Routing on the Internet

- **Scale Issues**
  - Network of IP Prefixes indicating different subnets: 150,000-200,000 so far
  - More than 20,000 organizations owning different “subnets”
  - Millions of different routes between and within subnets

- **Privacy Issues**
  - Organizations may not want to reveal internal topologies

- **Policy Issues**
  - No unique policy for determining the costs of links
  - Organizations may have preferred routes based on policies/relationships

Autonomous Systems

- **Definition**: A set of routers that has a single routing policy and runs under a single *technical* administration.

- **Internet Gateway Protocol**: An internal routing protocol run within the AS

- **Routing between ASes**: External Gateway Protocol at the early stages of internet
  - Replaced by the **Border Gateway Protocol (BGP)**
### Types of ASes

- **Stub AS**: An AS with only a single connection to one other AS (carries local traffic)
- **Multihomed AS**: An AS that has connections to more than one ASes but refuses to carry transit traffic
- **Transit AS**: An AS that has connections to more than one AS and carries both transit and local traffic

![Diagram of AS types](image)

### Drawbacks of Link-State, DV Routing

- Traffic is **restricted** to shortest path routes
- Not all nodes may assign the same metrics to links
  - Difficult to obtain the same global view of the network
  - Conflicting information may lead to routing oscillations
- Topology information **floated** in the entire network
- High processing overhead at each node to determine appropriate routes

### Path-Vector Routing

- **Key Idea**: Advertise the entire path to a destination (AS) rather than the cost of the path
  - Based on the Distance Vector algorithm
  - Send cost metrics to the destination along with the entire routing path to different ASes.

![Diagram of Path-Vector Routing](image)
Advantages of Path-Vector Routing

- Fast loop detection
  - AS can locate its own identifier in the path to a destination
  - Node can simply discard paths with loops
  - E.g. AS 5 discards route update from AS 1

Flexible Policy Implementation

- Each node can apply local policies
  - Path selection: Which path to use?
  - Path export: Which paths to advertise?

- Examples
  - Node 2 may prefer the path “2, 3, 1” over “2, 1”
  - Node 1 may not let node 3 hear the path “1, 2”

Border Gateway Protocol (BGP)

- Interdomain routing protocol for the Internet
  - Prefix-based path-vector protocol
  - Policy-based routing based on AS paths
  - Evolved during the past 15 years

- Establish a TCP connection on port 179
- Exchange information on all known active routes
- Exchange incremental updates
Storing and Learning Routes

- BGP router stores multiple routes toward a destination
  - Policies are applied for the selection of the “best” route
  - …and which routes shall be advertised

- Incremental updates
  - Announcement
    - New active routes toward destinations are updated by sending BGP updates
  - Withdrawal
    - If a route is no longer active, a withdrawal notification is sent to all neighbors

Routing in BGP

- Destination prefix based
- Includes route attributes such as next hop and AS path

```
128.122.56.0/16
AS 1
Next_Hop = 192.168.0.5
AS_path = AS 1

192.168.0.5
AS 2
Next_Hop = 12.127.2.1
AS_path = AS 2 AS 1

128.122.56.0/16
AS 3
Next_Hop = 101.108.2.2
AS_path = AS 3
```

BGP Attributes

- AS Path
  - Sequence of ASes that a route has traversed
  - Used for detecting loops and applying policies

- Next Hop
  - Next hop to reach a network

- Local Preference
  - Used to influence selection of paths
  - Path with highest local preference is selected

- Multi-Exit Discriminator (MED)
  - Use to influence exit (entry) point when multiple BGP routers exist in an AS
BGP Path Selection

- Follow a minimum hop approach
  - Shortest AS path
  - Arbitrary tie break
- Example
  - AS 2 routes to AS 6 through AS 3
- BGP is mainly policy-based routing

BGP Policy

- Import policy
  - Filter unwanted routes from neighbor
    - E.g., prefix that your customer doesn’t own
    - Manipulate attributes to influence path selection
      - E.g., assign local preference to favored routes
- Export policy
  - Filter routes you don’t want to tell your neighbor
    - E.g., don’t tell a peer a route learned from other peer
    - Manipulate attributes to control what they see
      - E.g., make a path look artificially longer than it is

Import Policy – Local Preference

- Favor one path over another
  - Used to override shortest AS path principle
  - Example: prefer a customer AS
Filtering

- Import Policy
  - Discard certain advertised paths
  - Recognize network or AS based on prefix
    - Discard routes that contain large ISPs (may induce longer delays)
    - Discard pre-fixes if not owned

- Export Policy
  - Limit propagation of routing information
  - Don’t announce routes from one peer to another
  - Don’t advertise routes containing untrusted ASes
  - Manipulate attributes to influence the way that other ASes perceive routes

BGP Policy Configuration

- Routing policy languages are vendor-specific
  - Not part of the BGP protocol specification
  - Different languages for Cisco, Juniper, etc.

- Still, all languages have some key features
  - Policy as a list of clauses
  - Each clause matches on route attributes
  - ... and either discards or modifies the matching routes

- Configuration done by human operators
  - Implementing the policies of their AS
  - Business relationships, traffic engineering, security, ...

Other Considerations

- AS_path can be misleading in terms of delay and # of hops
Treating Multiple Routers within an AS

- BGP routing info is disseminated within an AS using iBGP

Combination of eBGP with iBGP

Hot-potato (deflection) Routing

- Incentive
  - Minimize required storage, probability of buffer overflow
- Key idea: Get rid of packets as soon as possible
  - Packets may not necessarily follow optimal paths (deflect packets wherever possible)
  - E.g., if two packets must go out on the same link, deflect one of them to another at random

- Local pref = 90
- Local pref = 200