CS-340 Introduction to Computer Networking

Lecture 12: Broadcast

Steve Tarzia

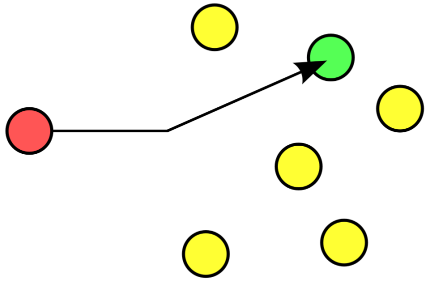
Last Lecture: BGP routing

- DV **Count to infinity** leads to slower convergence when links get worse.
  - Good news travels quickly, bad news travels slowly.
- Internet routing is hierarchical.
- **Autonomous systems** (ASes) are grouped routers with one routing policy.
- An **Interior Gateway Protocol** (IGP) (eg., OSPF) determines optimal routes within an AS.
  - Can use a centralized (Link State) shortest path algorithm, like Dijkstra's.
- The **Border Gateway Protocol** (BGP) determines routes between ASes.
  - Uses a distributed shortest-AS-hop path algorithm (Distance Vector).
- BGP **advertisement** includes a list of routes, each looking like:
  - {PREFIX: 43.5.0.0/16, AS-PATH: [AS4, AS65, AS1], NEXT-HOP: 5.6.7.200)}
  - This tells a neighboring AS that it can forward packets to the prefix.

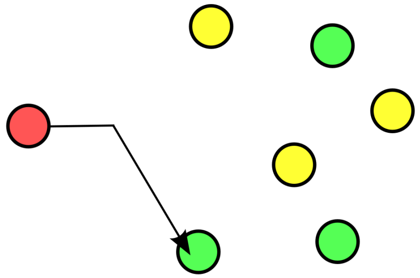
Generalized packet delivery

- IP networks were built for **unicast** addressing (sending a packet to a single destination address) but applications have more general needs:

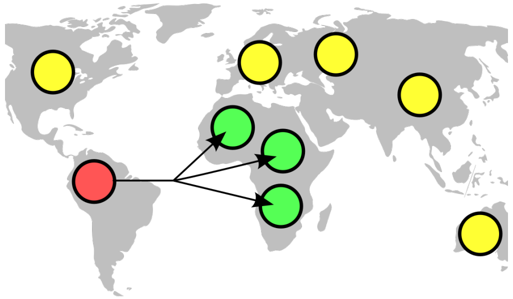
Unicast: to one node



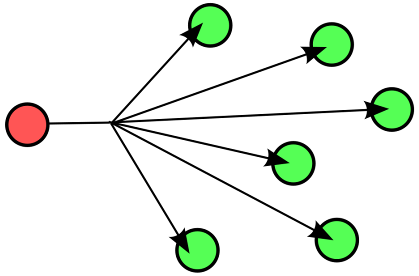
Anycast: to one of many possible nodes



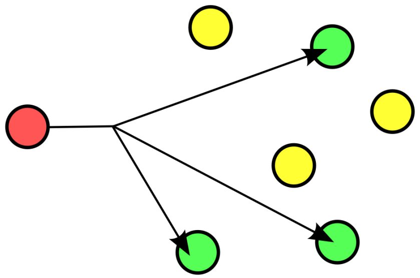
Geocast: to nodes in a specific location



Broadcast: to all nodes



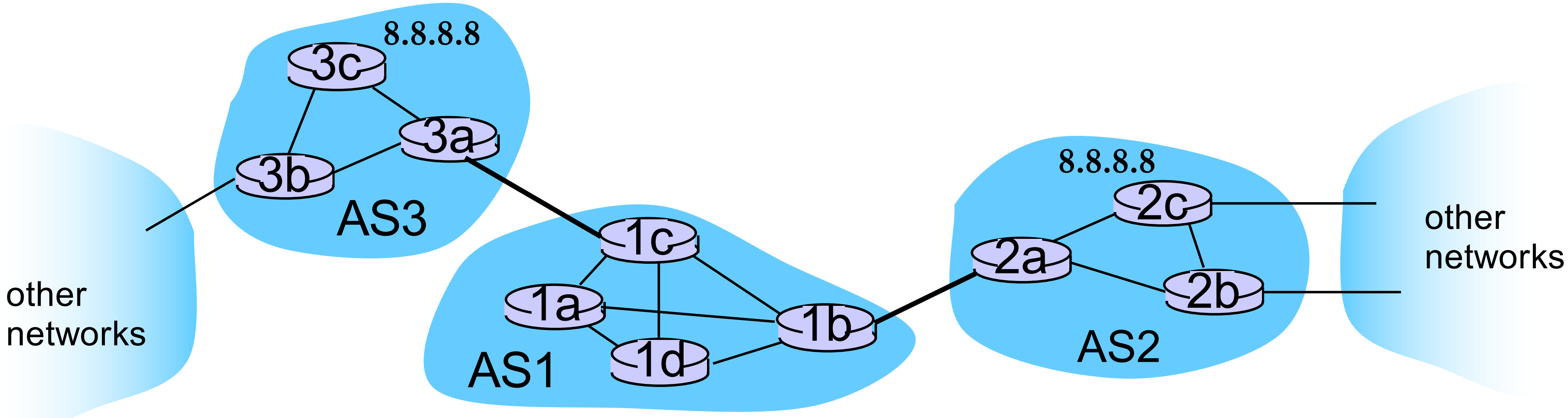
Multicast: to a set of nodes



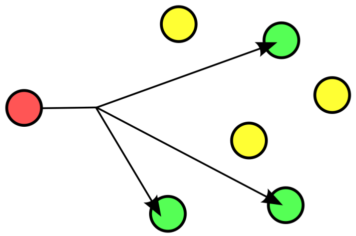
IPv4 anycast hack

Remember, this is Google's public DNS server

- The single IP address 8.8.8.8 maps to may different machines across the world, using a hack called **IPv4 anycast**.
- Multiple border gateways (routers) advertise short routes to 8.8.8.8.
- Within the AS, traffic is directed to the closest 8.8.8.8 instance.
- Below, there is nothing stopping two AS'es from advertising short routes to the same IP address, thus creating two copies of the same IP address.
- **BGP Hijacking** is the *malicious* takeover of another AS'es IP addresses.



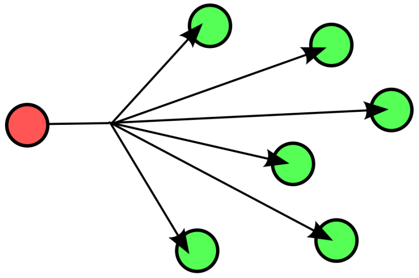
IP Multicast *(optional)*



- Certain IP addresses are reserved to identify **broadcast groups**.
  - Eg., 239.1.1.1. Traffic sent to that address is forwarded to every host that has subscribed to that broadcast group using Internet Group Management Protocol (IGMP).
- Adds burden to routers, since they must track all broadcasts active at that time. (To which neighbors should I forward multicast traffic?)
  - More routing info to track (in addition to BGP) and with finer granularity.
- Public Internet routers generally do **not** support IP Multicast.
- Used in special scenarios for streaming content to many viewers.
  - Eg., IPTV in hotels or video broadcast on a campus.
- Applications must use UDP, not TCP
  - (because sender would have to track ACK status of thousands of receivers)

Broadcast as a basic network building block

- IP (the Internet) does not support broadcast. Why?
  - No one message is relevant to all hosts on the entire Internet!
- Broadcasts are crucial in small-to-moderate sized **ad hoc networks**:
  - Networks of neighboring hosts that connect without prior planning.
  - Also called **mesh networking** and MANETS.
  - *Firechat* (2014–18) was a P2P chat app using bluetooth and ad-hoc WiFi.
- Broadcasts can be used to share information about network structure to later enable unicast routing.



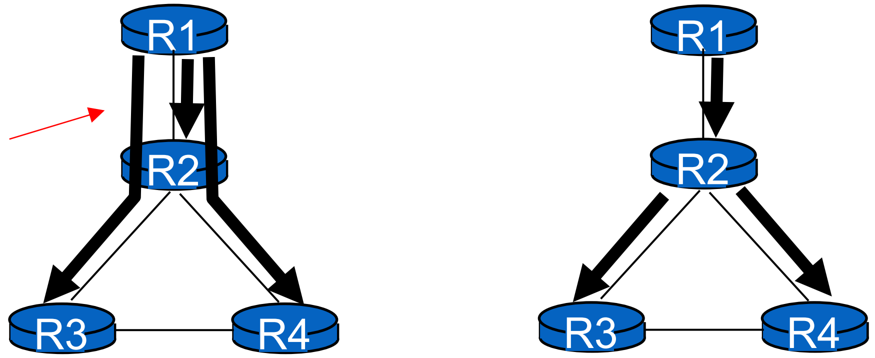
How to implement link state (LS) flooding?

Remember, a "link state" routing algorithms uses Dijkstra's algorithms to compute the shortest paths. Information on the whole network must be gathered in one place. Another way of looking at this is that any information about every link and node must be broadcast to everyone in the network.

Broadcasting (flooding)

- Naïve approach is to *unicast* the message to each node. Downsides?
  - Requires prior knowledge of the full network and it is inefficient:

First hop has duplicates



3x unicast   broadcast

- **Broadcast** allows a message to be **duplicated in transit** to reach different parts of the network.
- Must be careful to ensure:
  - **Correctness**: all nodes get the message.
  - **Efficiency**: ideally, message is received just once by each node
  - **Termination**: message is not retransmitted forever.

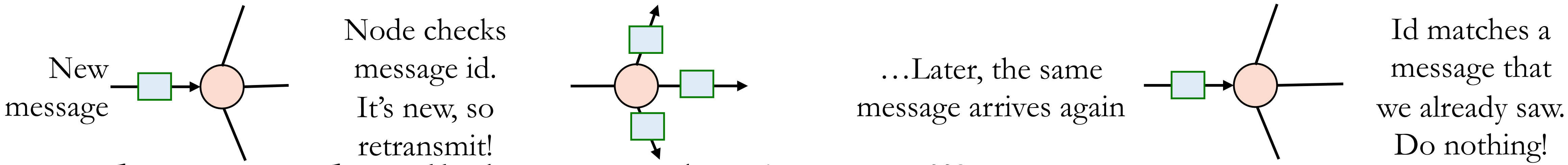
How would you implement broadcast?

Assume that each node just knows about its neighbors

What challenges arise?

Broadcast strategy #1: **Controlled Flooding**

- Nodes retransmit packet only the first time it is seen by a node.
  - Obviously, don't retransmit it on the link it was received on.



- How does a node tell that a packet is "new"?
  - Broadcast packets must contain a *unique identifier*.
  - Nodes must *remember* all recently-seen message identifiers.

- In **Link-State algorithm**, when flooding link state, messages include:
  ⟨link source, link destination, sequence number, link cost⟩

  uniquely identifies the message

LS flooding demo

Why do we need sequence numbers?

msg : <src, dst, seq, cost>

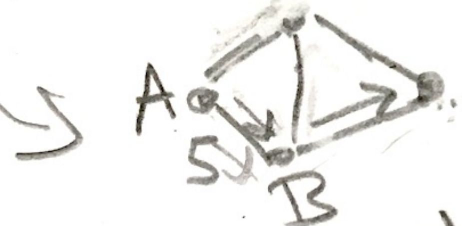
assume A sees linkupdate first (A,B)=3

A broadcast <A, B, ∅, 3>

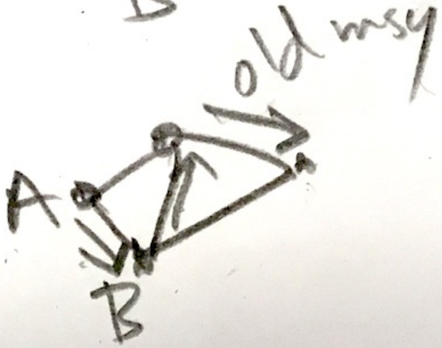
A also stores this msg as "received"

Eventually A gets its own msg back!

Why need a Seq Number?



A broadcast <A, B, 1, 5>



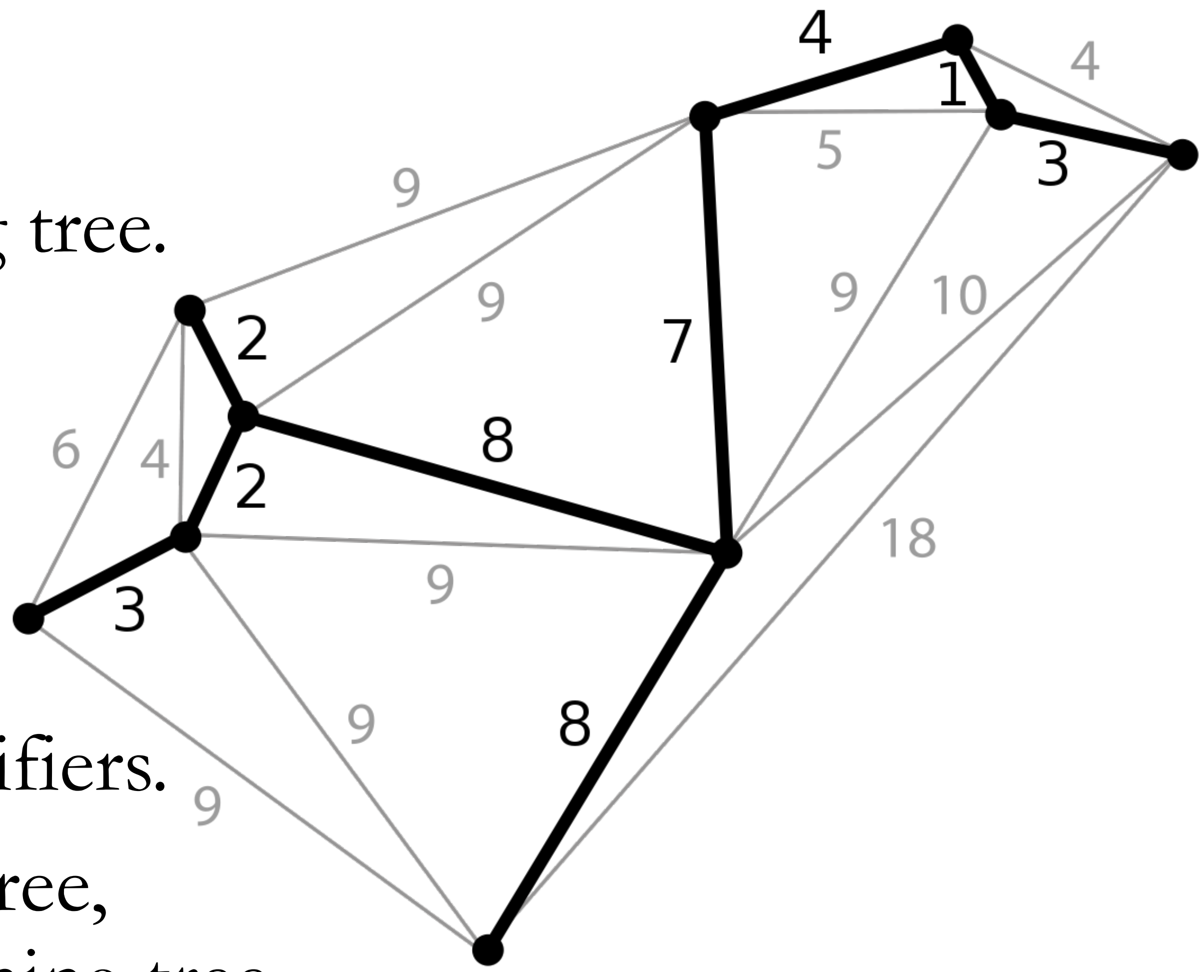
Broadcast strategy #2: **Spanning Tree**

- In a graph, a **spanning tree** is a subset of edges that connects all the nodes and has **no cycles** (loops).
- To implement broadcast efficiently, "flood" message only along spanning tree.

✓All nodes will be reached.

✓Lack of cycles prevents duplication and guarantees termination.

- No need to remember message identifiers.
- But must somehow find a spanning tree, and all nodes must use the *same* spanning tree.



Simplest spanning tree algorithm

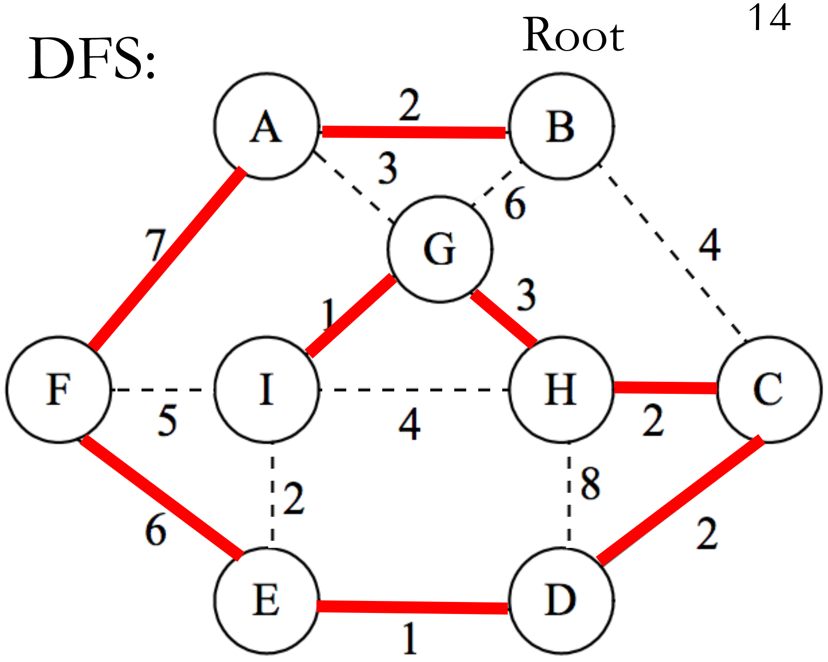
- |V| times, find an edge that does not create a **cycle** and add it to the tree.

In more detail:

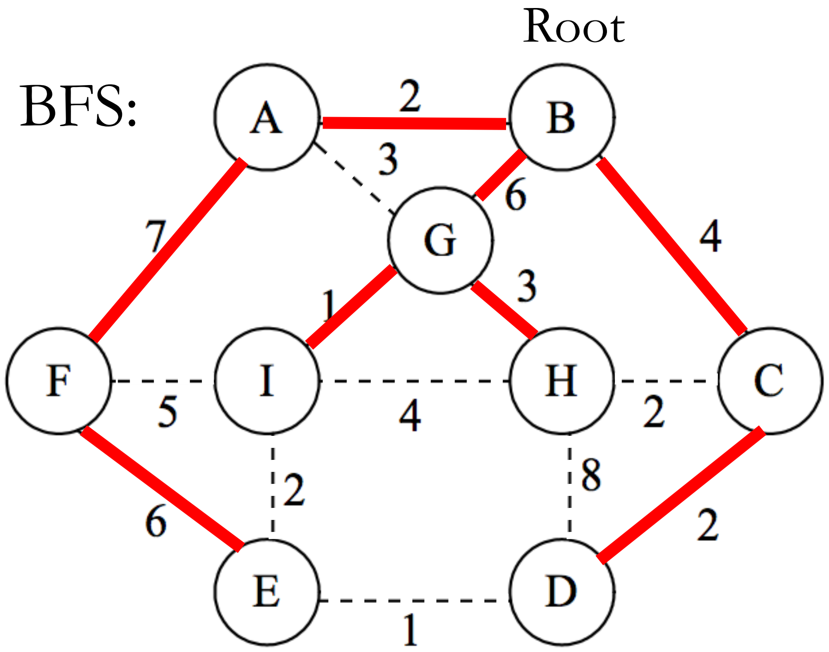
- Do a depth-first or breadth-first search for **unvisited** nodes, adding the connecting edges.
- Track the nodes already in the tree (visited) and ignore edges leading to those.

- Does **not** compute a *minimal-cost* spanning tree.

DFS:

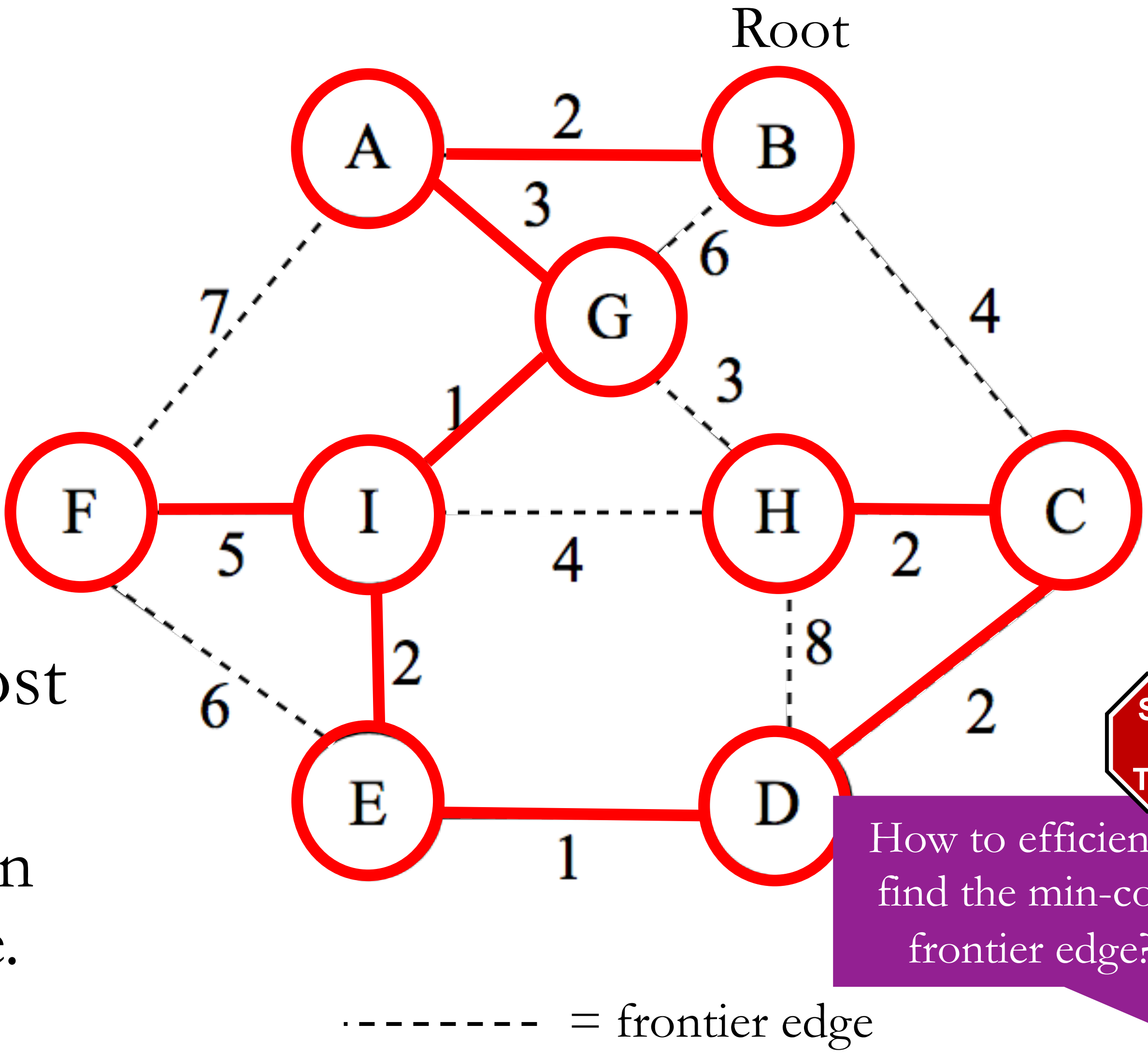


BFS:



Minimum Spanning Tree

- Find a spanning tree such that the sum of all edge costs is minimized.
- This minimizes the *total* communication cost for broadcast.
- ***Prim's algorithm***:
  - Start with a root node.
  - Repeatedly add the minimum-cost edge on the *frontier* of the tree.
    - A frontier edge connects a node in the tree to a node outside the tree.
- *See also:* Kruskal's algorithm



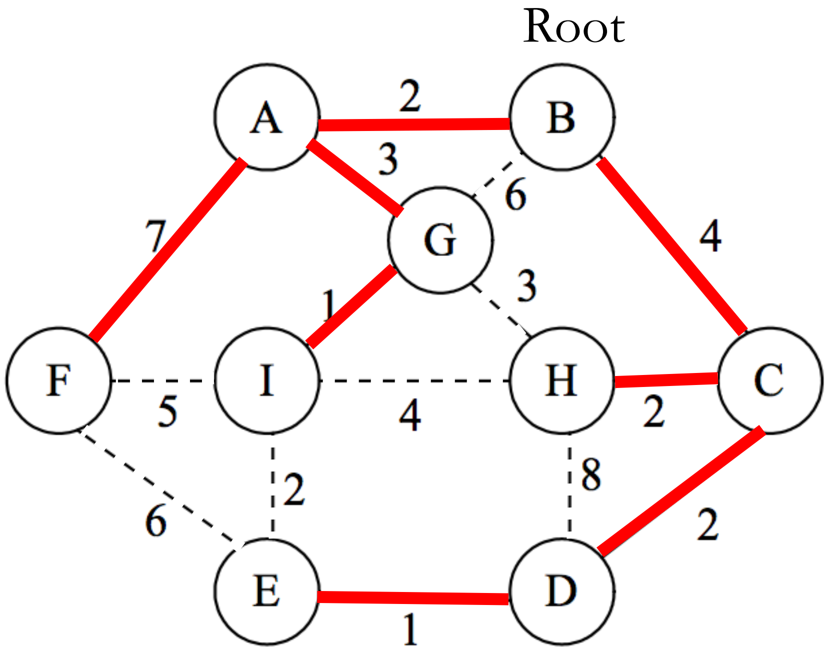
STOP and THINK

How to efficiently find the min-cost frontier edge?

Shortest Path Tree

Minimizes the **longest** path in the broadcast.

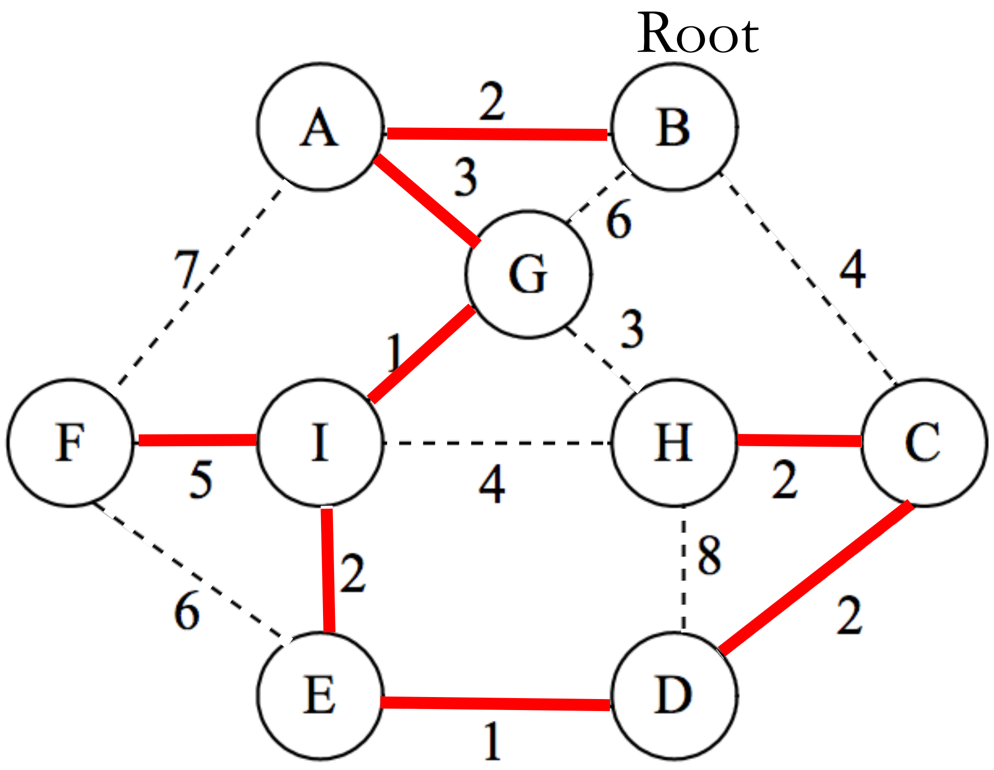
- A special spanning tree computed by Dijkstra's algorithm.
- Minimizes the path cost to every node
- Thus, minimizes the path cost to the farthest node from the root.
- If link costs represent latencies, shortest path tree minimizes the **broadcast latency**.
- *Not* a minimum spanning tree.
- *Total* edge "cost" may be greater, but this is irrelevant for parallel message latencies.



Tradeoff between total cost and longest path

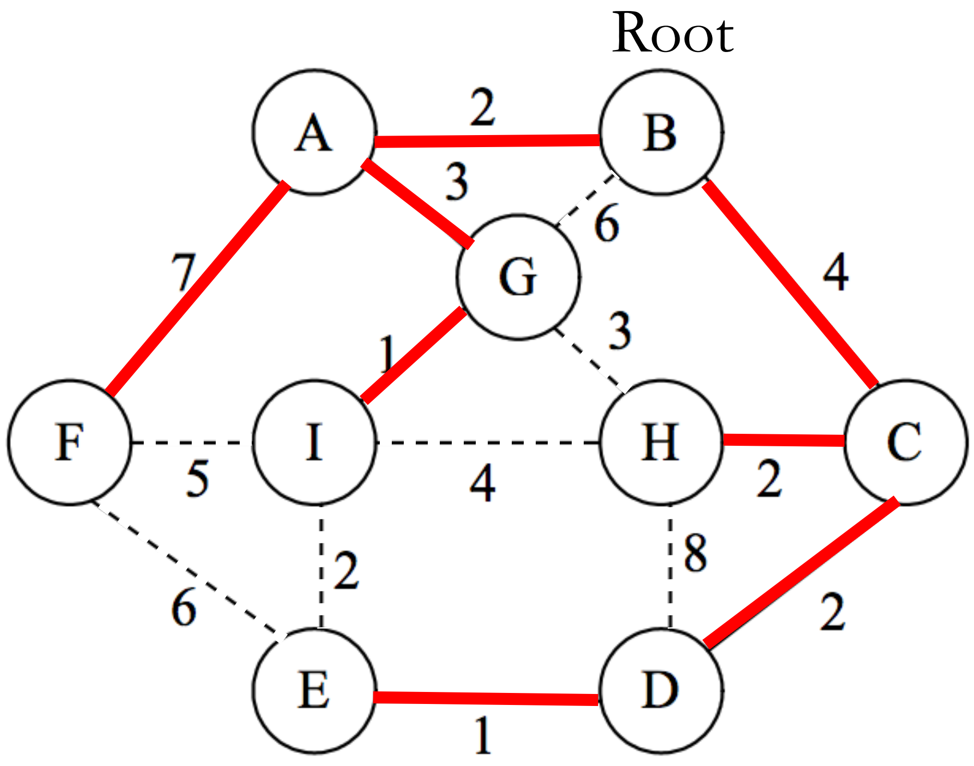
Minimum Spanning Tree (eg., Prim's alg.)

- Total cost: 18
- Longest path: 13
- Best choice if edge costs are cumulative (eg., dollars, energy).



Shortest Path Spanning Tree (Dijkstra)

- Total cost: 23
- Longest path: 9
- Best choice for broadcast if edge costs represent delays.



Broadcasting summary

Controlled flooding

- Nodes re-broadcast *new* messages.
- Must track recently-seen messages.
- Duplicate messages may be received.
- It's simple and it works!
- A totally **distributed** approach.

In the case of Internet routers, flooding is needed to share this full network knowledge.

Spanning tree

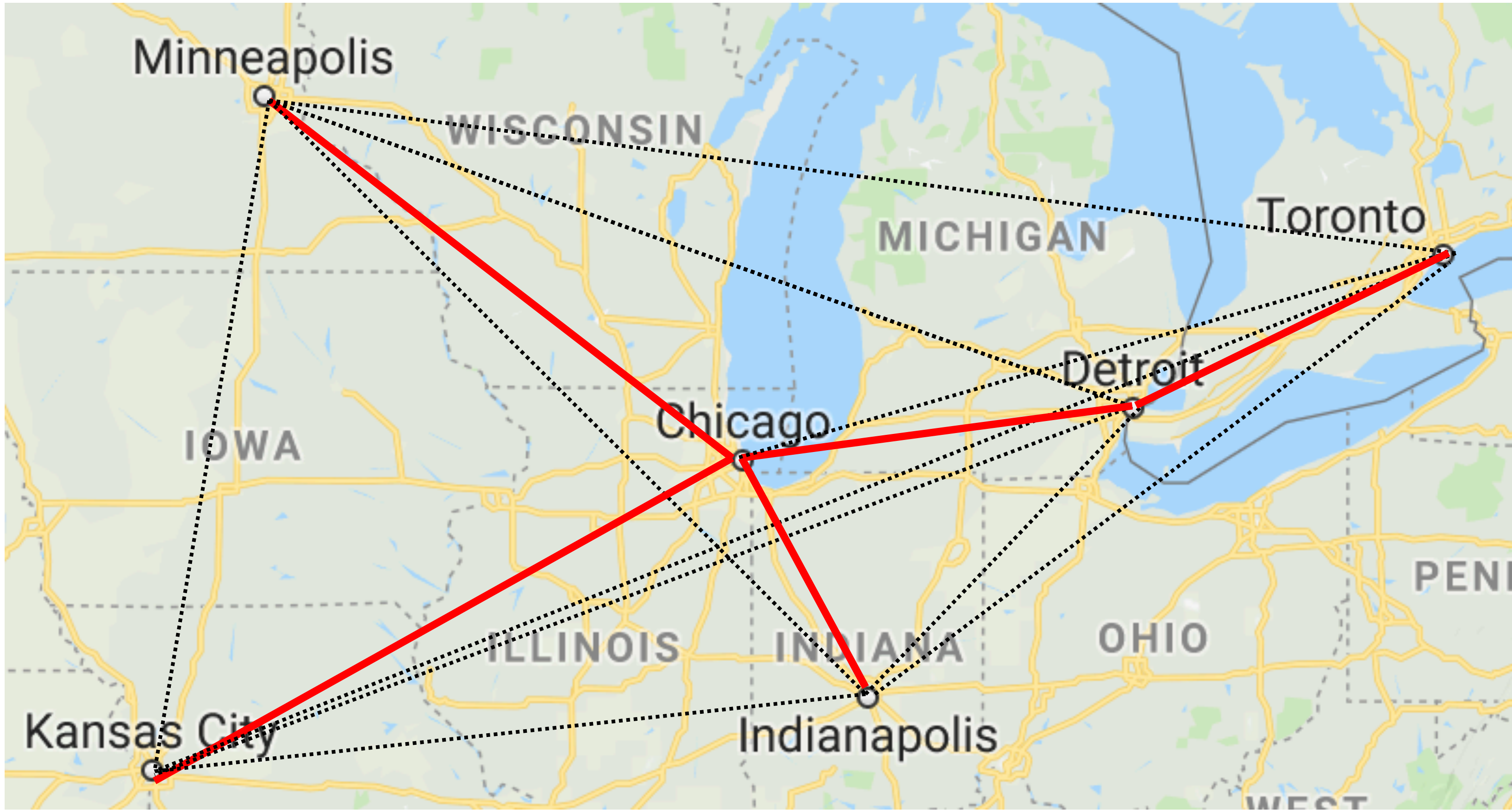
- Sends messages only along necessary paths.
- Nodes do not get duplicates.
- No need for message id's.
- Communication cost or delay can be minimized.
- Requires **centralized** knowledge of full network (to run opt. alg.)
- All nodes must agree on a spanning tree, so they must have the same information.

Network design problems

Would we ever want to build a **shortest path** tree network instead of an MST?

STOP and THINK

- MST can be used for other problems in networking (and elsewhere!)
- Eg., given a list of cities and their locations, what is the cheapest way to build a network that interconnects them?
- Cities are nodes and distance between all pairs of cities are weighted edges.
- MST is the cheapest connected graph (network) that can be built.



Recap

- BGP can be hacked to direct traffic for one IP address to two hosts.
- **Broadcasting** is sending a single message to every host.
  - **Multicast** is sending a single message to many (but not all) hosts.
- **Controlled flooding**: add a sequence number to messages, and retransmit only if you have not seen the received sequence number.
- **Spanning tree**: a graph without cycles that reaches all nodes. Broadcast can be done by transmitting along a spanning tree.
  - **Prim's algorithm** constructs a minimum-cost spanning tree
  - **Dijkstra's algorithm** constructs a shortest-path-from-root spanning tree