/MPLSView has to offer for the MPLS designing and planning field.

Wandl divides the planning to three different modules when speaking about the functionality of the IP/MPLSView: Simulation, design and optimization. Talking about the simulation, IP/MPLSView offers support at least for the following routing protocols: OSPF, ISIS, IGRP, EIGRP, RIP, BGP, MPLS, and GRE. It also supports the configuration of static and default routes, policy-based routing as well as access control lists. [22]

IP/MPLSView's simulation tools offer a way to run different kind of failure and bottleneck simulations. These situations are listed in the list below: [22]

- Multi-Layer failure simulations in LSP tunnels and IP layer.

- Failures in the network elements. This can be done straight from the topology map.
- “Batch” failure simulations. These cases include simulations for failures like failure of every node, site, interface, card etc.
- Multiple network element failures.
- Randomized daily failures.
- Peak/Worst case utilization analysis. The impact of traffic routing and rerouting during failures can be studied.
- Bottleneck analysis can be performed.

The last functionality, in the simulation tool, concerns validation of network changes. IP/MPLSView promises a safe environment to test the effects of the changes made to the network, before they actually are made. WANDL lists the following functionality for the simulation of network changes [22]:

- Performing of “What-if” scenario analyses.
- Making modifications to the network topology (add, delete, modify).
- Making modifications to any network elements (nodes, links, LSPs, VPNs).
- Change of routing protocols, protocol attributes, and metrics.
- Change of traffic load patterns, CoS policies and assignments.
- Performing of design (LSP tunnels, FRR, multicast) and viewing the impact caused to routing by this.

For the network topology design itself IP/MPLSView offers the following functionalities according to WANDL:

- Network can be designed from scratch (the network designing can be started from a blank sheet)
- Forecasted traffic can be utilized when designing a network incrementally on a top of an existing network configuration
- Tariffs and pricing data can be utilized for least-cost topology design
- Ensuring that the network design that can survive any level of user-defined resiliency requirements (node, link, card failure and SRLG)

IP/MPLSView also offers some other design and modeling functions which are not for topological designing but for traffic engineering. These functions are listed below:

- Designing of diverse paths protecting against site/link/SRLG failures
- Designing and simulation of LSP tunnels
- Designing of FRR tunnels
- Designing of diverse P2MP-TE multicast trees
- Performing of Route Reflector design and analysis

- Modeling of CoS classes, policies and different queuing schemes
- Modeling multicast distribution trees and routing
- Performing of VoIP backbone capacity design

IP/MPLSView offers simulation possibilities for many technologies and protocols used in MPLS networking. WANDL promises that the varying implementations of the standards used by different hardware vendors are also captured and modeled in by the IP/MPLSView routing and simulation engine. In the next list technologies simulated by IP/MPLSView are listed:

- IP core
- MPLS-TE
- MPLS-VPN
- VLAN
- BGP
- CoS
- Traffic load analysis
- Multicast
- VoIP
- IPv6

IP core simulation is carried by one specific module in IP/MPLSView. This module can construct the wanted network topology from the configuration files imported to IP/MPLSView. The module supports the following IGP: OSPF, ISIS, RIP, IGRP, and EIGRP. It also supports static routes and policy based routing. Routers can be modeled as logical routers and CoS traffic flows can be modeled with the module. The network topology is presented with Java based graphics.

For traffic engineering point, the IP/MPLS includes substantial amount of different functionalities. The LSP tunnels and statuses can be imported from already existing network or they can be designed from a start. The data can also travel to diverse direction: configuration tested in the simulator can be loaded back to the network simulated as ready configlets. When designing or simulating LSPs for a wanted network, IP/MPLSView includes a possibility to add a backup tunnels for the LSP created. Also the designing of the paths can be automated to reduce the work load of the designers. The paths can be redesigned afterwards for optimization if they have become suboptimal over time. IP/MPLSView also contains simulation/designing possibilities for most of the techniques added to MPLS functionality like FRR designing, DiffServ aware TE, P2MP-TE multicast tree computation. The impacts of adding or modifying tunnel or link properties can be viewed from a Java based GUI. Figure 4.3 illustrates the visualization of the simulator. In this case there are two backup tunnels generated for the primary LSP tunnel.

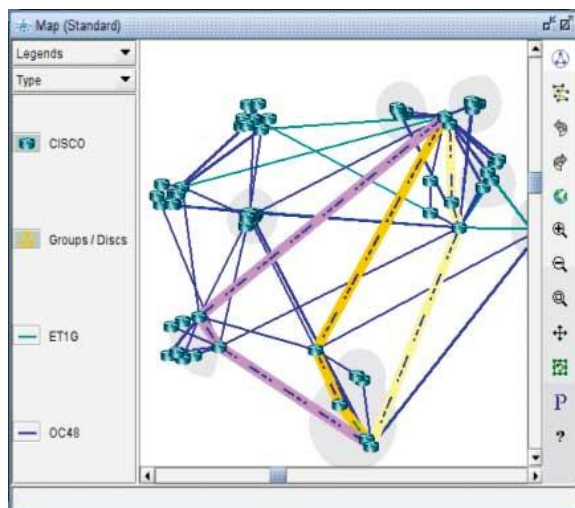


Figure 4.3 Tertiary LSP structure illustrated in IP/MPLSView [22]

IP/MPLSView supports wide variety of VPNs. It offers support for the VPNs of the following types:

- Layer 3 BGP/MPLS VPN (RFC2547bis) of full-mesh and hub-and-spoke
- VPLS-LDP
- VPLS-BGP
- Layer 2 VPN-Draft-Kompella(Layer 2 BGP)
- Layer 2 VPN-Draft-Martini(Layer 2 LDP)
- Layer 2 CCC
- TLS

The VPN module includes testing and monitoring possibilities: card failure simulations for to see the effects for those, integrity checks for the configuration files and VPN diagnostics and path tracing for VPN traffic.

For BGP analysis IP/MPLSView offers the following functionality:

- BGP routing studying with communities, admin weights, local preferences, MED and routing table calculations
- Performing of Route Reflector analysis
- Importing of BGP routing tables
- Performing of BGP peering analysis
- Performing of IBGP policy evaluation
- Different BGP map views with BGP neighbor relationships and Route Reflector hierarchy

For the CoS, DiffServ-TE is taken part of the IP/MPLSView design tool. The CoS module of the IP/MPLSView includes the following functionality:

- The CoS-module can model CoS classes, policies and different queuing schemes.
- Definition of application flows can be based on the CoS.
- Analysis of packet loss and delay statistics can be done for each CT separately.
- IP/MPLSView supports multiple class mapping, policy mapping, CBWFQ, Priority Queue CBWFQ, MDRR and Hierarchical WRR.

4.2. iVNT

iVNT (intelligent Virtual Network Topologies) is a software solution suite type of planning, forecasting and optimization tool made by Aria Networks. iVNT solution suite is divided to modules, which each handle certain type of functions and network types. The modules at the moment and their brief descriptions are listed below:

- iVNT MPLS-TE: Planning and optimisation software for MPLS-TE networks.
- iVNT Optical: Planning and optimisation software for Wavelength Division Multiplex networks.
- iVNT IP: Software for telecom carriers and enterprises to plan and design their network and services using Internet Protocol and Multi-Protocol Label Switching.
- iVNT Ethernet: A module directed to telecom carriers to design, plan, optimise and operate their network and services using Provider Backbone Bridging Traffic Engineering (PBB-TE), as at the moment being defined by IEEE 802.1Qay.
- iVNT Matrix-Maker: A module for telecom carriers to create traffic demand matrixes.
- iVNT VPN: A module for telecom carriers to plan and model IP-VPN services.
- IVNT TDM: A module for planning and optimisation of SDH/SONET networks and for migration of services to next generation networks.
- iVNT Server: Provides a central network planning platform for visualising, designing and planning the network for distributed network planners.
- Capacity Planner: A module for the service providers to see their resource consumption of planned services in devices, power, cards and ports.
- iAdapt: A framework for importing and exporting data to and from iVNT. It enables integration with one or more data sources (network equipment, network management systems, operational support systems, spreadsheets et al) using iAdapt Adapters.
- iCustomise: A customisation framework that enable iVNT to be customised and configured to meet specific planning and optimisation tasks, enables automation

of tasks that need to be performed on a regular basis, and supports a data transformation and manipulation capability.

- iVNT Inside: A module for embedding functionalities of products of other families to iVNT.
- iForecast: A business analytics and decision-support software to provide forecasting of next generation network infrastructure, customer services and capital and operational costs.

Because the focus of this research is MPLS planning, the modules of iVNT which have no direct use in MPLS network planning or design, are excluded from this study. The concentration is on explanation of the properties of the modules, which have more direct use in MPLS network planning and design [23].

The first module, which is under observation, is iVNT MPLS-TE which handles the traffic engineering simulation of MPLS networks in iVNT. The iVNT MPLS-TE module offers a good amount of functionality for the network planner or administrator to use. In the list below all the main features and functionality of the module are listed:

- Supported topologies: Linear, Line, Ring or Mesh networks (or combinations) comprising P, PE or CE routers.
- Supported services: P2P LSPs and P2MP LSPs. The P2MP LSPs can be optionally based on Steiner trees or least cost to destination trees. The module also supports Fast ReRoute for creating detour paths around failed links or failed nodes.
- MPLS-TE properties and Constrains: The module can optionally use the following properties and constrains when planning the network and LSPs: Bandwidth, Service Priority (DiffServ-TE), Link (IGP or metric) cost, Tariff Costs, Analytical costs, delay, distance, utilisation, maximum capacity fill policy, hop count, resource affinities (link colours) and over subscription policy
- Shared-Fate LSPs: The module provides the option to specify the use of shared-fate LSPs.
- End-to-End Protection: The module enables working and protection LSPs to be planned and provisioned, with full control over the required protection policy. Protection modes include unprotected LSPs, 1+1 protected LSPs, 1:1 protected LSPs and specific protection LSPs which utilise any combination of router and/or link diversity.
- Cost based planning: The module supports planning based on a variety of cost types which includes financial cost planning as well as metric-driven cost planning such as used by traffic-engineering metrics. The costs are divided into three categories in the module by the type of the cost: tariff costs, analytical costs and resource costs.

- Disjoint LSPs: iVNT MPLS-TE allows an operator to define LSP paths which do not share network resources with other user-specified LSPs. Like protection modes, Disjoint LSPs can be optionally specified and planned as path or subpath based.
- Constraint-Weights: The policies and the rules used to plan and optimise the network and LSPs can be decided by the user (network planner). These policies and rules can be controlled via weights to determine their relative importance, or if required to ignore a specific rule or policy.
- Traffic distribution: The module enables Least-Load Routing (LLR).
- Utilisation Thresholds: Thresholds can be set on resources to provide alarm information to inform operators when resource utilisation exceeds pre-defined levels during planning operations.
- Failure and event analysis: The module enables failure event analysis for single or multiple failure events and enables analysis of capacity, utilisation and impacted LSPs under conditions including: router failure analysis, inter-router failure analysis, vendor-specific failure analysis, Shared Resource Group analysis and automated and customisable failure event analysis.

This module has a multiplatform support for operational systems. At the time of writing the iVNT MPLS-TE module was functional all the major operational systems as well in the client side as on the server side. On Microsoft the module supports Vista, XP and 2003 Server. On Linux the Redhat Linux 5 and Debian Linux 5 are supported. On UNIX the Sun Solaris 10+ are supported. The module is also supported for deployment in Citrix environments and in virtualised (VMWare-based) platforms. [24]

The second module discussed is iVNT IP. This module is made, like already mentioned, for planning, design, modeling and optimization of IP and MPLS networks based on the metrics used in OSPF and IS-IS. [25]

The iVNT IP module works and operates in an IGP topology database, which is imported from network topology information as generated on a router by OSPF or IS-IS routing schemes. It also operates on a traffic demand matrix that represents IP traffic flows or a set of LSP based services like IP-VPN or VoIP. iVNT IP enables the user to define modeling objects, such as the target utilization envelope of routers and links, the failure scenarios of links, routers and SRGs (Shared Risk Groups) as well as it enables to define Equal Cost Multipath options and DiffServ drop policies. iVNT acts on the traffic matrix, the network topology and the modeling objectives to determine the IGP metrics which will optimize the traffic flows in the wanted network. In the next list the summary of the main features and capabilities is listed:

- Modeling features: The iVNT IP module offers tools for traffic demand planning and network capacity and utilization modeling, equal cost multipath modeling, DiffServ drop policy modeling, bandwidth over-subscription modeling,

modeling of current and future network topologies and services and it has features for OSPF and IS-IS IGP metric determination and modeling.

- **Analysis tools:** The module has features for single and multiple router, link and SRG failure simulation and analysis. Impact analysis of planned network outages and visualisation of all services transiting the network resources can be carried out with the module. The iVNT IP module has also a feature for scenario analysis – simulation, analysis and validation of network and traffic changes can be done with the functionality of the module.
- **Scalability:** Aria Networks promise scalability up to thousands of routers and hundreds of thousands of traffic flows.

The main focus on the iVNT functionality is upon the algorithms used in planning, forecasting and optimisation in the planning tool. Aria networks have developed a multi-algorithm software suite to do these tasks. This software suite is called DANI and it is used before network planning also in the other fields like the simulation and analysis process of pre-clinical drug discovery programs and for analysing complex diseases in the field of medicine. DANI is able to host, evolve and train many algorithms in parallel and includes a suite of public domain and proprietary patented algorithms. Types of algorithms that are presently supported by DANI include for example Dijkstra/SPF, Spanning tree algorithms, Steiner algorithms, Bayesian Net and Darwinian Neural Network. The work load of Dani can be distributed to several different processors or distributed computing environments. The iVNT and iForecast products both include a full implementation of Dani and utilise a mixture of algorithmic techniques to achieve their functions. [26]

5. EVALUATION OF THE ASPECTS OF PLANNING TOOLS

In this chapter all the information got from the planning tools studied in the process of making this thesis is collected. Because the material about the planning tools was scarce no deep analysis is made about the tools themselves, but the concentration is on the features which are done properly in the tools and which should be taken as referential model when planning a design tool for MPLS networks. The different phases of the MPLS network planning and how they should be implemented as a part of the planning tool are also discussed.

5.1. Functionality implementation in planning tools

In Chapter 3. the different phases in the MPLS network planning were discussed. In this chapter those phases are surveyed as a part of the functionality of planning tools. The phases illustrated in the Figure 3.1 are gone through one by one and the functionality implementation of every phase is discussed separately.

5.1.1. Forecasting in the MPLS planning tool

Like mentioned in the Subsection 3.2.1 there are many forecasting methods utilised in network planning and in forecasting in general. The traffic matrix can be formed to MPLS networks using many different algorithms. Like mentioned in the chapter 3.1.1 the algorithm which brings the best result for each individual network varies between networks.

If the results of the simulation run in the tool are wanted to be accurate and the simulated network to correspond the real implemented network as well as possible, there must be some kind of forecasting algorithm database implemented in the tool. The planning tool should be able to form traffic matrix models and store those different models. This must be done in a way which will decrease the CAPEX and OPEX of the network.

The computation based on the model has to be done within a limited time. This can be problematic in a real world MPLS core network. In the network, there can be thousands of routers and hundreds of thousands of LSPs. For this reason, the iteration of traffic matrix can take time. It would be suitable, if the calculation could be done

separated from the planning tool's other modules, in a way that would not harm the other processes used by the planning tool.

5.1.2. Making the network discovery using the planning tool

The two main ways how to do the network discovery are utilised in the most of the existing planning tools. Automated import and 'drag and drop' network discovery methods usually coexist side by side in the planning tools. And they should be. They both have certain purposes of use, in a way that the network planner cannot survive with either of them missing.

The most planning tools which were studied in this research have some kind of GUI feature for individual element addition to existing network topology. Of course some planning tools exist, like NS or TOTEM, where the addition of elements is done by inserting the topology info as text from a command line of the tool. This way of implementing, the addition is usually slower than via GUI and requires more expertise in using the planning tool. So for this reason, the GUI is very essential for the commercial use of planning tool.

The other aspect which is needed to be taken in to account is the addition of multiple elements to the network at the same time or importing the whole existing network topology to the planning tool. The importing of the existing network topology as whole is implemented in many of the planning tools existing. The IP/MPLSView, iVNT and MATE all have some kind of importing module for the import of network topology. The main difference between these importing modules is the way they do the import. For example MATE has two ways to import the existing topology as well as the configuration used in it. The routers used in the network can be polled via SNMP read-only polling and the topology is then discovered in the server dedicated to the task. The other way is to poll the Alcatel-Lucent 5620 SAM server which has the up to date information about the elements and their configuration stored. Using the latter way reduces the messaging needed and is this why implemented. In MATE there is also an offline discovery method which follows the steps used for the online discovery. The difference is that the network topology is parsed from the output of "show ospf" or "show is-is" command directed to one of the network routers. This output can be inserted to MATE offline. Also individual router configurations and traffic measurements can be imported to MATE offline [27]. Usually this automated network discovery is very crucial feature for the planning tool. The networks at commercial back bone, or even in the smaller scale, consist of so many elements that the network discovery cannot be done by adding one element at the time. The offline discovery can also be considered very desirable. When 'demoing' or studying with smaller networks, it can be so that there is no connection to the wanted network due of security or geographical issues (The network is in another continent where the demo, for example).

5.1.3. Simulation in planning tool

The simulation part in the planning tool is generally implemented with visualised events for different scenarios and functions. MPLS traffic has many different functions implemented in it, so there can be found many different visual simulations of different aspects of the network. For example the IP/MPLSView offers visual simulation for routing of different protocols, multi-layer failure situations, element failure situations, bottleneck analysis through simulation, adding or removing elements, VPN creation et cetera.

When talking about the MPLS planning tool simulation the following functionality is found from the planning tools surveyed in Chapter 4:

- LSP modelling
- Network visualisation
- “What-if” scenarios
- DiffServ traffic type flows
- Network failure simulations
- Routing metrics determination (the protocols available dependent on the tool)
- VPN modelling

On the basis of these results, one can say that at least these functionalities are required to get a thorough collection of simulations of the network simulated.

5.1.4. Optimisation with TE and configuration work in the planning tool

This is the part which can be considered probably the most important part of the planning tool functionality. In the optimisation part the capital spent to in the tool will pay itself back, if the tool’s functionality can provide major decreases to the CAPEX and OPEX directed to the network simulated.

In the planning tool the optimisation part is always based on the simulations run. If the simulations are precise and accurate, the optimisation will occur in a way which can be really said to optimise the network. If this is not the case, the optimisation will not succeed in the scale wanted. The optimisation is usually so large task and needs a lot of calculation power, that it should be somehow automated or at least some parts of the optimisation tasks should be automated. The usual network attributes which are optimised are the link metrics, LSP paths and the design of the network. The optimisation for these attributes of the network has to be done in a way which is the least-cost way for the service provider. Because of this, the tariffs and pricing data is good to be added as one parameter of the optimisation and the optimisation should also take this parameter in notice. The purpose of the tool dictates the value given to the cost parameter. If the planning tool is directed to educational use, the weight of the parameter can be marginal, but in the commercial planning tool the weight can be on

greater scale, and usually it is. For example the IP/MPLSView takes the tariffs and pricing data in notice in such way that the topology can be constructed and redesigned to be of least-cost type.

The TE part in the planning tool can have more functionality than just redesigning the LSP structure. Like mentioned before, the functionality demands, and with that the functionality, have increased for the MPLS networks a lot in recent years. For this reason, it is substantial that the planning tool can take all of the major functionality in use when engineering the traffic. The most important of these things are the different traffic types. The LSP engineering needs to be DiffServ capable, because the different class types have different priorities and so they cannot be treated as the same when redirecting traffic. With this, the planning tool should also be able to do FRR redesign in the same phase when the redesign of the LSPs is made. FRR is very important functionality in the case of high priority traffic. If the link or node malfunctions, the FRR should have the optimum backup LSP in use.

The configuration made in the optimisation process in the planning tool should be easily imported to the network which was simulated. The optimisation can involve thousands of elements in some cases and the configuration for all the elements can be time consuming and expensive, if done manually. For this reason the planning tool should have some kind of export module, where the configuration could be brought in to the 'real life' network as whole. All the planning tools discussed in this research (IP/MPLSView, iVNT and MATE) have some sort of exporting module/functionality for configuration exporting as part of them.

5.1.5. Network Provisioning in a planning tool

Like everything else in MPLS networks, the network provisioning has multiple aspects to take into account. The provisioning of the network does not only mean the network's devices or links which are needed to be provisioned, but as well the functionality like VPNs and VLANs need closer provisioning.

Again the one first thing to take into account when designing functionality for the MPLS designing tool is the scalability aspect of the function. The planning tool has to be designed in that manner, that it can be easily implemented to the use of larger networks which consist of thousands of network elements. For this reason, the provisioning has to be able to be automated if wanted. The automation of the provisioning helps also in other issues which can be risky, if done manually. It accelerates the deployment of the networks new services and revenues, reduces chance of misconfiguration, hides the network complexity due to high variety of services and protocols and also optimises the operations resources.

5.1.6. Network operations as a part of the planning tool

The network operations are constantly running in the 'background' of network planning process. It can be said that the network planning is never over due to the maintenance

work done after the planning phase of the network itself. But how should the planning tool be connected to the network operations? For this question, there are many views and opinions which vary on how the network operations should be handled.

Most of the network operators have a dedicated network operations centre (one or multiple) which handles the monitoring, configuring and maintenance of the network. These networks operations centres usually have some kind of maintenance/monitoring tool or application in use to help at the task. Some of the planning tools have also functionality for monitoring and maintenance, which can be utilised after the release.

For the monitoring and management, the tools provided at the time generally use the FCAPS model and framework. FCAPS is an acronym for Fault, Configuration, Accounting, Performance and Security, which are the categories of the framework. The principles of these different categories are defined in this framework. There are multiple solutions/tools on the market for each category. Some of the tools can handle multiple categories of FCAPS and some are even designed to handle all of them. This functionality is also added to some of the planning tools. For example IP/MPLSView is able to handle all of the fields in FCAPS.

When evaluating planning tools, one has to keep in mind that also the management and monitoring functions are accessible. If it is decided that the functionality of the planning tool is not wanted to stretch to the management and monitoring of the network, it has to be designed so that it will support the messaging with other platforms which has this functionality. The exporting and importing of data, has to be made in such manner that the data collected from the management platform will be easily adapted to the planner and vice versa.

6. CONCLUSIONS

The planning of the MPLS has many phases in it and each phase contains a lot of functionality requirements in it, regardless of the phase. These functionality requirements are to be modeled in the planning tool designed for the MPLS networks if the tool is wanted to have all the functionality which is needed to do the planning process of the network from start to finish.

The most important functionalities in the planning tool are that it has a way to simulate LSPs, MPLS-TE of certain type, different VPNs and DiffServ traffic flows, because these are the most important functionalities in the MPLS networks today. If this functionality is implemented and modeled in the planning tool and the planning tool has efficient optimization functionality, the CAPEX and OPEX of the network planned/simulated can be decreased. When designing a planning tool for the MPLS network, the scalability of the tool has to be kept in mind. The MPLS backbone networks today are in such a scale that the planning tool should have functionality which automates some processes in the planning cycle and so offers the opportunity to plan wider networks using MPLS. This functionality includes the automated TE engineering and importing and exporting of network topology as a whole.

The two planning tools studied in this research, the IP/MPLSView and iVNT, have both a wide range of functionality and can model and simulate MPLS networks of large scale. Both of them can be used for planning commercial MPLS backbone networks. The pricing of these tools is not known to the author, due to the companies' pricing and PR policy.

There are a couple of things which could be added to the list of additional research of this field. The first one is a survey directed to network planners of service providers which use MPLS in their networks. With this survey the real needs of the people working with MPLS networks can be charted and added as part of the research. This survey did not happen in this thesis, due to the lack of knowledge of where the survey was to be sent. The second thing is that the comparison of different planning tools could have been deeper and with that more comprehensive, if the information about the existing tools would have been available in greater scale. If additional research is made of this research field, these things should be taken into account.

REFERENCES

- [1] Cisco Systems, Inc., Multiprotocol Label Switching (MPLS), [WWW], [Cited 07/12/2009], Available at:
http://www.cisco.com/en/US/products/ps6557/products_ios_technology_home.html.
- [2] Minei, I., Lucek, J. MPLS-Enabled Applications, John Wiley & Sons Ltd, 2006.
- [3] Rosen, E., Viswanathan, A., Callon, R., Multiprotocol Label Switching Architecture, RFC 3031, January 2001.
- [4] Andersson, L., Doolan, P., Feldman, L., Fredette, A., Thomas, B., LDP Specifications, RFC 3036, January 2001.
- [5] Juniper Networks, Inc. How MPLS Works, [WWW], [Cited 14/01/2010], Available at:
<http://www.juniper.net/techpubs/software/erx/junose81/swconfig-bgp-mpls/html/mpls-config5.html>.
- [6] Koucheryavy, Y., Next Generation Networks and Systems, Lecture Notes: TLT-2686 Next Generation Networks and Systems, Technical University of Tampere, Spring 2002.
- [7] Jamoussi, B., Andersson, L., Callon, R., Dantu, R., Wu, L., Doolan, P., Worster, T., Feldman, N., Fredette, A., Girish, M., Gray, E., Heinänen, J., Kilty, T., Malis, A., Constraint-Based LSP setup using LDP, RFC 3212, IETF, January 2002.
- [8] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., Swallow, G., RSVP-TE: Extensions to RSVP for LSP Tunnels, RFC 3209, IETF, December 2001.
- [9] Rekhter, Y., Rosen, E., Carrying Label Information in BGP-4, RFC 3107, IETF, May 2001.
- [10] Le Faucheur, F., Wu, L., Davie, B., Davari, S., Vaananen, P., Krishnan, R., Cheval, P., Heinanen, J., Multi-Protocol Label Switching (MPLS) Support of Differentiated Services, RFC 3270, IETF, May 2002.

- [11] Le Faucheur, F., Lai, W., Requirements for Support of Differentiated Services-aware MPLS Traffic Engineering, RFC 3564, IETF, July 2003.
- [12] Laserre, M., Kompella, V., Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signaling, RFC 4762, IETF, January 2007.
- [13] Penttinen A., Chapter 10 – Network Planning and Dimensioning, Lecture Notes: S-38.145 - Introduction to Teletraffic Theory, Helsinki University of Technology, Fall 1999.
- [14] Evans, J., UK Network Operators' Forum 15 , Rochdale, United Kingdom, January 21, 2010. Available at: <http://www.uknof.org.uk/uknof15/Evans-TrafEng.pdf>.
- [15] Andersson, G., Johansson, M., Telkamp, T., Traffic Matrix Estimation on a Large IP Backbone – A Comparison on Real Data, Internet Measurement Conference, Taormina, Sicily, Italy, October 25-27, 2004. Available at: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.59.3152&rep=rep1&type=pdf>.
- [16] AT&T Inc., Global Customer Service Center, SLA Methodology, [WWW], [Cited 12/05/2010], Available at: <http://dedicated.sbcis.sbc.com/NDWS/sla/methodology.jsp>.
- [17] Morris, S.B., Network Management, MIBs and MPLS, Pearson Education, Inc., 2003. pp. 39-42.
- [18] Lancope Inc., Network Operations, [WWW], [Cited 14/05/2010], Available at: <http://www.lancope.com/solutions/network-ops.aspx>.
- [19] Veitch, P., Scalability and functionality challenges for MPLS VPN Networks, The Journal of The Communications Networks, Volume 6 Part 2, pp. 38-43, Communications Network, April-June 2007.
- [20] Wang, D., IP/MPLS Network Planning, Design, Simulation, Audit, and Management. Proceedings of the The 16th Annual Wireless and Optical Communications Conference, New Jersey, USA, April 27-28, 2007. Available at: http://www.wocc.org/wocc2007/doc/presentation/PlenarySpeaker-day1/wocc_DaveWang2.pdf.
- [21] Wandl, Inc., IP/MPLSView for Network Planning Brochure, [WWW], [Cited 03/03/2010], Available at: http://www.wandl.com/html/support/papers/IPMPLSView_Planning.pdf.

- [22] Wandl, Inc., Product IP/MPLSView , [WWW], [Cited 04/03/2010], Available at: http://www.wandl.com/html/mplsview/mplsview_simulate.php#route.
- [23] Aria Networks Ltd., iVNT Solution Suite Overview, [WWW], [Cited 08/05/2010], Available at: http://www.aria-networks.com/products/iVNT_Overview.html.
- [24] Aria Networks Ltd., iVNT MPLS-TE Product Brief, [WWW], [Cited 08/05/2010], Available at: http://www.aria-networks.com/products/iVNT_MPLS-TE.html.
- [25] Aria Networks Ltd., iVNT IP Data Sheet, [WWW], [Cited 08/05/2010], Available at: http://www.aria-networks.com/products/iVNT_IP.
- [26] Aria Networks Ltd., DANI – The Multi-Algorithm Software Suite, [WWW], [Cited 09/05/2010], Available at: http://www.aria-networks.com/technology/DANI_Multi-Algorithm_Software_Suite.html.
- [27] Cariden Technologies, Inc., MATE Overview Brochure, 2008.