Use cases and pitfalls

in MPLS/VPLS networks

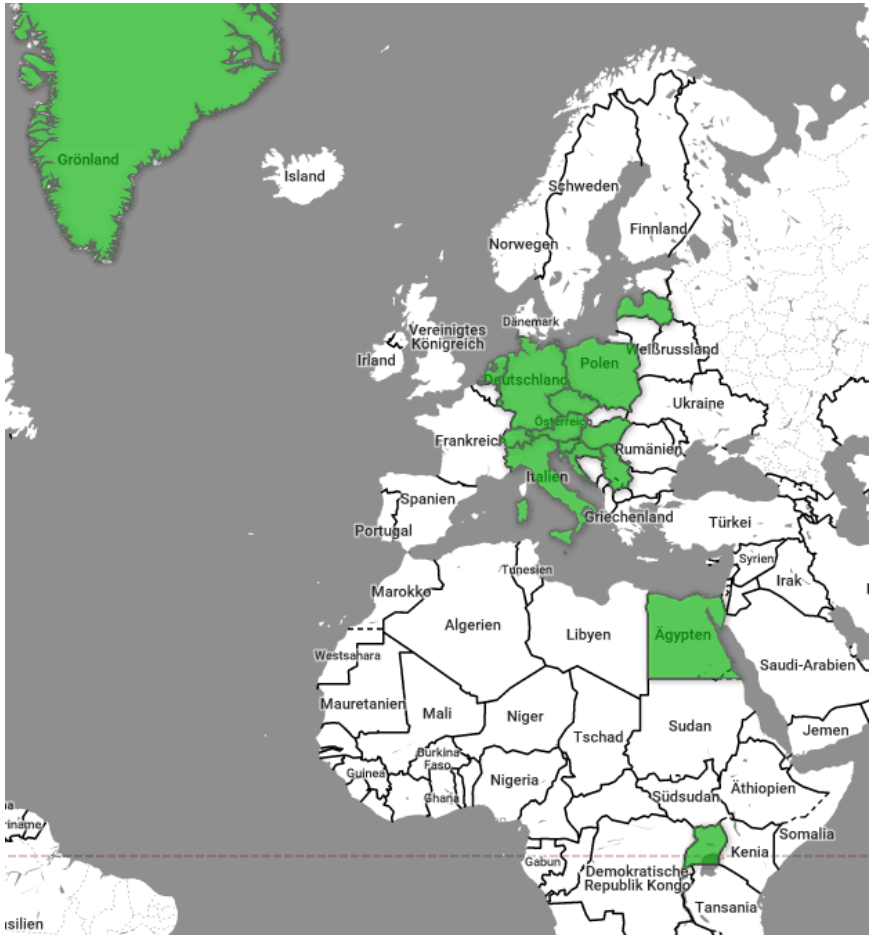
Contact

- Phone: +49 761 2926500
- Email: sales@fmsweb.de
- Shop: https://www.mikrotik-shop.de

- MikroTik Mirror: http://www.mikrotik-software.de
- Twitter: https://twitter.com/fmsweb_de
- Website: http://www.fmsweb.de
- Wiki: http://wiki.fmsweb.de
- Presentations: http://wiki.fmsweb.de/wiki/MUM-Presentations
- Facebook: https://www.facebook.com/fmsinternetservice

About me

- Sebastian Inacker <inacker@fmsweb.de>
- FMS Internetservice GmbH, Germany
- MikroTik Trainer TR0011 (May 2007)
- MTCNA, MTCRE, MTCTCE, MTCUME, MTCWE, MTCIPv6E, MTCINE



MikroTik trainings and workshops

- Own training center and on site

  (Austria, Germany, Greenland, Hungary, Luxembourg, Malta, Netherlands, Switzerland, Uganda)
- Inquiries: sales@fmsweb.de

Overview / big picture

“Implementing and running a MPLS/VPLS network is easy.

As long as it is running well.”

Topics:

- Typical use cases of our (ISP) customers
- Typical pitfalls
- Surprising pitfalls
- Real world examples

Overview / big picture

“Implementing and running a MPLS/VPLS network is easy.

As long as it is running well.”

Not *main* topics

- Step-by-step guide for *each* setup (focus on pitfalls)
- Reason for MPLS/VPLS (You *should* know, why)

Reasons for MPLS/VPLS

Ok, very short and incomplete…

Benefits of MPLS

- Routing more complex than MPLS
- Some future setups (L3 VPN, TE) require MPLS

Benefits of VPLS vs. EoIP

- VPLS: No fragmentation (if done right)
- EoIP: Big overhead (42 bytes) & might cause fragmentation

Overview / big picture

Pitfalls:

- Incomplete (of course)
- Not limited to MPLS/VPLS

Needs for VPLS

- MPLS
- Routing (OSPF here)
- Physical infrastructure

Warning / heads-up / caution

This presentation will include errors, mistakes and wrong configuration attempts *to show resulting errors*!

Examples are simplified.

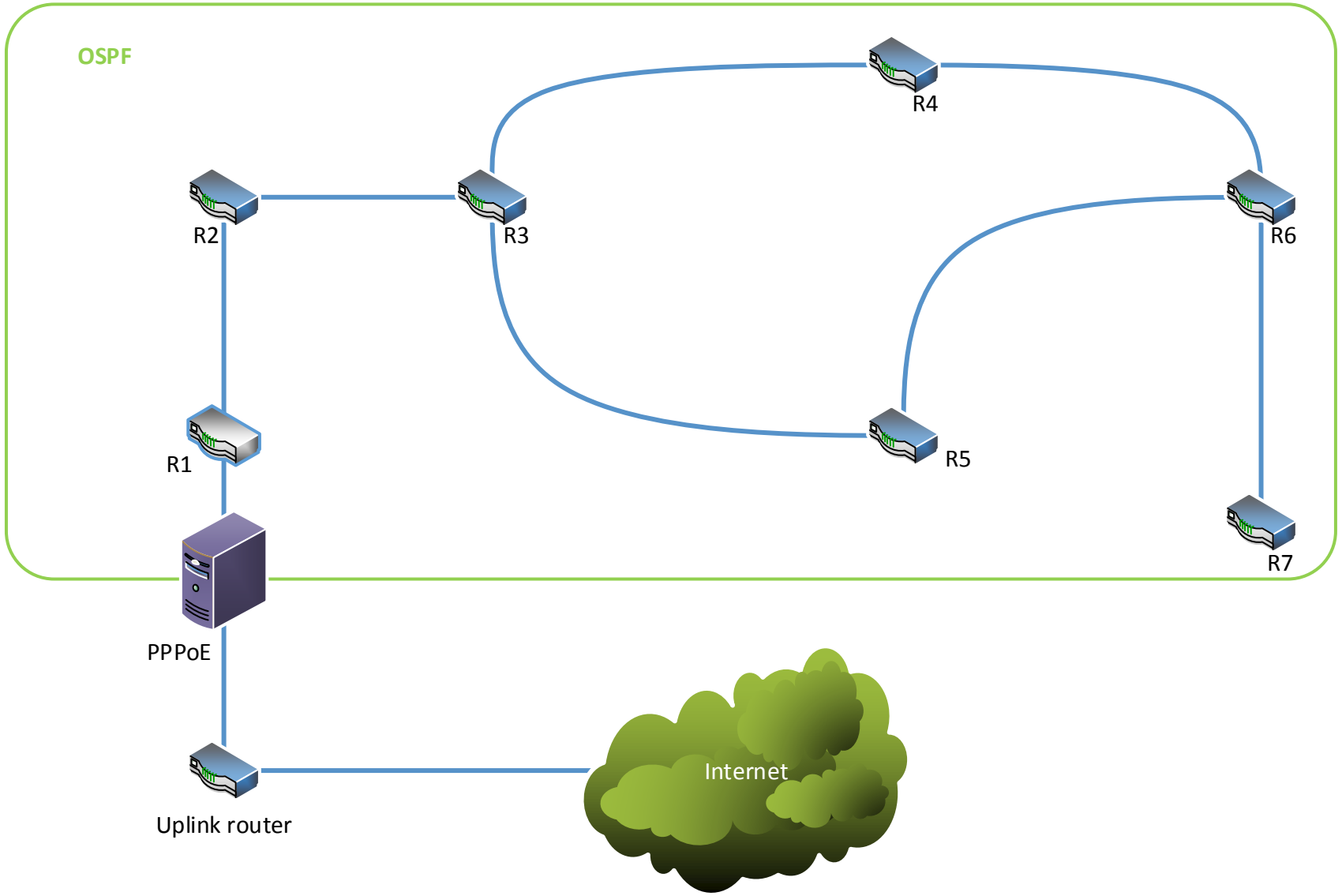
Keep that in mind.

The beginning

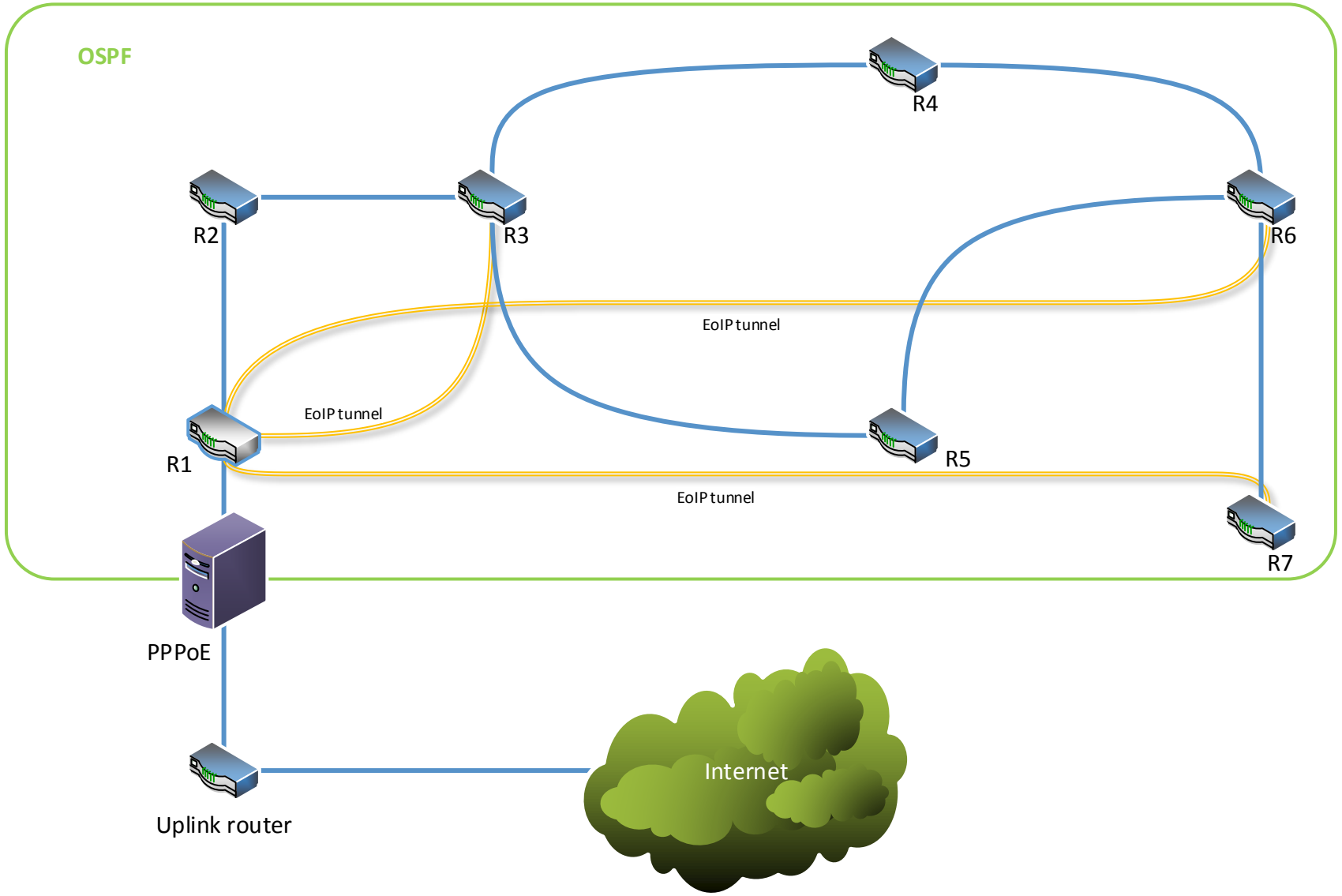
Existing setup

- Existing OSPF network
- One PPPoE server
- EoIP (L2 tunnel) for client connections

Existing setup



Existing setup



Requirements for MPLS

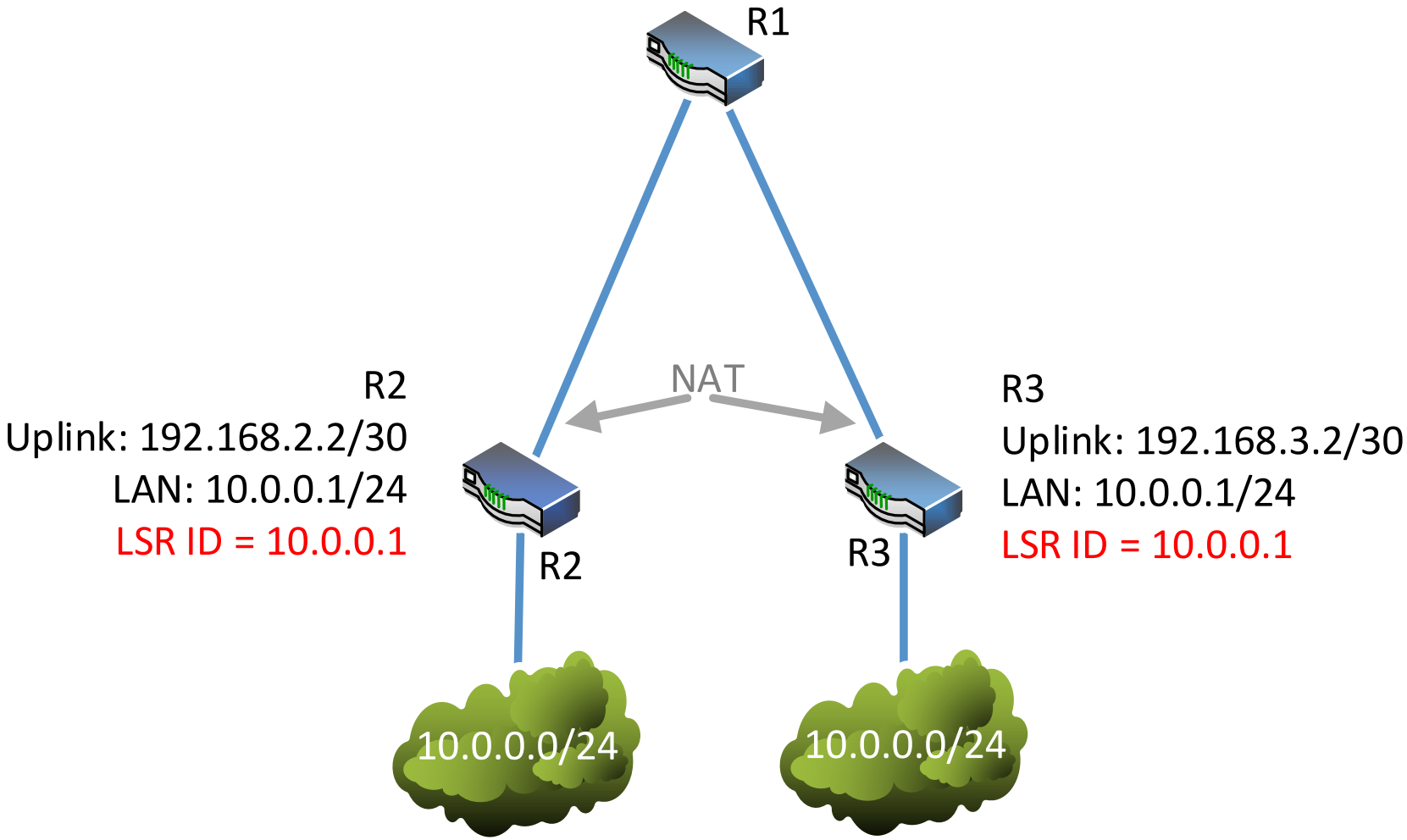
MPLS can be integrated without service disruption

Running MPLS on top of OSPF:

- Enable LDP (Label Distribution Protocol)
- Set
  - LSR ID (Label Switching Router's ID)
  - Transport Address

  If left unset: Lowest IP of router will be used
- Create LDP interfaces

1st possible issue: LSR ID not unique

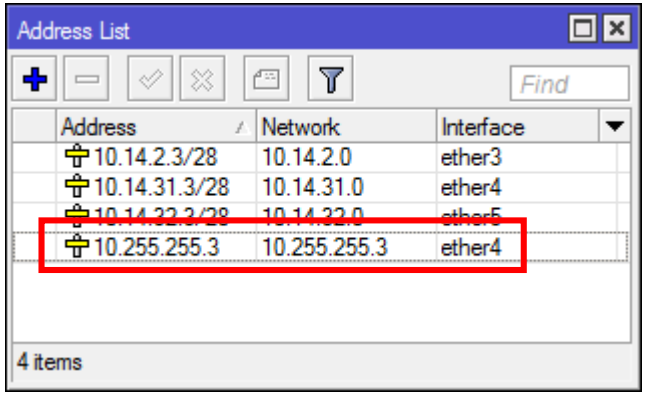


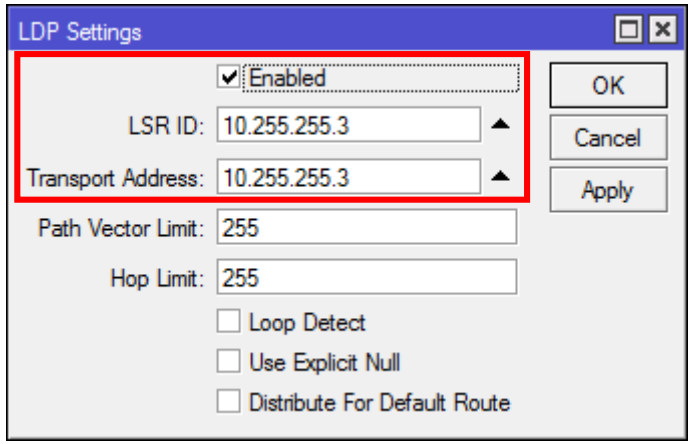
Unique IP for LDP

Unique IP for

- LDP (LSR ID and Transport Address)

Let‘s *try* 10.255.255.<*Router*>/32 on physical interface





Unique IP for OSPF

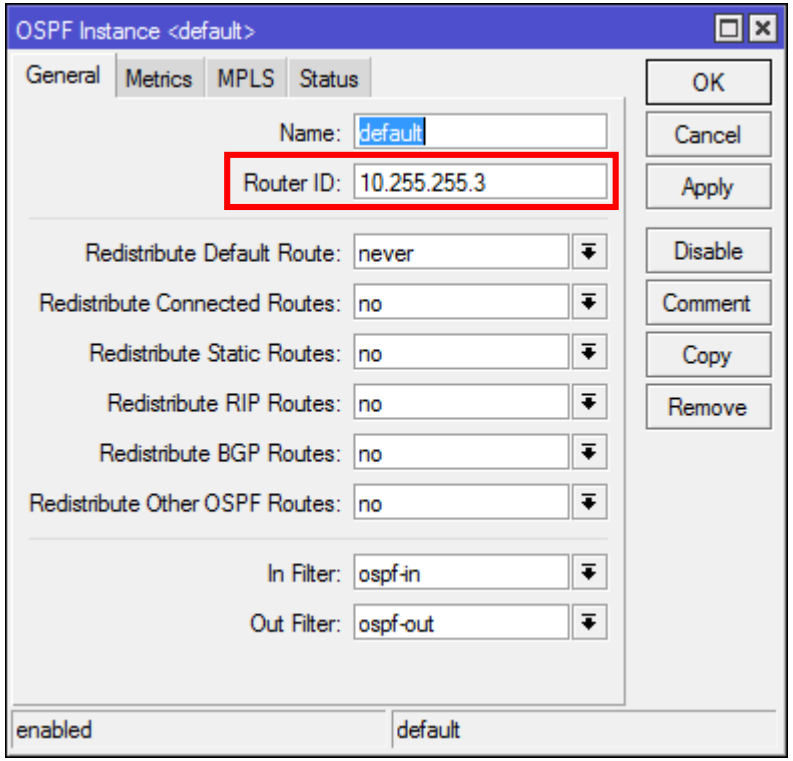
Unique IP for

- OSPF (Router ID) – same issue as with LSR ID

Take care: Setting of Router ID

- Restart of OSPF
- Loss of routing table

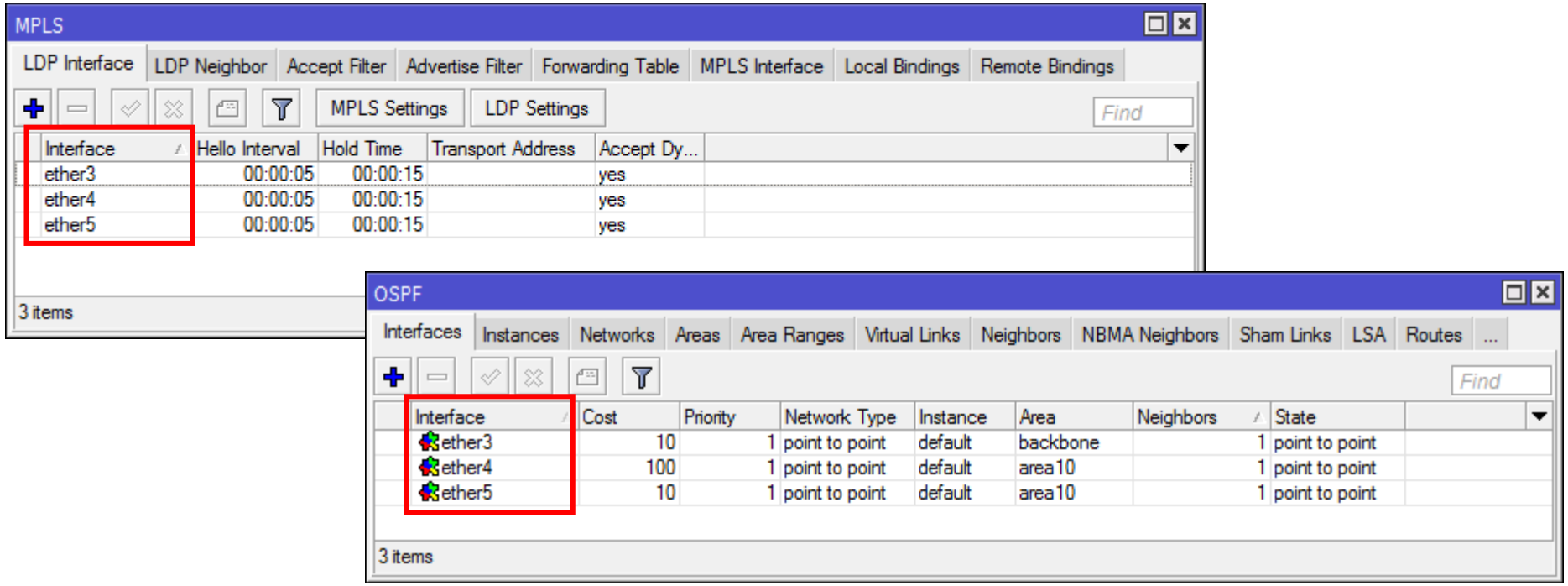
Service affecting action!



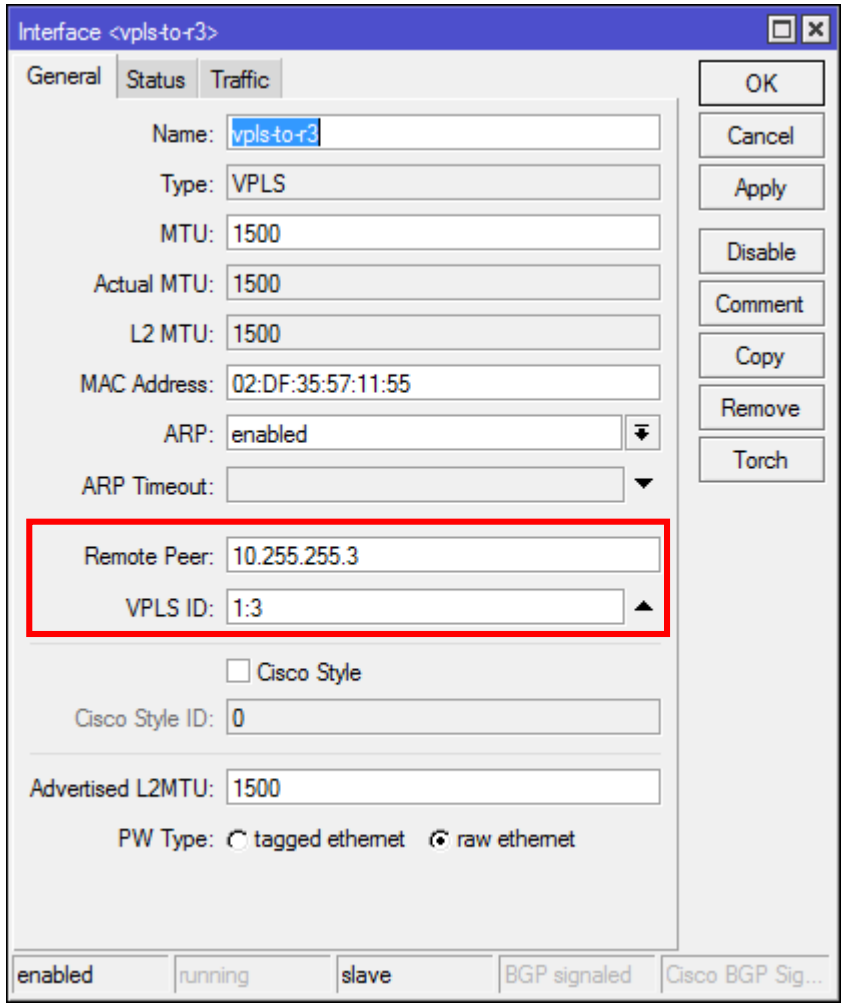
LDP interfaces

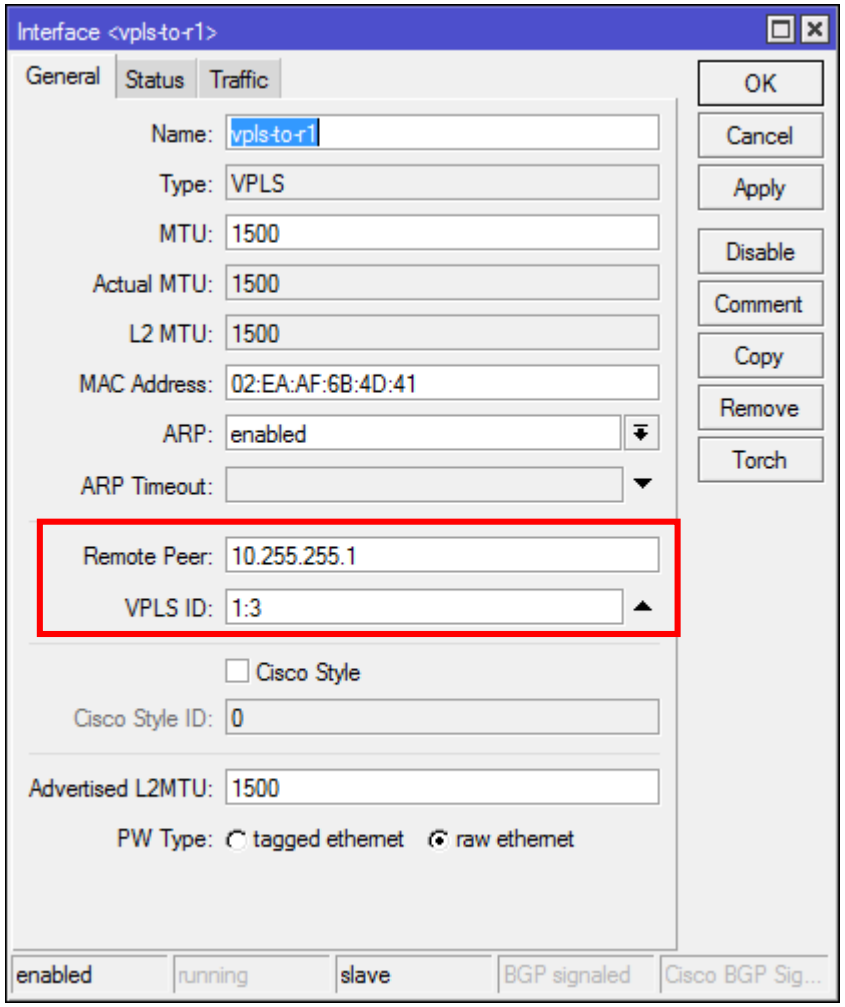
Set LDP interfaces

- Don‘t forget your backup path!
- Compare OSPF interfaces and LDP interfaces

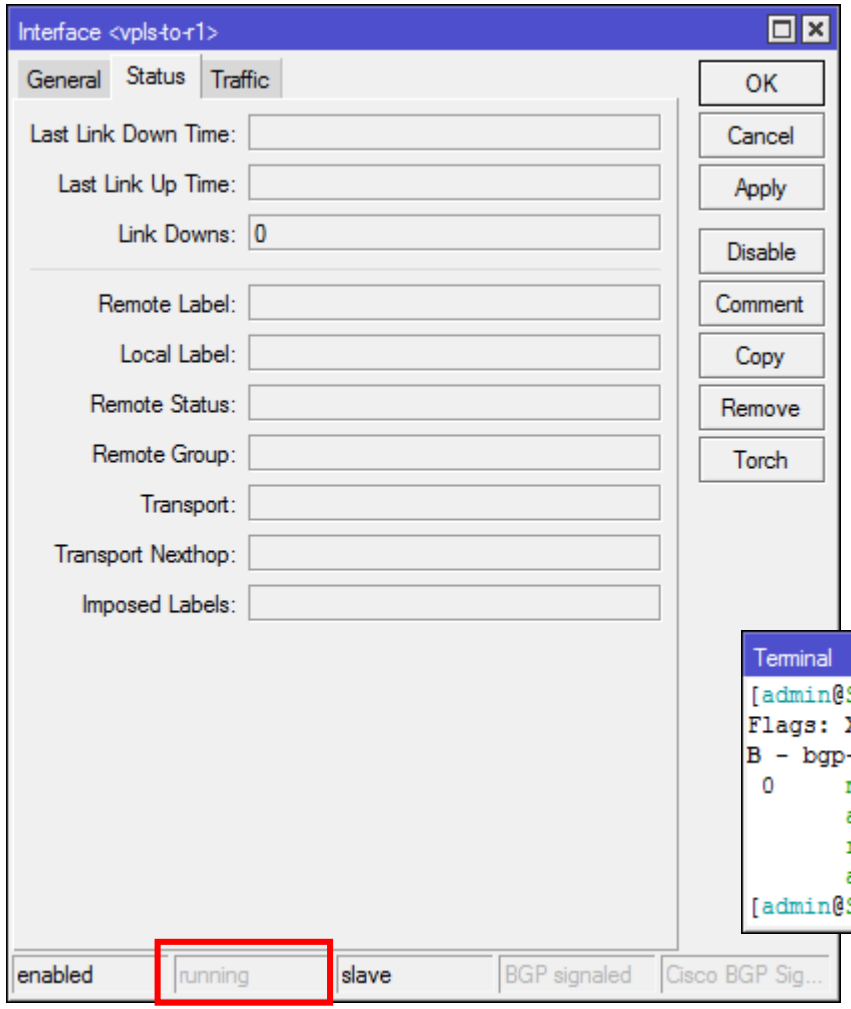


Create VPLS tunnels

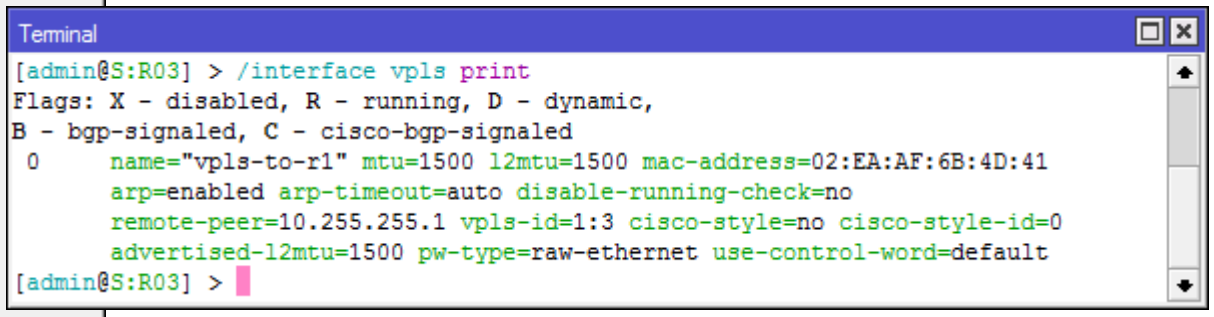




Check VPLS interface



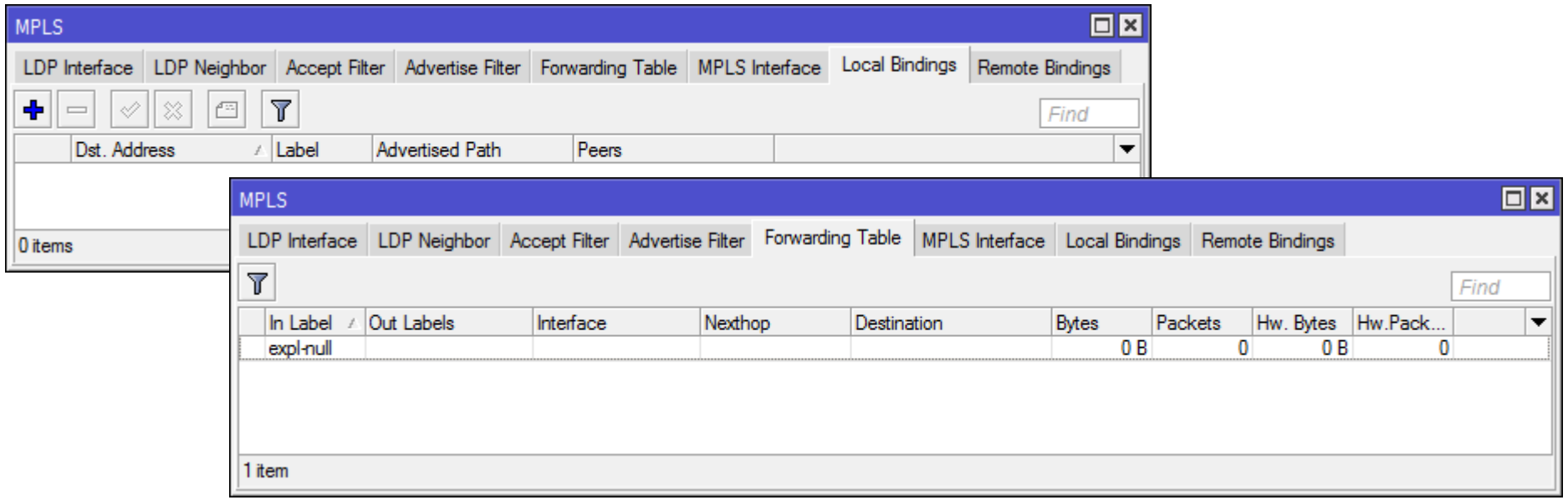
VPLS interface not running!



Check MPLS

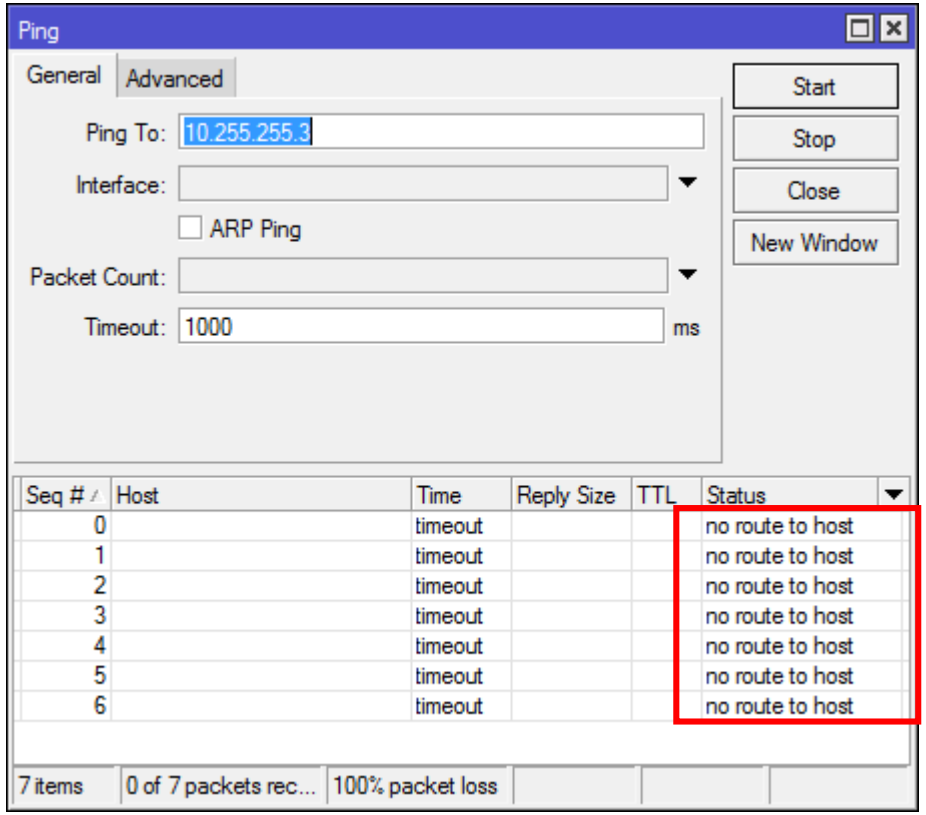
Empty:

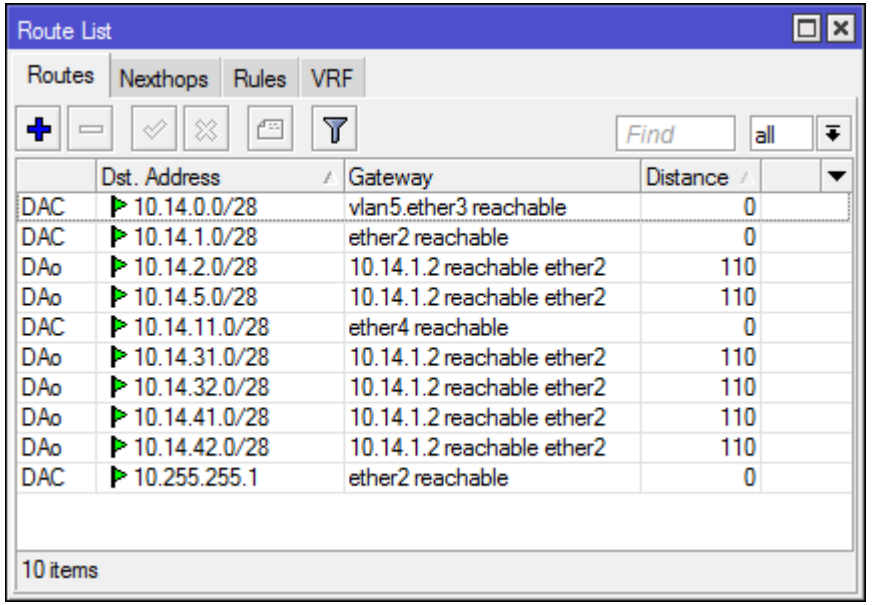
- MPLS Local Bindings
- MPLS Remote Bindings
- MPLS Forwarding Table



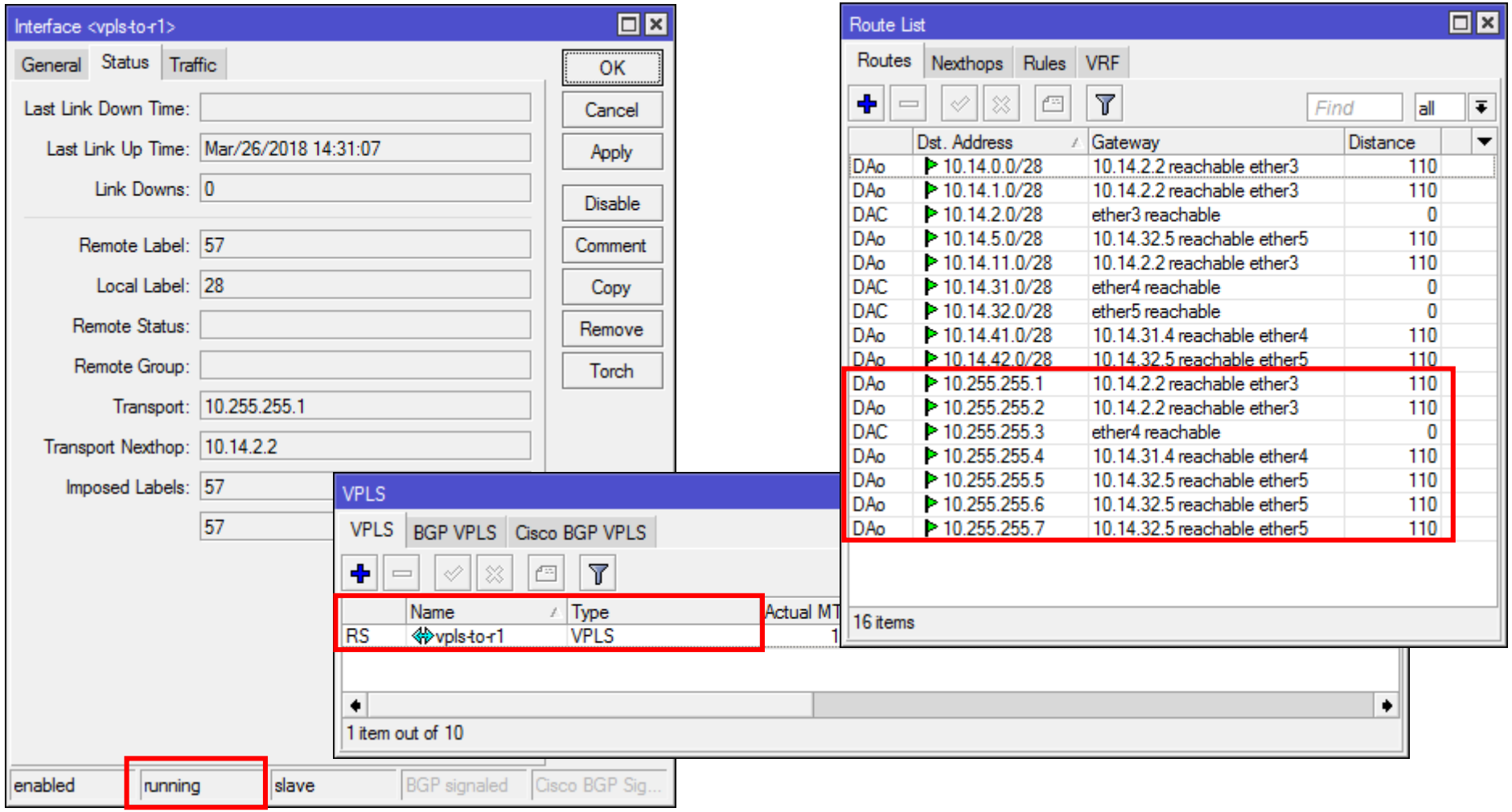
Check routing

IP routes to 10.255.255.x are missing





Routing ok, VPLS ok



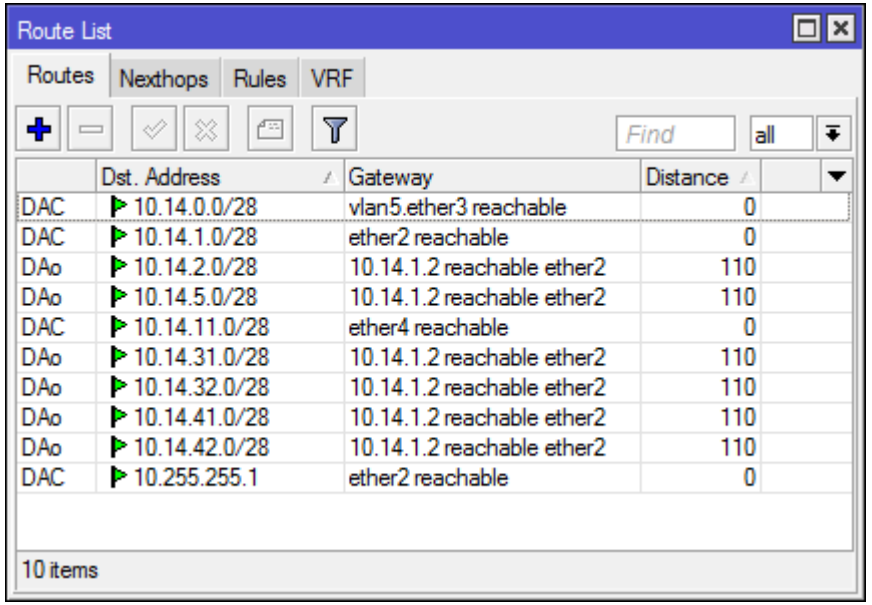
Lesson learned

MPLS is based on routing

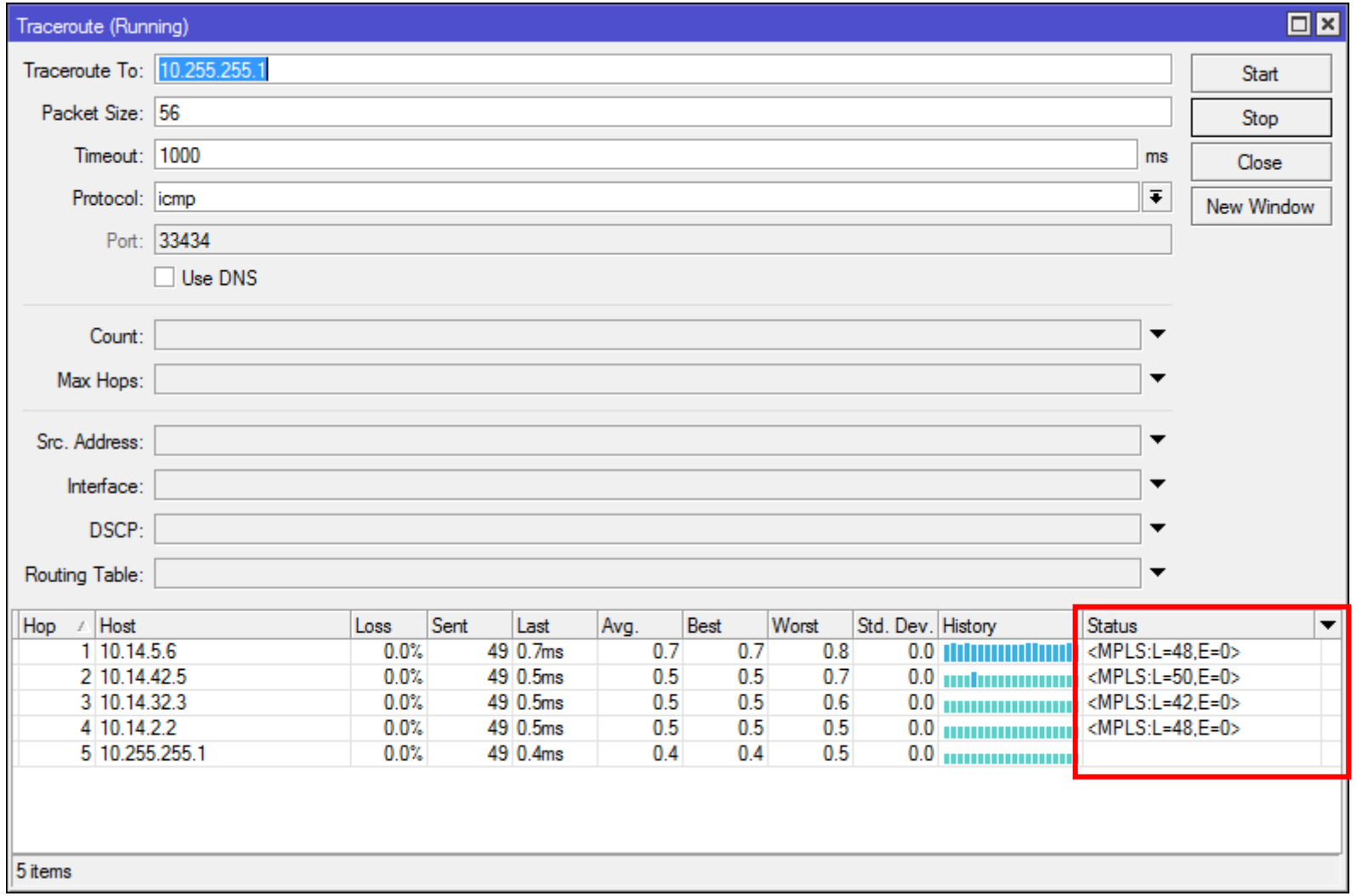
- Broken/incomplete routing, broken/incomplete MPLS
- Broken MPLS, broken VPLS

Debugging:

Consider dependencies!

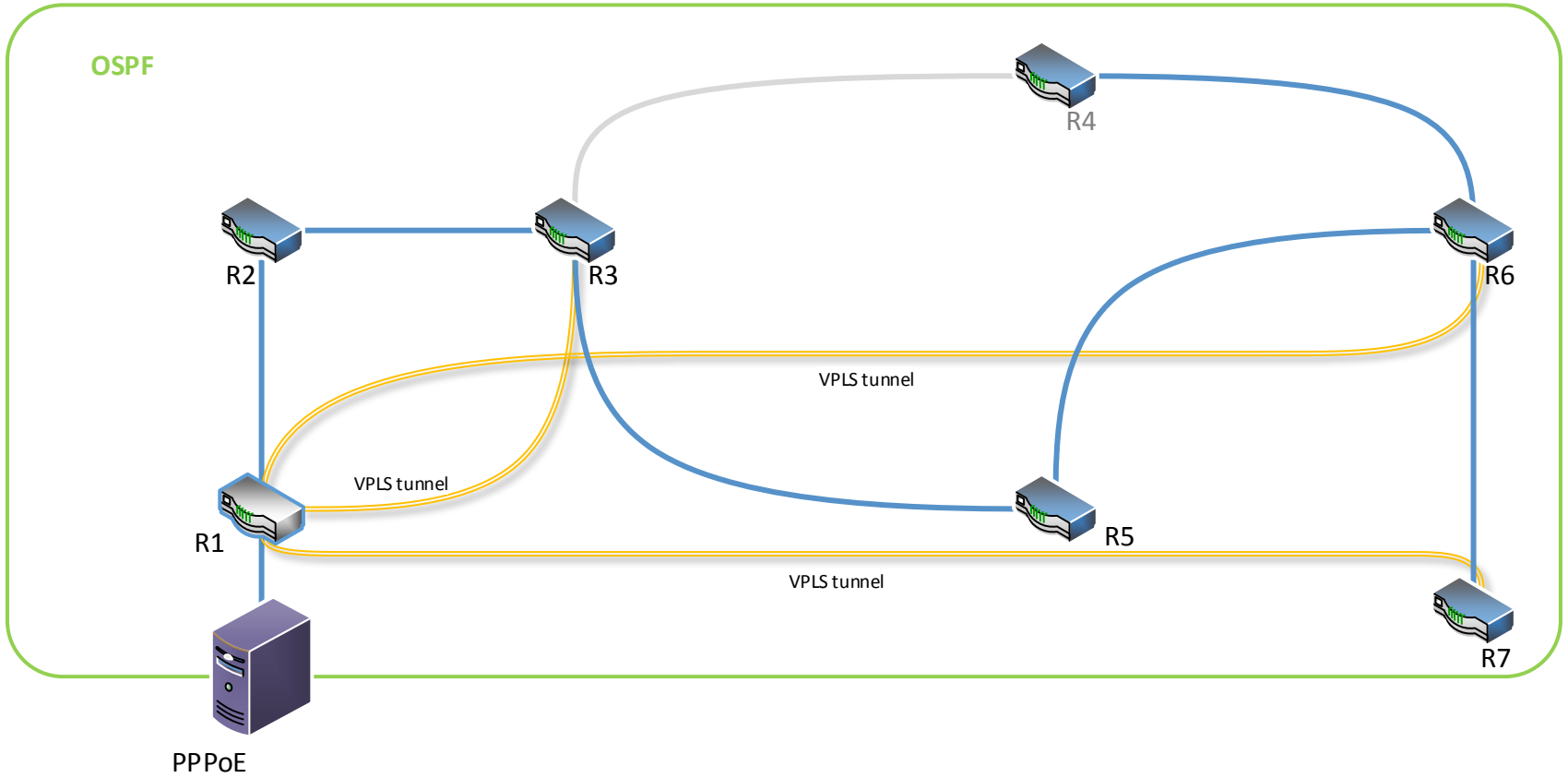


Working traceroute



Let‘s break things

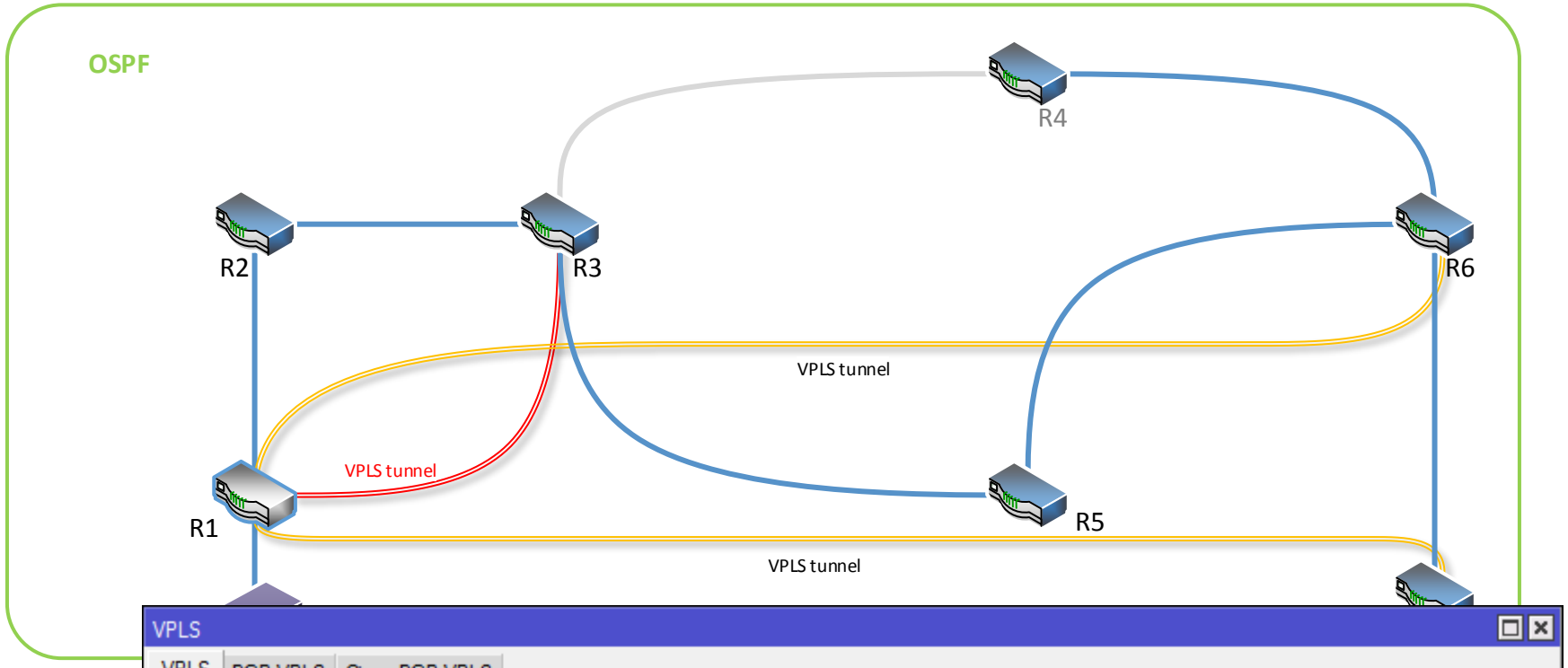
Maintenance at backup link



Maintenance at R4 (*backup link*). OSPF is going through R5.

Customers at R3 complain. Customers at R6, R7 are fine.

Maintenance at backup link



VPLS

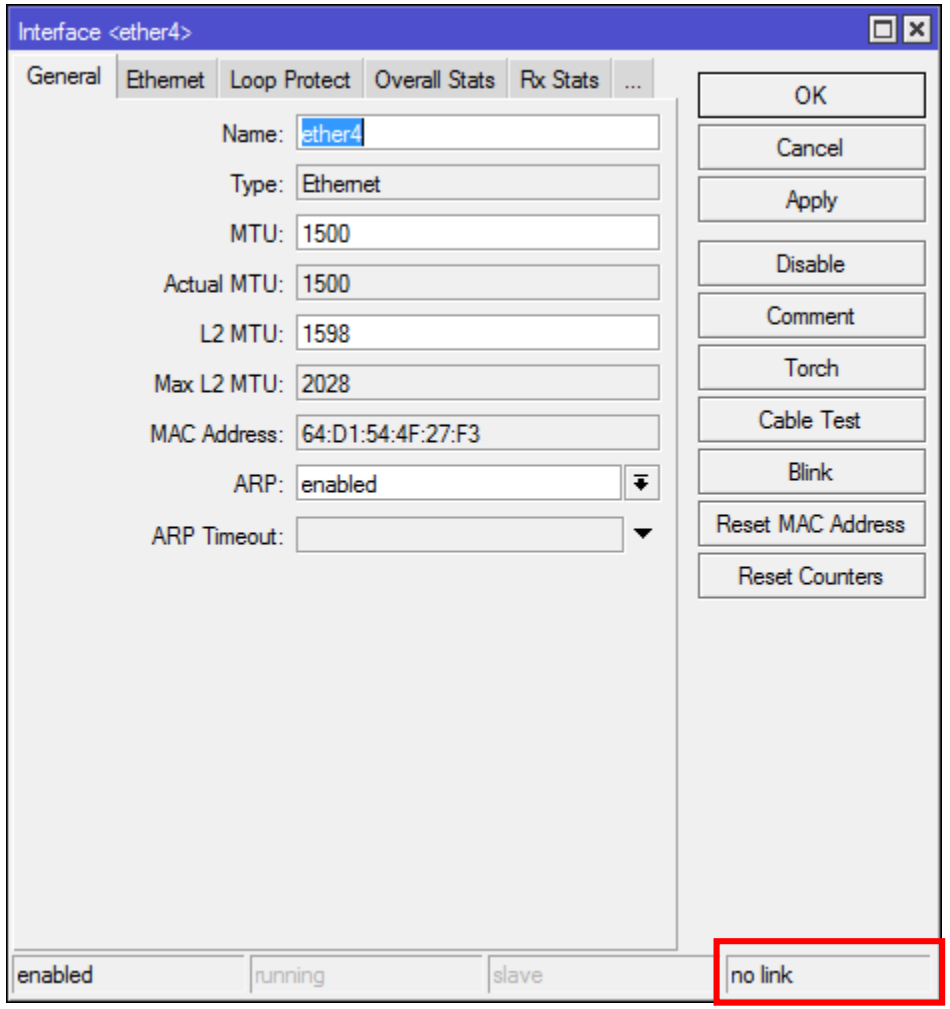
VPLS | BGP VPLS | Cisco BGP VPLS

Find

	Name	Type	Actual MTU	L2 MTU	Tx	Rx	Tx Packet (p/s)	Rx Pack
S	vpls-to-r3	VPLS	1500	1500	0 bps	912 bps	0	
RS	vpls-to-r6	VPLS	1500	1500	424 bps	0 bps	1	
RS	vpls-to-r7	VPLS	1500	1500	424 bps	0 bps	1	

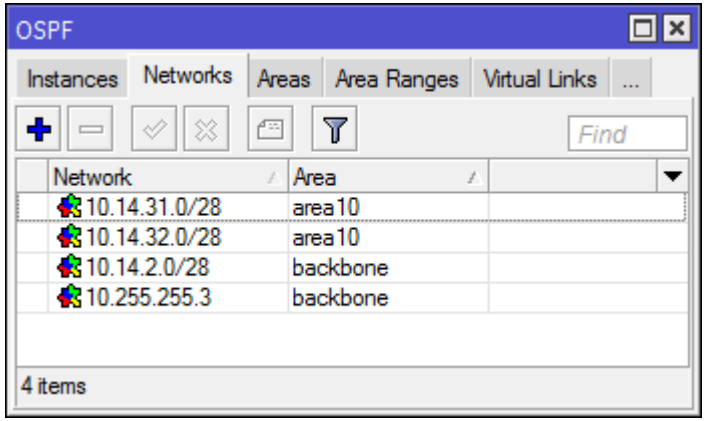
3 items out of 15

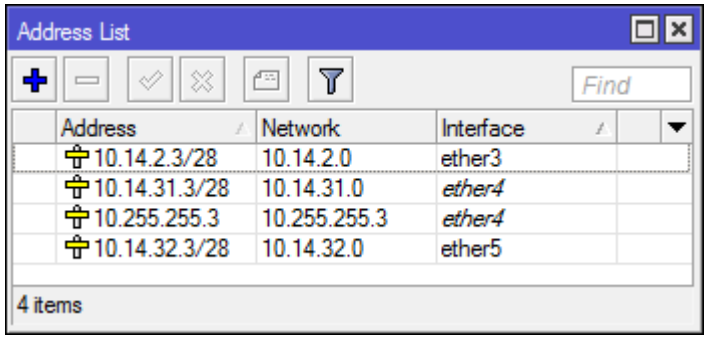
Maintenance at backup link



R3: No link on ether4

10.255.255.3/32 on ether4

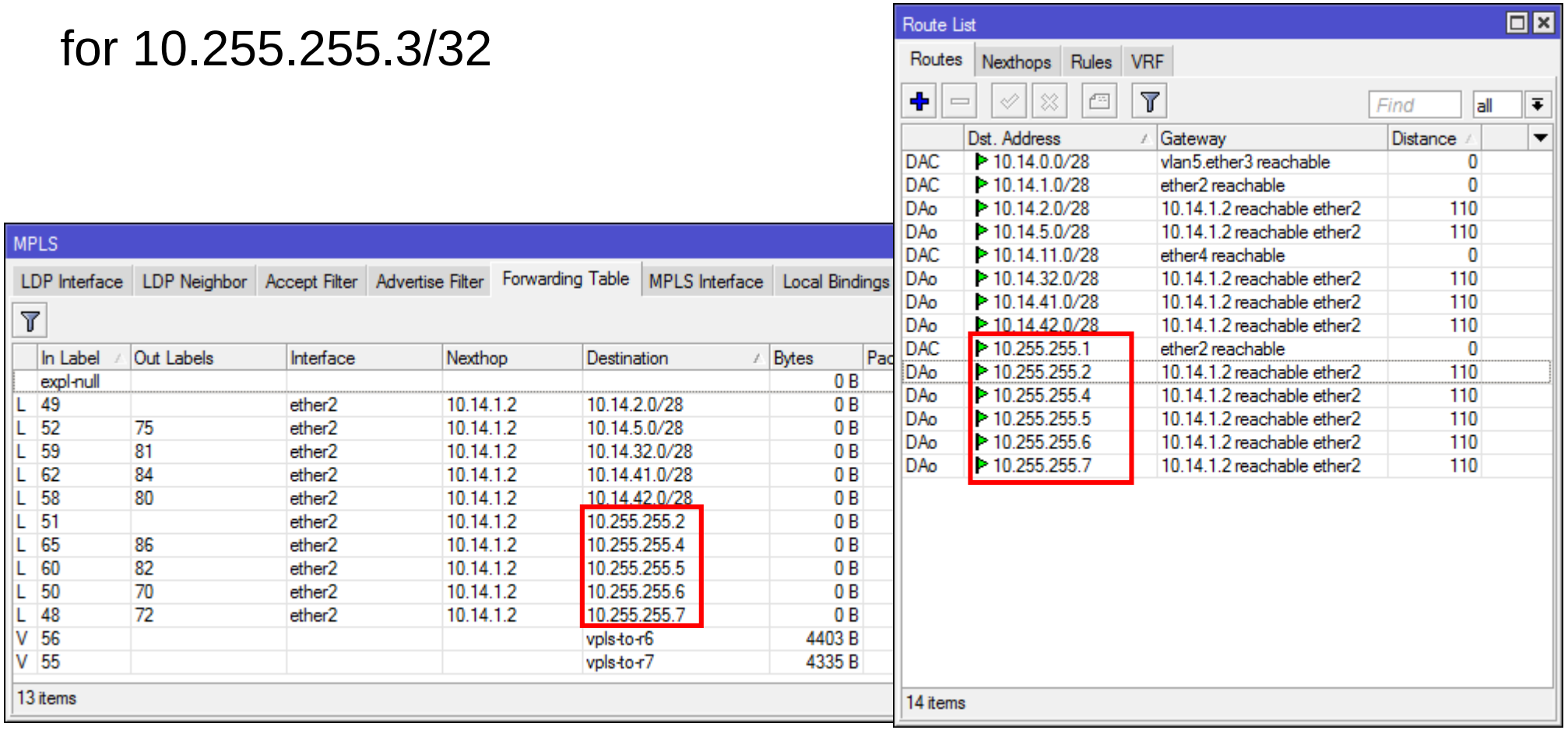




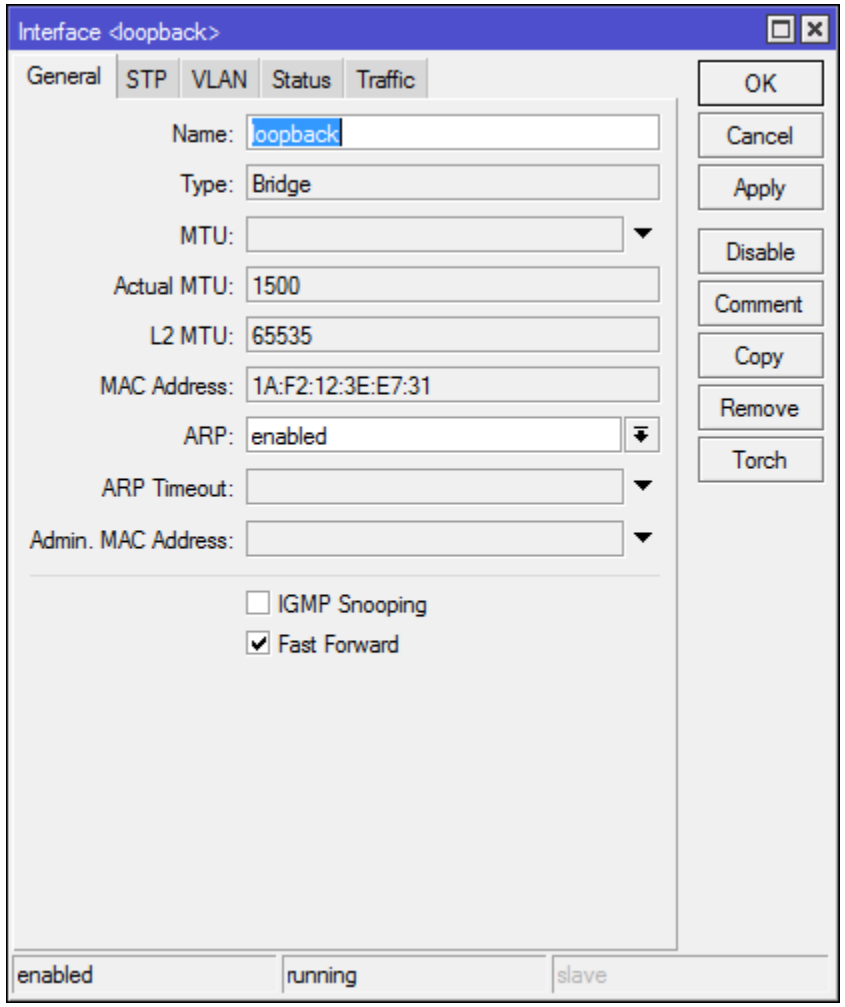
Maintenance at backup link

No MPLS Forwarding / IP Route

for 10.255.255.3/32



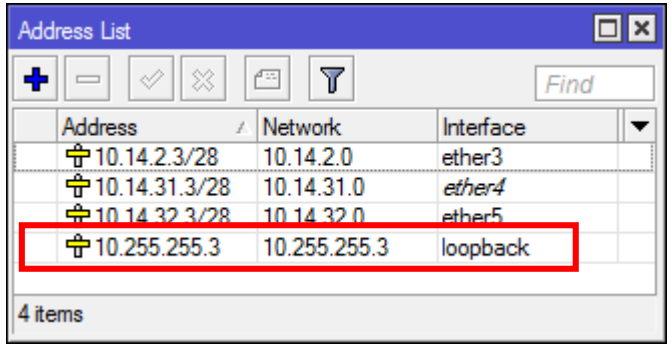
Loopback bridge



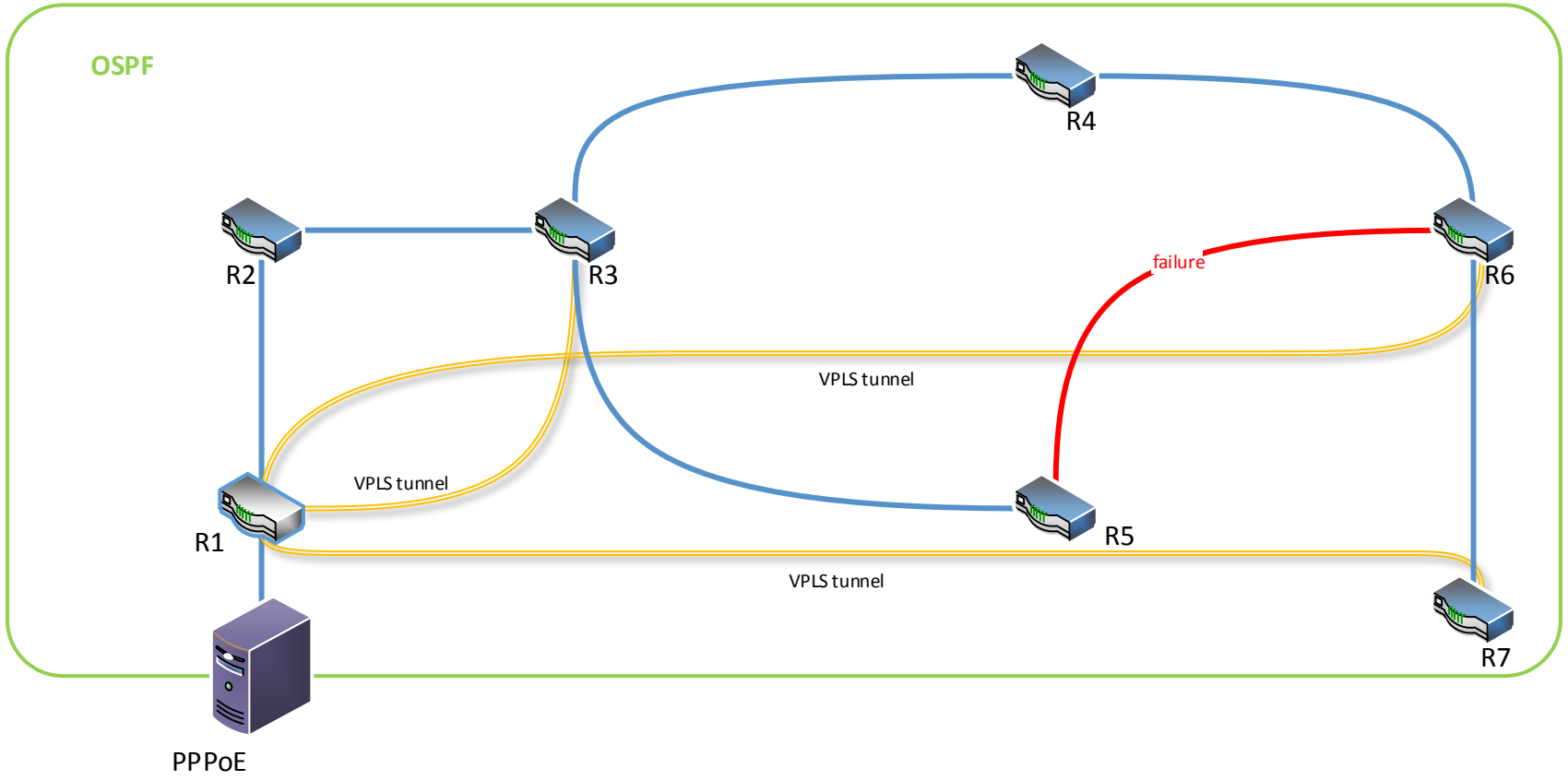
loopback bridge is a good idea

Loopback bridge:

Empty bridge with IP 10.255.255.x/32



Failure at main link



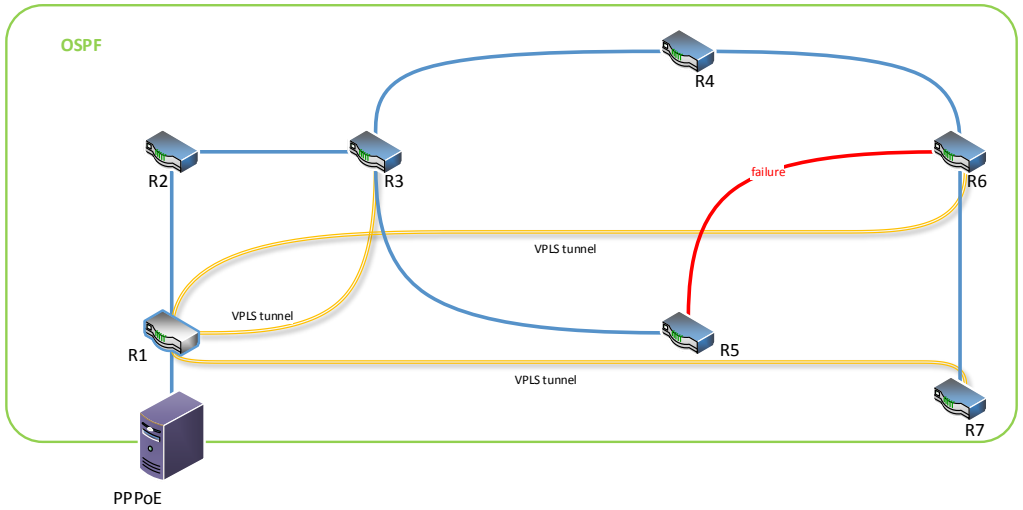
Failure at main link

Expected behaviour

- Routing through R4
- PPPoE customers at R3, R6, R7 online

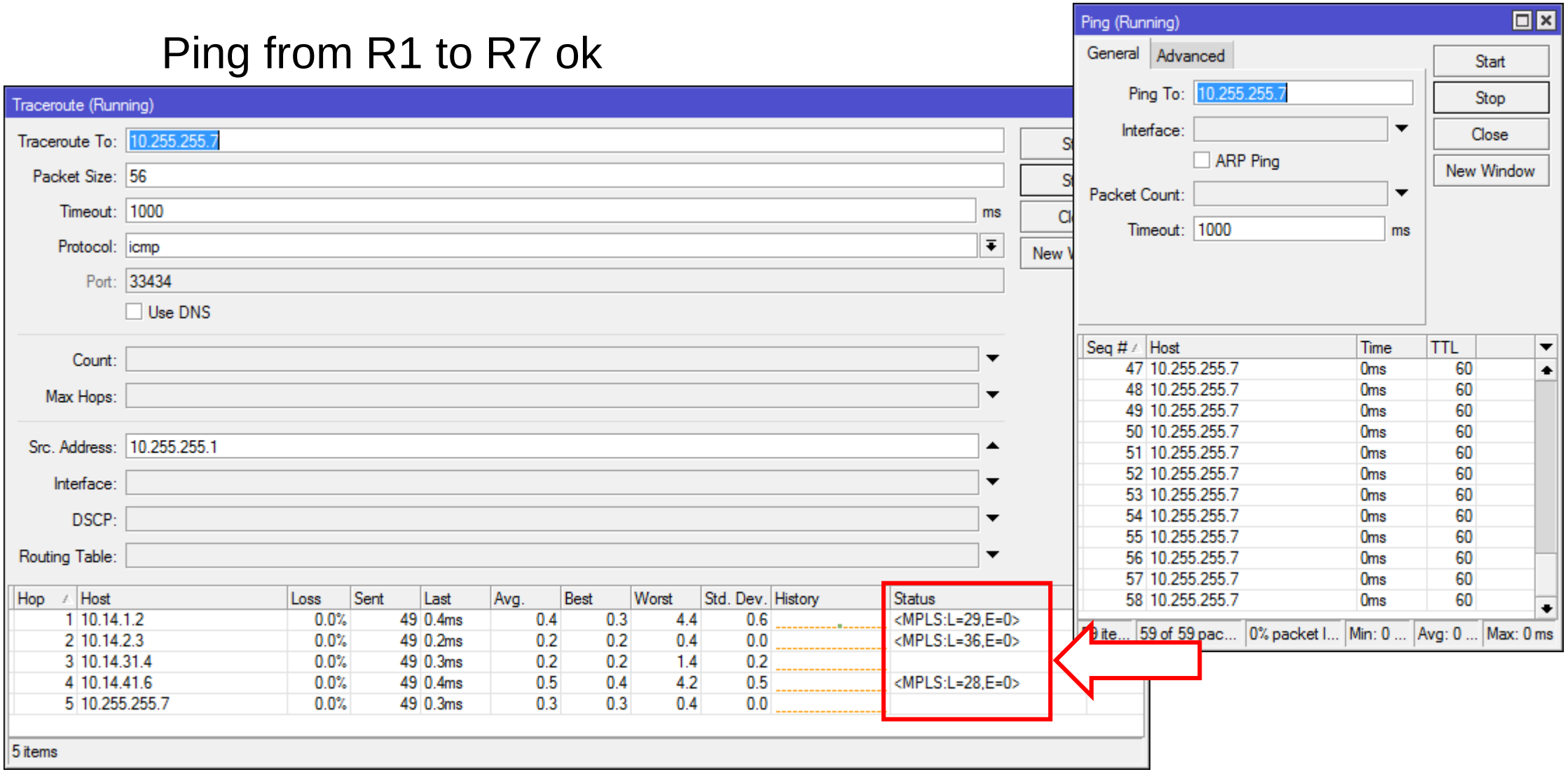
Observed behaviour

- Routing through R4
- PPPoE customers at R6, R7 *offline*



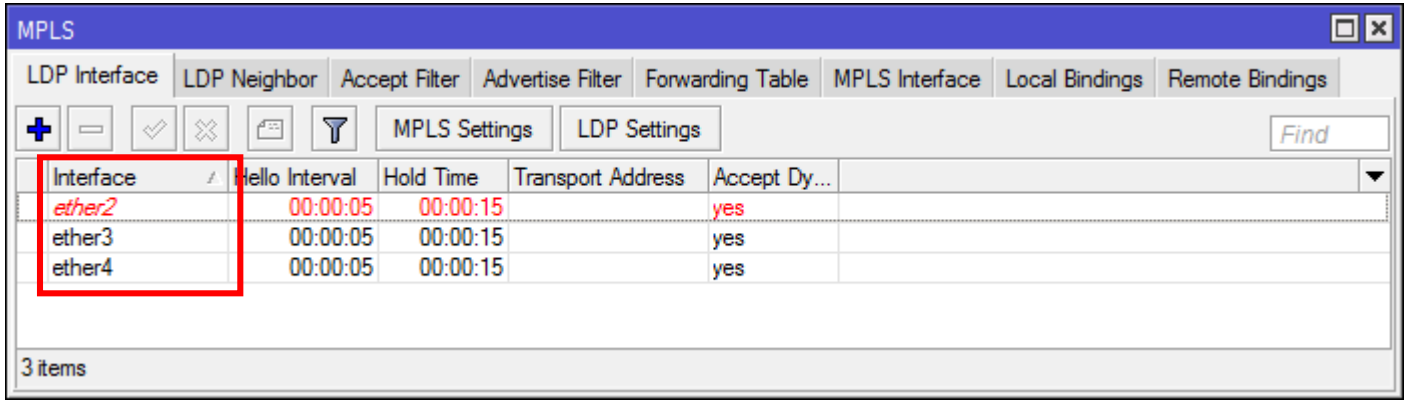
Failure at main link

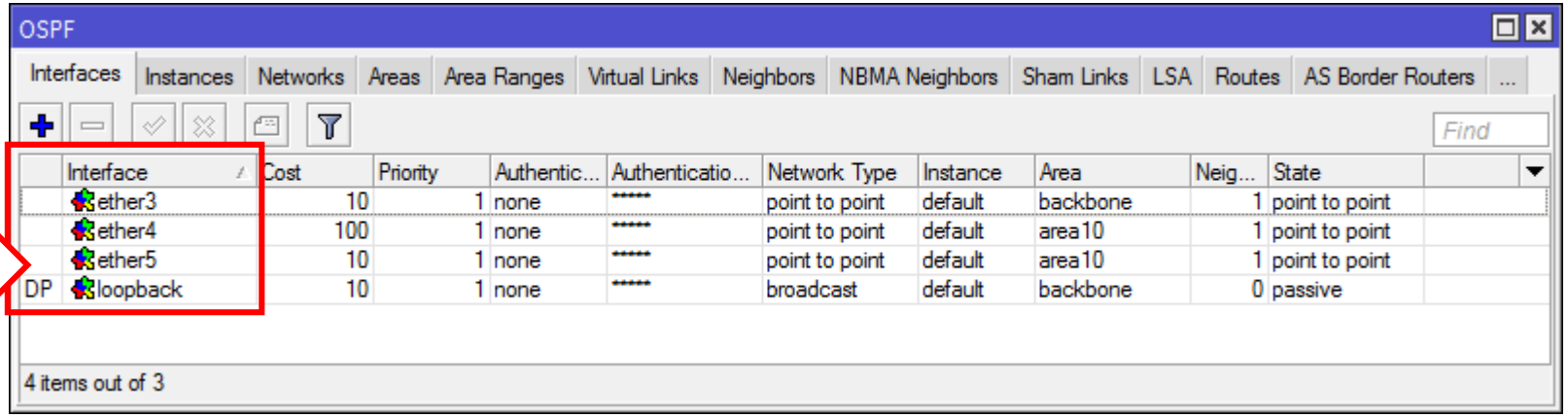
Ping from R1 to R7 ok



Wrong LDP interfaces at R3

- LDP:
  - *ether2*
  - ether3
  - ether4
- OSPF
  - ether3
  - ether4
  - ether5

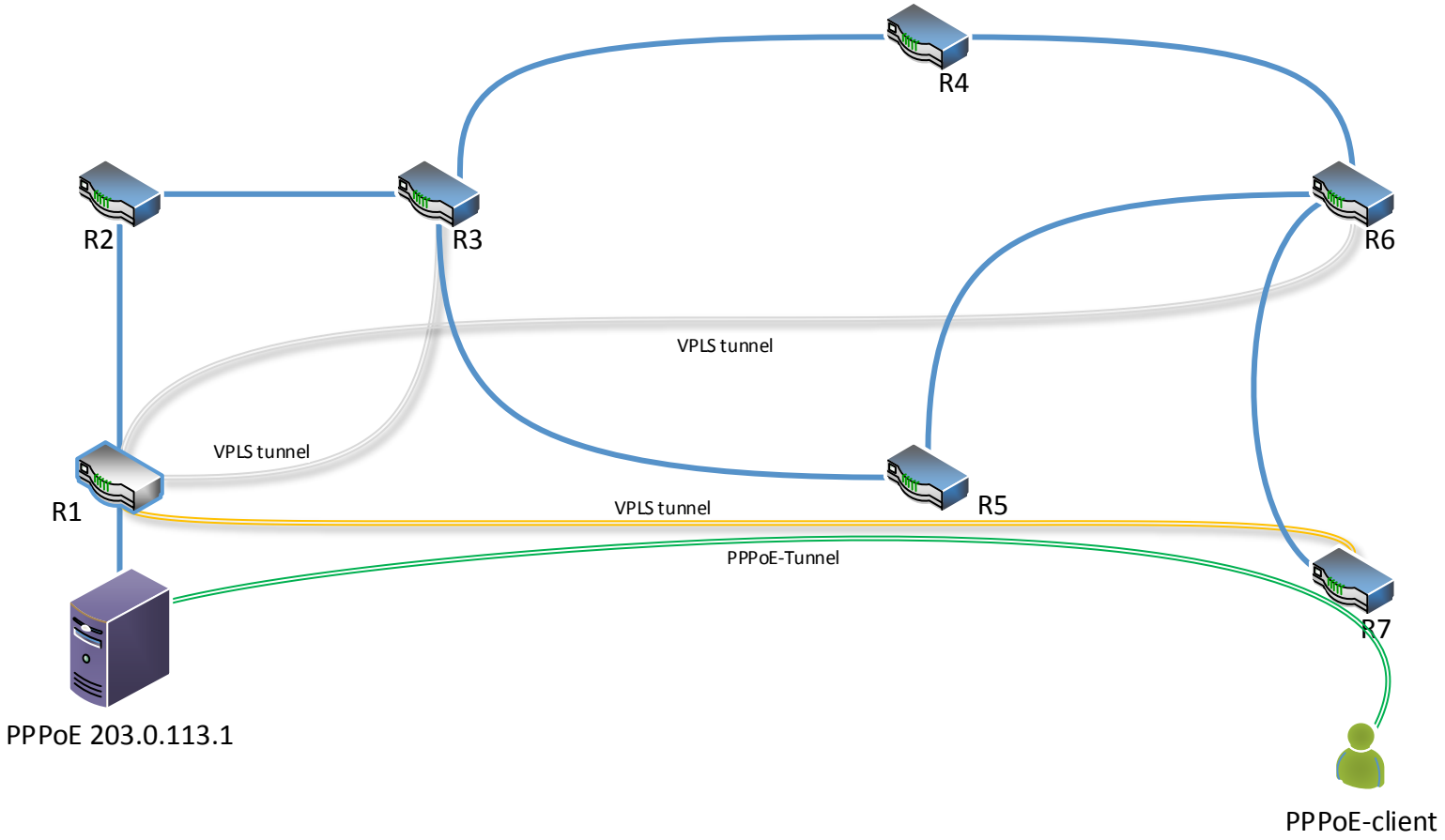




Examine setup

Monitor a PPPoE session

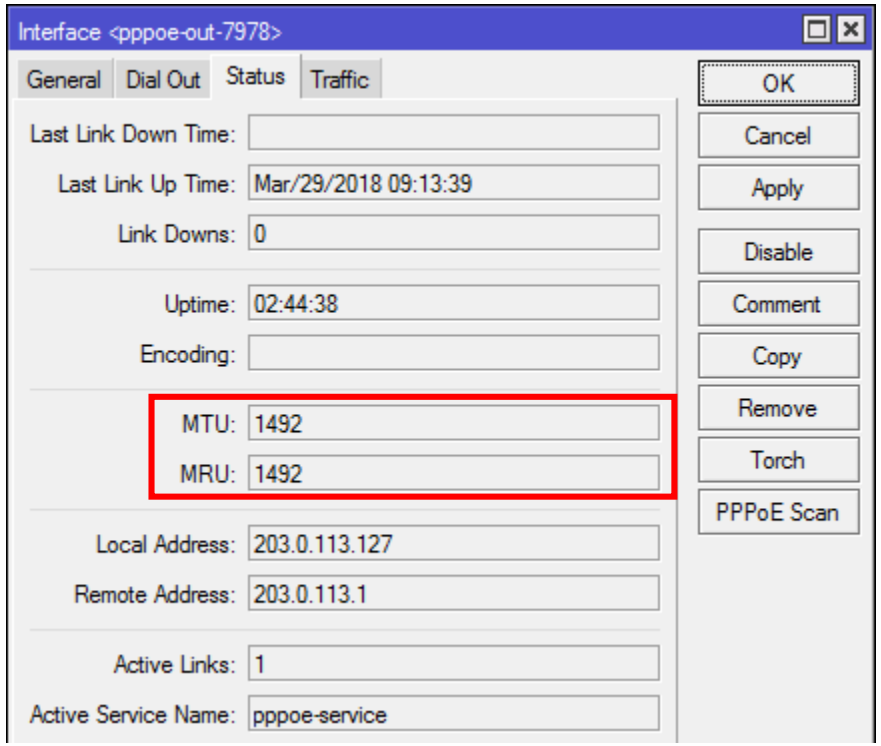
Bandwidth-test: PPPoE client to PPPoE server (download)

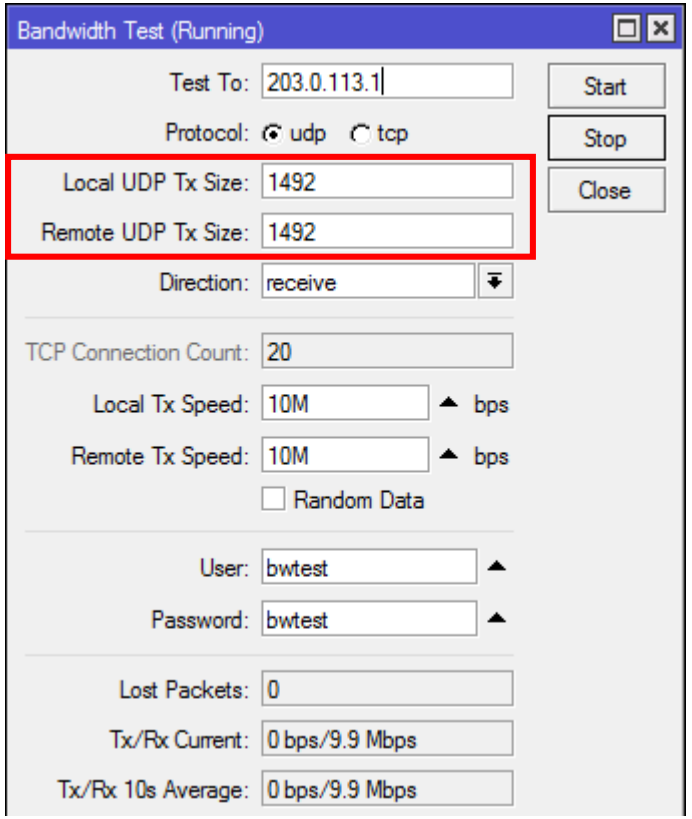


Monitor a PPPoE session

Bandwidth-test: PPPoE client to PPPoE server (download)

MTU PPPoE Client: 1492 → Bandwidth-test with 1492

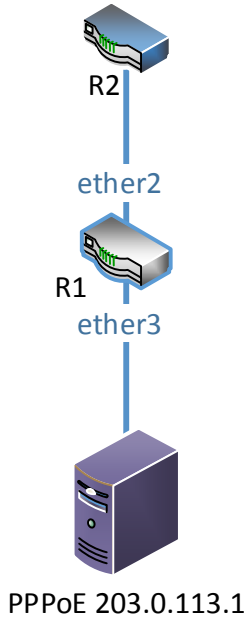


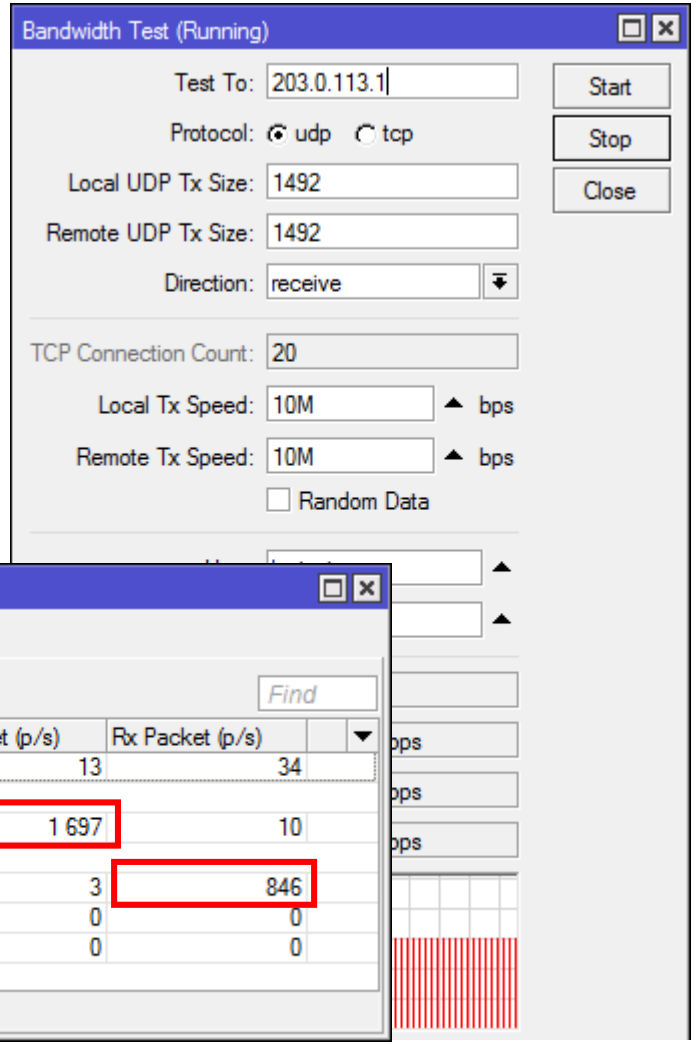


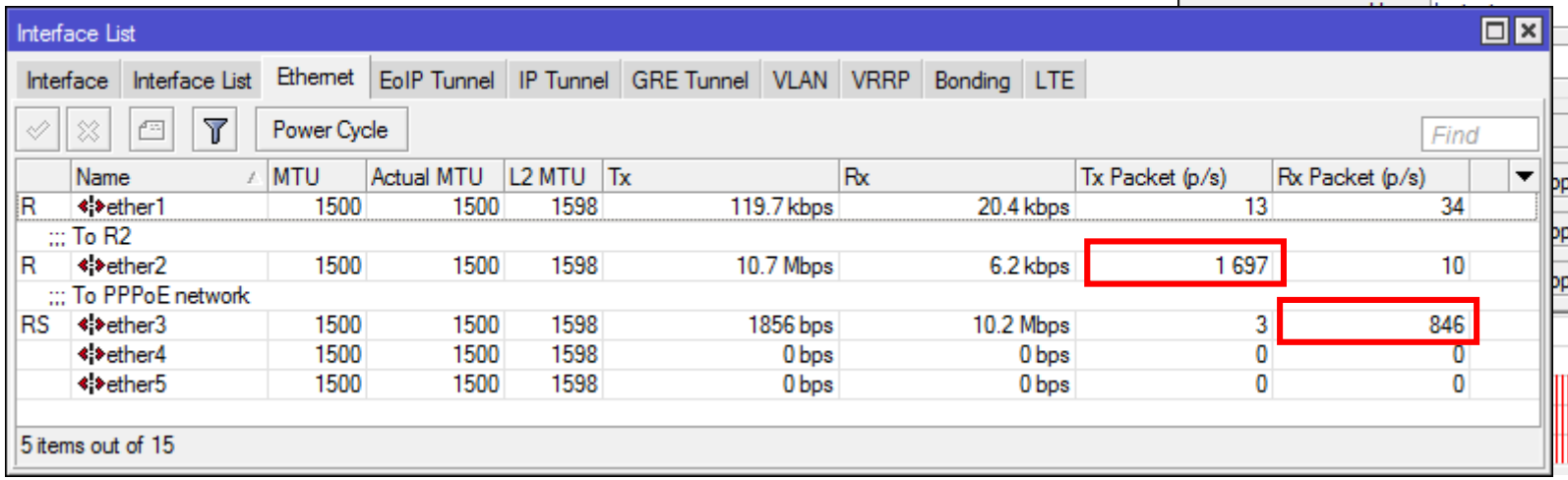
Monitor a PPPoE session

On R1

- Interface to R2: 1697 p/s
- Interface to PPPoE: 846 p/s





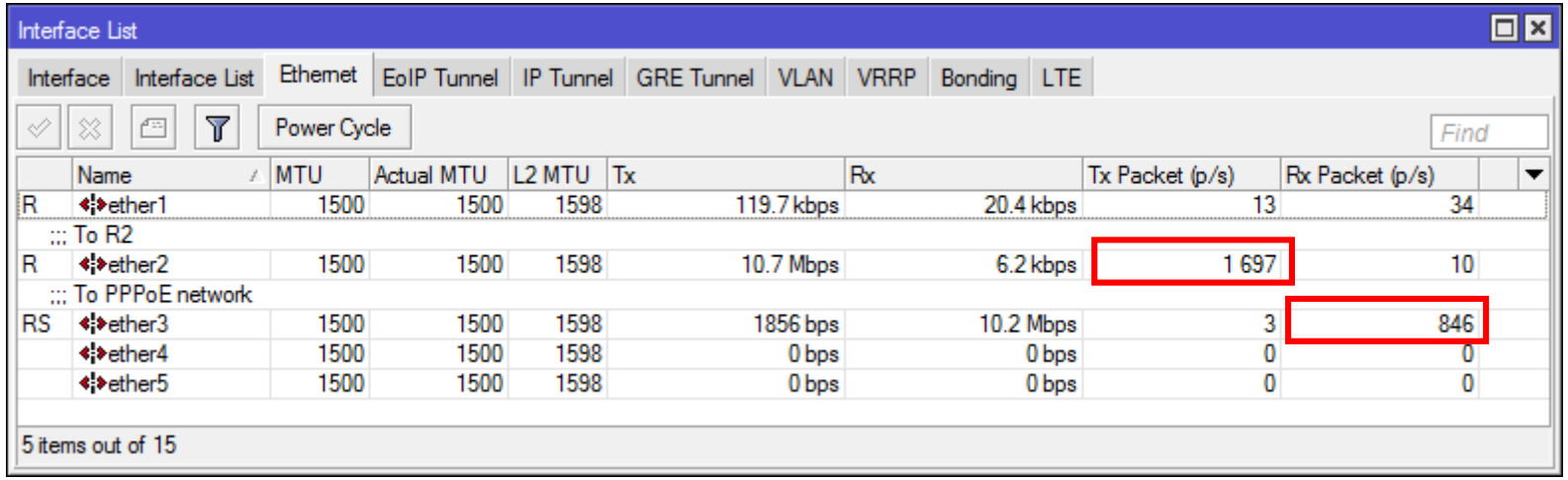


Fragmentation

Packet fragmentation?

Benefits of VPLS vs. EoIP

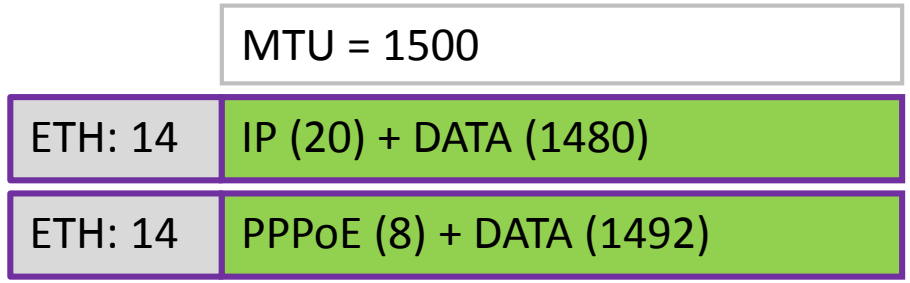
- VPLS: No fragmentation (*if done right*)



Packet sizes

Original frame

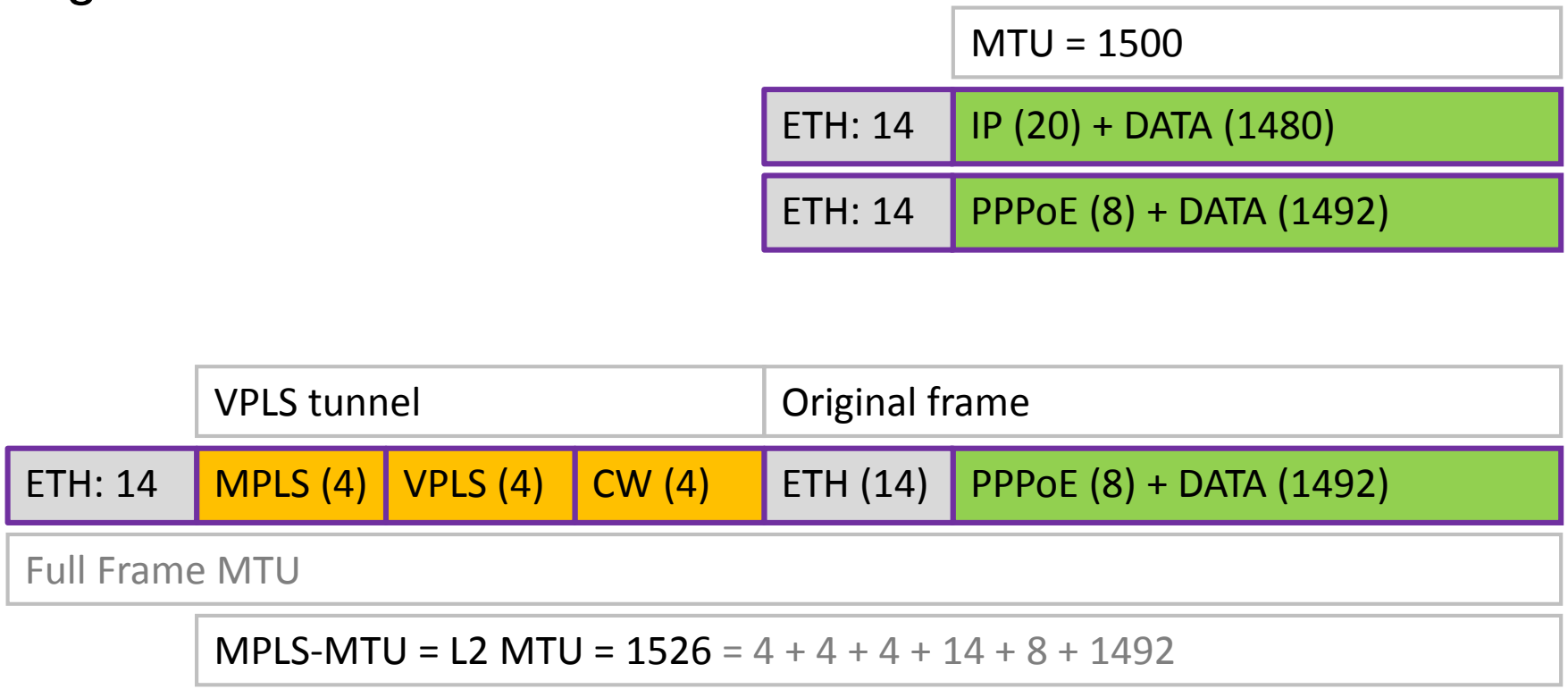
- L3 Size = 1500
- Full Frame Size = 1514



Packet sizes

Insertion of 1500 bytes (MTU) packet into VPLS tunnel:

No fragmentation

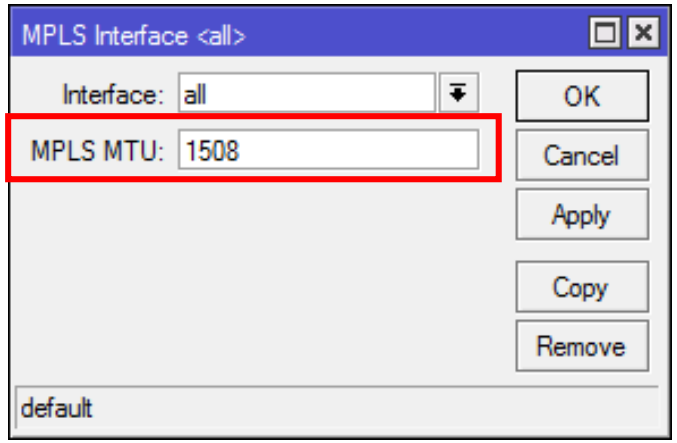


Packet sizes

VPLS packet is *fragmented* because:

Resulting MPLS-MTU: 1526

Interface MPLS MTU: 1508 (default)



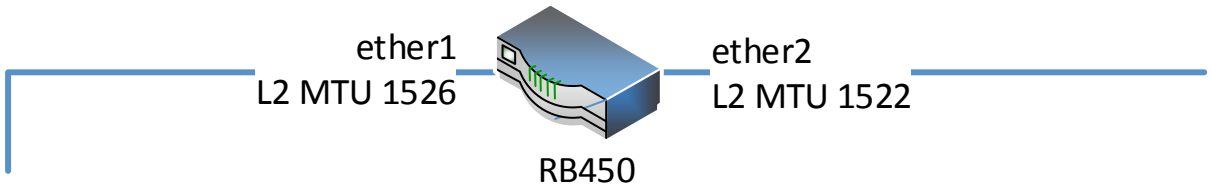
ETH: 14	MPLS (4)	VPLS (4)	CW (4)	ETH (14)	PPPoE (8) + DATA (1492)

MPLS-MTU = L2 MTU = 1526

Increase interface MPLS MTU

If hardware capable: Increase interface MPLS MTU

- L2 MTU (see Maximum Transmission Unit on RouterBoards)
  - RB433, RB450, RB493: ether1: 1526, ether2-last: *1522*
  - RB433GL, RB450G, RB493G: all interfaces: *1520*
  - *…*
- Switches, media converters, …

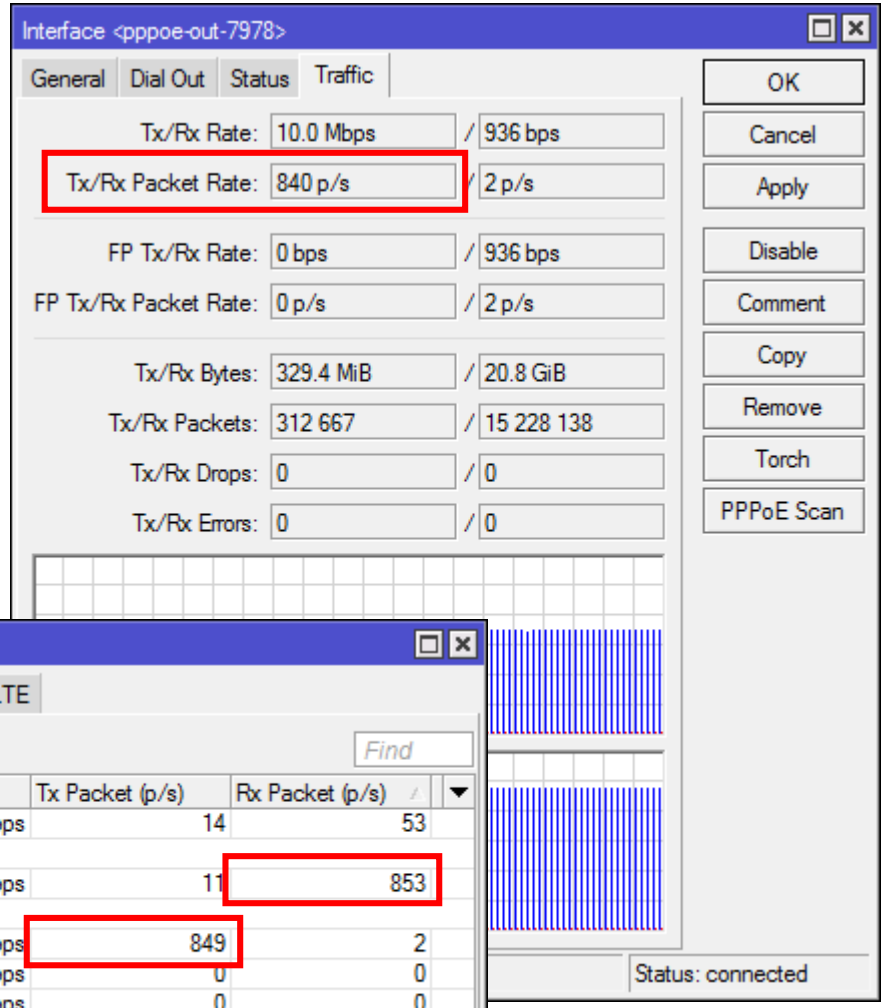


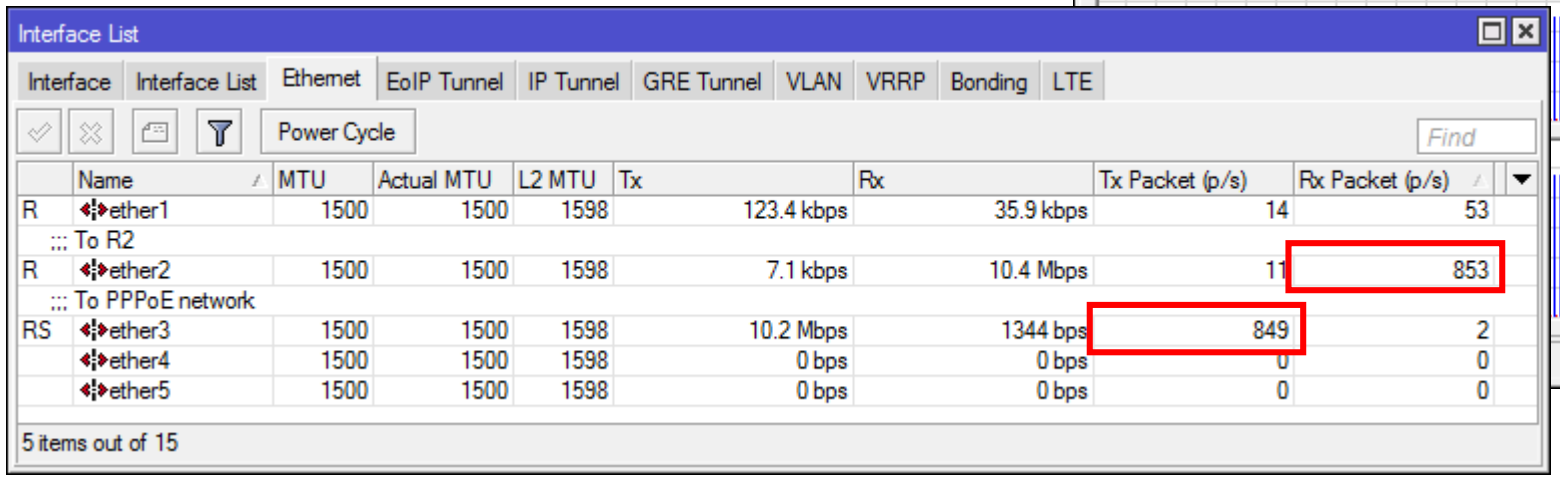
MPLS MTU set to 1526

MPLS Interface MTU: 1526 →

Corresponding packet counters

- PPPoE client
- Interface to backbone
- Interface to PPPoE server

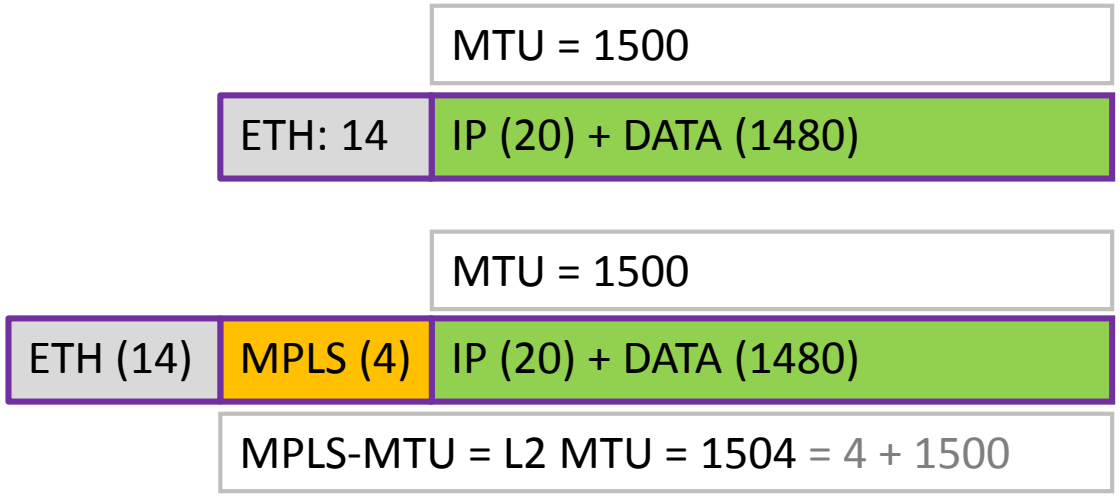




Why 1508?

1508 is enough for

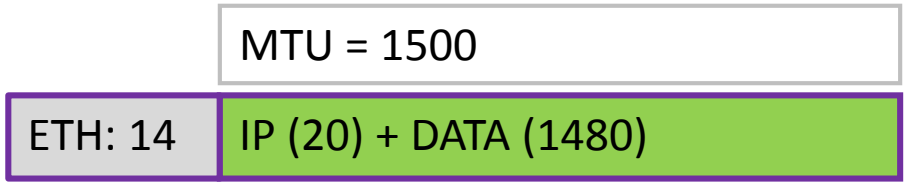
- MPLS for packet forwarding (1 MPLS label)



Why 1508?

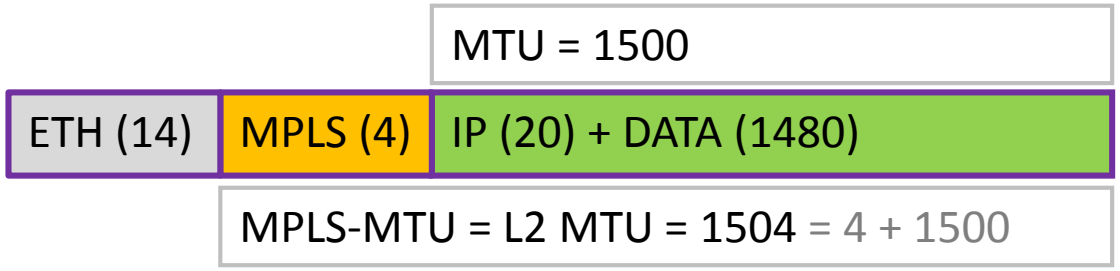
1508 is enough for

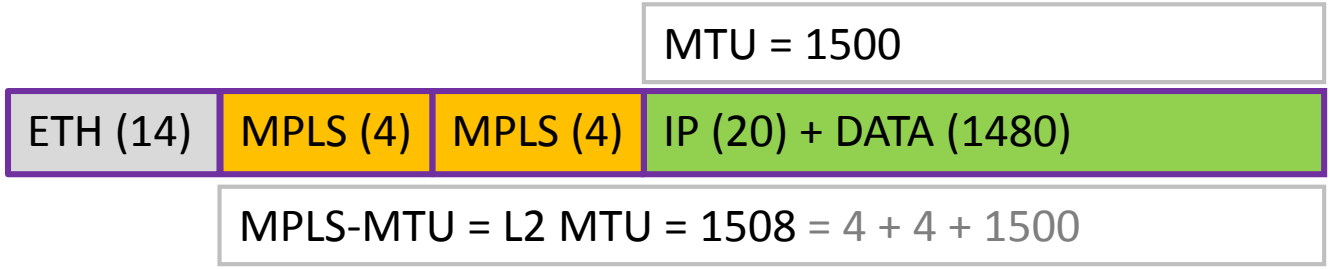
- MPLS for packet forwarding (1 MPLS label)
- Targeted LDP (2 MPLS labels)



Default 1526

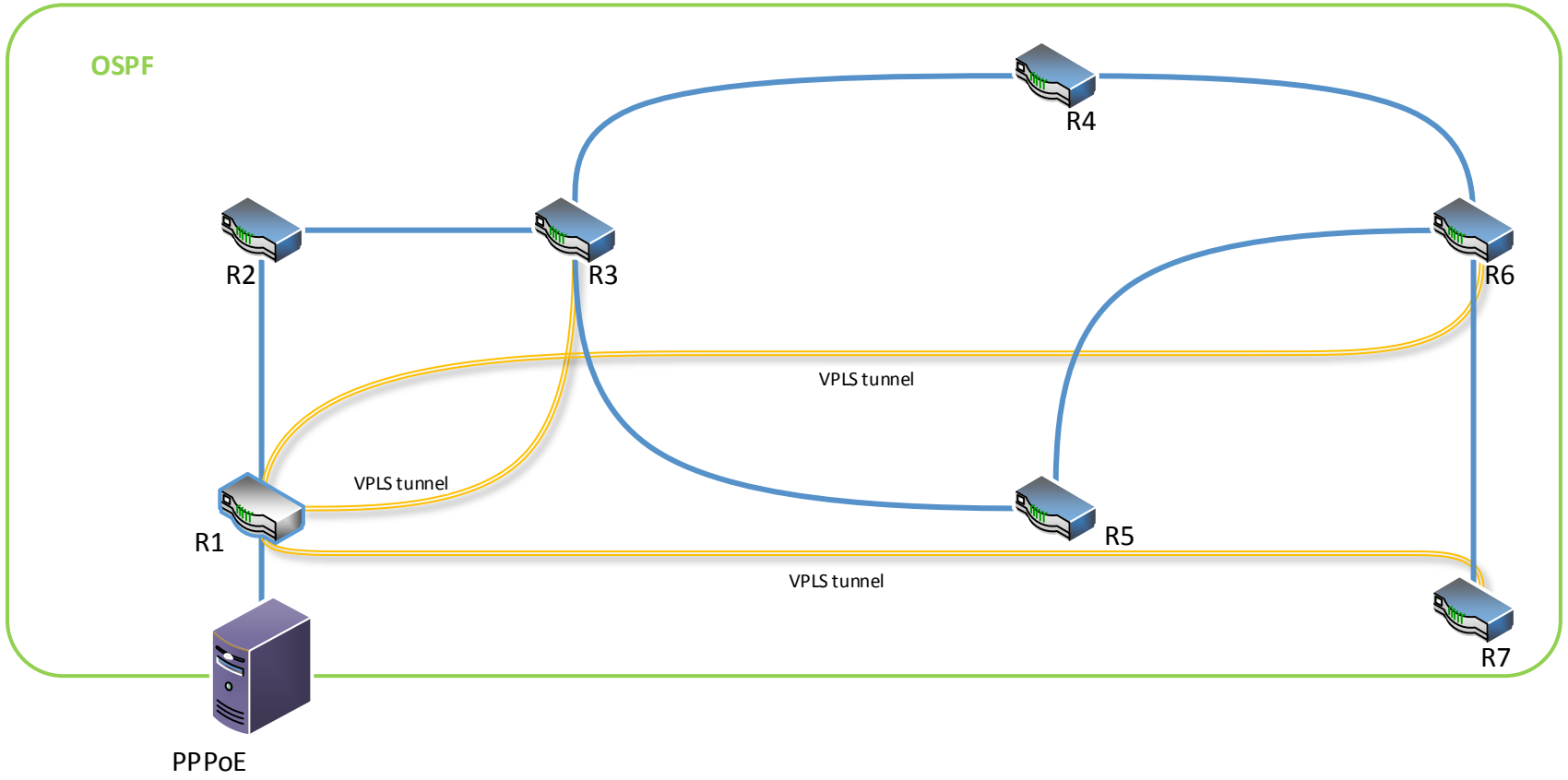
- Too large (?)





Network improvements

Current network



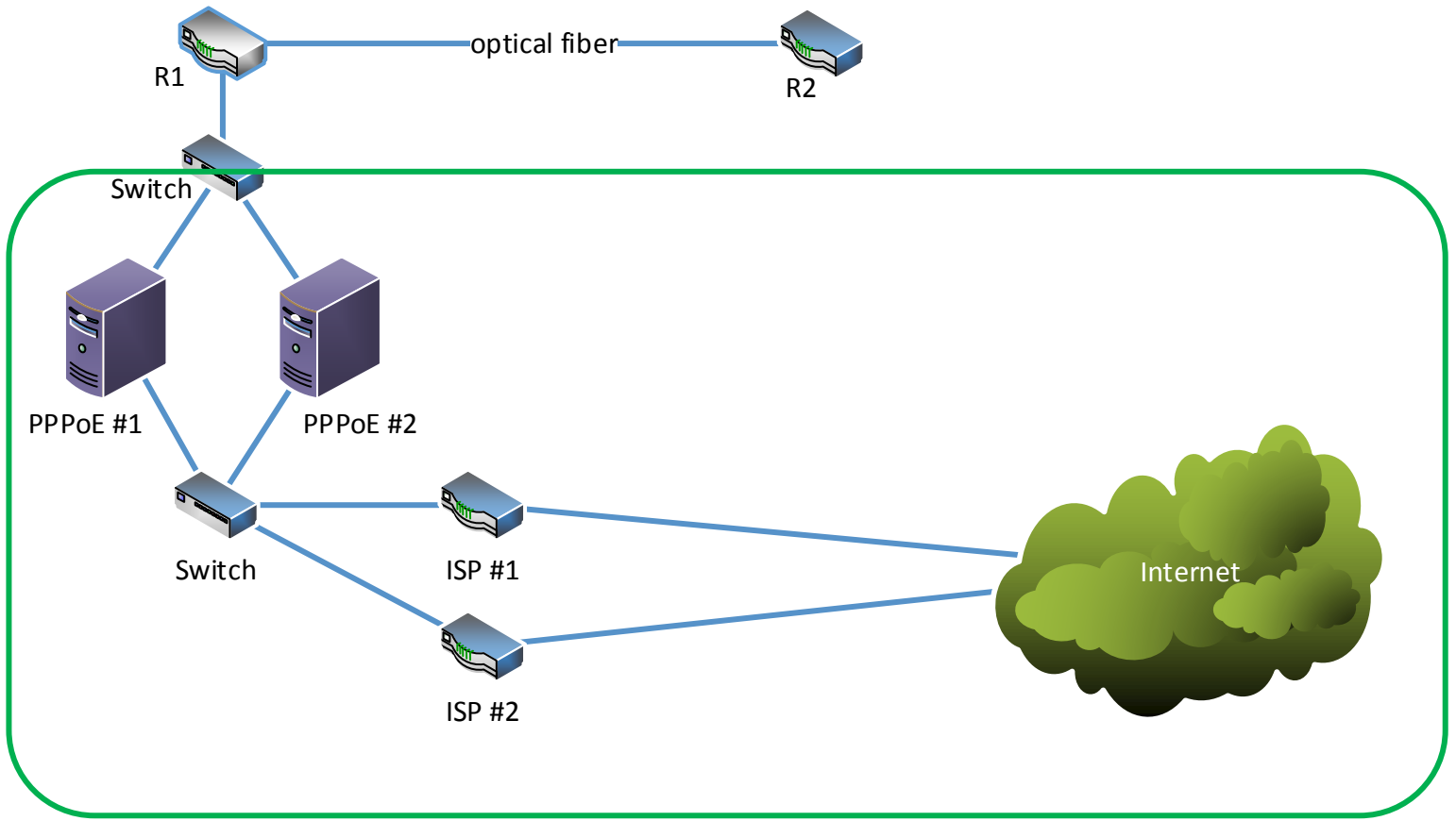
Redundancy

Redundancy:

- Type / coverage depends on
  - setup
  - needs
  - customer / network
- No claim for completeness
- Examples

Redundancy can become complex. Complexity can result in issues.

Redundancy at main site



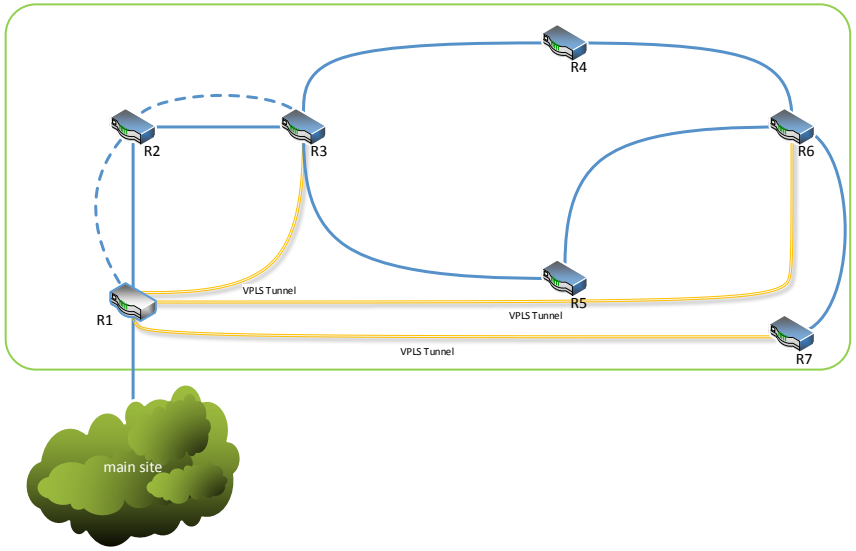
Green frame: See presentation of Patrik Schaub

(Access all FMS Internetservice presentations: click)

Redundancy at backbone

- Additional link / ip subnet between R1/R2 and R2/R3
- 2nd link is backup – same as on R3
  - OSPF interfaces:

    High(er) cost for backup link
- Don‘t forget to add LDP interface



OSPF

Interfaces | Instances | Networks | Areas | Area Ranges | Virtual Links | Neighbors | NBMA Neighbors | ...

Find

	Interface	Cost	Network Type	Instance	Area	Neighbors	State
	;;; From R1 to R2, main link						
	ether2	10	point to point	default	area5	1	point to point
	;;; From R1 to R2, backup link						
	ether5	100	point to point	default	area5	1	point to point
	;;; Loopback interface, passive						
P	loopback	10	broadcast	default	area5	0	passive

3 items

MPLS

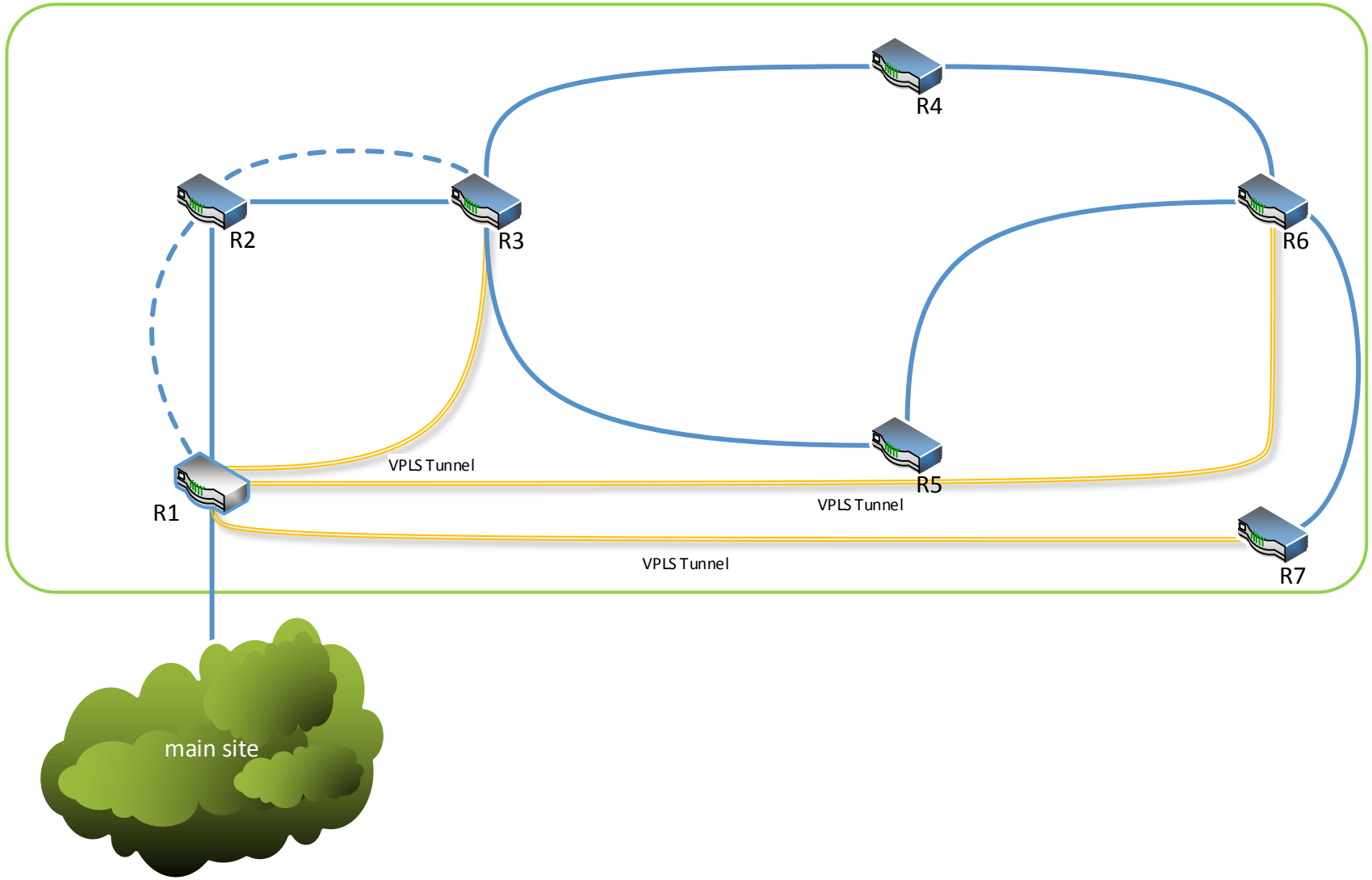
LDP Interface | LDP Neighbor | Accept Filter | Advertise Filter | Forwarding Table | ...

MPLS Settings | LDP Settings | Find

Interface	Hello Interval	Hold Time	Transport Address	Accept Dy...
ether2	00:00:05	00:00:15		yes
ether5	00:00:05	00:00:15		yes

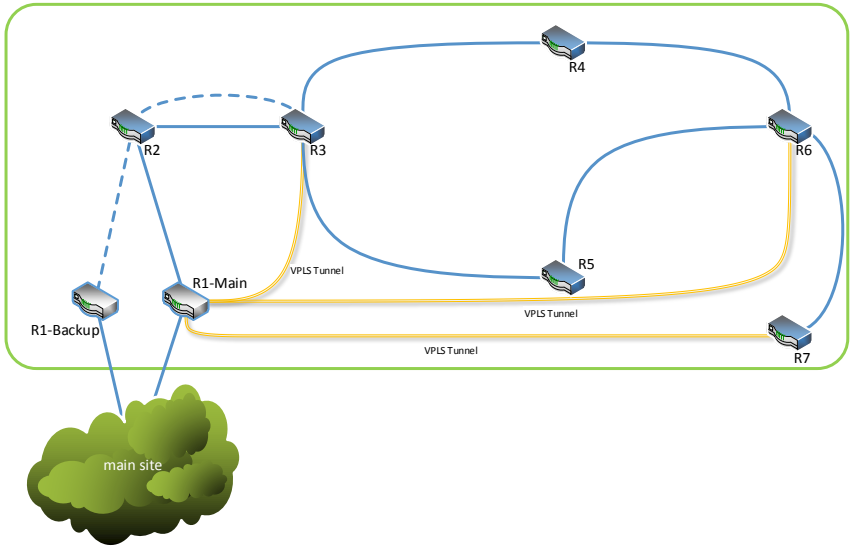
2 items

Redundancy at backbone

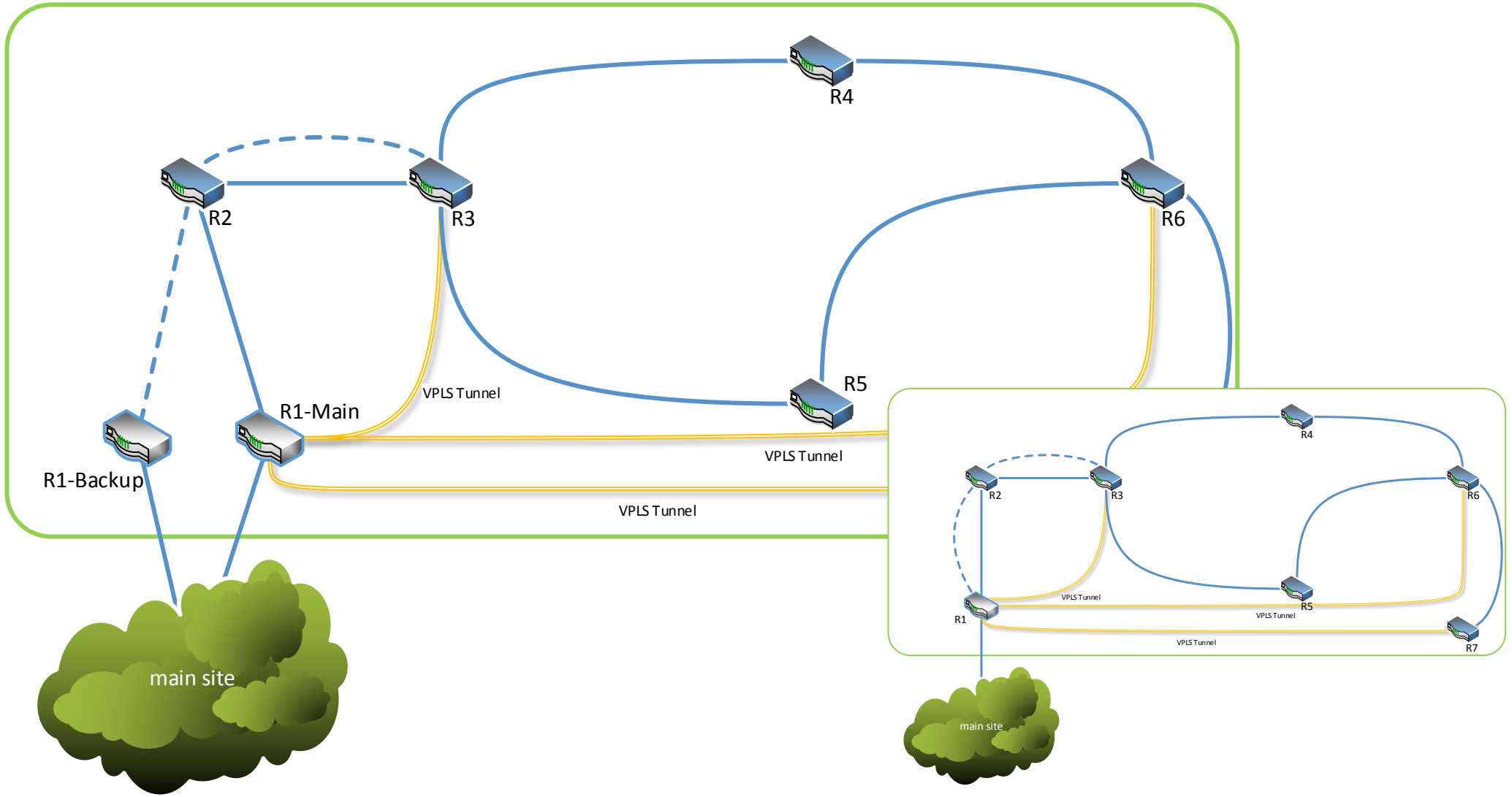


Redundancy for R1

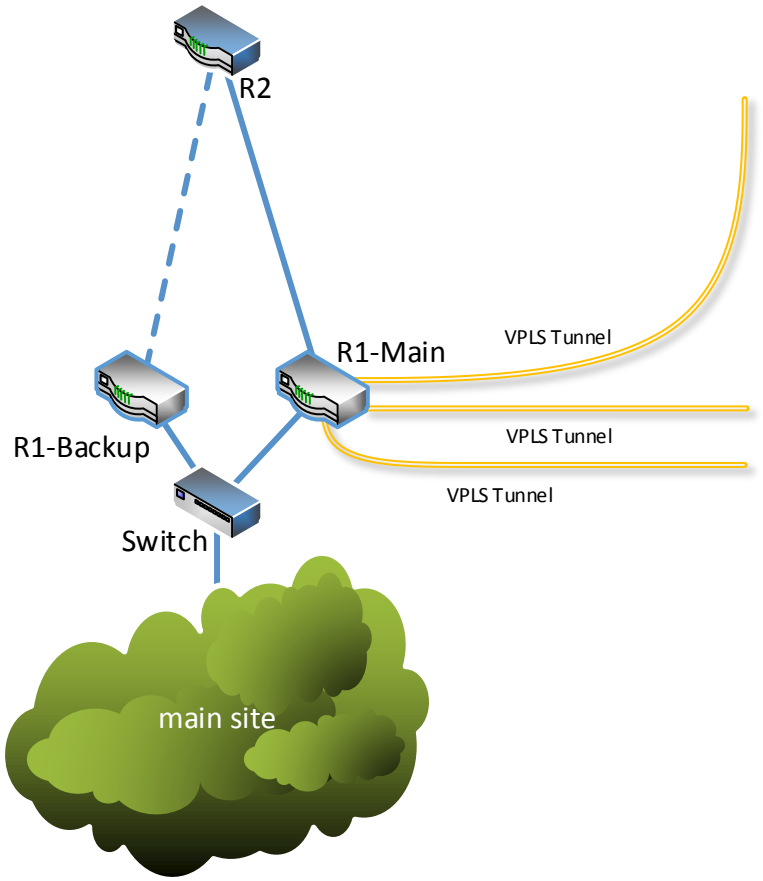
- Clone R1:
  - R1-Main (10.255.255.11)
  - R1-Backup (10.255.255.12)
- Main link connected to R1-Main
- Backup link connected to R1-Backup
- VPLS go to R1-Main (10.255.255.1)



Redundancy for R1



Redundancy for R1

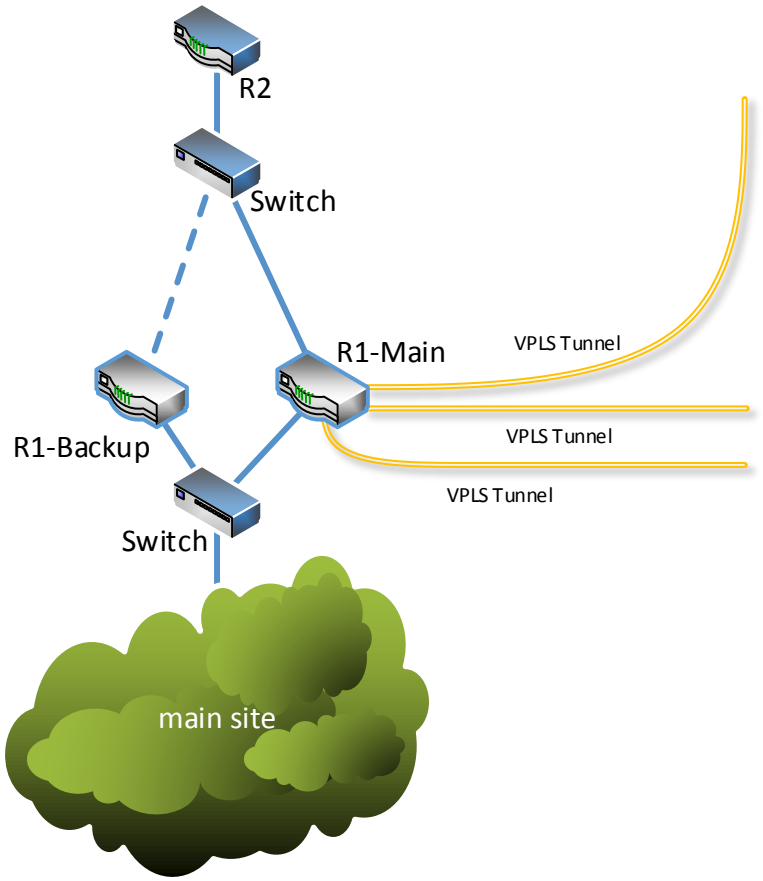


Who is R1-Main / R1-Backup?

Who is 10.255.255.1?

- No VRRP between Main / Backup on Interface to R2 (different L3 networks)

Redundancy for R1

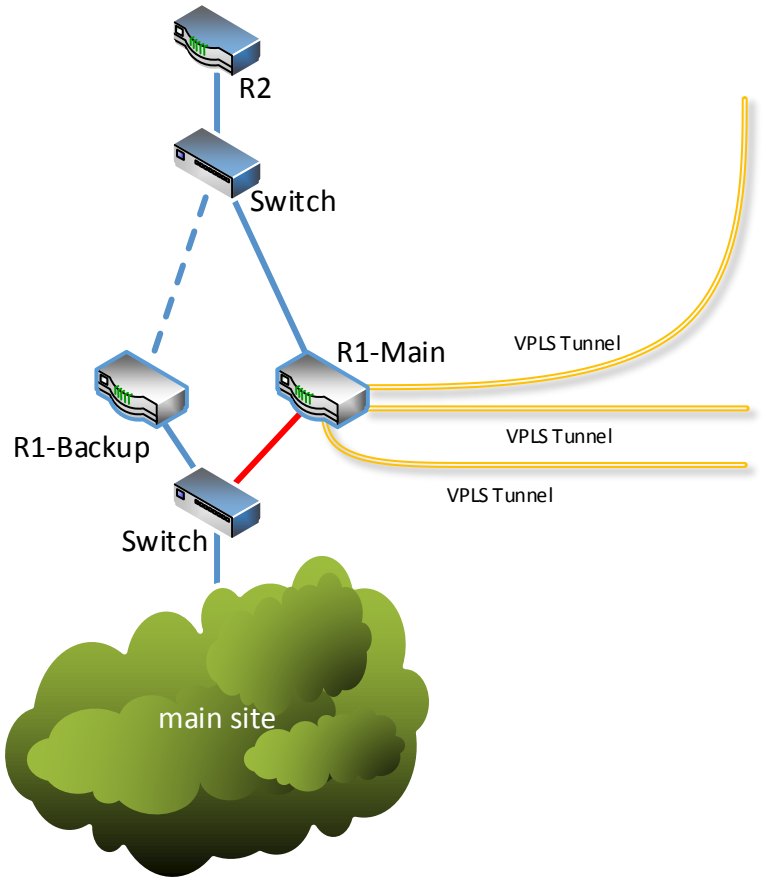


Who is R1-Main / R1-Backup?

Who is 10.255.255.1?

- Same L2 for R1-Main, R1-Backup and R2
  - VRRP on R2 side
  - Backup path: Decission by RSTP

Redundancy for R1

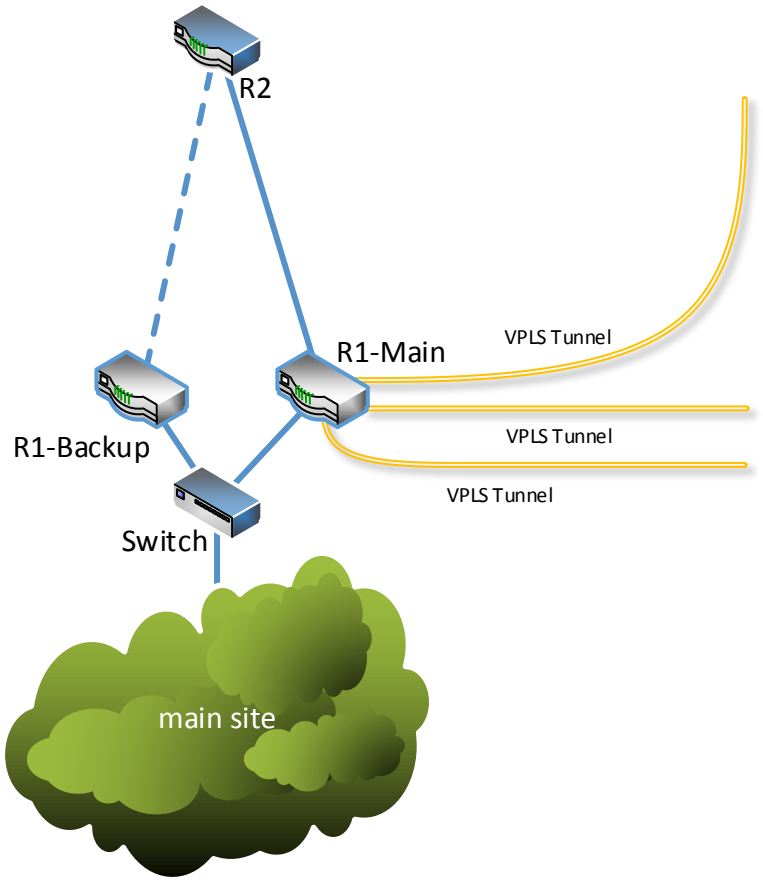


Who is R1-Main / R1-Backup?

Who is 10.255.255.1?

- Same L2 for R1-Main, R1-Backup and R2
  - VRRP on R2 side
  - Backup path: Decission by RSTP
- Failure on link to main site
  - VRRP is fine
  - Clients offline

Redundancy for R1



Who is R1-Main / R1-Backup?

Who is 10.255.255.1?

- R1-Main and R1-Backup:

  Connected to main site switch
- VRRP on this side
  - Management VLAN?

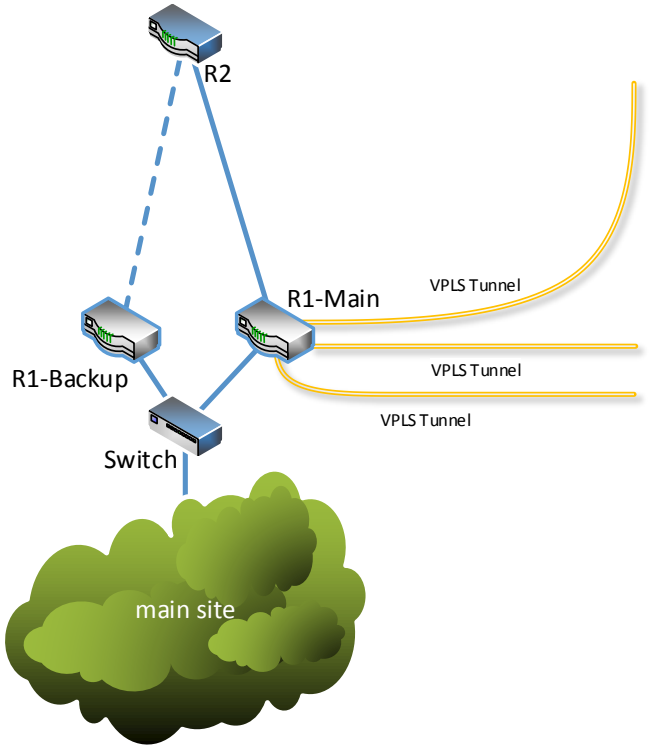
Redundancy for R1

VRRP and MPLS on R1-Main

- IP 10.255.255.1/32 on VRRP interface
- LSR ID = Transport address 10.255.255.1
- 10.255.255.11/32 on loopback, for OSPF

VRRP and MPLS on R1-Backup

- IP 10.255.255.1/32 on VRRP interface
- LSR ID = Transport address 10.255.255.1
- 10.255.255.12/32 on loopback, for OSPF



Let's ~~break~~ test things

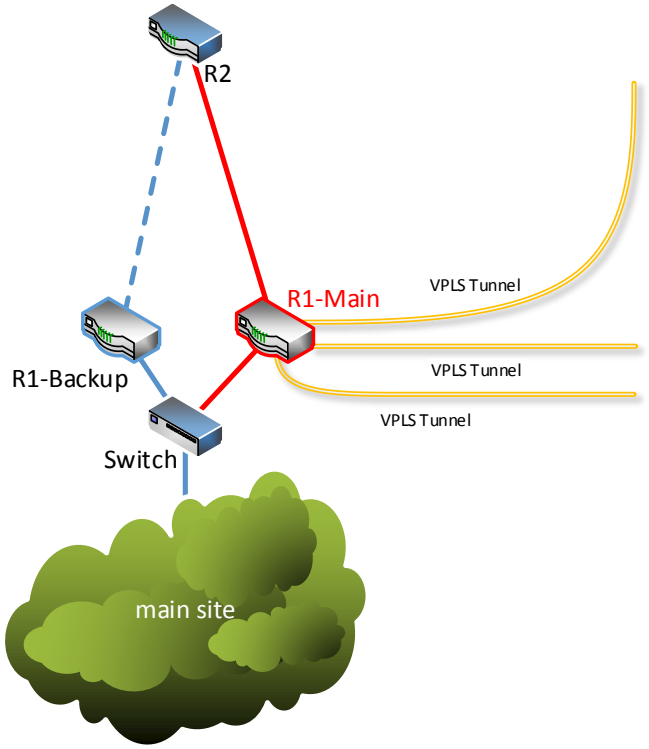
Failure of R1-Main or failure of link to main site

Expected behaviour

- 10.255.255.1 on R1-Backup
- VPLS tunnels to R1-Backup → up
- PPPoE clients reconnecting

Observed behavour

- Everything fine (~~stop testing!~~)

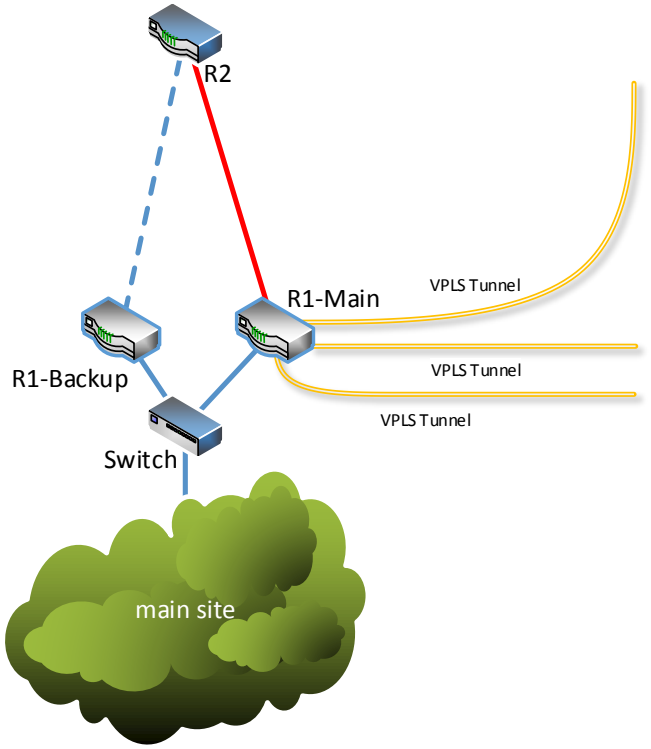


Let's ~~break~~ test things

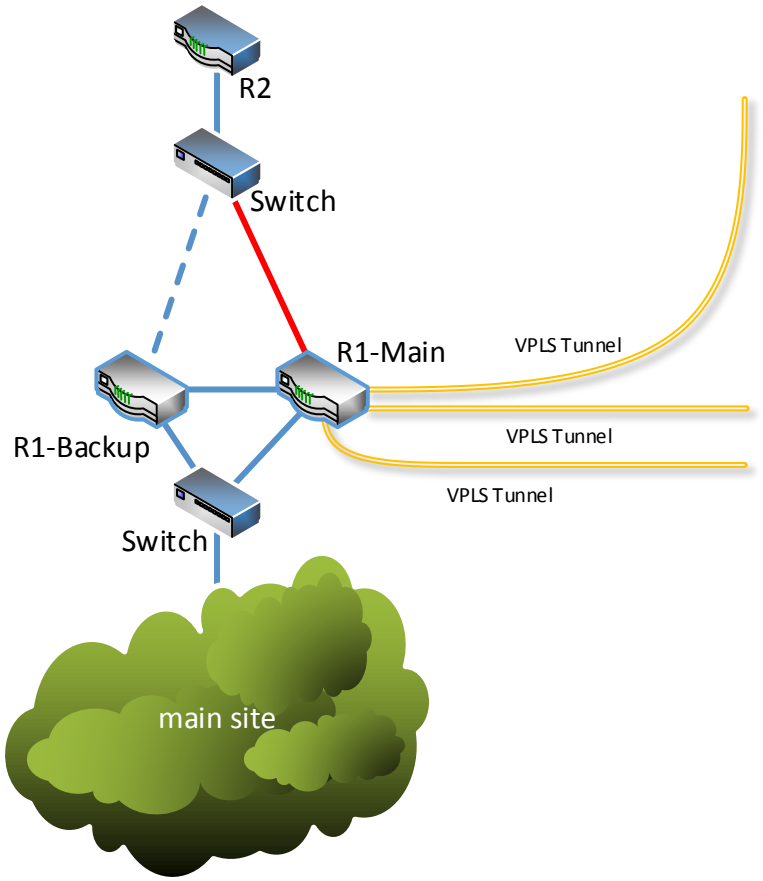
Failure of link R1-Main to R2

Expected behaviour

- 10.255.255.1 on R1-Main
- R1-Main VPLS master
- R2: *No route* to 10.255.255.1 (OSPF)
- Clients offline



Let's ~~break~~ test things



OSPF and LDP on crosslink

Expected behaviour

- 10.255.255.1 on R1-Main
- R2: route to 10.255.255.1
- VPLS ok & clients online

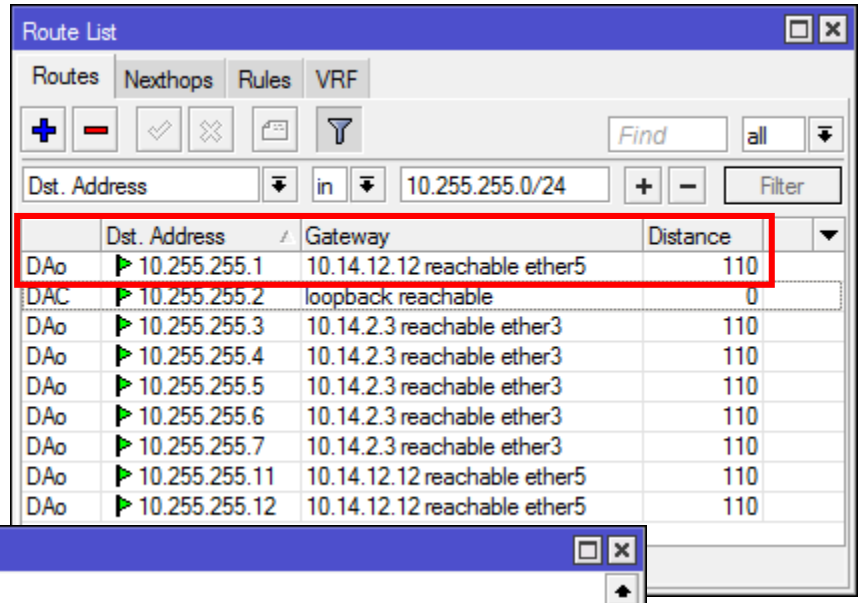
Observed behaviour

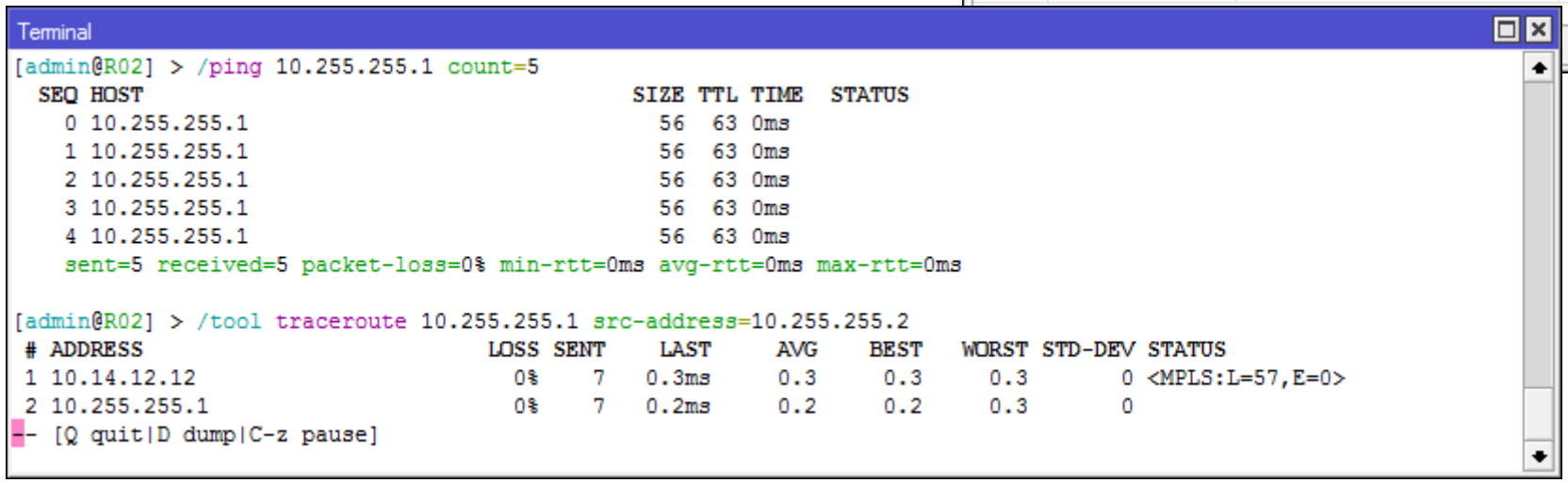
- Clients offline

Let's ~~break~~ test things

Tests from R2:

- Route via R1-Backup
- Ping to 10.255.255.1 ok
- Traceroute ok





Let's ~~break~~ test things

Tests from R7:

- Ping to 10.255.255.1 ok
- Traceroute… ?

Terminal

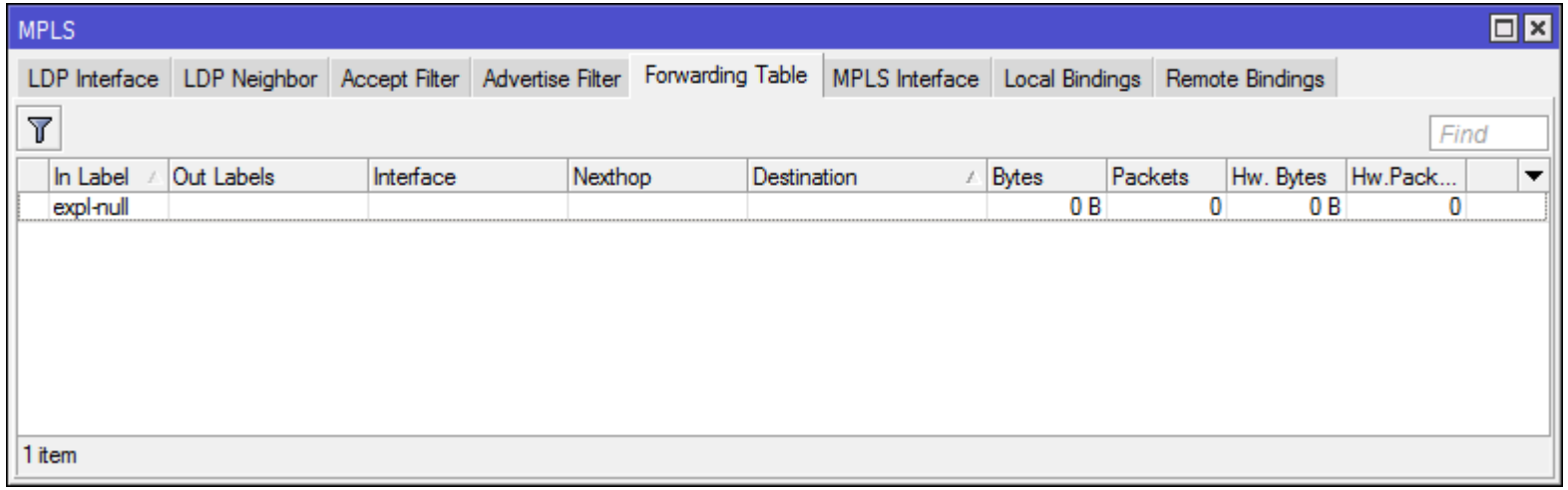
```
[admin@S:R07] > /ping 10.255.255.1 count=5
  SEQ HOST                                     SIZE TTL TIME  STATUS
    0 10.255.255.1                               56  59 0ms
    1 10.255.255.1                               56  59 0ms
    2 10.255.255.1                               56  59 0ms
    3 10.255.255.1                               56  59 0ms
    4 10.255.255.1                               56  59 0ms
    sent=5 received=5 packet-loss=0% min-rtt=0ms avg-rtt=0ms max-rtt=0ms

[admin@S:R07] > /tool traceroute 10.255.255.1 src-address=10.255.255.7
 # ADDRESS                                LOSS SENT    LAST     AVG    BEST   WORST STD-DEV STATUS
 1 10.14.5.6                               0%   14   0.5ms     0.5     0.5     0.6       0 <MPLS:L=51,E=0>
 2 10.14.41.4                              0%   14   0.4ms     0.4     0.4     0.5       0 <MPLS:L=61,E=0>
 3 10.14.31.3                              0%   14   0.4ms     0.4     0.3     0.7     0.1 <MPLS:L=53,E=0>
 4 10.14.2.2                               0%   14   0.4ms     0.3     0.3     0.4       0 <MPLS:L=50,E=0>
 5 10.14.12.12                             0%   14   0.3ms     0.3     0.3     0.3       0
 6 10.255.255.1                            0%   14   0.4ms     0.3     0.3     0.4       0
-- [Q quit|D dump|C-z pause]
```

Let's ~~break~~ test things

Routing between R1-Backup and R1-Main ok

MPLS/LDP broken on R1-Backup

- No forwarding Table

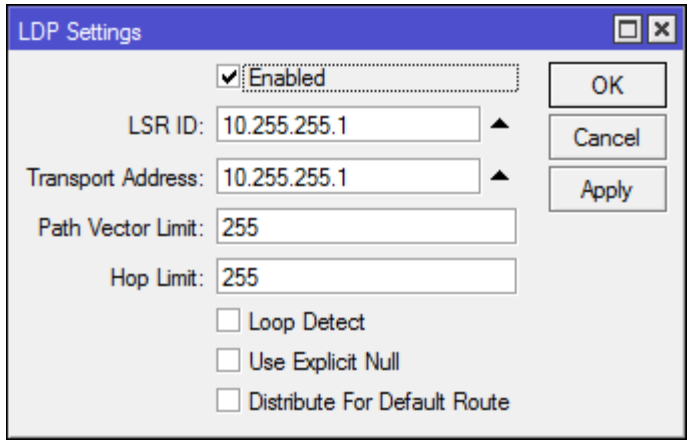


Routing is not enough for VPLS!

Let's ~~break~~ test things

Simple reason:

- LSR ID and Transport Address 10.255.255.1 is used on R1-Backup *and* R1-Main(!)
- IP 10.255.255.1 is active only on R1-Main (VRRP master)
- Duplicate ID (and transport address): Good idea? (No.)



Let’s fix things

(One possible) Solution:

- On VRRP Master:

  Set LSR ID and Transport Address to 10.255.255.1

- On VRRP Backup:

  Set LSR ID and Transport Address to router unique address

  (available on loopback)

Result: Working MPLS between routers

(OSPF was useing unique address as Router ID.)

Let’s fix things

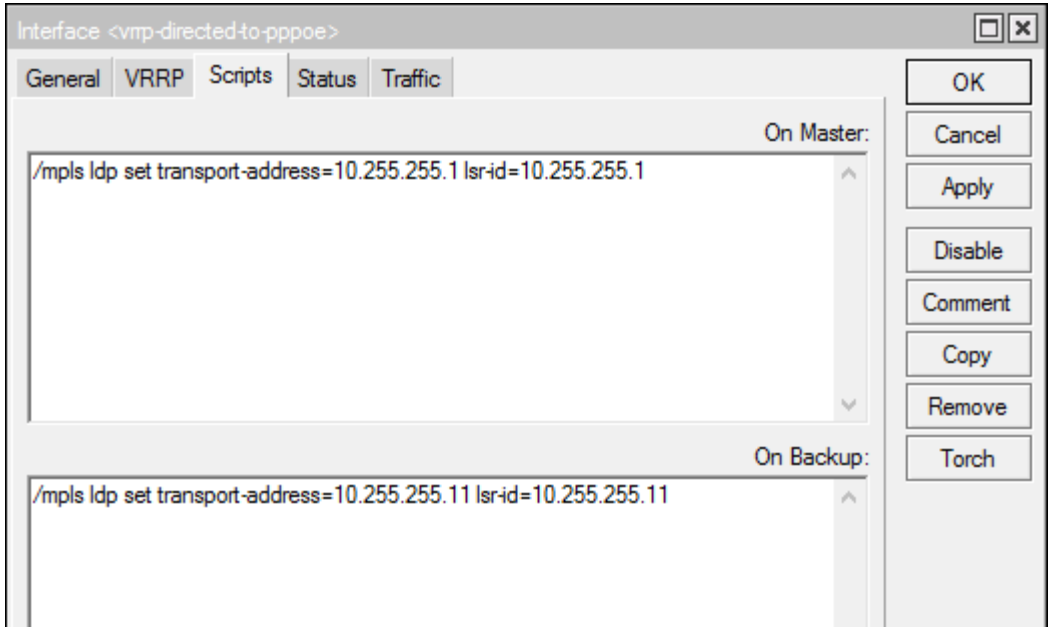
```
/interface vrrp
add interface=ether3 name=vrrp-directed-to-pppoe \
on-backup="/mpls ldp set transport-address=10.255.255.11 lsr-id=10.255.255.11" \
on-master="/mpls ldp set transport-address=10.255.255.1 lsr-id=10.255.255.1" \
preemption-mode=no vrid=5
```

R1-Main: 10.255.255.11

R1-Backup: 10.255.255.12

Note: Change of LSR ID

Service affecting



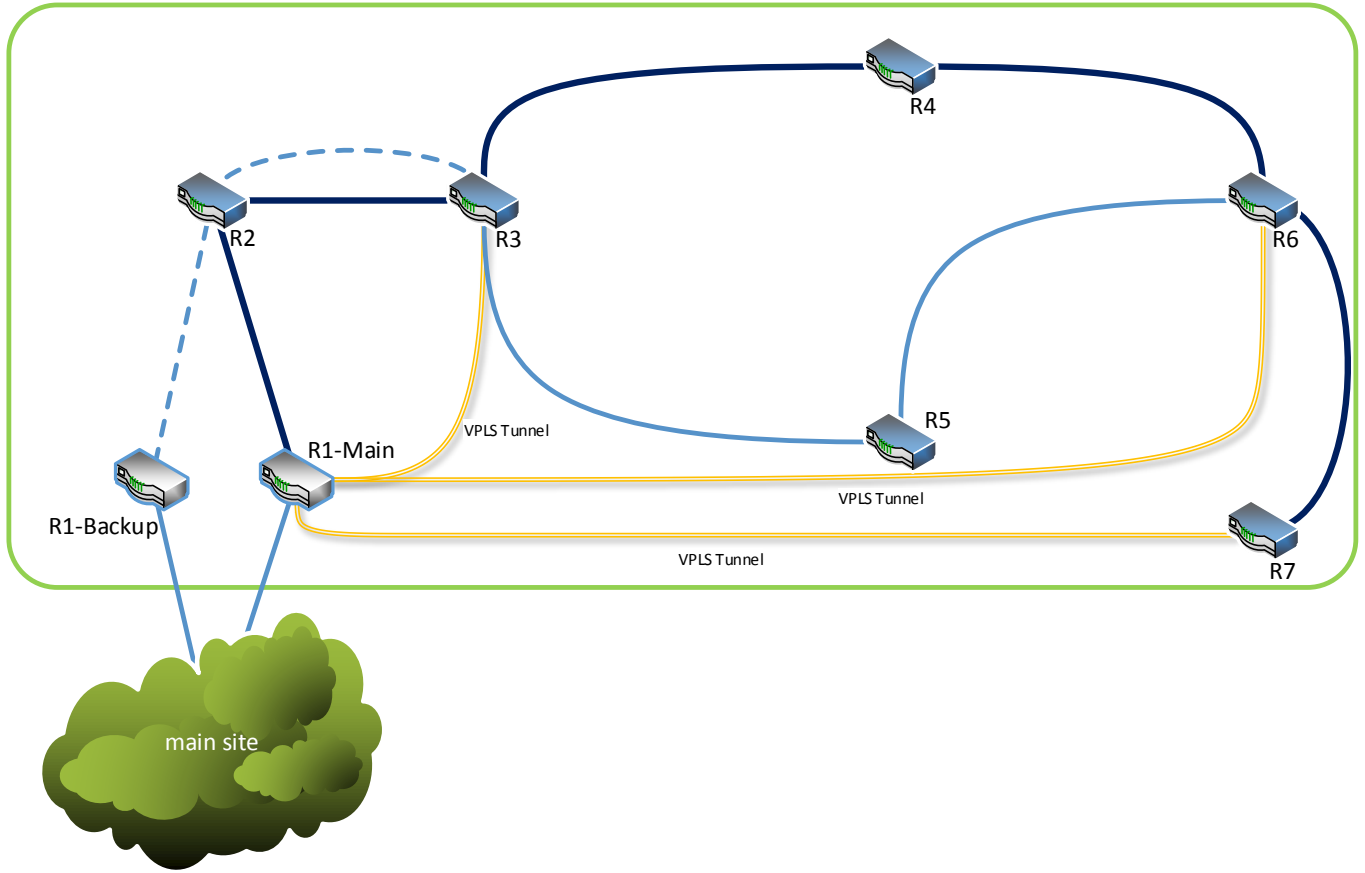
Traffic improvement

Use backup link

- Traffic from R7 to R1 through R4

But:

- OSPF goes through R5
- MPLS goes through R5
- VPLS goes through R5



Traffic engineering (TE) tunnel

- Enable TE support on all involved interfaces

For example on R3:

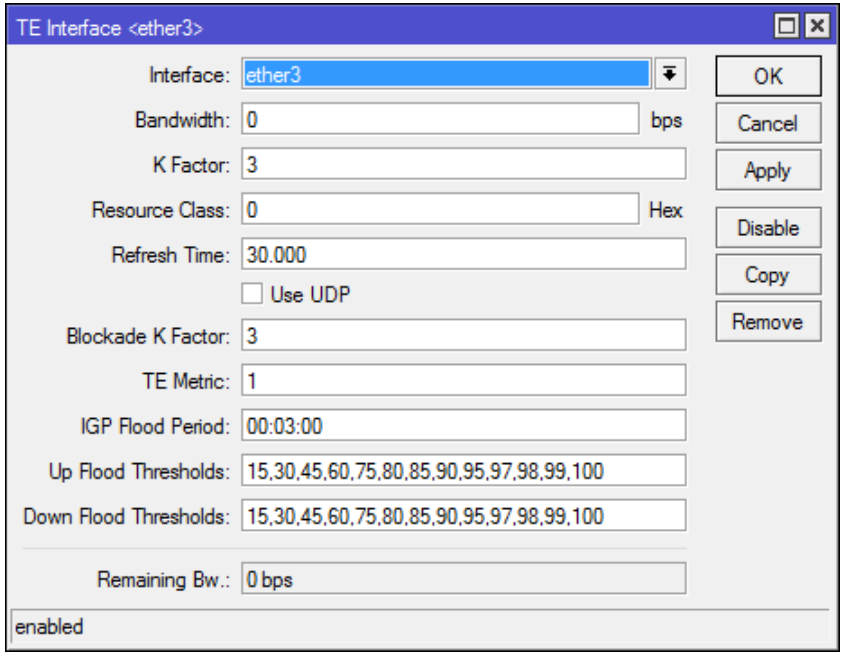
/mpls traffic-eng interface

add interface=ether3

add interface=ether4

add interface=ether5

(Compare with MPLS interfaces)



Traffic engineering (TE) tunnel

Use TE tunnel.

Here:

- No need for OSPF adjustments / single OSPF area
- No need for bandwith reservation / definition
- No need for Constrained Shortest Path First (CSPF)

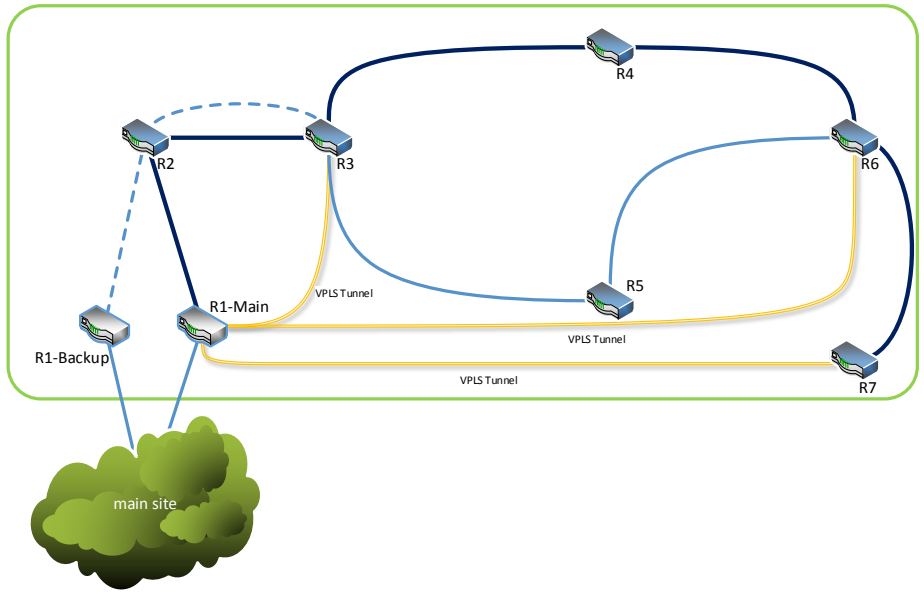
Traffic engineering (TE) tunnel

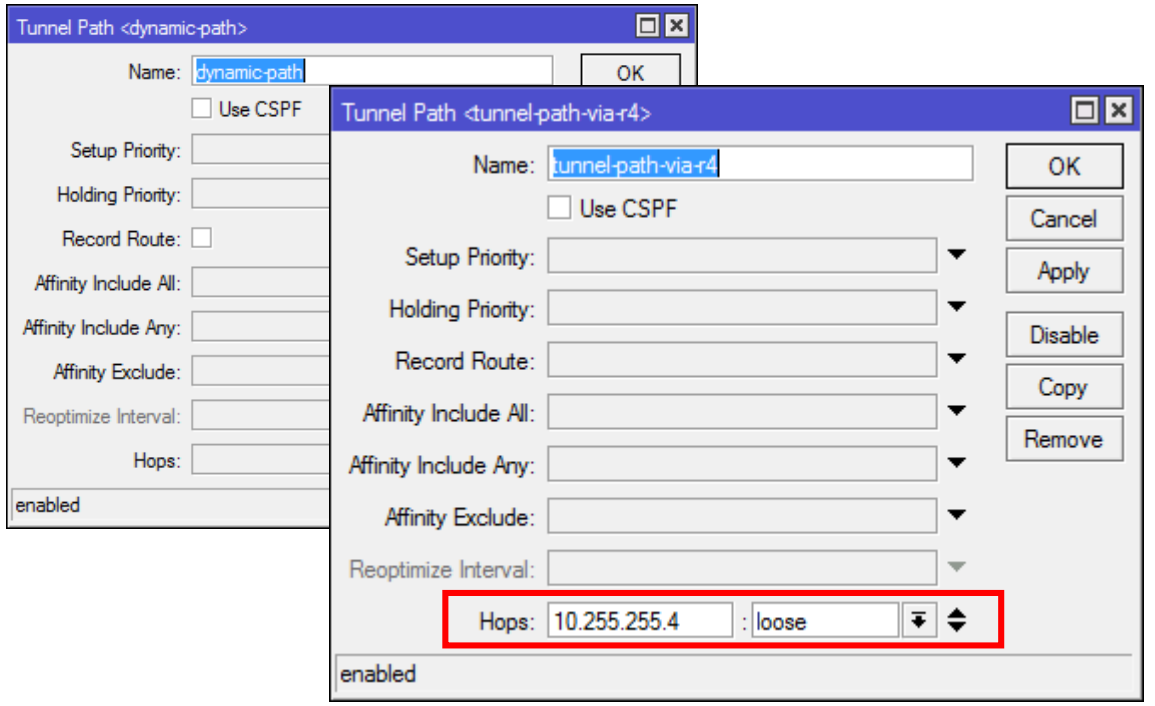
- Configure primary and secondary tunnel path (R1, R7)

/mpls traffic-eng tunnel-path

add name=tunnel-path-via-r4 use-cspf=no hops=10.255.255.4:loose

add name=dynamic-path use-cspf=no

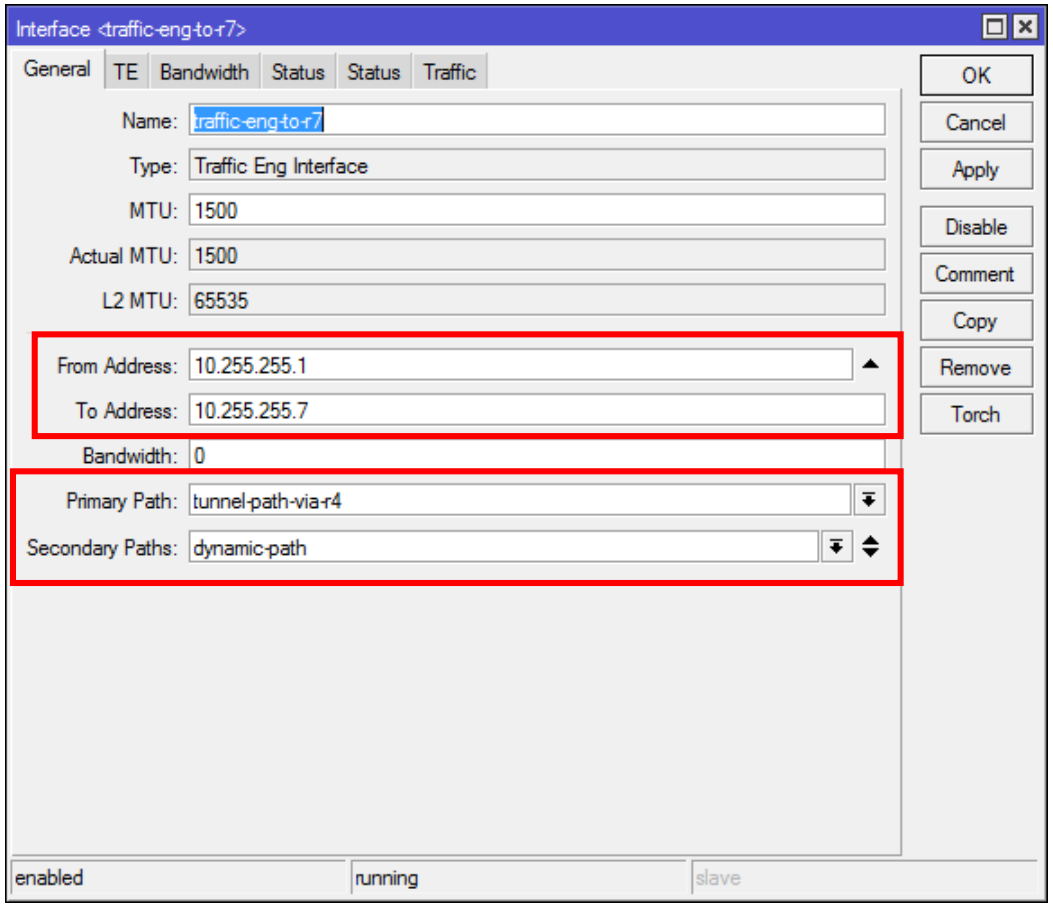




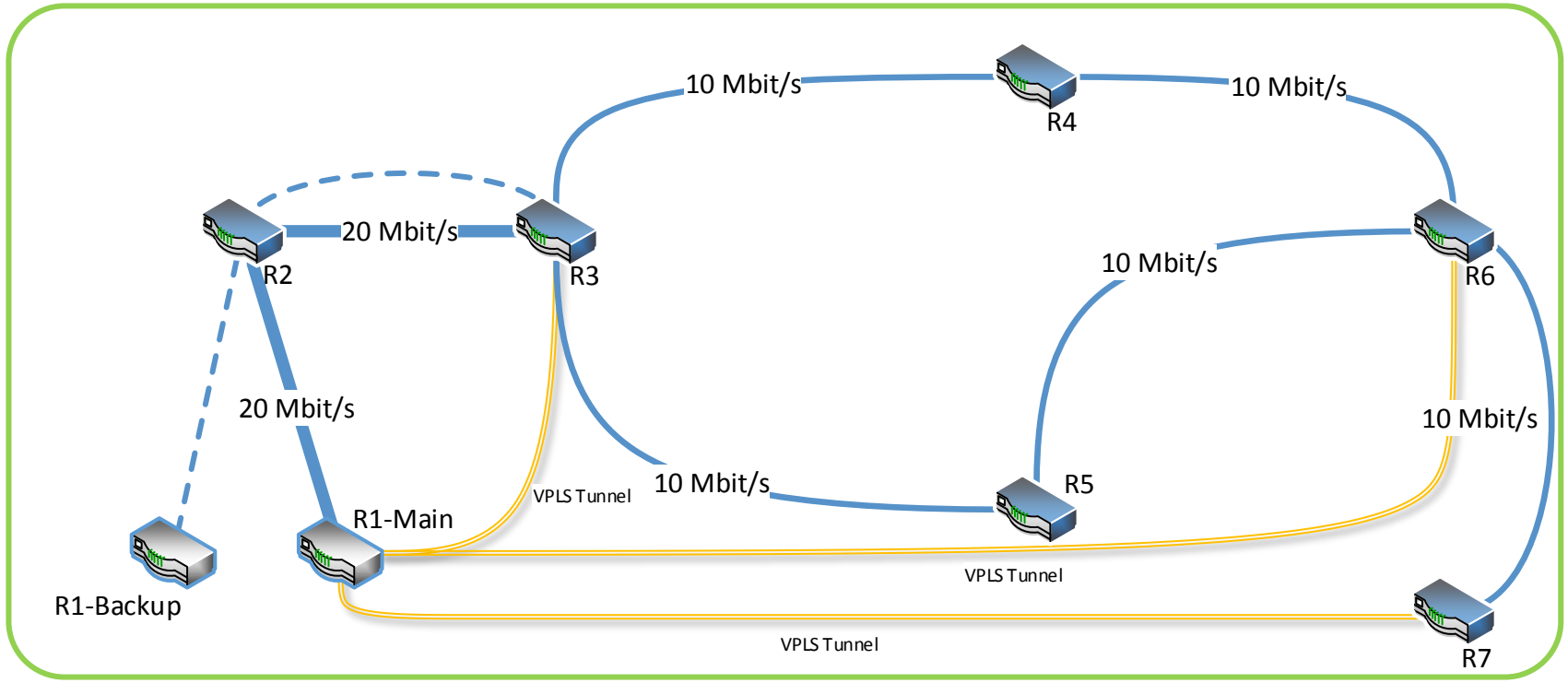
Traffic engineering (TE) tunnel

- Create TE Tunnel (R1, R7)

```
/interface traffic-eng add \
name=traffic-eng-to-r7 \
from-address=10.255.255.1 \
to-address=10.255.255.7 \
primary-path=tunnel-path-via-r4 \
secondary-paths=dynamic-path
```

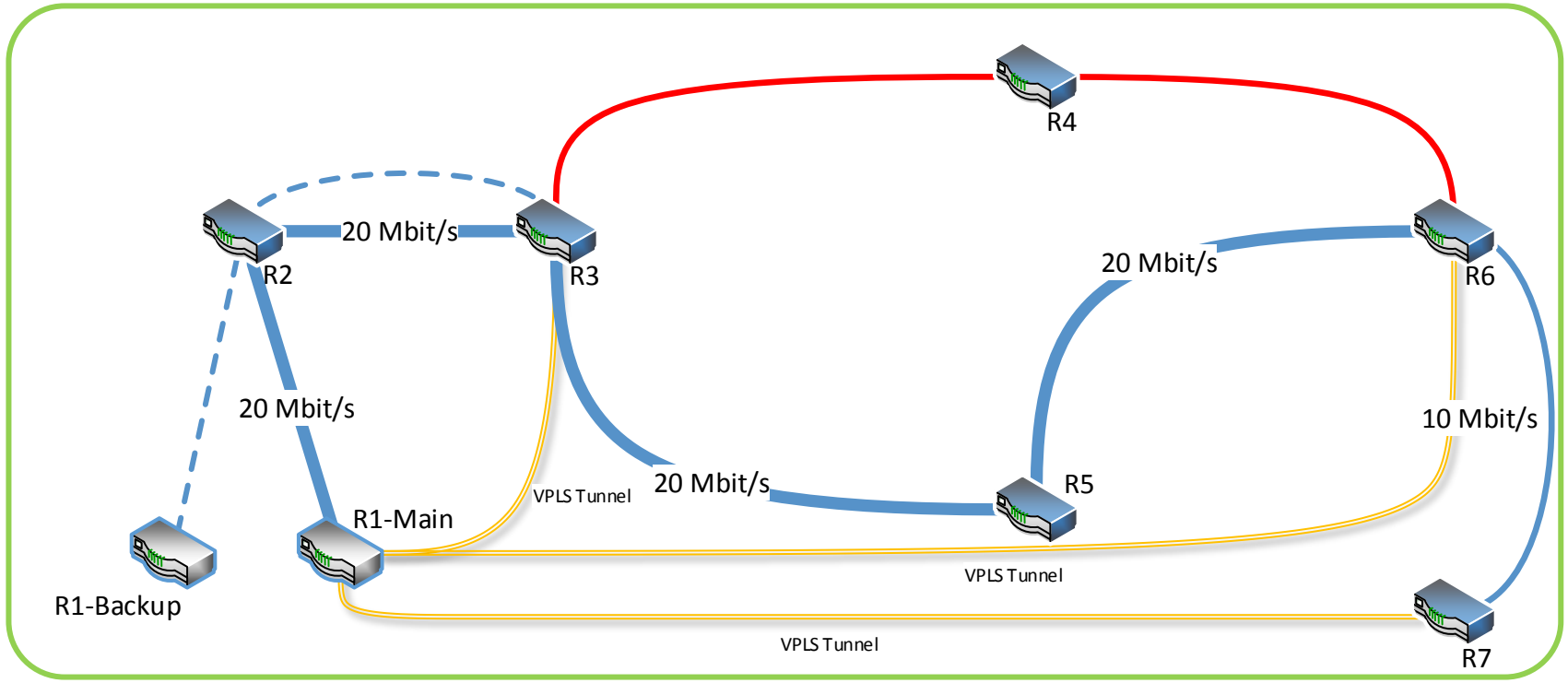


Result



10 Mbit/s to PPPoE client at R6 and R7

Result



10 Mbit/s to PPPoE client at R6 and R7

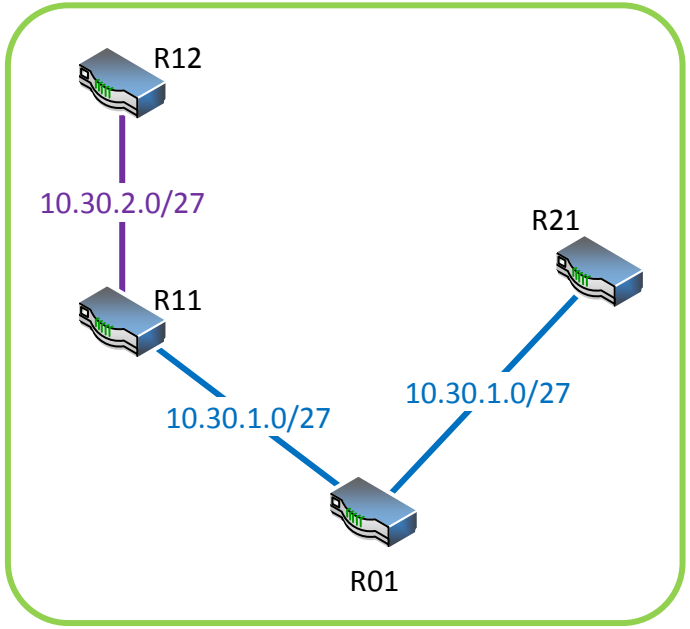
Failure of R4: Traffic through R5 (same for R5)

OSPF issue

OSPF setup (simplified)

R01, R11 and R21 on same subnet

- Bridge on R01
- Same horizon value
- R01 OSPF neigbors: R11, R21



Bridge

Bridge | Ports | VLANs | MSTIs | Port MST Overrides | Filters | NAT | Hosts | MDB

Find

#	Interface	Bridge	Horizon	Priority (h...	Path Cost	Role	Root Path Cost
0	ether5	bridge	5	80	10	designated port	
1	ether4	bridge	5	80	10	designated port	

2 items

OSPF

Virtual Links | Neighbors | NBMA Neighbors | Sham Links | LSA | Routes | AS Border Routers | ...

Find

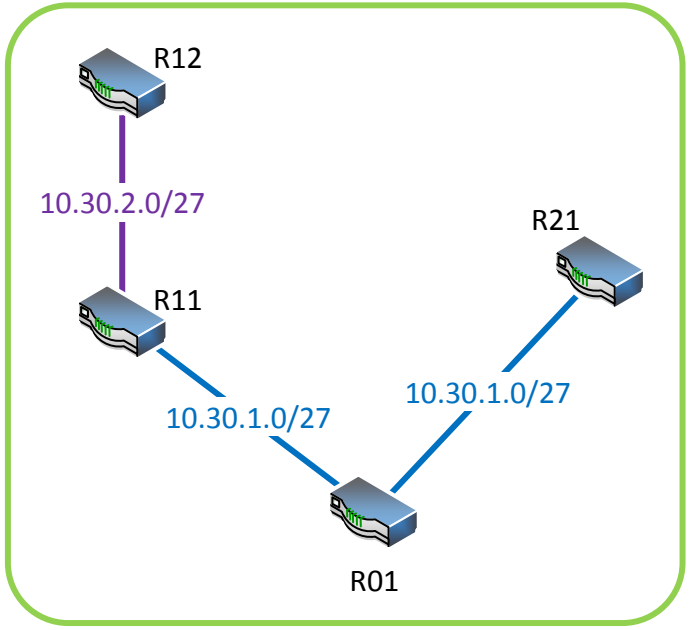
Instance	Router ID	Address	Interface	State Changes	State
default	10.10.10.21	10.30.1.21	bridge	11	Full
default	10.10.10.11	10.30.1.11	bridge	11	Full

2 items

OSPF setup (simplified)

Expected behaviour on R21

- OSPF neighbour (only) R01
- Route to 10.30.2.0/27



Observed behaviour

- As expected

OSPF

Virtual Links | Neighbors | NBMA Neighbors | Sham Links | LSA | Routes | AS Border Routers | Area Border Routers | ...

Find

Instance	Router ID	Address	Interface	DR Address	State Changes	State
default	10.10.10.1	10.30.1.1	ether5	10.30.1.1	5	Full

1 item

Route List

Routes | Nexthops | Rules | VRF

Find all

	Dst. Address	Gateway	Distance
DAC	10.10.10.21	loopback reachable	0
DAC	10.30.1.0/27	ether5 reachable	0
DAo	10.30.2.0/27	10.30.1.11 reachable ether5	110

3 items

OSPF setup (simplified)

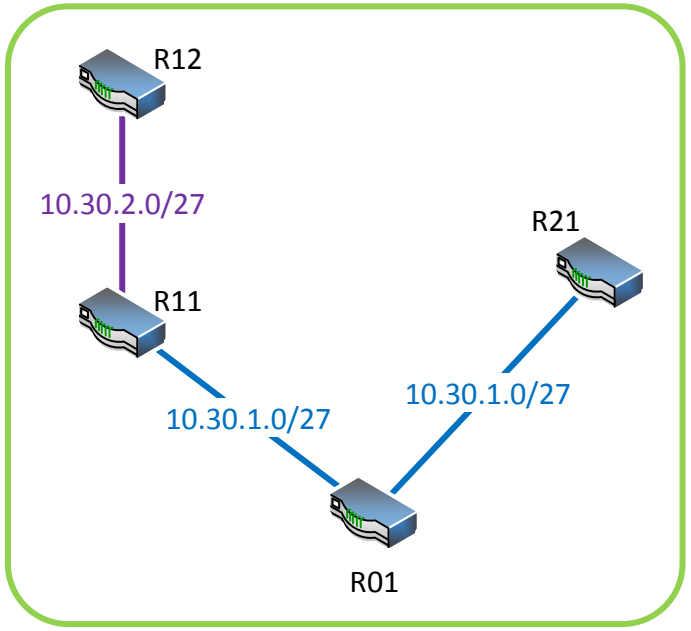
Reboot R01. No config change.

Expected behaviour on R21

- OSPF neighbour (only) R01
- Route to 10.30.2.0/27

Observed behaviour

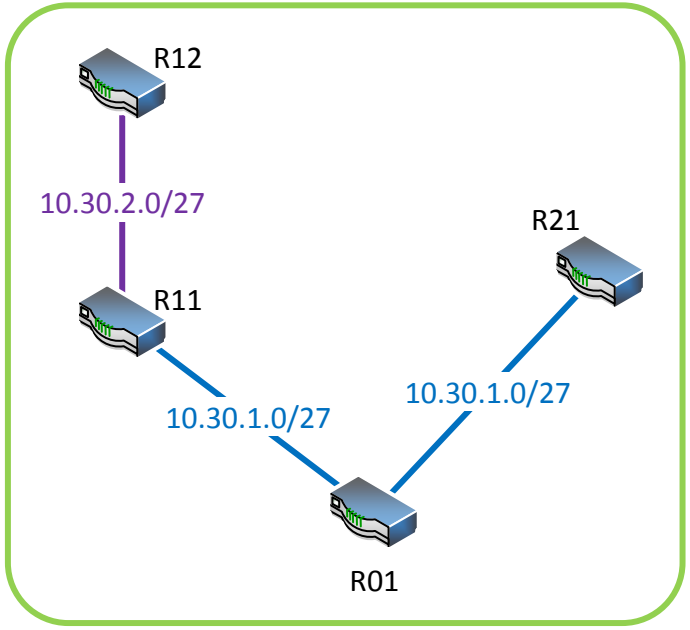
- 10.30.2.0/27 missing



Debug R21

Debug R21

- OSPF state to R01 full
- 10.30.2.0/27 missing



Route List

Routes | Nexthops | Rules | VRF

Find | all

	Dst. Address	Gateway	Distance
DAC	10.10.10.21	loopback reachable	0
DAC	10.30.1.0/27	ether5 reachable	0

2 items

OSPF

Virtual Links | Neighbors | NBMA Neighbors | Sham Links | LSA | Routes | AS Border Routers | Area Border Routers | ...

Find

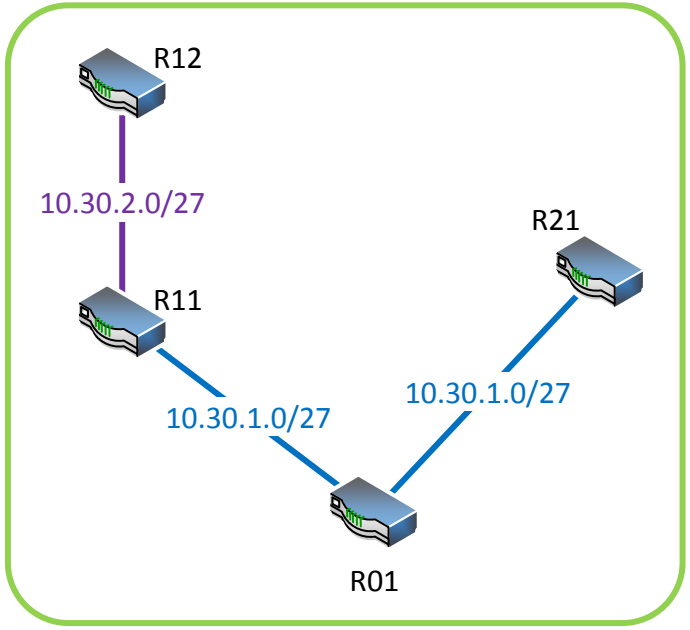
Instance	Router ID	Address	Interface	DR Address	State Changes	State
default	10.10.10.1	10.30.1.1	ether5	10.30.1.21	5	Full

1 item

Debug R01

Debug R01

- OSPF state to R11 & R21 full
- 10.30.2.0/27 missing



Route List

Routes | Nexthops | Rules | VRF

Find | all

	Dst. Address	Gateway	Distance
DAC	10.10.10.1	loopback reachable	0
DAC	10.30.1.0/27	bridge reachable	0

2 items

OSPF

Virtual Links | Neighbors | NBMA Neighbors | Sham Links | LSA | Routes | AS Border Routers | ...

Find

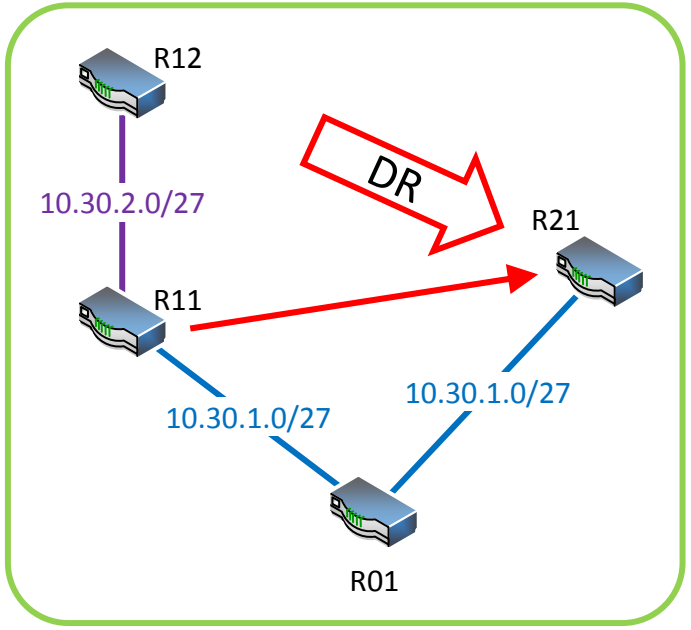
Instance	Router ID	Address	Interface	State Changes	State
default	10.10.10.11	10.30.1.11	bridge	6	Full
default	10.10.10.21	10.30.1.21	bridge	11	Full

2 items

OSPF Designated Router

OSPF with network type Broadcast will elect Designated Router (DR).

Who is DR? R21 is DR!

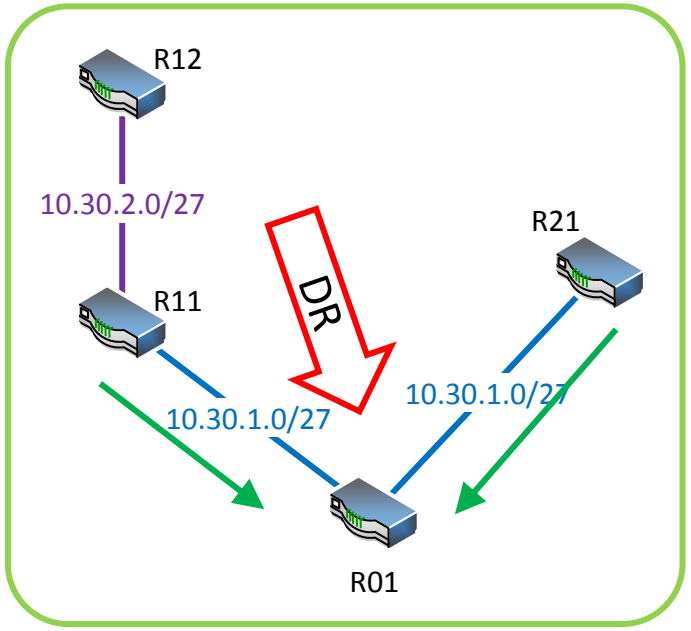


- R11 tries to update R21 - not allowed
  - by bridge horizon
  - or wireless default forward
  - or bridge filter
  - …

Possibilities

Possible solutions

- Force R01 to be DR
- Use network type ptmp



Thank you

FMS Internetservice GmbH

Phone: +49 761 2926500
Web: www.fmsweb.de
Shop: www.mikrotik-shop.de
Email: sales@fmsweb.de
Twitter: https://twitter.com/fmsweb_de