L70 - Multicasting Fundamentals

Multicasting Fundamentals

Agenda

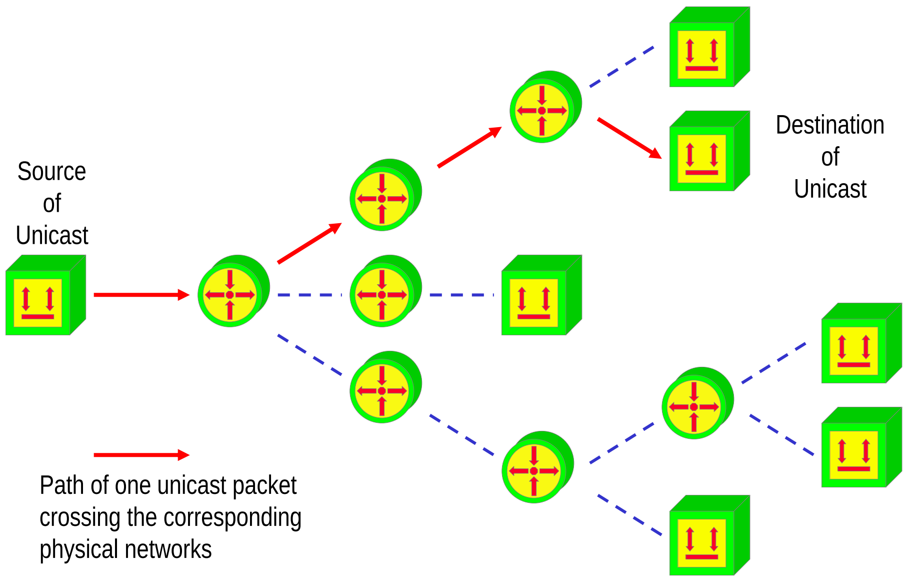
- **Introduction Multicasting**
- **Multicast Routing Principles**
  - Flooding
  - Spanning Tree
  - Reverse Path Forwarding (RPF)
  - Core Based Tree (CBT)
  - Dense Mode / Sparse Mode
  - Multicast Distribution Trees
    - Source Trees
    - Shared Trees
    - Bidirectional / Unidirectional Shared Trees

What is Multicasting?

- **Unicast transmission**
  - transmitting a packet to one receiver
  - point-to-point transmission
  - used by most applications today
- **Multicast transmission**
  - transmitting a packet to a group of receivers
  - point-to-multipoint transmission
  - used for example by multimedia applications in experimental environment nowadays (MBONE)
  - will be the future base-technology for multimedia applications (video conferences, legacy television etc.) in the Internet

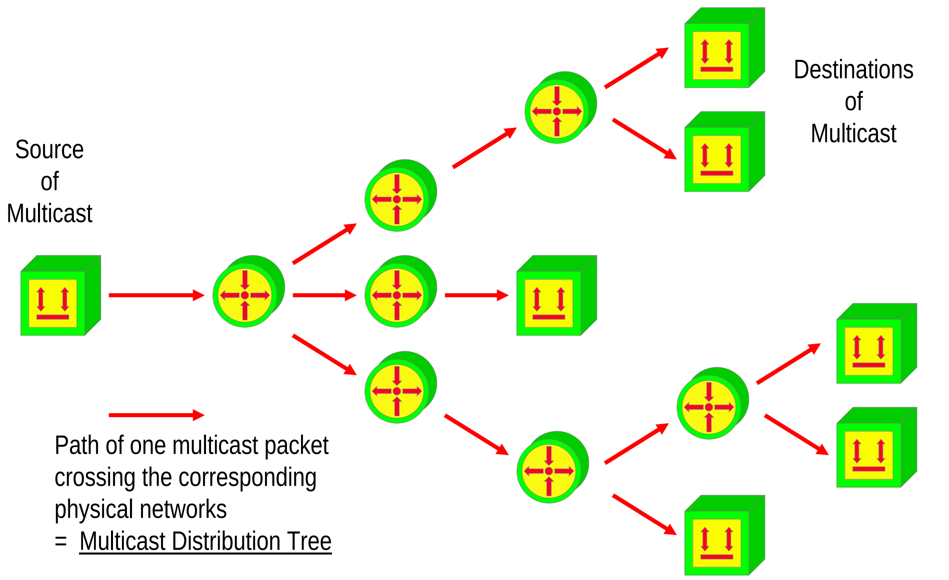
 MC_Fundamentals, v4.4 3

Unicast Transmission



 MC_Fundamentals, v4.4 4

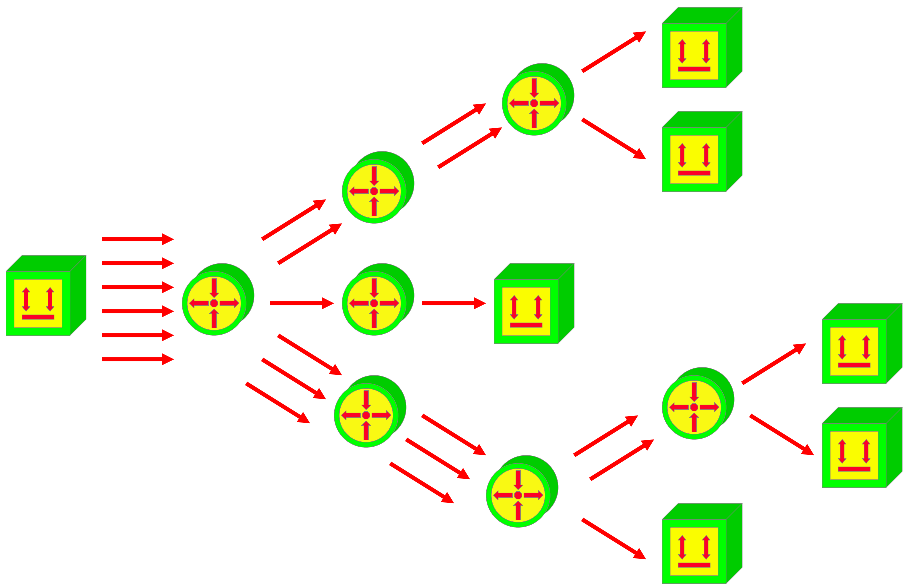
Multicast Transmission



Why Multicasting?

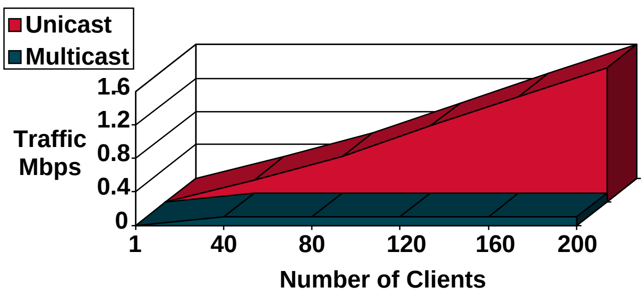
- **Sending to a group of receivers**
  - can be done in principle by sending multiple copies of the same packet using normal unicast transmission
    - multiple unicast packets
  - the same packet will appear more than once on one physical network
- **Compared to multiple unicast packets multicasting minimizes**
  - link bandwidth consumption
  - network load
  - sender and router processing
  - delivery delay

Multiple Unicast Packets



Comparison vs. Multicast vs. Unicast

- **Example: Streaming**
  - All clients listening to the same stream (e.g. 8 Kbps audio)



Why Multicasting?

- **Multicasting fits well**
  - with multimedia conferences
- **Multicasting allows**
  - easier handling of variable membership of listeners than doing it with a set of point-to-point addresses
- **Multicasting can be used**
  - when individual addresses of receivers are unknown or changeable
    - for locating services or functions
    - for advertising services or functions
    - resource discovery

Why Multicasting?

- **For such requirements multicasting is simpler and more robust than alternatives**
    - centralized directory servers
    - configuration files with high administrative overhead
    - exhaustive search
- **Multicasting has been used already**
  - for resource discovery in local networks
    - e.g. finding TFTP server, DNS server, BOOTP/DHCP server, all OSPF routers, all RIPv2 routers etc.
    - using limited IP broadcast (255.255.255.255) or certain class D IP addresses (224.x.x.x)
    - scope of these multicasts were limited to local network
      - TTL is set to 1

Components of Multicasting

- **Group address**
  - identifies multicast traffic for a group of receivers
- **Source host**
  - sender of multicast packets
  - using group address as destination address
- **Multicast routers**
  - responsible for forwarding of multicast packets to all networks where members of a certain group are located
  - duplicate packets if necessary
- **Destination hosts**
  - receiver of multicast packets
  - listen to group address

 MC_Fundamentals, v4.4 11

MBONE

- **MBONE**
  - experimental Multicast BackbONE of the Internet
    - interconnected set of multicast routers
    - virtual network overlaid on the Internet
    - overlay network using its own multicast routing protocols
  - multicast island
    - consists of multicast hosts and multicast routers
    - provides multicast service for testing multimedia applications
  - multicast islands are connected via tunnels
    - virtual point-to-point links using normal IP unicast routing for transport of IP encapsulated multicast traffic
    - tunnels carry multicast routing traffic too, in order to enable multicast routing in the MBONE

 MC_Fundamentals, v4.4 12

Multicast Routing

- **Multicast routers must be able to deliver multicast datagram's over the network**
  - several principle options for multicast routing were developed and tested in the MBONE during the last years in order to gain experience with IP multicasting
    - Flooding
    - Spanning Tree
    - Reverse Path Forwarding (RPF)
    - Core Based Tree (CBT)
    - Dense Mode / Sparse Mode
    - Multicast Distribution Trees

Agenda

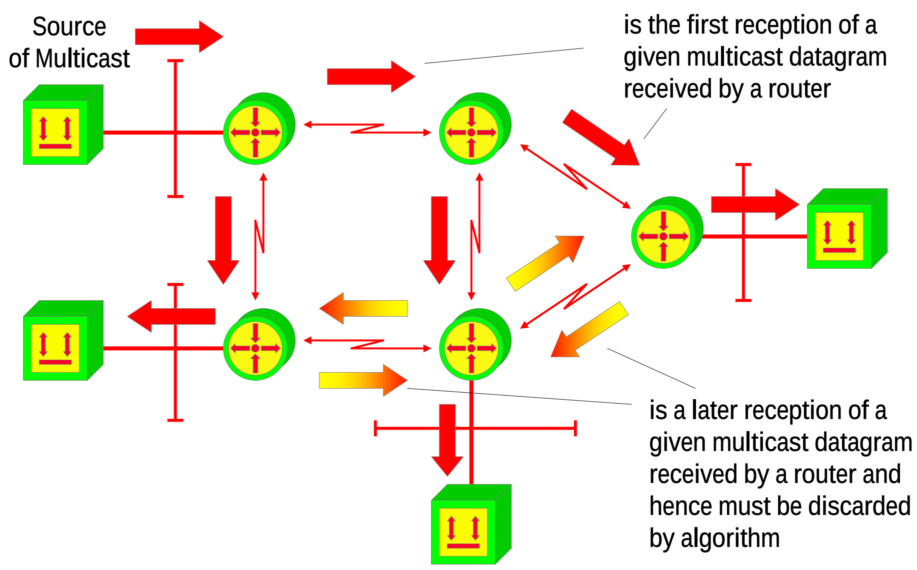
- **Introduction Multicasting**
- **Multicast Routing Principles**
  - Flooding
  - Spanning Tree
  - Reverse Path Forwarding (RPF)
  - Core Based Tree (CBT)
  - Dense Mode / Sparse Mode
  - Multicast Distribution Trees
    - Source Trees
    - Shared Trees
    - Bidirectional / Unidirectional Shared Trees

L70 - Multicasting Fundamentals

Flooding

- **simplest multicast routing algorithm**
  - node receives a packet sent to multicast destination
  - if particular packet is the first reception
    - forward a copy of the packet to all interfaces except the incoming interface
  - if particular packet is a duplicate
    - discard it
  - ensures that particular packet reaches all network destinations
- **well known technique**
  - e.g. used for propagation of LSA to all OSPF routers in an area
  - e.g. propagation of "Usenet" news

Flooding



Flooding

- **flooding principle**
  - does not depend on any kind of routing table
- **principle problem with flooding**
  - how to avoid looping and "broadcast-storm"?
    - either delete redundant paths (-> done with spanning tree) or delete duplicate packets
  - but how to test for first reception of a given packet?
- **OSPF solution**
    - LSA identified by sequence number (and age)
    - received LSA compared with entries in local link state database
    - sequence numbers of LSA's used for decision
      - what is first (new) LSA message
      - and what is duplicate (old, already stored in database) LSA message

Flooding

- **possible IP multicast solution**
  - list of recently seen (last-seen) multicast packets
    - containing IP identifier, group address of multicast packet
  - entry of list will be discarded after maximum lifetime of a packet
    - maximum lifetime of IP packets specified by TTL
  - last-seen lists
    - guarantee that a router will not forward a packet twice but do not guarantee that a router will receive a certain packet only once
  - but very memory greedy
    - last-seen list can be fairly long on a high speed network
  - and very transmission greedy
    - use of all available paths instead of just one
  - NOT A PRACTICABLE SOLUTION

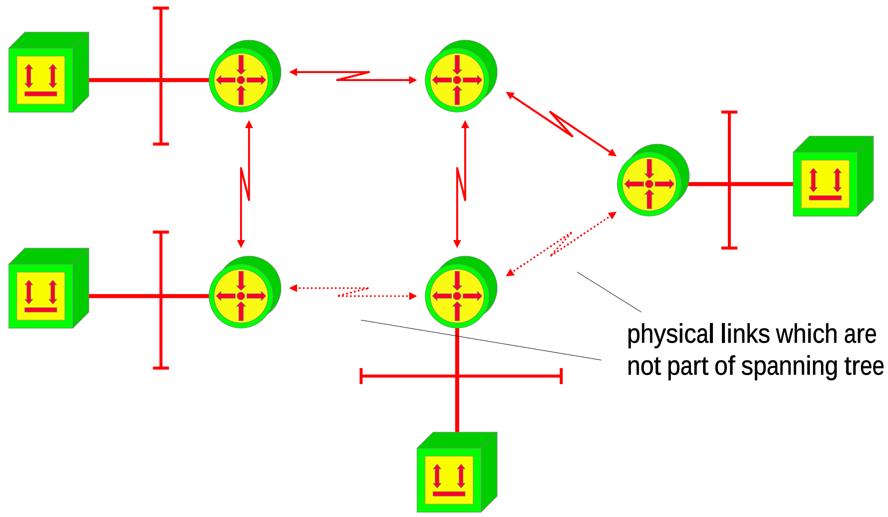
Agenda

- **Introduction Multicasting**
- **Multicast Routing Principles**
  - Flooding
  - Spanning Tree
  - Reverse Path Forwarding (RPF)
  - Core Based Tree (CBT)
  - Dense Mode / Sparse Mode
  - Multicast Distribution Trees
    - Source Trees
    - Shared Trees
    - Bidirectional / Unidirectional Shared Trees

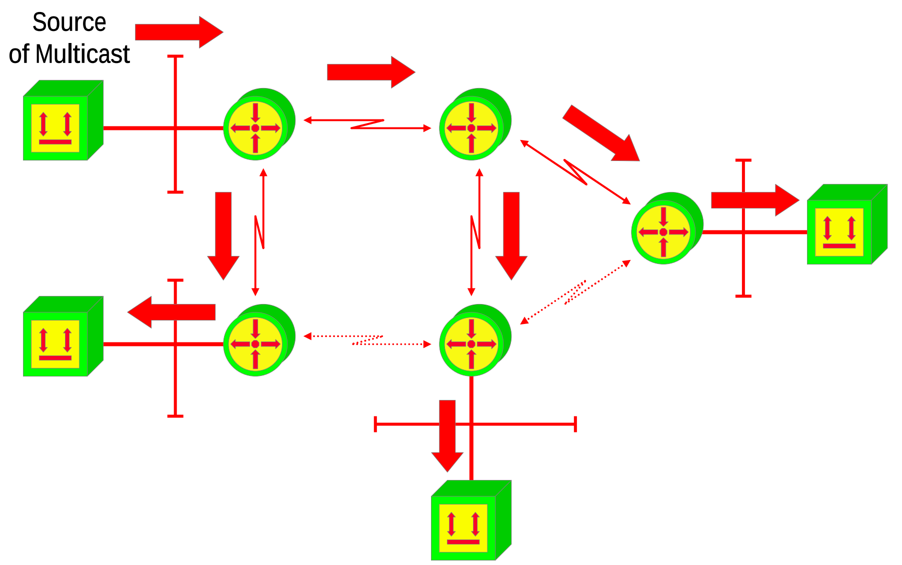
Spanning Tree

- **more efficient solution than flooding**
  - loop avoidance by marking some links as part of a tree and other links as unused
  - set of selected links forms a loop-less tree that spans to all nodes in the network
  - done by spanning tree protocol communication between multicast routers
  - multicast datagram's follow the tree
    - will not loop
    - all networks are reached
    - visit every network only once
  - does not depend on any kind of routing table

Spanning Tree



Spanning Tree

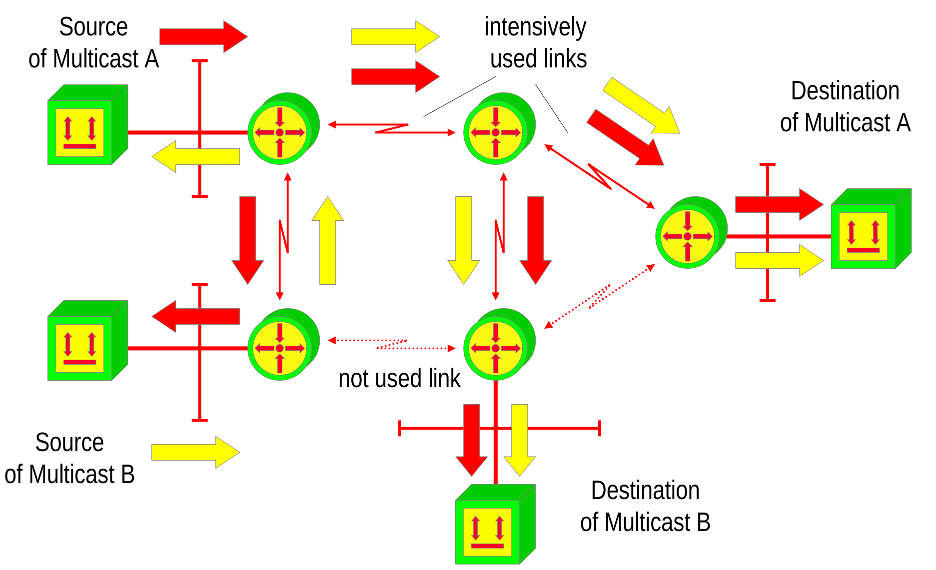


L70 - Multicasting Fundamentals

Spanning Tree

- **well known technique**
  - used for transparent IEEE 802.2 MAC bridges
  - spanning tree built during network initialization
    - first selecting a root bridge as center of the spanning tree
    - then marking links that are on the shortest path between any other bridge and that center
- **two drawbacks for multicasting**
  - concentrates traffic into same subset of network links
  - group membership of receivers is not taken into account
    - can lead to suboptimal propagation of packets

Spanning Tree Drawbacks



Agenda

- **Introduction Multicasting**
- **Multicast Routing Principles**
  - Flooding
  - Spanning Tree
  - Reverse Path Forwarding (RPF)
  - Core Based Tree (CBT)
  - Dense Mode / Sparse Mode
  - Multicast Distribution Trees
    - Source Trees
    - Shared Trees
    - Bidirectional / Unidirectional Shared Trees

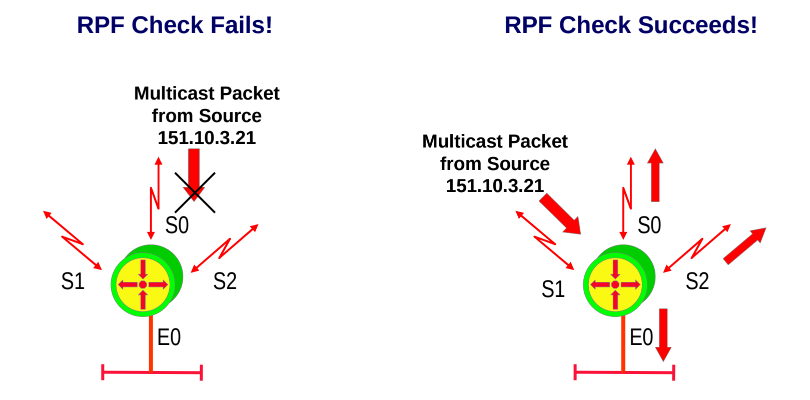
Reverse Path Forwarding (RPF)

- **drawbacks of spanning tree teach**
  - better to use an routing algorithm that computes tree according to location of the source -> source tree
  - leads to Reverse Path Forwarding (RPF)
- **principle of RPF**
  - if multicast packet is received
    - note source network (S) and incoming interface (I)
  - if interface I belongs to the shortest path towards S
    - forward multicast packet to all interfaces except I
  - otherwise
    - discard the packet

Reverse Path Forwarding (RPF)

- **RPF algorithm of a given multicast router**
  - uses a routing table to consider network topology
  - computes implicit spanning tree per network source
- **routing table can be**
  - separate multicast routing table
    - calculating shortest path from S to the router
    - necessary when links are not symmetric
      - forward path metric different from reverse path metric
    - separate routing process for multicasts needed
  - normal unicast routing table
    - calculating shortest path from the router to S
    - sufficient when links are symmetric
    - unicast routing process can be used for multicasts too

RPF-Check Based on Unicast Routing Table



Packet arrived on wrong interface!

Discard Packet.

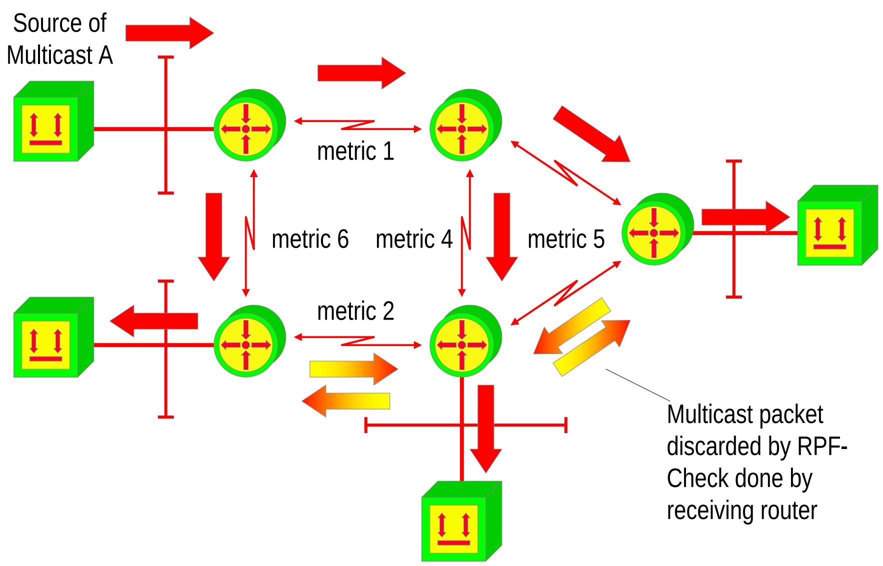
Routing Table

Network	Interface
151.10.0.0 /16	S1
198.14.32.0 /24	S0
204.1.16.0 /24	E0

Packet arrived on correct interface!

Forward Packet.

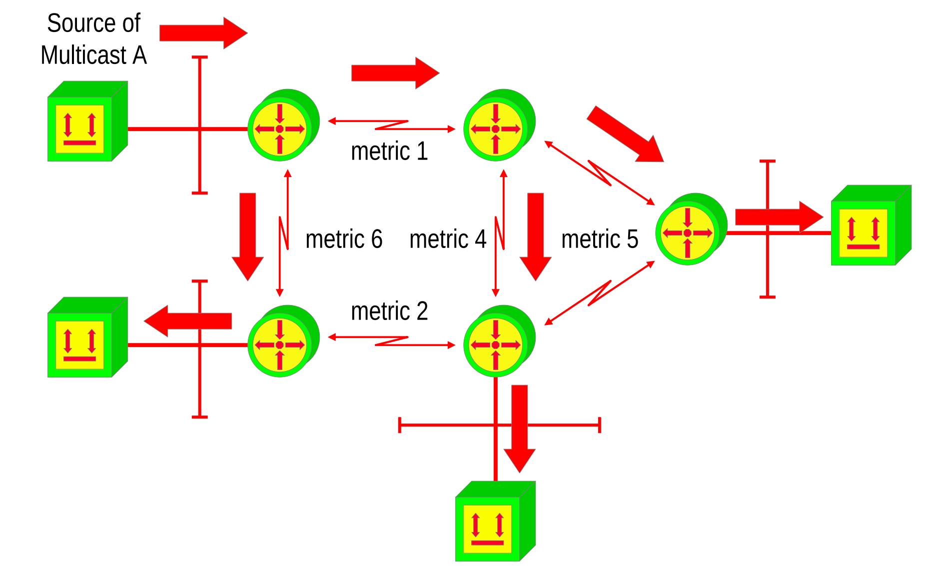
RPF Tree for Group A



Reverse Path Forwarding (RPF)

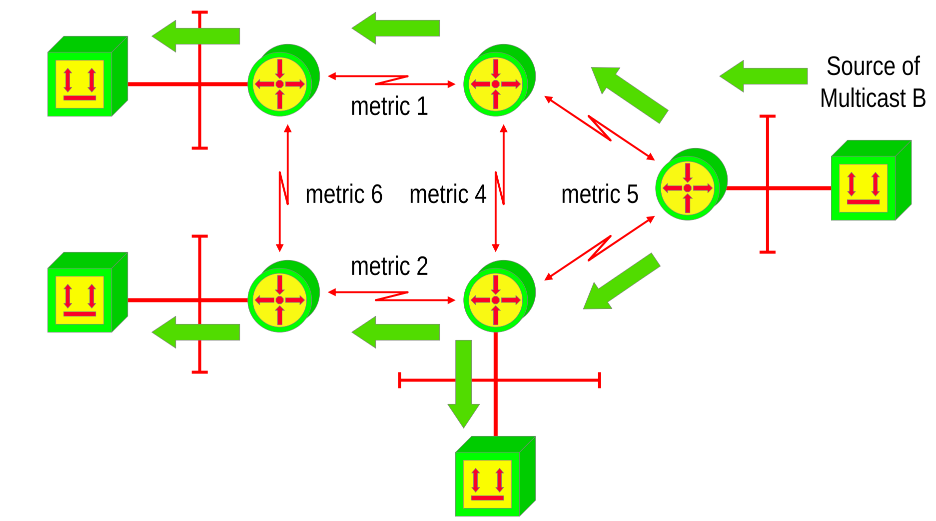
- **RPF algorithm could be improved**
  - if a router is able to look one-step-further
  - packets are forwarded to neighboring router only
    - when the local router is on the shortest path between source S and the neighboring router
  - is sometimes easily available
    - when using link state routing technique like OSPF
  - one-step-further information in OSPF
    - requires only one bit of additional storage per source and interface
  - one-step-further information improves link efficiency
    - by avoiding transmission of multicasts to neighbors which are farther away
      - compare it with flooding example

RPF Tree for Group A (one-step-further)



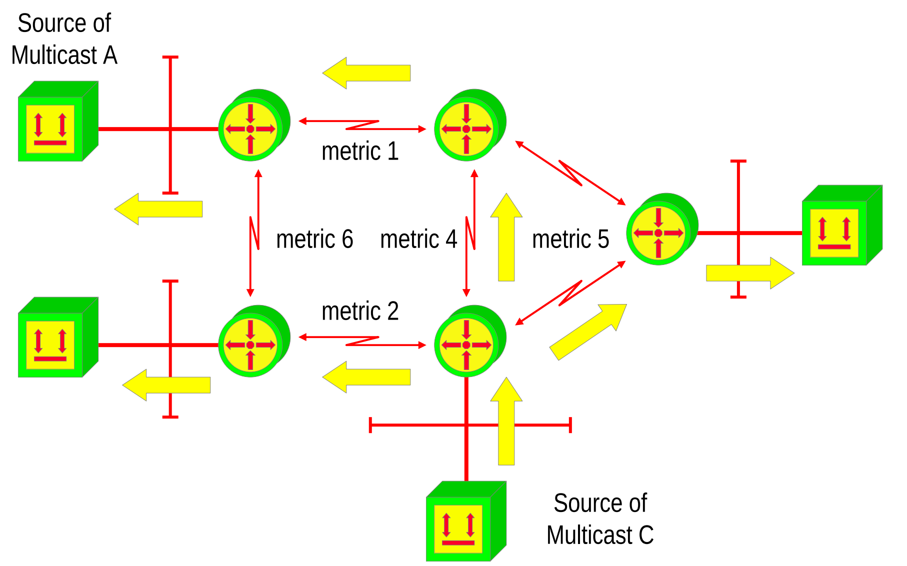
 MC_Fundamentals, v4.4 31

RPF Tree for Group B

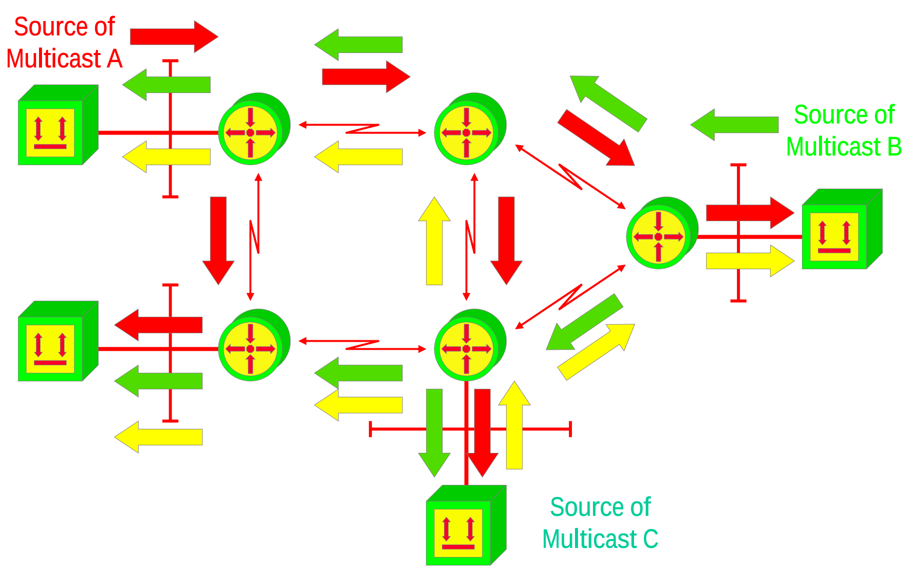


 MC_Fundamentals, v4.4 32

RPF Tree for Group C



RPF Tree for Group A, B, C



Reverse Path Forwarding (RPF)

- **RPF advantages**
  - fastest possible delivery as multicasting follows shortest path from source to destination
  - better network utilization compared to spanning tree because packets are spread over multiple links
- **remaining problem**
  - group membership is not taken into account
    - are there any group members downside the tree at all?
  - packets are flooded through whole network
- **general experience learned from tests**
  - multicast routers should know
    - if there are group members attached to the leaves of the network
    - if multicasts must be forwarded on branches of the tree

RPF and Truncated Broadcast

- **one method for membership discovery in case of IP multicasting**
  - IGMP (Internet Group Membership Protocol)
  - protocol between multicast router and multicast receiver on a local network
  - multicast router query periodically if there are members of a multicast group present on the local network or not
- **if there are no members in the leafs of the network**
  - multicasts for this group should not be transmitted to such a network
  - “truncated broadcast”

RPF with Flood and Prune

- **variant**
  - which also takes in account the group membership when building branches of the tree and use prune (cut off)-message in upstream direction in case no members are active in the leaf
- **principle**
  - first multicast packet of source S is propagated to all nodes (multicast router) of the network ⇨ flooding
  - nodes that have no point behind them in the RPF tree are called leaf nodes ⇨ border of the network
  - leaf nodes use IGMP to discover group members
  - if there is no group member attached to a leaf node a prune message is sent back to the upstream node

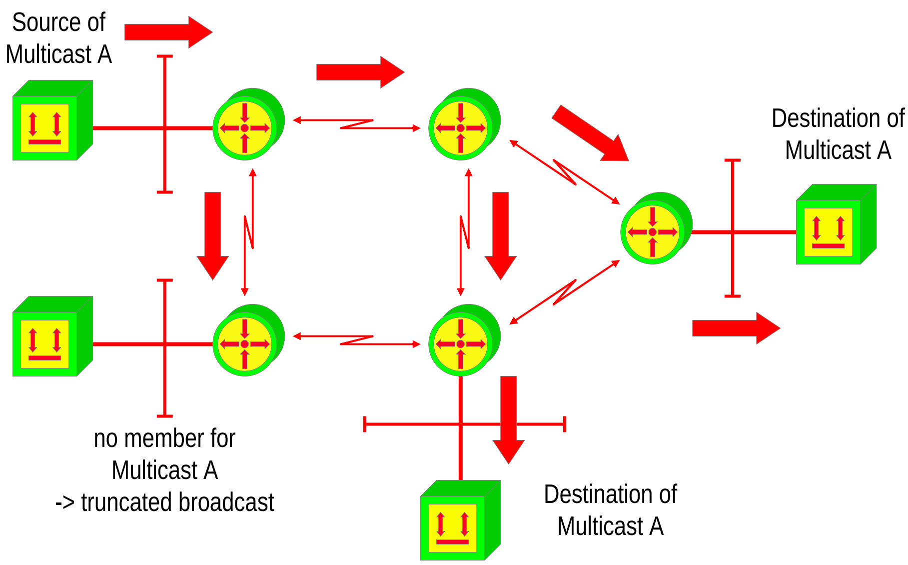
RPF with Flood and Prune

- **principle (cont.)**
  - on reception of a prune message upstream router will not send multicast packets for this group through this interface any longer
  - intermediate routers memorize reception of prune messages of a certain interface
    - a so called “soft” state is created for a specific time period
    - ages out after this period
  - if prune messages are received through all downstream interfaces of the tree intermediate router will sent a prune message to its upstream router too
  - prune method used to decide if multicast packets of a certain group must be forwarded on branches of the tree

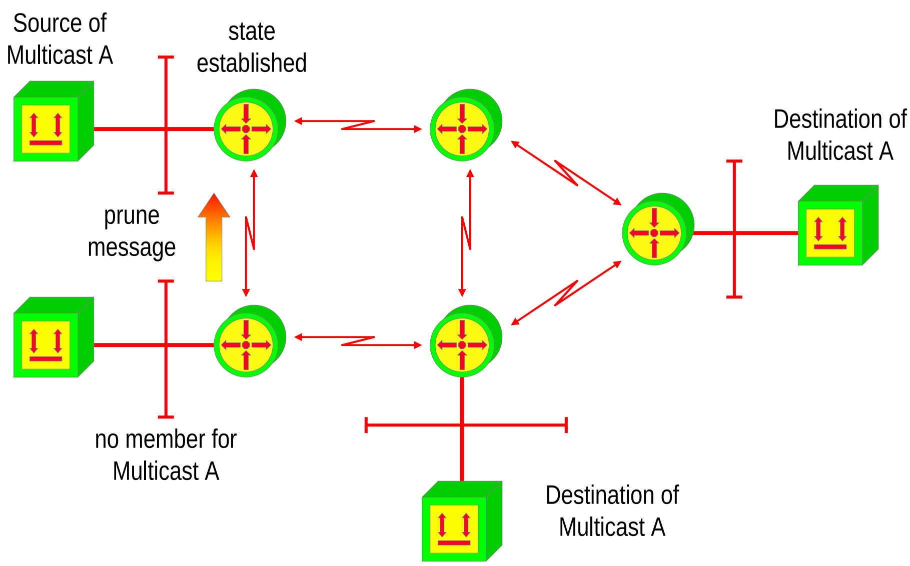
RPF with Grafting

- **joining of new members can be done**
  - either by grafting
    - if a router recognizes a certain group which was already pruned a graft message is sent to the upstream router
    - graft message will delete corresponding state for this group immediately
    - next multicast packet for this group will be transmitted to the downstream router
  - or by removing soft states
    - after timeout and flooding multicast packet again over the whole network
    - tree will be pruned according the new membership distribution

RPF: Flooding First Packet

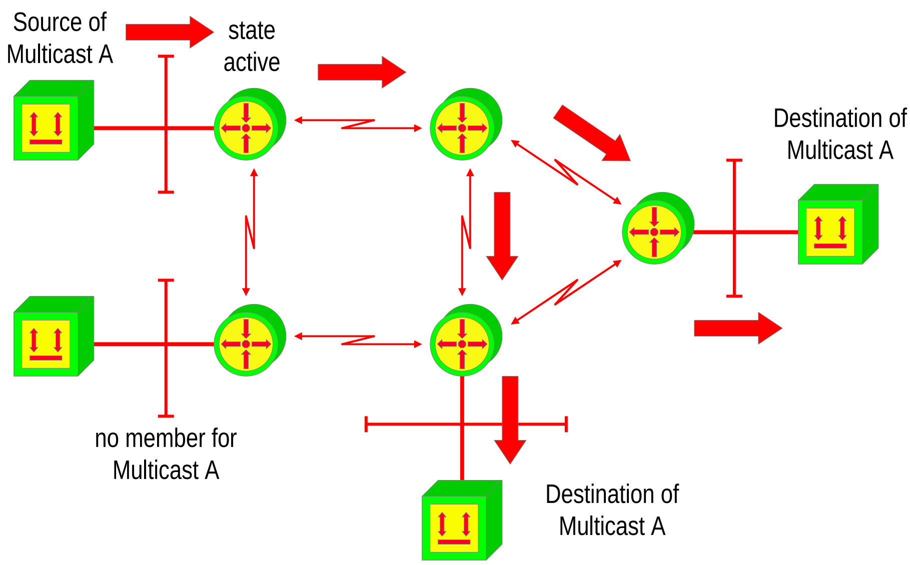


RPF: Pruning



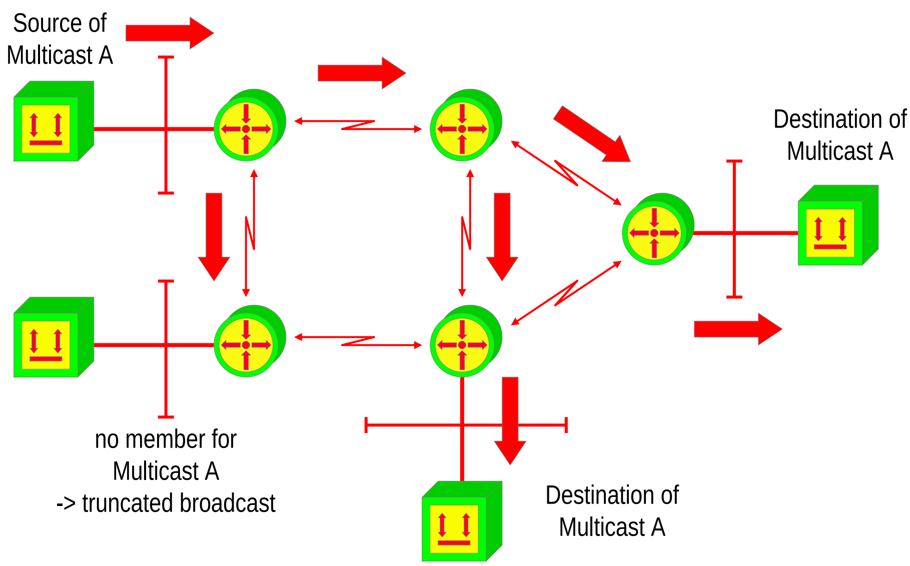
 MC_Fundamentals, v4.4 41

RPF: State Active



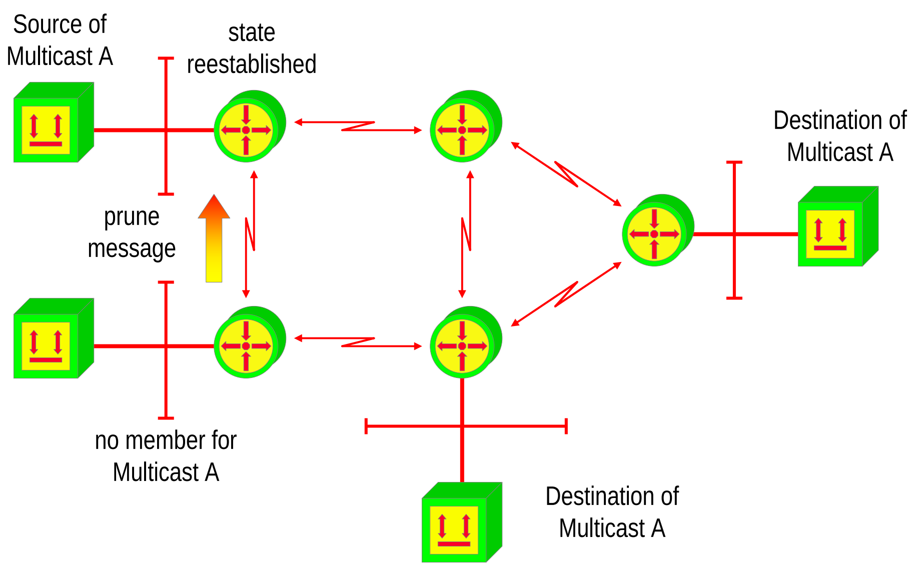
 MC_Fundamentals, v4.4 42

RPF: State Lifetime Exceeded



 MC_Fundamentals, v4.4 43

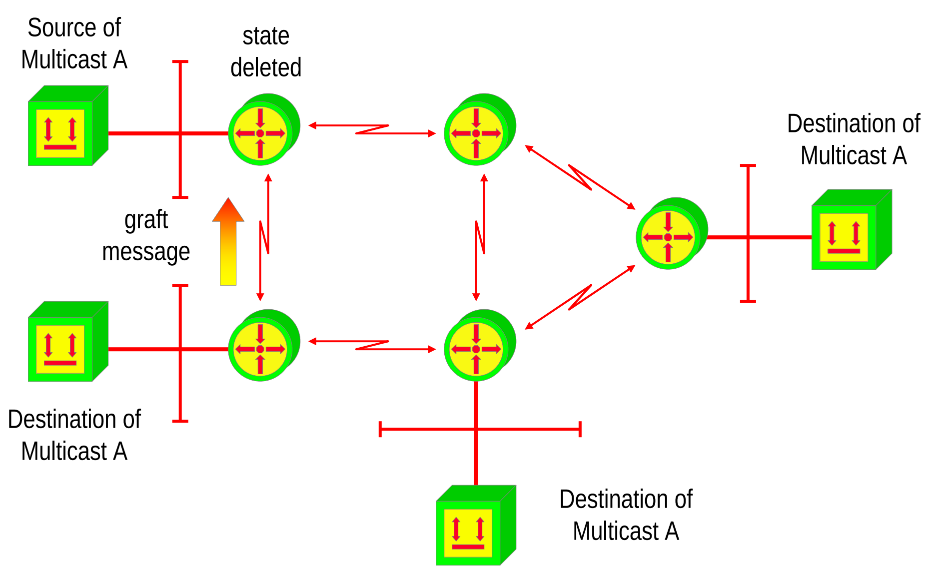
RPF: Pruning Repeated



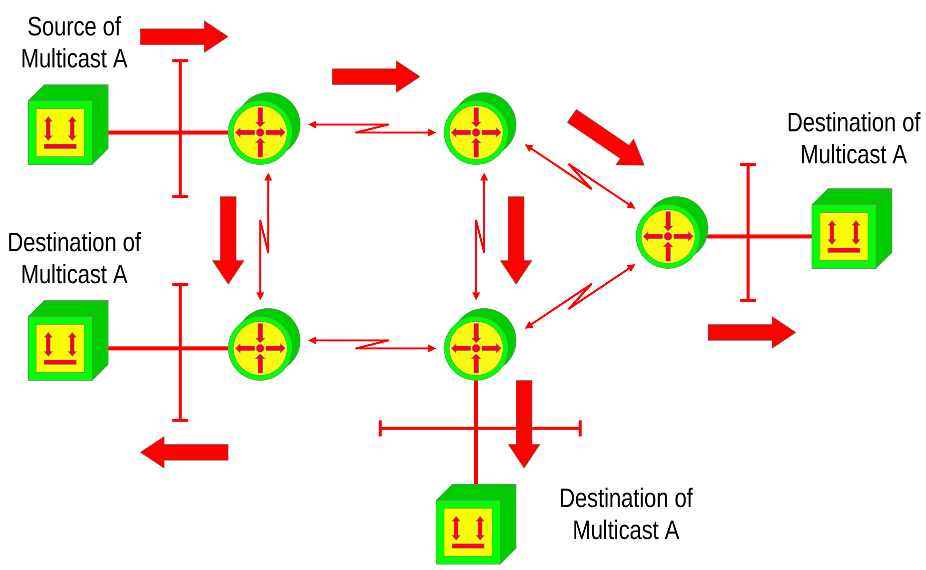
 MC_Fundamentals, v4.4 44

L70 - Multicasting Fundamentals

RPF with Grafting



RPF: Delivering Next Packet



RPF with Flood and Prune

- **drawbacks**
  - first packet is flooded to the whole network
  - routers must keep state information per group and source
  - state must be kept transient
    - varies with joining and leaving of members and with network topology
    - therefore every state information has only limited lifetime
    - if lifetime is exceeded the state is deleted
    - next multicast packet will be flooded again
    - "soft states" are created even if no receivers exist
  - multicast packets for a given source are flooded periodically to the whole network for refreshing or actualizing states

RPF and Rerouting

- **in case of topology change**
  - unicast routing or special multicast routing will converge
  - packets will follow the new shortest paths
  - states on old paths will age out
  - multicast packets are flooded
  - pruning will built new states along new paths

Agenda

- **Introduction Multicasting**
- **Multicast Routing Principles**
  - Flooding
  - Spanning Tree
  - Reverse Path Forwarding (RPF)
  - Core Based Tree (CBT)
  - Dense Mode / Sparse Mode
  - Multicast Distribution Trees
    - Source Trees
    - Shared Trees
    - Bidirectional / Unidirectional Shared Trees

 MC_Fundamentals, v4.4 49

Core Based Trees (CBT)

- **RPF with flood and prune**
  - periodic transmission to all network sites
    - in order to trigger pruning and to keep states active
  - CBT proposed a potential solution to this problem
- **original CBT principle**
  - a fixed point in the network (core-, rendezvous-point) is the center of a multicast group
  - a member sends a join message towards the core using the shortest path to the core
  - intermediate routers process join message
    - if group is already installed incoming interface is marked
    - if group does not exist, a "hard" state for this group is installed and a join message is forwarded one step further to the core

 MC_Fundamentals, v4.4 50

Core Based Trees (CBT)

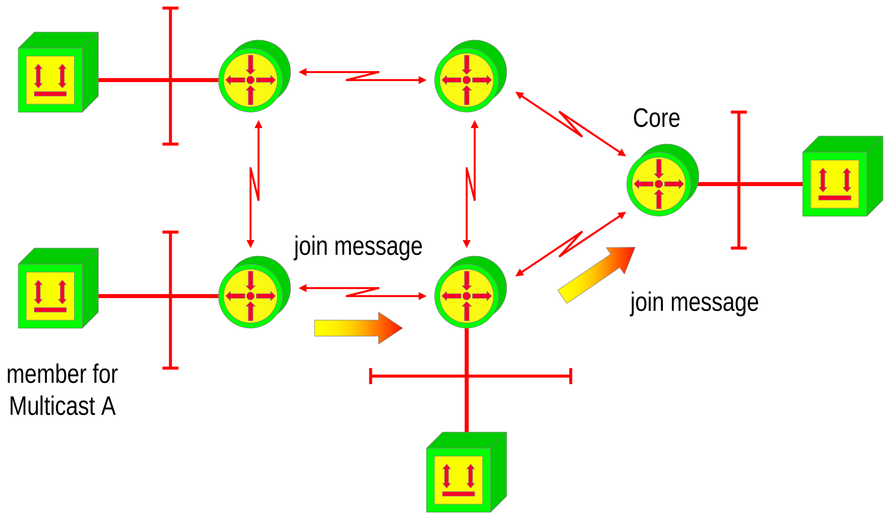
- **original CBT principle (cont.)**
  - in this way joining members create a core tree for a certain group
    - could be compared with establishing a source tree between the core as source and all members of a group
  - if a source transmits multicast packets
    - packets are encapsulated by first-hop multicast router in special register packets and are delivered hop-by-hop to the core using normal unicast transmission
  - core will decapsulate the multicast packet from the register packet
    - and multicasting will proceed along the already established core tree in the same way as it is done along a usual source tree

Core Based Trees (CBT)

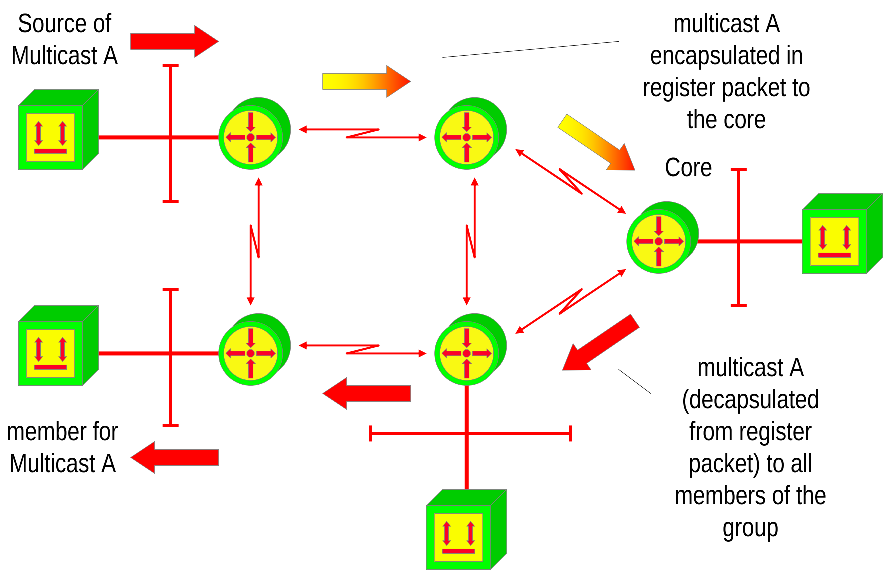
- **CBT advantages**
  - core tree is the same for all sources
    - therefore state information only per group is necessary (RPF needs state per group/source pair)
  - first packet will not be flooded to the whole network and also periodic flooding for refreshing is not needed
    - “hard-states” are acknowledged and repeated after a timeout
    - states are created on demand (if receivers exist)
  - may depend on unicast routing to calculate shortest path
- **CBT disadvantages**
  - path may be suboptimal for some multicast traffic
  - traffic of all sources uses same set of links which means multicast traffic concentration

L70 - Multicasting Fundamentals

CBT: Joining



CBT: Multicast Traffic



CBT and Rerouting

- **in case of a topology change**
  - unicast routing or special multicast routing will converge
  - join packets will be sent on the new shortest path towards the core
  - a new core tree will result
  - packets will follow the new tree

 MC_Fundamentals, v4.4 55

Agenda

- **Introduction Multicasting**
- **Multicast Routing Principles**
  - Flooding
  - Spanning Tree
  - Reverse Path Forwarding (RPF)
  - Core Based Tree (CBT)
  - Dense Mode / Sparse Mode
  - Multicast Distribution Trees
    - Source Trees
    - Shared Trees
    - Bidirectional / Unidirectional Shared Trees

 MC_Fundamentals, v4.4 56

Types of Multicast Protocols

- **Dense-mode**
  - Flood and Prune behavior
    - Multicast traffic is flooded through whole network
    - Routers with no listeners cut branches to stop traffic
    - Pruned branches can later be grafted to reduce join latency
  - Similar to radio broadcast
  - Best for large number of members in a large number of places
- **Sparse-mode**
  - Explicit Join behavior
    - Assumes no one wants traffic unless asked
    - Routers with active listeners build tree to receive traffic
  - Best for few members only in a small number of places

Agenda

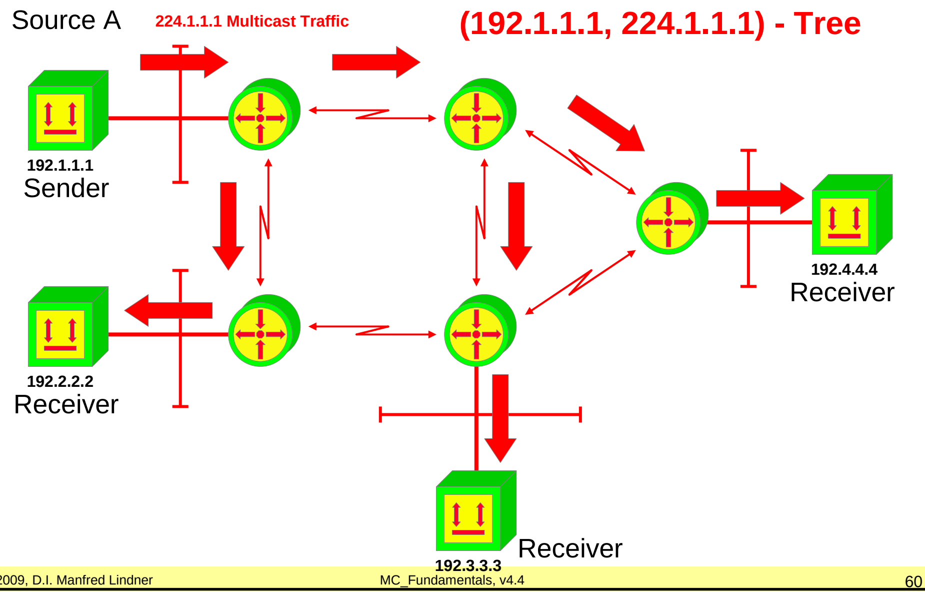
- **Introduction Multicasting**
- **Multicast Routing Principles**
  - Flooding
  - Spanning Tree
  - Reverse Path Forwarding (RPF)
  - Core Based Tree (CBT)
  - Dense Mode / Sparse Mode
  - Multicast Distribution Trees
    - Source Trees
    - Shared Trees
    - Bidirectional / Unidirectional Shared Trees

Multicast Distribution Trees - Source Trees

- **Source Trees**
  - each source is the root of a unidirectional spanning tree
  - uses the shortest path through the network
    - often called: SPT-Shortest Path Tree
  - separate SPT for every individual source
  - Notation: (S, G) *"S comma G"*
    - S ... IP address of the source
    - G ... multicast group address
  - typically used for RPF with flood and prune
  - note: router has to keep one state per source/group pair
- **are used in IP multicast routing protocols**
  - DVMRP, PIM

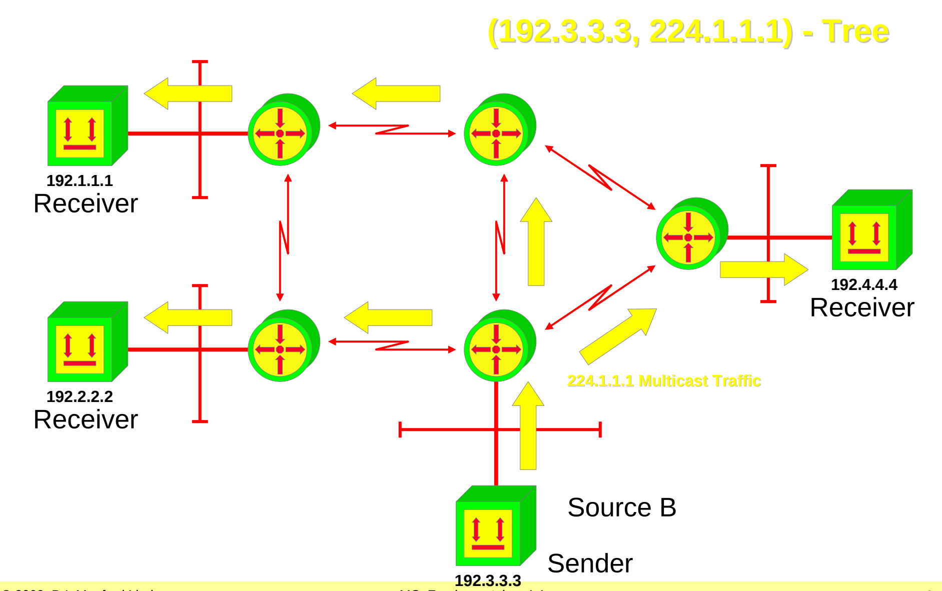
 MC_Fundamentals, v4.4 59

Source Tree for Source A (S,G)

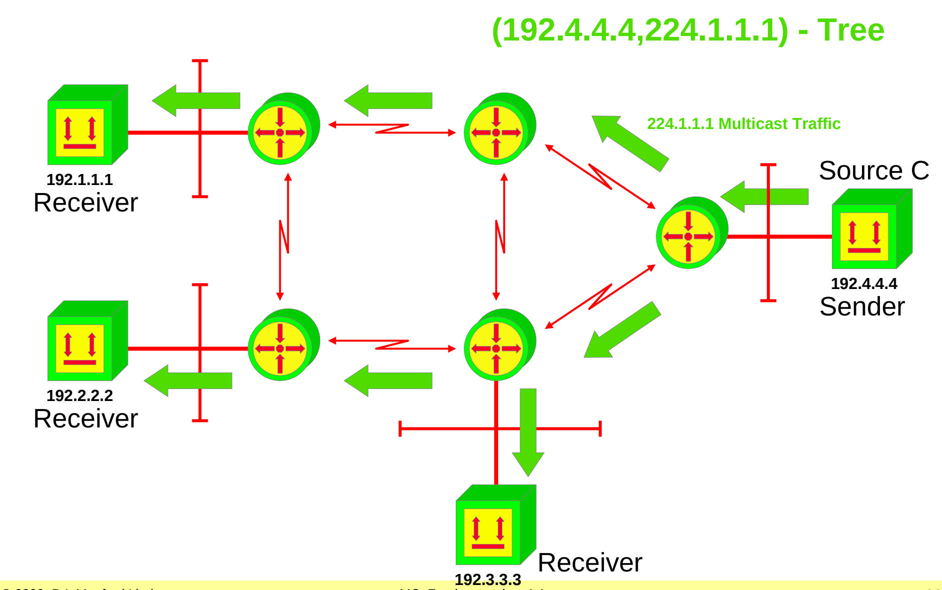


 MC_Fundamentals, v4.4 60

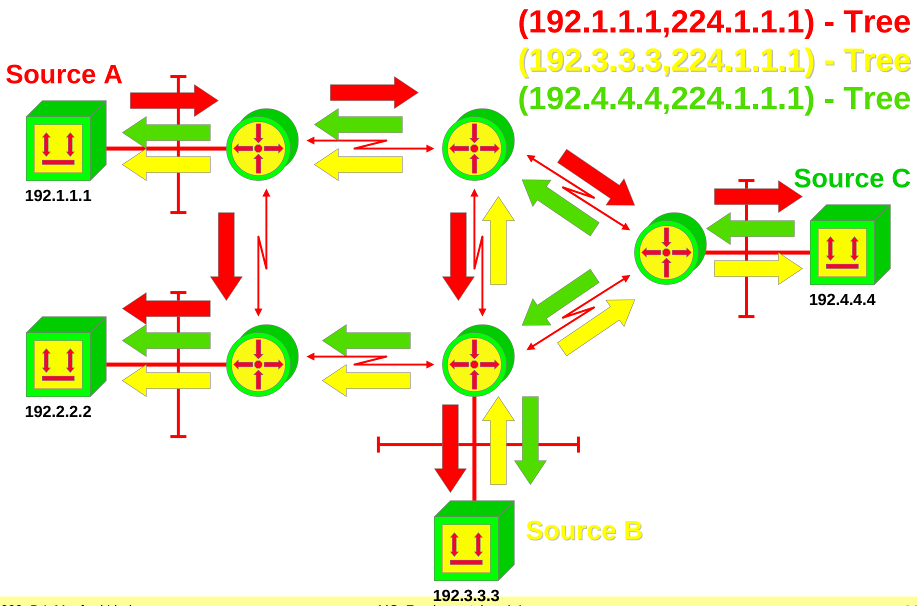
Source Tree for Source B (S,G)



Source Tree for Source C (S,G)



Source Trees for Source A, B, C



Agenda

- **Introduction Multicasting**
- **Multicast Routing Principles**
  - Flooding
  - Spanning Tree
  - Reverse Path Forwarding (RPF)
  - Core Based Tree (CBT)
  - Dense Mode / Sparse Mode
  - Multicast Distribution Trees
    - Source Trees
    - Shared Trees
    - Bidirectional / Unidirectional Shared Trees

Multicast Distribution Trees - Shared Trees

- **Shared Trees**
  - the source is not the root
  - root is somewhere in the network
  - this center is called "RP-Rendezvous Point" or "Core"
  - one tree per multicast group for all sources
    - all sources share the same tree
  - Notation: (*, G) *"Star comma G"*
    - * ... means all sources
    - G ... multicast group address
- **2 types:**
  - bidirectional (used in CBT-Core Based Tree)
  - unidirectional (used in PIM-SM)

Shared Trees - Principle

- **Principle**
  - a member sends a join message towards the root using the shortest path to the core
  - intermediate routers process join message
    - if group is already installed incoming interface is marked
    - if group does not exist, a state for this group is instantiated and a join message is forwarded one step further to the root
  - in this way joining members create a shared tree for a certain group

Shared Trees - Evaluation

- **Advantages**
  - shared tree is the same for all sources (*, G)
    - therefore state information only per group is necessary (RPF needs state per group/source pair)
  - first packet will not be flooded to the whole network and also periodic flooding for refreshing is not needed
    - "hard-states" are acknowledged and repeated after a timeout
    - states are created on demand (if receivers exist)
  - may depend on unicast routing to calculate shortest path
- **Disadvantages**
  - path may be suboptimal for some multicast traffic
  - placement of root is key issue to performance
  - traffic of all sources uses same set of links which means multicast traffic concentration near the root

Types of Shared Trees

- **Bidirectional Shared Trees**
  - multicast traffic flows up and down the tree
- **Unidirectional Shared Tree**
  - multicast traffic is first brought to root and then forwarded down the shared tree

Agenda

- **Introduction Multicasting**
- **Multicast Routing Principles**
  - Flooding
  - Spanning Tree
  - Reverse Path Forwarding (RPF)
  - Core Based Tree (CBT)
  - Dense Mode / Sparse Mode
  - Multicast Distribution Trees
    - Source Trees
    - Shared Trees
    - Bidirectional / Unidirectional Shared Trees

 MC_Fundamentals, v4.4 69

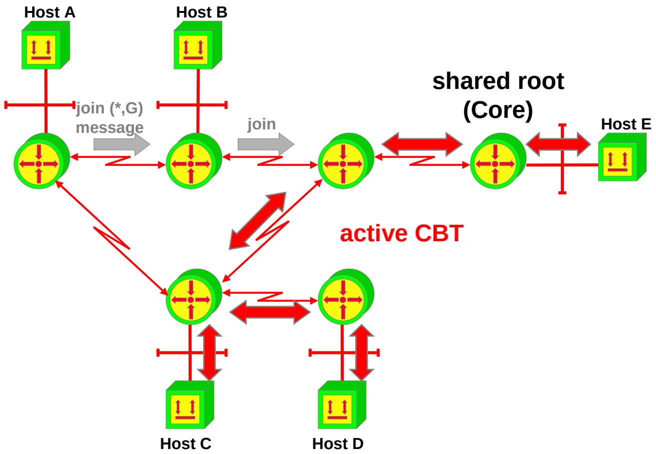
Bidirectional Shared Tree

- **Bidirectional Shared Tree**
  - multicast traffic is forwarded up and down the tree
  - router keeps list of participating interfaces of a group
    - incoming packet is forwarded to all other interfaces active for that group
  - source-only host, which has not joined group, unicasts traffic via IP-in-IP tunnel to core
  - suboptimal routes and increased latency, if core is placed badly
  - Core Based Tree, CBTv1, CBTv2, CBTv3
    - no actual network implementation

 MC_Fundamentals, v4.4 70

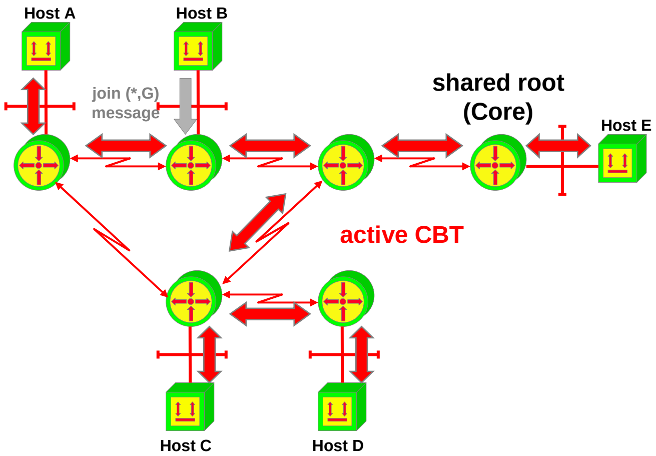
Bidirectional Shared Tree - CBT

Join mechanism: Host A joins CBT



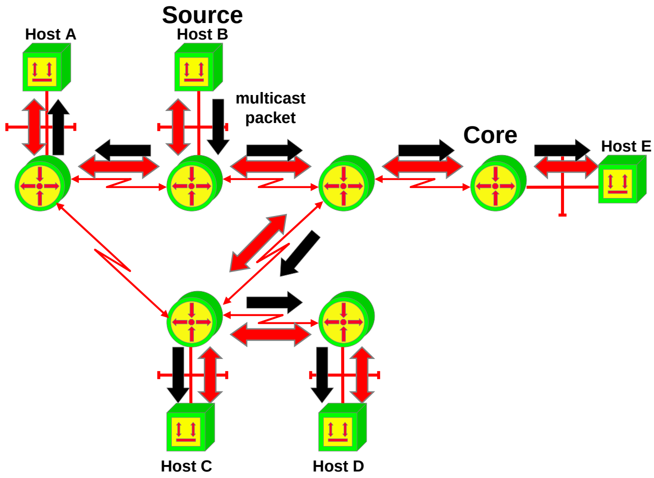
Bidirectional Shared Tree - CBT

Host A joined, Host B joins:



Bidirectional Shared Tree - CBT

final tree: Host B send multicast traffic



 MC_Fundamentals, v4.4 73

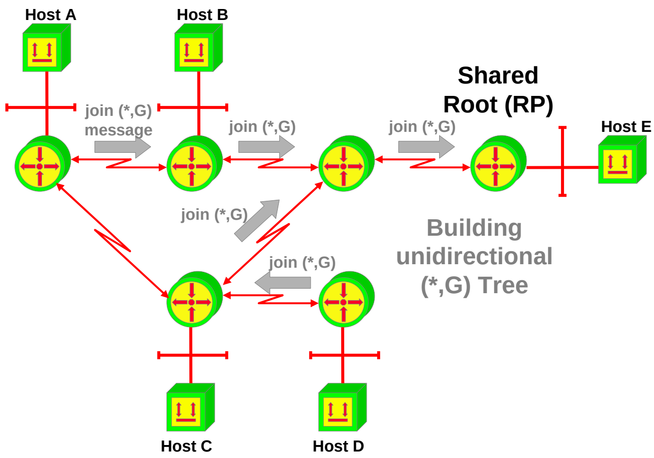
Unidirectional Shared Trees

- **Unidirectional Shared Tree**
  - multicast traffic is first forwarded to root, then flows down the shared tree to all members of the group
    - distribution tree could be compared with establishing a source based tree between the root as source and all members of a group. Remember one tree per group for every source!
  - How to get traffic to the root?
    - can be done by unicast IP-in-IP tunneling
      - packets are encapsulated by first-hop multicast router in special register packets and are delivered hop-by-hop to the core using normal unicast transmission.
    - or with a separate (S,G) - Tree
      - root joins a SPT towards the source to pull source's multicast traffic to the root. Used by Protocol Independent Multicast (PIM)
  - Root forwards multicast traffic down the tree

 MC_Fundamentals, v4.4 74

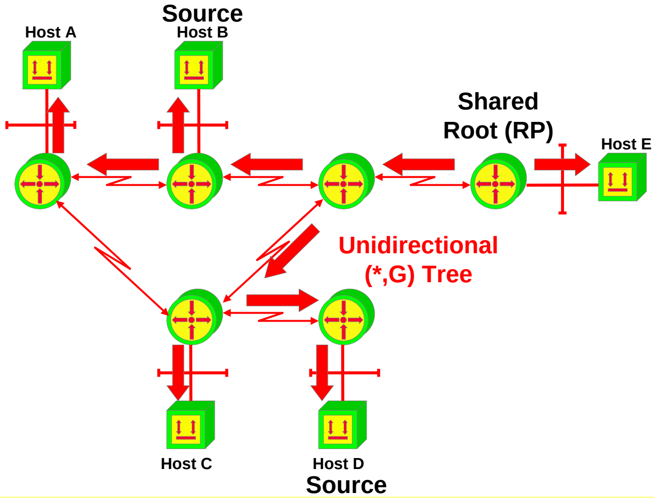
Unidirectional Shared Tree

Join process of receivers towards the shared root:



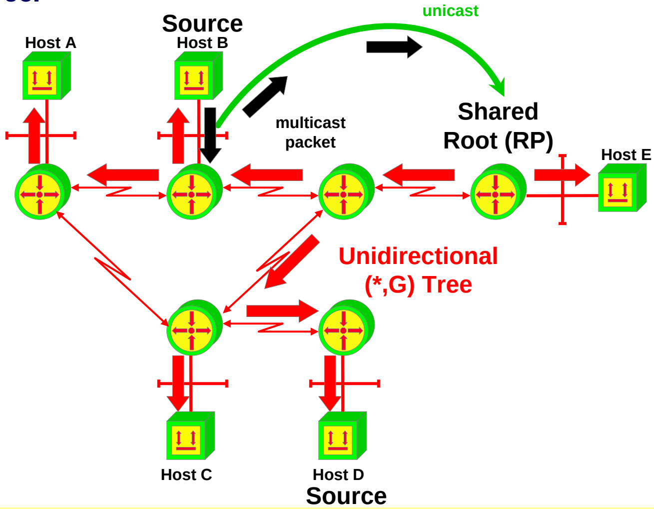
Unidirectional Shared Tree

final tree: every router sets according state



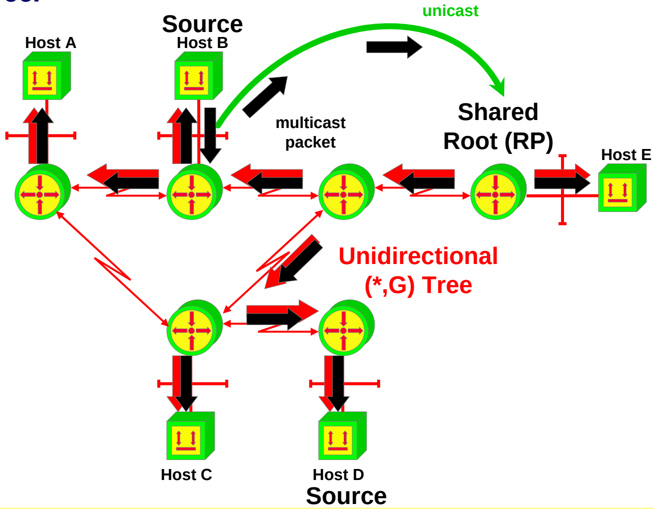
Unidirectional Shared Tree

Host B sends multicast, First hop router encapsulates multicast packet, forwards it to root. Root decapsulates and distributes traffic via (*,G) tree.



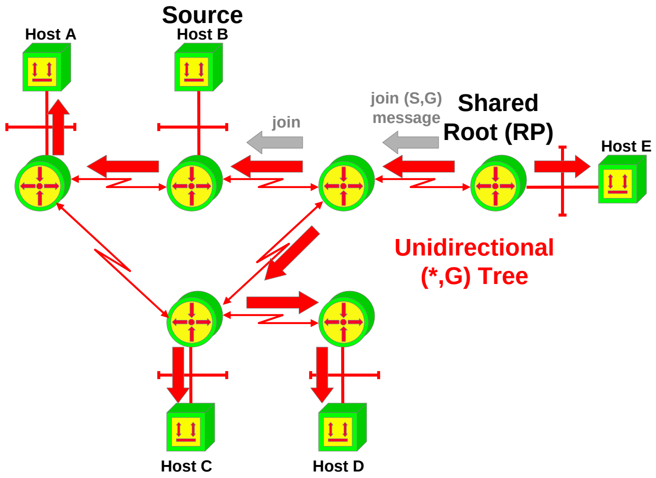
Unidirectional Shared Tree

Host B sends multicast, First hop router encapsulates multicast packet, forwards it to root. Root decapsulates and distributes traffic via (*,G) tree.



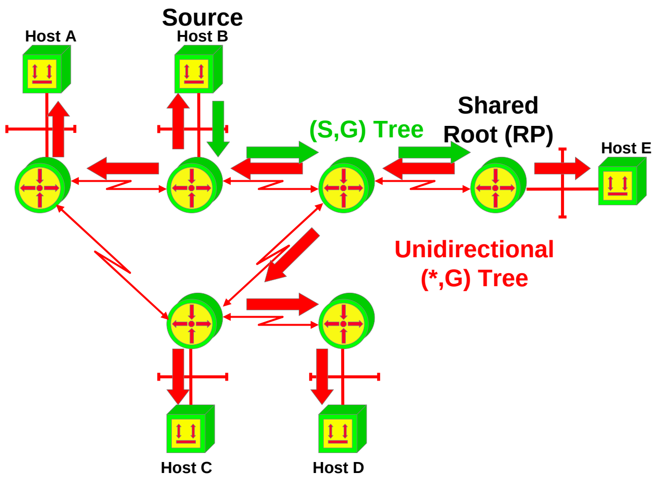
Unidirectional Shared Tree

Root joins SPT (S,G) with Host B as source



Unidirectional Shared Tree

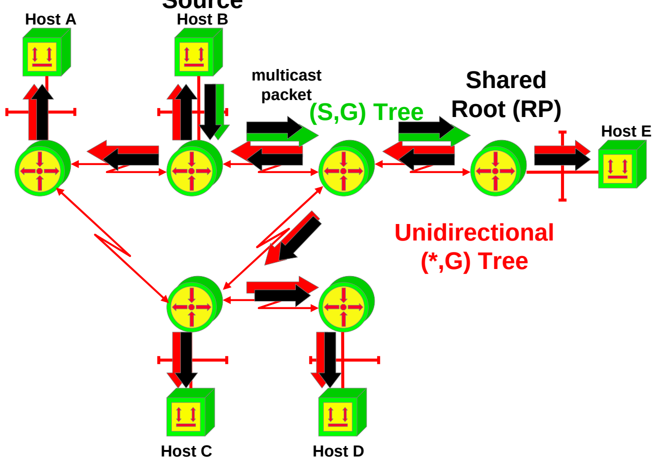
Root joined SPT with Host B as source



Unidirectional Shared Tree

Host B sends multicast packet.

up to root via (S,G)
down the tree via (*,G)



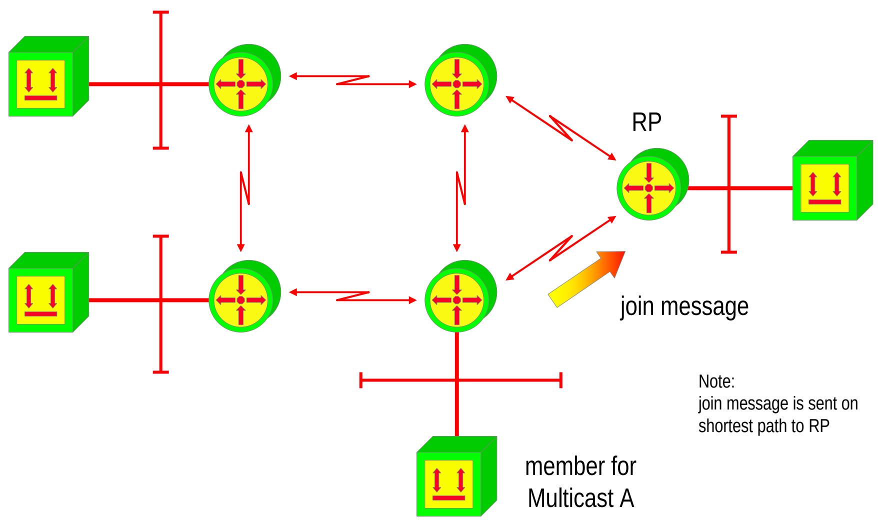
Shared Trees and Rerouting

- **in case of a topology change**
  - unicast routing or special multicast routing will converge
  - join packets will be sent on the new shortest path towards the core
  - a new (*, G) tree will result
  - packets will follow the new tree

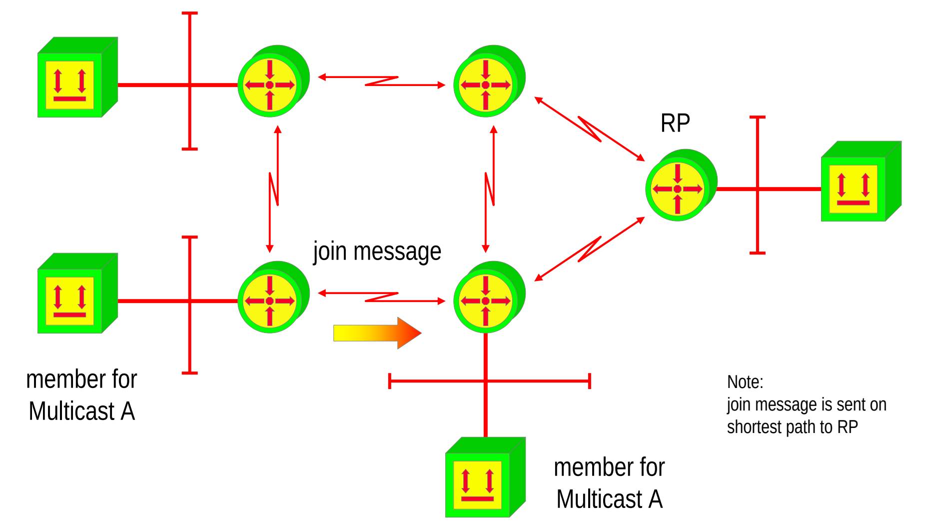
Additional Aspects of (S,G) and (*,G)

- in case of PIM-SM
  - first-hop multicast router of a multicast receiver can force building of source based tree (S, G)
    - new join message on shortest path to source and prune to RP
    - see following example
  - also any intermediate router can force building of a source-based distribution tree
    - e.g. if too much traffic from a given source coming along a suboptimal path is recognized
    - remember: multicast router are able to calculate shortest path to the RP and hence to any other network address
  - for such a behavior soft-states are used instead of hard-states
    - refreshed by periodic join messages
    - disappear if the information is not refreshed

PIM-SM: Joining 1

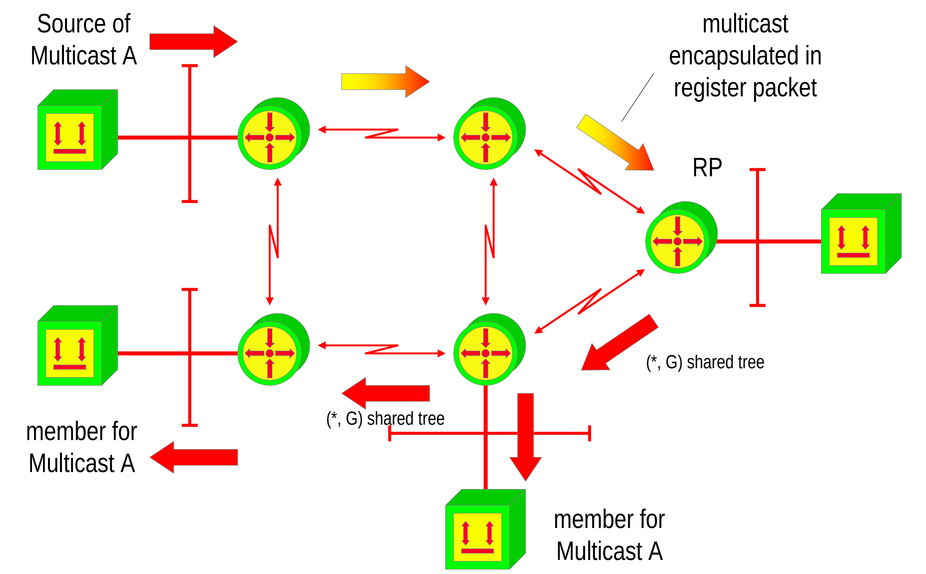


PIM-SM: Joining 2



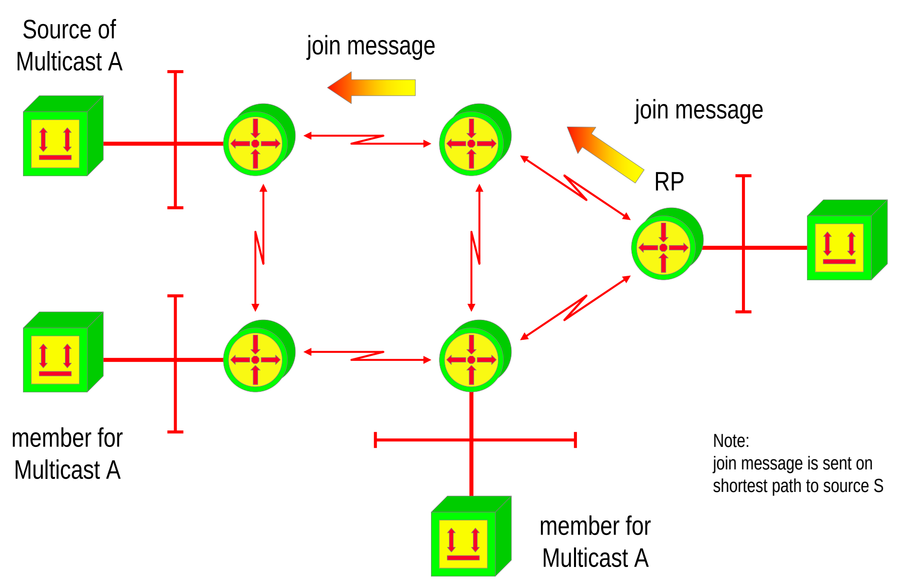
© 2009, D.I. Manfred Lindner MC_Fundamentals, v4.4 85

PIM-SM: Multicast Traffic



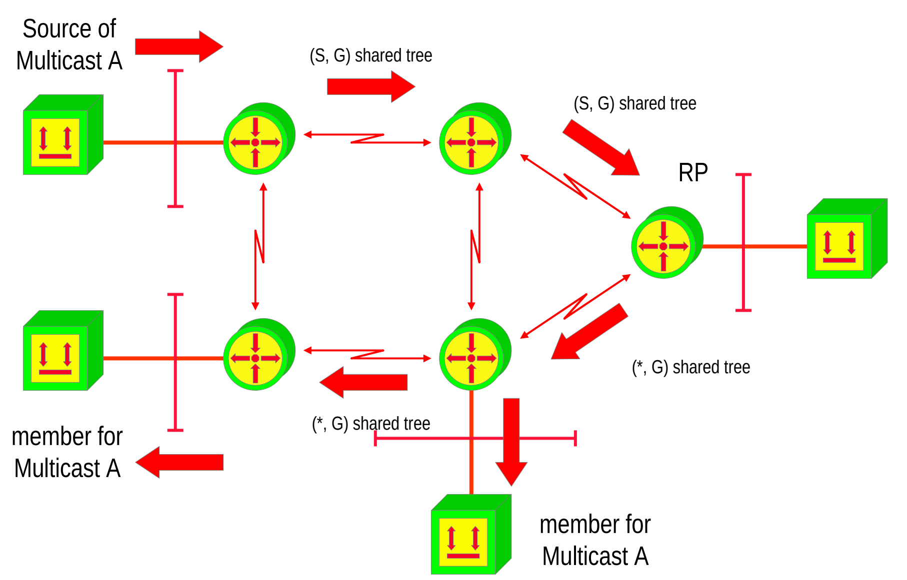
© 2009, D.I. Manfred Lindner MC_Fundamentals, v4.4 86

Example 1: Source Based Tree



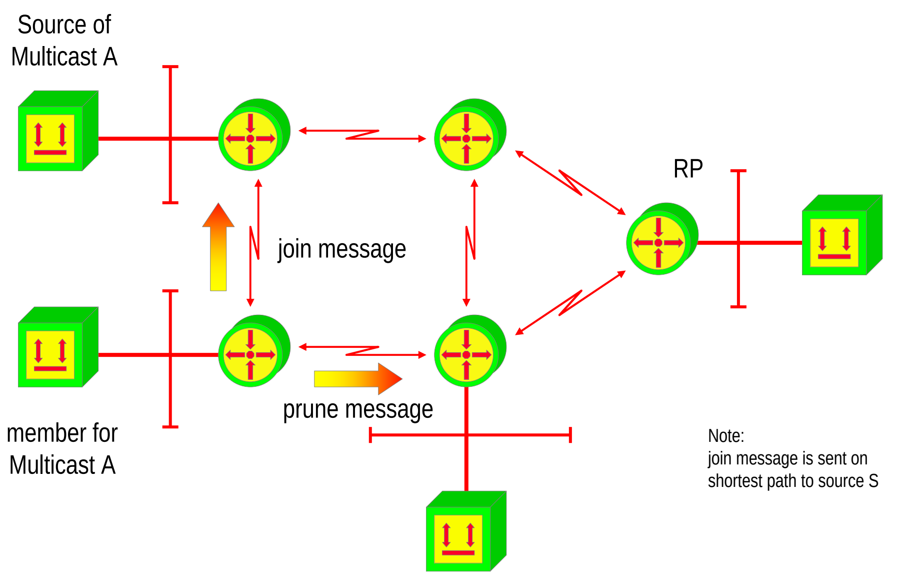
 MC_Fundamentals, v4.4 87

Example 1: Multicast Traffic



 MC_Fundamentals, v4.4 88

Example 2: Source Based Tree



Example 2: Multicast Traffic