Computer Networks and Data Systems

Interior Gateway Protocols (IGPs)

Note:
One of two critical subsystems

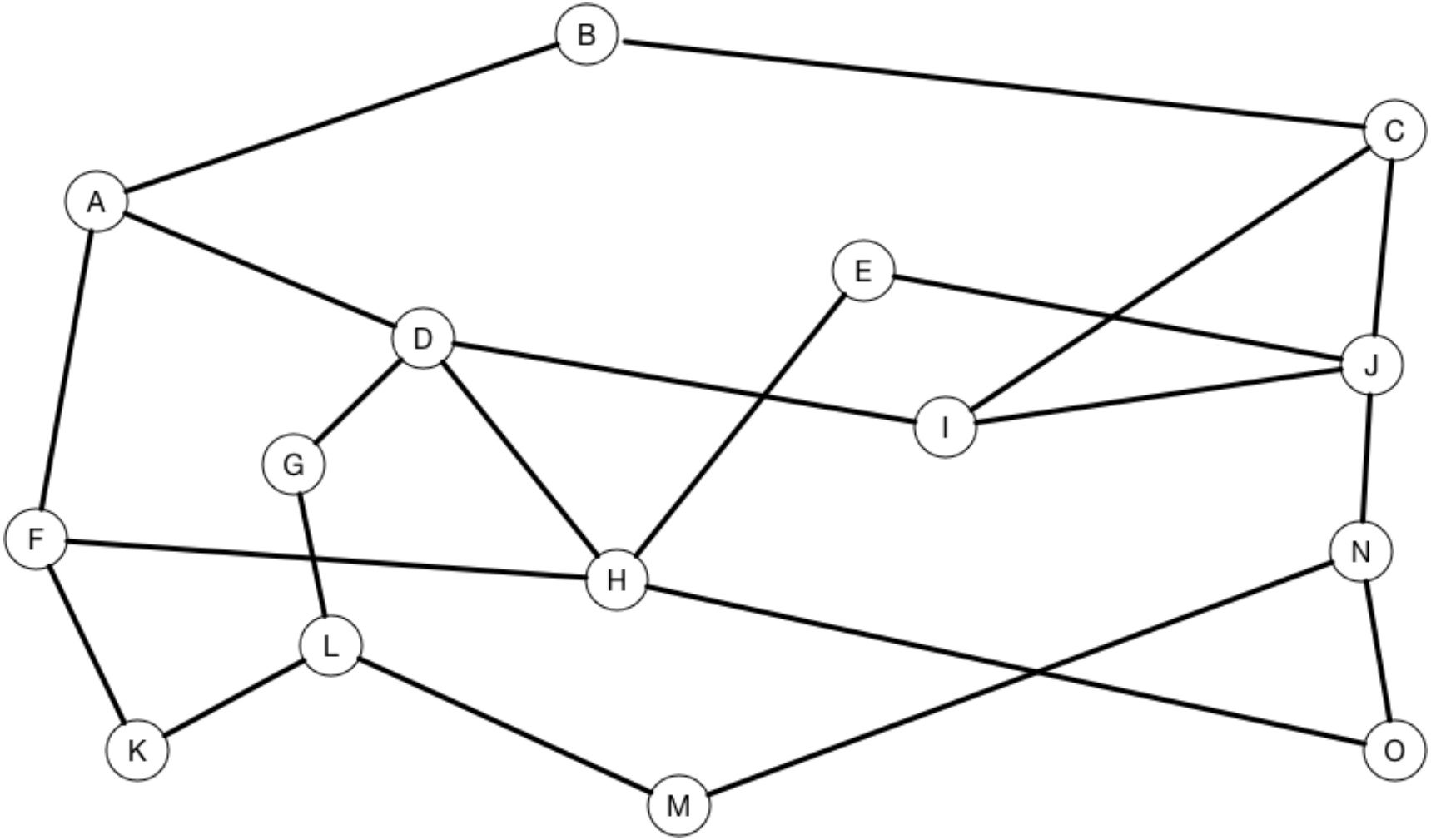
Routing (most importantly BGP) and naming (solely DNS) are, by far, the two most critical subsystems of the Internet infrastructure. Participation in and access to the routers themselves are generally, or rather should be, restricted to network administrators.

Determining a Route or Path

- Static (non-adaptive)
- Adaptive
  - Neighbor updates (distributed algorithm)
  - Network conditions
  - Per-packet detail
- Broadcast (explorer)

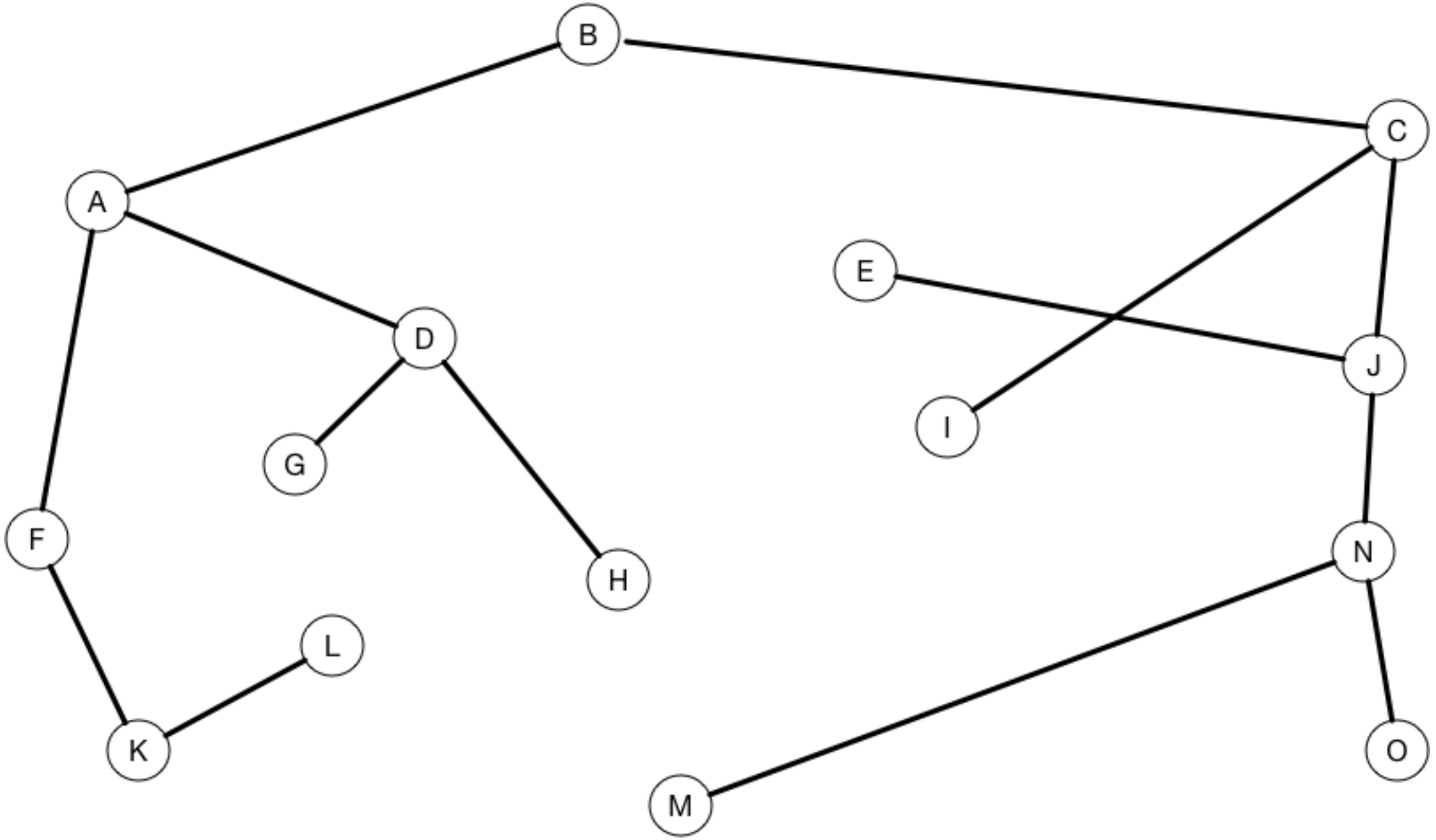
Sample Network

***adapted from Tannebaum, Computer Networks, Fig. 5-5(a)**



Sink Tree for Router B

*adapted from Tannebaum, Computer Networks, Fig. 5-5(b)



Factors in Path Computation: Metrics and Path Properties

- Cost
- Delay / performance / load
- Reliability / error rate
- Distance
- Policy (e.g. trustworthiness)

Route Maintenance Challenges

- Convergence time
- Fault discovery
- Timer values (static / adaptive)
- Link / path flapping
- Equal cost path selection
- Loop avoidance

What makes an IGP an IGP?

IGPs in Practice

- Internal
- Autonomy
- Common IP routing protocols used as IGPs
  - RIP (versions 1, 2, and ng)
  - OSPFv2 and OSPFv3
  - IS-IS
  - EIGRP
  - BGP

Distributed Routing Algorithms

- Two general types:
  - Distance vector
  - Link state

Distance Vector (DV) Routing

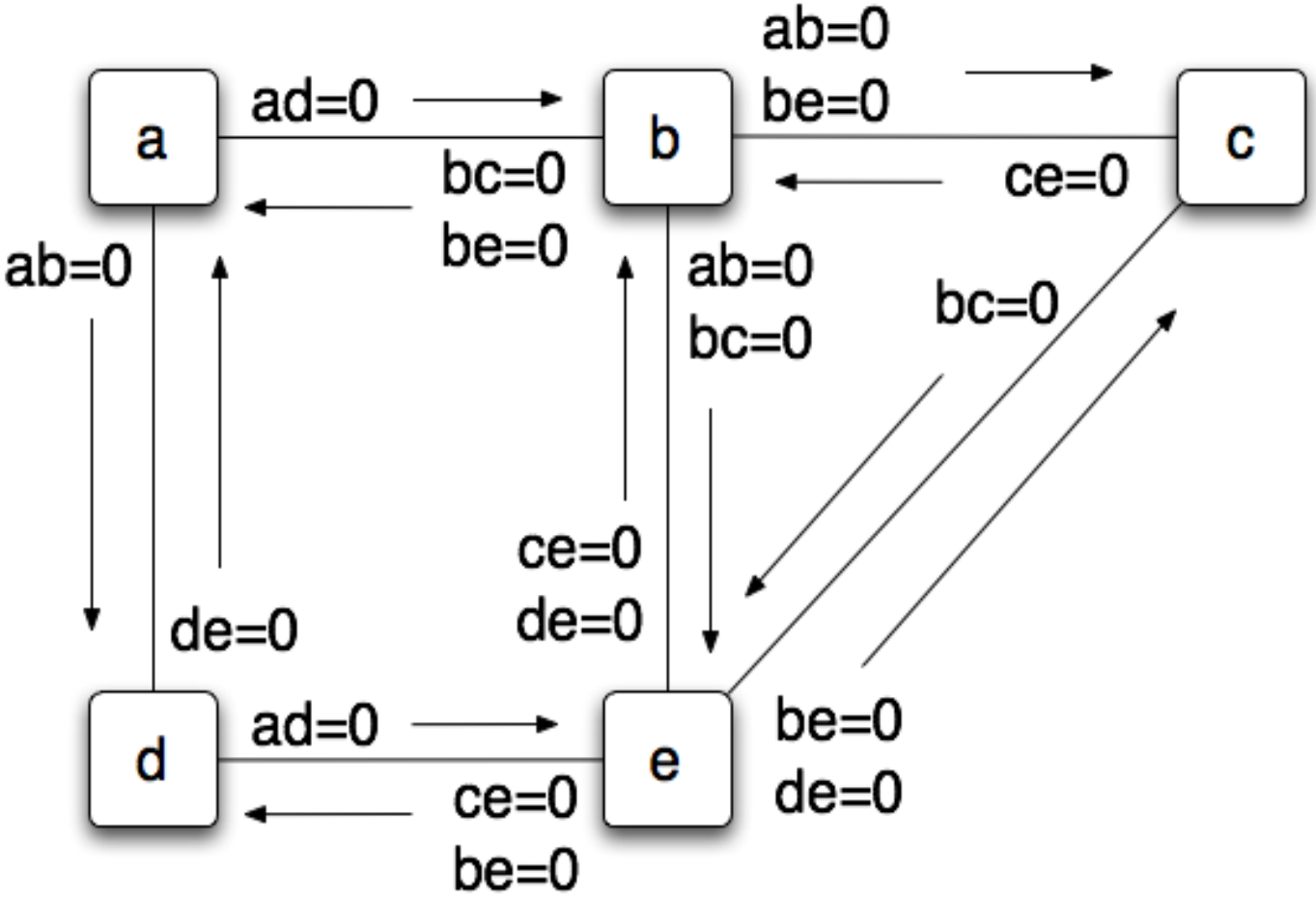
- Often known as Bellman-Ford
- Router neighbors gossip amongst themselves
  - kind of like the “telephone game”
- As announcement propagates, distance increases

DV Algorithm Summary

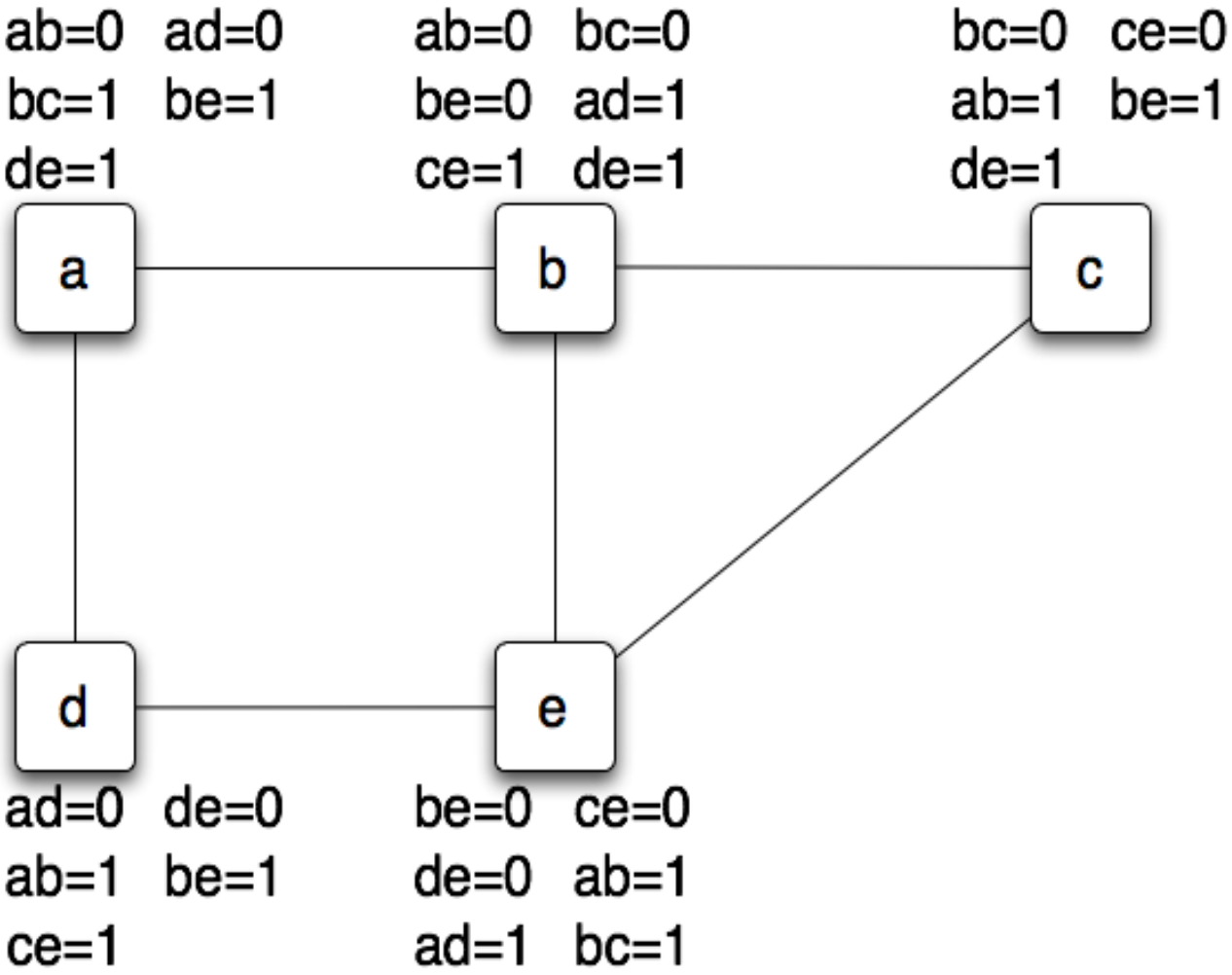
***adapted from Perlman, Interconnections**

- Each router initialized with it's own ID
- Each router link cost initialized
- Destination cost of infinity for everyone but itself
- Exchange DV (routes) known to neighbors
- Calculate DV based on min(cost) to destination
- Recalculation upon update or link failure

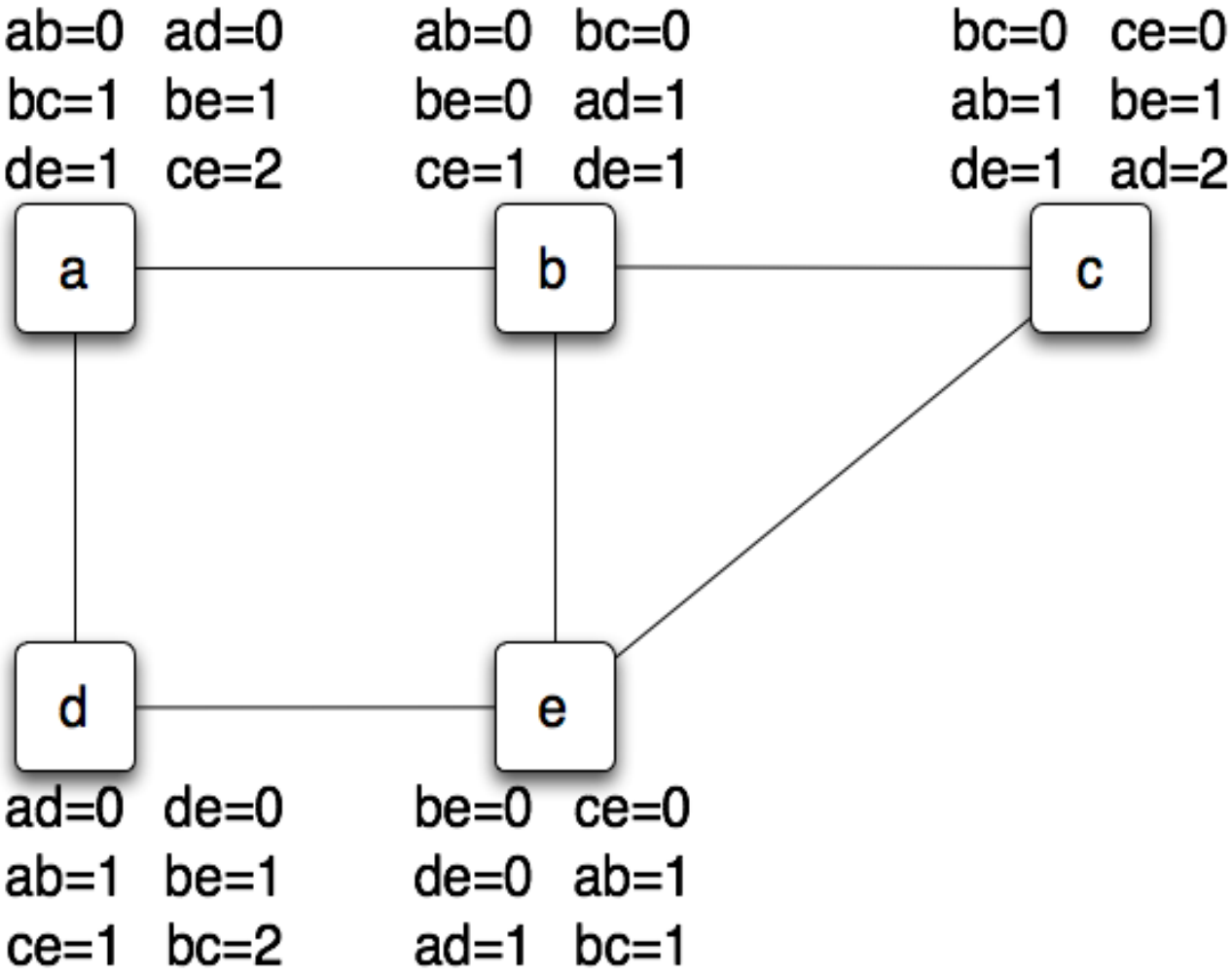
DV bootstrap 1



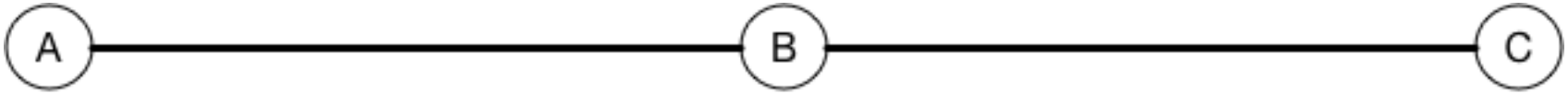
DV bootstrap 2

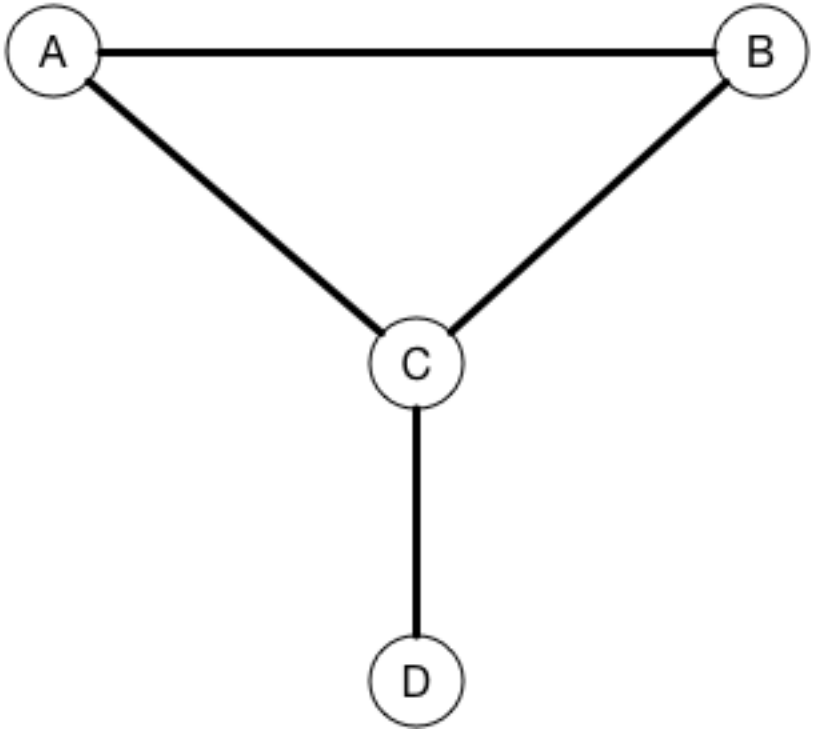


DV converged



Count to Infinity





DV Hacks and Tweaks

- Triggered updates
- Hold-down
- Reporting the entire path
- Split horizon
- Two metrics
- Poison Reverse

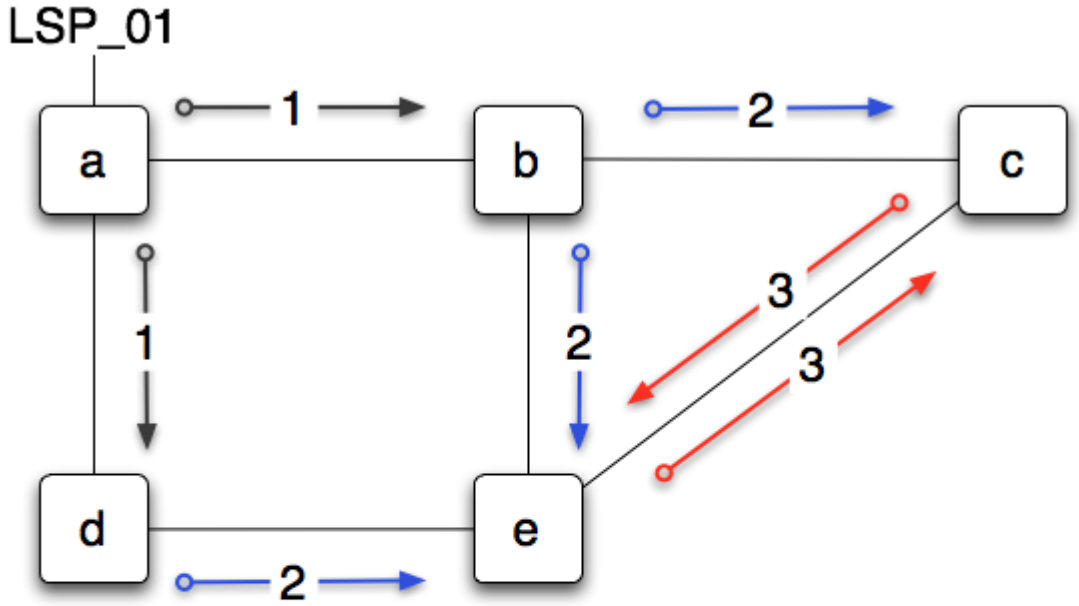
DV Summary

- Simple distance calculation determines path
- Periodic route updates sent to neighbors
- Convergence time can be slow
- Examples:
  - RIP (IP/IPX), RIPv2, RIPng, IGRP (cisco)
- In practice, a link state IGP is generally preferable

Link-state (LS) routing

- Algorithm commonly used is Dijkstra
- Routers exchange connectivity in LS packets
  - as opposed to exchanging the routing table
- A link-state packet (LSP) usually includes at least:
  - router id
  - sequence number
  - links and associated link costs
  - age

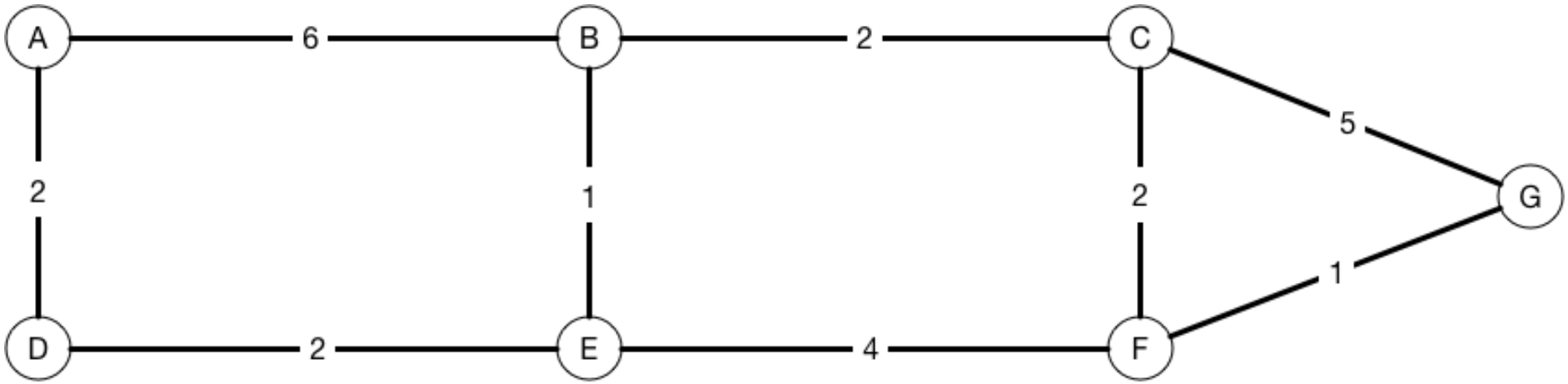
LS bootstrap



NOTE: sequence of events an example only

1) a floods LSP_01 to b and d
2) b floods LSP_01 to c and e
2) d floods LSP_01 to e, e ignores duplicate LSP
3) c or e flood LSP_01 to the other, whoever is faster

Route Cacluation with LS



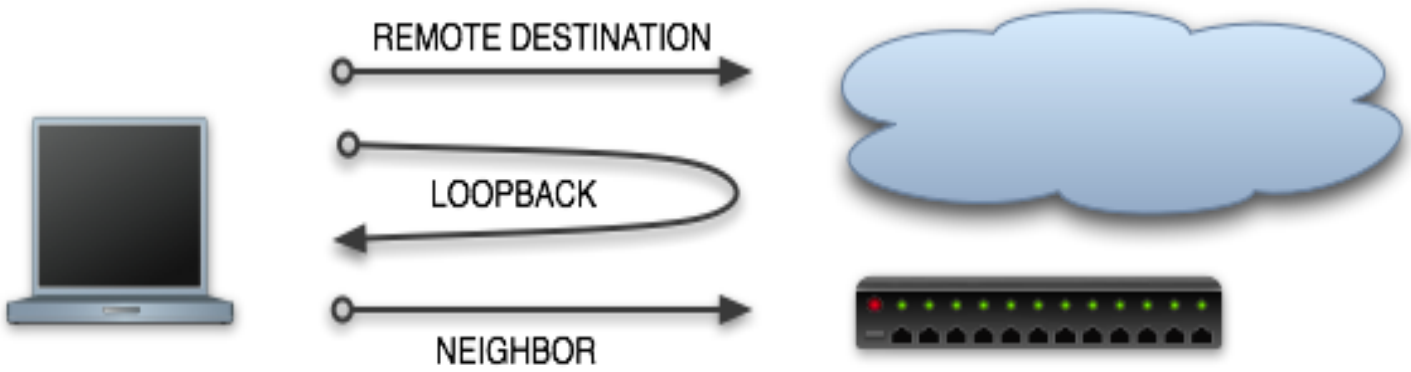
LS summary

- Each router builds their own map from LSPs
- Good convergence time
- Good loop avoidance
- Can be more complex and resource intensive
  - not really an issue these days in practice
- Generally preferable over distance vector
- Imagine cryptographically signed LSPs

Do all (IP) hosts route? Yes.

Most hosts make one of three routing decisions:

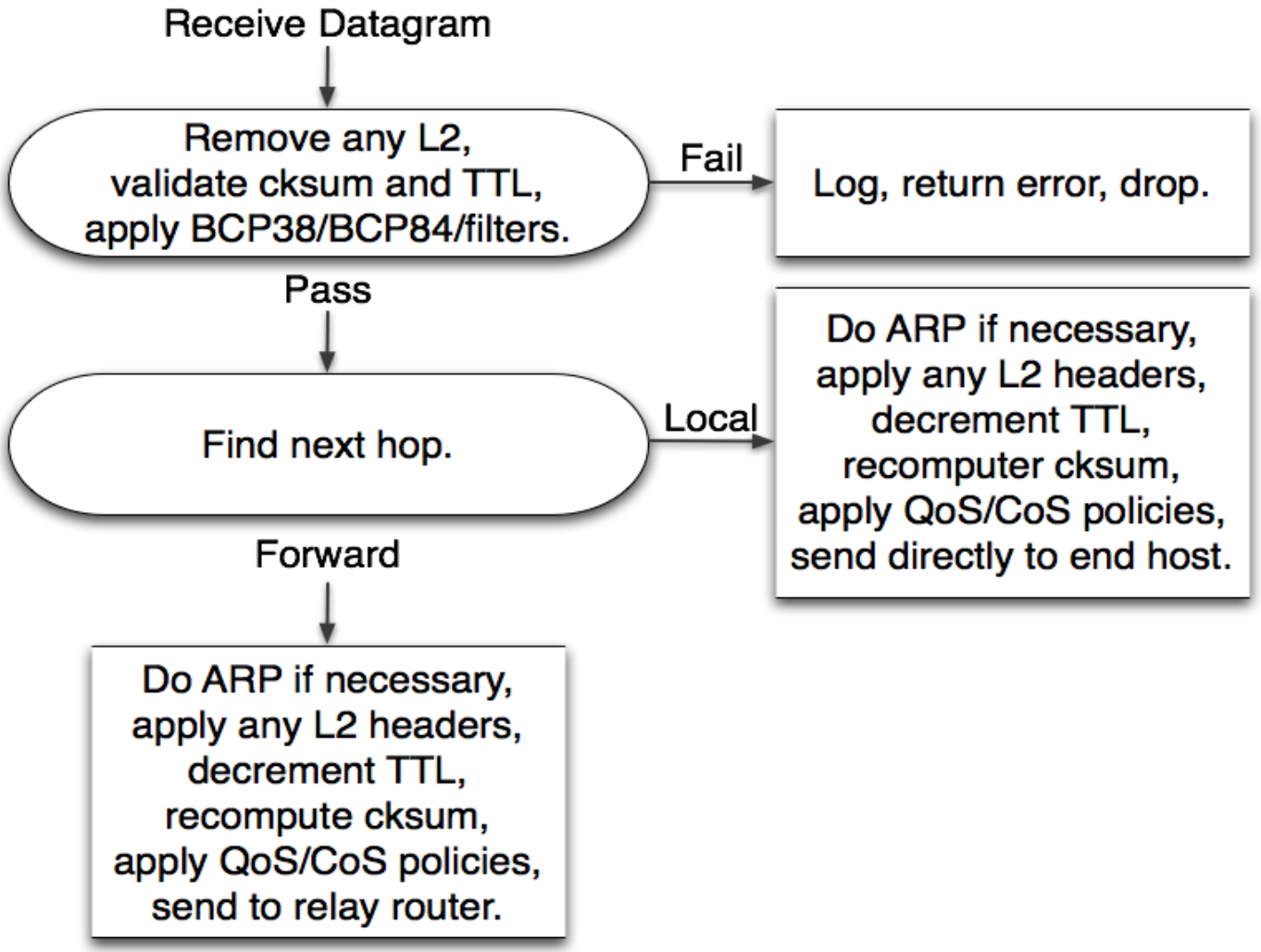
1) send packet to another via a relay
2) send packet to itself
3) send packet to a directly attached neighbor



Your end host != router

- Need to know your address, network and gateway
  - not so much a “routing system” process
  - this is your host's bootstrap challenge
- We don't tend to think of end hosts as routers
- How do they differ then?
  - network / interface attachments
  - distributed routing algorithms
  - forwarding packets on another's behalf

Real routers work more like this



IP's Best Match Forwarding

- Forward packet via the "most specific" route
- Most specific to least specific (IPv4 example):
  - host (/32) route, /31, /30, /29, ... default (/0)
- If no route, drop and return ICMP error to source

Routers as signposts

Key IPv4 field for routing: TTL

- More apt name today would be hop limit
  - in fact, that is just what it is called in IPv6 now
- This field prevents packets looping forever
- Other uses are secondary to this
  - traceroute
  - source OS fingerprint and distance detection
  - BGP peering hack (aka GTSM, RFC 3682)

Key IP field for routing: Destination Address

- Consists of both a...
  - host/interface identifier (usually unique) and
  - a network identifier (also usually unique)
- Combined, the daddr helps hosts and routers
  - get the packet to the correct network
  - and to the specific host on the correct network

A campus LS database

```
jtk> show ospf database

    OSPF database, Area 0.0.0.0
 Type      ID          Adv Rtr          Seq      Age  Opt  Cksum  Len
Router  192.0.2.7    192.0.2.7      0x8000ee56  2103  0x22 0x3ec6 876
Router  192.0.2.8    192.0.2.8      0x8000e729  2093  0x22 0x9cf6 1944
Router  192.0.2.10   192.0.2.10     0x80009d3a  1991  0x22 0x81b  1380
Router  192.0.2.12   192.0.2.12     0x80008467   930  0x22 0x7cd5 360
Router  192.0.2.13   192.0.2.13     0x80002042   262  0x22 0x8ab7 684
Router  192.0.2.14   192.0.2.14     0x80005f7a   516  0x22 0xe6ff  96
Router  192.0.2.15   192.0.2.15     0x8000a0a4   983  0x22 0x9e79  48

[...]
```

An example routing table

```
route-views.oregon-ix.net>show ip route
Codes: C - connected, S - static, R - RIP, M - mobile, B – BGP
       D - EIGRP, EX - EIGRP external, O – OSPF
       IA - OSPF inter area,  N1 - OSPF NSSA external type 1
       N2 - OSPF NSSA external type 2, E1 - OSPF external type 1
       E2 - OSPF external type 2, E - EGP i – IS-IS
       su - IS-IS summary, L1 - IS-IS level-1
       L2 - IS-IS level-2, ia - IS-IS inter area
       * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route

Gateway of last resort is 128.223.51.1 to network 0.0.0.0

B    216.221.5.0/24 [20/489] via 208.51.134.254, 18:06:49
B    210.51.225.0/24 [20/0] via 12.0.1.63, 18:07:52
B    210.17.195.0/24 [20/0] via 216.218.252.164, 18:08:11
B    209.136.89.0/24 [20/0] via 216.218.252.164, 18:08:21
B    209.34.243.0/24 [20/0] via 157.130.10.233, 17:59:49
B    205.204.1.0/24 [20/0] via 157.130.10.233, 18:00:57
B    204.255.51.0/24 [20/0] via 157.130.10.233, 17:59:44
B    204.238.34.0/24 [20/0] via 157.130.10.233, 18:00:28
```

Want router access?

- Telnet to route-views.routeviews.org
- Browse to http://routerproxy.grnoc.iu.edu/
- Go easy, don't ruin it for the rest of us please
  - notwithstanding potential bugs or attacks, by default access it intended to be limited (sorry, no “enable”), but they can still be **very** helpful for remote analysis and troubleshooting

You do have enable, kind of

- On Unix, Linux, Mac OS X
  - `netstat -arn`
- On Microsoft Windows
  - `route print`

Protocol encapsulation

- RIP uses UDP port 520 via IP broadcast/multicast
- IS-IS runs directly over layer 2 multicast
- OSPF is IP protocol 89, via IP multicast
- BGP uses TCP port 179, unicast

Router Security vs Route Security

- Router security
  - authentication, filtering, crypto... DONE!
  - uhm, no
- Route security
  - this is the old, “my security, depends on your ability to do security” problem
  - say you have and announce a /16
  - someone announces /24's in that /16.
  - uh-oh

Example Implementations

- Cisco IOS
- Juniper JunOS
- Zebra and derivatives
  - Quagga, Vyatta
- BIRD Internet Routing Daemon
- OpenBGPD
- MikroTik RouterOS

Stay Tuned For...

- BGP
- Policies and BGP peering
- RTBH techniques
- SAV and uRPF
- NetFlow
- flow-spec
- BGPSEC / RPKI