SRv6

Marian Rychtecký

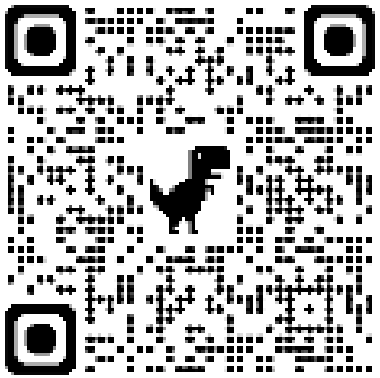
6. 6. 2025 - NTK

NIX·CZ

Poděkování

- Jakub Horn Cisco Systems
- V této prezentaci jsou použity vizualizace z Jakubovy skvělé prezentace.

Stažení možné zde



NIX·CZ

Doprava dat do cíle

- směrování dle cílové IP - Destination Address (DA)
  - přímé
  - tunelované (GRE, Geneve, VxLAN…)

- směrování dle štítků – Label (MPLS)

Destination Address

- Výhody
  - jednoduché (intuitivní)

- Nevýhody
  - žádný „TE“
  - neškáluje

MPLS

- Výhody
  - škálovatelnost / flexibilita
  - kompatibilita
  - bezpečnost
  - podpora více služeb

- Nevýhody
  - komplexita
  - omezená podpora pro cloudové služby
  - náročnost na zdroje
  - odhalování problémů

Segment routing

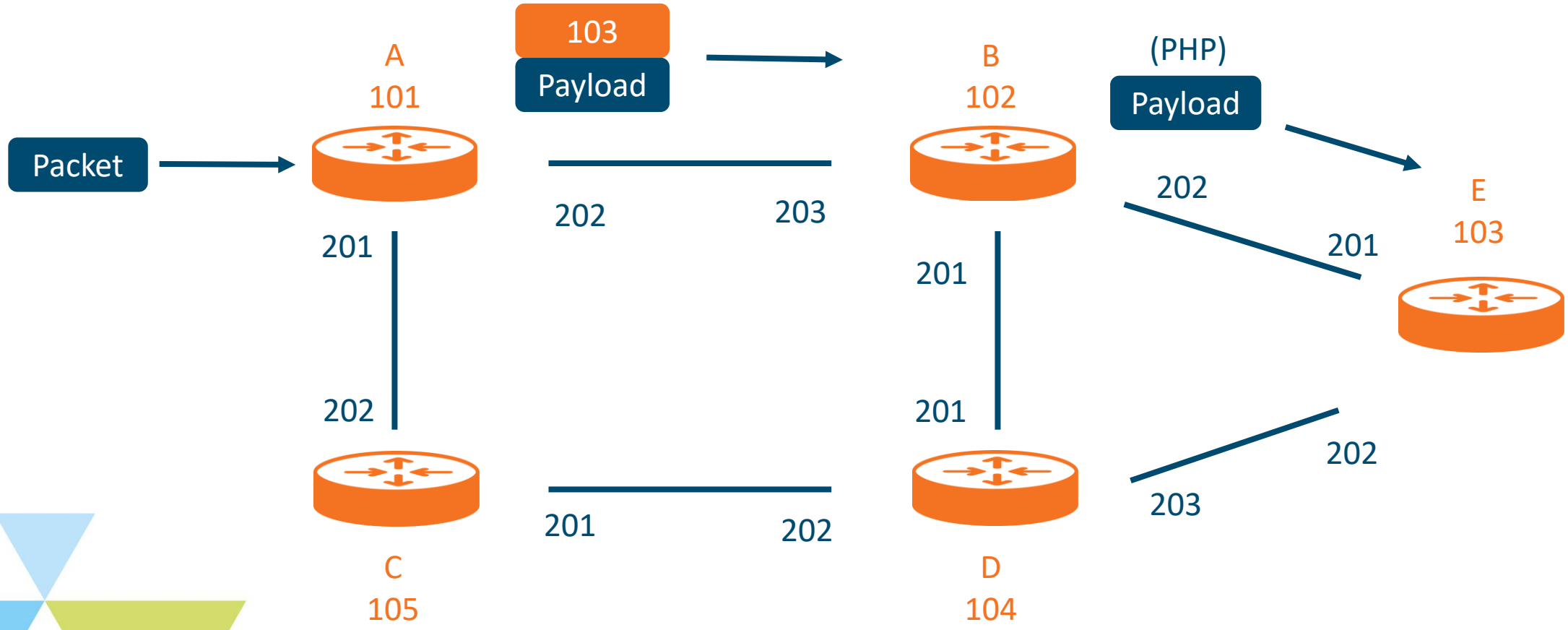
- cestu sítí určuje zdrojový nod – „source routing“
- cesta sítí obsažena v každém paketu
  - na ingress node se provede lookup IGP a vybere nejvýhodnější cesta skrz „segmenty“
  - každý segment má unikátní identifikátor (Segment ID - SID)
  - provoz v rámci segmentu je ve výchozím nastavení směrován nejkratší cestou

- kompatibilní s „MPLS label stack“
  - jednoduchá migrace

Segment routing – dvě instance

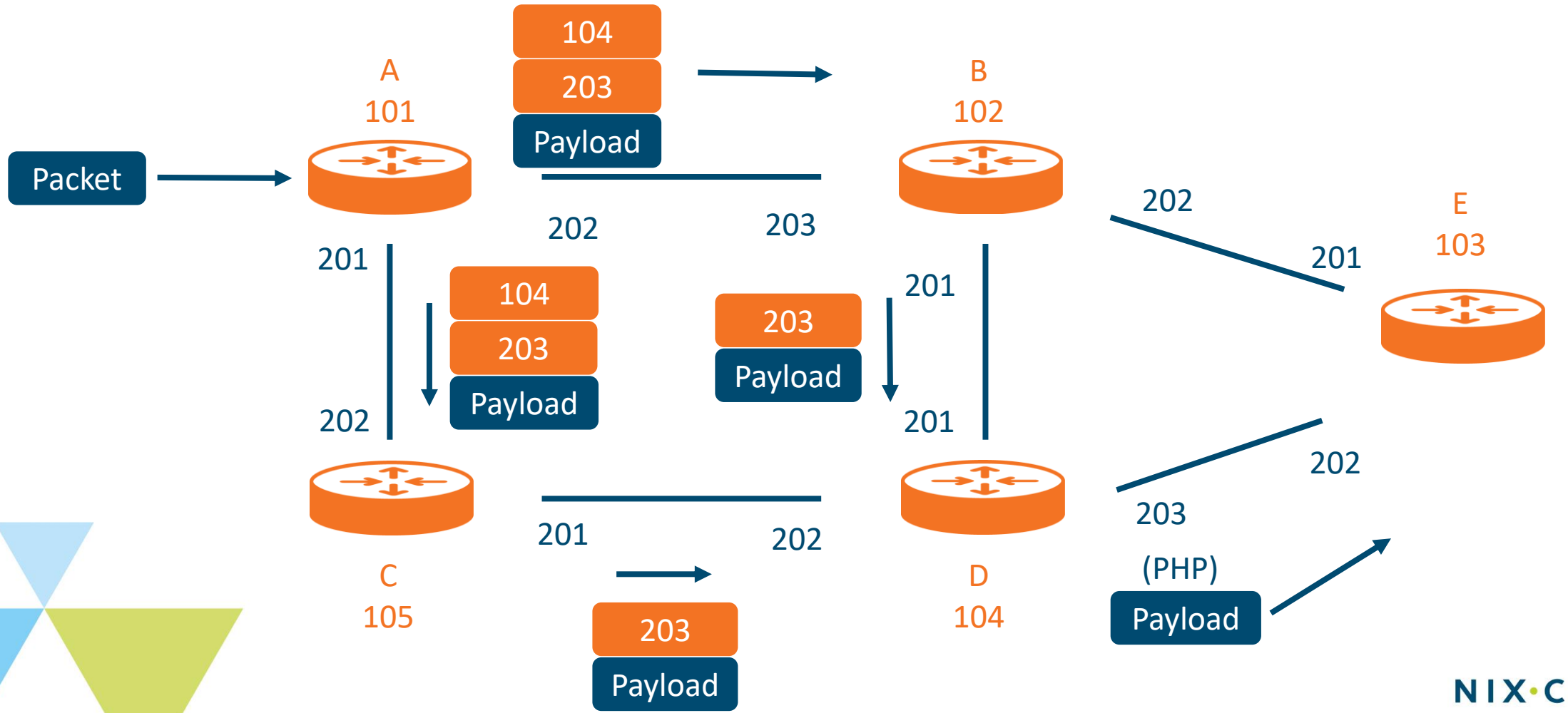
- SR-MPLS
  - segment je definován v MPLS „label“
- SRv6
  - data plane využívá IPv6
  - segment je kódovaný v IPv6 adrese

Segment routing – SR - MPLS



NIX·CZ

Segment routing - TE



NIX·CZ

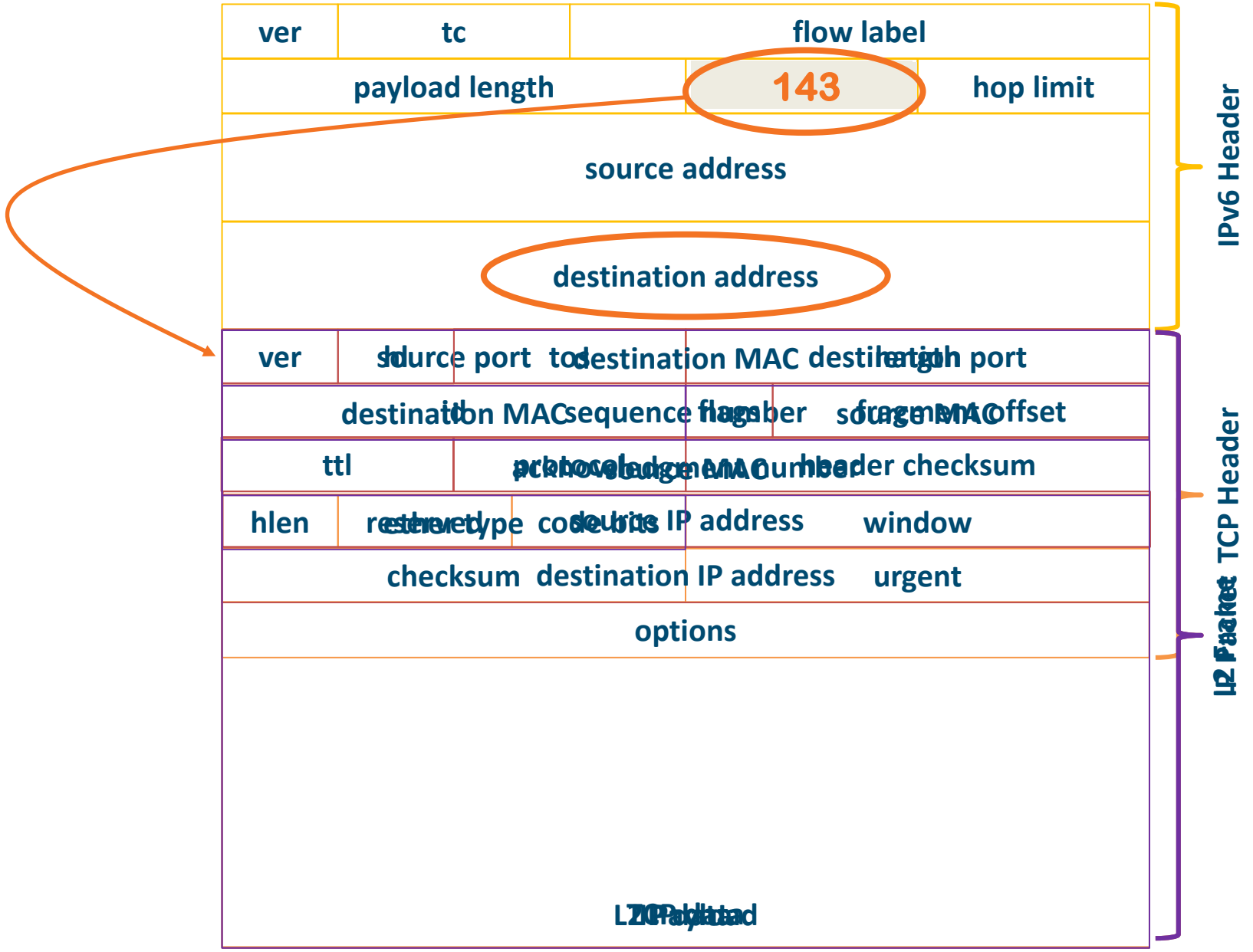
Segment routing

?

NIX·CZ

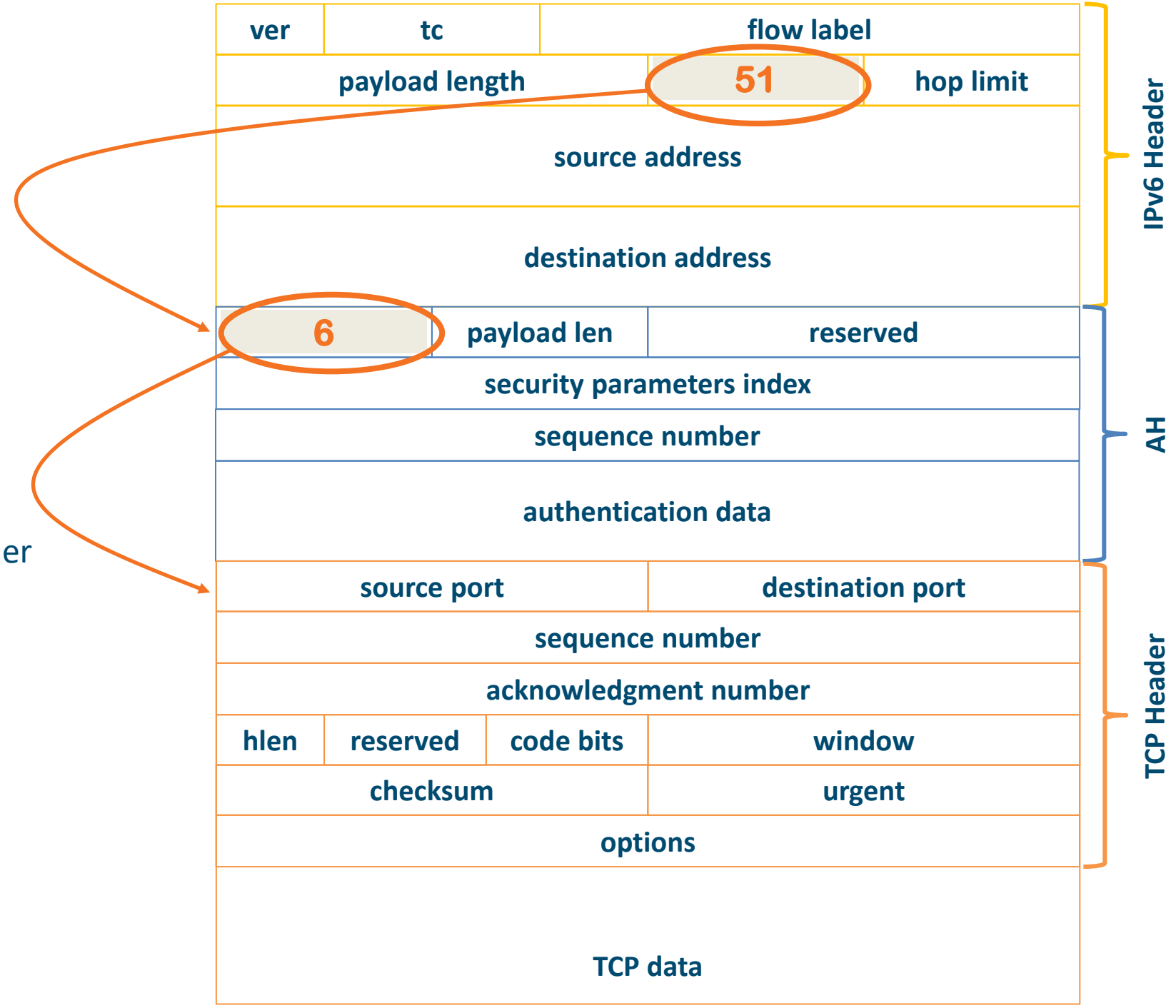
SRv6

- IPv6 Header
- Destination IP address
- Next header field:
  - TCP, UDP, ICMP....
  - IPv4, IPv6, L2



SRv6

- IPv6 Header
- Destination IP address
- Next header field:
  - TCP, UDP, ICMP....
  - IPv4, IPv6, L2
  - Hop by Hop, Dest. Options, Fragmentation, Authentication Header …



SRv6

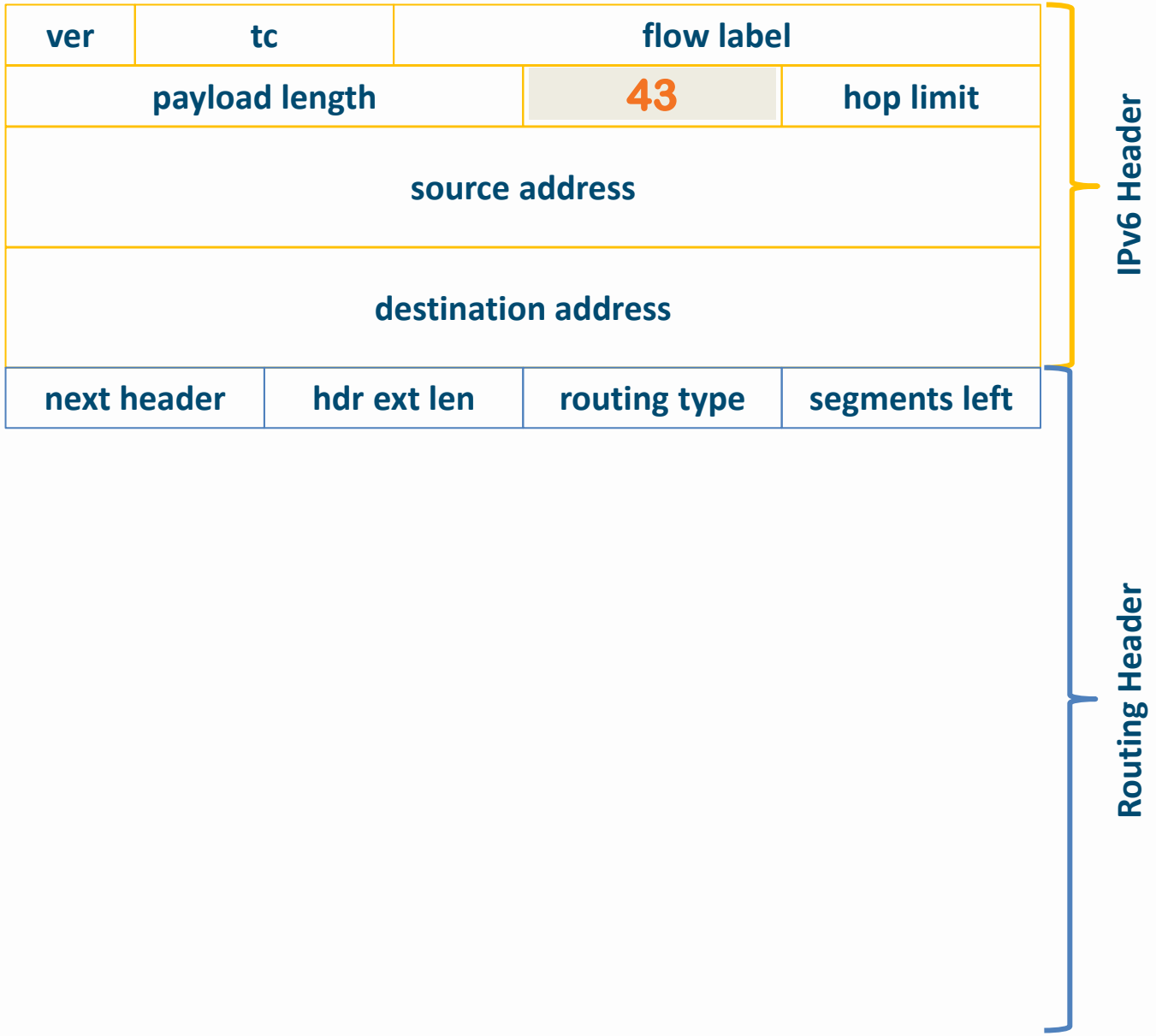
- IPv6 Header
- Destination IP address
- Next header field:
  - TCP, UDP, ICMP....
  - IPv4, IPv6, L2
  - Hop by Hop, Dest. Options, Fragmentation, Authentication Header …
  - **43 - IPv6 Routing Header**

    0 Source Route (deprecated)

    1 Nimrod (deprecated)

    2 Type 2 (RFC 6275)

    3 RPL (RFC 6554)



SRv6

- IPv6 Header
- Destination IP address
- Next header field:
  - TCP, UDP, ICMP....
  - IPv4, IPv6, L2
  - Hop by Hop, Dest. Options, Fragmentation, Authentication Header …
  - **43 - Routing Header**

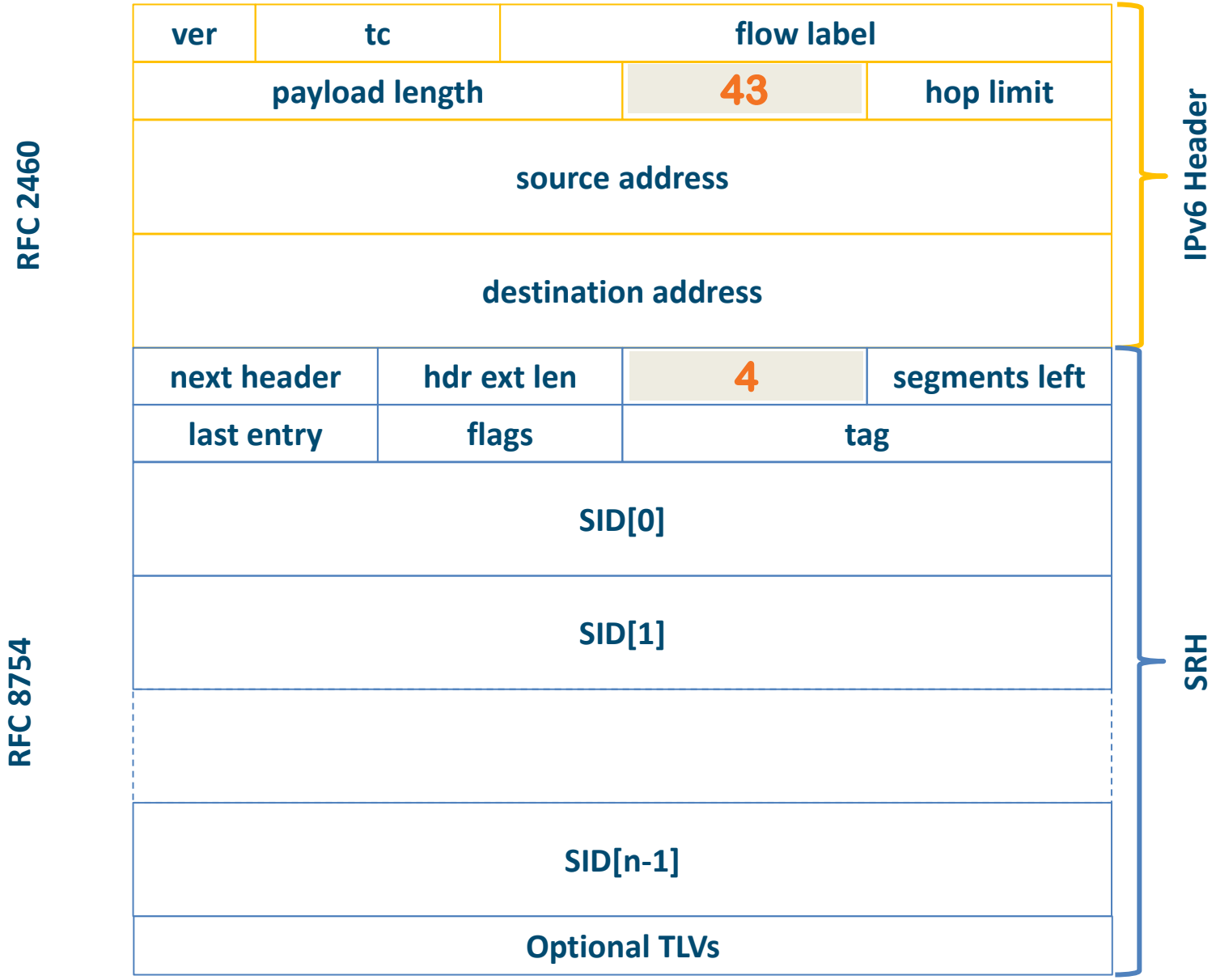
    0 Source Route (deprecated)

    1 Nimrod (deprecated)

    2 Type 2 (RFC 6275)

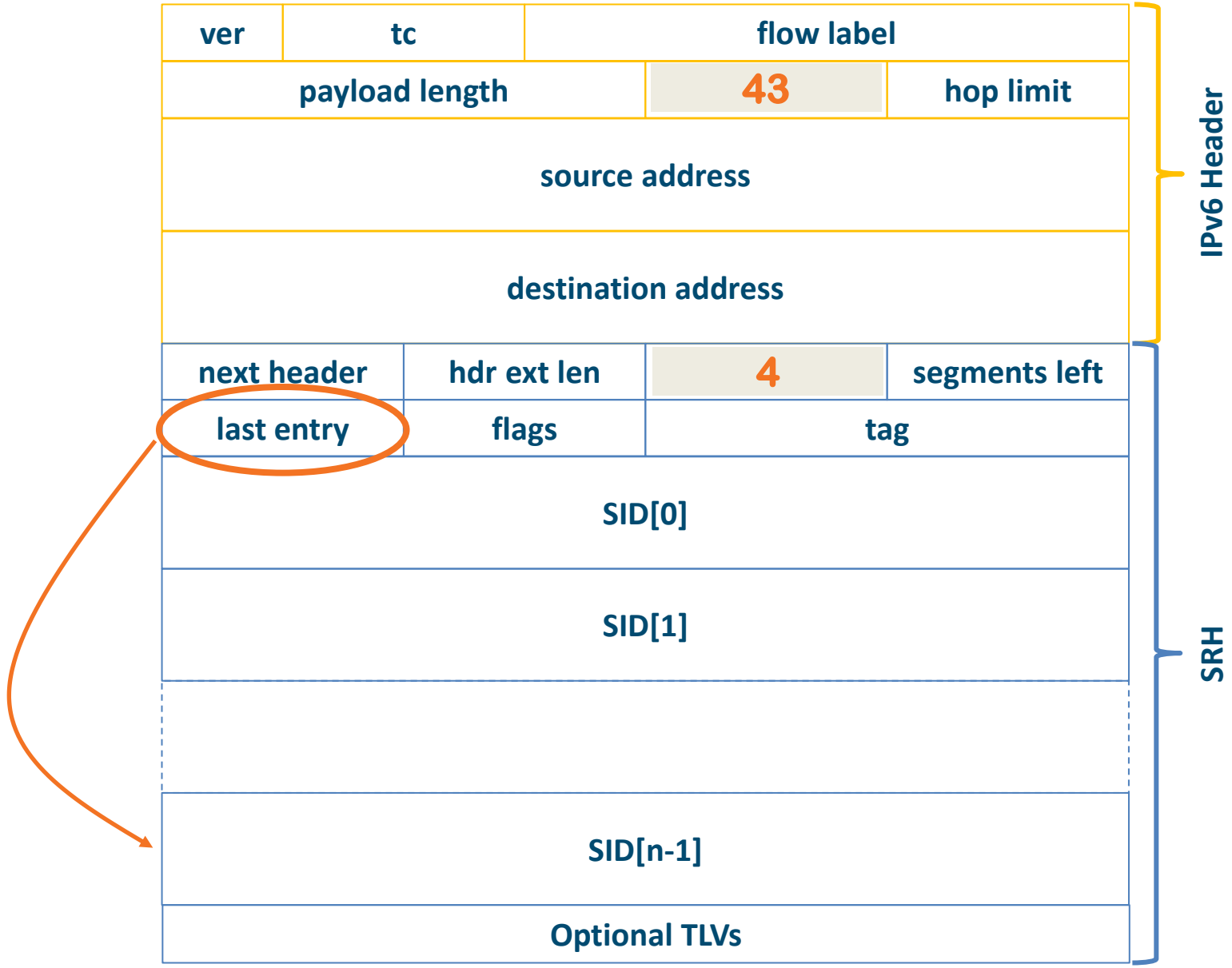
    3 RPL (RFC 6554)

    4 SRH (RFC 8754)



SRH

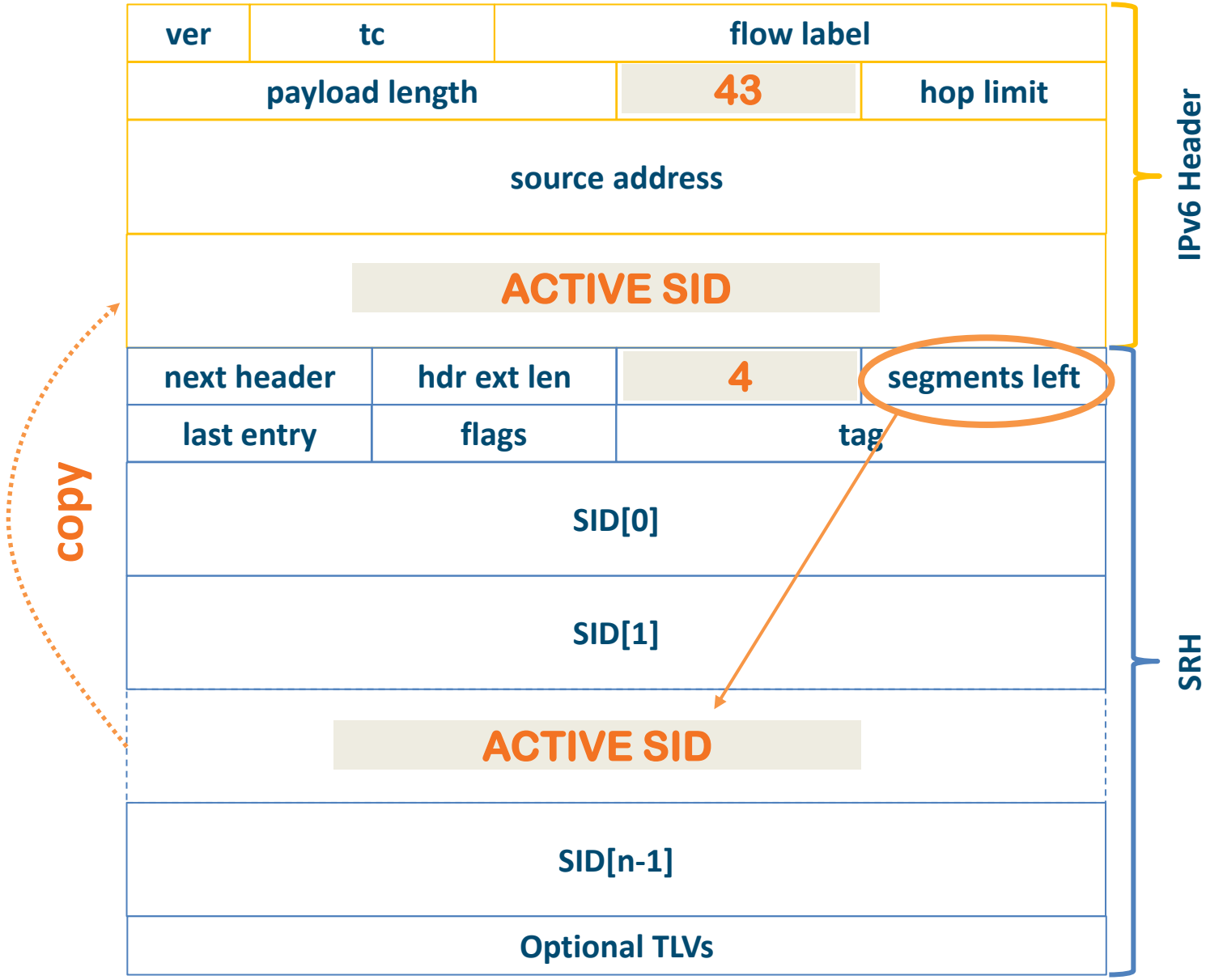
- Segment Routing Header
- First Segment
  - Pointer to very first SID



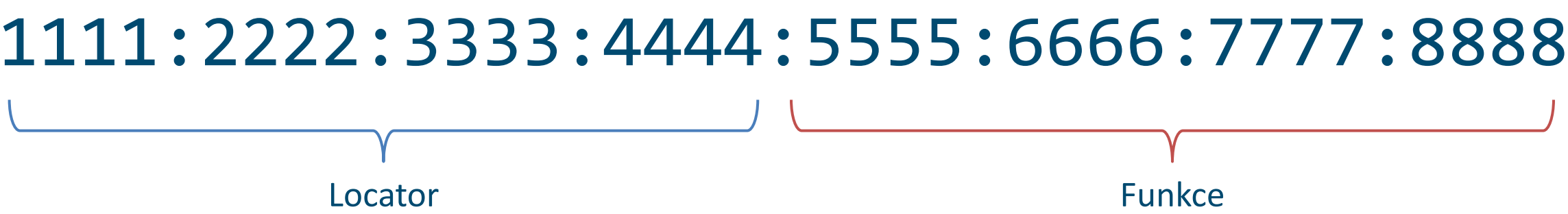
NIX·CZ

SRH

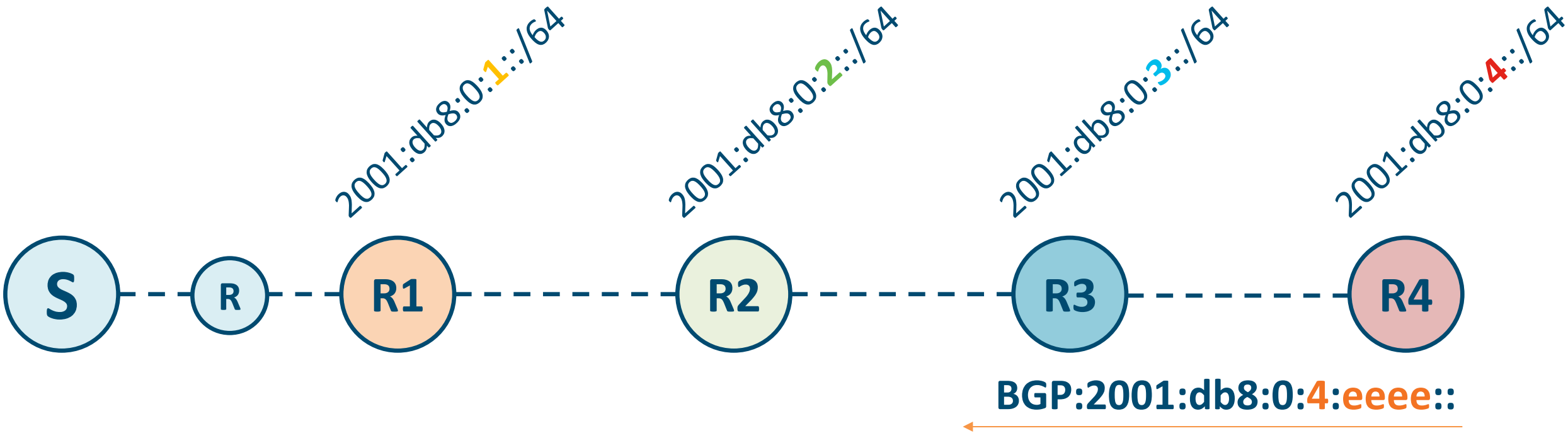
- Segment Routing Header
- First Segment
  - Pointer to very first SID
- Segments left
  - Pointer to Active SID
  - Active SID always in destination addr



Struktura SIDu - Locator



NIX·CZ




```
SA:2001::1
DA:2001:db8:0:1:1::
NH:RH
Type:4(SRH)
NH:IPv4|SL:3
Segment List:
[0]:2001:db8:0:4:eeee::
[1]:2001:db8:0:3:48::
[2]:2001:db8:0:2:1::
[3]:2001:db8:0:1:1::
```

```
SA:2001::1
DA:2001:db8:0:2:1::
NH:RH
Type:4(SRH)
NH:IPv4|SL:2
Segment List:
[0]:2001:db8:0:4:eeee::
[1]:2001:db8:0:3:48::
[2]:2001:db8:0:2:1::
[3]:2001:db8:0:1:1::
```

```
SA:2001::1
DA:2001:db8:0:3:48::
NH:RH
Type:4(SRH)
NH:IPv4|SL:1
Segment List:
[0]:2001:db8:0:4:eeee::
[1]:2001:db8:0:3:48::
[2]:2001:db8:0:2:1::
[3]:2001:db8:0:1:1::
```

```
SA:2001::1
DA:2001:db8:0:4:eeee::
NH:IPv4
Type:4(SRH)
NH:IPv4|SL:0
Segment List:
[0]:2001:db8:0:4:eeee::
[1]:2001:db8:0:3:48::
[2]:2001:db8:0:2:1::
[3]:2001:db8:0:1:1::
```

Struktura SID

NIX·CZ

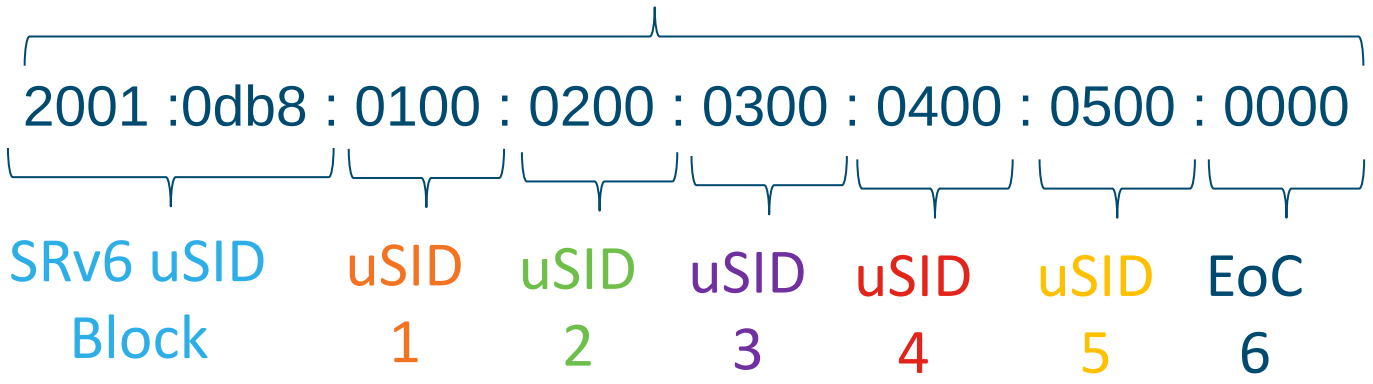
SRv6 - µSID (uSID)

„Zjednodušení“

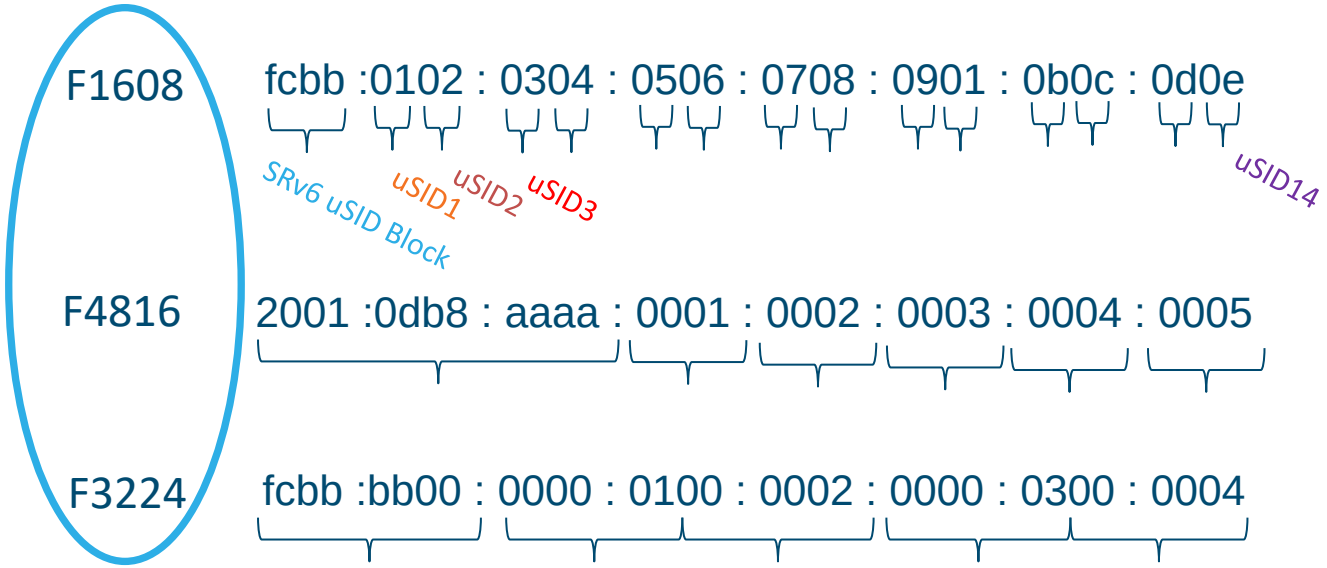
NIX·CZ

SRv6 uSID format

SRV6 uSID Container



FXXYY -> XX= Block size; Y=uSID size

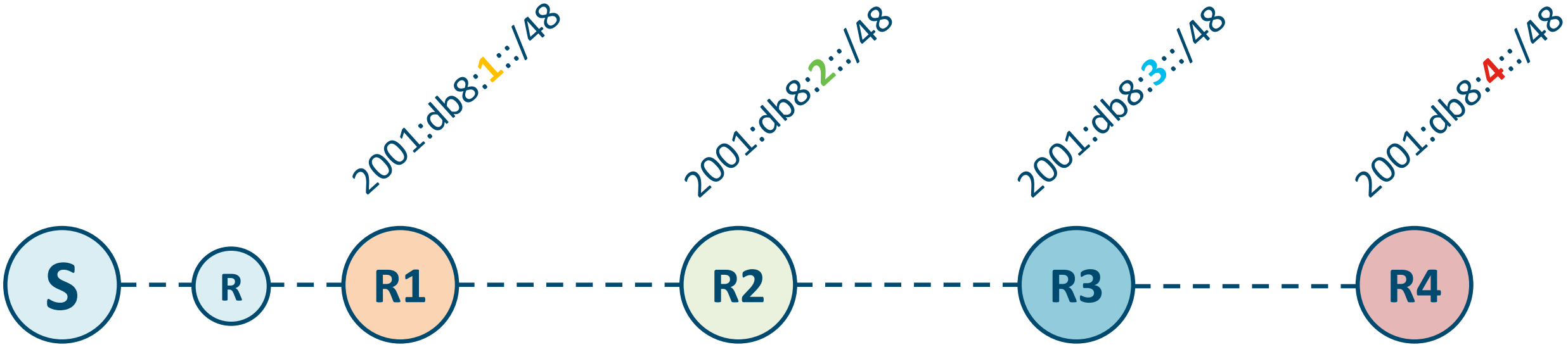


SRV6 Encapsulation

```
SA:2001::1
DA:2001:db8:0:4:1:0:0:0
NH:RH

Type:4(SRH)
NH:IPv4|SL:1
Segment List:
[0]: 2001:db8:0:5:45:0:0:0
[1]: 2001:db8:0:4:1:0:0:0
[2]: 2001:db8:0:3:48:0:0:0
[3]: 2001:db8:0:2:1:0:0:0
[4]: 2001:db8:0:1:42:0:0:0

SA:7.5.4.3
DA:11.6.19.71
Port:UDP

UDP Header/Data
```

SRV6 uSID Encapsulation

```
SA:2001::1
DA:2001:db8:100:200:300:400:500::
NH:IPv4

SA:7.5.4.3
DA:11.6.19.71
Port:UDP

UDP Header/Data
```



BGP:2001:db8:4:eeee::

```
SA:2001::1
DA:2001:db8:1:2:3:e000:4:eeee
NH:IPV4
```

```
SA:2001::1
DA:2001:db8:2:3:e000:4:eeee::
NH:IPV4
```

```
SA:2001::1
DA:2001:db8:3:e000:4:eeee::
NH:IPV4
```

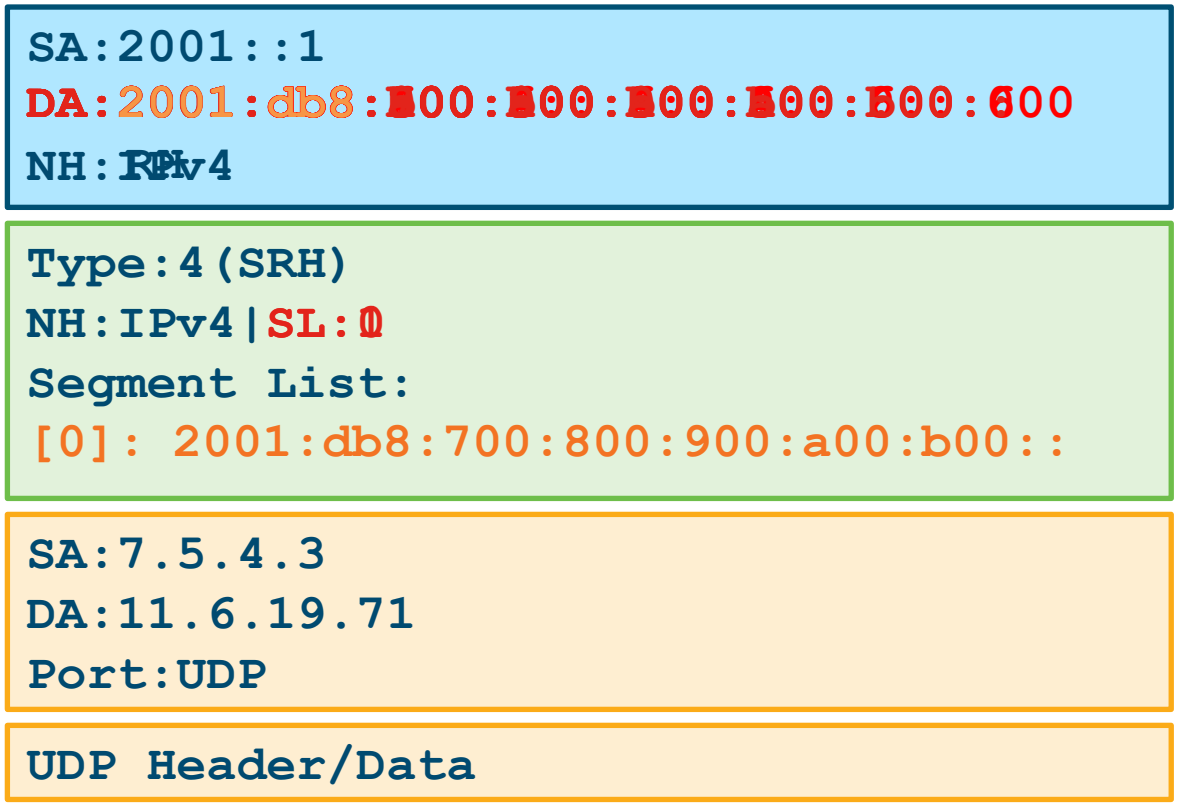
```
SA:2001::1
DA:2001:db8:4:eeee::
NH:IPV4
```

SRv6 uSID více než 6 SIDů?

100->200->300->400->500->600->700->800->900->a00->b00

Container 1 2001 : 0db8 : 0100 : 0200 : 0300 : 0400 : 0500 : 0600

Container 2 2001 : 0db8 : 0700 : 0800 : 0900 : 0a00 : 0b00 : 0000



Shift & Forward

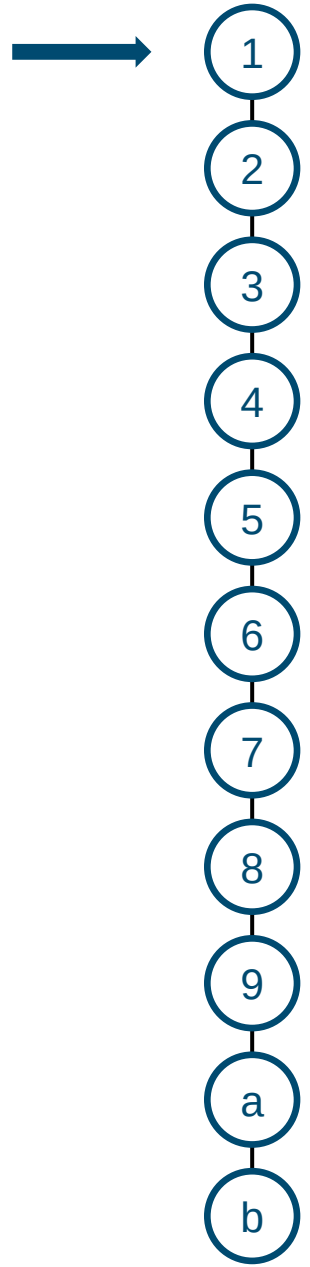
END of Container

-> is there SRH?

Decrement SL

Copy New SID (Container)

PSP

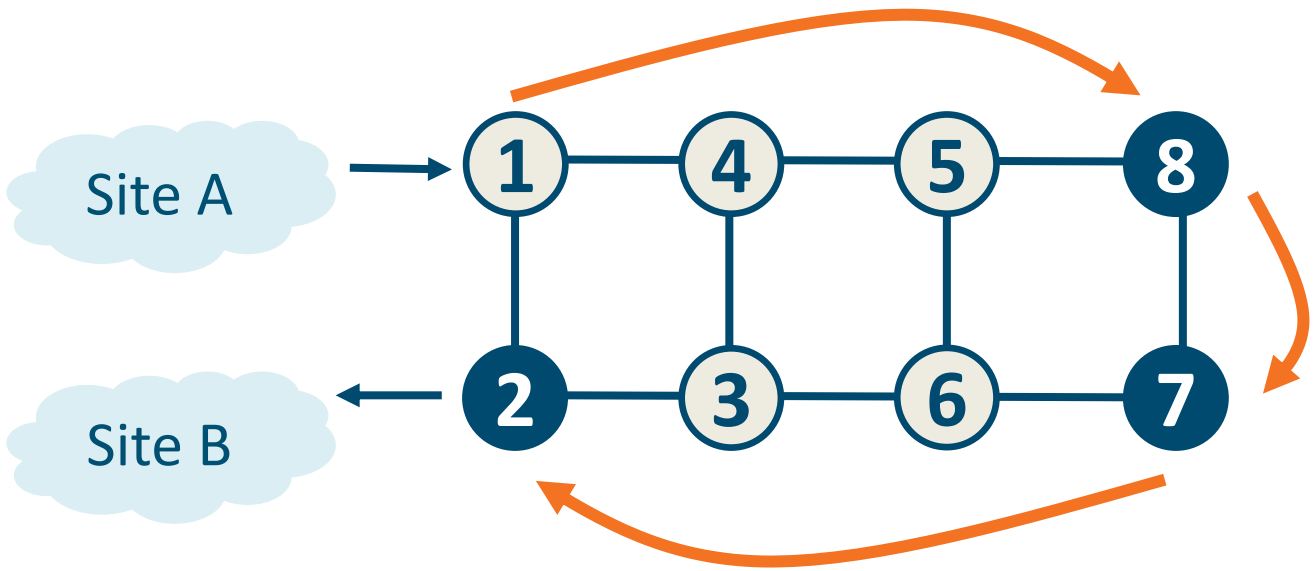


NIX·CZ

SRv6 functions: Řízení a služby / Instrukce

Codename		Behavior	
End	uN	Endpoint	[Node SID]
End.X	uA	Endpoint with Layer-3 cross-connect	[Adj SID]
End.B6.Insert	uB6.Insert	Endpoint bound to an SRv6 policy	[BSID]
End.B6.Encap	uB6.Encaps	Endpoint bound to an SRv6 encapsulation policy	[BSID]
End.DX6	uDX6	Endpoint with decapsulation and IPv6 cross-connect	[L3VPN Per-CE]
End.DX4	uDX4	Endpoint with decapsulation and IPv4 cross-connect	[L3VPN Per-CE]
End.DT6	uDT6	Endpoint with decapsulation and specific IPv6 table lookup	[L3VPN Per-VRF]
End.DT4	uDT4	Endpoint with decapsulation and specific IPv4 table lookup	[L3VPN Per-VRF]
End.DT46	uDT46	Endpoint with decapsulation and specific IPv4&v6 table lookup	[L3VPN Per-VRF]
End.DX2	uDX2	Endpoint with decapsulation and L2 cross-connect	[E-LINE]
End.DT2U/M	uDT2U/M	Endpoint with decapsulation and L2 unicast lookup / flooding	[E-LAN]
End.DTM	uDTM	Endpoint with decapsulation and MPLS table lookup	[Interworking]
H.Insert / H.Encaps		Headend with Insertion / Encapsulation of / into an SRv6 policy	[TiLFA]
H. Encaps.L2		H.Encaps Applied to Received L2 Frames	[L2 Port Mode]
H.Encaps.M		H.Encaps Applied to MPLS Label Stack	[Interworking]

The DA encodes the desired path , from Site A to Site B.



One single micro-program in the DA is enough
DA = BBBB:BBBB:0800:0700:0200:0000:0000:0000

- uN(8): "go via shortest path to node 8", identified by 0x0800 for human readability
- uN(7): "go via shortest path to node 7", identified by 0x0700 for human readability
- uN(2): "go via shortest path to node 2", identified by 0x0200 for human readability
- End of Container, identified by 0x0000
- End of Container, identified by 0x0000
- End of Container, identified by 0x0000

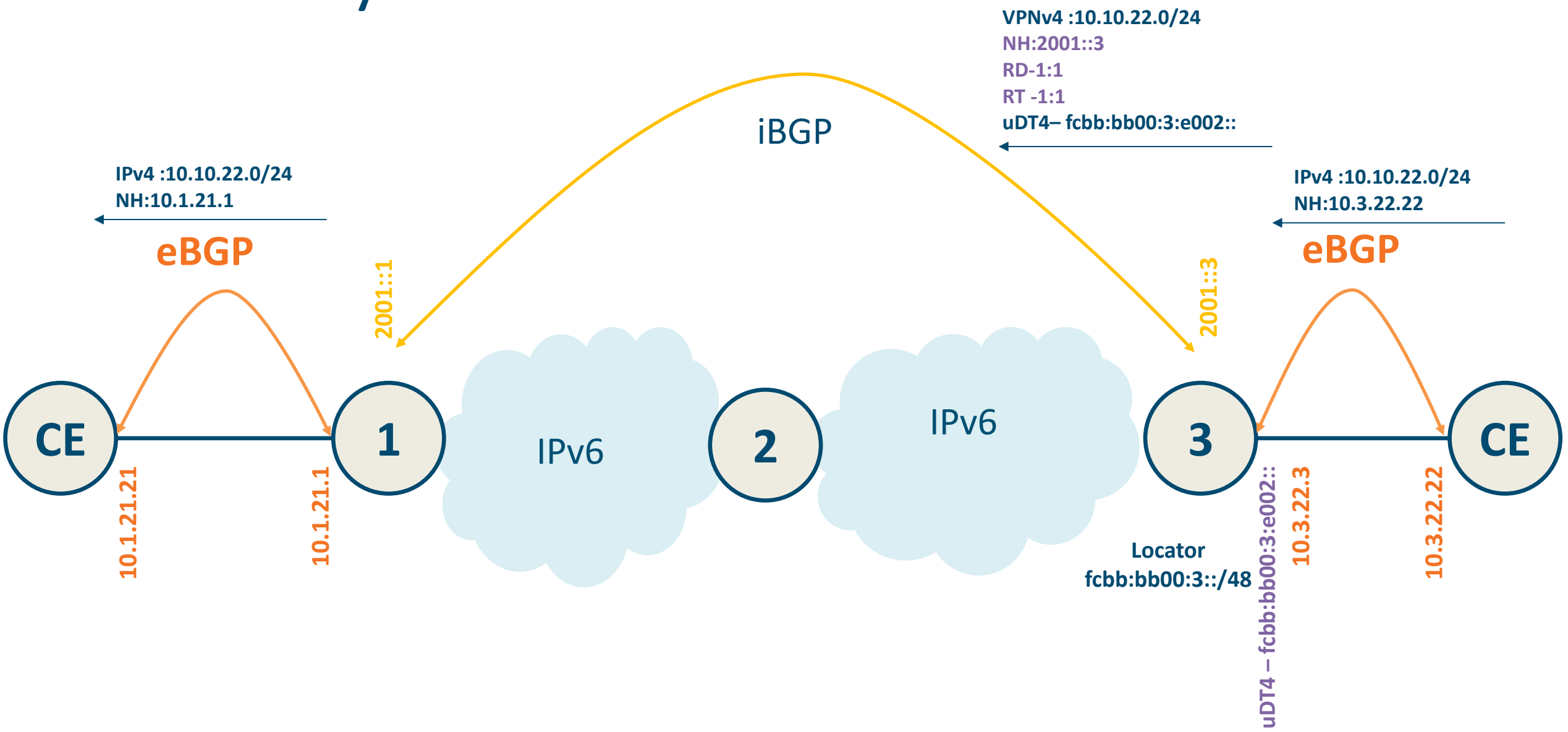
NIX·CZ

Control Plane

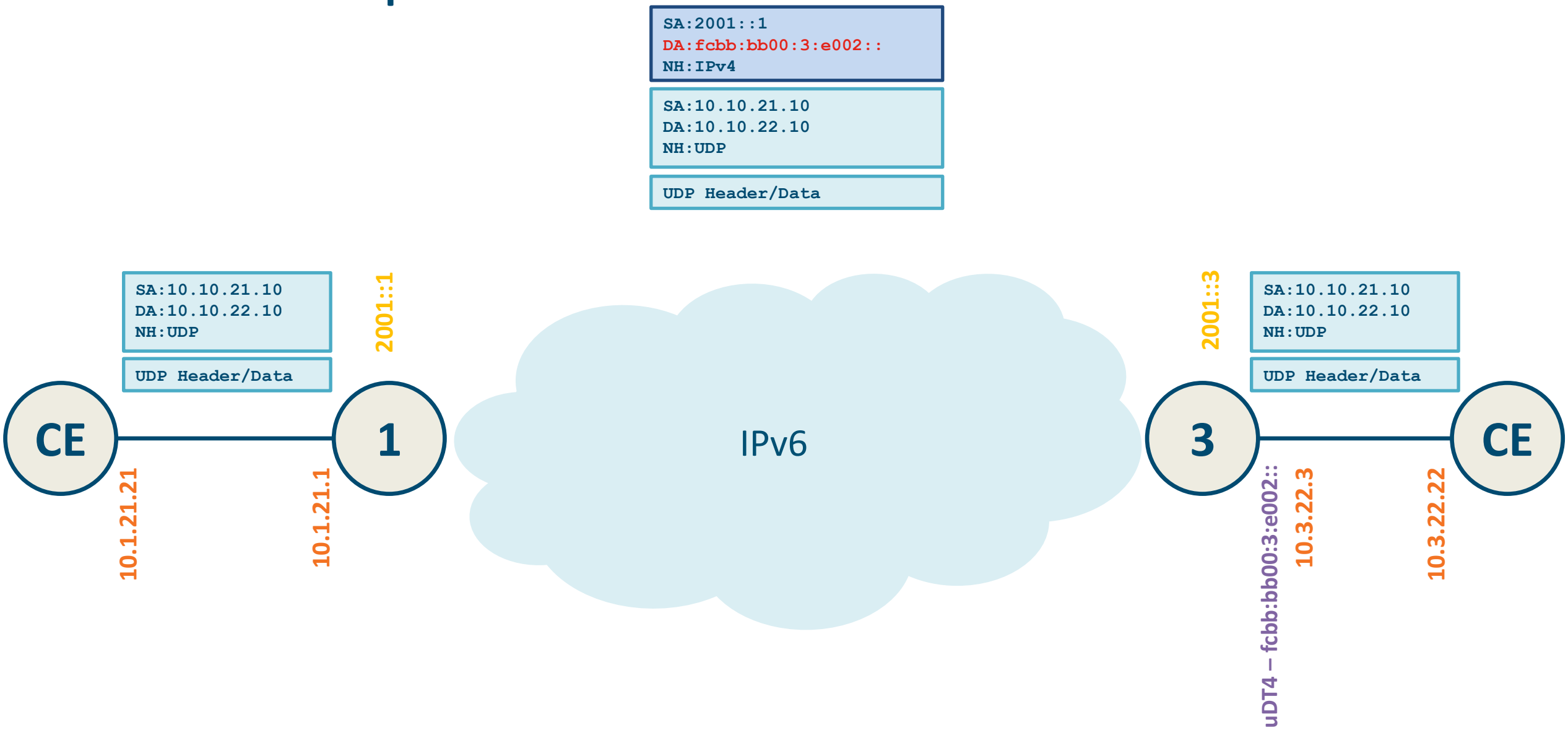
„Kam s ním?“

NIX·CZ

BGP Overlay

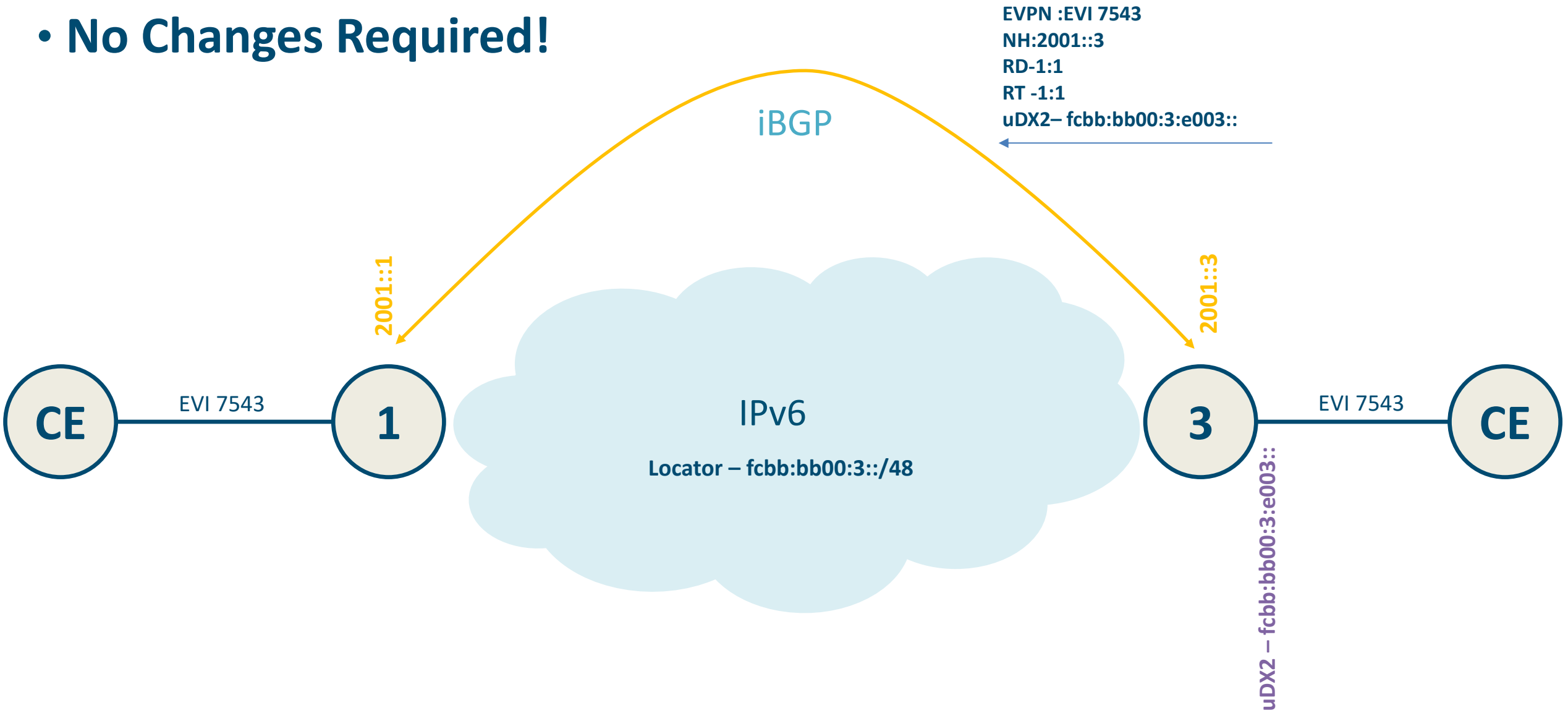

VPNv4 :10.10.22.0/24
NH:2001::3
RD-1:1
RT -1:1
uDT4– fcbb:bb00:3:e002::
iBGP
IPv4 :10.10.22.0/24
NH:10.1.21.1
eBGP
IPv4 :10.10.22.0/24
NH:10.3.22.22
eBGP
2001::1
2001::3
CE
1
IPv6
2
IPv6
3
CE
10.1.21.21
10.1.21.1
Locator
fcbb:bb00:3::/48
uDT4 – fcbb:bb00:3:e002::
10.3.22.3
10.3.22.22

L3 VPN Dataplane



EVPN

- **No Changes Required!**



EVPN Dataplane

ISIS for SRv6

LSP (Link State Packet):

TLVs:

Hostname: **R2**

Interfaces: Hu0/0/0/0
Hu0/0/0/1
Lo0

uA:fcbb:0:2:e001::
Structure: BL=32;NL=16;FL=16;AL=0

uA:fcbb:0:2:e002::
Structure: BL=32;NL=16;FL=16;AL=0

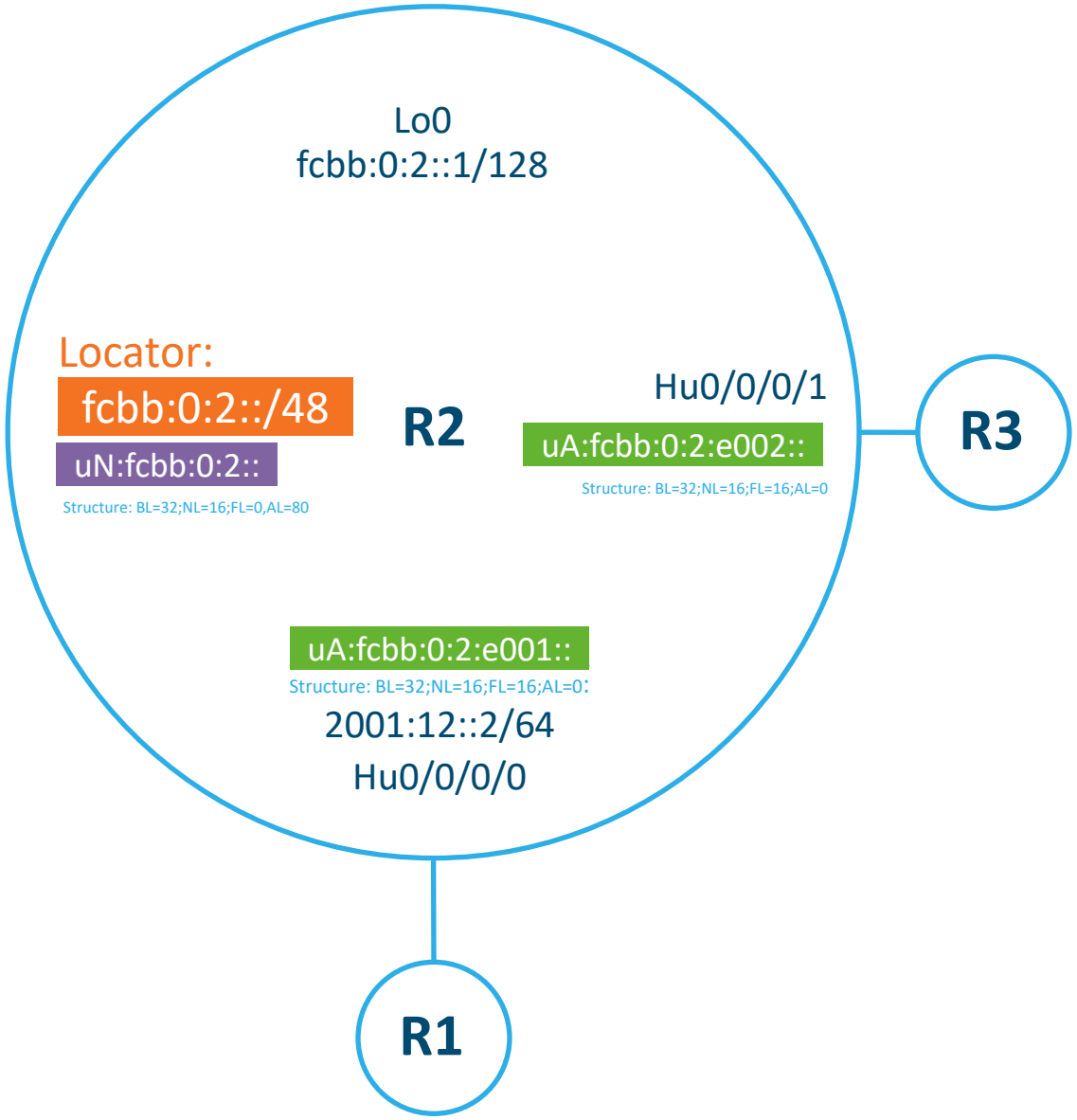
Neighbors: **R1**
R3

IP addresses: fcbb:0:2::1/128
2001:12::2/64

Locator: fcbb:0:2::/48
uN:fcbb:0:2::

Structure: BL=32;NL=16;FL=0,AL=80

Capabilities: Algorithms ……
SIDs can insert
SIDs can decap



ISIS LSP Example

```
IS-IS 1 (Level-2) Link State Database
LSPID                 LSP Seq Num  LSP Checksum  LSP Holdtime/Rcvd  ATT/P/OL
r2.00-00              0x00000009   0x4f06        1145 /1200         0/0/0
  Area Address:   49
  NLPID:          0x8e
  Hostname:       r1
  IPv6 Address:   2001::2
  Metric: 10         MT (IPv6 Unicast) IPv6 fcbb:bb00:2::1/128
    Prefix Attribute Flags: X:0 R:0 N:1 E:0 A:0
  Metric: 1          MT (IPv6 Unicast) IPv6 fcbb:bb00:2::/48
    Prefix Attribute Flags: X:0 R:0 N:0 E:0 A:0
  MT:             IPv6 Unicast                                0/0/0
  SRv6 Locator:   MT (IPv6 Unicast) fcbb:bb00:2::/48 D:0 Metric: 0 Algorithm: 0
    Prefix Attribute Flags: X:0 R:0 N:0 E:0 A:0
    END SID: fcbb:bb00:2:: uN (PSP/USD)
      SID Structure:
        Block Length: 32, Node-ID Length: 16, Func-Length: 0, Args-Length: 0
  Router Cap:     0.0.0.0 D:0 S:0
    IPv6 Router ID: 2001::2
    SR Algorithm:
      Algorithm: 0
      Algorithm: 1
    SRv6: O:0
    Node Maximum SID Depth:
      SRH Max SL:       3
      SRH Max End Pop:  3
      SRH Max T.insert: 3
      SRH Max T.encaps: 4
      SRH Max End D:    4
  Metric: 10         MT (IPv6 Unicast) IS-Extended r1.00
    Local Interface ID: 6, Remote Interface ID: 6
    Interface IPv6 Address: 2001:12::2
    Neighbor IPv6 Address: 2001:12::1
    END.X SID: fcbb:bb00:2:e001:: B:0 S:0 P:0 uA (PSP/USD) Alg:0
      SID Structure:
        Block Length: 32, Node-ID Length: 16, Func-Length: 16, Args-Length: 0
Total Level-2 LSP count: 1     Local Level-2 LSP count: 0
```

Locator

Capabilities

END

END.X

SID Structure

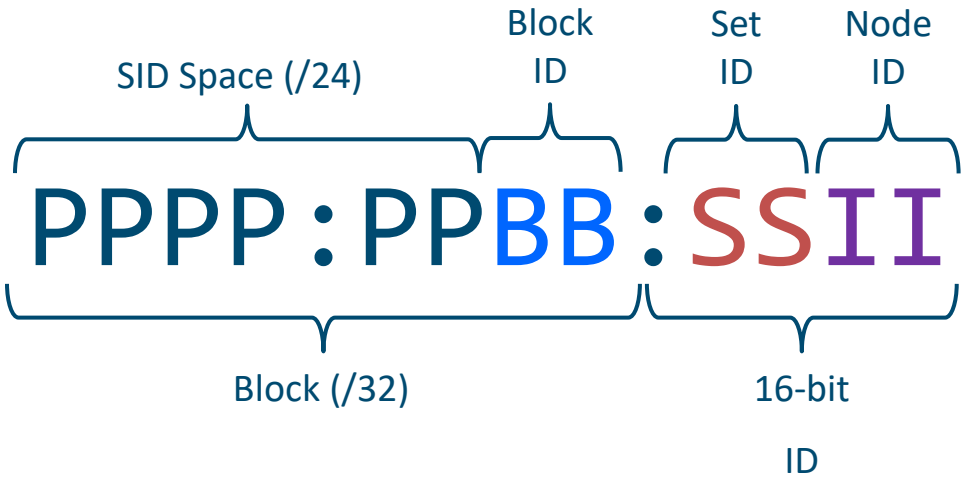
Adresování

NIX·CZ

Terminologie adresování

uSID F3216: uSID format with
- uSID Block size: 32 bits
- ID size: 16 bits

uSID F3216 structure:



NIX·CZ

1. Doporučené adresování ✓
   - Privátní alokace /24 z FC00::/8
   - FCCA:FE00::/24 – CA:FE jsem zvolil pro NIX.CZ

2. Podporované, ale nedoporučené ⚠
   - Alokace z veřejných GUA adres

3. SRv6 alokace ✓
   - Použití /24 z podsítě 5F00::/16

FCCA:FE00::/32

00 – Low-cost (Flex algo SPF)
01 – Low-delay (Flex algo latency)

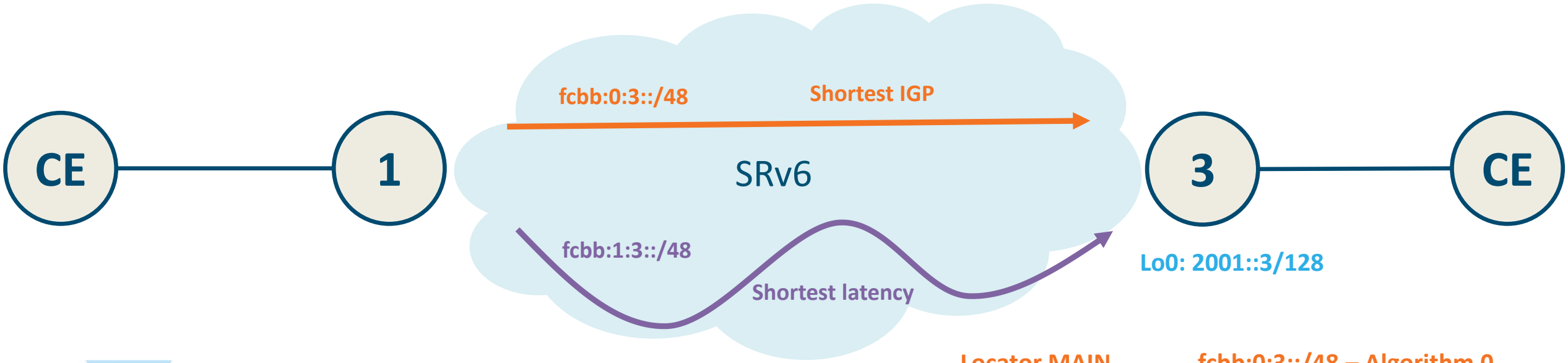
Flex Algo

NIX·CZ

Flexible Algorithm

- Zkráceně “Flex-Algo”
  - Definuje operátor sám na základě vlastních provozních pravidel
  - Výchozí algoritmus 0 uvažuje vše (routery, linky, IGP metriky)
- Flex-Algo K je definován jako
  - Minimalizuje některou z metrik: IGP(cestu), zpoždění, …
  - Vyřazení metrik z výpočtu cesty: souběh linek, SRLG, …
- Příklad
  - Operator1 definuje Flex-Algo 128 jako "minimalizuj IGP metriky a souběh linek s názvem "green"
  - Operator1 definuje Flex-Algo 128 jako "minimalizuj zpozdění a souběh linek s názvem "blue"

SRv6 Flex Algo IGP

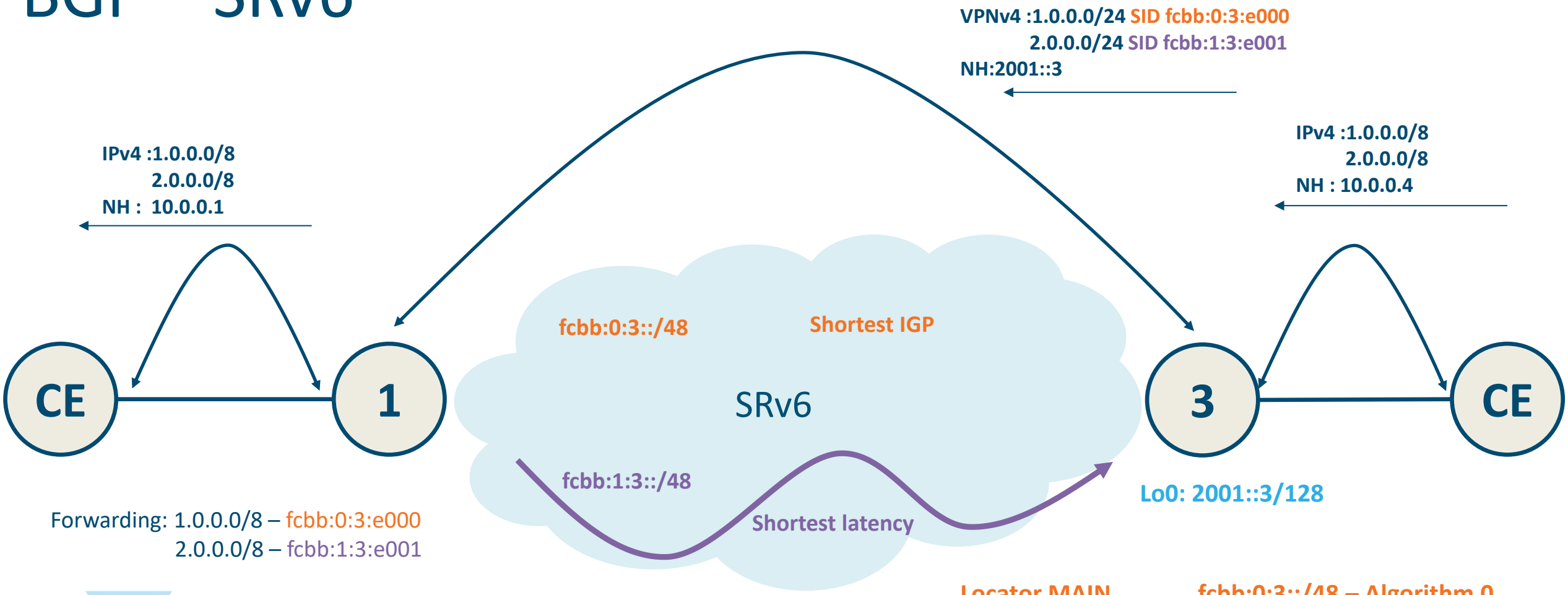


Locator MAIN fcbb:0:3::/48 – Algorithm 0

Locator LATENCY fcbb:1:3::/48 - Algorithm 128

NIX·CZ

BGP – SRv6

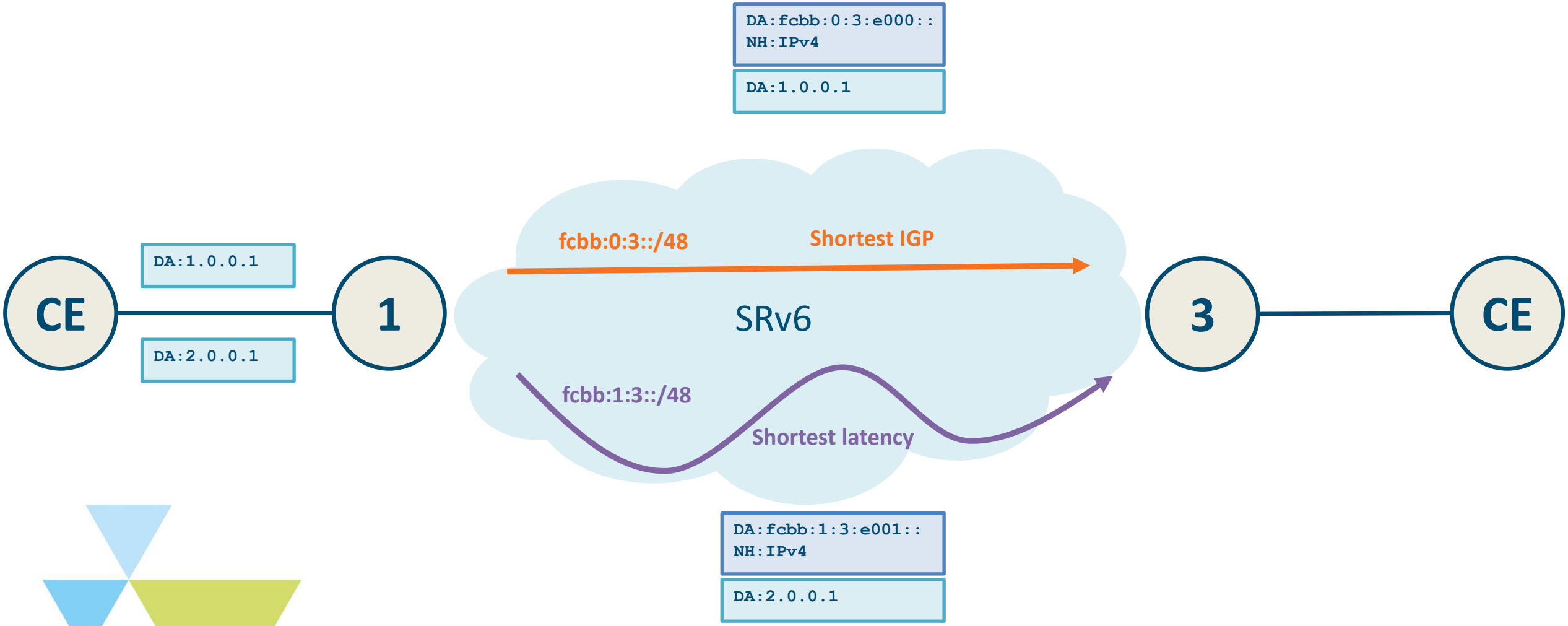


Forwarding: 1.0.0.0/8 – fcbb:0:3:e000
2.0.0.0/8 – fcbb:1:3:e001

Locator MAIN fcbb:0:3::/48 – Algorithm 0
Locator LATENCY fcbb:1:3::/48 - Algorithm 128

NIX·CZ

SRv6 DATAPLANE



NIX·CZ

SRv6 Flex Algo - IGP

```
segment-routing
 srv6
  locators
   locator LATENCY
    micro-segment behavior unode psp-usd
    prefix fcbb:bb01:1::/48
    algorithm 128
router isis 1
 flex-algo 128
  metric-type delay
  advertise-definition
 address-family ipv6 unicast
  segment-routing srv6
   locator LATENCY
```

New Locator Name

Locator Prefix (Different)

Flex Algo number 128-255

Definition of specific Flex Algo
Latency metric for 128

This Router will advertise
FA definition within the domain

This will result in:

- Locator is advertised +FA definition
- uN function is advertised - for FA
- uA for each ISIS interface is allocated and advertised for FA

SRv6 L3 VPN Flex Algo

```
router bgp 1
 address-family vpnv4 unicast
 vrf LowLatency
  rd 1:2
  address-family ipv4 unicast
   segment-routing srv6
    locator LATENCY
    alloc mode per-vrf
```

Name of the Locator

Single DT function is allocated per VRF and AF

This will result in:

- uDT4 function is allocated from LATENCY locator
- All prefixes in VRF are advertised with uDT4 function

SRv6 L3 VPN Multiple Algorithms in VRF

```
 route-policy MIX
   if destination in (1.1.1.1/32) then
     set srv6-alloc-mode per-vrf locator LATENCY
   else
     set srv6-alloc-mode per-vrf locator MAIN
   endif
 end-policy
router bgp 1
 vrf Both
  address-family ipv4 unicast
    segment-routing srv6
      alloc mode MIX
```

For prefix 1.1.1.1 we will allocate uDT from LATENCY

For all others uDT from MAIN

Route-Policy application

Latency Configuration

```
performance-measurement
 interface Gig0/0/0/0
  delay-measurement
   advertise-delay 7543
```


This will set latency of the link to 7543 **microseconds**

NIX·CZ

Děkuji za pozornost

mr@nix.cz

NIX·CZ