# CS120: Computer Networks

## Lecture 15 – Interdomain routing

Haoxian Chen

Slide adopted from: Zhice Yang

# Routing Protocols

- Routing Information Protocol (RIP)
  - Algorithm: Distance Vector
- Open Shortest Path First (OSPF)
  - Algorithm: Link State
- Border Gateway Protocol (BGP)

Intradomain Routing Protocol

Interdomain Routing Protocol

How do we scale to the global internet?

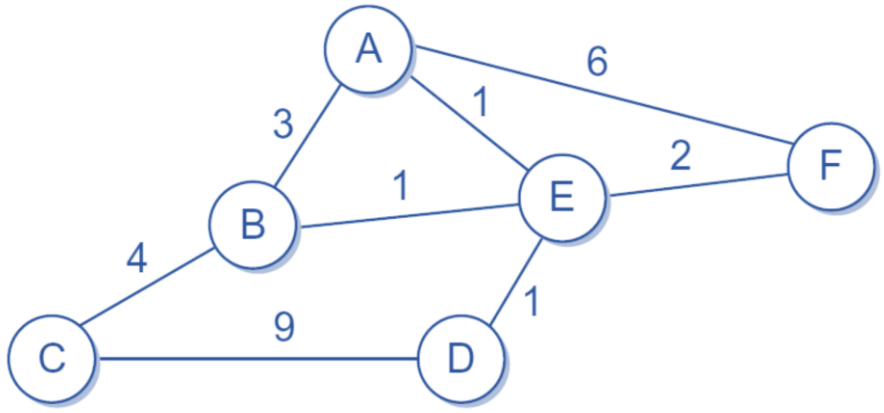
# Today's lecture

- Routing refresher
- Global internet hierarchy
- BGP basics
- Integrating BGP with intradomain routing
- Routing policies

# Network as a Graph

- The basic problem of routing is to find the lowest-cost path between any two nodes
  - Static approach has several shortcomings
    - Can not handle node or link failures
    - Can not handle addition of new nodes or links
    - Edge costs cannot change
  - Centralized solution does not scale

➢Distributed and dynamic protocol



# Distance Vector Algorithm

- x maintains its distance vector estimate $\boldsymbol{D}_x(y) = \{D_x(y): y \in N\}$
- x knows:
  - cost to each neighbor v: $c(x, v)$
  - neighbors' distance vectors estimate: $\boldsymbol{D}_v(y) = \{D_v(y): y \in N\}$
- Algorithm idea:
  - From time-to-time, each node sends its own distance vector estimate to neighbors
  - When x receives new distance vector estimate from neighbor, it updates its own distance vector estimate using Bellman-Ford equation
  - Under minor, natural conditions, the estimate $D_x(y)$ will converge to the actual lowest cost $d_x(y)$

# Link State Routing

- Network topology is known to all routers
  - Accomplished via broadcasting link state packets (LSP) to all routers
- Routing Algorithm: computes shortest paths from one node ('source') to all other nodes
  - Based on Dijkstra's Algorithm

# Distance vector vs. link state

| | Distance vector | Link state |
|---|---|---|
| Information maintained | | |
| Information exchange | | |
| Overhead | | |
| Convergence Speed | | |
| Examples | | |

# Distance vector vs. link state

| | Distance vector | Link - state |
|---|---|---|
| Information maintained | distance to each destination | complete network topology |
| Information exchange | router's own distance vector | router's neighbors and cost to each neighbor |
| Overhead | low | high |
| Convergence Speed | low | high |
| Examples | RIP | OSPF |

# Routing Protocols

- Routing Information Protocol (RIP)
  - Algorithm: Distance Vector
- Open Shortest Path First (OSPF)
  - Algorithm: Link State
- Border Gateway Protocol (BGP)

Intradomain Routing Protocol

Interdomain Routing Protocol

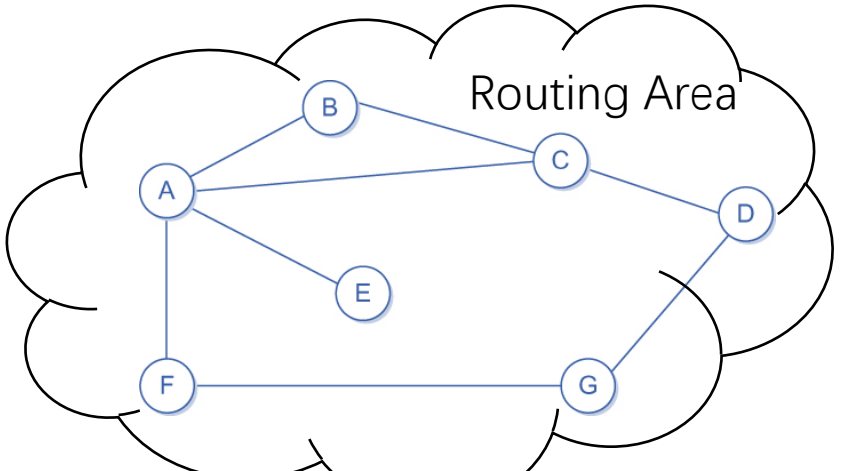
How do we scale to the global internet?

# Today's lecture

- Routing refresher
- **Global internet hierarchy**
- BGP basics
- Integrating BGP with intradomain routing
- Routing policies

# The Discussion on Routing So Far …

- Routers in the network are managed by the same administrator
  - e.g. Residential building, Campus, Network of a same ISP, Network of a big company, etc.
- Routers running same routing protocol
  - OSPF, RIP, etc.
  - These routers have certain coverage, called the routing area (see next slides)



# The Real Internet: Network of Network

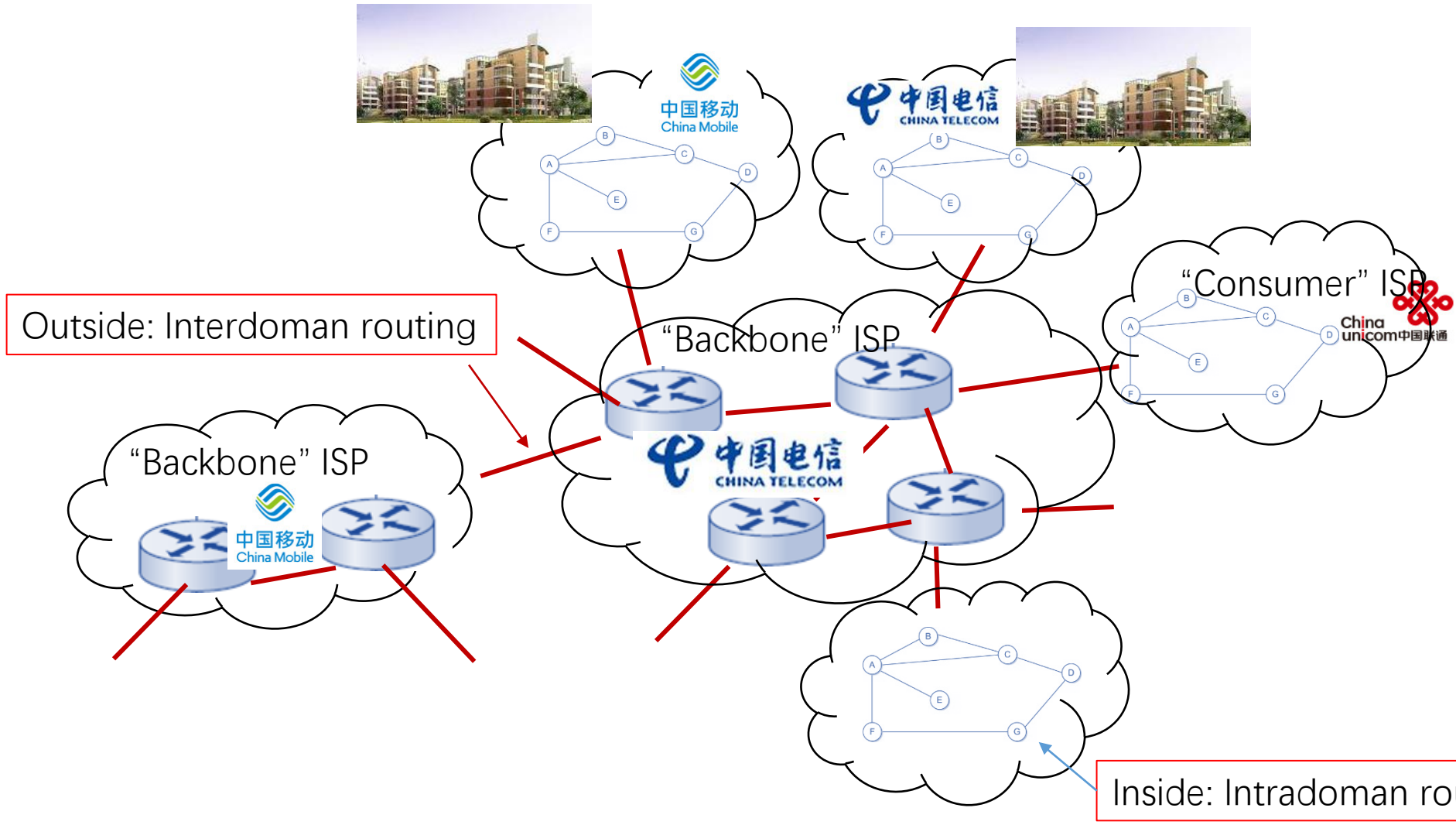
# Internet Service Providers

- End systems connect to Internet via “Consumer” ISPs (Internet Service Providers)
  - Residential, company and university ISPs
- “Consumer” ISPs are connected “Backbone” ISPs
  - Three Major Commercial ISPs in China




  - Other ISPs
    - e.g. cernet

# Network of Networks

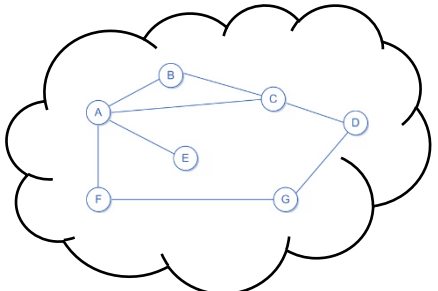


# Interdomain Routing Problems

- Scalability: More than 600 million destinations
  - Storage
    - Routing Table
  - Calculation
    - Shortest Path
  - Communication
    - Exchanges Routing Information
- Routing Management: Complex Routing Polices

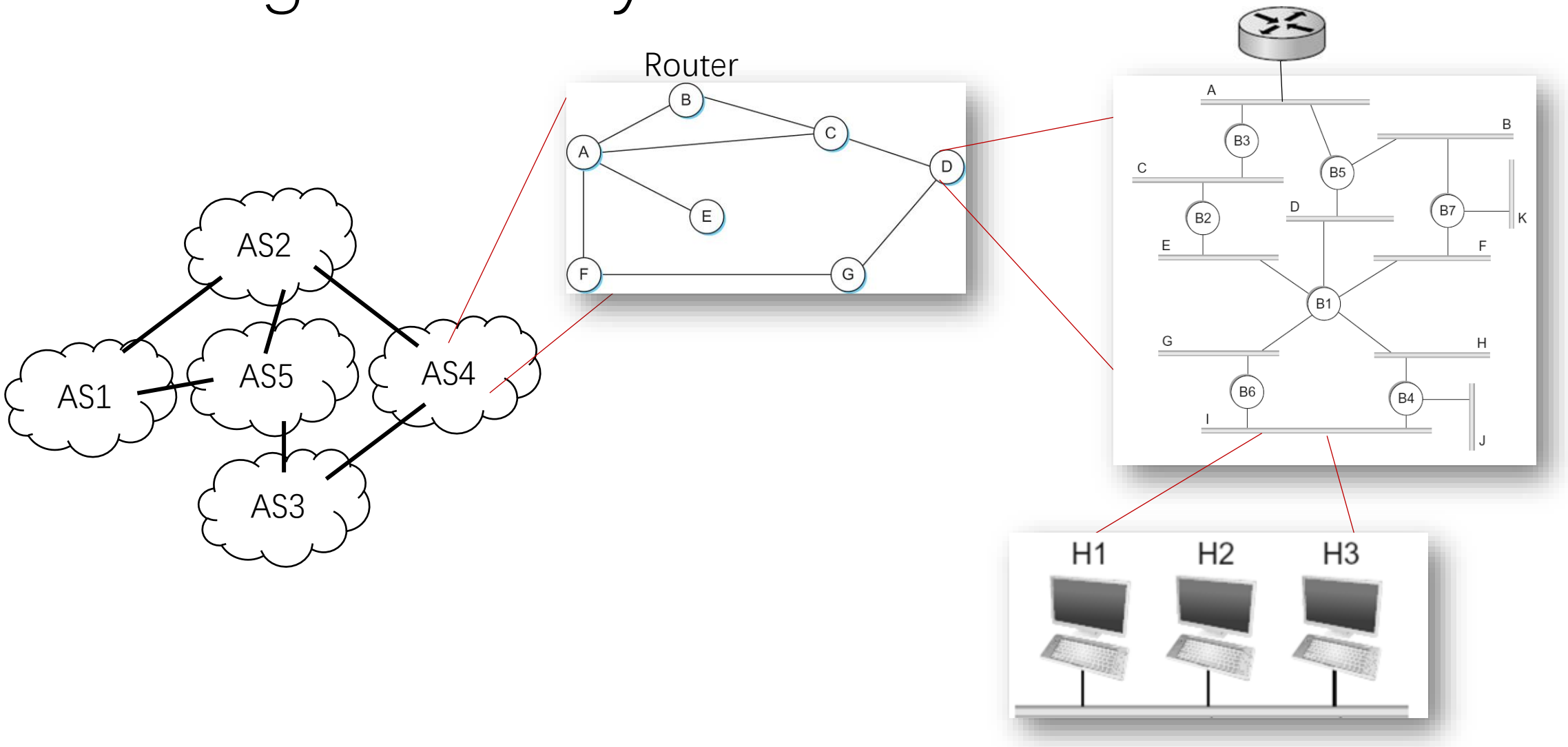
# Interdomain Routing: New Hierarchy

- Aggregate Routers into Logical Areas: Autonomous System
- Autonomous System (AS)
  - Corresponds to an administrative domain
  - e.g. University, company, backbone network
- Routers in same AS run same intradomain routing protocol
  - RIP, OSPF, etc.
- Routers in different AS run interdomain routing protocol
  - BGP, EGP
- Interdomain routing element: AS



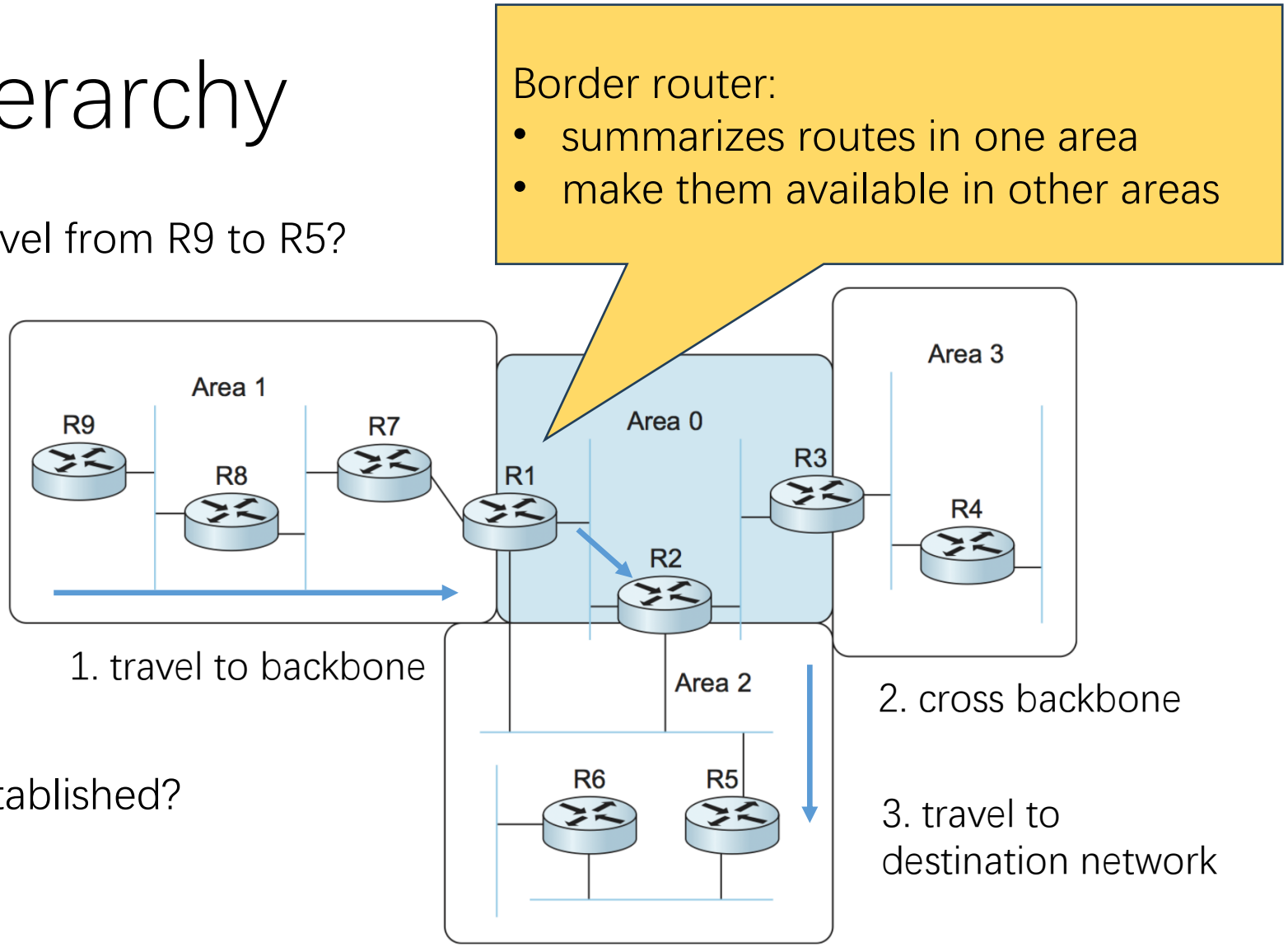

Autonomous System

# Routing Hierarchy



# Routing hierarchy

How does a packet travel from R9 to R5?




How are the routes established?

# Today's lecture

- Routing refresher
- Global internet hierarchy
- BGP basics
- Integrating BGP with intradomain routing
- Routing policies

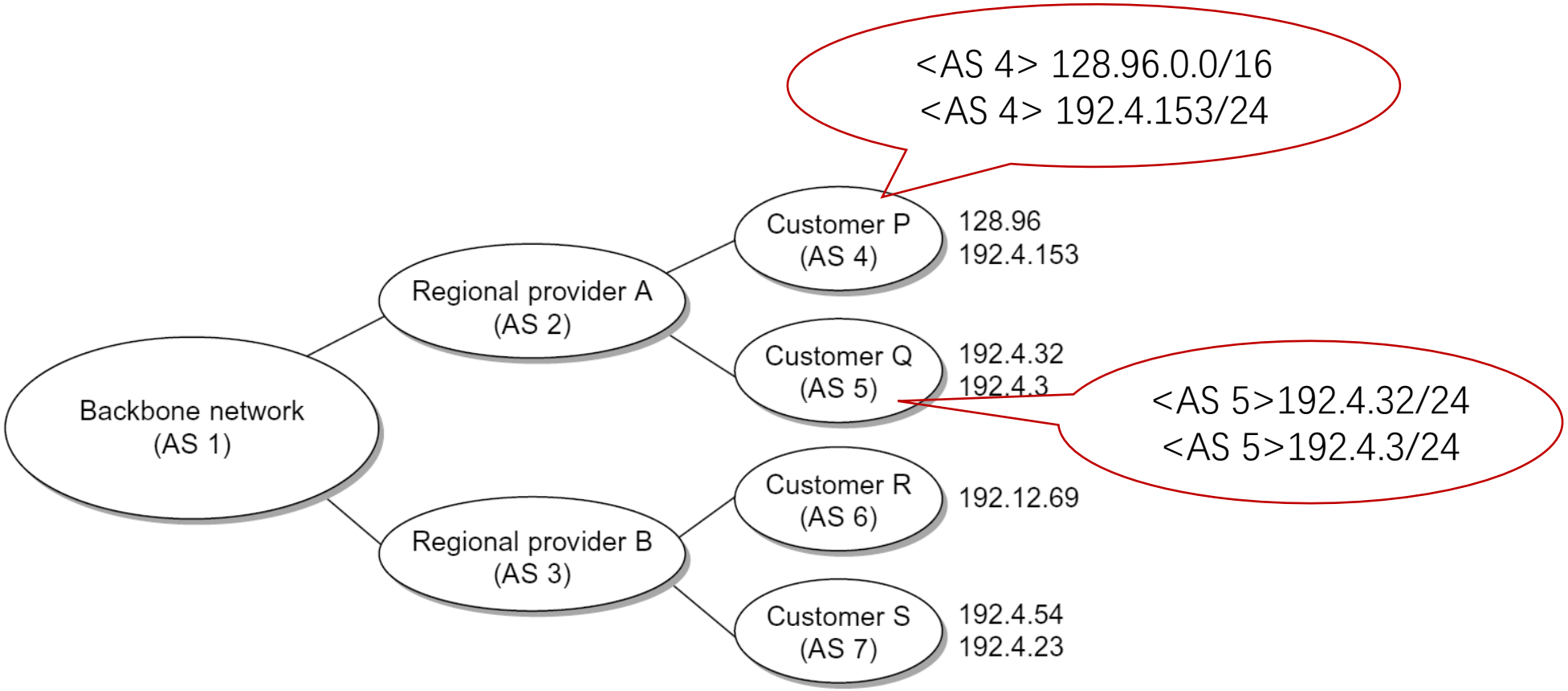
# Border Gateway Protocol (BGP)

- Widely-used Interdomain Routing Protocol
- Hardly be needed for small community or companies
- Routing Element: AS
- Routing Algorithm
  - Target on Reachability
    - Not the “best” route
  - Avoid Loop

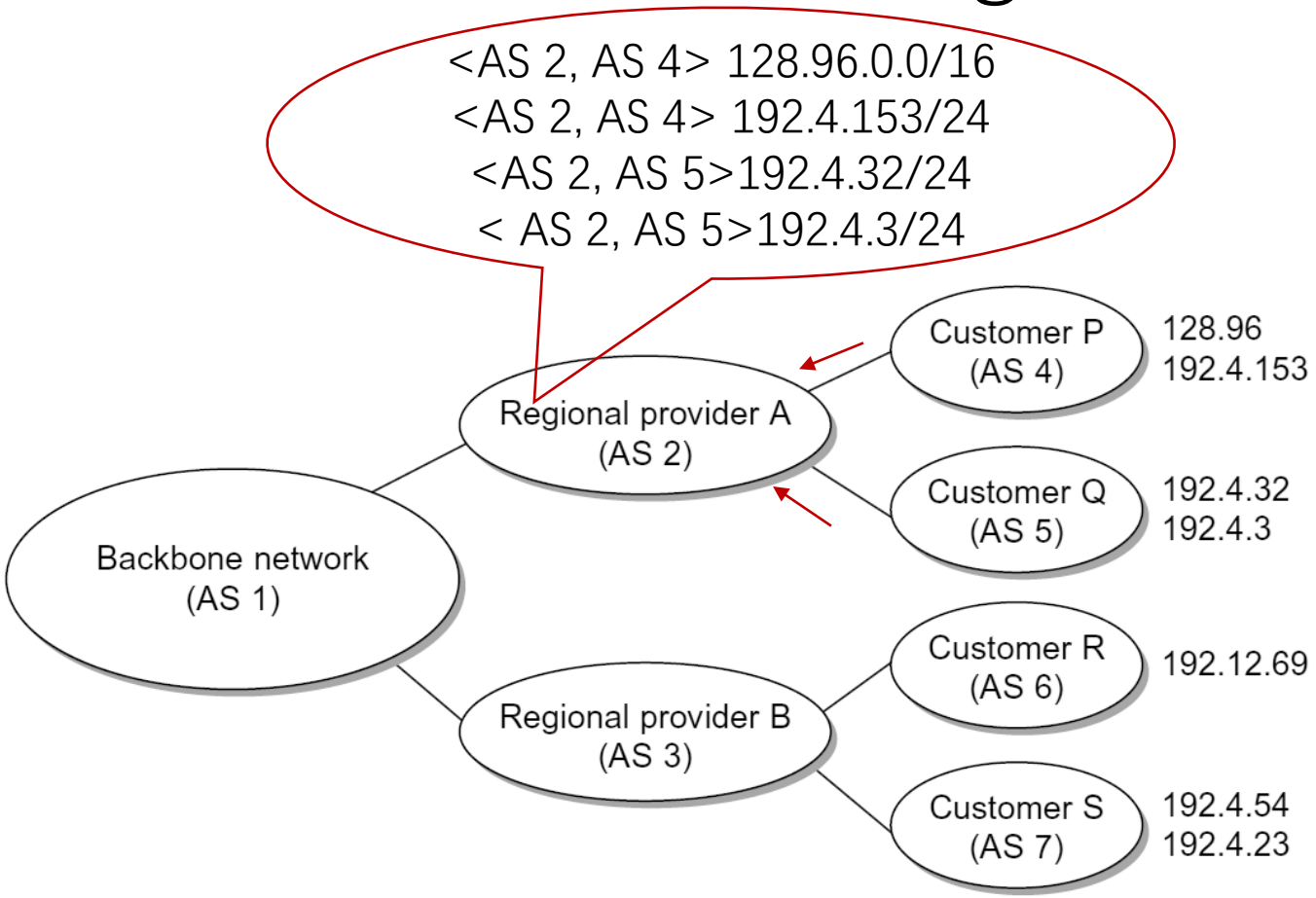
# Border Gateway Protocol (BGP)

- Broadcast route entries to neighbors
  - Similar to RIP
  - BGP route entry
    - AS path + network prefix+ next hop
    - eg. <AS a, AS b, AS c, ⋯>  128.96.0.0/16  12.5.6.1
  - Use AS number to detect cycles

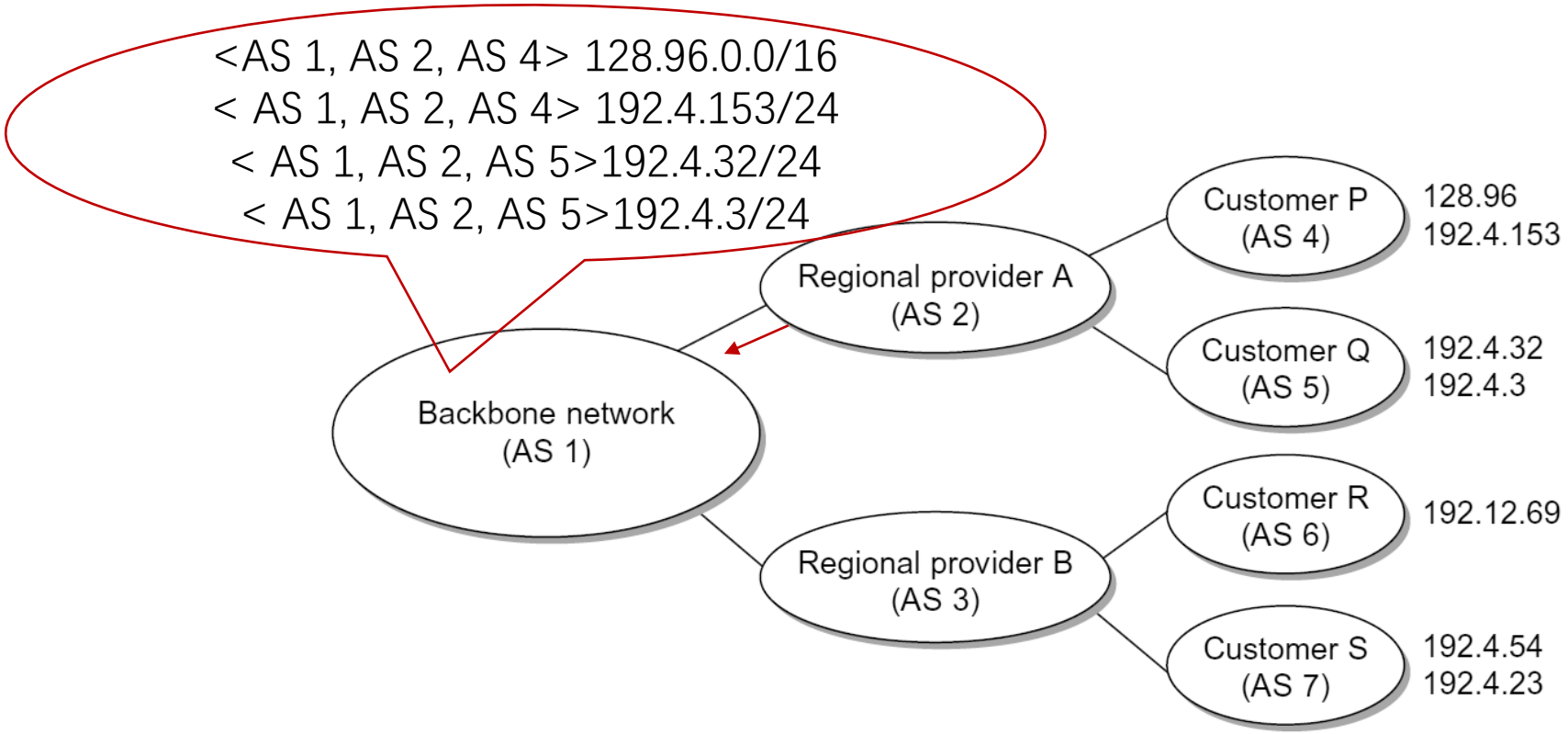
# BGP: Path Vector Routing



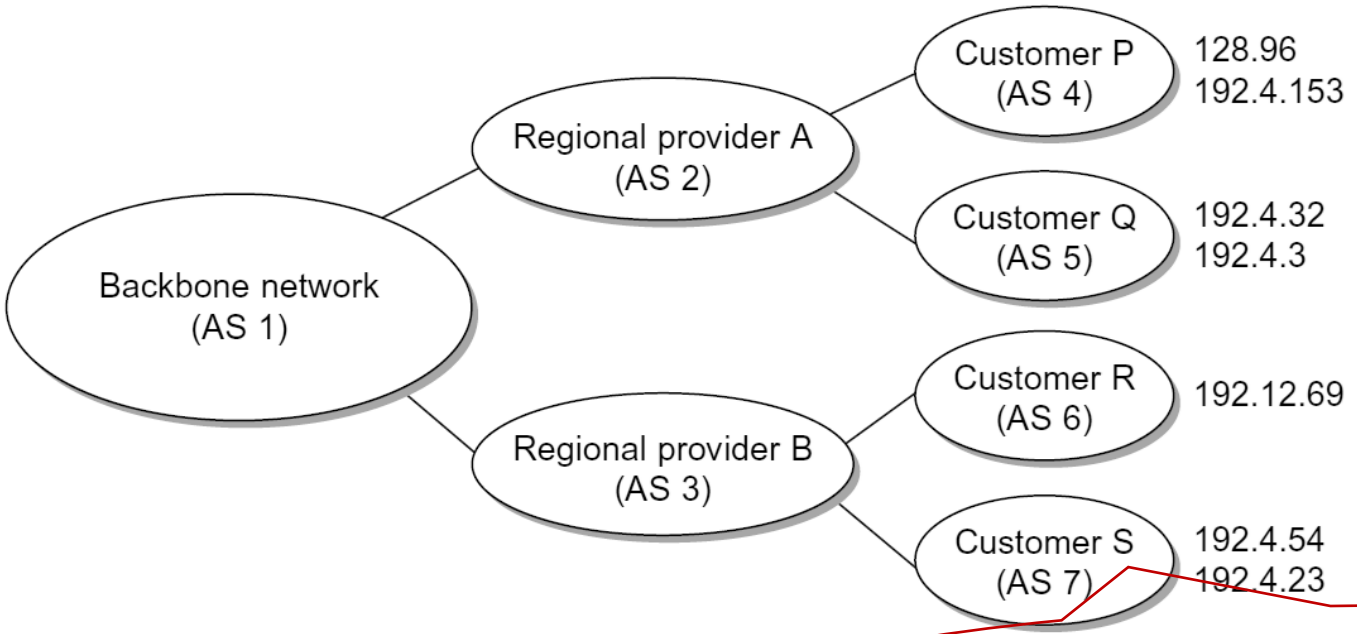
# BGP: Path Vector Routing



# BGP: Path Vector Routing

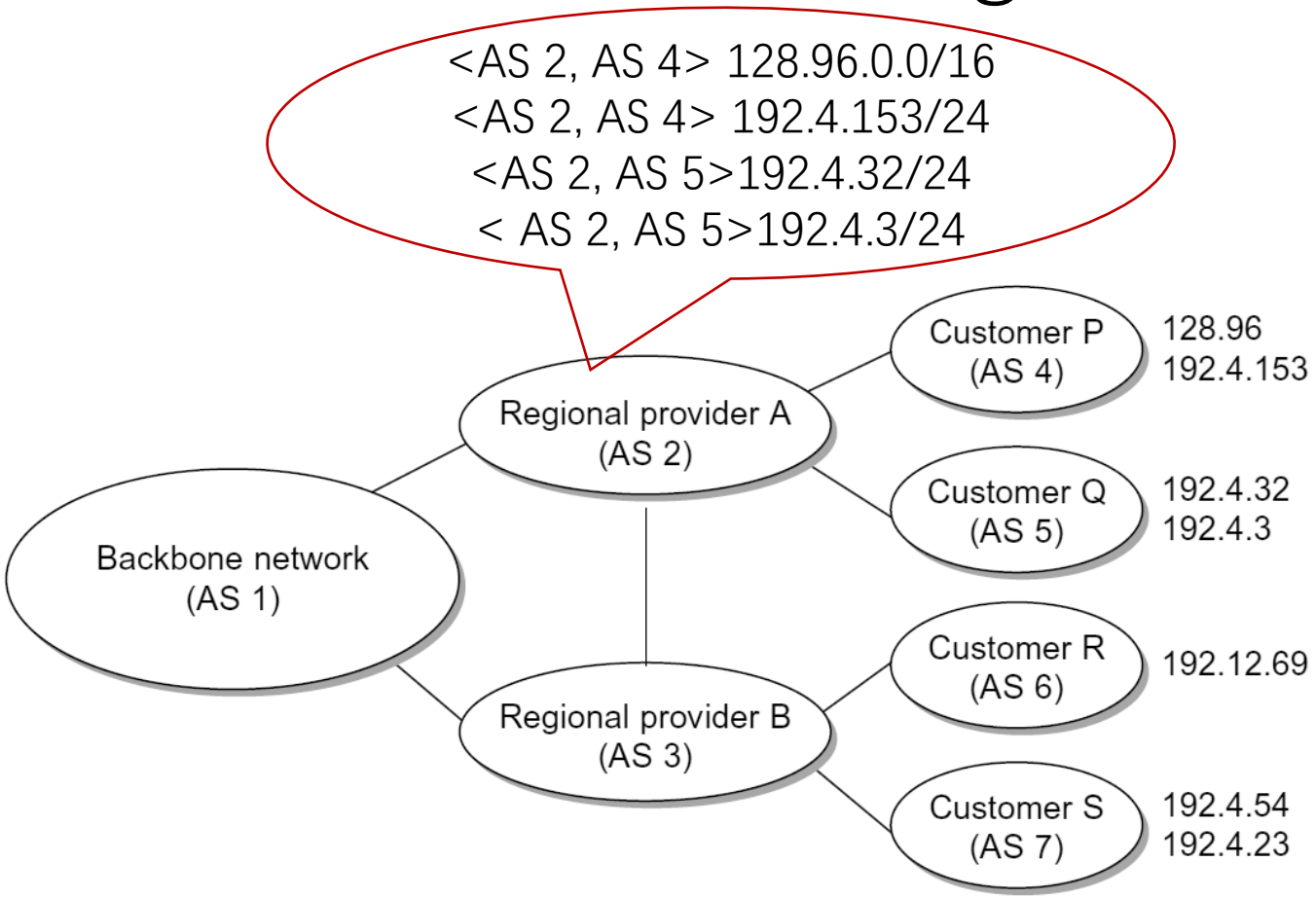


# BGP: Path Vector Routing

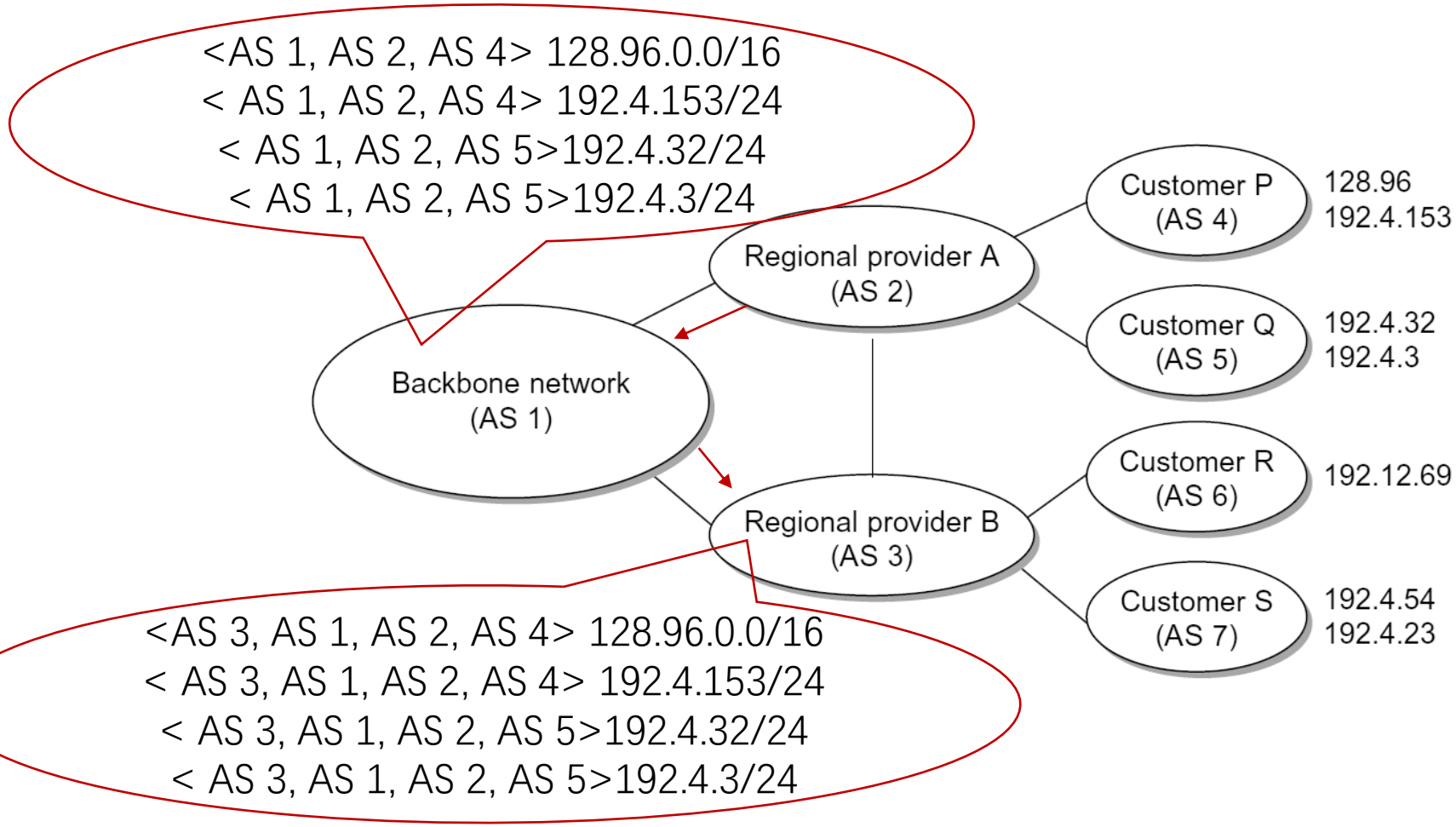



< AS 7, AS 3, AS 1, AS 2, AS 4> 128.96.0.0/16
< AS 7, AS 3, AS 1, AS 2, AS 4> 192.4.153/24
< AS 7, AS 3, AS 1, AS 2, AS 5>192.4.32/24
< AS 7, AS 3, AS 1, AS 2, AS 5>192.4.3/24

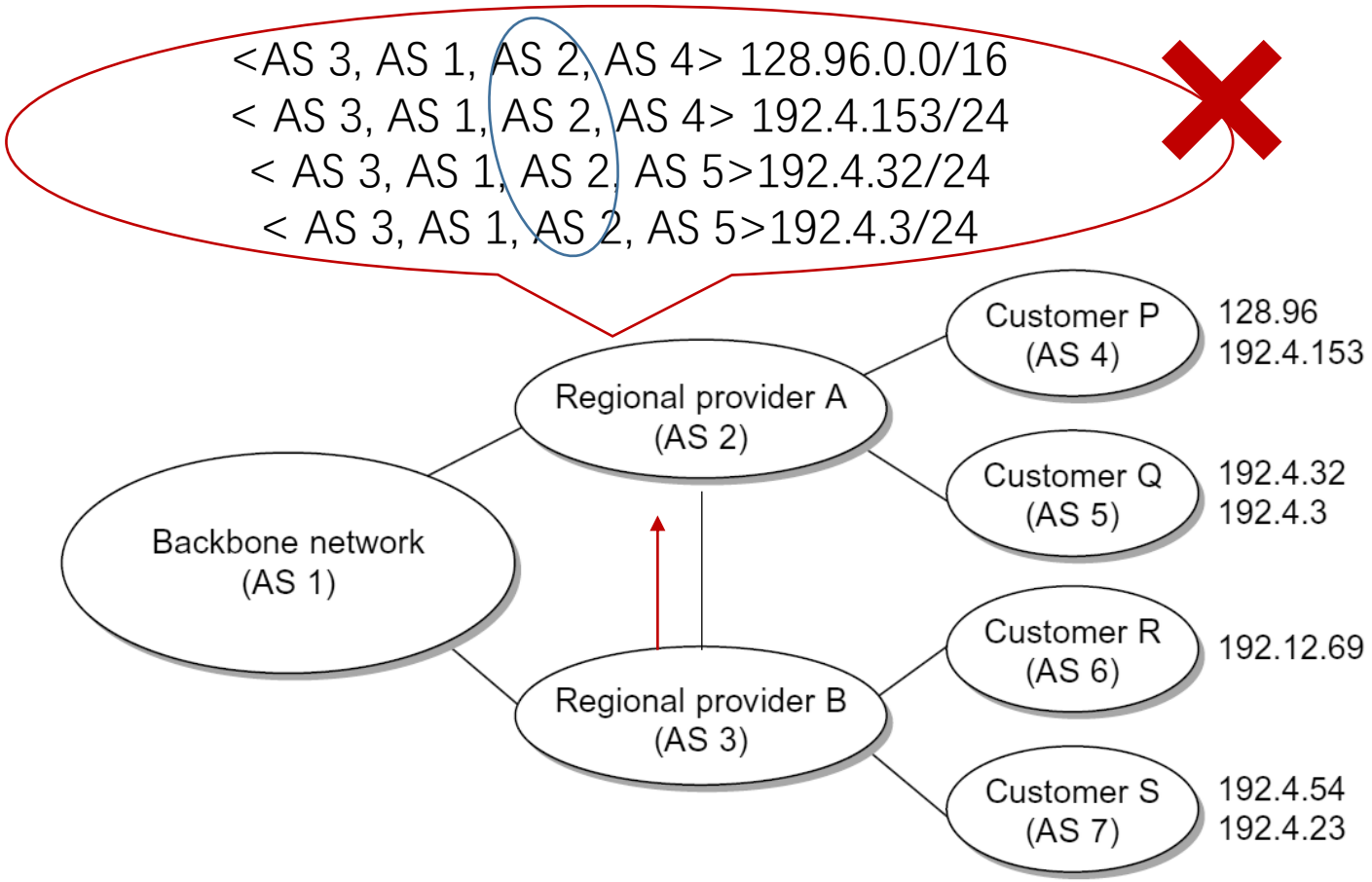
# BGP: Path Vector Routing



# BGP: Path Vector Routing



# BGP: Path Vector Routing




It's not a useful path

# Example

- AS route from telia Sofia to my computer
  - https://lg.telia.net/?type=bgp&router=sfia-b2&address=59.78.171.135

Dest Network

```
59.78.171.0/24 (2 entries, 1 announced)

*BGP    Preference: 170/-201
        Source: 2.255.253.187
        Protocol next hop: 2.255.254.180
        State: <Active Int Ext>
        Local AS:  1299 Peer AS:  1299
        Age: 4d 11:16:10        Metric: 100     Metric2: 2682
        AS path: 4637 4637 4637 4637 4538 4538 24364 I  (Originator)
```

Src Router

Border Router

Src AS

AS Route

AS1299

```
        Country: EU
Registration Date: 1993-09-01
      Registrar: ripencc
          Owner: TELIANET Telia Carrier, EU
```

AS4637

```
        Country: HK
Registration Date: 1995-10-30
      Registrar: apnic
          Owner: ASN-TELSTRA-GLOBAL Telstra Global, HK
```

AS4538

```
        Country: CN
Registration Date: 2002-08-01
      Registrar: apnic
          Owner: ERX-CERNET-BKB China Education and Research Network Center, CN
```

AS24364

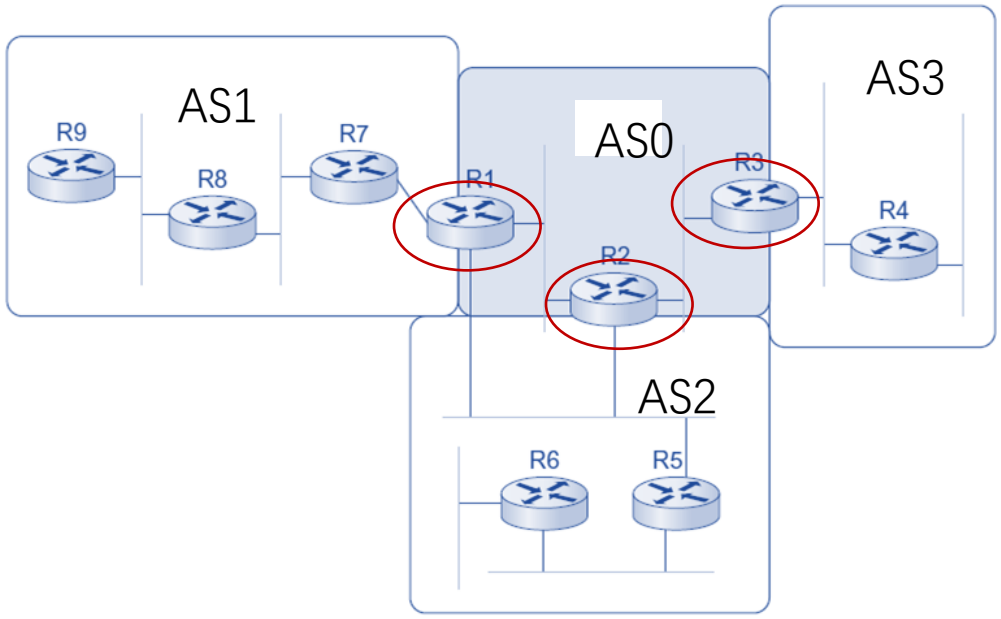
```
        Country: CN
Registration Date: 2005-03-03
      Registrar: apnic
          Owner: CNGI-SH-IX-AS-AP CERNET2 IX at Shanghai Jiaotong University, CN
```

# Today's lecture

- Routing refresher
- Global internet hierarchy
- BGP basics
- Integrating BGP with intradomain routing
- Routing policies

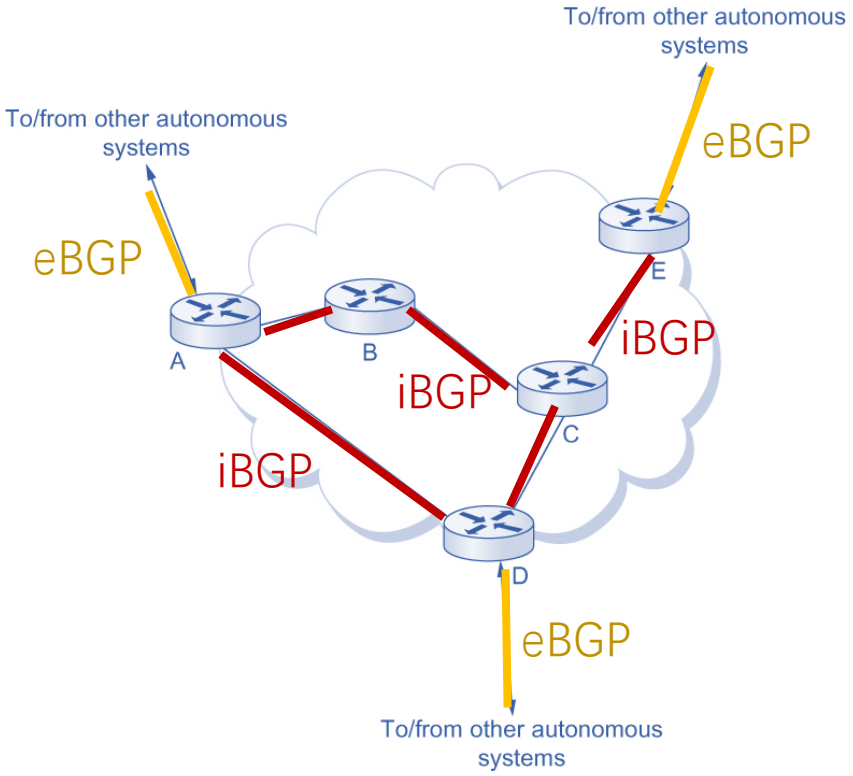
# BGP: Border Router

- Border Routers: connecting more than one ASs

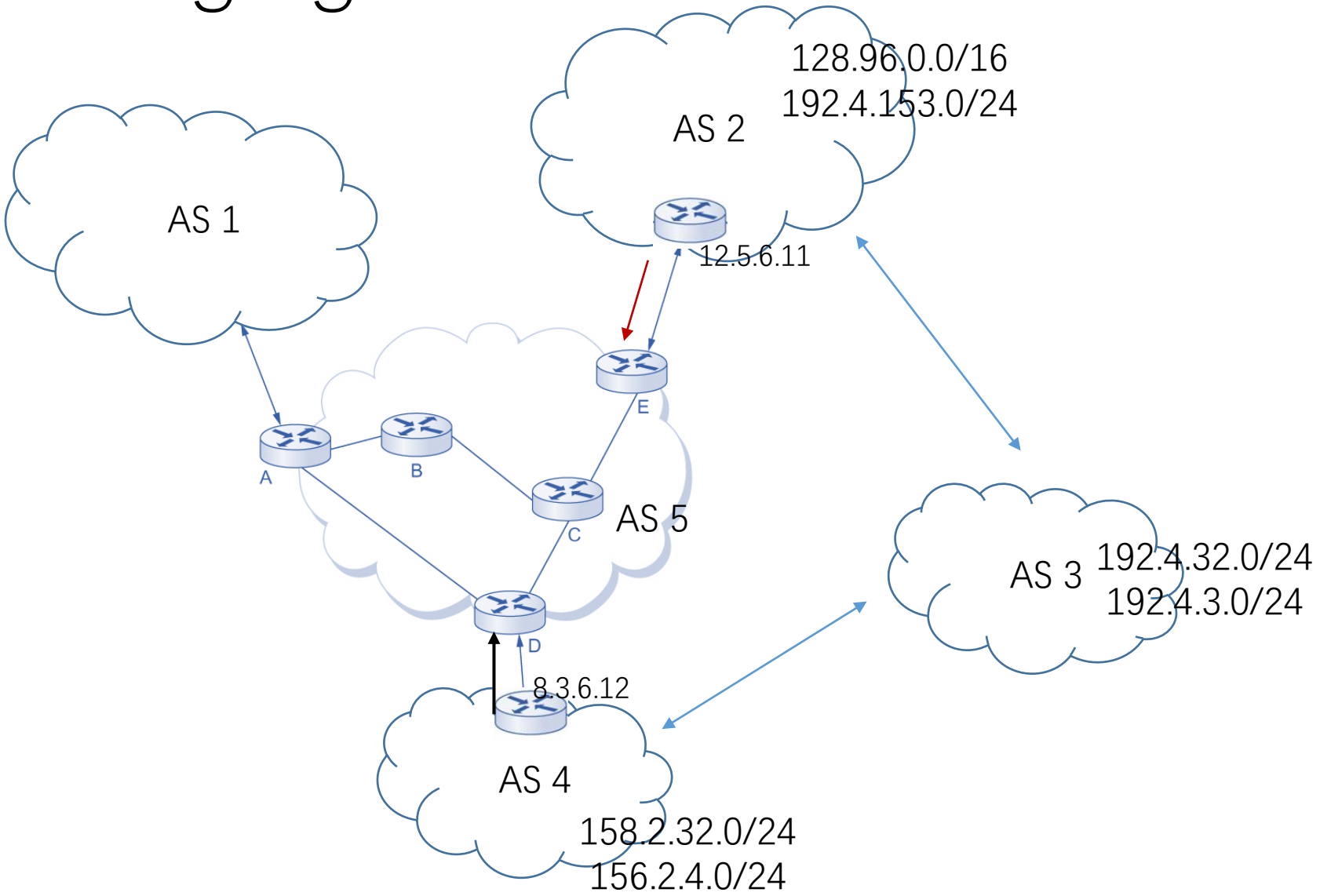


# BGP: Border Router

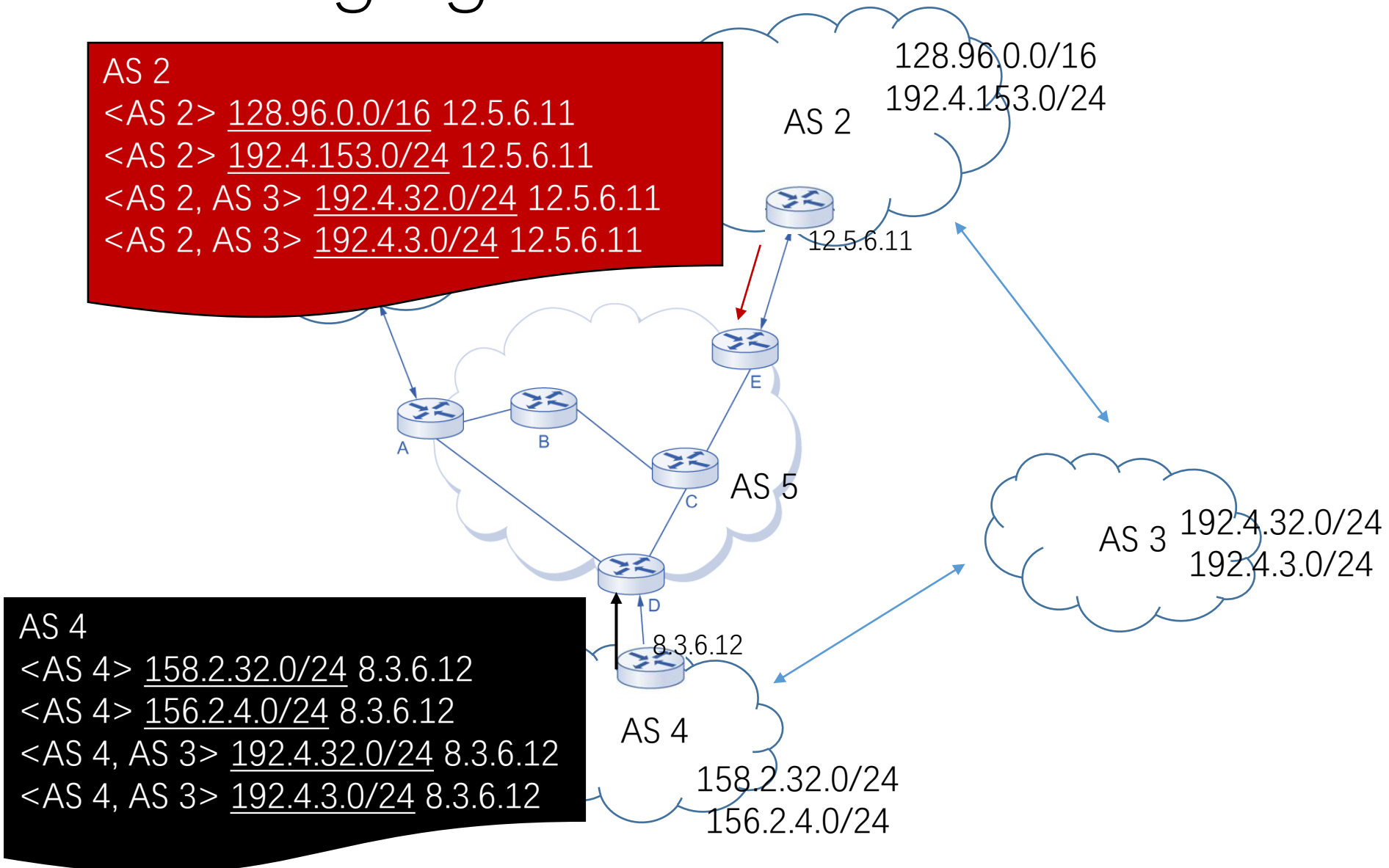
- Border Routers: connecting more than one ASs
  - Selected and configured by AS administrators
  - Routing entries exchanged with other Boarder Routers through exterior BGP (eBGP)
  - Routing entries exchanged with routers within the same AS through interior BGP (iBGP)



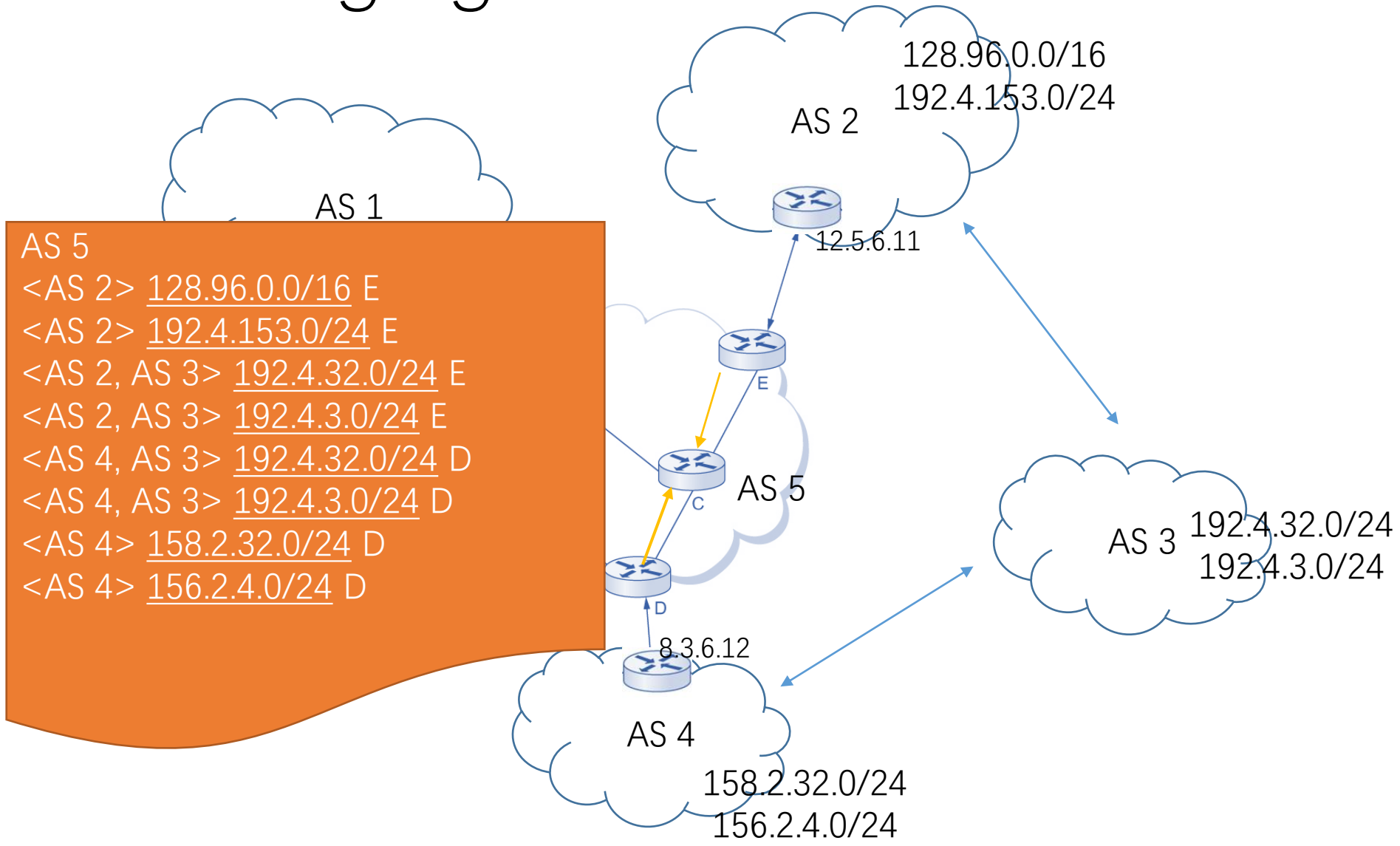
# BGP: Exchanging BGP entries



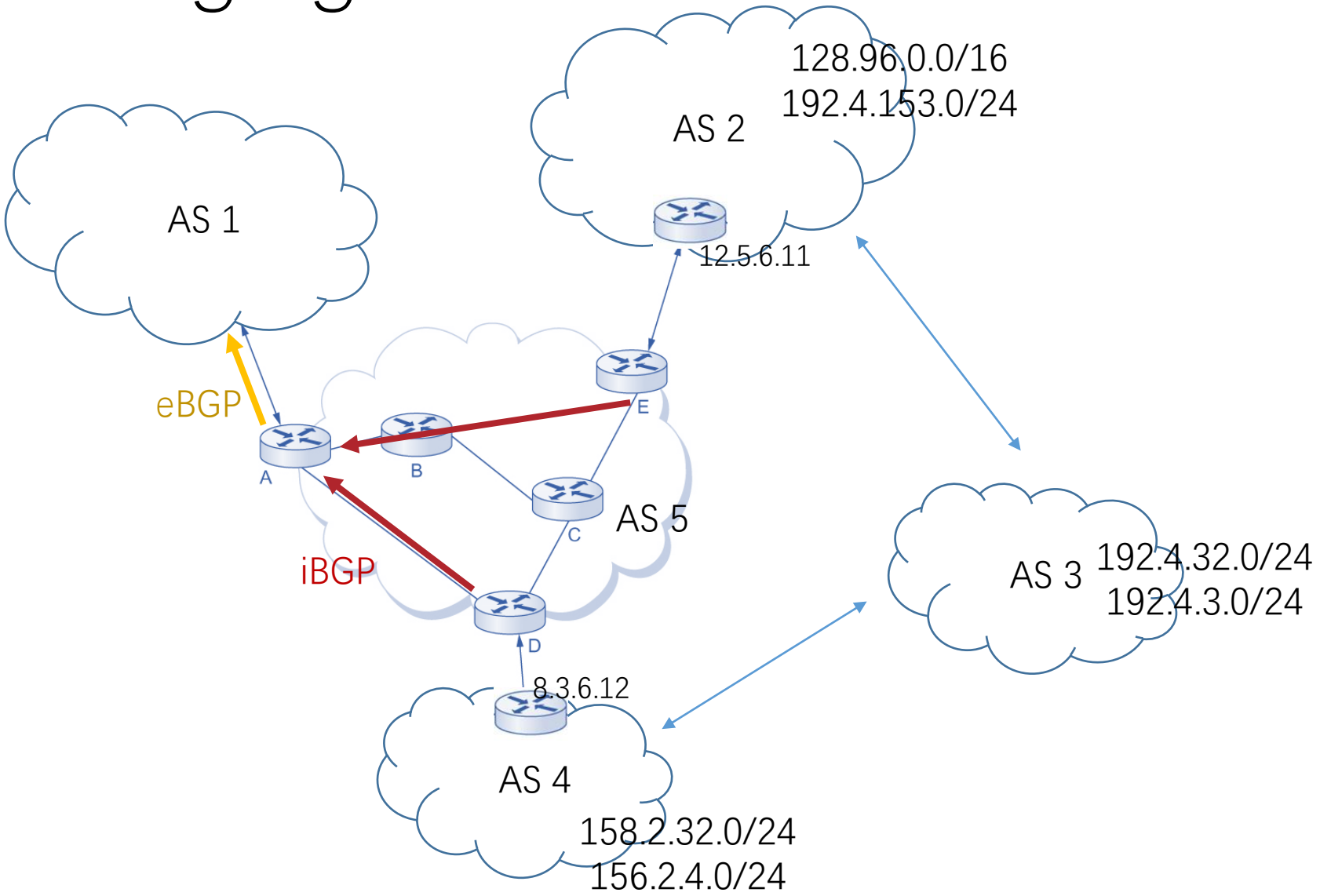
# BGP: Exchanging BGP entries



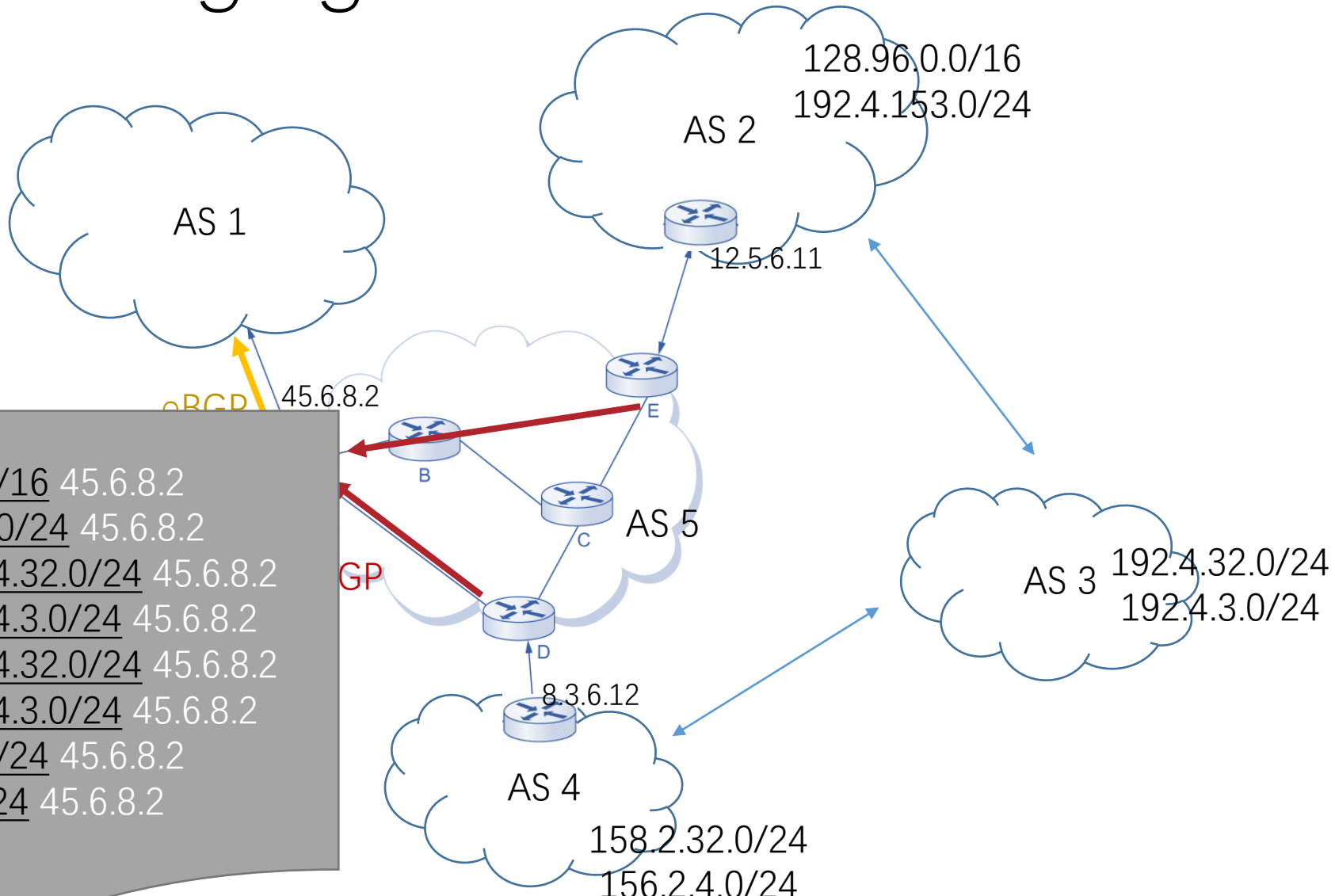
# BGP: Exchanging BGP entries



# BGP: Exchanging BGP entries

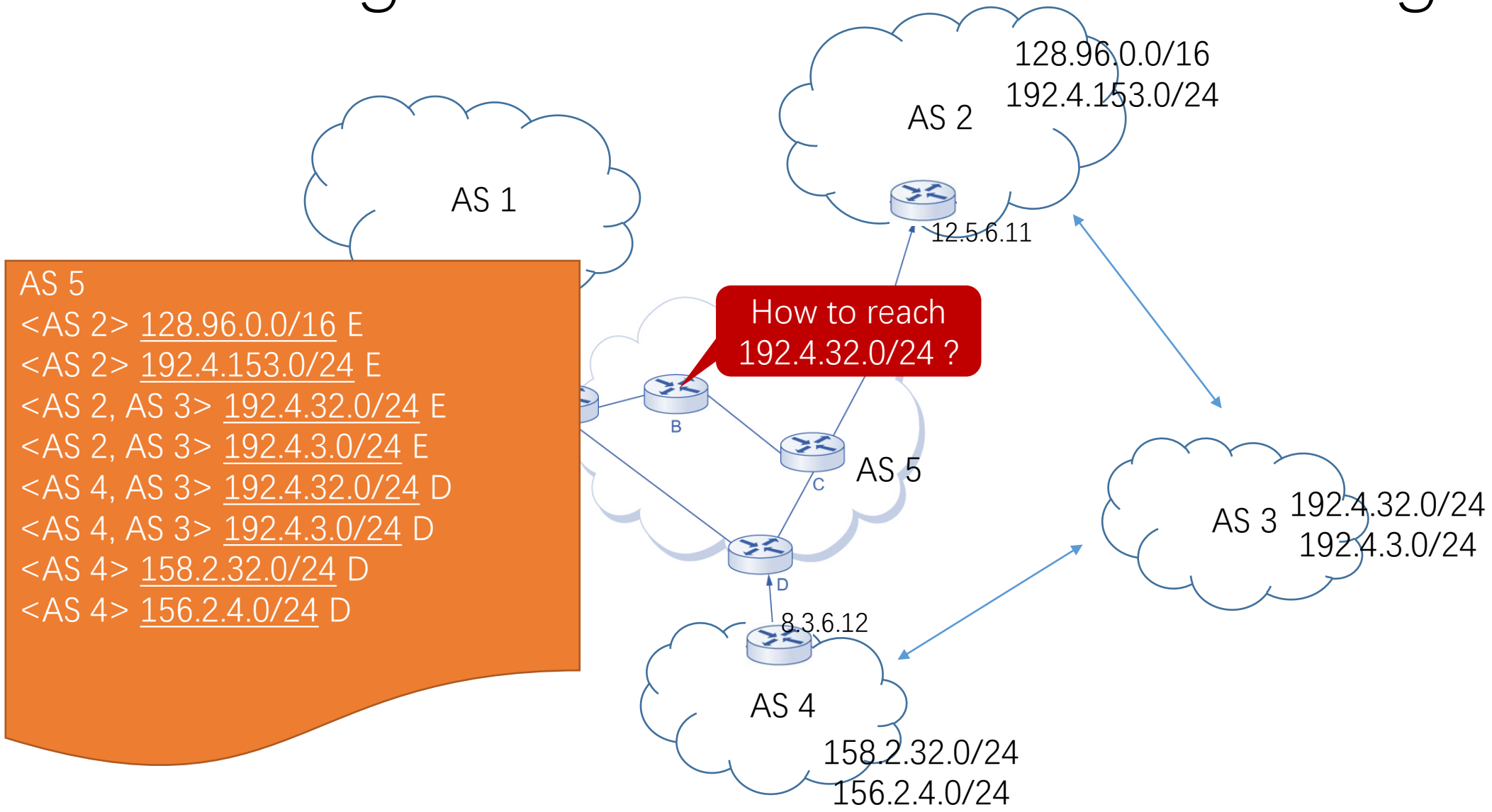


# BGP: Exchanging BGP entries




AS1
<AS 5, AS 2> 128.96.0.0/16 45.6.8.2
<AS 5, AS 2> 192.4.153.0/24 45.6.8.2
<AS 5, AS 2, AS 3> 192.4.32.0/24 45.6.8.2
<AS 5, AS 2, AS 3> 192.4.3.0/24 45.6.8.2
<AS 5, AS 4, AS 3> 192.4.32.0/24 45.6.8.2
<AS 5, AS 4, AS 3> 192.4.3.0/24 45.6.8.2
<AS 5, AS 4> 158.2.32.0/24 45.6.8.2
<AS 5, AS 4> 156.2.4.0/24 45.6.8.2

# BGP: Integrate with Intradomain Routing



# BGP: Integrate with Intradomain Routing

- Broadcast BGP entries in the AS through iBGP
- Roughly determine the best border router for certain prefix

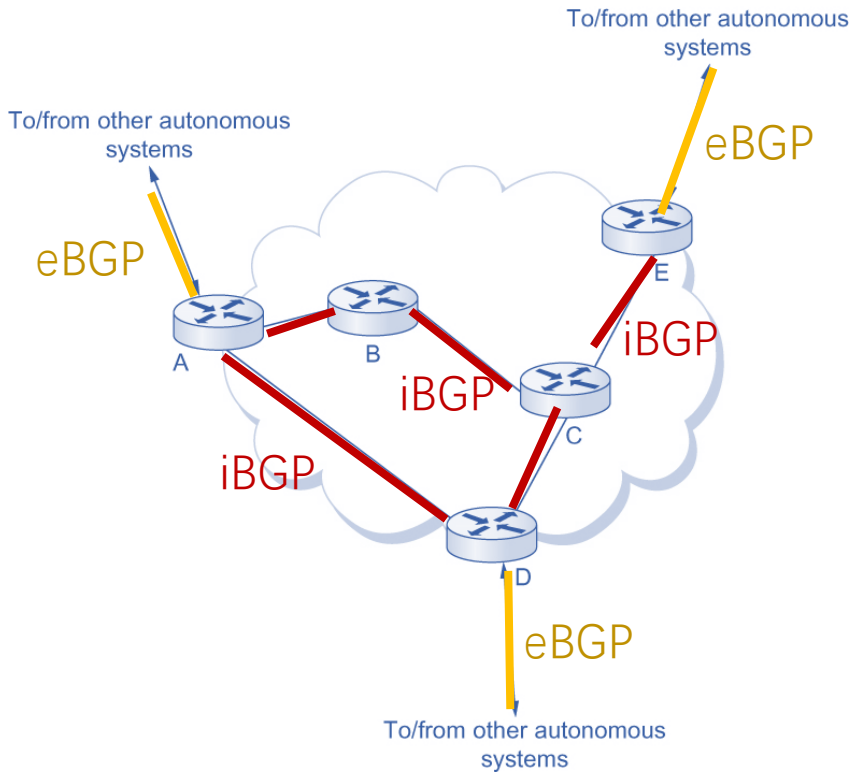
AS 5
<AS 2> 128.96.0.0/16 E
<AS 2> 192.4.153.0/24 E
<AS 2, AS 3> 192.4.32.0/24 E
<AS 2, AS 3> 192.4.3.0/24 E
<AS 4, AS 3> 192.4.32.0/24 D
<AS 4, AS 3> 192.4.3.0/24 D
<AS 4> 158.2.32.0/24 D
<AS 4> 156.2.4.0/24 D

BGP Table of B

| Dest | Next |
|---|---|
| A | A |
| C | C |
| D | C |
| E | C |

Routing Table of B

+




| Prefix | Next |
|---|---|
| 128.96.0.0/16 | C |
| 192.4.153.0/24 | C |
| 158.2.32.0/24 | C |
| 156.2.4.0/24 | C |

Part of Combined Table for B

# BGP: Integrate with Intradomain Routing

- Broadcast BGP entries in the AS through iBGP
- Roughly determine the best border router for certain prefix

AS 5
<AS 2> 128.96.0.0/16 E
<AS 2> 192.4.153.0/24 E
<AS 2, AS 3> 192.4.32.0/24 E
<AS 2, AS 3> 192.4.3.0/24 E
<AS 4, AS 3> 192.4.32.0/24 D
<AS 4, AS 3> 192.4.3.0/24 D
<AS 4> 158.2.32.0/24 D
<AS 4> 156.2.4.0/24 D

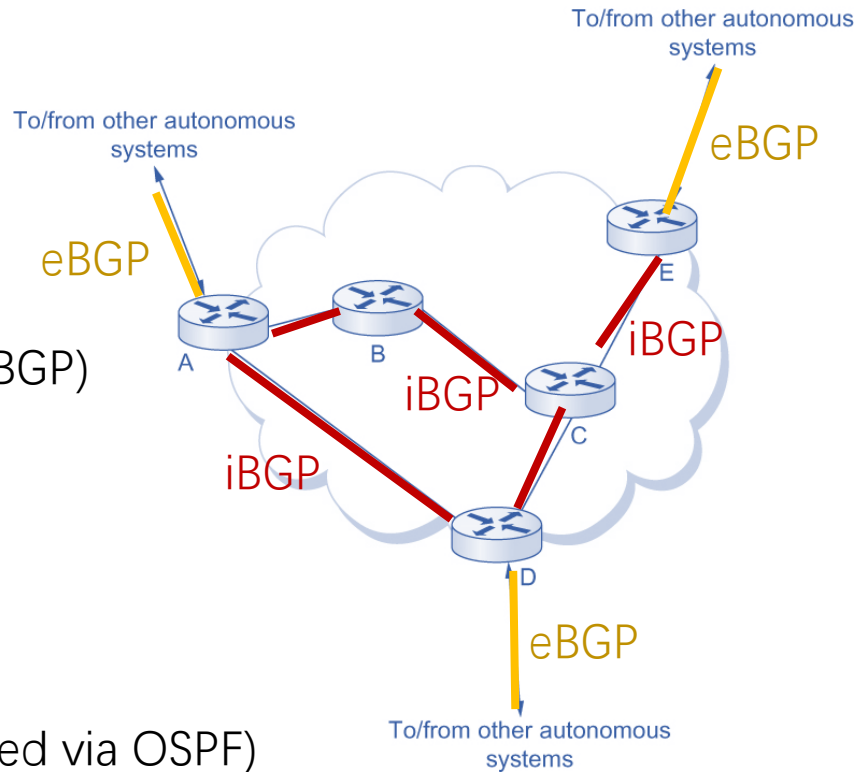
which one ?

BGP Table of B (obtained via BGP)

| Dest | Next |
|---|---|
| A | A |
| C | C |
| D | C |
| E | C |

Routing Table of B (e.g., obtained via OSPF)

+




| Prefix | Next |
|---|---|
| 128.96.0.0/16 | C |
| 192.4.153.0/24 | C |
| 158.2.32.0/24 | C |
| 156.2.4.0/24 | C |

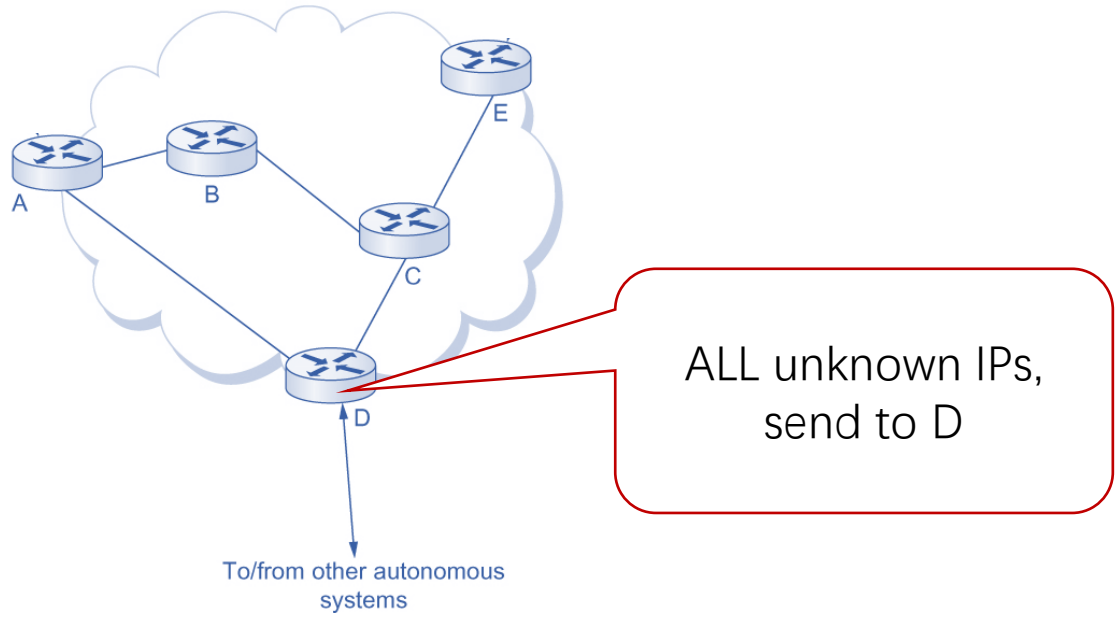
Part of Combined Table for B

# BGP: Integrate with Intradomain Routing

- Broadcast BGP entries are broadcast in the AS through iBGP
- Roughly determine the best border router for certain prefix
  - Selection Priority
    - Local Preference
    - AS hops
    - Distance to the border router
    - BGP ID

# BGP: Integrate with Intradomain Routing

- Other Methods: Default Gateway Router (Static Method)
  - Inject a default router entry into all routers in the AS through intradomain routing protocol



# Today's lecture

- Routing refresher
- Global internet hierarchy
- BGP basics
- Integrating BGP with intradomain routing
- Routing policies

# Interdomain Routing Problems

- Scalability: More than 600 million destinations
  - Storage
    - Routing Table
  - Calculation
    - Shortest Path
  - Communication
    - Exchanges Routing Information

➢Routing Management: Complex Routing Polices

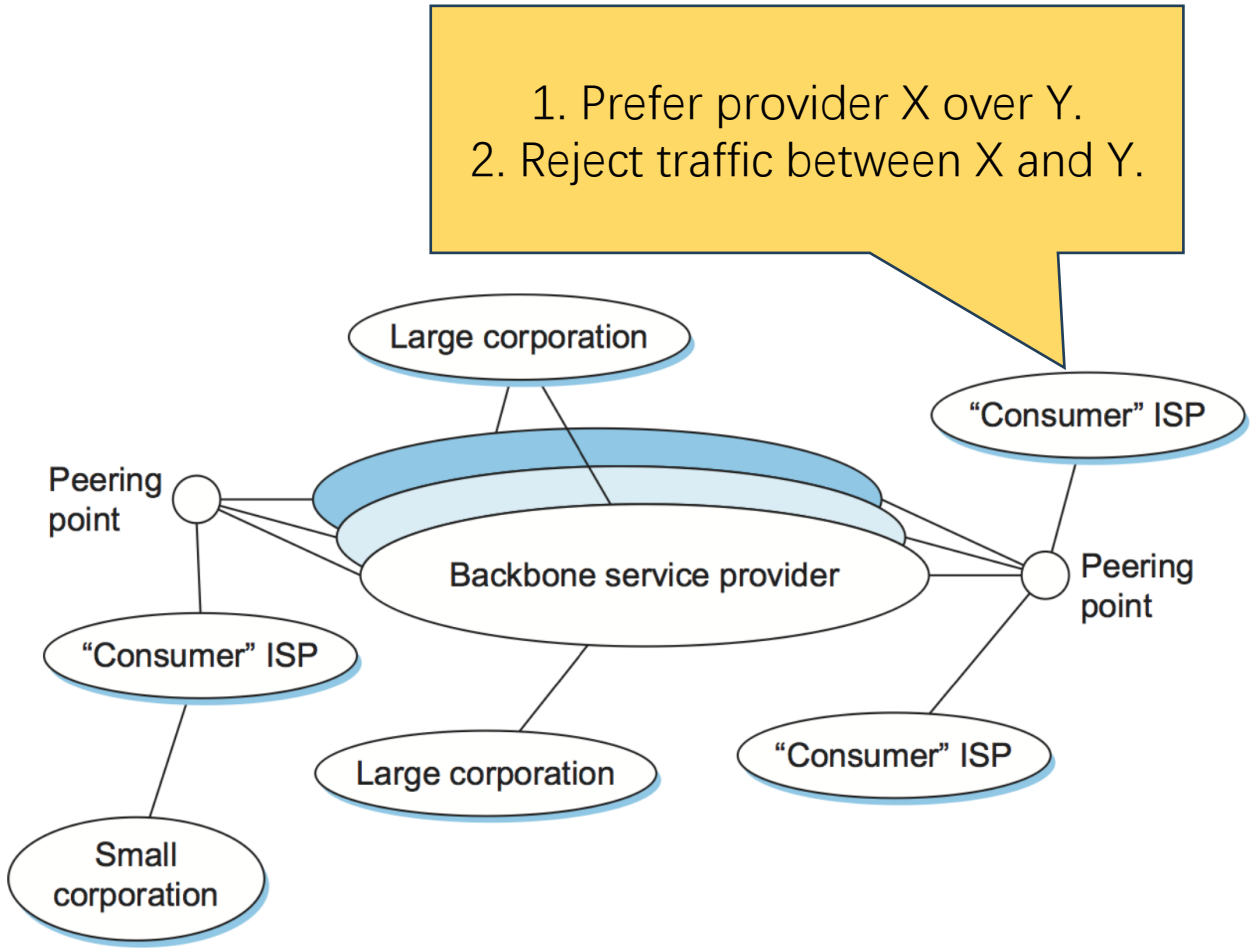
# Policies

We have established connectivity of the global internet!

But connectivity is not the only objective.

Business drives complex policies.

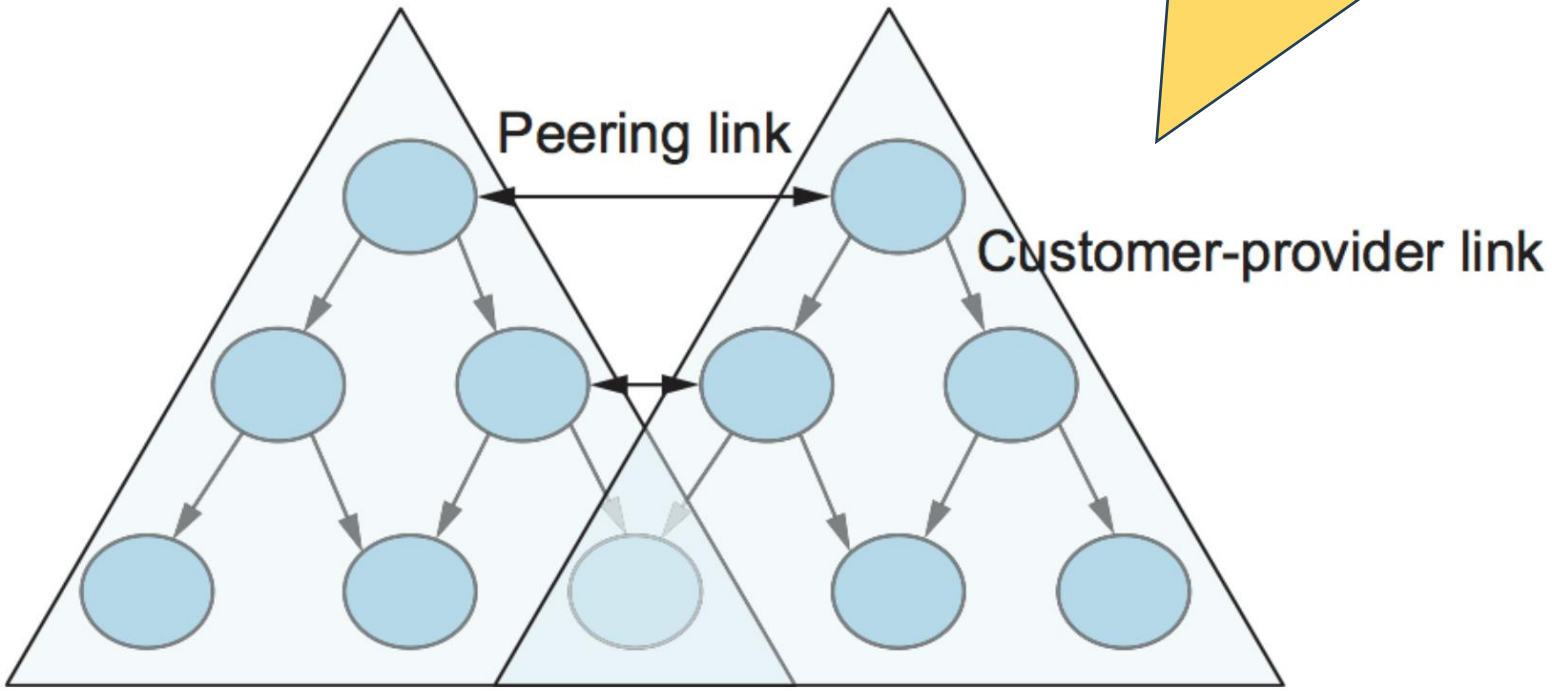
ASes need to implement policies without the help of each other.



# Policies

Provider to customer

advertise
- all the routes I know about to my customer
- routes learned from my customers to everyone



# Policies

Customer wants to:

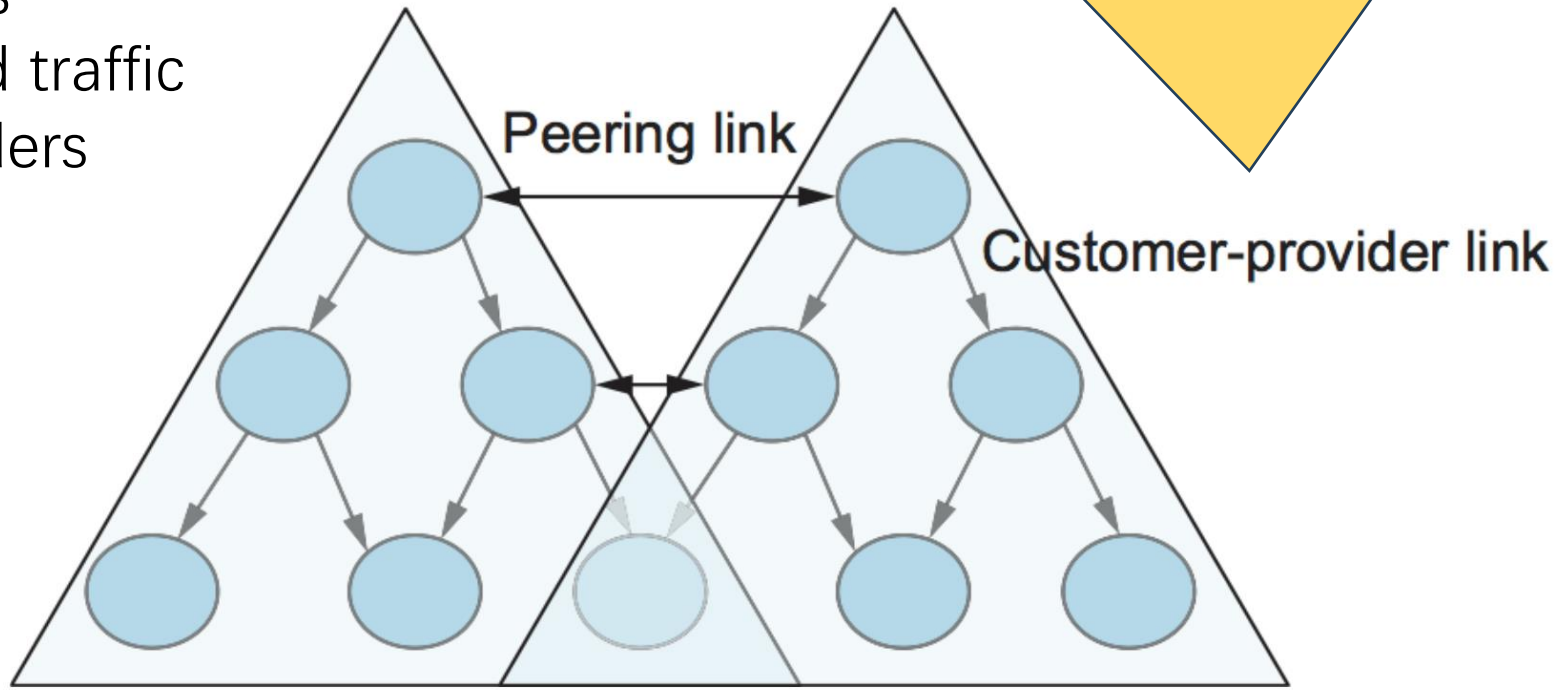
- get traffic directed to him and his customers
- be able to send traffic through its providers

advertise

- my own prefixes to my provider
- routes learned from my customers to my provider
- routes learned from my provider to my customers

don't advertise

- routes learned from one provider to another provider

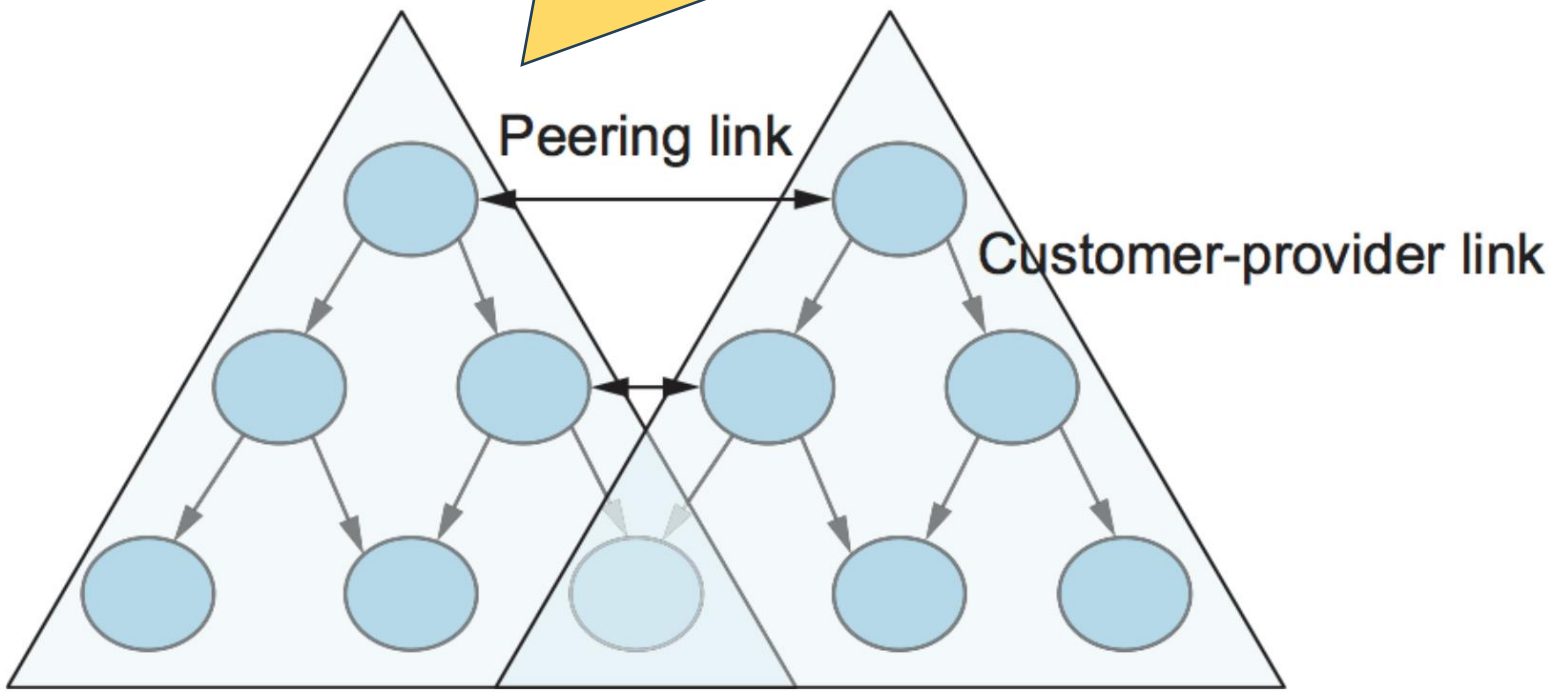


# Policies

Peers:
want to get access to each other's customers without paying other providers.

advertise:
- routes learned from my customers
- routes learned from my peer

don't advertise:
- routes from my peer to any provider



# Policy implementation: BGP attributes

BGP uses various attributes associated with each route to make routing decisions

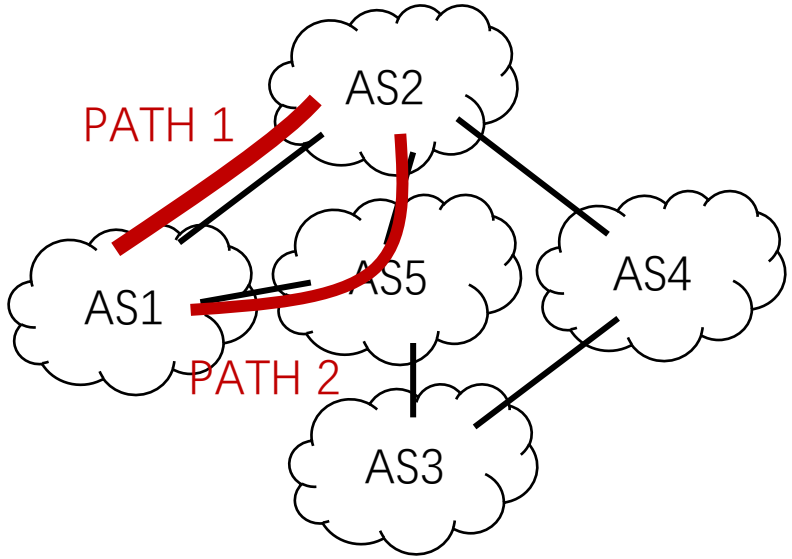
- **AS_PATH**: Represents the path the route has taken through different autonomous systems.
- **NEXT_HOP**: Indicates the IP address of the next-hop router to reach the destination.
- **LOCAL_PREF**: Used in the BGP path selection process within an AS.
- **MED** (Multi-Exit Discriminator): Influences routing decisions between neighboring ASes.

# AS path prepending

Lowers a path's preference with longer AS path.

- AS5 helps AS1 to forward traffic
  - ISP negotiation
  - Priority
    - eg. AS path prepending
- AS5 blocks AS1 traffic to AS2
  - Does not broadcast AS2 entry to AS1

```
AS path: 4637 4637 4637 4637 4538 4538 24364
```



# Takeways

- How to scale networks: introduce hierarchy
- Participants of BGP: Autonomous systems
  - much smaller than the number of networks in the world
- Hierarchy: tradeoff between scalability and optimality
  - BGP only finds policy-compliant, loop-free routes, not ”optimal” route

# Reference

- Textbook 4.1
- http://www.ciscopress.com/articles/article.asp?p=24090