# Multi-domain VPNs

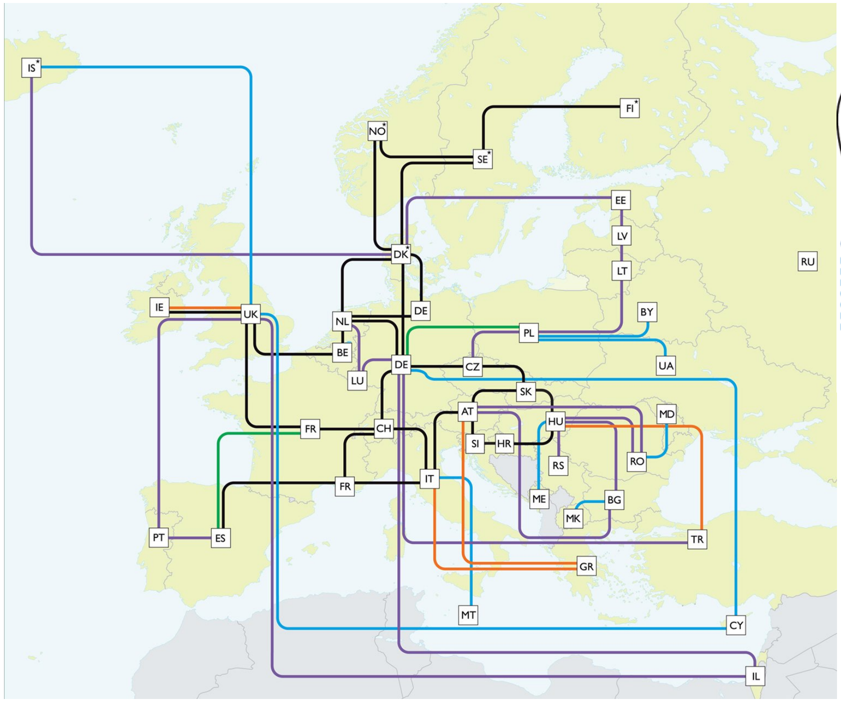
A practical approach to enable end-to-end services over multiple domains

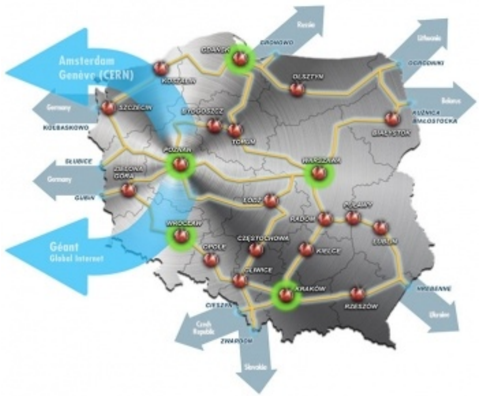
DENOG7, Darmstadt

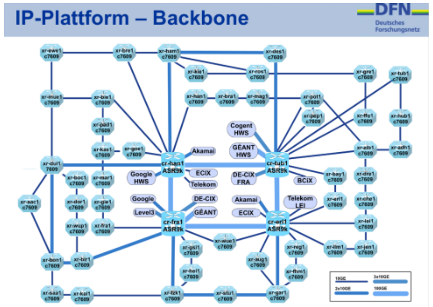
Thomas Schmid, schmid@dfn.de

# The research network landscape

GÉANT



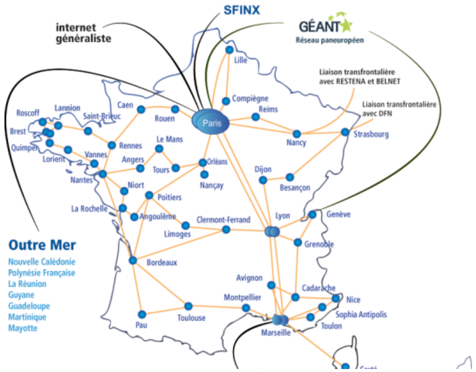


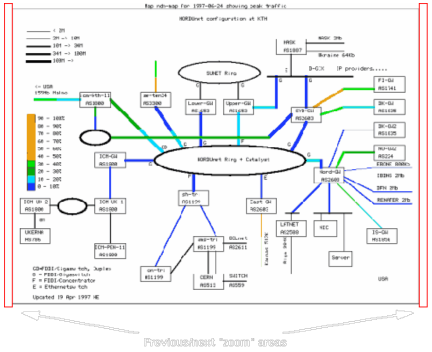


NRENs are in general interconnected via the GÈANT network.
No end-users are connected to GÈANT.

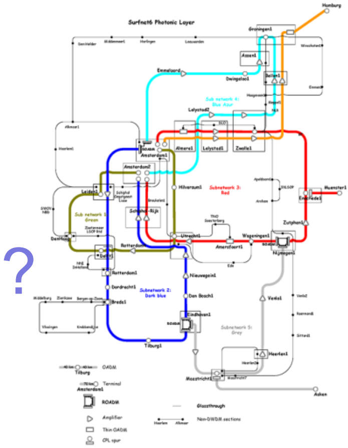
# The NREN challenge

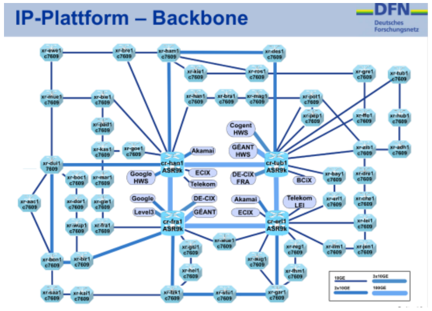
- All NRENs are created unequal
  - Multi-vendor
  - Pure IP
  - IP+MPLS
  - PBB
  - MPLS-TP
  - MEF
  - Transport technologies
  - …





Amsterdam
Genève (CERN)
Géant
Global Internet
Germany
Russia
Lithuania
Belarus
Ukraine
Czech Republic
GDAŃSK
KOSZALIN
OLSZTYN
SZCZECIN
BYDGOSZCZ
TORUŃ
BIAŁYSTOK
POZNAŃ
ZIELONA GÓRA
WARSZAWA
ŁÓDŹ
WROCŁAW
RADOM
PUŁAWY
LUBLIN
OPOLE
CZĘSTOCHOWA
KIELCE
GLIWICE
KRAKÓW
RZESZÓW
CIESZYN



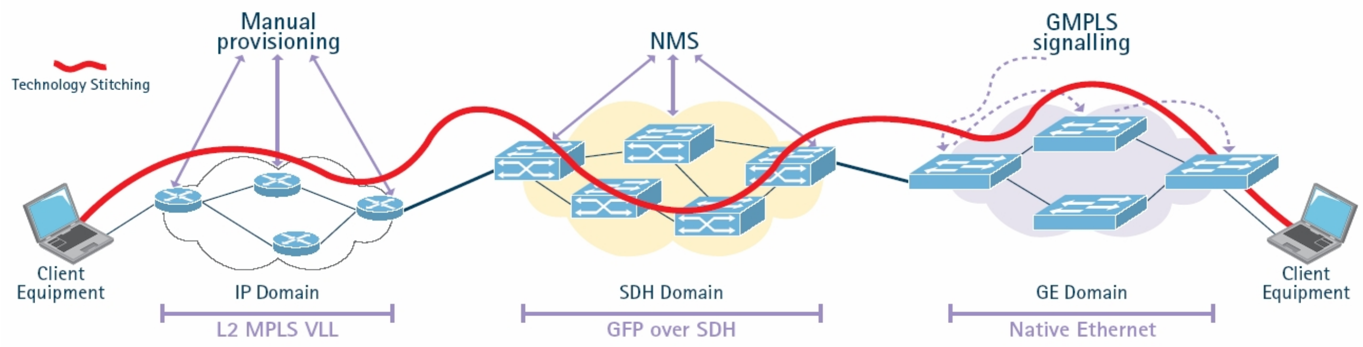


How to offer private e2e services?

# A brief history of private inter-domain connections

- 90s:
  - ATM SVCs, SDH: Not operated by the NRENs
- 00s: NG-SDH, Ethernet, MPLS back-to-back, MPLS-TE tunnel stitching
- 10s: Lambdas, OTN, Ethernet, MPLS ubiquitous

Example: BoD (Bandwidth on Demand)



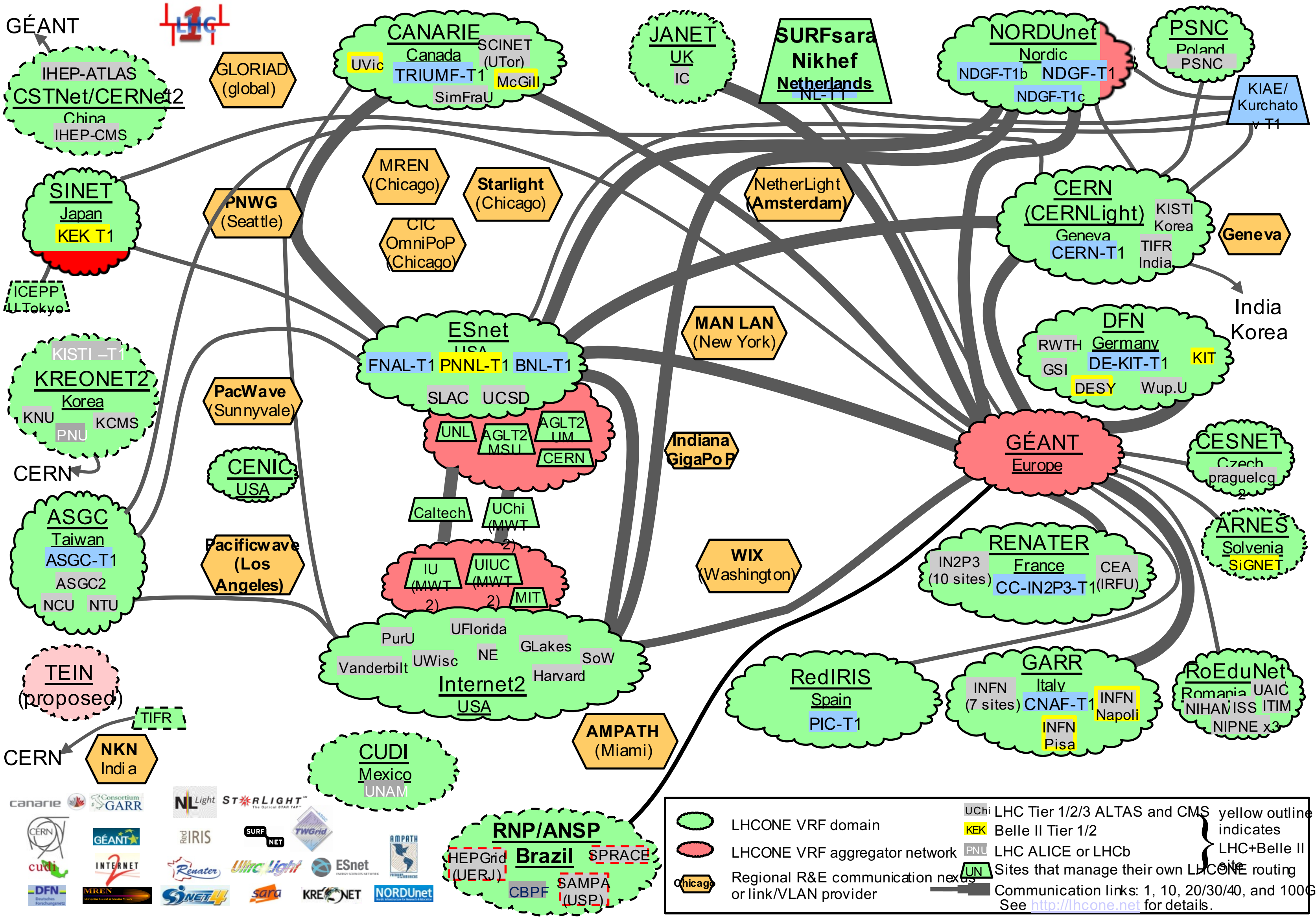
Complex: Topology databases, PCEs etc. http://services.geant.net/bod/Pages/Home.aspx

Stitching technologies ☹

# Example: LHCONE

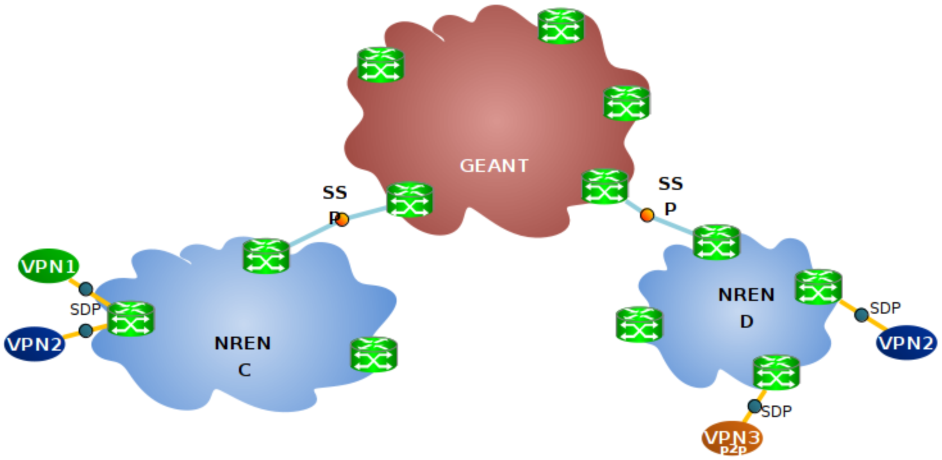
- LHCONE: Large Hadron Collider Open Network Environment
- Private Network to distribute data from the large hadron collider at CERN among data centers (↔ LHCOPN mostly for traffic CERN-Tier1 data centers)
- One VRF per domain
- Domains interconnected via normal IP, no labels involved: back-to-back VPNs (→ no support for L2VPNs)
- In some parts separate physical/logical infrastructure reserved for LHCONE traffic

# LHCONE: A global infrastructure for the High Energy Physics (LHC and Belle II) data management



28 May 2015 – WEJohnston, ESnet, wej@es.net

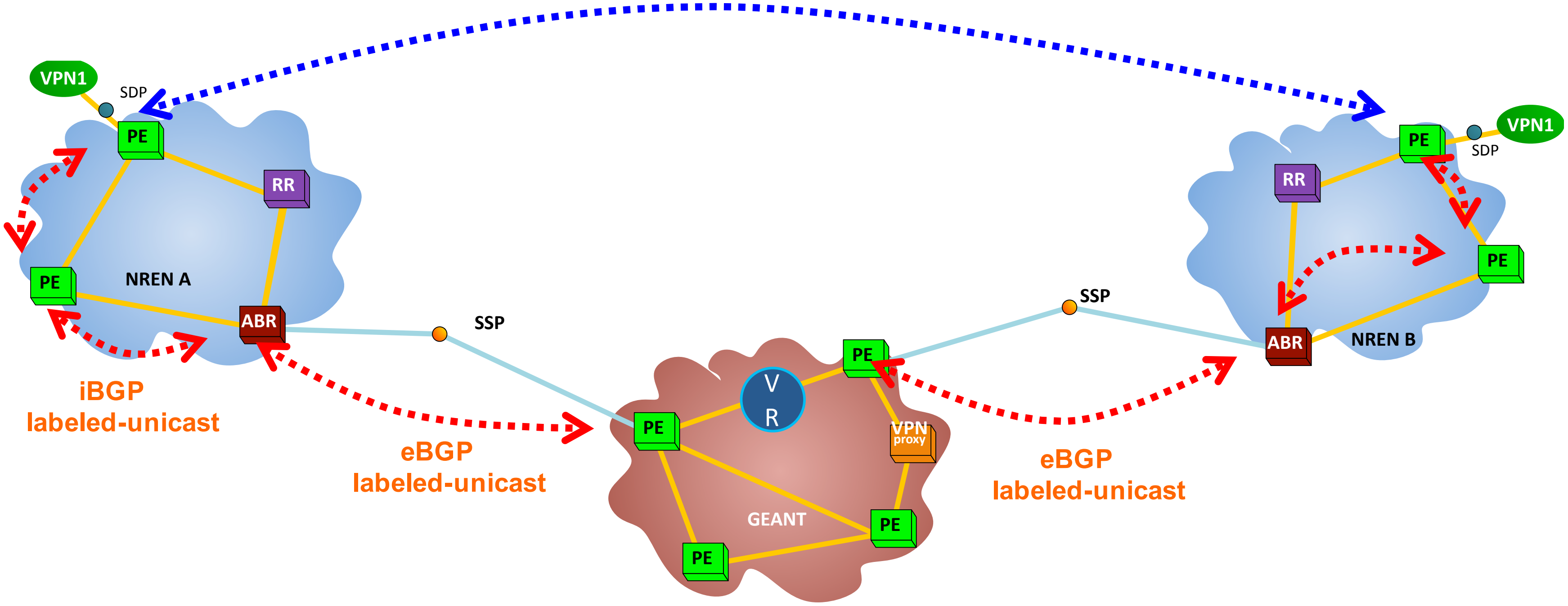
# Carrier-support-carrier for hierarchical VPNs



- RFC4364 Option 10.c (2006!)
- Means to provide seamless end-to-end MPLS services over multiple domains
- No stitching
- Hierarchical architecture: GÈANT is Carrier-of-Carrier
- No CAPEX
- Supported on almost all router hardware
- → MDVPN: multidomain VPN
- But: no user community
  - No large scale implementation according to vendors

# MDVPN: tLDP-signalling L2 circuit

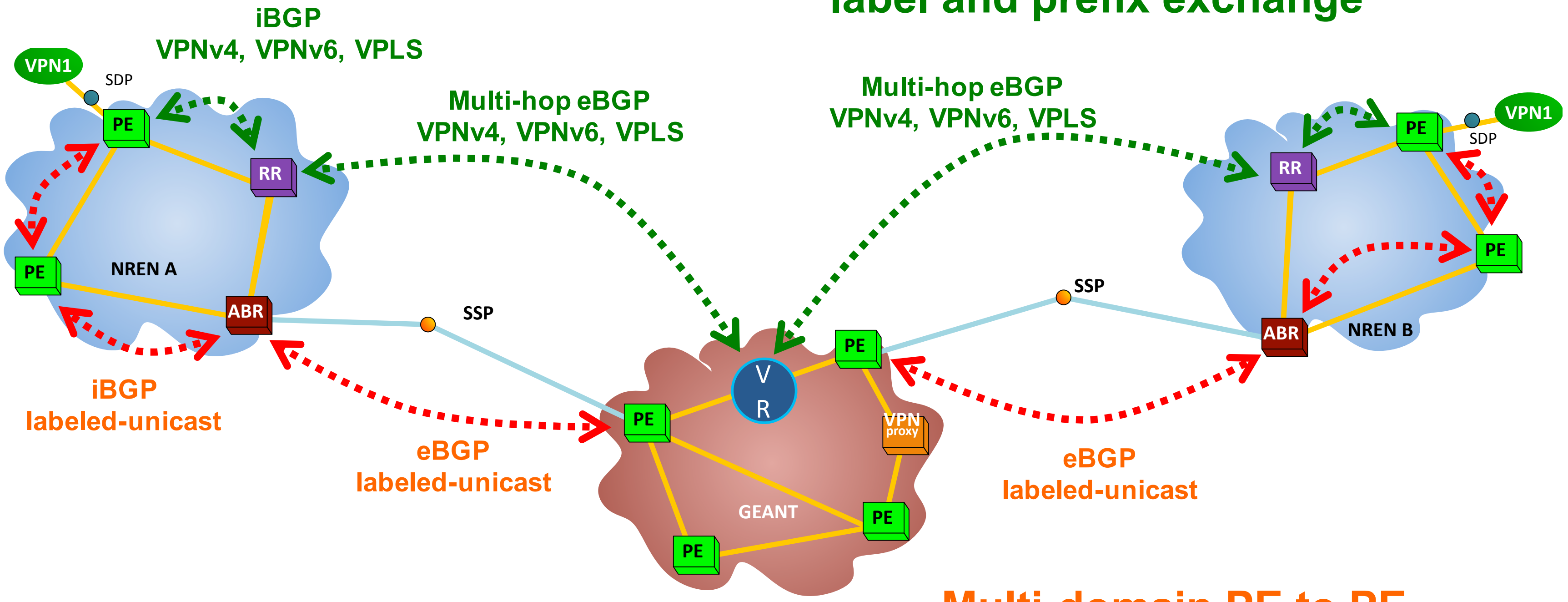
Targeted LDP -signaled L2 circuit
label exchange



Multi-domain PE to PE MPLS path

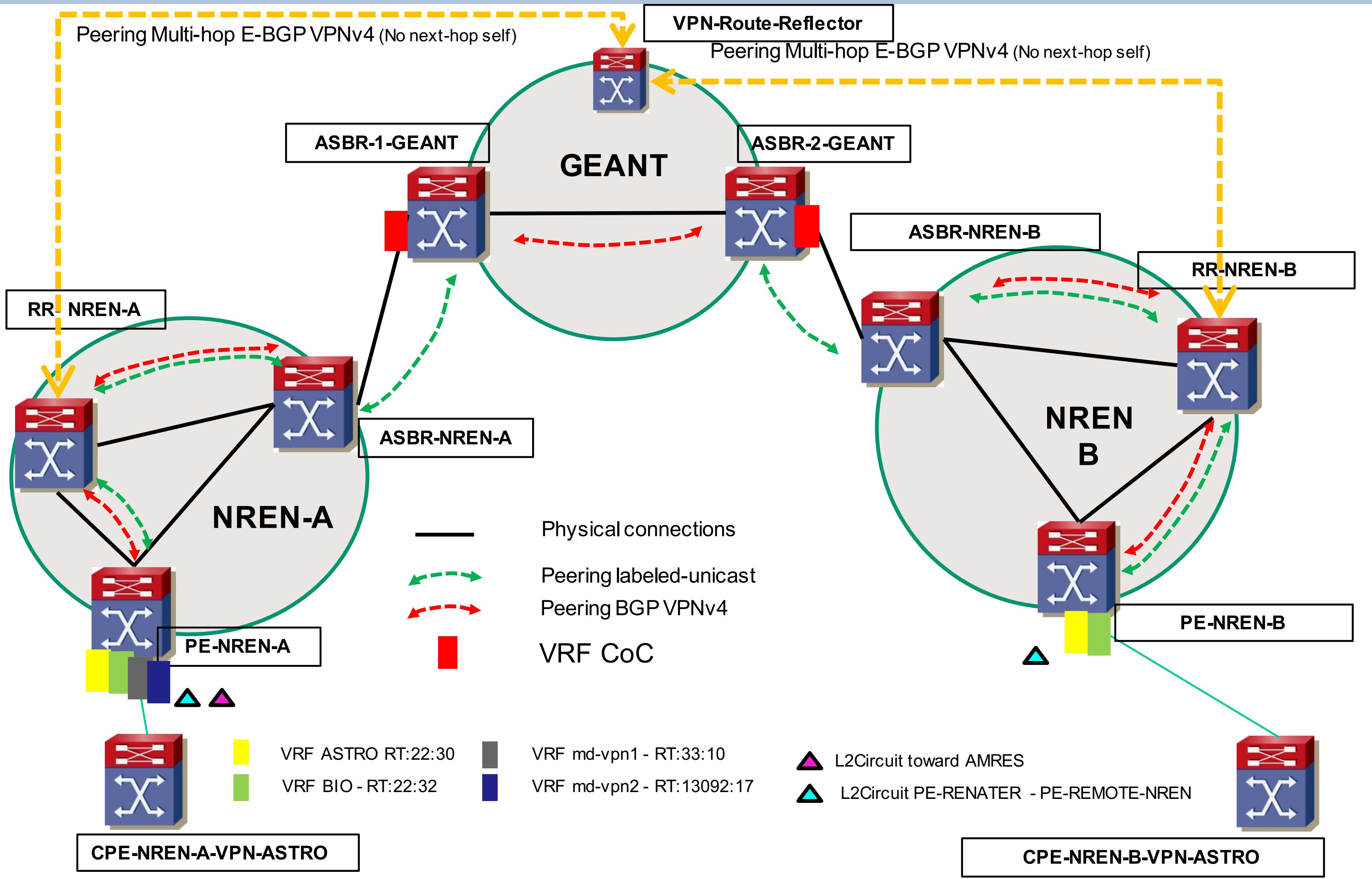
# MDVPN: BGP-signalling L2VPN, L3VPN

BGP-signaled L2VPN and L3VPN label and prefix exchange



Multi-domain PE to PE MPLS path

# Standard deployment



Peering Multi-hop E-BGP VPNv4 (No next-hop self)

VPN-Route-Reflector

Peering Multi-hop E-BGP VPNv4 (No next-hop self)

| | |
|---|---|
| — | Physical connections |
| ⇠⇢ | Peering labeled-unicast |
| ⇠⇢ | Peering BGP VPNv4 |
| 🟥 | VRF CoC |

| | | | |
|---|---|---|---|
| 🟨 | VRF ASTRO RT:22:30 | ■ | VRF md-vpn1 - RT:33:10 |
| 🟩 | VRF BIO - RT:22:32 | 🟦 | VRF md-vpn2 - RT:13092:17 |

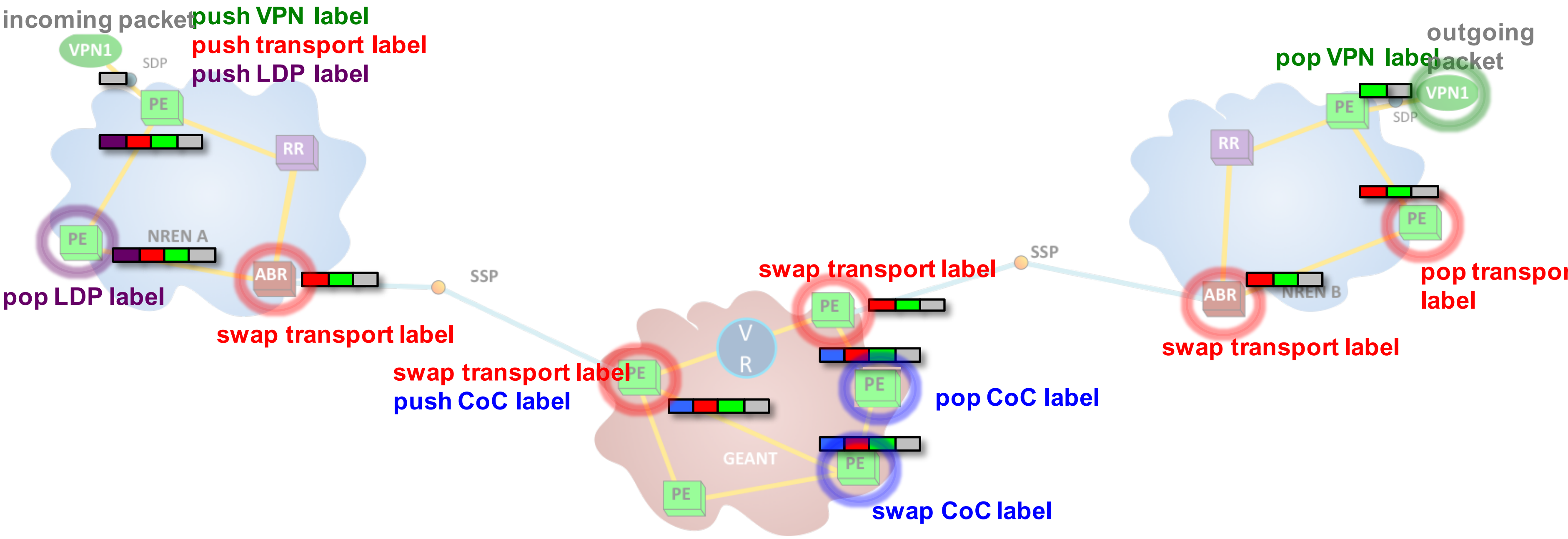
△ L2Circuit toward AMRES

△ L2Circuit PE-RENATER - PE-REMOTE-NREN

# In short

- GÈANT: Carrier-of-Carrier
  - only sees the /32s of the PEs with labels
  - Transparent to configured VPNs between NRENs
  - MDVPN runs in separate VRF (for monitoring/accounting purposes)
- ASBR-ASBR BGP LU session: distribute Loopback addresses (/32s) of PEs with labels
  - No LDP required here
- VPN route-reflector: distribute BGP routes used e.g. in L3VPNs
  - Signalling: not in the forwarding path - Could be anywhere
  - For practical reasons run by GÉANT
- Traffic uses shared infrastructure
  - Logical separation in VRF over VLAN on ASBR
  - Dedicated infrastructures or bandwidth reservation optional
- Easy to extend into regional metronets

# MDVPN data plane label operations



incoming packet

push VPN label
push transport label
push LDP label

pop LDP label

swap transport label

swap transport label
push CoC label

swap transport label

pop CoC label

swap CoC label

swap transport label

pop VPN label

outgoing packet

pop transpor
label

MDVPN packets labels:

| | | | |
|---|---|---|---|
| LDP label | Transport label | VPN label | Data |
| CoC label | Transport label | VPN label | Data |

With the courtesy of Jani Myyry **(Funet)**

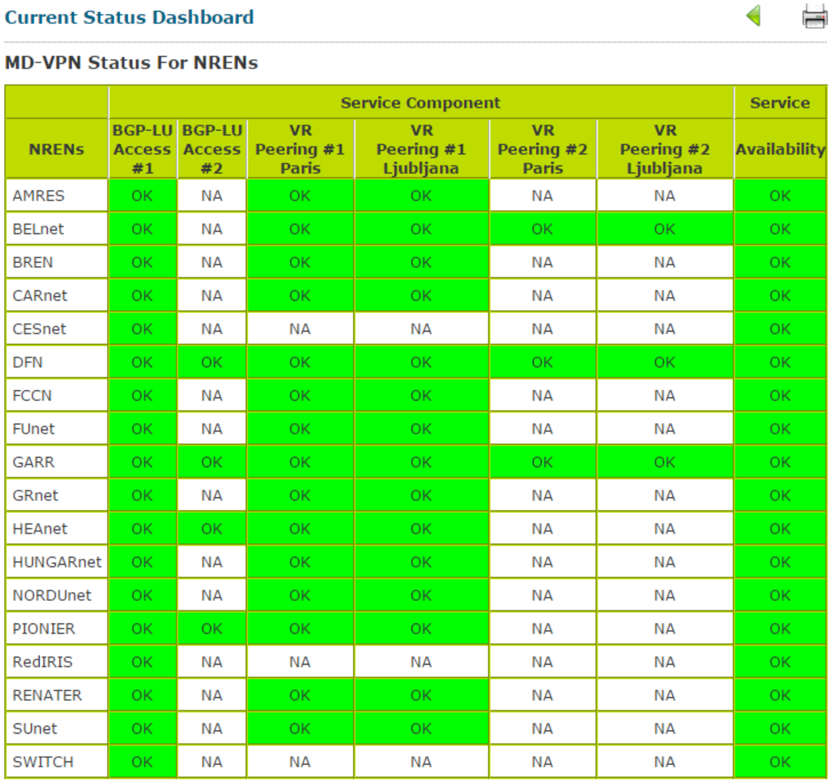
# Operation

Implement new service: one phone call and then…

```
routerA#conf t
routerA(conf)>interface TengigE1/1
routerA(conf-if)>xconnect <IP of remote PE> 123 encap mpls
```

Done ☺

- Great tool to easily deploy VPN services
  - Technology transparent for customers
- Support for all kind of VPN technologies
  - L2 VPN ☑
  - L3 VPN incl. 6VPE ☑
  - VPLS ☑
    - Even with autodiscovery ☑
  - EVPN (currently testing – looks good)
  - Multicast: in theory yes
- Implementation of new services over multiple domains is as easy as in the own domain
- Monitoring:
  - Signalling plane: routing protocols
  - Forwarding plane: ping-VPN (PEs)



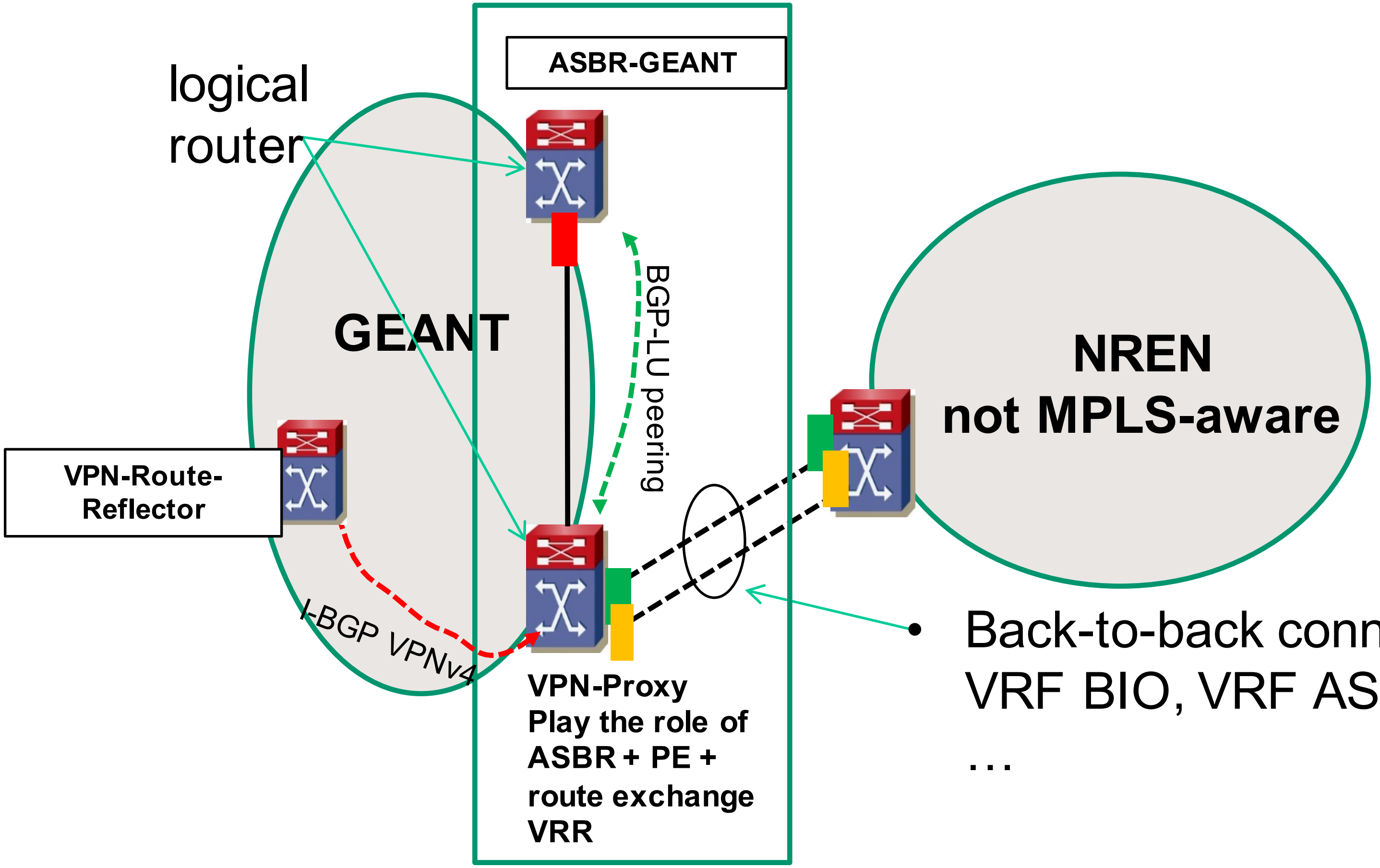
Current Status Dashboard

MD-VPN Status For NRENs

| | Service Component | | | | | | Service |
|---|---|---|---|---|---|---|---|
| NRENs | BGP-LU Access #1 | BGP-LU Access #2 | VR Peering #1 Paris | VR Peering #1 Ljubljana | VR Peering #2 Paris | VR Peering #2 Ljubljana | Availability |
| AMRES | OK | NA | OK | OK | NA | NA | OK |
| BELnet | OK | NA | OK | OK | OK | OK | OK |
| BREN | OK | NA | OK | OK | NA | NA | OK |
| CARnet | OK | NA | OK | OK | NA | NA | OK |
| CESnet | OK | NA | NA | NA | NA | NA | OK |
| DFN | OK | OK | OK | OK | OK | OK | OK |
| FCCN | OK | NA | OK | OK | NA | NA | OK |
| FUnet | OK | NA | OK | OK | NA | NA | OK |
| GARR | OK | OK | OK | OK | OK | OK | OK |
| GRnet | OK | NA | OK | OK | NA | NA | OK |
| HEAnet | OK | OK | OK | OK | NA | NA | OK |
| HUNGARnet | OK | NA | OK | OK | NA | NA | OK |
| NORDUnet | OK | NA | OK | OK | NA | NA | OK |
| PIONIER | OK | OK | OK | OK | NA | NA | OK |
| RedIRIS | OK | NA | NA | NA | NA | NA | OK |
| RENATER | OK | NA | OK | OK | NA | NA | OK |
| SUnet | OK | NA | OK | OK | NA | NA | OK |
| SWITCH | OK | NA | NA | NA | NA | NA | OK |

# VPN-Proxy implementation

- Solution for NRENs that don't support MPLS in their network
- Implemented with the help of logical routers available in Juniper



- Back-to-back connection, VRF BIO, VRF ASTRO, …

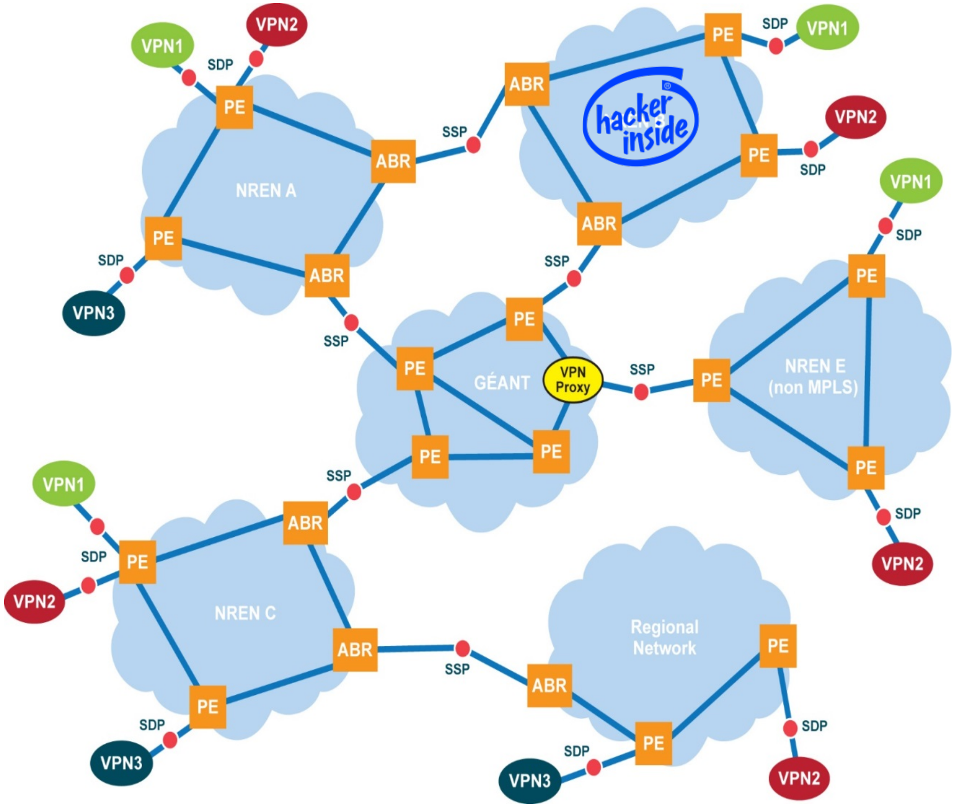
# Gory details

- MTU discovery not working
  - Juniper doesn‘t signal MTU to Cisco
- Control label distribution between own network and GÉANT
  - Internal: labels for Loopbacks in IGP ↔ BGP towards GÉANT
- E.g. IOS-XR: wtf - „ebgp-multihop mpls“ required on CRS-1, not on ASR (took the TAC one month)
- IOS-XR needs static hostroute on ASBR interface for conected ASBR address
  - LSPs must always be built on /32s
- Don‘t change next-hop
- VPLS site-IDs: different formats, no autonegotiation
- Security
  - BGP Signalling standard security mechanisms
  - Limit targeted LDP Sessions: difficult on Cisco → use packet filters on ASBR (not very elegant compared to Juniper: implicit deny)
- Missing filter options for inner labels between domains

# Attack scenario

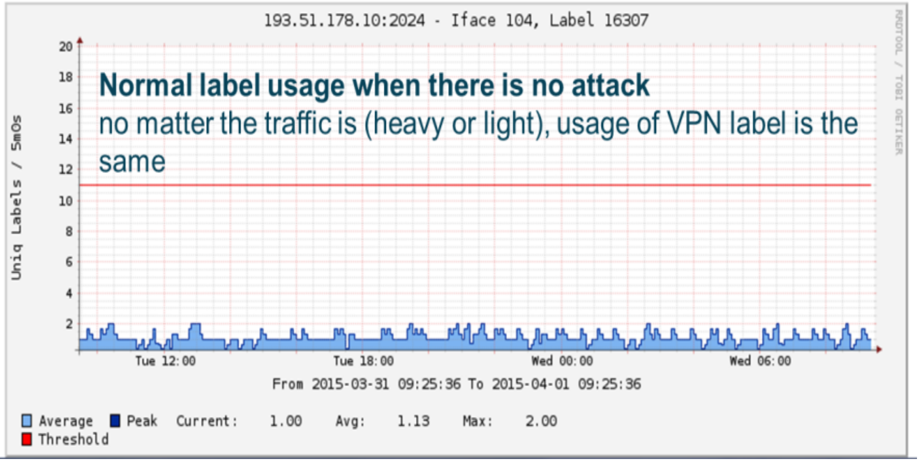
- MDVPNs are all in the same trust domain
- But: internal VPNs are vulnerable too!
  - Unless they‘re on a separate infrastructure
- Attacker has to:
  - Control a router in an NREN
  - Guess the inner VPN label
  - Guess the IP addresses in the attacked VPN
- Then he can inject packets into the internal VPN
  - Will he ever know it worked?
  - Do the usual hacking stuff
  - Perhaps will even get a response

⇨Takes a large amount of packets!

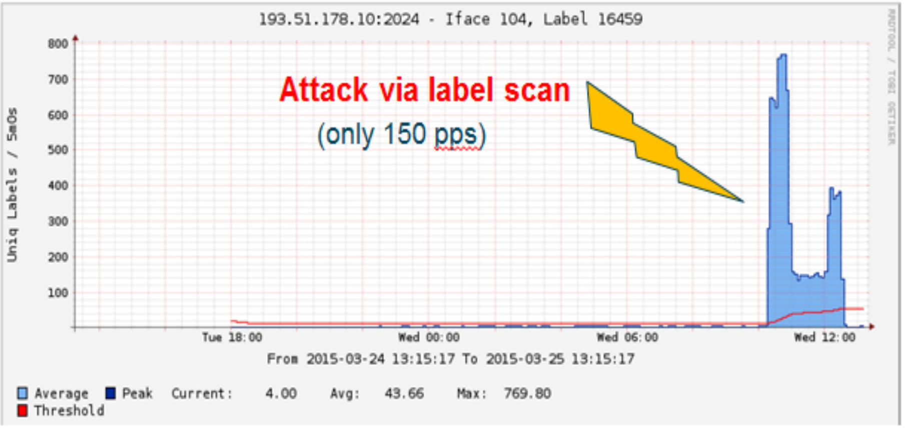


# Dealing with attacks

- Vendors don‘t support filters for inner labels
  - Also hard to keep track of internal inner label usage
- Therefore try to detect the attack and take appropriate measures
  - E.g. automatic shut down BGP LU peering with NREN
- Analyze netflow data (e.g. on GÉANT ASBR):



**Normal label usage when there is no attack**
no matter the traffic is (heavy or light), usage of VPN label is the same

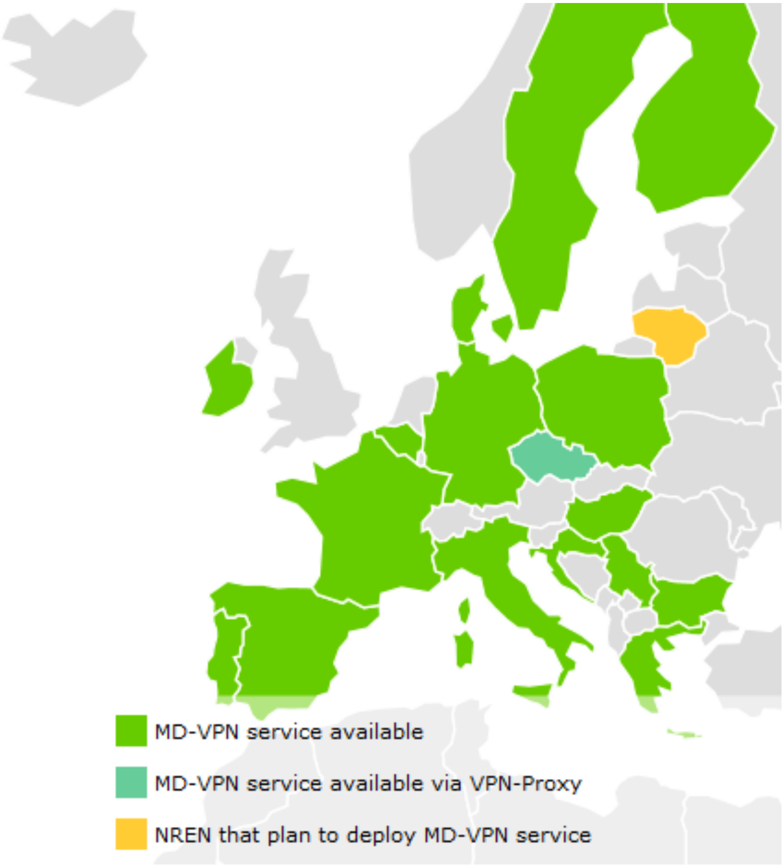


**Attack via label scan**
(only 150 pps)

- **2015/03/25 10:21:39 ALARM 193.51.178.10:29770 (#49), interface 104, label {16459 0}, threshold reached, 409 unique labels, 13 labels is allowed**
- **2015/03/25 10:21:39 ALARM 193.51.178.10:2024 (#17), interface 104, label {16459 0}, threshold reached, 416 unique labels, 13 labels is allowed**

# Deployment status and outlook

- 18 NRENs connected
- More than 450 PEs


🟩 MD-VPN service available
🟩 MD-VPN service available via VPN-Proxy
🟨 NREN that plan to deploy MD-VPN service

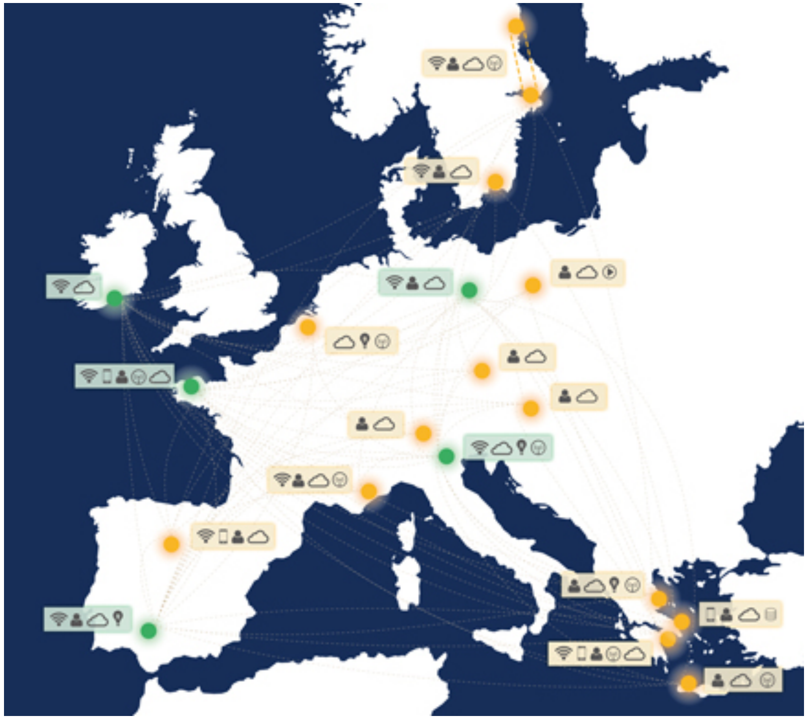
Future development:

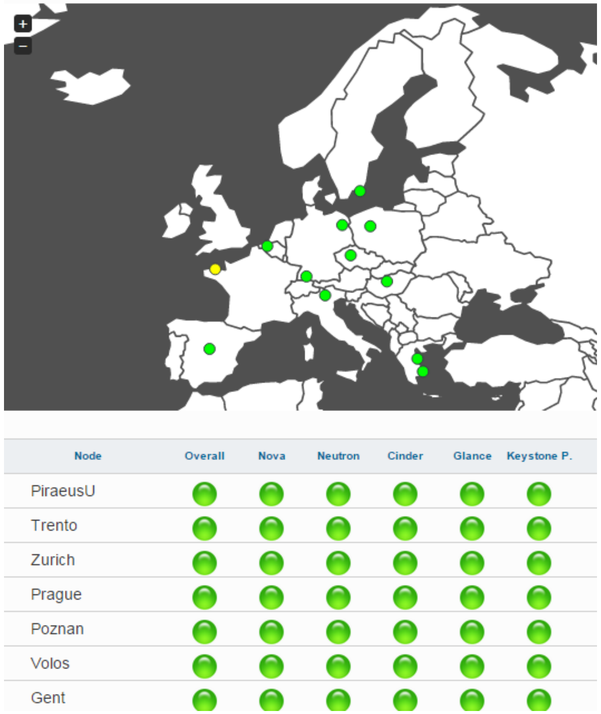
- „last mile problem“: crossing the campus network to reach the researchers
  - NTTL: network-to-the-lab. Small router using downstream label on demand with tunnels.
- Automation
- Integration with other services
  - E.g. Science DMZ
- EVPN
- ASBR inner label filter (cooperation with DELL)

# XiFi: A scientist project using MD-VPN for production

- **16 sites connected in 12 countries**

  https://www.fi-xifi.eu/federation.html
- Using all types of connection:
  - Direct connection
  - Via VPN-Proxy
  - Private companies not connected to any NREN





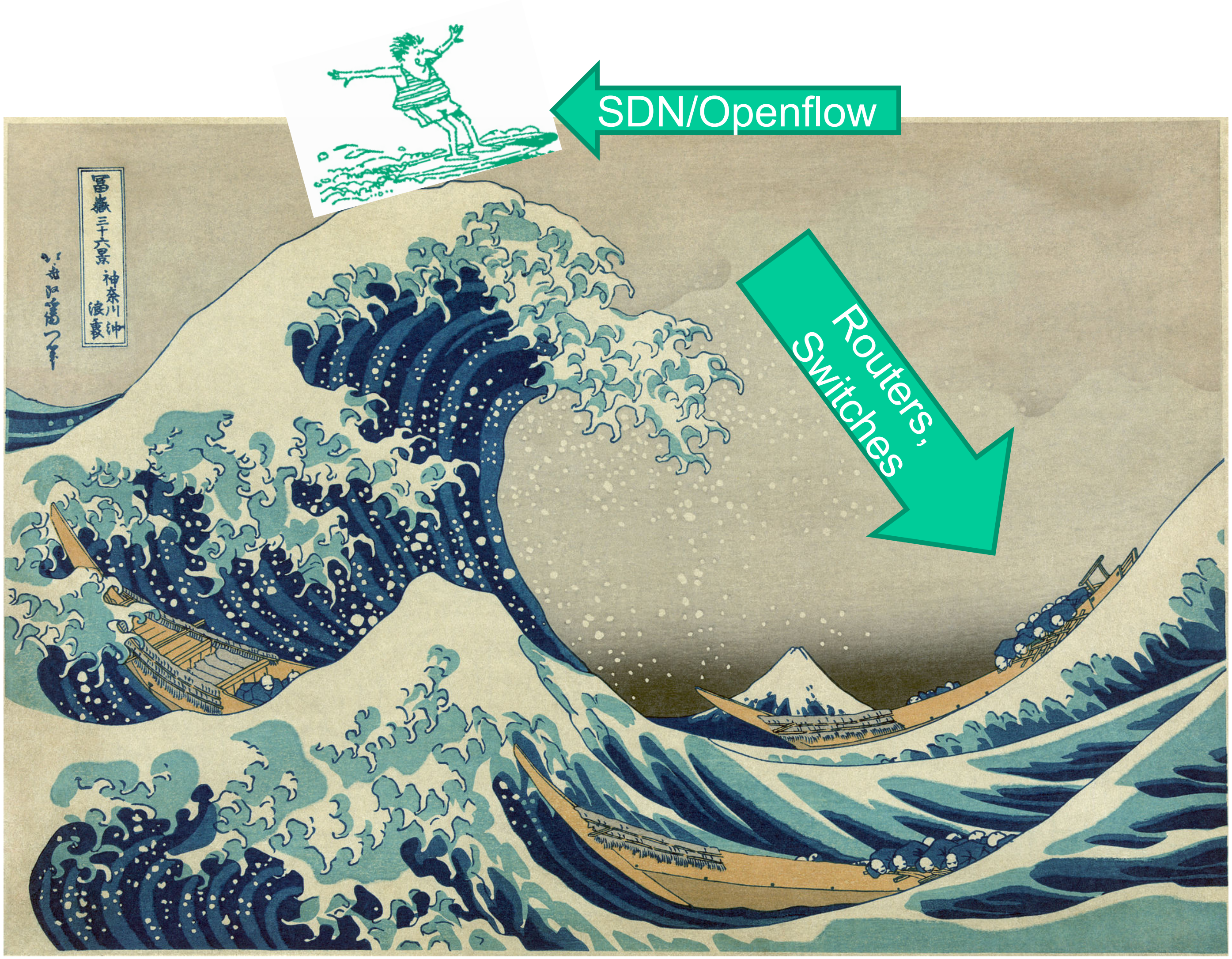
| Node | Overall | Nova | Neutron | Cinder | Glance | Keystone P. |
|---|---|---|---|---|---|---|
| PiraeusU | ● | ● | ● | ● | ● | ● |
| Trento | ● | ● | ● | ● | ● | ● |
| Zurich | ● | ● | ● | ● | ● | ● |
| Prague | ● | ● | ● | ● | ● | ● |
| Poznan | ● | ● | ● | ● | ● | ● |
| Volos | ● | ● | ● | ● | ● | ● |
| Gent | ● | ● | ● | ● | ● | ● |

http://infographic.lab.fi-ware.org/status

A first scientist project FIWARE

FIWARE is a project of the European Public-Private-Partnership on Future Internet (FI-PPP) programme

# Future?



SDN/Openflow

Routers, Switches

# The team

Work carried out with support from EU (GN3 project SA3T3)

**A small team, very small amount of manpower … but highly motivated and skilled**

- Tomasz Szewczyk (PSNC)
- Thomas Schmid (DFN)
- Magnus Bergroth (NORDUnet)
- Daniel Lete (HEAnet)
- Carlos Friacas (FCCN)
- Jani Myyry (Funet)
- Bojan Jakovljevic (AMRES)
- Miguel Angel Sotos (RedIRIS)
- Niall Donaghy (DANTE)
- Xavier Jeannin (RENATER)

- With the support of
  Brian Bach Mortensen (DiEC)

# QUESTIONS?