Interoperability (and migration) from IPv4 to IPv6-only Backbone with SR-MPLS

ITNOG 8 - 7/5/2024

Nicola Modena - CCIE #19119 / JNCIE-SP #986
nicola@modena.to - http://tierzero.it

What I will talk about

- This talk is about
  adoption of IPv6 and MPLS in service provider transport infrastructures,
  reflecting my point of view and experience in ipv6-only infrastructures.
  I try to be as technological and less fanatical as possible,
  but certainly, more pragmatic.

About me

- **Nicola Modena** - CCIE #19119 / JNCIE-SP #986 Emeritus
  Independent Network Architect

  More than 25 years experience designing and implementing service provider and large enterprise networks.
  https://tierzero.it | nicola@modena.to

Why use an IPv6 Backbone ?

False myths?

- MPLS is dead
- IPv4 addresses are over
- I want to provide IPv6 services
- There are some (big) advantages using IPv6

ISO/OSI Layers 8,9,10 -> (economic, politics, religion)

DISCLAIMER: I am pro IPv6, this doesn't imply being against IPv4 !

Where do we start from ?

hypothetical customer: "I'm using a dual-stack solution with OSPF and OSPFv3, MPLS with LDP, RSVP for link protection and some traffic engineering … but honestly no one really knows how it works…, BGP with two separate sessions for IPv4 and IPv6 …and 6RD for broadband users …with addresses as /32 in OSPF…"

There are ~~two~~ three options:

forget my presentation and spend all your time managing it

PowerPoint

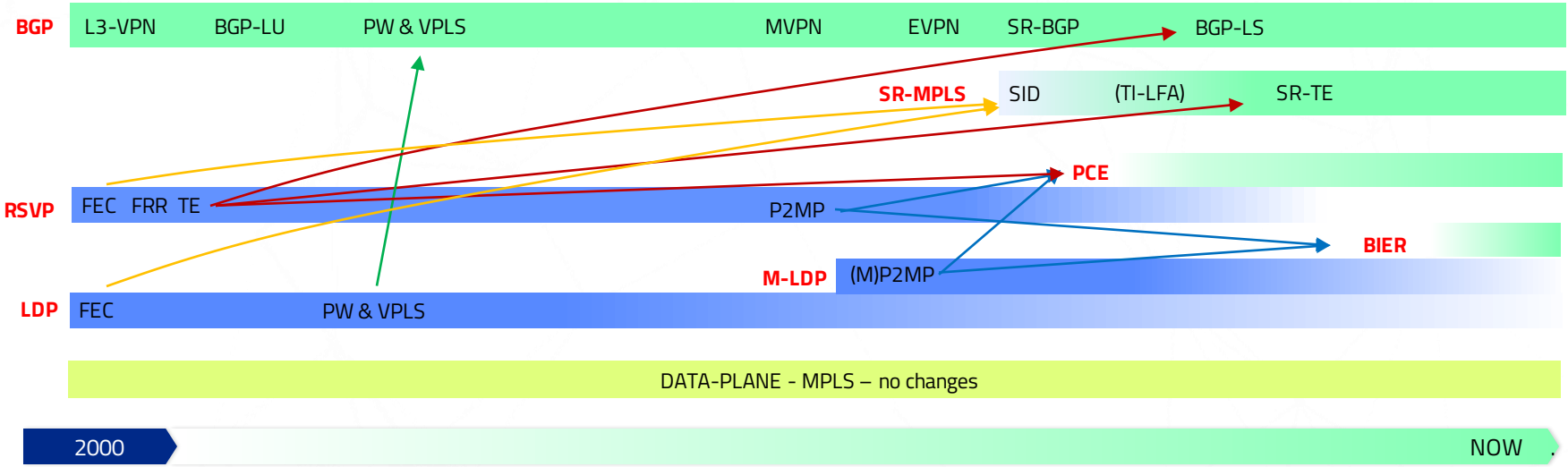
adopts a new SDN controller
…with a new overlay technology
…managed by ~~ML block-chain~~ **AI**
…fully automated …intent based
…from the public cloud …green
…with a fancy license model

simplify your infrastructure, enjoy your free time and be ready for an ipv6-only infrastructure

IGP, MPLS and IP Addressing

Where is technology converging ?

MPLS evolution



- In more than 25 year the data-plane remained almost unchanged
- The control plane has shifted from LDP/RSVP to Segment-Routing (SPRING)
- Traffic Engineering is now signaled with SR-TE or with external controller based on PCE
- All the services signaled with MP-BGP
- Multicast is still signaled with MLDP or PCE, with a slow BIER adoption
- Seamless solutions with BGP-LU allow for great scalability

SR-MPLS - Source Packet Routing - Spring

- **IGP (IS-IS or OSPF) signals Segment-IDs encoded as MPLS labels**
  -> (no more LDP, RSVP, synchronization, distributed states, etc)
- New concept of **Global Labels** – unique and static values in the entire domain
- **BGP-LU** with Global Labels
- Useful for **troubleshooting**, invaluable for **observability**
- **TI-LFA** – Topology Independent Loop Free Alternate for traffic protection
- SR-TE - Traffic Engineering capability without external controller
- Moving states from Network to Packet for high scalability
- Usually fewer labels allocated (pro) but deeper label stacks (cons)

May coexist and interact with exiting signaling protocols like LDP and RSVP (useful for migrations)

SR-MPLS configuration

With IS-IS & SR-MPLS moving from IPv4 to Dual-Stack to IPv6-Only it's just 1 line configuration change:

Single-Stack IPv6

example: lo0.0 **2001:db8::1/128**

```
[edit protocols isis source-packet-routing]
srgb start-label 16000 index-range 8000;
node-segment {
    ipv6-index 601;
}
```

Dual-Stack IPv4 + IPv6

ex: lo0.0 **100.0.0.1/32** & **2001:db8::1/128**

```
[edit protocols isis source-packet-routing]
srgb start-label 16000 index-range 8000;
node-segment {
    ipv4-index 1;
    ipv6-index 601;
}
```

Single-Stack IPv4

example: lo0.0 **100.0.0.1/32**

```
[edit protocols isis source-packet-routing]
srgb start-label 16000 index-range 8000;
node-segment {
    ipv4-index 1;
}
```

resulting label: SRGB start + ipv4/6-Index

```
nmodena@PTX-04> show isis database detail
…
MX-01.00-00 Sequence: 0x9, Checksum: 0x5a43, Lifetime: 755 secs
  IPV4 Index: 1, IPV6 Index: 601
  Node Segment Blocks Advertised:
    Start Index : 0, Size : 8000, Label-Range: [ 16000, 23999 ]

nmodena@PTX-04> show route 100.0.0.1
…
100.0.0.1/32       *[L-ISIS/14] 00:04:35, metric 20
                    >  to 100.0.0.3 via et-0/0/0.0, Push 16001

nmodena@PTX-03> show route 2001:db8::1
…
2001:db8::1/128    *[L-ISIS/14] 00:00:37, metric 20
                    >  to fe80::5200:ff:fe06:4 via et-0/0/3.0, Push 16601
```

JUNOS

IS-IS takes care of all:

Interface configuration

Use IPv4 unnumbered and IPv6 link-local on backbone intf.

Dual-Stack (and MPLS)

```
interface HundredGigE 0/0/0/0
 description "---- Core link ----"
 mtu 9000
 ipv4 point-to-point
 ipv4 unnumbered Loopback0
 ipv6 enable
!
```

Single-Stack IPv6 (and MPLS)

```
interface HundredGigE 0/0/0/0
 description "---- Core link ----"
 mtu 9000
 ipv4 forwarding
 ipv6 enable
!
```

Required for IPv4 PHP

PRO

- Save (private) IPv4 addresses and simplify configurations and provisioning

CONS

- Neighbor are not reachable without IGP adjacency
  - -> use SSH and PING over IPv6 link local address, and an out-of-band management network
  - -> traceroute works just fine
- use Adjacent-SID for Strict SR-TE Policy (abstract from address-family)

Dual-Stack vs Single-Stack

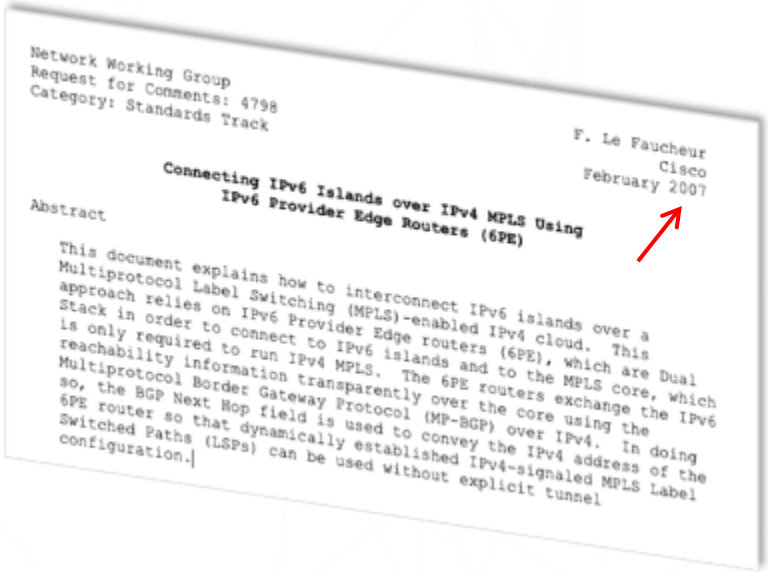
or single-stack from IPv4-Only to IPv6-Only

Single Stack (IPv4 or IPv6)

- Simpler
- Uses less resources
- Uniform and consistent
- Easier to manage and secure
- **Must be able to provide any type of services (IPv4 and IPv6)**

Can we do everything just with IPv4 and MPLS ?

Single Stack IPv4/MPLS



Network Working Group
Request for Comments: 4798
Category: Standards Track

F. Le Faucheur
Cisco
February 2007

Connecting IPv6 Islands over IPv4 MPLS Using IPv6 Provider Edge Routers (6PE)

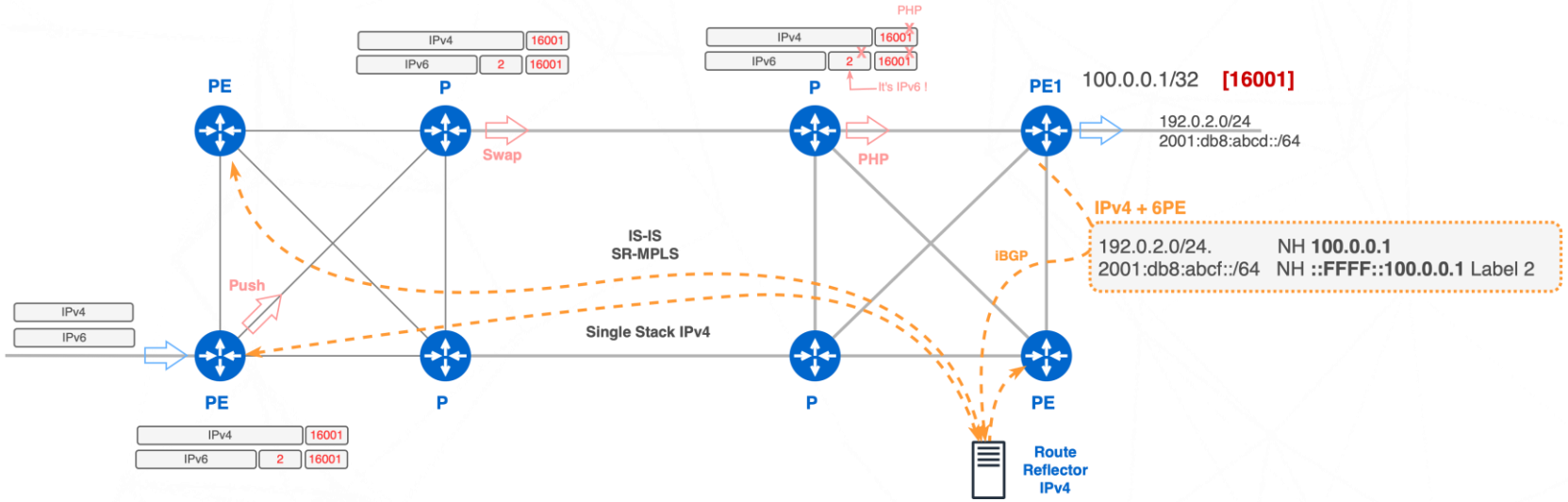
Abstract

This document explains how to interconnect IPv6 islands over a Multiprotocol Label Switching (MPLS)-enabled IPv4 cloud. This approach relies on IPv6 Provider Edge routers (6PE), which are Dual Stack in order to connect to IPv6 islands and to the MPLS core, which is only required to run IPv4 MPLS. The 6PE routers exchange the IPv6 reachability information transparently over the core using the Multiprotocol Border Gateway Protocol (MP-BGP) over IPv4. In doing so, the BGP Next Hop field is used to convey the IPv4 address of the 6PE router so that dynamically established IPv4-signaled MPLS Label Switched Paths (LSPs) can be used without explicit tunnel configuration.

- IPv4 + MPLS (with SR-MPLS or whatever)
- MP-BGP only over IPv4 sessions
- All the usuals MPLS services (L2,L3VPN,EVPN, FlowSpec, Multicast, etc)
- IPv6 services with **6PE & 6VPE** using BGP-LU (Labeled Unicast) - **RFC 4798**

Widely adopted production-proofed

6PE - Connecting IPv6 islands over IPv4 MPLS



Problem nr. 1

- BGP IPv6 NLRI must have an IPv6 Next-Hop
- BGP is over IPv4 sessions, Next-Hop is IPv4

Solution:

- NH as IPv4 mapped IPv6 address RFC 4291 - 2.5.5.2

Problem nr. 2

- Allocation of a dedicated label for each IPv6 prefix (historical)
- High resource usage when using many IPv6 prefix

Solution:

- Use IPv6 explicit null label (value 2) with IPv6 Labeled Unicast

Single Stack IPv6 + MPLS (updated 05/2024)

- IPv6 + MPLS with IS-IS and SR-MPLS (anyone investing in OSPFv3 and LDPv6 ?)
- MP-BGP only over IPv6 sessions
- All the usuals MPLS services (L2,L3VPN,EVPN, FlowSpec, Multicast, etc)
- IPv4 services with **something like "4PE & 4VPE"** **(-> do they exist ?)**

-> What does it work with BGP IPv6 signaling and IPv4 services ?

Please, it's 2024!

-> Can you provide IPv4 services with just MAP-T/E, 464XLAT, DS-Lite, ~~Layer-2~~ ?

-> Do you have Enterprise customers with IPv4 ?

Connecting IPv4 Islands over IPv6 Core using IPv4 Provider Edge Routers 4PE

RFC 8950 – IPv4 NLRI with IPv6 Next-Hop

- Advertise IPv4 NLRI with with IPv6 Next-Hop
- Requires to explicitly set IPv6 next-hop with a policy

Most vendors start to support it

draft-mishra-idr-v4-islands-v6-core-4pe-06

- Use IPv4 AFI/SAFI Labeled or Unlabeled or both
- Impose IPv6 next-hop as RFC 8950
- May impose IPv4 explicit/implicit null label or even an arbitrary topmost label

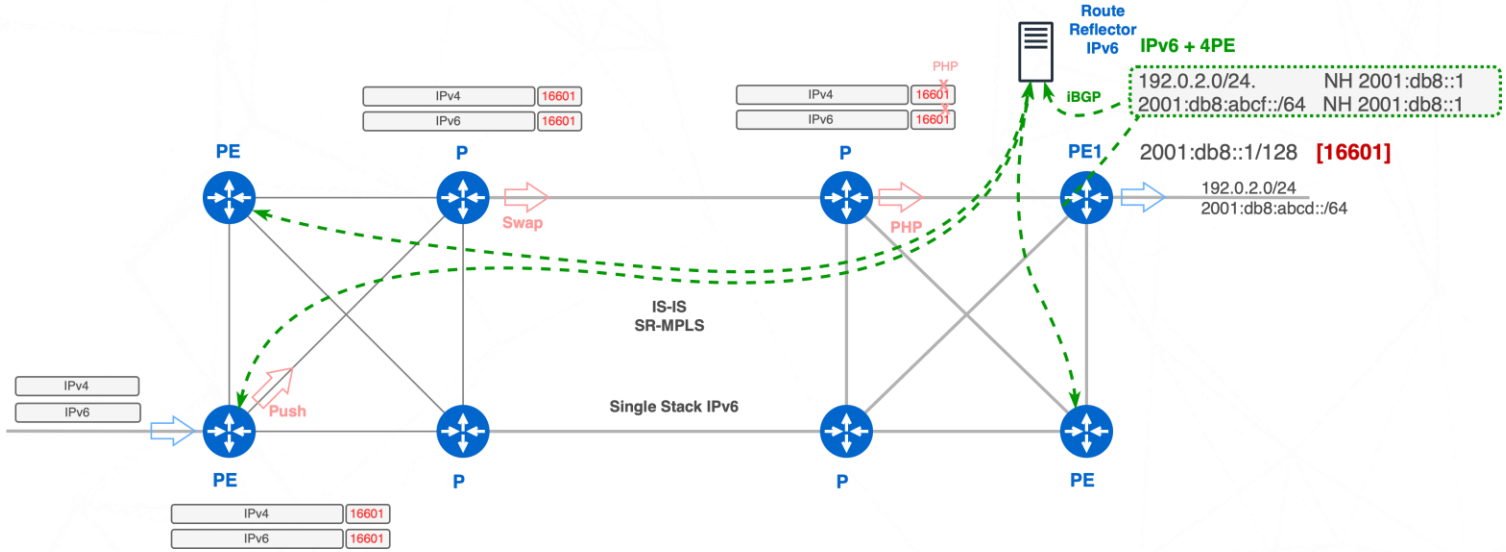
Still each vendor has its own options and defaults

In my interoperability tests the more compatible and efficient combination is:

- BGP IPv4 Address Family (without label) and RFC 8950
- MPLS encapsulation without Explicit/Implicit NULL, just IPv6-node index label (*)

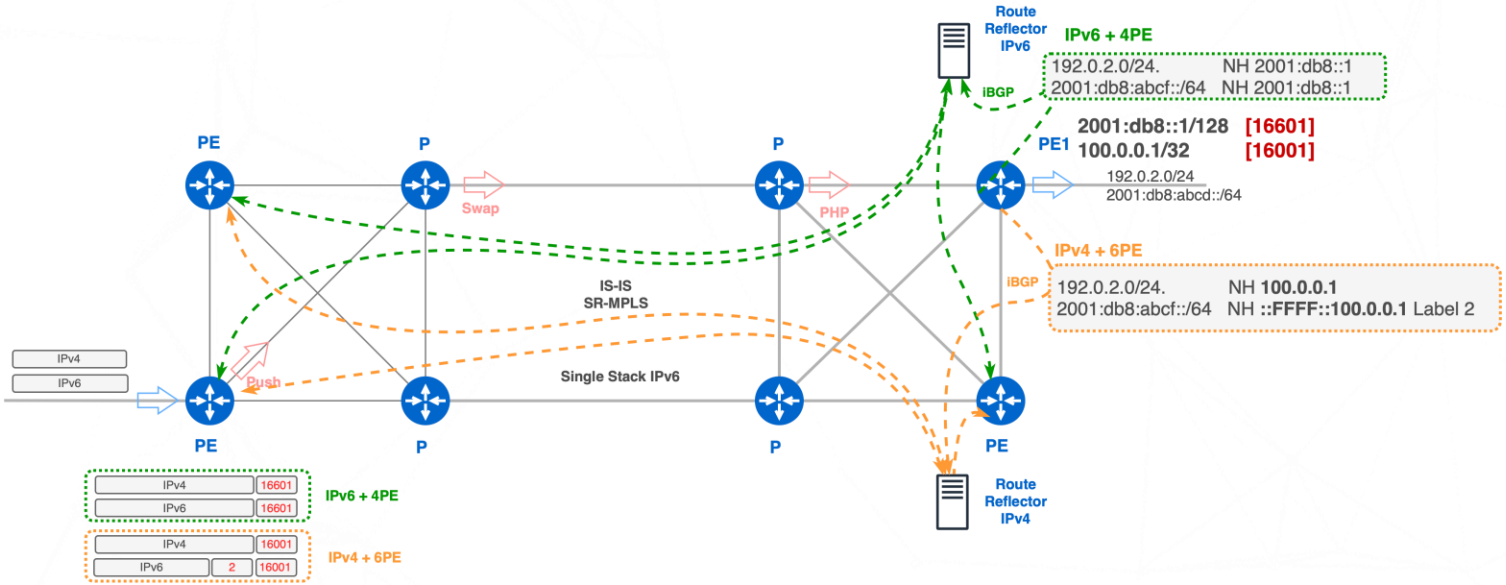
(*) this requires the capacity to recognize packet encapsulation (IPv4) when performing PHP

Connecting IPv4 Islands over IPv6 Core using IPv4 Provider Edge Routers 4PE



- Route Reflectors are IPv6 Only
- Both IPv6 and IPv4 prefixes are advertised with an IPv6 next-hop
- MPLS forwarding using MPLS label corresponding to IPv6 next-hop (from IS-IS and SR-MPLS)
- PHP router must identify IPv4 or IPv6 encapsulation

IPv4-IPv6 CP interoperability & migration



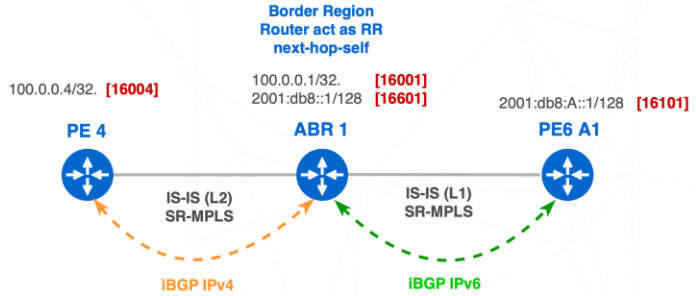
- **IPv4 and IPv6 BGP control-plane may coexist on the same backbone for interoperability and migration**
- Next-hop (label) will identify LSP availability of destination PE (IPv4/dual-stack/IPv6 from IS-IS & SR-MPLS)
- Test and migrate single service by just changing route preference
- P-routers are completely transparent to service migration

Advanced topics

It seems perfect but we are still in PowerPoint, where is the catch?

EVPN Services

```
routing-instances {
    …
        protocols {
            evpn {
                encapsulation mpls | mpls-inet6;
    …
}}}}
```




```
nmodena@MX-01> show route protocol evpn table CUST-B detail | match "^3|INGRESS"
3:100.0.0.1:4::10::2001:db8::1/248 IM (1 entry, 1 announced)
                PMSI: Flags 0x20: Label 18: Type INGRESS-REPLICATION 2001:db8::1
```

JUNOS

- EVPN configuration requires to explicitly define AFI in configuration because:
- Type-3 and Type-4 NLRIs encode PE loopback address in the signaling
- A simple NH rewrite does not ensure interoperability
- Border GW must recreate the advertisement and manage BUM data-plane interoperability
- Currently some vendor start to support Type-5 translation between IPv4 and IPv6 Control-Plane

Conclusion

Let's try to be pratical and constructive

Uses cases

Large SP Backbone with simple and predefined services

- comes from experience of managing IPv6 only access
- private IPv4 address are over

Large SP Mergers with overlapping private address space

- Limited number of services
- Seamless & Border Gateway

Where Layer 8,9,10 are predominant

- Greenfield with few devices, few services, selected vendor
- Basic IPv4 services
- But usually, this combination leads to other technology

My Advice

Start seriously and consciously to embrace your ipv6 adoption

Be realistic:

- What are you really using ?
- What it's working in your backbone?
- What do you really need ?

Forget "we've always done it this way" approach, technology evolves!

And, even more important:

"Start from requirements not from technology!"

Simplify your backbone:

- Single-Stack (Ipv4 and Ipv6 in the future)
- Consider a migration to IS-IS and SR-MPLS
- Shrink your IGP – use it just for loopbacks -
- Use BGP for everything
- Use seamless instead multi-area/multi-level (but only if you are big enough)

Backup Slides

Some further insights

MPLS Encapsulation

Its power comes from simplicity

* Lookup only on the ingress and forwarding trough an LSP
* it's a 32 **(20) bit index** – the simpler and more efficient encoding format
* **Dedicated ethertype**, enabled only on core facing interfaces
* **Just 4 byte x label** vs IPv4 (20+[0-40]) GRE (20+[0-16]) VXLAN (20+8+8) IPv6 (40 + [n * EH])
* Perfect for creating efficient **hierarchical solution with label stacks**
* Few special purpose label – **almost unchanged in 25+ years**

IS-IS & Multi-Topology

IS-IS does not relay on IP and can route ipv4 and/or ipv6

It can operate in two modes:

IS-IS Single-Topology (IPv4/IPv6)

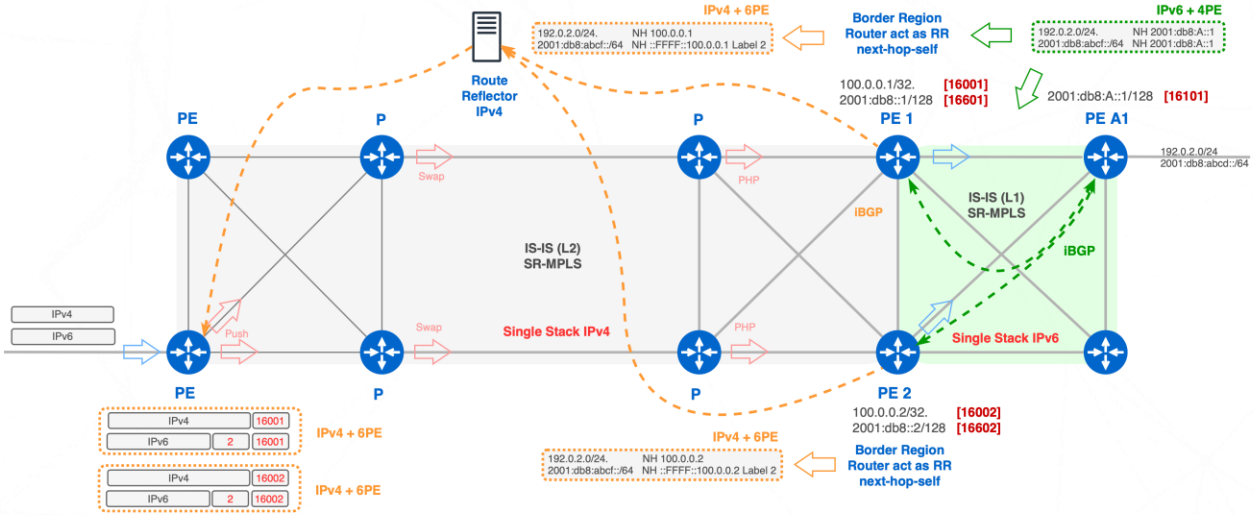
- Requires both IPv4 and IPv6 active on interfaces (with exceptions)
- Run a single SPF for both address-family
- Typically used in DUAL-STACK scenario

IS-IS Multi-Topology (default)

- each address-family has its own topology and SPF
- should be used for single-stack solutions

Check for command, guideline and best practice for a safe migration

MPLS Seamless Architecture



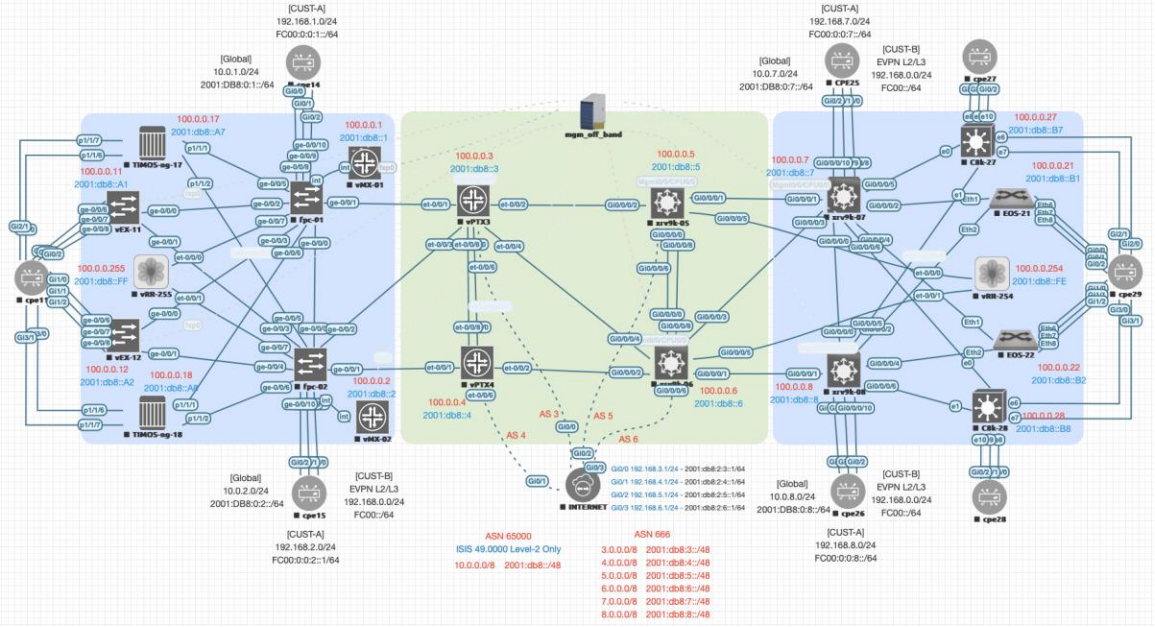
- Single Stack IPv4-only backbone (with 6PE)
- Single Stack IPv6-only region (example: greenfield extension)
- Border Router act as Route Reflectors imposing IPv4 or IPv6 next-hop
- IPv4 and IPv6 prefix does not allocate labels as Border Router perform IP lookup (draft 4PE)

What about SRv6 ?

- SRv6 propose to replace MPLS dataplane with a native IPv6 Encapsulation
- Currently two (incompatible) encapsulation revision: SID -> uSID
- And an upcoming new encapsulation SRm6 that try to “solve intrinsic security and efficiency problems”
- Security problems that requires to enforce ACL on all untrusted interfaces (by design)
- Efficiency problems that requires HW upgrade even between SID and uSID
- Currently no advantages regards MPLS in backbone infrastructure
- Still no multicast support
- Differences in vision between vendors
- No or very limited interoperability with MPLS and existing infrastructure

In any case, all developments at the IETF level to implement IPv6 control-plane always consider MPLS and SRv6 encapsulation, if in a few years some advantage emerges in changing encapsulation, the control-plane will still be IS-IS & MP-BGP and you will be ready.

My Lab topology



I use: EVE-NG, Containerlab and NET-LAB.

Devices: Juniper PTX,MX,EX,vRR – Cisco XRv9k, Cat 8K – Arista vEOS – Nokia VSR NG

I share some topology and configuration on https://github.com/nmodena, or just drop me an email

Any questions ?

you can find me:

nicola@modena.to

linkedin.com/in/nmodena

Telegram: it-nog & ipv6-italia channels

This presentation (and future updates) at https://github.com/nmodena/blog

Special thank to Ivan Pepelnjak and Massimo Magnani for the revision