

S-38.3191 Verkkopalvelujen tuotanto S-38.3191 Network Service Provisioning

Lecture 5: Multicast
23.10.2008

Juha Järvinen @ netlab.tkk.fi

Mostly based on material by
Lic.Tech. Marko Luoma

Communication modes

- **Unicast**

- One sender
- One receiver

- **Broadcast**

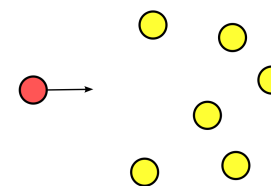
- One sender
- All receive

- **Multicast**

- One sender
- N receivers

- **Anycast**

- One sender
- One receiver on a group



Forwarding modes

- **Point-to-point**

- From single ingress to single egress
- 'link'

- **Point-to-multipoint**

- From single ingress to multiple egress
- 'tree'

- **Multipoint-to-multipoint**

- From multiple ingress to multiple egress
- 'mesh'

Multicast

- **It is all about determining where**

- are the receivers
 - Is the knowledge of receivers maintained
 - Broadcast vs multicast
- the replication of packets should be done
 - Is there knowledge about service within the network
 - Source replication vs network replication

Multicast forwarding

▪ Problem

- How to form a point-to-multipoint structure that is efficient
 - e.g. Shortest path tree

▪ Reverse path forwarding

- Logic:
 - Send out replicate of the packet to all interfaces except to the one with the shortest path to the sender
 - That is where the packet should have come in the first place
 - Dropped duplicates along the way on the network
 - Wasted resources
 - From the sender to all other stations on the network
 - Wasted resources

Forwarding

▪ Unicast forwarding

- Forwarding of packets towards destination via the shortest path
 - SPF is based on weighting of IGP

▪ Reverse path forwarding

- Forwarding of packets away from the shortest path to the sender
 - No knowledge of the destination only about the source

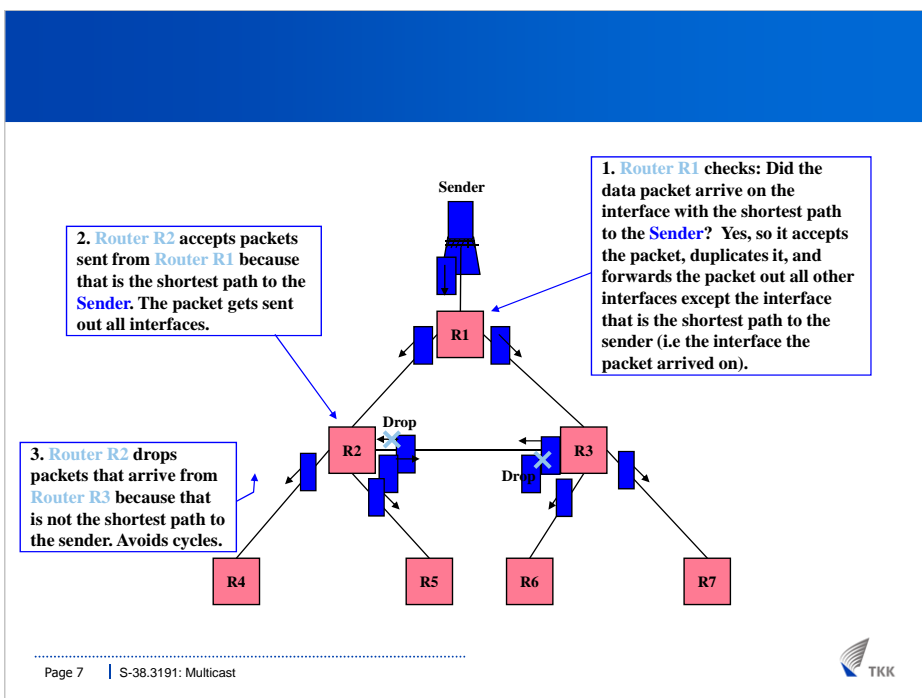
▪ Reverse path forwarding check

- Accept packets only from the link which points to SPF towards source

Building an intelligent tree

▪ Requirements

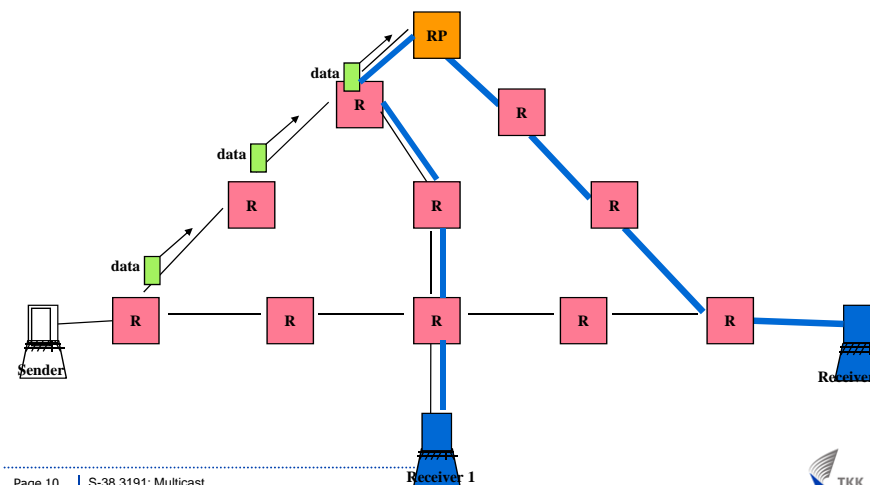
- We know where the receivers are
 - Group management protocol (IGMP)
 - Receivers express their existence to the network
 - join, leave
- We are able to communicate that knowledge within the network
 - Multicast routing protocol (PIM)
 - Activation of RPF on selected downstream links



Building an intelligent tree

- If we want to minimize the state information within the core we can build a **shared tree** for each sender

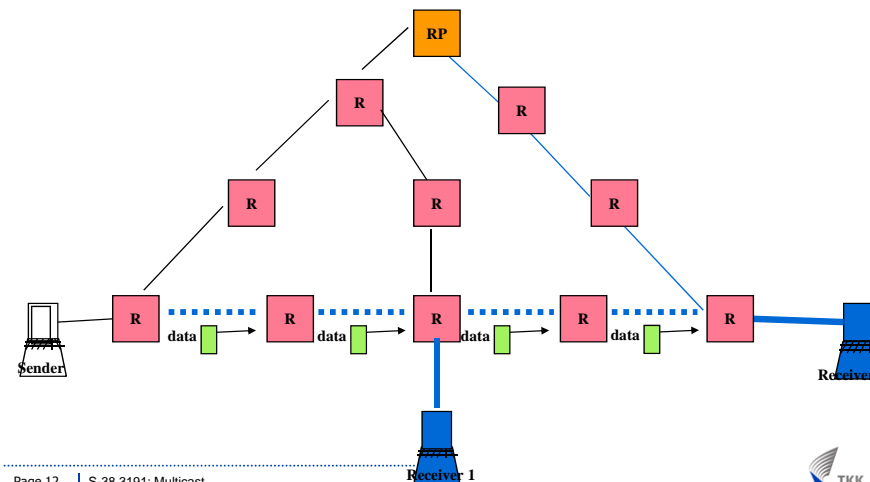
- A single point within the network acts as a relay agent
 - Traffic in unicast sent to relay point which then multicasts it forward in RPF-tree
- Efficiency depends on the distribution of sources and receivers in relation to location of relay-point
 - Usually very unnefficient due to sparse nature of receivers and variable locations of sources



Building an intelligent tree

- If we want to achieve maximal efficiency, we must use **shortest paths toward source** of the information not towards common replication point

- Use explicit source rooted RPF-tree
 - Joining/grafting to multicast tree at the lowest level possible
 - Pruning from the tree if there are no receivers below

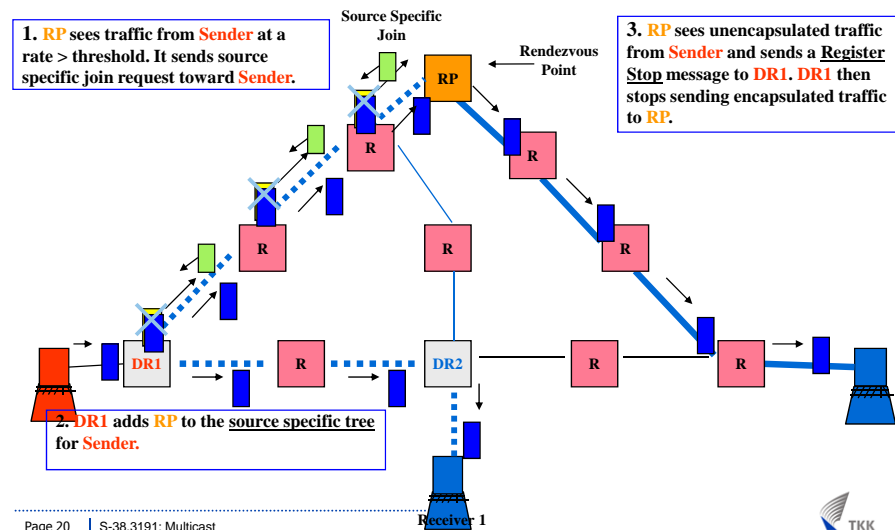
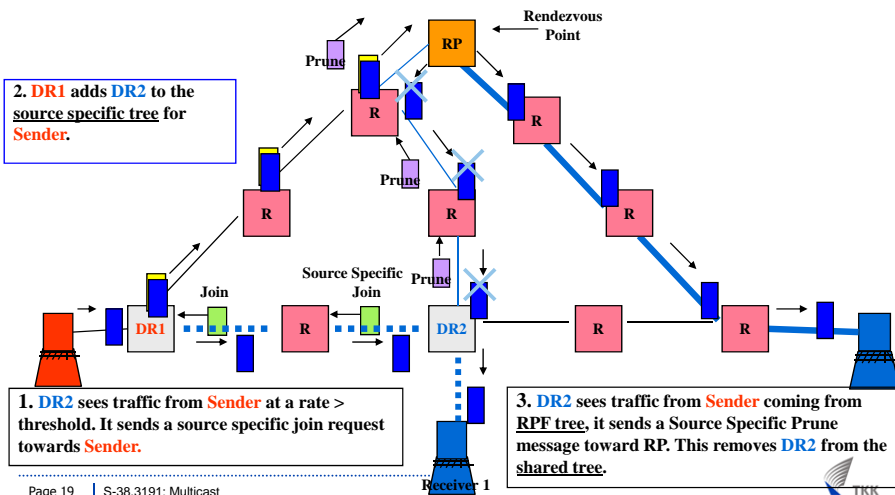
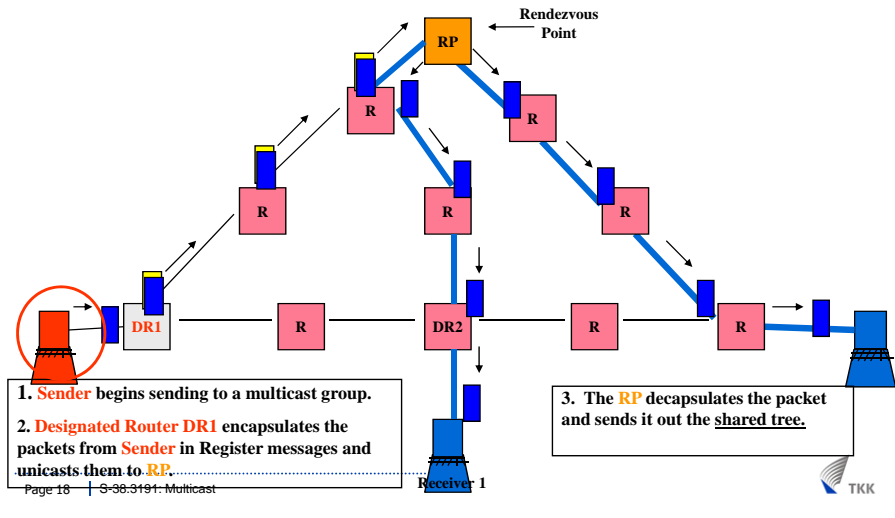
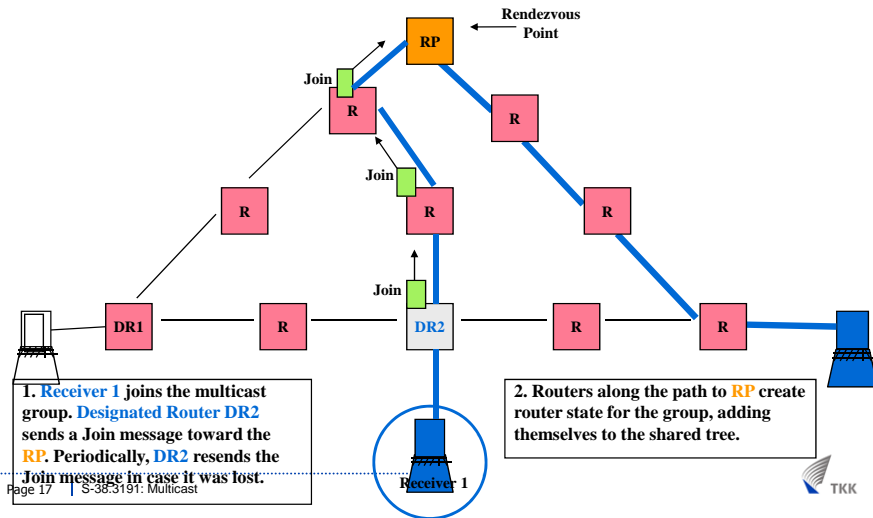


- **Is a combination of shared tree and source specific tree operation**
 - Infrequent packets are replicated in a single point (shared tree)
 - Rendezvous point
 - High data volumes are transferred in more optimal way (source specific tree)
 - Decision based on predefined policy
 - Data rate greater than x
 - Amount of bytes greater than x

- **Different Modes**
 - Sparse Mode (SM)
 - is the most widely deployed as of 2006
 - Dense Mode (DM)
 - Source Specific Mode (SSM)
 - Bidirectional Mode (Bidir) [Also commonly known as Sparse-Dense Mode (SDM)]
- **SSM and Bidir are simpler and more scalable variations developed more recently and gaining in popularity**

- **Is a protocol for efficiently routing to multicast groups that may span wide-area (and inter-domain) internets.**
- **It is not dependent on any particular unicast routing protocol for topology discovery**
 - sparse-mode because it is suitable for groups where a very low percentage of the nodes (and their routers) will subscribe to the multicast session.
 - Unlike earlier dense-mode multicast routing protocols such as DVMRP and PIM-DM which flooded packets everywhere and then pruned off branches where there were no receivers, PIM-SM explicitly constructs a tree from each sender to the receivers in the multicast group. Multicast packets from the sender then follow this tree.

- **Shared trees**
 - Uses less resources from the network
 - Amount of state information is dependent of $N(*,G)$
 - All sources (*) are mapped into single distribution tree per group (G)
 - Usually suboptimal path from the source to all receivers within the group
 - » Unless RP is rooted at the source
- **Source specific tree**
 - Takes more resources from the network
 - Amount of state information is dependent of $N(S,G)$
 - Every source (S) has its own distribution tree per group (G)
 - Optimal path from the source to all receivers within the group
 - Minimizes the IGP cost towards source from each receiver
 - Optimal for one-to-many distribution (point-to-multipoint)



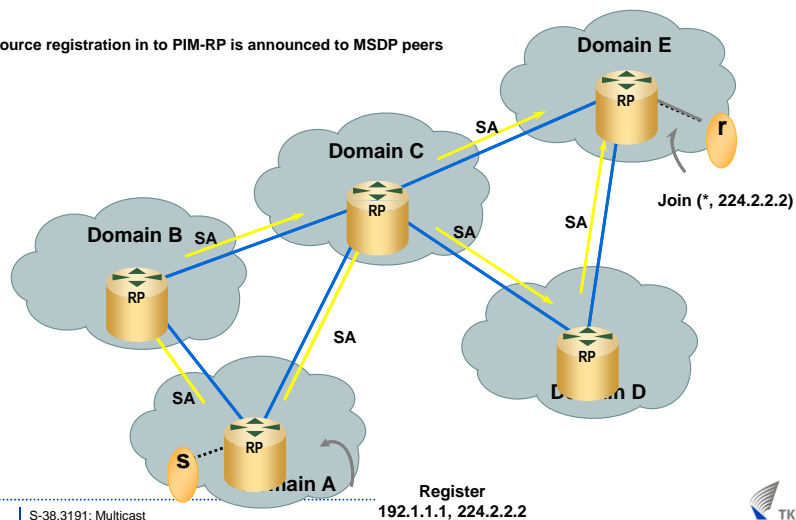
Rendezvous Point (RP)

- **Location of RP can be based on**
 - Static configuration of served groups
 - Bootstrap process with priorities
 - Anycast operation
- **Bootstrap Router (BSR)**
 - Dynamically elected (like OSPF DR election process)
 - Constructs a set of RP IP addresses based on received Candidate-RP messages

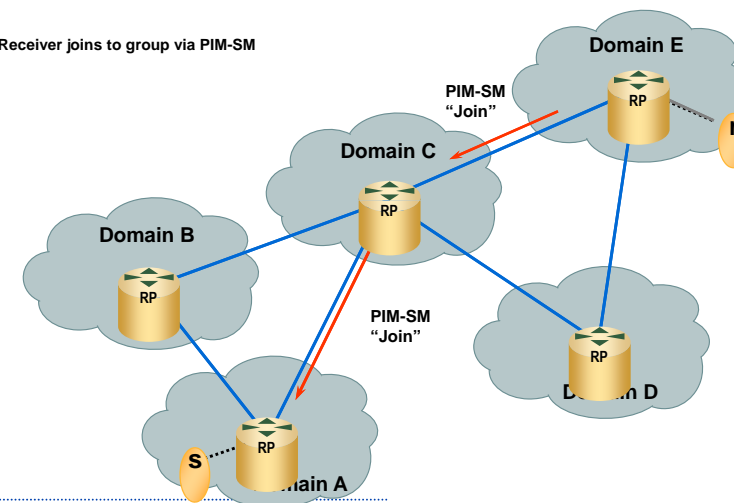
Interdomain multicast

- **Need to transfer multicast routing information across domain borders**
 - Active groups (senders and receivers)
 - Multicast Source Discovery Protocol (MSDP)
 - Unicast routes for RPF
 - Multicast BGP (MBGP)

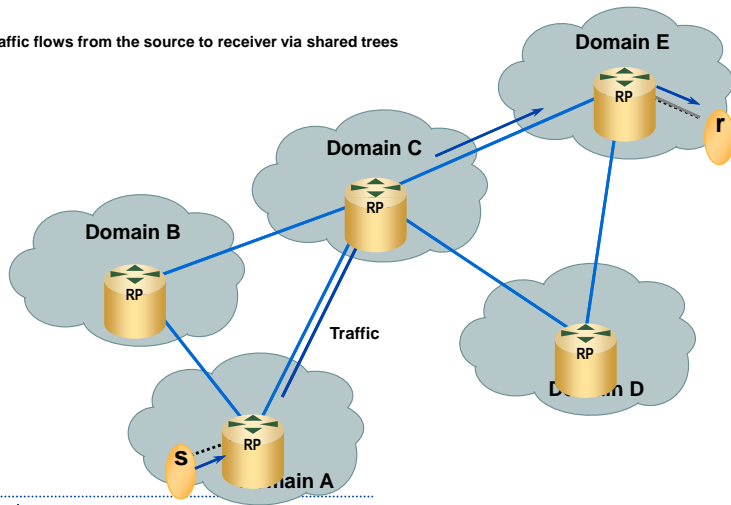
1. Source registration in to PIM-RP is announced to MSDP peers



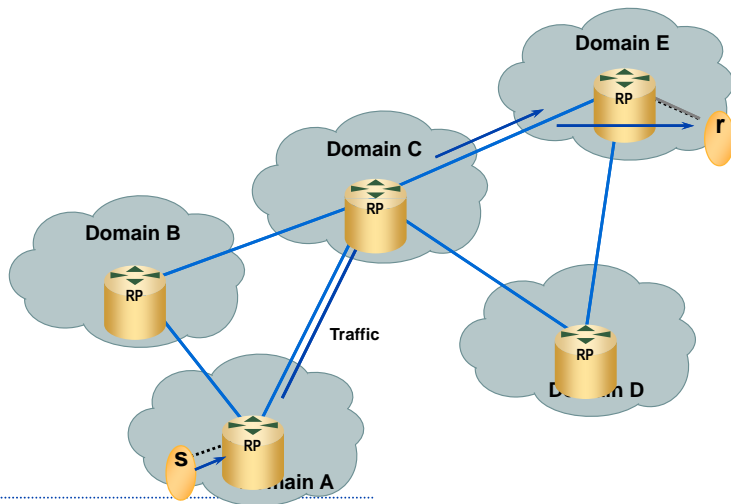
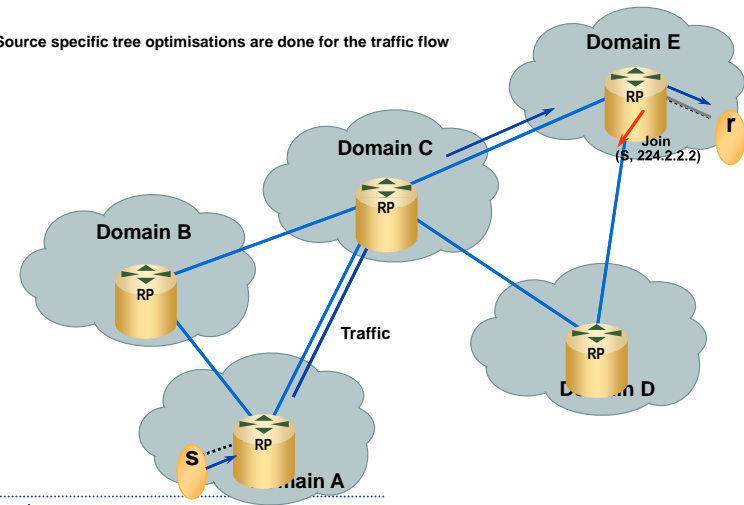
2. Receiver joins to group via PIM-SM



3. Traffic flows from the source to receiver via shared trees



4. Source specific tree optimisations are done for the traffic flow



MBGP

▪ Multiprotocol Extensions to BGP (RFC 2283).

- Tag unicast prefixes as multicast source prefixes for intra-domain multicast routing protocols to do RPF checks.
- Why same routes two times
 - Allows for inter-domain RPF checking where unicast and multicast paths are non-congruent
 - Inter-provider relationships
 - Policies ;-)

MBGP

- **Different NLRI (Network Layer Reachability Information) for unicast and multicast routes**
 - Address Family Information (AFI) = 1 (IPv4)
 - SAFI = 1 (NLRI is used for unicast)
 - SAFI = 2 (NLRI is used for multicast RPF check)
 - SAFI = 3 (NLRI is used for both unicast and multicast RPF check)
- **Allows for different policies between multicast and unicast**
 - For example different ingress/egress point for unicast and multicast traffic



Multicast over Layer 3 VPNs

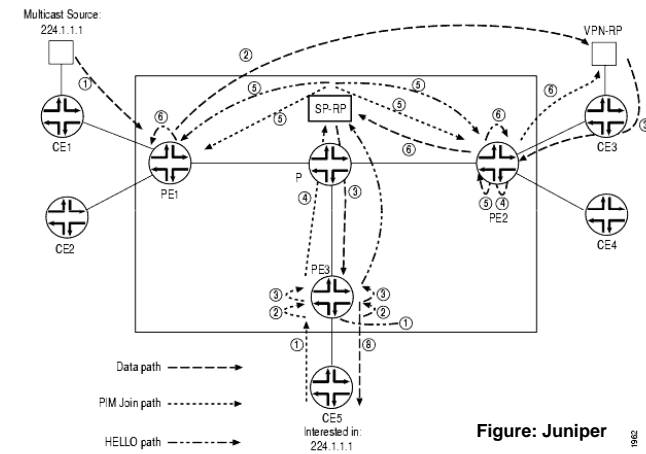


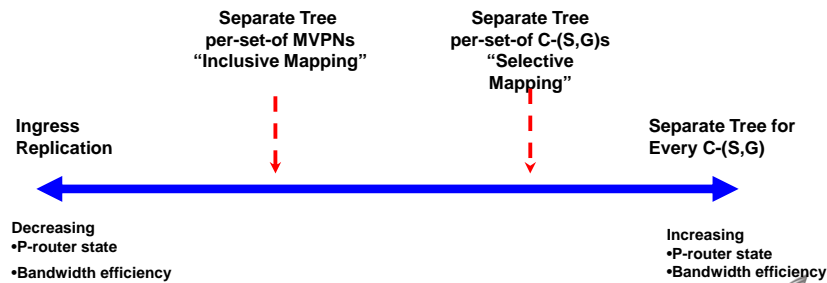
Figure: Juniper

<http://www.juniper.net/techpubs/software/junos/junos62/feature-guide-62/html/fg-l3vpn-multicast2.html>



MPLS VPN multicast

- **Multicasting in MPLS VPN is not straight forward**
 - Optimal replication requires state in core routers (P)



MPLS VPN multicast

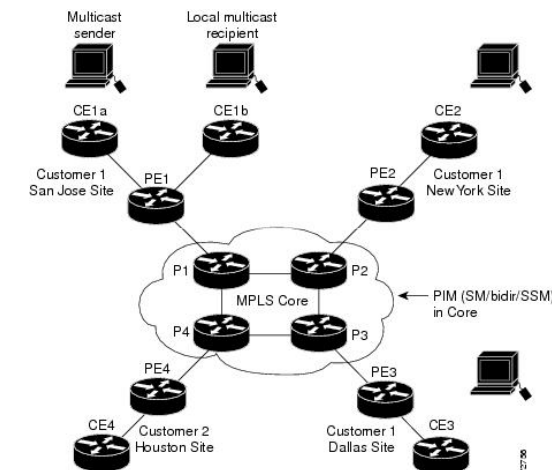


Figure: Cisco

http://www.cisco.com/en/US/docs/ios/12_2t/12_2t13/feature/guide/ftmltvpn.html

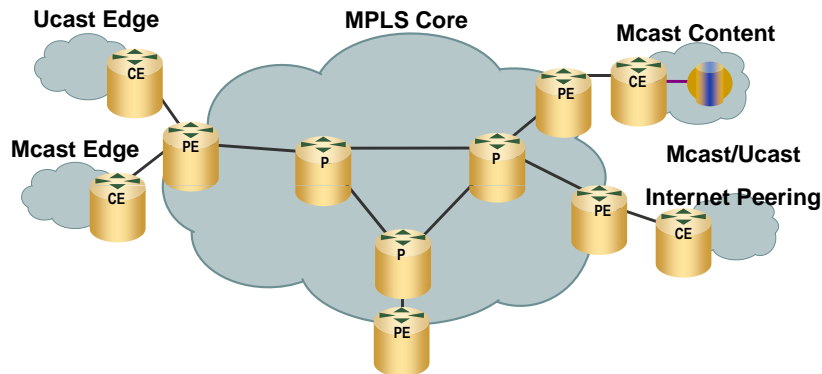


Scalability

- **Optimizing performance**
 - A given customer (multicast) packet should traverse a given service provider link at most once
 - Deliver customer multicast traffic to only PEs that have (customer) receivers for that traffic
 - Deliver customer multicast traffic along the "optimal" paths within the service provider (from the ingress PE to the egress PEs)
- **Optimizing state-space**
 - The amount of state within the service provider network required to support Multicast in 2547 VPN service should be no greater than what is required to support unicast in 2547 VPN service
 - The overhead of maintaining the state to support Multicast in 2547 VPN service should be no greater than what is required to support unicast in 2547 VPN service

Data plane

- **How to tunnel multicast traffic in Service Provider network**
 - By using p2mp LSP's
 - Static distribution trees
 - GRE encapsulation and IP multicast forwarding
 - Multicast distribution trees
 - Aka multicast service
 - Unicast tunneling with ingress replication
- **How to exchange multicast routing information with service provider network**
 - MBGP
 - PIM
 - MSDP
 - Tunneling through



Multicast in use

- **Wikipedia:** "Recently the BBC has begun encouraging UK-based ISPs to adopt Multicast onto their networks by providing BBC Radio at higher quality than is available via their Unicast delivered services. This has also been supported by a variety of commercial radio networks - including Virgin Radio, GCAP and EMAP."
- Some operators use Multicast as an internal service.
- More in the future?

Summary

- **“Multicast is a network addressing method for the delivery of information to a group of destinations simultaneously using the most efficient strategy to deliver the messages over each link of the network only once, creating copies only when the links to the multiple destinations split (typically network switches and routers). Multicast is often used for streaming media and Internet television applications.”**
- **Path between source and destination is not always optimal**
- **It is possible**
 - to use Multicast between different domains
 - to create Multicast VPNs