Link-state Routing (SPF)

- Routers distribute link cost and topology information to all other routers in their *area*
- All routers have complete information about the network
- Each router computes its own optimal path to destinations
- Ensures loop free environments
- Pair of switches periodically:

 - Test link between them

 - *Broadcast* link status message

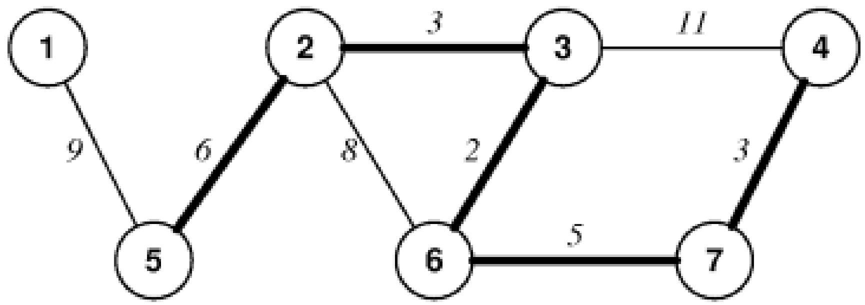
- Switch:

 - Receives status message

 - Builds a graph of the network

 - Uses Dijkstra's algorithm with itself as the source node to build routing table

Link-state Routing Example



- Assume nodes 2 and 3:
  - Test link between them
  - *Broadcast* link state information
- Each node (switch):
  - Receives broadcast state information
  - Recomputes routes as needed

Dijkstra's Shortest Path Algorithm

- Input:
  - Graph with weighted edges
  - Source node, *n*
- Output:
  - Set of shortest paths from *n* to each node
  - Cost of each path
- Called *Shortest Path First* (SPF) algorithm

Shortest Path First Algorithm

- Start with self as source node
- Move outward
- At each step:

 - Find node *u* such that it:

   * Has not been considered

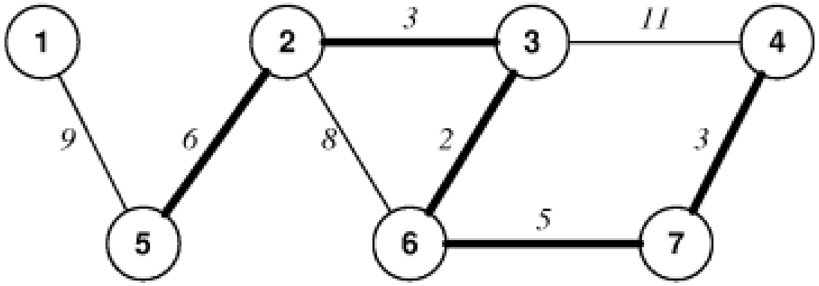
   * Is ``closest'' (weight-wise) to source

 - Compute:

   * Distance from *u* to each neighbor *v*

   * If distance shorter, make path from *u* go through *v*

Result of SPF Algorithm



u	D[u]	R[u]
2	8	6
3	2	6
7	5	6
2	5	3
4	13	3
5	11	3
4	8	7

- Example routes from node 6:

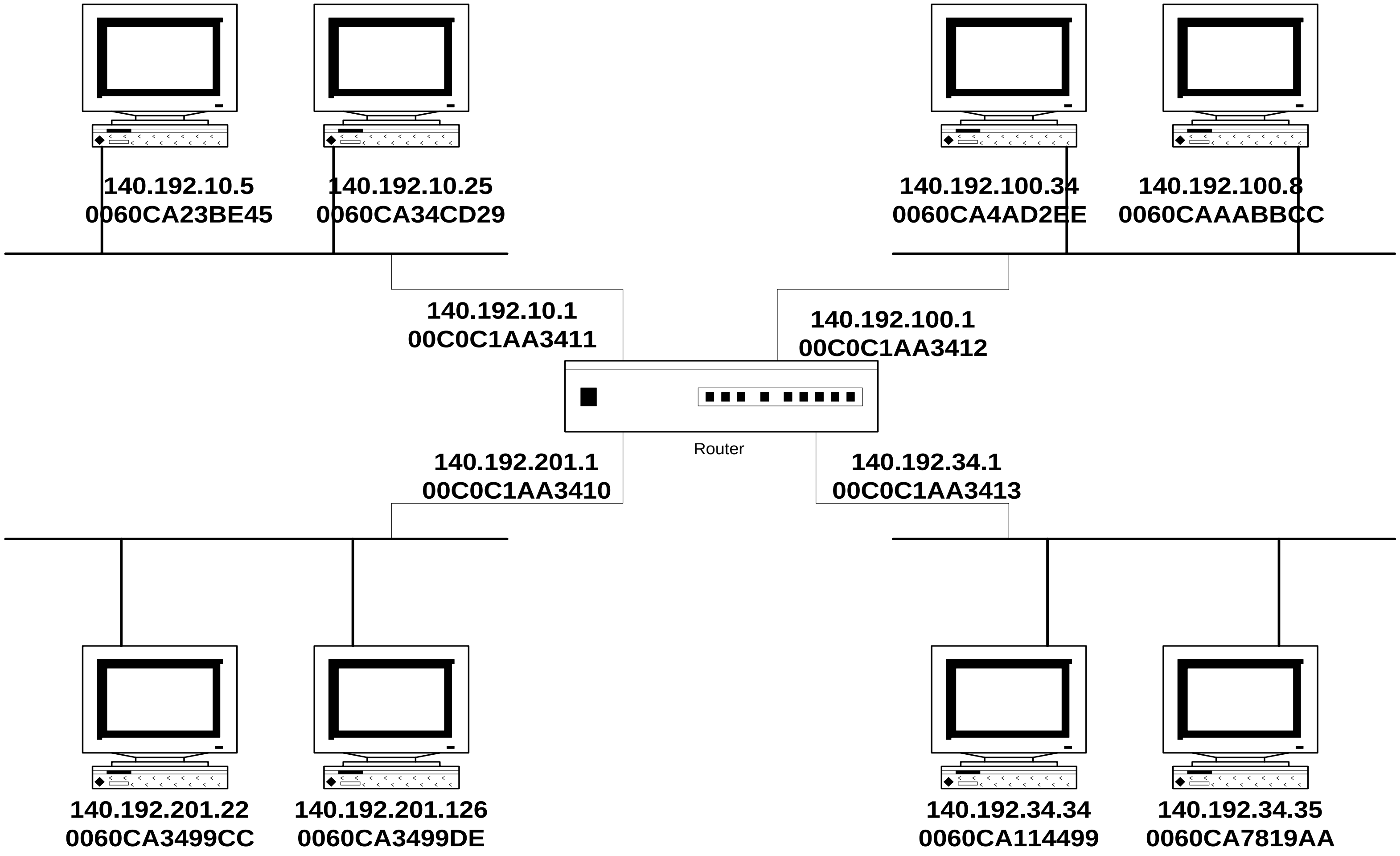
- To 3, next hop = node 6, cost = 2
- To 2, next hop = node 3, cost = 5
- To 5, next hop = node 3, cost = 11
- To 4, next hop = node 7, cost = 8

TCP/IP Routing Overview

- Strip off layer 2 headers/trailers
- Extract destination address field, *D*
- Look up *D* in the routing table
- Find next hop address, *N*
- Send datagram to *N*
- Add on layer 2 headers/trailers

TCP/IP Routing (1/4)

IP Routing



TCP/IP Routing (2/4)

Layer 2

DA	SA	Protocol	P. DA	P. SA	Data	FCS

Layer 3

P. DA	P. SA	Data

Layer 2

00C0C1AA3413	0060CA114499	IP	140.192.10.5	140.192.34.34	Data	FCS

Layer 3

140.192.10.5	140.192.34.34	Data

TCP/IP Routing (3/4)

From 140.192.34.34 to 140.192.10.5
140.192.34.34 knows that 140.192.10.5 isn't on the same net and sends it to router at 140.192.34.1
Note DA for layer 2

00C0C1AA3413	0060CA114499	IP	140.192.10.5	140.192.34.34	Data	FCS

Inside the router the Layer 2 headers and trailers are removed leaving only the layer 3 packet.
The router looks up the packet's DA in the routing table and forwards to the appropriate interface.

140.192.10.5	140.192.34.34	Data

At the interface, layer 2 headers and trailers are added back.
DA is the address of the destination host.
SA is the address of the router.
FCS is recalculated.

0060CA23BE45	00C0C1AA3411	IP	140.192.10.5	140.192.34.34	Data	FCS

TCP/IP Routing (4/4)

Routing Table

Network	Interface
140.192.10.0	0
140.192.100.0	1
140.192.201.0	2
140.192.34.0	3

Layer 2 <--> Layer 3 Table
ARP Table

Network.Host	Layer 2
140.192.10.5	0060CA23BE45
140.192.10.25	0060CA34CD29
140.192.100.34	0060CA4AD2EE
140.192.100.8	0060CAAABBCC
140.192.201.22	0060CA3499CC
140.192.201.126	0060CA3499DE
140.192.34.34	0060CA114499
140.192.34.35	0060CA7819AA

Internet Routing

- Autonomous System (AS):

- Routers in the Internet are divided into groups, where each group is named an AS
- Routers within an AS exchange routing information, which is then summarized before being passed to another AS
- No routing protocol can scale to support the Internet
- Each AS under one administrative authority (AA)
- No exact meaning for AA:
  * University
  * Organization
  * Multiple sites of the same organization

Internet Routing Protocols

- Interior Gateway Protocols (IGPs):
  - Routers within an AS use an IGP to exchange routing information
  - An AS must be able to isolate itself from other sites; it must be able to keep its local internets operating even when other parts of the Internet have failed
  - Sites want full administrative control over their routers and networks and may not want to run the same routing protocols as other sites
- Exterior Gateway Protocols (EGPs):
  - A router in one AS uses an EGP to exchange routing information with a router in another AS
  - More complex, more flexible, less traffic than IGPs
  - To save traffic an EGP summarizes routing information from the AS before passing it to another AS
  - Policy constraints over released routing information

Internet Routing Metrics

- Routing metric:

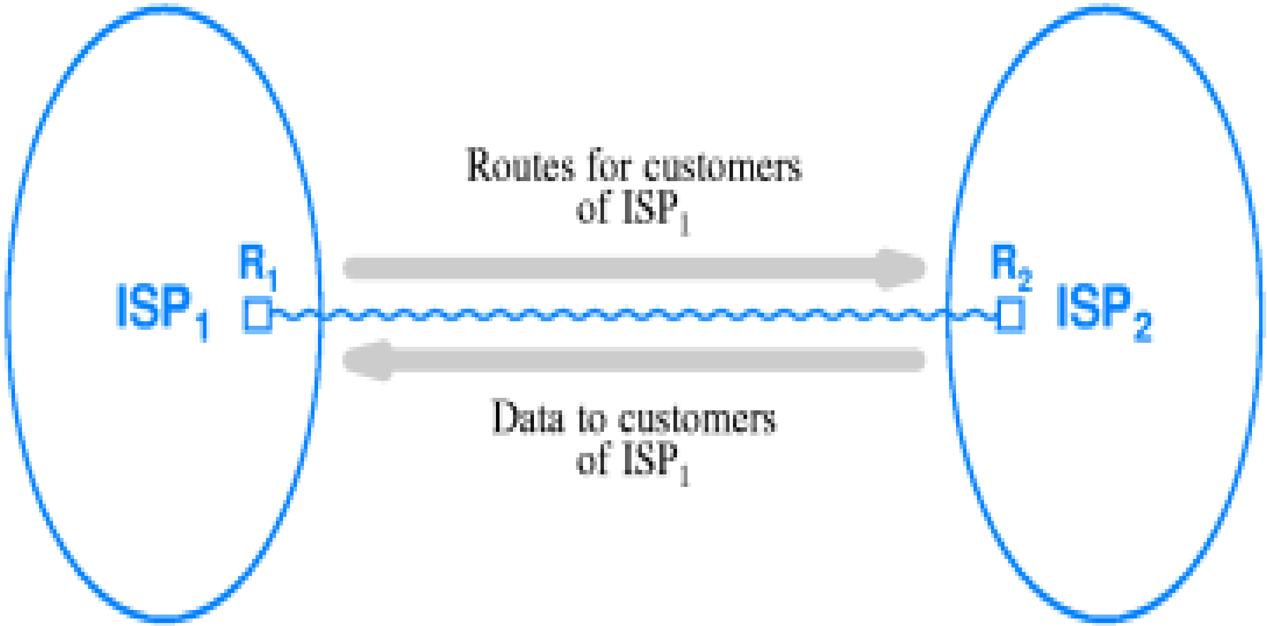
  - A measure of the path that routing software uses when choosing a route
  - Internet routing uses a combination of two metrics:

    * Administrative cost: Assigned manually according to policy

    * Hop count: Number of intermediate networks (or routers)

- IGPs: Use metrics to choose optimal paths within an AS
- EGPs: Merely find a path to each destination; cannot find optimal path

Route and Data Traffic



- Each ISP is an AS that uses an EGP to advertise its customers’ networks to other ISPs
- After an ISP advertises destination D, datagrams destined for D can begin to arrive

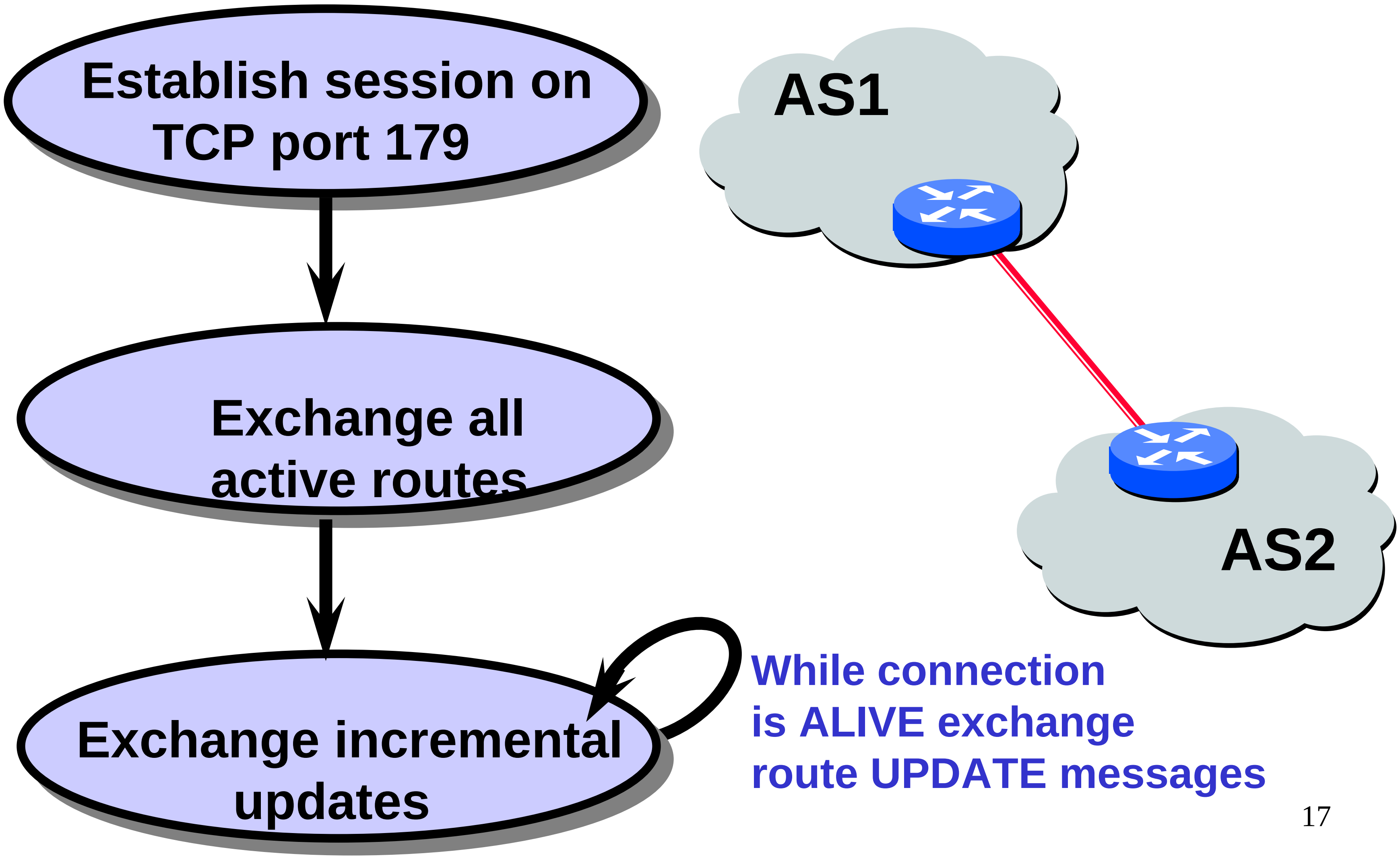
Border Gateway Protocol (BGP)

- BGP-4 (most popular EGP):
  - Routing among autonomous systems, no metrics, no details about routers within an AS
  - Policy support, restrict route advertisements to outsiders
  - Reliable transport: BGP uses TCP
- All major ISPs use BGP to exchange routing information
- *Routing arbiter system*:
  - A distributed database with all possible destination autonomous systems in the Internet
  - Each copy runs on a separate *route server*
  - ISPs use BGP to receive information from one of the route servers

BGP Operation (1/2)

- Distance vector algorithm with extra information:
  - For each route, store the complete path (ASs)
  - No extra computation, just extra storage
- Advantages:
  - Can make policy choices based on set of ASs in path
  - Can easily avoid loops

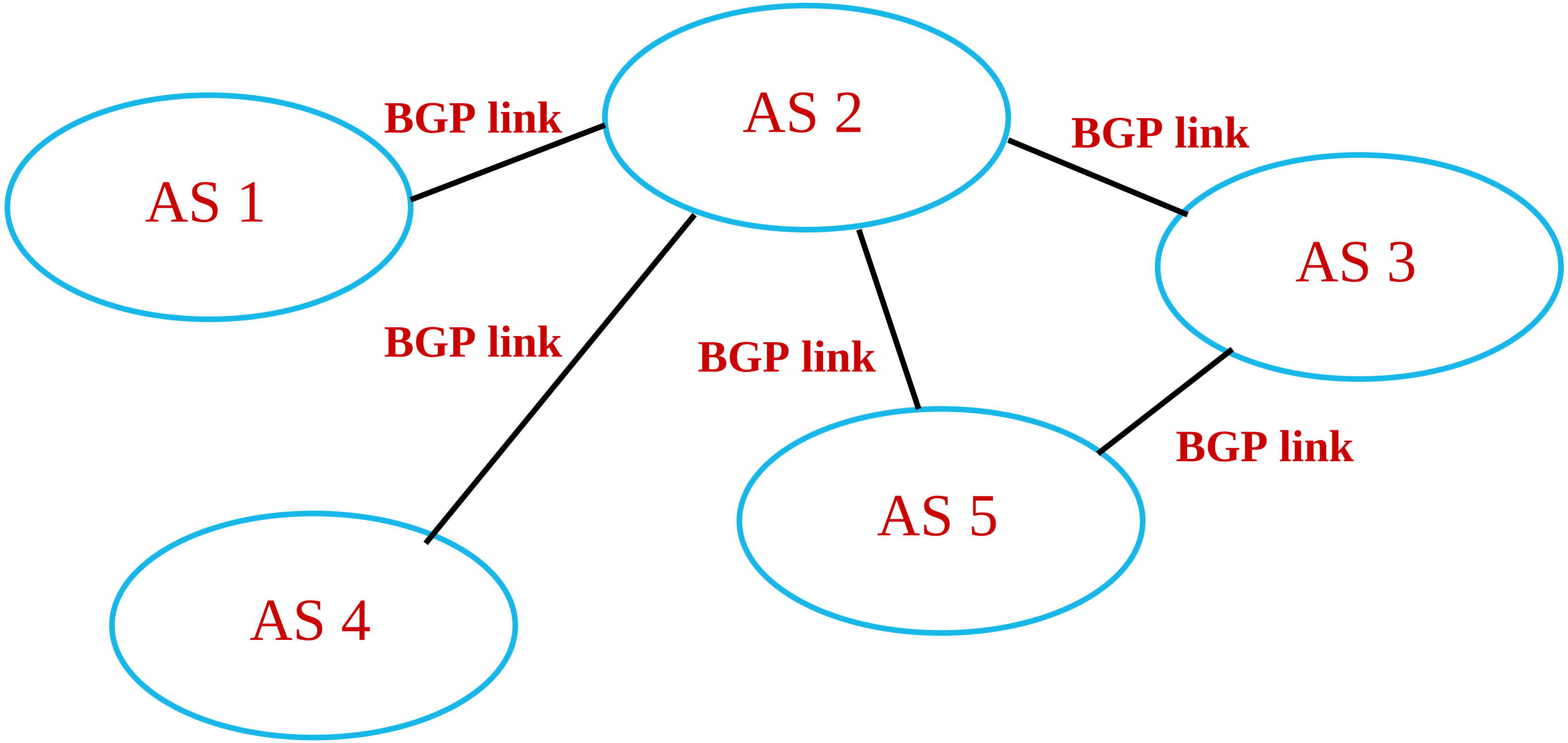
BGP Operation (2/2)



BGP Types of Messages

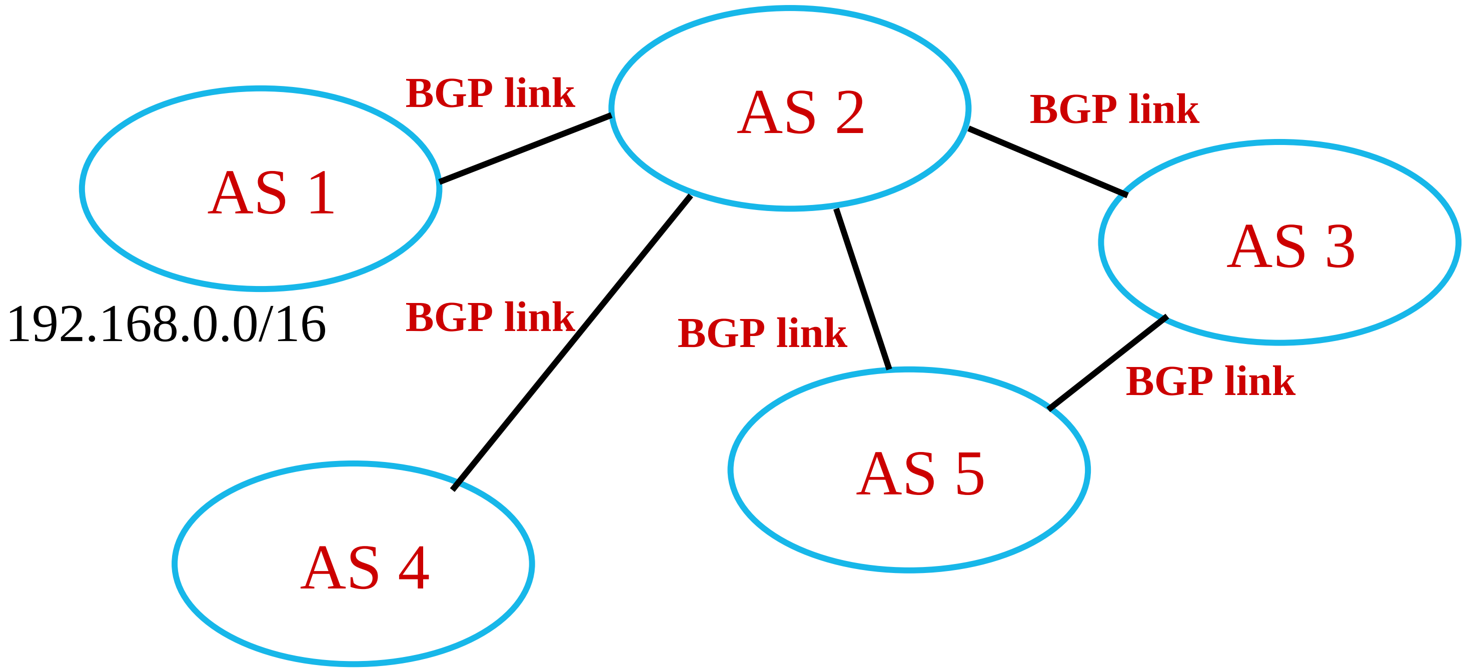
- **Open**: Establish a new peer session
- **Keep alive**: Handshake at regular intervals
- **Notification**: Shuts down a peer session
- **Update**: Announcing new routes or withdrawing previously announced routes
- Routes are specified in terms of prefix

BGP AS Links and Path Tree



- BGP provides routes to other ASs (address prefixes)
- BGP builds a graph of ASs
- Graph derived from routing information
- BGP sees the entire Internet as a graph
- BGP skips intermediate routers within ASs

BGP Route Advertising



	192.168.0.0/16 is Reachable through this AS:	
• AS 1 advertises:	192.168.0.0/16 through	AS 1
• AS 2 advertises:	192.168.0.0/16 through	AS 1 and AS 2
• AS 3 advertises:	192.168.0.0/16 through	AS 1 and AS 2 and AS 3
• AS 5 sees:	192.168.0.0/16 through	AS 1 and AS 2
• AS 5 sees:	192.168.0.0/16 through	AS 1 and AS 2 and AS 3

Routing Information Protocol (RIP)

- RIP is an IGP:
  - Routing within an autonomous system
  - Distance measured in network hops; *origin-one counting*
  - Unreliable transport: RIP uses UDP
  - Broadcast (v1) or multicast (v2) message delivery
  - Default route advertisements; can be installed to all routers of an organization
  - Distance vector routing approach
  - Passive RIP:
    * Only routers can propagate routing information
    * Hosts can listen passively and update their routing tables

RIP Routes Propagation

- Each outgoing message contains an advertisement that lists all the networks the sender can reach

  - Entry: (destination network, distance)

    * [IP address of network | Subnet mask | Next hop | Distance]

  - Distance in Internet hops (networks)
- Receiver:

  - If it does not have a route to destination, or
  - If distance shorter than the distance of the current route
  - Replaces its route, else it ignores the pair
- Chief advantage: Simplicity

  - Little configuration
  - Default route propagation; configure one router to ISP

RIP Disadvantages

- Each message contains the complete list of destination and distances a router knows

 - Large messages

- Receiver must compare all entries of an incoming message to the current route for the destination

 - Consumes CPU cycles

 - Introduces delay: Route changes propagate slowly, one router at a time

- Therefore, RIP does not scale to a large internet
- 15 Hops
- RIPv1: Classful routing.
- RIPv2 supported CIDR