# ROUTING, PART 3

**Internet Protocols**

*CSC / ECE 573*

*Fall, 2005*

*N. C. State University*

## Announcements

I. *No class or office hours on Tuesday, Oct. 20*
   I. *Result → short summary only of multicast*




## Today's Lecture

I. BGP-4 Basics
II. Routing Decisions
III. Interior BGP (IBGP)



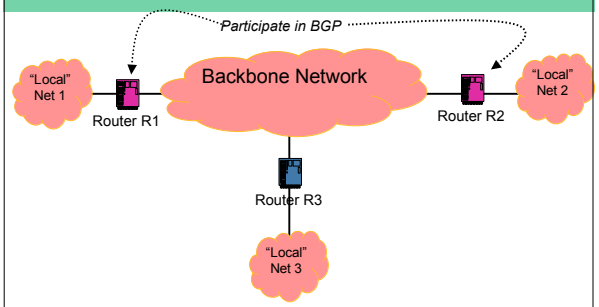

## BGP-4 BASICS

## Two Levels: Local, and Global

- Independent domains must export information to backbone networks
  - e.g., the "hidden network" problem
- The backbone must export info into the independent domain
  - e.g., the "extra hop" problem
- Needed: a router that can "speak" both interior and exterior routing protocols



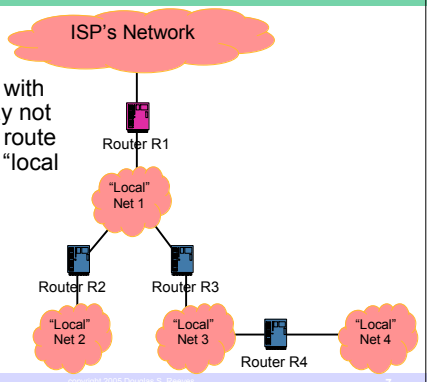

## "Extra Hops"




- Without info from the backbone network, router R3 may send traffic to router R1 to reach local network 2

## "Hidden Networks"

- Without exchanging information with R1, ISP may not know of (or route packets to) "local net 4"







## BGP-4 (RFC 1771)

- Exterior gateway (inter-domain) protocol
  - the "workhorse" protocol of the Internet
  - regards Internet as a network of ASes
- A *path-vector* (compared to distance-vector) protocol
  - collection of path information forms a route to a destination
  - routes "dynamically injected" into BGP-4 from IGPs (RIP, OSPF, ISIS)
- Supports AS-specific policies




## Traffic Types

- *Local* traffic originates and/or ends within this AS
- *Transit* traffic just passes through
- Goal (for many network operators): reduce transit traffic!




## Types of Networks

- *Stub* networks have only one connection to the BGP graph, never carry transit traffic
  - has a single exit/entry point
  - all that is needed is a default route
- *Multi-homed* networks could be used for transit traffic, but they refuse
  - advertises only its own routes, none from other ASes
- *Transit networks* will carry some "through" traffic
  - advertises to other AS'es the routes that it learned from another AS




## BGP-4 Message Types

- Smallest BGP message 19 bytes, largest 4096 bytes
- OPEN message establishes connection between BGP peers, includes AS number
- NOTIFICATION message conveys error messages
- KEEPALIVE message maintains the connection if there are no other messages being exchanged




## BGP-4 Message Types (cont'd)

- UPDATE message has reachability information, path attributes, unreachable (withdrawn) routes
  - no distance information exchanged!
- Reachability info = prefix + length (e.g., 192.168/16)
  - encoded in 1-5 bytes (1 byte length, 1-4 bytes prefix)
  - instead of 8 bytes
- Withdrawal of routes -- "bad news" can travel quickly!

## BGP-4 Communication

- BGP routers communicate using TCP (reliable delivery)
- BGP router's neighbors are configured by the network administrator, not discovered
- When routers first boot up, they contact neighbors (establish a session) and exchange full routing table information
  - network prefixes
  - path *attributes*

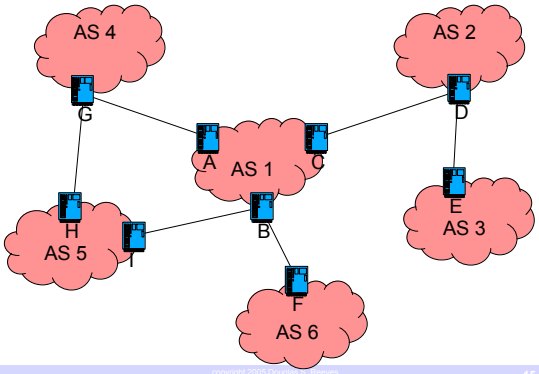



## BGP-4 Communication (cont'd)

- Periodically thereafter they only exchange updates, not full routing tables
- At termination of session, delete info from routing table learned from the other peer




## A Sample AS Configuration

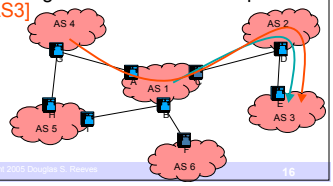






## Path Vector Protocol

- Each BGP router keeps track of the exact path to each destination
  - also includes this path in updates sent to neighbors
- Example (destination = AS 3's network prefix)
  - Router G in AS4 accepts and installs route update from router A, having AS path [AS1, AS2, AS3]
  - G transmits route to neighbor router H with AS path [AS4, AS1, AS2, AS3]

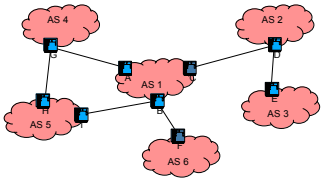






## Path Vector Protocol (cont'd)

- Advantages
  - easily eliminates loops (how?)
  - allows policy decisions to be made based on the entire set of ASes in the path







## AS-Specific Policies

- Policy restrictions expressed in terms of the ASes on the path
- Examples…
  1. "never use AS X, for any destination"
  2. "don't use AS X to get to a destination in AS Y"
  3. "don't use AS X unless it is the only possible path"
  4. "only accept an advertised route with a prefix in the range x.x.x.x-y.y.y.y if it originated with AS Z"
- Policies of different ASes may conflict!
  - e.g., AS5 prefers to use AS4 to reach AS1, and AS4 prefers to use AS5 to reach AS1

## BGP Path *Attributes*

- *Mandatory* attributes must be provided in every UPDATE message
- Optional but *transitive* attributes should be passed on to other BGP routers




## List of Attributes (Partial)

| Attribute | Category | Purpose |
|---|---|---|
| Origin | Mandatory | Identify source of a route |
| AS Path | Mandatory | Identifies the AS'es on the path to the destination |
| Next Hop | Mandatory | IP address of interface of next router in the path (may be different than the router that provided this UPDATE) |




## List of Attributes (cont'd)

| Attribute | Category | Purpose |
|---|---|---|
| Multi-Exit Discrimator | | When neighboring AS'es are connected at multiple points, identifies "best" entry point |
| Local Pref | | BGP routers advertising higher values are preferred as exits from AS (used in advertisements in IBGP) |
| Aggregator | Transitive | Identity of router which combines a collection of routes into a single aggregate |
| Community | Transitive | Identify collection of prefixes as a group for purposes of applying policies (ex.: NO_ADVERTISE) |




## ROUTING DECISIONS

## Routing Decisions

- Decision process invoked when…
  - new BGP updates received, or
  - BGP session with neighbor terminates, or
  - configured policies change




## Determining Routes

- Each router applies a scoring function to the routes received
- The scoring function is not part of BGP; it is left as a local decision
  - BGP has no globally agreed upon metric
  - allows for significant degree of autonomy in selecting routes
- Examples
  - minimize the number of ASes traversed
  - assign weights to ASes, use the maximum-weight path

## The "Best Path Algorithm" (most important to least important)

- Prefer the path with the largest WEIGHT
- Prefer the path with the largest LOCAL_PREF
- Prefer the path that was locally originated (manually configured by AS administrator on this router)
- Prefer the path with the shortest AS_PATH
- Prefer the route with the lowest origin type
- Prefer the path with the lowest MED
- .... (only a partial list)




## AT&T Generalization of Best Path

1. Highest local preference (assigned by the route import policy and conveyed to other routers by IBGP)
2. Shortest AS path
3. eBGP over iBGP (prefer routes learned externally to those learned internally, i.e., get data out of the AS as quickly as possible)
4. Lowest iBGP metric (select closest exit point)
5. Lowest router ID (break ties)




## Complications to Aggregation

- Sender should only advertise paths that traffic is "encouraged" to follow
  - administrator can control how routes are aggregated
- The “IP address portability” problem
  - i.e., ISP customer moves to a new ISP, keeps old address block
  - more routes must be installed in routing tables
- Multihoming
  - difficulty of aggregating addresses that are multihomed




## BGP Aggregation Problems

- What path gets advertised for an aggregated or summarized route?
- Ex.: AS 1 wants to advertise paths to two destinations (prefixes): 192.9.0.0/18 and 200.16.64.0/18.
- AS 1 routing table:

| Prefix | AS path |
|---|---|
| 192.9.3.0/24 | [AS2, AS3] |
| 192.9.17.0/24 | [AS2] |
| 200.16.67.0/24 | [AS4] |
| 200.16.68.0/24 | [AS4] |
| 200.16.75.0/24 | [AS5] |
| 200.16.80.0/24 | [AS2, AS3] |
| 200.16.92.0/24 | [AS2] |



## BGP Aggregation Problems (cont’d)

- Dilemma: the routes being aggregated use different paths
- Solution: advertise [required or common] and {possible} routers on the path
  - 192.9.0.0/18 advertised with path [AS1, AS2]{AS3}
  - 200.16.64.0/18 advertised with path [AS1]{AS2, AS3, AS4, AS5}



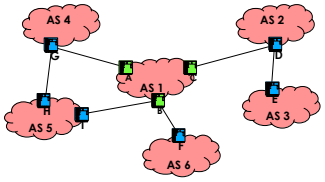

## Other Features

- Reliability
  - uses TCP for transport; reliable delivery ensured
  - since complete network topology not exchanged, router must store *alternative routes* in case of failures
  - compare with RIP: compute best route only
- Security (optional)
  - MD5 message digest in first 16 bytes (header) of BGP message
  - provides message authentication using shared key

## INTERIOR BGP (IBGP)

## IBGP

- For communicating information between routers in the same AS
  - ex.: when router C learns routing info from AS2, C must communicate to A and B so they can advertise to AS4, AS6, and AS5

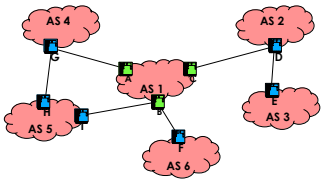






## IBGP (cont'd)

- Tunnel through which BGP info flows between boundary routers of an AS
- Info needed to determine best entry and exit points for a non-transit network







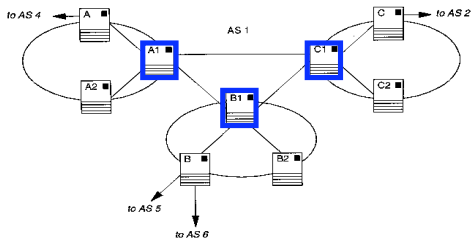
## IBGP (cont'd)

- Problem: does every BGP router have to peer with every other interior neighbor?
  - ensures there will be no routing loops
  - would require *full-mesh* connectivity of interior neighbors




## IBGP: Using Hierarchy

- AS divided into *regions*
- One router in each region designated a *Route Reflector*






## IBGP: Using Hierarchy (cont'd)

- Reflectors for a region peer with…
  - other reflectors for this AS (in a full mesh)
  - non-reflector routers in its region
  - (but non-reflectors peer only with the reflector for their region)
- Updates received by a reflector…
  - from router in its region are "reflected" to other routers in the region, and to other reflectors
  - from another reflector are forwarded to all routers in its region

## Summary

1. BGP-4 is universally used for intra-AS routing
2. Each AS can have its own policies, configure routes its own way
3. BGP-4 is a “path-vector” routing protocol
4. IBGP propagates routing information between the boundary routers of an AS
   – may use hierarchy to reduce communication




## Next Lecture

- Multicasting and IGMPv3