# Routing IPv6

**Rocky Mountain IPv6 Task Force Summit**

**Denver**

21 April, 2009

Jeff Doyle
President
Jeff Doyle and Associates, Inc.

# Routers Supporting IPv6

- 6Wind
- AddPac Technologoes
- ALAXALA Technology
- AlaxalA Networks
- Alcatel/Lucent
- Allied Telesis
- Allied
- Telesyn
- Alpha Networks
- Beijing Jiaxun Feuihong Networks
- Billion Electric
- BITWAY Networking Technology
- BSD (KAME)
- Cisco Systems
- Delta Networks
- Digital China Networks
- D-Link
- DrayTek
- Extreme Networks
- Fiberhome Networks
- Fujitsu
- FugtureSoft
- Harbour Networks
- Hitachi
- Huawei
- IP-infusion
- Juniper Networks
- LG-Nortel
- Linux (USAGI)
- Mercury Corporation
- Microsoft
- NEC
- Netgenetech
- NextHop
- Nokia
- Nortel
- Novell
- Omron
- Panasonic
- Ruijie Networks
- Samsung
- Shanghai Baud Data Communication
- Tsinghua Unisplendor Bitway Networking
- Xoprt
- Yamaha
- Zebra
- ZTE Corporation
- ZyXEL

# MTU Path Discovery

- IPv6 routers do not fragment packets
- IPv6 MTU must be at least 1280 bytes
  - Recommended MTU: 1500 bytes
- Nodes should implement MTU PD
  - Otherwise they must not exceed 1280 bytes
- MTU path discovery uses ICMP "packet too big" error messages
  - Be sure to consider them when filtering ICMPv6

# Static Routes

- Static route configuration syntax is the same as IPv4
- Except prefix and next hop are IPv6
- Next hop address can be global or link local
  - ICMPv6 Redirect messages need link-local address

prefix → next-hop address

# Static Routes on Cisco and Juniper

- Juniper Networks syntax
  - IPv4 static route:

```
[edit]
set routing-options static route [ipv4_prefix/prefix_length] next-hop
    [ipv4_if_address]
```

  - IPv6 static route:

```
[edit]
set routing-options rib inet6.0 static route [ipv6_prefix/prefix_length] next
   -hop [ipv6_if_address]
```

- Cisco Systems syntax:
  - IPv4 static route:

```
ip route [ipv4_prefix] [ipv4_address_mask] [ipv4_if_address]
```

  - IPv6 static route:

```
ipv6 route [ipv6_prefix/prefix_length][outgoing interface] [ipv6_if_address]
```

# RIPng

- RFC 2080 describes RIPngv1, not to be confused with RIPv1
- Based on RIP Version 2 (RIPv2)
- Uses UDP port 521
- Operational procedures, timers and stability functions remain unchanged
- RIPng is not backward compatible to RIPv2
- Message format changed to carry larger IPv6 addresses

# Juniper and Cisco RIPng Configurations

**Juniper Networks example:**

```
[edit protocols ripng]
jeff@Juniper1# show
group Peers {
    export prefixes;
    neighbor fe-0/0/0.0;
}
[edit policy-options]
jeff@Juniper1# show
policy-statement prefixes {
    from protocol direct;
    then accept;
}
```

**Cisco Systems example:**

```
interface Ethernet1/0
 ipv6 address 2001:1100:A:B::1/64
 ipv6 enable
 ipv6 rip Demo enable
!
ipv6 router rip Demo
!
```

# Cisco EIGRP

- Supported as of IOS 12.4(6)T
- Same DUAL convergence algorithm
- Simple addition of TLVs to support IPv6
- Differences from EIGRP for IPv4:
  - Configured directly on interface
  - No `network` statement
  - Requires Router ID

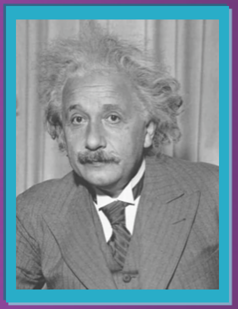
# Cisco EIGRP Configuration Example

```
ipv6 unicast-routing
interface e0
  ipv6 enable
  ipv6 eigrp 1
  no shutdown
!
ipv6 router eigrp 1
  router-id 10.1.1.1
  no shutdown
```

# IS-IS

- RFC 5308, Routing IPv6 with IS-IS (5 pages)
- 2 new TLVs are defined:
  - IPv6 Reachability (TLV type 236)
  - IPv6 Interface Address (TLV type 232)
- IPv6 NLPID = 142

**"Make it as simple as possible, but no simpler."**
— *Albert Einstein*



# IS-IS on Juniper Routers

```
[edit]
jeff@Juniper1# show interfaces
fe-0/0/0 {
    unit 0 {
        family iso;
        family inet6 {
            address 2001:2200:a:1::2/64;
        }
    }
}
[edit]
jeff@Juniper1# show protocols isis
interface fe-0/0/0.0;
interface fe-0/0/1.0;
interface fe-0/0/2.0;
```

# IS-IS on Cisco Routers

```
interface ethernet-1
    ip address 10.1.1.1 255.255.255.0
    ipv6 address 2001:0001::45c/64
    ip router isis
    ipv6 router isis

 interface ethernet-2
    ip address 10.2.1.1 255.255.255.0
    ipv6 address 2001:0002::45a/64
    ip router isis
    ipv6 router isis

 router isis
     address-family ipv6
     exit-address-family
     net 42.0001.0000.0000.072c.00
```

# OSPFv3

- Unlike IS-IS, entirely new version required
- RFC 2740
- Fundamental OSPF mechanisms and algorithms unchanged
- Packet and LSA formats are different

# OSPFv3 Differences from OSPFv2

- Runs per-link rather than per-subnet
  - Multiple instances on a single link
- More flexible handling of unknown LSA types
  - More network changes without adjacency disruptions possible
- Link-local flooding scope added
  - Similar to flooding scope of type 9 Opaque LSAs
  - Area and AS flooding remain unchanged
- Authentication removed
  - Uses IPv6 Authentication (AH) extension header instead
- Neighboring routers always identified by RID
- Removal of addressing semantics
  - IPv6 addresses not present in most OSPF packets
  - RIDs, AIDs, and LSA IDs remain 32 bits

# OSPFv3 LSAs

| Type | Description |
| --- | --- |
| 0x2001 | Router-LSA |
| 0x2002 | Network-LSA |
| 0x2003 | Inter-Area-Prefix-LSA |
| 0x2004 | Inter-Area-Router-LSA |
| 0x4005 | AS-External-LSA |
| 0x2006 | Group-Membership-LSA |
| 0x2007 | Type-7-LSA (NSSA) |
| 0x0008 | Link-LSA |
| 0x2009 | Intra-Area-Prefix-LSA |

# OSPFv3: Intra-Area-Prefix LSA

- OSPFv2:
  - Prefixes are advertised in Router (Type 1) LSAs
    - Primary purpose of Type 1 LSAs is to compute SPF tree
  - Any addition/deletion/change of prefix requires flood of new Type 1 LSA
    - Yet prefix change does not affect SPF tree
  - SPF re-calculation is needlessly triggered
  - Partial Route Calculation (PRC) cannot help OSPFv2 to scale
- OSPFv3:
  - Prefixes are advertised in Intra-Area-Prefix LSAs
    - Not Router LSAs
  - Intra-Area-Prefix LSAs do not trigger SPF run
  - Scalability much improved in very large areas
    - More comparable to IS-IS
    - PRC becomes useful for OSPFv3

# Juniper and Cisco OSPFv3 Configuration

**Cisco Systems example**

```
interface Ethernet1/0
 ipv6 address 2001:1100:A:B::1/64
 ipv6 enable
 ipv6 ospf 1 area 0.0.0.0
!
```

**Juniper Networks example**

```
[edit]
jeff@Juniper1# show protocols ospf3
area 0.0.0.0 {
    interface fe-0/0/0.0;
    interface fe-0/0/1.0;
    interface fe-0/0/2.0;
}
```

# Multiprocotol BGP-4

- MP-BGP defined in RFC 2283
- Two BGP attributes defined:
  - Multiprotocol Reachable NLRI advertises arbitrary Network Layer Routing Information
  - Multiprotocol Unreachable NLRI withdraws arbitrary Network Layer Routing Information
  - Address Family Identfier (AFI) specifies what NLRI is being carried (IPv6, IP Multicast, L2VPN, L3VPN, IPX...)
- Use of MP-BGP extensions for IPv6 defined in RFC 2545
  - IPv6 AFI = 2
- BGP TCP session can be over IPv4 or IPv6
- Advertised Next-Hop address must be global IPv6 address
  - And can be followed by a link-local IPv6 address
  - Resolves conflicts between IPv6 rules and BGP rules

# M-BGP for IPv6 on Cisco

```
router bgp 1
 no synchronization
 no bgp default ipv4-unicast
 bgp log-neighbor-changes
 neighbor 2001:2200:A:1::2 remote-as 200
 no auto-summary
 !
address-family ipv6
 neighbor 2001:2200:A:1::2 activate
 exit-address-family
 !
```

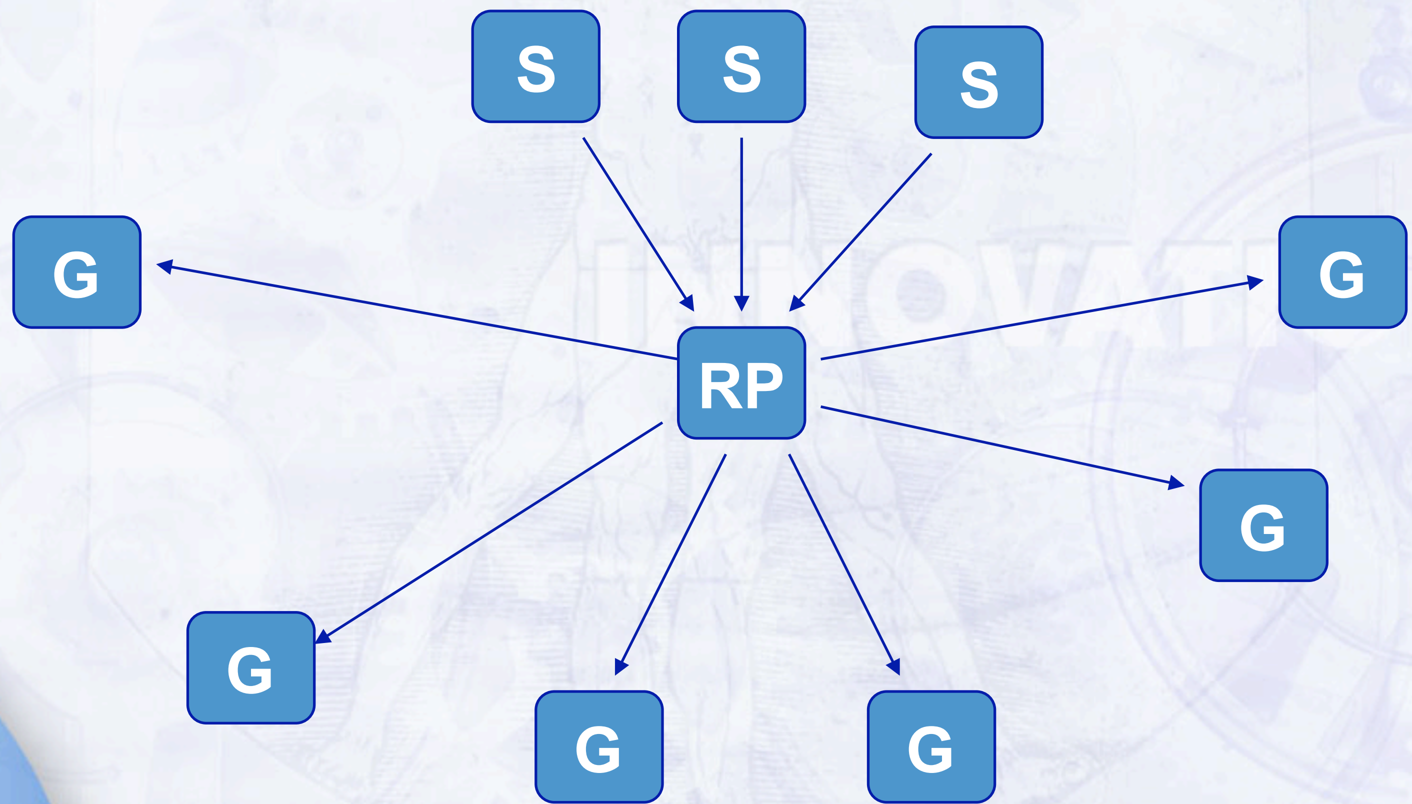
# M-BGP for IPv6 on Juniper

```
[edit]
jeff@Juniper1# show protocols bgp
group v6-peers {
    type external;
    family inet6 {
        unicast;
    }
    export v6-export;
    peer-as 200;
    neighbor 2001:2200:a:2::2;
}
```

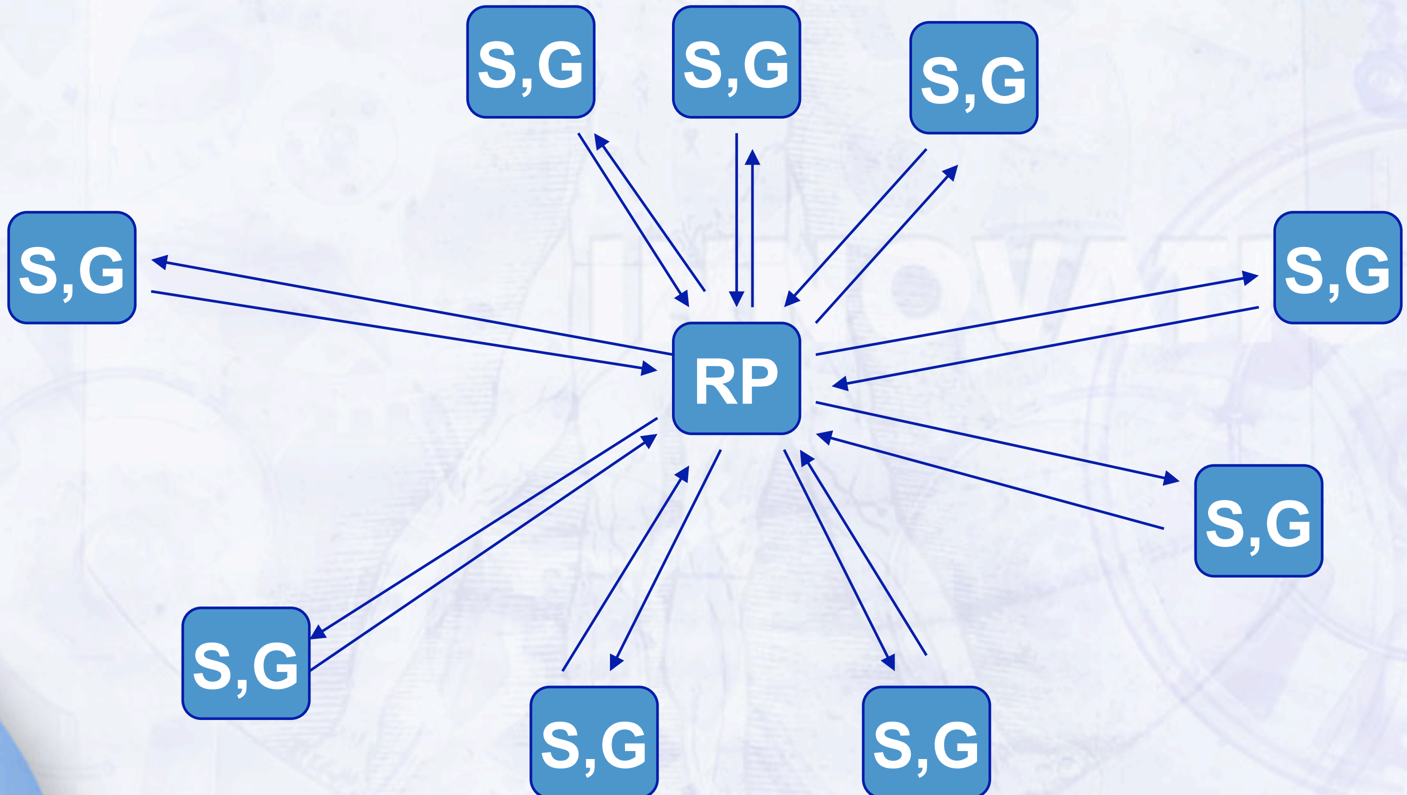
# IPv6 Multicast Routing

- PIM-SM
  - "Basic" PIM-SM
  - PIM-Bidir
  - PIM-SSM
- MP-BGP
- Legacy protocols not supporting IPv6:
  - DVMRP
  - PIM-DM

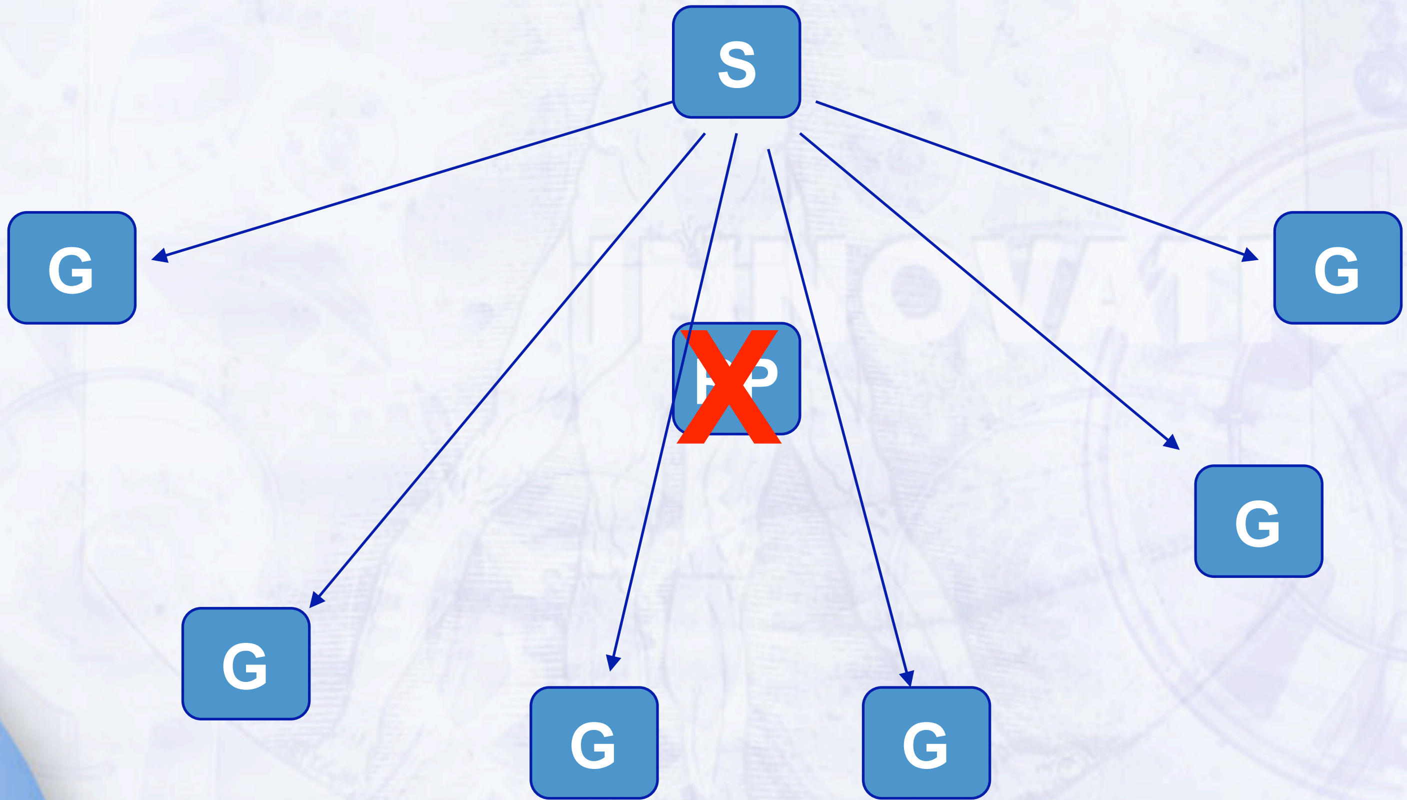
# PIM-SM



# PIM-BIDIR



# PIM-SSM



# Multicast Operational Models

- Any-Source Multicast (ASM)
  - Basic PIM-SM
    - Smaller-scale many-to-many applications
    - “Few-to-many” applications
    - Examples: Conferencing, small chat rooms, data distribution
  - Bidirectional PIM (PIM-Bidir)
    - Larger-scale many-to-many applications
    - Examples: Full-participation voice/video/multimedia conferencing, massively multiplayer gaming, large chat rooms
- Single-Source Multicast (SSM)
  - PIM-SSM
    - Single-to-many applications
    - Examples: Audio, video content distribution
    - Requires MLDv2 (equivalent to IGMPv3 for IPv4)

# Rendezvous Point (RP) Discovery

- PIM-SM, PIM-Bidir require RP for shared trees
  - PIM-SSM does not require RP
- Static RP Configuration
  - Currently most widely used method for IPv4 multicast
  - But will it scale operationally?
- Bootstrap Router (BSR) protocol
- Embedded RP addresses
  - Promising for automated RP discovery without added mechanism
- No Auto-RP for IPv6
  - Never widely deployed anyway

# Embedded RP Addresses: RFC 3306

Standard IPv6 Multicast Address Format

| 0xFF | Flags (4 bits) | Scope (4 bits) | Group ID (112 bits) |
| --- | --- | --- | --- |

- Leverages Unicast-Prefix-Based Multicast Addresses
  - RFC 3306
  - Format is intended for dynamic IPv6 multicast address allocation
  - Can support both ASM and SSM models

Unicast-Prefix-Based Multicast Address Format

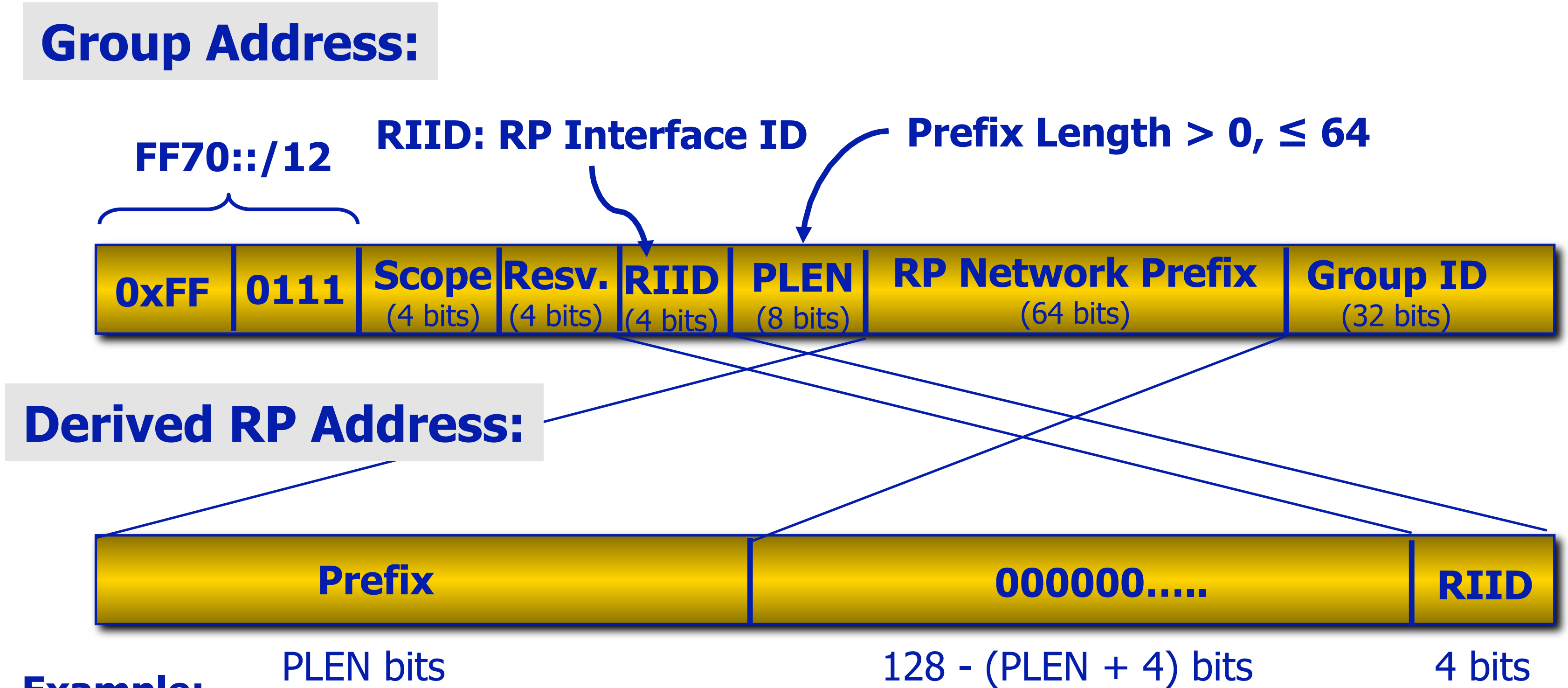
| 0xFF | Flags (4 bits) | Scope (4 bits) | Reserved (8 bits) | PLEN (8 bits) | Network Prefix (64 bits) | Group ID (32 bits) |
| --- | --- | --- | --- | --- | --- | --- |

**0RPT**

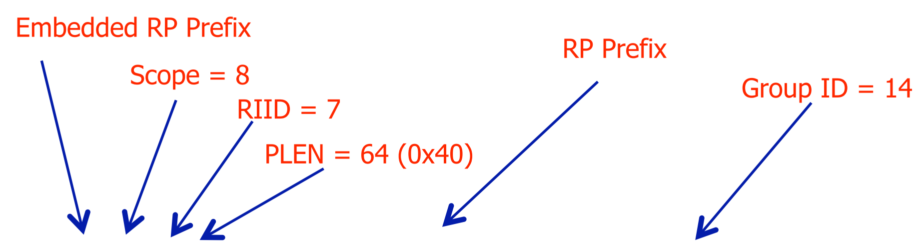
R = Embedded RP Address
P = Multicast address based on network prefix
T = Transient address

PLEN (Prefix Length) = 0 for SSM

# Embedded RP Addresses




**Example:**
RP Prefix = 2001:DB8:ABCD:1234::/64
RP Interface ID: 7
RP Address = 2001:DB8:ABCD:1234::7
Scope = Organization Local (8)
Group ID = 0x14




Group Address with Embedded RP Address = FF78:740:2001:DB8:ABCD:1234::14

# Inter-Domain IPv6 Multicast

- MP-BGP
- SSM models with PIM-SSM
- ASM models problematic
  - No IPv6 version of MSDP
  - Embedded RP might help here
  - For now, “big SSM communities” will work
    - But need a more scalable solution for the long run

# Conclusions

- Unicast IPv6 routing essentially the same as unicast IPv4
  - If you understand IPv4 routing, you "have it made"
- OSPFv3 is a big improvement over OSPFv2
  - Changes based on 10 years' experience
  - Discussions underway to extend OSPFv3 for IPv4
- Simple IPv6 multicast very similar to IPv4 multicast
  - "Simple" is mostly what is in use now
- Complex (large scale and/or interdomain) IPv6 multicast still needs work
  - But, then, so does large-scale IPv4 multicast
  - IPv6 solutions should prove to be simpler in the long run

# Thank You

jdoyle@doyleassociates.net

+1 (303) 428-4680

www.doyleassociates.net