Routing(2) – Inter-domain Routing

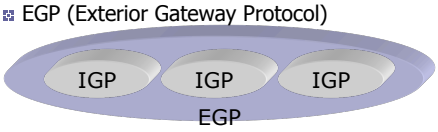
Information Network I
Youki Kadobayashi

Outline

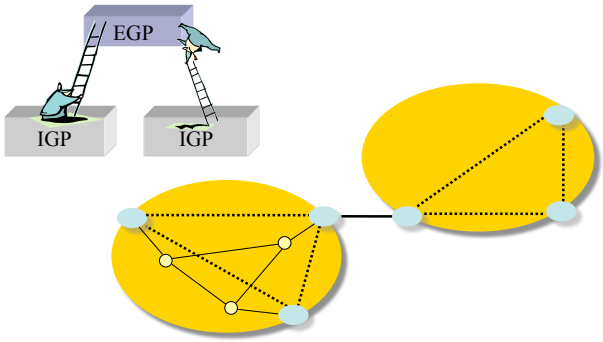
- Distance vector routing
- Link state routing
- IGP and EGP
  - Intra-domain routing protocol, inter-domain routing protocol
- **Path vector routing**
- BGP: Border Gateway Protocol
- Route aggregation

Hierarchical Routing

- Routing domain
  - Defines the boundary between domains
  - Fault isolation, route aggregation
- Distinction between intra-domain routing protocol and inter-domain routing protocol
  - IGP (Interior Gateway Protocol)
  - EGP (Exterior Gateway Protocol)



Hierarchical Routing: two-tier routing



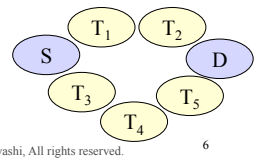
Hierarchical Routing – IGP and EGP

- IGP
  - RIP-2, RIPng: distance vector routing
  - OSPF, IS-IS: link state routing
  - Focus on propagating the state of each link/router as fast as possible
- EGP
  - BGP4, BGP4+: path vector routing
  - Focus on the routing stability of the whole internet

Path Vector Routing

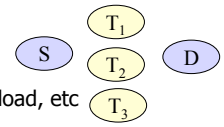
- Derived from Bellman-Ford algorithm
- Information exchange in distance vector routing: (prefix, metric)
- Information exchange in path vector routing: (prefix, path, attributes)
- Assigns distance as well as path information to the route information
  → Embodies "routing without loops"
  This protocol prioritizes route that has the shortest path vector.

Route expression from S to D
- (D, { T_1, T_2 }, ...)
- (D, { T_3, T_4, T_5 }, ...)



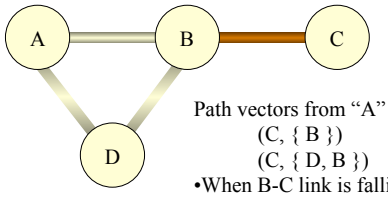
Path Vector Routing: Background

- Multiple alternative routes
  - Dense connections between ISPs
  - Which route should we prioritize?
    - constrained by cost, contract, load, etc



- Routing policy
  - Encodes the intention of the intermediate ISPs
  - Route selection policies enable each domain to select a particular route among multiple routes
    - ⇨ Policy can't be expressed by scalar cost.
- Cost of loops
  - Convergence time from transient state ～ RTT

Loop Avoidance in Path Vector Routing



Path vectors from "A" to C
(C, { B })
(C, { D, B })

•When B-C link is falling down, B is deleted from path vector.
•Rejects path vector that include the router itself
•Loop avoidance

Q&A

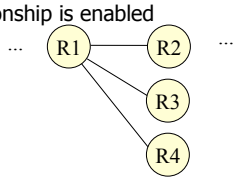
BGP
Border Gateway Protocol

BGP: Border Gateway Protocol

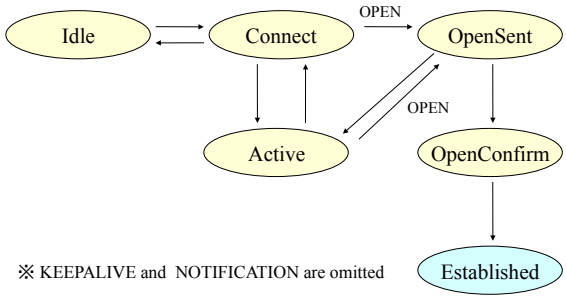
- Algorithm
  - Path vector
- Transport
  - TCP
  - TCP provides retransmission and acknowledgement.
- **Adjacency relationship and state transition**
- Routing information
- Topology

Adjacency Relationship of BGP

- Adjacency relationship is defined by ISP operator
- Adjacency relationship must be explicitly configured
- Why？
  - C.f. OSPF : if parameters match, adjacency relationship is enabled



State Transition of BGP: Establishment of Adjacency Relationship



BGP: Border Gateway Protocol

- Algorithm
- Transport
- Adjacency relationship and state transition
- **Route information**
  - **efficient path vector expression**
- Topology

Expression of Path Vector in BGP

- AS (Autonomous System) is expressed as an AS number
  - AS: routing domain that is operated by single policy
- BGP collects and encodes (prefix, AS-path, attributes)
  - (AS-path, attributes, { prefix1, prefix2, ... })
  - Reduction of traffic

Example of Path Vector in BGP

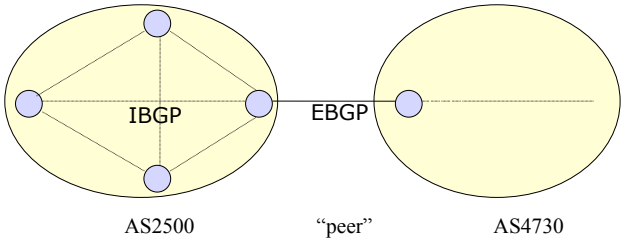
```
show ip bgp 163.221.0.0
BGP routing table entry for 163.221.0.0/16, version 30149334
Paths: (6 available, best #3)
  Not advertised to any peer
  6461 2516 2500 2500
    208.185.175.169 from 208.185.175.169 (216.200.254.220)
      Origin IGP, metric 0, localpref 400, valid, external
  6461 2516 2500 2500, (received-only)
    208.185.175.169 from 208.185.175.169 (216.200.254.220)
      Origin IGP, metric 0, localpref 100, valid, external
  2500, (received & used)
    203.181.70.232 (metric 2) from 203.181.70.232 (203.181.70.227)
      Origin IGP, metric 505, localpref 400, valid, internal, best
      Originator: 203.181.70.227, Cluster list: 0.0.0.16
  2500, (received & used)
    203.181.70.233 (metric 2) from 203.181.70.233 (203.181.70.227)
      Origin IGP, metric 505, localpref 400, valid, internal
      Originator: 203.181.70.227, Cluster list: 0.0.0.16
  2500
    202.249.2.1 from 202.249.2.1 (203.178.136.4)
      Origin IGP, metric 1010, localpref 400, valid, external
  2500, (received-only)
```

Q&A

BGP: Border Gateway Protocol

- Algorithm
- Transport
- Adjacency relationship and state transition
- Route information
- **Topology**

Topology of BGP: IBGP and EBGP



Topology Constraints in IBGP and EBGP

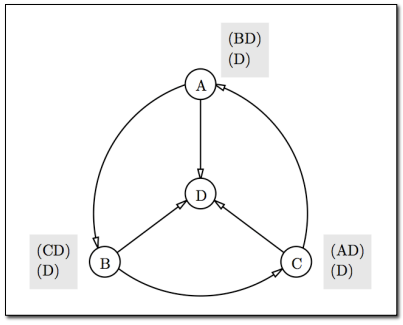
- IBGP
  - must establish peering between all IBGP routers in same AS.
  - doesn't need physical adjacency.
  - All IBGP routers has same routing information.
- EBGP
  - Requires physical adjacency in principle
  - Routers don't share the same route information with adjacent ASes.

Q&A

Policies in BGP

- Routing policy
  - Encode policies of transit providers
  - Express route selection policy among alternative routes
- Policies transmitted to other providers
  - MED
- Policies within IBGP
  - Local preference, admin distance, route map, etc.

Limitation of policy in BGP: dispute wheel



Rectangle denotes ordered list of path preferences

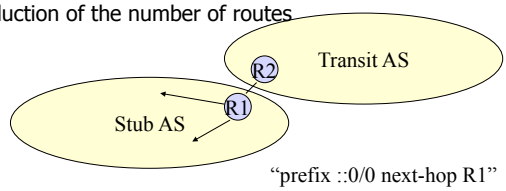
Can we reach D?

N. Feamster et al., "Implications of Autonomy for the Expressiveness of Policy Routing", SIGCOMM'05.

Reduction and aggregation of routes

Route Reduction : Default route

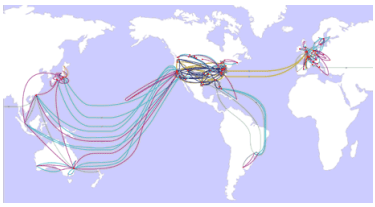
- 0.0.0.0/0 (IPv4), ::0/0 (IPv6)
- Longest prefix match
  → matches in the end of route search
- Results to hiding of routes and reduction of the number of routes



"prefix ::0/0 next-hop R1"

When and where routes cannot be reduced?

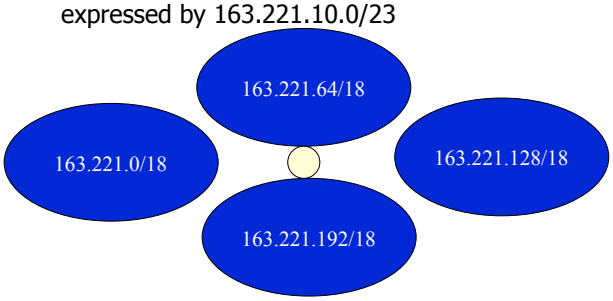
- Default-free
  - No default route
- Tier-1 ISPs, North America backbone



(source: UUNet network maps, www.uu.net)

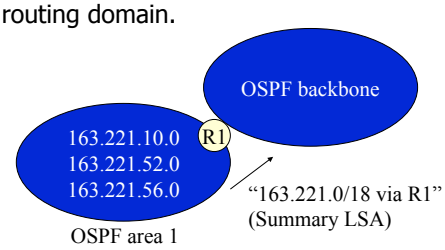
Aggregation

- 163.221.10.0/24 and 163.221.11.0/24 are expressed by 163.221.10.0/23

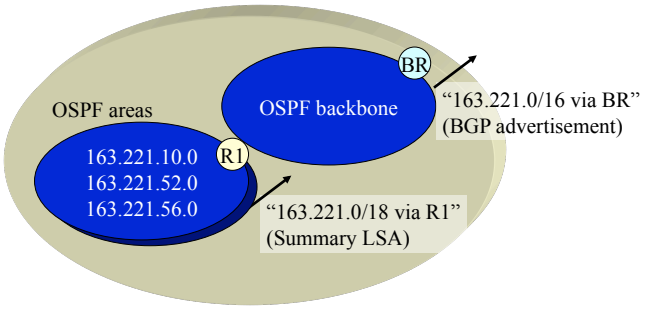


Route Aggregation at Domain Edge

- Routes are aggregated at the edge of the routing domain.



Hierarchical Aggregation of Routes



Problems in Route Aggregation

- Route aggregation depends on address assignment.
  - Planned address assignment is important for route aggregation.
- Can we make predictions of the future number of departments of NAIST?
  - Can we make predictions of the growth of an ISP?
- → prefix renumbering
  - renumbering to aggregable addresses
  - development of technology
- Manual operation is necessary for route aggregation
  - internet full-route : 250,000
  - BGP table growth trends - Telstra

Q&A

Summary

- Hierarchical routing concepts
  - IGP, EGP
- Path Vector Routing
  - Loop-free, policy-aware
- BGP
  - State transition, route information and topology
  - Limitation of policy-based routing
- Route Aggregation
  - Aggregation concepts, challenges

Assignment

- Choose two web sites, investigate inter-domain routes by looking glass, and then visualize AS paths.
- Mapping AS number to provider name:

```
$ whois -h radb.ra.net. AS2500
aut-num:    AS2500
as-name:    WIDE
descr:      WIDE Project in Japan
```

- Optionally, pick a slow web site and investigate its reason, by using above tools and tcpdump/wireshark
- Deadline: 5/18 17:00
- Post to A3F Internet Engineering Lab