

Distance Vector vs. Link State

- **Messages per router**

- DV: $O(d)$ where d is # of neighbors of node
- LS: $O(n \cdot d)$ for n nodes in system
- Sizes?

- **Computation**

- DV: Could count all the way to ∞ if loop
- LS: $O(n^2)$

- **Robustness from malfunctioning router?**

- DV: Node can advertise incorrect *path* cost
- DV: Costs used by others, errors propagate through net
- LS: Node can advertise incorrect *link* cost

Route Propagation

- **Know a smarter router**
 - hosts know local router
 - local routers know site routers
 - site routers know core router
 - core routers know everything
- **Introduce notion of *Autonomous System (AS)***
- **Two-level route propagation hierarchy**
 - interior gateway protocol (each AS selects its own)
 - exterior gateway protocol (Internet-wide standard)

Autonomous systems

- **Corresponds to an administrative domain**
 - Internet is not a single network
 - ASes reflect organization of the Internet
 - *E.g., Brown, IBM, Sprint etc.*
- **Goals:**
 - ASes want to choose their own local routing algorithm
 - ASes want to set policies about non-local routing
- **Each AS assigned unique 16-bit number**

Types of AS

- *Local traffic* – packets with src or dst in local AS
- *Transit traffic* – passes through an AS
- *Stub AS*
 - Connects to only a single other AS
- *Multihomed AS*
 - Connects to multiple ASes
 - Carries no transit traffic
- *Transit AS*
 - Connects to multiple ASes and carries transit traffic

EGP: Exterior Gateway Protocol

- **Overview**

- designed for tree-structured Internet
- concerned with reachability, not optimal routes

- **Protocol messages**

- neighbor acquisition: one router requests that another be its peer; peers exchange reachability information
- neighbor reachability: one router periodically tests if another is still reachable; exchange HELLO/ACK messages; uses a k -out-of- n rule
- routing updates: peers periodically exchange their routing tables (distance-vector)

Today: BGP-4

- **Goal: Share connectivity information across ASes**
 - Don't strive for "optimal" routes—too hard.
 - Different ASes may have different notions of cost.
 - *Policies* may dictate suboptimal routes.
- **BGP used by two types of routers:**
 - *edge* routers, connecting organization to world
 - *core* routers, making up backbone
- **Within ASes, can use any routing protocol**
 - But backbones too big for RIP or OSPF
 - So *internal* BGP (IBGP) variant for use within ASes

Choice of Routing Algorithm

- **Constraints:**

- Scaling
- Autonomy (policy and privacy)

- **Link-state?**

- Requires sharing of complete network information
- Information exchanges don't scale
- Can't express policy (or must share widely)

- **Distance Vector?**

- Scales and retains privacy
- Can't implement policy
- Can't avoid loops if shortest paths not taken

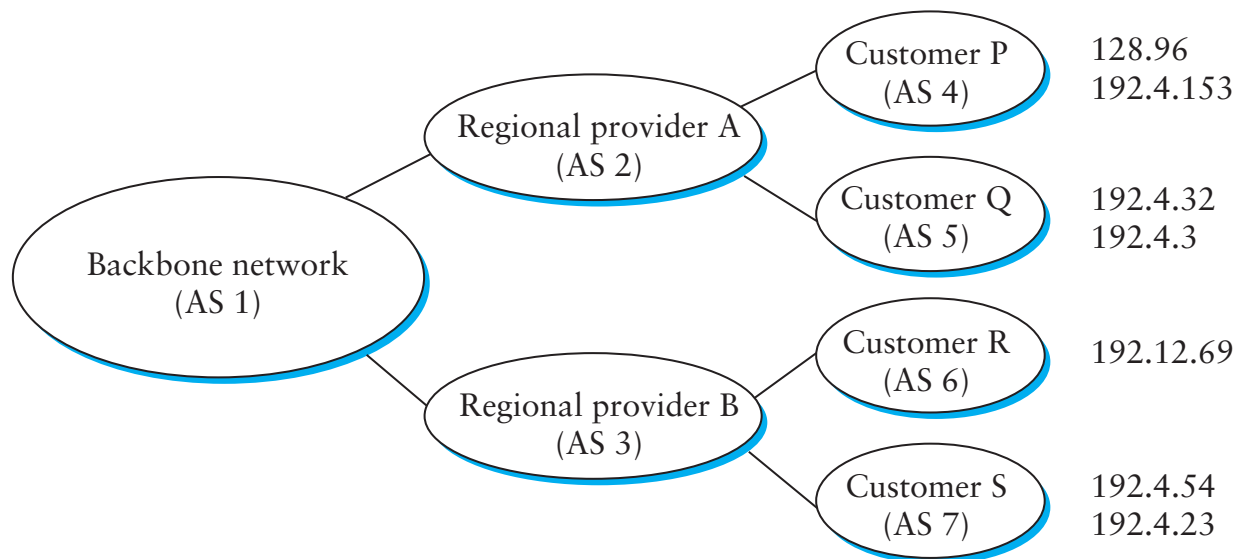
Path Vector Protocol

- **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
- **In addition, separate *speaker* & *gateway* roles**
 - *speaker* talks BGP protocol to other ASes
 - *gateways* are routers that border other ASes
 - Can have more gateways than speakers
 - Speaker can reach gateways over local network

BGP Example

- **Speaker for AS2 advertises reachability to P and Q**

- network 128.96, 192.4.153, 192.4.32, and 192.4.3, can be reached directly from AS2



- **Speaker for backbone advertises**

- networks 128.96, 192.4.153, 192.4.32, and 192.4.3 can be reached along the path (AS1, AS2).

- **Speaker can cancel previously advertised paths**

Basic BGP Messages

- **Open:**
 - Establishes BGP session (uses TCP port #179)
- **Notification:**
 - Report unusual conditions (message header error, ...)
- **Update:**
 - Inform neighbor of new routes that become active
 - Inform neighbor of old routes that become inactive
- **Keepalive:**
 - Inform neighbor that connection is still viable

Attributes of BGP routes

- **AS path**
- **Origin**
 - Who originated the announcement?
 - IGP, EGP, or “incomplete” (for static routes)
- **Multi-Exit Discriminator (MED)**
 - How close prefix is to link it is announced on
 - Used if ASes *A* & *B* connect at multiple points
- **Local preference**
 - Used in IBGP to select (or give preference to) a particular exit for a particular prefix

Multicast - Sending message to many

- **Internet radio**
- **Stock quote information**
- **Internet multi-way chat / video conferencing**
- **Multi-player games**

What's wrong with sending data to each recipient? (link *stress*)

A Multicast *service model*

- **Receivers join a multicast group G**
- **Senders send packets to address G**
- **Network routes and delivers packets to all members of G**
- **Class D addresses (start 1110) – 224-239.x.x.x**

LAN Multicast

- Easy on a shared medium.
- Ethernet multicast address range
00:00:5e:[0-7]x:xx:xx
- Set low 23-bits of Ethernet address to low bits of IP address

What about the Internet?

Use trees to scale to limit stress

- Each recipient forwards data to several others
- What about receivers with little upstream b/w?
- Messages travel over more hops for many recipients
- Added latency is known as *stretch*

Eliminating stretch

- Optimize topology of your tree. How?
- Make *routers* form the branch points.
- This is what IP Multicast does.

Source Specific vs Shared Trees

- **Source-specific trees — best tree for each source**
- **Shared trees — single spanning tree over recipient**
- **Hard to find one shared tree that's best for many senders**
- **State in routers much larger for source-specific**

Multicast Routing: LS

- Each host on a LAN periodically announces the groups it belongs to using IGMP
- Augment update message (LSP) to include set of groups that have members on a particular LAN
- Each router uses Dijkstra's algorithm to compute shortest-path spanning tree for each source/group pair
- Each router caches tree for currently active source/group pairs

Multicast Routing: DV

- **Reverse Path Broadcast**
 - Each router already knows that shortest path to S goes through router N
 - When receive multicast packet from S, forward on all outgoing links (except one it arrived on), iff packet arrived from N
 - Eliminate duplicate broadcast packets by letting only “parent” for LAN (relative to S) forward
- **shortest path to S (learn from distance vector)**
- **smallest address to break ties**

DV (cont)

- **Reverse Path Multicast**
- **Goal: prune networks that have no hosts in group G**
- **Step 1: determine if LAN is a leaf w/ no members in G**
 - leaf if parent is only router on the LAN
 - determine if any hosts are members of G using IGMP
- **Step 2: propagate “no members of G here” information**
 - augment (destination, cost) update sent to neighbors with set of groups for which this network is interested in receiving multicast packets
 - only happens when multicast address becomes active

PIM-SM: Internet Scale

- **Protocol Independent Multicast — Not linked to DV/LS**
- **Sparse Mode – Scales well for small groups**
- **Name an explicit Rendezvous Point (RP)**
- **Send Join message (*,G) to the RP**
- **Routers note the join.**
- **Source specific Joins (S,G) when needed**

IP Multicast Problems

IP Multicast is not widely deployed.

- **Address space**
- **Economic**
- **Lack of “killer app”**
- **Unreliable**
- **Poor building block**

MBone

- Connect Multicast enabled regions with *tunnels*.
- Encapsulate multicast packets in unicast.
- An overlay network

Coming up

- **Today: 4:30pm Interview Panel (Cisco, Google, VMWare)**
- **Today: 6:15pm Cisco Tech Presentation [B&H]**
- **Tue: Wireless Routing Protocols**
- **Tue, 4:30pm: Apple recruiting event [Lubrano]**
- **TCP!**