



SOI-Asia, Feb. 4 – Feb. 12, 2008



Toward IP Multicast Deployment

Lecturer: Hitoshi Asaeda, Ph.D.
Graduate School of Media and Governance
Keio University

Copyright © Hitoshi Asaeda

1

First Day

- IP multicast communication architecture
 - Concept and components
 - Addressing architecture
 - Routing path construction
 - Reverse Path Forwarding
 - Communication protocols
 - PIM-SM, MSDP, MBGP, Embedded RP
 - Source-Specific Multicast (SSM)

Copyright © Hitoshi Asaeda

2

Second Day

- IP multicast communication architecture
 - Communication protocols
 - Source-Specific Multicast (SSM)
 - IGMPv3, MLDv2, LW-IGMPv3/LW-MLDv2
 - IGMP/MLD Proxy
 - Automatic multicast tunnel (AMT)
 - Session announcement
 - SDP
 - SAP
 - Channel Reflector
- Multicast security

Copyright © Hitoshi Asaeda

3

Third Day

- Operation and management
 - WIDE backbone, AI3 backbone, and other International connectivity
 - Router configurations
 - Cisco/XORP
 - Monitoring tools
 - dbeacon,
 - ssm ping and mtrace2
 - Member counting
 - Direct counting (RTCP, PIM extension) and estimation

Copyright © Hitoshi Asaeda

4

Fourth Day

- Applications and services
 - RTP/RTCP
 - Well-known applications
 - DVTS, VLC, VIC/RAT
 - Application development
 - MSF APIs
 - Debugging tool
 - mtest, mcastread/mcastsend

Copyright © Hitoshi Asaeda

5

If Time Remains...

- Algorithms and advanced technologies
 - Multihoming
 - Multicast over UDL
 - PIM-SM over satellite networks
 - Reliable multicast data transmission

Copyright © Hitoshi Asaeda

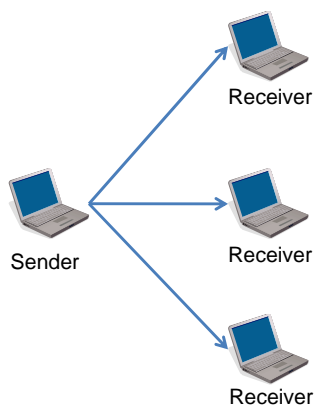
6

Concept

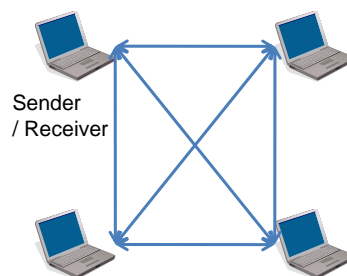
Copyright © Hitoshi Asaeda

7

Communication Style



One-to-many communication



Many-to-many communication

Copyright © Hitoshi Asaeda

8

Unicast/Broadcast Communication

- Unicast
 - One-to-one communication, like an HTTP access, telnet, ftp
 - Reliable or non-reliable communication
- Broadcast
 - One-to-many or many-to-many data flooding
 - Data is flooded only on the link (with TTL=1)
 - No receiver management
 - All receiver on the same link receives data
 - Non-reliable communication

Copyright © Hitoshi Asaeda

9

IP Multicast Communication

- Concept
 - Multicast data sender sends the data only once, and only the intended recipients (who want to receive the data) receive the data
 - IP multicast provides one-to-many or many-to-many communication effectively
 - Each data (i.e. multicast stream) is classified by multicast address (and source address if SSM is used)
 - Non-reliable communication (i.e. on top of UDP)
 - IP multicast is basically applied to real-time applications

Copyright © Hitoshi Asaeda

10

IP Multicast Communication

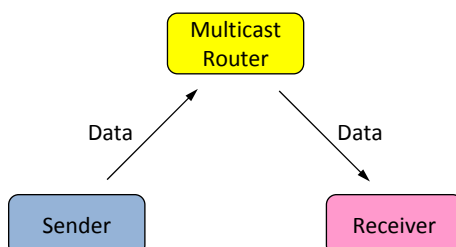
- Advantages
 - Effective data forwarding to a large number of receivers
 - Effective use of computing and network resources
- Disadvantages
 - Various “difficulties” and “limitations”
 - This lecture details these difficulties and limitations, and proposes the future or candidate solutions if possible

Copyright © Hitoshi Asaeda

11

Data Flow

- Data sender
 - Sender sends data once
- Data receiver
 - Receiver that has requested getting the data receives the data
- Multicast routers
 - Router copies and forwards the data only toward the data receivers

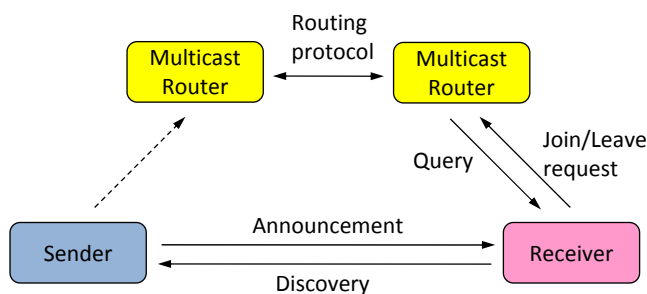


Copyright © Hitoshi Asaeda

12

Communication Flow

- Control messages
 - Sender announces the session information or receivers discover the session information
 - Each receiver requests to start and stop receiving data by “join and leave” operations
 - Multicast routers maintains membership state by having reports



Copyright © Hitoshi Asaeda

13

Terminologies

- Group address (or multicast address)
 - Used for destination address
- Join and leave
 - Data reception state requested by receiver hosts
- Join and prune
 - Data reception state requested by routers
- $(*,G)$ and (S,G)
 - Notation of source address and group address in join-and-leave (or join-and-prune) state

Copyright © Hitoshi Asaeda

14

Terminologies

- Scope
 - Expected data distribution area
 - Classified by multicast address or TTL
- TTL (Time To Live) or Hop limit
 - Expected maximum hop count of each packet
- IIF and OIF
 - IIF: Incoming interface from which data is received
 - OIF: Outgoing interface to which data is sent

Copyright © Hitoshi Asaeda

15

Terminologies

- Multicast session
 - Multicast data stream classified by the “multicast address” is called “multicast session”
- Multicast channel
 - Multicast data stream explicitly classified by the pair of “multicast address” and “source address” is called “multicast channel”
 - Used for SSM

Copyright © Hitoshi Asaeda

16

Multicast Address Assignment

Copyright © Hitoshi Asaeda

17

IP Multicast Address

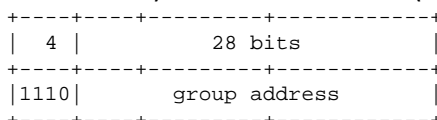
- IP multicast address
 - IPv4: 224.0.0.0 – 239.255.255.255
 - IPv6: FFx0::1
 - MUST be specified as a destination address
 - MUST NOT be specified as a source address
- Dynamic address assignment
 - Regular applications select their multicast addresses dynamically
 - Some multicast addresses are assigned by IANA for special uses

Copyright © Hitoshi Asaeda

18

IPv4 Multicast Addresses

- IPv4 multicast address
 - 224.0.0.0 (0xe0000000) - 239.255.255.255 (0xefffffff)



- Administrative scope [RFC2365]
 - Local address (224/24)
 - Administrative scope (239/8)
 - Organization-Local (239.192/14)
- GLOP address (233/8) [RFC3180]
- EGLOP (233.252.0.0 - 233.255.255.255) [RFC3138]
- SSM address (232/8) [RFC]

Copyright © Hitoshi Asaeda

19

IPv6 Multicast Addresses

- IPv6 multicast address: FFxx::



- Flags
 - 000T (T=1: transient, T=0: well-known)
- Scope
 - 0x1: Interface Local
 - 0x2: Link-Local
 - 0x3: Subnet-Local
 - 0x4: Admin-Local
 - 0x5: Site-Local
 - 0x8: Organization-Local
 - 0xE: Global
- SSM address (FF3x::/32 (or 96)) [RFC3306]

Copyright © Hitoshi Asaeda

20

Embedded Multicast Address

- Dynamic address assignment gives the difficulties in operation and policy management
 - No AS or site dependency in multicast address assignment
 - Embedded multicast address proposal
- GLOP/EGLOP/Unicast-prefix-based address
 - ASN or unicast address is mapped in a multicast address
 - Beneficial for operation and management
 - Possibility for the address or prefix aggregation

Copyright © Hitoshi Asaeda

21

Embedded Multicast Address for IPv4

- GLOP address (233/8) [RFC3180]

```

0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-----+-----+-----+-----+-----+-----+-----+-----+
| 1 1 1 0 | 1 0 0 1 |           16 bits AS           | local bits |
+-----+-----+-----+-----+-----+-----+-----+-----+
    
```

- EGLOP (233.252/14 (233.252.0.0 - 233.255.255.255)) [RFC3138]
 - To support longer ASN (e.g. private ASN (64512 – 65535))
 - Regional Registry (RIR) assigns address blocks, instead of IANA

```

0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-----+-----+-----+-----+-----+-----+-----+-----+
| 1 1 1 0 | 1 0 0 1 | 1 1 1 1 | 1 1 |           18 bits space |
+-----+-----+-----+-----+-----+-----+-----+-----+
    
```

Copyright © Hitoshi Asaeda

22

Embedded Multicast Address for IPv4

- Problems in future
 - 16 bit ASN will become unavailable
 - January, 2009, the RIRs will only give them out by request.
 - January 2010, they will not distinguish between the 16 and 32 bit pools.
 - In practice, 32 bit ASN will be given by IANA.
 - There is no GLOP/EGLOP space for 32 bit ASN.
 - No solution is provided yet ...

Copyright © Hitoshi Asaeda

23

Embedded Multicast Address for IPv6

- Unicast-prefix-based address (IPv6) – [RFC3306]

8	4	4	8	8	64	32
-----+-----+-----+-----+-----+-----+-----						
11111111	flgs	scop	reserved	plen	network prefix	group ID
-----+-----+-----+-----+-----+-----+-----						

- Example: A network with a unicast prefix of 3FFE:FFFF:1::/48 would also have a unicast prefix-based multicast prefix of FF3x:0030:3FFE:FFFF:0001::/96 (where “x” is a valid scope value).
- FF3x::/32 for SSM
 - plen = 0, network prefix = 0

Copyright © Hitoshi Asaeda

24

Multicast Address Assignment Problems – Summary

- No GLOP/EGLOP use with 32 bit ASN
- IPv6 multicast is simpler to use
- Need dynamic address assignment protocol
 - Multicast Address Dynamic Client Allocation Protocol (MADCAP) is defined [RFC2730] and implemented on Windows, but not commonly used

Copyright © Hitoshi Asaeda

25

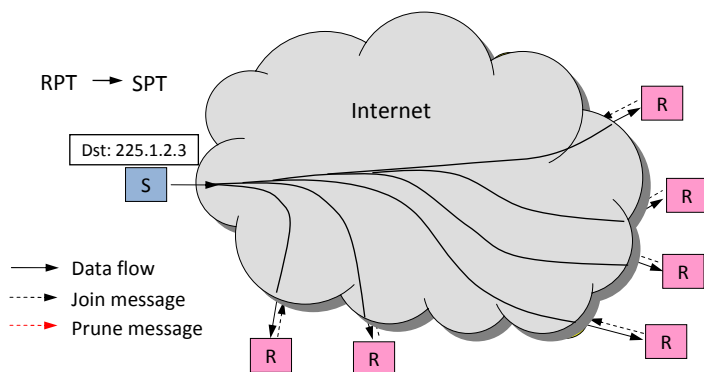
Multicast Routing Protocols

Copyright © Hitoshi Asaeda

26

Tree Coordination

- Routing tree coordination is complex?
- Routing tree coordination is not scalable?



Copyright © Hitoshi Asaeda

27

Multicast Routing Table

- Routing entries
 - Source address
 - When (*,G) tree, source address is NULL
 - Multicast address
 - Incoming Interface (IIF)
 - Outgoing Interface (OIF)

Copyright © Hitoshi Asaeda

28

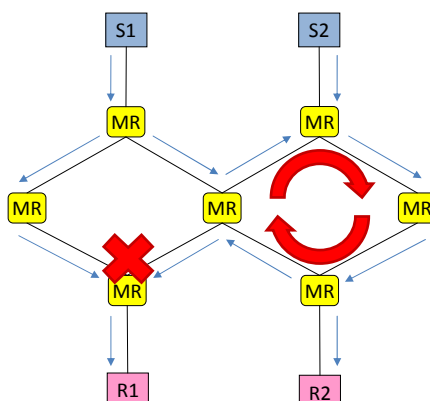
Reverse Path Forwarding

- Reverse Path Forwarding (RPF)
 - The multicast RPF algorithm allows a multicast router to accept a multicast datagram only on the interface where it would send a unicast datagram to the source of that datagram
 - To define incoming and outgoing interfaces
 - To avoid routing loop
 - To avoid forwarding duplicate packets

Copyright © Hitoshi Asaeda

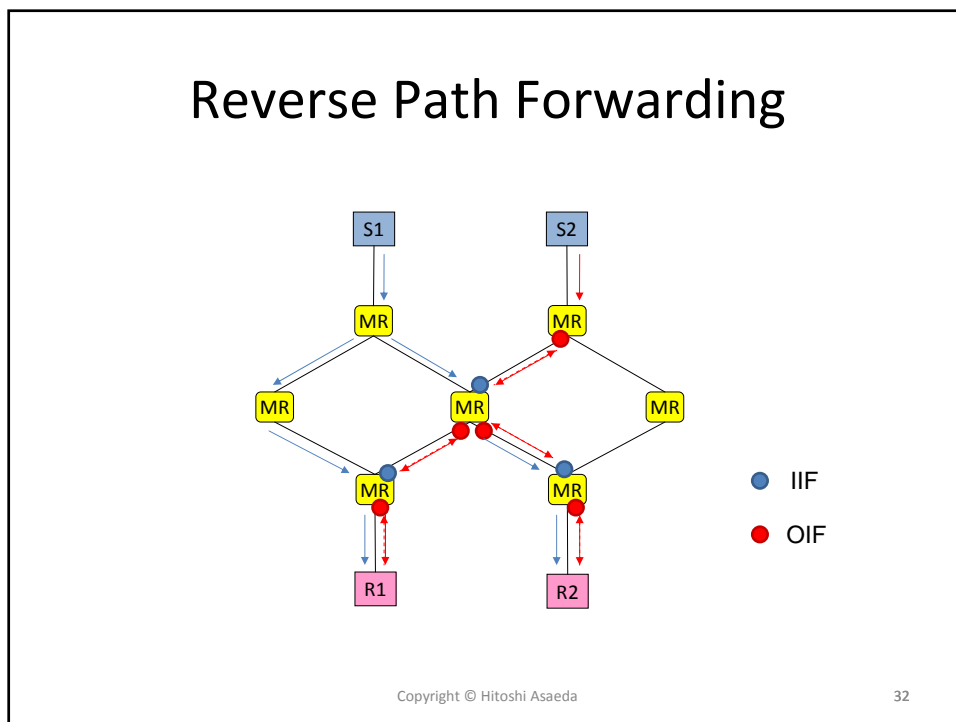
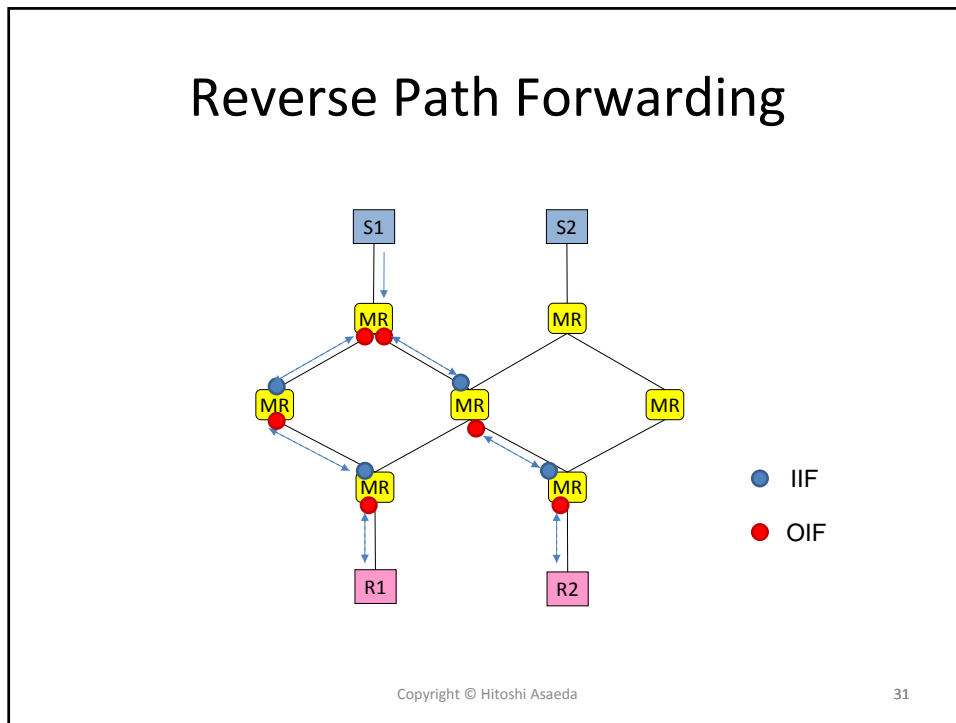
29

Reverse Path Forwarding



Copyright © Hitoshi Asaeda

30



Multicast Routing Protocols

- DVMRP
 - Broadcast-and-prune type protocol
 - Used in the traditional MBone
- MOSPF
 - Link state type protocol
- CBT
 - Shared tree
- PIM-DM
 - Broadcast-and-prune type protocol
- PIM-SM
 - Shared tree and Shortest-path tree

Copyright © Hitoshi Asaeda

33

Additional Protocols and Functions

- MBGP
 - Used for policy definition
 - Define upstream multicast routers
- MSDP
 - Create peering connections with multiple RPs
 - IPv4 only (not used in IPv6)
- Embedded RP
 - Used with PIM-SM in IPv6

Copyright © Hitoshi Asaeda

34

PIM-SM

- Protocol Independent Multicast Routing Protocol
 - Sparse Mode (PIM-SM)
 - PIM-SMv2 standard protocol defined in [RFC4601]
 - Explicit-join type routing protocol
 - Components
 - Bootstrap router (BSR) [RFC5059]
 - Rendezvous Point router (RP)
 - Designated router (DR)
 - PIM Multicast Border Router (PMBR)
 - (IGMP/MLD querier)
 - Router that has lower IP address becomes querier

Copyright © Hitoshi Asaeda

35

PIM Messages

- PIM has own IP protocol number, 103
 - IP + PIM message (not with UDP, IGMP, etc.)
- ALL-PIM-ROUTERS address
 - 224.0.0.13 for IPv4 and ff02::d for IPv6
- Types

– 0 = Hello	Multicast to ALL-PIM-ROUTERS
– 1 = Register	Unicast to RP
– 2 = Register-Stop	Unicast to source of Register packet
– 3 = Join/Prune	Multicast to ALL-PIM-ROUTERS
– 4 = Bootstrap	Multicast to ALL-PIM-ROUTERS
– 5 = Assert	Multicast to ALL-PIM-ROUTERS
– 6 = Graft (used in PIM-DM only)	Unicast to RPF'(S)
– 7 = Graft-Ack (used in PIM-DM only)	Unicast to source of Graft packet
– 8 = Candidate-RP-Advertisement	

Copyright © Hitoshi Asaeda

36

Bootstrap Router (BSR)

- Cand-BSR sends candidacy by PIM Hello
- BSR is selected from Cand-BSR with the priority value in the scope
- Receive Candidate-RP-Advertisement message from Cand-RP
- Send Candidate-RP-Set message to all PIM routers

Copyright © Hitoshi Asaeda

37

Rendezvous Point Router (RP)

- RP is a core router that becomes a root of multicast routing tree
- Each RP defines group prefix the RP supports
- Sender's data is encapsulated and transmitted to RP
- Receiver's join is forwarded toward RP
- RP resolves data senders' addresses and join requests sent by receivers

Copyright © Hitoshi Asaeda

38

RP Configuration

- Static Configuration
 - A PIM router MUST support the static configuration of group-to-RP mappings.
 - Such a mechanism is not robust to failures, but does at least provide a basic mechanism.
- Cisco's Auto-RP
 - This mechanism is not useful if PIM-DM is not being run in parallel with PIM-SM, and was only intended for use with PIM-SMv1.
 - No standard specification currently exists.

Copyright © Hitoshi Asaeda

39

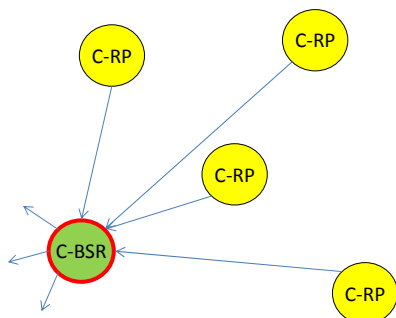
RP Configuration

- Dynamic configuration
 - RP-set is announced by Bootstrap Router (BSR)
 - A bootstrap mechanism based on the automatic election of BSR
 - $\text{Value}(G,M,C(i)) = (1103515245 * ((1103515245 * (G\&M)+12345) \text{ XOR } C(i)) + 12345) \text{ mod } 2^{31}$
 - Any router in the domain that is configured to be a possible RP reports its candidacy to the BSR, and then a domain-wide flooding mechanism distributes the BSR's chosen set of RPs throughout the domain.

Copyright © Hitoshi Asaeda

40

Candidate RPs



- Cand-RP sends C-RP-Adv as the candidacy to BSR by unicasts
- RP-set is calculated by BSR and periodically announced in a Bootstrap message
- In many cases or usually, Cand-BSR is also Cand-RP

Copyright © Hitoshi Asaeda

41

RP Configuration

- Embedded-RP
 - Embedded-RP defines an address allocation policy in which the address of the Rendezvous Point (RP) is encoded in an IPv6 multicast group address.
- Anycast RP
 - RP address is announced with anycast address
 - Anycast RP for PIM and MSDP [RFC3446]
 - MSDP is the requirement
 - Only for IPv4 because Embedded RP take its role for IPv6

Copyright © Hitoshi Asaeda

42

Designated Router (DR)

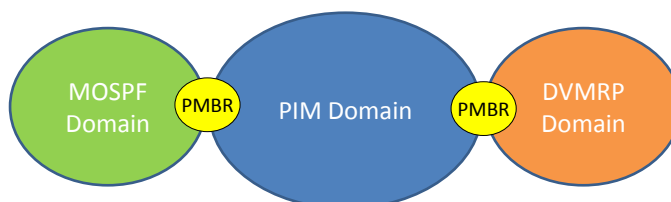
- DR is a single router that performs to send PIM Join/Prune and Assert messages
- DR election is performed using PIM Hello messages
- DR is selected with the priority value and IP address

Copyright © Hitoshi Asaeda

43

PIM Multicast Border Router

- PMBR provides interconnection with domains using other routing protocols
 - [RFC2715] is not a specification of PMBR but is useful for this topic
- Two tasks
 - Ensure that traffic from sources outside the PIM-SM domain reaches receivers inside the domain.
 - Ensure that traffic from sources inside the PIM-SM domain reaches receivers outside the domain.



Copyright © Hitoshi Asaeda

44

Shared Tree

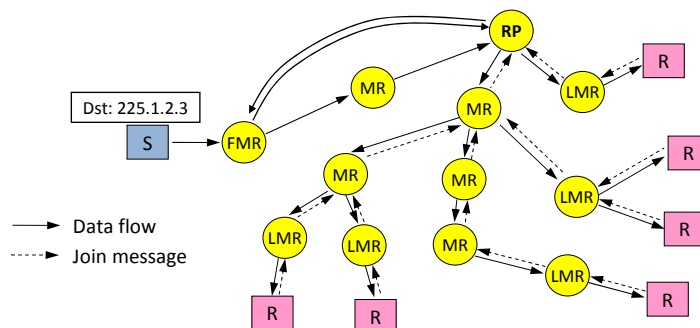
- Concept
 - Routing tree is rooted at a core router (or RP)
 - Join messages from receivers for a group are sent towards the RP
 - Data from senders is sent to the RP so that receivers can discover who the senders are and start to receive traffic destined for the group.
 - Shared tree (or RPT) is created multicast address prefix
 - Enable (*,G) join/leave
 - Source address discovery

Copyright © Hitoshi Asaeda

45

Rendezvous Point Tree

- Join messages and forwarded data along Rendezvous Point Tree (RPT)



Copyright © Hitoshi Asaeda

46

Shortest-Path Tree

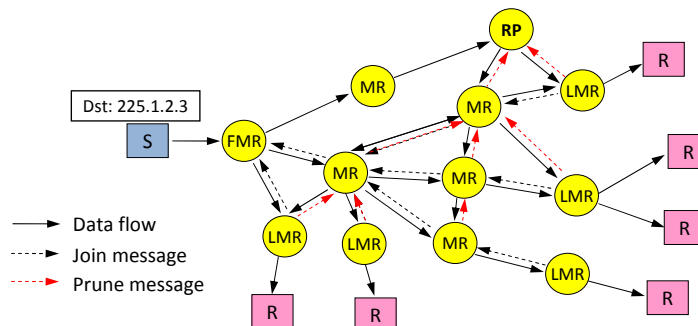
- Concept
 - Source-based tree
 - Routing tree is constructed for each source
 - Routing tree is rooted at each source
 - Optimized tree (since the tree is coordinated with the shortest path)

Copyright © Hitoshi Asaeda

47

Shortest-Path Tree

- Join and prune messages and forwarded data over SPT
- PIM-SM switches from RPT to SPT



Copyright © Hitoshi Asaeda

48

Bi-Directional PIM

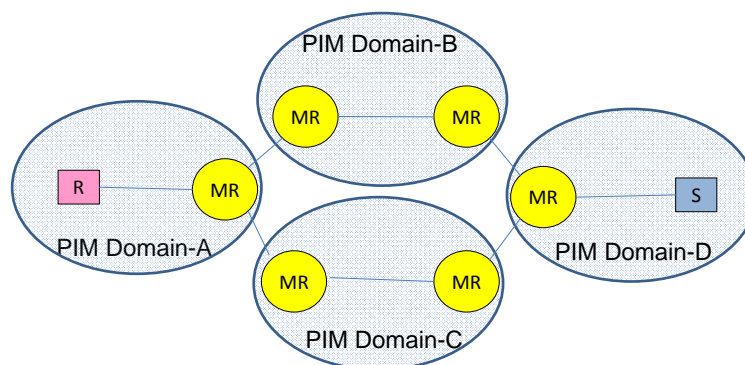
- Defined in [RFC5015]
- Not switched from RPT to SPT
 - No SSM transition
- Similar to CBT concept

Copyright © Hitoshi Asaeda

49

MBGP

- Multiprotocol Extensions for BGP-4 [RFC4760]
 - Implemented with MRIB



Copyright © Hitoshi Asaeda

50

Multicast Routing Information Base

- MRIB is used to define routing policy
 - Define incoming interface for RPF calculation

Copyright © Hitoshi Asaeda

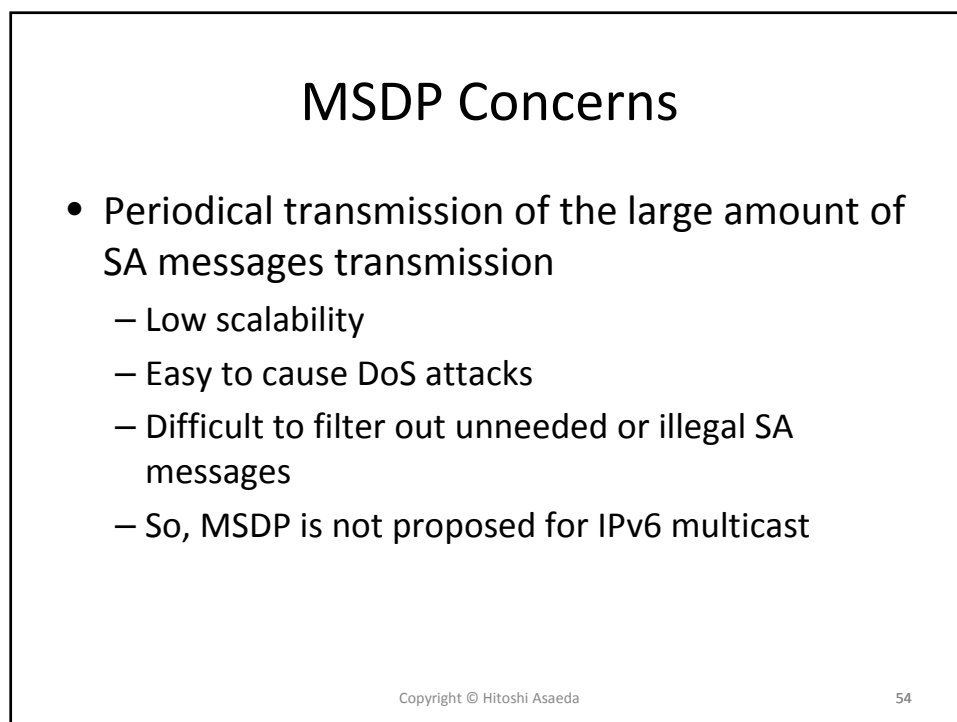
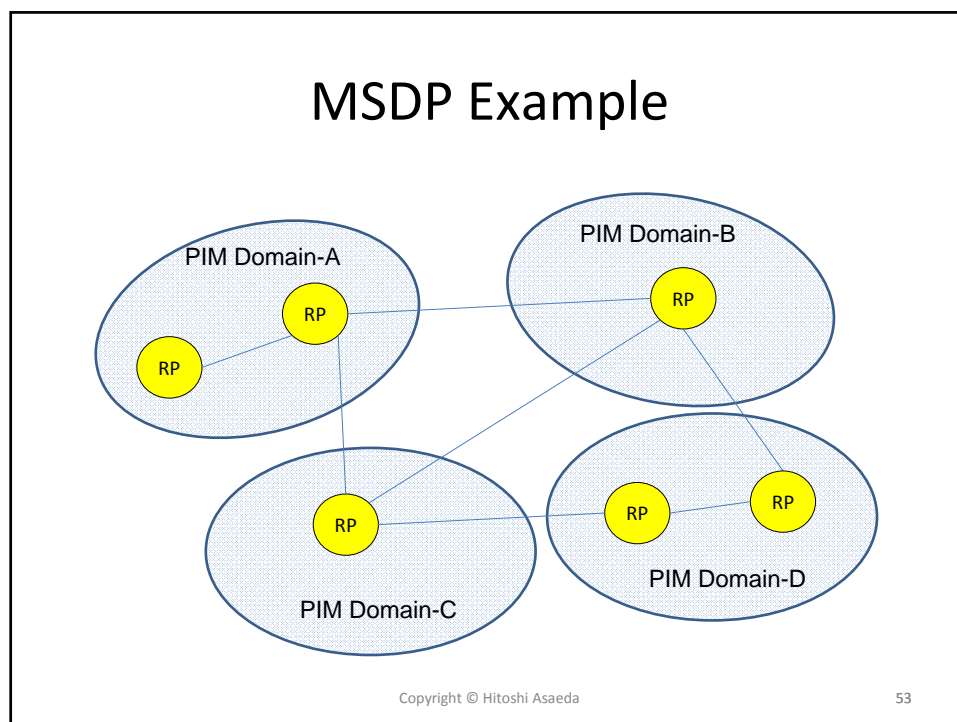
51

MSDP

- Multicast Source Discovery Protocol
 - Defined in [RFC3618]
 - MSDP is a mechanism to connect multiple IPv4 PIM-SM domains together
 - Each PIM-SM domain uses its own independent RP and needs to discover sources in other PIM domains
 - MSDP creates peering relationship that is made up of a TCP connection

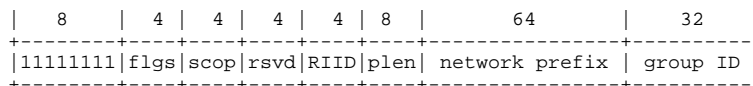
Copyright © Hitoshi Asaeda

52



Embedded RP

- Defined in [RFC3956]
- Extension of unicast-prefix-based address (IPv6) [RFC3306]

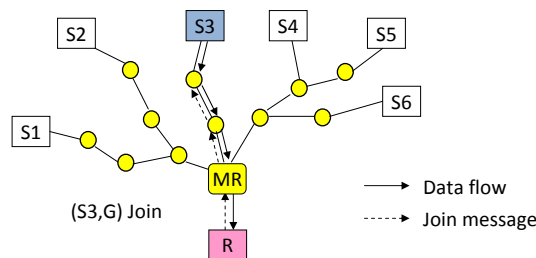


– plen MUST NOT be 0

- Example: If RP address is 2001:db8:1234::5 , then group address the RP supports is ff7e:0530:2001:db8:1234:5678::8000

Source-Specific Multicast (SSM)

- Host specifies (S,G) addresses to join/leave a session
- No Rendezvous Point router (RP)
- Simple tree coordination
 - Known as a deployable multicast communication model



That's all for today

Copyright © Hitoshi Asaeda

57