MPLS
Multi-Protocol Label Switching

CERN - 18 June 2004

edoardo.martelli@cern.ch

Agenda

* History

* MPLS technology

* MPLS applications:

  - Traffic Engineering

  - Virtual Private Networks

* GMPLS

* MPLS at CERN

History (I)

In mid 1990s, service providers were concerned that the current routers would be unable to scale to meet increasing traffic demands.

At the same time, ATM switches throughput was an order of magnitude higher than that of routers.

Some vendors thought of using ATM switching capabilities in conjunction with IP routers.

History (II)

In 1996, an IETF working group called MPLS was formed.

Soon the initial concern was over, since the routers forwarding capabilities reached and overtaken that of ATM switches.

But MPLS was versatile, and ready for being adopted by new applications.

Multi Protocol Label Switching

MPLS is a mechanism for engineering traffic, independently from routing table.

MPLS performs analysis of a packet's destination just once (ingress), assigns it a label, then places it in a preconfigured tunnel (Label Switched Path LSP).

At each hop through the LSP, MPLS handles the packet at Layer 2 only.

MPLS packet

When a packet enter an LSP, the ingress router examines the packet and assigns it a ***label*** based on the destination, placing a 32 bit header in front of the IP packet's header.

The label transforms the packet from one that is forwarded based on IP addressing, in one that is forwarded based on its MPLS label.

L2 header	**32 bit MPLS header**	**IP packet**

MPLS header

Label	Exp	S	TTL
20 bits	3 bits	1bit	8 bits

LABEL: Identifies the packet. It is changed at every hop.

EXP (experimental): used for queuing the packets in different classes of services.

S (Stacking bit): indicates if the packet has more than one MPLS labels.

TTL (Time To Live): limit on the number of router hops the packets can travel through

MPLS label

A label is a short, fixed length, locally significant identifier which is used to identify a FEC (Forwarding Equivalence Class) or an LSP (Label Switched Path).

Labels from 0 to 15 are reserved.

0: IPv4 explicit null. It indicates that the label stack must be popped, and the forwarding of the packet based on the IPv4 header.

1: router alert label. When a received packet contains this label, it is delivered to a local software module for processing.

3: implicit null label. It can be assigned and distributed, but actually it never appears: when a LSR would push this label, it pops the MPLS stack instead. Used for the Penultimate Hop Popping technique.

LSP - Label Switched Path

MPLS is responsible for directing a flow of IP packets along a predetermined path across a network.

This path is the ***LSP (Label Switched Path)*** that is unidirectional, from the ingress router to the egress one.

An LSP can be established across multiple Layer 2 transports such as ATM, Frame Relay or Ethernet. Thus, MPLS is able to create end-to-end circuits, with specific performance characteristics, across any type of transport medium.

LSP provisioning

LSPs can be provisioned in different ways:

Statically, configuring each hop with the labels to be used.

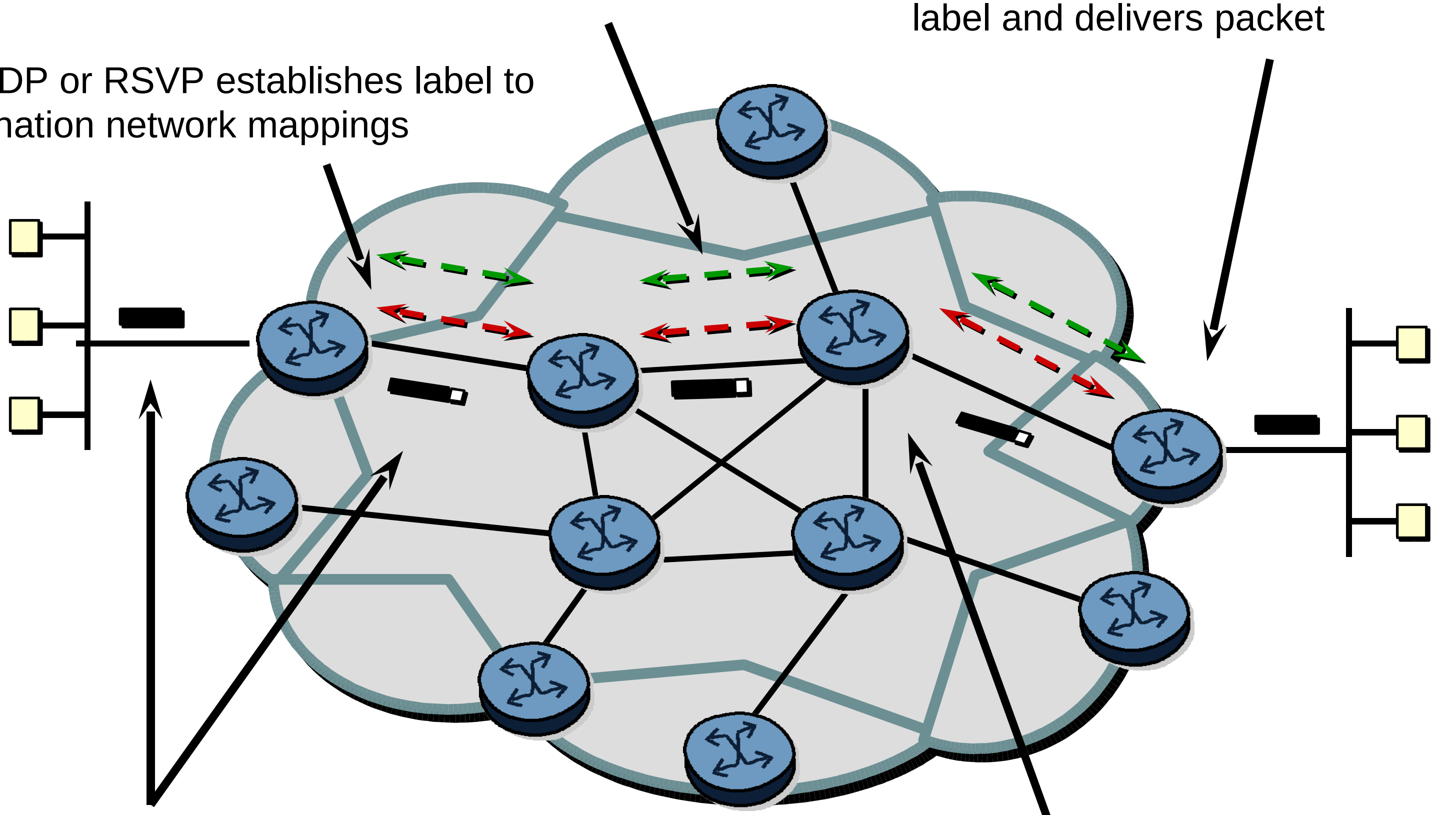
Dinamically, using a signaling protocol:
- *RSVP* (Resource Reservation Setup Protocol)
- *LDP* (Label Distribution Protocol).
An LSP is configured on the ingress router only.

MPLS operations

1a. Existing routing protocols (e.g. OSPF, IS-IS) establish reachability to destination networks

4. Edge LSR at egress removes label and delivers packet

1b. LDP or RSVP establishes label to destination network mappings

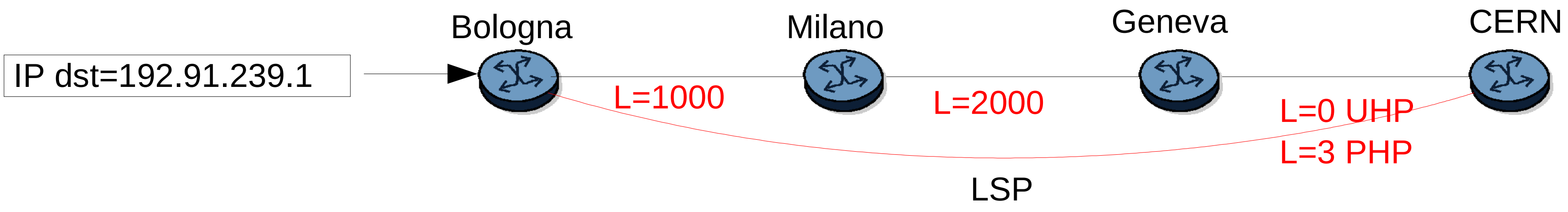


2. Ingress Edge Label Switch Router receives packet, performs Layer 3 value-added services, and "labels" packets

3. LSRs switch labelled packets using label swapping

LSR: Label Switching Router

Packet processing example (I)



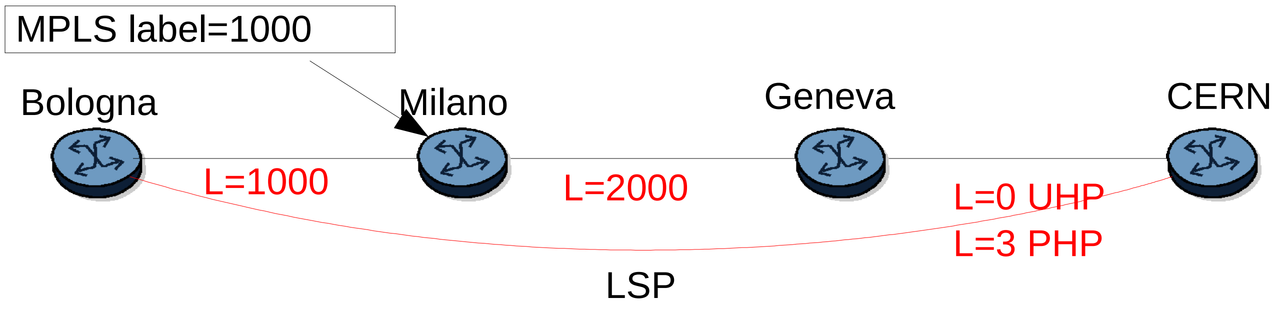
Bologna is the ***ingress*** router for an LSP to CERN.

An IP packet addressed to 192.91.239.1 arrives in Bologna.

Bologna has a route for 192.91.239.0/26 with next hop the LSP to CERN.

Bologna pushes the label 1000 and sends the packet to Milano

Packet processing example (II)

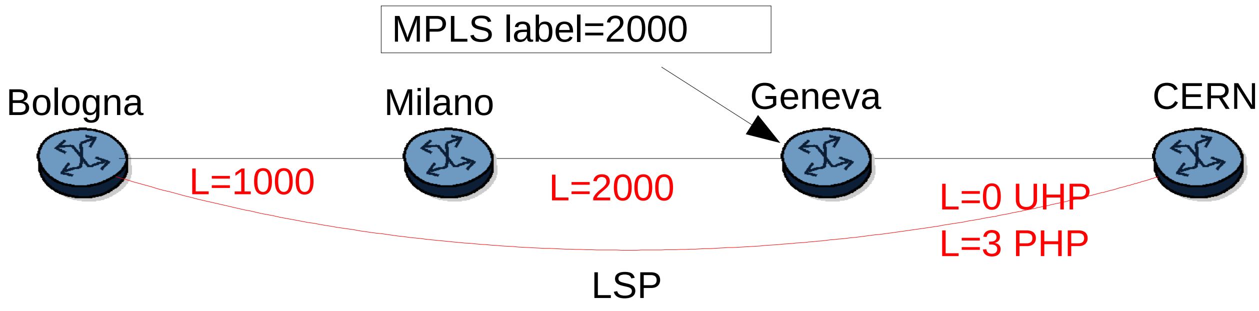


Milano is a ***transit*** router.

Milano receives a packet with an MPLS label, so it forwards it using the MPLS forwarding table and not the normal IP routing table.

Milano pops the label 1000 and pushes the label 2000, then forwards the packet to Geneva.

Packet processing example (III)

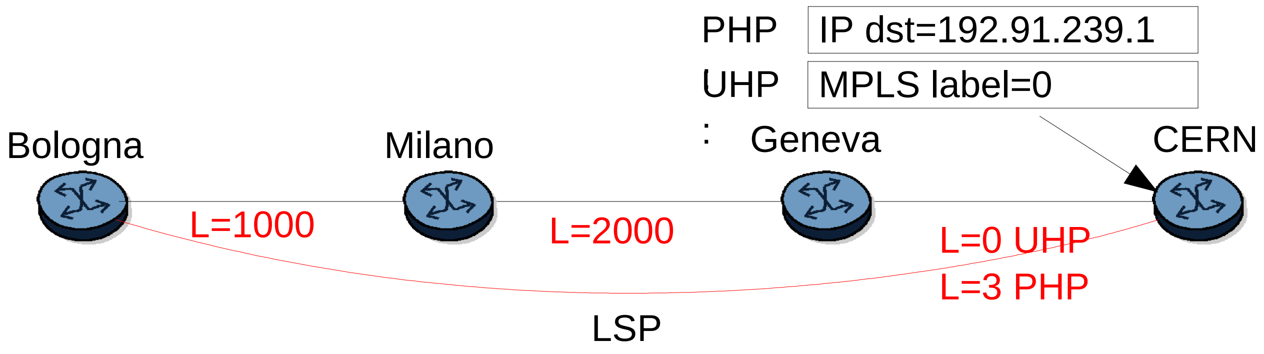


Geneva is a transit router, and also the ***penultimate*** router.

Geneva receives a packet with label 2000.
It either can pop the label stack (implicit null label, 3), or swap it with the label 0 (explicit null label).

The ***Penultimate Hop Popping*** technique is used to reduce the load on the egress routers, usually endpoints for several LSPs.

Packet processing example (IV)



CERN is the ***egress*** router.

In case of PHP, CERN receives a normal IP packet, and routes it accordingly to its IP routing table.

In case of UHP, CERN receives a MPLS packet with label 0, so it pops the MPLS header and then makes a regular IP forwarding decision.

MPLS Applications

Traffic Engineering

Layer 3 VPN (Virtual Private Networks)
- BGP/MPLS VPN (RFC2547bis)
- Virtual Router (draft-ietf-l3vpn-vpn-vr-01.txt)

Layer 2 Transport
- AToM
- CCC
- VPLS (Virtual Private LAN Services)

TE – Traffic Engineering

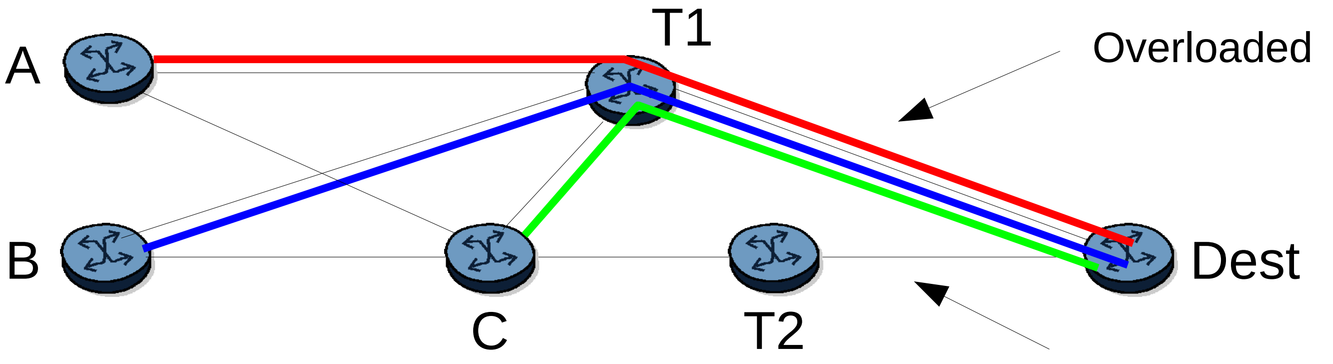
A major goal of Internet Traffic Engineering is to facilitate efficient and reliable network operations while simultaneously optimizing network resource utilization and traffic performance.

Traffic engineering refers to the process of selecting the paths chosen by data traffic in order to balance the traffic load on the various links, routers, and switches in the network.

Traffic engineering is most important in networks where multiple parallel or alternate paths are available.

TE – Traffic Engineering

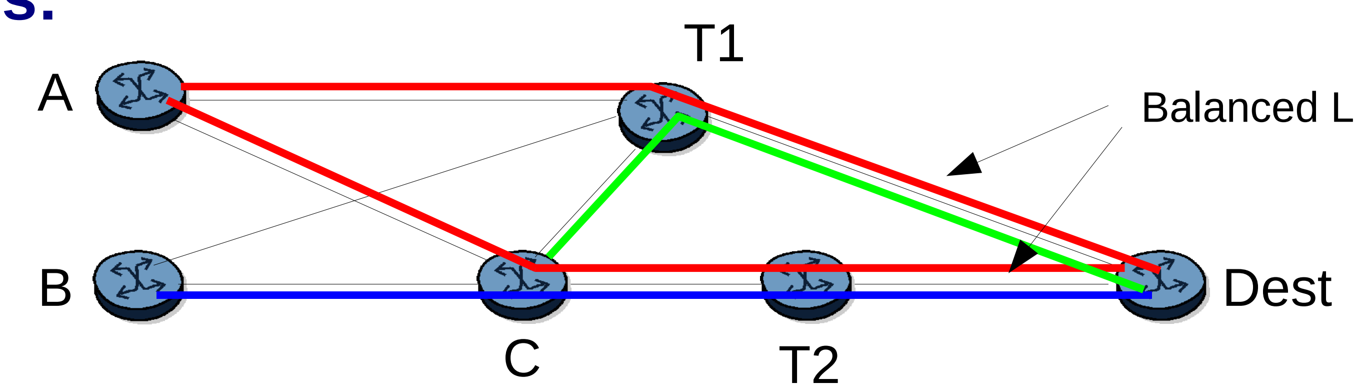
IGP paths:



Overloaded

Unused bandwidth

TE paths:



Balanced Load

IGP: Interior Gateway Protocol (es. ISIS, OSPF)

TE – Traffic Engineering

The goal of TE is to compute a path from one given node to another, such that the path does not violate the constraints (e.g. Bandwidth/administrative requirements...) and is optimal with respect to some scalar metric.

Once the path is computed, TE (a.k.a. Constraint based routing) is responsible for establishing and maintaining forwarding state along such a path.

RSVP - Resource Reservation Protocol

Usually, RSVP is used for MPLS Traffic Engineering.

RSVP was originally designed to allow end hosts to request specific qualities of service (QoS) from the network. RSVP was used by routers to deliver QoS requests to all nodes along the path(s) of the flows and to establish and maintain state to provide the requested service.

RSVP was designed to support extensible mechanisms, so it was adopted by the MPLS developers and modified in order to carry the necessary information for LSPs set up, labels distributions, and related resource reservations informations.

ERO – Explicit route objects

RSVP supports ***Explicit Route Objects*** (EROs), i.e. a way of forcing the routing of an LSP over one or more specified transit point.

The use of EROs allows the LSP to be routed over a path that would have not been used following the normal IP routing.
Hence it allows for *traffic engineering*.

EROs can be ***Loose*** or ***Strict***. A strict ERO specifies all the hops. A loose ones only some, relying on the normal routing to reach them.

VPN – Virtual Private Networks

A VPN is a set of sites that are allowed to communicate to each other.

Since MPLS allows for the creation of "virtual circuits" or tunnels, across an IP network, it was a logical consequence to use MPLS to provision Virtual Private Network services.

VPN services isolate customers traffic across the provider's IP network and provide secure end-to-end connectivity for customer sites.

It should be noted that using MPLS for VPNs simply provides traffic isolation, much like an ATM or Frame Relay service. MPLS currently has no mechanism for packet encryption, so if customer requirements included encryption, some other method, such as IPsec, would have to be employed.

BGP/MPLS VPN - RFC2547bis

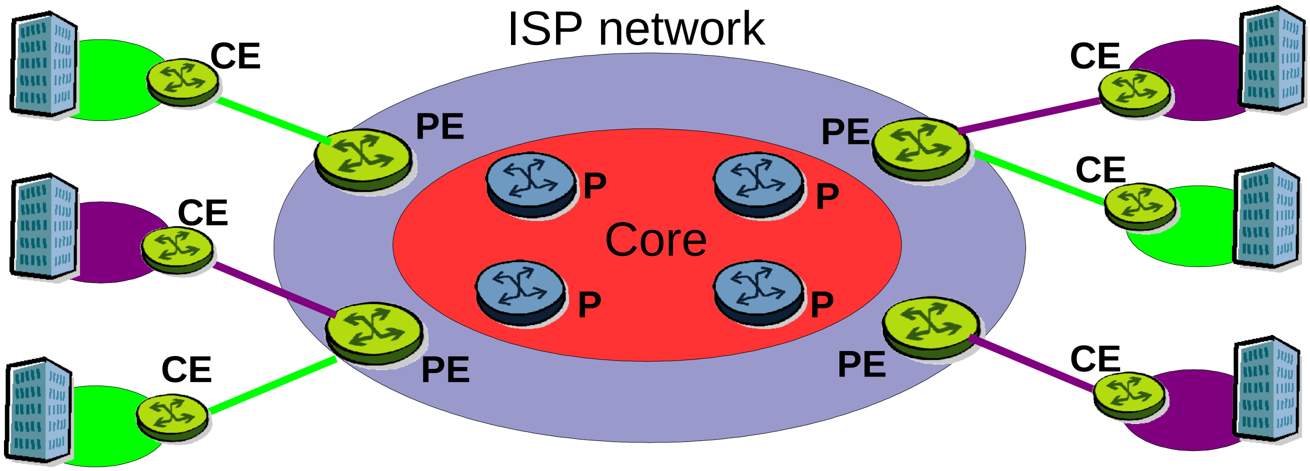
RFC 2547bis defines a mechanism that allows service providers to use their IP backbone to provide VPN services to their customers.

RFC 2547bis VPNs are also known as BGP/MPLS VPNs because BGP is used to distribute VPN routing informations across the provider's backbone and because MPLS is used to forward VPN traffic from one VPN site to another.

From the enterprise customer's perspective, an MPLS IP VPN looks like a private IP cloud connecting multiple sites. The enterprise has complete control over the network, including IP addressing route advertisement.

RFC: Request For Comment
BGP: Border Gateway Protocol

MPLS/BGP VPN connection model

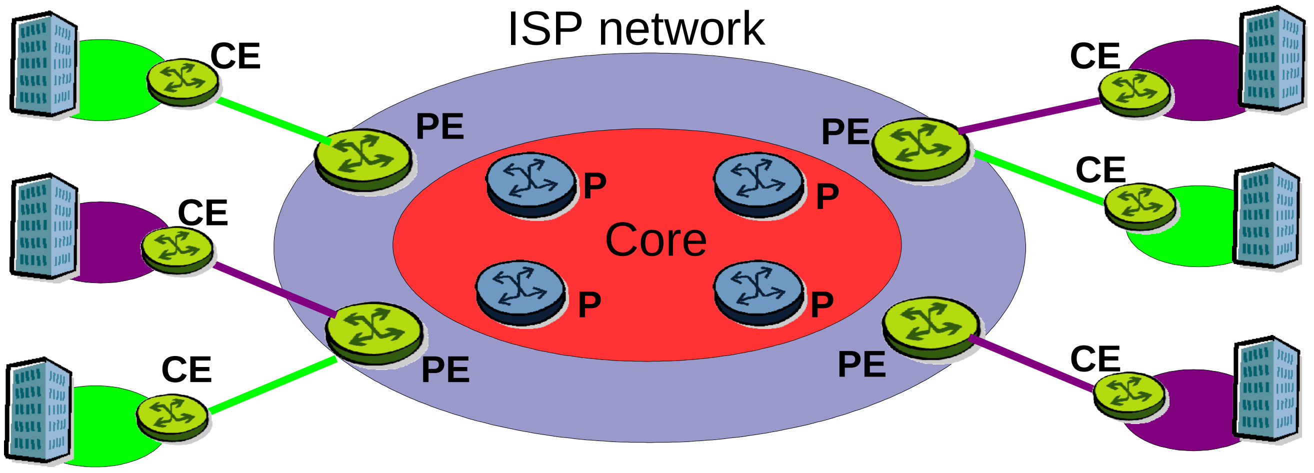


P: Provider routers*.* They populate the core of the service provider network.

PE: Provider Edge routers*.* They face the customers routers and are connected to the core.

CE: Customer Edge routers*.* Connect the customer sites to the ISP network.

MPLS/BGP VPN connection model - Core



PE routers are MP-iBGP fully meshed. They use MPLS with the core and plain IP with the customers.

PE distribute VPN information through MP-BGP to other PE routers.

P and PE routers share a common IGP (OSPF or ISIS)

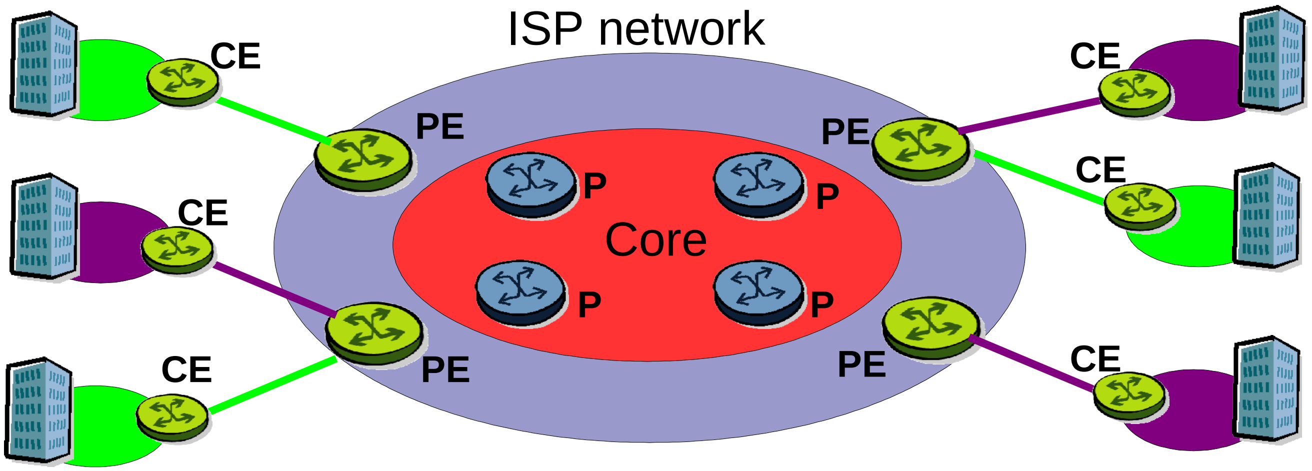
P routers use MPLS.

P routers do not run BGP and do not have any VPN knowledge

Labels are distributed through LDP

MP-BGP: Multi Protocol Border Gateway Protocol

MPLS/BGP VPN connection model - Edge



PE and CE routers exchange routing informations through eBGP or an IGP or static routing

CE router runs IP and standard routing software, so they don't need MPLS or any other special feature.

PE routers maintain separate routing tables:

The global routing table

- exchanged with all PE and P routers
- populated by the backbone IGP (ISIS or OSPF)

VRF (VPN Routing and Forwarding) routing tables

- Routing and Forwarding table associated with one or more directly connected sites
- populated by the routing protocol between CE and PE

LDP – Label Distribution Protocol

LDP provides a standard methodology for dynamic label distribution in an MPLS network by assigning labels to routes that have been chosen by the underlying routing protocols. The resulting labeled paths, the LSPs, forward label traffic across an MPLS backbone to particular destinations.

LDP provides the means for LSRs to request, distribute, and release label prefix binding information to peer routers in a network. LDP enables LSRs to discover potential peers and to establish LDP sessions with those peers for the purpose of exchanging label binding information.

—

LSP: Label Switched Path

LSR: Label Switching Router

FEC – Forwarding Equivalence Class

Forwarding Equivalence Class (FEC) is a set of packets which will be forwarded in the same manner (e.g., over the same path with the same forwarding treatment). Typically packets belonging to the same FEC will follow the same path in the MPLS domain.

While assigning a packet to an FEC, the ingress LSR may look at the IP header and also some other information such as the interface on which this packet arrived. The FEC to which a packet is assigned is identified by a label.

L2 VPN

Layer 2 VPNs allow service providers to provision Layer 2 services such as Frame Relay, ATM and Ethernet between customer locations over an IP/MPLS backbone. Service providers can thus provision Layer 2 services over their IP networks, removing the need to maintain separate IP and Frame Relay/ATM network infrastructures.

Layer 2 VPNs are an extension of the work being undertaken in the PWE3 working group.

L2 VPN standards

There are two IETF working groups defining standards for the support of Layer 2 services over IP networks:

PWE3 (Pseudo Wire Emulation Edge to Edge) is a working group responsible for defining solutions for the support, encapsulation and service emulation of pseudo wires over packet-based IP networks. The Internet draft documents previously known as Draft-Martini have now been adopted by the PWE3 working group. This approach is sometimes referred to as "Any Transport (or Protocol) over MPLS (AToM/APoM).

PPVPN (Provider Provisioned VPN) is the working group responsible for defining and specifying a set of solutions for supporting provider-based VPN implementations. This group has defined a number of draft documents dealing with Layer 2 VPNs, known as Kompella's Drafts. They describe the support of Layer 2 VPNs in a similar fashion presently being used to support Layer 3 VPNs or IP VPNs as defined under RFC 2547bis.

Martini vs Kompella

Both methods specify a way to group individual connection at different routers into a flat network. The ways the packets are moved across the MPLS network, and the ways the L2 frames are encapsulated in MPLS packets are almost the same for both methods.

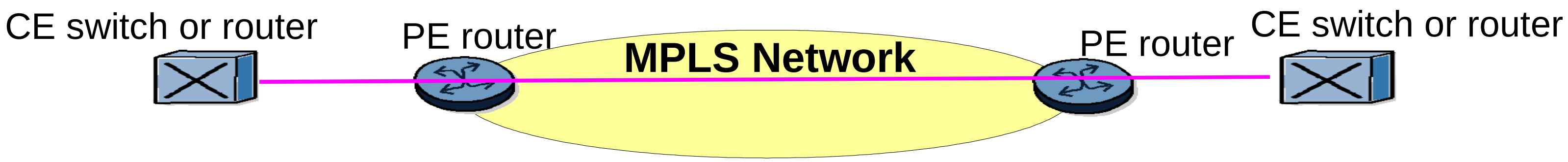
The ***Martini drafts*** are named after Luca Martini, a Cisco fellow.
Under Martini, VPNs are built from VC (Virtual Circuit) IDs tags that identify the virtual channels. These tags are distributed using LDP. Circuits are only point-to-point.

The ***Kompella drafts*** are named after Kireeti Kompella, a distinguished engineer at Juniper.
Under Kompella, VPNs are built using a BGP attribute. Circuits are both point-to-point and point-to-multipoint.

Kompella argues that provisioning Martini's VPN is more complicated, since it must be fully meshed. And also troubleshooting is more complex.
Martini argues that LDP converges more quickly than BGP and that LDP is more secure.
Kompella says that with BGP, VPNs scale better in a multi AS (Autonomous System) scenario, since LDP is an intra AS protocol.

VPWS – Virtual Private Wire Services

Virtual Private Wire Service (VPWS) provides point-to-point connectivity between customer sites, where the service provider network emulates a set of wires between the customer's sites over the underlying MPLS network.



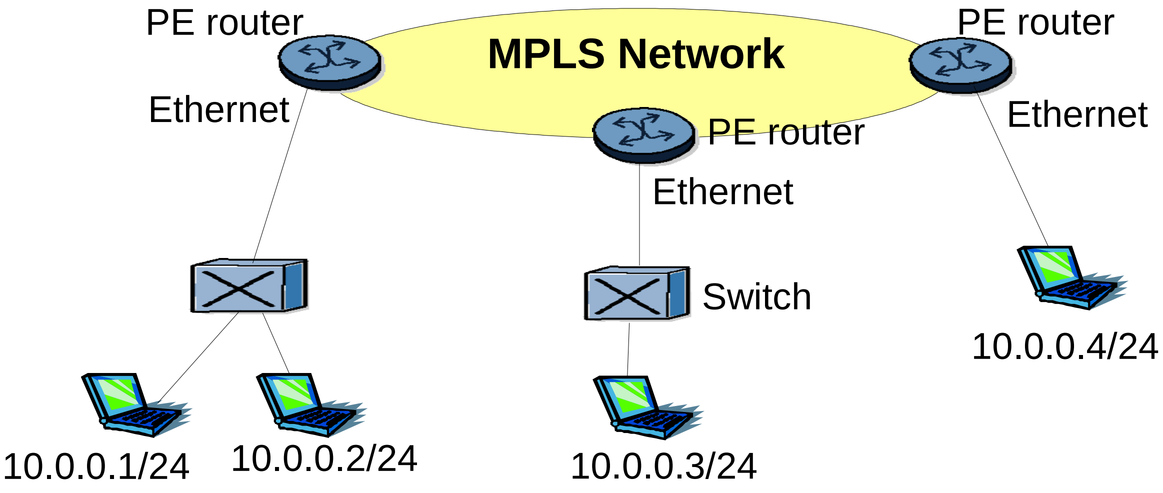
AToM - CCC

AToM (Any Transport over MPLS) and ***CCC*** (Circuit Cross Connection) are solutions for transporting Layer 2 packets over an MPLS backbone. They enables service providers to supply connectivity between customer sites with existing data link layer (Layer 2) networks by using a MPLS backbone. They provide only point-to-point connection.

AToM is a Cisco technology that is becoming an IETF standard.
CCC is a Juniper technology.

VPLS – Virtual Private LAN Services

VPLS refers to a method for using MPLS to create virtual LAN services based on Ethernet. In this type of service, all edge devices are connected to the same LAN, i.e. they maintain MAC address tables for all reachable end nodes, much in the same way as a LAN switch.



QoS – Quality of Service

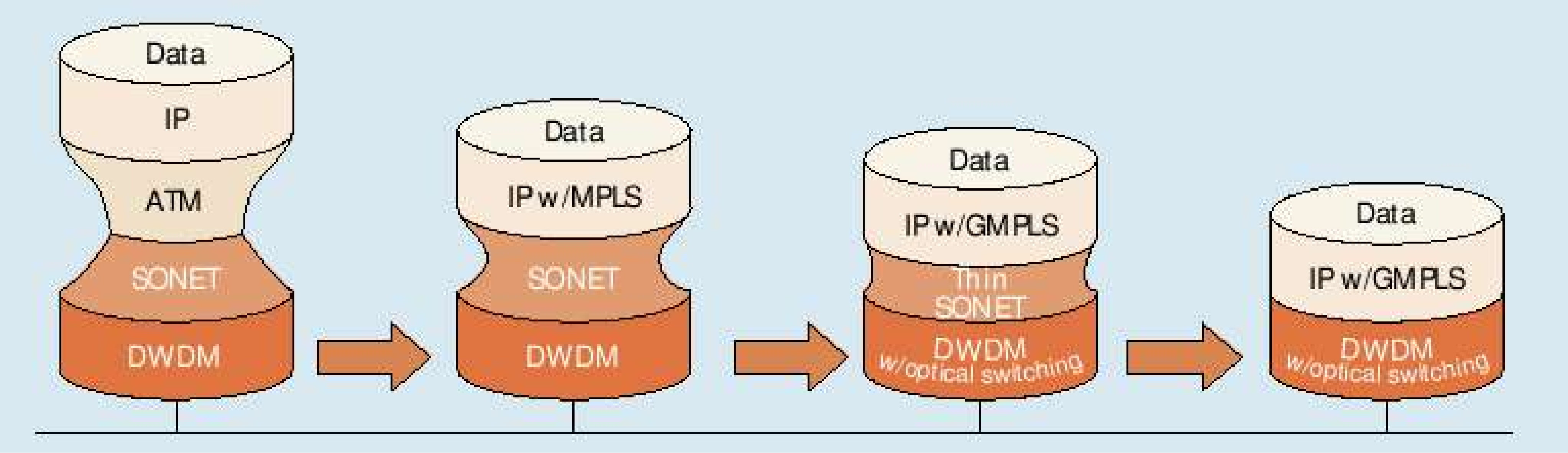
MPLS supports the same QoS as IP. These mechanisms are IP Precedence, Committed Access Rate (CAR), Random Early Detection (RED), Weighted RED, Weighted Fair Queuing (WFQ), Class-based WFQ, and Priority Queuing. Proprietary and non-standard QoS mechanisms can also be support but are not guaranteed to interoperate with other vendors.

Since MPLS also supports reservation of Layer 2 resources, MPLS can deliver finely grained quality of service, much in the same manner as ATM and Frame Relay.

GMPLS – Generalized MPLS

GMPLS, also referred to as multiprotocol lambda switching, is an extension of MPLS for supporting not only packet switching devices, but also those devices that perform switching in the time, wavelength, and space domains, such as optical switches, TDM muxes, and SONET/ADMs.

Evolution toward photonic networking



GMPLS features

GMPLS supports several features including:
- Link Bundling - the grouping of multiple, independent physical links into a single logical link
- Link Hierarchy - the issuing of a suite of labels to support the various requirements of physical and logical devices across a given path
- Unnumbered Links - the ability to configure paths without requiring an IP address on every physical or logical interface
- Constraint Based Routing - the ability to automatically provision additional bandwidth, or change forwarding behavior based on network conditions such as congestion or demands for additional bandwidth.

GMPLS introduces a new protocol called LMP (Link Management Protocol). LMP runs between adjacent nodes and is responsible for establishing control channel connectivity as well as failure detection. LMP also verifies connectivity between channels.

GMPLS Models

GMPLS supports two methods of operation, peer and overlay.

In the ***peer model***, all devices in a given domain share the same control plane. This provides true integration between optical switches and routers. Routers have visibility into the optical topology and routers peer with optical switches.

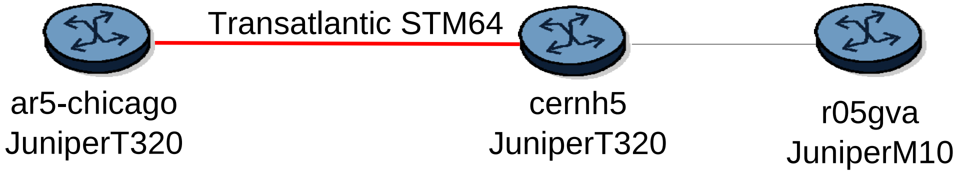
In the ***overlay model***, the optical and routed (IP) layers are separated, with minimal interaction.

MPLS at CERN

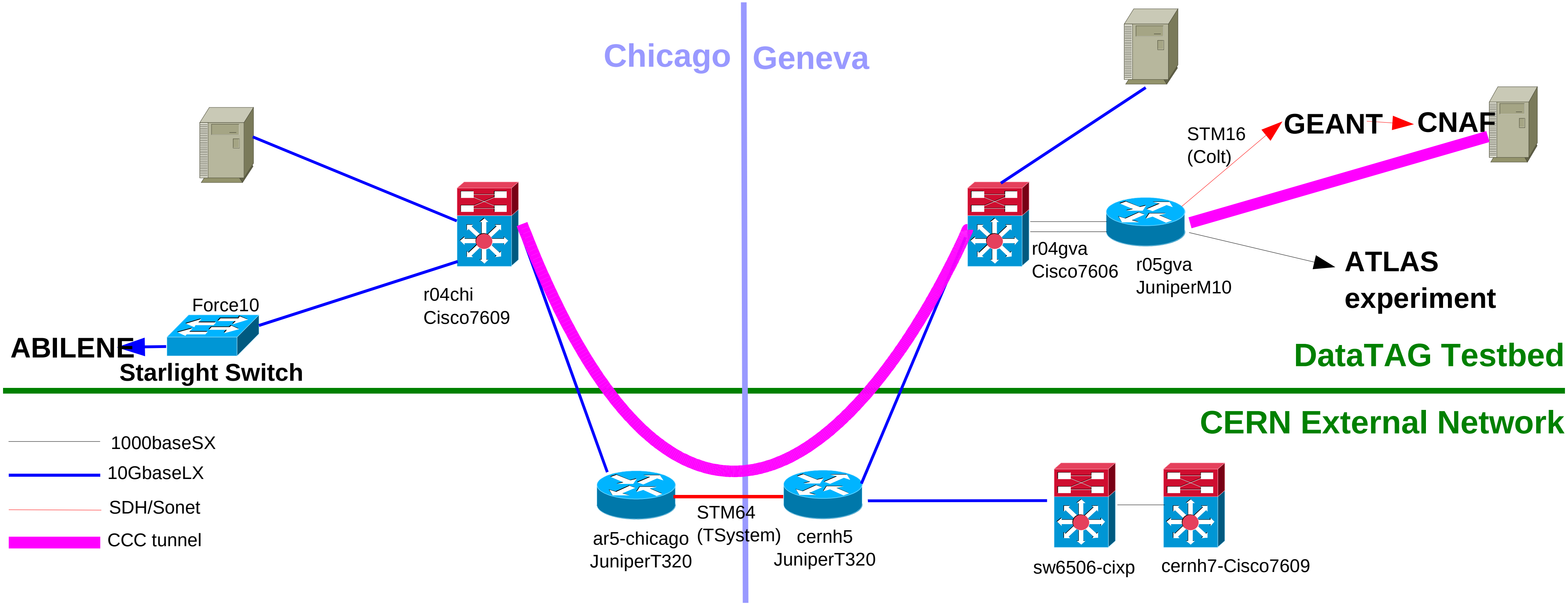
CERN has implemented MPLS in the external network, but only on the Juniper routers.

CCC is used to provision end to end Layer 2 connections on demand.

CERN MPLS backbone



MPLS at CERN: DataTAG testbed

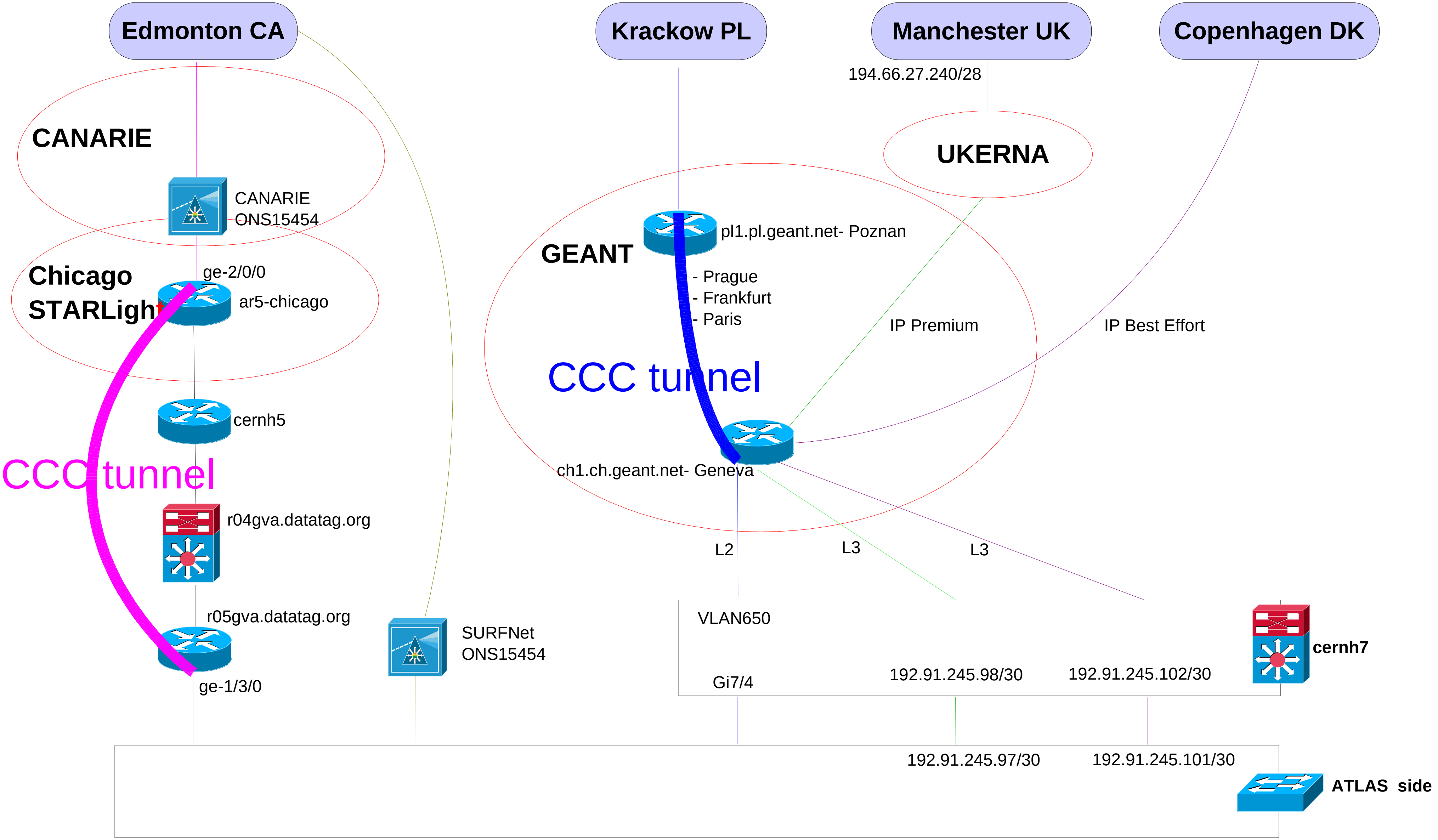


- CCC tunnel via the production network to connect the two testbeds at CERN and Chicago Starlight

- CCC tunnel via GEANT to connect the DataTAG testbed with one in Italy

- LAN distributed among Chicago, Geneva and Bologna

MPLS at CERN: ATLAS remote farms



Resources

Introduction to Juniper Networks Routers - slides.
MPLS VPNs – Telenor Nextra - slides.
http://www.mplsrc.com/
http://www.cisco.com/
http://www.juniper.net/
http://www.telecommagazine.com/default.asp?journalid=3&func=articles&page=0304t
http://www.calient.net/files/GMPLS.pdf
http://www.juniper.net/solutions/literature/white_papers/200012.pdf