Routing Protocols ---
Exterior Gateway Protocol

Linda Wu
(CMPT 471 • 2003-3)

Content
- Limiting router interaction
- Autonomous system
- BGP protocol
- BGP messages
- Other issues on BGP

Reference: chapter 15

Limiting Router Interaction
- Two issues in router interaction
  - Delay
  - Overhead
- Deciding router numbers
  - A simple general heuristic
    - It is safe to allow up to a dozen routers to participate in a single routing protocol across a wide area network; approximately five times as many can safely participate across a set of local area network
  - Traffic monitor
    - Observe network trend over long period
    - Determine whether too many routers participate in a single routing protocol

Limiting Router Interaction (cont.)
- Consequences of limiting router number
  - Only a small group of routers participate in a routing protocol
  - Some routers will be outside of the group
  - Non-participating router can only make a group member as a default route
    - Routing will be suboptimal \(\rightarrow\) extra hop problem
    - A mechanism is needed that allows non-participating routers to learn routes from participating routers
Limiting Router Interaction (cont.)

- R1 & R2: Participating router
- R3: Non-participating router, with R1 as its default

Extra hop problem occurs when:
- R3 sends datagram destined for net 2

Limiting Router Interaction (cont.)

- Hidden network

- Net 4 was just installed
- R1 does not know about Net 4
- From the viewpoint of backbone network, Net 4 is hidden behind Net 1

Autonomous System

- Autonomous system (AS)
  - Autonomous system is a group of networks and routers controlled by a single administrative authority
  - Interior routing: routing inside an AS
    - Routers within an AS are free to choose their own routing protocol
  - Exterior routing: routing between ASs
    - Only one exterior routing protocol is chosen to handle routing between ASs
Autonomous System (cont.)
- Routing protocols
  - Exterior Gateway Protocol (EGP)
    Protocol used to pass routing information between two autonomous systems
    - BGP (Border Gateway Protocol)
  - Interior Gateway Protocol (IGP)
    Protocol used by interior routers to exchange network reachability and routing information
    - RIP (Routing Information Protocol)
    - OSPF (Open SPF)
    - HELLO

Autonomous System (cont.)
- From a core to independent ASs
  - Each AS must advertise its network to other ASs
  - Each AS is assigned an autonomous system number (AS#) by a central authority
  - When routers exchange information, the messages carry the AS# of the system each router represents

BGP Protocol
- Border Gateway Protocol
  - Currently used in most TCP/IP internets
  - Current version: BGP-4
  - Based on path vector routing method
  - When 2 ASs agree to exchange routing information
    - Each designates a router that will speak BGP on its behalf: BGP speaker
    - The selected router is usually near the "edge" of the AS: border router / gateway
    - The two routers are said to be BGP peer of one another

BGP Protocol (cont.)
- BGP characteristics
  - Inter-autonomous system communication
  - Coordination among multiple BGP speakers
  - Propagation of reachability information
  - Next-hop paradigm
  - Policy support
  - Reliable transport
  - Path information
  - Incremental updates
  - Support for classless addressing
  - Route aggregation
  - Authentication
**BGP Protocol (cont.)**

- **Path vector routing**
  - Each entry in the routing table contains:
    - Destination network
    - Next hop
    - The path to destination: a list of autonomous systems that a packet should travel through to reach the destination

<table>
<thead>
<tr>
<th>Destination</th>
<th>Next hop</th>
<th>Path</th>
</tr>
</thead>
<tbody>
<tr>
<td>Net 1</td>
<td>R1</td>
<td>AS14, AS20, AS67</td>
</tr>
<tr>
<td>Net 2</td>
<td>R5</td>
<td>AS22, AS67, AS03, AS80</td>
</tr>
<tr>
<td>Net 3</td>
<td>R6</td>
<td>AS33, AS12, AS19, AS04</td>
</tr>
<tr>
<td>Net 4</td>
<td>R3</td>
<td>AS16, AS02, AS06</td>
</tr>
</tbody>
</table>

- **Path vector message**
  - When a router receives a path vector message: (destination, next hop, path)
    - It verifies that the advertised path is in agreement with its policy
    - If it is, the router updates its routing table and modifies the message:
      - Adding its AS# to the path
      - Replace next-hop field with its own identification
    - Send the message to the next router

- **Loop prevention**
  - When a router receives a path vector message, it checks whether its AS# is in the path list; if it is, looping is involved and the message is ignored

- **Policy routing**
  - When a router receives a message, it checks the path
  - If one of the ASs in the path is against its policy
    - Ignores the path and destination
    - Does not update its routing table
    - Does not send the message to its neighbors
BGP Protocol (cont.)

- Path attributes
  - Well-known attribute: must be recognized by every BGP router
    - Well-known mandatory attribute: must appear in every update message
      - e.g., ORIGIN, AS_PATH, NEXT_HOP
    - Well-known discretionary attribute: not required to be included every update message
  - Optional attribute: need not be recognized by every BGP router
    - Optional transitive attribute: must be passed to the next router if the current router has not implemented this attribute
    - Optional nontransitive attribute: should be discarded if not implemented by current router

BGP Protocol (cont.)

- BGP message types

<table>
<thead>
<tr>
<th>Type Code</th>
<th>Message Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>OPEN</td>
<td>Initialize communication</td>
</tr>
<tr>
<td>2</td>
<td>UPDATE</td>
<td>Advertise or withdraw routes</td>
</tr>
<tr>
<td>3</td>
<td>NOTIFICATION</td>
<td>Response to an incorrect message</td>
</tr>
<tr>
<td>4</td>
<td>KEEPALIVE</td>
<td>Actively test peer connectivity</td>
</tr>
</tbody>
</table>

BGP Protocol (cont.)

- How BGP works
  - BGP connection establishing
    - Routers wait for BGP connection on the well-known TCP port 179
    - One BGP router opens TCP connection toward the port on its peer router
    - Each peer sends an OPEN message to negotiate the parameters for the connection
    - Each peer sends a KEEPALIVE message as response to OPEN message to confirm the connection

BGP Protocol (cont.)

- BGP information exchanging
  - Two peers use UPDATE messages to exchange routing information
  - Routers adjust their routing tables based on routes learned via the connection
  - Initial exchange
    - Full exchange of reachability information between BGP routers (i.e., the whole routing table is exchanged)
  - Incremental update
    - Only changes are reported
    - Advantage: conserve bandwidth and processing power
BGP Protocol (cont.)

- BGP connection maintaining
  - BGP peers periodically exchange KEEPALIVE messages to test peer connectivity

- BGP connection closing
  - One peer router sends NOTIFICATION message to terminate the connection

BGP Messages

- BGP message header
  - All BGP messages share the same common header to identify message type
  - Length
    - Total message length in bytes
  - Type
    - Message type (1 ~ 4)

<table>
<thead>
<tr>
<th>0</th>
<th>16</th>
<th>24</th>
<th>31</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marker</td>
<td>(16 bytes)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

BGP Messages (cont.)

- Marker
  - A value that both sides agree to use to mark the beginning of a message
  - In the initial message, the marker consists of all 1s
  - If the peers agree to use an authentication mechanism, the marker contains authentication information
    - MD5: message digest algorithm version 5
  - The receiver is required to verify that the marker value is intact before processing the message

BGP Messages (cont.)

- OPEN message

<table>
<thead>
<tr>
<th>0</th>
<th>8</th>
<th>16</th>
<th>31</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common header</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19 bytes, Type = 1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Version</td>
<td></td>
</tr>
<tr>
<td>AS#</td>
<td>Hold Time</td>
<td></td>
</tr>
<tr>
<td>BGP Identifier</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parm.Len</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optional Parameters (variable)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
BGP Messages (cont.)

- Version
  - The version of BGP; current version: 4
- AS#
  - Autonomous system number of the sender's system
- Hold Time
  - The maximum time that the receiver should wait for a message from the sender
  - The timer is reset each time a message arrives
  - If timer expires, the receiver assumes the sender is no longer available

BGP Messages (cont.)

- BGP Identifier
  - To uniquely identify the sender
  - One of the IP addresses of the router
  - If a router has several peers, it must use the same identifier in all communication
- Parm.Len
  - Length of optional parameters in bytes
  - If no optional parameters, Parm.Len = 0
- Optional Parameters
  - A list of parameters
  - Each parameter has 3 subfields: (type, length, value)
  - Only one parameter type is defined so far: type 1 for authentication

BGP Messages (cont.)

- KEEPALIVE message
  - Two peers periodically exchange KEEPALIVE messages to test network connectivity and to verify both peers continue to function
  - "Hold Time" in OPEN message
    - = 0: no KEEPALIVE message
    - > 0: KEEPALIVE interval = 1/3 * Hold Time

```
0 8 16 24 31
Common header
19 bytes, Type = 4
```

BGP Messages (cont.)

- UPDATE message
  - Withdrawn Len & Withdrawn Destinations
    - To withdraw previous advertisements when a destination becomes unreachable
    - Withdrawn Len: size of Withdrawn Destinations field
  - Path Len & Path Attributes
    - To advertise new destinations that are reachable
    - Path Len: size of Path Attributes field
  - Destination networks
    - Networks actually advertised by this message
BGP Messages (cont.)

- **Path Attributes**
  - Attributes apply to all destination networks advertised in the message
  - If different attributes apply to some destinations, they must be advertised in a separate UPDATE message
  - This field contains a list of items
  - Each item consists of a triple
    - (type, length, value)
    - Type: 2 bytes
    - Length: 1~2 bytes (length of value)

- **Compressed mask-address pairs**
  - BGP sends an address mask with each IP address
  - *Withdrawn Destinations & Destination Networks* fields use a compressed representation to reduce message size
    - LEN: number of 1s in the mask
    - IP address: only the bytes (\[
        \left\lceil \frac{\text{LEN}}{8} \right\rceil
      \right)
    - covered by the mask are included
    - LEN = 0: no address byte follows; for default route

Format of *Type* subfield

<table>
<thead>
<tr>
<th>Bits</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0: well-known attribute; 1: optional</td>
</tr>
<tr>
<td>1</td>
<td>0: transitive attribute; 1: nontransitive</td>
</tr>
<tr>
<td>2</td>
<td>0: optional transitive attribute is complete; 1: partial</td>
</tr>
<tr>
<td>3</td>
<td>0: length field is 1 byte; 1: 2 bytes</td>
</tr>
<tr>
<td>5~7</td>
<td>Unused (must be zero)</td>
</tr>
</tbody>
</table>

1: ORIGIN
2: AS_PATH
3: NEXT_HOP

BGP Messages (cont.)

- NOTIFICATION message
  - Used for control or when error occurs
  - Once a router detects an error
    - Sends a NOTIFICATION message
    - Closes the TCP connection

```
0  8  16  31
   common header
   19 bytes, Type = 3
```

- Error subcode
- Error data (variable)

BGP Messages (cont.)

- Error code
  - Error code description
  - Error subcode
  - 1 Message header error
  - 2 OPEN message error
  - 3 UPDATE message error
  - 4 Hold timer expired
  - 5 Finite state machine error
  - 6 Cease

- Terminate connection; no subcode defined

Other Issues on BGP

- Encapsulation
  - BGP messages are encapsulated in TCP segments using well-known port 179
  - No need for error control and flow control

- Route selection
  - BGP can only specify whether a path to a given destination exists
  - BGP cannot select the shortest path
    - The cost of routes across autonomous systems is unknown to BGP
  - BGP is a reachability protocol rather than a routing protocol
    - A receiver can implement route policy constraints, but cannot choose a least cost route
    - A sender must only advertise paths that traffic should follow
Other Issues on BGP (cont.)

- Consequences
  - At any given instance, all traffic routed from a computer in one AS to a network in another will traverse one path, even if multiple physical connections are present
  - BGP alone is inadequate for optimal routing
  - BGP alone will not guarantee global consistency

Routing Arbiter (RA) system

- Routing arbiter database
  - Database of reachability information
  - Updates to database are authenticated: only AS that owns a given network is allowed to advertise reachability

- Route server (RS)
  - Multiple route servers maintain copies of routing arbiter database and run BGP
  - Each ISP designates a BGP border router, which maintains a connection to the route server