IPv6 Courses

December 20, 2010

Table of Contents

1 IPv6 Routing

G6

- Group of IPv6 actors in France (researchers, engineers...)
- Academic & industrial partners
  - CNRS, Institut TELECOM, INRIA, Universities...
  - AFNIC, 6Wind, Bull...
- Launched in 1995 by:
  - Alain Durand
  - Bernard Tuy
- Is today a legal association under French Law (1901)
  - Laurent Toutain, President
- For further information: http://www.g6.asso.fr/

G6Charter

- Share experience gained from IPv6 experimentations and deployment
- Spread IPv6 information
  - Tutorials and trainings (ISPs, Engineers, netadmins...)
  - Online book (in French), "IPv6, Théorie et pratique": http://livre.g6.asso.fr/
- Initiate research activities around IPv6
- Active in RIPE & IETF working groups
- Promotion of IPv6: French Task Force

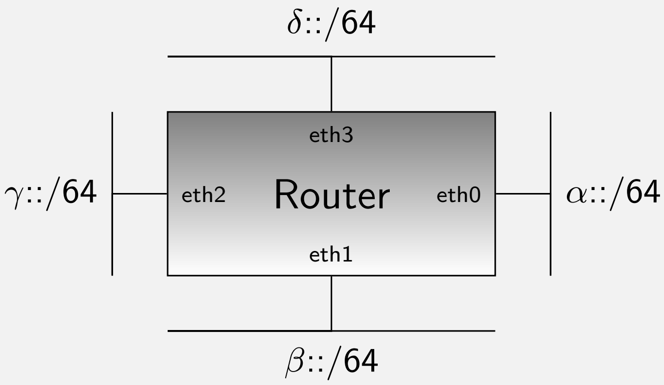
Hypertext Symbols

- Several symbols are used in this document:
  - All RFCs and Internet Drafts are hypertext links.
    - Check that there is no more recent version of the document.
  - is a link to a *Techniques de l'Ingénieur* article on the subject (in French, access may be restricted).
  - is a link to the online edition of *IPv6, Théorie et Pratique* (in French)
  - is a link to other information on the web.
- Material concerning IPv6 is taken from the G6 tutorial and copyrighted from G6.

Router configuration

Router's interface must be manually configured

- assign an address + a prefix length
- routers install automatically in the FIB the prefix assigned to the interface



FIB

Prefix	Next Hop
α	eth0
β	eth1
γ	eth2
δ	eth3

Example : Cisco router

IPv6 Routing
Concepts & Generality
IGP
EGP
Usage

```
cisco_showroom#   show ipv6 route

IPv6 Routing Table - 7 entries

Codes: C - connected, L - Local, S - Static, R - RIP, B - BGP
U - Per-user Static route, I1 - ISIS L1, I2 - ISIS L2, IA - ISIS interarea,
IS - ISIS summary, O - OSPF intra, OI - OSPF inter, OE1 - OSPF ext 1,
OE2 - OSPF ext 2, ON1 - OSPF NSSA ext 1, ON2 - OSPF NSSA ext 2



S ::/0 [1/0] via FE80::216:17FF:FE87:A7, Vlan338
C 2001:660:7301:3303::/64 [0/0] via ::, Vlan333
L 2001:660:7301:3303::1/128 [0/0] via ::, Vlan333
C 2001:660:7301:3308::/64 [0/0] via ::, Vlan338
L 2001:660:7301:3308:20D:29FF:FE75:43C4/128 [0/0] via ::, Vlan338
L FE80::/10 [0/0] via ::, Null0
L FF00::/8 [0/0] via ::, Null0

....
```


Example: Linux

IPv6 Routing
Concepts & Generality
IGP
EGP
Usage

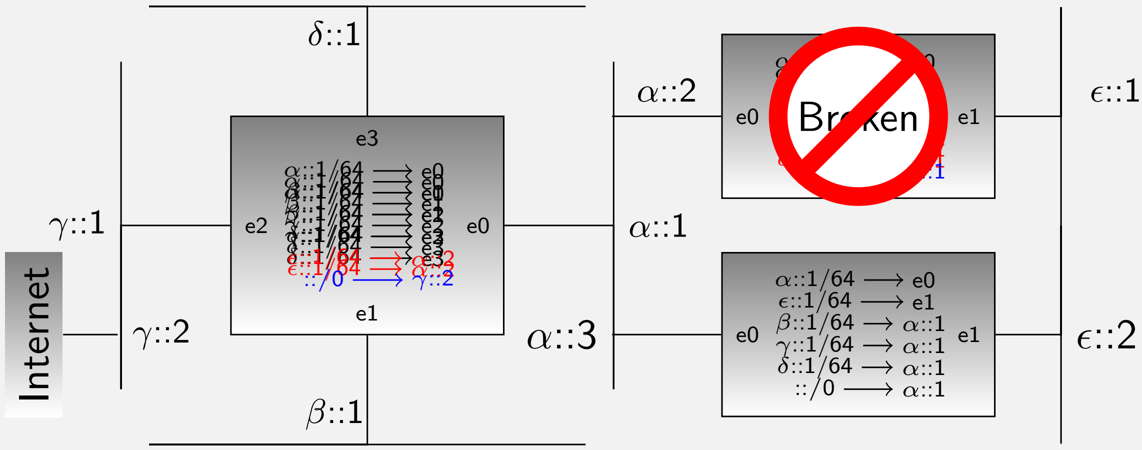
netstat -rn ip -6

Table de routage IPv6 du noyau

Destination	Next Hop	Flags	Metric	Ref	Use	Iface
::1/128	::	U	0	6	1	lo
2001:660:7301:3302::/128	::	U	0	0	2	lo
::/0	fe80::213:c4ff:fe69:5f49	UG	1	34532	0	eth1
2001:660:7301:3303::/64	fe80::20d:29ff:fe75:43c4	UG	1024	6708480	0	eth0.338

Static route

- Manually configured routes



- simple to configure, but subject to errors (loops)
- cannot find another path if router fails

Example : commands

- BSD:
  - `route add -inet6 default fe80::216:17ff:fe87:a7`
  - `route add -inet6 2001:660:7301:3305::1 -prefixlen 64 fe80::216:17ff:fe87:a7`
- Linux:
  - `route -A inet6 add default gw fe80::216:17ff:fe87:a7 dev eth0`
  - `ip -6 route add default via fe80::216:17ff:fe87:a7/64 dev eth0`
- Cisco:
  - `ipv6 route ::/0 vlan 338 fe80::216:17ff:fe87:a7`
  - `ip -6 route add 2001:660:7301:3303::/64 via fe80::20d:29ff:fe75:43c4 dev eth0.338`

Example : Cisco router

IPv6 Routing
Concepts & Generality
IGP
EGP
Usage

```
cisco_showroom#  show ipv6 route
IPv6 Routing Table - 7 entries

Codes: C - Connected, L - Local, S - Static, R - RIP, B - BGP
I1 - ISIS L1, I2 - ISIS L2, IA - ISIS interarea, IS - ISIS summary
O - OSPF intra, OI - OSPF inter, OE1 - OSPF ext 1, OE2 - OSPF ext 2
ON1 - OSPF NSSA ext 1, ON2 - OSPF NSSA ext 2, U - Per-user Static route

S ::/0 [1/0] via FE80::216:17FF:FE87:A7, Vlan338
C 2001:660:7301:3303::/64 [0/0] via ::, Vlan333
L 2001:660:7301:3303::1/128 [0/0] via ::, Vlan333
C 2001:660:7301:3308::/64 [0/0] via ::, Vlan338
L 2001:660:7301:3308:20D:29FF:FE75:43C4/128 [0/0] via ::, Vlan338
L FE80::/10 [0/0] via ::, Null0
L FF00::/8 [0/0] via ::, Null0
...
```


Routing Protocol

IPv6 Routing
Concepts & Generality
IGP
EGP
Usage

Definition

A routing protocol is an application sharing knowledge among routers to construct FIB

- The application may have its own database different from the FIB
- Two families of routing protocol are defined:
  - Interior Gateway Protocol
  - Exterior GAteway Protocol

IGP

Interior Gateway Protocol

- Simple configuration (even if protocol can be complex)
- Discover other routers, and exchange information
- E.g, RIPng (Distance Vector), OSPFv3 or IS-IS (Link State)

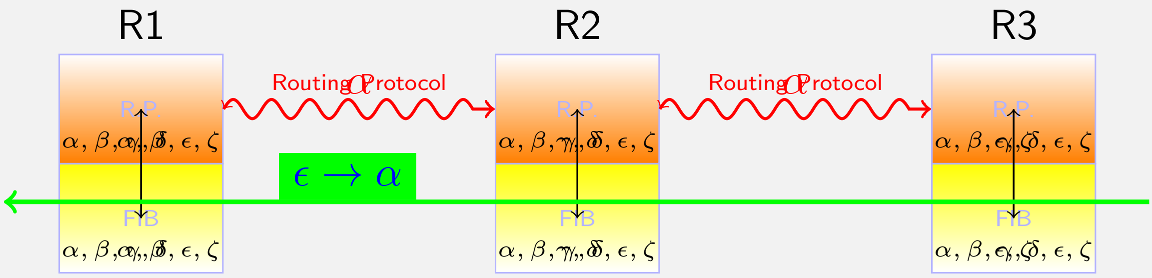
EGP

Exterior GAteway Protocol

- A lot of controls, nothing is automatically learnt
- Complex configuration to hide or show prefixes
- Used mainly between providers
- Only one protocol used : Border Gateway Protocol (with its extensions: MBGP)

Router's behavior

- All routers have a manual configuration of their interfaces
- Prefixes learnt from this configuration are spread among other routers
- Routers select announcements and add these prefixes in their FIB
- Routing announcements and traffic (forwarding) flow in opposite directions



RIPng

Distance Vector Algorithm

IPv6 Routing
Concepts & Generality
IGP
EGP
Usage

- Very simple algorithm
- Routing Protocol database and FIB are common
  - A cost is just added to each prefix
    → Usually the number of routers to cross
- Current implementation:
  - IPv4: RIPv2 (RFC 1723)
  - IPv6: RIPng (RFC 2080)

Distance Vector Algorithm

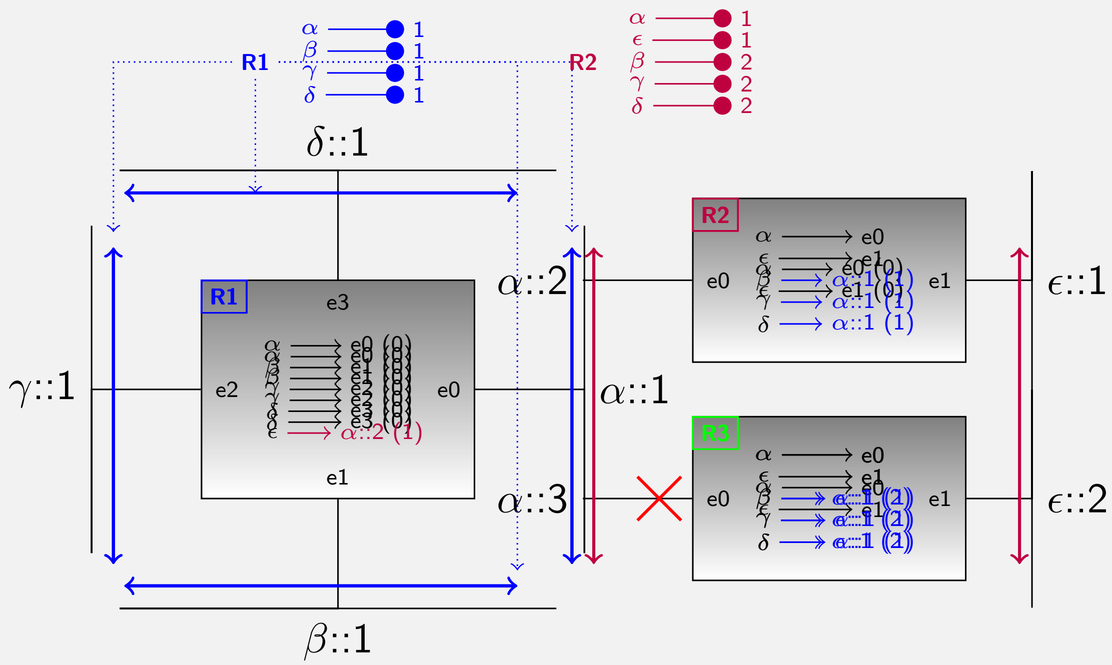
IPv6 Routing
Concepts & Generality
IGP
EGP
Usage

Principle

Routers periodically broadcast their FIB on each link they are connected to:

- If a router find a new prefix in broadcasted information, ...
- else if an broadcasted prefix as a lower cost than the one already stored
  - This entry is added/changed in the FIB
  - Next Hop is set to the IP address of the broadcasting router
  - Cost is increased by one
- Otherwise the information is ignored

Distance Vector Example (Add Route)

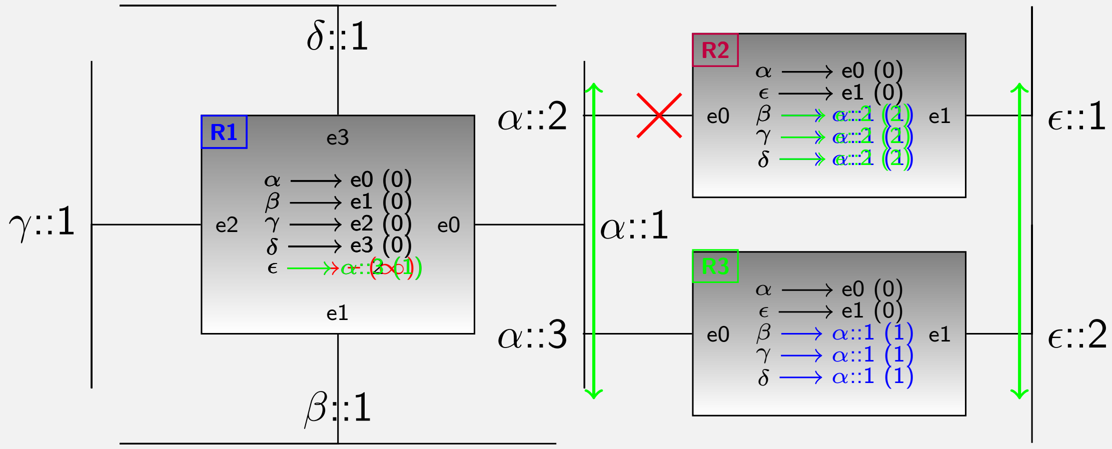


Distance Vector: Periodic Refreshing

IPv6 Routing
Concepts & Generality
IGP
EGP
Usage

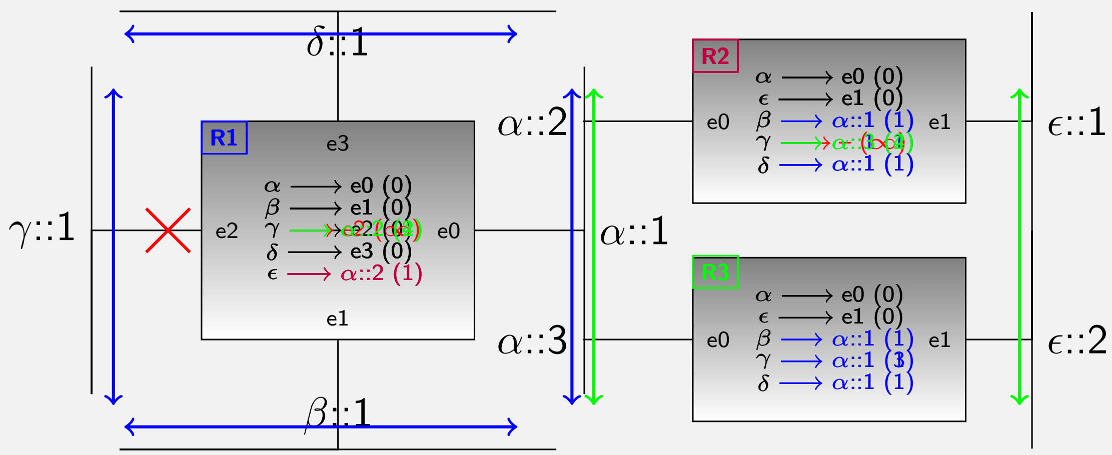
- Distance Vector (DV) implies periodic refreshing
  - To discover best path
  - To discover dead routers
- In RIP Routing Table are flooded every 30 seconds
- No specific message to remove an entry
  - If an entry is not seen during X flooding, this entry is removed
    - In RIP $X = 3$ (90 s)
  - Remove is similar to set cost to ∞
- DV cannot be used if Routing Table contains a lot of entries
  - For instance core network routers may contain 220 000 entries
- DV can only be used in very small networks
- DV have also bad convergence performances...

Distance Vector Example (Change Route)



Distance Vector Example (Bad Convergence)

IPv6 Routing
Concepts & Generality
IGP
EGP
Usage

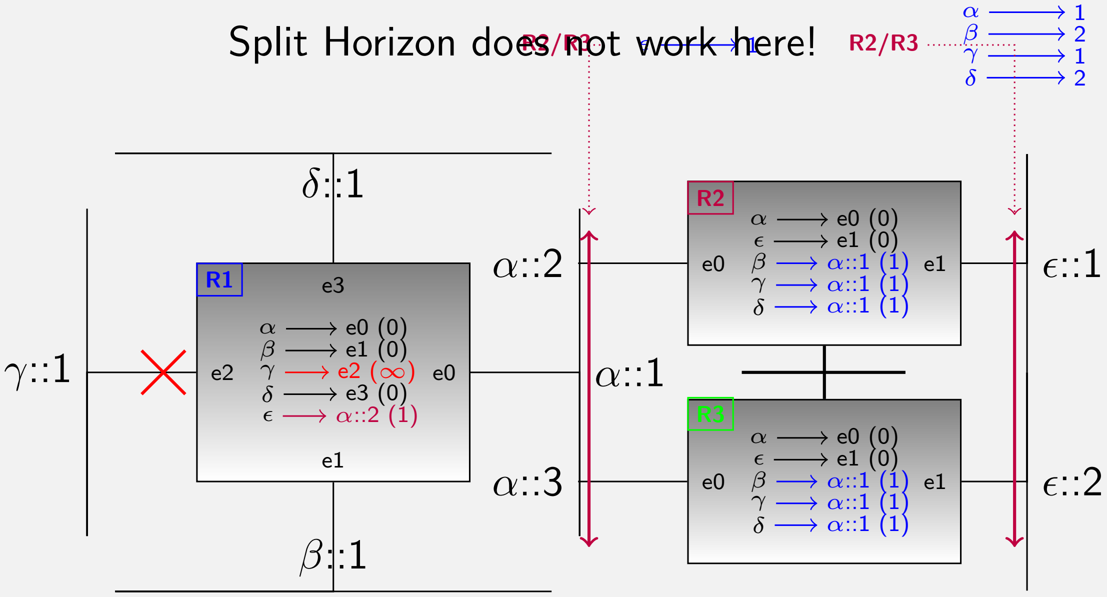


Counter Measures

- 3 ways to reduce this impact :
  - $16 = \infty$:
    - Not only 15 routers in the network but :
    - 15 routers between two links
  - Poisoning reverse:
    - When receiving a route marked ∞, mark it immediately ∞
    - Propagate rapidly route withdraw to avoid wrong routing messages
  - Split Horizon:
    - Do not announce through an interface, routes that use this interface.

Distance Vector Example (Split Horizon)

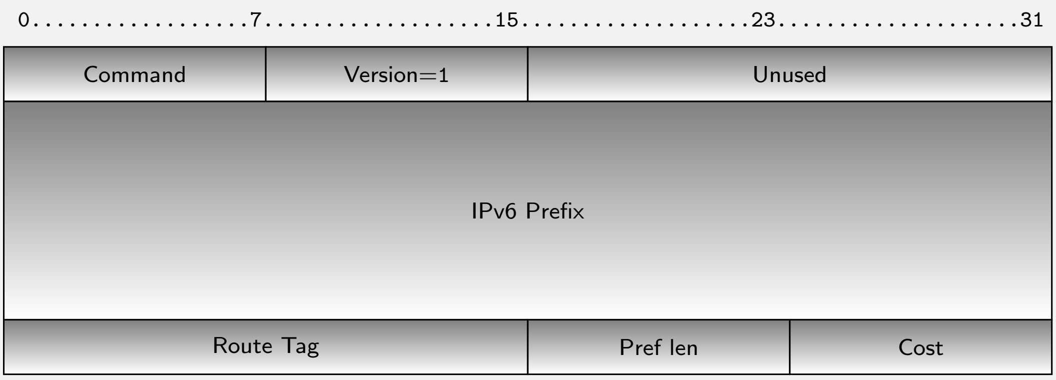
Split Horizon does not work here!



Distance Vector : Summary

- Distance Vector is very simple to understand, manage, implement.
- Performances are weak:
  - Shortsighted network vision: based on summary made by other routers
    - Like *concierges exchanging gossips*
  - Can lead to routing loops or long delays to reconnect parts of the network.
  - Periodically all routing table must be sent on the network:
    - to detect dead routers
    - to detect better paths
  - Generates network and processing load
- Distance Vector must be limited to small network auto-configuration
- RIPng implement this algorithm

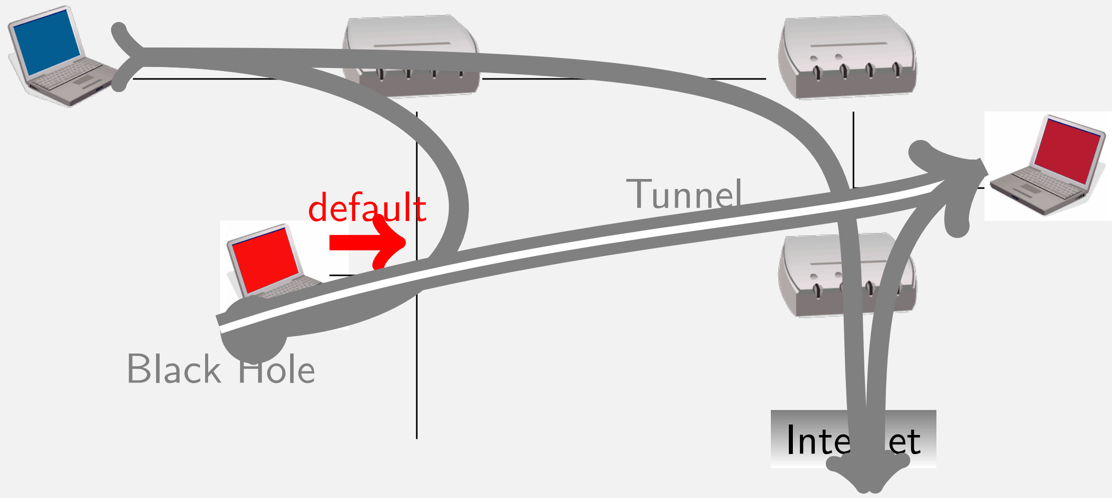
RIPng



- Multicast address : `ff02:2::9`, UDP Port Number 521
- Next Hop : instead of the source address in the IP header
- Route Tag : Differentiate internal and external routes

Securing Routing Protocols

- Routing messages can be viewed as configuration messages:
  - Forging wrong routing messages can highjack some traffics
  - announcing default route may block communications



IPv6 Routing
Concepts & Generality
IGP
EGP
Usage

Securing Routing Protocol

- include a secret password on each message
  - protect from configuration mistakes
  - weak security, since packets can be eavesdropped and secret learnt
- use cryptographic
  - RIPng uses built-in IPsec extentions
  - RIPv2 uses its own mechanisms
    - Originally based on MD5 algorithm (RFC 2082)
    - Obsoleted by RFC 4822, recommending SHA1 algorithm.

IPv6 Routing
Concepts & Generality
IGP
EGP
Usage

OSPFv3

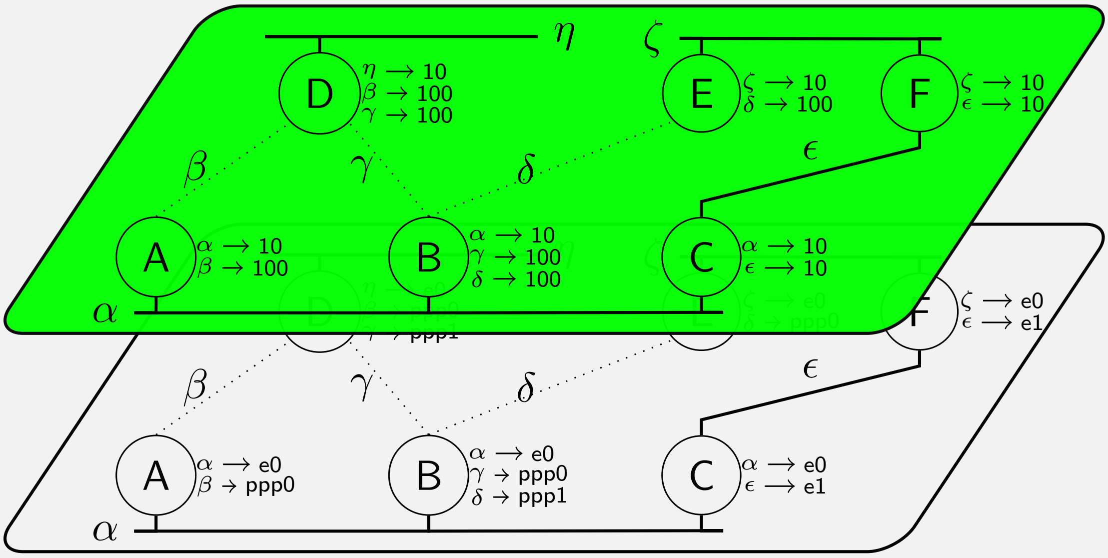
Link State Algorithm

IPv6 Routing
Concepts & Generality
IGP
EGP
Usage

- Distance Vector leads to a shortsighted view of network topology
  - each router processes routing messages
  - split horizon help for the first next hop
  - periodically flush all routing information
  - can be compared to caretaker exchanging gossip!
- Link State Protocols are divided in several states
  - learn network topology
  - exchange/flood local configuration
  - compute shortest path to a destination prefix
  - fill the FIB with Next Hop
  - less traffic, incremental updates
  - More a database synchronization algorithm than a routing protocol
- Two implementations
  - OSPF : v2 for IPv4 and v3 for IPv6
  - IS-IS : IP agnostic

Example: Database initialization

A cost is associated to each interface (for example related to the speed)



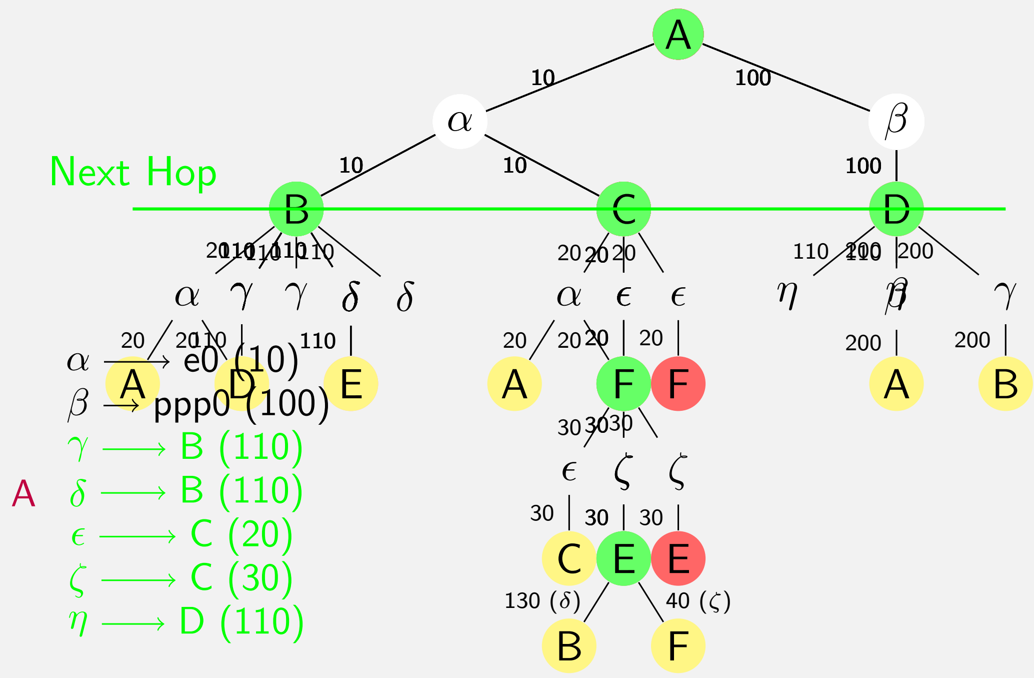
Example: Database state after synchronization

- After a flooding period (explained latter) all routers have exchanged their information
- Each router has all entries in its database

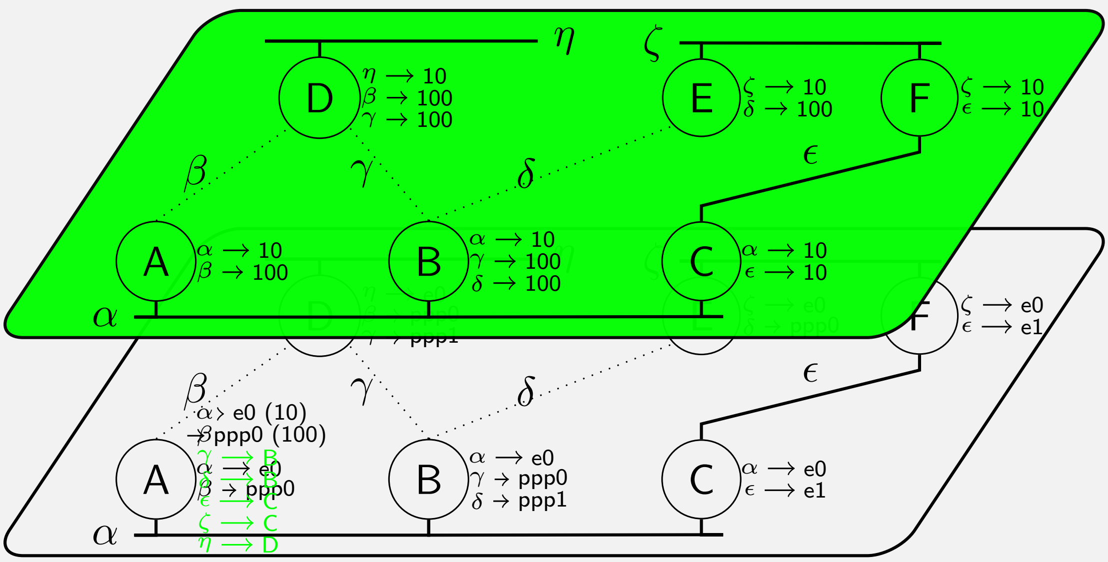
A: $\alpha \rightarrow 10$, $\beta \rightarrow 100$
B: $\alpha \rightarrow 10$, $\gamma \rightarrow 100$, $\delta \rightarrow 100$
C: $\alpha \rightarrow 10$, $\epsilon \rightarrow 10$
D: $\eta \rightarrow 10$, $\beta \rightarrow 100$, $\gamma \rightarrow 100$
E: $\zeta \rightarrow 10$, $\delta \rightarrow 100$
F: $\zeta \rightarrow 10$, $\epsilon \rightarrow 10$

- Entries are called Link State
- These entries gives a full knowledge of the network topology
- can be viewed as a graph with nodes (router) and vertex (prefixes)
- find the shortest paths from the root (router doing computation) to prefixes
  - based on Dijkstra's algorithm
  - explore the graph always using the shortest path
  - complexity is $O(N^2)$

Example: Shortest Path First Algorithm

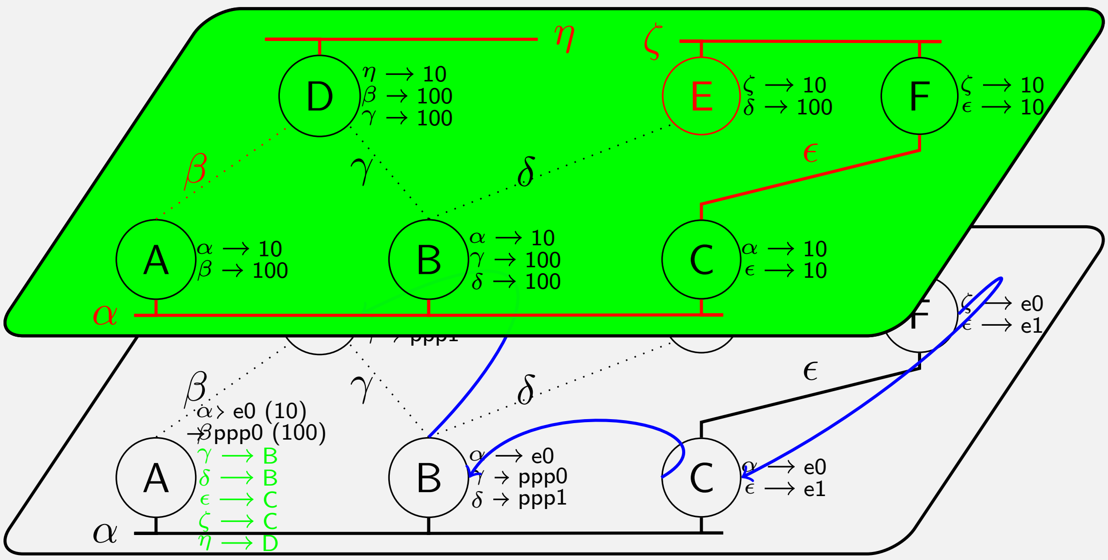


Example: FIB entries



Warning: SPT is not forwarding path

Path is different, but cost is the same



reducing SPF computation

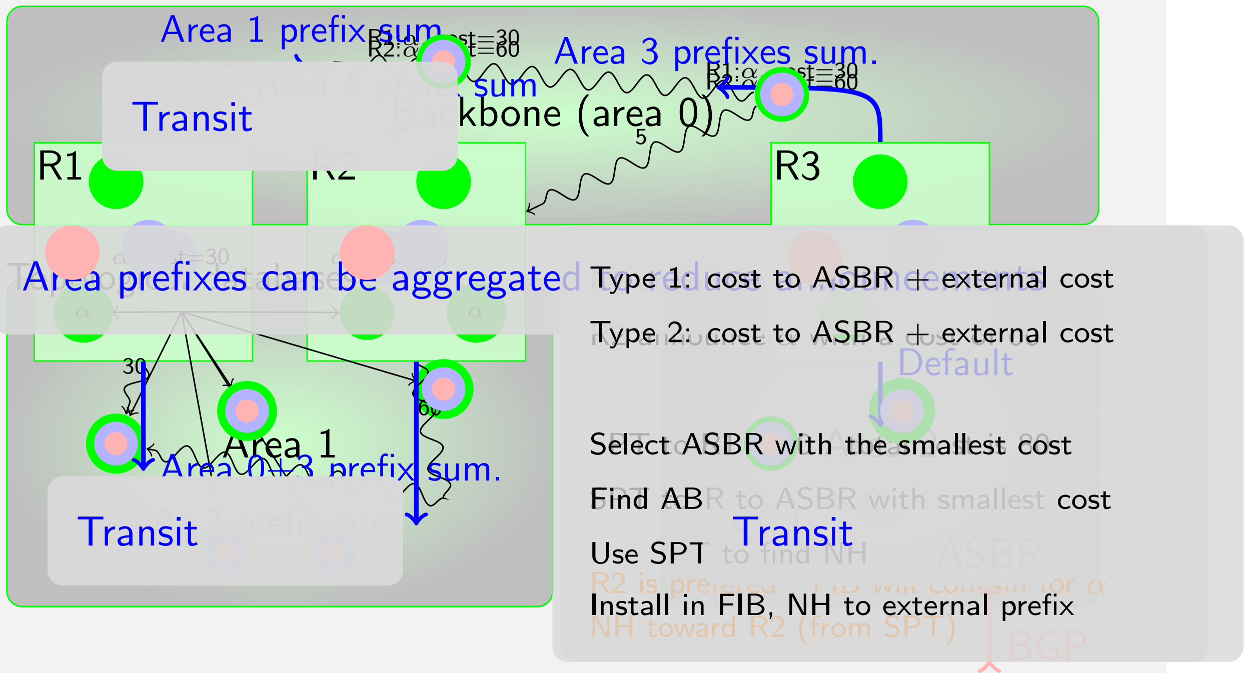
- SPF algorithm is complex ($O(N^2)$)
- every-time a router detect a topology change:
  - change is flooded to all routers
  - All router must rerun SPF algorithm

Definition

A network can be divided into several areas. All OSPF network contains

- A mandatory connexe backbone (Area 0)
- Optional areas, directly connected to the backbone through ABR (Area Border Router)
- ABR belongs to backbone and one (several) area(s)
- ABR sends summary (prefix+cost) of each area to the other ones.

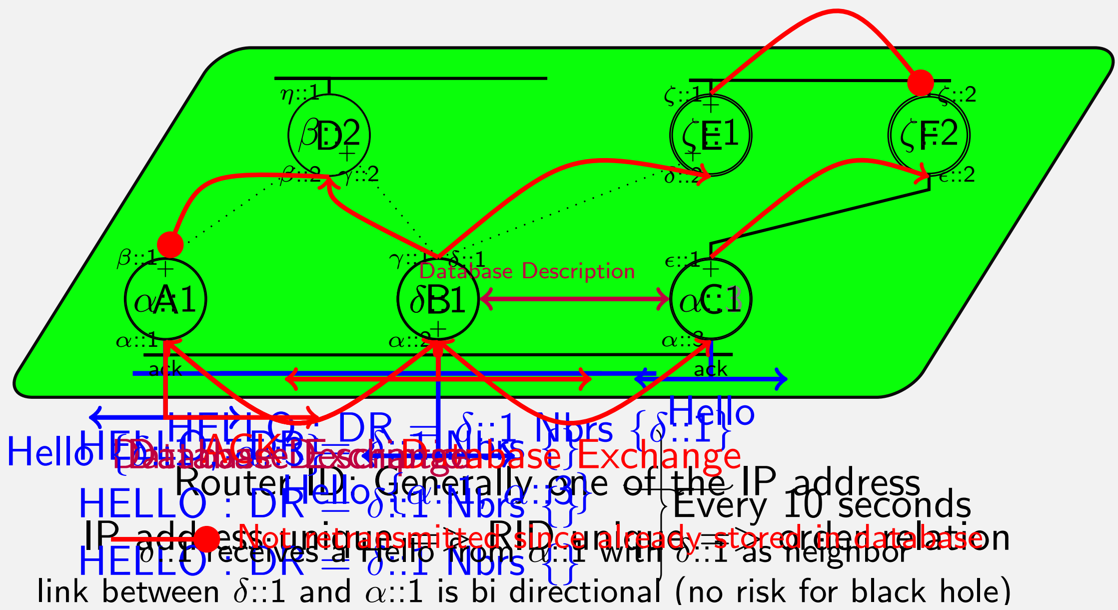
Division into Areas



Database synchronization

- OSPF maintain several kind of record (topological, summary, external prefixes, ASBR,...);
- All routers must share the same information
- Incremental update to reduce the bandwidth
  - compare to RIP, dumping database every 30 seconds
- exchange must be reliable in OSPF
- done is several steps and several protocols:
  - HELLO protocol: discover peers, elect a Designated Router
  - Database description: Describe information contained in each router's database
  - Database exchange: request and transfer information missing or more recent

Example: Flooding



Link State Announcement Format (RFC 2740)

0................8...............16................24...............32

LS Age	LS Type
LS ID	
Advertising Router	
LS sequence number	
LS checksum	Length

LS Age
The time in seconds since the LSA was originated.
After 40 min a new LS must be generated even if there is no change.

LS Type
The LS type field indicates the function performed by the LSA. The high-order three bits of LS type encode generic properties of the LSA, while the remainder (called LSA function code) indicate the LSA's specific functionality.

LS ID
Together with LS type and Advertising Router, uniquely identifies the LSA in the link-state database.

LS Type

0 ... 16

U	S2	S1	LSA Function Code

U-bit	LSA Handling
0	Treat the LSA as if it had link-local flooding scope
1	Store and flood the LSA, as if type understood

S2	S1	Flooding Scope
0	0	Link-Local Scoping. Flooded only on link it is originated on.
0	1	Area Scoping. Flooded to all routers in the originating area
1	0	AS Scoping. Flooded to all routers in the AS
1	1	Reserved

LSA function code

LSA function code	LS Type	Description
1	0x2001	Router-LSA
2	0x2002	Network-LSA
3	0x2003	Inter-Area-Prefix-LSA
4	0x2004	Inter-Area-Router-LSA
5	0x4005	AS-External-LSA
6	0x2006	Group-membership-LSA
7	0x2007	Type-7-LSA
8	0x0008	Link-LSA
9	0x2009	Intra-Area-Prefix-LSA

Link State Announcement Format

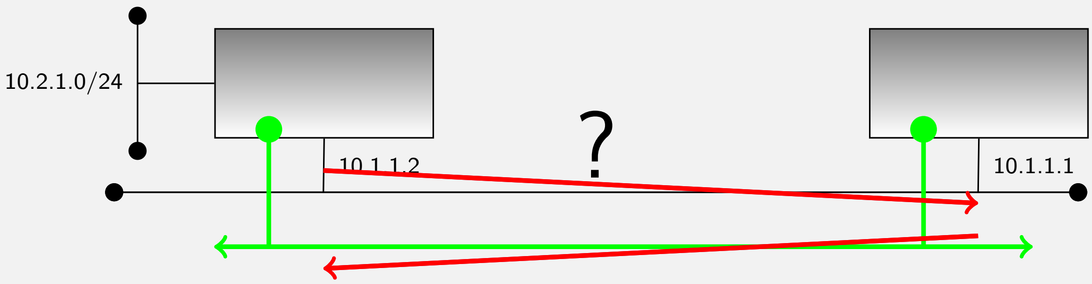
LS Age		001	1
Link State ID 2001:660:7301:3728::6265			
Advertising Router 2001:660:7301:3728::6265			
LS sequence number			
LS checksum		Length	
0	wveb	Options	
Type	0	metric	
Interface ID			
Neighbor Interface ID			
Neighbor Router ID			
...			
Type	0	metric	
Interface ID			
Neighbor Interface ID			
Neighbor Router ID			
...			

Router-LSAs have LS type equal to 0x2001. Each router in an area originates one or more router-LSAs distinguished by their Link-State IDs. Router-LSAs originated by the router describe the state and cost of the router's interfaces to the area.

Example


```
{ S 80000002 age 5 rtr 10.1.1.2 }
{ E S 80000002 age 3:09 rtr 10.1.1.1 }
{ E S 80000001 age 2:49 sum ... }
{ E S 80000003 age 2:44 sum ... }
{ E S 80000001 age 2:59 abr... }
```

```
{ E S 80000002 age 3:09 rtr 10.1.1.1 }
{ E S 80000001 age 2:49 sum 10.1.2.0 abr 10.1.1.1 }
{ E S 80000003 age 2:44 sum 10.1.100.0 abr 10.1.1.1 }
{ E S 80000001 age 2:59 abr 10.1.1.1 rtr 10.1.1.1 }
{ S 80000002 age 5 rtr 10.1.1.2 }
```




```
10.1.1.1 > 224.0.0.5:  OSPFv2-hello 44:
          area 0.0.0.1 E mask 255.255.255.0 int 10 pri 5 dead 40 dr 10.1.1.1

          nbrs ;
```

- One router (IP address: 10.1.1.1) on 10.1.1.0/24 link
- network is in area 1
- router knows no neighbor
- Hello period is 10s
- After 40s a router is supposed dead
- priority 5 is used when electing DR

10.1.1.1 > 224.0.0.5: OSPFv2-hello 44:
area 0.0.0.1 E mask 255.255.255.0 int 10 pri 5 dead 40 dr 10.1.1.1

IS-IS

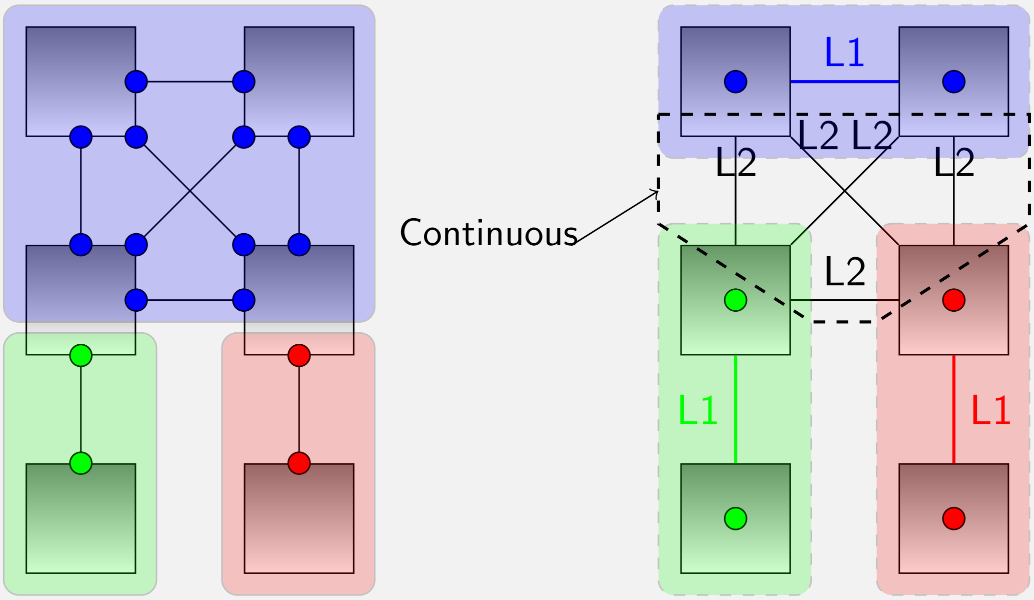
Intermediate System to Intermediate System

IPv6 Routing
Concepts & Generality
IGP
EGP
Usage

- Originally designed to route OSI Datagram protocol CLNP (Connection Less Network Protocol)
- Adapted to route both CLNP and IPv6 protocol.
- IS-IS don't use IP to carry routing messages
- Based on Shortest Path First algorithm to compute routes
  - OSPF derived from IS-IS, but was claim to be Open.
  - now-days IS-IS is also a IETF working group
- IS-IS is still widely used in provider backbones

Areas versus Levels

IPv6 Routing
Concepts & Generality
IGP
EGP
Usage



◯area assigned to interfaces

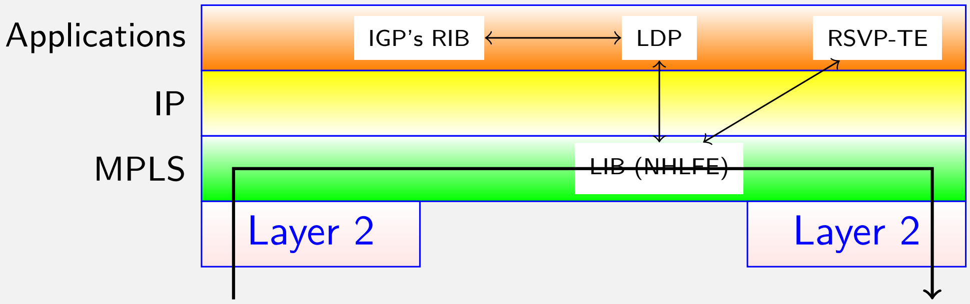
◯area value included into address

IPv6 Routing
Concepts & Generality
IGP
EGP
Usage

MPLS

MPLS: Multi Protocol Label Switching

- Goal:
  - Use routing protocols as a *signaling plane*
  - Switch the data plane
    - Use of virtual path called Label Switched Path
    - A switching table: Next Hop Label Forwarding Element
  - Offer more flexibility for traffic engineering
- LDP Label Distribution Protocol convert IGP routing into LSP
- RSVP-TE can be used to by-pass IGP routing and assign bandwidth

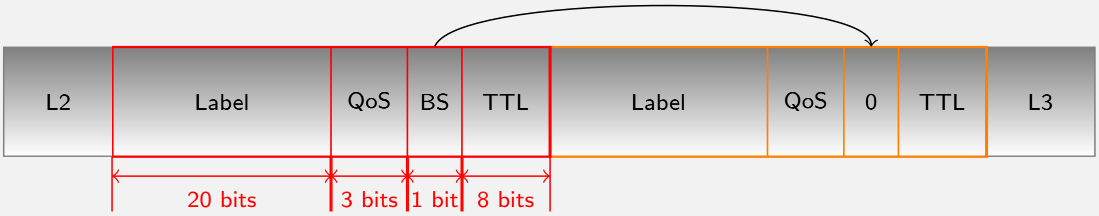


MPLS: Multi Protocol Label Switching

IPv6 Routing
Concepts & Generality
IGP
EGP
Usage

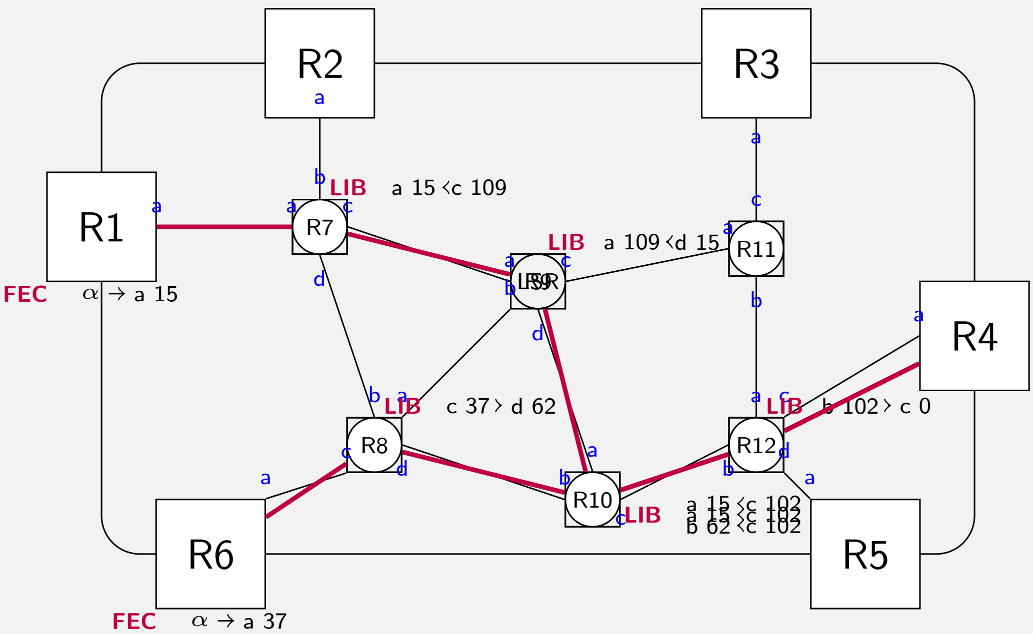
- Multi Protocol :
  - Layer 2
    - Any layer 2 with VPs like ATM, Frame Relay
    - Point-to-Point Networks
    - Ethernet
    - Extension for Optical Networks (GMPLS)
  - Layer 3
    - Forwarding is under Layer 3
    - IPv4 (with public or private addresses), IPv6
    - Ethernet / Bridging
- Done by hardware: more efficient than an IP tunnel.

Label Format [RFC3032]



- ATM and Frame relay: Label in copied in the VPI/VCI or DLCI
- Ethernet: Ethertype 8847 (Unicast), 8848 (Multicast)
- PPP: protocol : 281 (Unicast), 282 (Multicast)
- Label can be :
  - per platform: unique for each LSR
  - per interface: unique for one interface (may be reused elsewhere)
  - Special labels :
    - 0: IPv4 Explicit NULL Label = POP and forwarding
    - 2: IPv6 Explicit NULL Label = POP and forwarding
    - 3: Implicit NULL Label = POP
    - 14: OAM Alert Label [RFC3429]

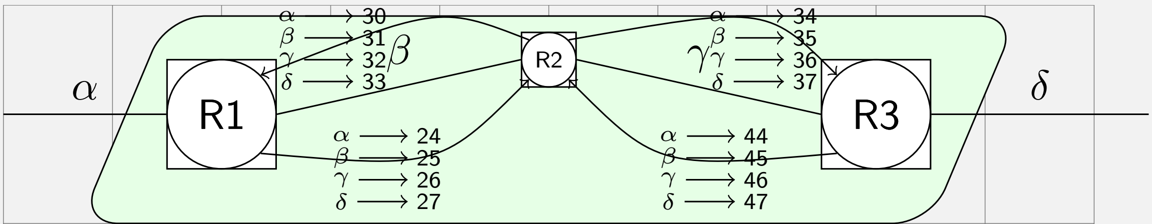
switching example



How to build FEC and LIB ?

- Manually
- Using Label Distribution Protocol
  - works with an IGP (prefix discovery and route selection)
  - two label distribution modes :
    - independent : each LSR works independently on FEC
    - ordered : downstream LSR has to establish the FEC first
  - two label retention modes :
    - conservative : keep in memory just useful labels
    - liberal : memorize label not useful for the LSP (faster reroute)
- RSVP-TE
  - independent of IGP: allow the provider to force paths and reserve resources
  - allow rerouting in case of link failure
- MP-BGP
  - Dissociate internal routing and external routing
  - VPN, IPv6 over IPv4, IPv4 over IPv6

Example : LDP usage



FIB	R1	α → e0 β → e1	R2	α → e0 γ → e1	R3	γ → e0 δ → e2
FEC Labels	R1	γ → e1 32 δ → e1 33	R2 e0	α → 30 β → 31 γ → 32 δ → 33 e0 30 → e0 34 e0 31 → e0 25	R3 e0	α → 40 β → 41 β → 42 δ → 43 e0 34 e0 35
LIB	R1 e1	α → 24 e1 24 → pop e1 25 → pop δ → 27	R2 e1	e0 32 → e1 46 e0 33 → e1 47 e1 34 → e0 24 e1 35 → e0 25 e1 36 → e1 46 e1 37 → e1 47	R3	e0 46 → pop e0 47 → pop

BGP

What protocol for small and big network ?

How to manage a dual stack network ?