Introduction to Computer Networks

CS640

Inter-domain Routing

https://pages.cs.wisc.edu/~mgliu/CS640/F22/

Ming Liu
mgliu@cs.wisc.edu

Today

Last lecture

- How to decide the forwarding path among routers?

Today

- How to decide the forwarding path among routers **at scale**?

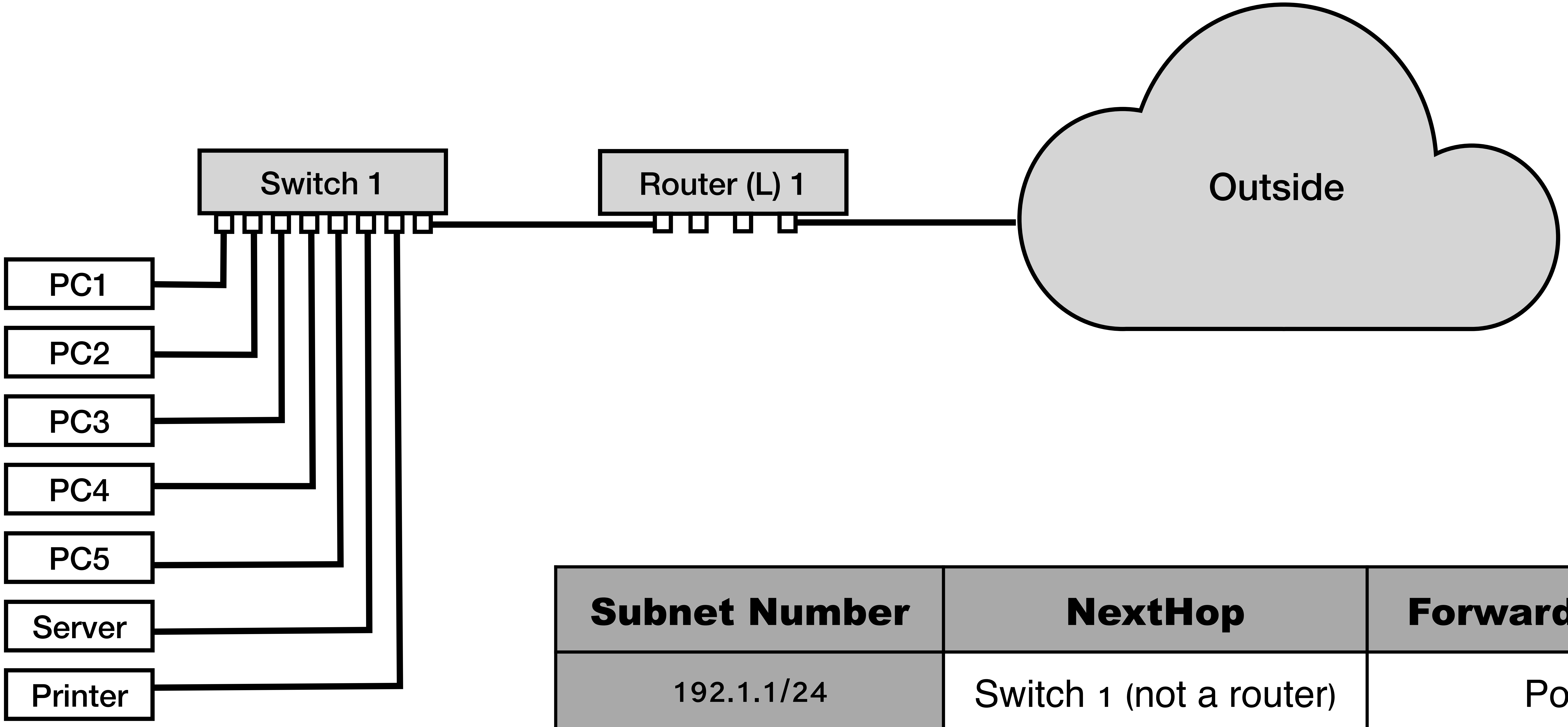
Announcements

- Lab3 is due 11/04/2022, 11:59 PM

Suppose you are building networks for your startup to satisfy the following requirements: (1) hosts within the startup can communicate with each other; (2) each host can talk to the outside.

- 5 Desktops
- 1 Printer
- 1 Web server

Device	# Ports	Per-port BW (Gbps)	Table size (#Entries)	Cost ($)
Low-end Router	4	1	128	2K
High-end Router	32	1	64K	100K
Ethernet Switch	8	1	512	1K

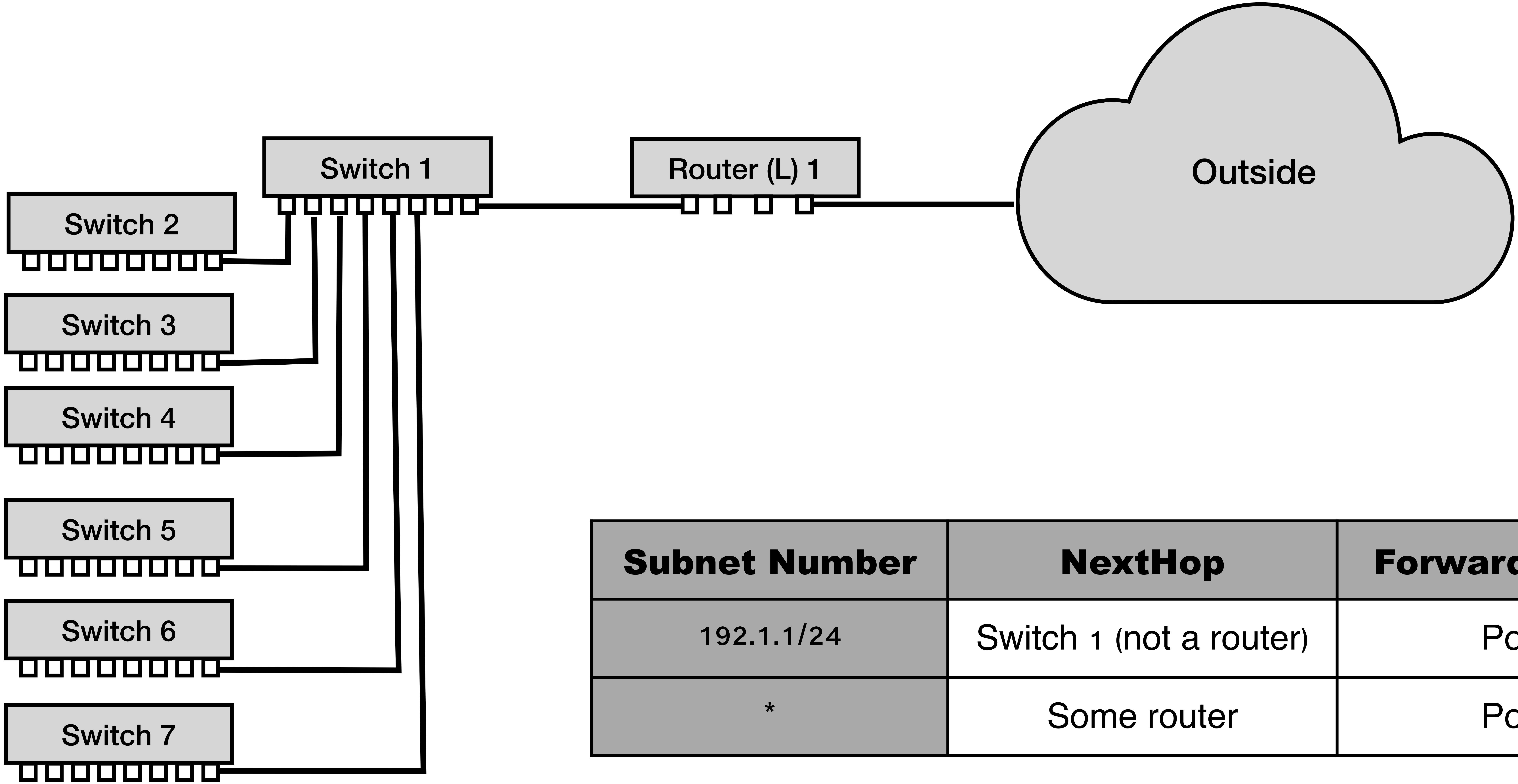


Subnet Number	NextHop	Forwarding Port
192.1.1/24	Switch 1 (not a router)	Port 0
*	Some router	Port 3

Suppose your startup grows and you need to provide more desktops for employees. Still, you are building networks to satisfy the following requirements: (1) hosts within the startup can communicate with each other; (2) each host can talk to the outside.

- 40 Desktops
- 1 Printer
- 1 Web server

Device	# Ports	Per-port BW (Gbps)	Table size (#Entries)	Cost ($)
Low-end Router	4	1	128	2K
High-end Router	32	1	128K	100K
Ethernet Switch	8	1	512	1K

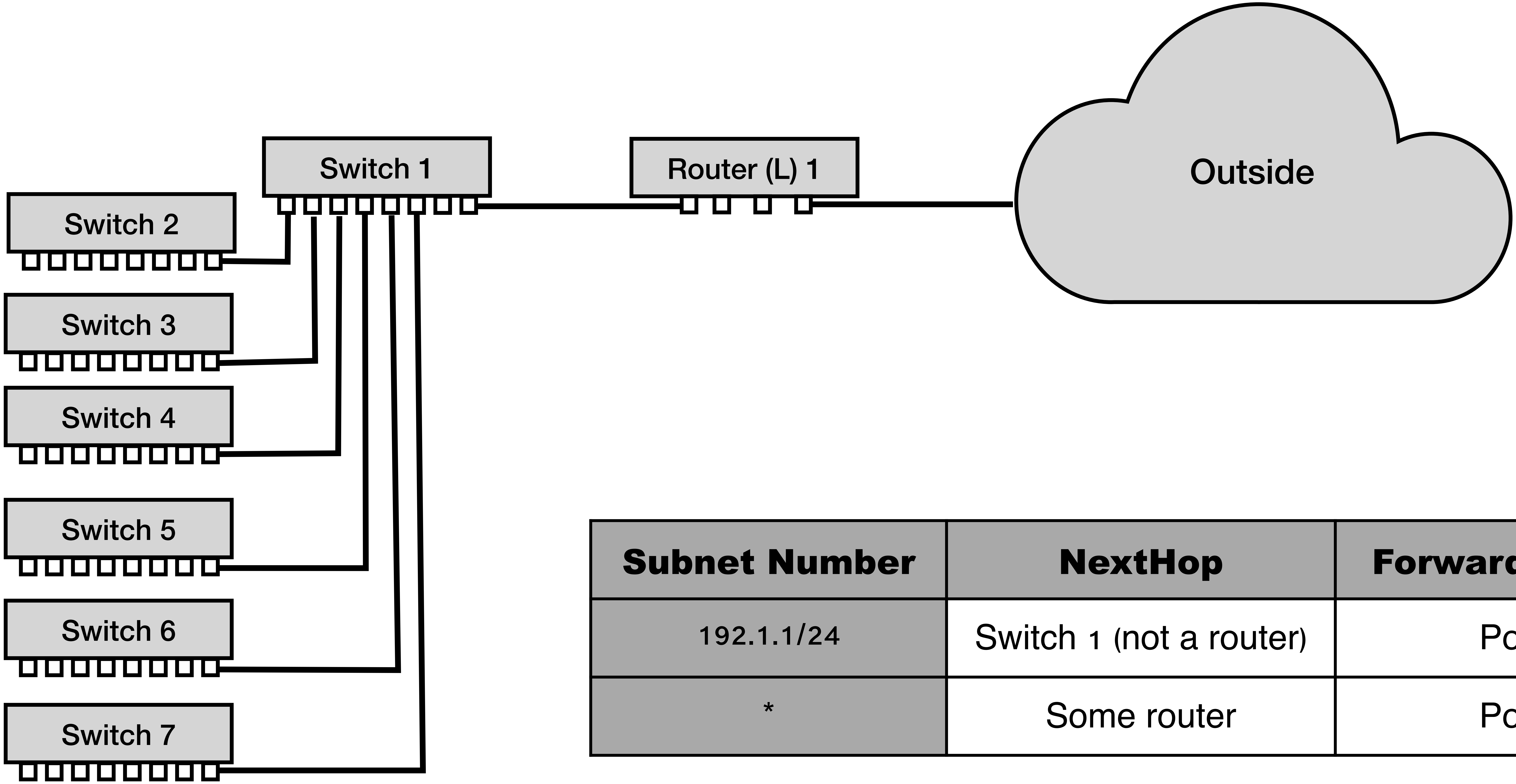


Subnet Number	NextHop	Forwarding Port
192.1.1/24	Switch 1 (not a router)	Port 0
*	Some router	Port 3

Suppose your startup continues to grow. So you decide to split the company into two groups: group A focuses on R&D; group B focuses on sales. You apply three class C addresses for two groups. Still, you are building networks to satisfy the following requirements: (1) hosts within the startup can communicate with each other; (2) each host can talk to the outside.

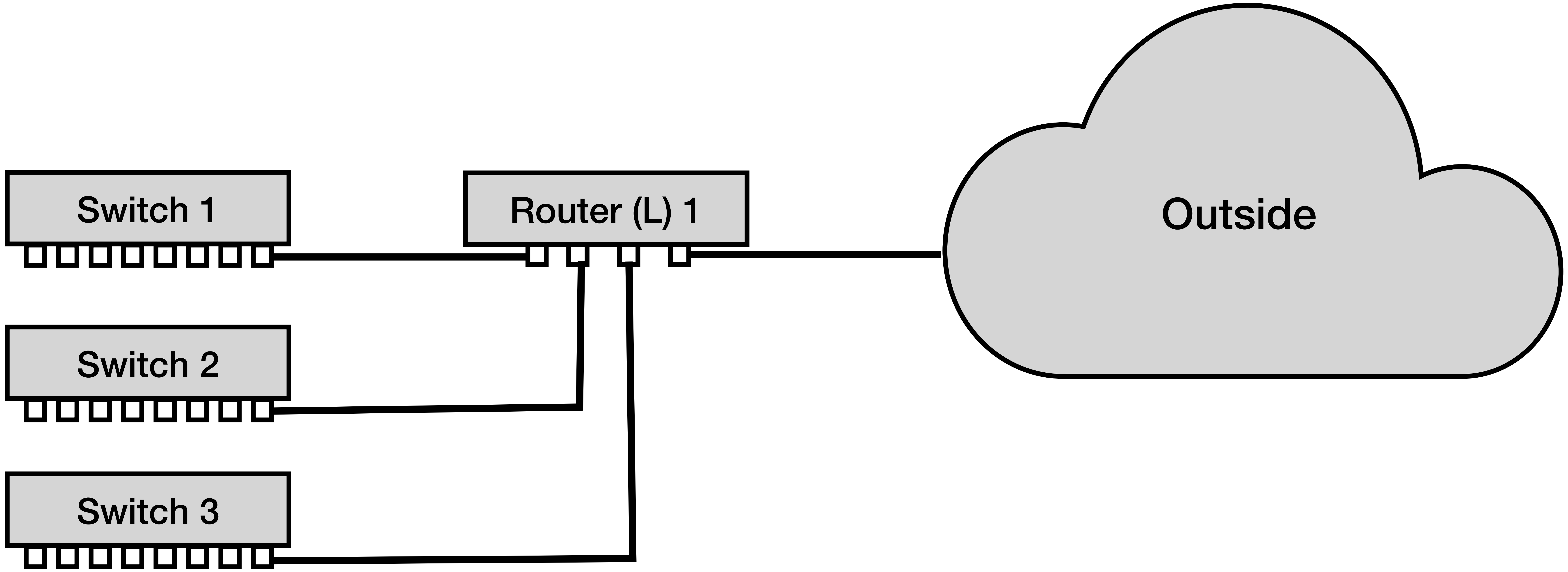
- 400 Desktops (group A) + 100 Desktops (group B)
- 1 Printer (group B)
- 1 Web server (group B)

Device	# Ports	Per-port BW (Gbps)	Table size (#Entries)	Cost ($)
Low-end Router	4	1	128	2K
High-end Router	32	1	128K	100K
Ethernet Switch	8	1	512	1K



Subnet Number	NextHop	Forwarding Port
192.1.1/24	Switch 1 (not a router)	Port 0
*	Some router	Port 3

More switches?

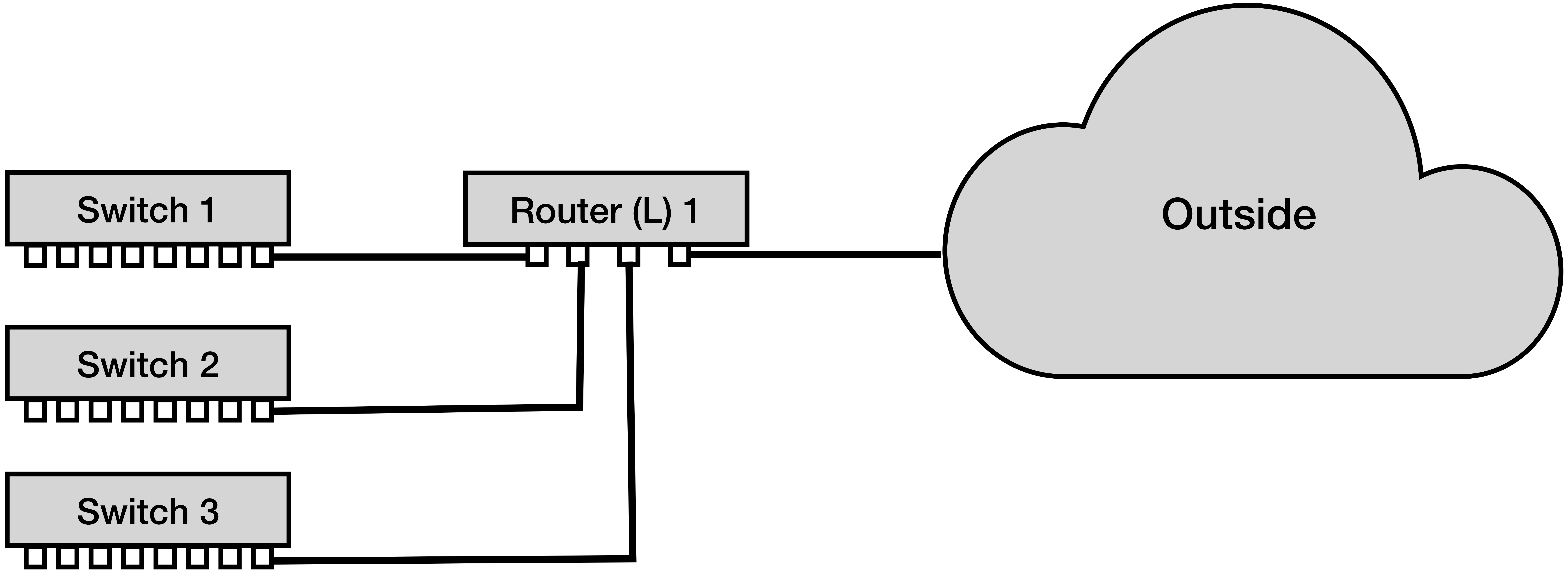


Subnet Number	NextHop	Forwarding Port
192.1.1/24	Switch 1 (not a router)	Port 0
192.1.2/24	Switch 2 (not a router)	Port 1
192.1.3/24	Switch 3 (not a router)	Port 2
*	Some router	Port 3

Suppose your startup expands significantly. There are 10 subdivisions that share 200 class C addresses. Still, you are building networks to satisfy the following requirements: (1) hosts within the startup can communicate with each other; (2) each host can talk to the outside.

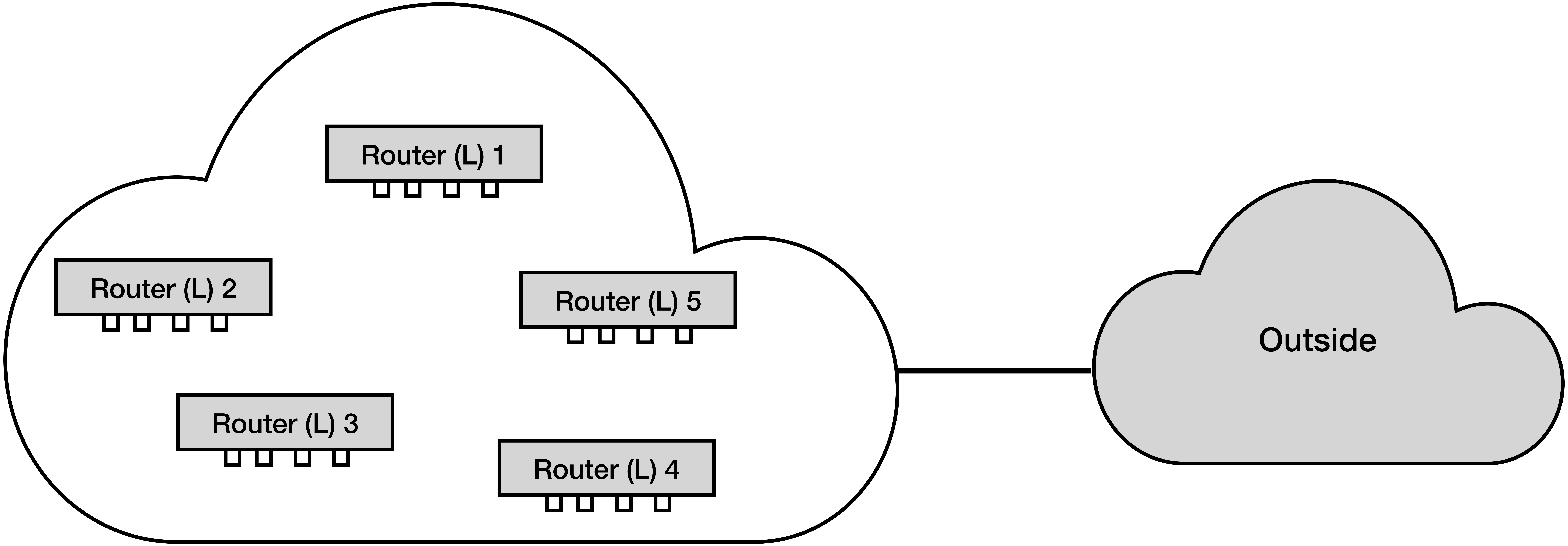
- 10^4 Desktops
- 10 Printers
- 10^2 Web servers

Device	# Ports	Per-port BW (Gbps)	Table size (#Entries)	Cost ($)
Low-end Router	4	1	128	2K
High-end Router	32	1	128K	100K
Ethernet Switch	8	1	512	1K



Subnet Number	NextHop	Forwarding Port
192.1.1/24	Switch 1 (not a router)	Port 0
192.1.2/24	Switch 2 (not a router)	Port 0
192.1.3/24	Switch 3 (not a router)	Port 0
*	Some router	Port 3

Does this work?


Router (L) 1
Router (L) 2
Router (L) 5
Router (L) 3
Router (L) 4
Outside

Q: What factors decide the scale of a network?

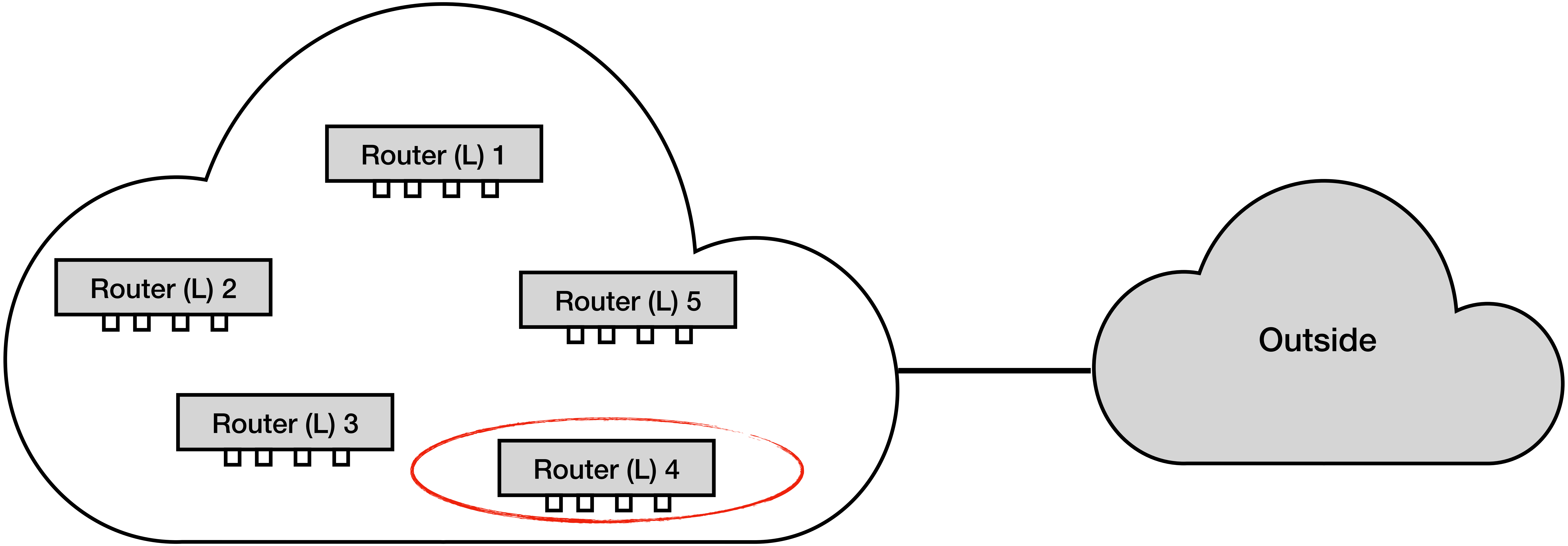
Q: What factors decide the scale of a network?

A: Four factors

- #1: The number of hosts
- #2: The aggregated size of all forwarding tables
- #3: Bandwidth requirement of host-host communications
- #4: The number of subnets

L6

Q: How to transmit a packet reliably between two NICs in a small-scaled network?



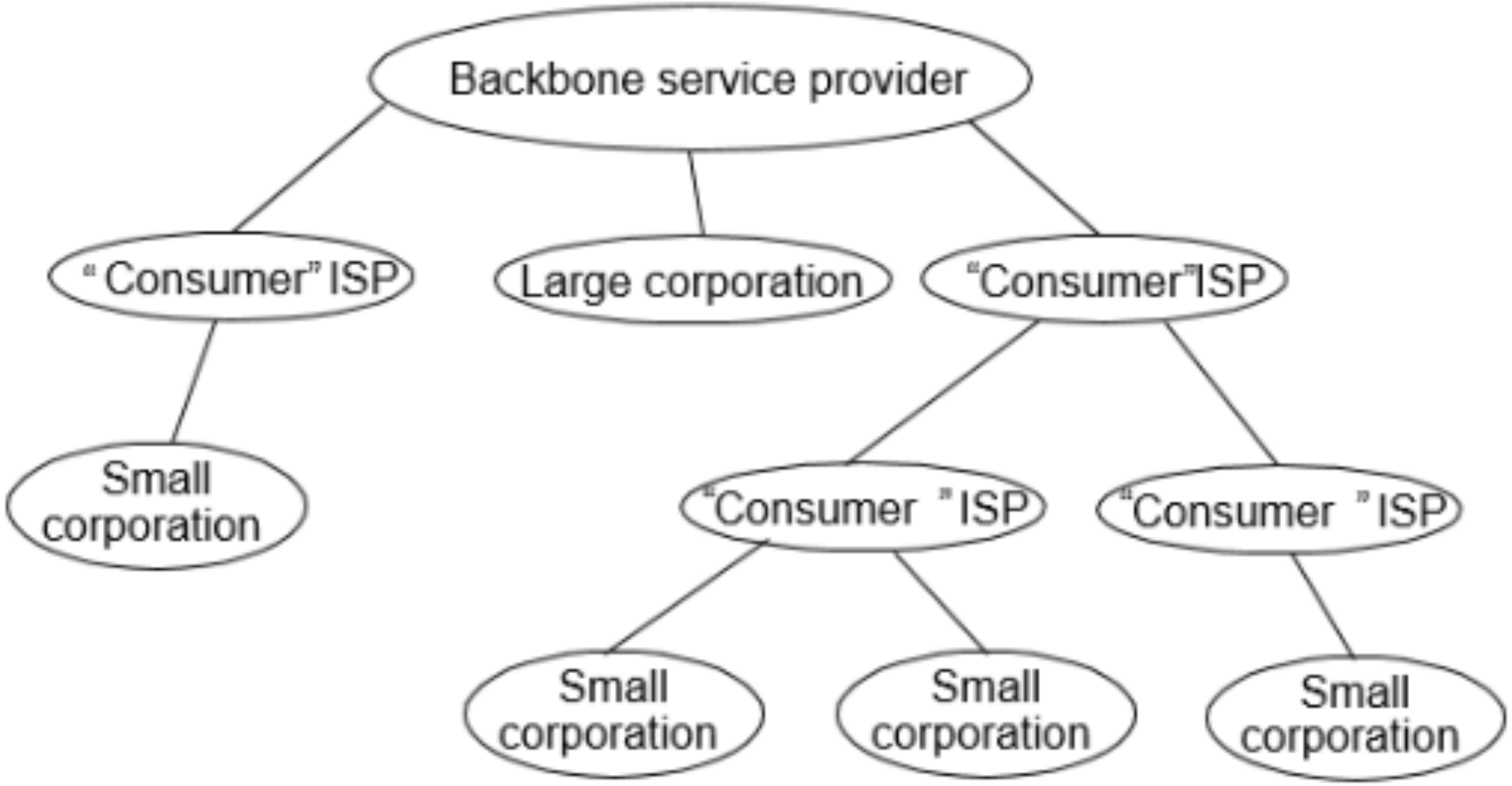
One or several special routers (or gateways) forwarding both internal/external traffic

Suppose the outside is also one such startup. How do we build networks between these two startups?

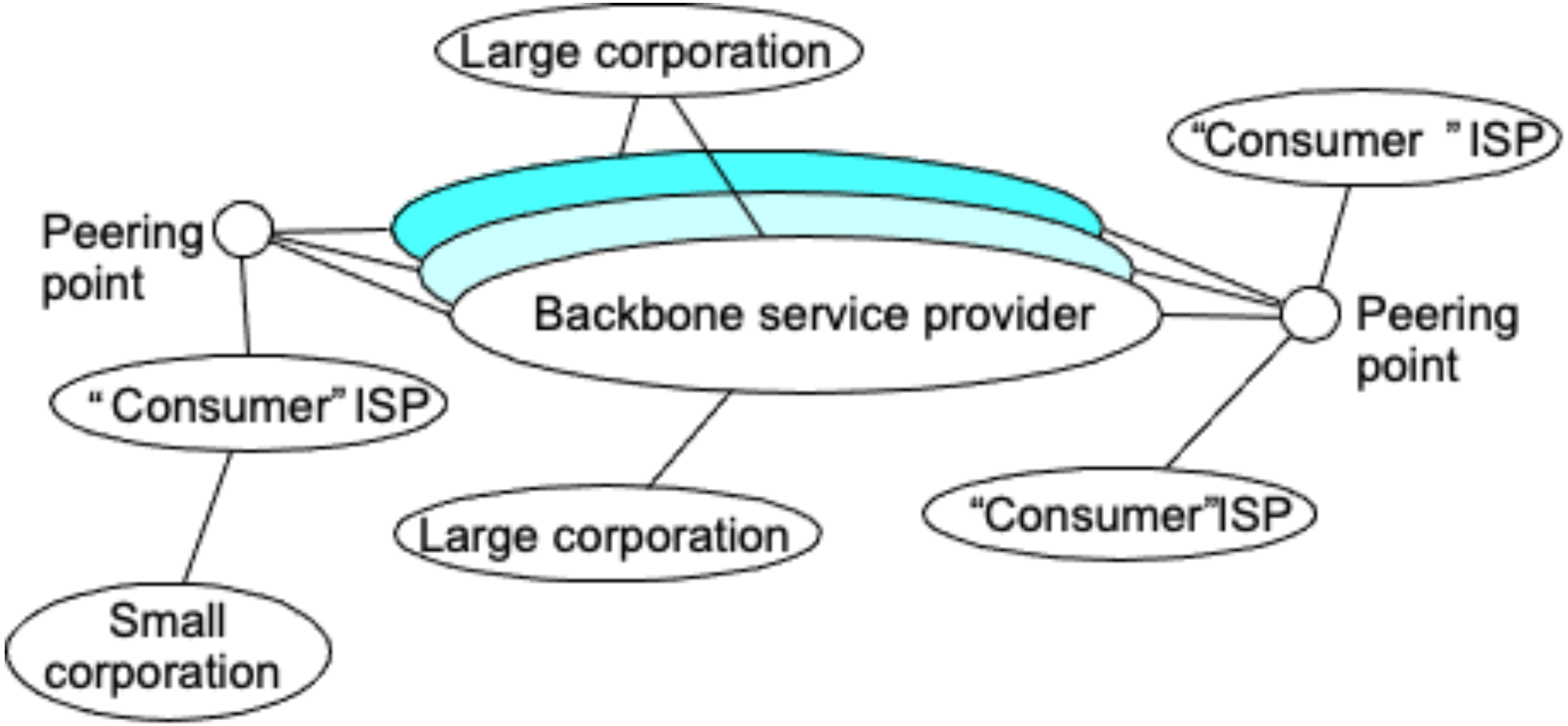
Suppose the outside is millions of such startups. How do we build networks among them?

Device	# Ports	Per-port BW (Gbps)	Table size (#Entries)	Cost ($)
Low-end Router	4	1	128	2K
High-end Router	32	1	128K	100K
Ethernet Switch	8	1	512	1K

Internet Structure (Original Idea)



Internet Structure (Today)



Q: Can we use RIP/OSPF to achieve routing at such a scale?

Q: Can we use RIP/OSPF to achieve routing at such a scale?

A: No.

- #1: Scalability — a huge amount of routers involved
- #2: Privacy— Networking hardware has ownership

L2

Networking hardware has ownership

- The fabric is build and maintained by network providers

Routing in the Internet

Autonomous System (AS)

- Corresponds to an administrative domain
- Examples: University, company, backbone network, your startup,…
- Assign each AS a 16-bit number

Two-level routing hierarchy

- interior gateway protocol (each AS selects its own)
- exterior gateway protocol (Internet-wide standard)

Key Idea of Route Propagation in the Internet

Route information is propagated at various levels

- Hosts know local router
- Local routers know site routers
- Site routers know core router
- Core routers know everything

Popular Interior Gateway Protocols

RIP: Router Information Protocol

- Distance-vector algorithm
- Cost is based on #hops

OSPF: Open Shortest Path First

- Link-state algorithm
- Supports load balancing and authentication

Border Gateway Protocol

BGP-1 was developed in 1989 to address problems with EGP (Exterior Gateway Protocol)

Current version: BGP-4

Assumption: The Internet is an arbitrarily interconnected set of ASes

Autonomous System (AS)

AS traffic types

- Local: starts or ends within an AS
- Transit: passes through an AS

AS types

- stub AS: has a single connection to one other AS
  - carries local traffic only
- multi-homed AS: has connections to more than one AS
  - refuses to carry transit traffic
- transmit AS: has connections to more than one AS
  - carries both transmit and local traffic

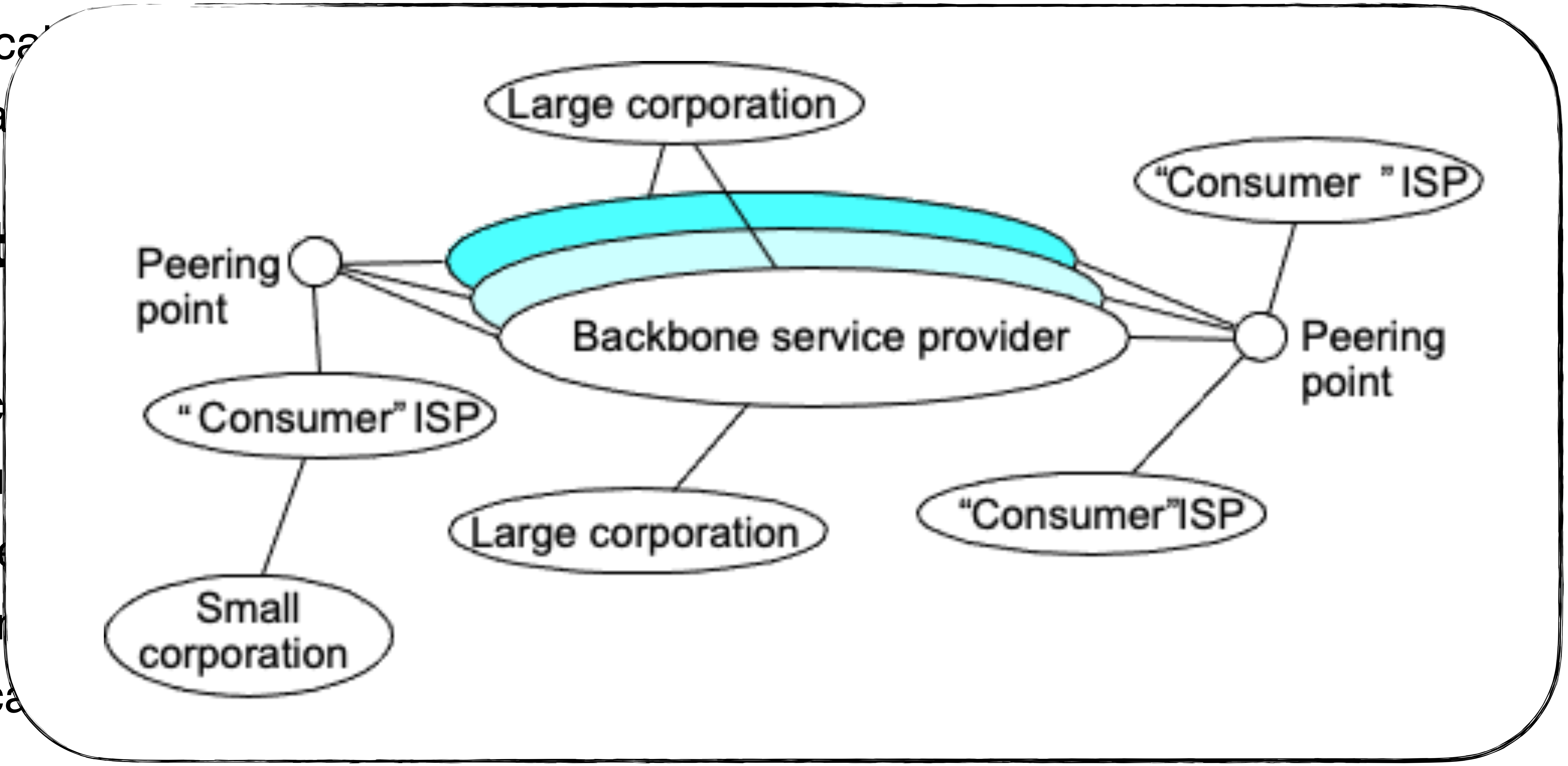
Autonomous System (AS)

AS traffic types

- Loca
- Tra

AS t

- stu
  - c
- mu
  - re
- tra
  - ca



AS Characteristics

#1: Each AS has one or more border routers

- Handles inter-AS traffic

#2: At least one BGP speaker for an AS that participates in routing

- Border routers might or might not be BGP speakers

AS Characteristics

#3: BGP speaker establishes BGP sessions with peers and advertises route information

- Local network names
- Other reachable networks (transit AS only)
- Give path information – AS Path, or Path vector
- Withdraw routes

AS Characteristics

#3: BGP speaker establishes BGP sessions with peers and advertises route information

- Local network names
- Other reachable networks (transit AS only)
- Give path information – AS Path, or Path vector
- Withdraw routes

- Peers: neighbor routers exchange routing information
- Advertises: an AS publicizes its learned routing information

AS Characteristics

#3: BGP speaker establishes BGP sessions with peers and advertises route information

- Local network names
- Other reachable networks (transit AS only)
- Give path information – AS Path, or Path vector
- Withdraw routes

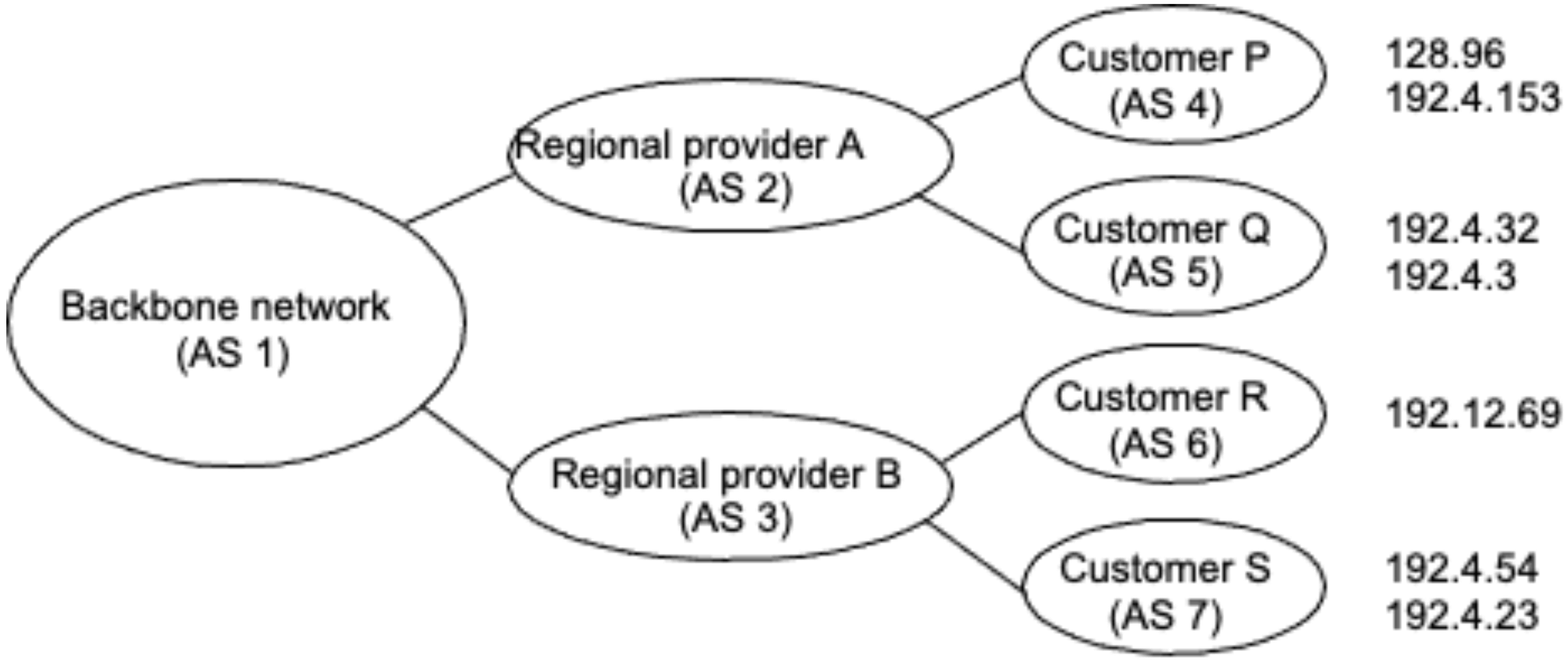
- Peers: neighbor routers exchange routing information
- Advertises: an AS publicizes its learned routing information

- Unlike RIP and OSPF, BGP advertises complete path as an enumerated list of autonomous systems to reach a particular network

BGP Example

Speaker for AS2 advertises reachability to P and Q

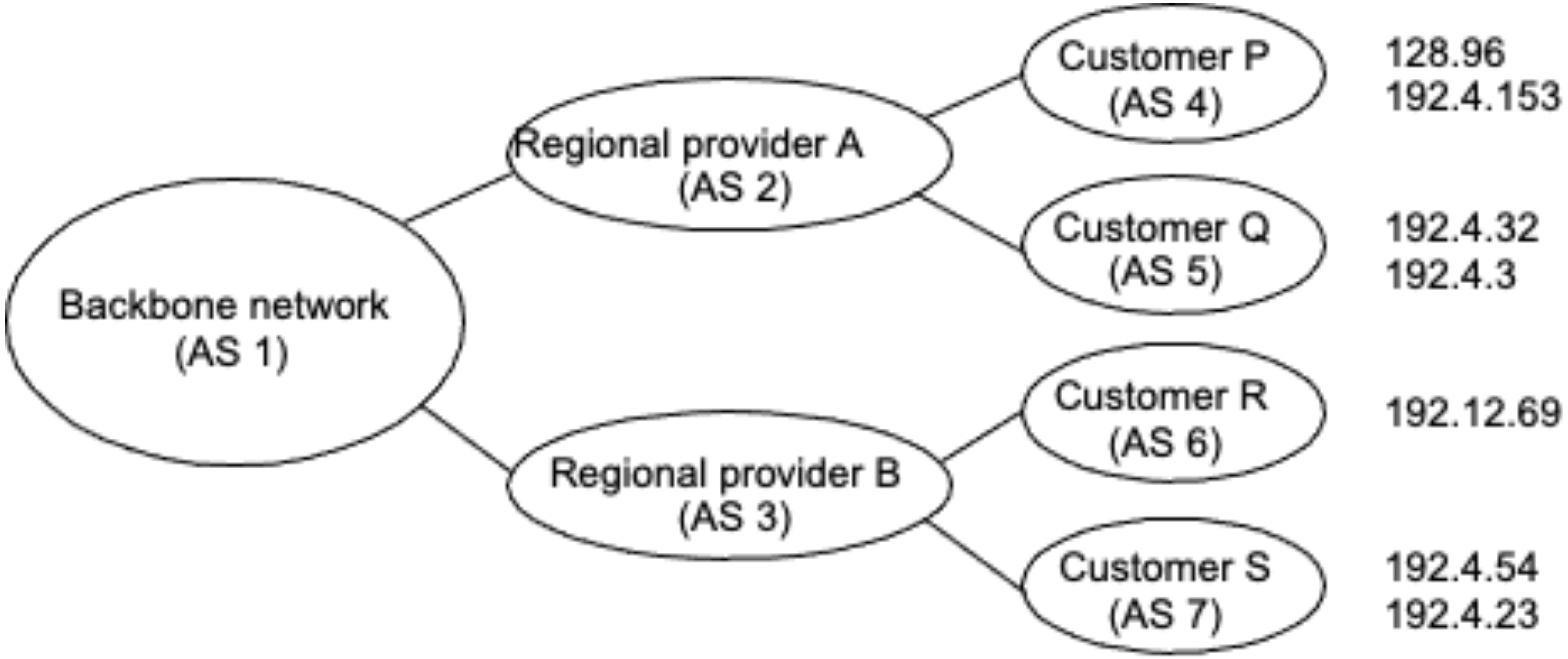
- Network 128.96, 192.4.153, 192.4.32, and 192.4.3, can be reached directly from AS2



BGP Example

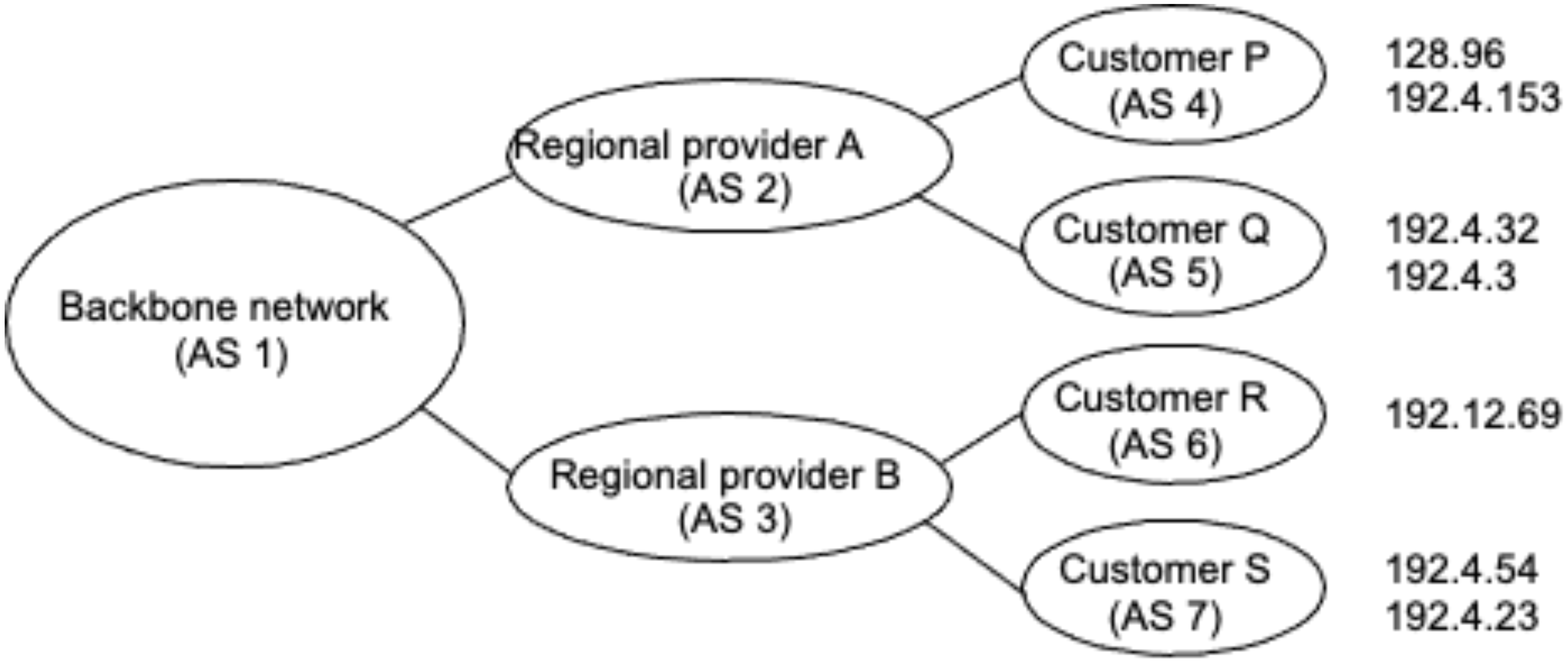
Speaker for backbone advertises

- Network 128.96, 192.4.153, 192.4.32, and 192.4.3, can be reached directly from (AS1, AS2)



BGP Example

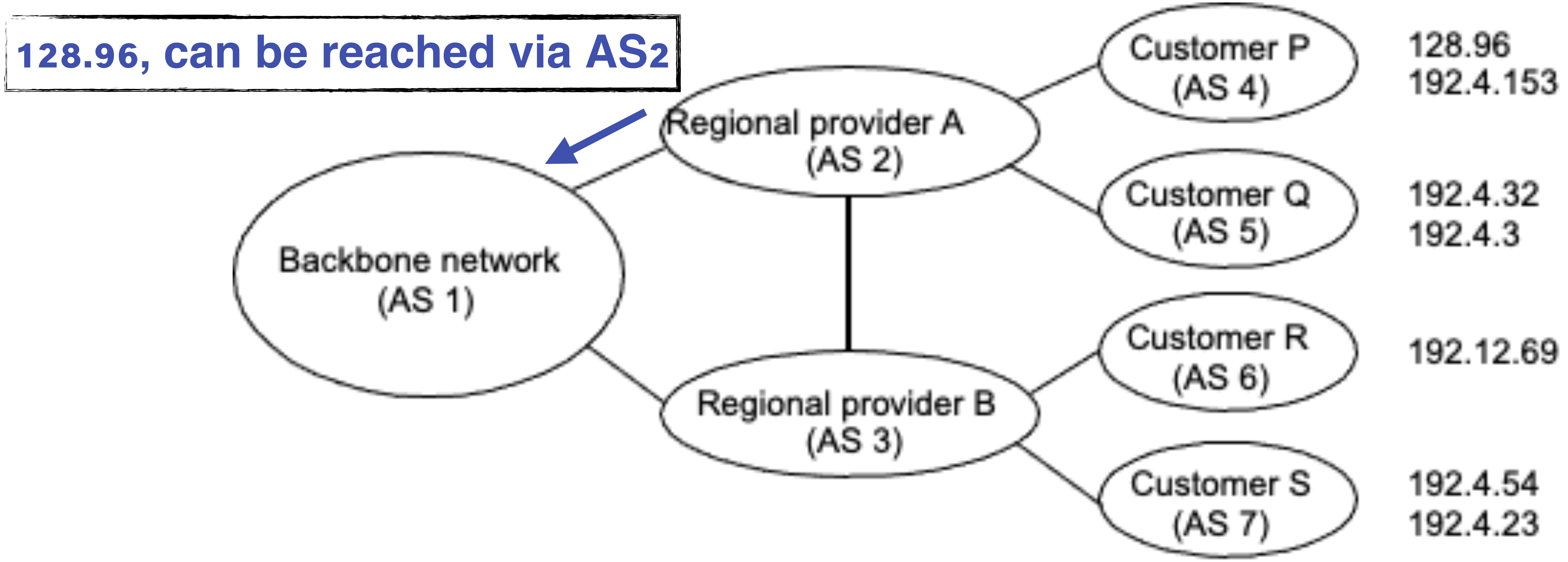
Speaker can cancel previously advertised paths



BGP Goal

Find loop free paths between ASes

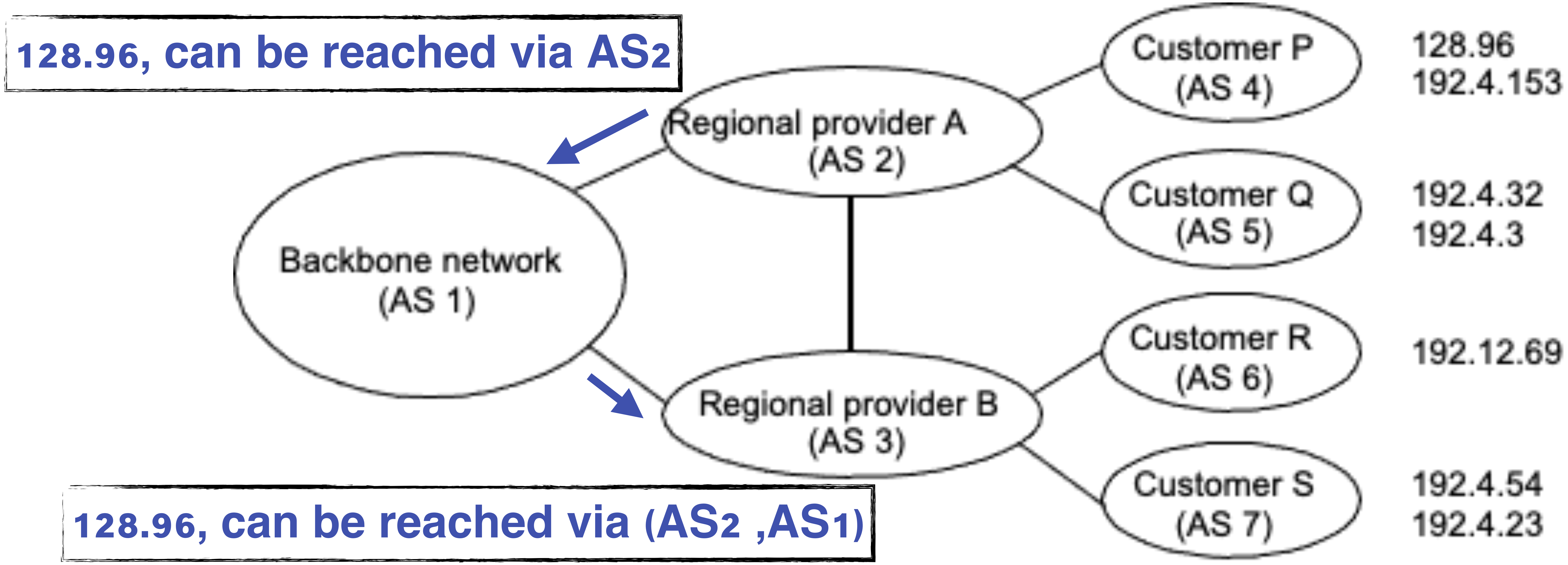
- Optimality is secondary goal
- It's neither a distance-vector nor a link-state protocol



BGP Goal

Find loop free paths between ASes

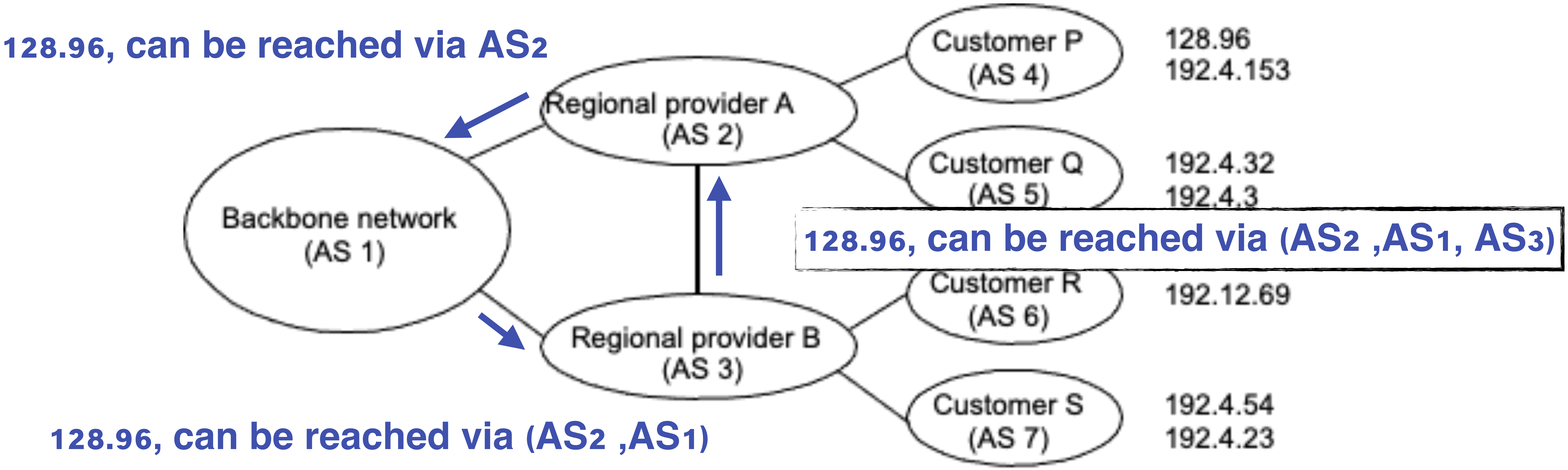
- Optimality is secondary goal
- It's neither a distance-vector nor a link-state protocol



BGP Goal

Find loop free paths between ASes

- Optimality is secondary goal
- It's neither a distance-vector nor a link-state protocol

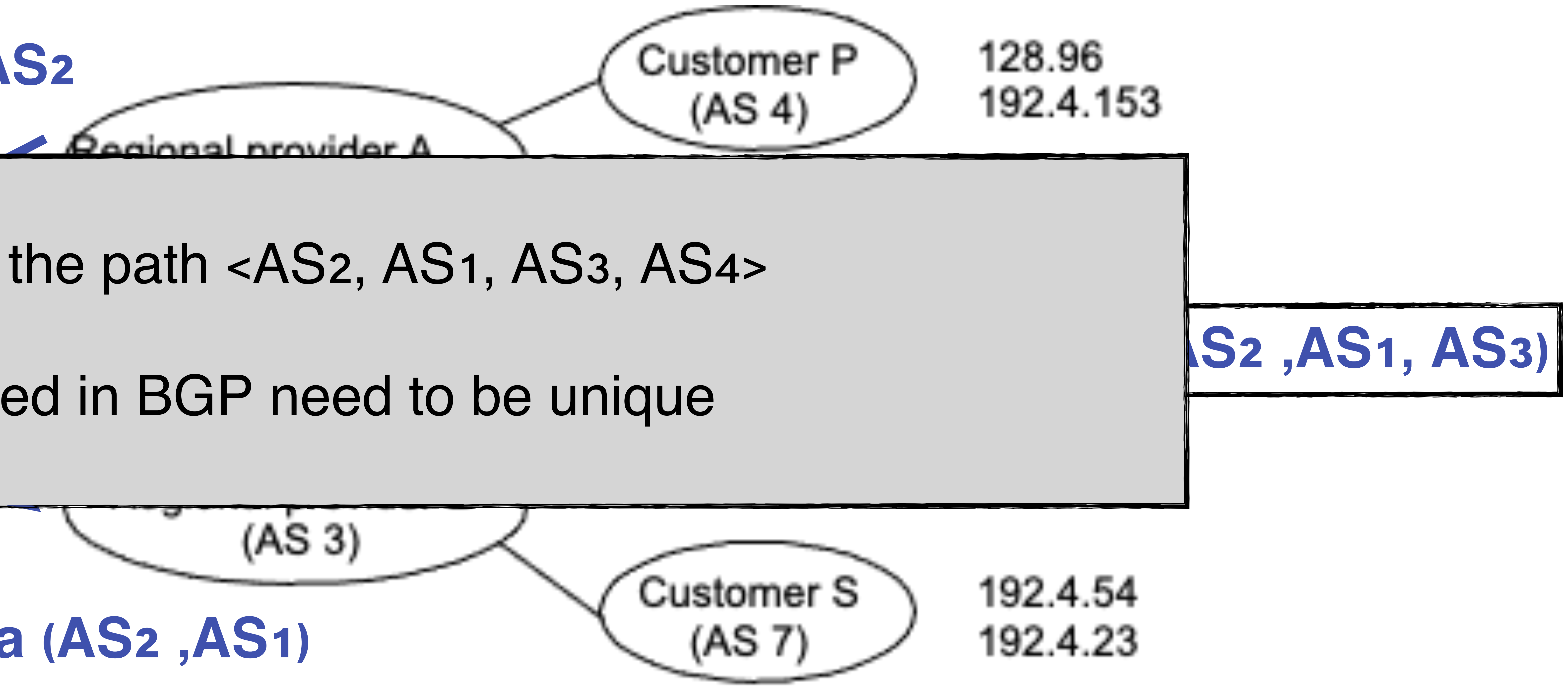


BGP Goal

Find loop free paths between ASes

- Optimality is secondary goal
- It's neither a distance-vector nor a link-state protocol

128.96, can be reached via AS2



AS2 sees itself in the path <AS2, AS1, AS3, AS4>

AS numbers carried in BGP need to be unique

128.96, can be reached via (AS2 ,AS1)

BGP Goal

Find loop free paths between ASes

- Optimality is secondary goal
- It's neither a distance-vector nor a link-state protocol

Challenges

- Internet's size (~12K active ASes) means large tables in BGP routers
- **Policy-compliant** path (not just scalar cost of a path)
- Autonomous domains mean different path metrics
- Trust among different ASes

Q: How does BGP work?

Q: How does BGP work?

A: Policy management

- #1: Learn — Import routing information from my neighbors
- #2: Speak — Export routing information to my neighbors

Policy in BGP

BGP provides the capability for enforcing policies

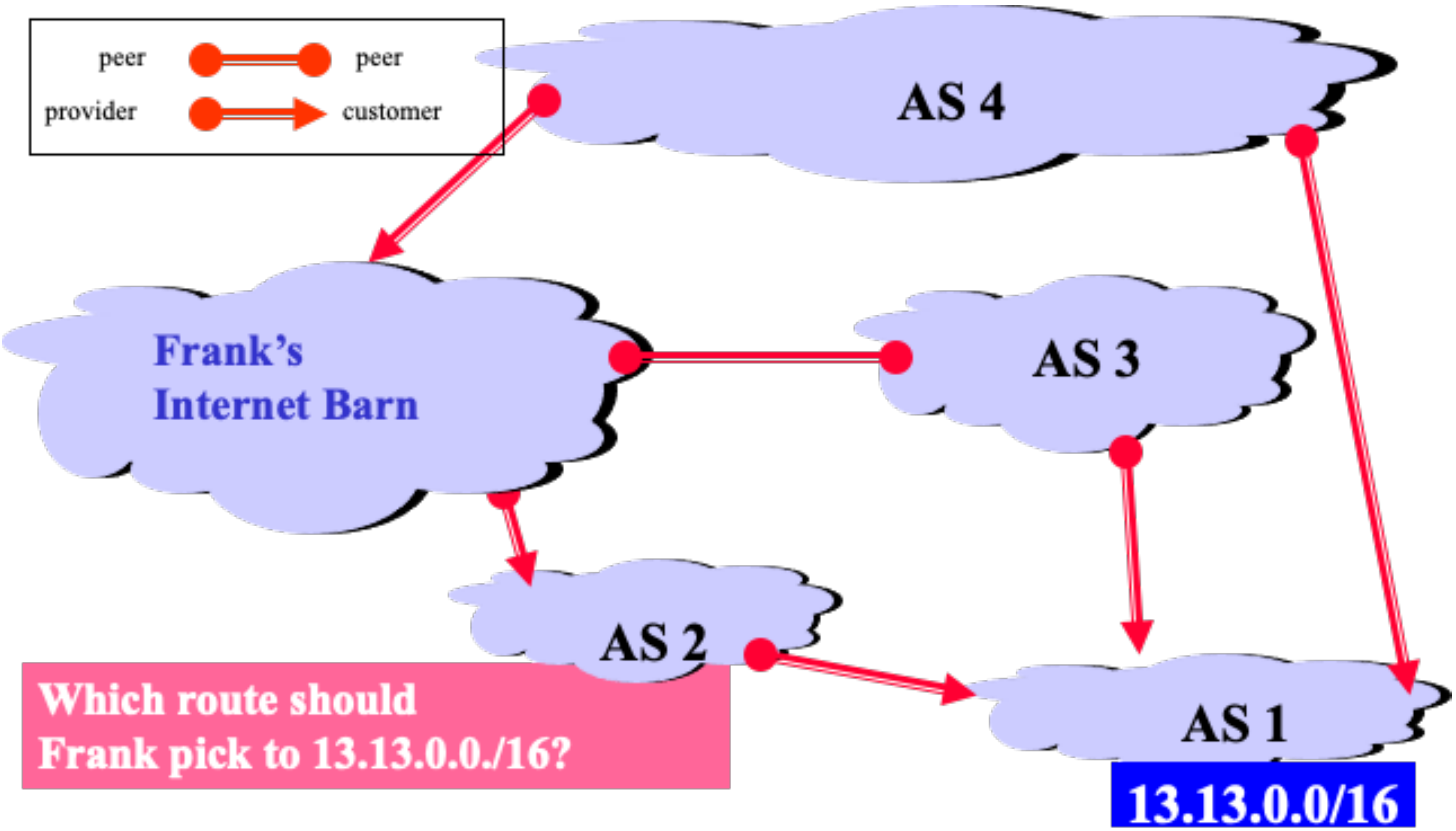
- Policies are not part of BGP. They are provided to BGP for routing configuration.

Policy enforcement:

- Import: choosing appropriate paths from multiple alternatives
- Export: controlling advertisement to other ASes

Policies can be arbitrarily complex. There are some common ones.

BGP Policy Example



Peering and Customer-Provider

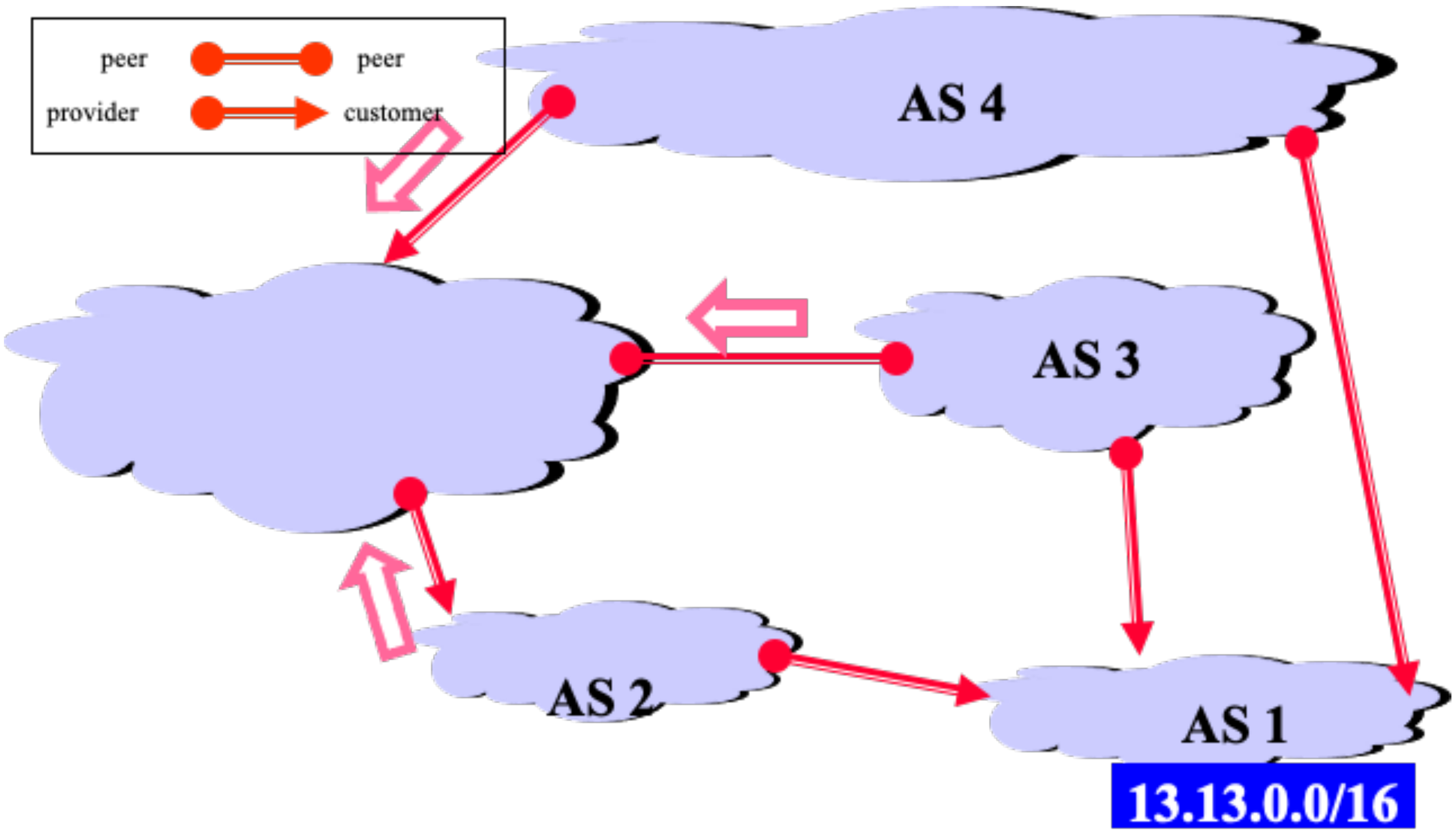
Peering relationship

- Peers provide transit to each other
- Peering relationships are free and involve no cost

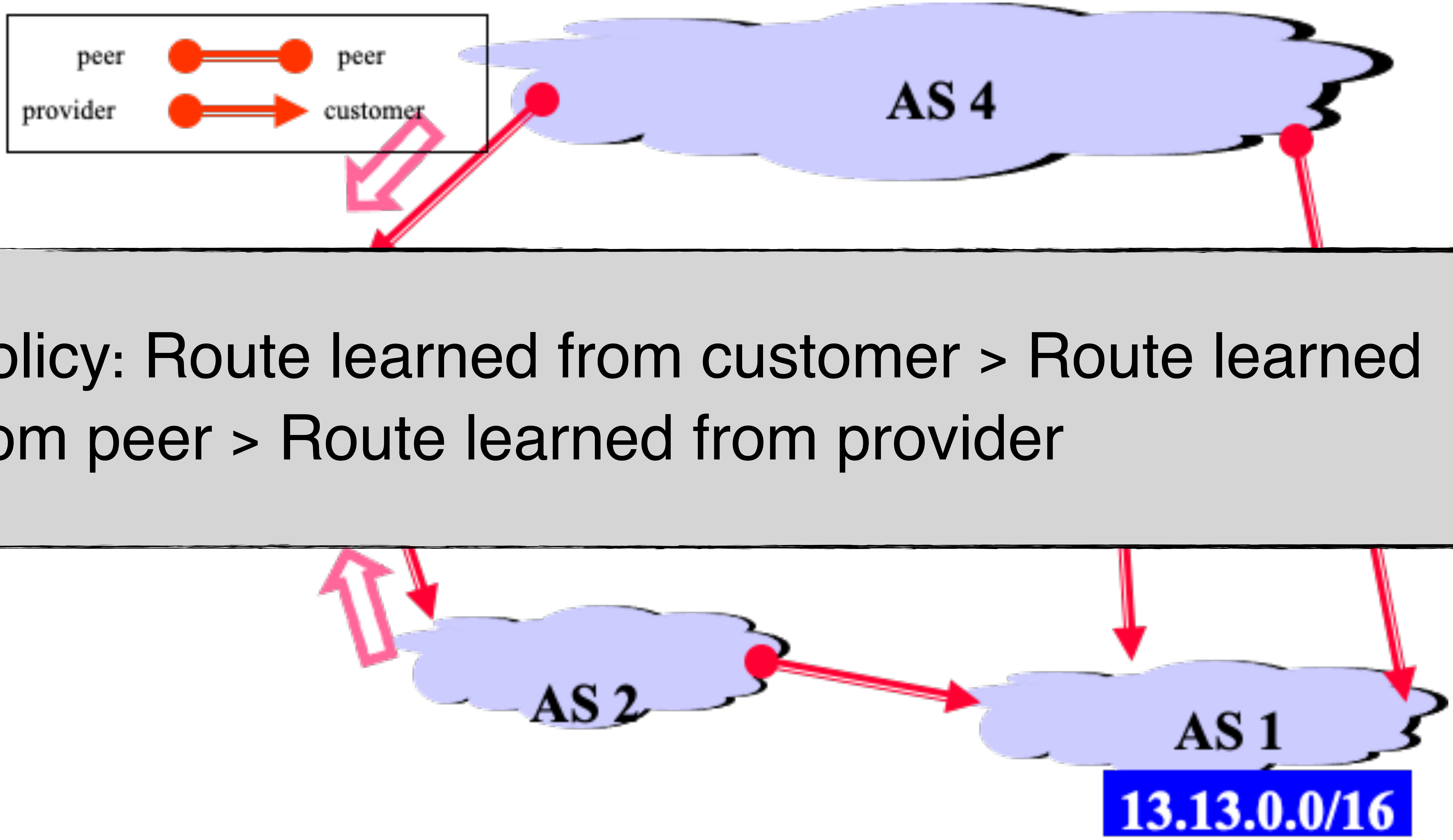
Customer-Provider relationship

- Customers use providers to reach the rest of the Internet
- Customers pay providers for this

Import Policy: Prefer Customer Routing

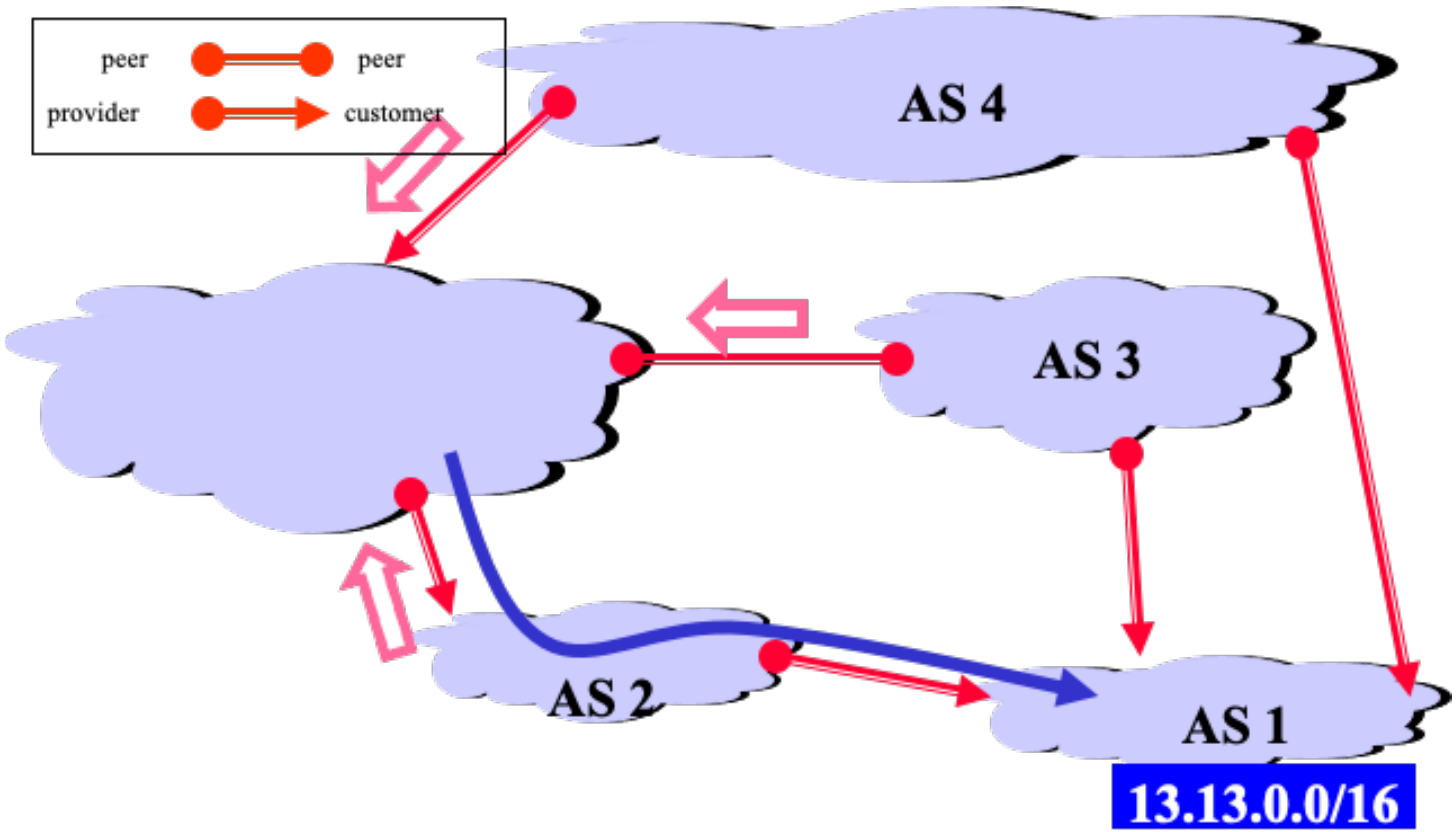


Import Policy: Prefer Customer Routing



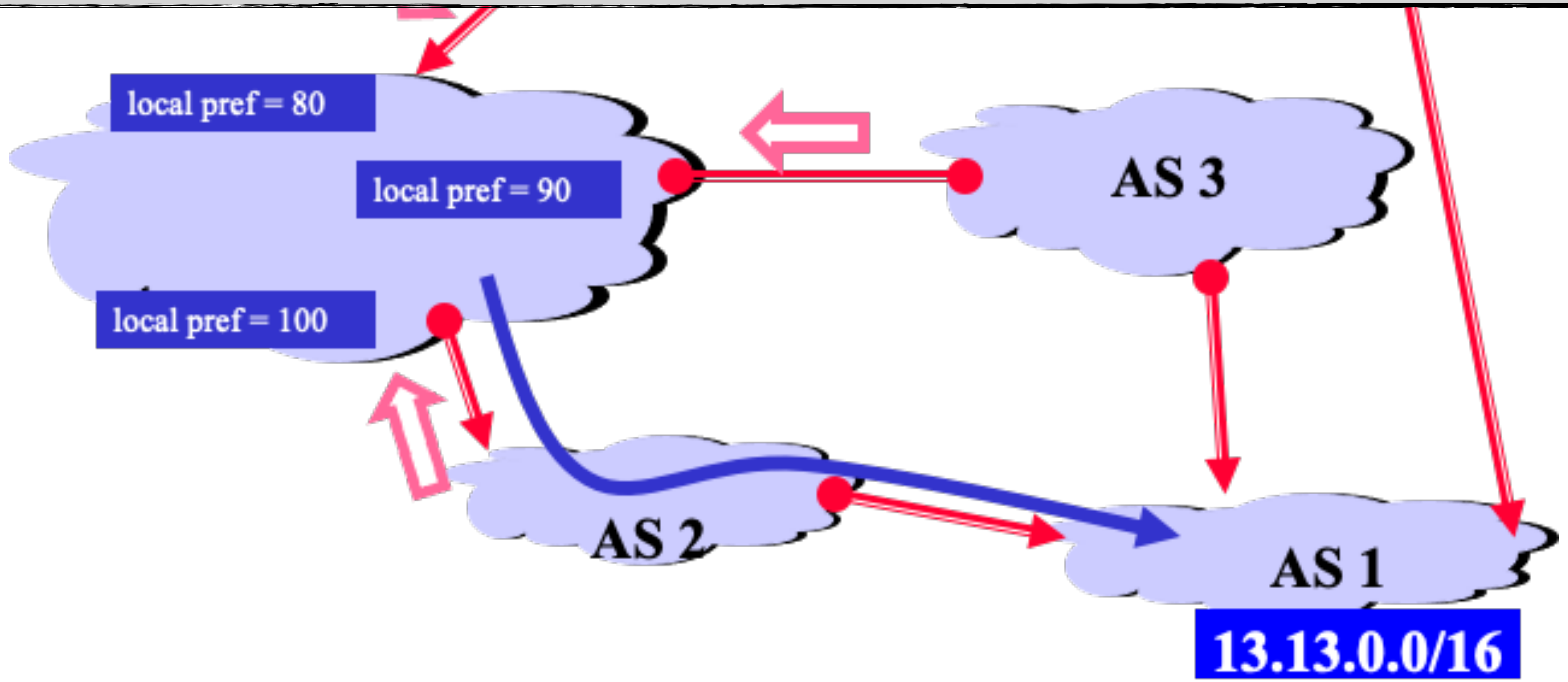
Policy: Route learned from customer > Route learned from peer > Route learned from provider

Import Policy: Prefer Customer Routing

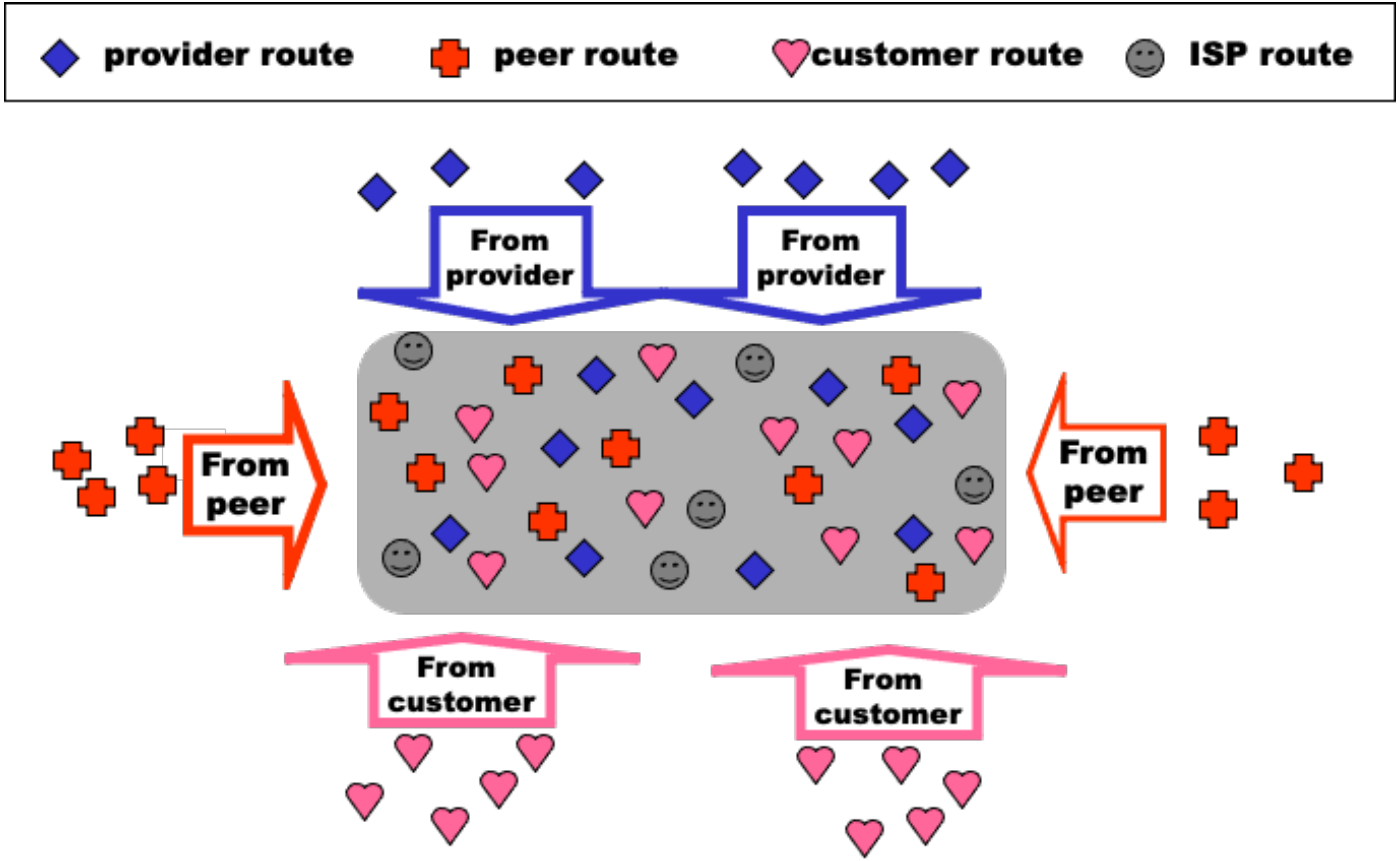


Import Policy: Prefer Customer Routing

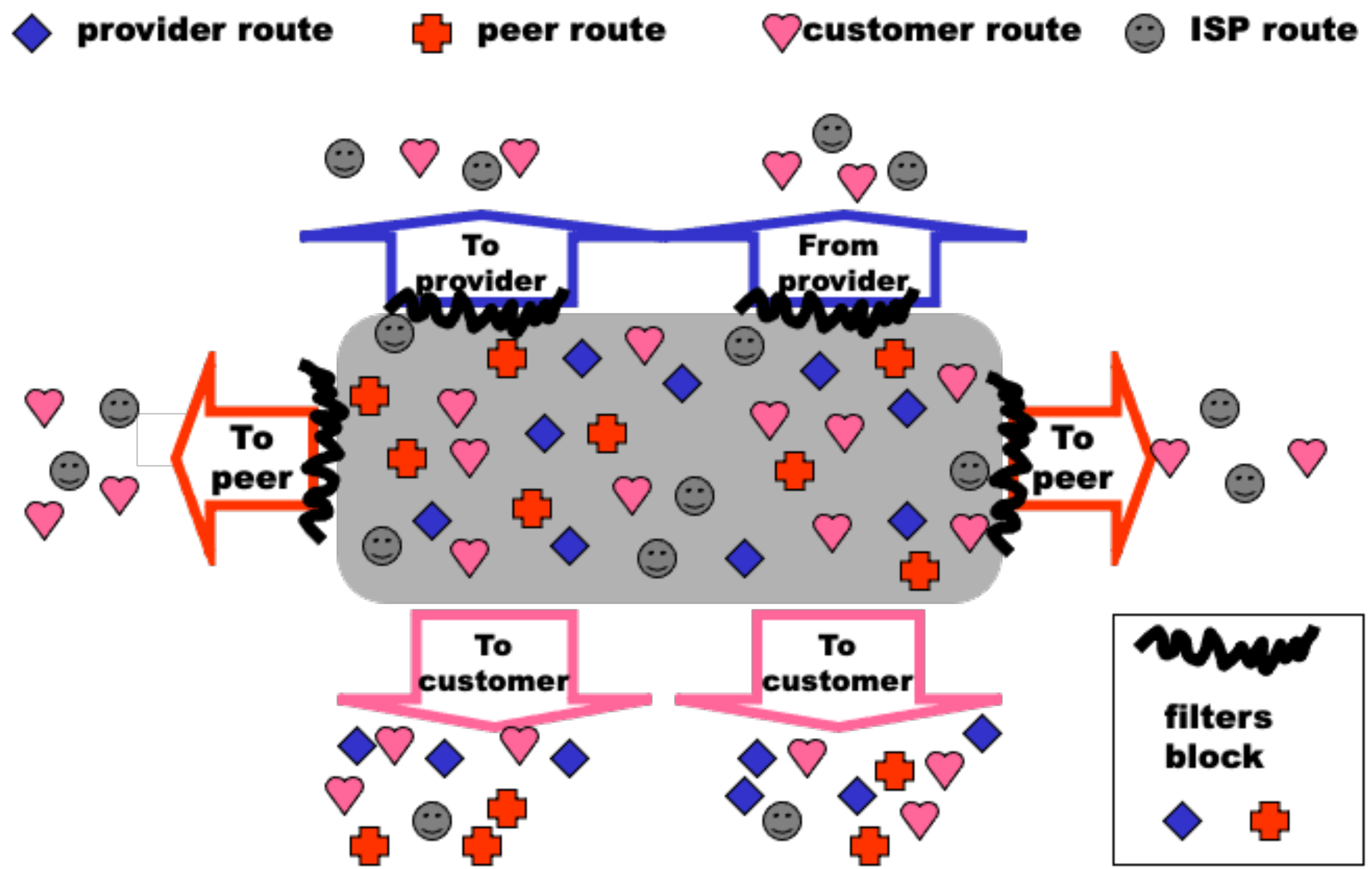
Set appropriate "local pref" to reflect preferences: higher local preference values are preferred.



Import Routes



Export Routes



BGP Export Policies

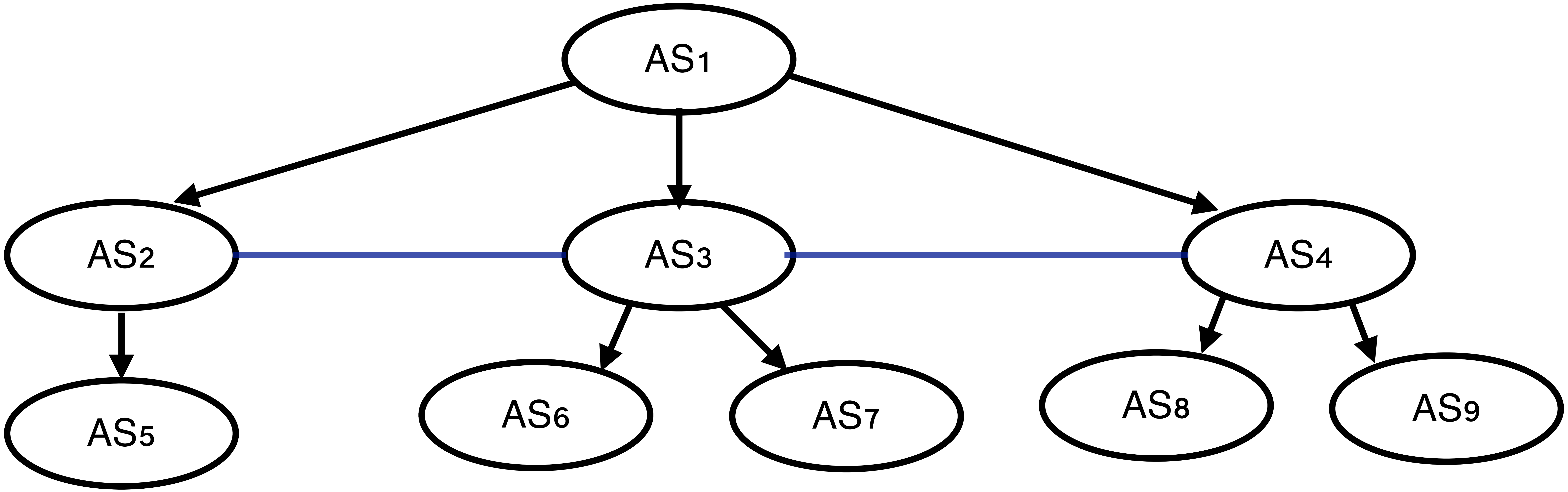
Route learned From ↓	Advertise to → Customer	Provider	Peer
Customer	✓	✓	✓
Provider	✓	✗	✗
Peer	✓	✗	✗

A BGP Example

Consider a network with 9 ASes. They have the following relationships:

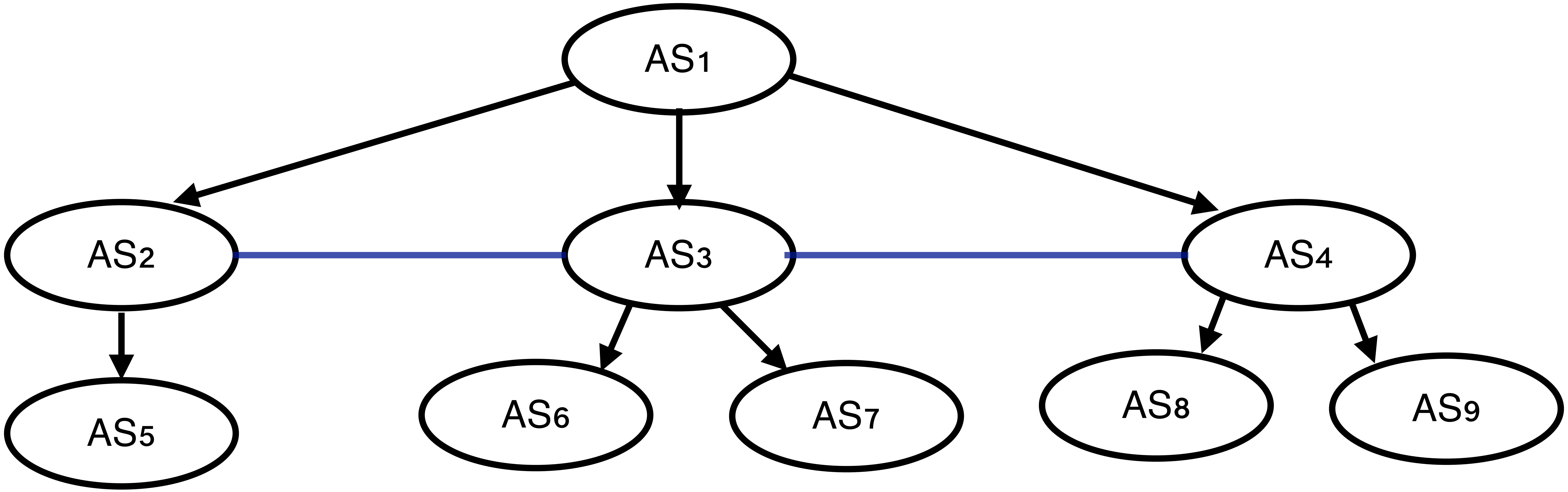
- AS_1 is the provider for AS_2, AS_3, and AS_4
- AS_2 is the provider for AS_5
- AS_2 and AS_3 are peers; AS_3 and AS_4 are peers
- AS_3 is the provider for AS_6 and AS_7
- AS_4 is the provider for AS_8 and AS_9

A BGP Example



A BGP Example (1)

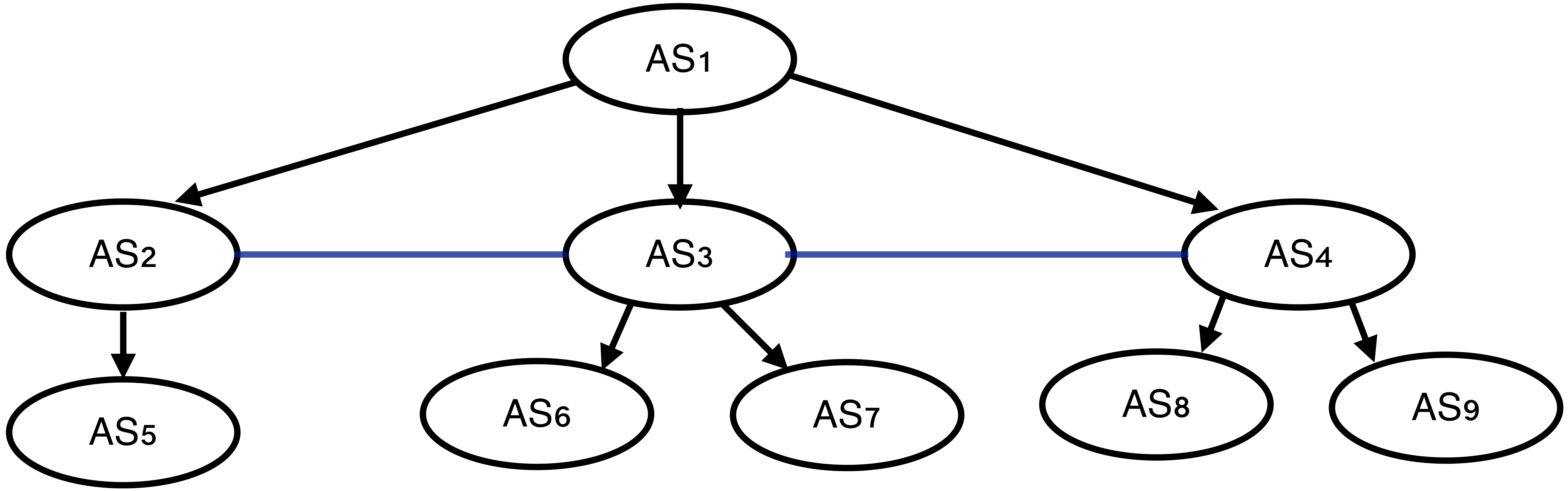
What is the AS path used for AS8-> AS7?



A BGP Example (1)

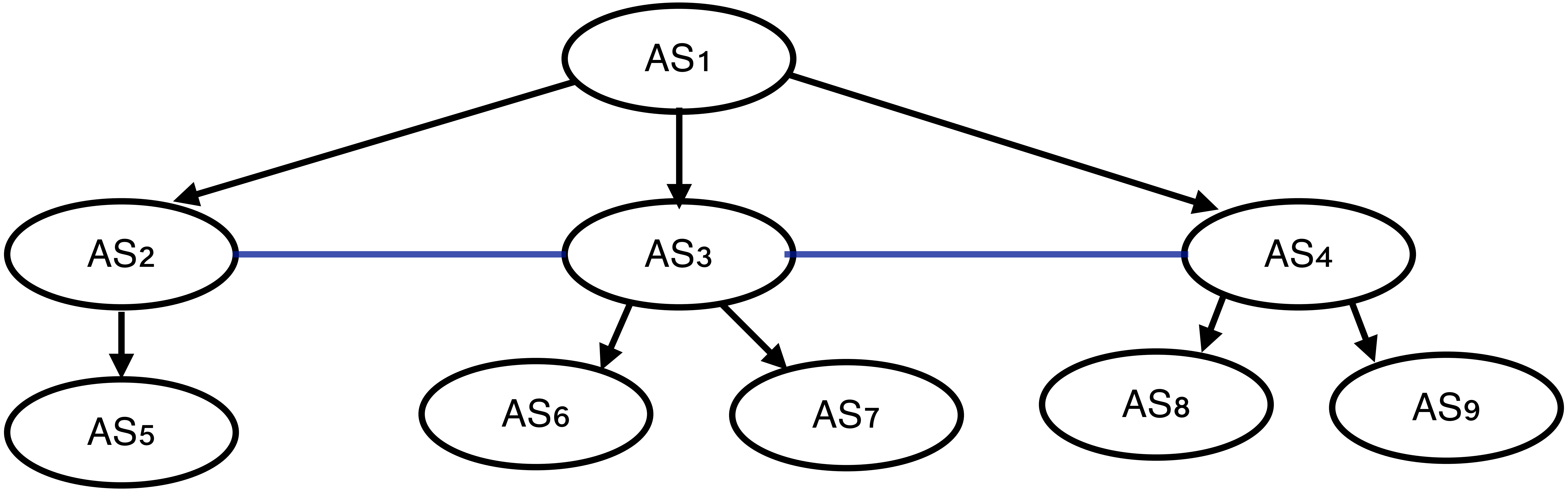
What is the AS path used for AS8-> AS7?

- AS8 -> AS4 -> AS3 -> AS7



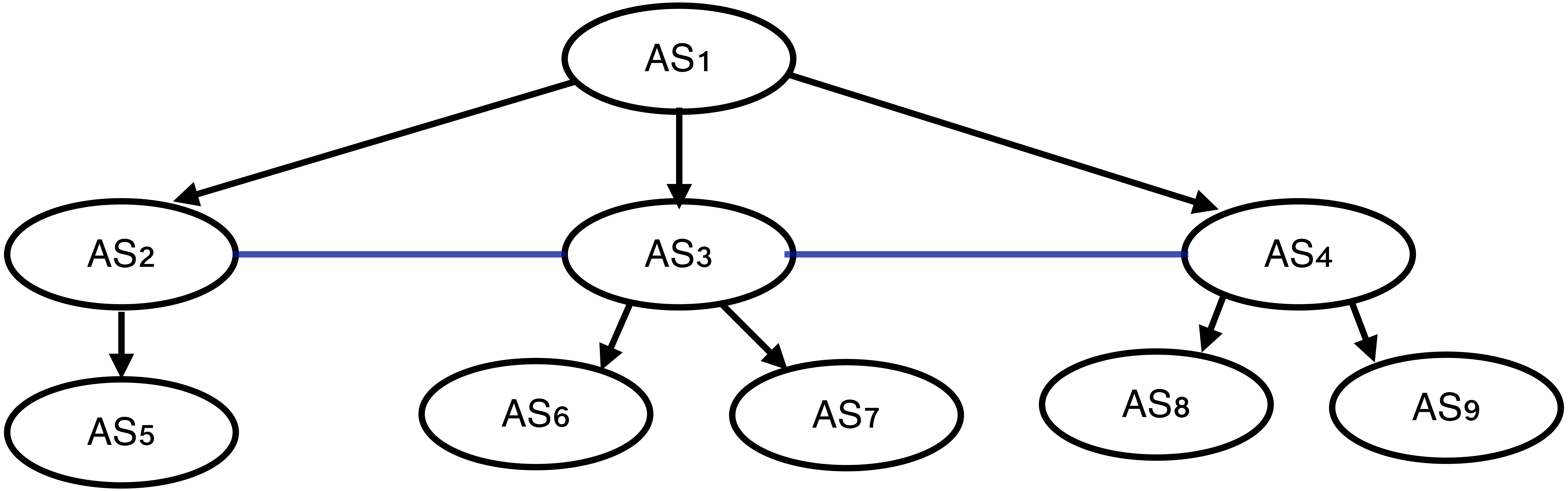
A BGP Example (2)

Is (AS5, AS2, AS3, AS4, AS8) a valid path to go from a host in AS5 to a host in AS8?



A BGP Example (2)

Is (AS5, AS2, AS3, AS4, AS8) a valid path to go from a host in AS5 to a host in AS8? => No!



BGP in Reality

AS 7007 incident

From Wikiped

The AS 7007

sometimes ir

Probably bec

and had the

problems tha

of these facto

Analysis

How Pakistan knocked YouTube offline (and how to make sure it never happens again)

Suspicious event hijacks Amazon traffic for 2 hours, steals cryptocurrency

Almost 1,300 addresses for Amazon Route 53 rerouted for two hours.

DAN GOODIN - 4/24/2018, 2:00 PM

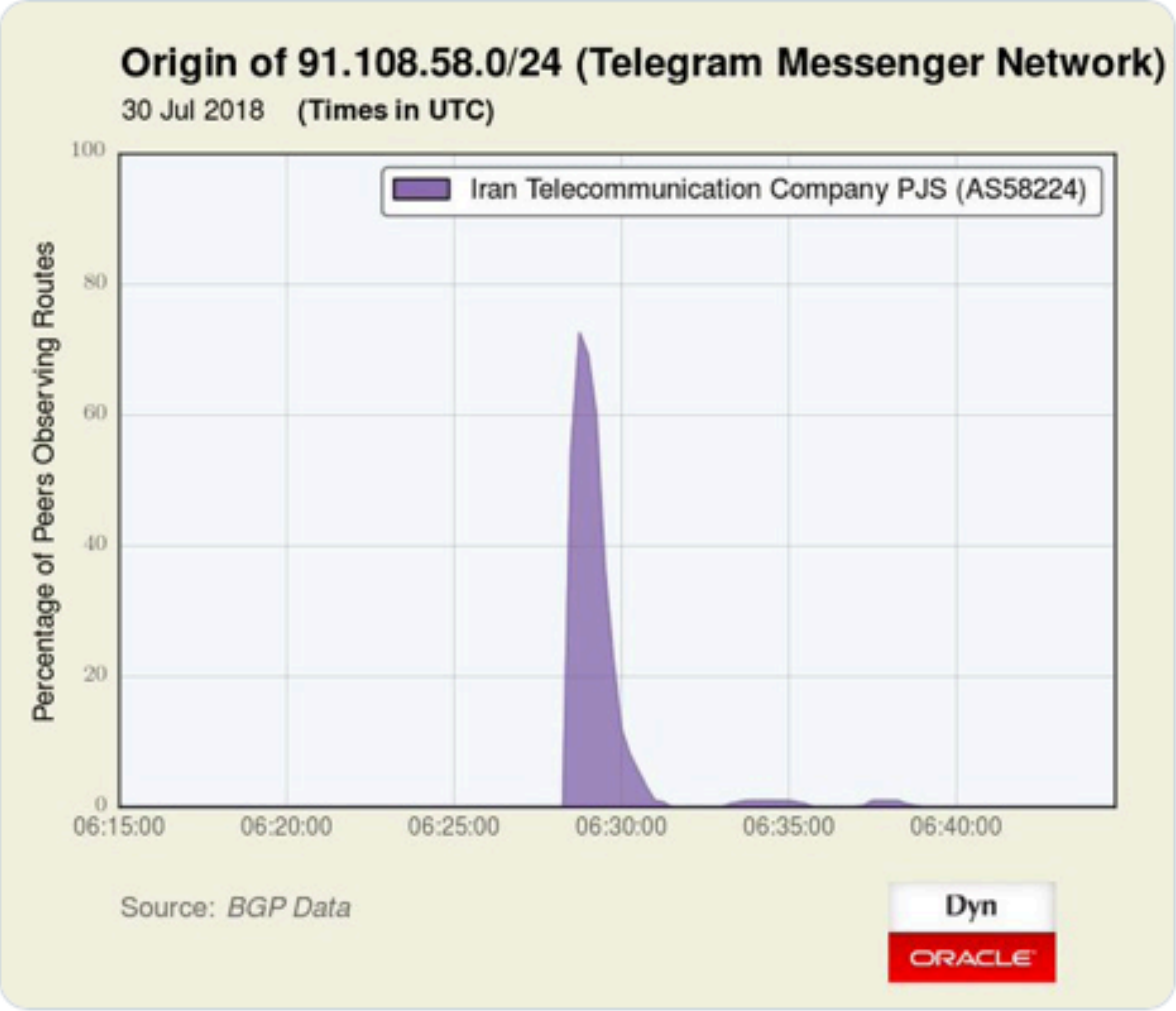
amazon.com

125

Amazon lost control of a small number of its cloud services IP addresses for two hours on Tuesday morning when hackers exploited a known Internet-protocol weakness that let them to redirect traffic to rogue destinations. By subverting Amazon's domain-resolution service, the attackers masqueraded as cryptocurrency website MyEtherWallet.com and stole about $150,00 in digital coins from unwitting end users. They may have targeted other Amazon customers as well.

InternetIntelligence
@InternetIntel

At 06:28 UTC earlier today (30-Jul), an Iranian state telecom network briefly leaked over 100 prefixes. Most were Iranian networks, but the leak also included 10 prefixes of popular messaging app @telegram (8 were more-specifics).



BGP in Reality

AS 7007 incident

From Wikipedi

The AS 7007

sometimes in

Probably bec

and had the

problems tha

of these facto

Analysis

How Pak

offline (a

happens

Suspicious

for 2 hours

Almost 1,300 addresses f

DAN GOODIN - 4/24/2018, 2:00 PM

125

Amazon lost

Tuesday morn

redirect traffi

attackers mas

in digital coin

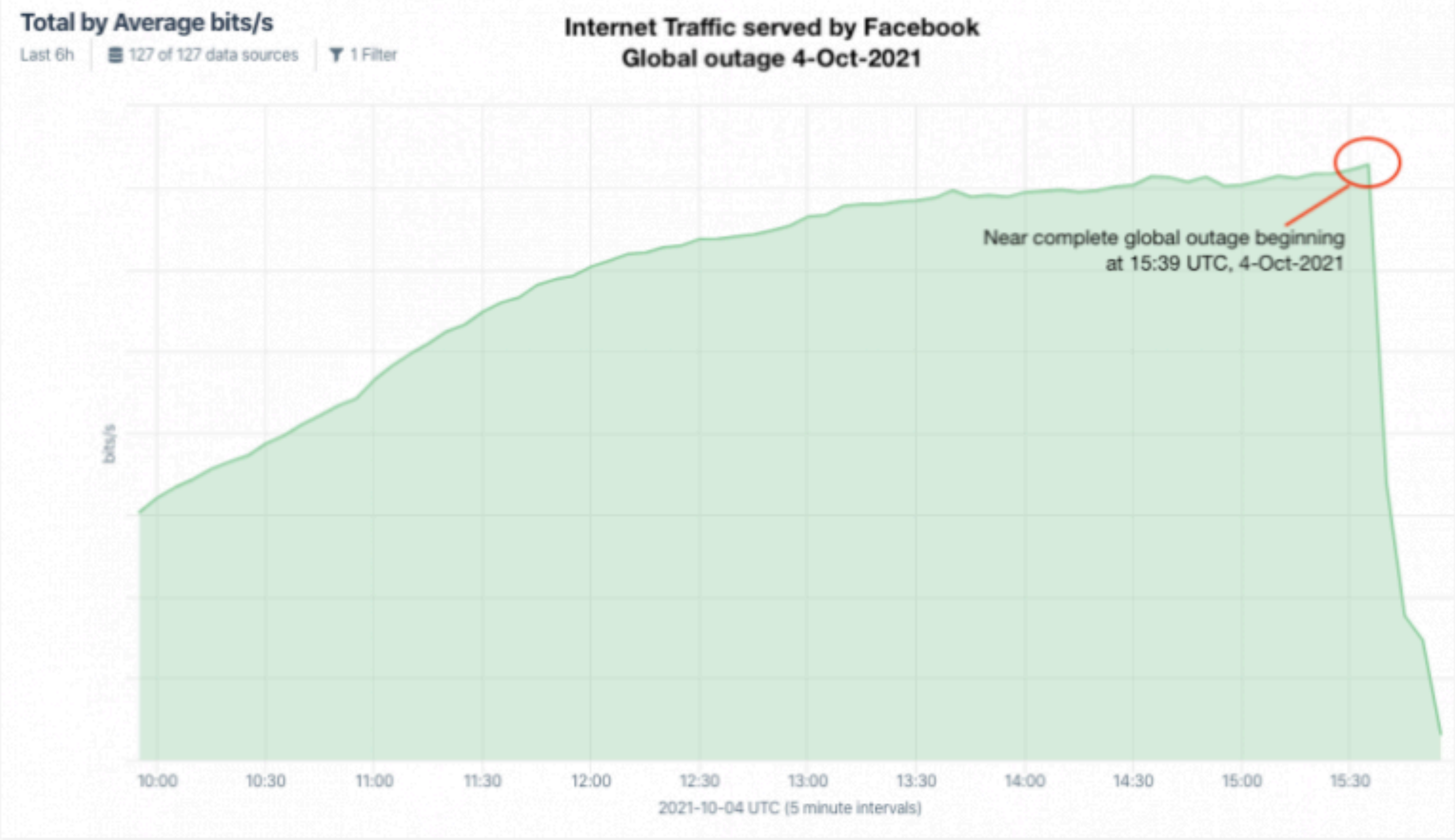
well.

What Happened to Facebook, Instagram, & WhatsApp

October 4, 2021

124 Comments

Facebook and its sister properties **Instagram** and **WhatsApp** are suffering from ongoing, global outages. We don't yet know why this happened, but the how is clear: Earlier this morning, something inside Facebook caused the company to revoke key digital records that tell computers and other Internet-enabled devices how to find these destinations online.



-Jul), an Iranian state

over 100 prefixes.

t the leak also

messaging app

fics).

ram Messenger Network)

Terminology

1. Host
2. NIC
3. Multi-port I/O bridge
4. Protocol
5. RTT
6. Packet
7. Header
8. Payload
9. BDP
10. Baud rate
11. Frame/Framing
12. Parity bit
13. Checksum
14. Ethernet
15. MAC
16. (L2) Switch
17. Broadcast
18. Acknowledgement
19. Timeout
20. Datagram
21. TTL
22. MTU
23. Best effort
24. (L3) Router
25. Subnet mask
26. CIDR
27. Converge
28. Count-to-infinity
29. Line card
30. Network processor
31. Gateway

Principle

1. Layering
2. Minimal States
3. Hierarchy

Technique

1. NRZ Encoding
2. NRZI Encoding
3. Manchester Encoding
4. 4B/5B Encoding
5. Byte Stuffing
6. Byte Counting
7. Bit Stuffing
8. 2-D Parity
9. CRC
10. MAC Learning
11. Store-and-Forward
12. Cut-through
13. Spanning Tree
14. CSMA/CD
15. Stop-and-Wait
16. Sliding Window

16. Fragmentation and Reassembly
17. Path MTU discovery
18. DHCP
19. Subnetting
20. Supernetting
21. Longest prefix match
22. Distance vector routing (RIP)
23. Link state routing (OSPF)
24. Boarder gateway protocol (BGP)

Summary

Today's takeaways

#1: BGP enables routing across ASes by enforcing import/export policies

#2: Common policies

- Import: Route learned from customer > Route learned from peer > Route learned from provider
- Export: BGP export policy matrix

Next lecture

- IP Potpourri