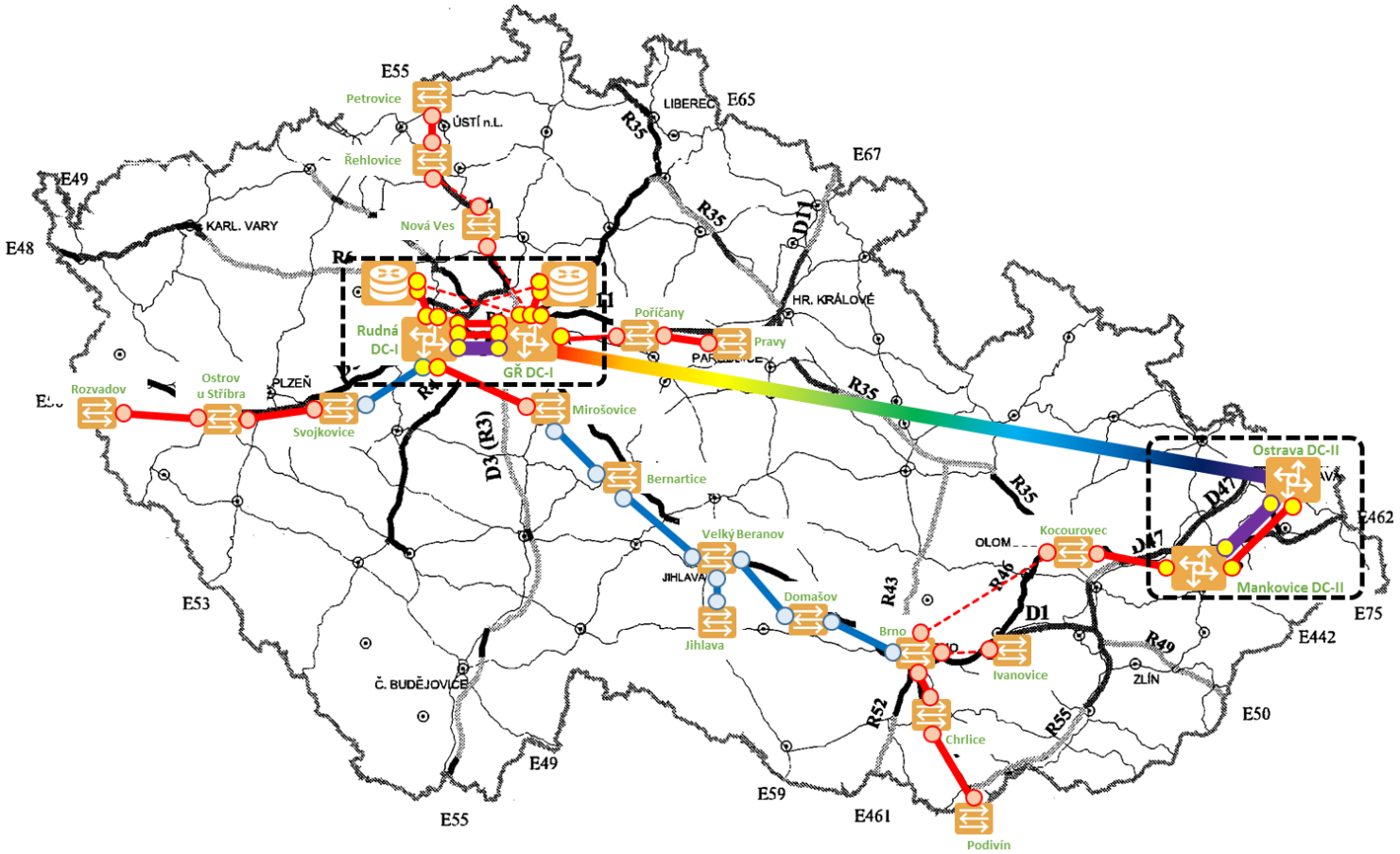
Technické zajímavosti z implementace MPLS VPN s multicast VPN

Radim Roška / radim.roska@huatech.cz

CSNOG 2019

About project

- Topology follows highway network
- Interconnect RSD remote sites distributed around highways
- MPLS services – mainly L3VPN and multicast
- Services:
  - Internal RSD data services (data, voice, etc)
  - Multicast traffic from cameras



HUATECH

Technical requirements

- Basic requirements
  - Various L3 VPN services between all sites and core DCs (green field MPLS network)
  - Cable-cuts are regular during road works ☺ - design solution to overcome this
  - High reliability is needed (traffic monitoring, especially in tunnels)
- Starting with RFP requirements and evolving during discussions with customer – some interesting extra requirements:
  - Multicast inside L3VPN service
  - Way more complicated routing on some of the sites
  - L2 encryption on some links
  - Various RPs (rendezvous points) for any-source multicast & RP redundancy

Deployed platform

- Challenging requirements – router could be a first choice
- Core sites – chassis switch S12700 (100G between core sites)
- For all other sites - high-end GE switch surprisingly matched all items in RFP and gave us edge over competition
- We selected very flexible switch - S5720-HI
  - GE access SW (combo interfaces)
  - 4x 10G uplinks further extended by 4x 10G in extension card

Technical introduction

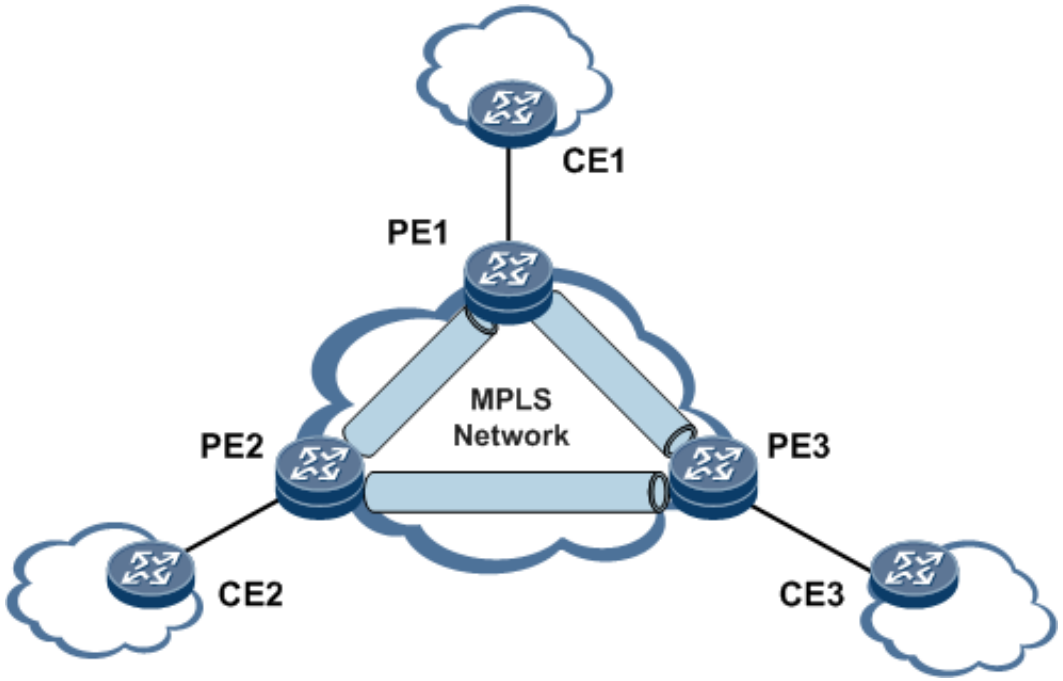
High level design in its essence

- Underlay
  - `undo portswitch` => backbone interfaces are routed, no need for VLANIF
  - Single area OSPF with BFD, GR, tuned timers
  - MPLS and MPLS LDP
  - MP-BGP with redundant router reflectors (RR)
  - PIM SM for multicast
- Services
  - Traditional MPLS based L3VPN
  - L2VPN (neither Martini, nor Kompella) => BGP AD VPLS
  - Multicast & Multicast VRF
  - Multi vrf OSPF to ensure dynamic routing where needed
  - VRRP for default GW redundancy

OSPF & MPLS & MPLS LDP

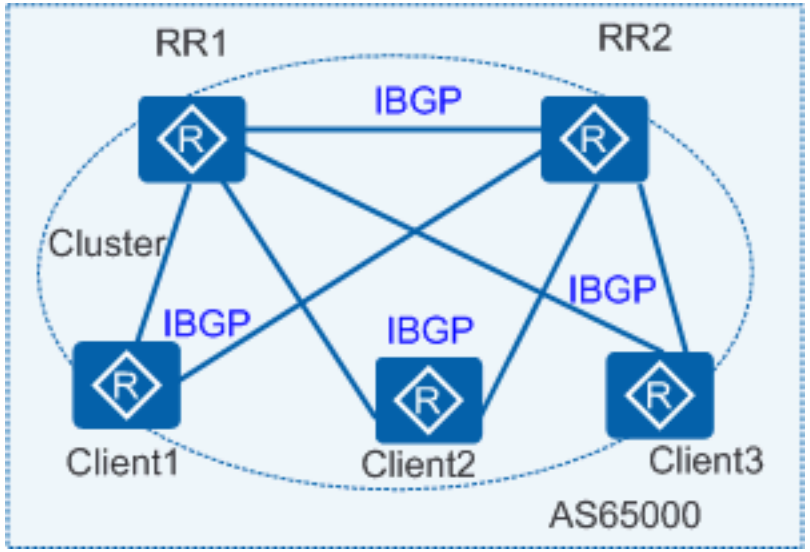
- OSPF
  - ensure reachability of all PEs and its loopbacks (router ID)
  - network-type – p2p
- MPLS LDP creates FEC/LSPs for mentioned loopbacks
- In this project P & PE & CE is converged

  = MPLS LSR / BGP peer && site default GW

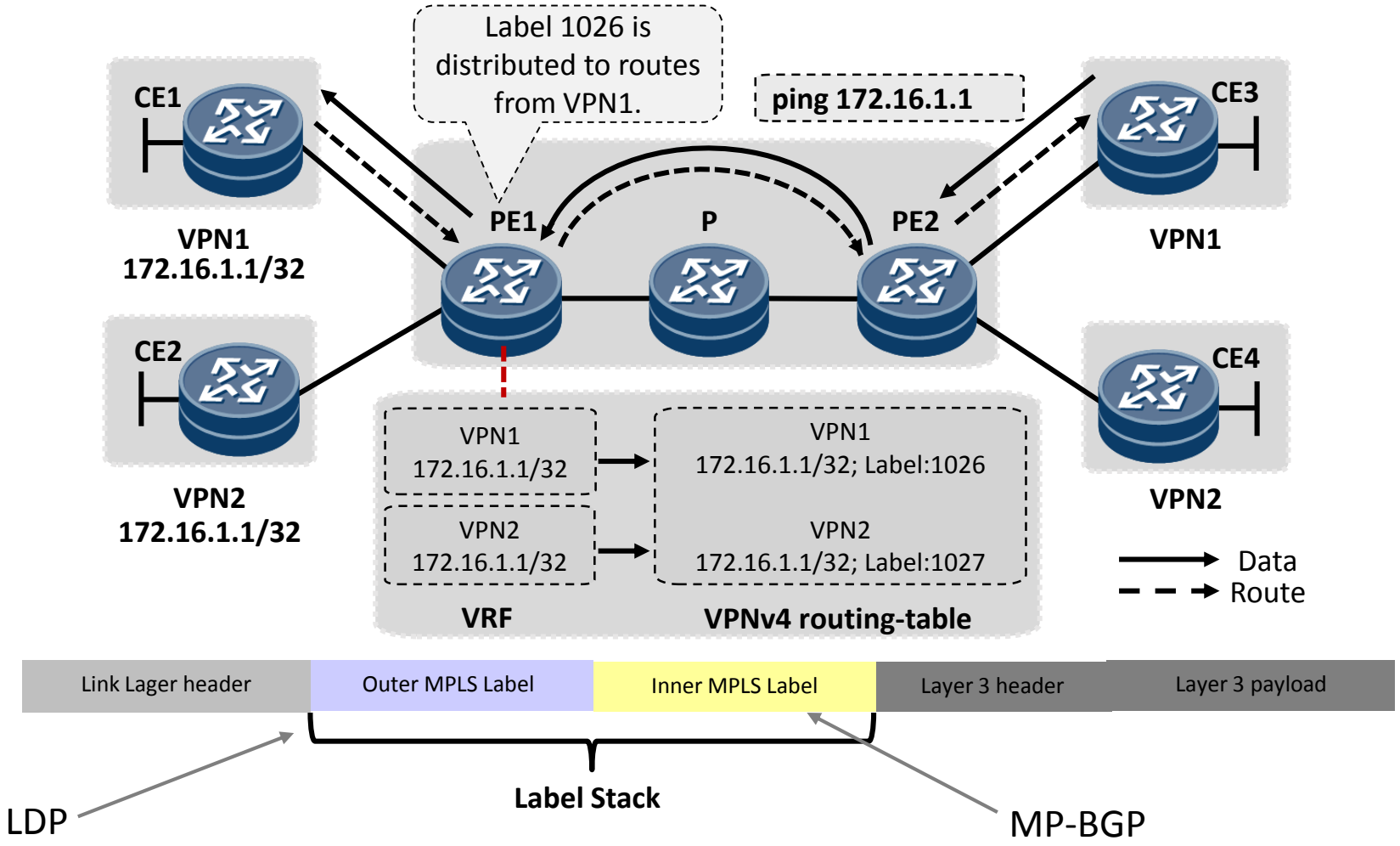


MP-BGP

- Enable route exchange from VRF routing tables between PEs
- Exchange MPLS labels for specific services (VRFs)
- To avoid full-mash BGP sessions => use RR design (route reflector)



MPLS VPN Label Nesting



L3VPN routing table on PE

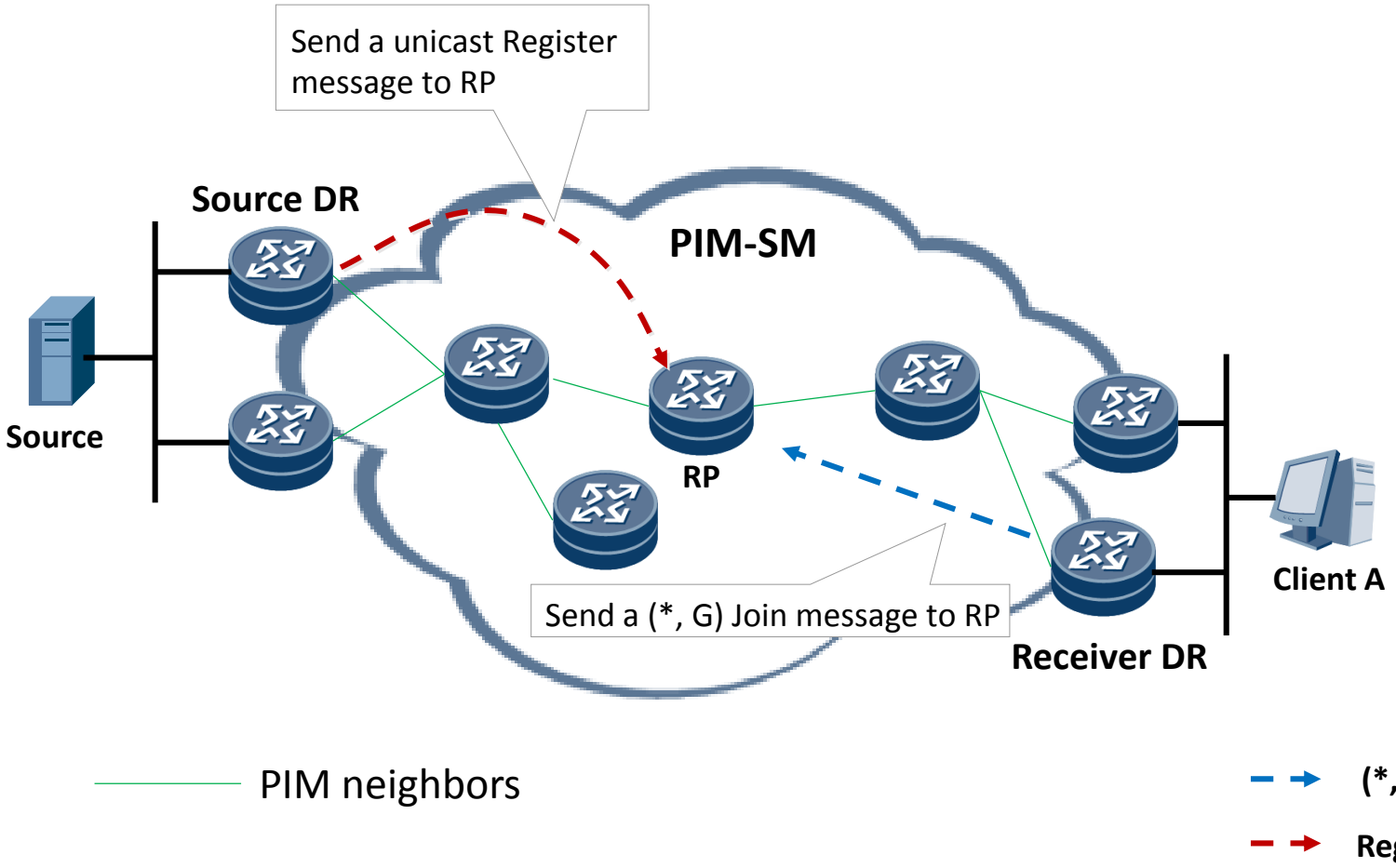
```
<XXX-PE01>display ip routing-table vpn-instance SERVICE
Route Flags: R - relay, D - download to fib
------------------------------------------------------------------------------
Routing Tables: SERVICE
         Destinations : 677        Routes : 677

Destination/Mask    Proto   Pre  Cost      Flags NextHop         Interface

        0.0.0.0/0   IBGP    100  0           RD  x.x.1.1         XGigabitEthernet0/2/1
        x.x.x.x/24  IBGP    100  0           RD  x.x.6.1         XGigabitEthernet0/2/1
         . . . .
```

Multicast

Multicast service requirements

- Customer has big number of multicast sources all over the network (monitoring cameras)
- Potential need to allow separation of multicast to different L3VPNs
- Some cameras are placed in 3rd party network with RP placed there as well

Multicast (any source) needs RP



Multicast SOURCE

HUATECH

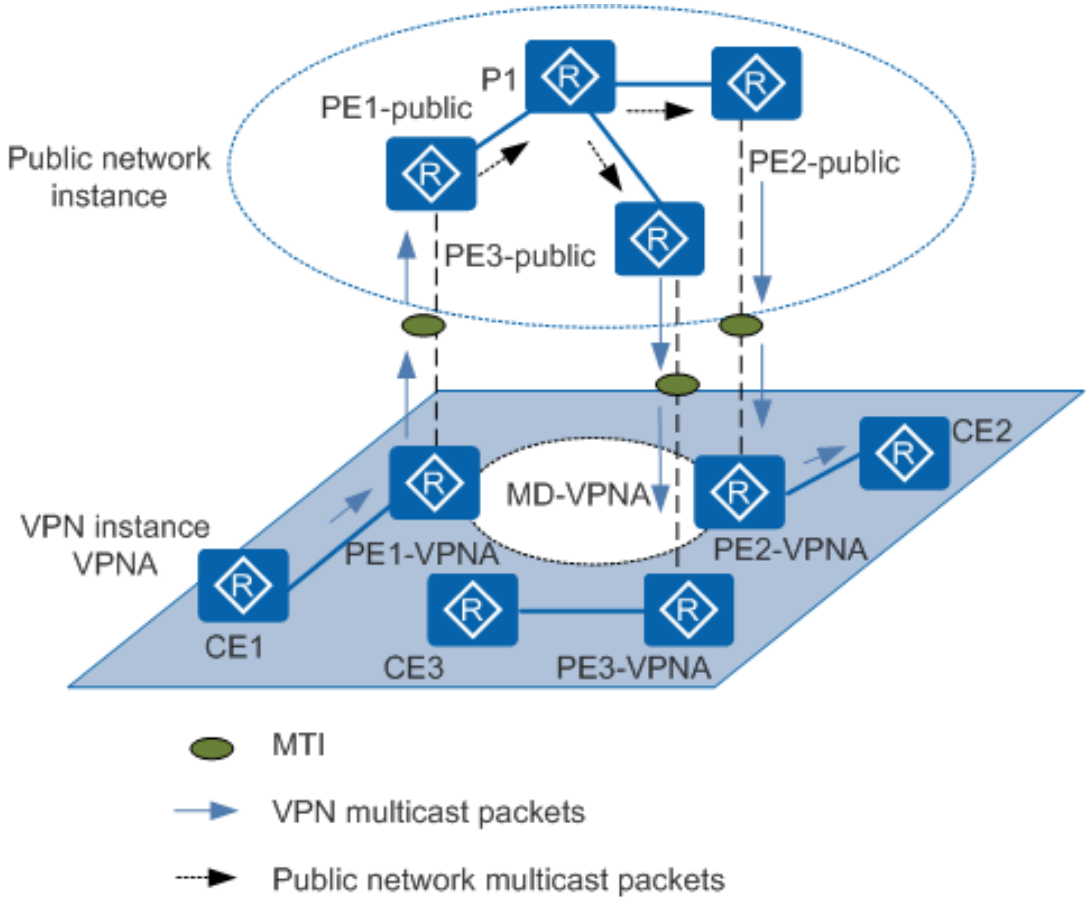
Multicast principles - RP

- Standard any source multicast uses PIM SM with RP
  - RP (rendezvous point)
    - 2 RPs dynamically set across network (primary & secondary)
    - Each RP pair handles specified multicast groups (MC IPv4 addresses)

```
<xxx-PE01>dis pim vpn-instance XXX-Service rp-info
 VPN-Instance: XXX-Service
 PIM-SM BSR RP Number:8
 Group/MaskLen: 224.x.0.0/16
     RP: x.67.0.1
     Priority: 1
     Uptime: 10w:6d
     Expires: 00:01:57
                                        Group/MaskLen: 224.x.0.0/16
                                            RP: x.67.0.2
                                            Priority: 5
                                            Uptime: 10w:3d
                                            Expires: 00:01:57
                                       . . . . .
```

Multicast VPN = Multicast within Multicast

- PIM SM in public VRF distributes multicast between PEs
- L3VPN = multicast service in public multicast (multicast GRE) between related PEs
- PIM SM in VRF – distributes specific MC services in
- No MPLS involved



Multicast VPN

```
ip vpn-instance XXXX
 description XXXX VRF
 ipv4-family
  . . . .
  multicast routing-enable
  multicast-domain source-interface LoopBack1
  multicast-domain share-group 239.0.0.2 binding mtunnel 0   // public multicast => mcast GRE
  multicast-domain switch-group-pool 239.2.37.0 255.255.255.240 threshold 100
                  //use multicast addresses from this pool for SPT switchover
```

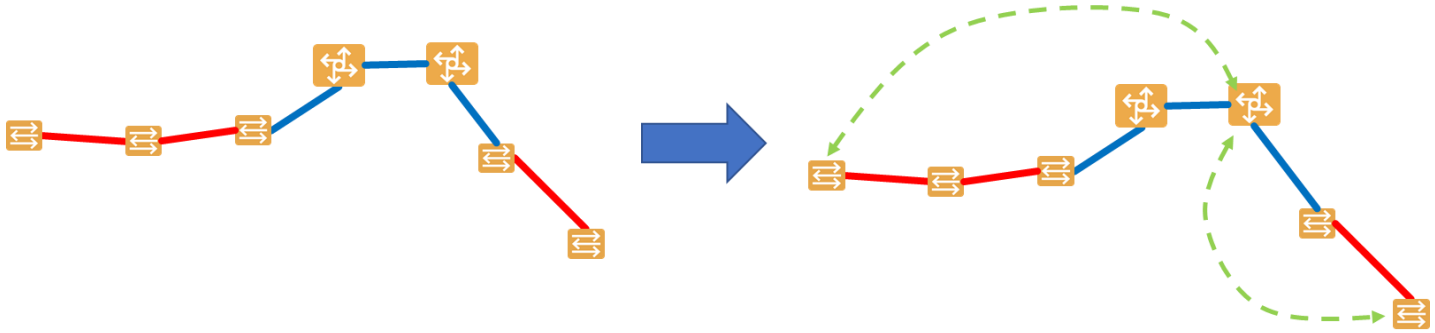
```
<xxx-PE01>dis pim routing-table  brief
 VPN-Instance: public net
 Total 1 (*, G) entry; 18 (S, G) entries

  00001.(*, 239.0.0.2)
       Upstream interface:XGE0/2/1
       Number of downstream:1

  00002.(x.x.1.1, 239.0.0.2)
       Upstream interface:XGE0/2/1
       Number of downstream:1
 . . . .
```

```
<xxx-PE01>dis pim vpn XXXX routing-table  brief
 VPN-Instance: XXXX
 Total 3 (*, G) entries; 140 (S, G) entries

  00001.(x.x.7.18, 224.18.0.28)
       Upstream interface:MTun0
       Number of downstream:2

  00002.(x.x.7.18, 224.18.0.29)
       Upstream interface:MTun0
       Number of downstream:3
  . . . .
```

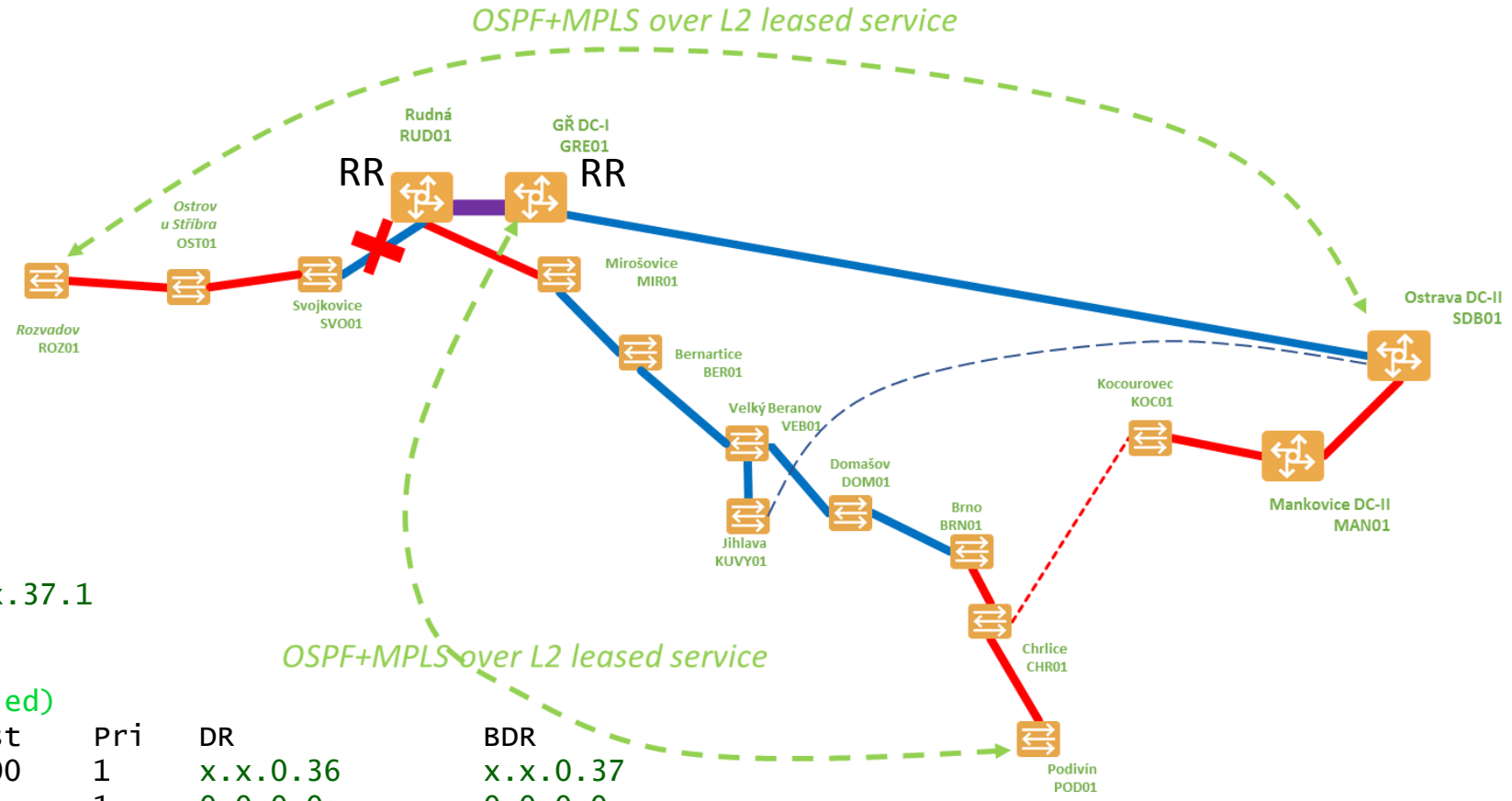
Extra requirements

How to keep service ON during fiber cut on star topology?

- Use existing leased L3VPN services for backup – very complicated
- STAR => RING and extend MPLS
  - Initially *MPLS over Ethernet over GRE* over L3 leased backup service
  - Luckily T-Mobile agreed to reconfigure services and provide L2 p2p services connecting end-points with core sites.



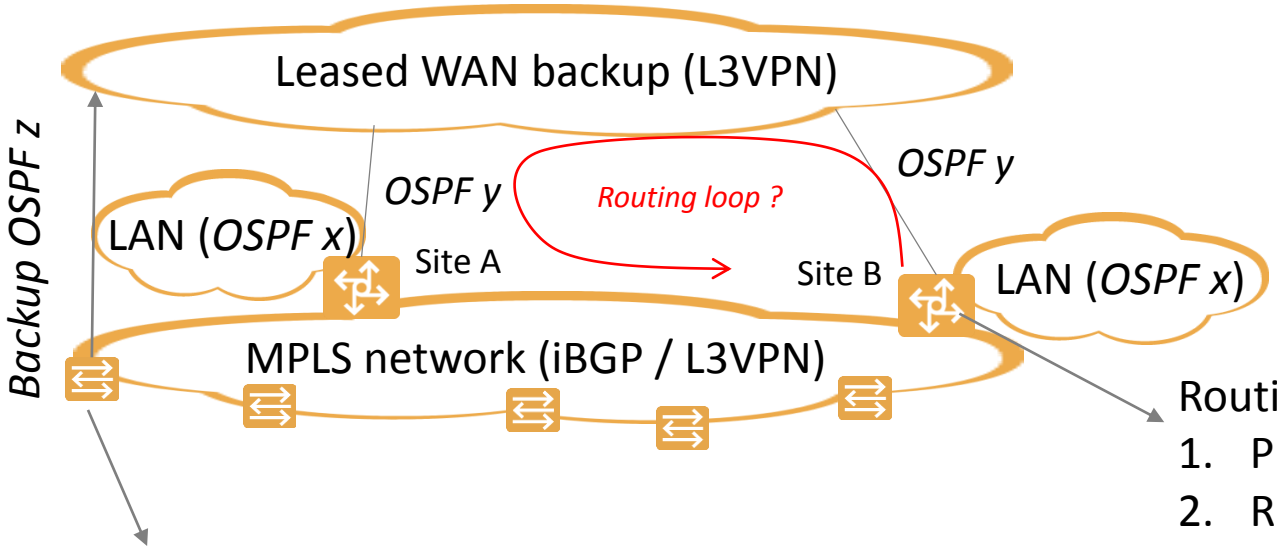
How to keep service ON during fiber cut on star topology?




```
<xxx-PE01>display ospf interface

        OSPF Process 1 with Router ID x.x.37.1
                Interfaces

 Area: 0.0.0.0           (MPLS TE not enabled)
 IP Address      Type            State     Cost    Pri   DR              BDR
 x.x.0.36        Broadcast       DR        5000    1     x.x.0.36        x.x.0.37
 x.x.37.1        P2P             P-2-P     0       1     0.0.0.0         0.0.0.0
 x.x.0.5         P2P             P-2-P     10      1     0.0.0.0         0.0.0.0
```


HUATECH

Route control on core sites



Non-core site - simple:

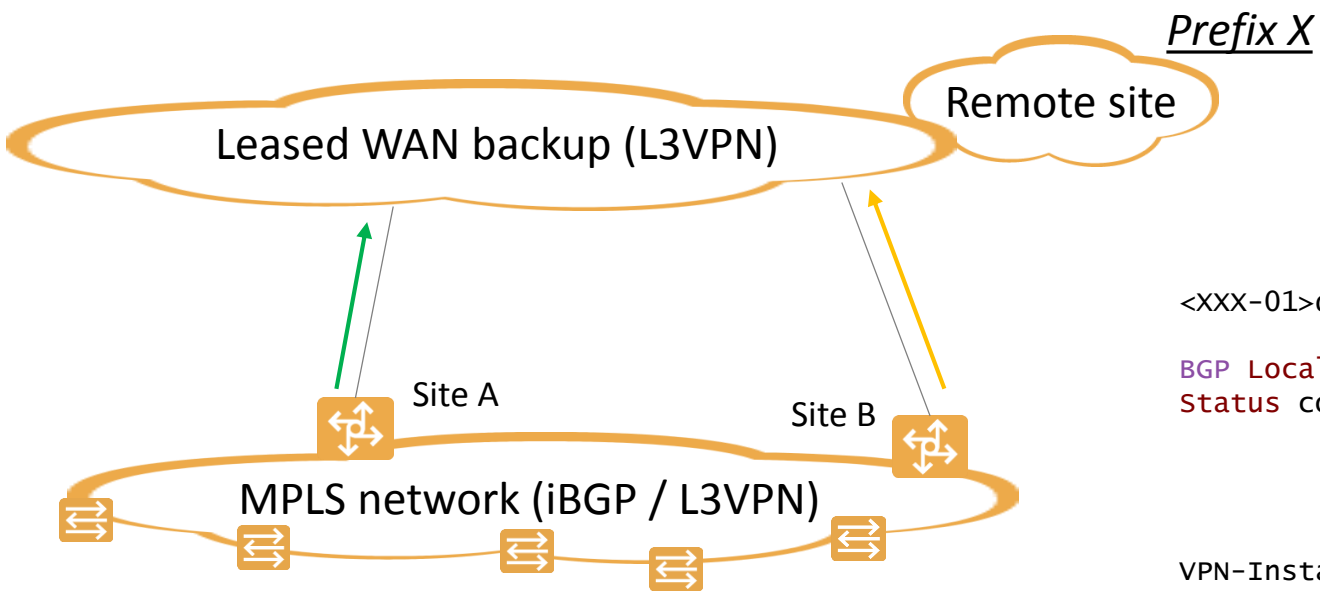
1. iBGP (MPLS L3VPN) primary path
2. *OSPF z* – only backup => not redistributed to iBGP && lower preference than iBGP

Routing table v L3VPN *service_x*:

1. Preference: OSPF x > iBGP > OSPF y
2. Risk of routing loop
   - *site A: iBGP => OSPF y => WAN => site B: OSPF y => iBGP*
3. Control redistribution (route-policy/route-maps)
   - Plan redistribution carefully
   - Use tags to control, that the route is not coming back

TIP: if in similar situation, try to use BGP instead of OSPF ;)

Route control on core sites




```
<XXX-01>dis bgp vpnv4 vpn-instance SERVICE_X routing-table

BGP Local router ID is x.x.23.1
Status codes: * - valid, > - best, d - damped,
              h - history,  i - internal, s - suppressed, S - Stale
              Origin : i - IGP, e - EGP, ? - incomplete


VPN-Instance PROV, Router ID x.x4.23.1:

Total Number of Routes: 667
     Network            NextHop        MED        LocPrf     PrefVal Path/Ogn

. . . . . .
*>i  x.x.18.16/30       x.x.1.1        1          120         0      ?
* i                     x.x.6.1        1          100         0      ?
```

Prefix X is reachable through Site A and Site B.

Site A is preferred from all sites.

Easy with BGP – modify *Local Preference* BGP attribute using route map.

Conclusion

- Initial concerns from some of new requirements turned to fun ☺
- Great vendor documentation
- Once more routing protocols meet on one router, try to reduce them to rely on single protocol metrics
- Be aware of routing loops
- Try to avoid multicast VPNs ☺

Thank you, questions?

www.huatech.cz

Huatech a.s., Na Maninách 1092/20
Praha 7 – Holešovice, 170 00

+420 220 611 111
info@huatech.cz