CS4450

Computer Networks:
Architecture and Protocols

Lecture 9
Spanning Tree Protocol
Why Network Layer?

Rachit Agarwal

CORNELL UNIVERSITY FOUNDED A.D. 1865

Goals for Today's Lecture

- Finish **Spanning Tree Protocol**

- **Why** do we need network layer?
    - **Why** not just use switched Ethernet across the Internet?

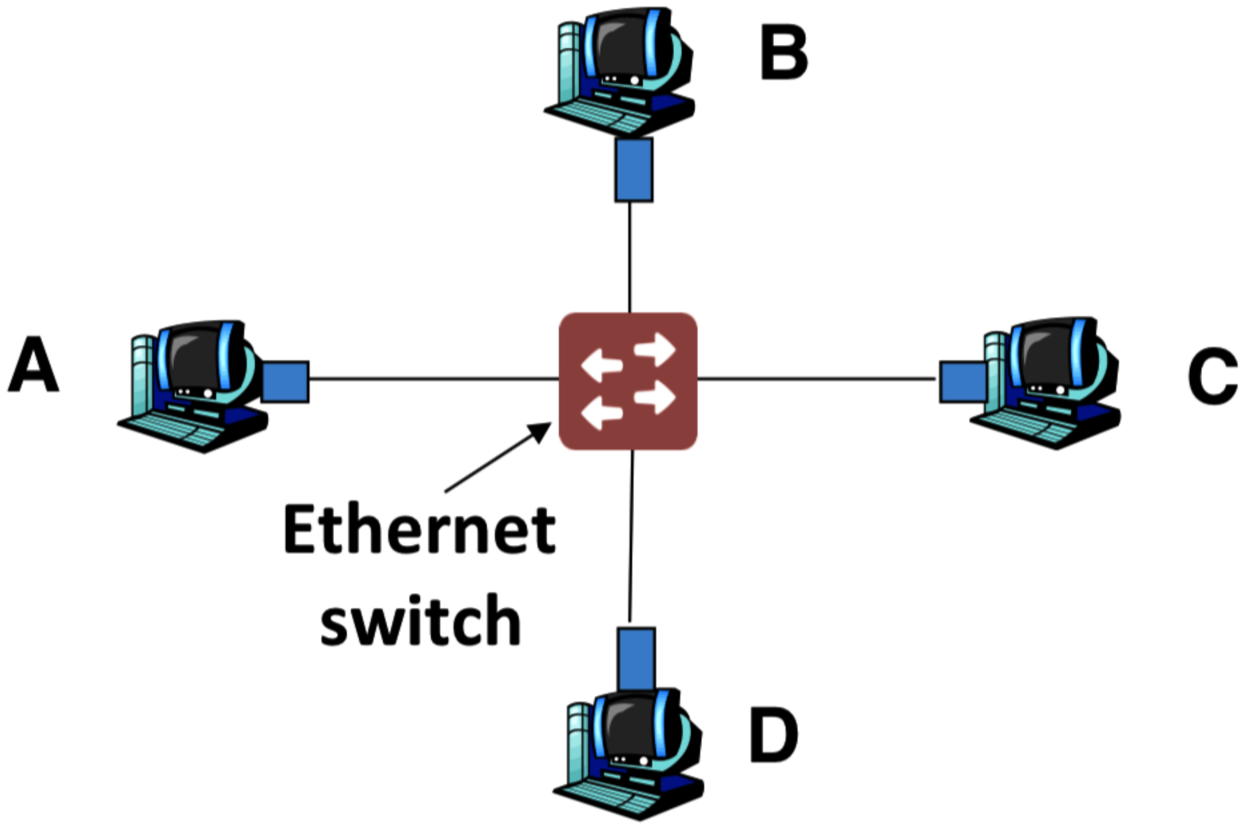
Recap of Link Layer so far

Recap: Link layer

- **Traditional Link Layer: Broadcast Ethernet**
  - CSMA/CD
    - Random access on a broadcast channel
    - Exponential Backoff
  - Why Frames?
    - To incorporate sentinel bits for identifying frame start/end
    - To incorporate link layer source and destination names
    - To incorporate CRC for checking correctness of received frames

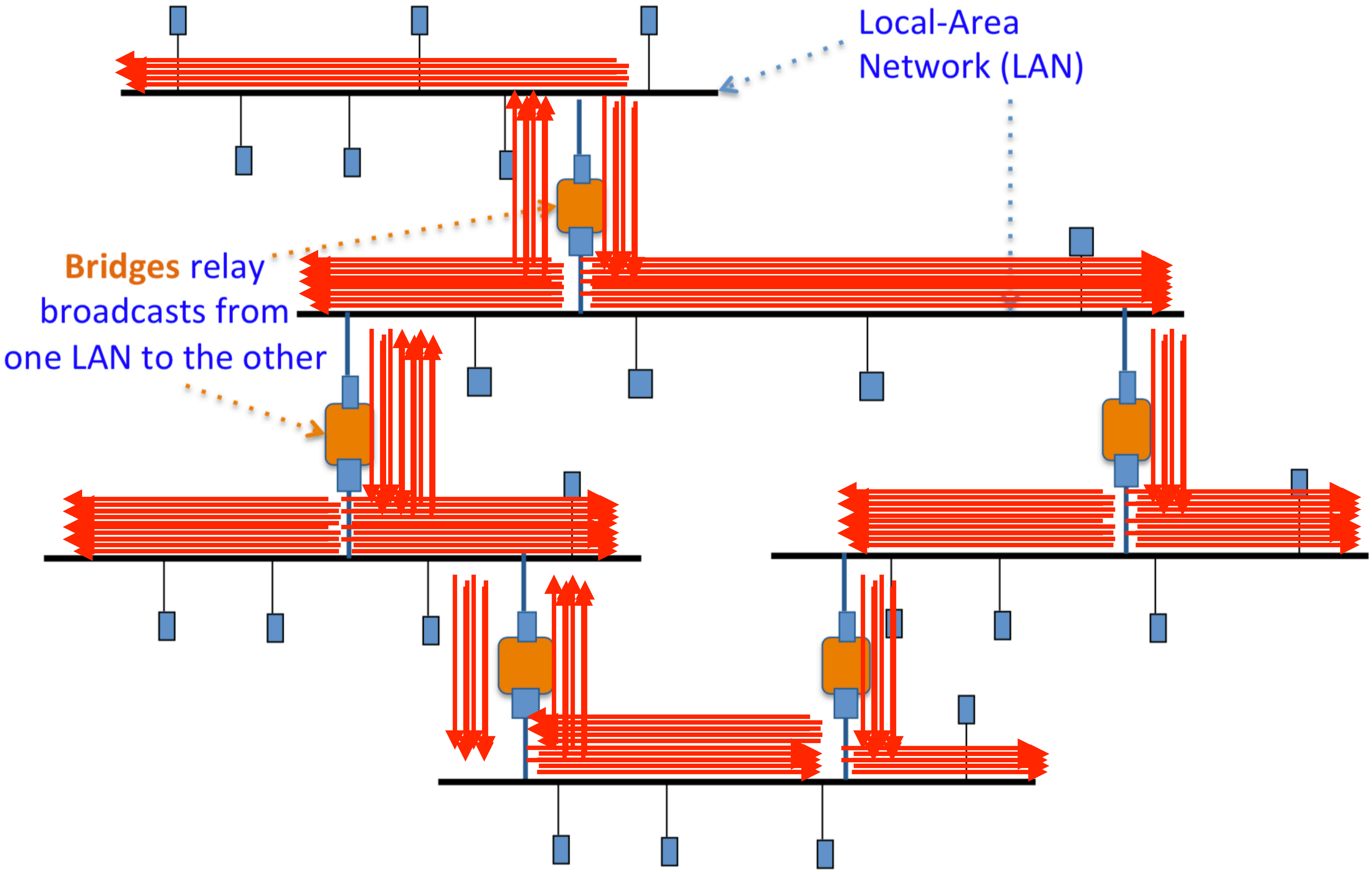
- **Modern Link Layer: Switched Ethernet**
  - Why?
    - Scalability limits of traditional Ethernet
    - Why?
      - Detecting collisions on a broadcast channel

Recap: Switched Ethernet



- Enables concurrent communication
    - Host A can talk to C, while B talks to D
    - No collisions -> no need for CSMA, CD
    - No constraints on link lengths or frame size

Recap: Broadcast storm



Local-Area Network (LAN)

Bridges relay broadcasts from one LAN to the other

Recap: Avoiding Broadcast Storm Problem

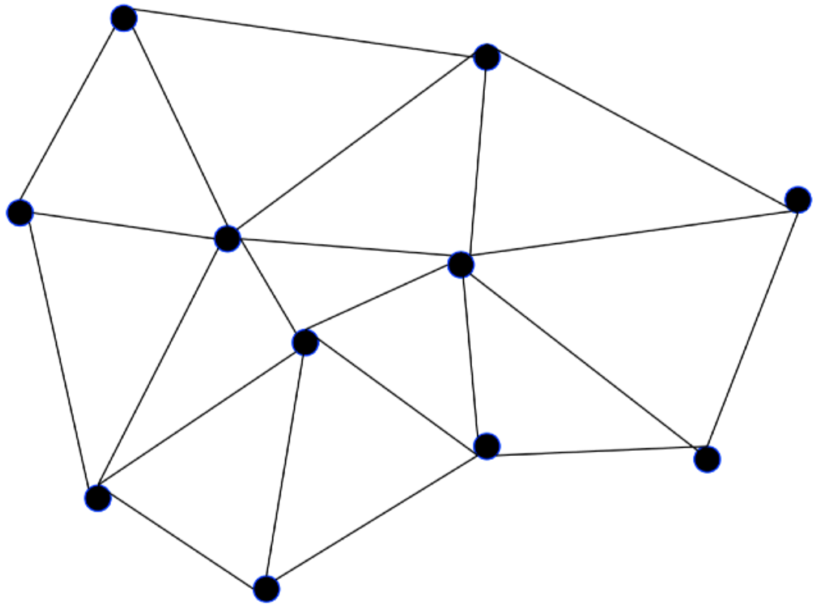
Getting rid of loops

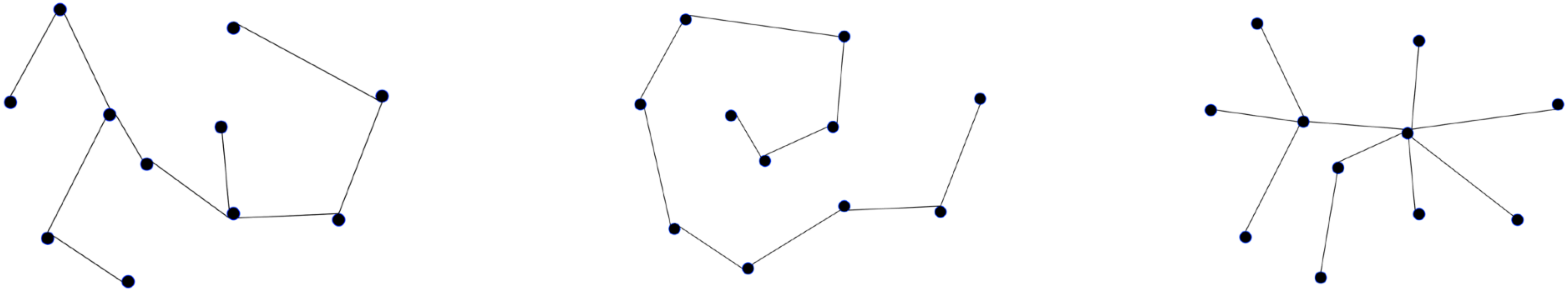
using

Spanning Tree Protocol

Recap: Spanning Tree definition

- **Subgraph that includes all vertices but contains no cycles**
    - Links not in the spanning tree are not used in forwarding frames





Questions?

Algorithm has Two Aspects…

- Pick a root:
    - Pick the one with the smallest identifier (MAC name/address)

- Compute the shortest paths to the root
    - Only keep the links on the shortest path
    - Break ties in some way
        - so we only keep one shortest path from each node
        - **One approach:** Choose the path via neighbor switch with the smallest identifier

- Ethernet’s spanning tree construction does both with a single algorithm

Steps in Spanning Tree Protocol (assuming all delays = 1)

- **Messages (Y,d,X)**
    - Proposing root Y; from node X; advertising a distance d to Y

- Initially each switch proposes itself as the root
    - that is, switch X announces (X,0,X) to its neighbors

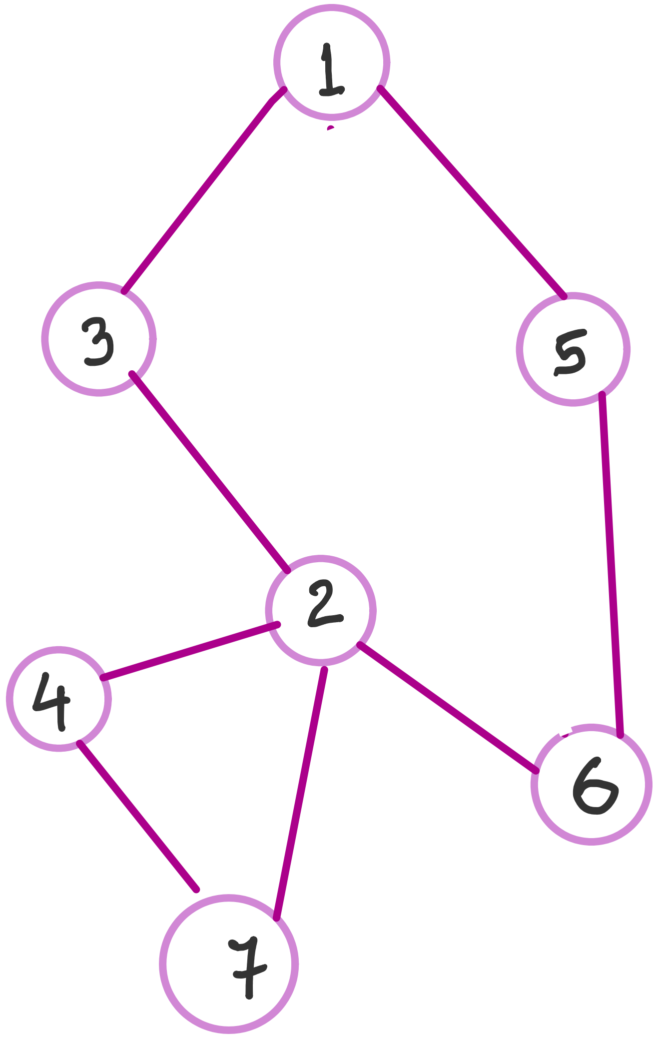
- Switches update their view; each switch Z:
    - Upon receiving message (Y,d,X) from X, check Y's id
    - If Y's id < current root: set root = Y
    - Set next-hop = X

- Switches compute their distance from the root; each switch Z:
    - Shortest distance to root = d + distance from X = d + 1

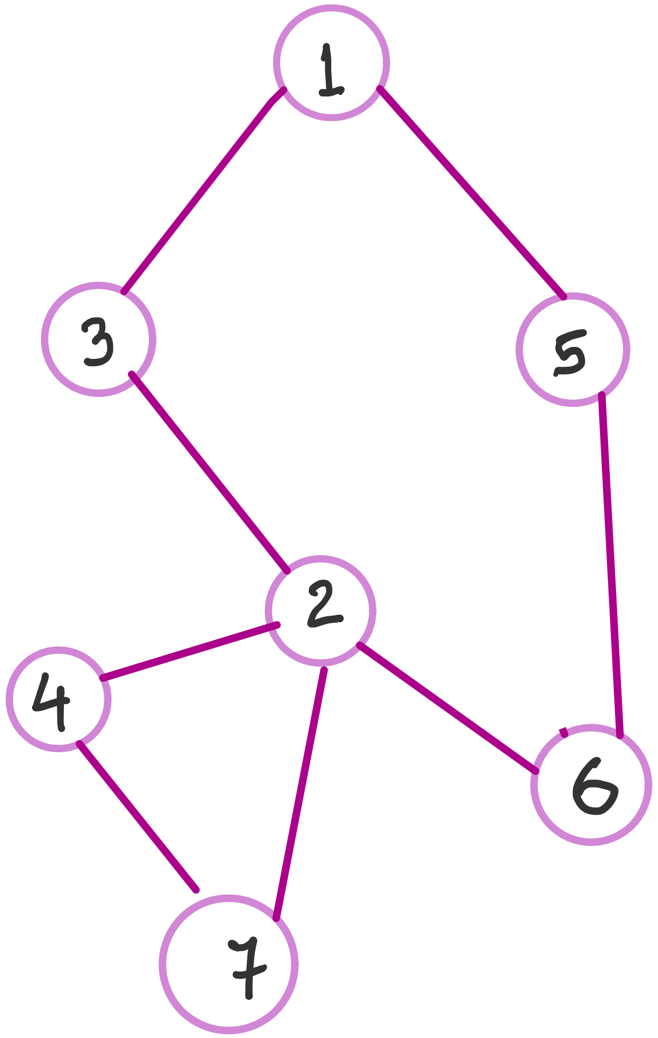
- If **root changed OR shortest distance to the root changed:**
    - **Switch Z** sends neighbors updated message (Y, d+1, Z)

Lets run the Spanning Tree Protocol on this example

(assume all links have “distance” 1)

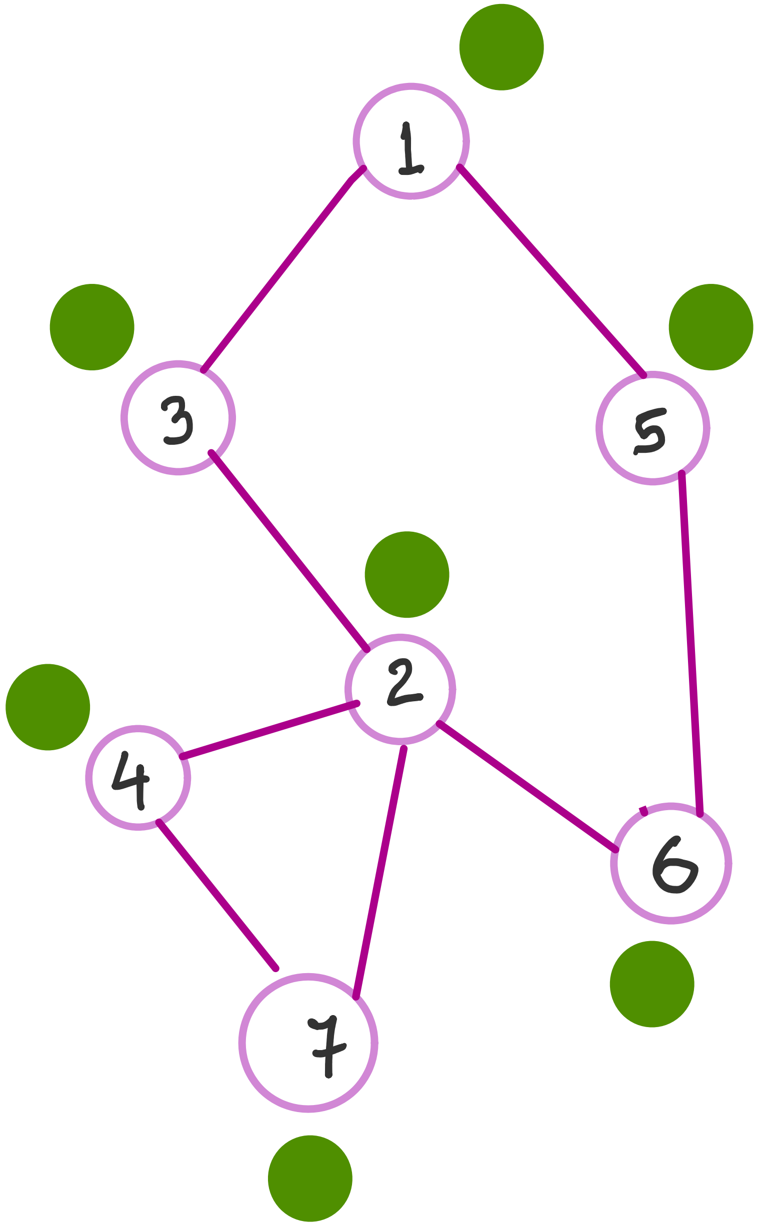


Round 1



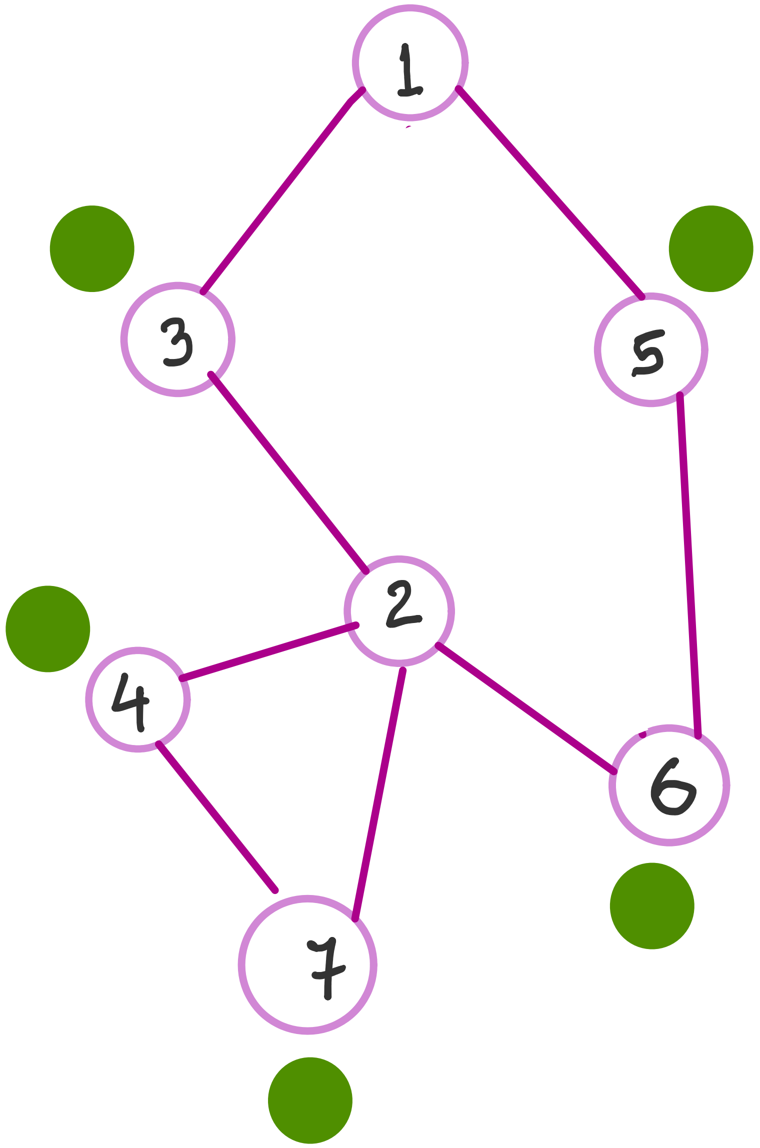
	Receive	Send	Next-hop
1		(1, 0, 1)	1
2		(2, 0, 2)	2
3		(3, 0, 3)	3
4		(4, 0, 4)	4
5		(5, 0, 5)	5
6		(6, 0, 6)	6
7		(7, 0, 7)	7

Round 2



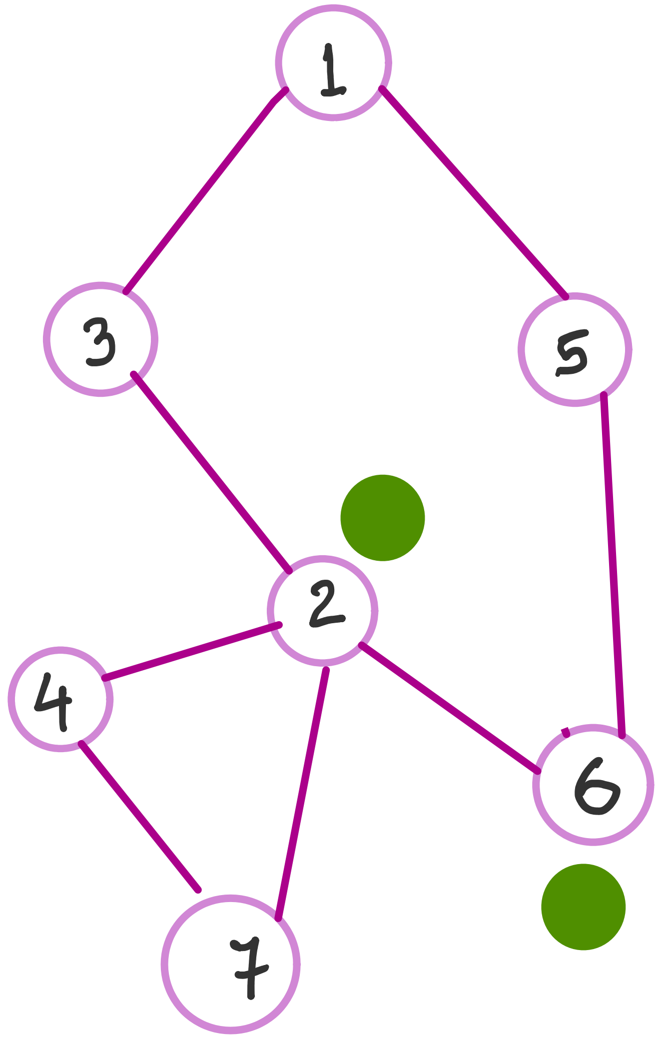
	Receive	Send	Next hop
1 (1, 0, 1)	(3, 0, 3), (5, 0, 5)		1
2 (2, 0, 2)	(3, 0, 3), (4, 0, 4), (6, 0, 6), (7, 0, 7)		2
3 (3, 0, 3)	(1, 0, 1), (2, 0, 2)	**(1, 1, 3)**	**1**
4 (4, 0, 4)	(2, 0, 2), (7, 0, 7)	**(2, 1, 4)**	**2**
5 (5, 0, 5)	(1, 0, 1), (6, 0, 6)	**(1, 1, 5)**	**1**
6 (6, 0, 6)	(2, 0, 2), (5, 0, 5)	**(2, 1, 6)**	**2**
7 (7, 0, 7)	(2, 0, 2), (4, 0, 4)	**(2, 1, 7)**	**2**

Round 3



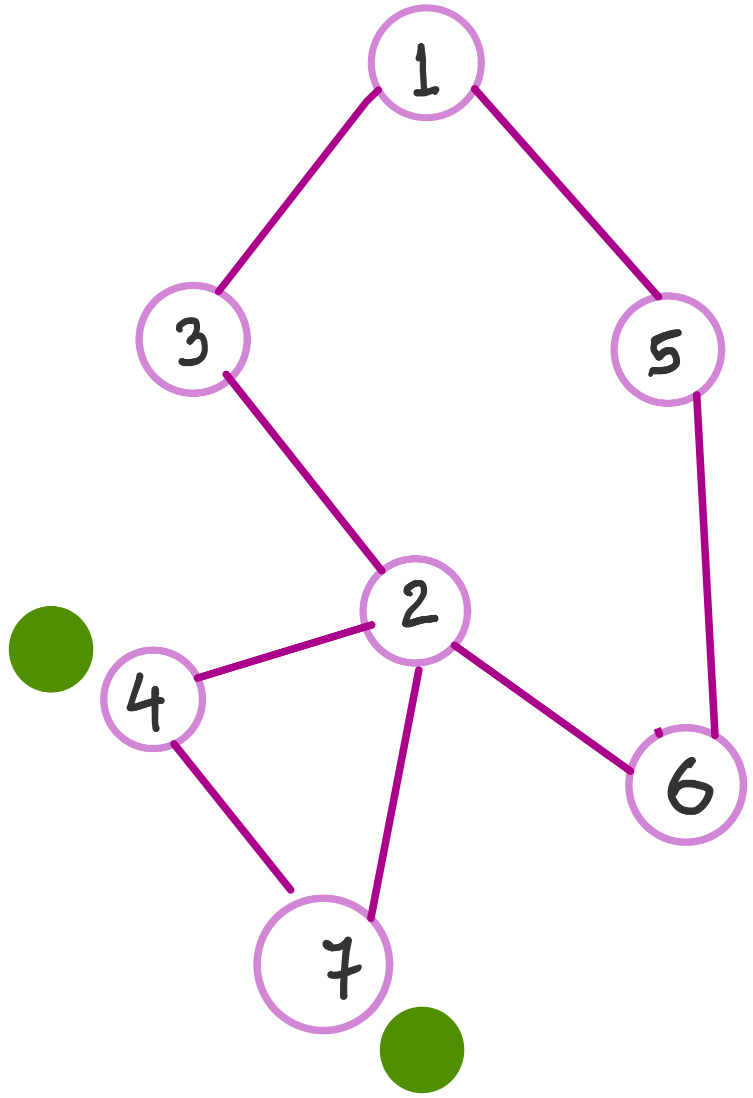
	Receive	Send	Next hop
1	(1, 1, 3), (1, 1, 5)		1
2	(1, 1, 3), (2, 1, 4), (2, 1, 6), (2, 1, 7)	**(1, 2, 2)**	**3**
3 (1, 1, 3)			1
4 (2, 1, 4)	(2, 1, 7)		2
5 (1, 1, 5)	(2, 1, 6)		1
6 (2, 1, 6)	(1, 1, 5)	**(1, 2, 5)**	**5**
7 (2, 1, 7)	(2, 1, 4)		2

Round 4



	Receive	Send	Next hop
1			1
2 (1, 2, 2)	(1, 2, 6)		3
3	(1, 2, 2)		1
4	(1, 2, 2)	**(1, 3, 4)**	**2**
5	(1, 2, 6)		1
6 (1, 2, 6)	(1, 2, 2)		5
7	(1, 2, 2)	**(1, 3, 7)**	**2**

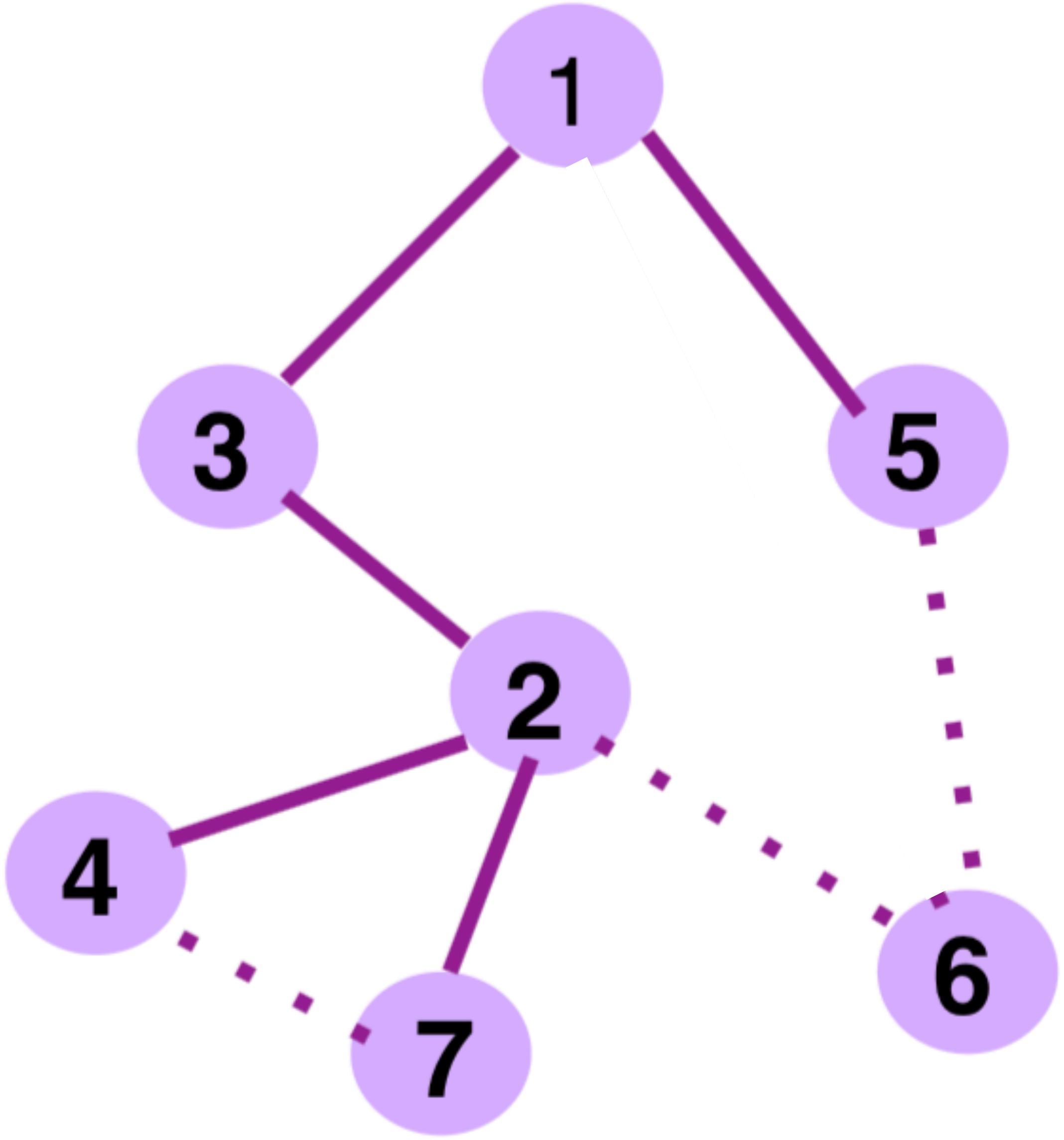
Round 5



	Receive	Send	Next hop
1			1
2	(1, 3, 4), (1, 3, 7)		3
3			1
4 (1, 3, 4)	(1, 3, 7)		2
5			1
6			1
7 (1, 3, 7)	(1, 3, 4)		2

After Round 5: We have our Spanning Tree

- 3-1
- 5-1
- 6-1
- 2-3
- 4-2
- 7-2



Questions?

Spanning Tree Protocol (++ Incorporating distances)

- **Messages (Y,d,X)**
  - Proposing root Y; from node X; advertising a distance d to Y

- Initially each switch proposes itself as the root
  - that is, switch X announces (X,0,X) to its neighbors

- Switches update their view; each switch Z:
  - Upon receiving message (Y,d,X) from X, check Y’s id
  - If Y’s id < current root: set root = Y
  - Set next-hop = X

- Switches compute their distance from the root; each switch Z:
  - Shortest distance to root = d + distanceTo(X)

- If **root changed OR shortest distance to the root changed:**
  - switch Z sends neighbors updated message (Y, d+distanceTo(X), Z)

Spanning Tree Protocol ++ (incorporating failures)

- Protocol must react to failures
    - Failure of the root node
    - Failure of switches and links

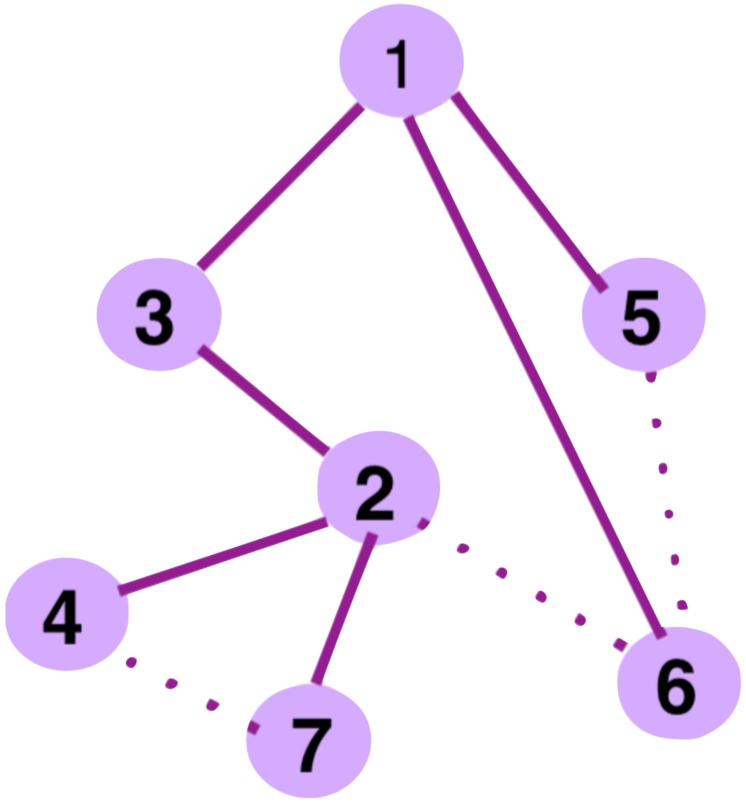
- **Root node sends periodic announcement messages**
    - Few possible implementations, but this is simple to understand
    - Other switches continue forwarding messages

- Detecting failures through timeout (soft state)
    - If no word from root, time out and send a (Y, 0, Y) message to all neighbors (in the graph)!

- **If multiple messages with a new root received, send message (Y, d, X) to the neighbor sending the message**

Example: Suppose link 2-4 fails

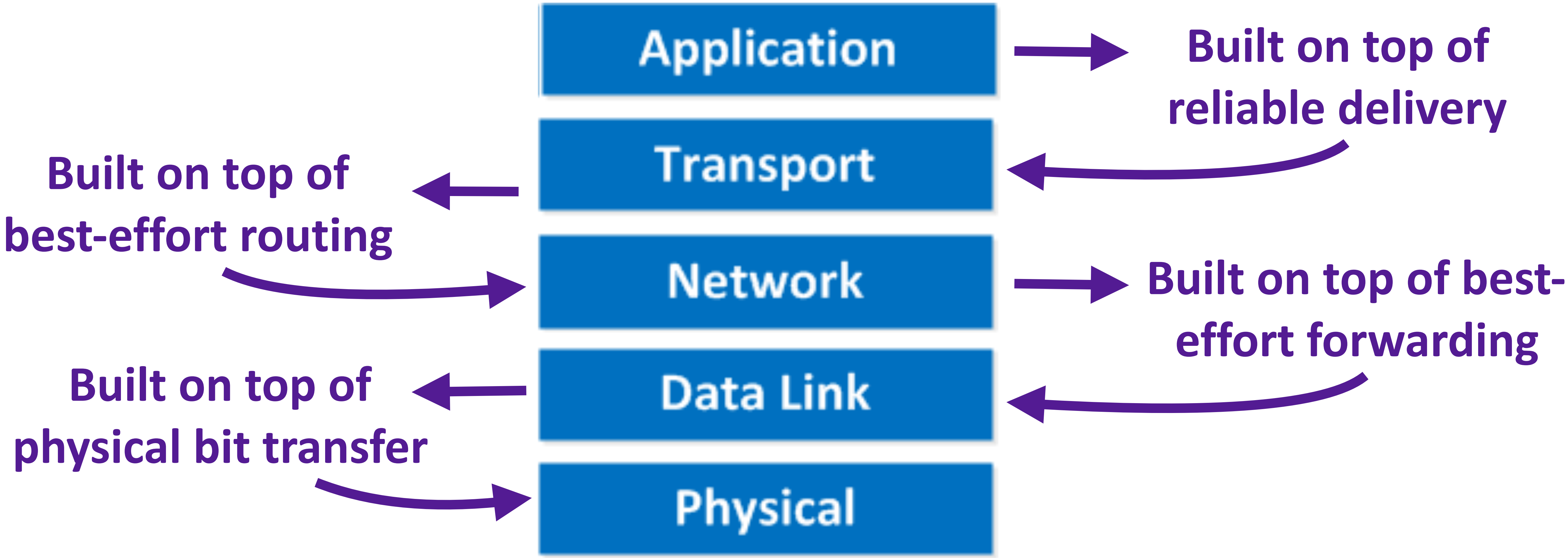
- 4 will send (4, 0, 4) to all its neighbors
    - 4 will stop receiving announcement messages from the root
    - Why?
- At some point, 7 will respond with (1, 3, 7)
- 4 will now update to (1, 4, 4) and send update message
- New spanning tree!



Questions?

The end of Link Layer ….

And the beginning of network layer :-D



Application

Transport

Network

Data Link

Physical

Built on top of reliable delivery

Built on top of best-effort routing

Built on top of best-effort forwarding

Built on top of physical bit transfer

Why do we need a network layer?

- Why not just use spanning trees across the entire network?

- Easy to design routing algorithms for (spanning) trees
    - **Nodes can "flood" packet to all other nodes**

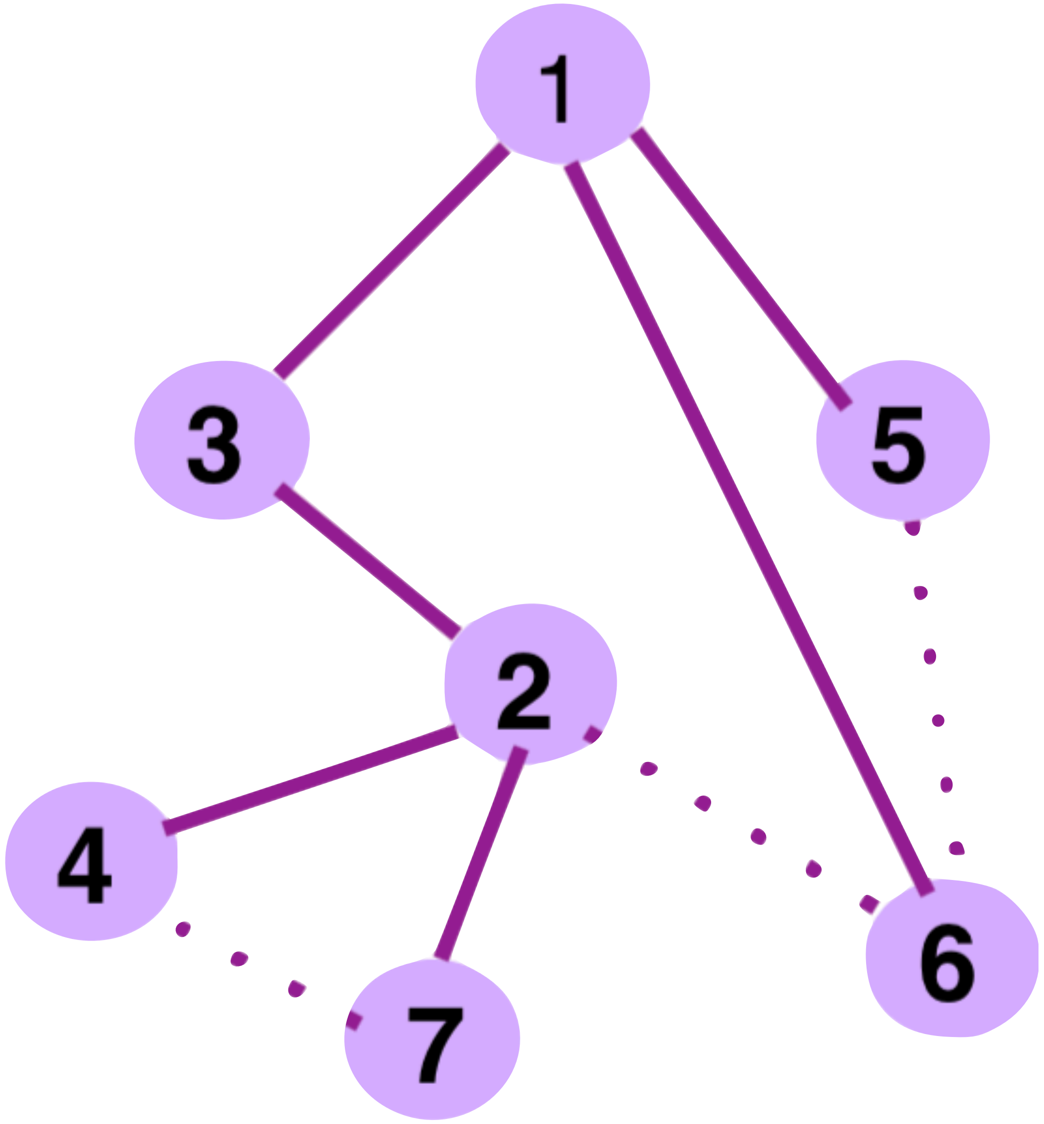
Flooding on a Spanning Tree

- Sends packet to ***every*** node in the network

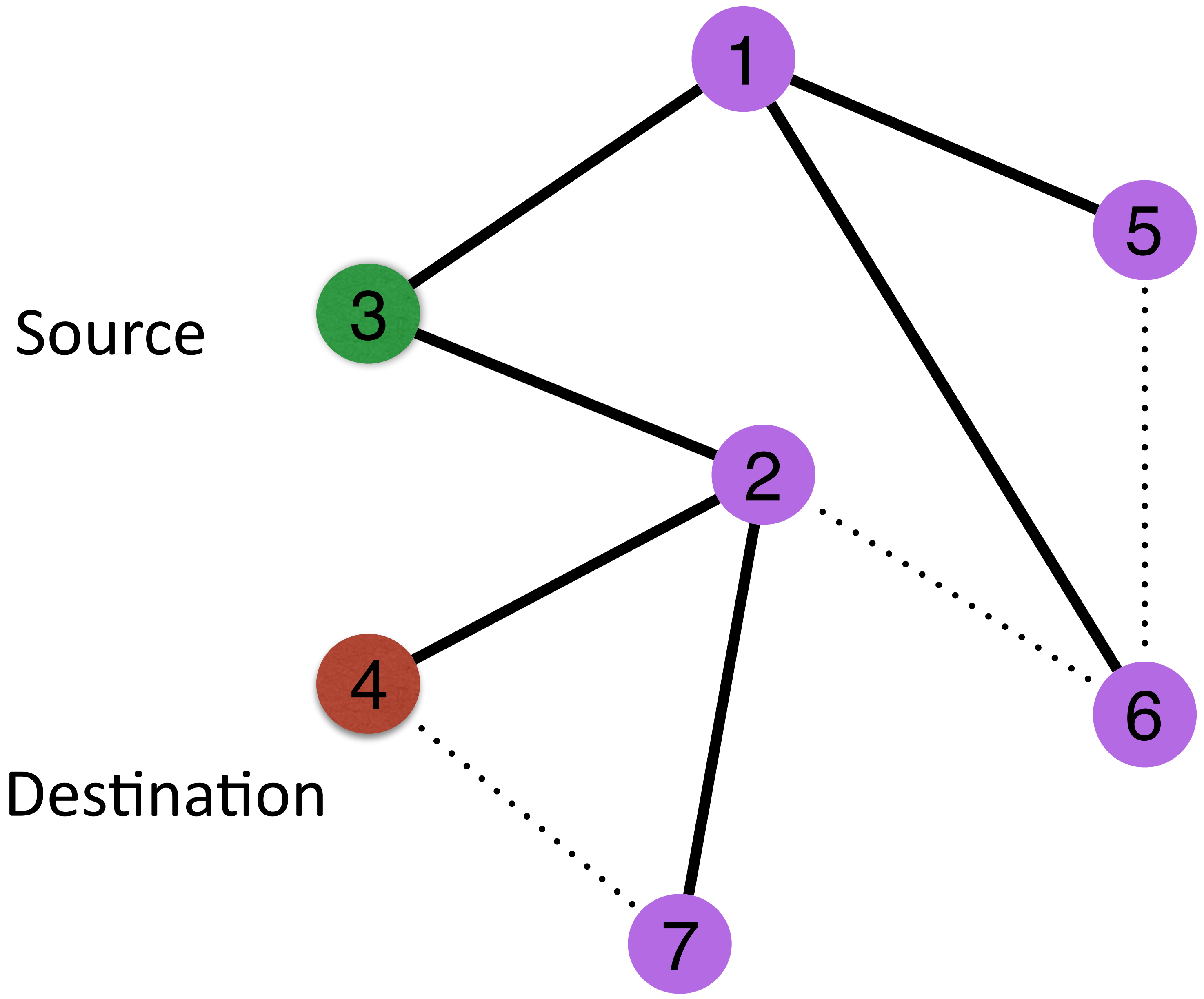
- **Step 1**: Ignore the links not belonging to the Spanning Tree

- **Step 2**: Originating node sends “flood” packet out every link (on spanning tree)

- **Step 3**: Send incoming packet out to all links **other than the one that sent the packet**

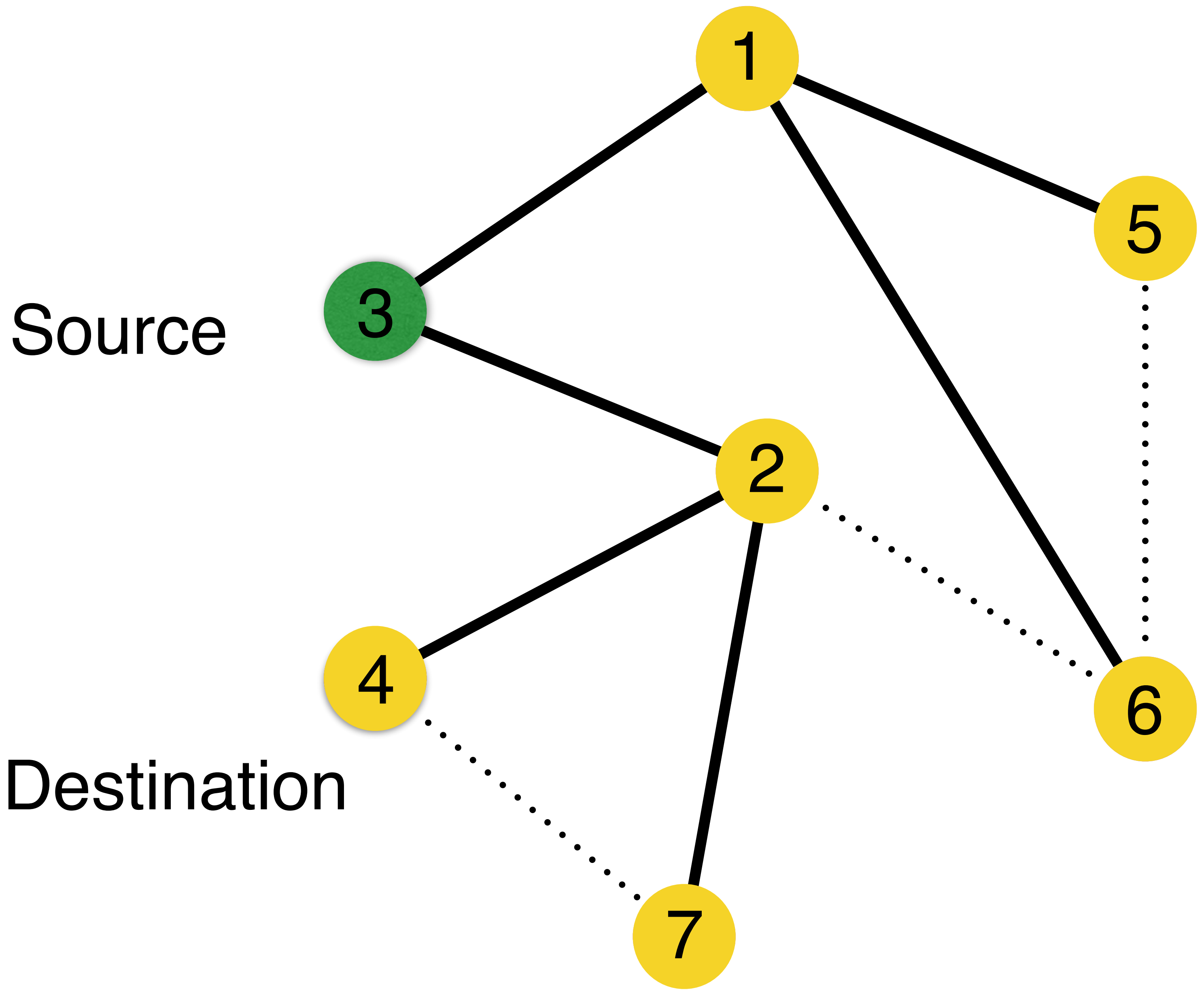


Flooding Example



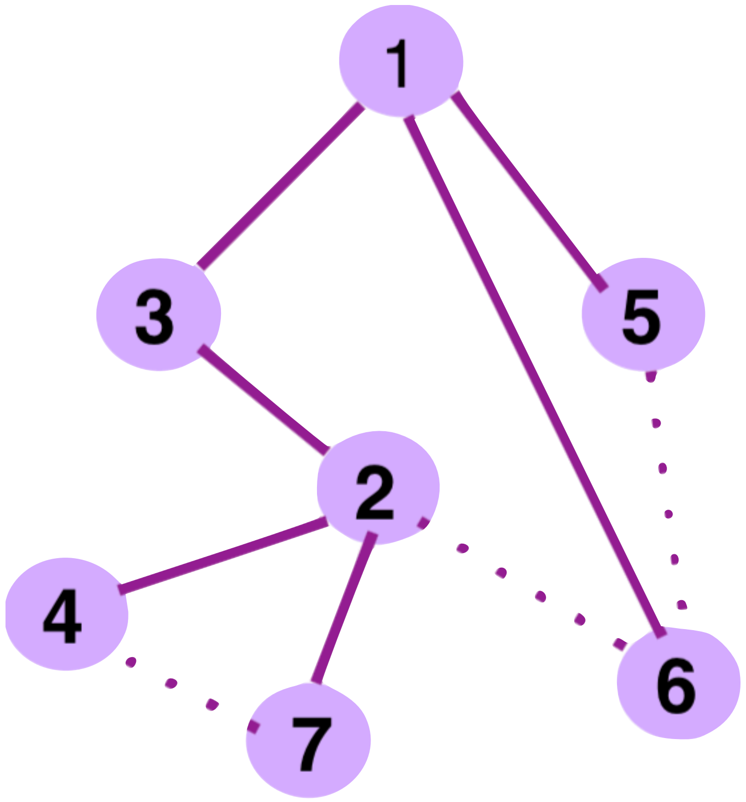
Flooding Example

Eventually all nodes are covered



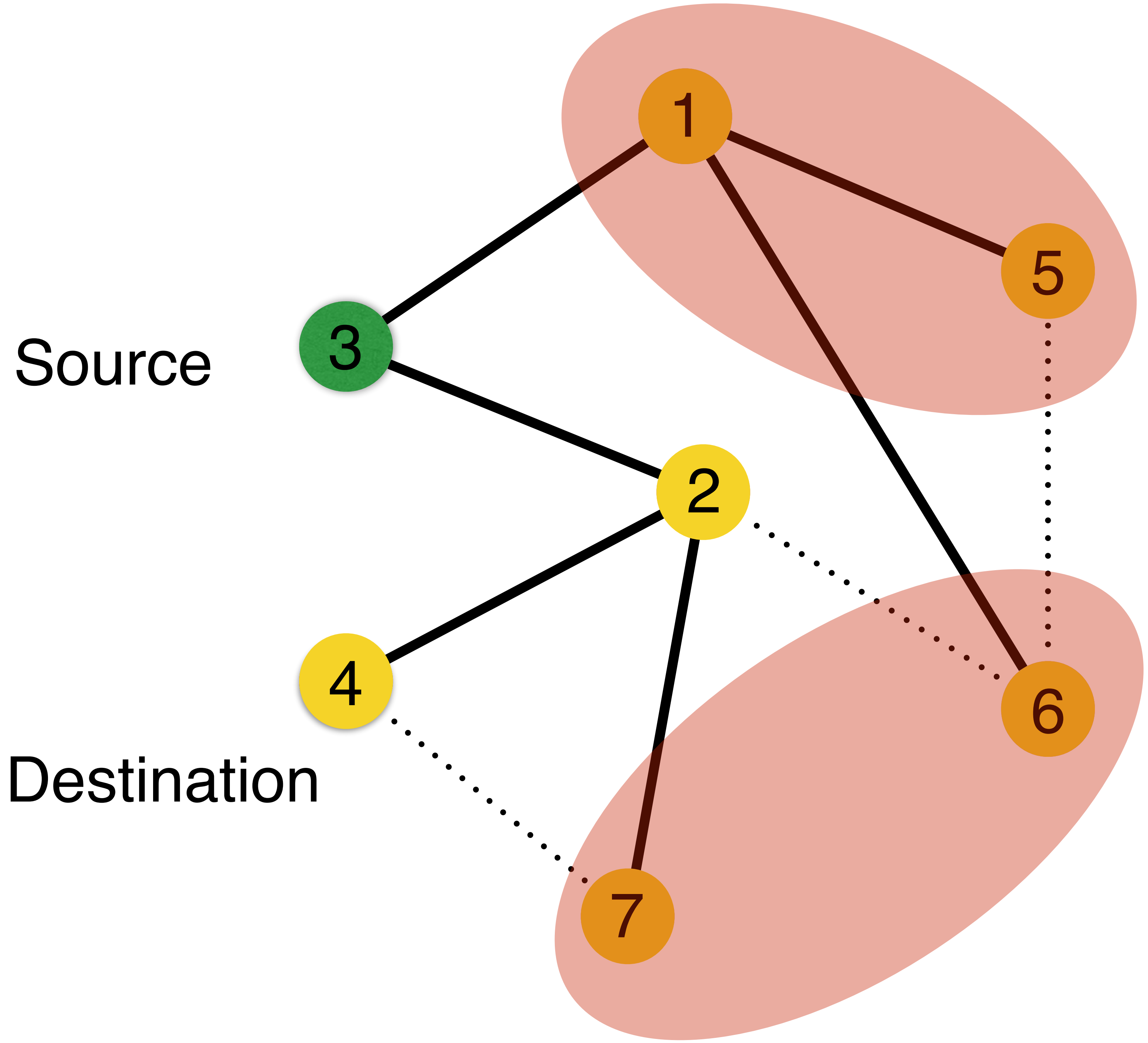
One copy of packet delivered to destination

Routing via Flooding on Spanning Tree …



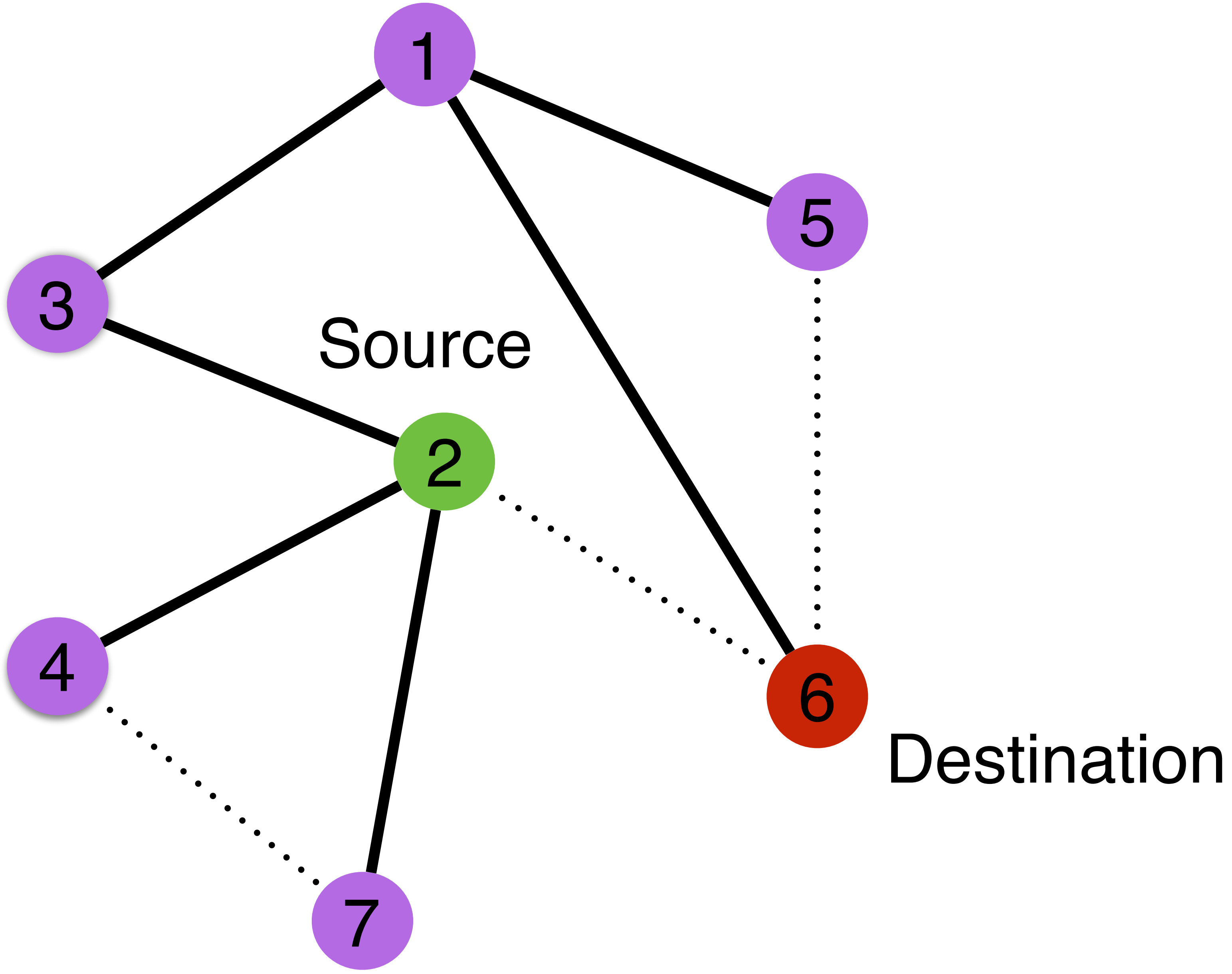
- There's only one path from source to destination
- How do you find that path? Ideas?
- Easy to design routing algorithms for trees
    - **Nodes can "flood" packet to all other nodes**
- Amazing properties:
    - No routing tables needed!
    - No packets will ever loop.
    - At least (and exactly) one packet must reach the destination
        - Assuming no failures

Three fundamental issues!



Issue 1: Each host has to do unnecessary packet processing!
(to decide whether the packet is destined to the host)

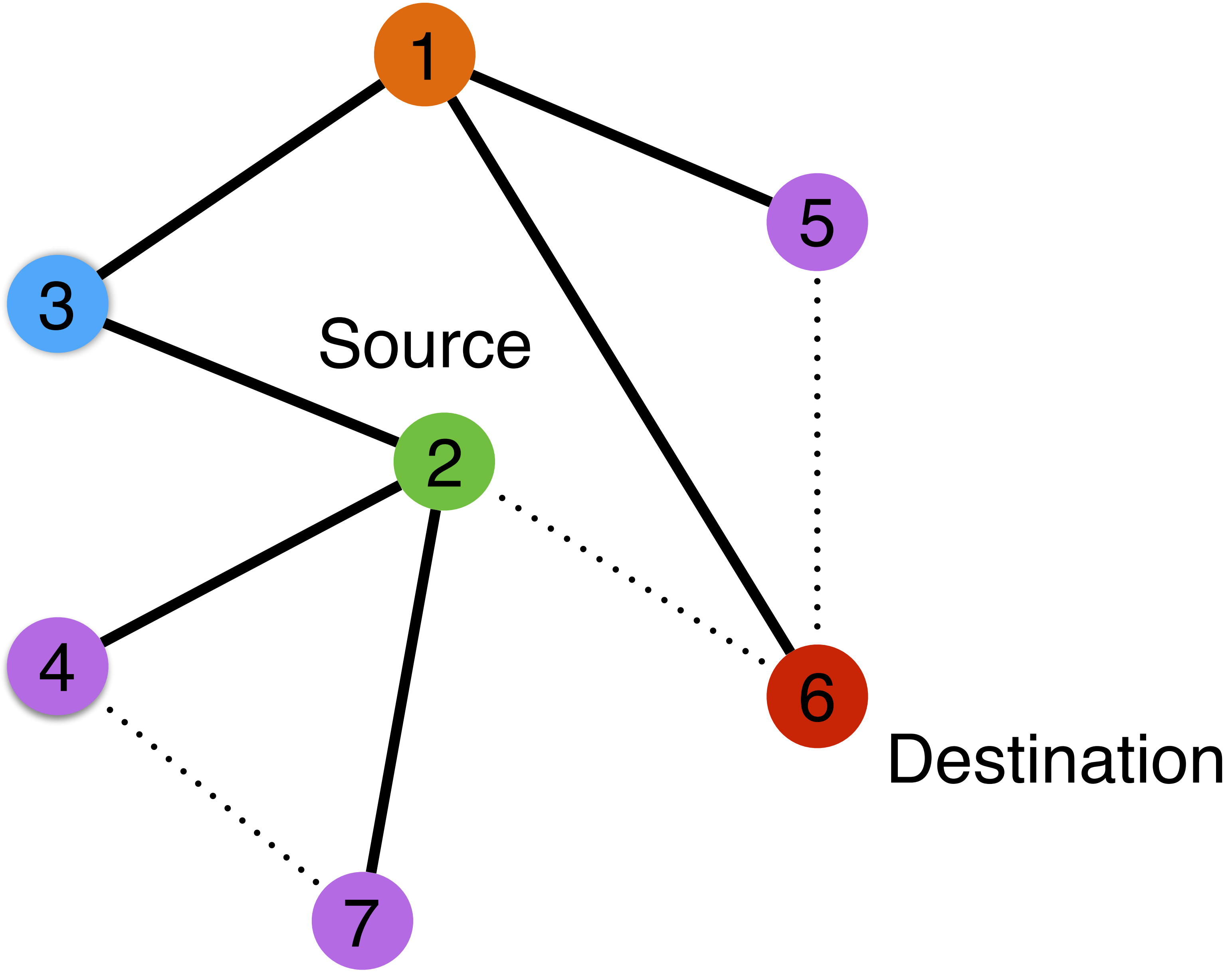
Three fundamental issues!



Issue 2: Higher latency!

(The packets unnecessarily traverse much longer paths)

Three fundamental issues!



Issue 3: Lower bandwidth availability!
(2-6 and 3-1 packets unnecessarily have to share bandwidth)

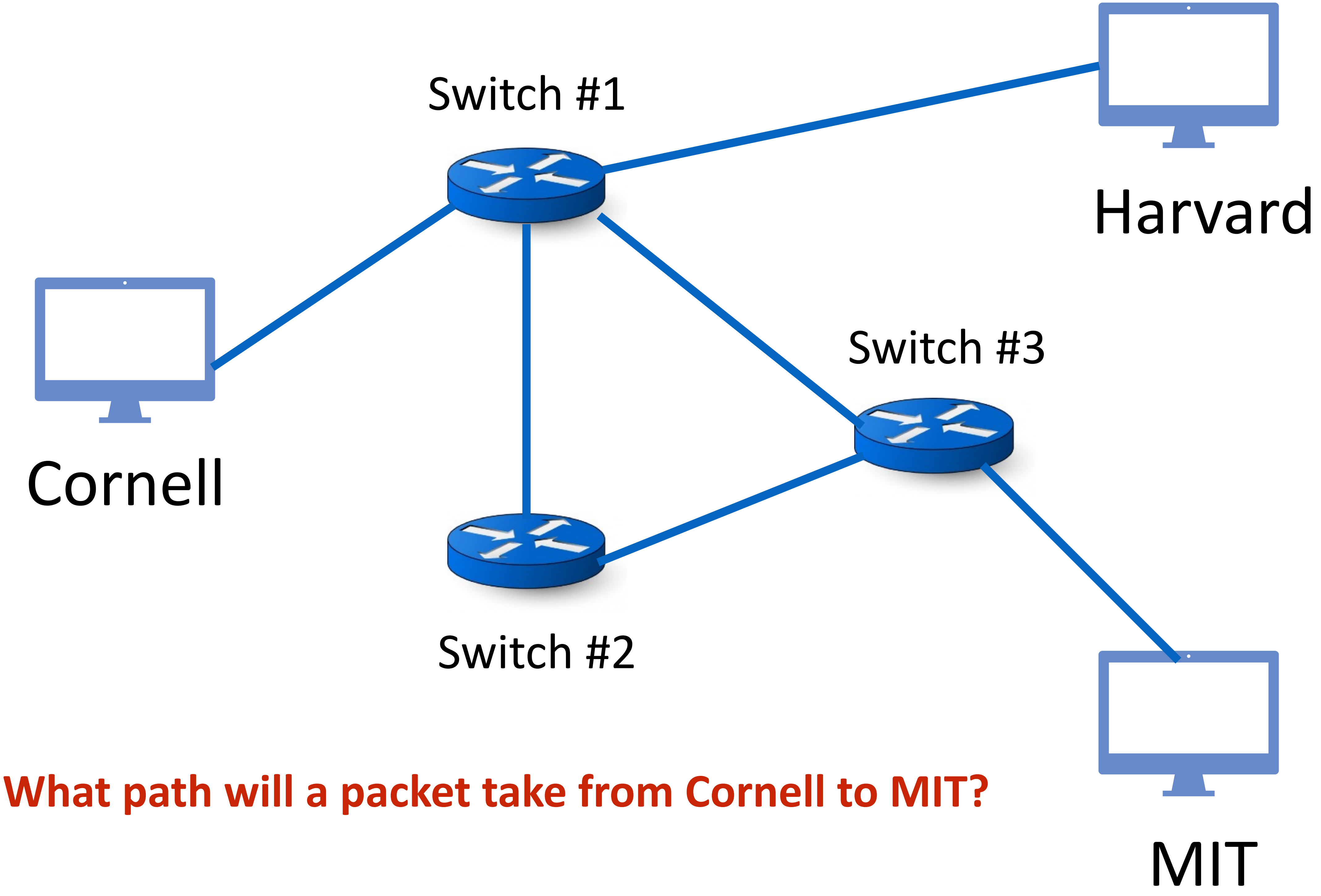
Questions?

Why do we need a network layer?

- Network layer performs “routing” of packets to alleviate these issues
- Uses routing tables
- Lets understand routing tables first

Routing Packets via Routing Tables

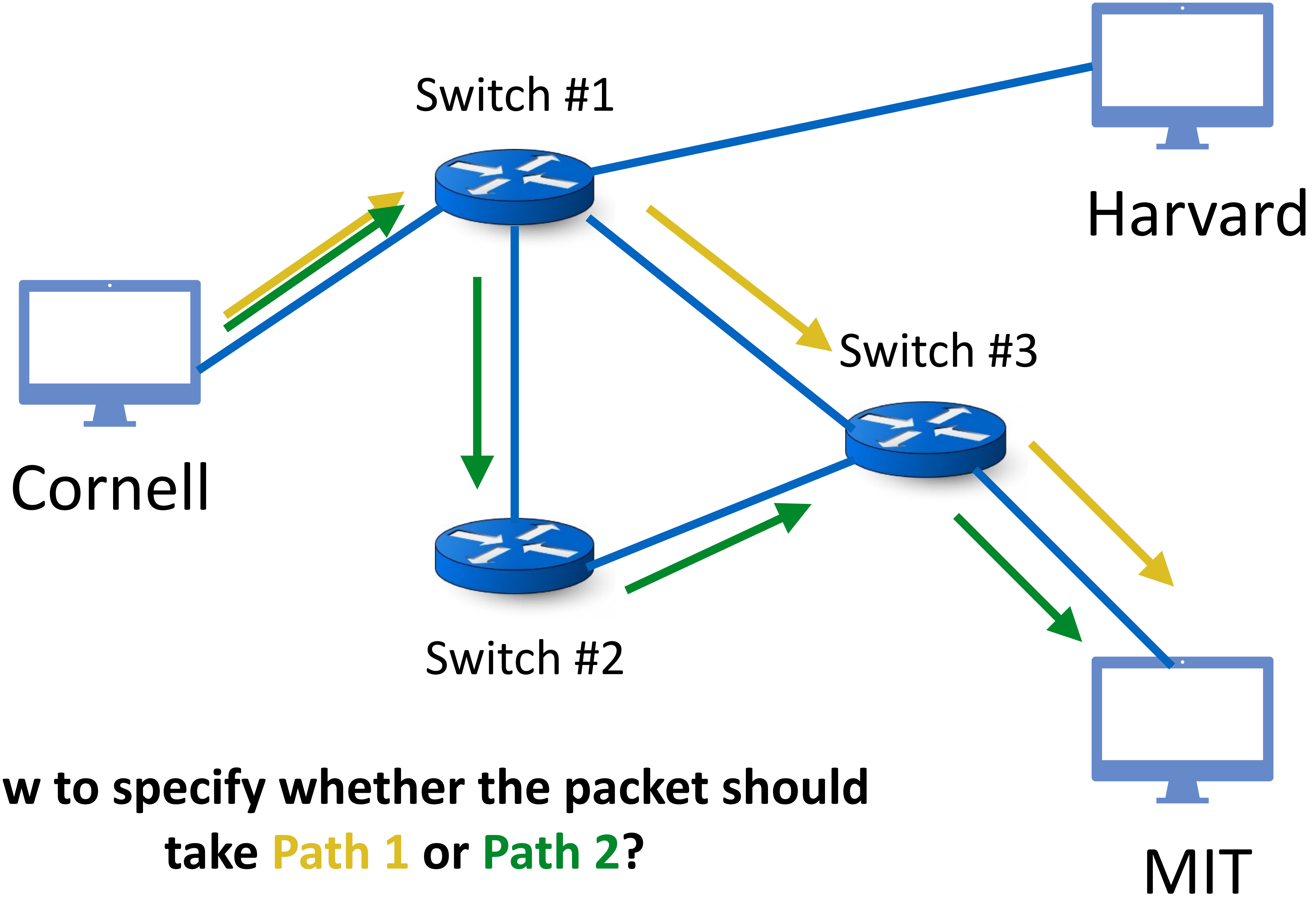
- **Routing tables allow finding path from source to destination**



What path will a packet take from Cornell to MIT?

Routing Packets via Routing Tables

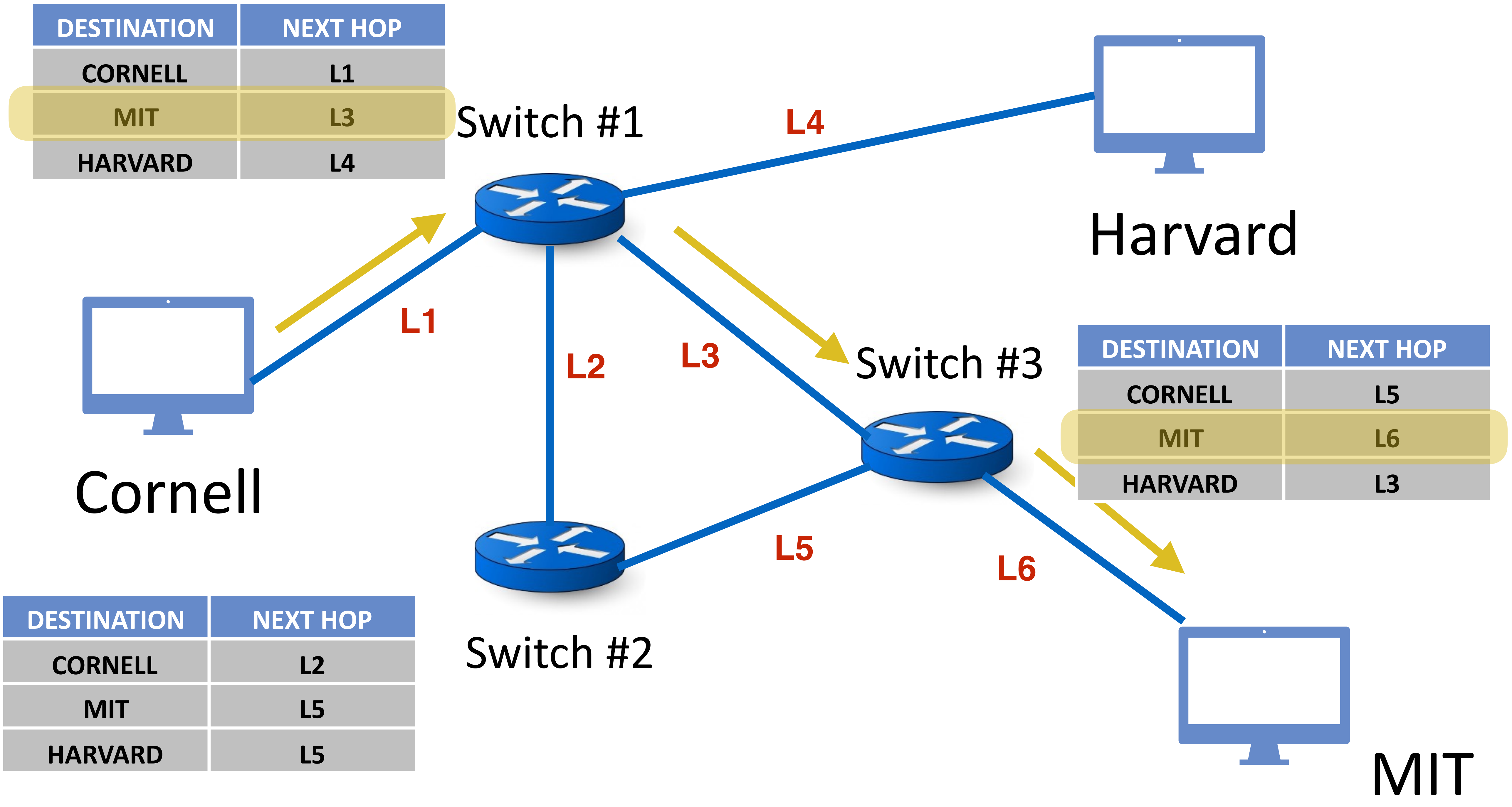
- **Finding path for a packet from source to destination**



How to specify whether the packet should take Path 1 or Path 2?

Routing Table

- **Suppose packet follows Path 1: Cornell - S#1 - S#3 - MIT**



DESTINATION	NEXT HOP
CORNELL	L1
MIT	L3
HARVARD	L4

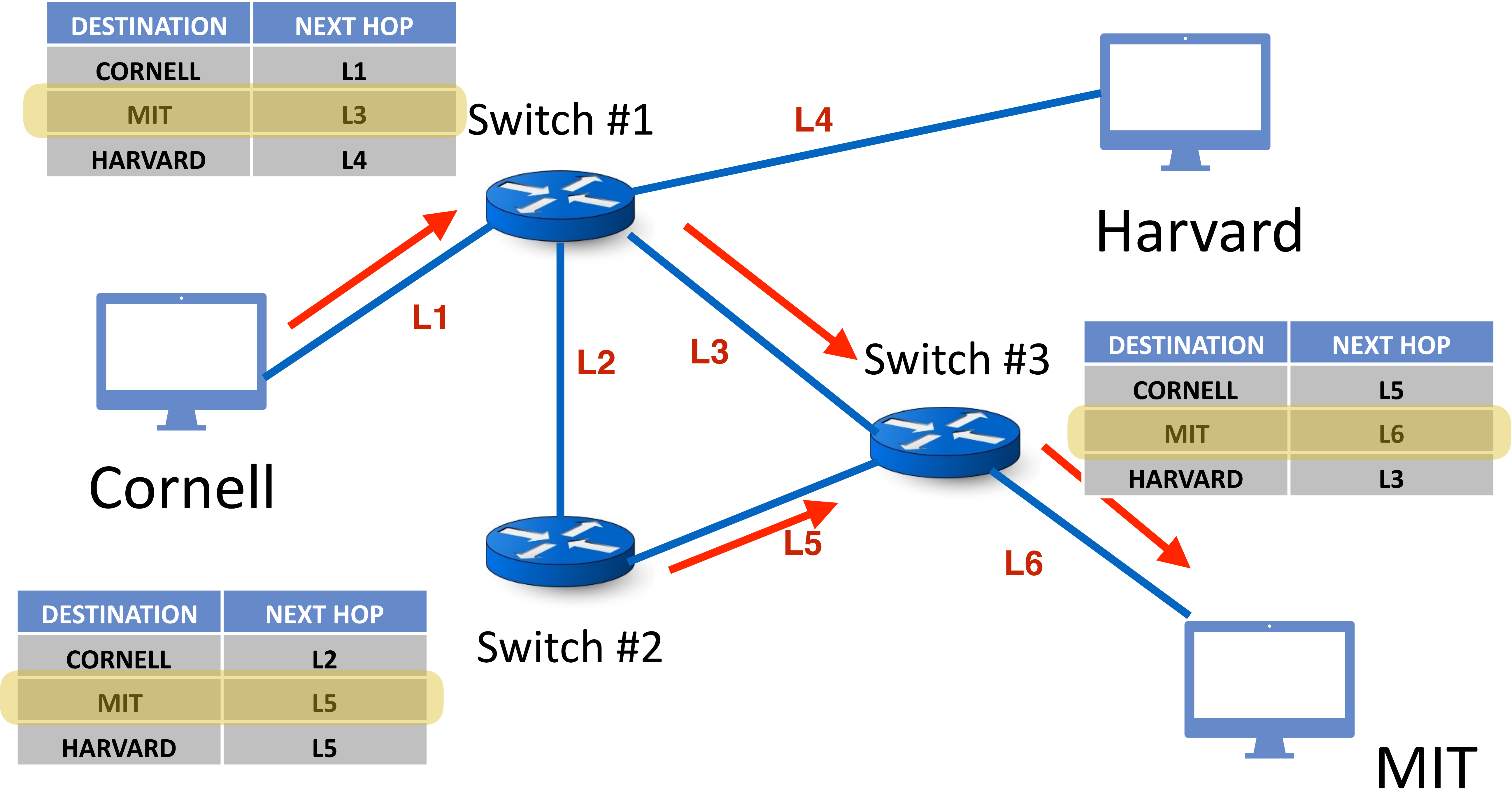
DESTINATION	NEXT HOP
CORNELL	L5
MIT	L6
HARVARD	L3

DESTINATION	NEXT HOP
CORNELL	L2
MIT	L5
HARVARD	L5

Each Switch stores a table indicating the next hop for corresponding destination of a packet (called a routing table)

Routing Table: The right way to think about them

- Lets focus on one destination - MIT



DESTINATION	NEXT HOP
CORNELL	L1
MIT	L3
HARVARD	L4

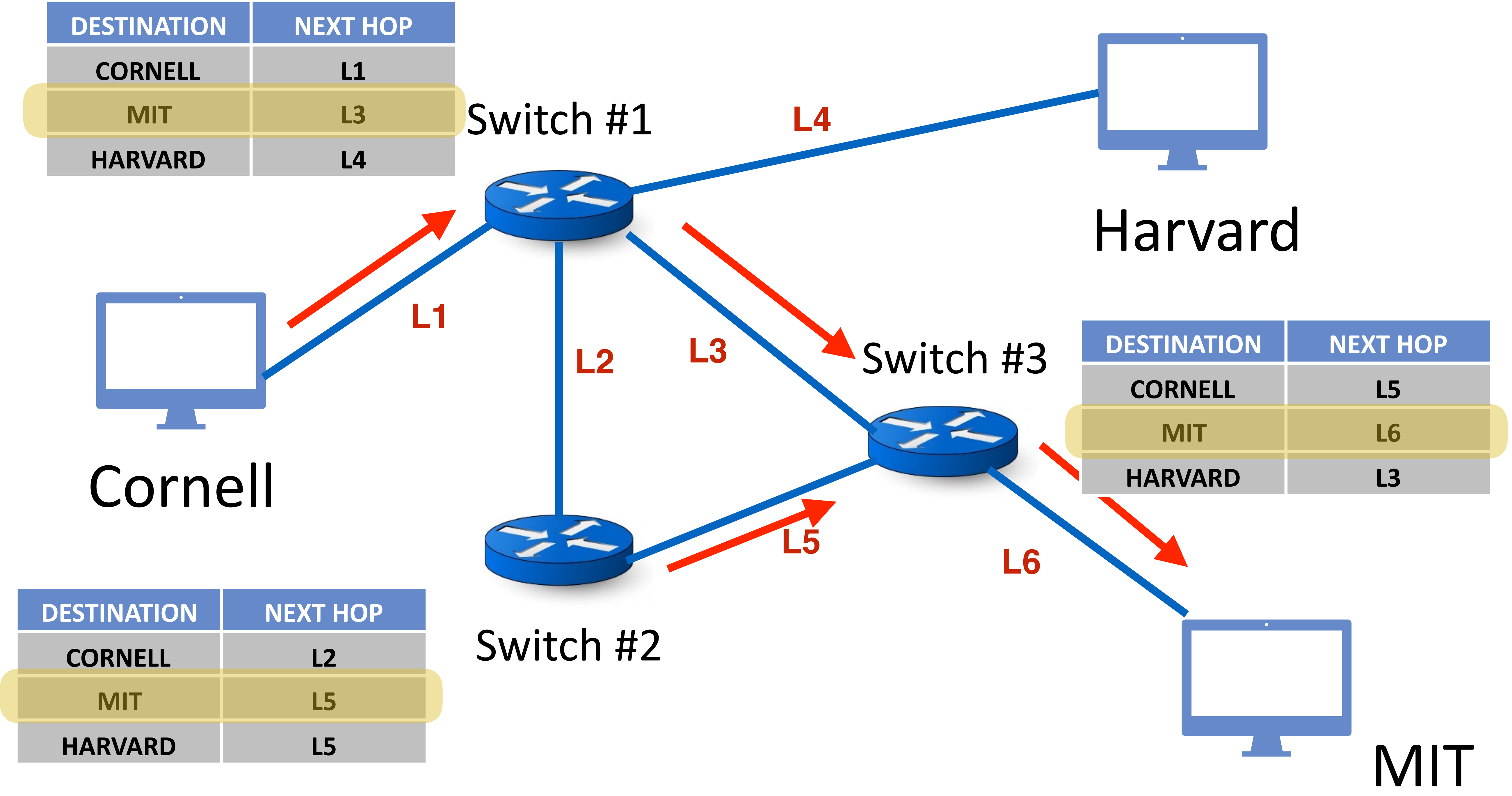
DESTINATION	NEXT HOP
CORNELL	L5
MIT	L6
HARVARD	L3

DESTINATION	NEXT HOP
CORNELL	L2
MIT	L5
HARVARD	L5

See something interesting?

Routing Table: The right way to think about them

- Lets focus on one destination - MIT



DESTINATION	NEXT HOP
CORNELL	L1
MIT	L3
HARVARD	L4

DESTINATION	NEXT HOP
CORNELL	L5
MIT	L6
HARVARD	L3

DESTINATION	NEXT HOP
CORNELL	L2
MIT	L5
HARVARD	L5

Routing table entries for a particular destination form a (directed) spanning tree with that destination as the root!!!!

Routing Table: The right way to think about them

- Routing tables are nothing but ….
    - A collection of (directed) spanning tree
    - One for each destination

- **Routing Protocols**
    - “n” spanning tree protocols running in parallel

“Valid Routing Tables” (routing state)

- Global routing state is valid if:
    - it **always** results in deliver packets to their destinations

- **Goal of Routing Protocols**
    - Compute a valid state
    - But how to tell if a routing state is valid?…
    - Think about it, what could make routing incorrect?

Validity of a Routing State

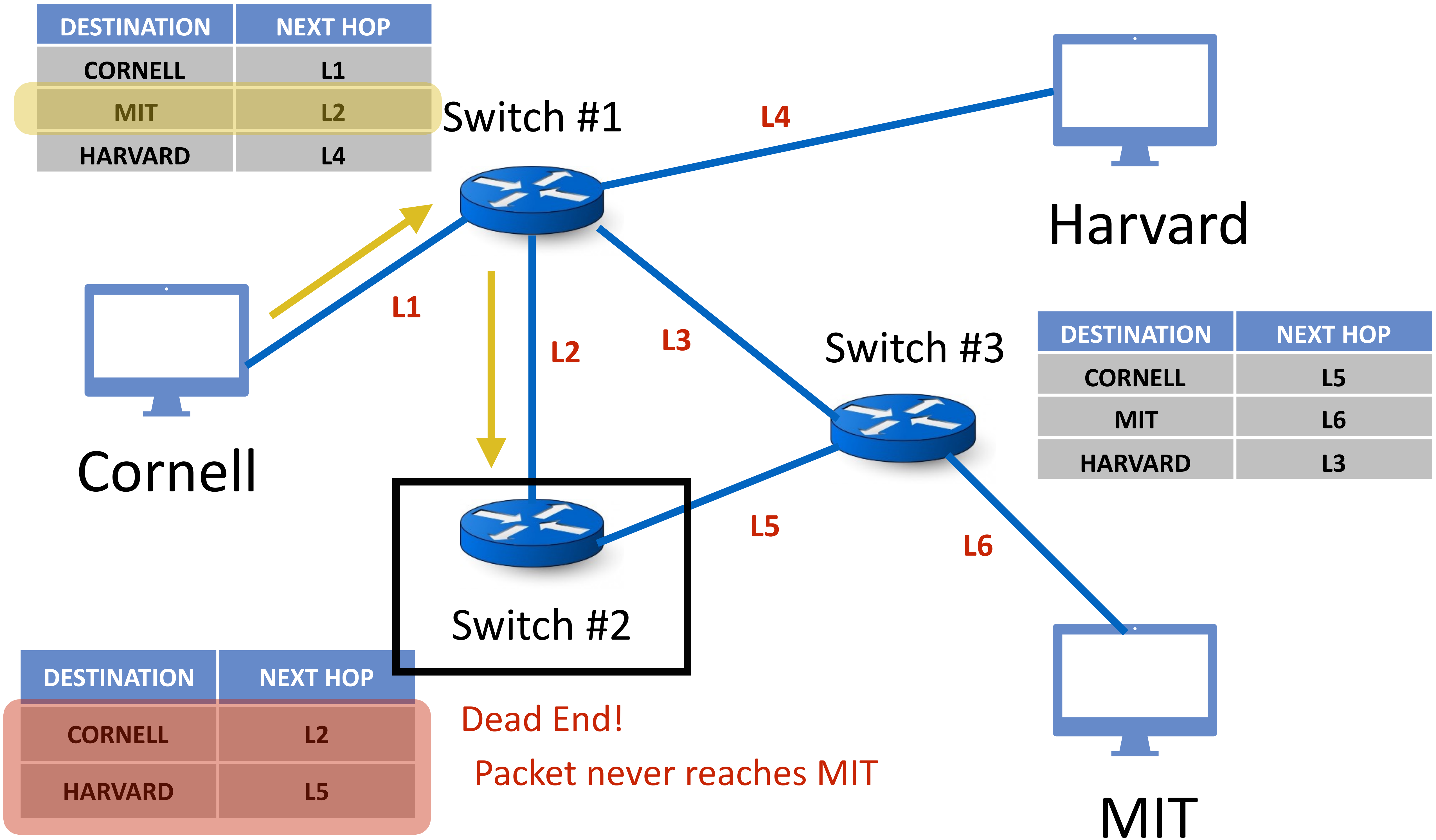
- Global routing state valid **if and only if**:
  - There are no **dead ends** (other than destination)
  - There are no **loops**

- A **dead end** is when there is **no outgoing link**
  - A packet arrives, but ..
    - the routing table does not have an outgoing link
    - And that node is not the destination

- A **loop** is when a **packet cycles around** the same set of nodes forever

Example: Routing with Dead Ends

- Suppose packet wants to go from Cornell to MIT using given state:



DESTINATION	NEXT HOP
CORNELL	L1
MIT	L2
HARVARD	L4

DESTINATION	NEXT HOP
CORNELL	L5
MIT	L6
HARVARD	L3

DESTINATION	NEXT HOP
CORNELL	L2
HARVARD	L5

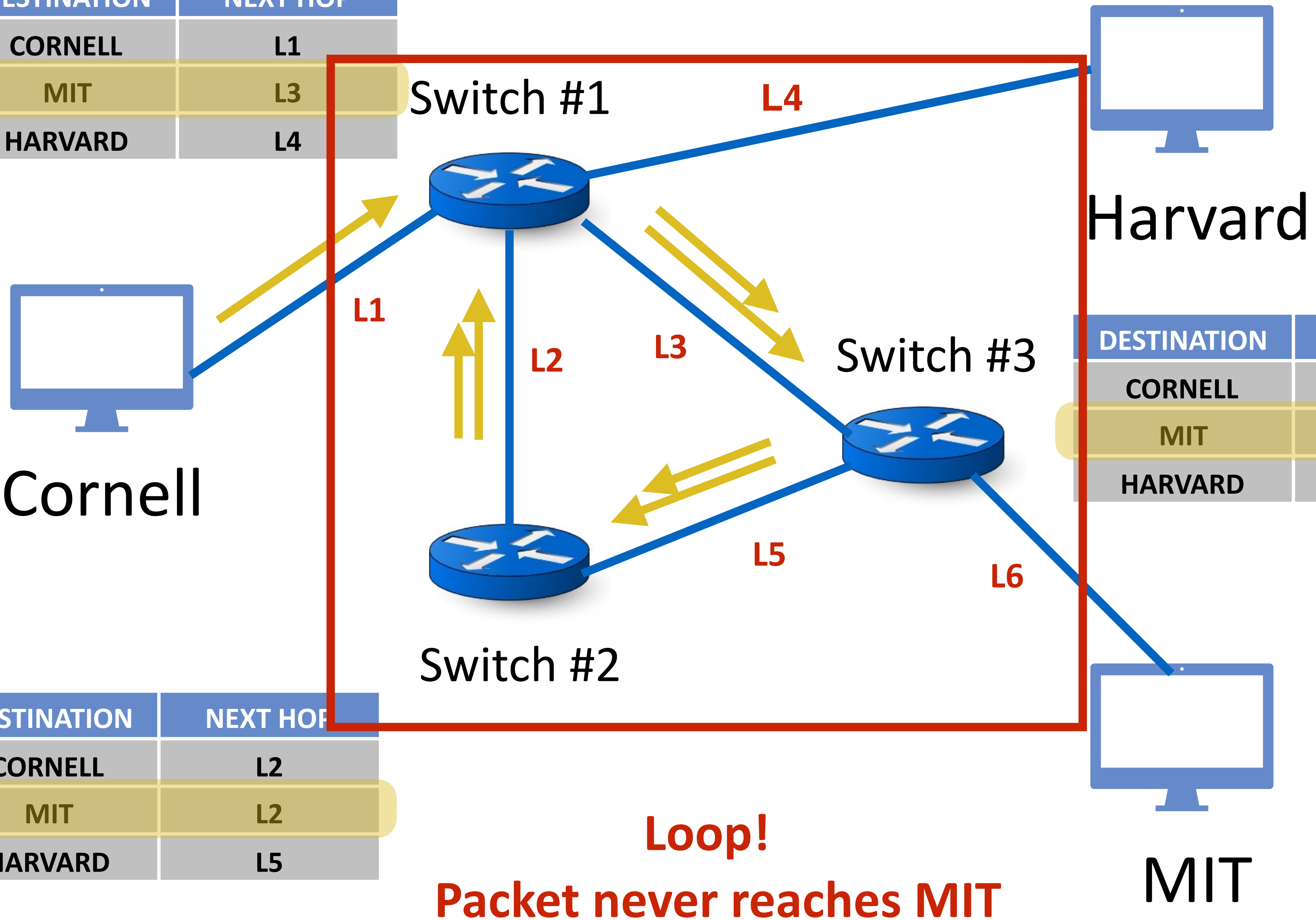
Dead End!

Packet never reaches MIT

No forwarding decision for MIT!

Example: Routing with Loops

- Suppose packet wants to go from Cornell to MIT using given state:



DESTINATION	NEXT HOP
CORNELL	L1
MIT	L3
HARVARD	L4

DESTINATION	NEXT HOP
CORNELL	L5
MIT	L5
HARVARD	L3

DESTINATION	NEXT HOP
CORNELL	L2
MIT	L2
HARVARD	L5

Loop!

Packet never reaches MIT

Two Questions

- How can we **verify** given routing state is valid?
- How can we **produce** valid routing state?

Checking Validity of a Routing State

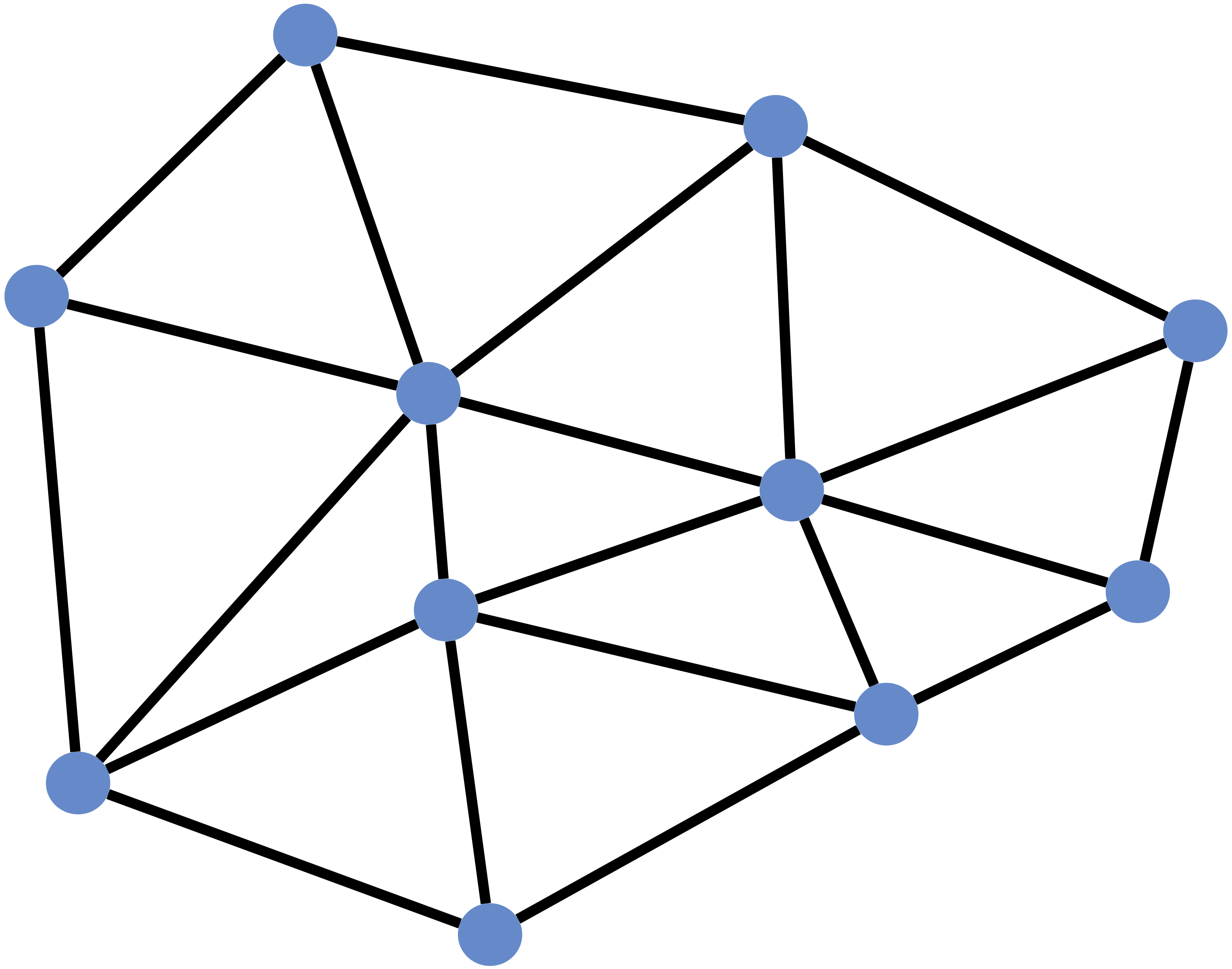
- Check validity of routing state for one destination at a time…

- For each node:
  - Mark the outgoing link with arrow for the required destination
  - There can only be one at each node

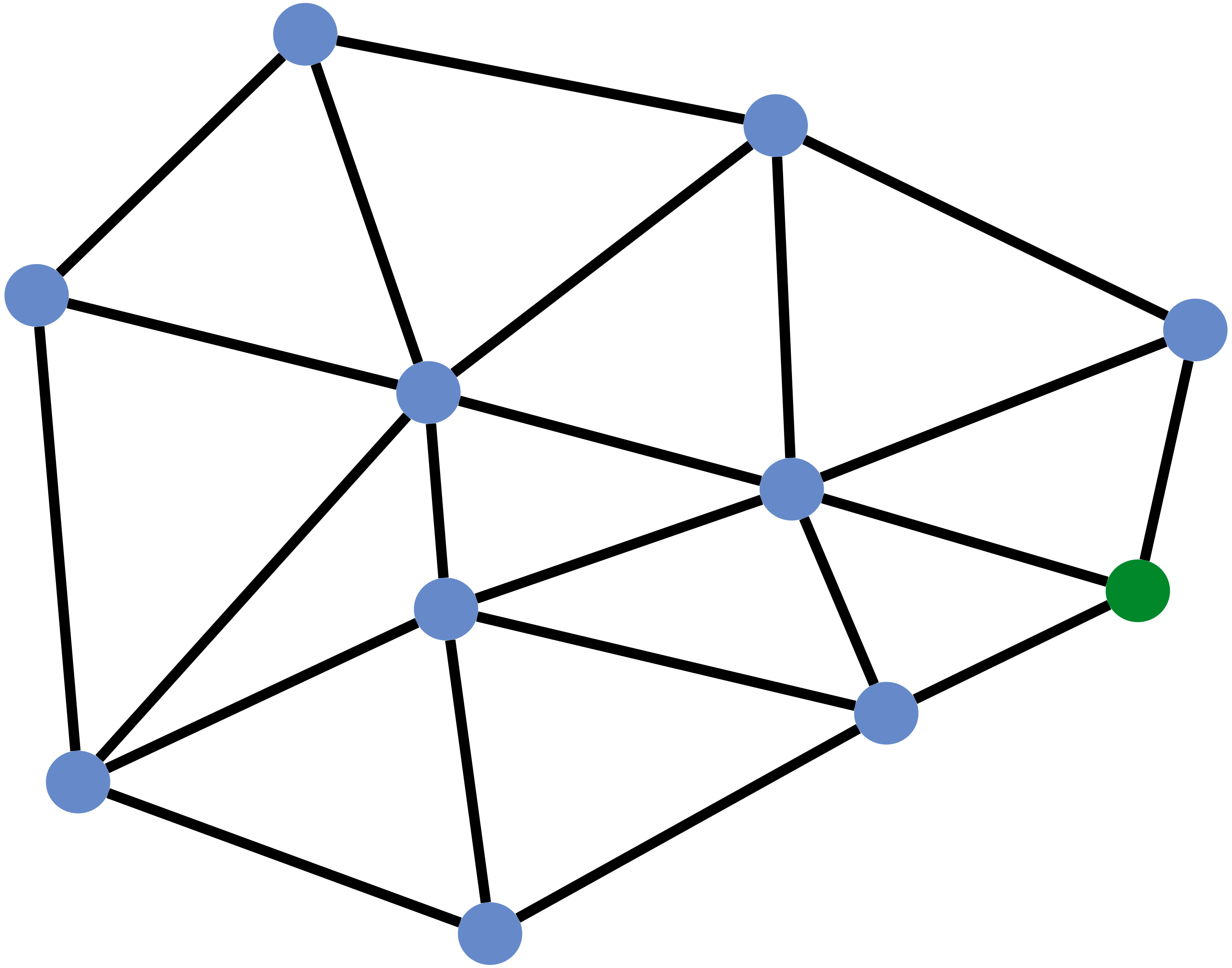
- Eliminate all links with no arrows

- Look what’s left. **State is valid if and only if**
  - Remaining graph is a spanning tree with destination as sink
  - Why is this true?
    - Tree -> No loops
    - Spanning (tree) -> No dead ends

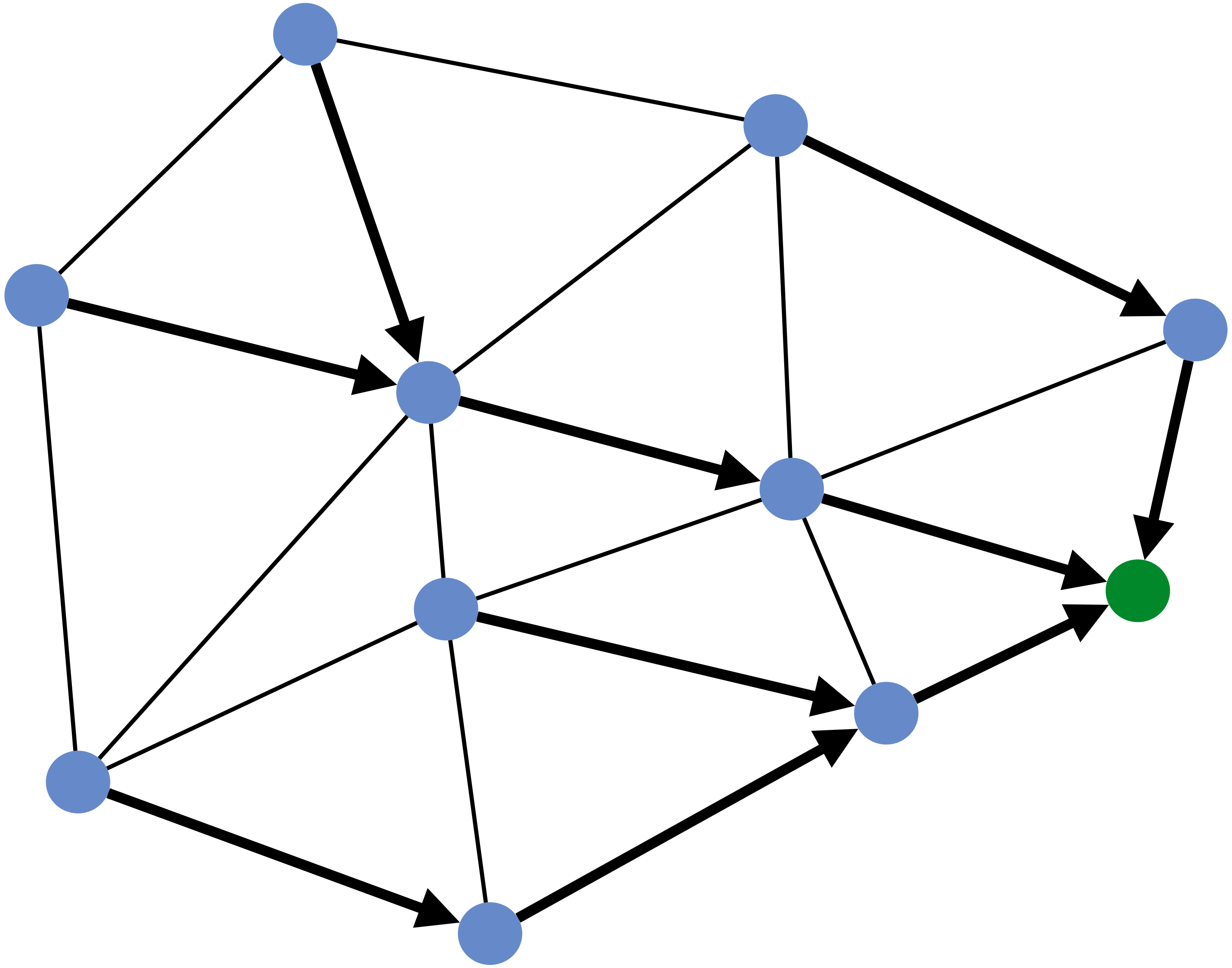
Example 1



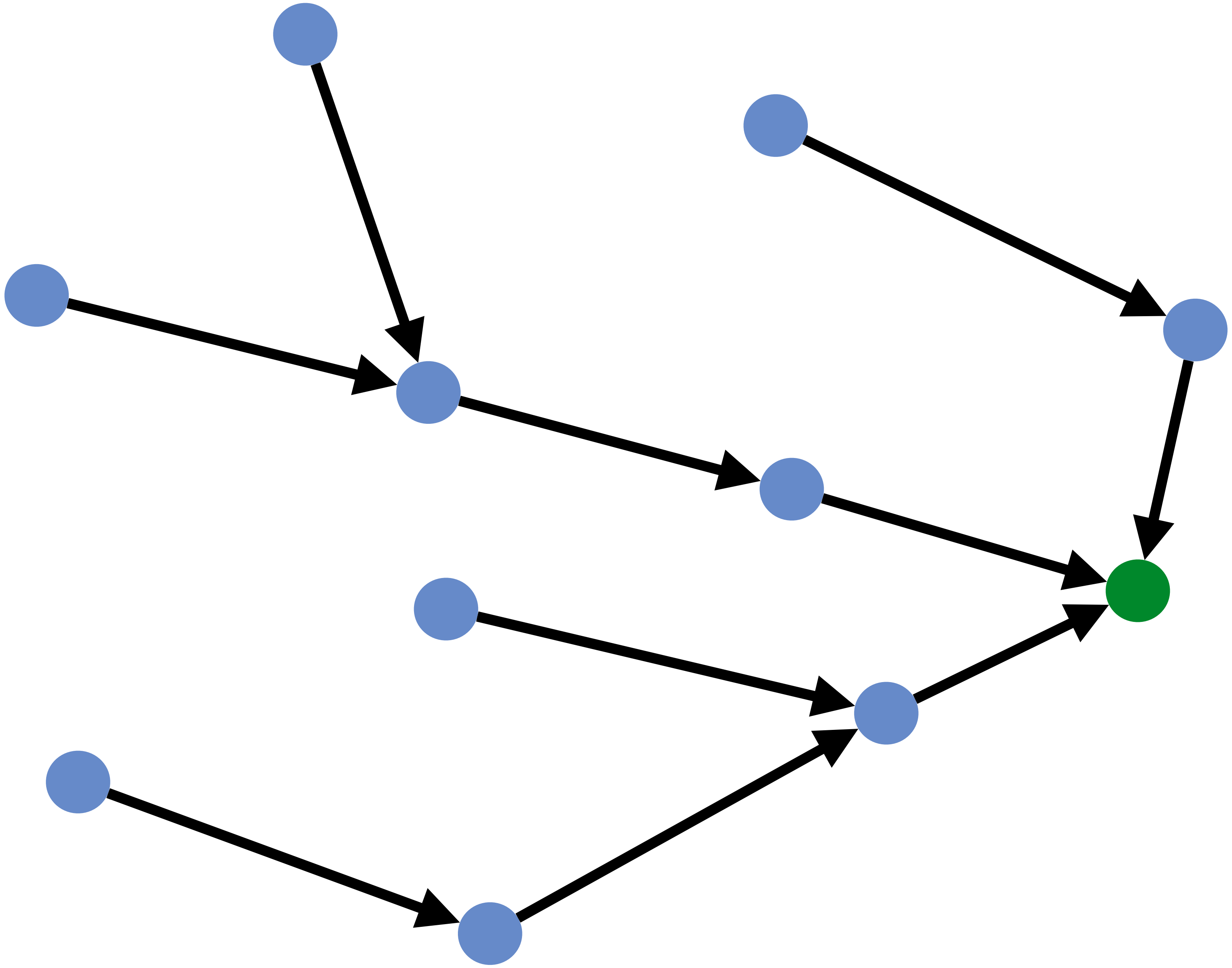
Example 1: Pick Destination



Example 1: Put Arrows on Outgoing Ports

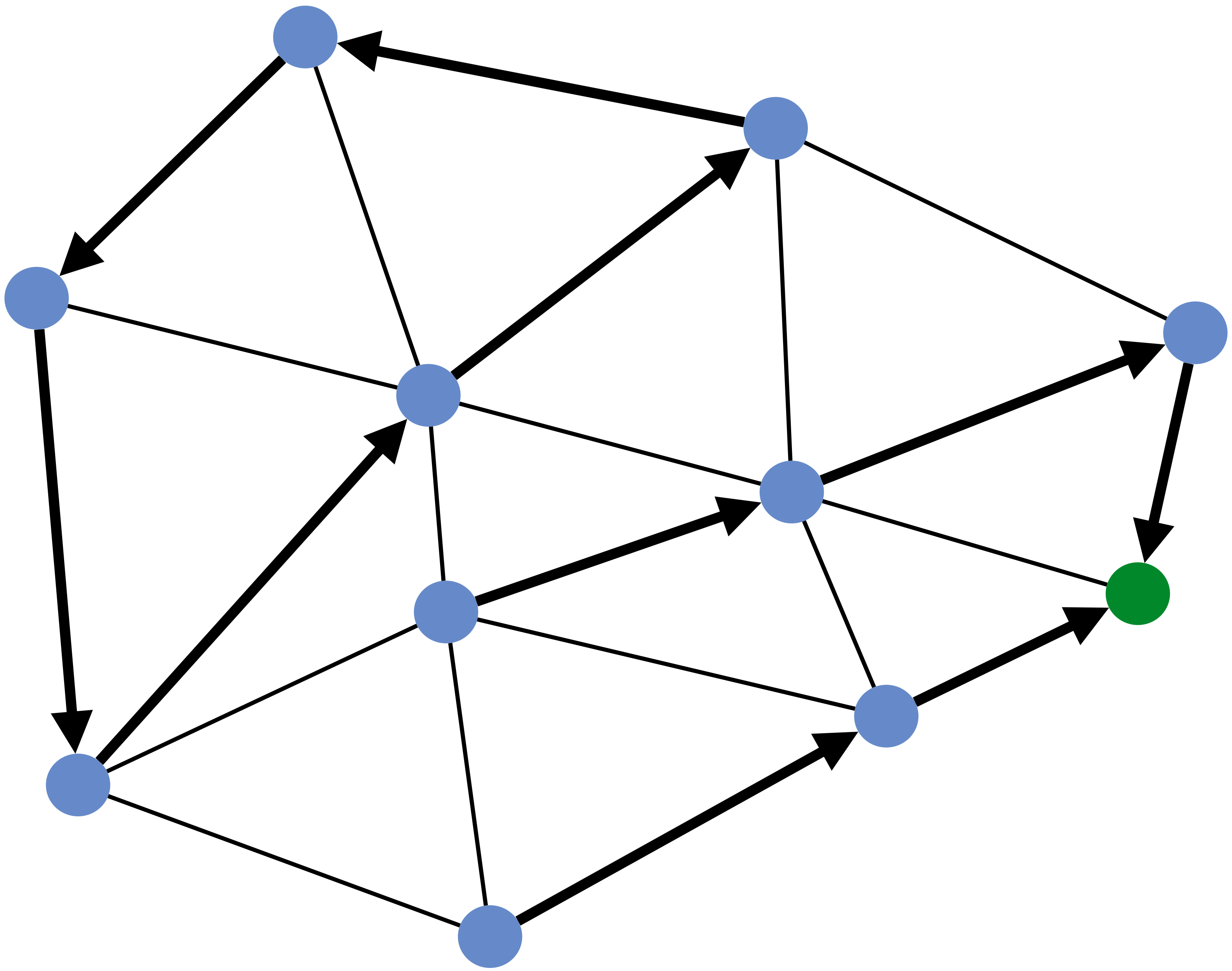


Example 1: Remove unused Links

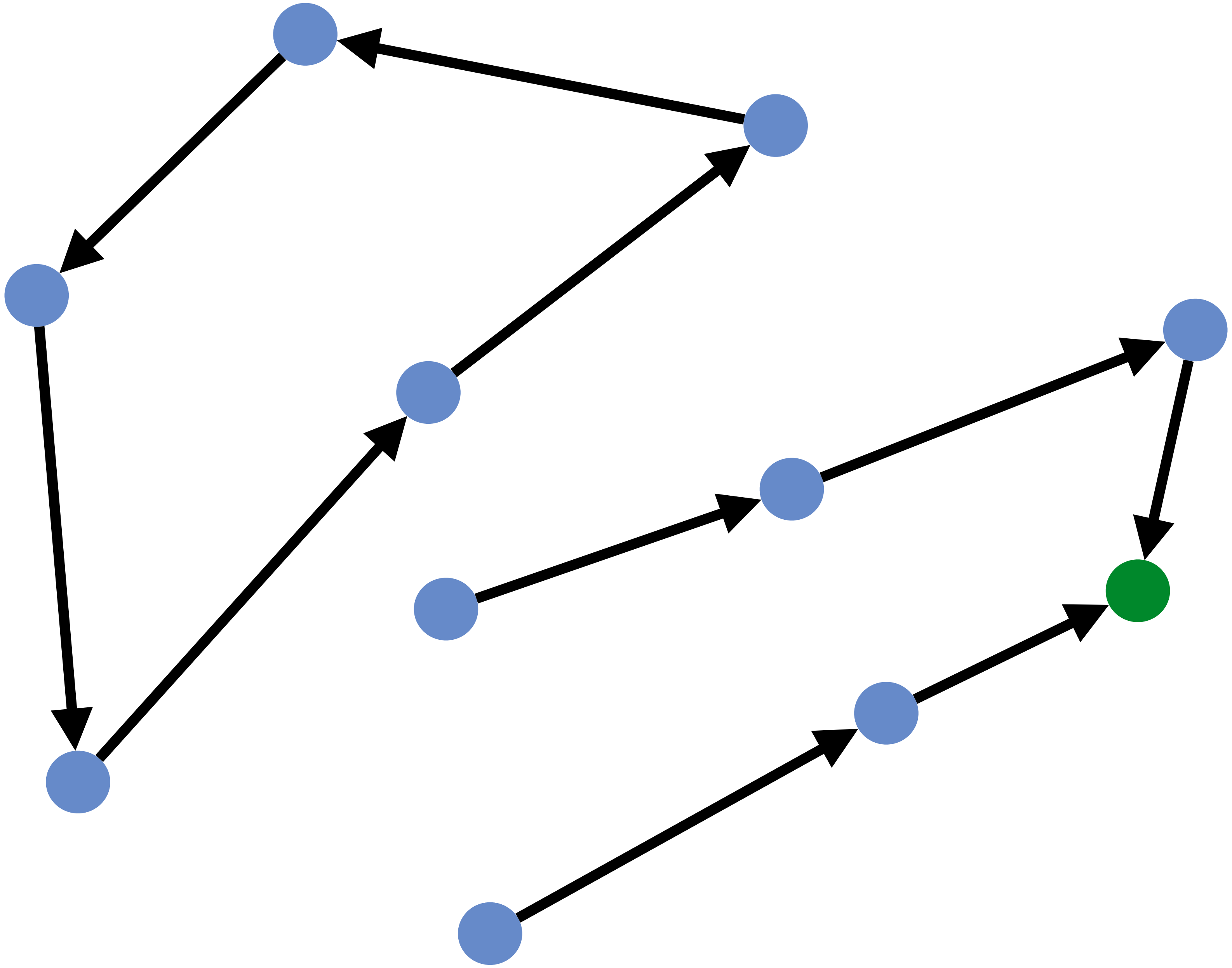


Leaves Spanning Tree: Valid

Example 2:

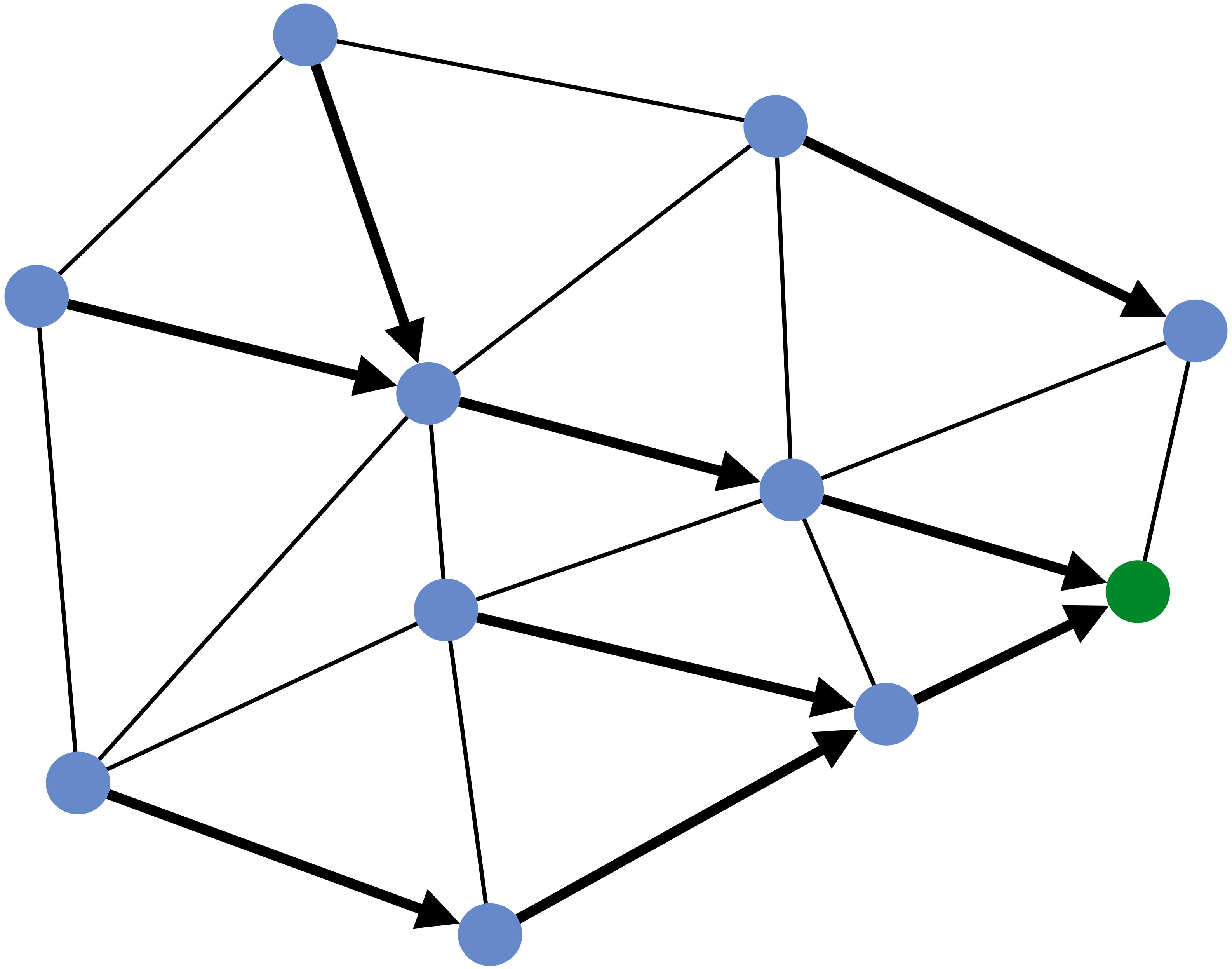


Example 2:

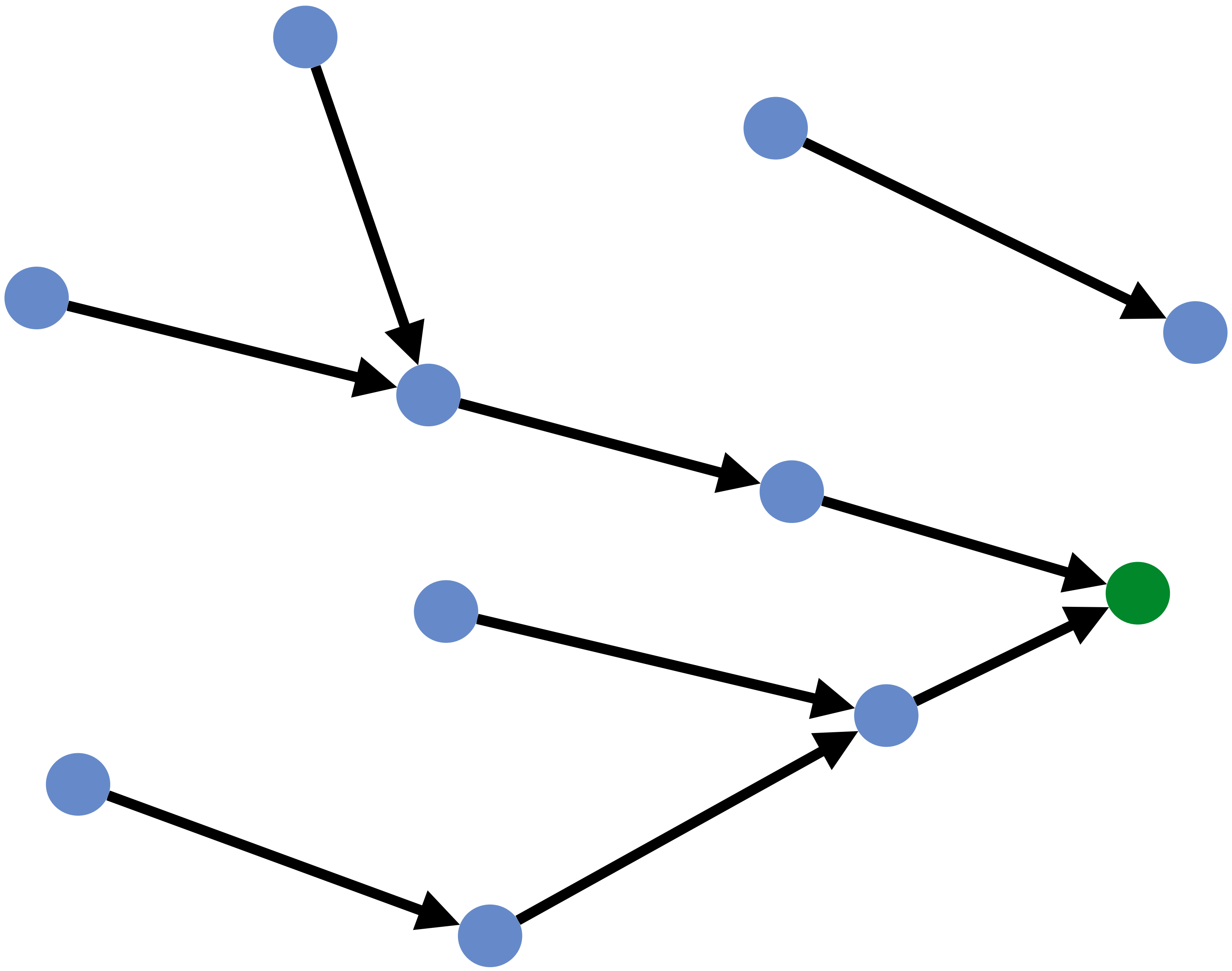


Is this valid?

Example 3:



Example 3:



Is this valid?

Checking Validity of a Routing State

- Simple to check validity of routing state for a particular destination

- Dead ends: nodes without arrows

- Loops: obvious, disconnected from destination and rest of the graph

Two Questions

- How can we **verify** given routing state is valid?

- How can we **produce** valid routing state?

Creating Valid Routing State

- Easy to avoid dead ends

- Avoiding loops is hard

- **The key difference between routing protocols is how they avoid loops!**

- Try to think a loop avoidance design for five minutes

#1: Create Tree Out of Topology

- Remove enough links to create a tree containing all nodes

- Sounds familiar? Spanning trees!

- If the topology has no loops, then just make sure not sending packets back from where they came
    - That causes an immediate loop

- Therefore, if no loops in topology and no formation of immediate loops ensures valid routing

- However… three challenges
    - Unnecessary host resources used to process packets
    - High latency
    - Low bandwidth (utilization)

#2: Obtain a Global View

- A global view of the network makes computing paths without loops easy
  - Many graph algorithms for computing loop-free paths
  - For e.g., Dijkstra’s Algorithm

- Getting the global view of network is challenging!

#3: Distributed Route Computation

- Often getting a global view of the network is infeasible
    - Distributed algorithms to compute feasible route

- **Approach A**: Finding optimal route for maximizing/minimizing a metric

- **Approach B**: Finding feasible route via exchanging paths among switches

Welcome to the Network Layer!

- THE functionality: **delivering the data**

- **THE protocol: Internet Protocol (IP)**
    - To achieve its functionality (delivering the data), IP protocol has **three** responsibilities

- **Addressing (next lecture)**
- **Encapsulating data into packets (next lecture)**
- **Routing (using a variety of protocols; several lectures)**

Next lecture!