



Cisco Expo  
2008

## Bi-Directional PIM (Bi-Dir)



**Vladimir Settey**

**Systems Engineer, Cisco Slovakia**

**Enable Your Network  
Empower Your Business**

# Multicast Application Categories

- One-to-Many Applications
  - Video, TV, Radio, Concerts, Stock Ticker, etc.
- Few-to-Few Applications
  - Small (<10 member) Video/Audio Conferences
- Few-to-Many Applications
  - TIBCO RV Servers (Publishing)
- Many-to-Many Applications
  - Stock Trading Floors, Gaming
- Many-to-Few Applications
  - TIBCO RV Clients (Subscriptions)

# Multicast Application Categories

## PIM-SM (S, G) State

- One-to-Many Applications
  - Single (S,G) entry
- Few-to-Few Applications
  - Few (<10 typical) (S,G) entries
- Few-to-Many Applications
  - Few (<10 typical) (S,G) entries
- Many-to-Many Applications
  - Unlimited (S,G) entries*
- Many-to-Few Applications
  - Unlimited (S,G) entries*

# Many-to-Any State Problem

- Creates huge amounts of (S,G) state

  - State maintenance workloads skyrocket

    - High OIL fanouts make the problem worse

  - Router performance begins to suffer

- Using Shared-Trees only.

  - Provides some (S,G) state reduction

    - Results in (S,G) state only along SPT to RP

    - Frequently still too much (S,G) state

    - Need a solution that only uses (\*,G) state

# Bidirectional (Bidir) PIM

- Idea:

  - Use the same tree for traffic from sources towards RP and from RP to receivers

- Benefits:

  - Less state in routers

    - Only (\*, G) state is used

    - Source traffic follows the Shared Tree

      - Flows up the Shared Tree to reach the RP.

      - Flows down the Shared Tree to reach all other receivers.

# Bidirectional (Bidir) PIM

- Bidirectional Shared-Trees

  - Violates current (\*,G) RPF rules

    - Traffic often accepted on *outgoing* interfaces.

    - Care must be taken to avoid multicast loops

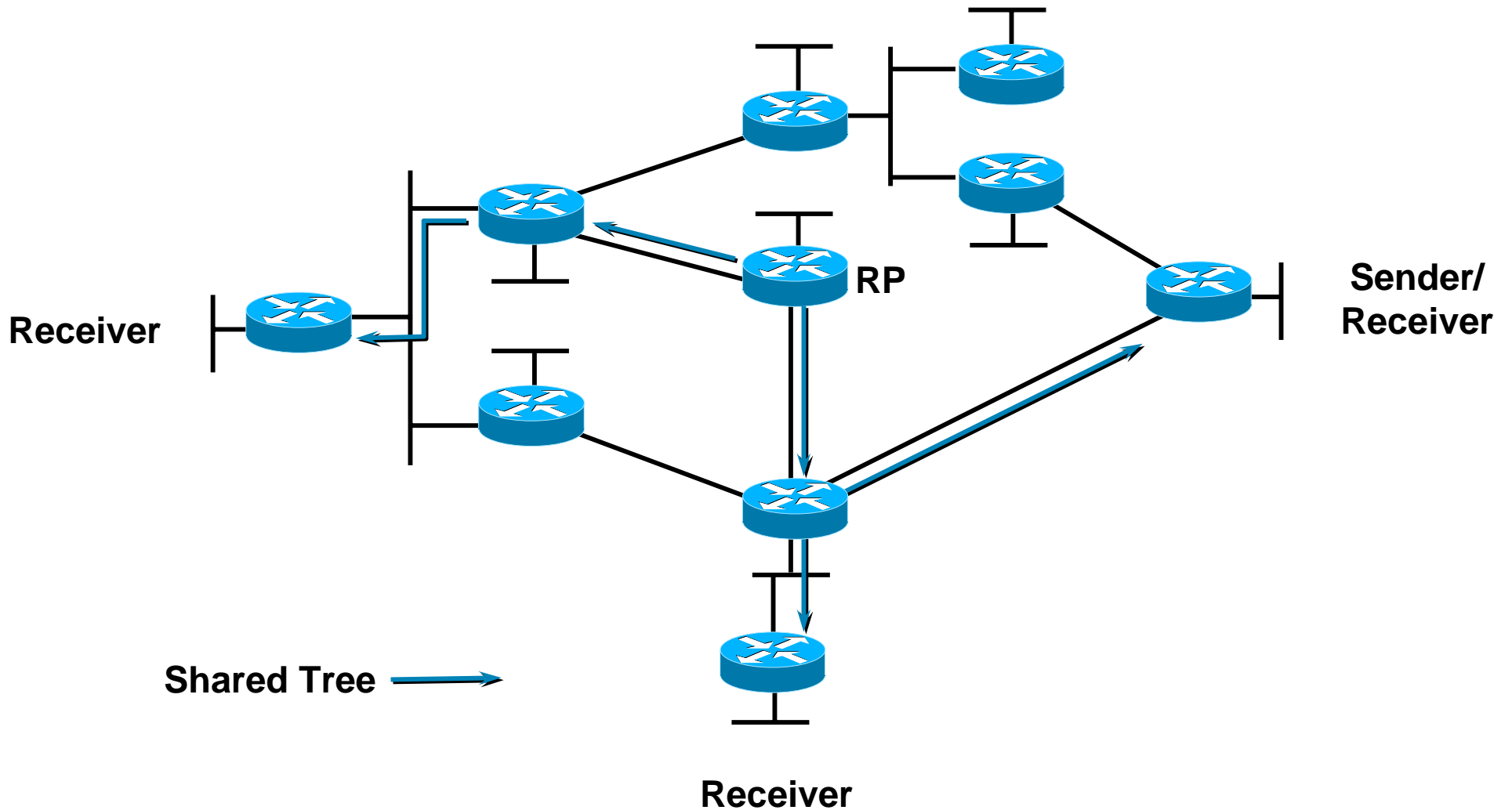
  - Requires a Designated Forwarder (DF)

    - Responsible for forwarding traffic up Shared Tree

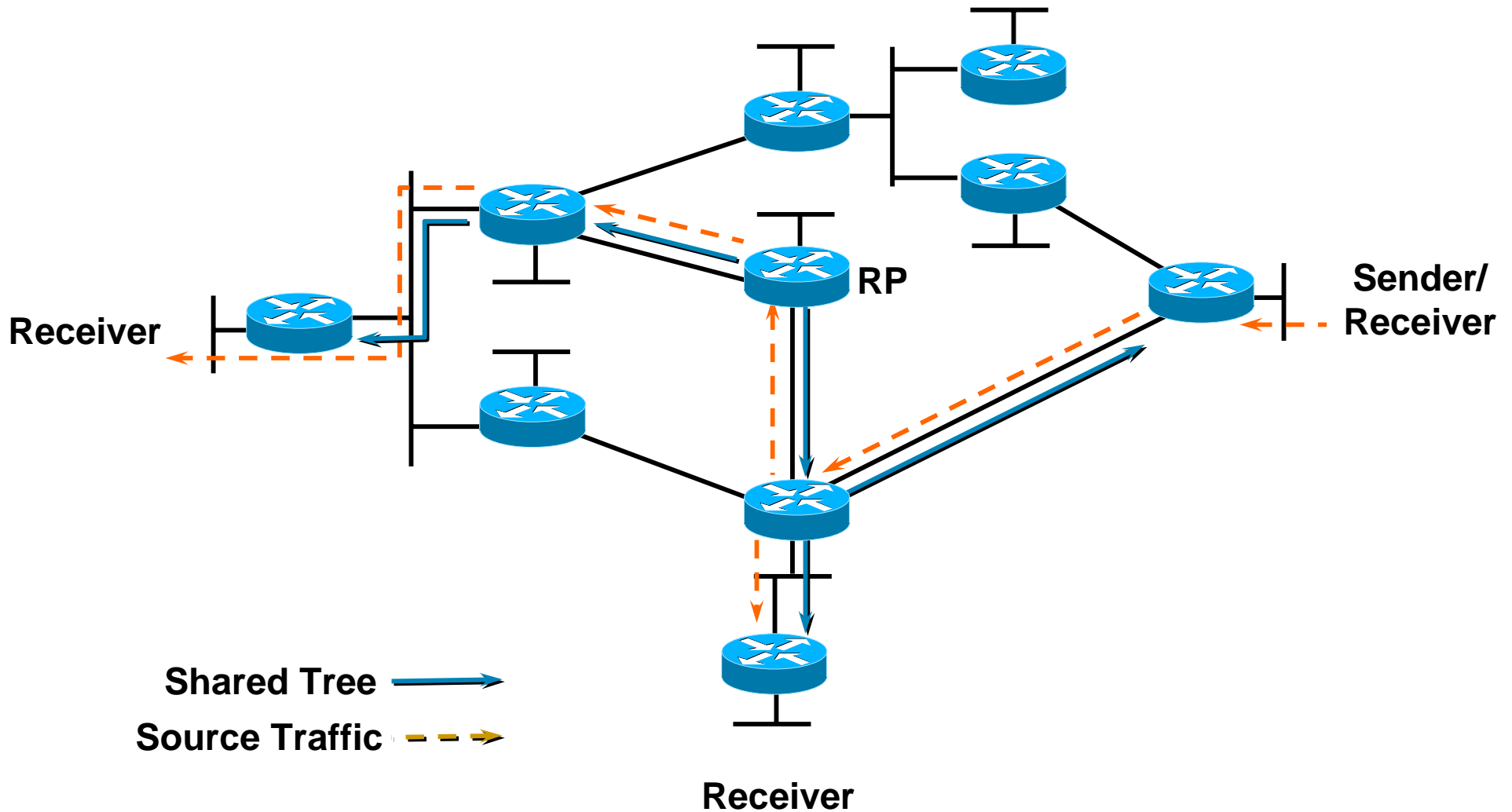
      - DF's will accept data on the interfaces in their OIL.

      - Then send it out all other interfaces. (Including the IIF.)

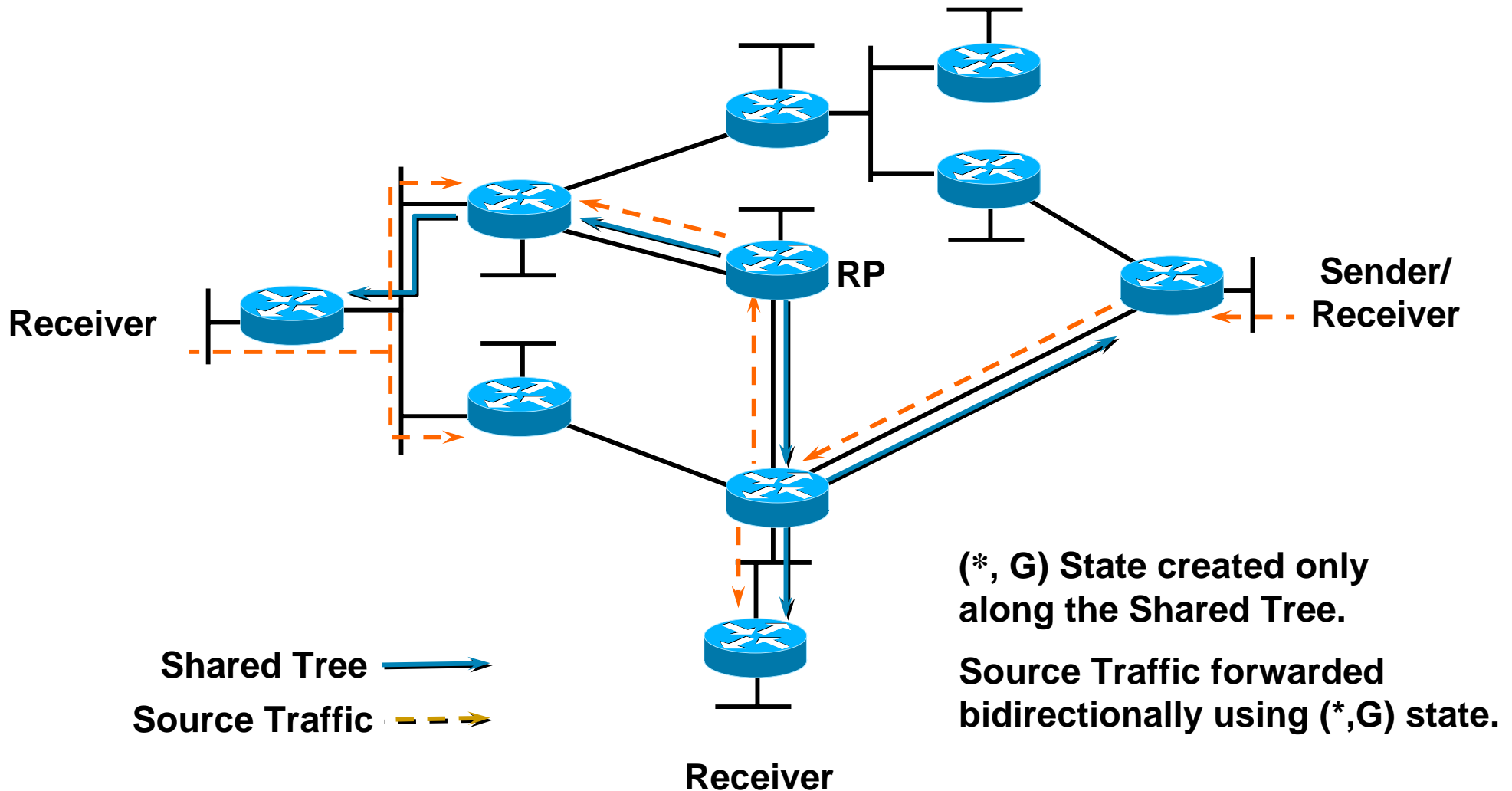
# Bidirectional PIM – Overview



# Bidirectional PIM – Overview



# Bidirectional PIM – Overview



# PIM Modifications for Bidir Operation

- Designated Forwarders (DF)

One DF per link.

Router with best path to the RP is elected DF.

Note: Designated Routers (DR) are not used for bidir groups.

In addition to normal (\*,G) forwarding rules:

Accepts traffic on outgoing interfaces.

Forwards traffic out all other interfaces

# Designated Forwarder Election

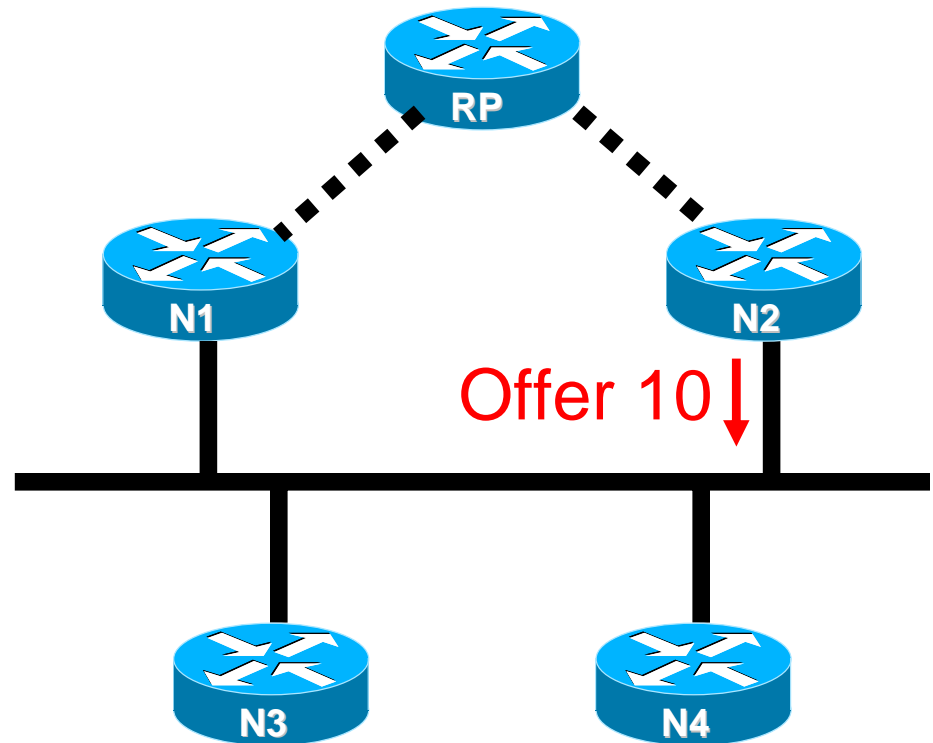
- Automatically performed on every link.
  - When Bidir Group-range/RP is learned or configured.
  - Router with the best path to the RP elected DF.
  - Uses assert-like metric comparison to pick best path.
- Purpose:
  - Ensures all routers on link agree on who is DF.
  - Prevents route loops from forming.

# DF Election Messages

- Offer – used to advertise local metrics to reach the RP
- Winner – used by a DF announcing or re-asserting its status
- Backoff – used by a DF upon receipt of a better Offer
- Pass – used by an acting DF to pass the DF responsibility to a better candidate

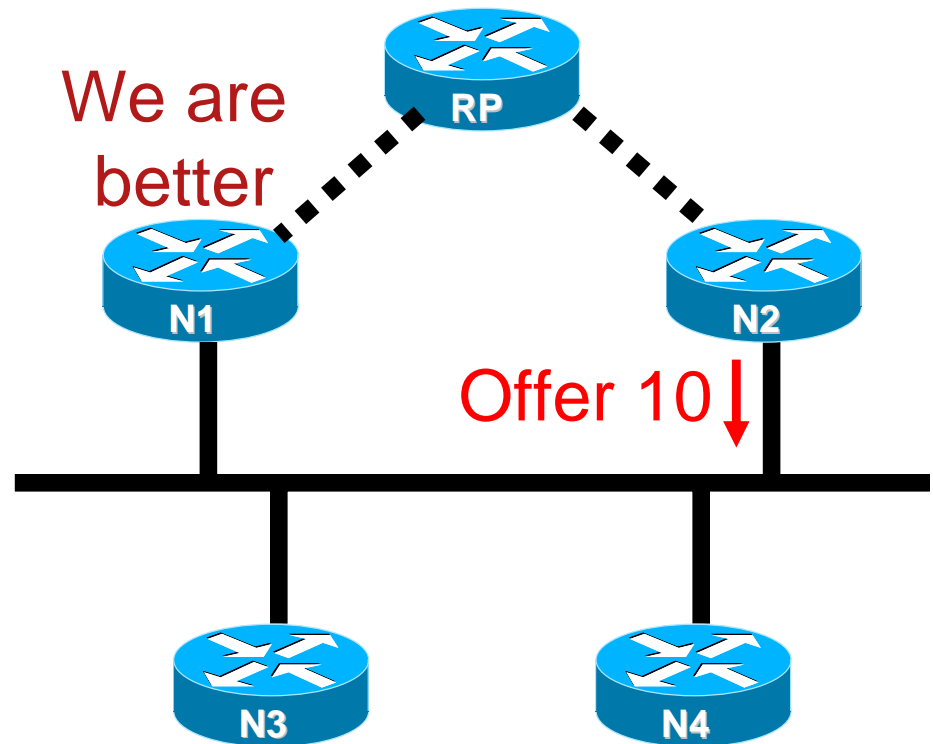
# Initial Election

- On RP discovery send offer with RP metric



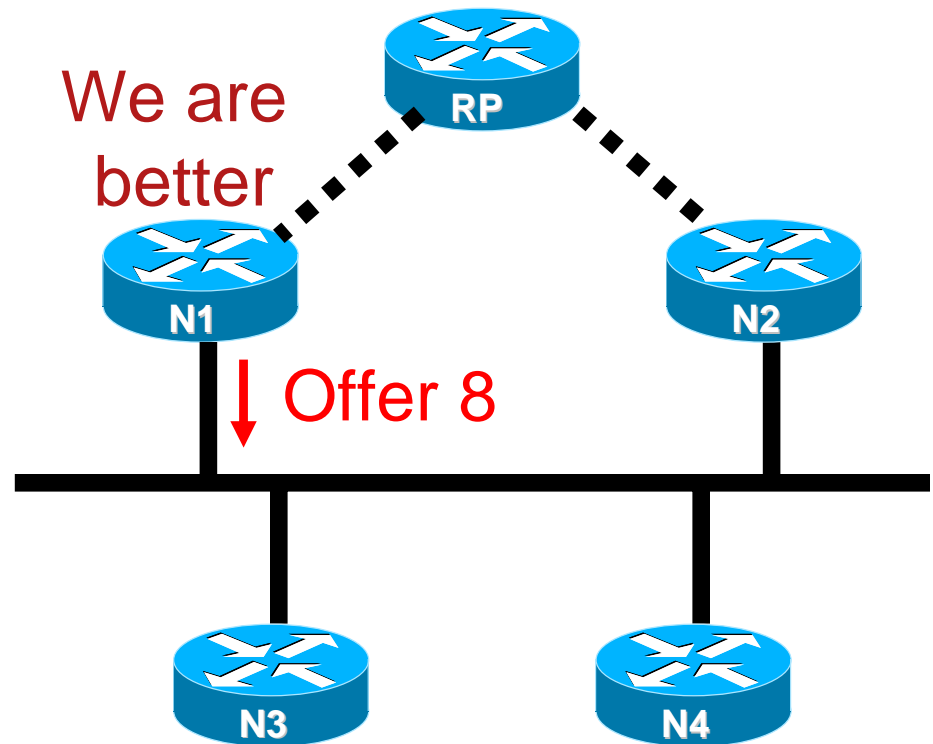
# Initial Election (Cont.)

- On RP discovery send offer with RP metric
- Neighbors compare with own metric and back off unless better



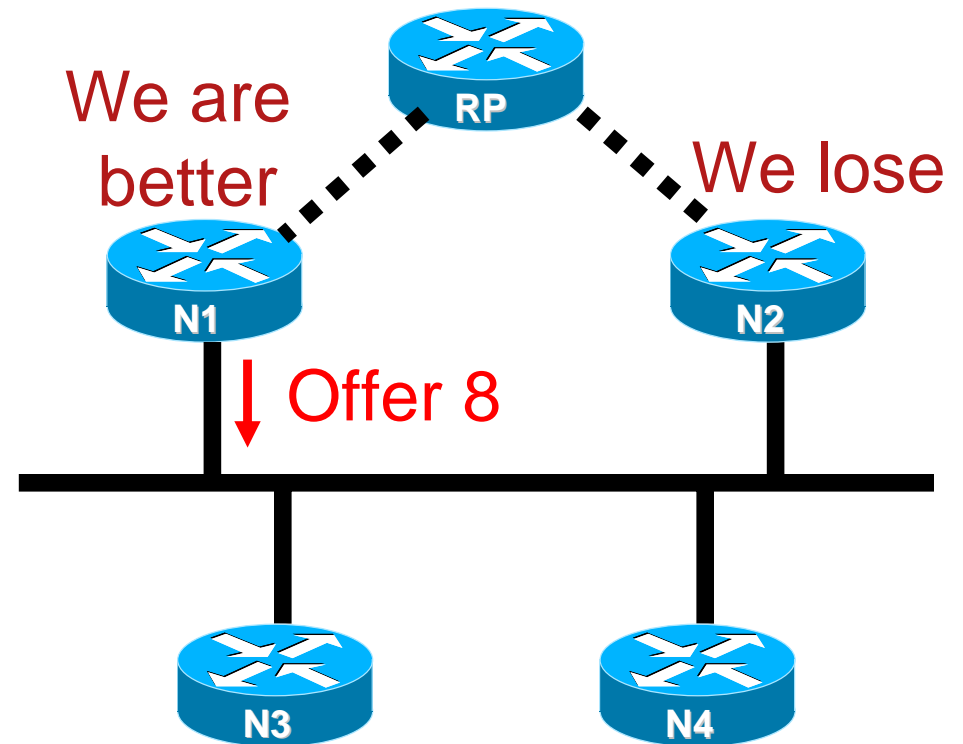
# Initial Election (Cont.)

- On RP discovery send offer with RP metric
- Neighbors compare with own metric and back off unless better



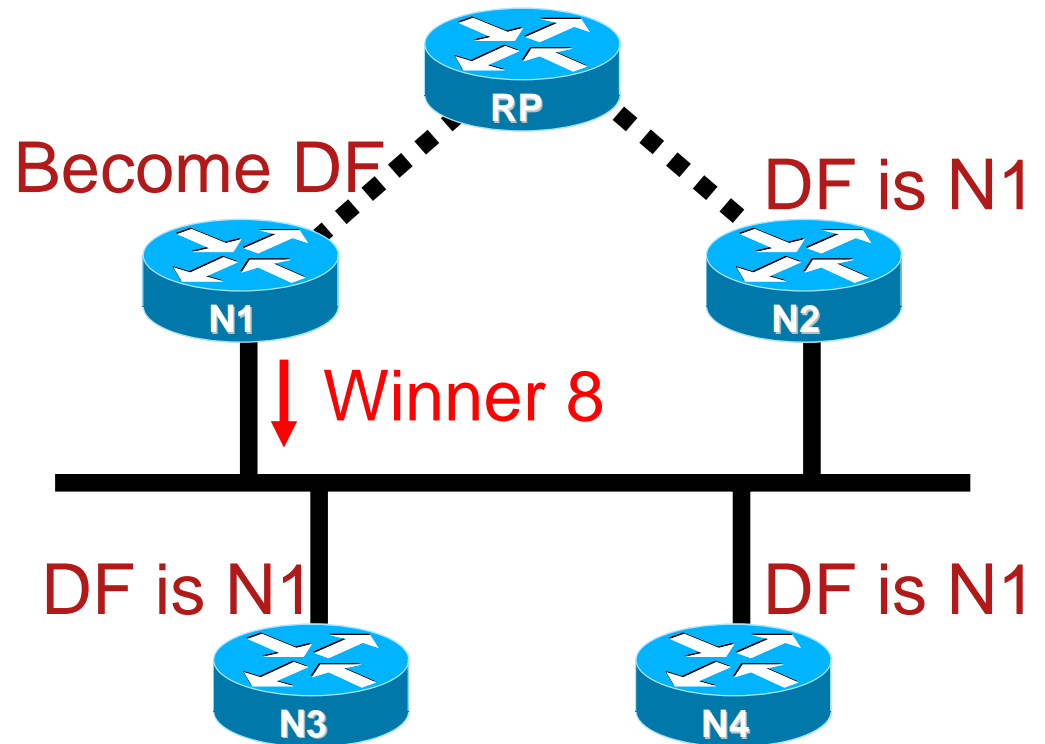
# Initial Election (Cont.)

- On RP discovery send offer with RP metric
- Neighbors compare with own metric and back off unless better



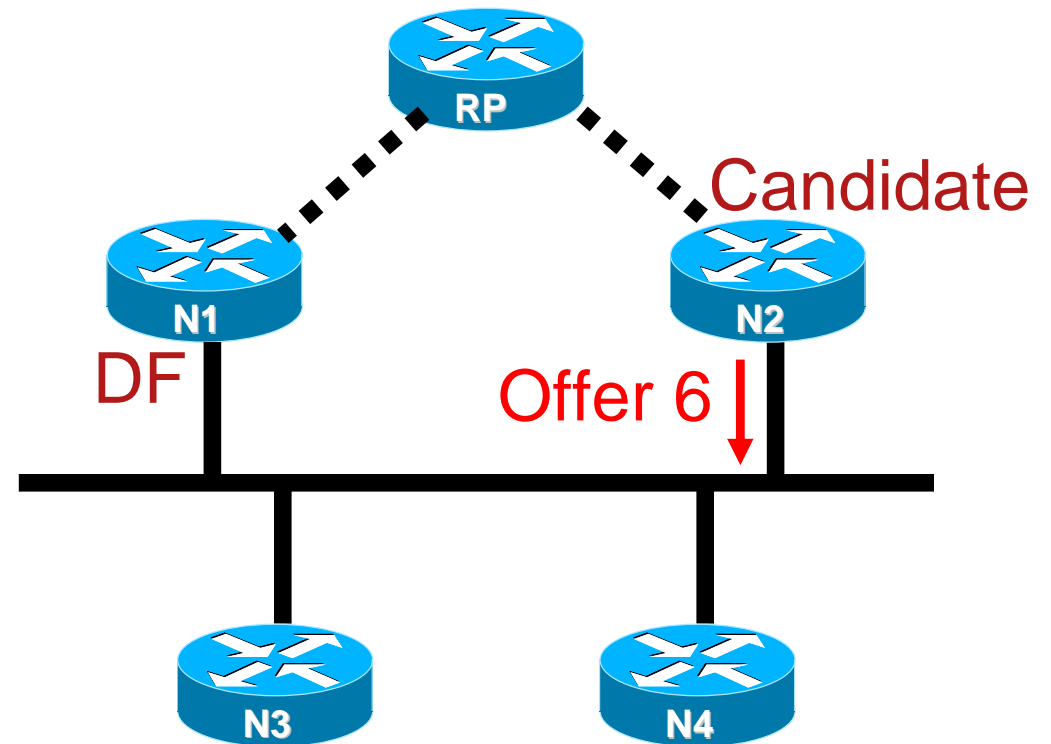
# Initial Election (Cont.)

- On RP discovery send offer with RP metric
- Neighbors compare with own metric and back off unless better
- After repeating 3 uncontested Offers, send a Winner and assume DF role



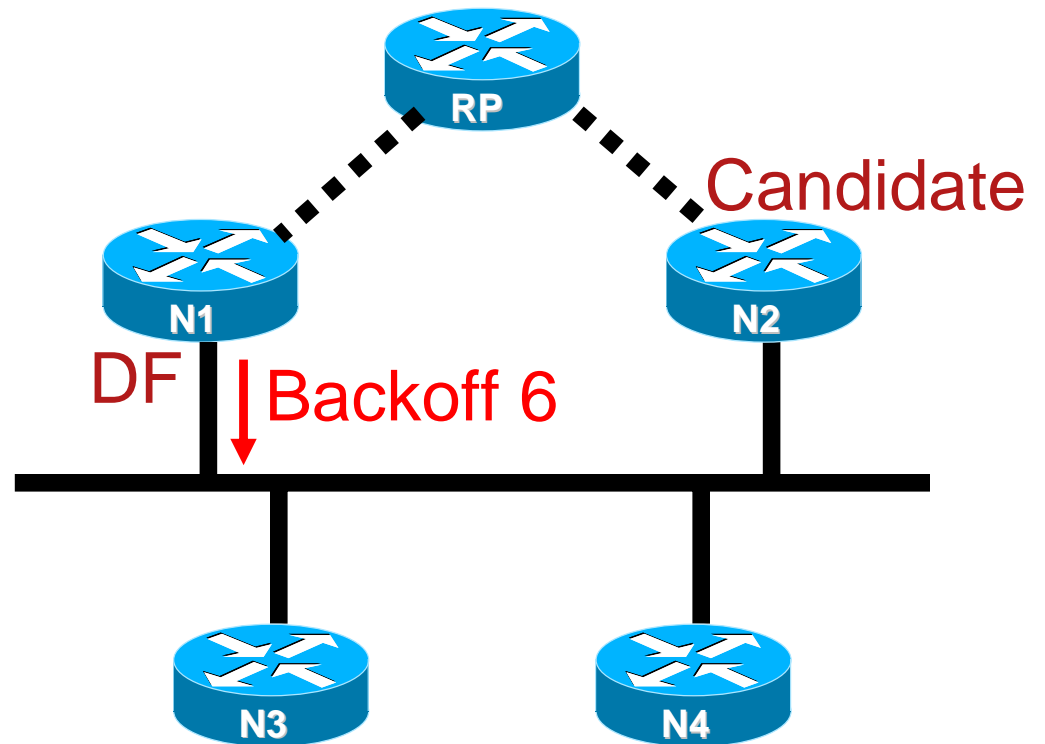
# Non-DF Metric Becomes Better

- New candidate sends improved Offer.



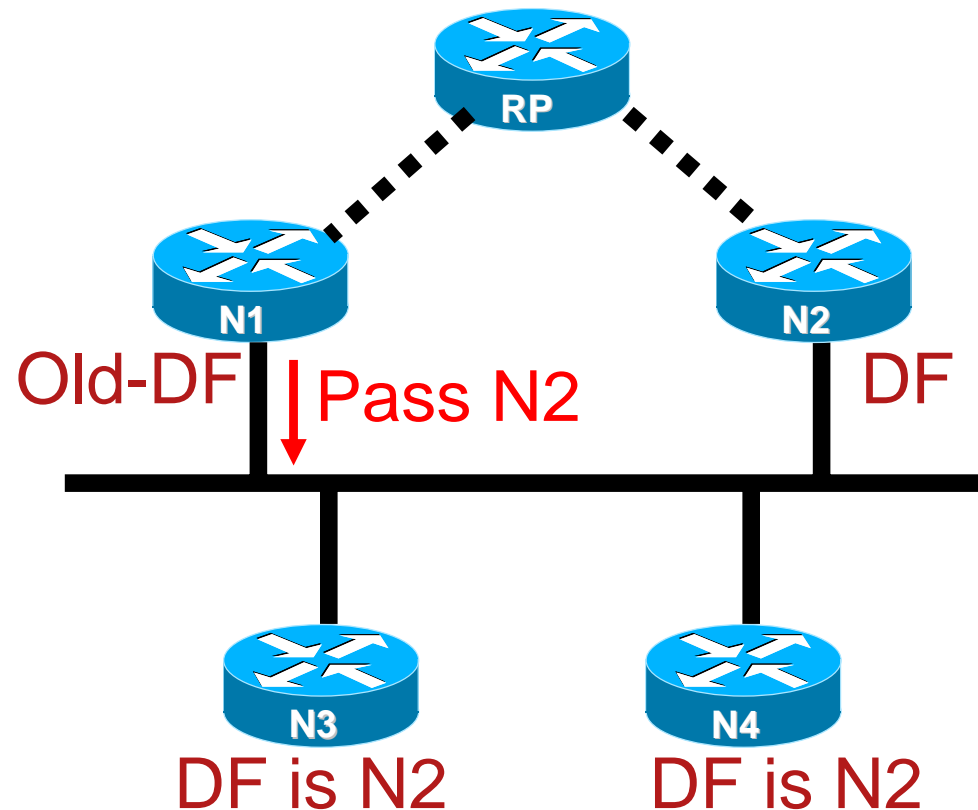
# Non-DF Metric Becomes Better (Cont.)

- New candidate sends improved Offer.
- DF responds with Backoff instructing candidate to wait.



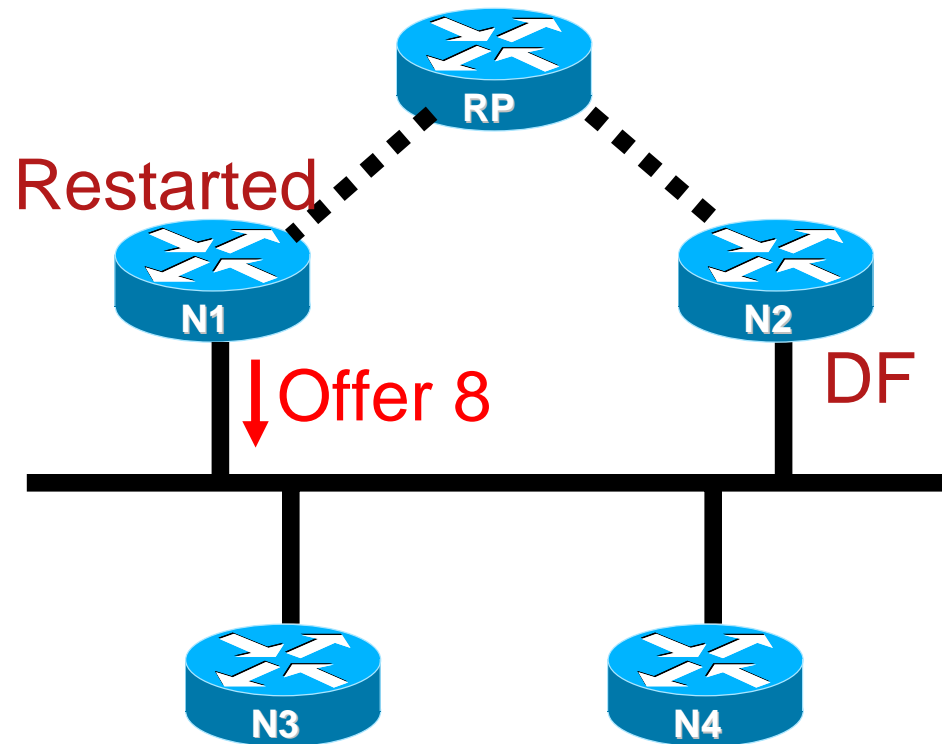
# Non-DF Metric Becomes Better (Cont.)

- New candidate sends improved Offer.
- DF responds with Backoff instructing candidate to wait.
- Before backoff period expires, old DF stops forwarding and sends Pass. On receipt candidate becomes DF.



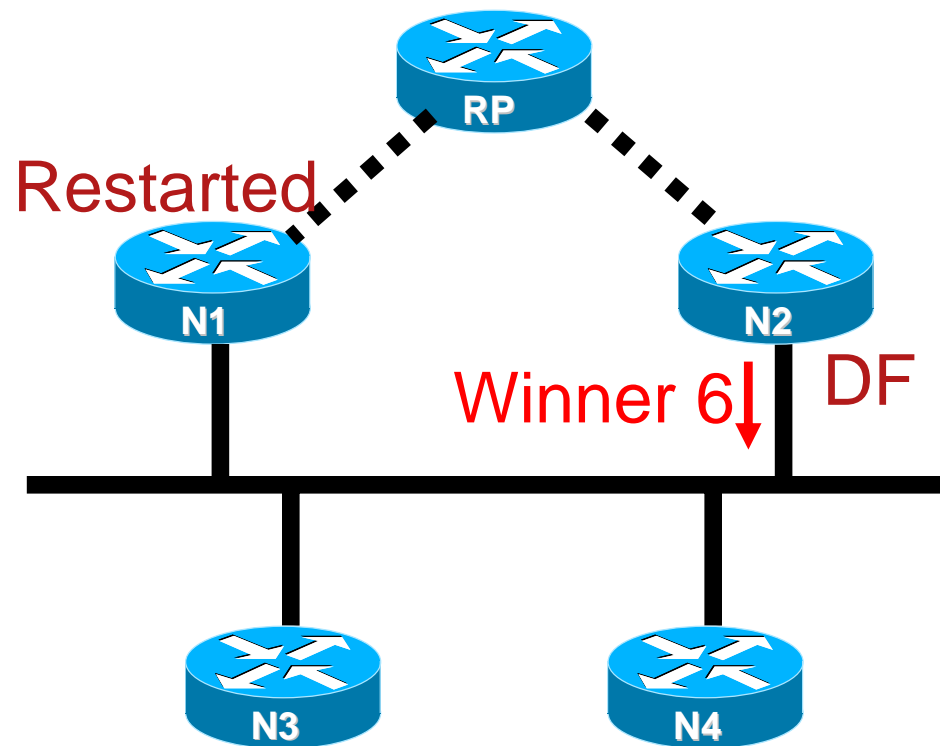
# New PIM Neighbor Startup

- Router N1 restarts and has no knowledge of DF
- On RP discovery it sends an Offer



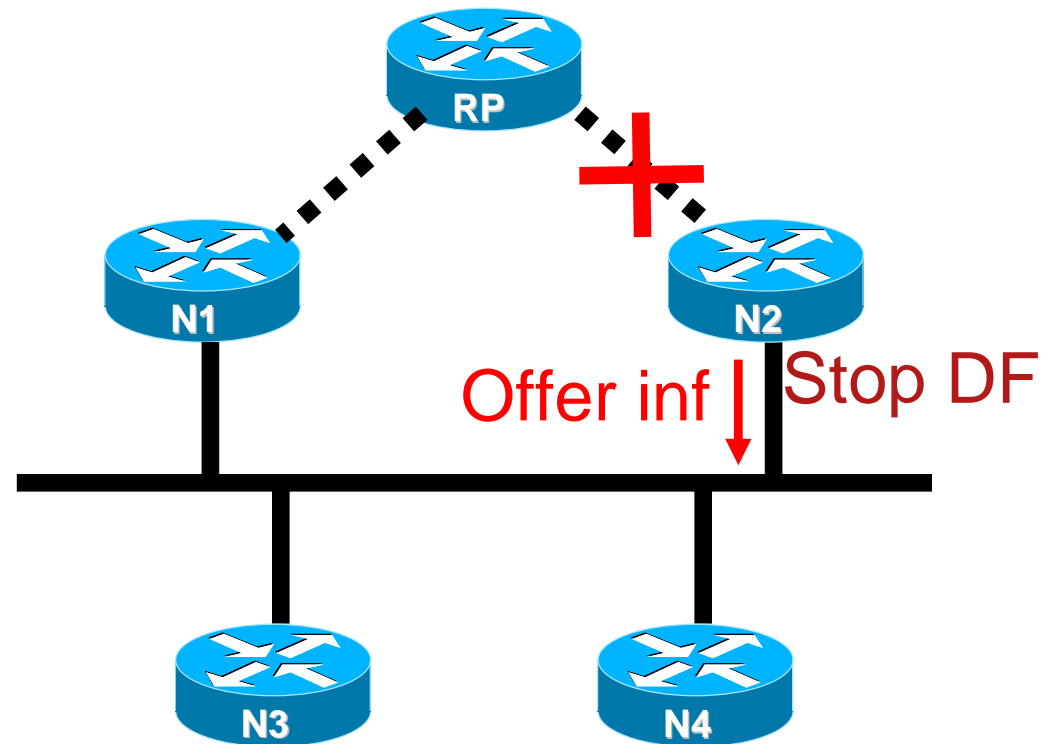
# New PIM Neighbor Startup (Cont.)

- Router N1 restarts and has no knowledge of DF
- On RP discovery it sends an Offer
- Acting DF responds with Winner or Backoff depending on metric comparison



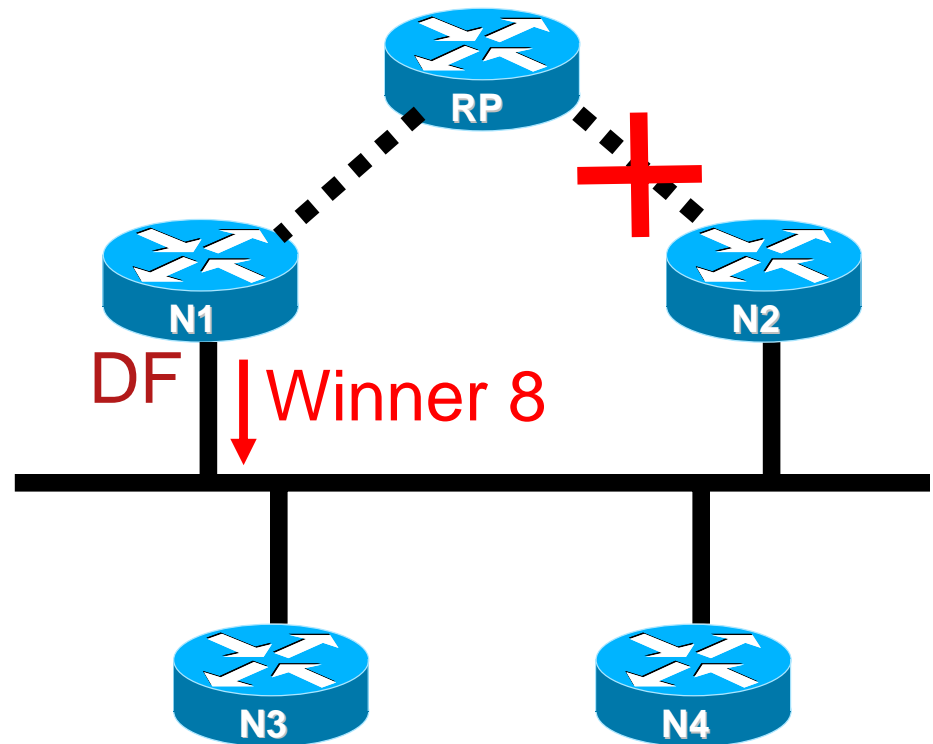
# DF Loses Path to the RP

- Immediately stop acting as the DF and send Offer with infinite metric to trigger re-election.



# DF Loses Path to the RP (Cont.)

- Immediately stop acting as the DF and send Offer with infinite metric to trigger re-election.
- Other candidates respond with real Offers and eventually best candidate takes over with a Winner message.



# DF Dies

DF is no longer available:

- Downstream routers notice a change in the RPF information (different next-hop) and can trigger a re-election.
- If no downstream routers are available, the PIM neighbor timeout will trigger a re-election.

# Other Metric Changes

- RP metric at a non-DF router changes to a value that is still worse than that of the acting DF - no action is taken.
- Metric at the DF improves - a Winner message may be sent to update information in neighboring routers.
- Metric at the DF becomes worse - three Winner messages are sent to give the opportunity to a better candidate to respond with an Offer.

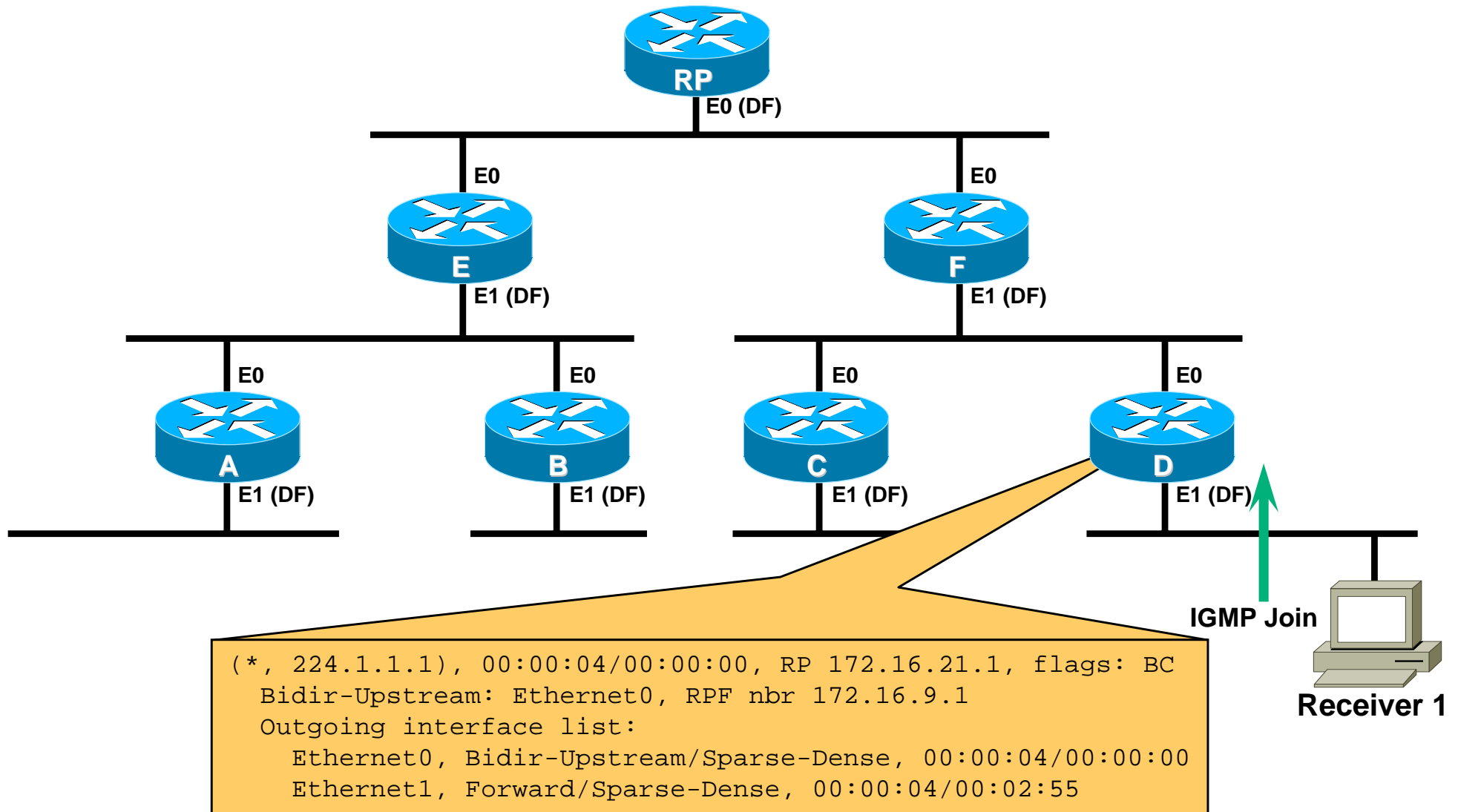
# Additional Robustness

- DF re-announces sending Winner message when new PIM neighbors are discovered
- Periodic Winner messages can be sent for RPs with active groups

# DF Advantages

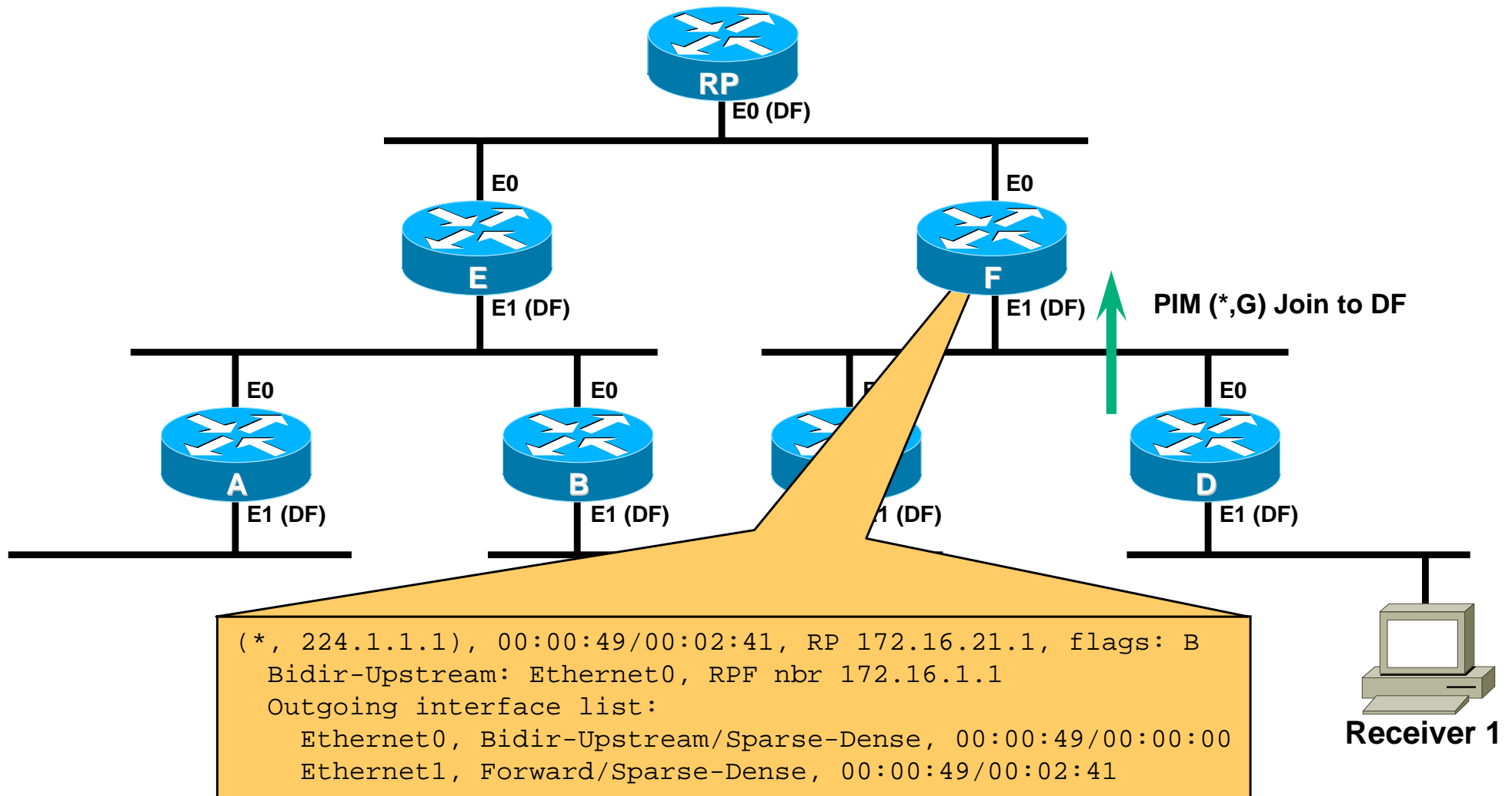
- DF election enforces a single forwarder for traffic in both directions between a link and the RP.
- DF is responsible for originating Joins for local receivers.
- DF eliminates loops that can result from misconfiguration (static multicast routes).

# Forwarding / Tree Building



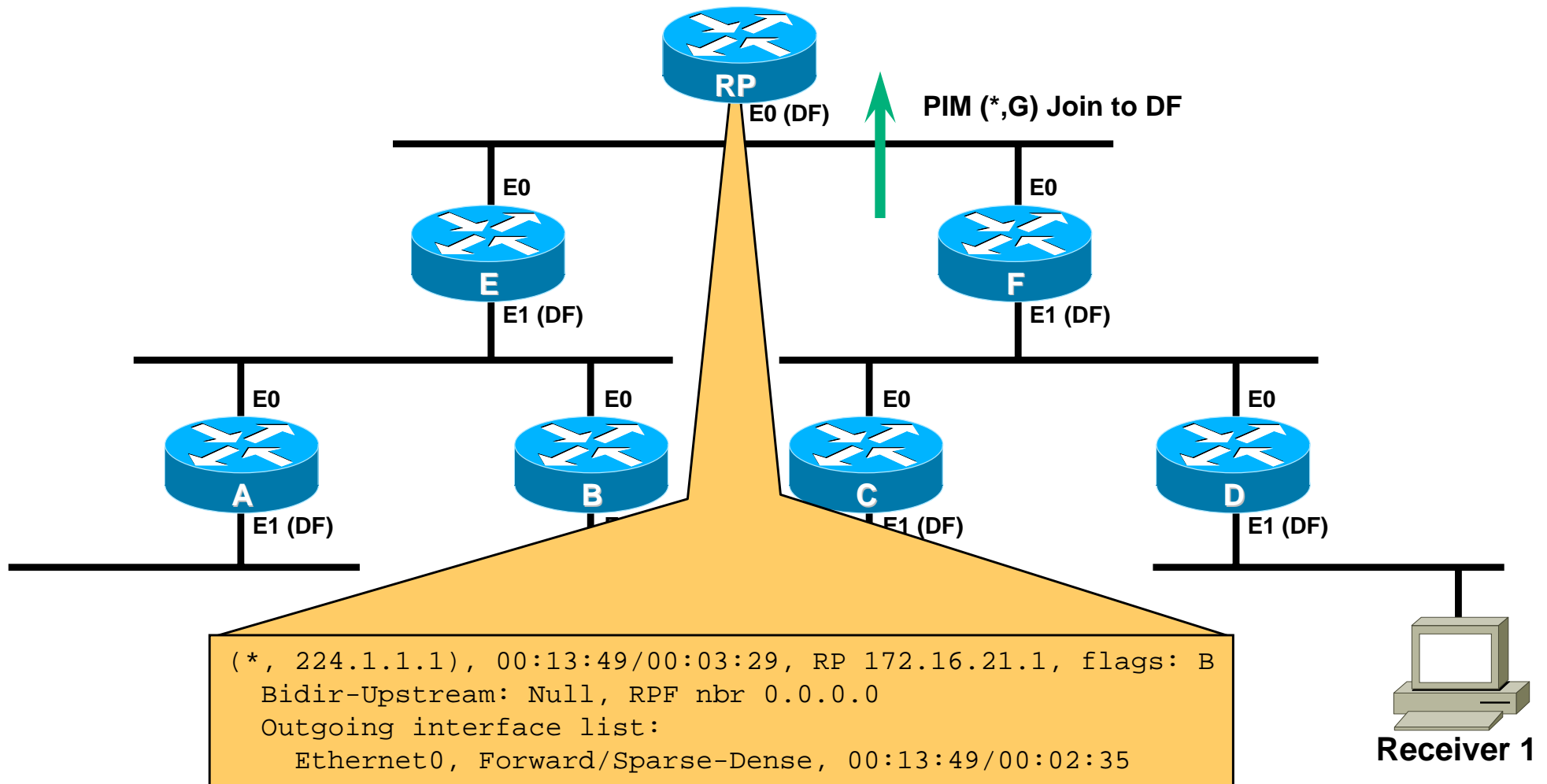
**Receiver 1 joins group causing router “D” to create (\*, G) state.**

# Forwarding / Tree Building



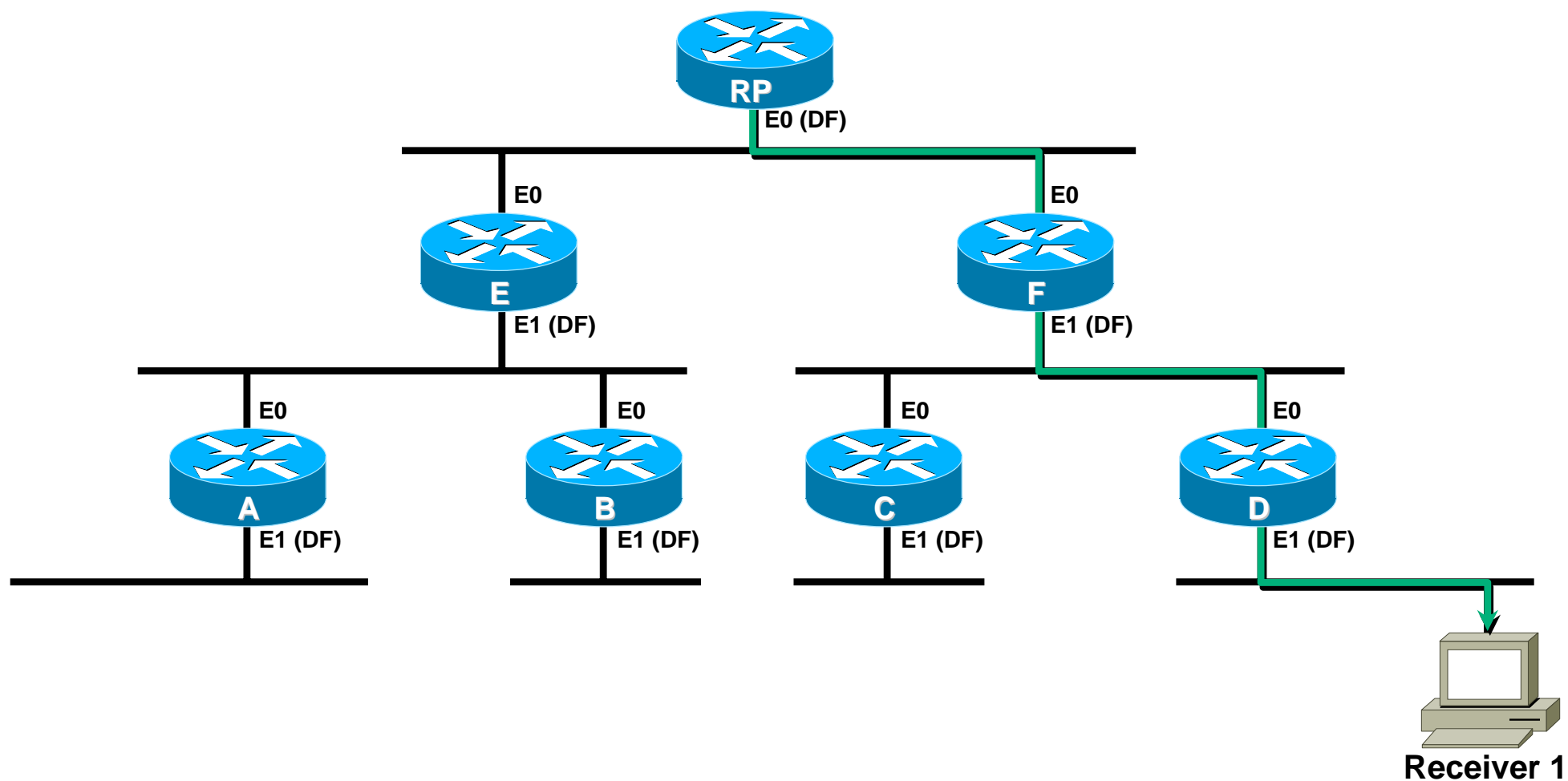
**Router “D” sends (\*, G) Join to router “F” (DF) causing it to create (\*, G) state.**

# Forwarding / Tree Building



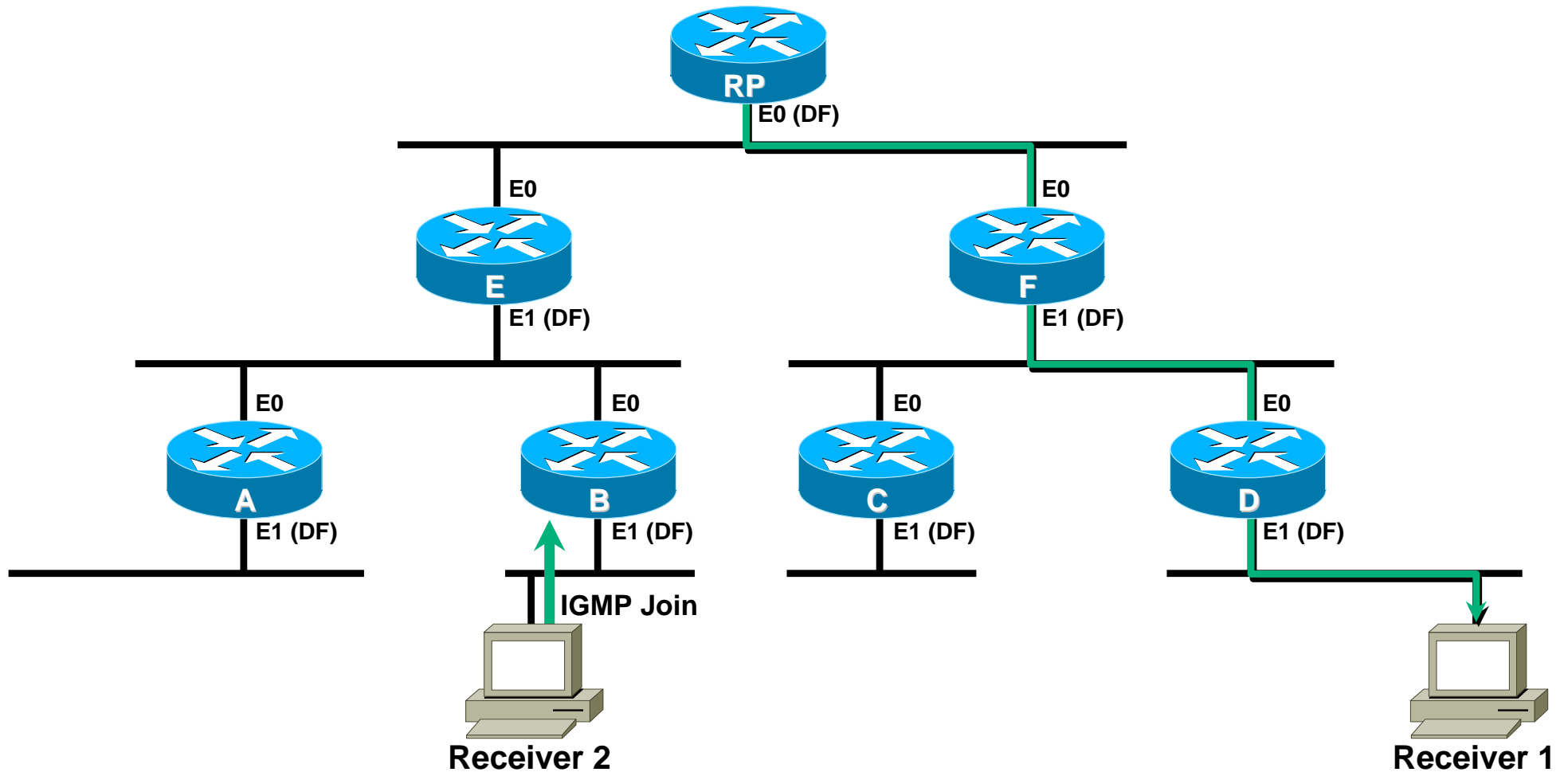
Router “F” sends (\*, G) Join to “RP” causing it to create (\*, G) state.

# Forwarding / Tree Building



**Branch of Shared Tree is now built down to Receiver 1.**

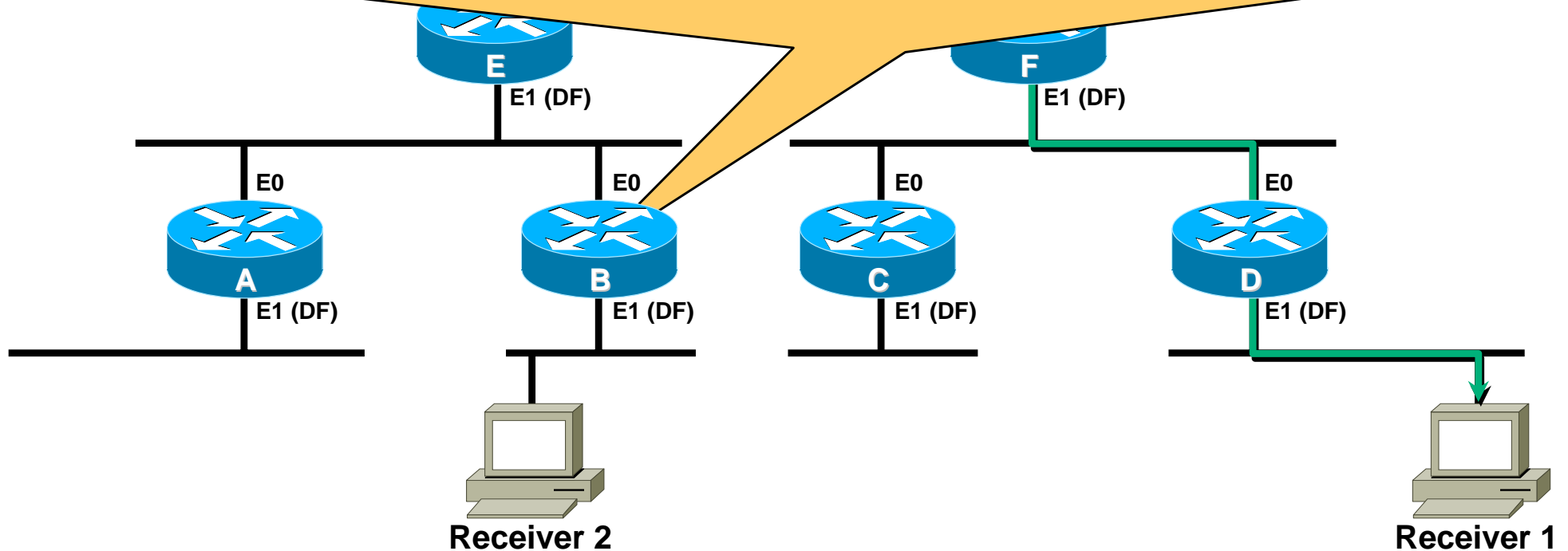
# Forwarding / Tree Building



**Receiver 2 also joins group.**

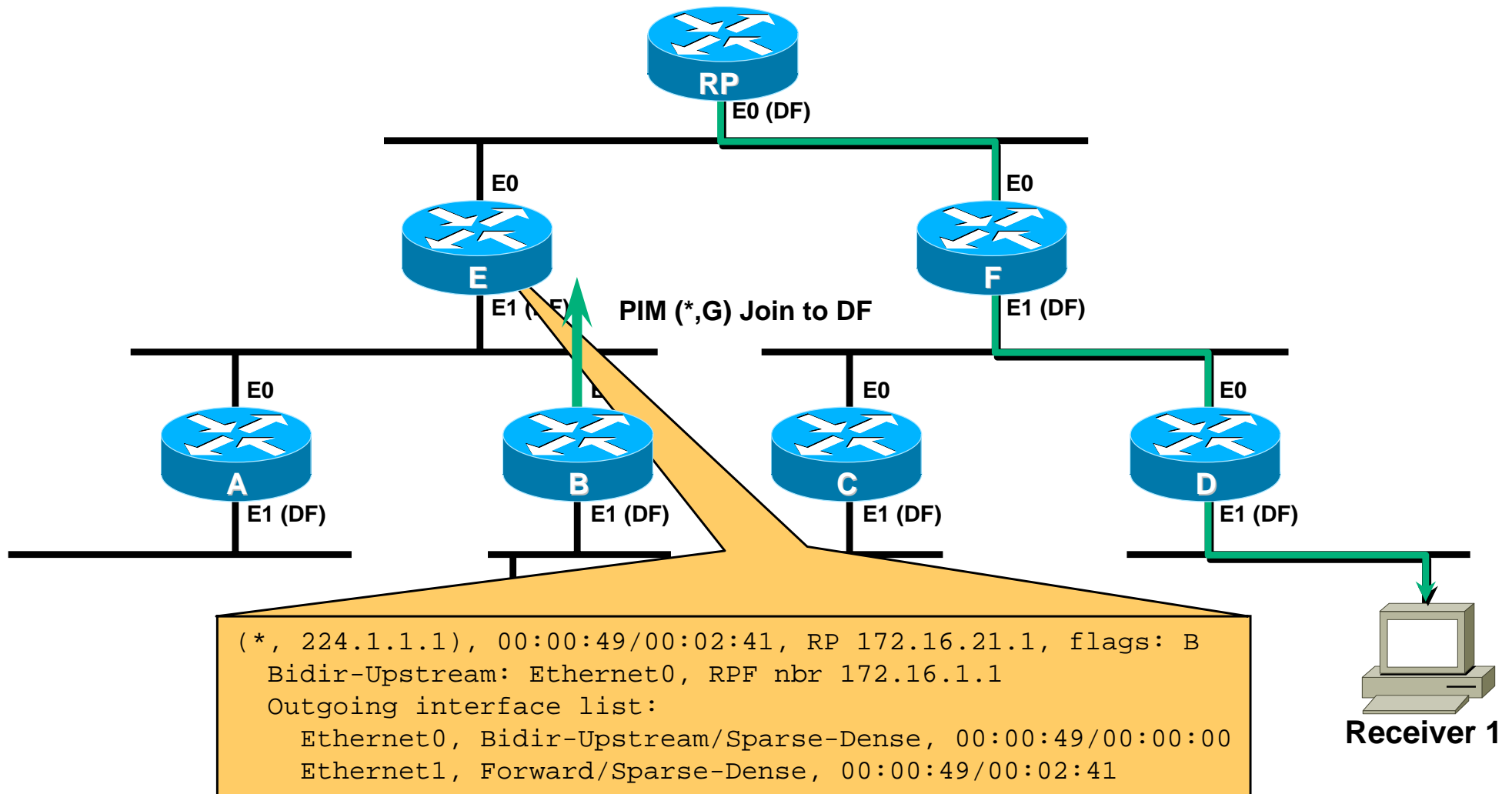
# Forwarding / Tree Building

```
(*, 224.1.1.1), 00:00:04/00:00:00, RP 172.16.21.1, flags: BC  
Bidir-Upstream: Ethernet0, RPF nbr 172.16.9.1  
Outgoing interface list:  
  Ethernet0, Bidir-Upstream/Sparse-Dense, 00:00:04/00:00:00  
  Ethernet1, Forward/Sparse-Dense, 00:00:04/00:02:55
```



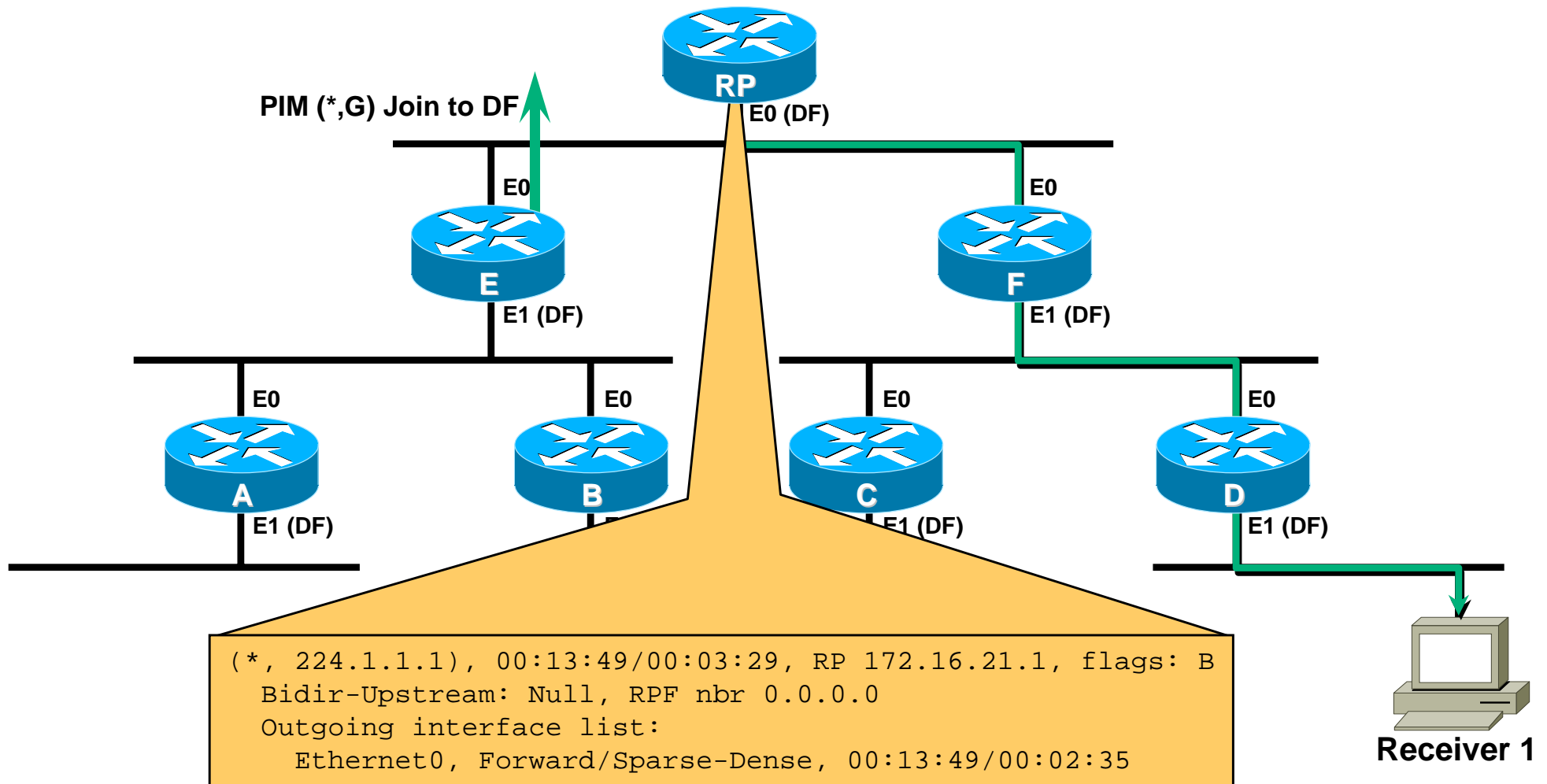
**Router "B" creates (\*, G) state.**

# Forwarding / Tree Building



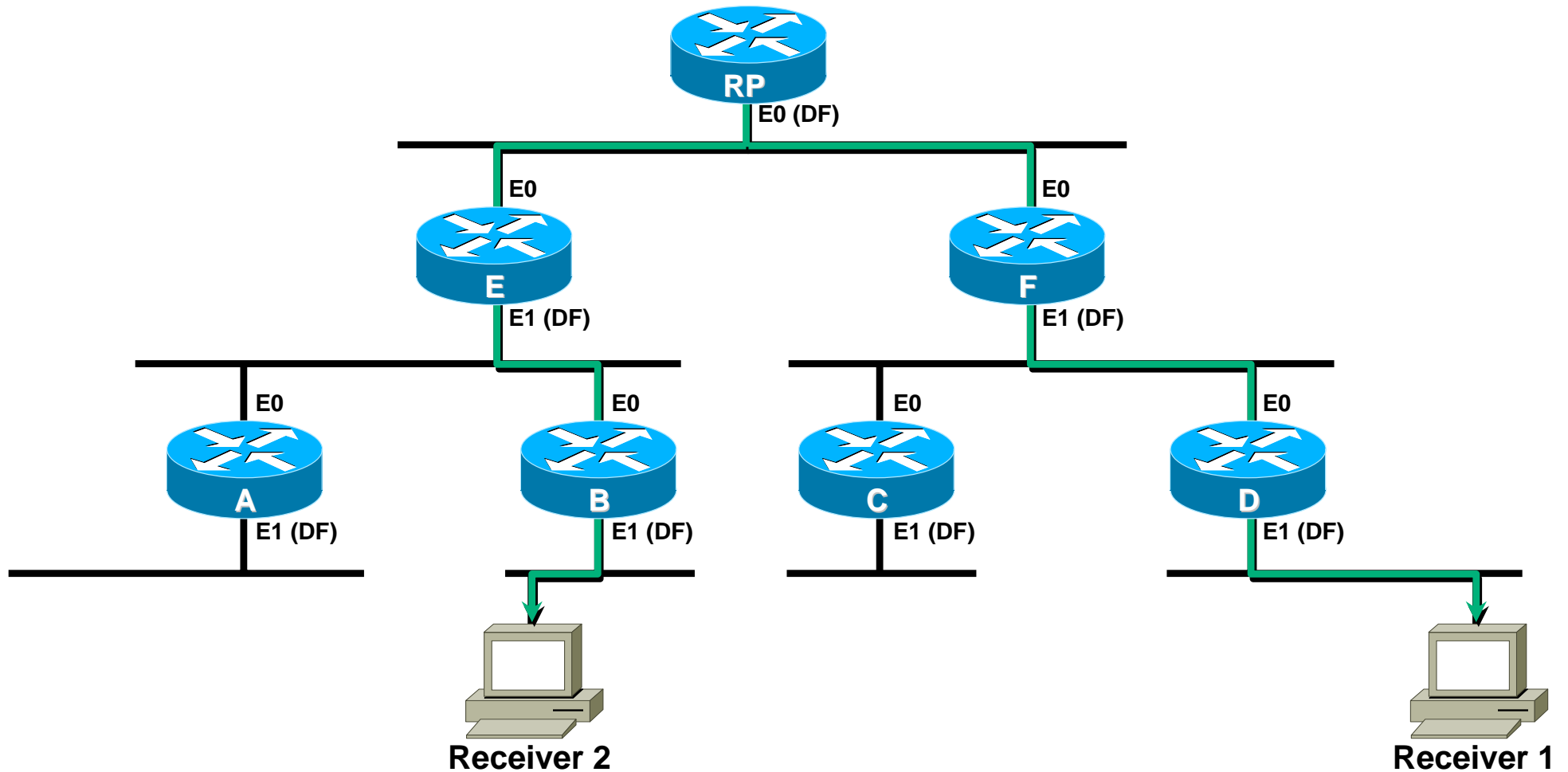
**Router “B” sends (\*, G) Join to “E” (DF) causing it to create (\*, G) state.**

# Forwarding / Tree Building



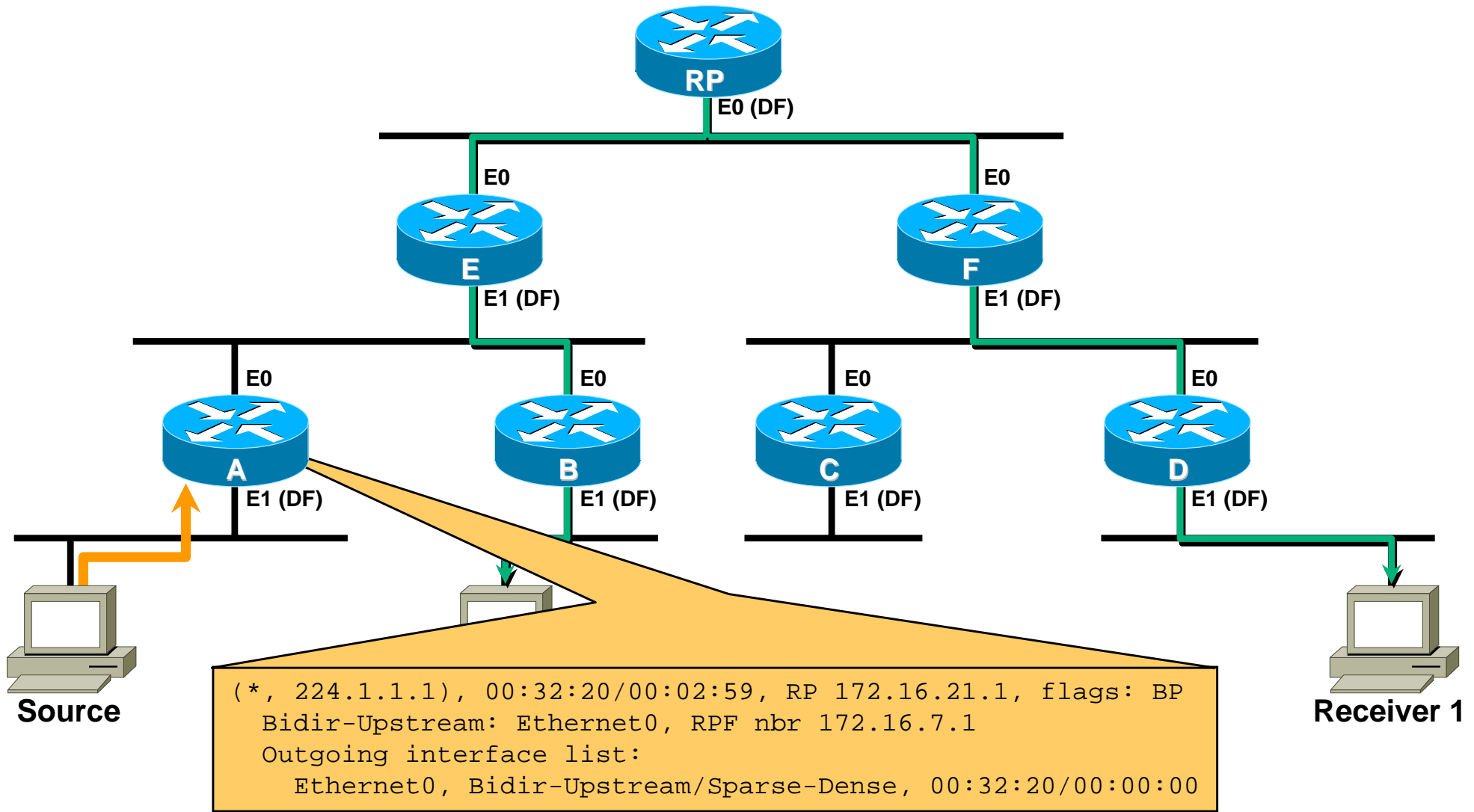
Router “E” sends (\*, G) Join to “RP”. (State on RP remains unchanged.)

# Forwarding / Tree Building



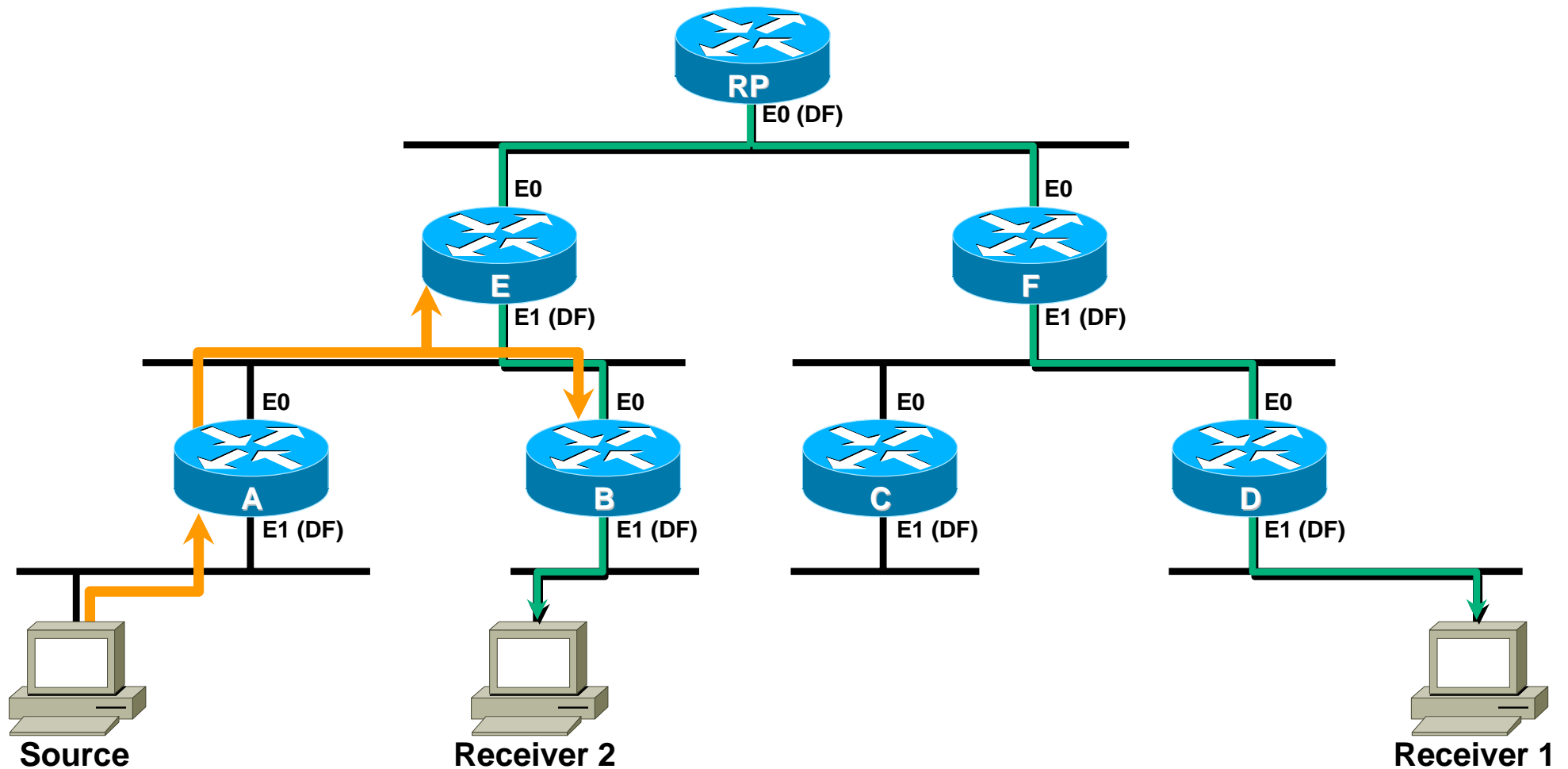
**New branch of Shared Tree is built to Receiver 2.**

# Forwarding / Tree Building



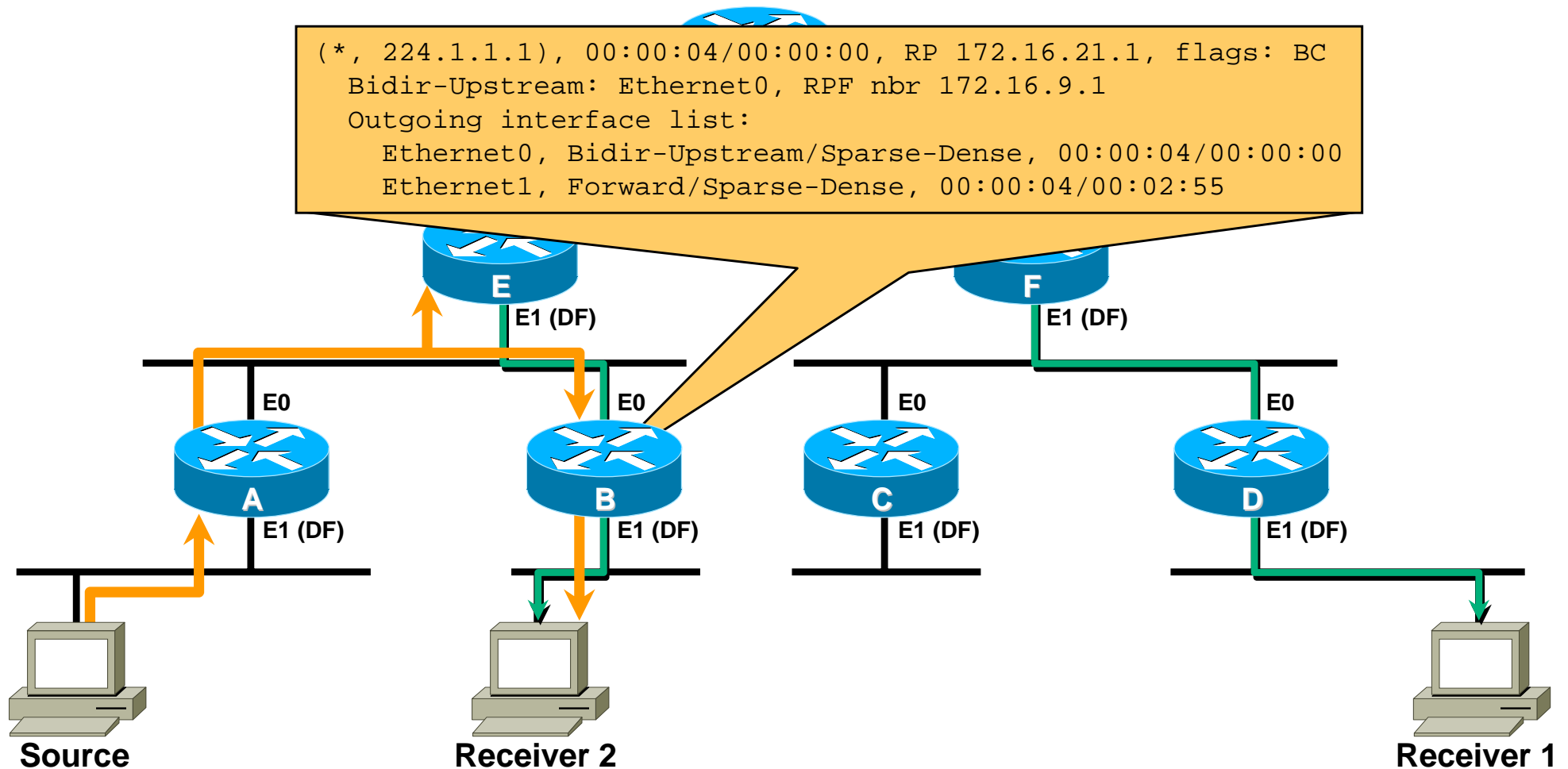
Arriving traffic from Source causes router “A” to create (\*, G) state.

# Forwarding / Tree Building



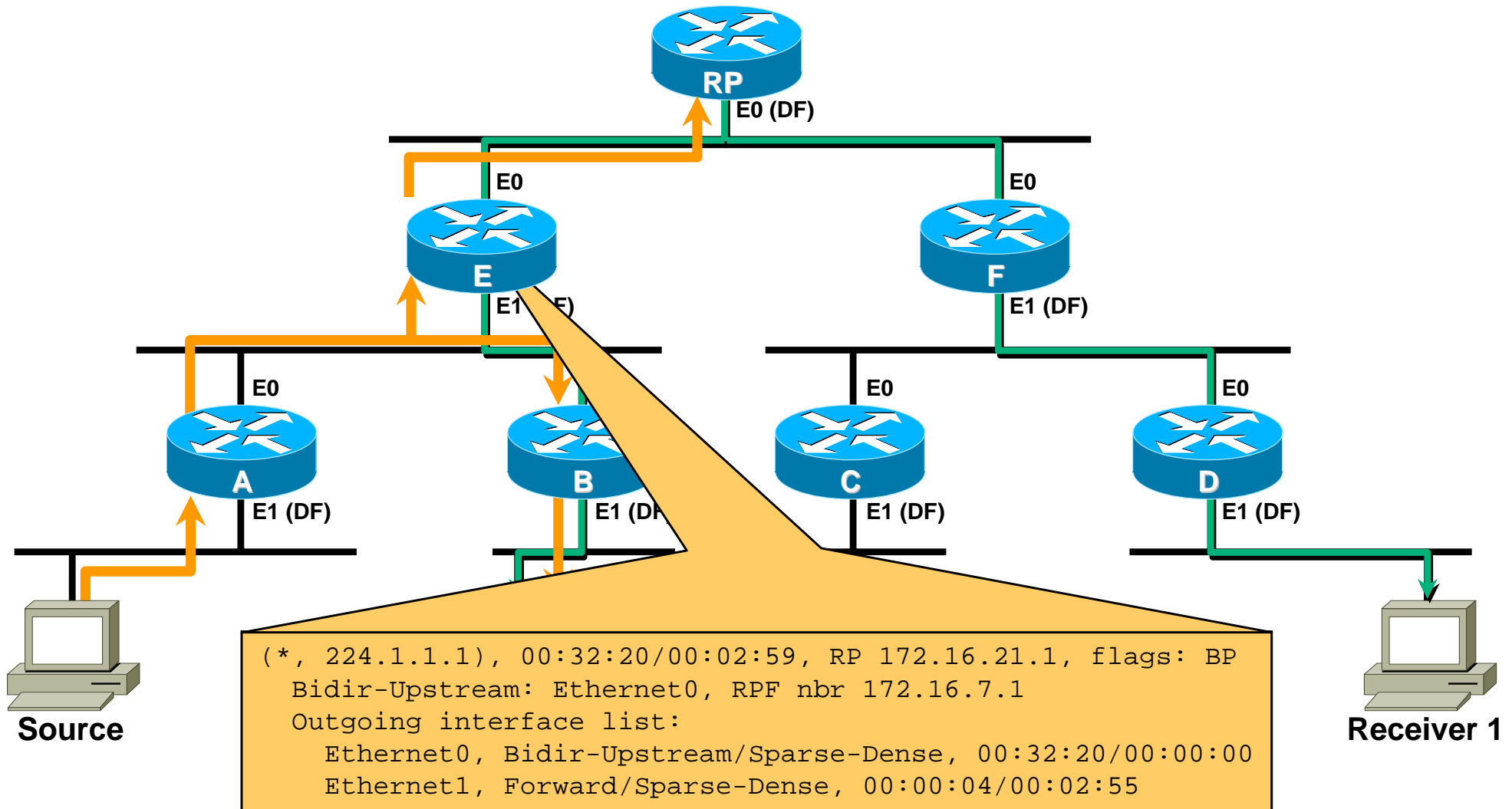
Traffic is forwarded toward router “E” and also arrives at IIF of router “B”.

# Forwarding / Tree Building



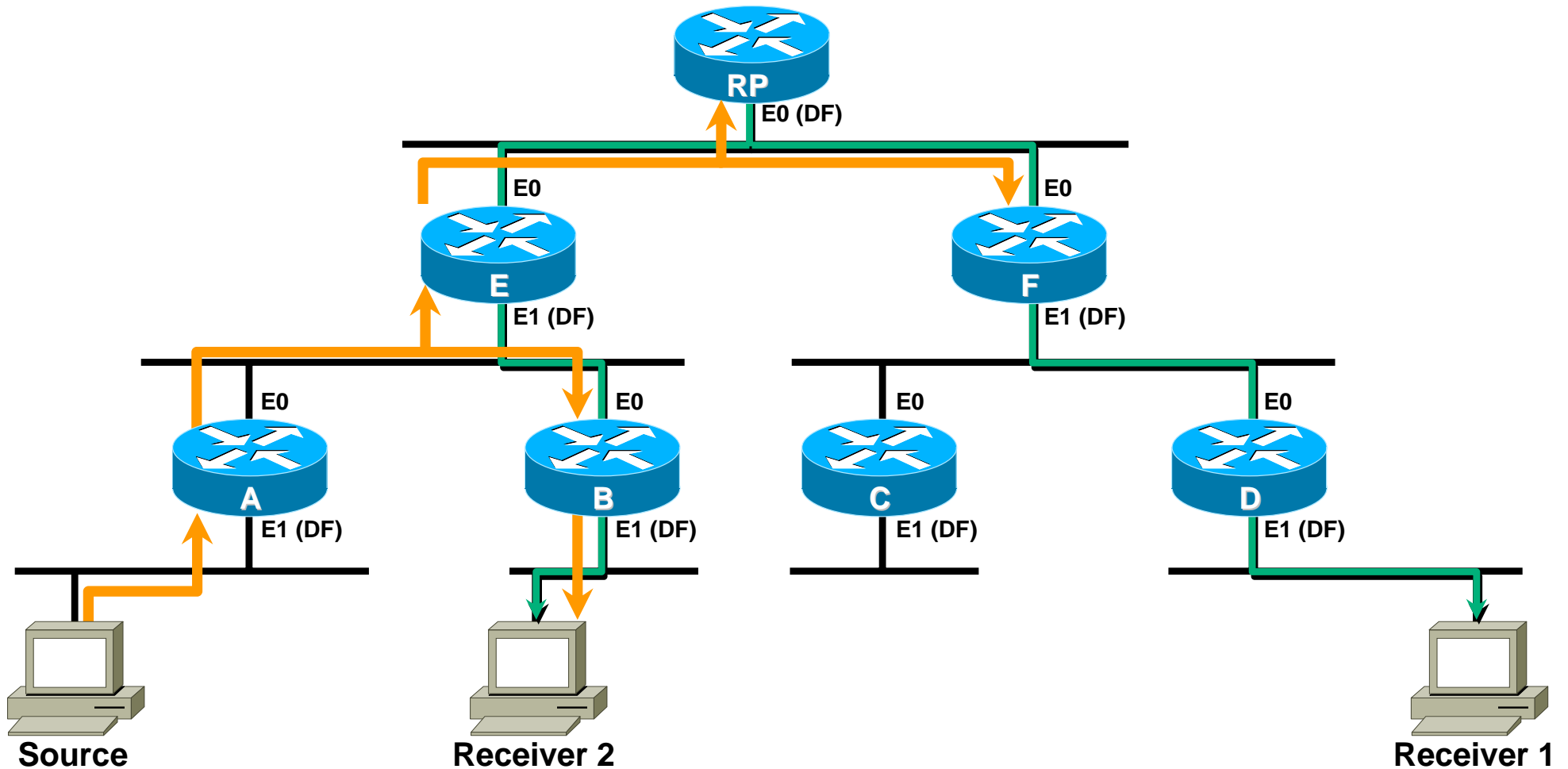
**Router “B” forwards traffic back down Shared Tree ala normal PIM-SM.**

# Forwarding / Tree Building



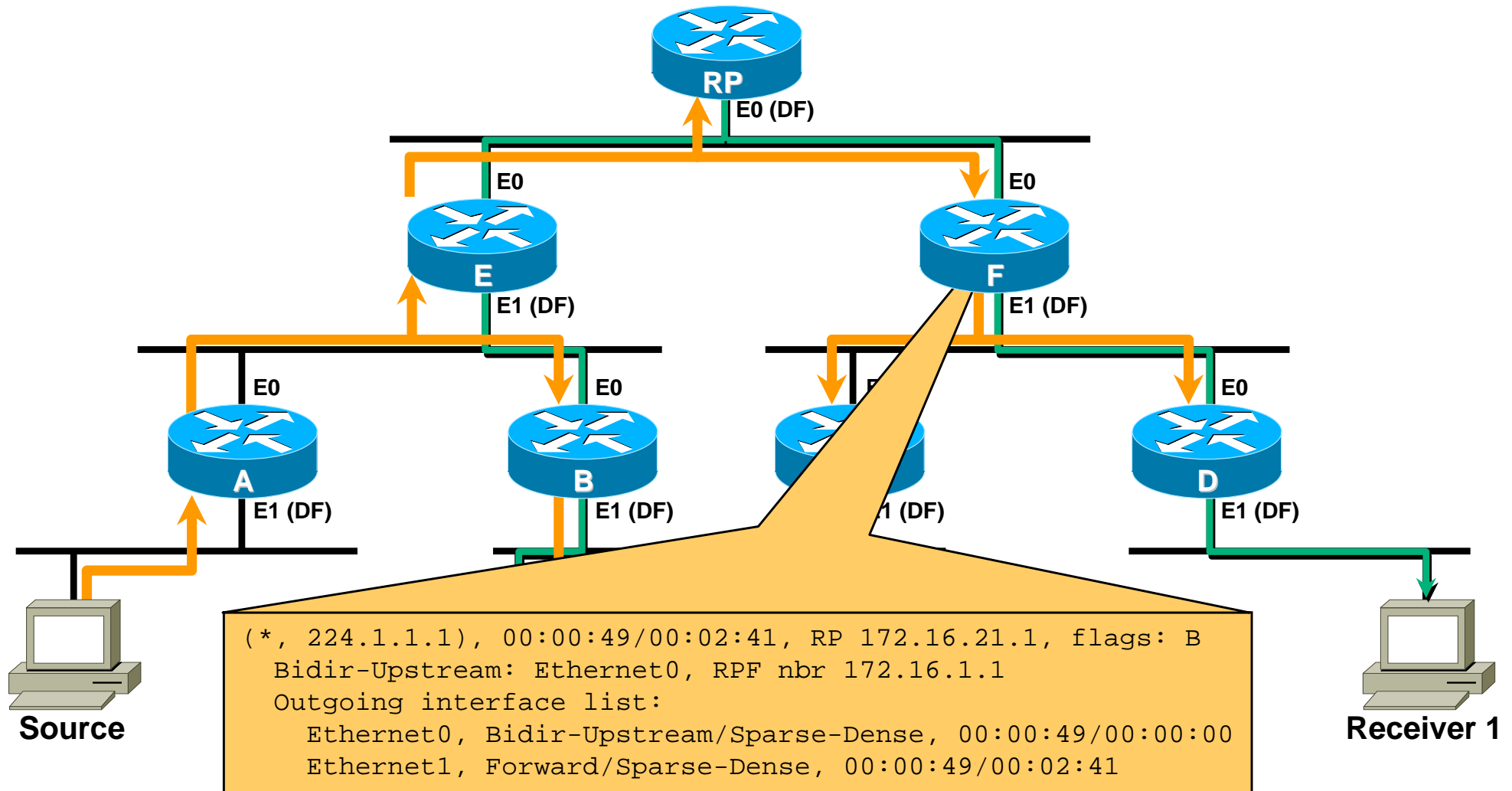
**Router "E" forwards traffic on toward RP .**

# Forwarding / Tree Building



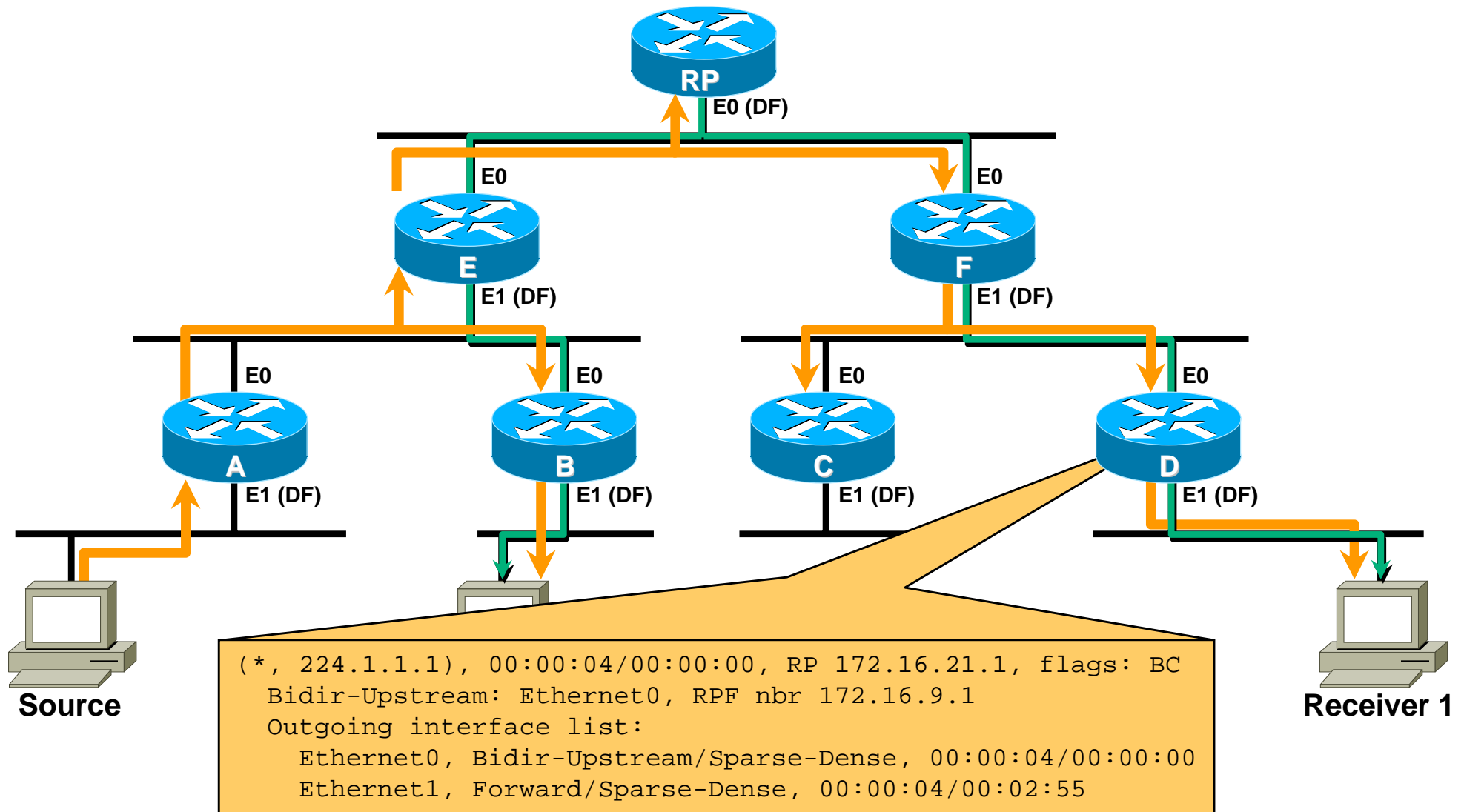
Traffic forwarded toward RP also arrives at the IIF of router “F”.

# Forwarding / Tree Building



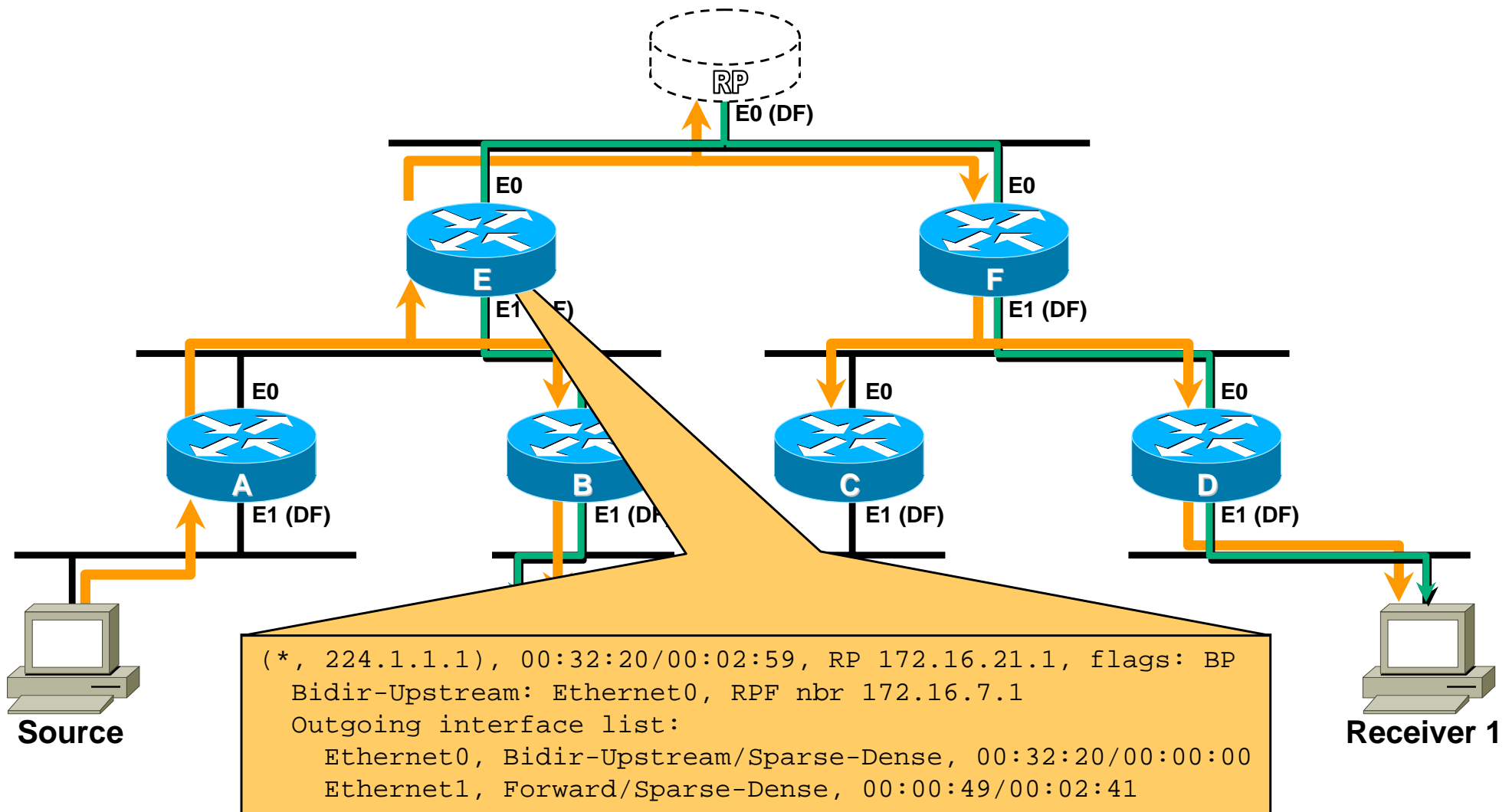
**Router “F” forwards traffic on down the Shared Tree ala normal PIM-SM.**

# Forwarding / Tree Building



**Router "D" forwards traffic to Receiver 1 via the Shared Tree.**

# Bidir PIM – Phantom RP



**Router “E” forwards traffic onto core LAN segment.**

# Configuring PIM Bi-Directional Mode

- Define Candidate RP and groups / modes it is willing to serve

```
ip pim rp-address Loopback0 45 bidir
```

```
ip pim rp-address Loopback1 46
```

```
! Two loopbacks needed due to a nature of ACLs (permit, deny)
```

```
access-list 45 permit 224.0.0.0 0.255.255.255
```

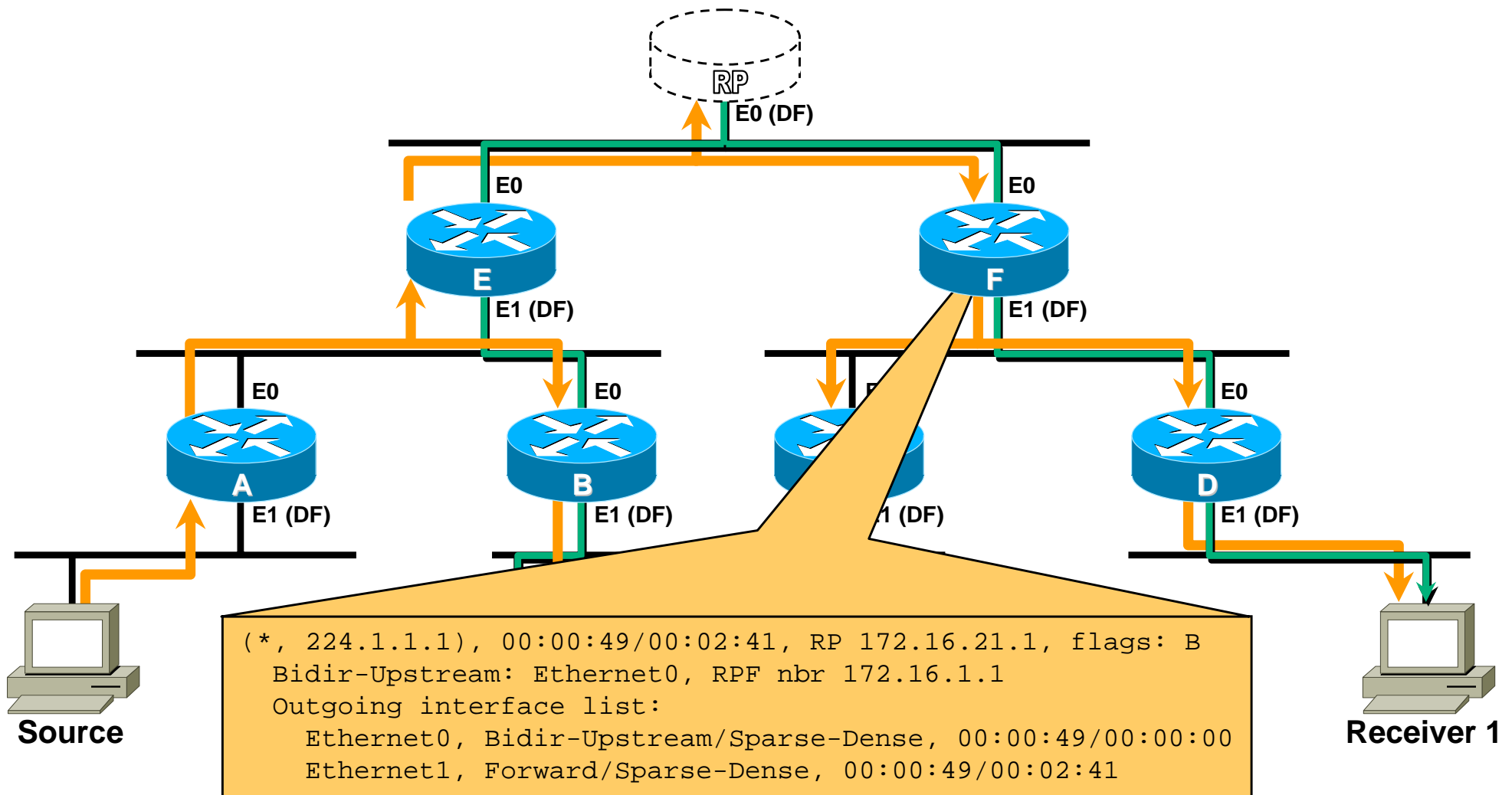
```
access-list 45 permit 227.0.0.0 0.255.255.255
```

```
! Those two groups will be PIM SM bi-directional
```

```
access-list 46 permit 226.0.0.0 0.255.255.255
```

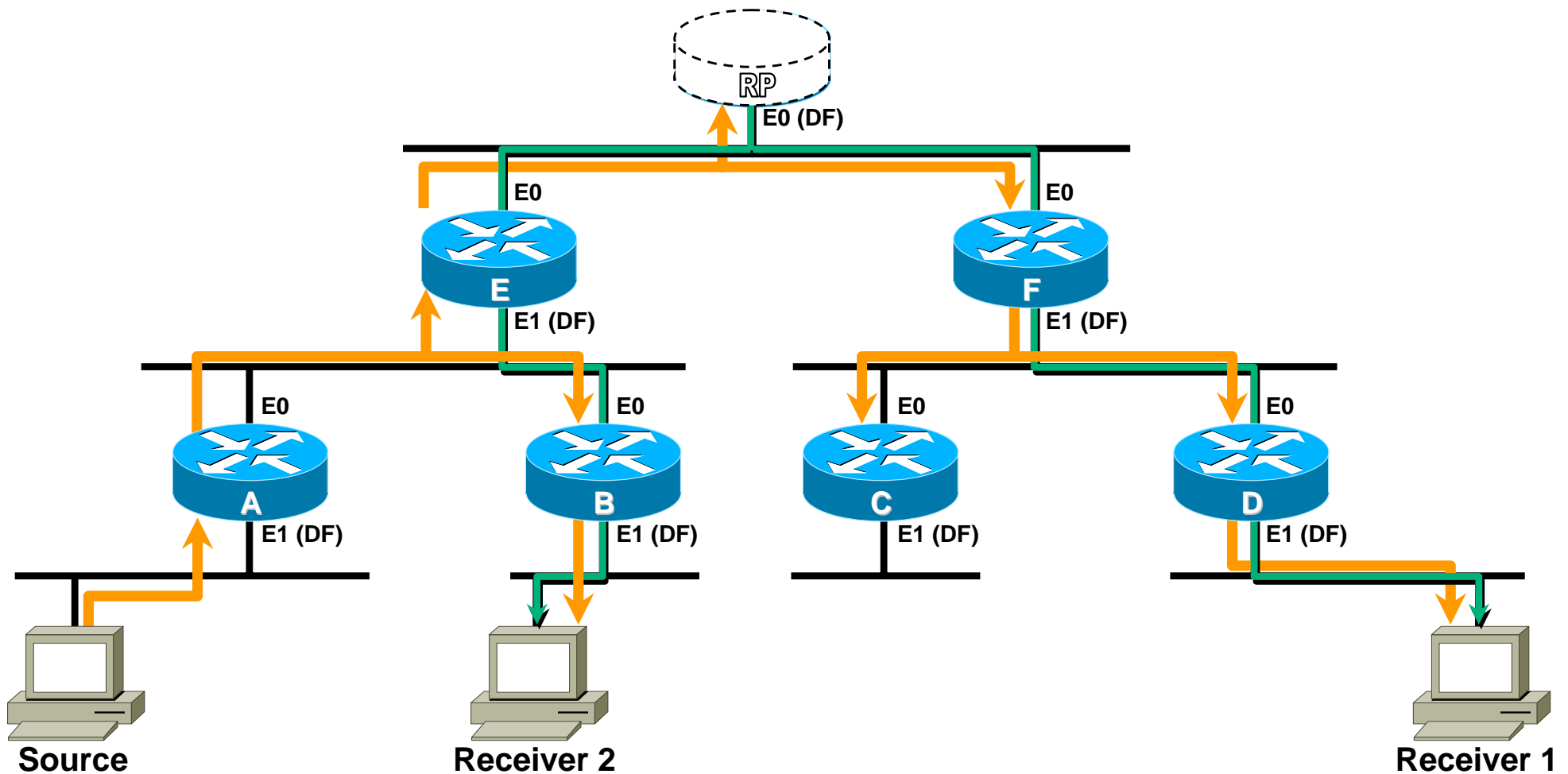
```
! This group will be PIM SM
```

# Bidir PIM – Phantom RP



**Router “F” forwards traffic on down the Shared Tree ala normal PIM-SM.**

# Bidir PIM – Phantom RP

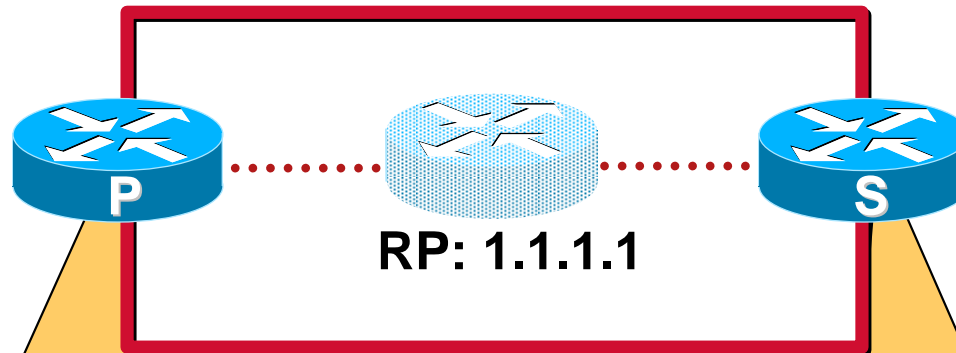


Question: Does a Bidir RP even have to physically exist?

Answer: No. It can just be a phantom address.

# Phantom RP on Point-to-Point Core

## Static Route Method



```
ip multicast-routing

interface Loopback0
 ip address 11.0.0.1 255.255.255.255
 ip pim sparse-mode

router ospf 11
 redistribute static subnets

ip route 1.1.1.1 255.255.255.255 Loopback0

ip pim bidir-enable
ip pim rp-address 1.1.1.1 bidir
```

```
ip multicast-routing

interface Loopback0
 ip address 11.0.0.2 255.255.255.255
 ip pim sparse-mode

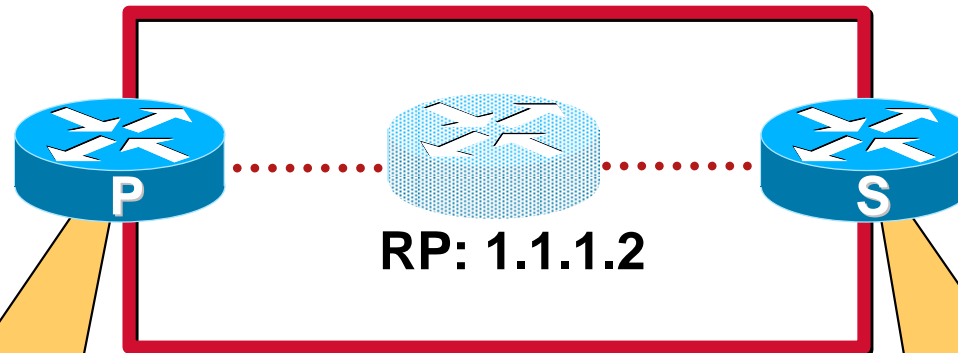
router ospf 11
 redistribute static subnets

ip route 1.1.1.0 255.255.255.254 Loopback0

ip pim bidir-enable
ip pim rp-address 1.1.1.1 bidir
```

# Phantom RP on Point-to-Point Core

## Netmask Method



```
ip multicast-routing
!  
interface Loopback0  
ip address 1.1.1.1 255.255.255.252  
ip pim sparse-mode  
ip ospf network point-to-point  
!  
router ospf 11  
network 1.1.1.0 0.0.0.3 area 0  
network 10.1.1.0 0.0.0.255 area 0  
network 10.1.2.0 0.0.0.255 area 0  
!  
ip pim bidir-enable  
ip pim rp-address 1.1.1.1  
ip pim rp-address 1.1.1.2 bidir
```

```
ip multicast-routing
!  
interface Loopback0  
ip address 1.1.1.1 255.255.255.248  
ip pim sparse-mode  
ip ospf network point-to-point  
!  
router ospf 11  
network 1.1.1.0 0.0.0.7 area 0  
network 10.1.1.0 0.0.0.255 area 0  
network 10.1.2.0 0.0.0.255 area 0  
!  
ip pim bidir-enable  
ip pim rp-address 1.1.1.1  
ip pim rp-address 1.1.1.2 bidir
```

# Bidir PIM – Summary

- Uses Shared Trees only.

Single (\*, G) forwarding entry per group.

Source traffic flows up and down Shared Tree.

- Drastically reduces network mroute state.

Eliminates ALL (S,G) state in the network.

By eliminating SPT between source and RP.

Allows Many-to-Any applications to scale.

Permits virtually an unlimited number of sources.

# Q & A



