CS144: An Introduction to Computer Networks

Routing: How do packets know the way?

Today:

BGP: The Border Gateway Protocol

(Part 3 of 3)

Nick McKeown

BGP: The Border Gateway Protocol

Border Gateway Protocol

BGP does not use Bellman-Ford or Dijkstra algorithm

BGP routers advertise routes to their neighbors, containing:

- A prefix (a contiguous set of destination addresses)
- The list of AS's a packet will traverse to reach its destination.

Example of path advertisement:

"The network 171.64/16 can be reached via the path {AS1, AS5, AS32, AS13}"

Q: Why advertise a path of AS's for each prefix, rather than

a. The next hop for each prefix

b. The path of IP addresses

Border Gateway Protocol

"The network 171.64/16 can be reached via the path {AS1, AS5, AS32, AS13}"

Paths with loops are detected locally and ignored.

A BGP router may connect to several peers and receive multiple different advertised paths for the same prefix.

Local policies chosen by the AS administrator pick the preferred path.

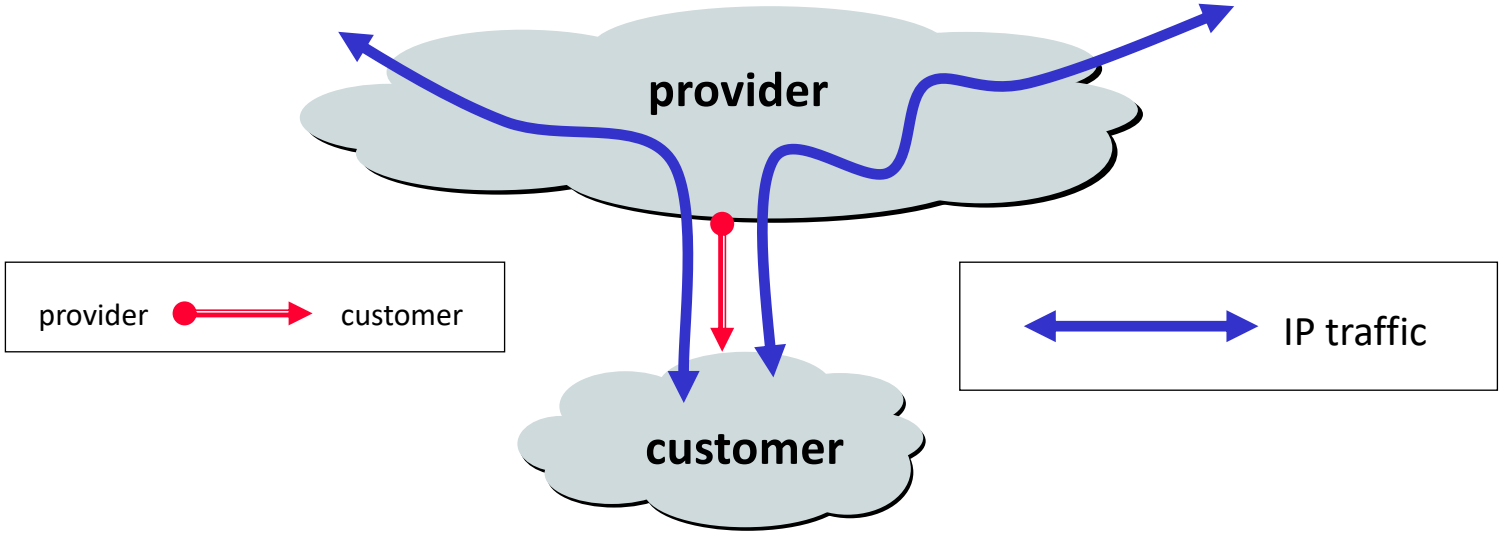
Border Gateway Protocol: Details

- BGP neighbors ("peers") establish a TCP connection.
- The TCP connection is manually configured at both ends.
- Neighbors send "keepalive" messages every 60 seconds.

- BGP is sometimes called a "Path vector" algorithm.
- It is not a link-state or a distance-vector routing protocol.

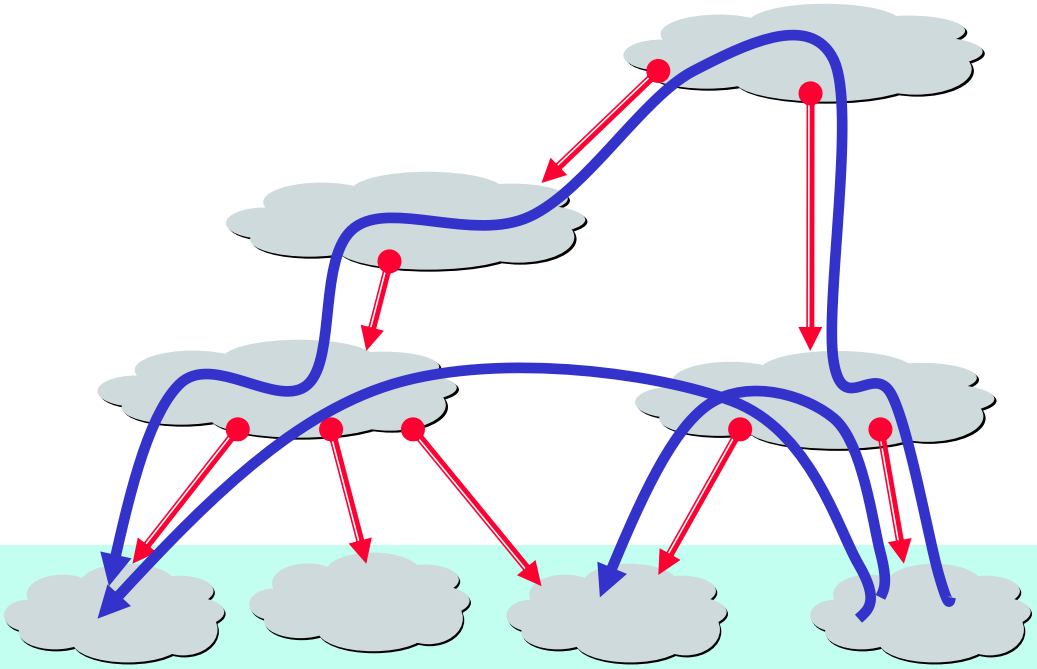
- When an advertised path changes, the path vector is first "withdrawn", then the new one is advertised.

Customers and Providers



Customers pay providers to carry their packets.

Customer-Provider Hierarchy



Transit AS's

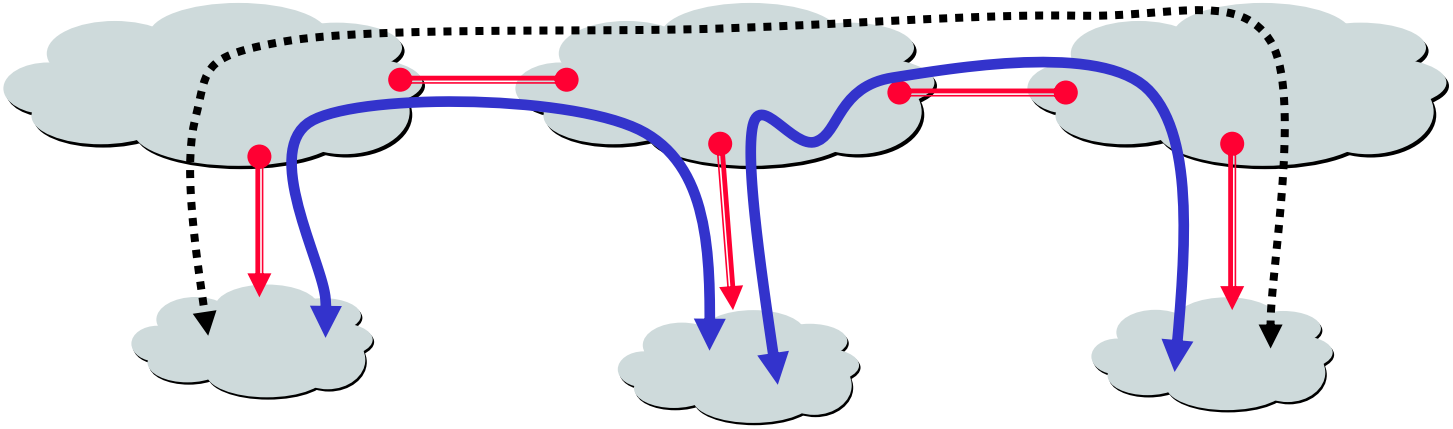
Stub AS's

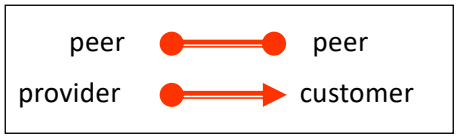
Routers inside Stub AS's can use a "default route" for unknown IP addresses. The default route is the Border router.

provider → customer

IP traffic ←

The Peering Relationship





traffic allowed

Transit traffic NOT allowed

Peers provide transit between their respective customers

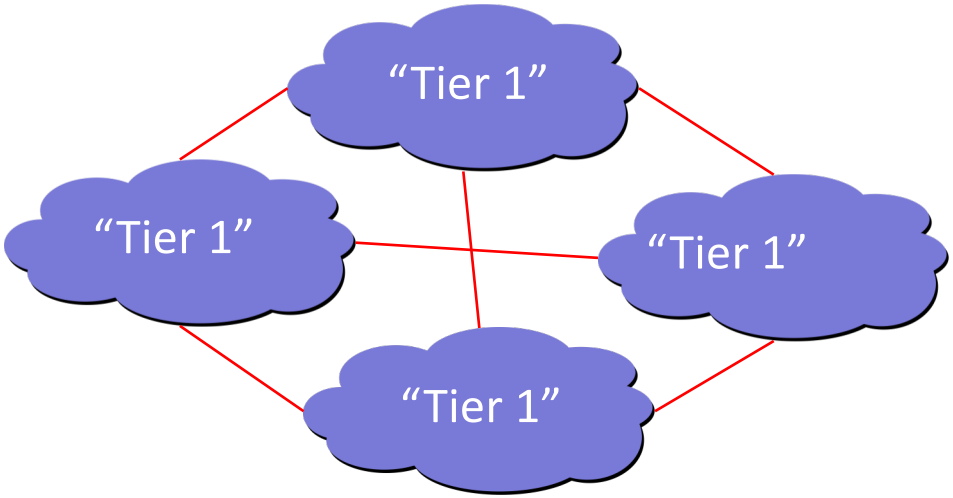
Peers do not provide transit between peers

Peers (typically) do not exchange $$$

So how does traffic from the left side reach the right side?

“Tier 1” Providers

A **Tier 1** network is a transit-free network that peers with every other tier 1 network



Tier 1 ISPs

Definition: A **Tier 1 ISP** has access to the entire ***Internet Region*** solely via its free and reciprocal peering agreements.

Definition: An **Internet Region** is a portion of the Internet network typically bounded by a country's geographical boundaries.

Each Internet Region has its own set of "**Tier 1 ISPs**."

The litmus test:

"*Does an ISP pay anyone to reach any destination in the Internet Region*?"
If the answer is "No" then it is a **Tier 1 ISP**, and
If the answer is "Yes" then it is a **Tier 2 ISP**.

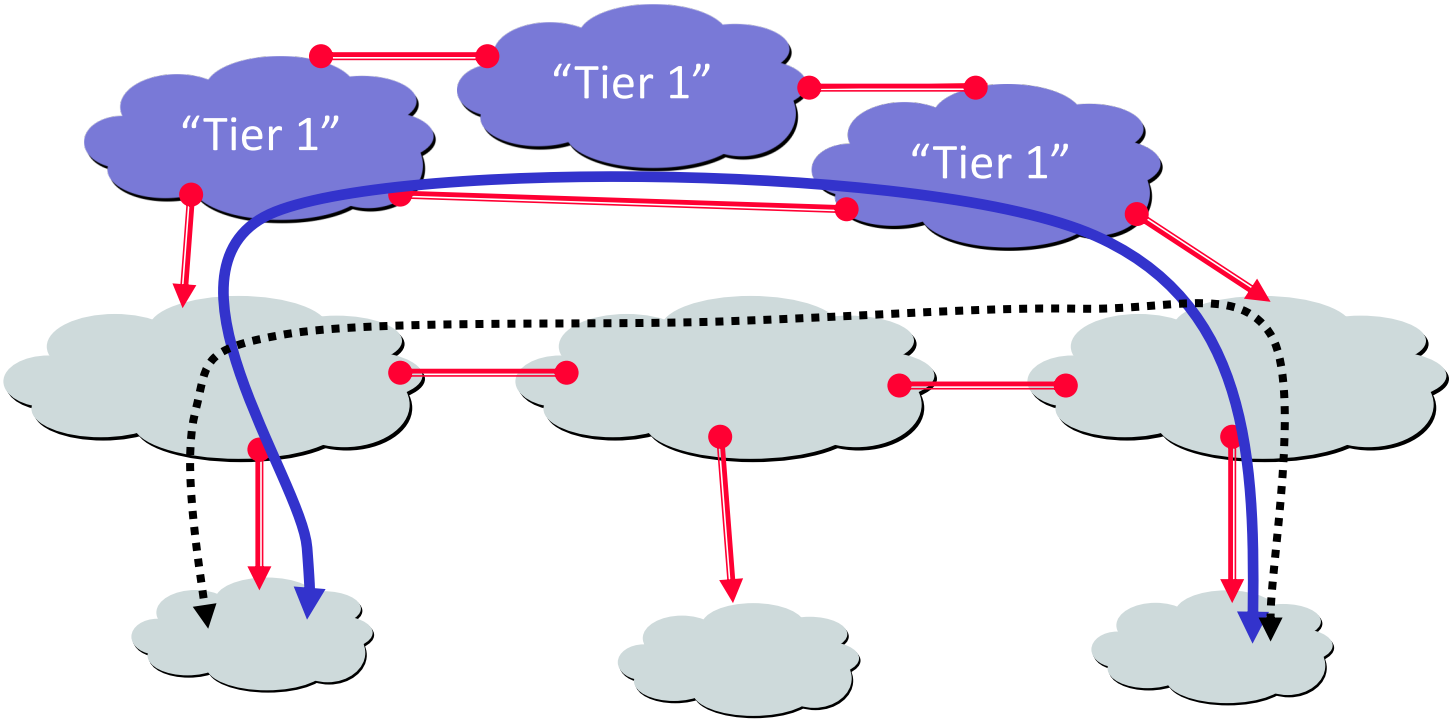
Tier 1 ISPs by country

The U.S. Internet Region Tier 1 ISPs

1. AT&T
2. Verizon
3. Sprint (Softbank Broadband)
4. Century Link (Qwest)
5. Level 3 (with Global Crossing now)
6. NTT/Verio
7. Cogent

The Japan Internet Region Tier 1 ISPs

1. NTT
2. Japan Telecom (Softbank)
3. KDDI
4. IIJ
5. Powered.com



Peers provide transit between their respective customers

Peers do not provide transit between peers

Peers (typically) do not exchange $$$

Other types of routing

Multicast

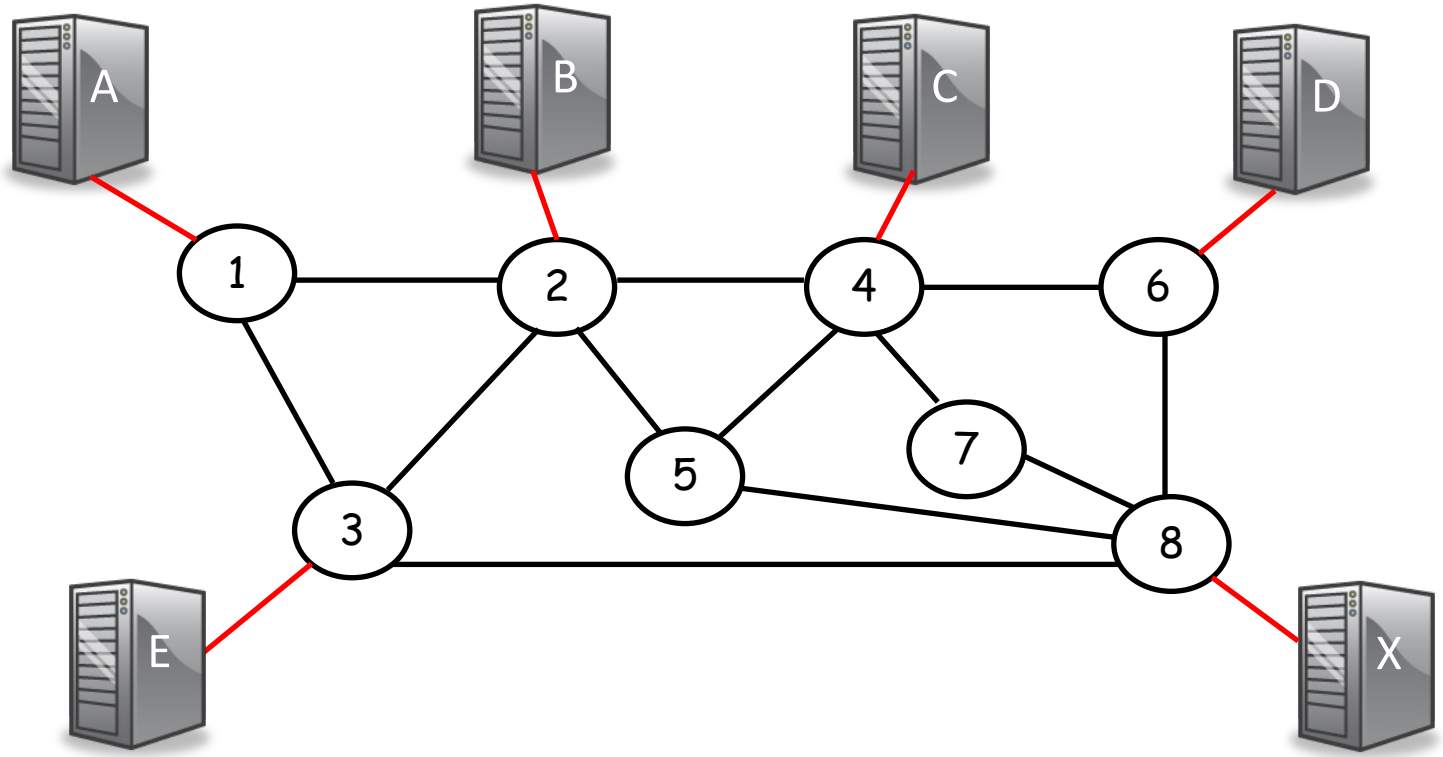
Anycast

Multicast

IP multicast datagram
to IP DA = Group 1
from IP SA = A

Group 1 = A, B, C, E, X

Data	A	Group 1

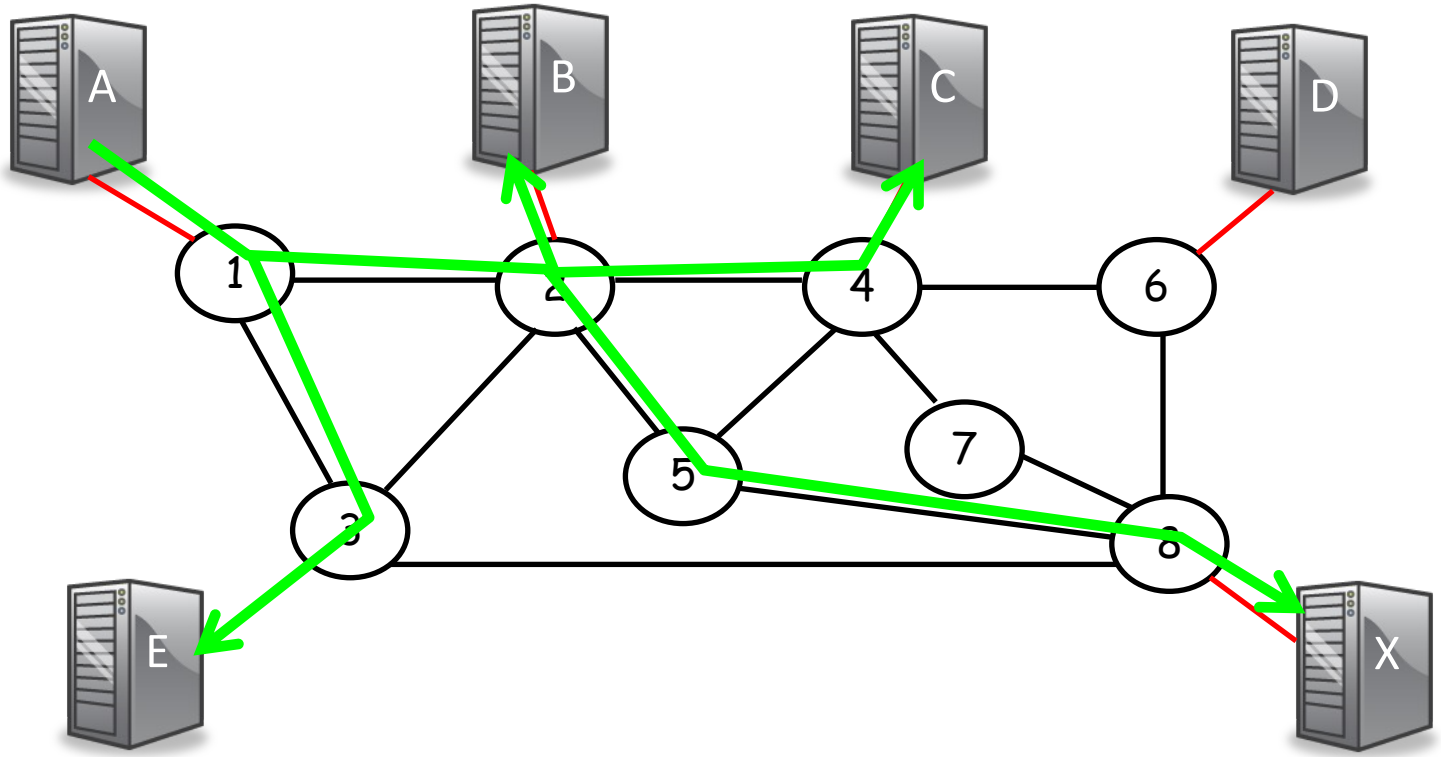


Multicast

IP multicast datagram
to IP DA = Group 1
from IP SA = A

Group 1 = A, B, C, E, X

Data	A	Group 1



Multicast

Technique

- Reverse Path Broadcast (RPB) and Pruning

Practice

- IGMP – group management
- DVMRP – the first multicast routing protocol
- PIM – protocol independent multicast

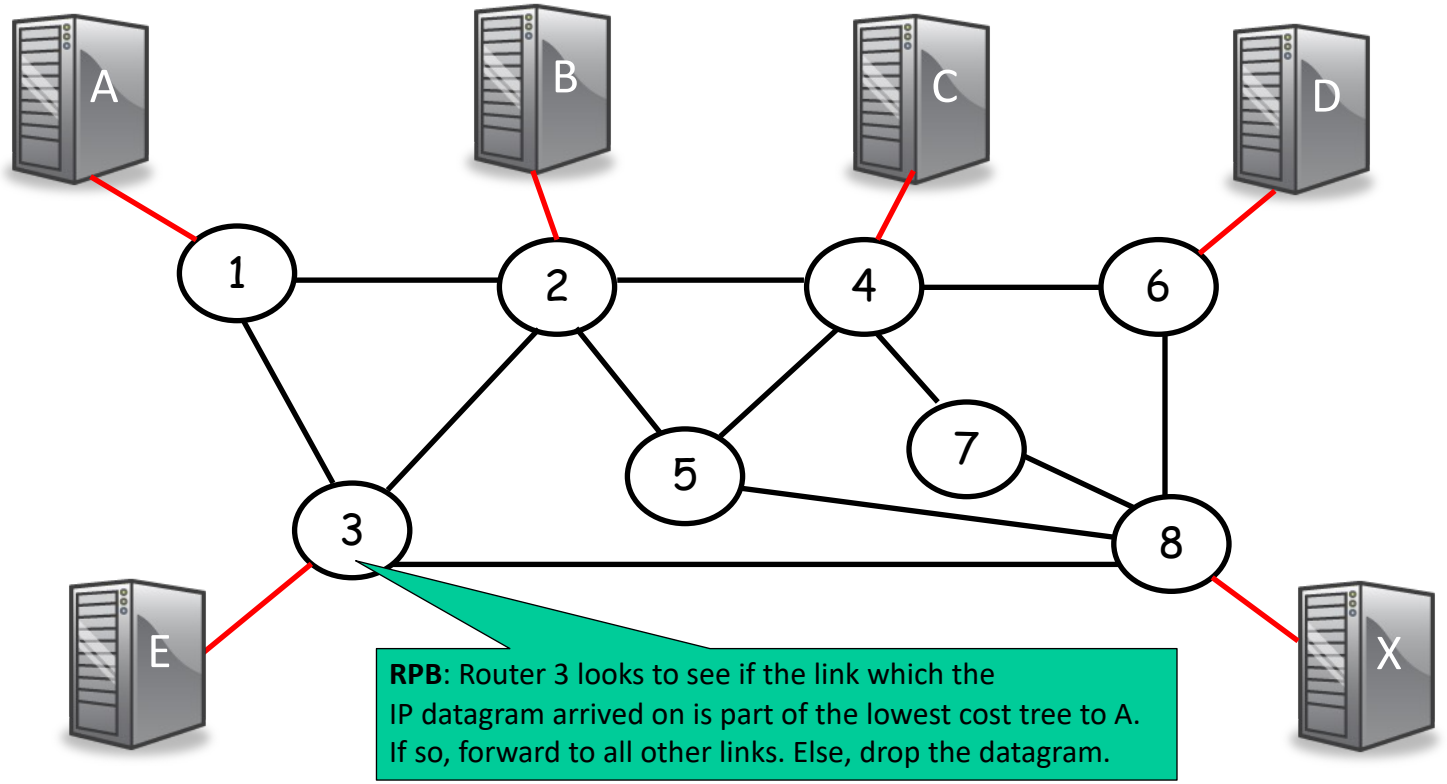
Reverse Path Broadcast (RPB)

aka Reverse Path Forwarding (RPF)

IP multicast datagram
to IP DA = Group 1
from IP SA = A

Group 1 = A, B, C, E, X

Data	A	Group 1



RPB + Pruning

1. Packets delivered loop-free to every end host.
2. Routers with no interested hosts send prune messages towards source (IGMP).
3. Resulting tree is the minimum cost spanning tree from source to the set of interested hosts in the multicast group.