

## KarBol Project

### Layer 2 Circuit between Karlsruhe and Bologna crossing DFN, GÉANT and GARR domains

Authors: Laura Leone, GARR ([laura.leone@garr.it](mailto:laura.leone@garr.it))  
Marco Marletta, GARR ([marco.marletta@garr.it](mailto:marco.marletta@garr.it))

---

#### Abstract:

*The present document describes, from GARR network perspective, the setup of an end-to-end layer 2 connection between two sites crossing different routing and administrative domains. Multidomain and interoperability issues encountered during the implementation are also described.*

---

<i>Creation date:</i>	January 01, 2004
<i>Last modified:</i>	March 11, 2004
<i>Distribution level:</i>	Public
<i>File name:</i>	GARR-04-001.PDF

---

<b>1. INTRODUCTION.....</b>	<b>3</b>
<b>2. TECHNOLOGIES TO CREATE THE L2 CIRCUIT .....</b>	<b>3</b>
<b>3. TOPOLOGY.....</b>	<b>4</b>
<b>4. IMPLEMENTATION ISSUES.....</b>	<b>5</b>
<b>5. CIRCUIT SETUP.....</b>	<b>7</b>
<b>6. SUMMARY AND FUTURE STEPS.....</b>	<b>7</b>
<b>7. TROUBLESHOOTING THE SETUP.....</b>	<b>8</b>
<b>8. ACKNOWLEDGEMENTS.....</b>	<b>9</b>
<b>9. REFERENCES.....</b>	<b>9</b>
<b>APPENDIX A – CONFIGURATIONS.....</b>	<b>10</b>
<b>1 Configuration on equipments outside GARR domain .....</b>	<b>10</b>
<b>2 GARR-side configurations.....</b>	<b>12</b>

## 1. Introduction

GRID projects and applications are more and more often requesting end-to-end capacity services, with feature and performance constraints, such as:

- possibility for software to use computing and storage resources as if they were located in the same LAN;
- a dedicated infrastructure, without the cost and the operational overhead of a dedicated link between the end sites;
- robustness of the infrastructure with respect to security, and inherent improvement of performances due to minor security measures to interpose.

In order to investigate which networking technical solutions fit the requirements, a test case has been implemented in the GARR, GEANT and DFN networks between August 2003 and February 2004 to set-up a L2 circuit, configuring for this purpose the GARR backbone. This circuit is now up and running.

The test bed is an end-to-end L2 path between an Italian site, INFN-CNAF Bologna and a German one, FZK Karlsruhe. Both sites are involved in GRID projects.

Only a strong and continuous co-operation between all the entities involved has made possible this test case to succeed.

This note describes the GARR configuration, but for completeness also the whole set-up that has been detailed in a Dante note [1].

## 2. Technologies to create the L2 Circuit

The simplest way to create an end to end connection is a dedicated circuit using an SDH channel or a single wavelength (WDM). It is very often the most expensive and less flexible solution in Europe. When the two sites are far hundreds of Kilometres and the circuit has to cross many administrative domains, the provisioning may become lengthy and difficult.

If a physical infrastructure is already in place there are technologies, which allow providing logical circuits on top of physical connections.

A layer 2 (L2) circuit is a logical or physical pipe connecting two sites using a shared physical infrastructure and data link layer. The pipe may in addition provide bandwidth assurances.

Multi Protocol Label Switching (MPLS) technology creates a packet based, connection oriented virtual circuit and it may be used to build a L2 path.

Among the others we investigated Ethernet over MPLS (EoMPLS) technology because it gives the possibility to create an Ethernet virtual local area network (VLAN) among geographically separated sites. These sites can operate together transparently over an MPLS network as they were on a common Ethernet network.

EoMPLS does not perform MAC learning or MAC look up for forwarding packets from the Ethernet interface.

MPLS forwards packets based on switching labels and not relying on IP information. The labels are assigned when the packets enter into the network at the ingress label switching router (LSR).

The virtual circuit is called a Label Switched Path (LSP), which acts as a tunnel between the ingress and the egress points. To create a bi-directional circuit two LSPs are needed, each uni-directional LSP is used to transport L2 PDUs in each direction. The signalling protocols such as reservation protocol (RSVP) and label distributed protocol (LDP) are implemented to signal LSPs. Signalled LSPs have dynamically assigned MPLS labels and are configured only on the ingress routers and not on each router all along the path.

This mechanism allows the device at the ingress point to perform a single IP address lookup on a packet and then attach one or more MPLS labels to it. The attached labels provide enough information to transport the packet along the circuit made by the concatenation of LSPs up to the egress point. The MPLS routing table is independent on the IP routing table.

The implementation of Traffic Engineering mechanism on the backbone based on MPLS and RSVP gives the possibility to route traffic on the network using informations on usable links/nodes, LSPs priority and link bandwidth booking.

Once MPLS is enabled on the switching hardware, it is possible to offer services to create L3 VPNs (RFC 2547) or L2 VPNs both point-to-point and multipoint (many drafts on this topic are being produced by IETF working groups).

### **3. Topology**

Figure 1 depicts the detail of the created circuit. Table 1 shows the administrative responsibility for creating all the LSPs which concatenate to produce the end to end circuit. The L2 circuit configuration needs to be carefully matched between all domains.

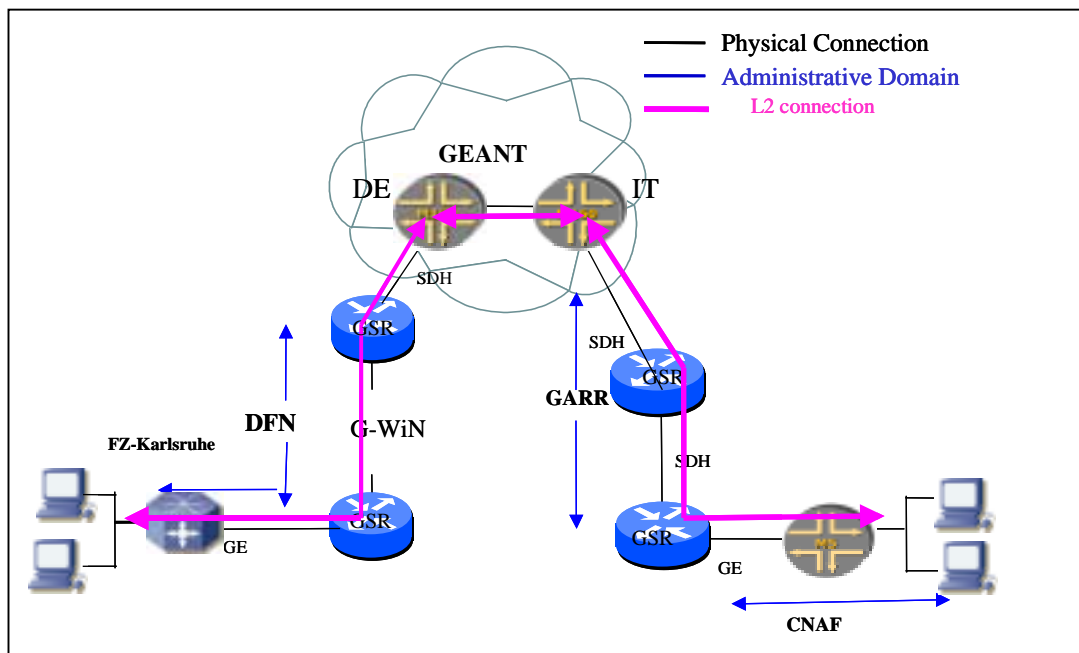


Figure 1 - Circuit topology

LSP	Name	Configuration
LSP1 Karlsruhe PE to GEANT –DE	<a href="#">Kربول_Karlsruhe_Frankfurt</a>	DFN
LSP2 GEANT-DE to GEANT –IT	<a href="#">Kربول_Frankfurt_Milan</a>	DANTE
LSP3 GEANT-IT to CNAF	<a href="#">Kربول_Milan_Cnaf</a>	DANTE
LSP4 CNAF to GEANT-IT	<a href="#">Kربول_Cnaf_Milan</a>	GARR
LSP5 GEANT-IT to GEANT-DE	<a href="#">Kربول_Milan_Frankfurt</a>	DANTE
LSP6 GEANT-DE to Karlsruhe PE	<a href="#">Kربول_Frankfurt_Karlsruhe</a>	DFN

Table 1 – LSP naming and configuration responsibilities

## 4. Implementation issues

The L2 circuit between INFN-CNAF and FZK Karlsruhe domains has been set-up crossing three additional different administrative domains. The GARR domain is based on Cisco 12400 GSRs, the GEANT domain on Juniper M160 and the GWin (DFN) one, again with Cisco 12400 GSRs. The end-sites platforms are a Juniper M5 in CNAF and a Cisco 6500 Catalyst in Karlsruhe.

The complexity of such set-up moved us to consider and investigate two main issues:

- the multidomain transit
- the interoperability between different platforms

These issues were already encountered in October 2002, in the frame of DataTag project, when a L2 circuit was realized, between INFN-CNAF in Italy and CERN in Switzerland, crossing GARR, GEANT and CERN domains.

In that case the two end-points were based on Juniper hardware and we had interoperability requirements only for the transit routers, but since MPLS and RSVP are standard we had not significant troubles. Furthermore the JunOS signalling protocol for L2circuit TE-enabled (CCC) is RSVP so we had implemented it all along the path routers.

We solved the multidomain issue through the implementation of RSVP extensions so that the signalling is made router-to-router and is MPLS compliant.

The RSVP extensions support traffic engineering features.

RSVP signalling include objects (messages) and, in particular, the explicit route object (ERO) used to tell downstream routers how to set up the LSP to the destination. The ERO can be either *strict* if the next hop is directly connected or *loose* otherwise. The use of this object allows you to force the LSP to choose particular links/nodes along the path and allow the ingress node to control the routing table of the LSP independent from the IGP preferred path. A path with a complete list of strict EROs succeeds even if the IGP is *broken*, as the ERO in this case strictly controls the routing: it verifies that the next-hop is directly connected.

The RSVP implementation described is also the one we used for KarBol project setup.

The interoperability between different vendors has been more deeply investigated in the KarBol case-study.

Cisco and Juniper, on the end-sites, both implement the draft Martini (while draft Kompella is only supported by Juniper) to create an Ethernet VLAN to be transported over an MPLS backbone.

Any Transport over MPLS (AToM) is the Cisco naming of the Draft-Martini solution for transporting L2 packets over a MPLS backbone.

AToM uses a directed Label Distribution Protocol (LDP) session between edge routers for setting up and maintaining connections. Forwarding occurs through the use of two level labels, switching between the edge routers. The external label (tunnel label) routes the packet over the MPLS backbone to the egress router at the ingress one. The VC label determines the egress interface, and it binds the L2 egress interface to the tunnel label. To make Cisco and Juniper implementations compliant we implemented LDP on Bologna end-site node and we signalled the path between domains using RSVP with ERO extensions to create the egress LSP. We deployed the “transport” of LDP in RSVP all along the path.

The MTU issue is another topic we investigated. When enabling MPLS, the size of a packet depends on the number of labels in the label stack written in front of the IP packet. The total size of one label is four bytes; the total size of a label stack is  $n \times 4$  bytes. If a label stack is formed, the frames can exceed the medium MTU (the default Ethernet frame MTU is 1514 Bytes). For this reason an Ethernet MTU size of 1526 byte has to be taken into account, adding one 4-byte LDP label and one 4-byte RSVP label to 1518-byte VLAN-tagged Ethernet frame.

Issues like Authorization and Authentication are not yet considered in this first step.

## 5. Circuit setup

The CNAF-Karlsruhe L2 bi-directional circuit is realised through the creation and concatenation of three LSPs in each direction. This concatenation of LSP is routed on a primary path. A secondary, backup path has been also pre-computed for the LSP and inserted in the MPLS routing tables. In case of primary path failure, the LSP automatically follows the secondary path.

The connections of LSPs are made on the GEANT Juniper router through a mechanism called “LSP stitching”.

This test focused on the set-up and troubleshooting of the LSPs to create a L2 circuit.

Both MPLS and RSVP have been configured all along the path in each domain, while LDP has been configured at the end-points to signal the L2 PDUs between ingress and egress routers, through the Virtual Circuit made up of the concatenation of all single LSPs. Only the routers at each end know about the created virtual circuit (VC) for transporting the L2 VLAN traffic. All other routers in the path just switch the frames based on their MPLS labels. A dedicated virtual local area network (VLAN) in each end user site has been configured, using the same value of VLAN-ID to group the hosts related to the project. The hosts in the same VLAN have IP addresses which belongs to the same subnet, at the moment a CERN assigned subnet registered for DataTag project.

To summarise, on the CE router the Ethernet frames are tagged with a VLAN-ID, are LDP labelled and finally one MPLS label is added. These frames are then transported over the L2 circuit.

## 6. Summary and future steps

We successfully configured a multidomain and multivendor L2 circuit.

Many solutions exist to solve the problem and were considered. One possibility was to create a direct LSP from one end user site to the other one, transparently crossing the GEANT domain. For monitoring reasons this solution was not chosen.

The adopted solution creates the circuit connecting (“stitching”) several LSPs, configured in each domain. This solution lets each administrative authority monitor and manage the created LSPs and impose the end-user the respect of the selected policies.

To implement such a solution, the collaboration and interaction between all involved parties is mandatory.

The future development of this research will focus on higher multiplicity of L2 circuits between different domains, including more VLANs in the set-up.

This first step of the project did not consider security issues between parties because of the complexity of such a dynamic set-up. A specific, manual agreement between domains was chosen, but it will not be generally applicable to a general service

because it does not scale. Authentication and authorisation issues have to be investigated in the future.

Another topic to be covered is the dynamic allocation of resources, based on the monitoring of the available bandwidth, using a bandwidth broker.

Other open issues are:

- the monitoring of MPLS backbone
- configurations and interoperability between vendors' protocols implementation.

At the time of writing, GARR is implementing LSPs monitoring using SNMP queries to router SNMP agents which implement new and evolving MIBs.

## 7. Troubleshooting the setup

```
lab@PICASSO> show mpls lsp
```

*Ingress LSP: 3 sessions*

To	From	State	Rt	ActivePath	P	LSPname
62.40.103.89	131.154.97.253	Up	0	path_cnaf_geantMilan	*	datatag_cnaf_milan
62.40.103.89	131.154.97.253	Up	0	path_cnaf_geantMilan	*	Karbol_Cnaf_Milan
192.91.239.253	131.154.97.253	Up	0	datatag_path_cnaf_cern	*	datatag_cnaf_cern

Total 3 displayed, Up 3, Down 0

*Egress LSP: 3 sessions*

To	From	State	Rt	Style	Labelin	Labelout	LSPname
131.154.97.253	62.40.102.23	Up	0	1 FF	3	-	Karbol_Milan_Cnaf
131.154.97.253	62.40.102.23	Up	0	1 FF	3	-	datatag_milan_cnaf
131.154.97.253	192.91.239.253	Up	0	1 FF	3	-	datatag_cern_cnaf

Total 3 displayed, Up 3, Down 0

*Transit LSP: 0 sessions*

Total 0 displayed, Up 0, Down 0

```
lab@PICASSO> show l2circuit connections
```

*L-2 Circuit Connections:*

*Legend for connection status (St)*

*EI -- encapsulation invalid    NP -- interface h/w not present*  
*MM -- mtu mismatch            Dn -- down*  
*EM -- encapsulation mismatch    VC-Dn -- Virtual circuit Down*  
*CM -- control-word mismatch    Up -- operational*  
*VM -- vlan id mismatch        CF -- Call admission control failure*  
*OL -- no outgoing label        XX -- unknown*  
*NC -- intf encaps not CCC/TCC*  
*CB -- rcvd cell-bundle size bad*

*Legend for interface status*

*Up -- operational*

*Dn -- down*

*Neighbor: 188.1.16.5*

<i>Interface</i>	<i>Type</i>	<i>St</i>	<i>Time last up</i>	<i># Up trans</i>
<i>ge-0/1/0.570(vc 1)</i>	<i>rmt</i>	<i>Up</i>	<i>Feb 19 12:17:34 2004</i>	<i>1</i>

*Local interface: ge-0/1/0.570, Status: Up, Encapsulation: VLAN*  
*Remote PE: 188.1.16.5, Negotiated control-word: No*  
*Incoming label: 100000, Outgoing label: 149*

## **8. Acknowledgements**

Thanks to Mauro Campanella, Christian Cinetto and Gloria Vuagnin for the document review, to Fabio Palozza for his precious technical support, and to everyone involved in the KarBol project at DANTE, DFN, CNAF and FZK.

## **9. References**

[1] [Bologna – Karlsruhe L2 connection by Otto Kreiter \(Dante\) - \( Dante internal note\)](#)

[2] <http://www.juniper.net/techpubs/software/junos/junos53/swconfig53-mpls-apps/html/>

[3] <ftp://ftp.ietf.org/internet-drafts/draft-martini-l2circuit-trans-mpls-13.txt>

[4] [www.cisco.com/en/US/products/sw/iosswrel/ps1829/products\\_feature\\_guide09186a008016102a.html#1130490](http://www.cisco.com/en/US/products/sw/iosswrel/ps1829/products_feature_guide09186a008016102a.html#1130490)

[5] [Building Core Networks with OSPF , BGP and MPLS Boot Camp. by Cisco System](#)

[6] [MPLS Training course documentation by Fabio Palozza \(Juniper Networks - Italy\)](#)

## Appendix A – Configurations

### 1 Configuration on equipments outside GARR domain

#### *Cisco PE router*

- Configure a loopback interface with /32 mask (optional)
- Configure RSVP on the PE on all interfaces participating in the tunnel.

*ip rsvp bandwidth*

By default 75% of the physical bandwidth of the interface is allocated, otherwise the value must be specified:

*ip rsvp [bandwidth the\_bandwidth\_in\_kilobits]*

- Configure a tunnel from the PE to the border Géant router
  - Enable mpls traffic engineering globally:

*mpls traffic-eng tunnels*

- Enable mpls traffic-engineering per interface :

*mpls traffic-eng tunnels*

- Configure a tunnel to the Géant border router:

```
interface Tunnel X  
description PE1-Géant  
ip unnumbered Loopback0  
no ip directed-broadcast  
load-interval 30  
tunnel destination [Géant-interface address-toward-the-NRN]  
tunnel mode mpls traffic-eng  
tunnel mpls traffic-eng path-option 1 explicit identifier 3 verbatim  
tunnel mpls traffic-eng record-route
```

Note: the *verbatim* command disallows the IGP checking

- Setup an explicit path:

```
ip explicit-path identifier 3 enable  
next-address yyyy  
next-address [Géant-interface address-toward-the-NRN]
```

- Configure a pseudo-wire class with tunnel selection

```
pseudowire-class www
encapsulation mpls
preferred-path TunnelX
```

- Configure the interface vs CE router

```
int gigabitethernet1/0.XX
encapsulation dot1Q XX
xconnect [The_Loopback_address_of_the_remote_PE] 1 pw-class www
```

Note: 1 is the VC ID

### ***Juniper router not involved in the stitching (P router)***

- Enable rsvp

```
rsvp {
  interface all;
}
```

- Enable mpls

```
mpls {
  interface all;
}
```

### ***Juniper router involved in the stitching***

- Disallow the IGP checking

```
no-cspf
```

- Configure the LSPs between Géant routers

```
mpls {
  label-switched-path Géant1-Géant2 {
    to [loopback_address_of_the_Géant2_router];
  }
}
```

- Configure the LSP between Géant and PE routers

```
mpls {
  label-switched-path Géant1-PE1 {
    from [interface_address_toward_NRN];
    to [PE1_loopback_address];
    no-cspf;
    primary path-to-PE1;
  }
  path path-to-PE1 {
    ....
    [ip_addresses_of_the_hops] strict;
    ....
  }
}
```

}

}

- Stitching together LSPs

```
connections {
  lsp-switch PE1-Géant1-Géant2 {
    transmit-lsp Géant1-Géant2;
    receive-lsp PE1-Géant1;
  }

  lsp-switch Géant2-Géant1-PE1 {
    transmit-lsp Géant1-PE1;
    receive-lsp Géant2-Géant1;
  }
}
```

## 2 GARR-side configurations

### *CNAF Juniper router (CE)*

<JunOS 5.6R2.4>

```
lab@PICASSO>
ge-0/0/0 {
  description "to Cisco 12416 GARR BO";
  unit 0 {
    family inet {
      address 193.206.128.38/30;
    }
    family mpls;
  }
}
```

- Configuration vs LAN

```
ge-0/1/0 {
  description "to switch 3com CNAF";
  vlan-tagging;
  encapsulation vlan-ccc;
}
unit 570 {
  encapsulation vlan-ccc;
  vlan-id 570;
}
```

- RSVP configuration

```
protocols {
  rsvp {
    interface all;
  }
}
```

- MPLS configuration

```
protocols {
  mpls {
    interface all;
  }
}
```

- LSP definition

```
label-switched-path Karbol_Cnaf_Milan {
  from 131.154.97.253;
  to 62.40.103.89;
  install 188.1.16.5/32;
  no-cspf;
  primary path_cnaf_geantMilan;
  secondary path_cnaf_geantMilan_10G;
}
```

- PATH definition

```
path path_cnaf_geantMilan {
  193.206.128.37 strict;
  193.206.134.21 strict;
  62.40.103.89 strict;
}
path path_cnaf_geantMilan_10G {
  193.206.128.37 strict;
  193.206.134.21 strict;
  62.40.103.189 strict;
}
```

- LDP configuration

```
ldp {
  interface ge-0/0/0.0;
  interface lo0.0;
}
```

- Stitching configuration

```
connections {
  remote-interface-switch CCC-CNAF-KARLSRUHE {
    interface ge-0/1/0.570;
    transmit-lsp Karbol_Cnaf_Milan ;
    receive-lsp Karbol_Milan_Cnaf;
  }
}
```

- L2 neighbor definition

```

l2circuit {
  neighbor 188.1.16.5 {
    interface ge-0/1/0.570 {
      virtual-circuit-id 1;
      control-word;
    }
  }
}

```

### **Bologna Cisco Router (P)**

<IOS 12.0(25)S1>

```

!
hostname RTG_BOLOGNA
!
mpls ldp logging neighbor-changes
mpls traffic-eng tunnels
no mpls traffic-eng auto-bw timers frequency 0
!
interface POS2/0
description Bologna - Milano 2.5 Gbps (TD 029750/07)
ip address 193.206.134.22 255.255.255.252
mpls traffic-eng tunnels
 ip rsvp bandwidth 1200000 1200000
 ip rsvp signalling hello
!
!
interface GigabitEthernet3/0/1
description TEST DATATAG CNAF
mtu 4470
ip address 193.206.128.37 255.255.255.252
mpls traffic-eng tunnels
 ip rsvp bandwidth 1000000 1000000
 ip rsvp signalling hello
!
router ospf 137
mpls traffic-eng router-id Loopback0
mpls traffic-eng area 0
!
ip rsvp signalling hello
ip rsvp signalling hello statistics
!

```

### **Milano Cisco Router (P)**

<IOS 12.0(25)S1>

```

!
hostname RTG_MILANO
!
mpls ldp logging neighbor-changes

```

```
mpls traffic-eng tunnels
no mpls traffic-eng auto-bw timers frequency 0
!
interface POS12/0
description Milano - Bologna 2.5 Gbps (TD 029750/07)
ip address 193.206.134.21 255.255.255.252
mpls traffic-eng tunnels
tag-switching mtu 4458
ip rsvp bandwidth 1200000 1200000
ip rsvp signalling hello
!
interface POS12/2
description GEANT access 2.5Gbps
ip address 62.40.103.90 255.255.255.252
mpls traffic-eng tunnels
tag-switching mtu 4458
ip rsvp bandwidth 1000000 1000000
ip rsvp signalling hello
!
router ospf 137
router-id 193.206.129.251
mpls traffic-eng router-id Loopback0
mpls traffic-eng area 0
!
ip rsvp signalling hello
ip rsvp signalling hello statistics
!
```