The University Of Alabama in Huntsville Computer Science

cs670

Routing Algorithm review

The University Of Alabama in Huntsville Computer Science

cs670

Short review of routing algorithm types

- There are two general classes of routing algorithms
  - Distance Vector (RIP1, RIP2, BGP*)
  - Link State (OSPF, IS-IS)

*BGP uses a variant of the DV algorithm called the Path Vector algorithm

The University Of Alabama in Huntsville Computer Science

cs670

Distance Vector Routing

- The idea:
  - Each router knows the cost to each of its immediate neighbors
  - Each router builds a "distance vector" that contains the total cost of the best-known route to every destination (initial costs = ∞)
  - At intervals, each router sends its DV to all neighbors
  - When a router *R* receives a DV from a neighbor *N*, *R* scans the table to see if there are any cases where, for Destination *D*:

    N*'s cost to get to *D* + *R*'s cost to get to *N* < *R*'s current cost to get to *D

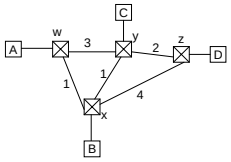
    If there are any such cases, R updates its table so that future traffic for D is sent to N.

Example:
- Assume router R knows a path to D with a cost of 25.
- R's neighbor N knows a path to D with a cost of 20.
- If the cost from R to N is 3, then R can get to D through N with a total cost of 23.
- Since this is less than the current cost, R will update its routing table so that all traffic for D goes to N.

The University Of Alabama in Huntsville Computer Science

cs670

A Distance Vector example



- The tables below show each router's DV after all routers have exchanged DVs.
- Bold entries show changes since last exchange

Initial

	A	B	C	D
w	**0-**	**99-**	**99-**	**99-**
y	**99-**	**99-**	**0-**	**99-**
z	**99-**	**99-**	**99-**	**0-**
x	**99-**	**0-**	**99-**	**99-**

After first exchange

	A	B	C	D
w	0-	**1x**	**3y**	99-
y	**3w**	**1x**	0-	**2z**
z	99-	**4x**	**2y**	0-
x	**1w**	0-	**1y**	**4z**

"Router x knows a path to A through router w with cost 1."

After second exchange

	A	B	C	D
w	0-	1x	3y	**5y**
y	**2x**	1x	0-	2z
z	**5y**	**3y**	2y	0-
x	1w	0-	1y	**3y**

After third exchange

	A	B	C	D
w	0-	1x	**2x**	**4x**
y	2x	1x	0-	2z
z	**4y**	3y	2y	0-
x	1w	0-	1y	3y

The Link State routing algorithm

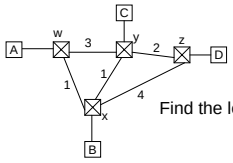
- The idea:
  - Each router discovers cost to immediate neighbors
  - At intervals, this info is flooded to all other routers in a "Link State Packet". This gives all routers a map of the network and link costs.
  - Each router runs a part-finding algorithm (e.g, Dijkstra's Shortest Path Algorithm) to calculate least-cost paths.

Dijkstra's Shortest Path Algorithm

- To find the least-cost path from node S to node D:

1. Call S the "working node". Tentatively label all other nodes (∞, -).
2. Tentatively label each neighbor of the working node (W, x) where W is the name of the working node and x is the total cost from the source node.
3. Examine the entire graph and find the tentatively-labeled node with the smallest cost in its label. Change this to the permanent label of the node. This is the new working node.
4. Repeat 2-4 until D is permanently labeled
5. Record the name of D. Call D the copying node.
6. Record the name W, where (W,x) is the label of the copying node.
7. W is the new copying node. If W <> S, repeat 5-6.
8. The least-cost path is the reverse order of the recorded node names.

Dijkstra's Shortest Path Example



Find the least-cost path from A (w) to D (z)

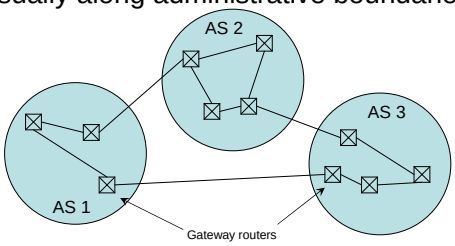
	w	x	y	z	
1	**(0 A)**	(inf -)	(inf -)	(inf -)	w is the starting node
2	**(0 A)**	**(1 w)**	(3 w)	(inf -)	Calc total costs to x and y. x is low cost, so it is new working node.
3	**(0 A)**	**(1 w)**	**(2 x)**	(5 x)	The path to y through x is lower cost, so we replace y's label.
4	**(0 A)**	**(1 w)**	**(2 x)**	**(4 y)**	z is permanently labeled, so stop

Recorded labels are: z y x w, so the path is w x y z

Border Gateway Protocol (BGP)

Autonomous Systems (AS)

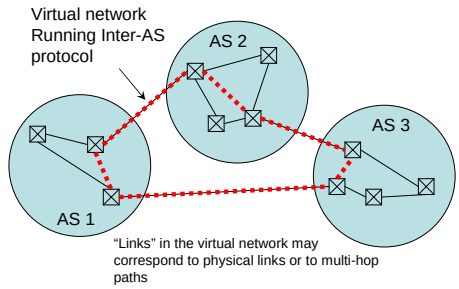
- aka “Routing Domains”
- Large networks are divided into AS’s, usually along administrative boundaries



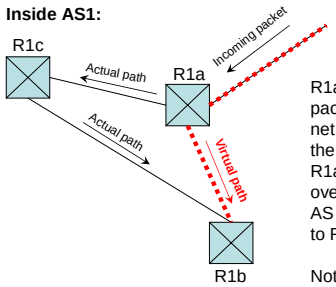
AS concepts

- Within an AS, all routers run the same routing protocol and share the same information
- Gateway routers (“Boundary routers”):
  - Run their local AS’s protocol and,
  - Run a separate routing protocol to route between the Gateway routers (“Inter-AS Routing protocol”)

Inter-AS routing



How routing packets flow over the virtual Inter-AS network



R1a receives a routing packet on the Inter-AS net. The next node on the virtual net is R1b. R1a hands the packet over to the AS1 Intra-AS protocols to send it to R1b.

Note that routers outside of AS1 require no knowledge of AS1’s internals

BGP

- BGPv4 is the standard Inter-AS routing protocol for the Internet
- RFC 1771 (+ see 1772, 1773)
- A BGP route is based on reachability, not (necessarily) cost

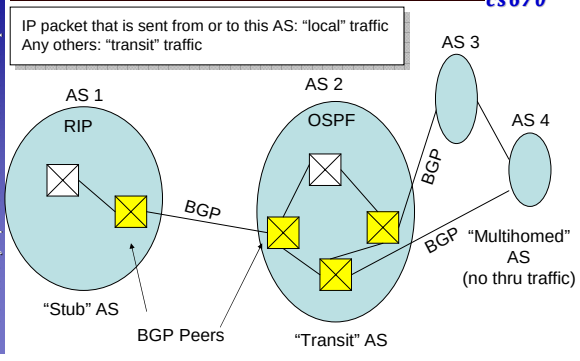
BGP protocol

- Based on DV concepts
- A "Path Vector" protocol
  - Routers advertise complete routes to destination AS's, not costs (normally)
  - Being able to control which AS your route goes through is very important to the government
    (ensure traffic from Vandenberg to the White House is not routed through another country)

Path vector problems?

- Resolves the Count-to-Infinity problem with Split Horizon algorithm (a router receiving a path vector can check to see if it is on the path)

Terminology



Speakers and gateways

- Admin of each AS declares one or more BGP routers to be "BGP speaker:
  - Establish BGP sessions to other AS's
  - Determine routes to advertise
- "Gateways" are BGP routers through which traffic enters and leaves the AS
  - Speaker is not necessarily a gateway

Routing

- BGP routers exchange route info
  - Inter-AS – exchange done directly
  - Intra-AS – exchange via internal AS paths
- Note that BGP routes to networks, not individual destinations or routers

Route Advertisement

- Sent between BGP routers
- Only when something changes
- Consists of:
  - Network address in CIDR format (e.g. 127.04.114/24)
  - Attributes (we'll talk about those later)

BGP route advertising

- Administrator is free to set own policies for advertising routes
  - For example:
    - Particular AS's preferred
    - Particular AS's avoided

Processing route advertisements

- Advert treated as a "contract" from the peer router to forward traffic to the ntwk
- BGP router can ignore adverts:
  - If own AS number is in the path
  - If forbidden AS is in the path
  - etc.

Choosing between paths

- A BGP router may receive several paths to a single destination, but uses only one
- Major selection criteria:
  - Preferences
    - set locally by administrator (metric, filters)
    - suggested by neighboring routers (local_pref, MED)
  - Minimum number of AS's crossed
  - Route Filters
    - Prefix List, AS-Path, Community
- Routes not selcetd are saved for backup

Observations

- Other AS's have to trust that a speaker will advertise the best route (and "best" may mean something different to each)
- A speaker does not have to advertise a route, even if it knows one
  - Can refuse to provide transit
  - Can try to block access to other AS's

How are BGP messages transmitted?

- TCP
- Port 179

BGP message types

- OPEN
  - Establishes link to a BGP peer, Authenticates
- UPDATE
  - Contains a route being advertised (no more than one)
  - Optionally, provides the address of a network for which the sender wants to withdraw a previously-advertised route (multiple)
- KEEPALIVE
  - ACK or "I'm still alive" (sent to neighbors on a regular basis: typ 30sec)
- NOTIFICATION
  - Error or other BGP control message

Some Attributes (in UPDATE messages)

- ORIGIN: Source of the route information
- AS-PATH: List of AS's to traverse on the way to the network
- NEXT_HOP: IP addr of router to start AS_PATH
- UNREACHABLE: Previously-advertised route has become unreachable

BGP variants

- eBGP
  - The common form of BGP. What we've been talking about.
- iBGP
  - A modified "internal use" variation of BGP.
  - Used among edge routers of the same AS to synchronize external routes.

BGP AS-Path Padding

- Append the local AS multiple times upon advertisement
- "Scruffy" fix to emulate Distance-Vector
- Big corporations (WorldCom, Sprint, etc) use this in their Internet routing tables for ISPs

Ex:

Original AS Path	= 100 300 500
..with Padding	= 100 100 100 300 500

AS Path 100 now has a distance of 3.