4over6 Transit using Encapsulation and BGP-MP Extension

----4over6 proposal for Mesh Problem

Jianping Wu, Yong Cui, Xing Li

Tsinghua University (CERNET)

Feb 23, 2006

Contact: yong@csnet1.cs.tsinghua.edu.cn

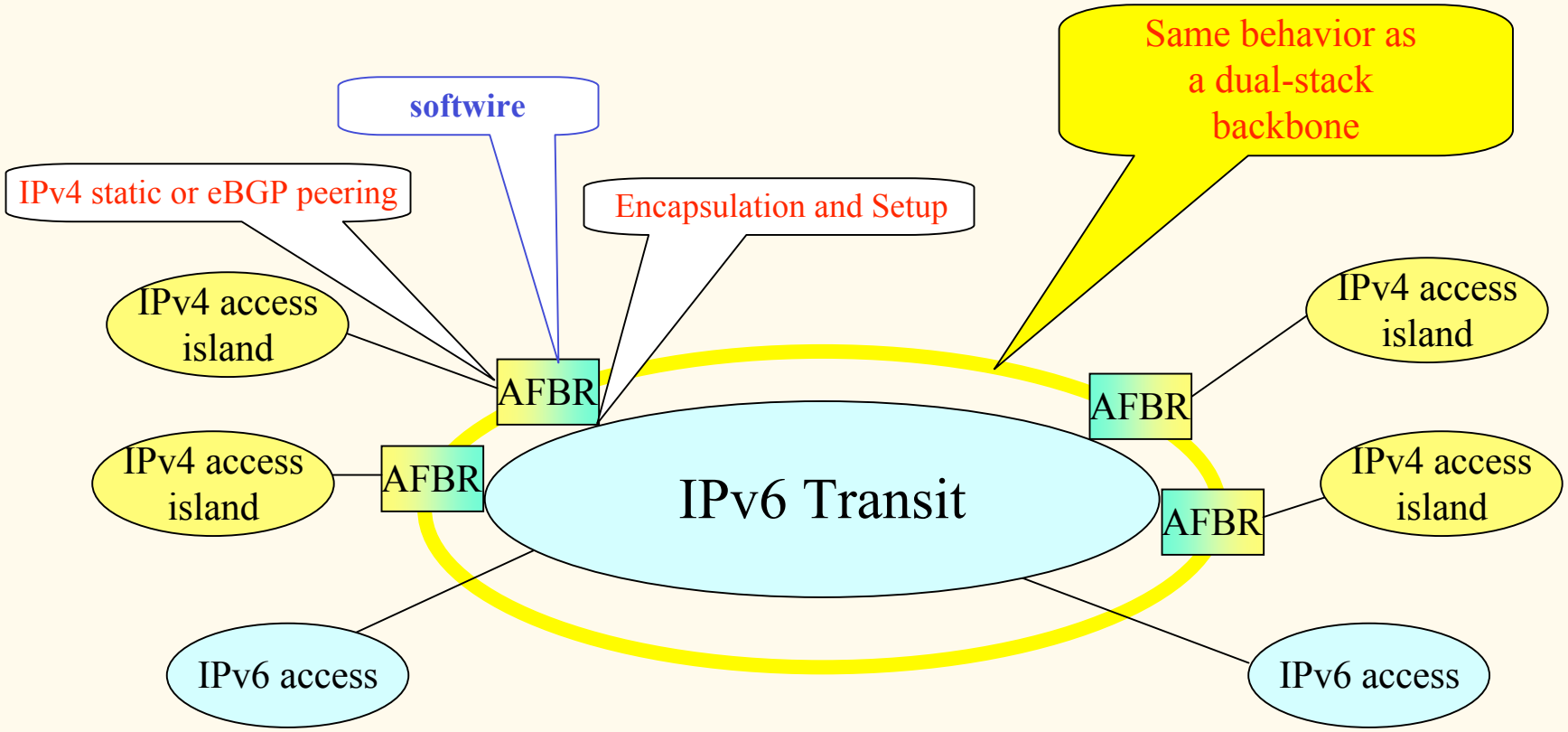
Content

- Mesh Problem
- 4over6 proposal
  - 4over6 framework
  - Packet forwarding
  - BGP-MP 4over6 extension
    - Protocol definition
    - AFBR routing behavior
  - Example of 4over6
  - Implementation framework
  - Criteria discussions
- Extension to IPv6 over IPv4
- Conclusion

Mesh Problem

- Description
  - Core network problem
  - ISP initiated
  - complex routing topology
- Applicability
  - ISPs (or large enterprise networks acting as ISP for their internal resources) establish connectivity to 'islands' of networks of one address family type across a transit core of a differing address family type.

4over6 Framework for Mesh Problem

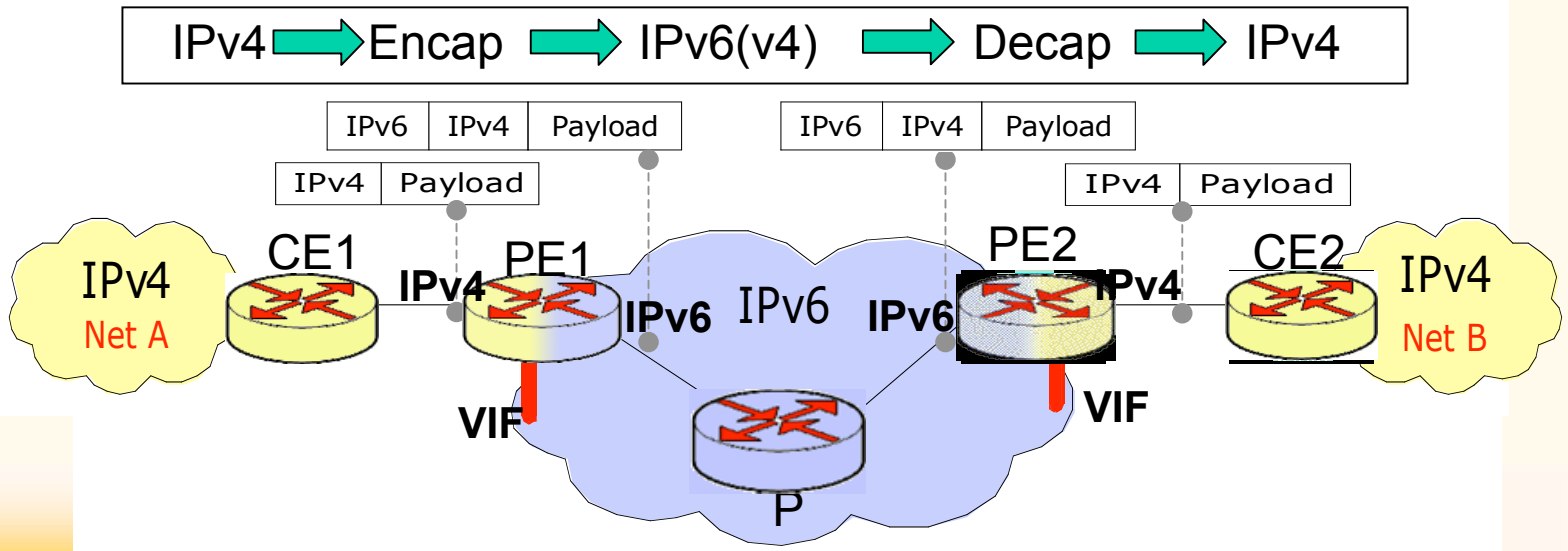


Framework Functionalities

- Mesh problem statement
  - Core (consisting of P routers) provides transit in one address family
  - Access networks are in another address family
  - Therefore, PE routers are dual-stack and provide Functionalities of softwires
- Proposed solution for mesh problem
  - Data plane of PE routers
    - Encapsulation (GRE, IP-IP, IP over UDP over IP, etc.)
  - Control plane of PE routers
    - End point discovery

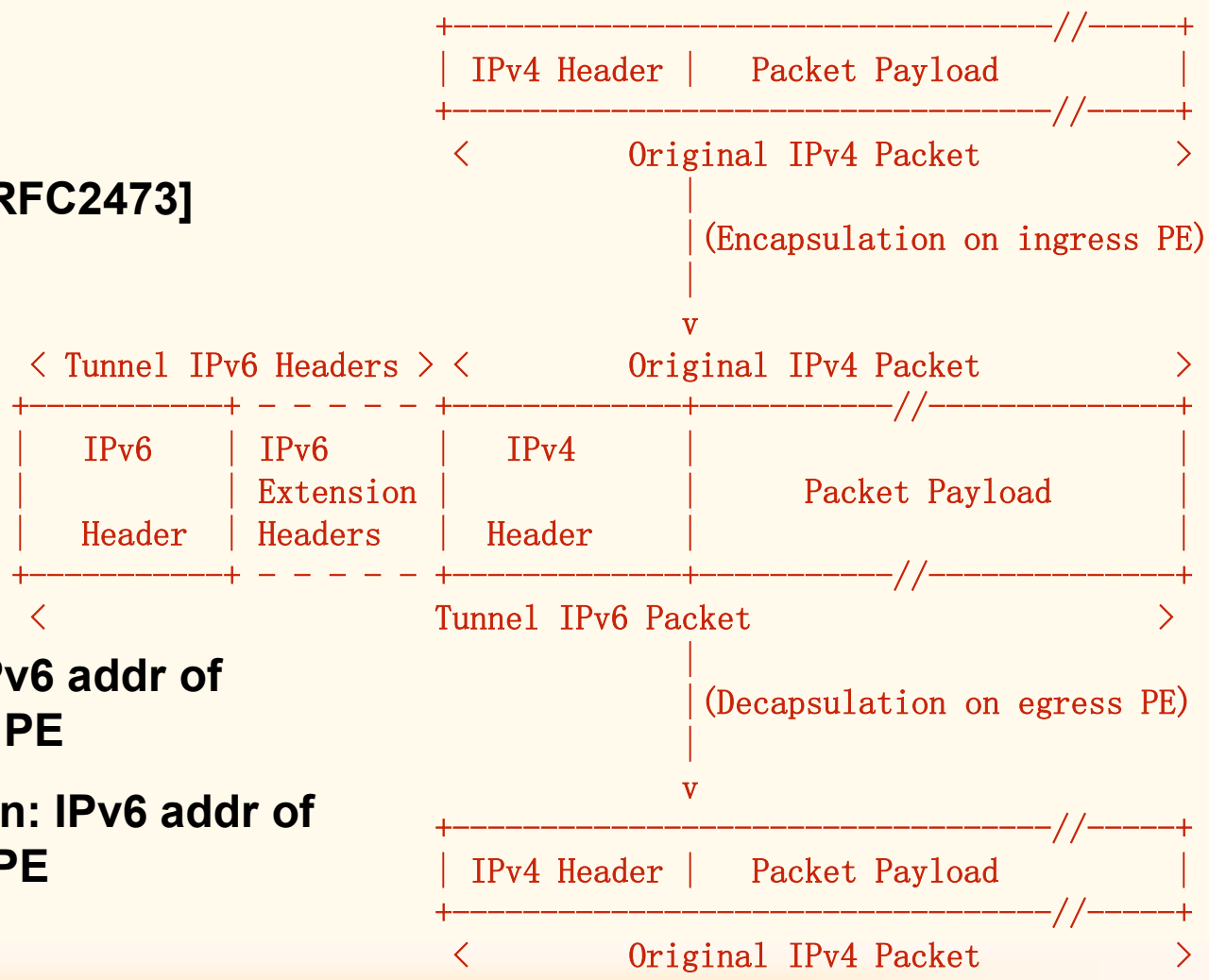
Packet Forwarding

- 4over6 packet forwarding
  - Encapsulation on ingress PE
  - Transmission of encapsulated packet in IPv6 Core via P routers
  - Decapsulation on egress PE back to IPv4 edge
- Reuse existing encapsulation technologies
  - GRE [2784], IP over IP [2473], IP over UDP over IP[RFC 3142]
  - Emerging technologies
- 4over6 VIF
  - 4over6 virtual interface on PE with both IPv4/v6 addresses



Example of IPv4 over IPv6 Encapsulation and Decapsulation

By reusing [RFC2473]



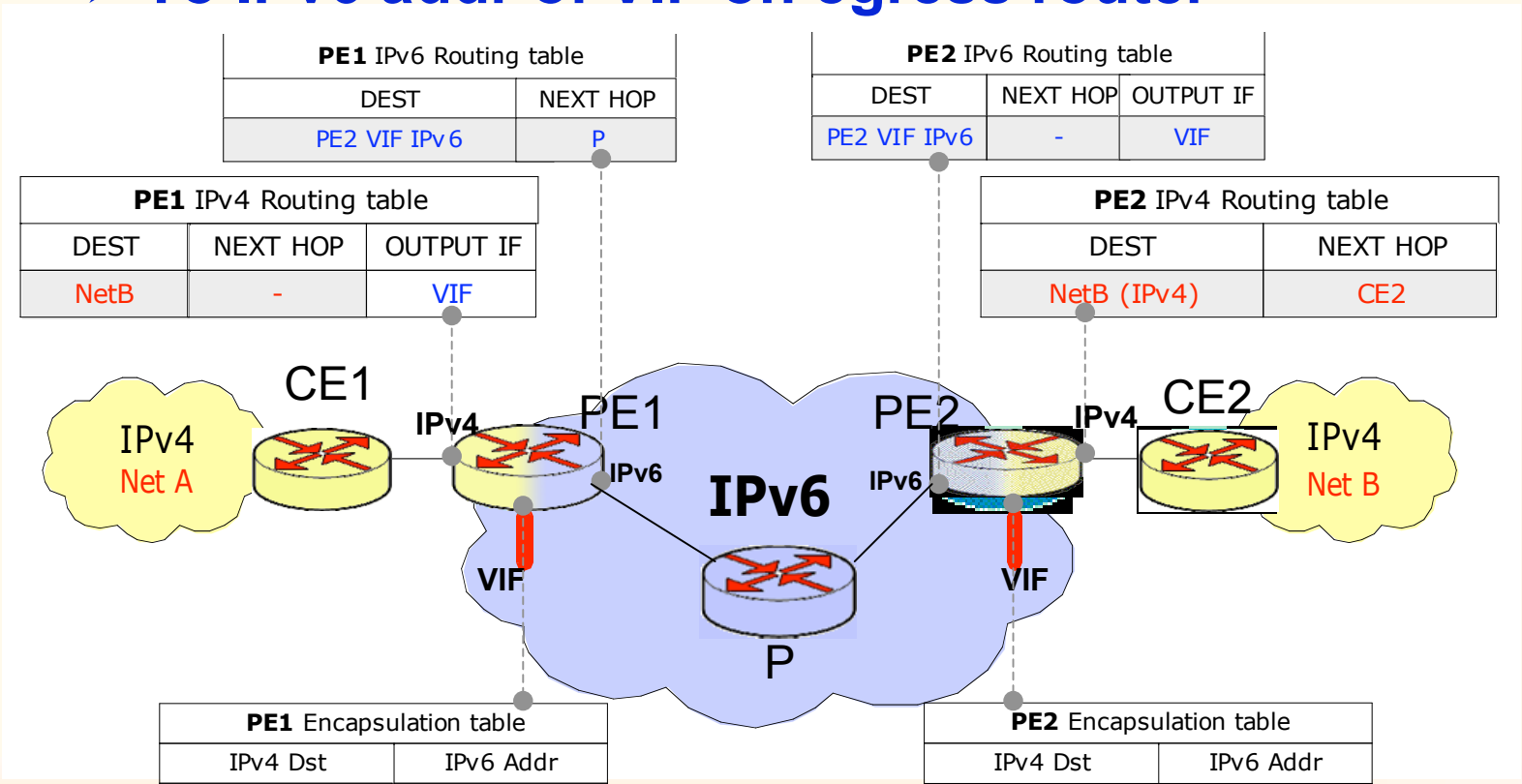
IPv6 source: IPv6 addr of VIF on ingress PE

IPv6 destination: IPv6 addr of VIF on egress PE

Encapsulation table

❑ Mapping

- ➢ From IPv4 dst edge networks with prefixes
- ➢ To IPv6 addr of VIF on egress router



Encapsulation table (cont.)

- Characteristics of encapsulation table
  - Contains the mapping from IPv4 to IPv6
  - Multiple dest to One VIF
  - Routing info about egress PE and dst networks
  - Use for encapsulation on ingress PE(AFBR)
  - Currently no automatic scheme for endpoint discovery
- How to construct Enc Tab?
  - Transmit Network Reachability info from egress PE to ingress PE
- Why use BGP?
  - Have similar extensions with BGP-MP
  - Setup a peering relationship between PEs

BGP-MP 4over6 Protocol Definition

- **BGP-MP Objective**
  - Peering between AFBR (PE)
  - Encapsulation table
    - From IPv4 edge network addresses with prefix
    - To egress VIF IPv6 address
- **BGP-MP 4over6 extension**
  - OPEN message indicates the capability of BGP entity by AFI and SAFI
  - BGP UPDATE Message includes routing info (Next Hop, NLRI) with AFI and SAFI

Address Family Identifier

Number	Description		Reference
0	Reserved		
1	IP (IP version 4)		
2	IP6 (IP version 6)		
3	NSAP		
4	HDLC (8-bit multidrop)		
5	BBN 1822		
6	802 (includes all 802 media plus Ethernet "canonical format")		
7	E.163		
8	E.164 (SMDS, Frame Relay, ATM)		
9	F.69 (Telex)		
10	X.121 (X.25, Frame Relay)		
11	IPX		
12	Appletalk		
13	Decnet IV		
14	Banyan Vines		
15	E.164 with NSAP format subaddress	[UNI-3.1]	[Malis]
16	DNS (Domain Name System)		
17	Distinguished Name		[Lynn]
18	AS Number		[Lynn]
19	XTP on IP version 4		[Saul]
20	XTP on IP version 6		[Saul]
21	XTP native mode XTP		[Saul]
22	Fibre Channel World-Wide Port Name		[Bakke]
23	Fibre Channel World-Wide Node Name		[Bakke]
24	GWID		[Hegde]
65535	Reserved		

Use: IP=1 for IPv4 edge networks

SAFI

Value	Description	Reference
0	Reserved	
1	Network Layer Reachability Information used for unicast forwarding	[RFC2858]
2	Network Layer Reachability Information used for mulitcast forwarding	[RFC2858]
3	Network Layer Reachability Information used for both unicast and multicast forwarding	[RFC2858]
4	Network Layer Reachability Information (NLRI) with MPLS Labels	[RFC3107]
5-63	Unassigned	
64	Tunnel SAFI	[Nalawade]
65	Virtual Private LAN Service (VPLS)	[Kompella]
66	BGP MDT SAFI	[Nalawade]
67-127	Unassigned	
128	MPLS-labeled VPN address	
129-255	Private Use	

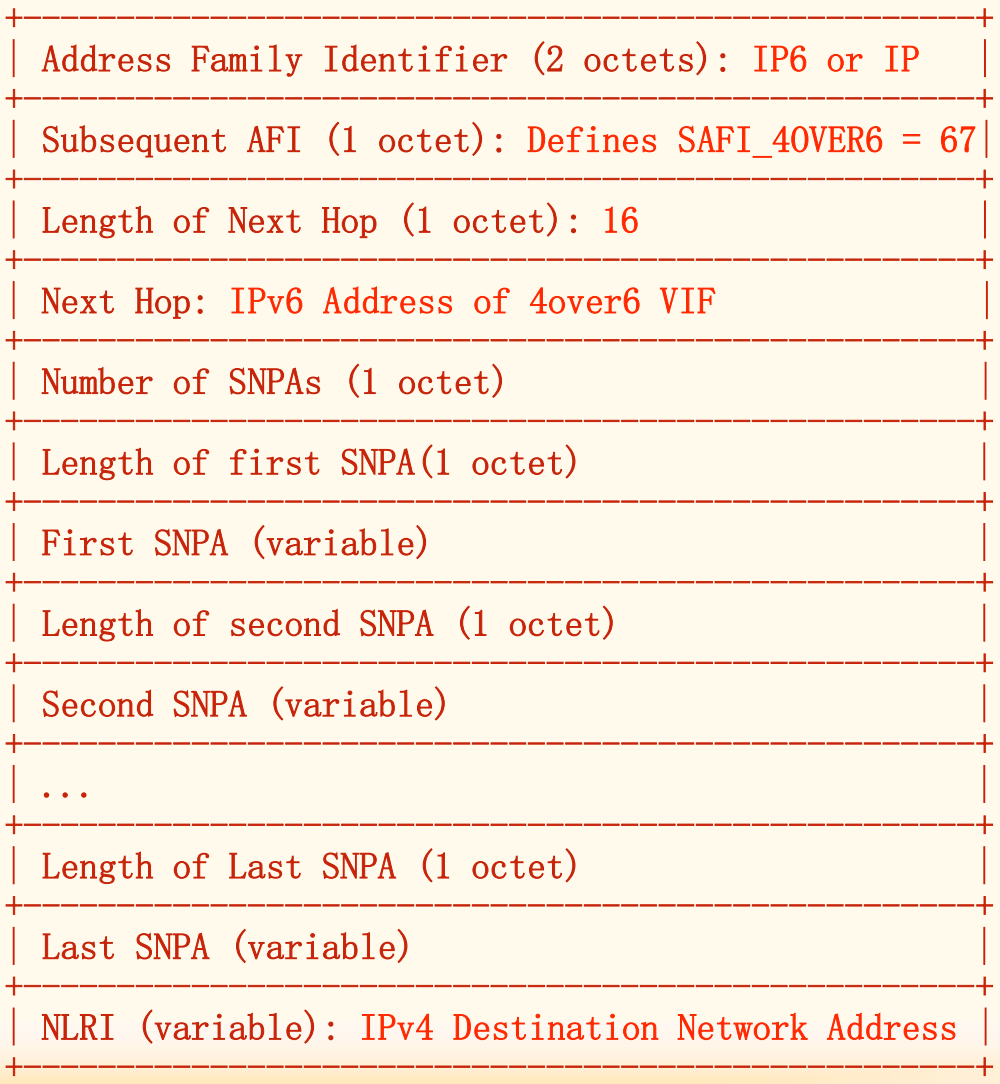
Define: SAFI_4over6 = 67 (FCFS for 64-128)

Indicate 4over6 capability

BGP-MP 4over6 Protocol Definition

UPDATE Message

IPv4 over IPv6



AFI_IP=1

SAFI_4OVER6 = 67

Length of IPv6

IPv6 VIF on PE

Dst IPv4 network addr
With prefix length

AFBR Protocol Behavior

- **Behavior overview**
  - **On 4over6 PE routers**
  - **Routing between PE <-> CE**
    - Make PE learn edge routing info of local edge network
    - RIP, OSPF, I-BGP, E-BGP, static, etc.
  - **Routing between PE <-> PE**
    - I-BGP peering with each other
    - Use BGP-MP 4over6 extension
  - **4over6 virtual interface on PE**
    - Configure addresses in both IPv4/v6

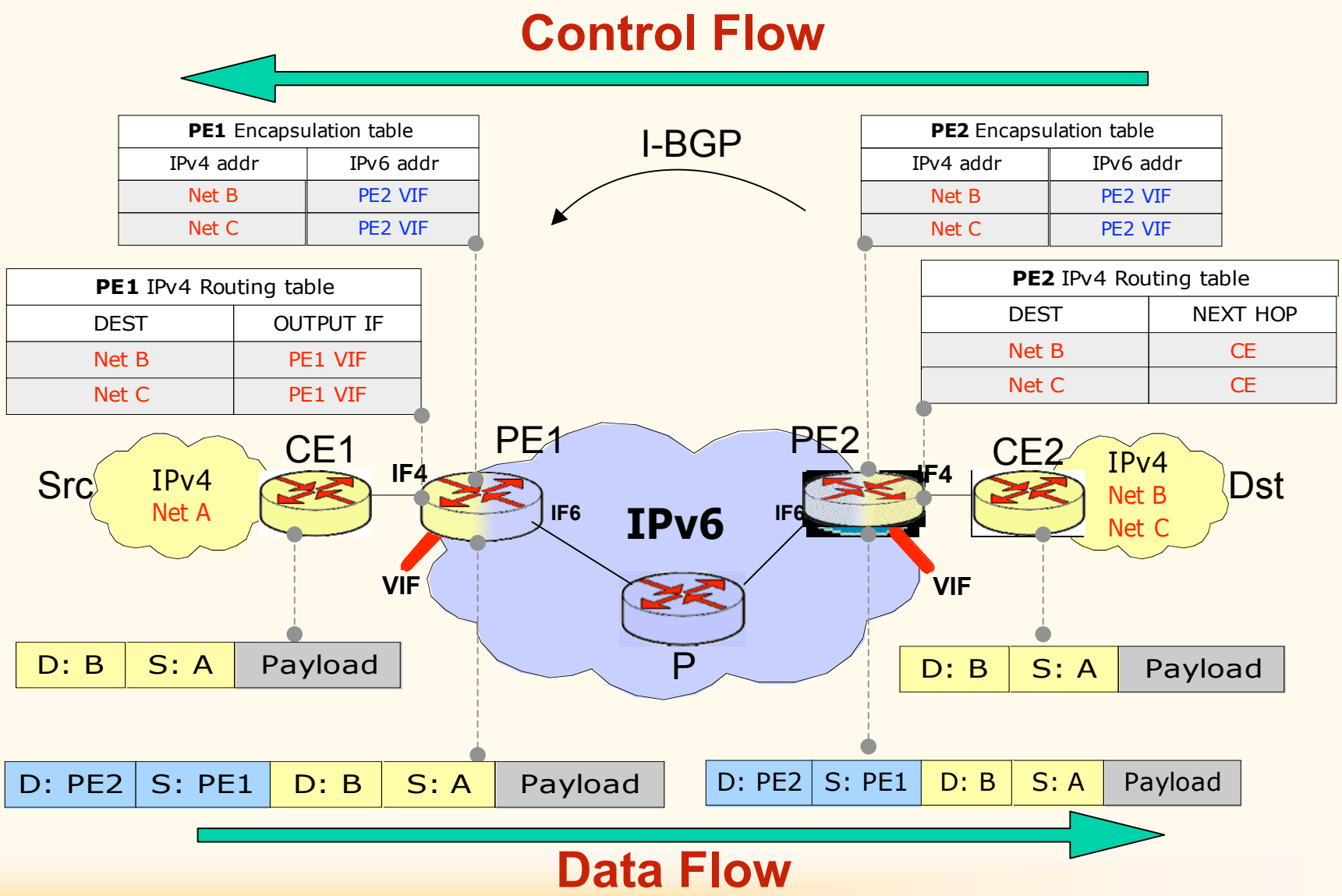
Protocol Behavior of BGP-MP 4over6 Extension

- For routing info received from CE (static)
  - Construct the encapsulation table
    - IPv4 Network addr with prefix
      - Should be the original edge destination
    - Corresponding IPv6 addr
      - should be the address of PE's 4over6 VIF
  - 4over6 I-BGP entity sends to its peers on core network
    - Taking AFI as edge IPv4 AFI
    - Taking SAFI as SAFI_4OVER6 = 67
    - Destination (in IPv4 edge AF)
      - Should be the original edge destination with prefix
    - Nexthop (in IPv6 core AF)
      - should be the address of its 4over6 VIF

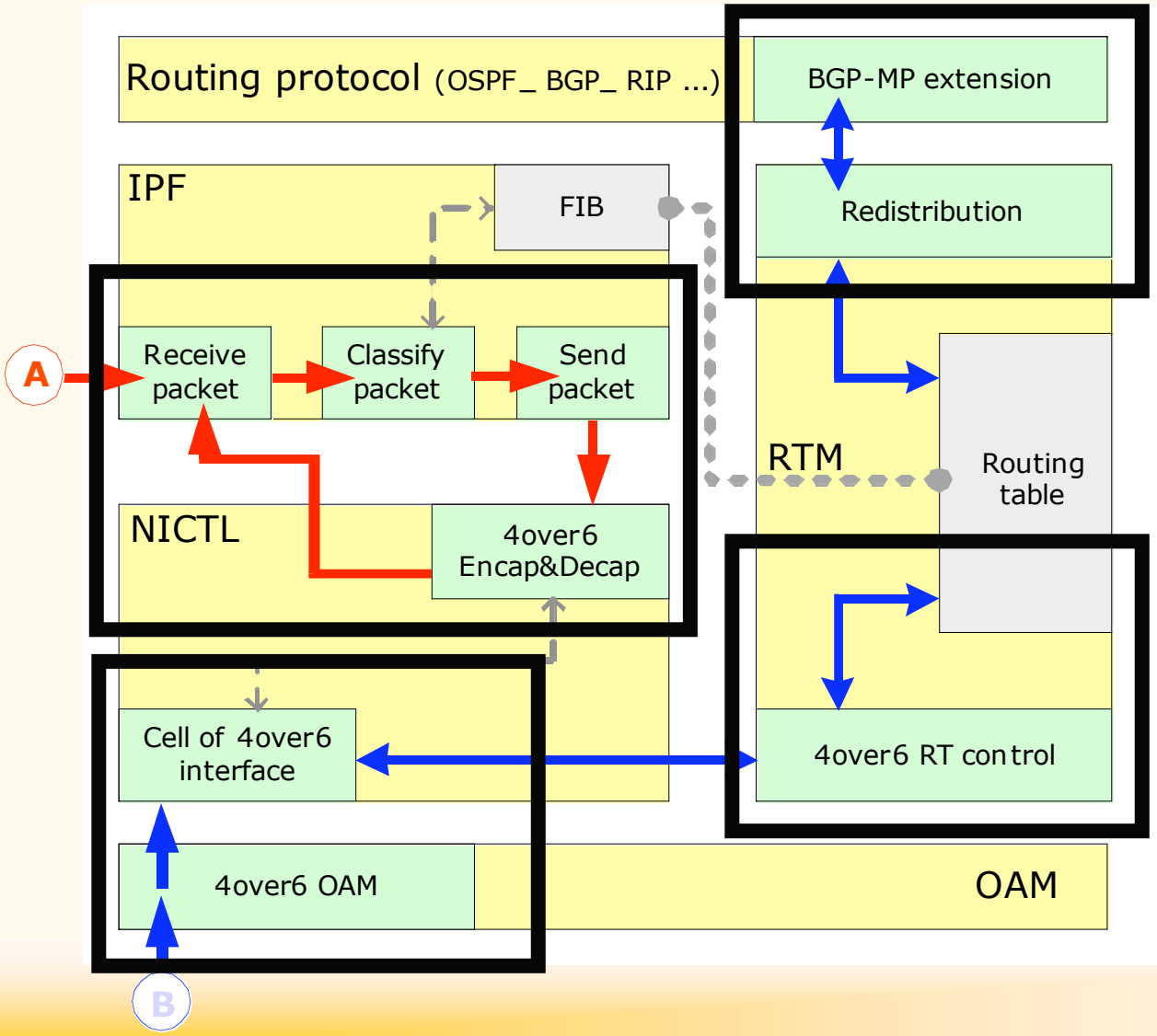
Protocol Behavior of BGP-MP 4over6 Extension

- For routing info received from other PE
  - Confirm the routing type
    - Edge AFI (IPv4) and SAFI_4OVER6
    - Destination is in edge AF format (IPv4)
    - Next hop is in core AF format (IPv6)
  - Construct the encapsulation table
    - IPv4 network addr/prefix is NLRI in UPDATE message
    - Mapped IPv6 addr is NEXTHOP in UPDATE message
  - Set IPv4 routing table
    - Keep the original destination in Edge AF (IPv4)
    - Take OUTPUT IF as 4over6 VIF

Example of 4over6



Implementation Framework



****

1) 4over6 OAM
2) 4over6 RT control
3) BGP-MP Ext
4) 4over6 VIF

****

Technical Criteria - Scalability

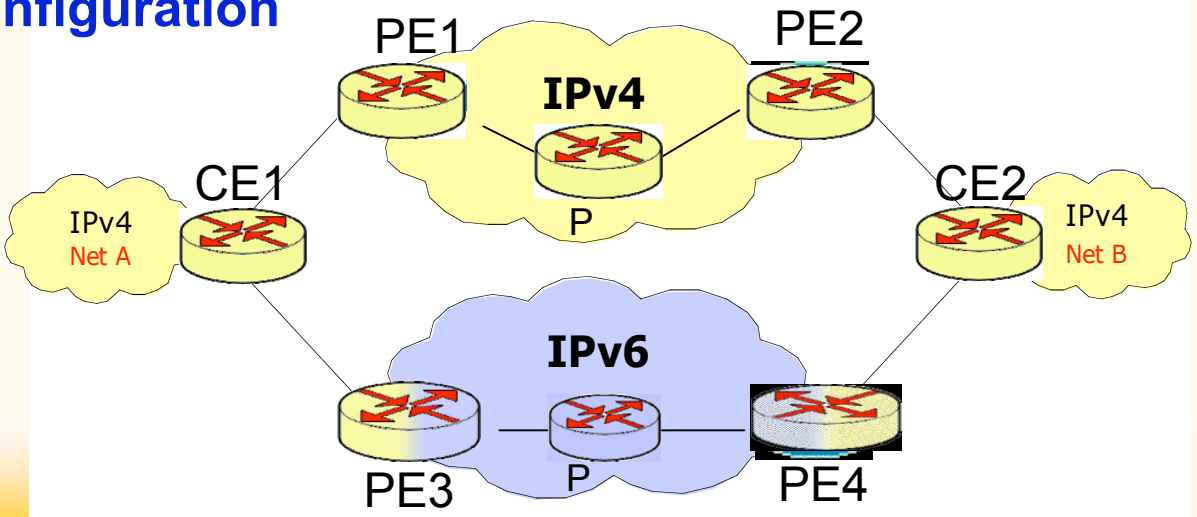
- Advantage
  - Single stack P routers construct a transit “dual-stack” core
  - Only PE needs to be extended
  - Only PE maintains the edge routing info
  - No per flow state or resource allocation
- Disadvantage
  - Similar to ASBR
  - 4over6 AFBR routers need full mesh connection for I-BGP
  - Router Reflector may be used
- Scalability
  - Number of AFBRs
    - Same as ASBR
    - Unlimited in theory with RR
    - Dozens of AFBRs without RR
  - Routing table size
    - Same on P routers, additional 4over6 routes on AFBR for reachable access networks
  - Number of network peers
    - Thousand access networks

Technical Criteria - Security

- **Security**
  - **No per flow state maintenance to alleviate DDoS attacks**
  - **Integration with deployed solutions**
    - Encapsulation techniques are widely implemented
    - BGP-MP is widely deployed
  - **Control session**
    - Support IPSec between BGP-MP peers
  - **Encrypted data**
    - Support IPSec in tunnel data transmission

Technical Criteria - Multihoming

- Multihoming problem
  - Edge networks access multiple backbones especially in different AFs
- CE select PE on particular AF
  - Default routing or policy routing
  - Preference should be along the same AF
  - CEs don't learn routes from PE
- PE learn routes from CE
  - Only the routes to edge networks by routing protocol or configuration



Technical Criteria - Multicast

- PE support multicast in edge AF with CE
  - PIM-SM supports tunnel interfaces
    - RFC 2362: Hello Join/Prune Message with edge AF addr
  - Tunnel mechanisms can be applied to multicast
    - E.g. RFC 2473
  - Multicast duplication before encapsulation
    - PE1 receives a multicast packet, looks up the multicast forwarding table, and sends one copy of multicast packet to the virtual interface
    - Encapsulates the multicast packet in a unicast packet and sends it to PE2
  - Multicast duplication after decapsulation
    - PE2 decapsulate the received encapsulated packet
    - The original multicast packet is delivered to the multicast module in PE2
- P doesn't support multicast

Other Technical Criteria

- Support Mesh cases
  - Announce reachability of prefixes of one AF across a network of another AF
  - AFBRs perform dual-stack functionality
- Available Encapsulations
  - Support IPv4 over IPv6, e.g. GRE, IPIPv6, etc.
- OAM
  - Usage accounting
    - Need to be defined
  - End point failure detection
    - By BGP sessions
  - Path failure detection
    - By BGP UPDATE message
- Does solution enable L2 and L3 connectivity
  - Enable L3 connectivity

Other Technical Criteria (cont.)

- IANA considerations
  - New SAFI needs to be defined: SAFI_4OVER6=67
- Encapsulation type
  - Recognition is self contained in an encapsulated packet
  - Default preference to IPIPv6
- BGP convergence
  - No convergence issue within one AS
- BGP Reopen for 4over6 enable
  - Long-lived 4over6

Non-technical Criteria

0) Reused existing technology
   - Existing and future Encap & Decap
   - BGP-MP in RFC 2858

1) Is the solution documented (published)?
   - Submitted on Feb 22 as an individual draft

2) Are there any known issues in the solution (completeness)?
   - MIB, accounting, etc.

3) Has the solution been fully implemented (status idea)?
   - Yes, we have a prototype in the University Lab

Non-technical Criteria

4) Do two independent, commercially supported, inter-operable implementations of all the components of the underlying technology exist (interop)?
   - Bitway company will implement it in March
   - Looking for other commercial implementations
5) Have ISPs experimented with all the components of the solution successfully (deployment)?
   - CERNET2 will test the solution in March
   - CERNET2 will deploy the solution in June

Extension to IPv6 over IPv4

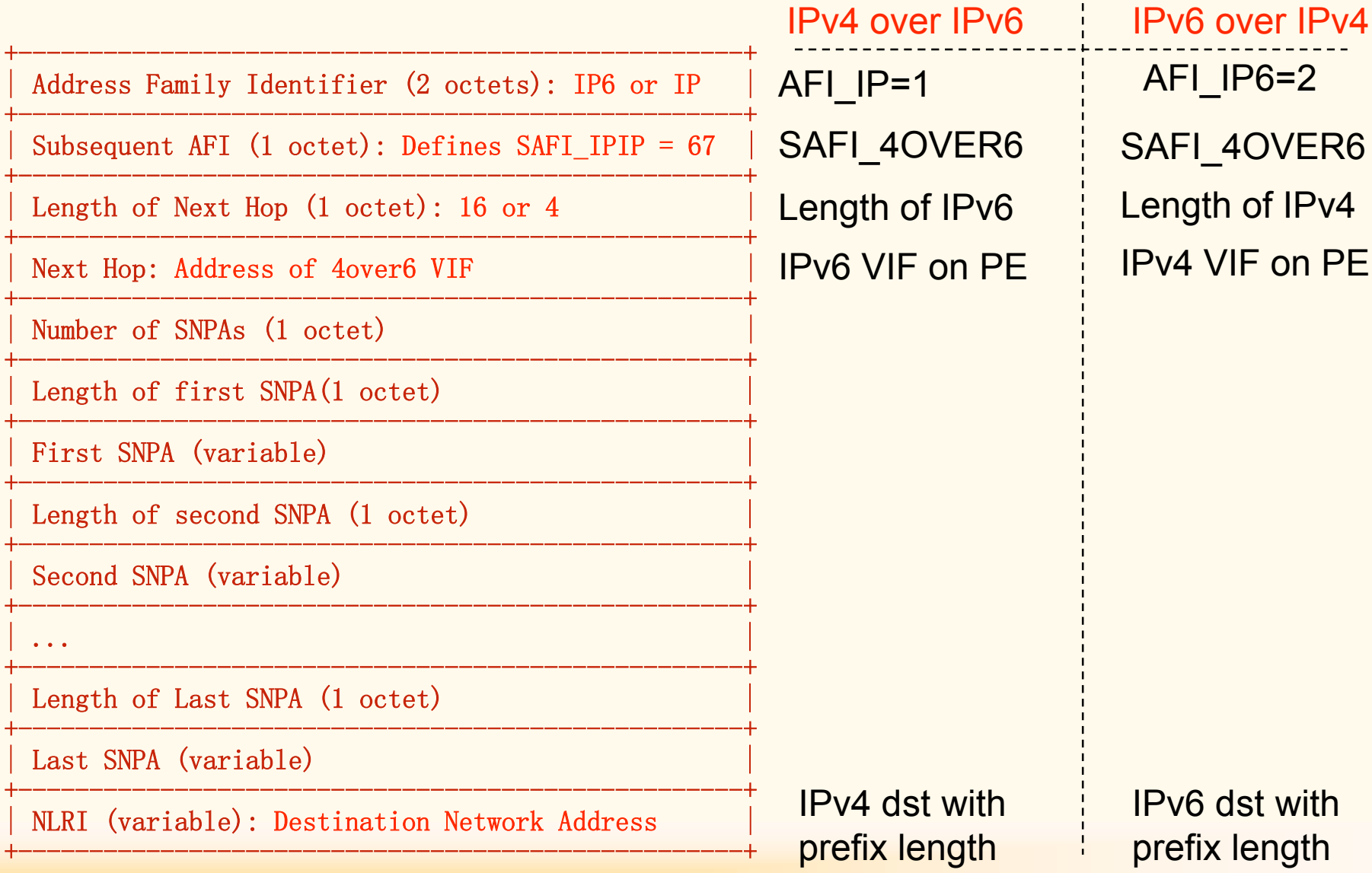
- Address format
  - 4over6 proposal uses IPv4/IPv6 address in an equal position rather than coding one addr to another
- Encapsulation table
  - IPv4 dst -> IPv6 4over6 VIF
  - IPv6 dst -> IPv4 4over6 VIF
- Encapsulation techniques
  - GRE[2784], IPv6 over IPv4 [2893], etc.

Extension to IPv6 over IPv4 (AFI)

Number	Description		Reference
0	Reserved		
1	IP (IP version 4)		
2	IP6 (IP version 6)		
3	NSAP		
4	HDLC (8-bit multidrop)		
5	BBN 1822		
6	802 (includes all 802 media plus Ethernet "canonical format")		
7	E.163		
8	E.164 (SMDS, Frame Relay, ATM)		
9	F.69 (Telex)		
10	X.121 (X.25, Frame Relay)		
11	IPX		
12	Appletalk		
13	Decnet IV		
14	Banyan Vines		
15	E.164 with NSAP format subaddress	[UNI-3.1]	[Malis]
16	DNS (Domain Name System)		
17	Distinguished Name		[Lynn]
18	AS Number		[Lynn]
19	XTP on IP version 4		[Saul]
20	XTP on IP version 6		[Saul]
21	XTP native mode XTP		[Saul]
22	Fibre Channel World-Wide Port Name		[Bakke]
23	Fibre Channel World-Wide Node Name		[Bakke]
24	GWID		[Hegde]
65535	Reserved		

Use: IP=1 for IPv4 edge networks
IP6=2 for IPv6 edge networks

Extension to IPv6 over IPv4 Protocol Definition



Extension to IPv6 over IPv4 Protocol Behavior

- Similar to 4over6
- Receiving from CE
  - Construct encapsulation table
    - Dst IPv6 network address with prefix length
    - 4over6 IPv4 VIF
  - Send to other PEs
    - AFI=IP, SAFI=SAFI_4OVER6
- Receiving from PE
  - Identify the type with AFI=IP, SAFI=SAFI_4OVER6
  - Set the encapsulation table and IPv6 RT

Conclusion

- 4over6 proposal for Mesh Problem
  - IPv4 over IPv6 backbones
  - IPv6 backbones act as dual-stack core
- Packet encapsulation is reused
  - Encapsulation and Decapsulation
- BGP-MP 4over6 extension is defined
  - New SAFI: SAFI_4OVER6 = 67
  - Protocol behavior is defined
- Advantage
  - Only PE router needs to be extended to maintain routing info of access networks
  - Core networks and custom networks are not aware of 4over6
  - Simple extension and configuration
  - Easy to extend to IPv6 over IPv4

Q and A

Thanks