

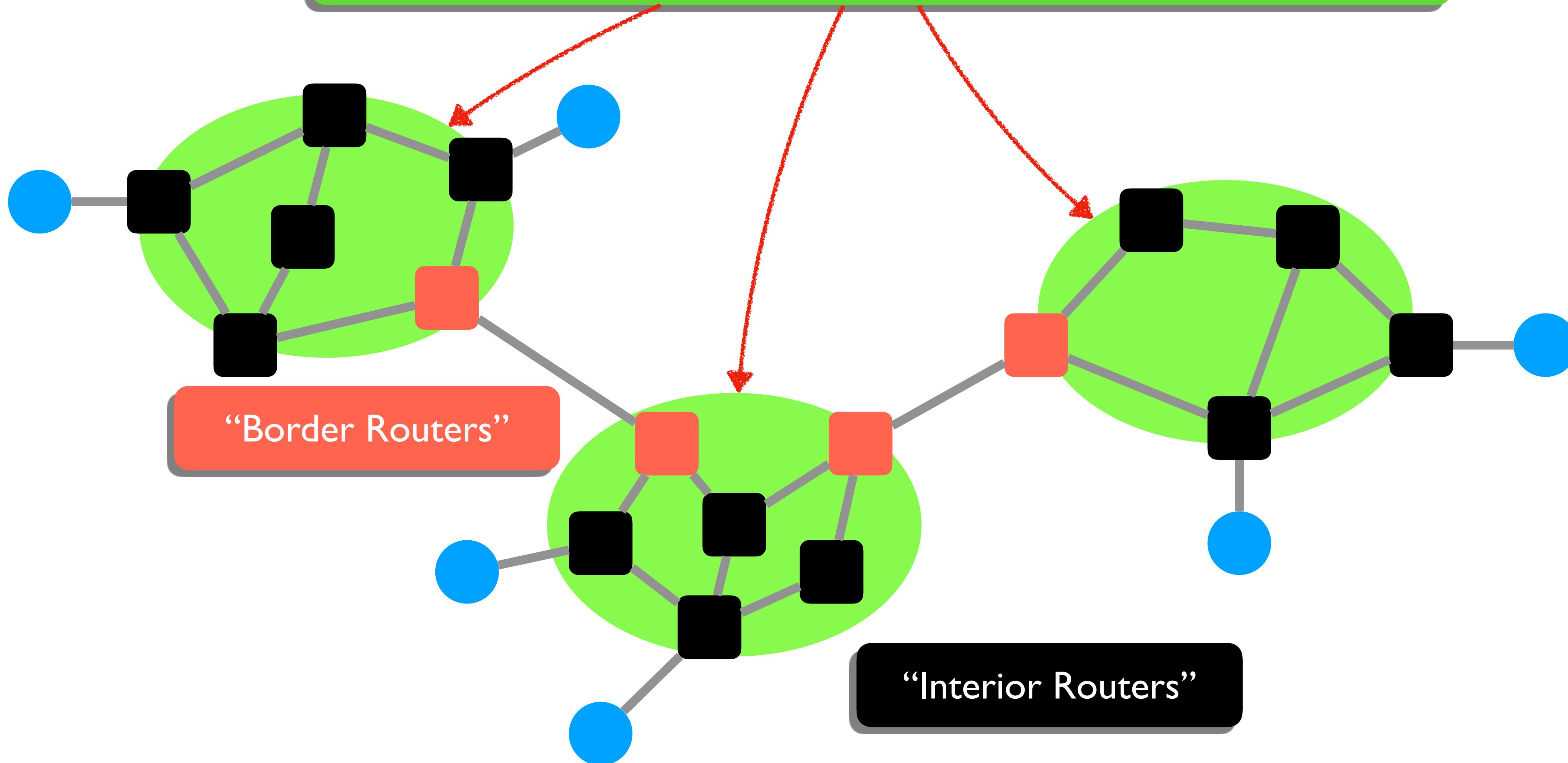
BGP

CPSC 433/533, Spring 2021
Anurag Khandelwal

Administrivia

- I have 1-2 slots still open for research over semester/summer
 - If you like system hacking and looking for projects, let me know!

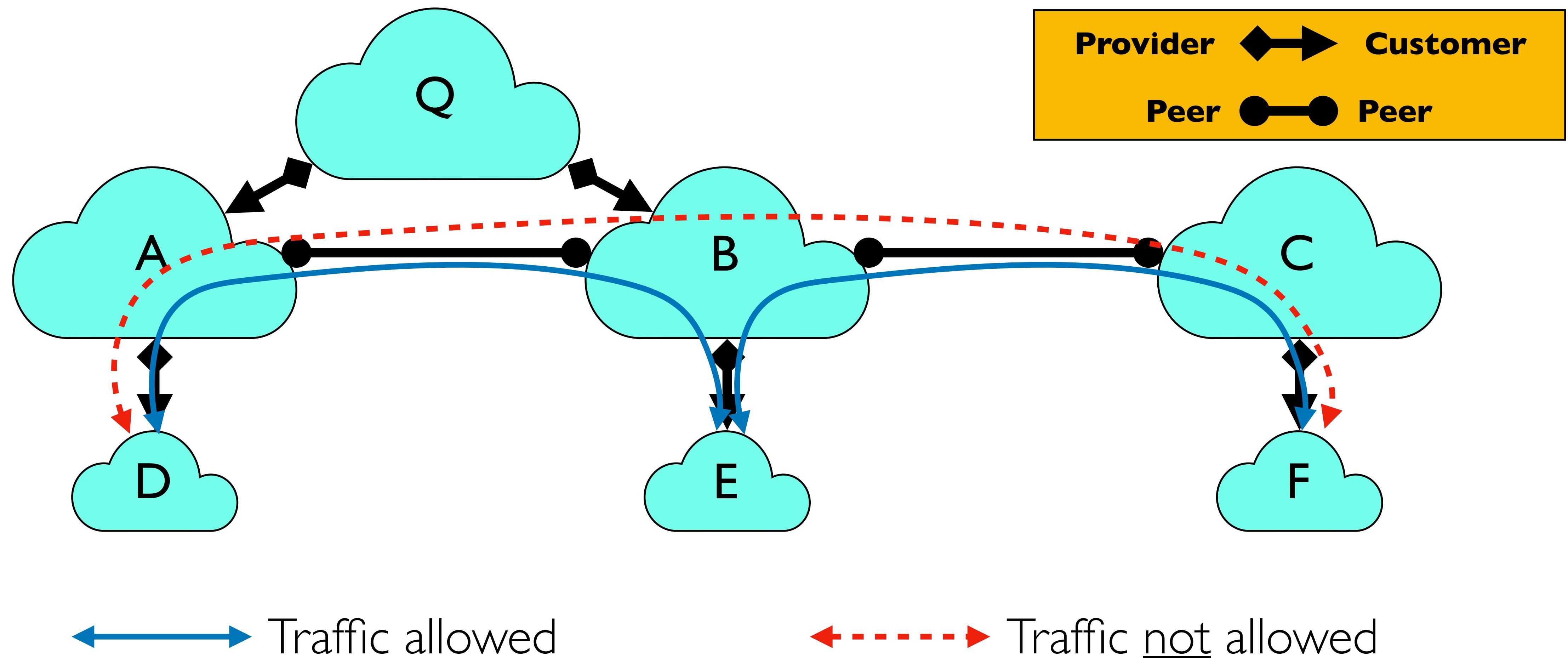
“Autonomous System (AS)” or “Domain”
Region of a network under a single administrative authority



Topology and Routes shaped by business relationships between ASes

- **Three basic kinds of relationships between ASes**
 - AS A can be AS B's customer
 - AS A can be AS B's provider
 - AS A can be AS B's peer
- **Business implications**
 - Customer pays provider
 - Peers don't pay each other
 - Exchange roughly equal traffic

Routing Follows the Money!



Interdomain Routing: Setup

Interdomain Routing: Setup

- Destinations are IP prefixes ($12.0.0.0/8$)

Interdomain Routing: Setup

- Destinations are IP prefixes ($12.0.0.0/8$)
- Nodes are Autonomous Systems (ASes)
 - Internals of each AS are hidden

Interdomain Routing: Setup

- Destinations are IP prefixes ($12.0.0.0/8$)
- Nodes are Autonomous Systems (ASes)
 - Internals of each AS are hidden
- Links represent both physical links and business relationships

Interdomain Routing: Setup

- Destinations are IP prefixes ($12.0.0.0/8$)
- Nodes are Autonomous Systems (ASes)
 - Internals of each AS are hidden
- Links represent both physical links and business relationships
- Border Gateway Protocol (BGP) is the Interdomain routing protocol
 - Implemented by AS border routers

BGP: Basic Idea

BGP: Basic Idea

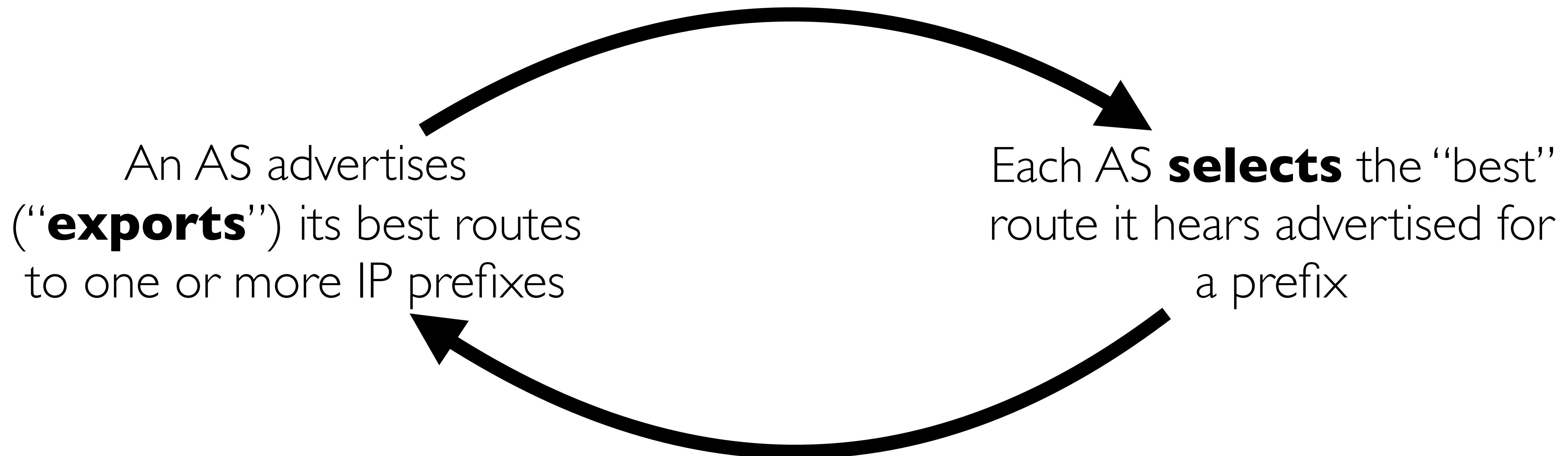
An AS advertises
("“**exports**”") its best routes
to one or more IP prefixes

BGP: Basic Idea

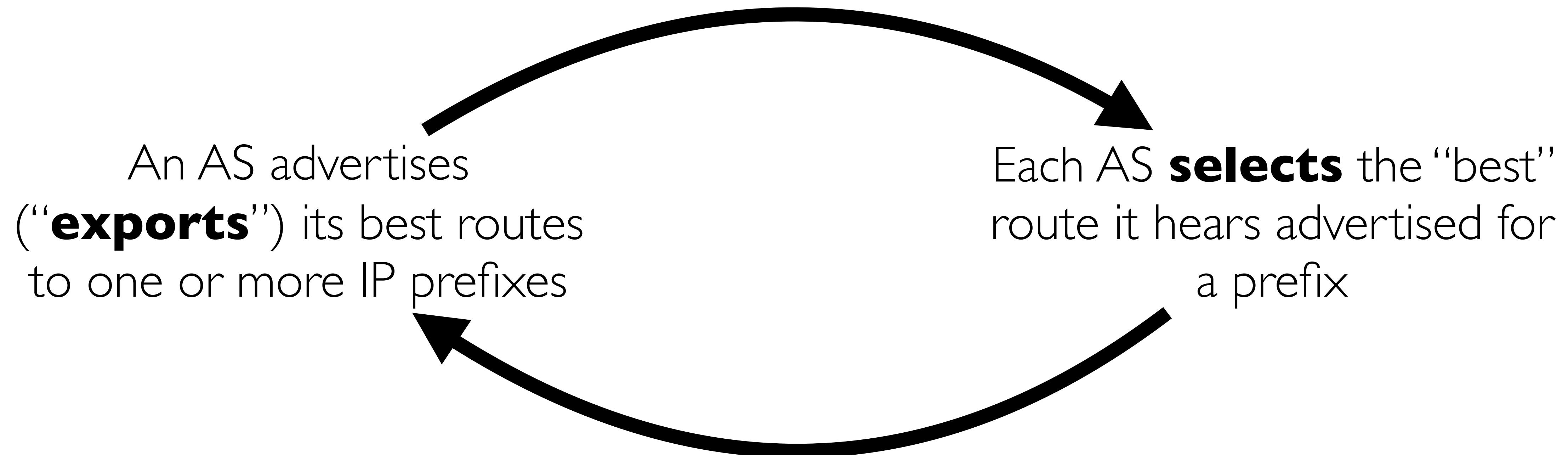
An AS advertises (“**exports**”) its best routes to one or more IP prefixes

Each AS **selects** the “best” route it hears advertised for a prefix

BGP: Basic Idea

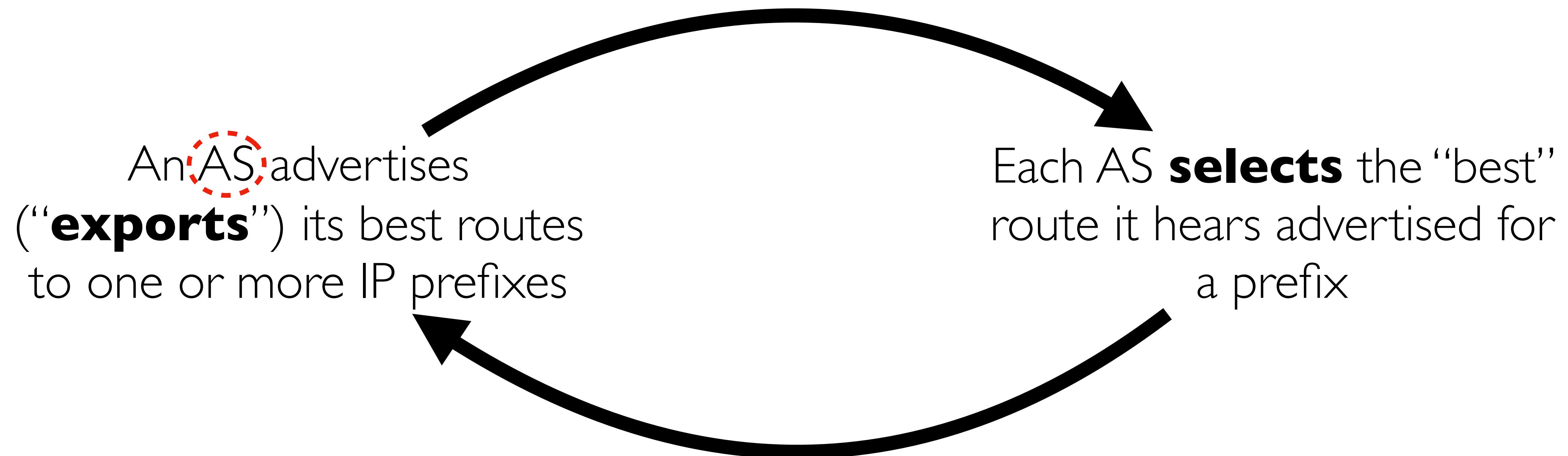


BGP: Basic Idea



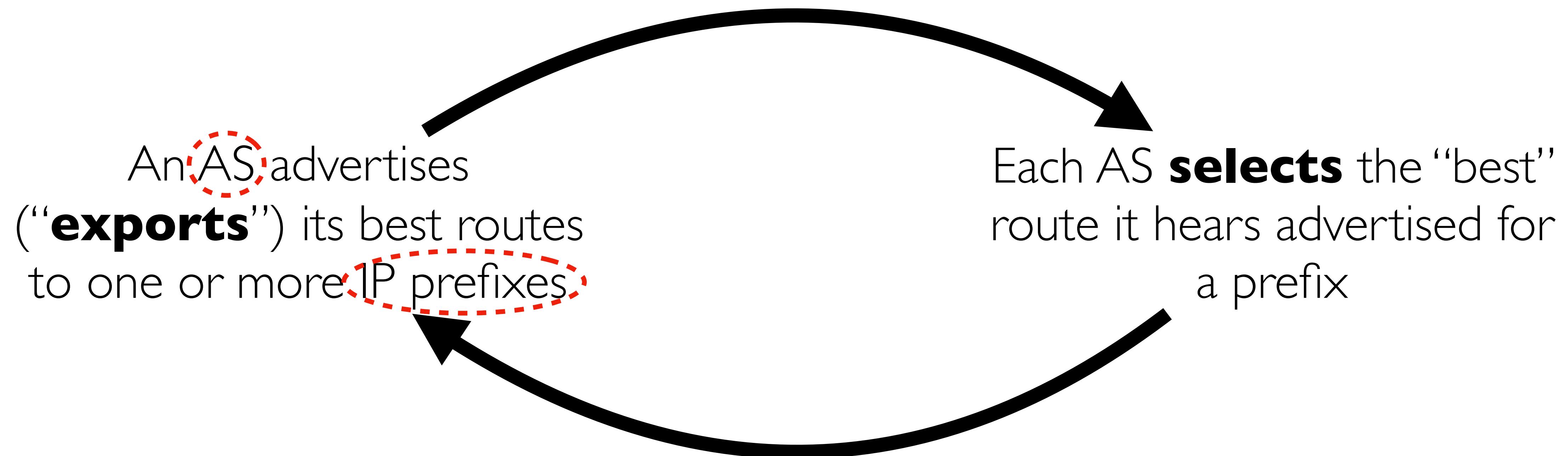
You've heard this story before!

BGP: Basic Idea



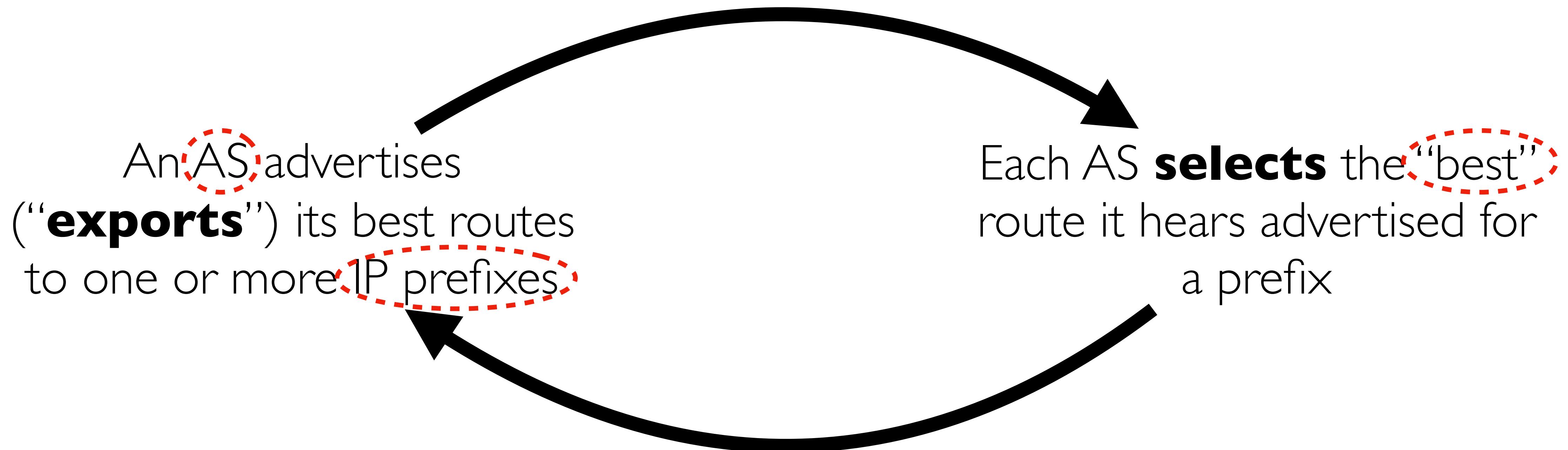
You’ve heard this story before!

BGP: Basic Idea



You've heard this story before!

BGP: Basic Idea



You’ve heard this story before!

BGP inspired by Distance Vector

BGP inspired by Distance Vector

- Per-destination route advertisements

BGP inspired by Distance Vector

- Per-destination route advertisements
- No global sharing of network topology information

BGP inspired by Distance Vector

- Per-destination route advertisements
- No global sharing of network topology information
- Iterative and distributed convergence on paths

BGP inspired by Distance Vector

- Per-destination route advertisements
- No global sharing of network topology information
- Iterative and distributed convergence on paths
- *With four crucial differences*

Differences between BGP and DV

(I) not picking the shortest path routes

Differences between BGP and DV

(I) not picking the shortest path routes

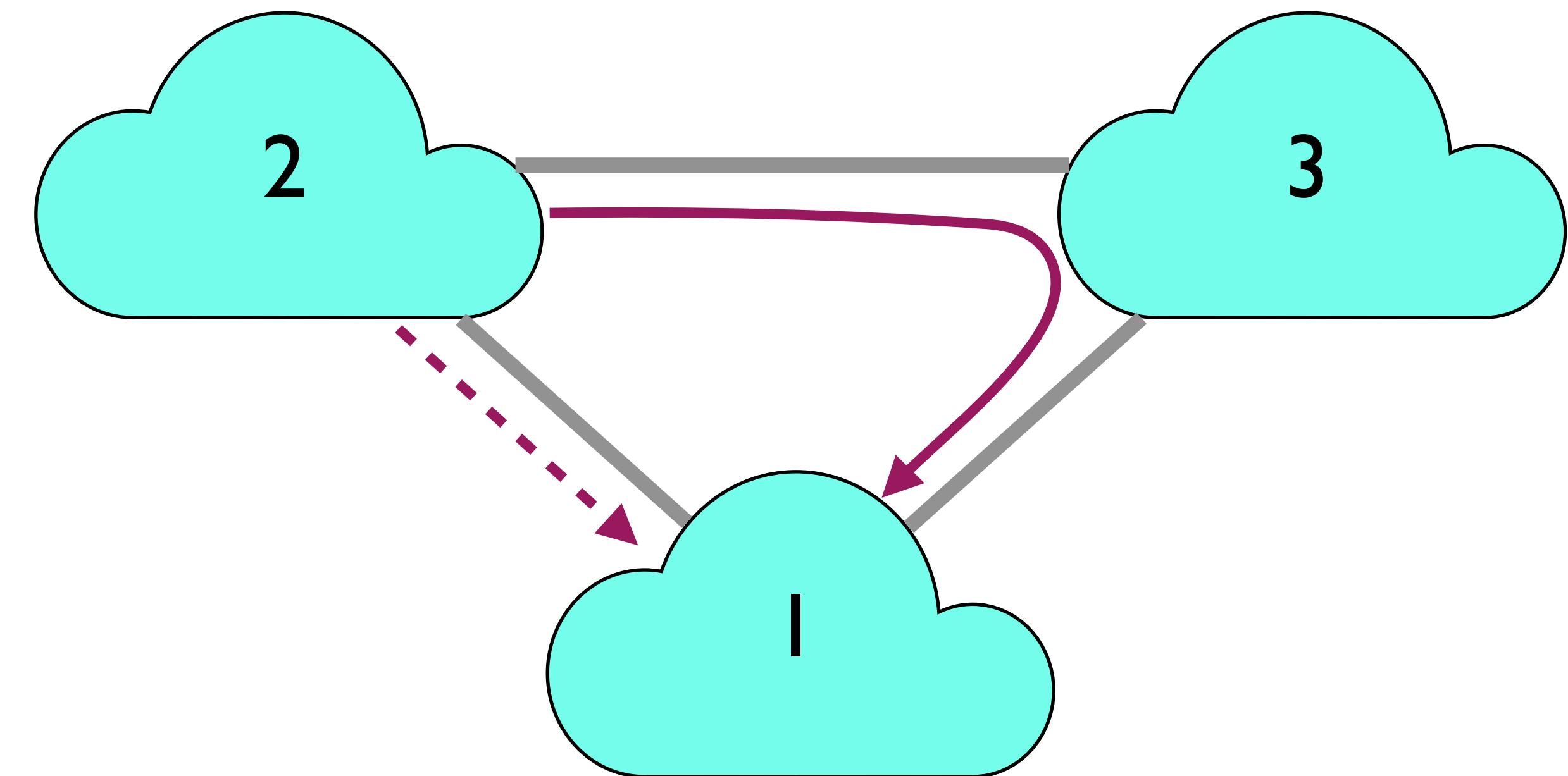
- BGP selects the east route based on policy, not shortest distance (least cost)

Differences between BGP and DV

(I) not picking the shortest path routes

- BGP selects the east route based on policy, not shortest distance (least cost)

**Node 2 may prefer
“2, 3, 1” over “2, 1”**

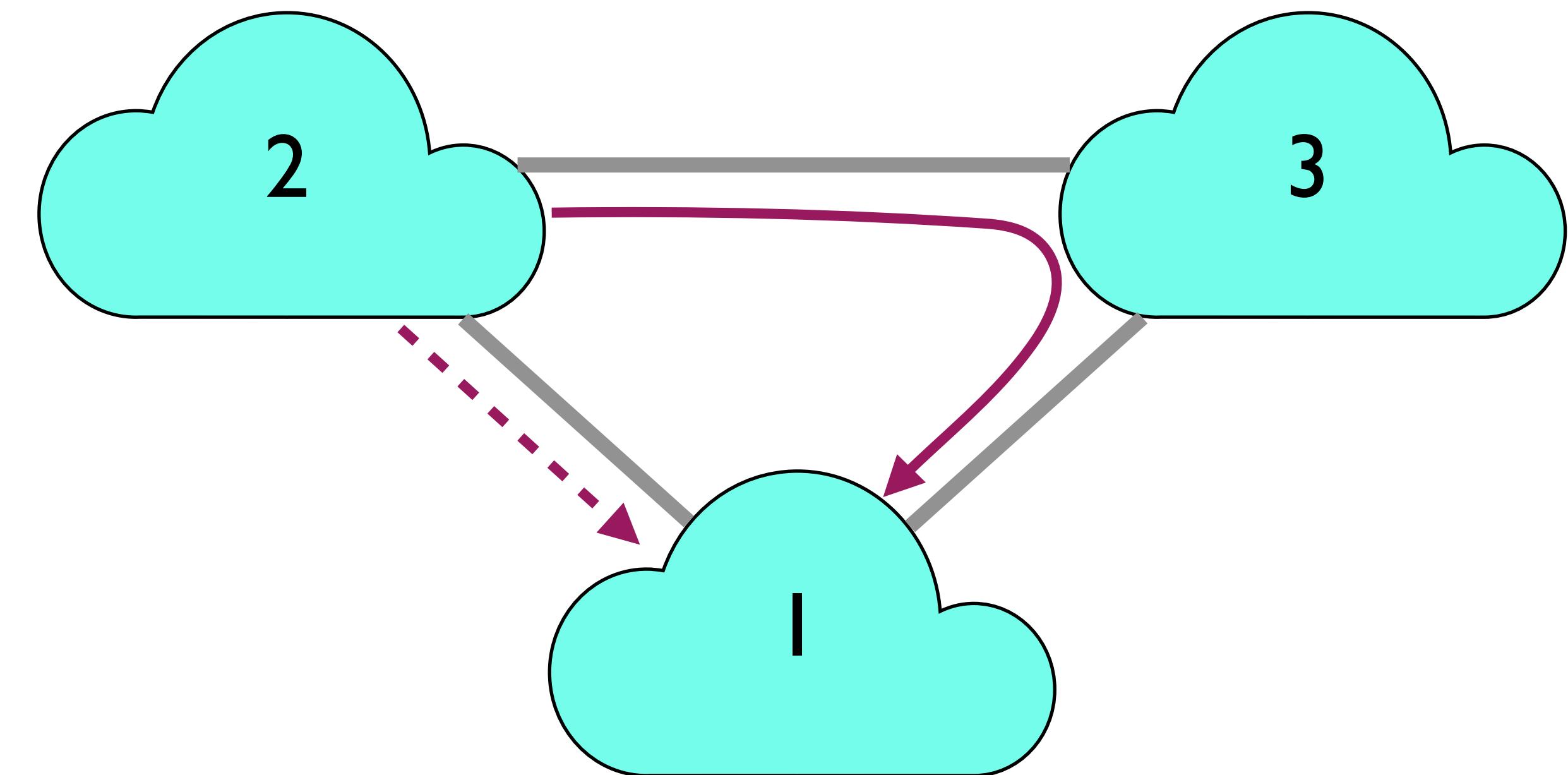


Differences between BGP and DV

(I) not picking the shortest path routes

- BGP selects the east route based on policy, not shortest distance (least cost)

**Node 2 may prefer
“2, 3, 1” over “2, 1”**



- How do we avoid loops?

Differences between BGP and DV

(2) path-vector routing

Differences between BGP and DV

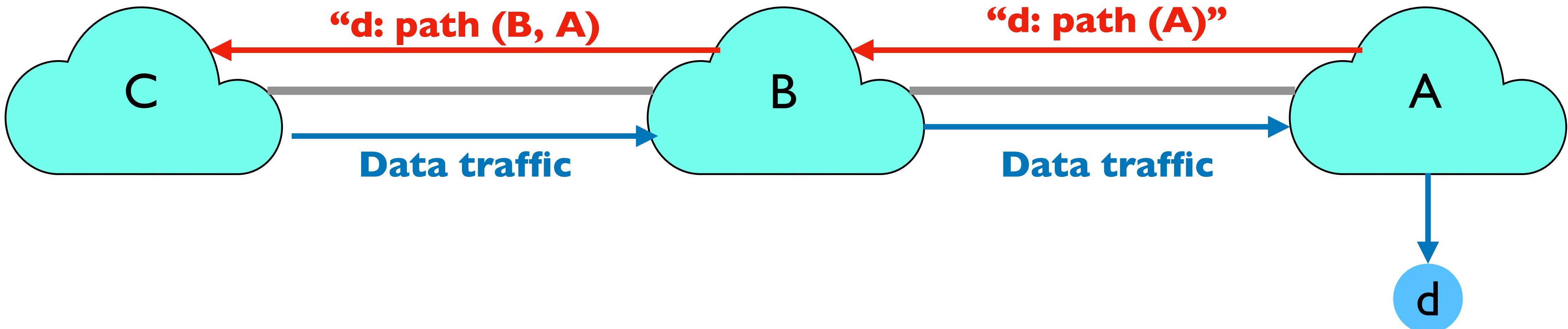
(2) path-vector routing

- **Key idea:** advertise the entire path
 - Distance vector: send distance metric per destination d
 - Path vector: send the entire path for each destination d

Differences between BGP and DV

(2) path-vector routing

- **Key idea:** advertise the entire path
 - Distance vector: send distance metric per destination d
 - Path vector: send the entire path for each destination d



Differences between BGP and DV

(2) path-vector routing

- **Key idea:** advertise the entire path

- Distance vector: send distance metric per destination d
- Path vector: send the entire path for each destination d

- **Benefits**

- Loop avoidance is easy

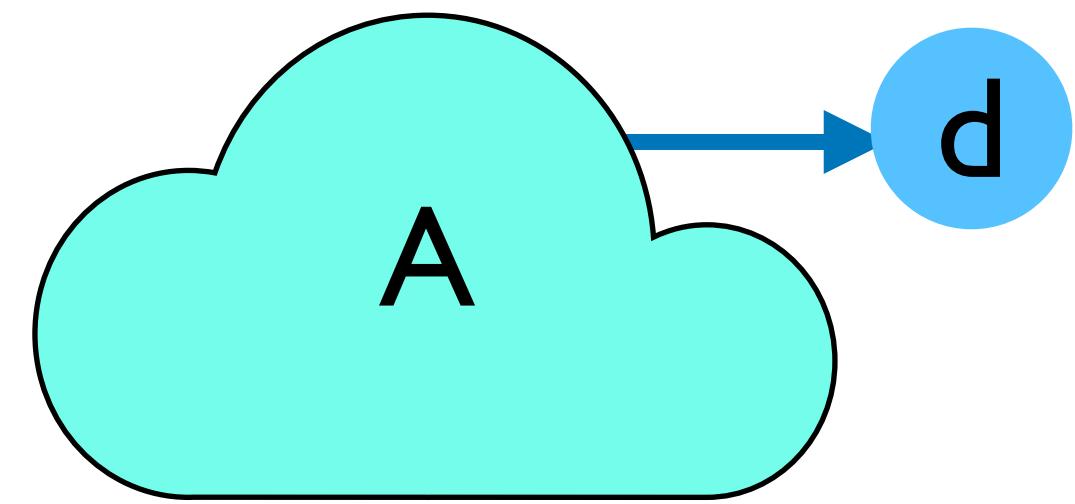
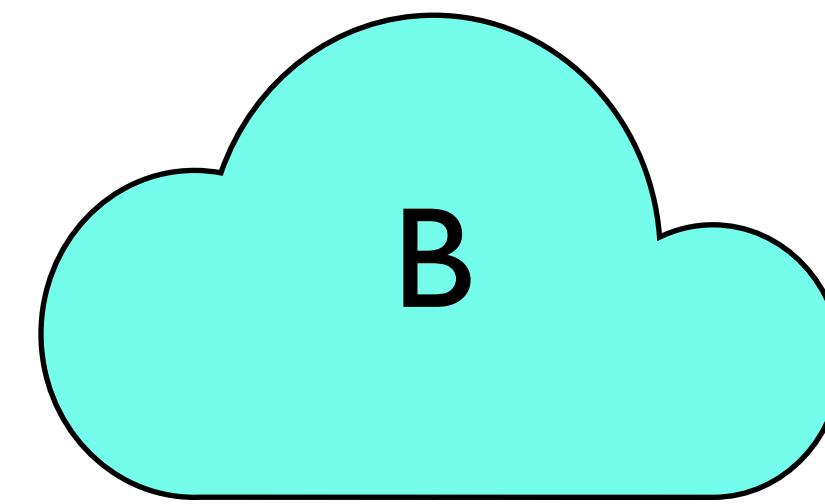
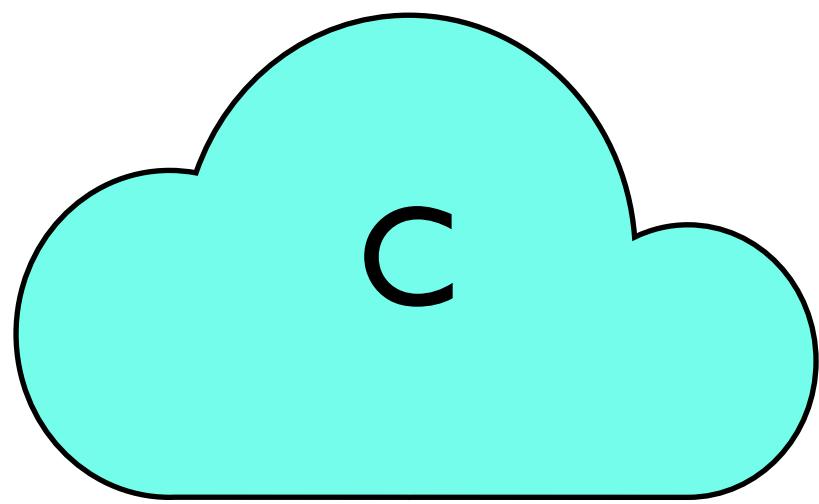
Loop Detection with Path-Vector

Loop Detection with Path-Vector

- Node can easily detect a loop
 - Look for its own node identifier in the path

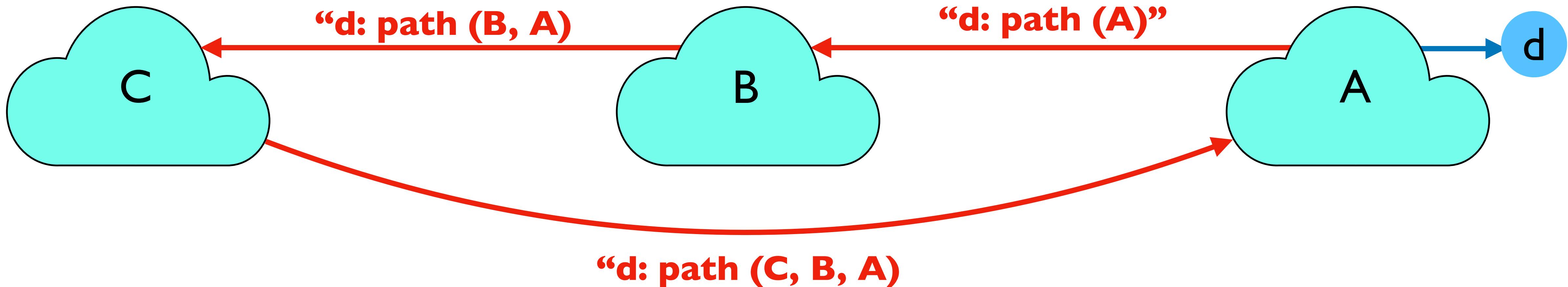
Loop Detection with Path-Vector

- Node can easily detect a loop
 - Look for its own node identifier in the path



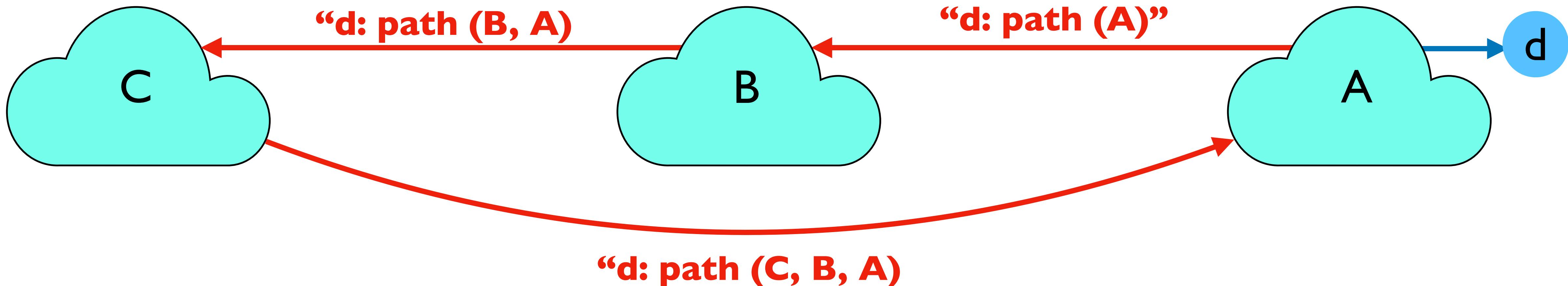
Loop Detection with Path-Vector

- Node can easily detect a loop
 - Look for its own node identifier in the path



Loop Detection with Path-Vector

- Node can easily detect a loop
 - Look for its own node identifier in the path
- Node can simply discard paths with loops
 - E.g., node A sees itself in path “C, B, A” & discards the advertisement



Differences between BGP and DV

(2) path-vector routing

- **Key idea:** advertise the entire path

- Distance vector: send distance metric per destination d
- Path vector: send the entire path for each destination d

- **Benefits**

- Loop avoidance is easy
- Flexible policies based on entire path

Differences between BGP and DV

(3) selective route advertisement

Differences between BGP and DV

(3) selective route advertisement

- For policy reasons, an AS may choose not to advertise a route to a destination

Differences between BGP and DV

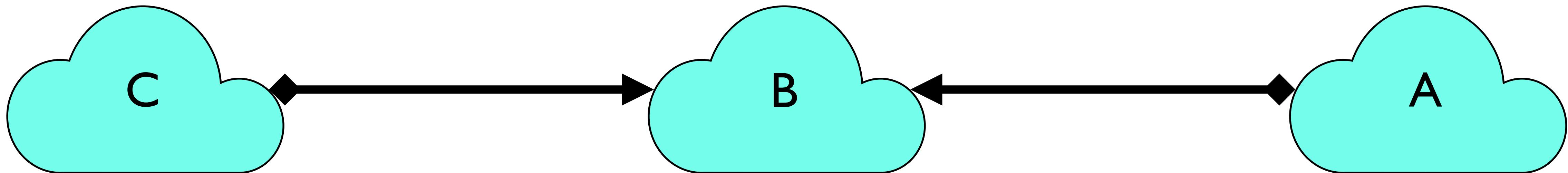
(3) selective route advertisement

- For policy reasons, an AS may choose not to advertise a route to a destination
- Hence reachability is not guaranteed even if the graph is connected

Differences between BGP and DV

(3) selective route advertisement

- For policy reasons, an AS may choose not to advertise a route to a destination
- Hence reachability is not guaranteed even if the graph is connected



Example: AS B does not want to carry traffic between AS A and AS C

Differences between BGP and DV

(4) BGP may *aggregate routes*

Differences between BGP and DV

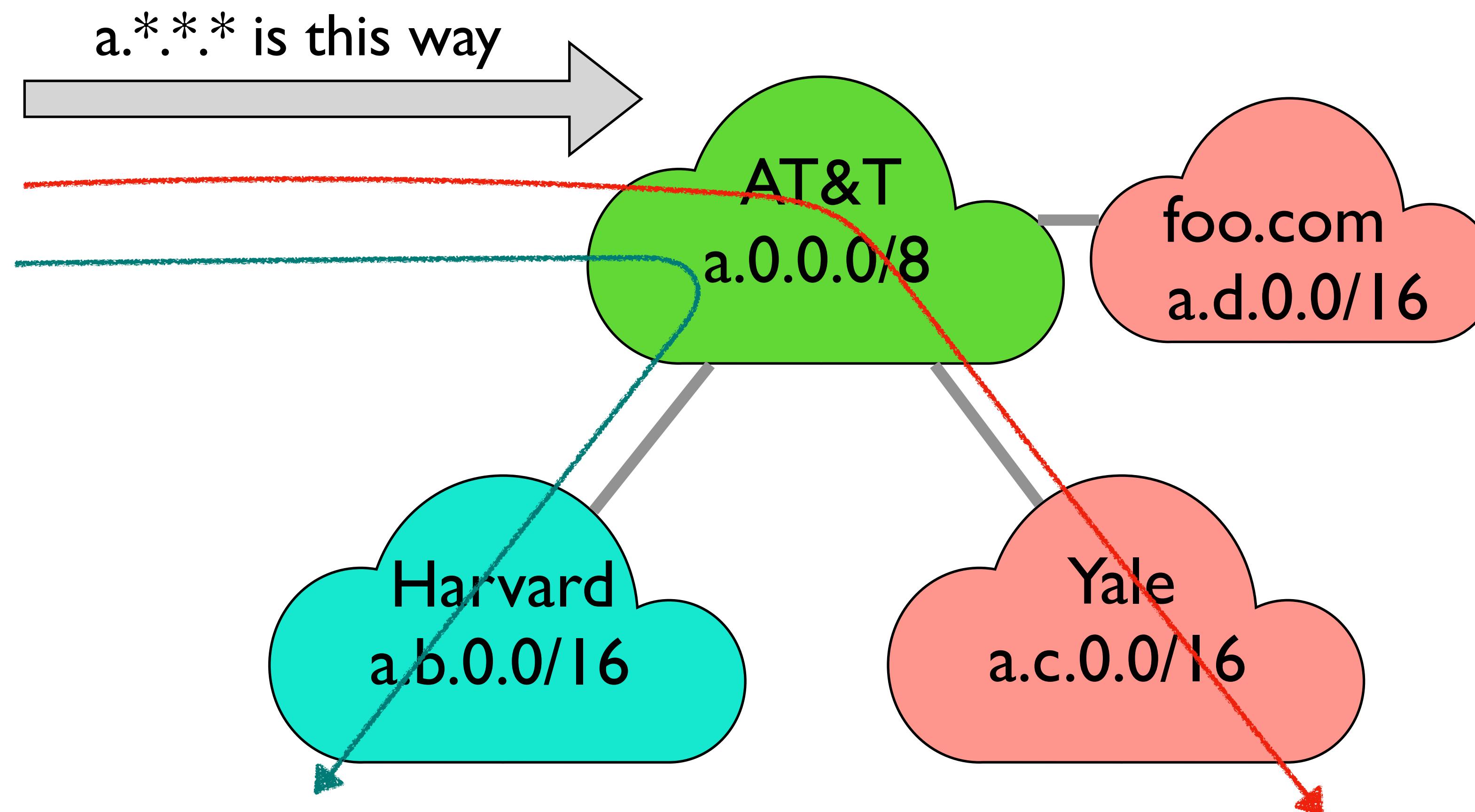
(4) BGP may *aggregate routes*

- For scalability, BGP may aggregate routes for different prefixes

Differences between BGP and DV

(4) BGP may *aggregate routes*

- For scalability, BGP may aggregate routes for different prefixes



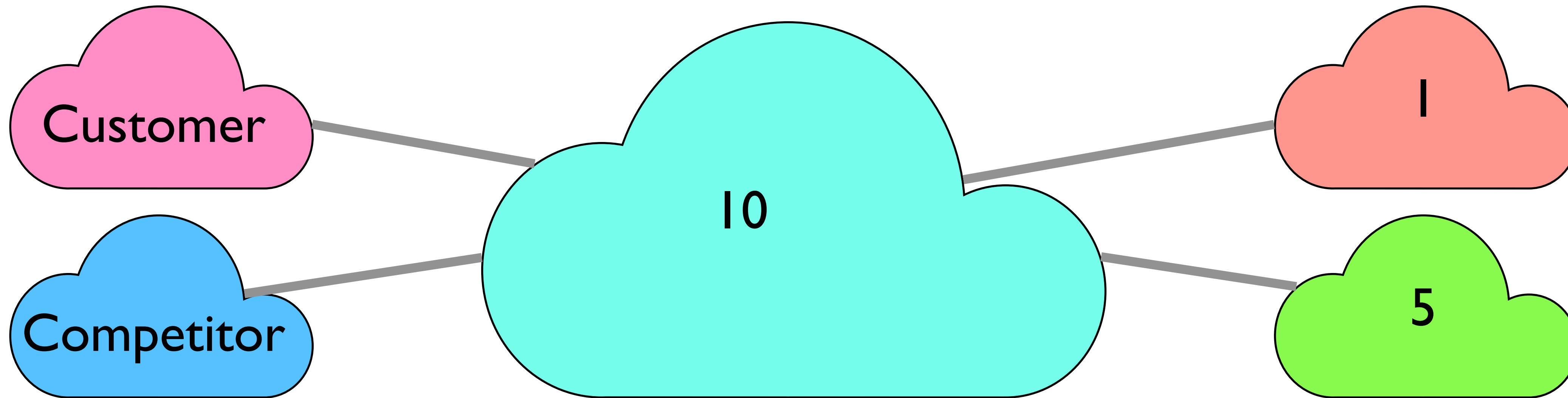
Questions?

BGP Outline

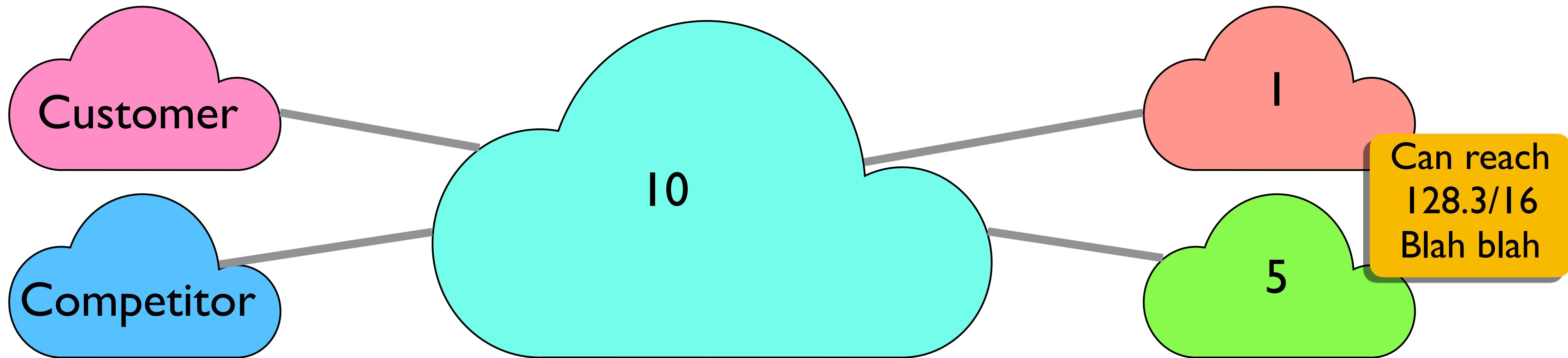
- **BGP Policy**
 - Typical policies, how they are implemented
- **BGP Protocol Details**
- **Issues with BGP**

Policy imposed in how routes are selected and exported

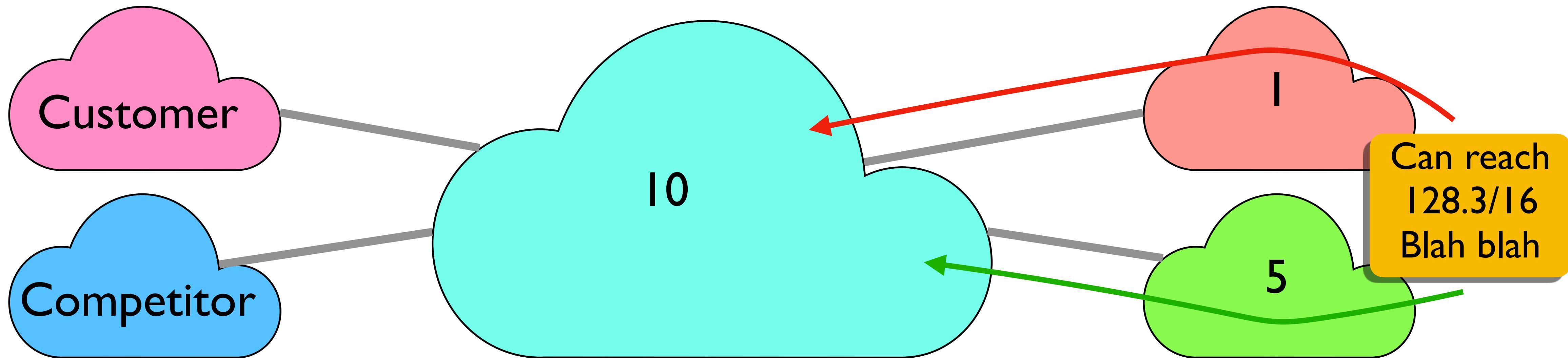
Policy imposed in how routes are selected and exported



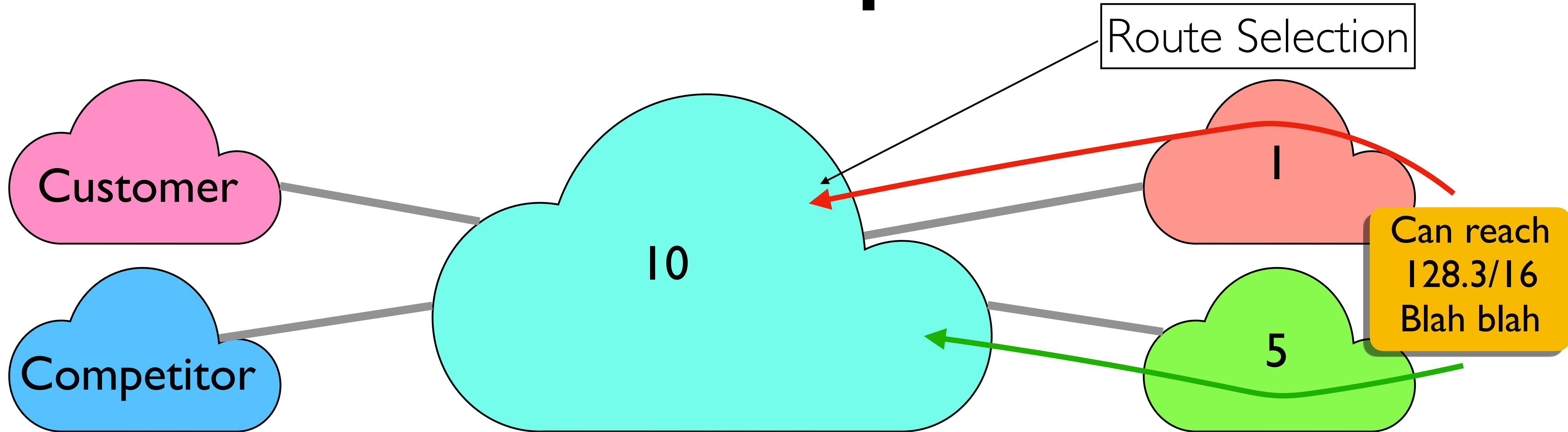
Policy imposed in how routes are selected and exported



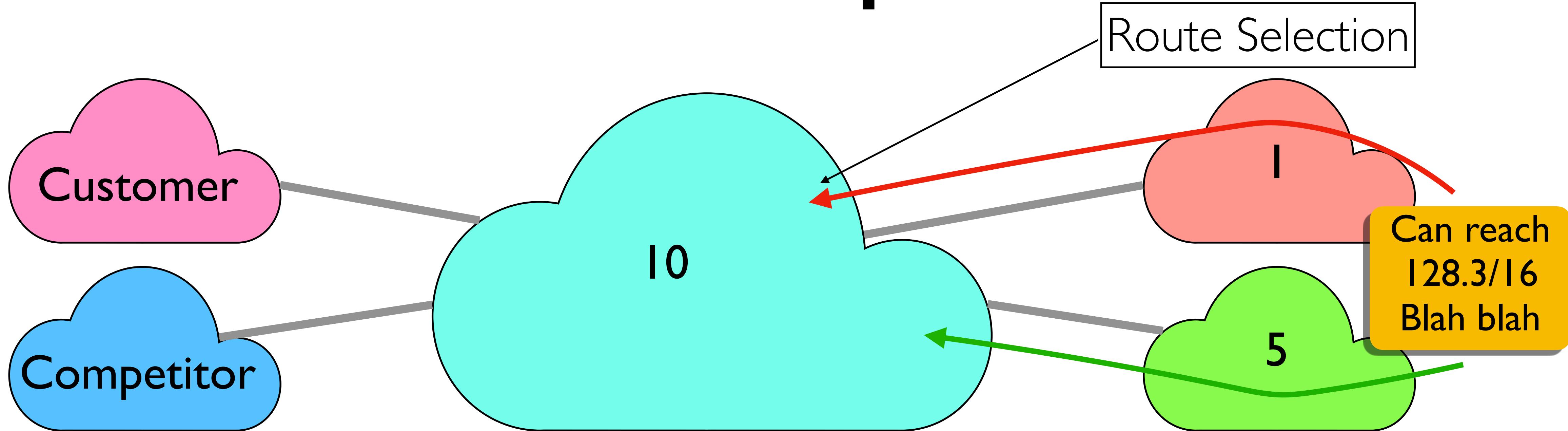
Policy imposed in how routes are selected and exported



Policy imposed in how routes are selected and exported

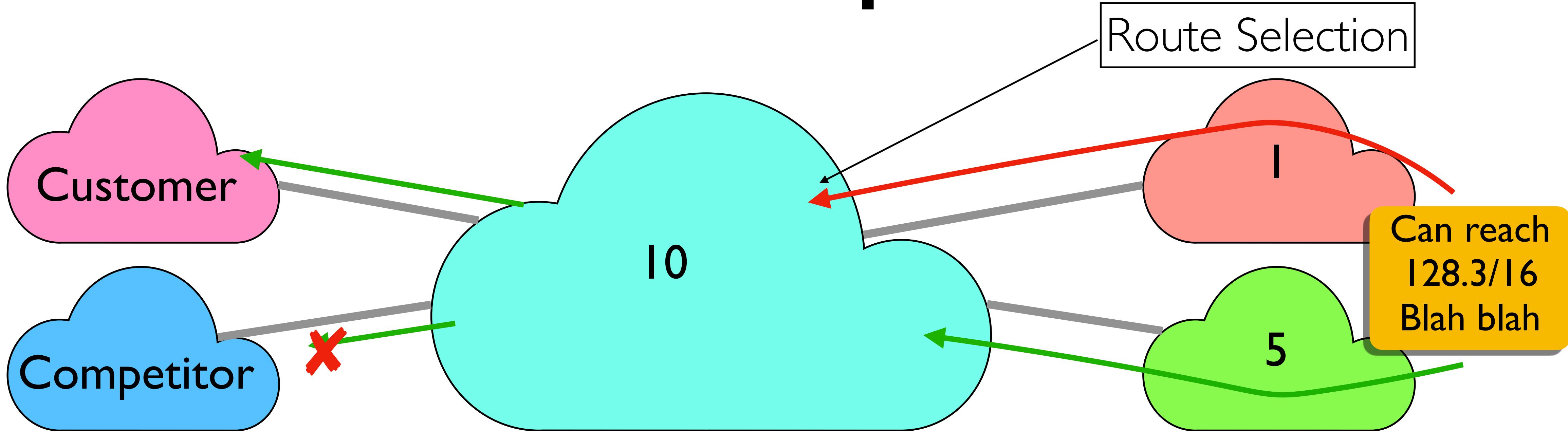


Policy imposed in how routes are selected and exported



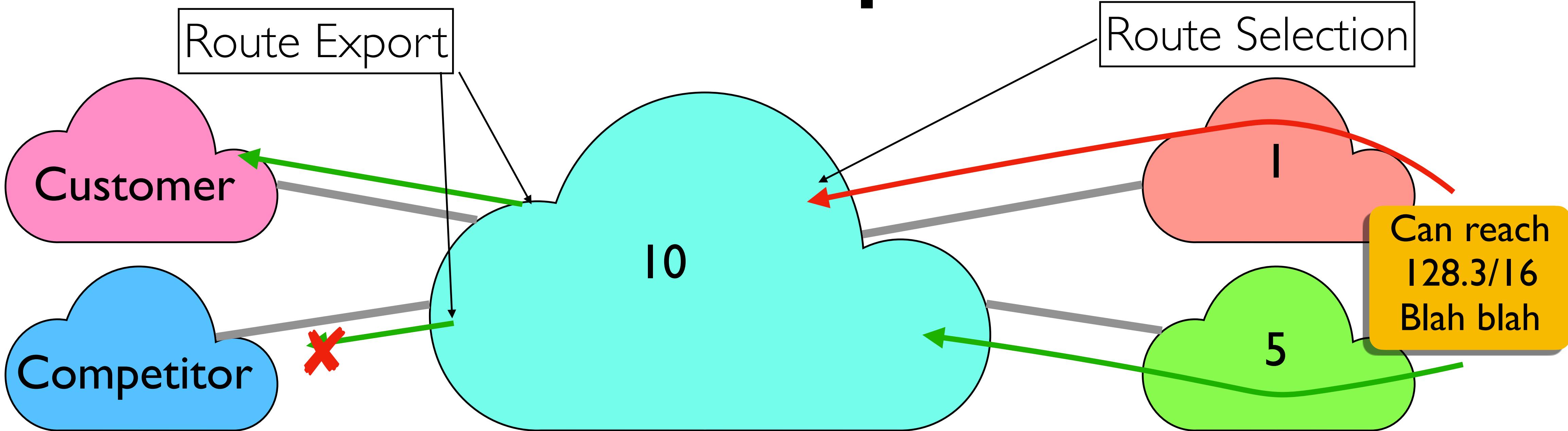
- **Selection:** Which path to use?
 - Controls whether/how traffic **leaves** the network

Policy imposed in how routes are selected and exported



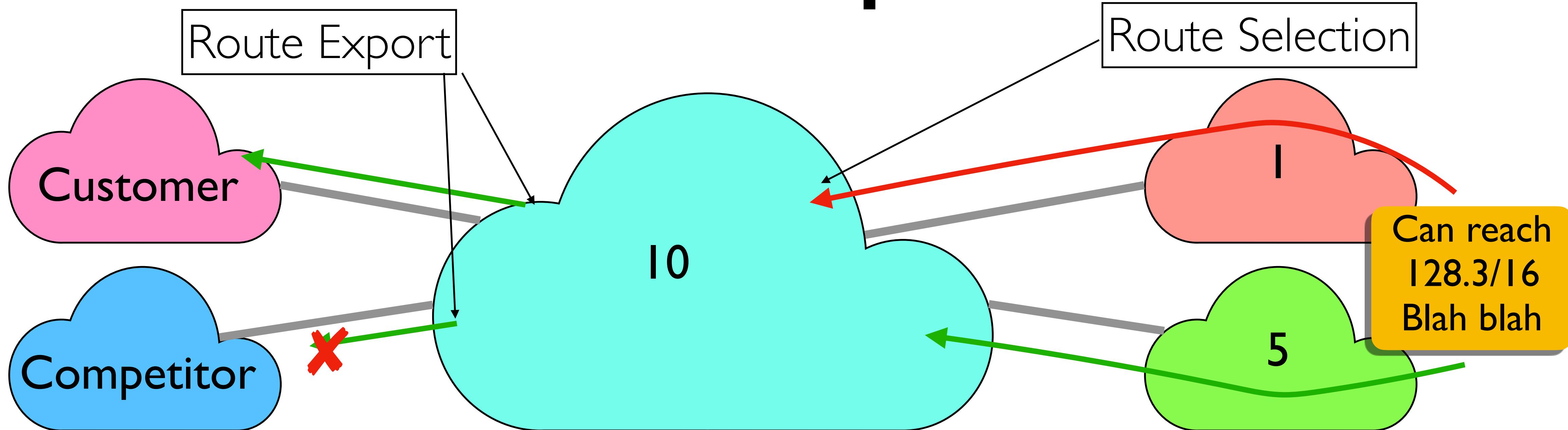
- **Selection:** Which path to use?
 - Controls whether/how traffic **leaves** the network

Policy imposed in how routes are selected and exported



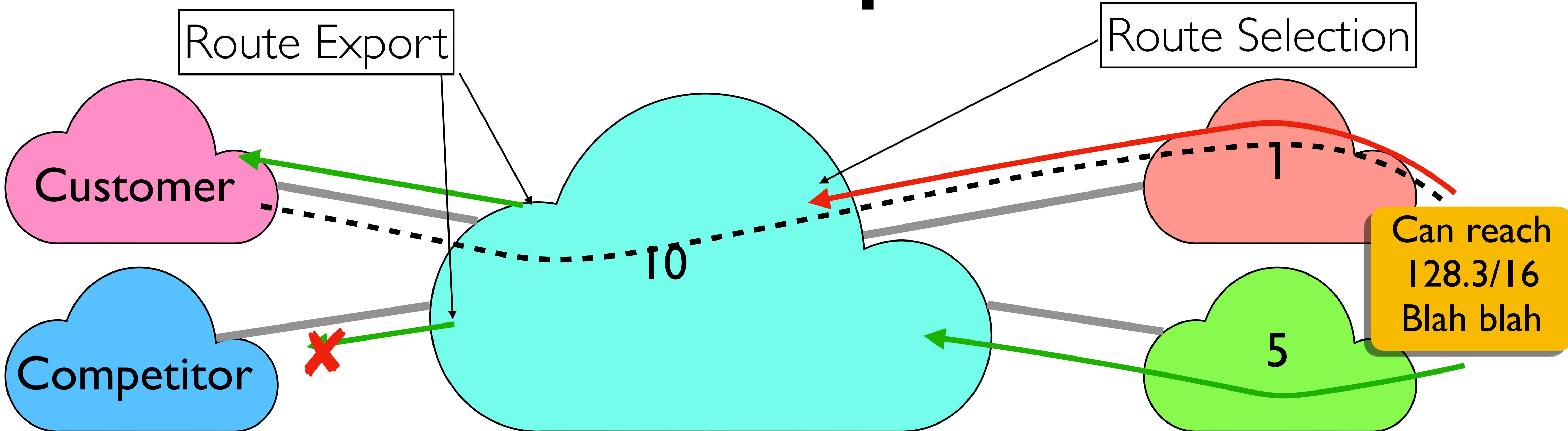
- **Selection:** Which path to use?
 - Controls whether/how traffic **leaves** the network

Policy imposed in how routes are selected and exported



- **Selection:** Which path to use?
 - Controls whether/how traffic **leaves** the network
- **Export:** Which paths to advertise?
 - Controls whether/how traffic **enters** the network

Policy imposed in how routes are selected and exported



- **Selection:** Which path to use?
 - Controls whether/how traffic **leaves** the network
- **Export:** Which paths to advertise?
 - Controls whether/how traffic **enters** the network

Typical Selection Policy

Typical Selection Policy

- In decreasing order of priority
 - Make/save money (send to customer > peer > provider)
 - Maximize performance (small AS path length)
 - Minimize use of my network bandwidth (“hot potato”)
 - ...

Typical Selection Policy

- In decreasing order of priority
 - Make/save money (send to customer > peer > provider)
 - Maximize performance (small AS path length)
 - Minimize use of my network bandwidth (“hot potato”)
 - ...
- BGP uses something called route “attributes” to implement the above (soon...)

Typical Export: Peer-Peer Case

- Peer exchange traffic between their customers



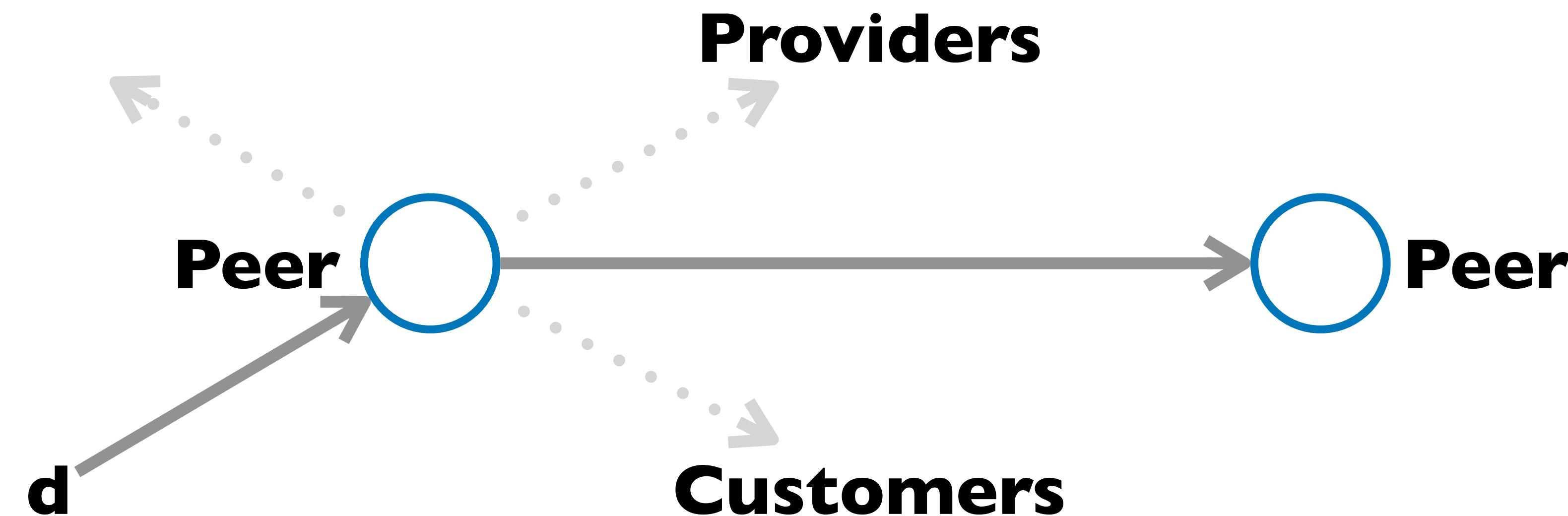
Typical Export: Peer-Peer Case

- Peer exchange traffic between their customers
 - AS exports **only customer routes to a peer**



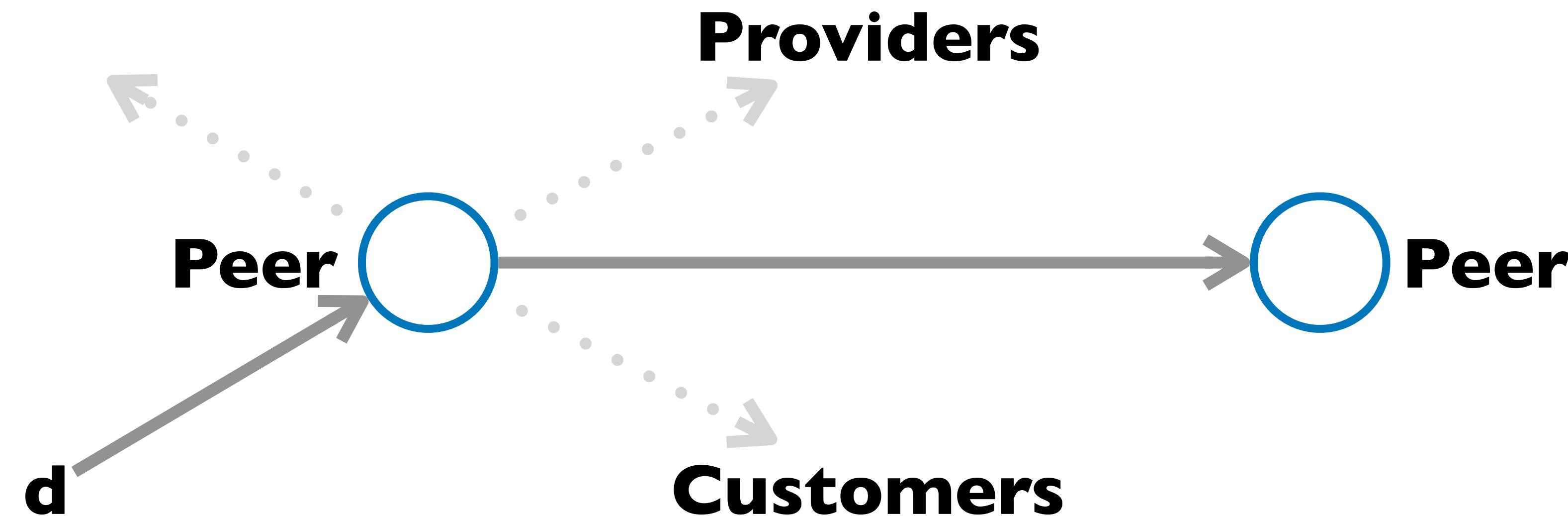
Typical Export: Peer-Peer Case

- Peer exchange traffic between their customers
 - AS exports **only customer routes to a peer**



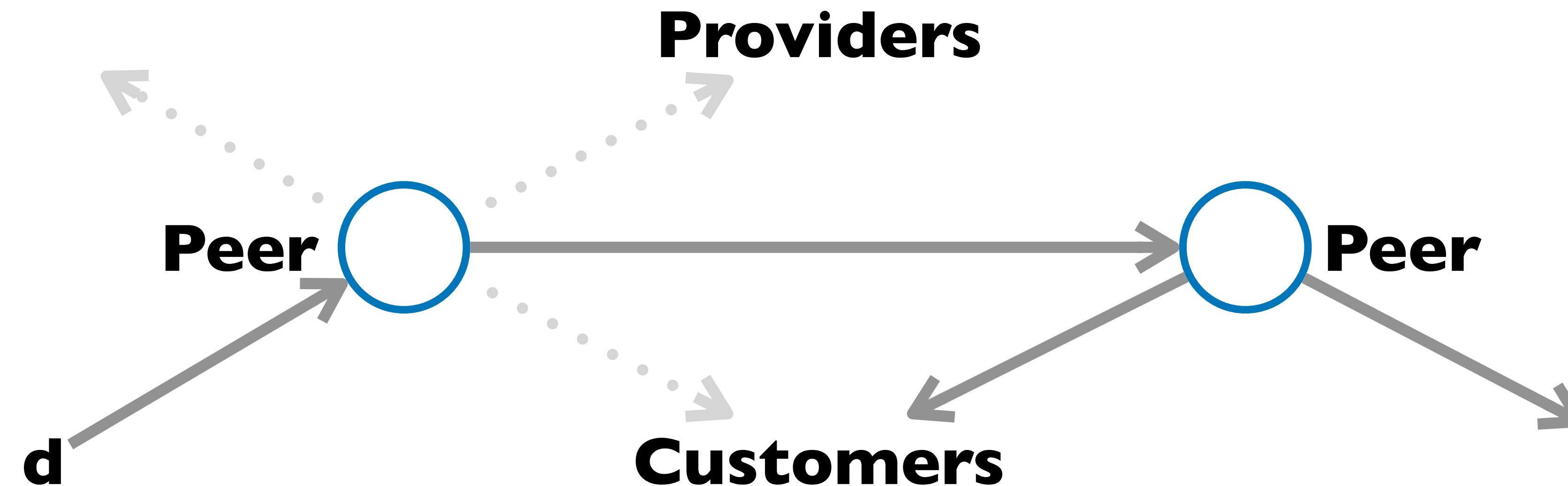
Typical Export: Peer-Peer Case

- Peer exchange traffic between their customers
 - AS exports **only customer routes to a peer**
 - AS exports **a peer's routes only to its customers**



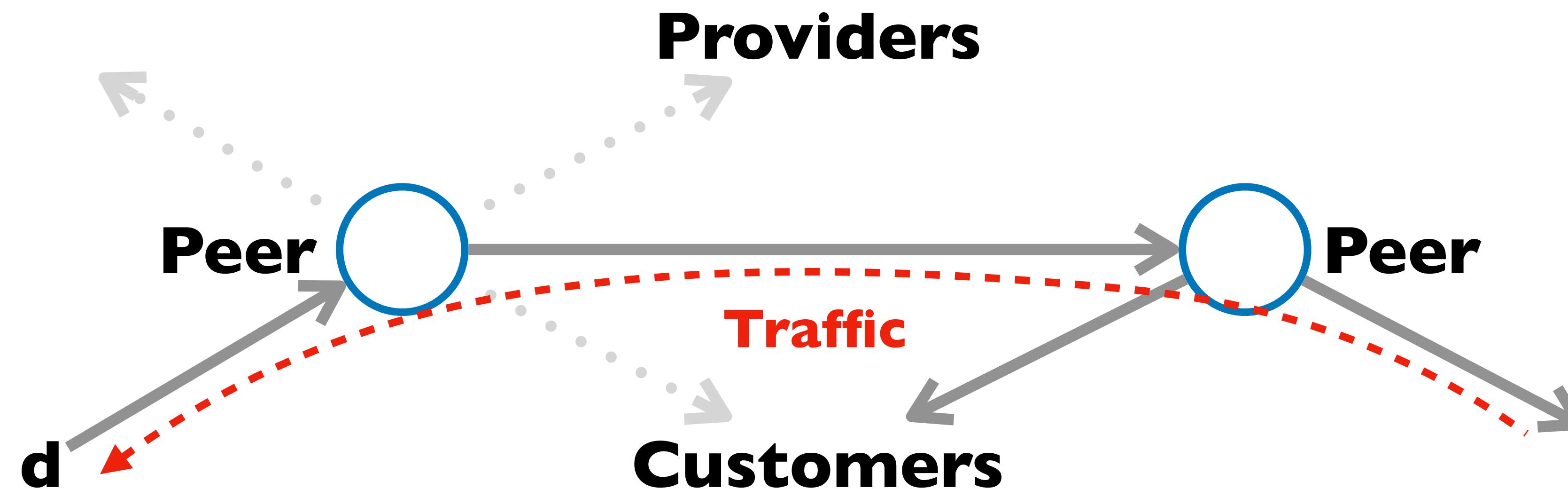
Typical Export: Peer-Peer Case

- Peer exchange traffic between their customers
 - AS exports **only customer routes to a peer**
 - AS exports **a peer's routes only to its customers**



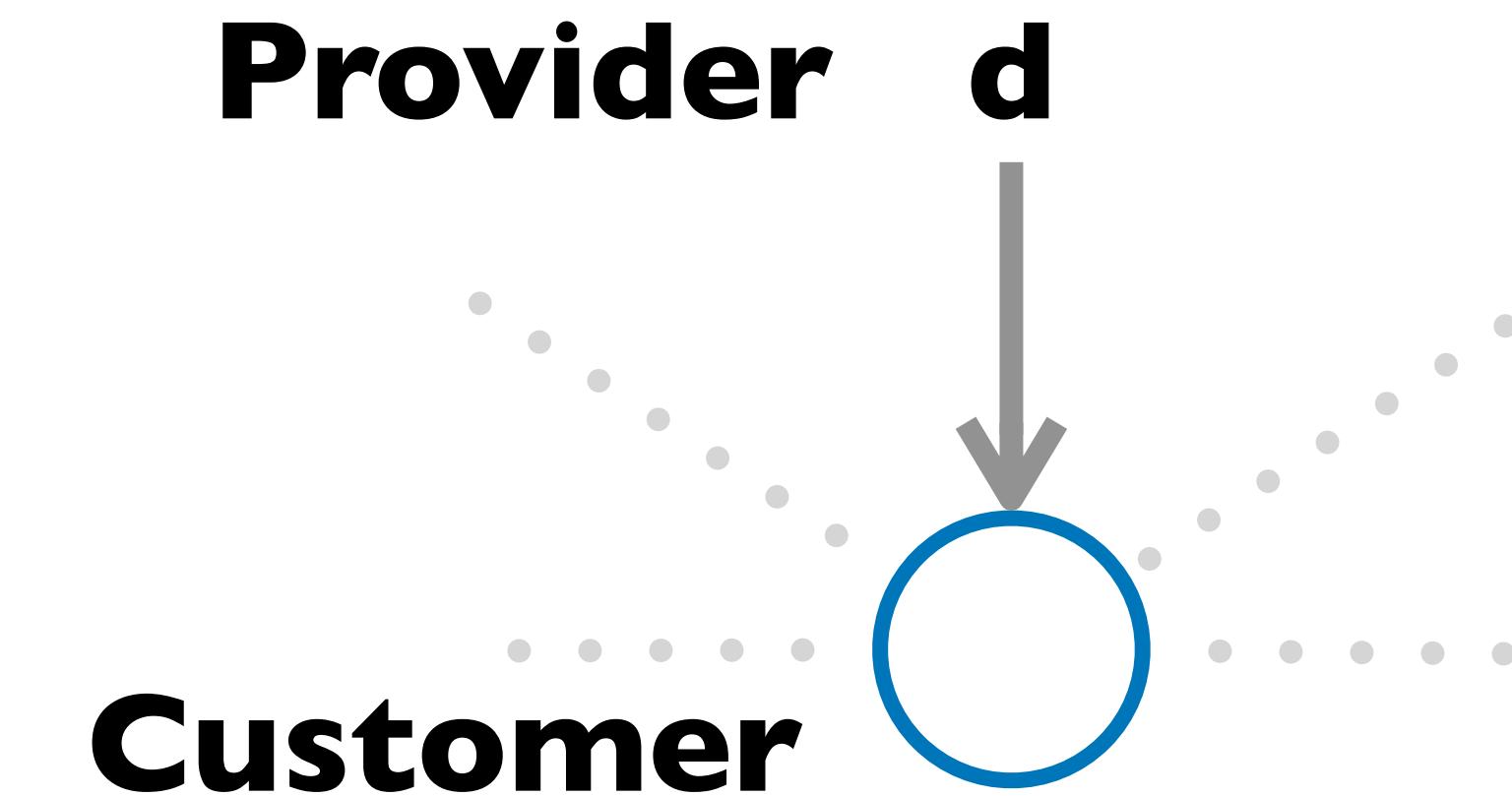
Typical Export: Peer-Peer Case

- Peer exchange traffic between their customers
 - AS exports **only customer routes to a peer**
 - AS exports **a peer's routes only to its customers**



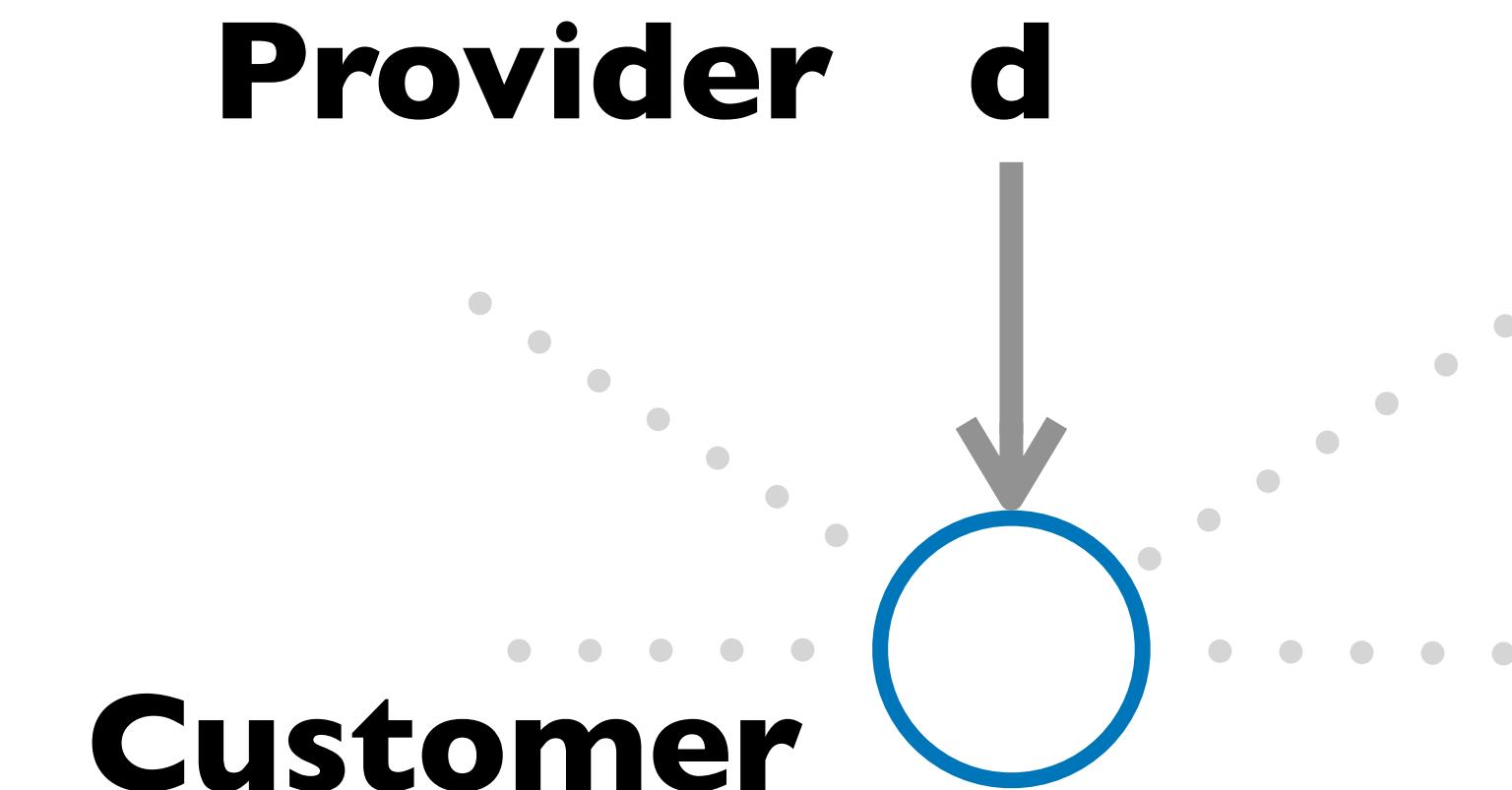
Typical Export: Customer-Provider Case

- Customer pays provider for access to Internet



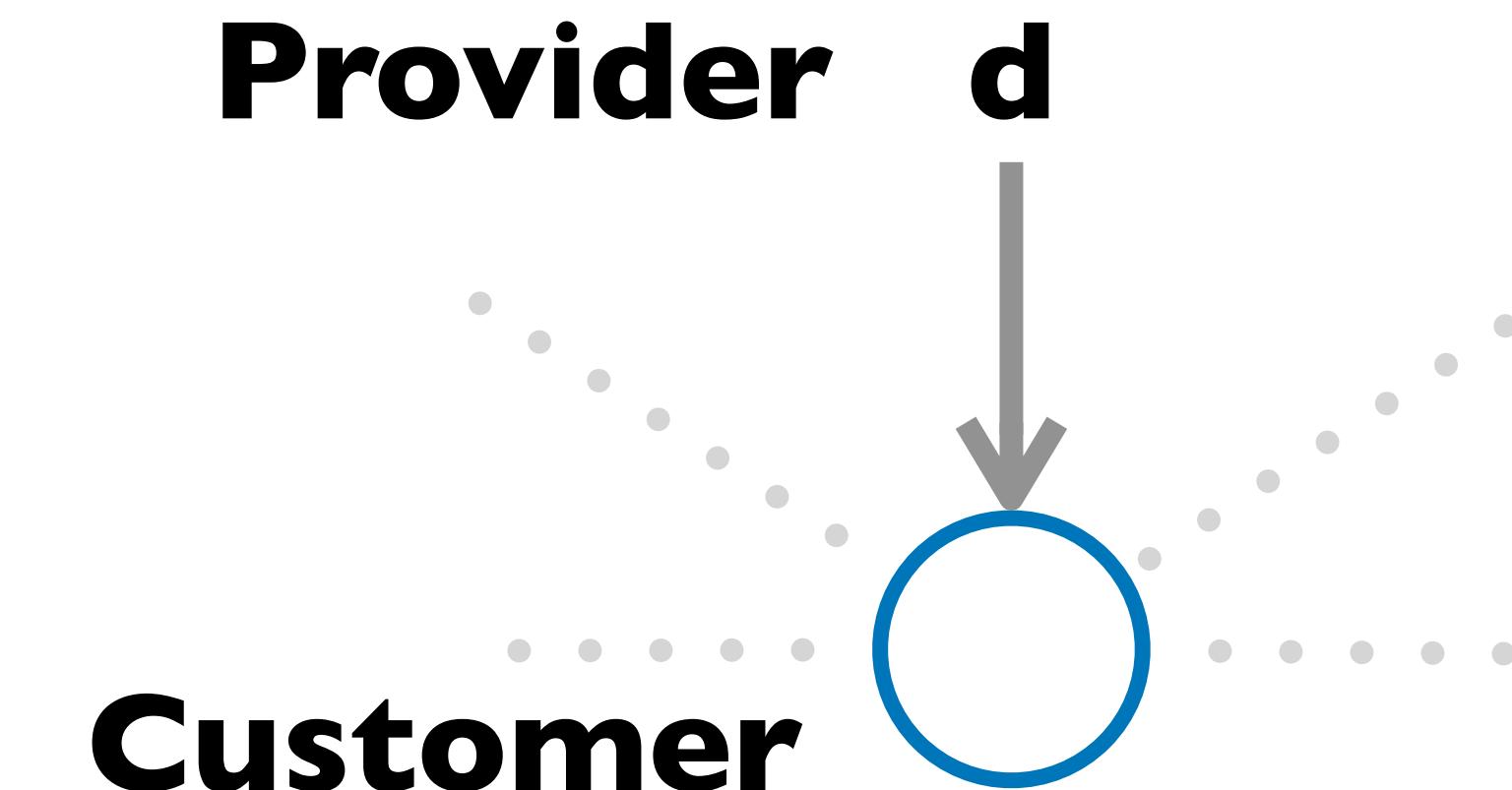
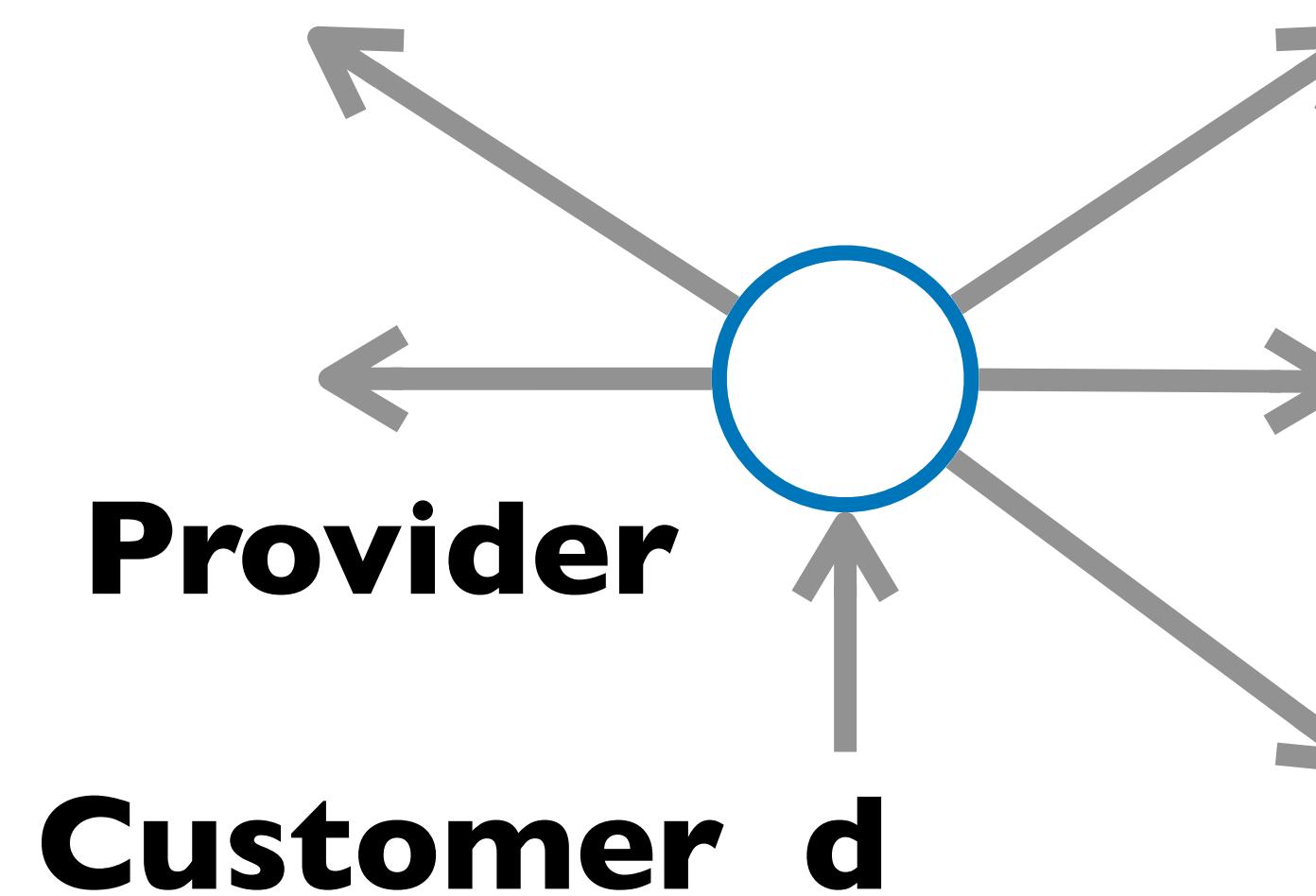
Typical Export: Customer-Provider Case

- Customer pays provider for access to Internet
 - Provider exports **its customer routes to everybody**



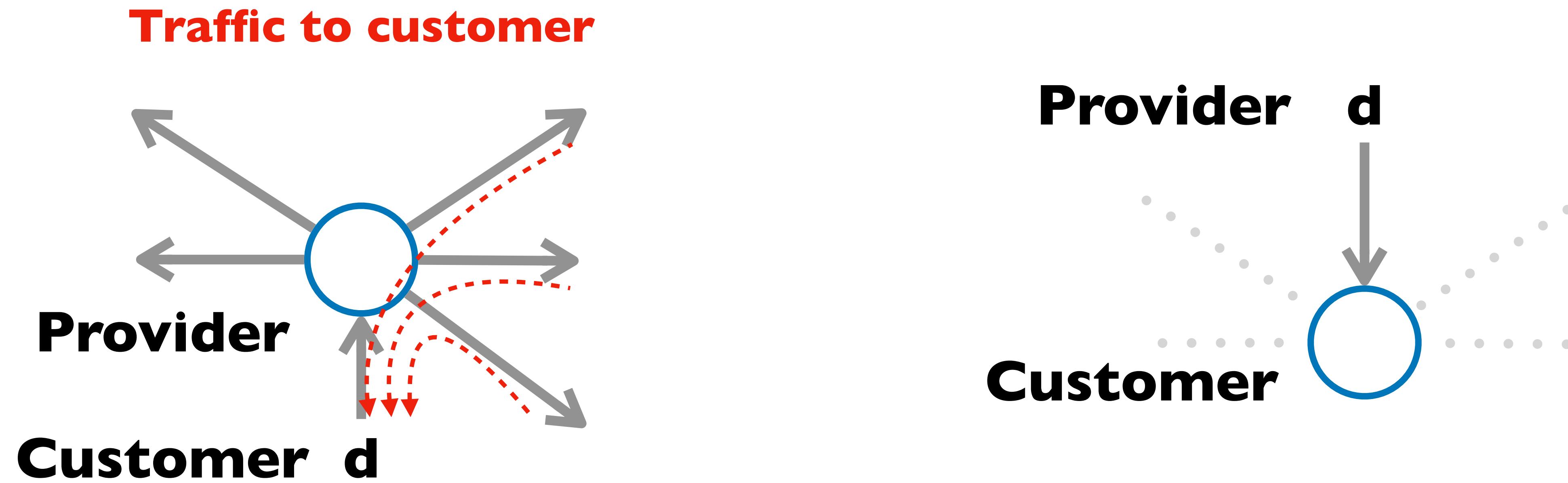
Typical Export: Customer-Provider Case

- Customer pays provider for access to Internet
 - Provider exports **its customer routes to everybody**



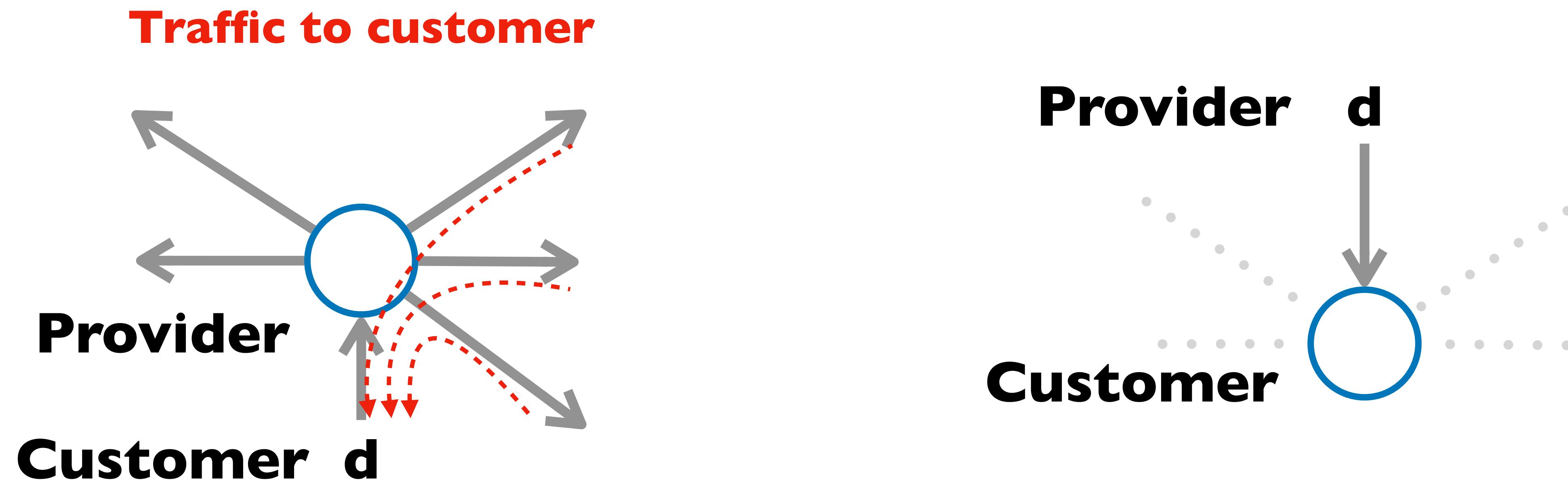
Typical Export: Customer-Provider Case

- Customer pays provider for access to Internet
 - Provider exports **its customer routes to everybody**



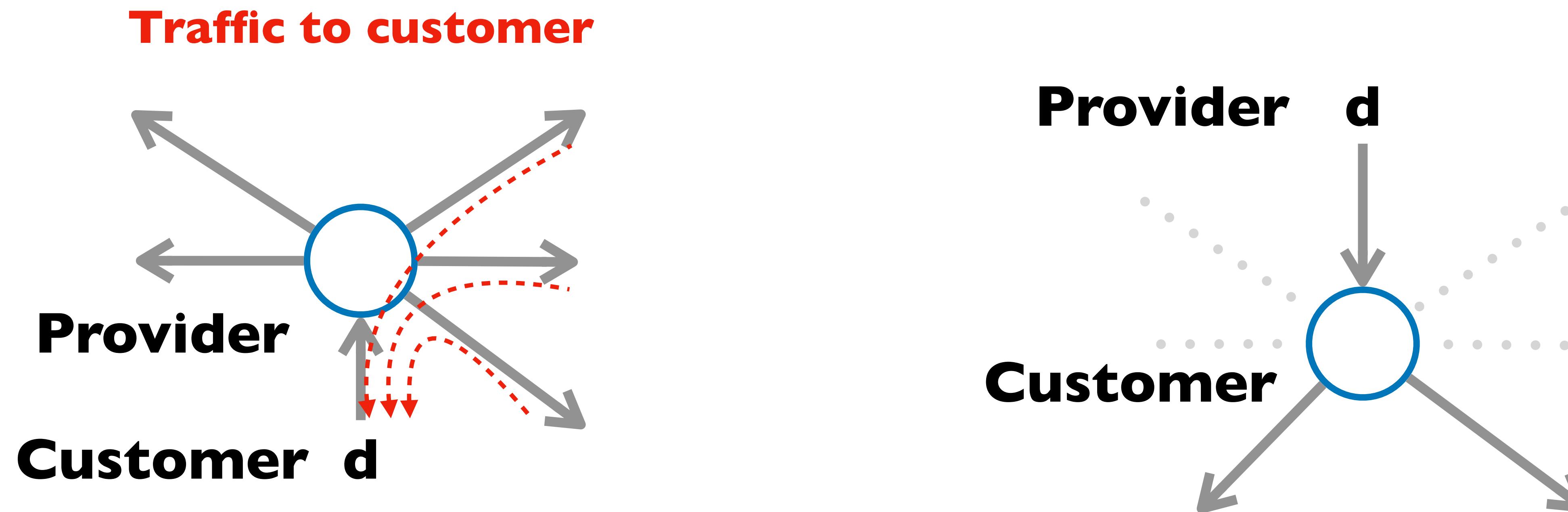
Typical Export: Customer-Provider Case

- Customer pays provider for access to Internet
 - Provider exports **its customer routes to everybody**
 - Customer exports **a provider routes only to its customers**



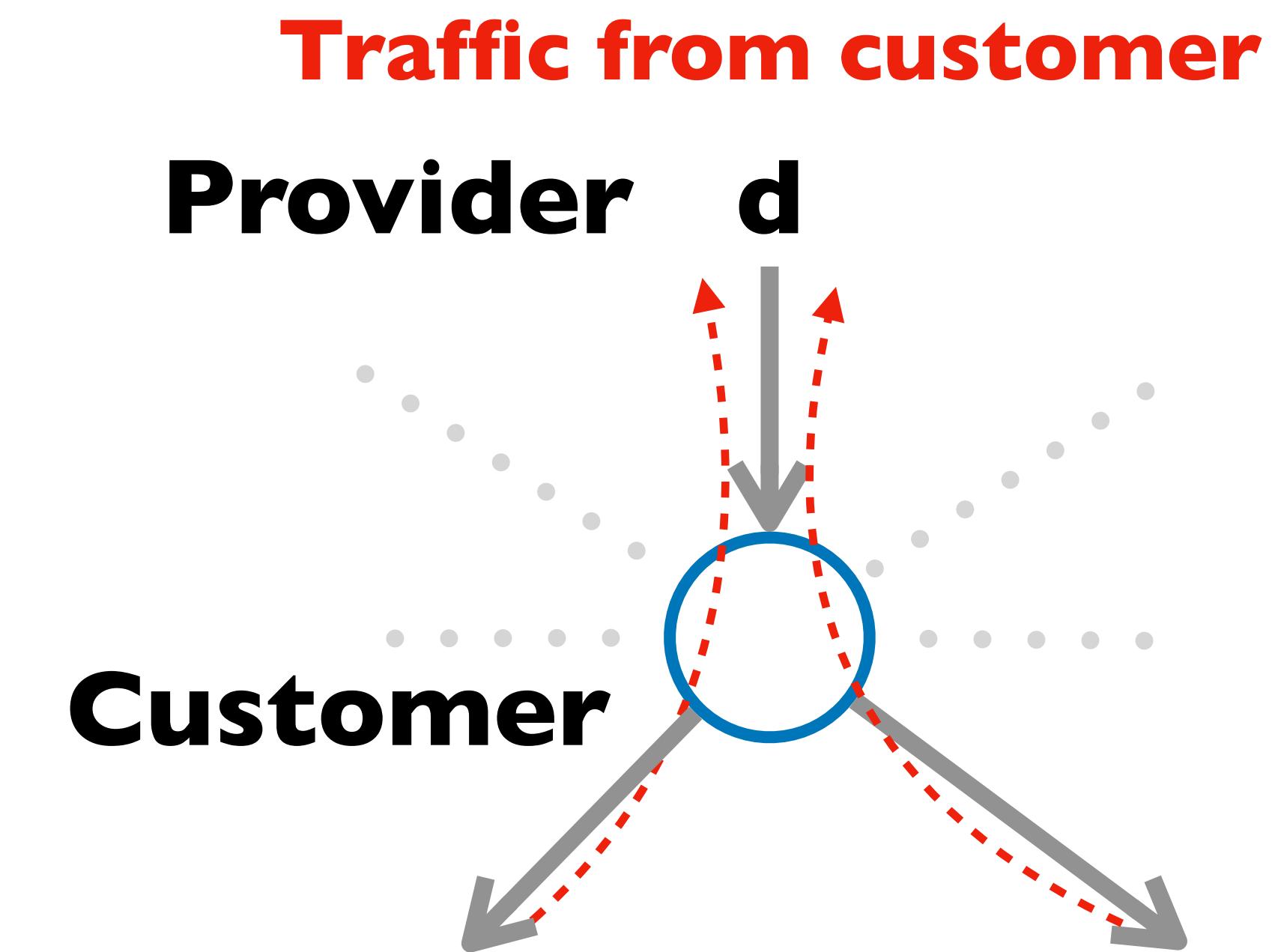
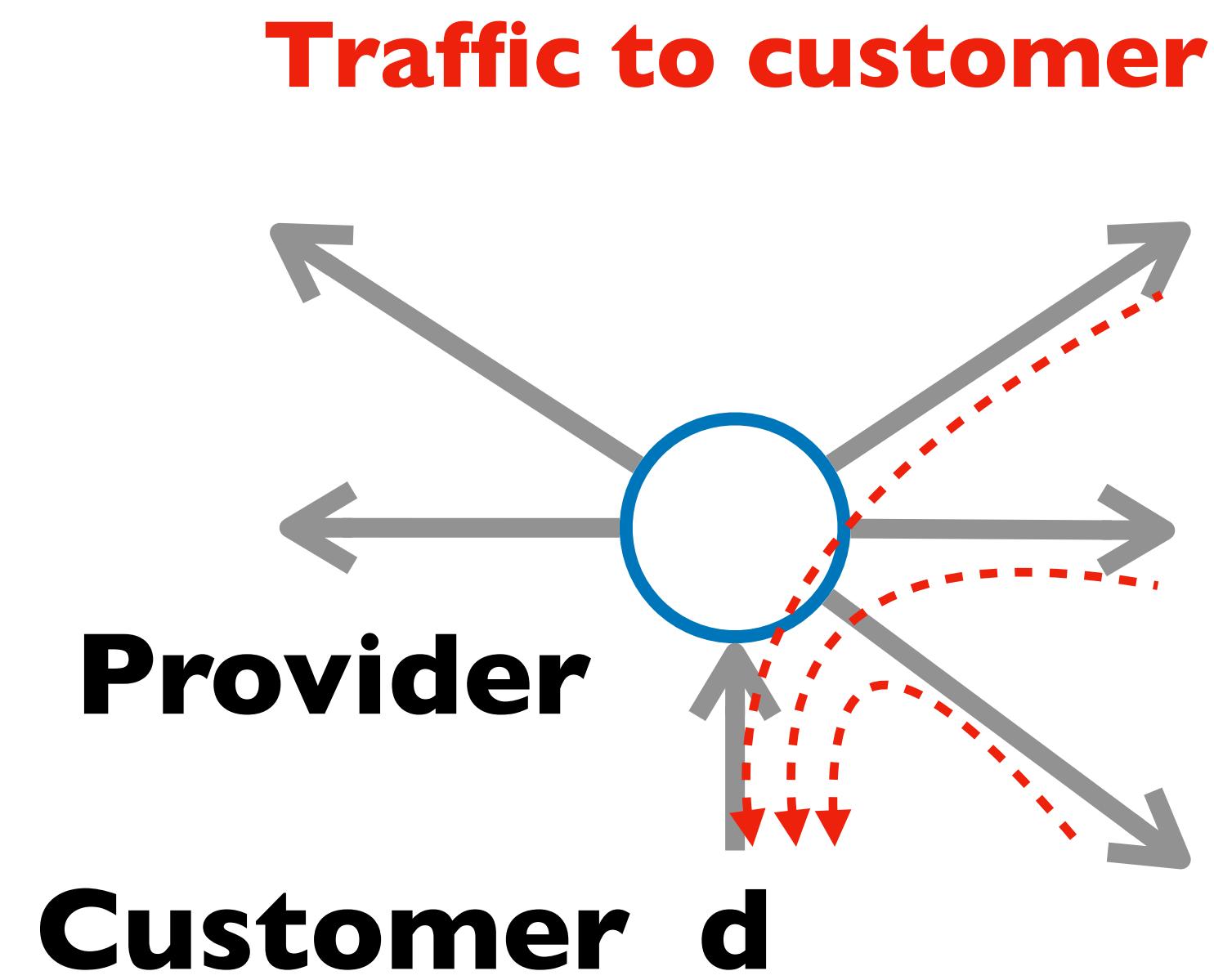
Typical Export: Customer-Provider Case

- Customer pays provider for access to Internet
 - Provider exports **its customer routes to everybody**
 - Customer exports **a provider routes only to its customers**



Typical Export: Customer-Provider Case

- Customer pays provider for access to Internet
 - Provider exports **its customer routes to everybody**
 - Customer exports **a provider routes only to its customers**



Typical Export Policy

Destination prefix advertised by...	Export route to...
Customer	Everyone (providers, peers, other customers)
Peer	Customers
Provider	Customers

We'll refer to these as the “Gao-Rexford” rules
(Capture common — but not required! — practice)



Gao-Rexford





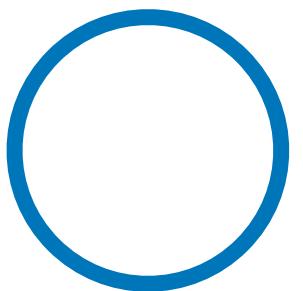
Gao-Rexford



Providers

Peers

Customers





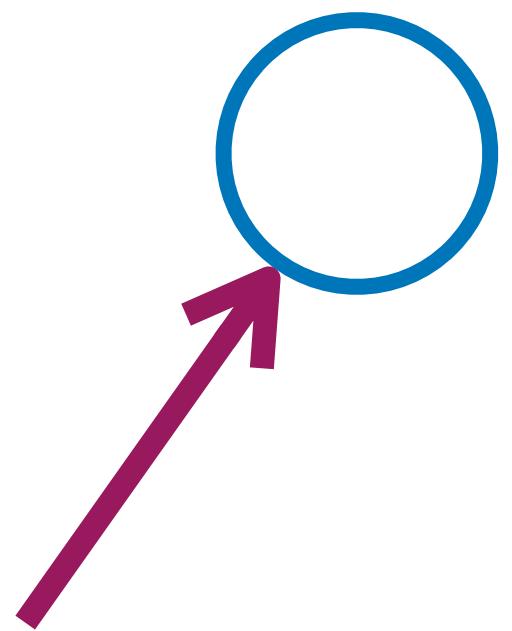
Gao-Rexford



Providers

Peers

Customers





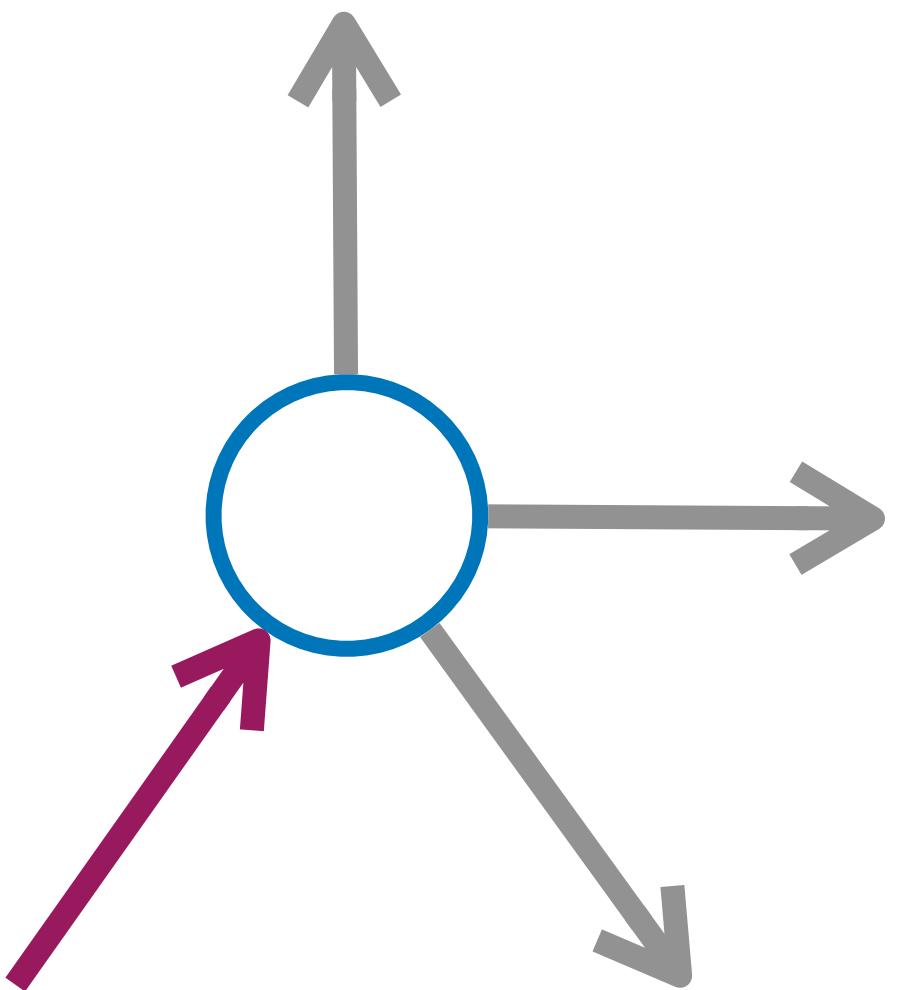
Gao-Rexford



Providers

Peers

Customers





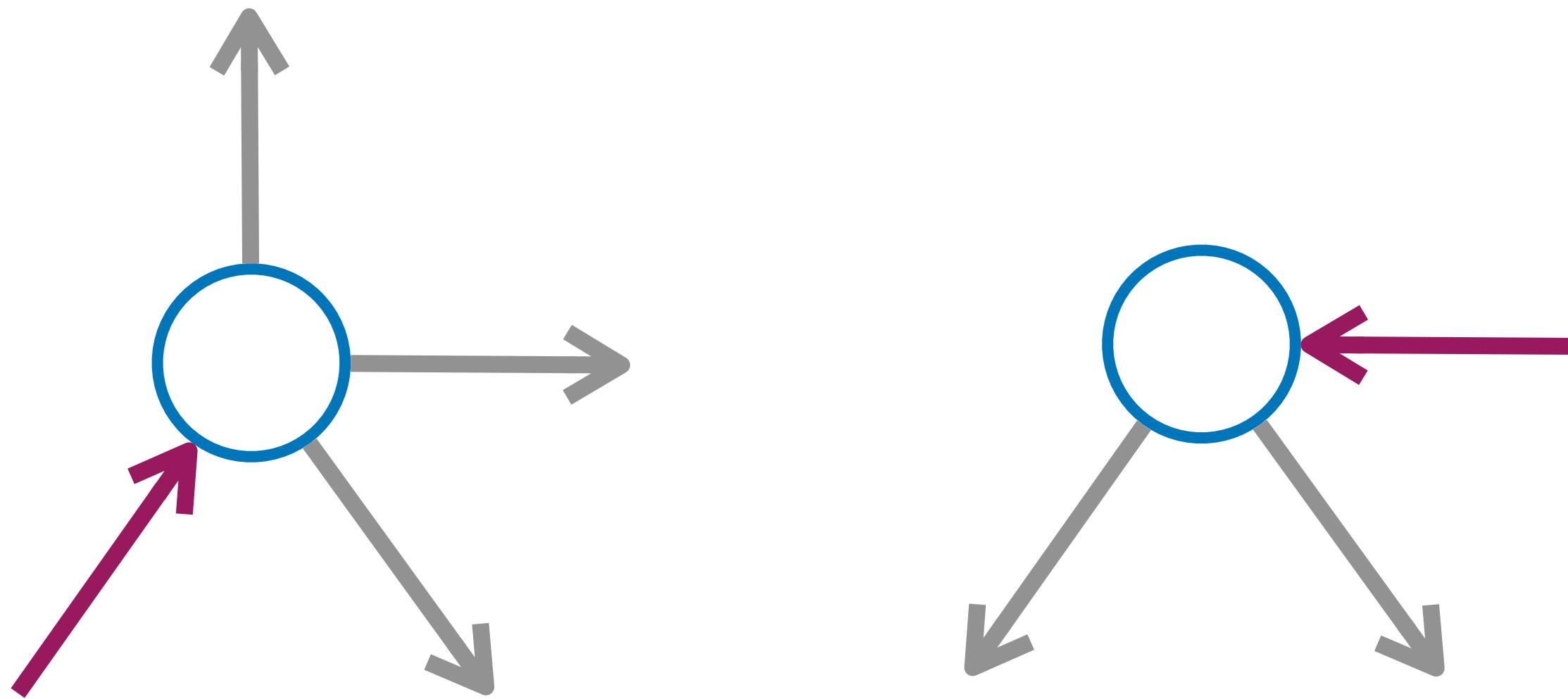
Gao-Rexford



Providers

Peers

Customers





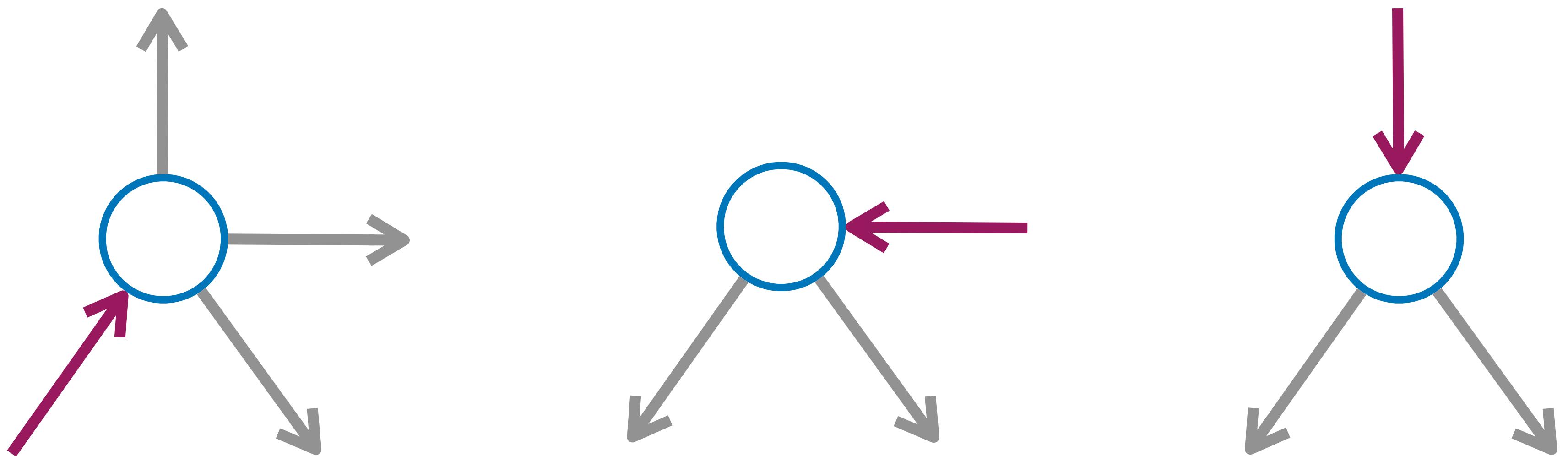
Gao-Rexford



Providers

Peers

Customers





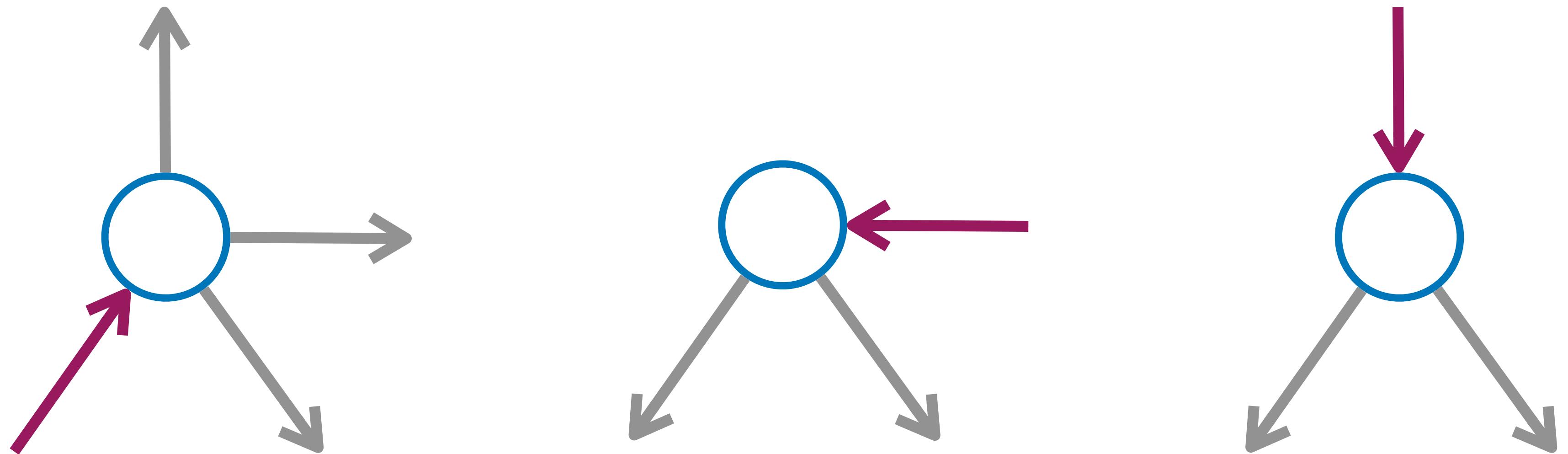
Gao-Rexford



Providers

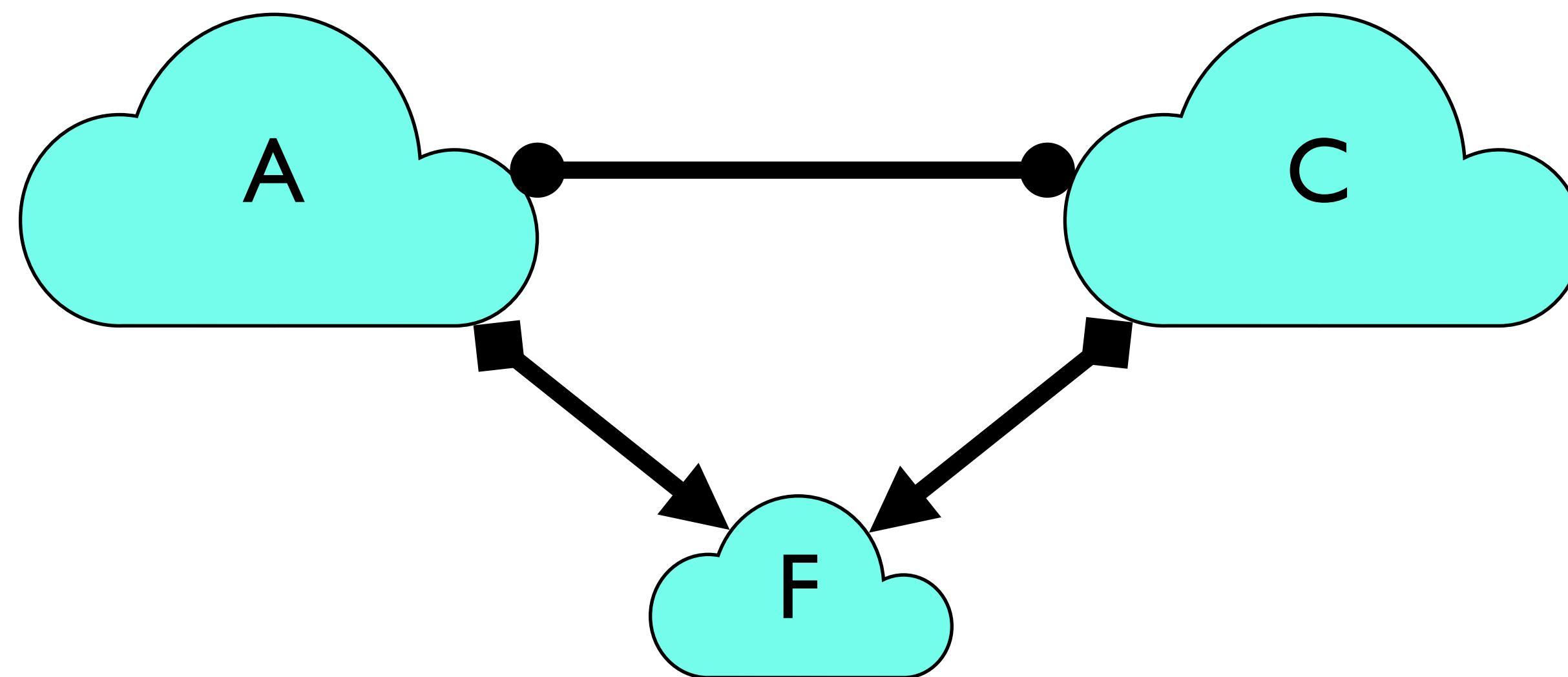
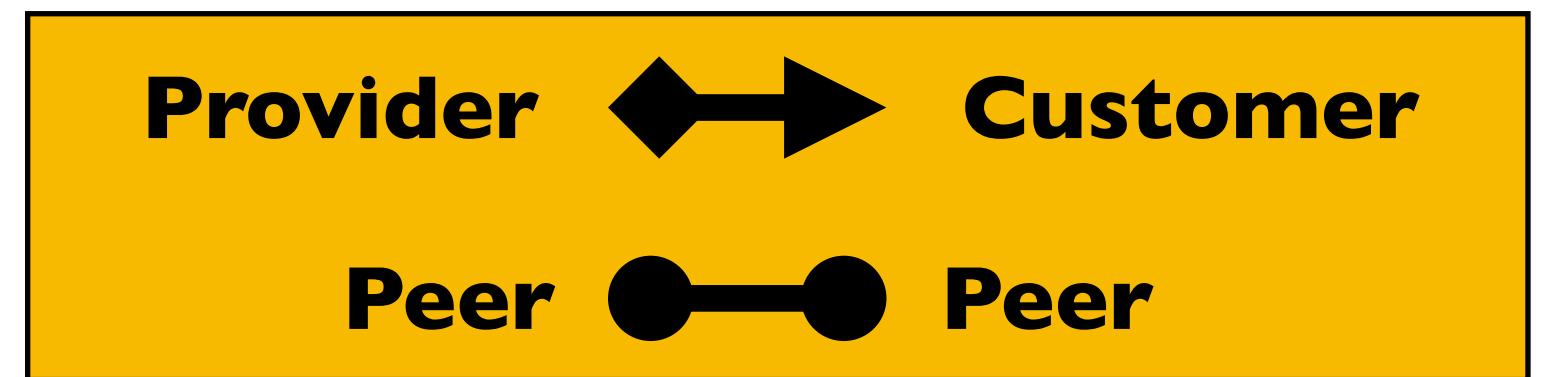
Peers

Customers

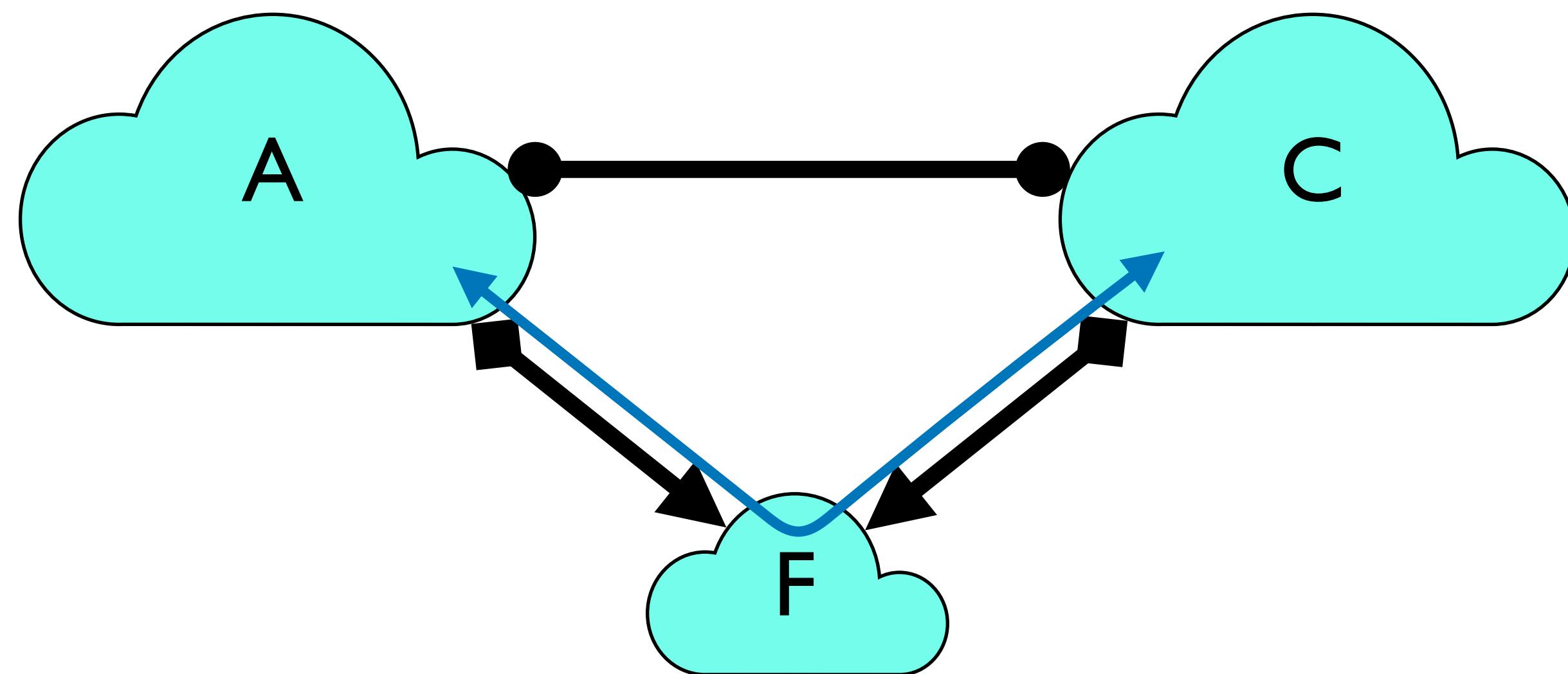
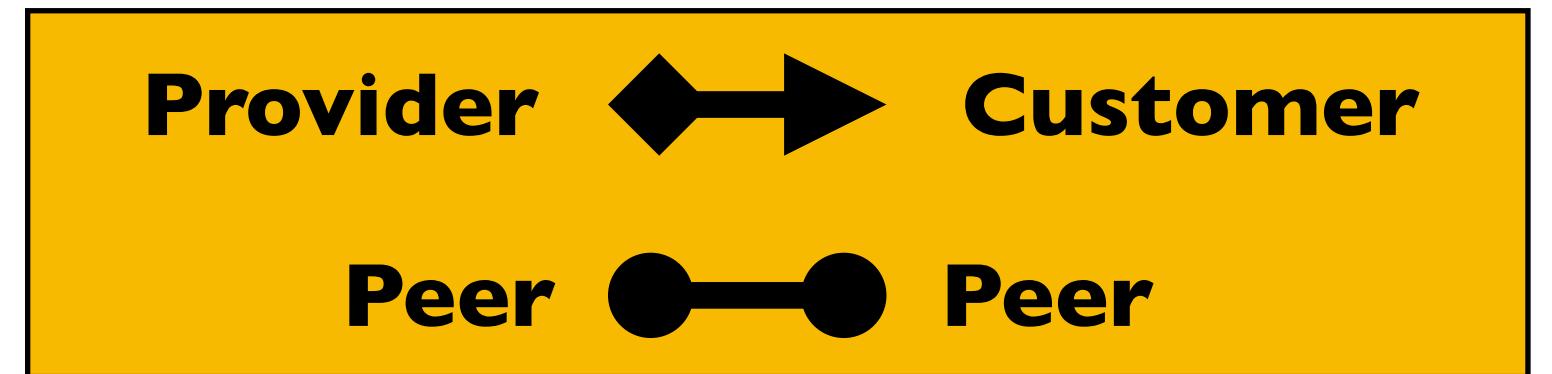


**With Gao-Rexford, the AS policy graph is a DAG
(directed acyclic graph) and routes are “valley free”**

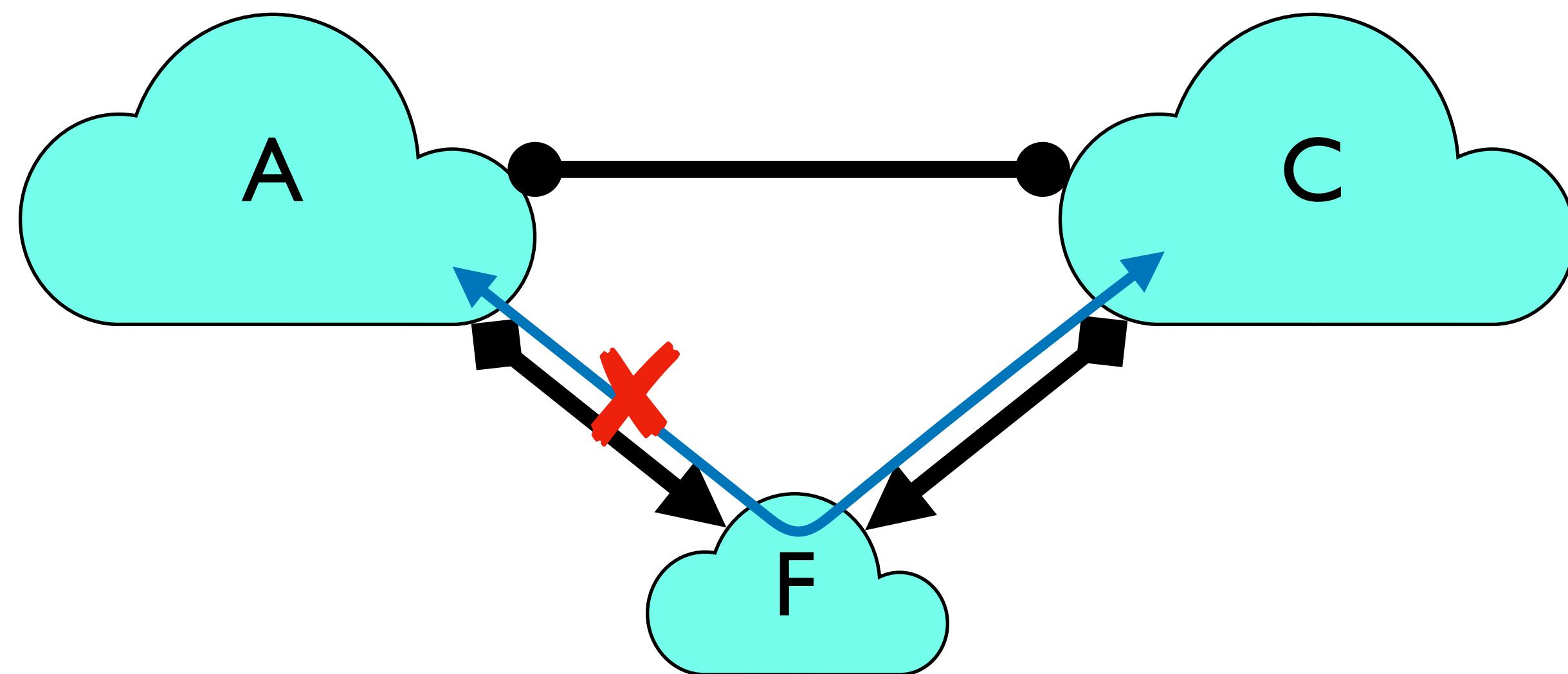
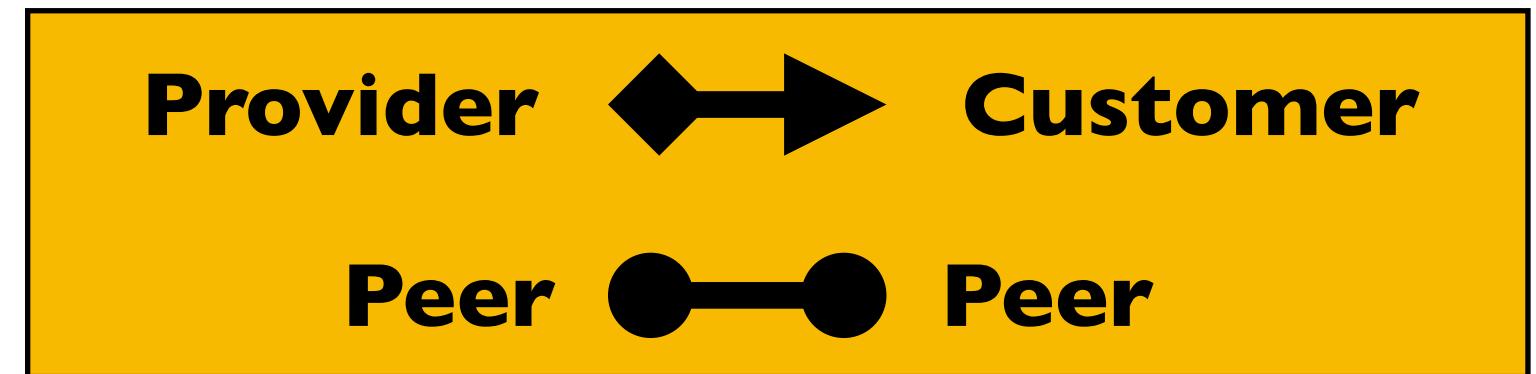
Valley Freedom



Valley Freedom



Valley Freedom



Questions?

BGP Outline

- **BGP Policy**

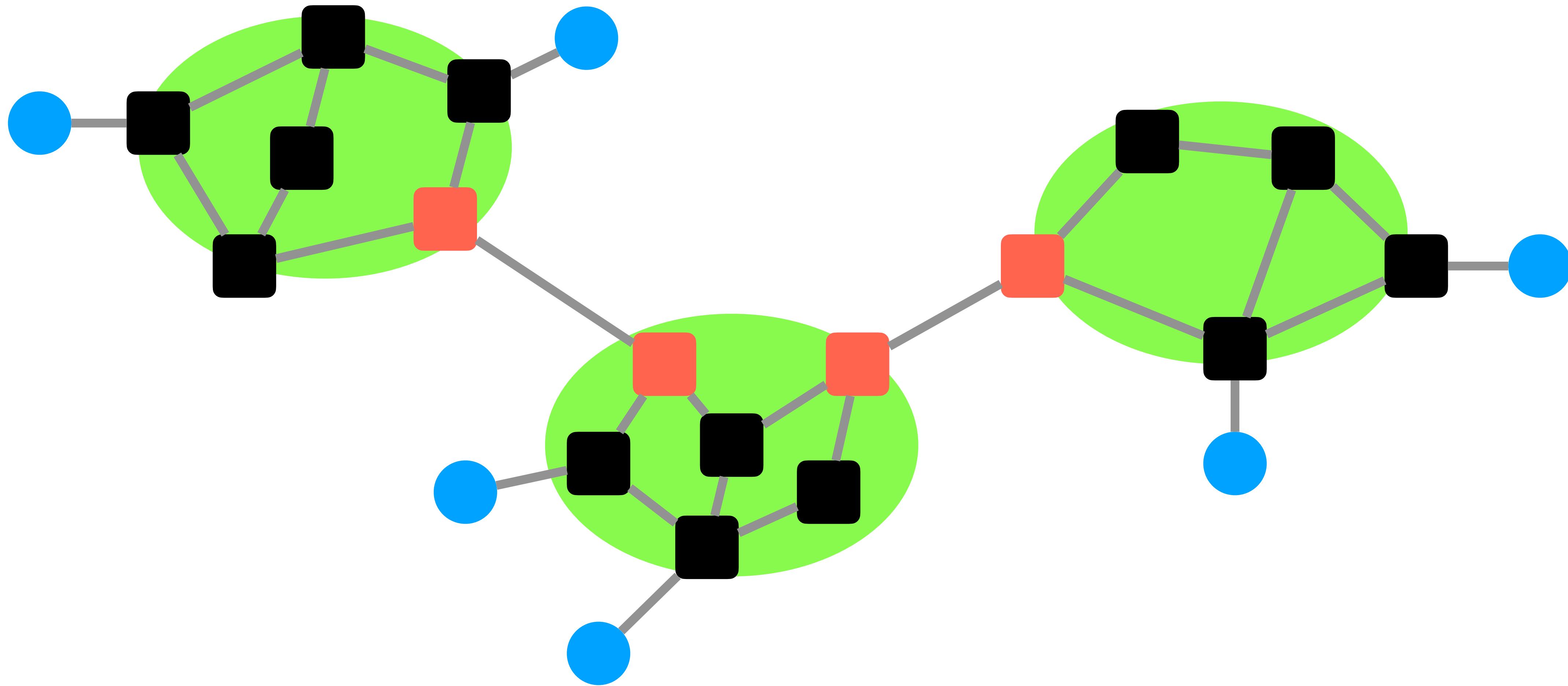
- Typical policies, how they are implemented

- **BGP Protocol Details**

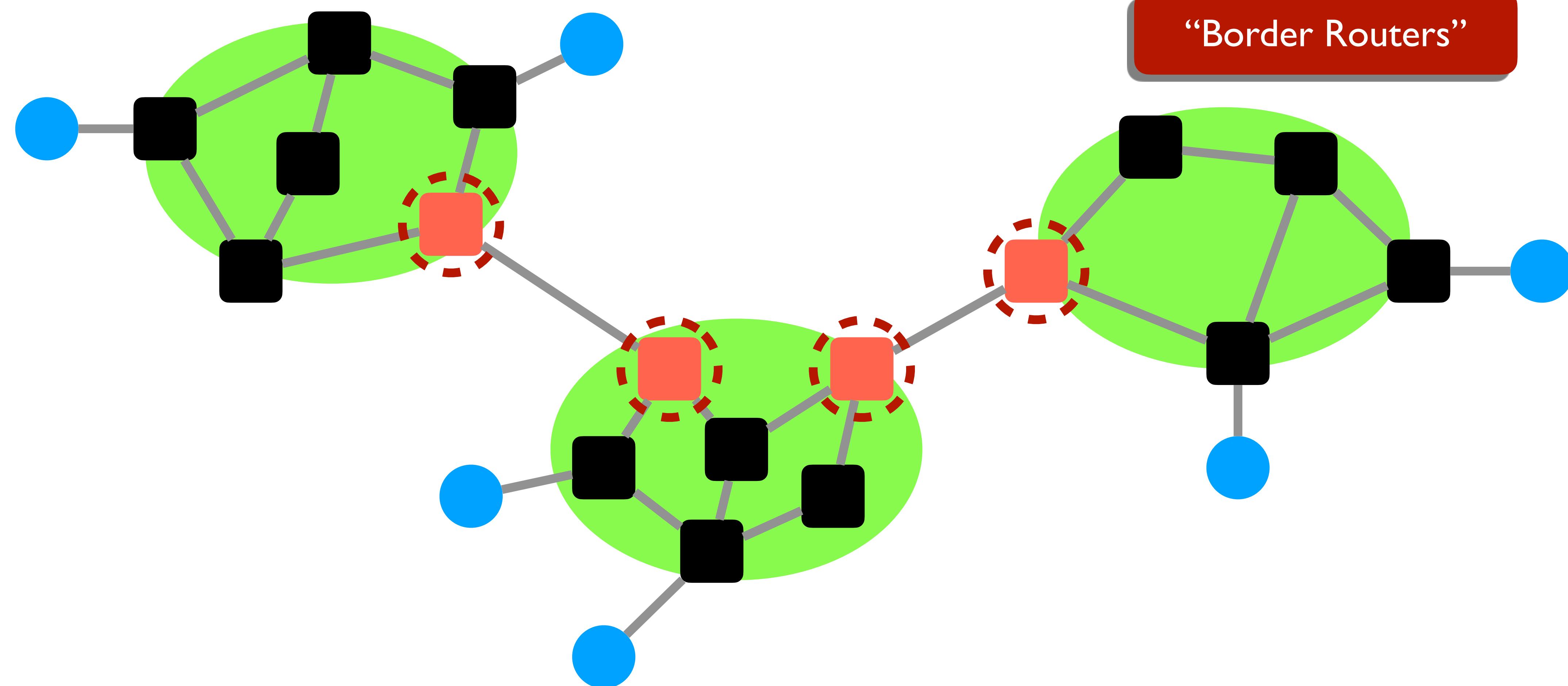
- Stay awake as long as you can...

- **Issues with BGP**

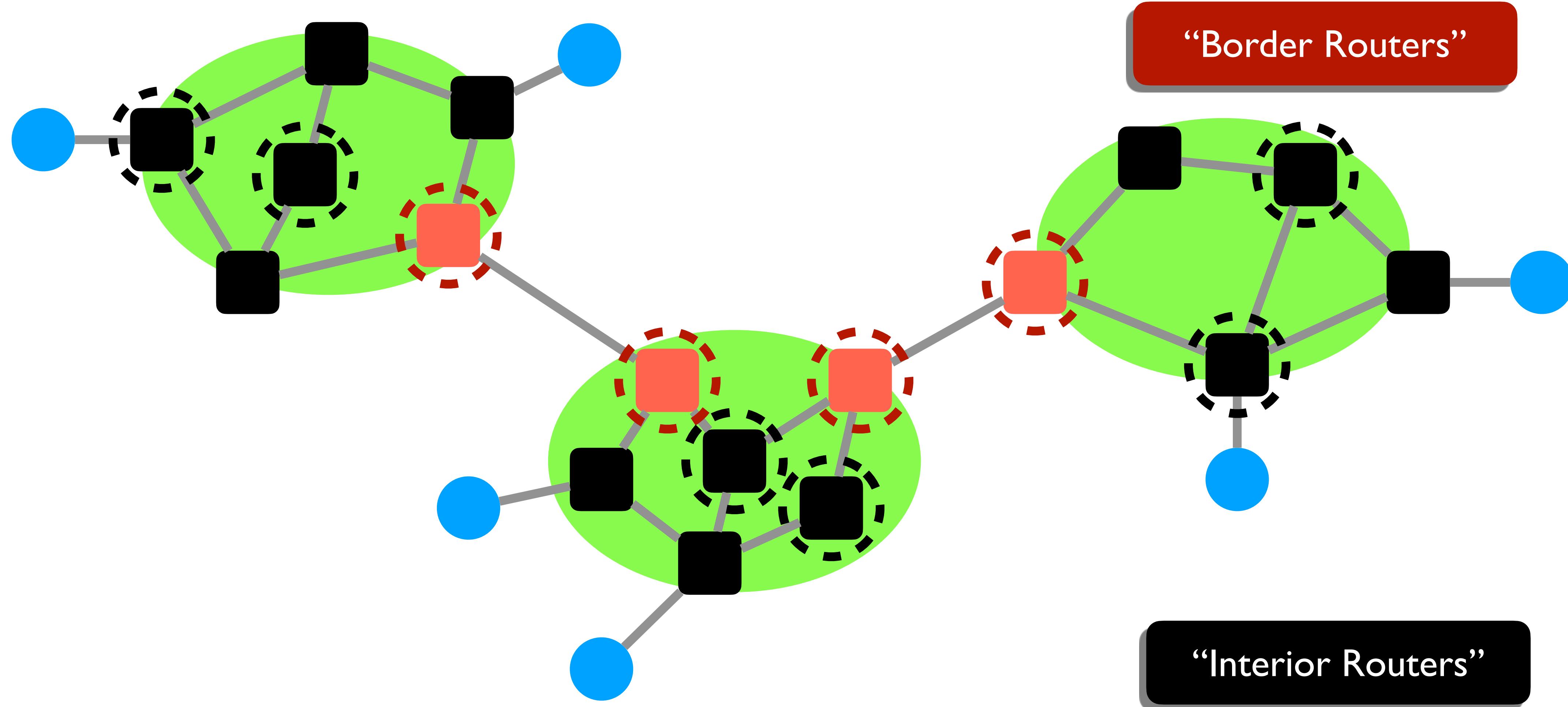
Who Speaks BGP?



Who Speaks BGP?



Who Speaks BGP?



What does “speak BGP” mean?

What does “speak BGP” mean?

- **Implement the BGP protocol standard**
 - Read more here: <http://tools.ietf.org/html/rfc4271>

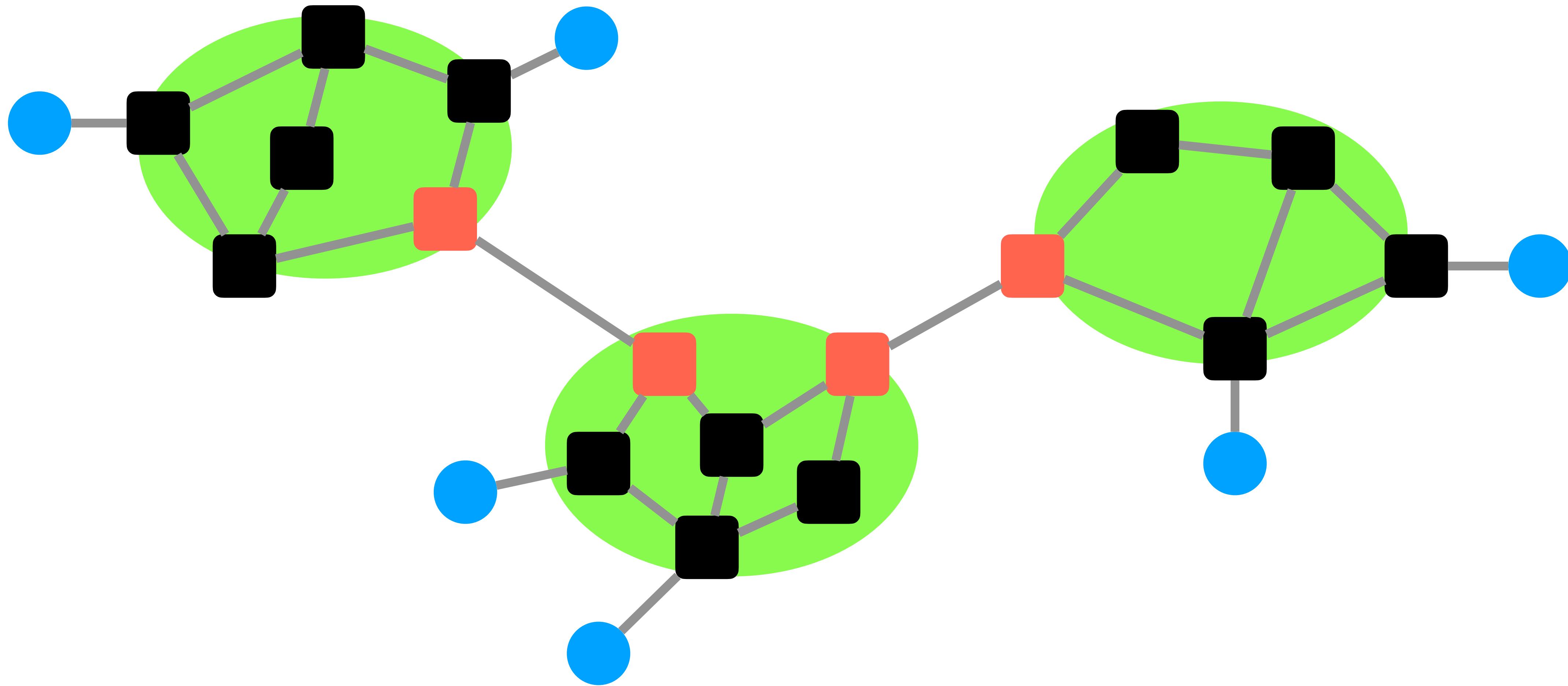
What does “speak BGP” mean?

- **Implement the BGP protocol standard**
 - Read more here: <http://tools.ietf.org/html/rfc4271>
- **Specifies what messages to exchange with other BGP “speakers”**
 - Message types (e.g., route advertisements, updates)
 - Message syntax

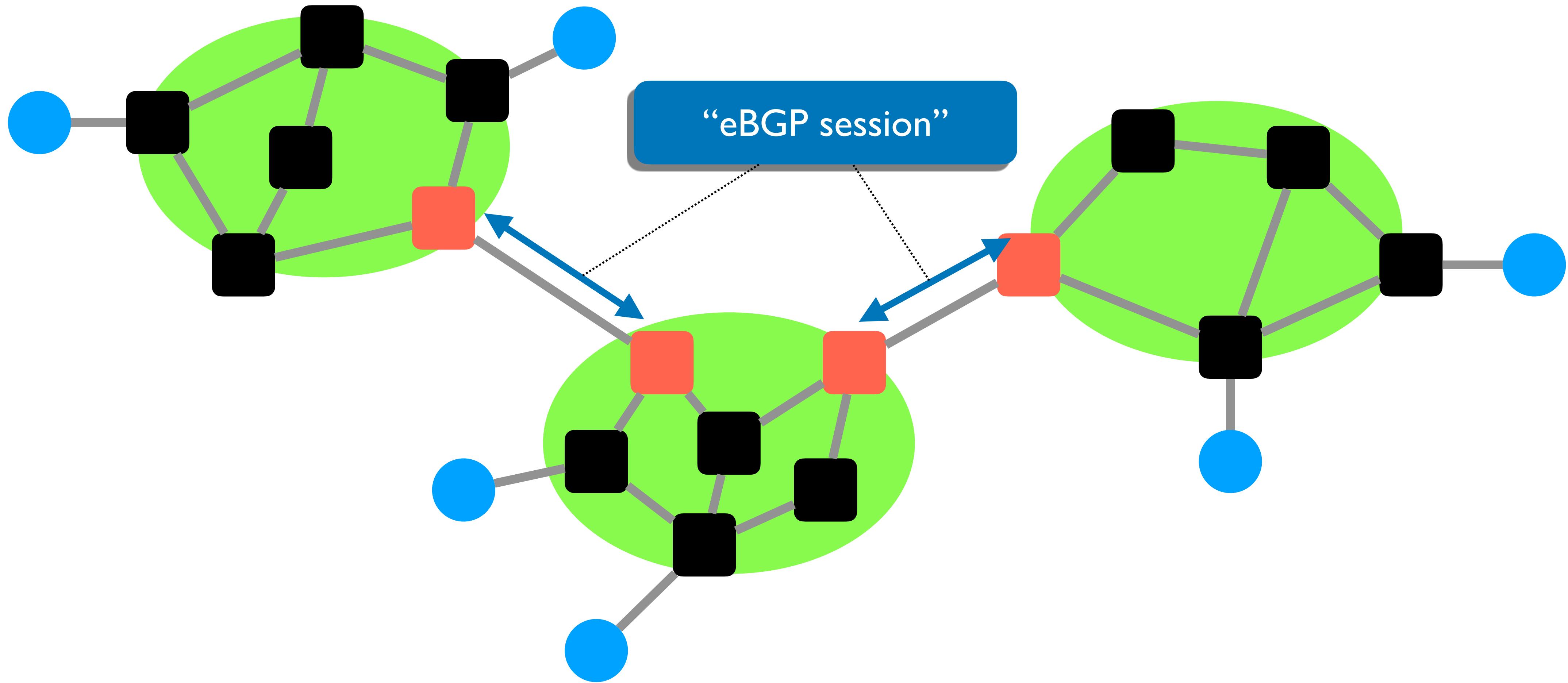
What does “speak BGP” mean?

- **Implement the BGP protocol standard**
 - Read more here: <http://tools.ietf.org/html/rfc4271>
- **Specifies what messages to exchange with other BGP “speakers”**
 - Message types (e.g., route advertisements, updates)
 - Message syntax
- **And how to process these messages**
 - e.g., “when you receive a BGP update, do...”
 - Follows BGP state machine in the protocol spec. + policy decisions, etc.

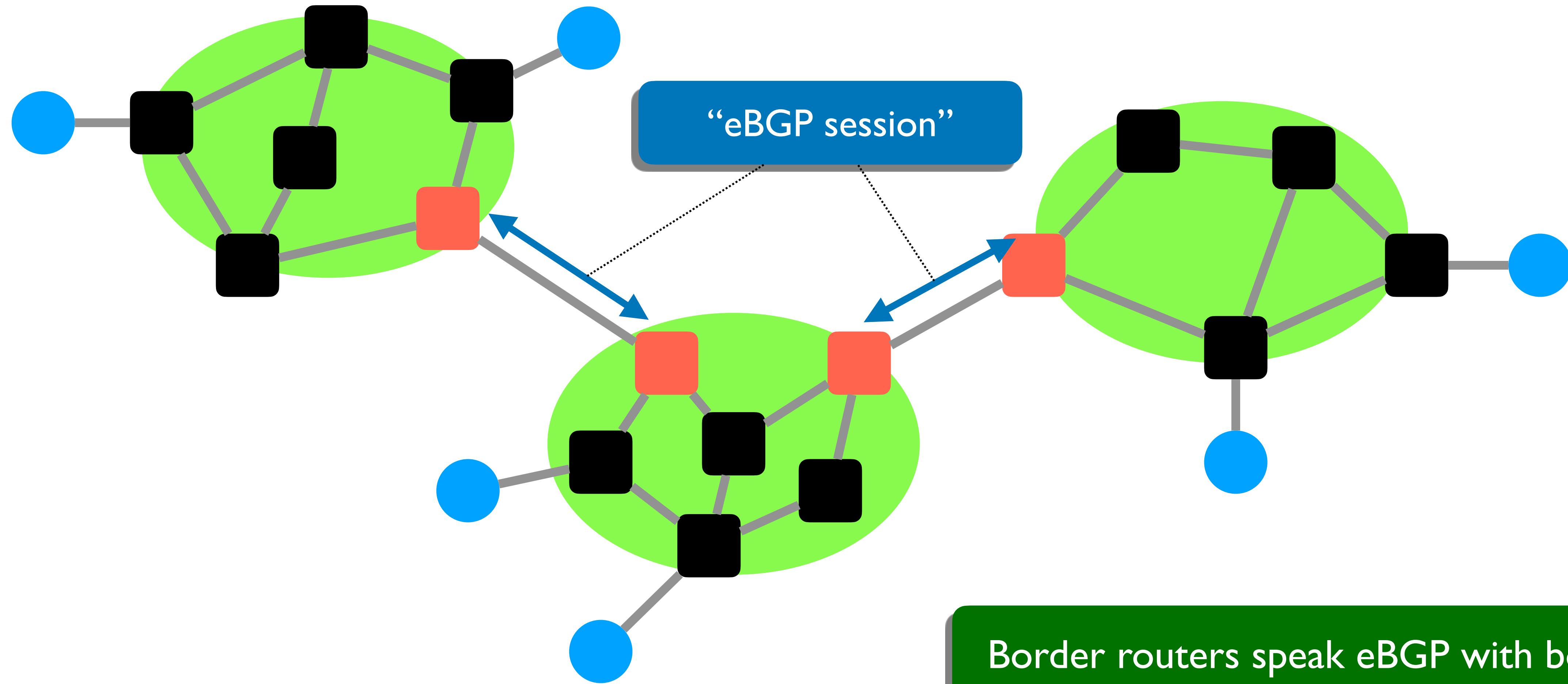
BGP “sessions”



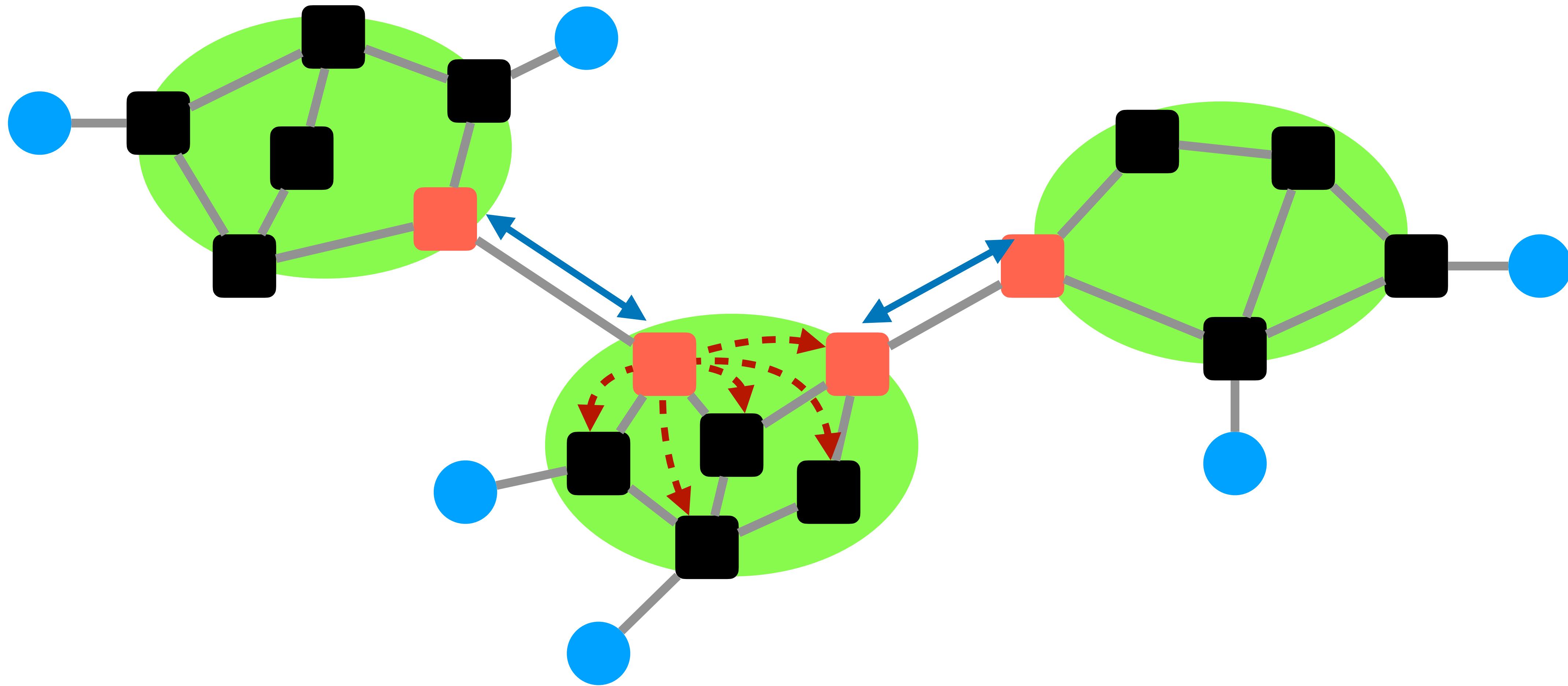
BGP “sessions”



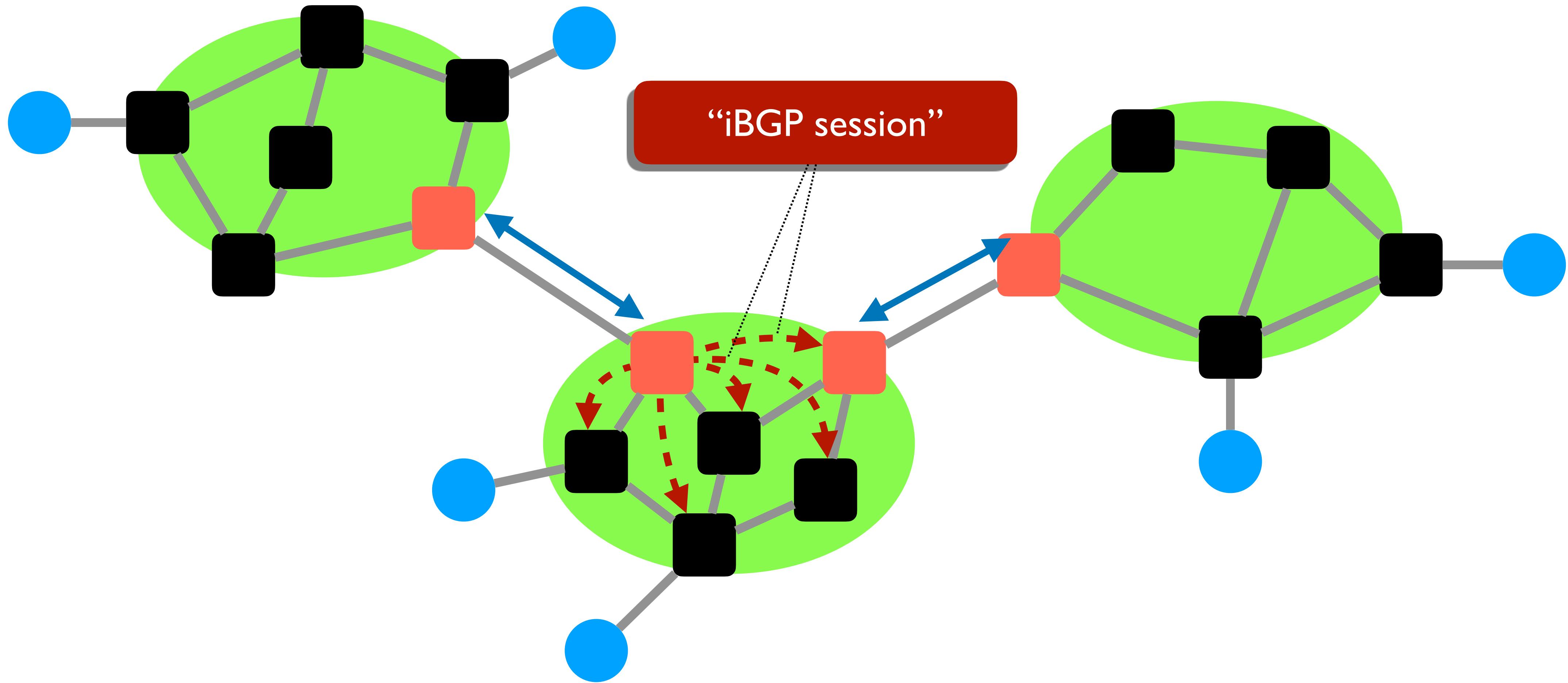
BGP “sessions”



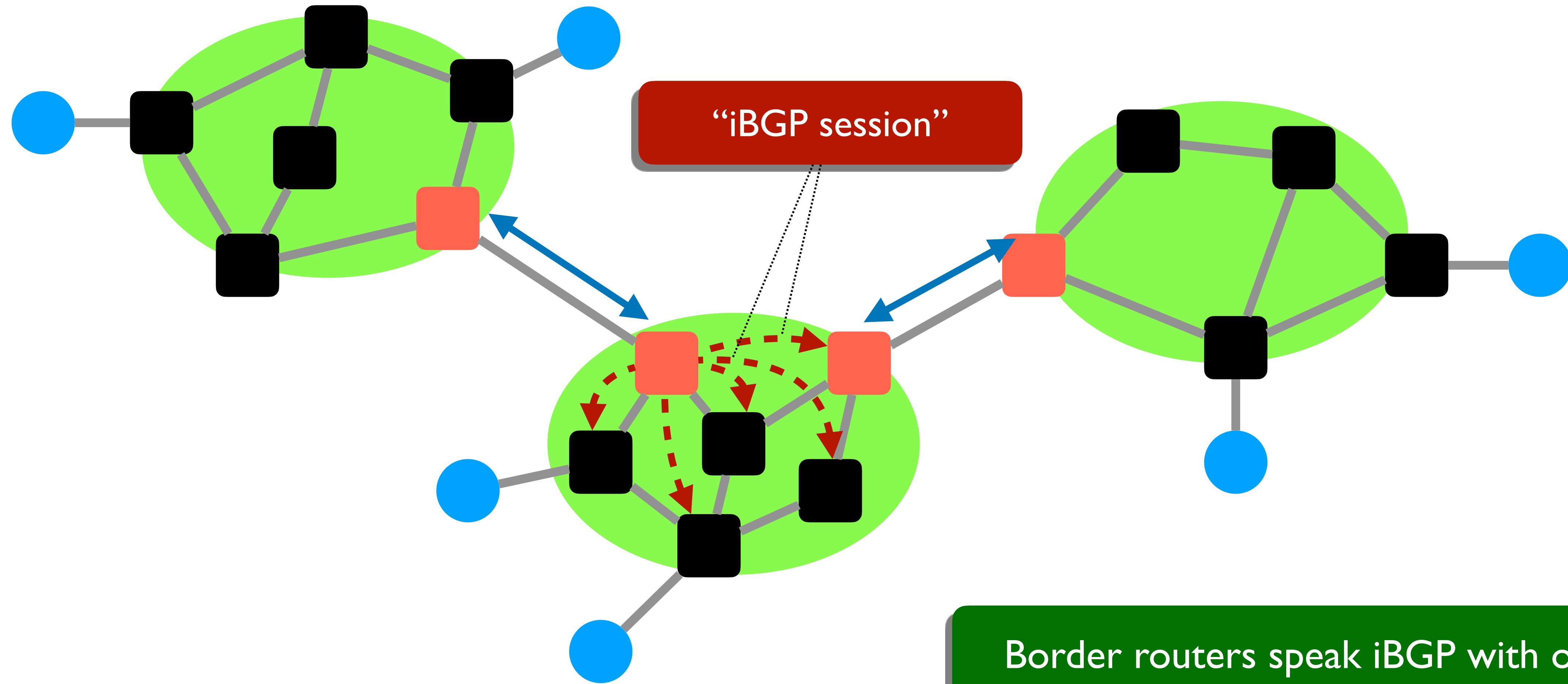
BGP “sessions”



BGP “sessions”



BGP “sessions”



Border routers speak iBGP with other (interior and border) routers in its own AS

eBGP, iBGP, IGP

eBGP, iBGP, IGP

- **eBGP: BGP sessions between border routers in different ASes**
 - Learn routes to external destinations

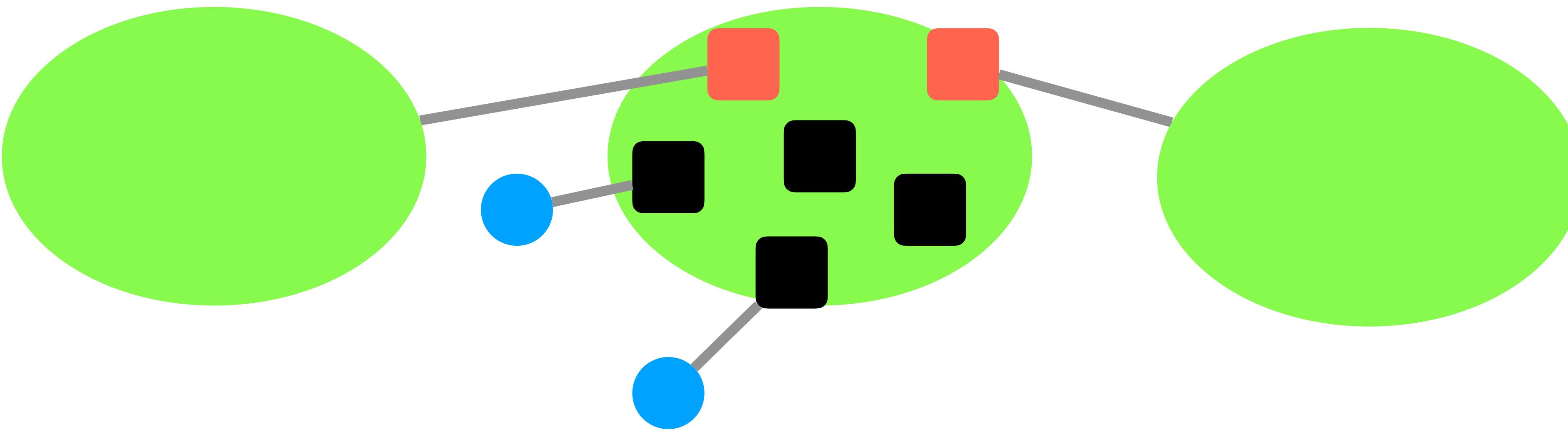
eBGP, iBGP, IGP

- **eBGP: BGP sessions between border routers in different ASes**
 - Learn routes to external destinations
- **iBGP: BGP sessions between border router and other routers within the same AS**
 - Distribute externally learned routes internally

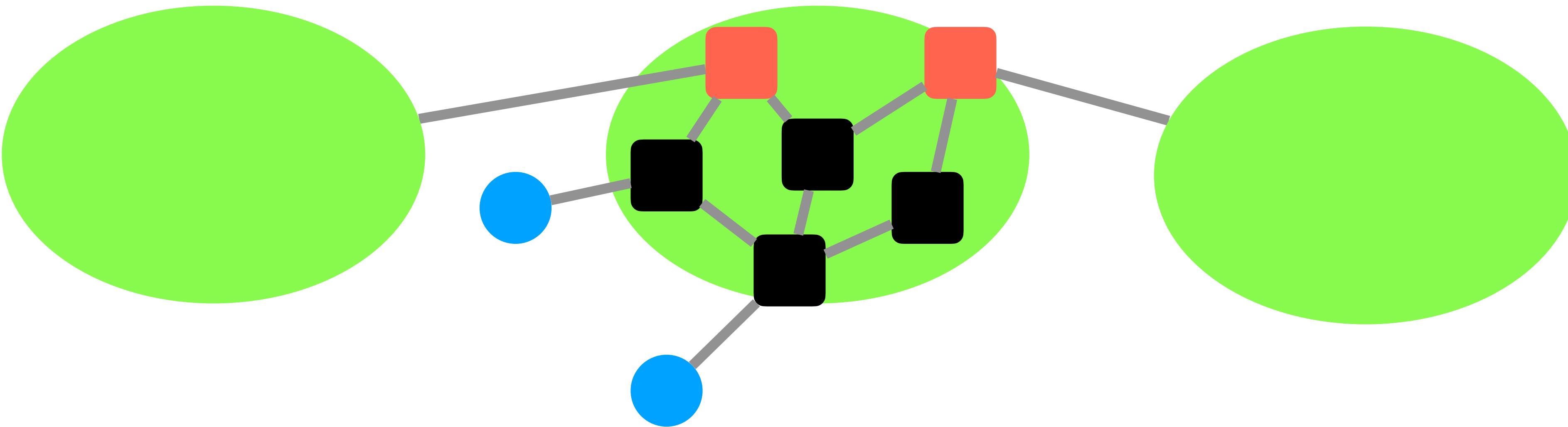
eBGP, iBGP, IGP

- **eBGP: BGP sessions between border routers in different ASes**
 - Learn routes to external destinations
- **iBGP: BGP sessions between border router and other routers within the same AS**
 - Distribute externally learned routes internally
- **IGP: “Interior Gateway Protocol” = Intradomain routing protocol**
 - Provide internal reachability
 - e.g., OSPF, RIP

BGP “sessions”

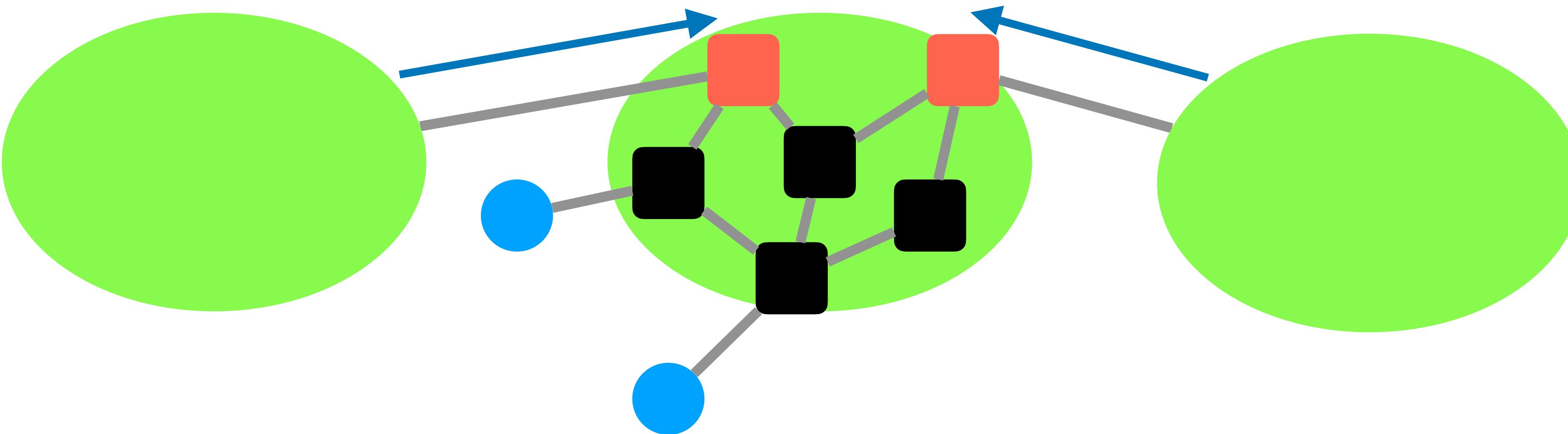


BGP “sessions”



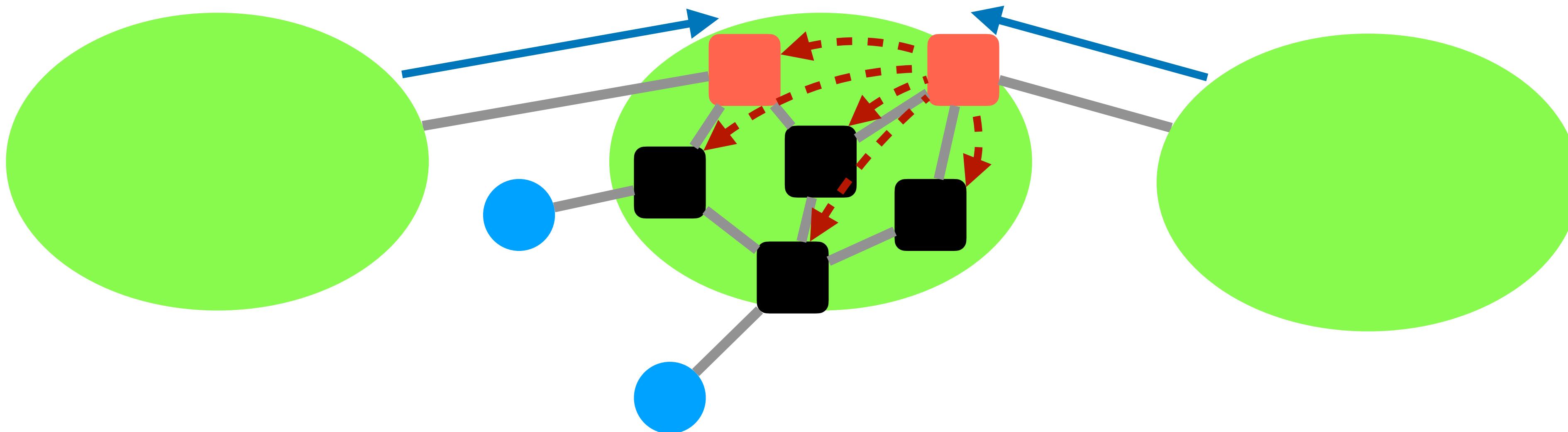
- I. Provide internal reachability (**IGP**) —

BGP “sessions”



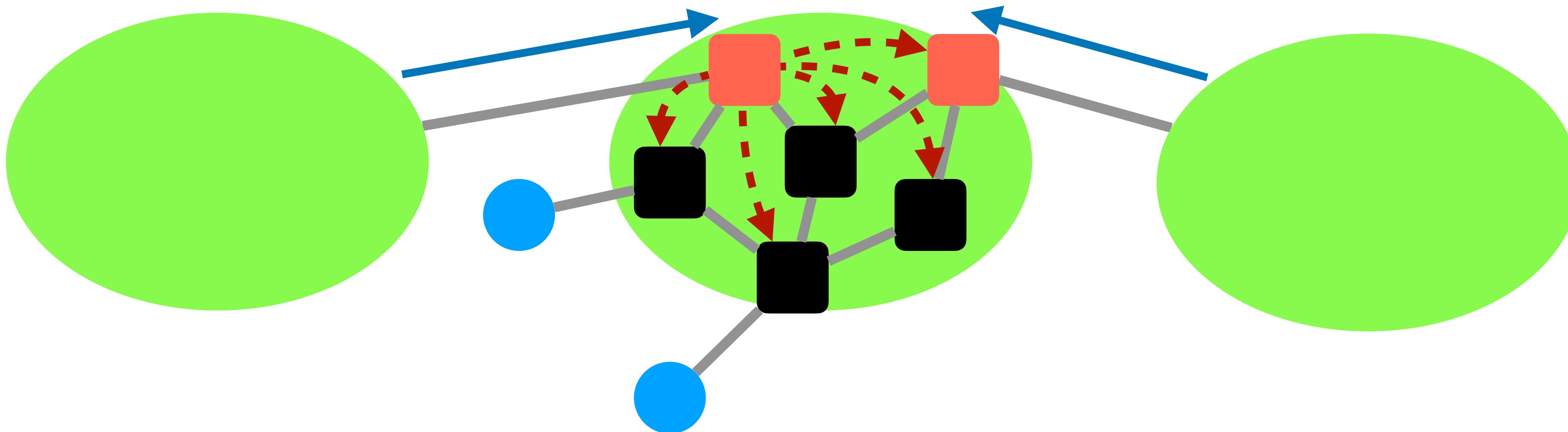
1. Provide internal reachability (**IGP**) —
2. Learn routes to external destinations (**eBGP**) →

BGP “sessions”



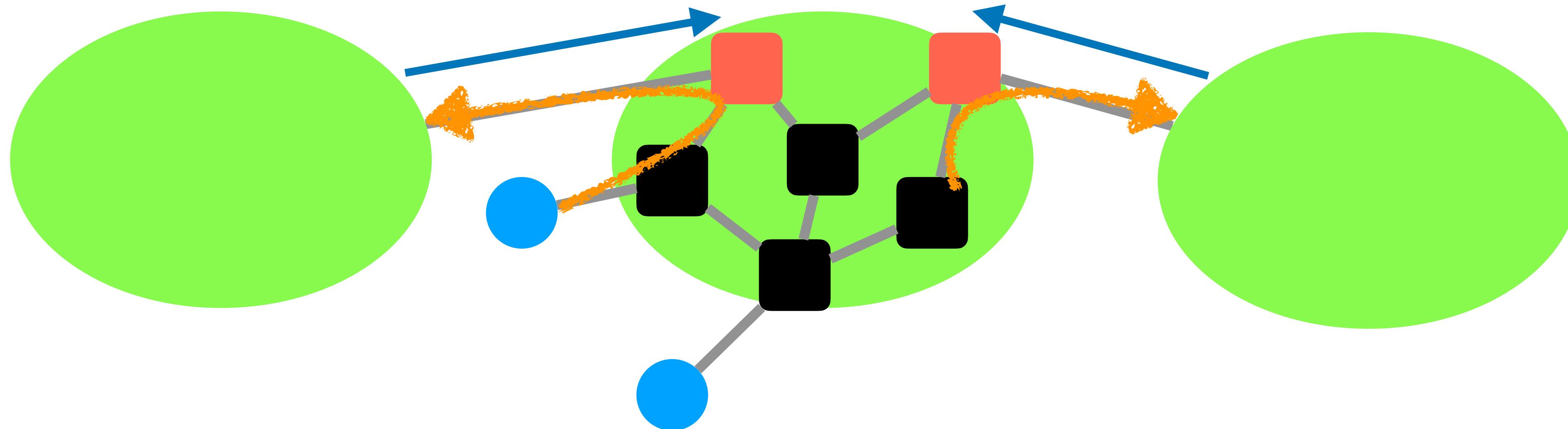
1. Provide internal reachability (**IGP**) —
2. Learn routes to external destinations (**eBGP**) →
3. Distribute externally learned routes internally (**iBGP**) ⤵

BGP “sessions”



1. Provide internal reachability (**IGP**) —
2. Learn routes to external destinations (**eBGP**) →
3. Distribute externally learned routes internally (**iBGP**) ⤵

BGP “sessions”



1. Provide internal reachability (**IGP**) —
2. Learn routes to external destinations (**eBGP**) →
3. Distribute externally learned routes internally (**iBGP**) ⤵ →
4. Travel shortest path to egress (**IGP**) →

Basic Messages in BGP

Basic Messages in BGP

- **Open**

- Establishes BGP session
- BGP uses TCP [*will make sense in 1-2 weeks*]

Basic Messages in BGP

- **Open**

- Establishes BGP session
- BGP uses TCP [*will make sense in 1-2 weeks*]

- **Notification**

- Report unusual conditions

Basic Messages in BGP

- **Open**

- Establishes BGP session
- BGP uses TCP [*will make sense in 1-2 weeks*]

- **Notification**

- Report unusual conditions

- **Update**

- Inform neighbor of new routes
- Inform neighbor of old routes that became inactive

Basic Messages in BGP

- **Open**

- Establishes BGP session
- BGP uses TCP [*will make sense in 1-2 weeks*]

- **Notification**

- Report unusual conditions

- **Update**

- Inform neighbor of new routes
- Inform neighbor of old routes that became inactive

- **Keepalive**

- Inform neighbor that connection is still viable

Route Updates

Route Updates

- **Format <IP Prefix: Route Attributes>**
 - Attributes describe properties of the route

Route Updates

- **Format <IP Prefix: Route Attributes>**

- Attributes describe properties of the route

- **Two kinds of updates**

- *Announcements*: new routes or changes to existing routes
- *Withdrawal*: remove routes that no longer exist

Route Attributes

Route Attributes

- **Routes are described using attributes**
 - Used in route selection/export decisions

Route Attributes

- **Routes are described using attributes**
 - Used in route selection/export decisions
- **Some attributes are local**
 - i.e., private within an AS, not included in announcements

Route Attributes

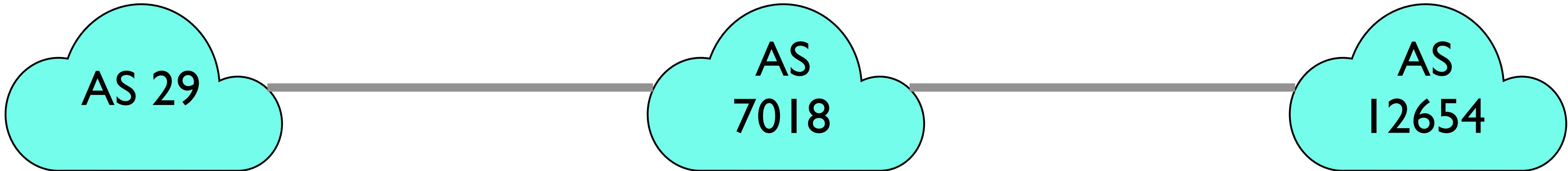
- **Routes are described using attributes**
 - Used in route selection/export decisions
- **Some attributes are local**
 - i.e., private within an AS, not included in announcements
- **Some attributes are propagated with eBGP route announcements**

Route Attributes

- **Routes are described using attributes**
 - Used in route selection/export decisions
- **Some attributes are local**
 - i.e., private within an AS, not included in announcements
- **Some attributes are propagated with eBGP route announcements**
- **There are many standardized attributes in BGP**
 - We will discuss a few

Attributes (I):ASPATH

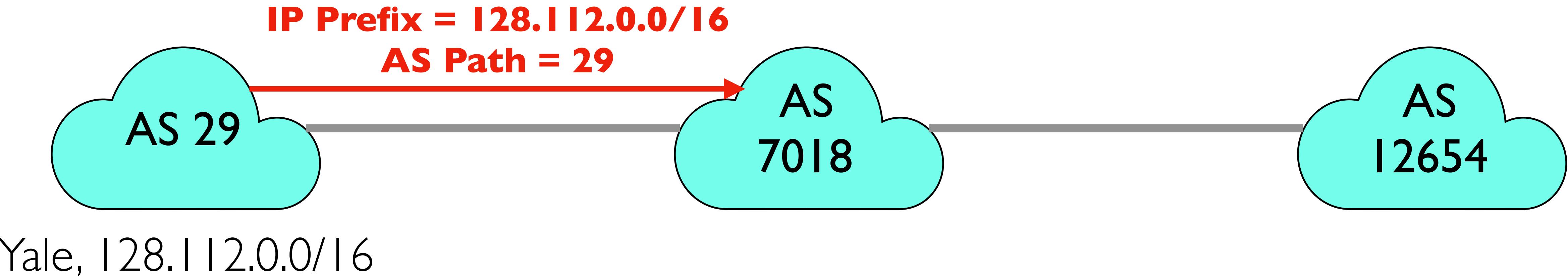
- Carried in route announcements
- Vector that lists all the ASes a route advertisement has traversed (in reverse order)



Yale, 128.112.0.0/16

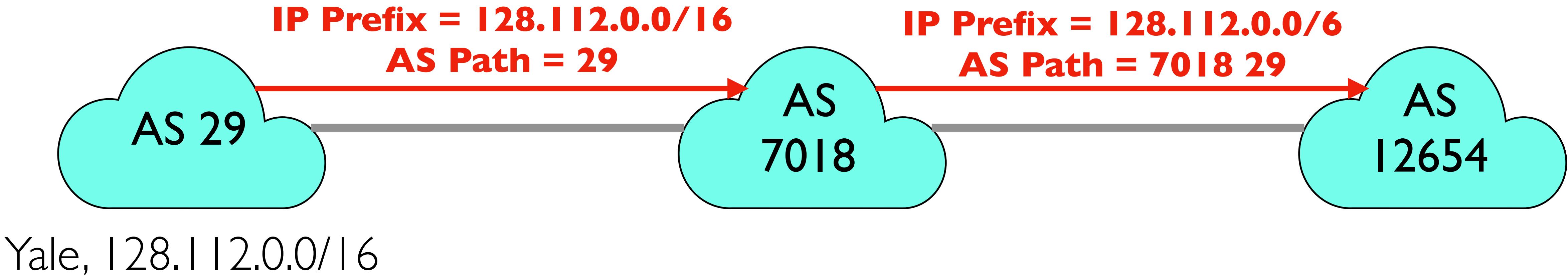
Attributes (I):ASPATH

- Carried in route announcements
- Vector that lists all the ASes a route advertisement has traversed (in reverse order)



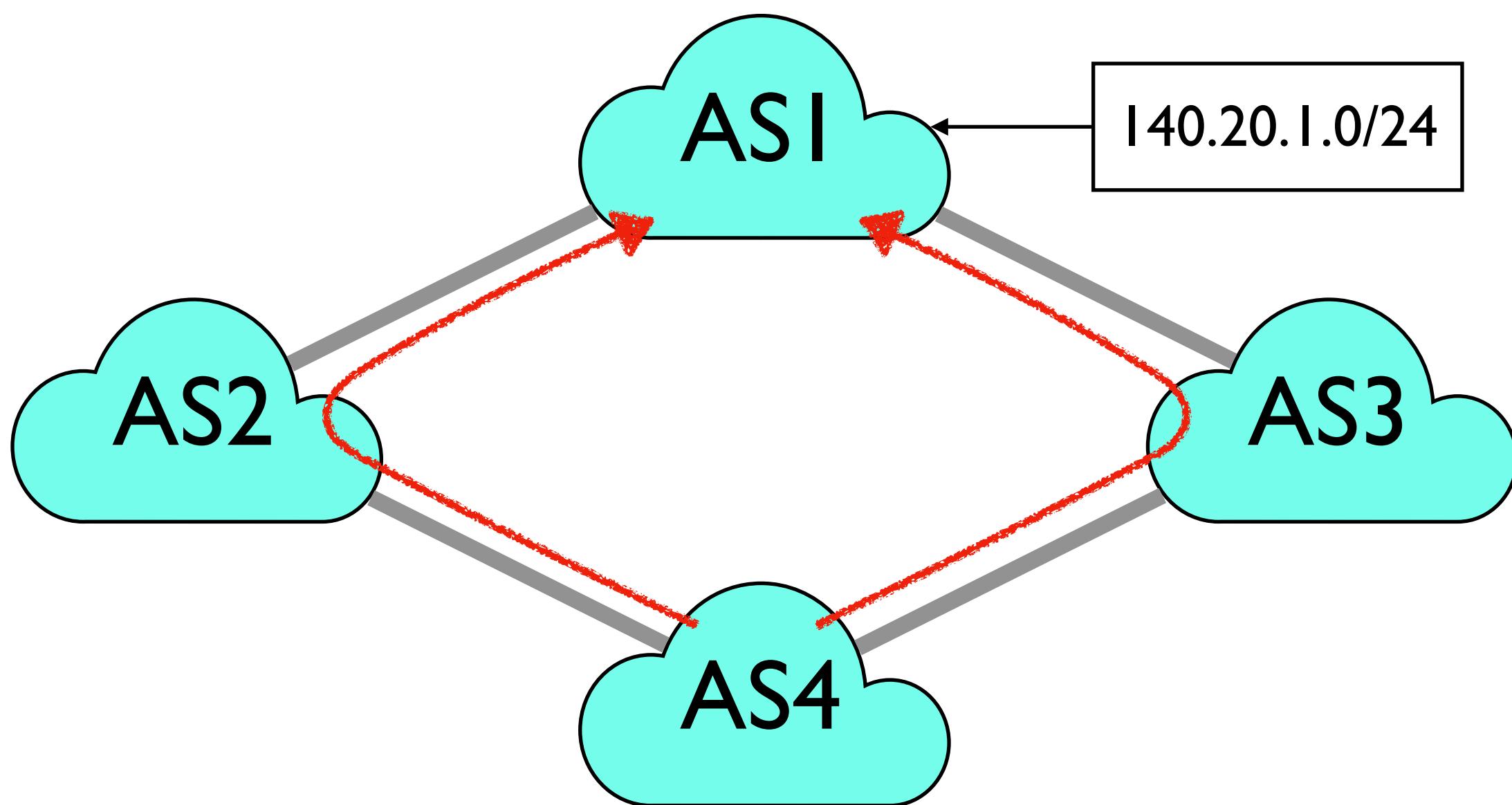
Attributes (I):ASPATH

- Carried in route announcements
- Vector that lists all the ASes a route advertisement has traversed (in reverse order)



Attributes (2): LOCAL PREF

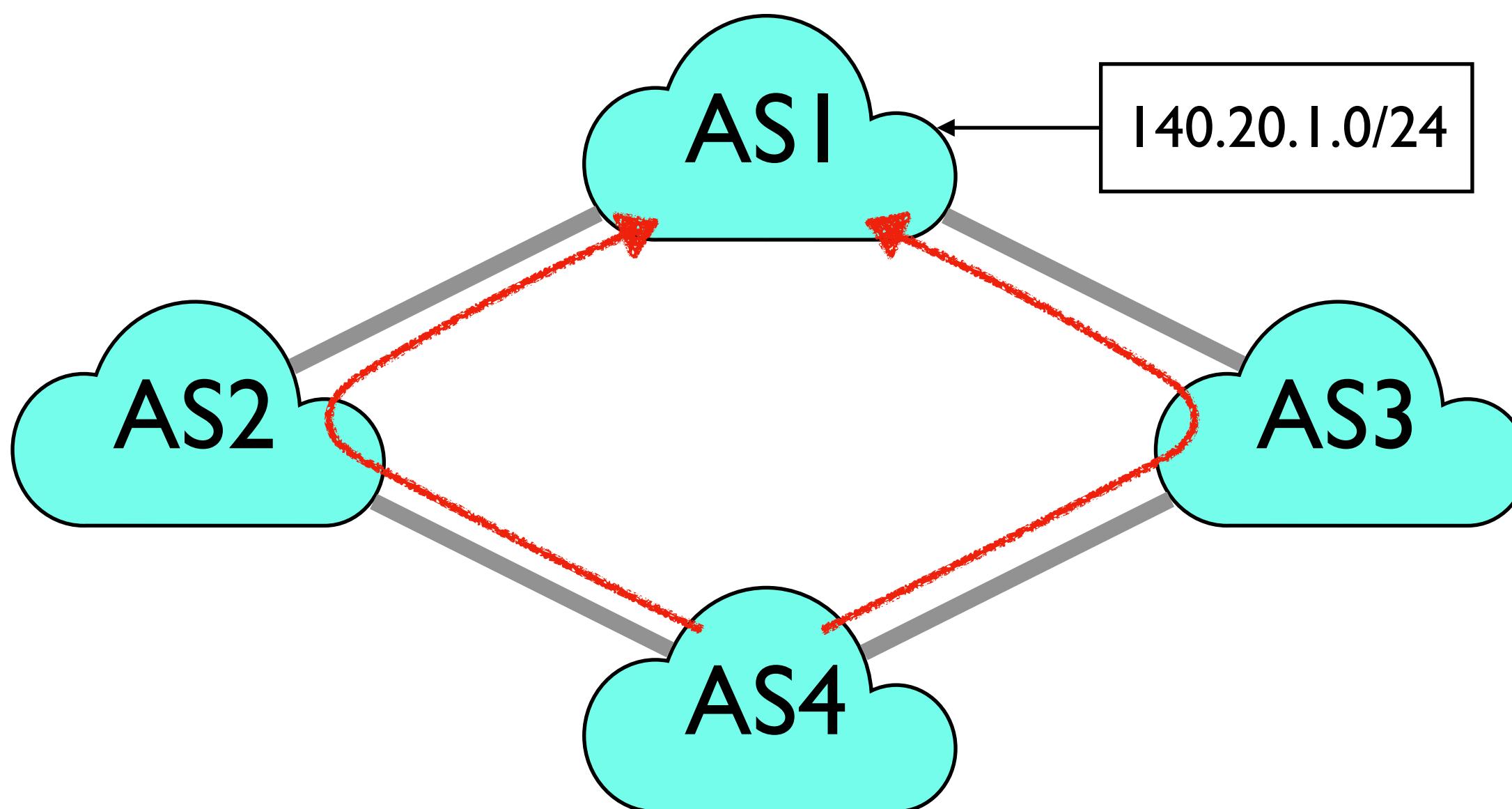
- “Local Preference”
- Used to choose between different AS paths
- The higher the value the more preferred



Destination	AS Path	Local Pref
140.20.1.0/24	AS3 AS1	300
140.20.1.0/24	AS2 AS1	100

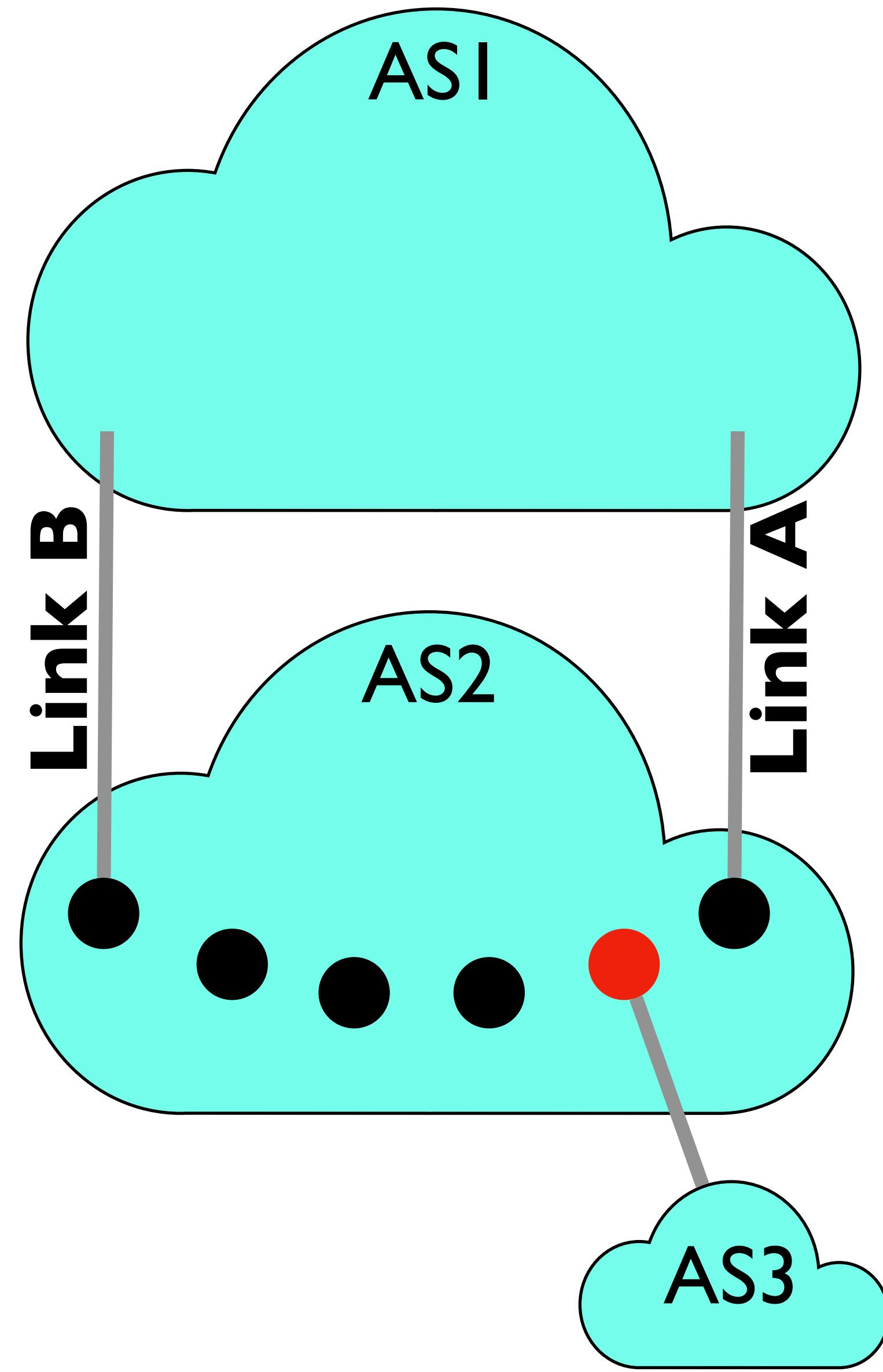
Attributes (2): LOCAL PREF

- “Local Preference”
- Used to choose between different AS paths
- The higher the value the more preferred
- Local to an AS; carried only in iBGP messages



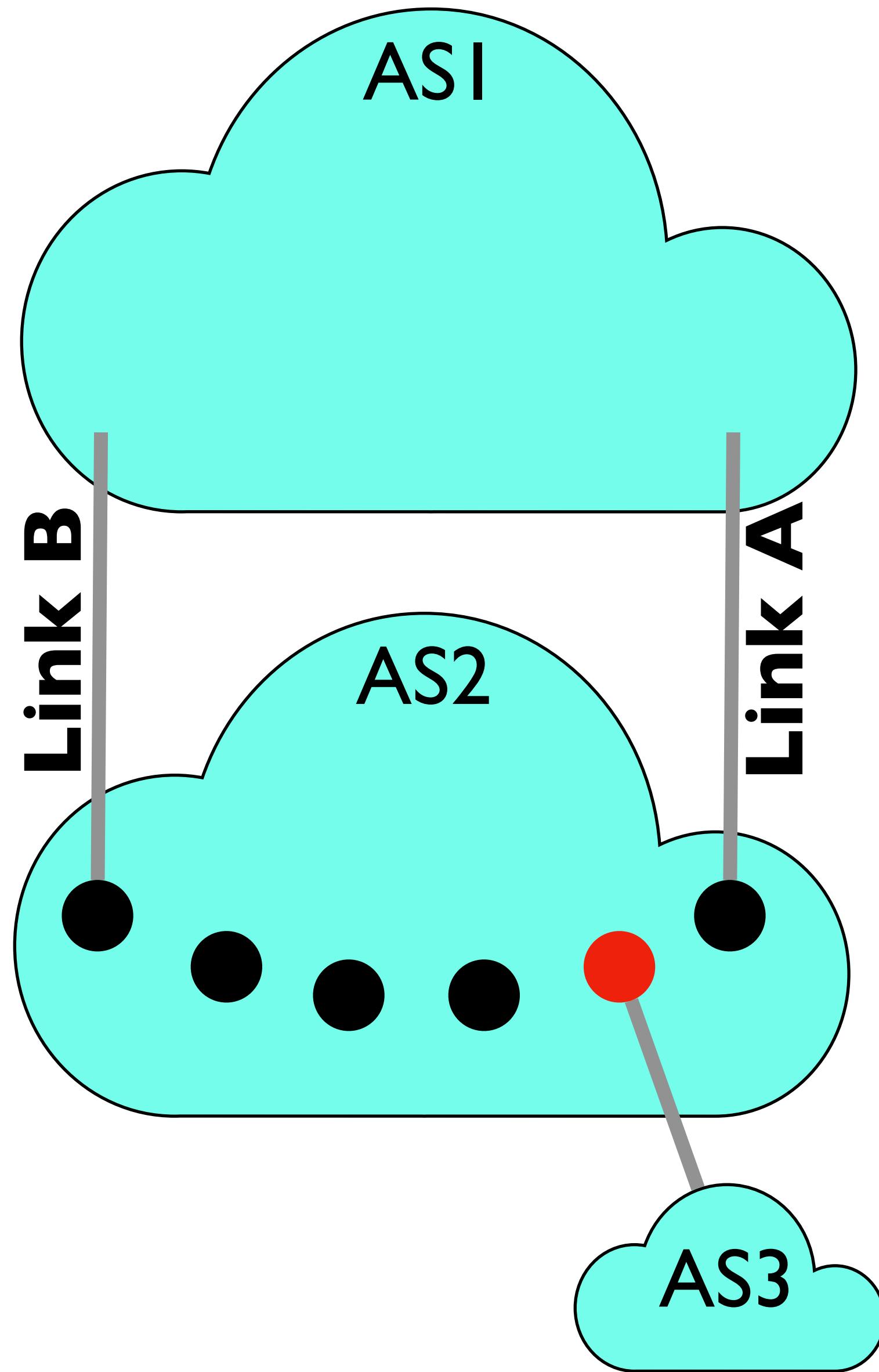
Destination	AS Path	Local Pref
140.20.1.0/24	AS3 AS1	300
140.20.1.0/24	AS2 AS1	100

Attributes (3): MED



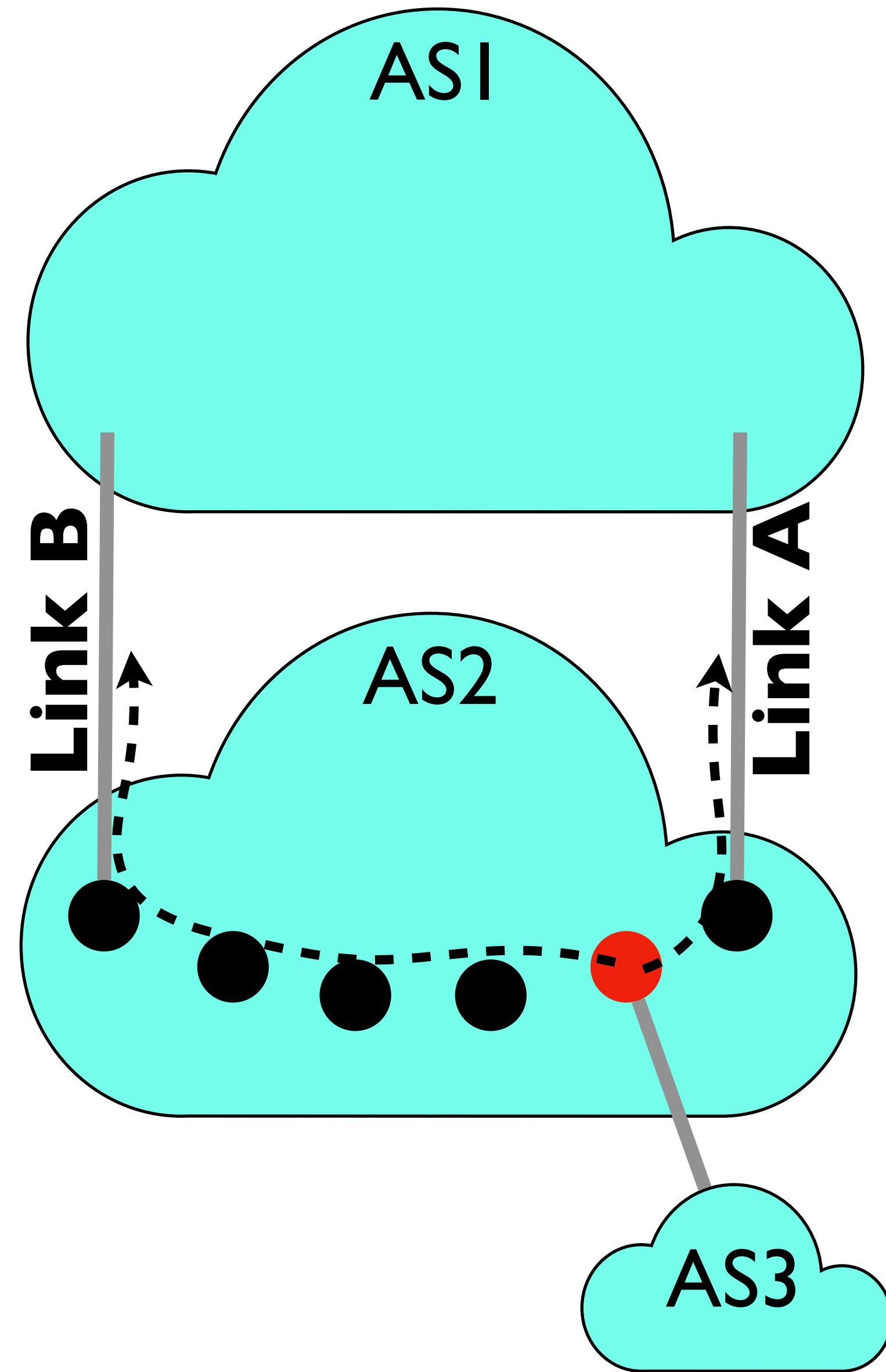
Attributes (3): MED

- “Multi-Exit Discriminator”



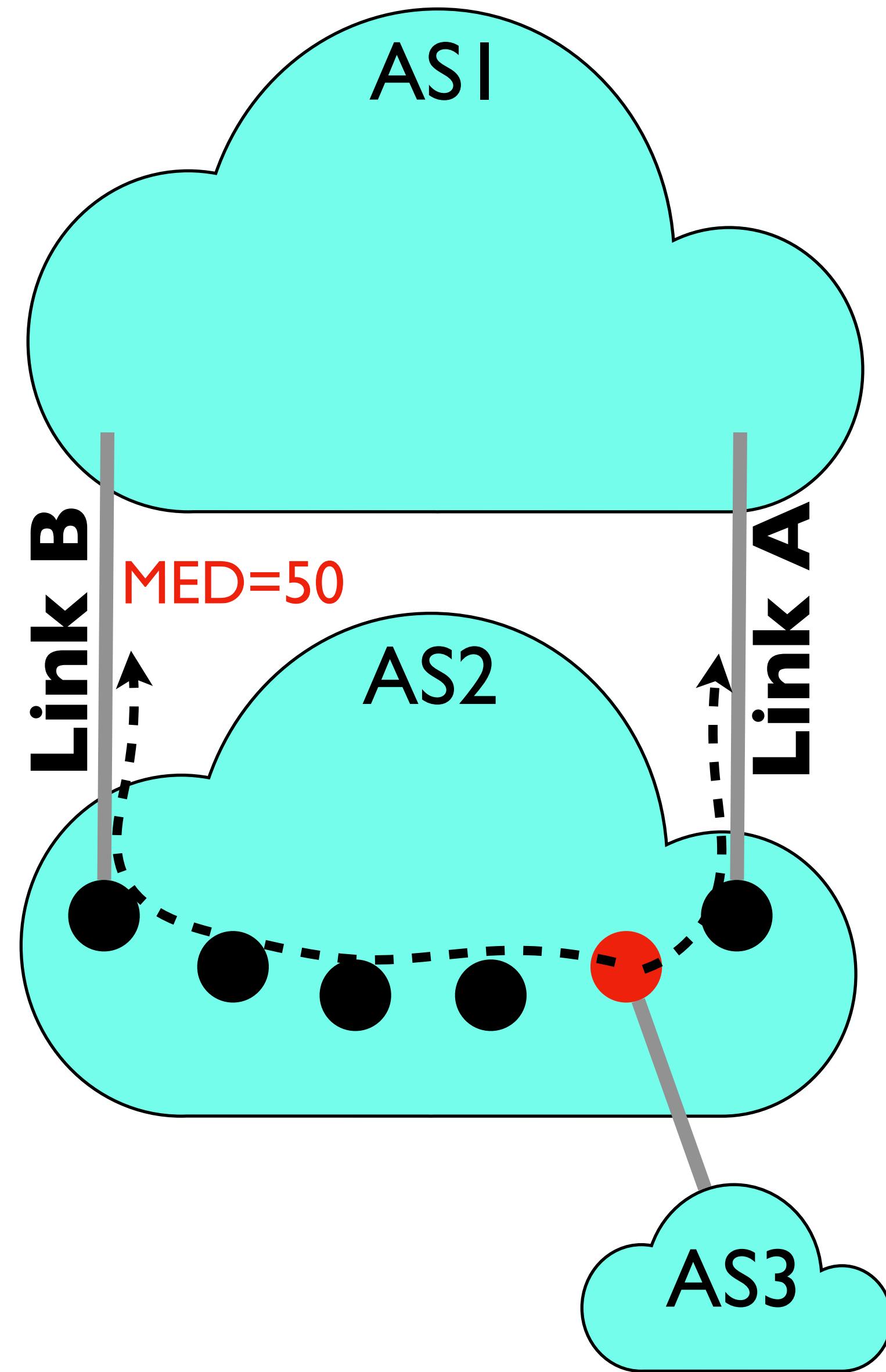
Attributes (3): MED

- “Multi-Exit Discriminator”



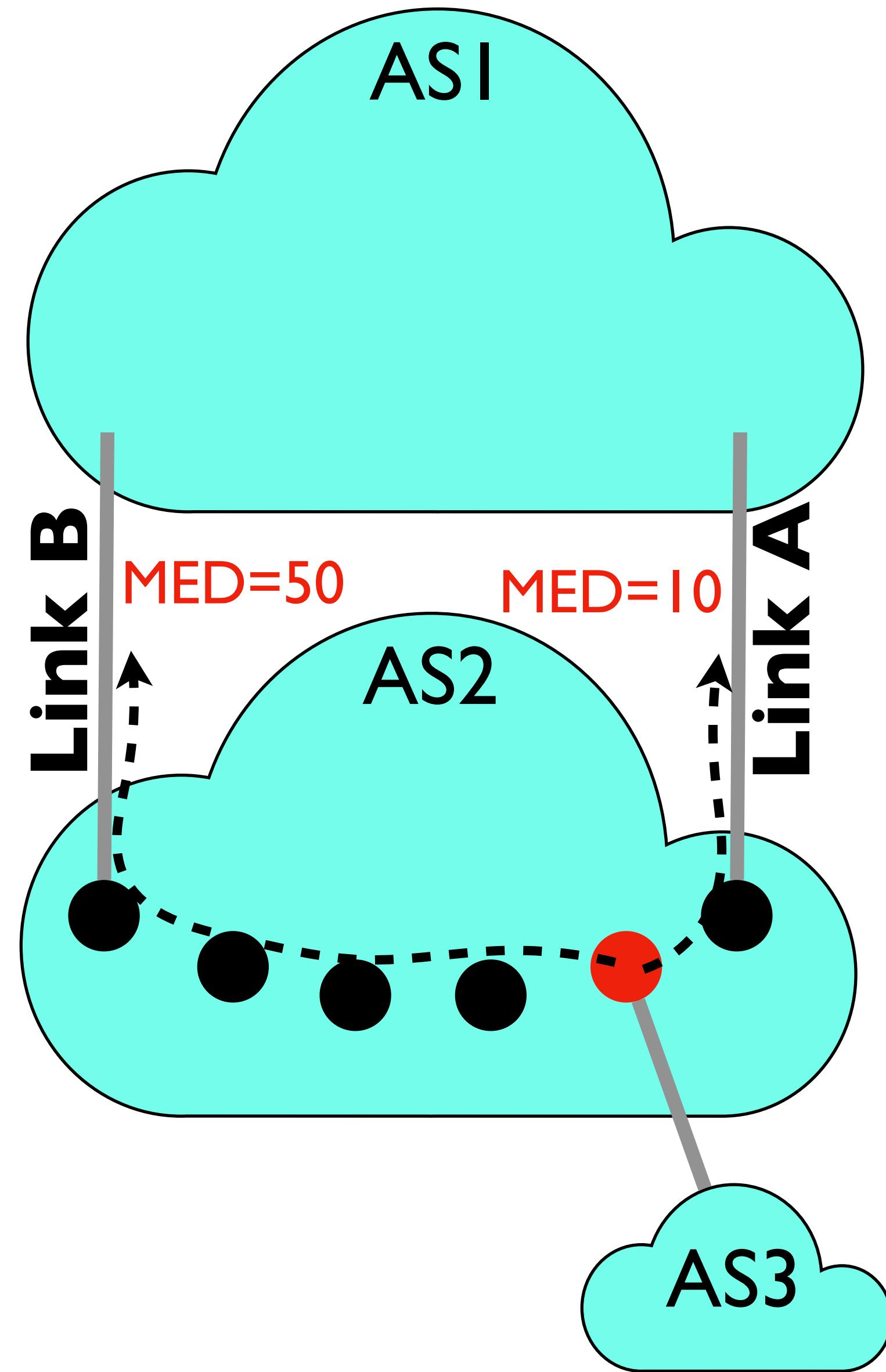
Attributes (3): MED

- “Multi-Exit Discriminator”



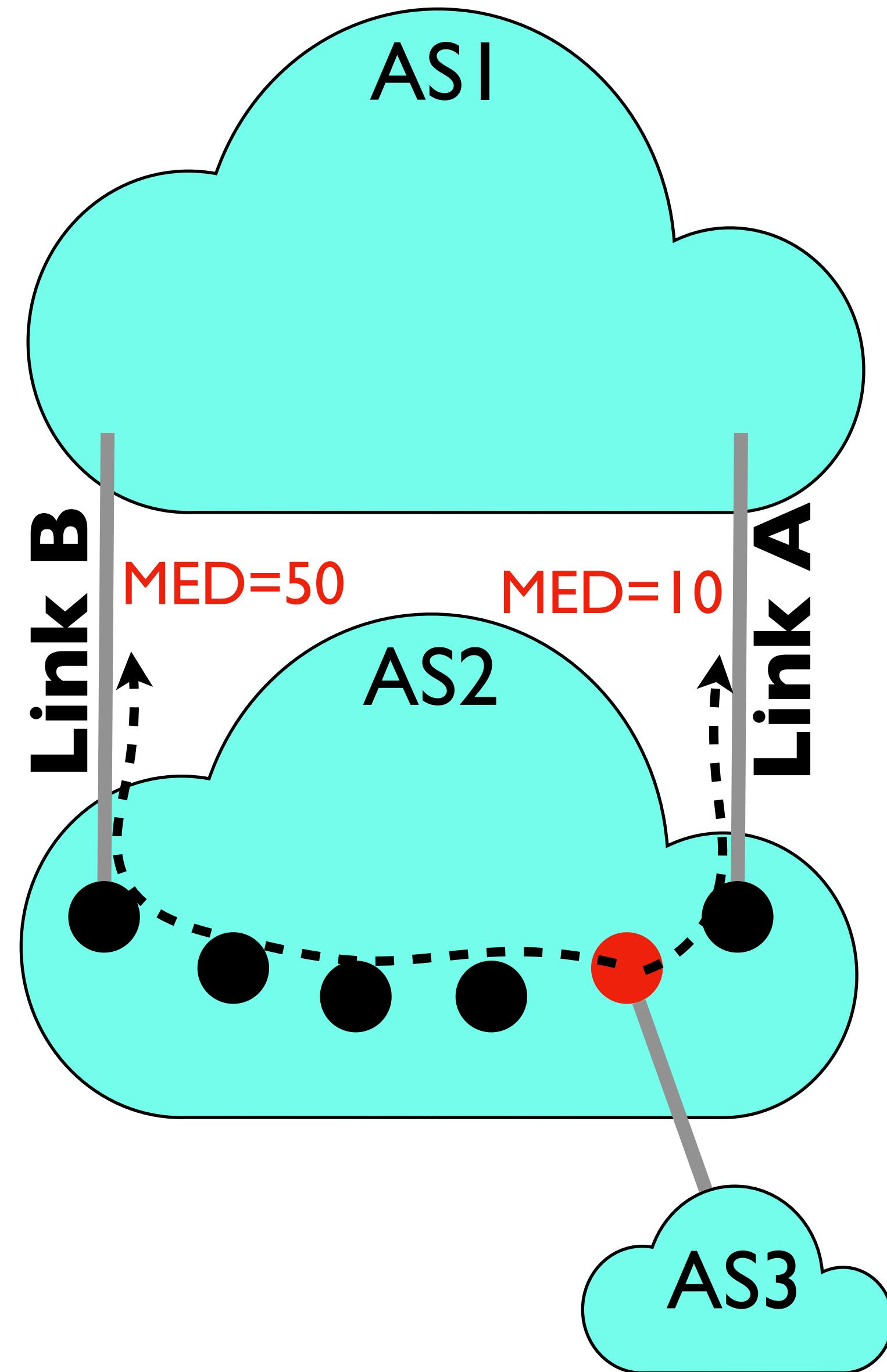
Attributes (3): MED

- “Multi-Exit Discriminator”



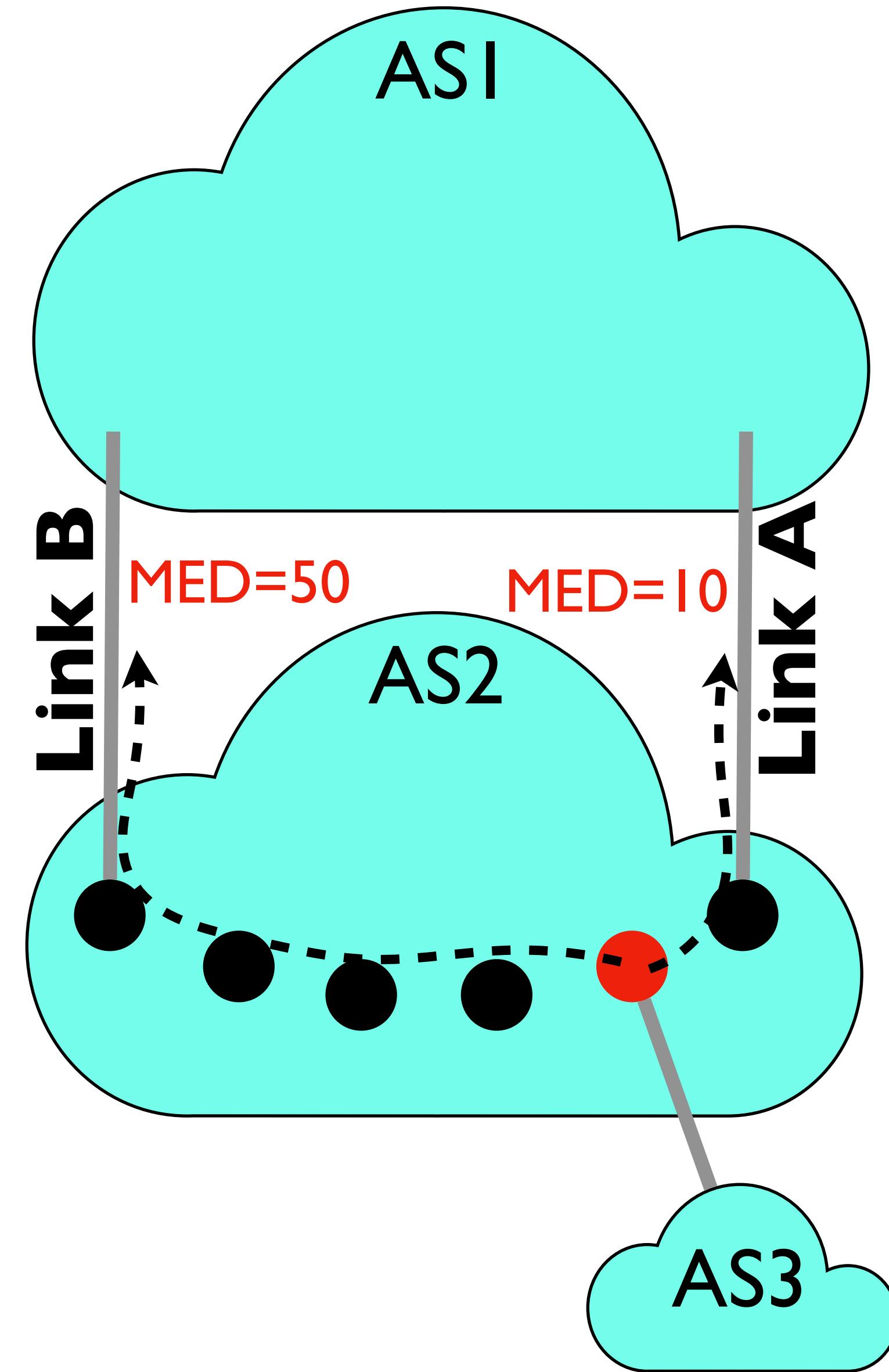
Attributes (3): MED

- “Multi-Exit Discriminator”
- Used when two ASes are interconnected via two or more links to specify how close a prefix is to the link it is announced on
- Lower is better



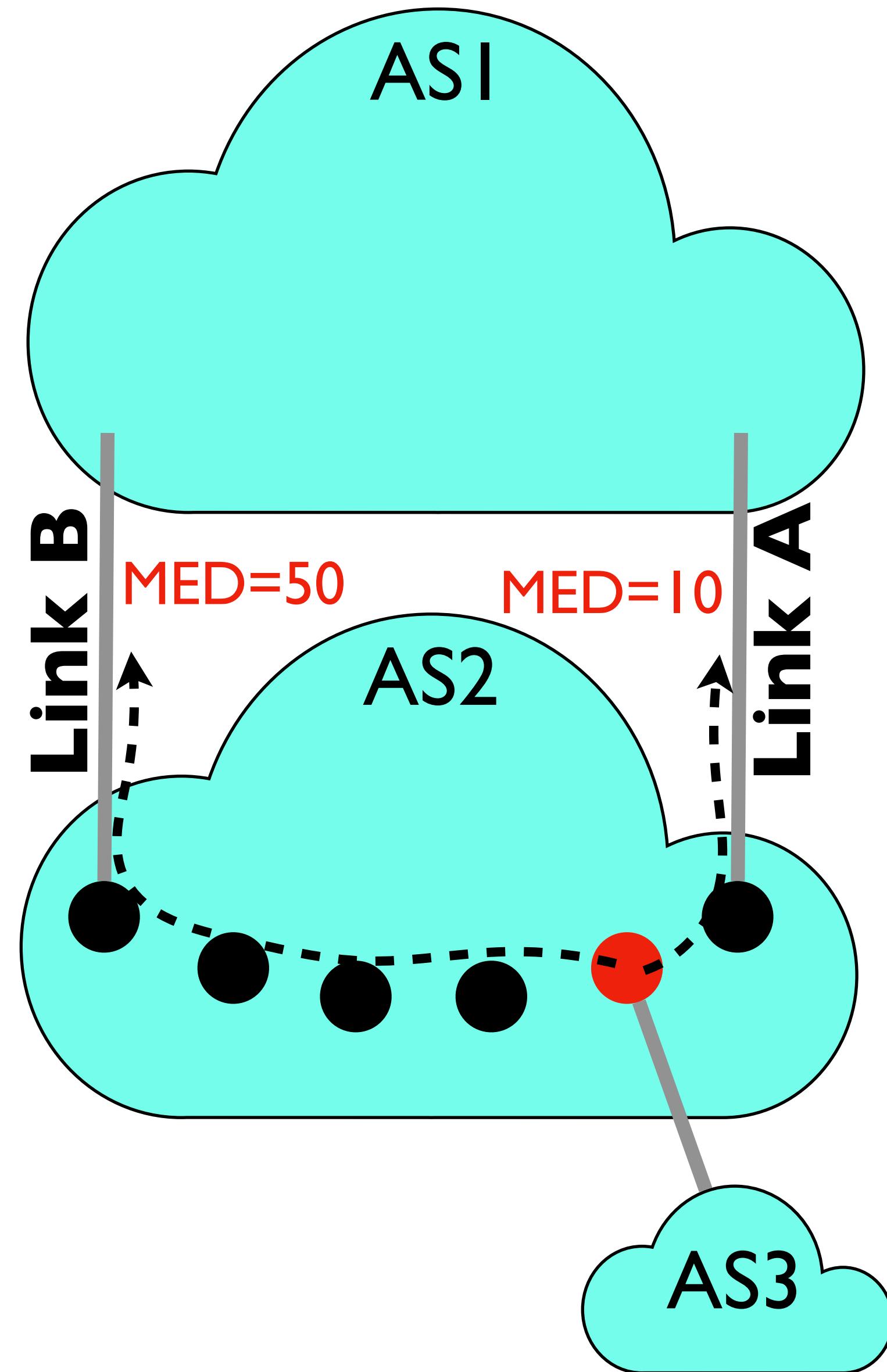
Attributes (3): MED

- “Multi-Exit Discriminator”
- Used when two ASes are interconnected via two or more links to specify how close a prefix is to the link it is announced on
- Lower is better
- AS announcing prefix sets MED



Attributes (3): MED

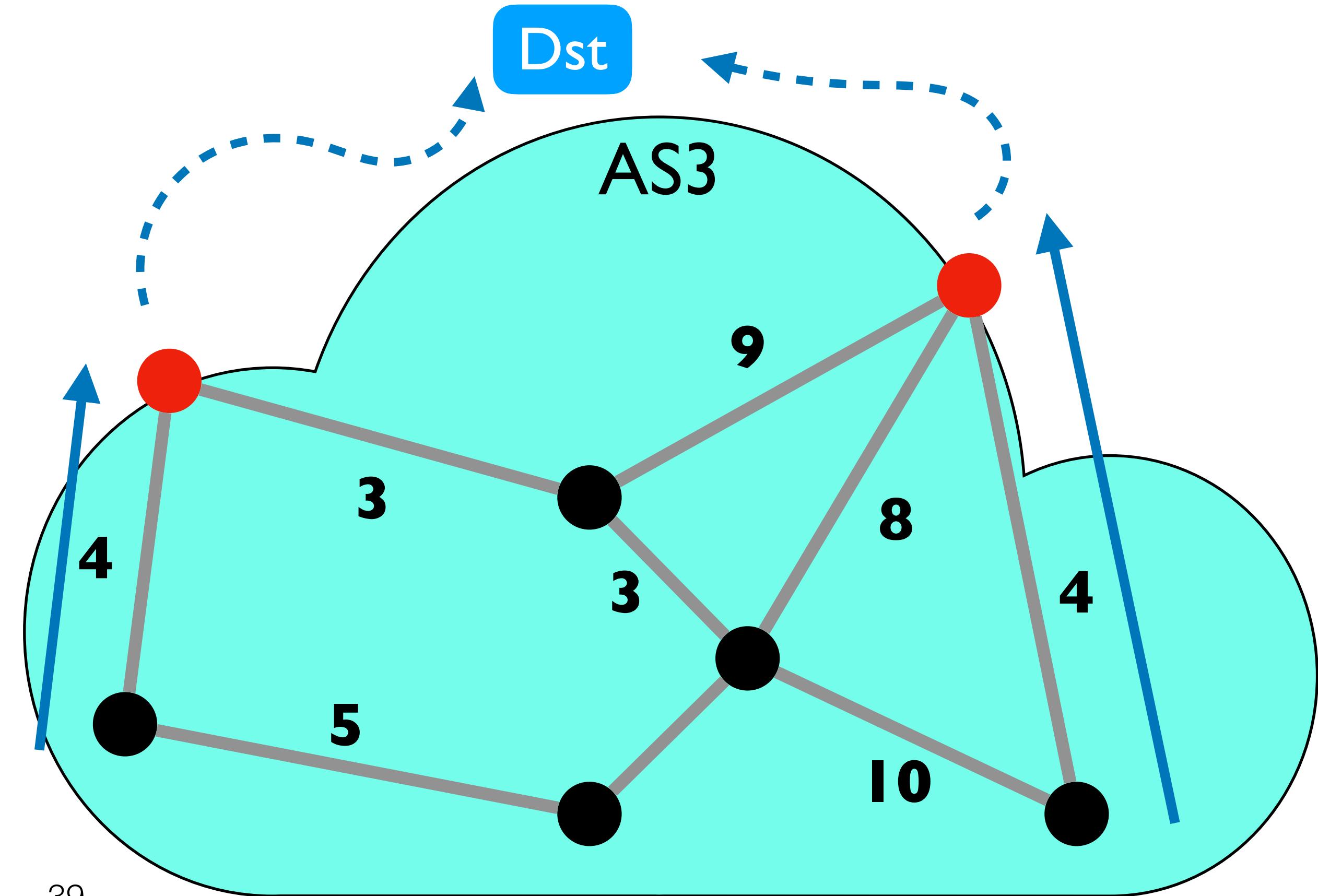
- “Multi-Exit Discriminator”
- Used when two ASes are interconnected via two or more links to specify how close a prefix is to the link it is announced on
- Lower is better
- AS announcing prefix sets MED
- AS receiving prefix (optionally!) uses MED to select link



Attributes (4): IGP Cost

- **Used for hot potato routing**

- Each router selects the closest egress point based on the path cost in intradomain protocol



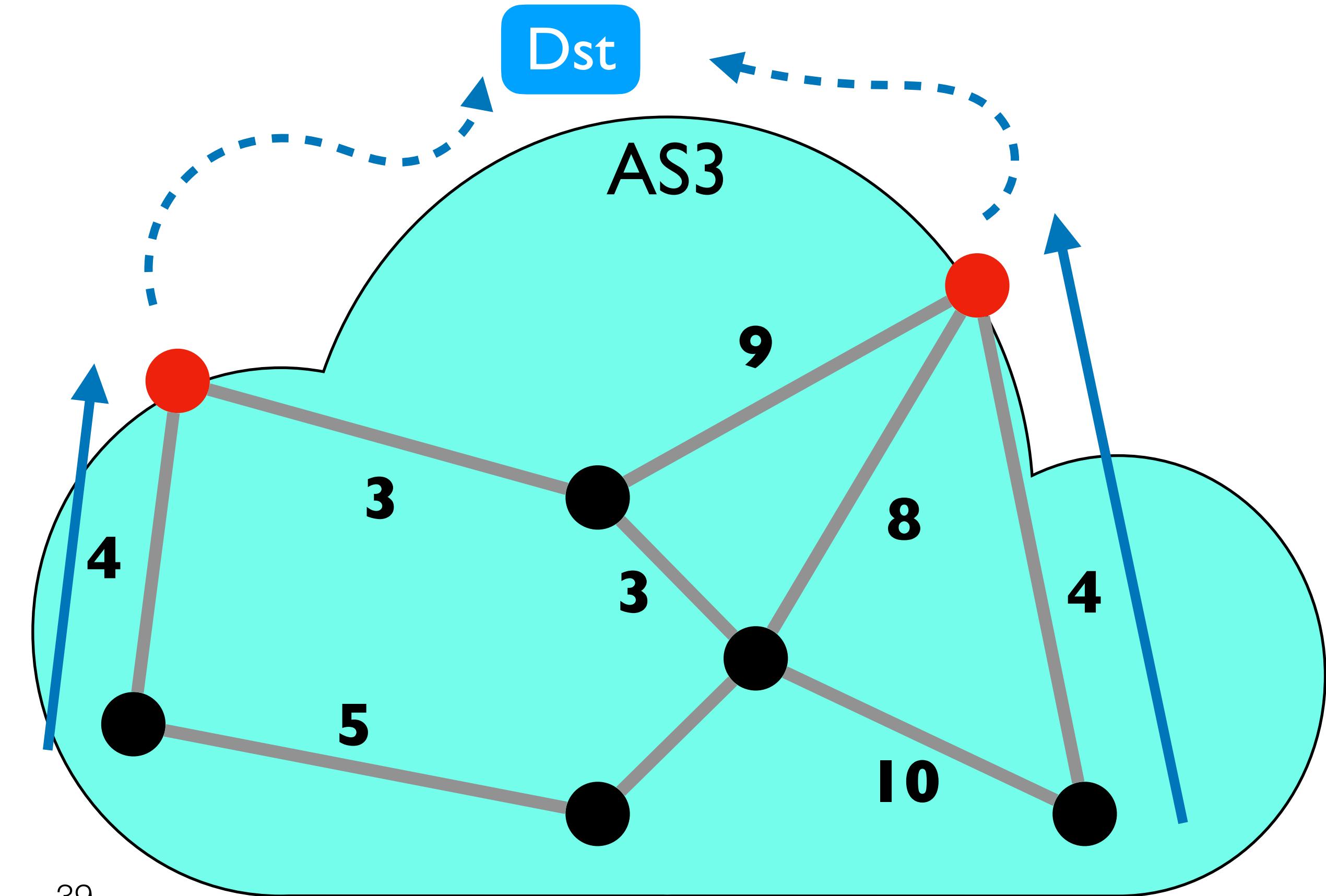
Attributes (4): IGP Cost

- **Used for hot potato routing**

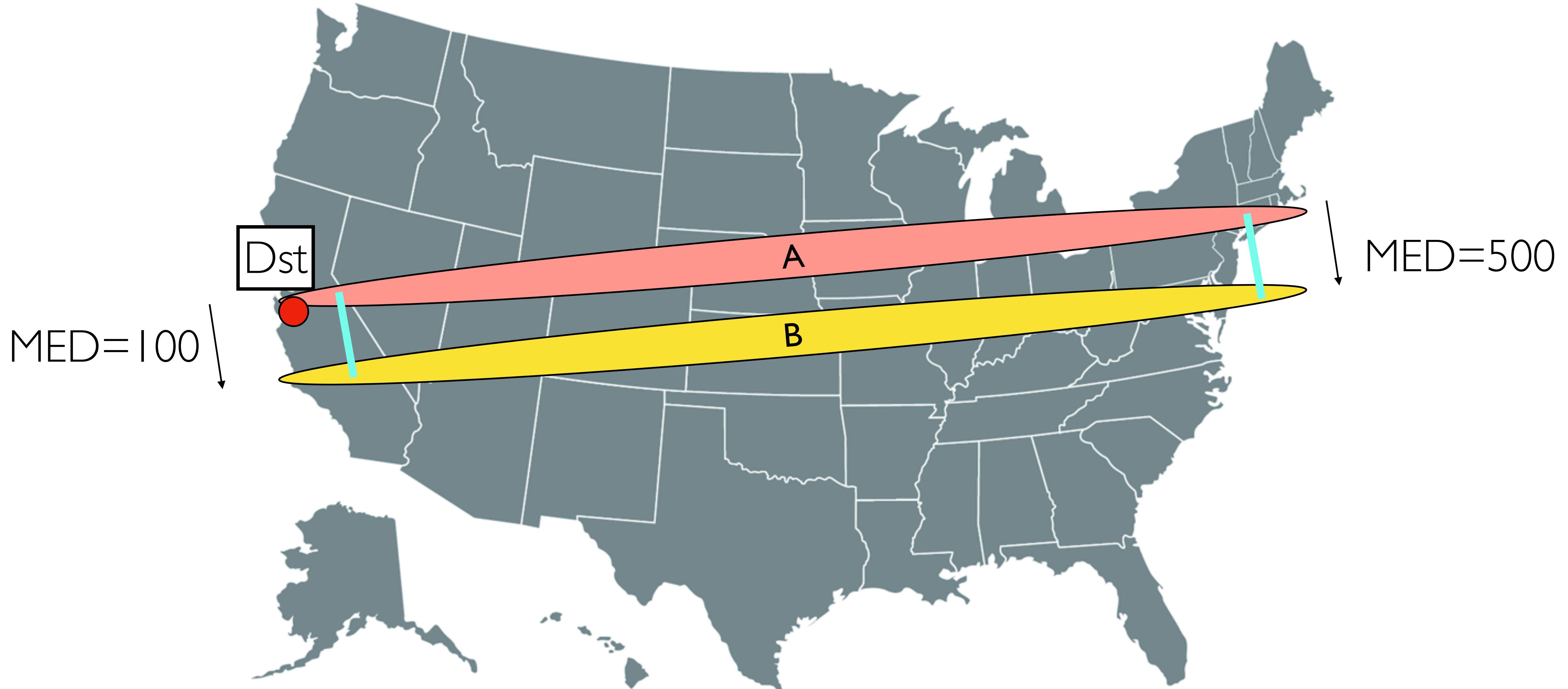
- Each router selects the closest egress point based on the path cost in intradomain protocol



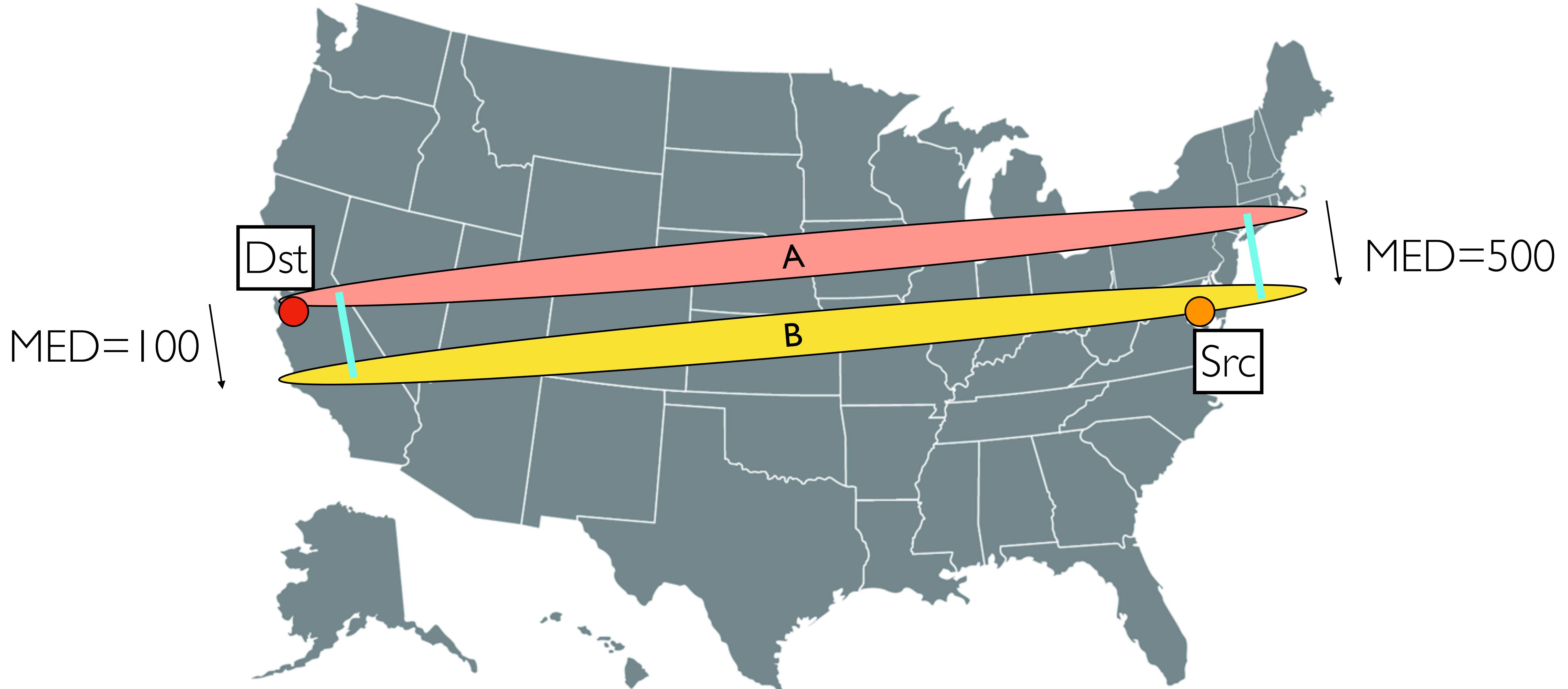
Hot Potato



IGP may conflict with MED



IGP may conflict with MED



Typical Selection Policy

Typical Selection Policy

- In decreasing order of priority
 - Make/save money (send to customer > peer > provider)
 - Maximize performance (small AS path length)
 - Minimize use of my network bandwidth (“hot potato”)
 - ...

Using Attributes

- Rules for route selection in priority order

Priority	Rule	Remarks
1	LOCAL PREF	Pick highest LOCAL PREF
2	ASPATH	Pick shortest ASPATH length
3	MED	Lowest MED preferred
4	eBGP > iBGP	Did AS learn route via eBGP (preferred) or iBGP?
5	Hot Potato	Lowest IGP cost to next hop (egress router)
6	Router ID	Smallest next-hop router's IP address as tie-breaker

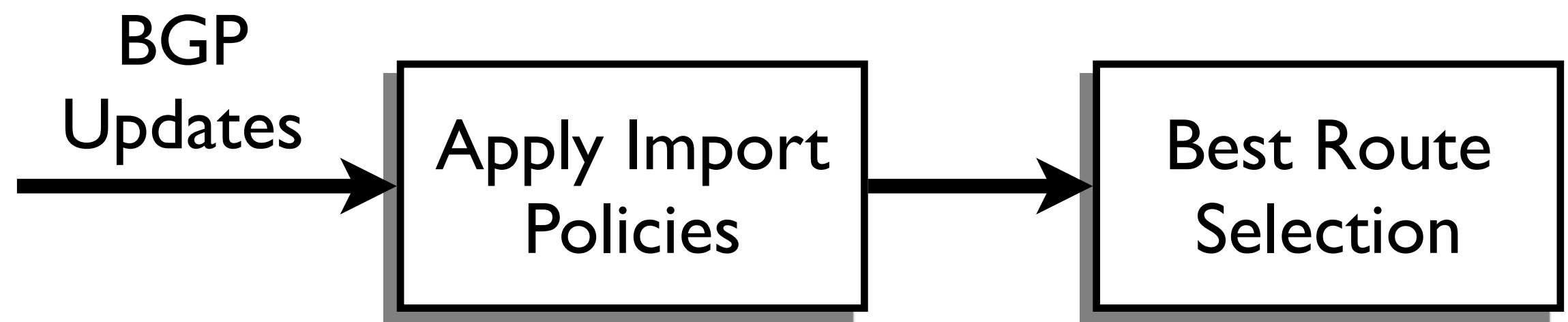
BGP UPDATE Processing

BGP UPDATE Processing

BGP
Updates



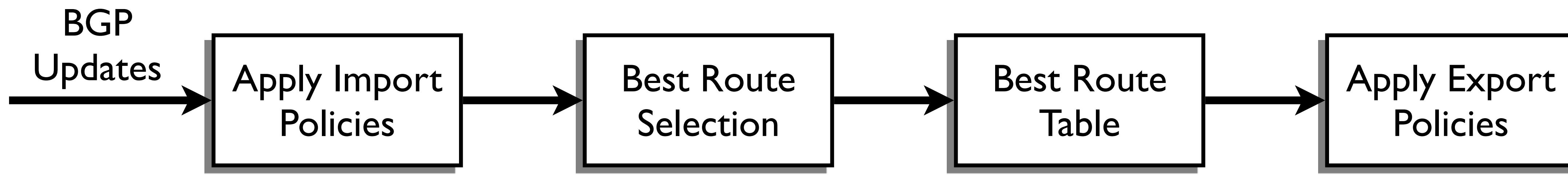
BGP UPDATE Processing



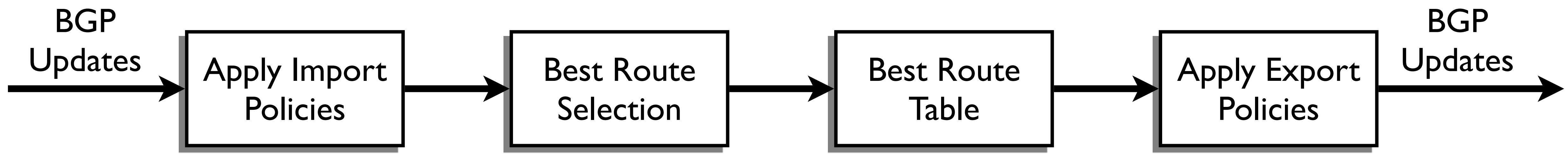
BGP UPDATE Processing



BGP UPDATE Processing

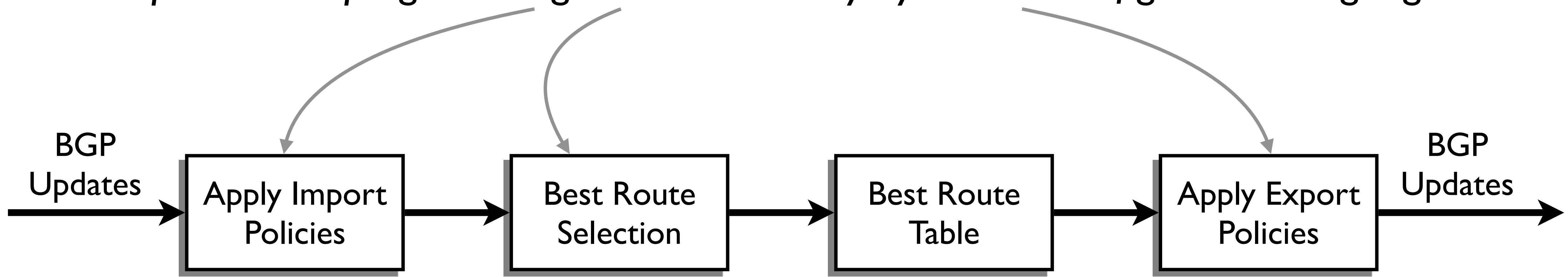


BGP UPDATE Processing

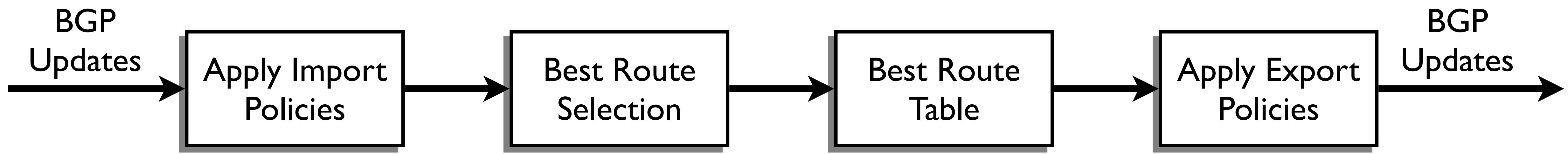


BGP UPDATE Processing

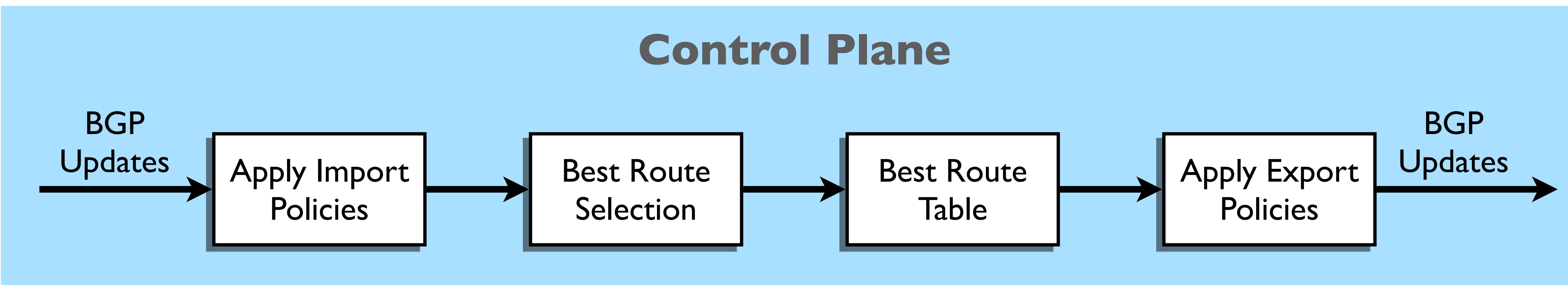
Open-ended programming: constrained only by vendor configuration language



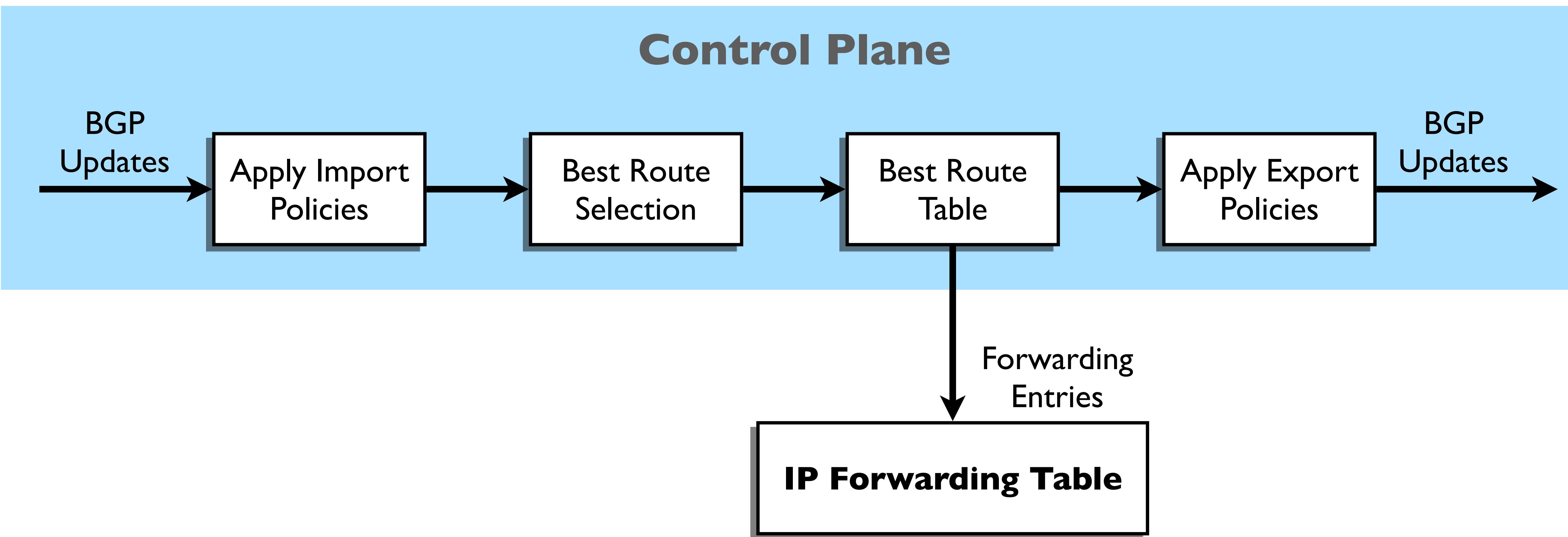
BGP UPDATE Processing



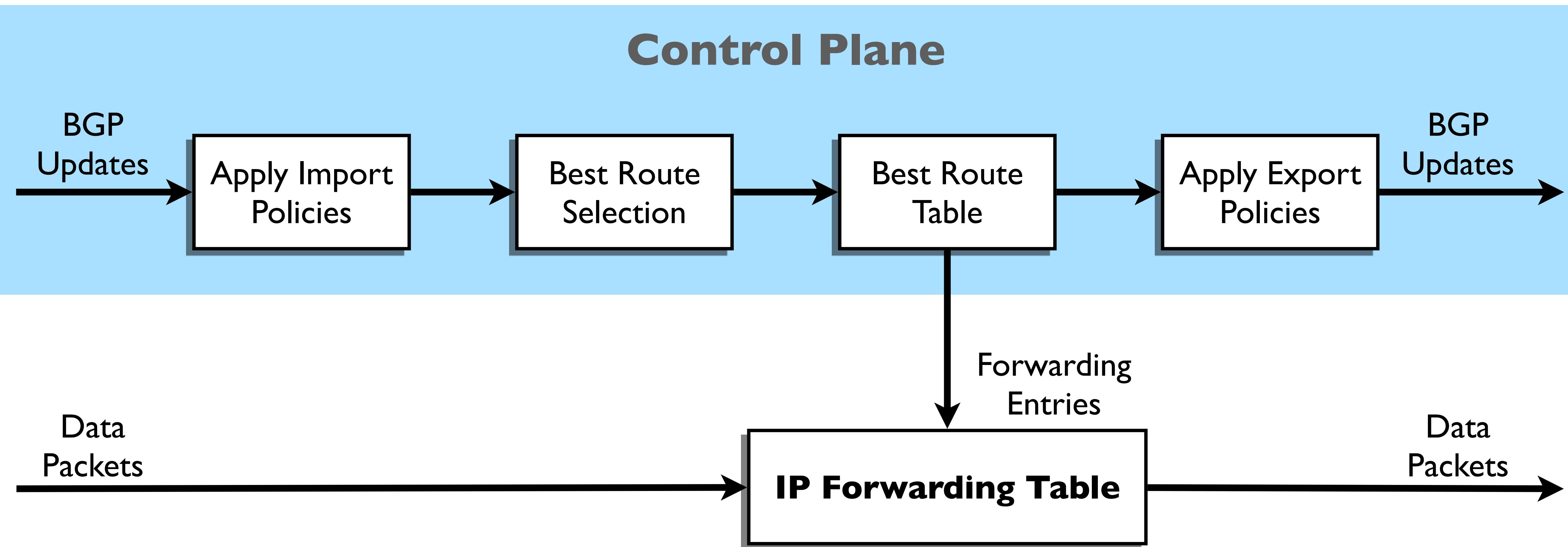
BGP UPDATE Processing



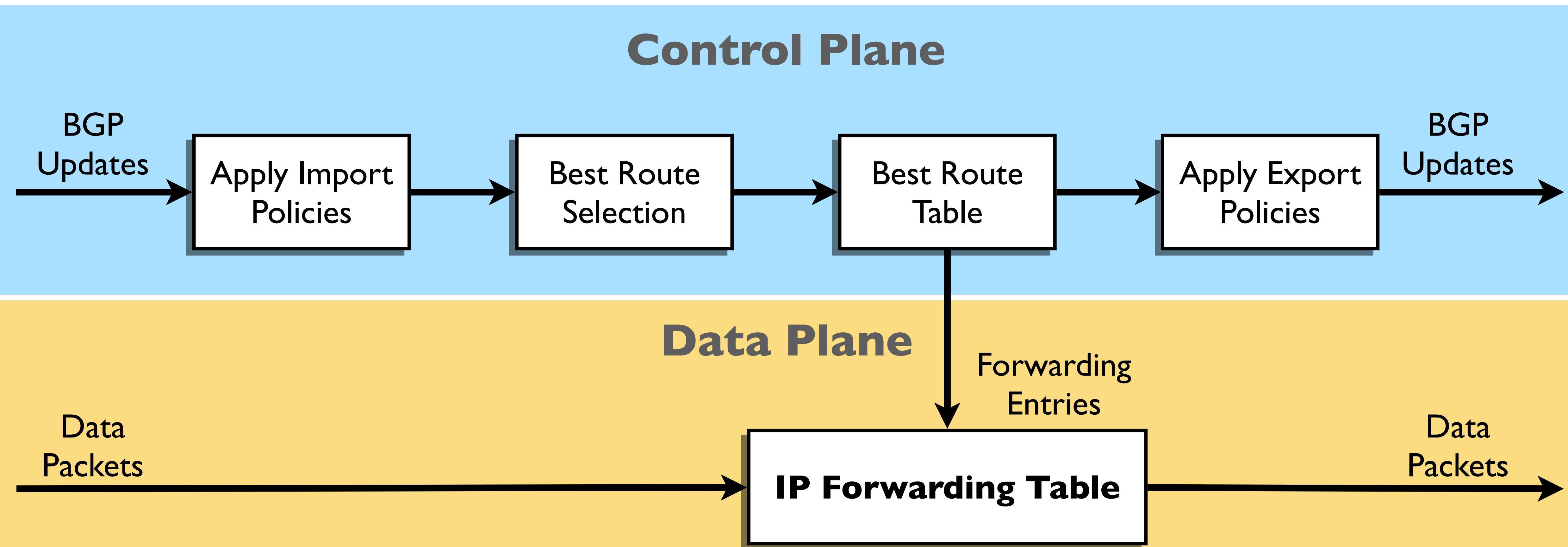
BGP UPDATE Processing



BGP UPDATE Processing



BGP UPDATE Processing



Questions?

BGP Outline

- **BGP Policy**

- Typical policies, how they are implemented

- **BGP Protocol Details**

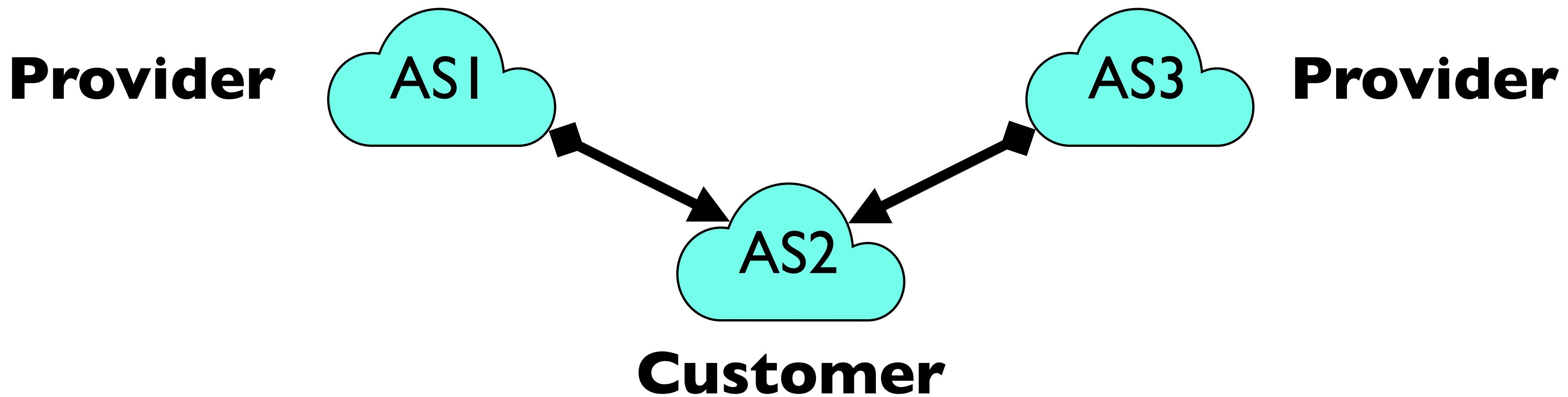
- **Issues with BGP**

Issues with BGP

- Reachability
- Security
- Convergence
- Performance
- Misconfigurations

Reachability

- In normal routing, if graph is connected, then reachability is assured
- With policy routing, this does not always hold



Security

Security

- **An AS can claim to serve a prefix that they actually don't have a route to (blackholing traffic)**

- Problem not specific to policy or path vector
- Important because of AS autonomy
- *Fixable: make ASes “prove” they have a path*

Security

- **An AS can claim to serve a prefix that they actually don't have a route to (blackholing traffic)**
 - Problem not specific to policy or path vector
 - Important because of AS autonomy
 - *Fixable: make ASes “prove” they have a path*
- **Note: AS may forward packets along a route different from what is advertised**
 - Tell customers about fictitious short path...
 - Much harder to fix!

Convergence

Convergence

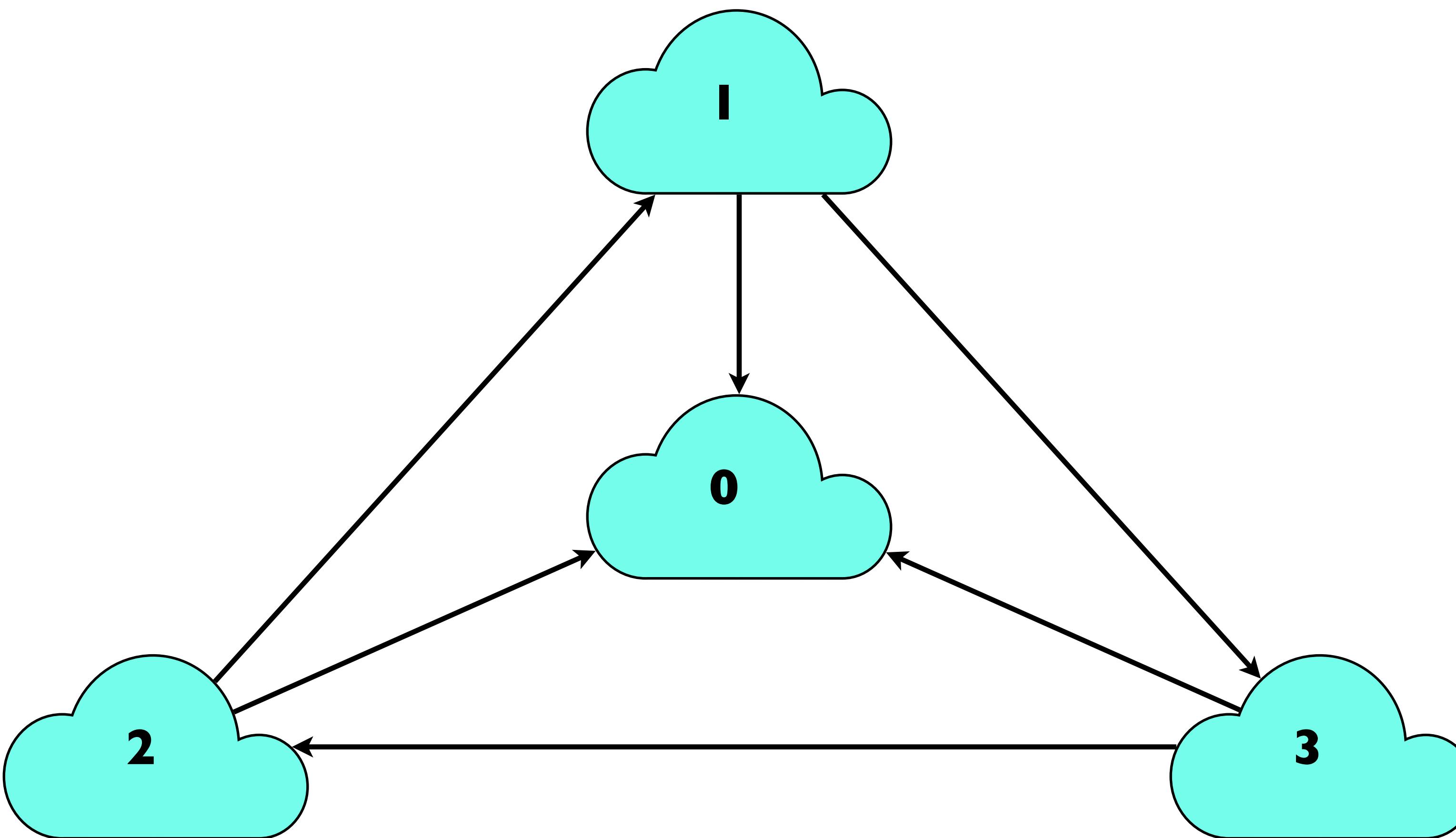
- **Result:** If all AS policies follow “Gao-Rexford” rules, BGP is guaranteed to converge (safety)

Convergence

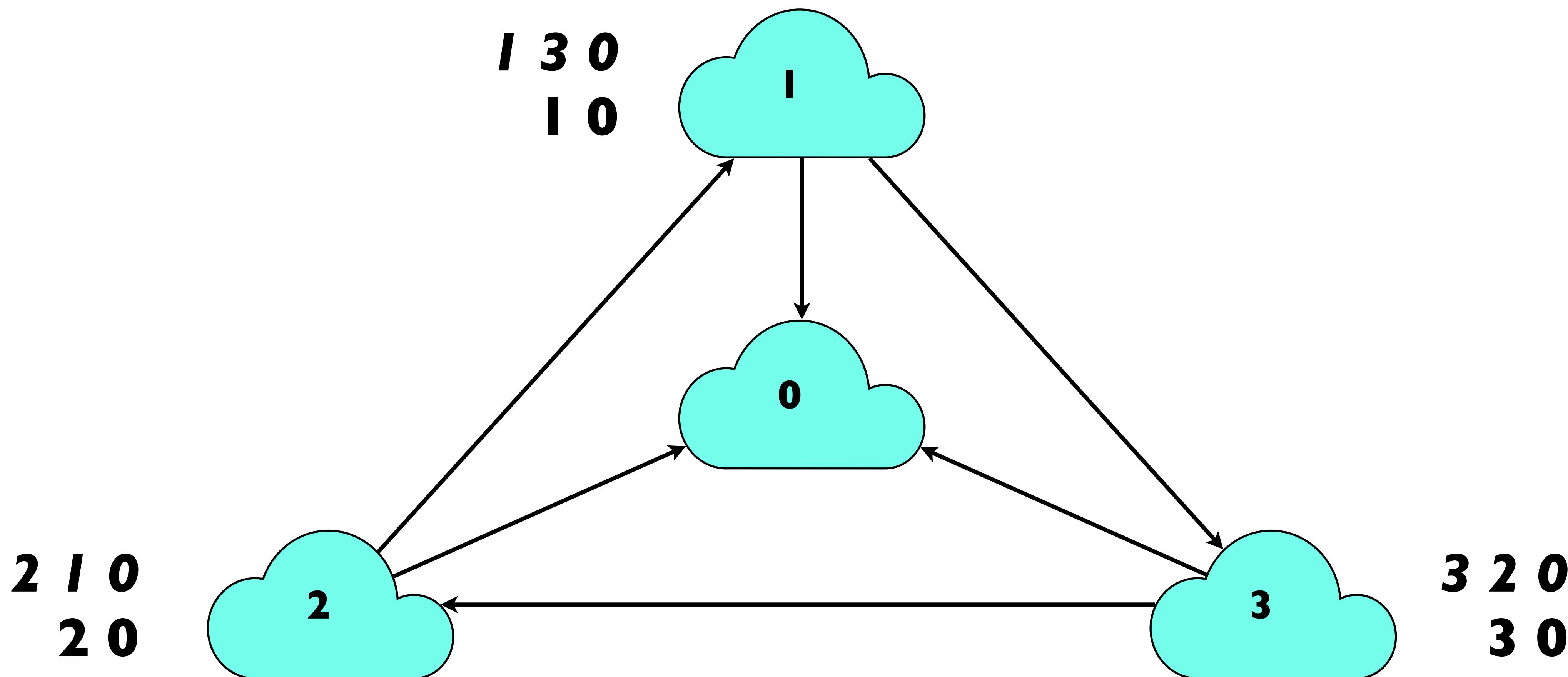
- **Result:** If all AS policies follow “Gao-Rexford” rules, BGP is guaranteed to converge (safety)
- For arbitrary policies, BGP may fail to converge!

Example of Policy Oscillation

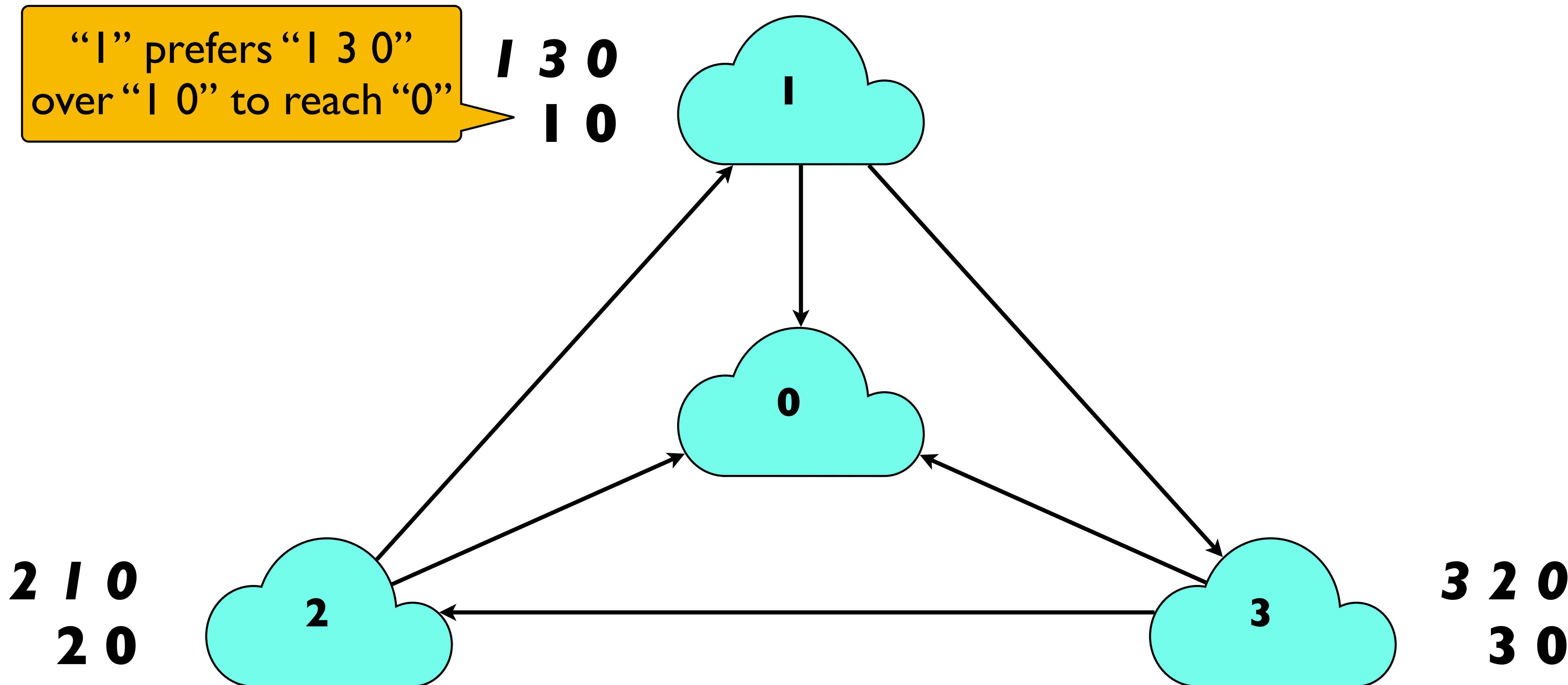
Example of Policy Oscillation



Example of Policy Oscillation

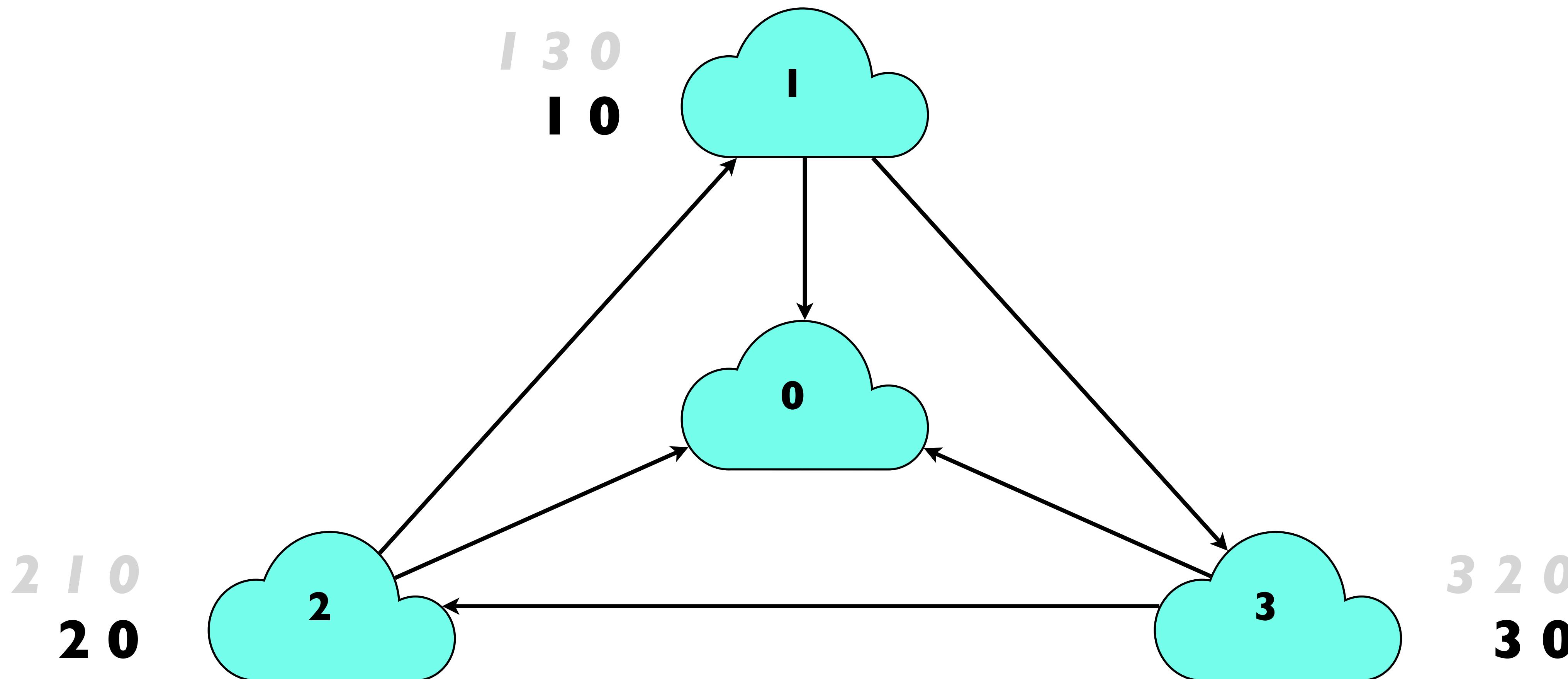


Example of Policy Oscillation



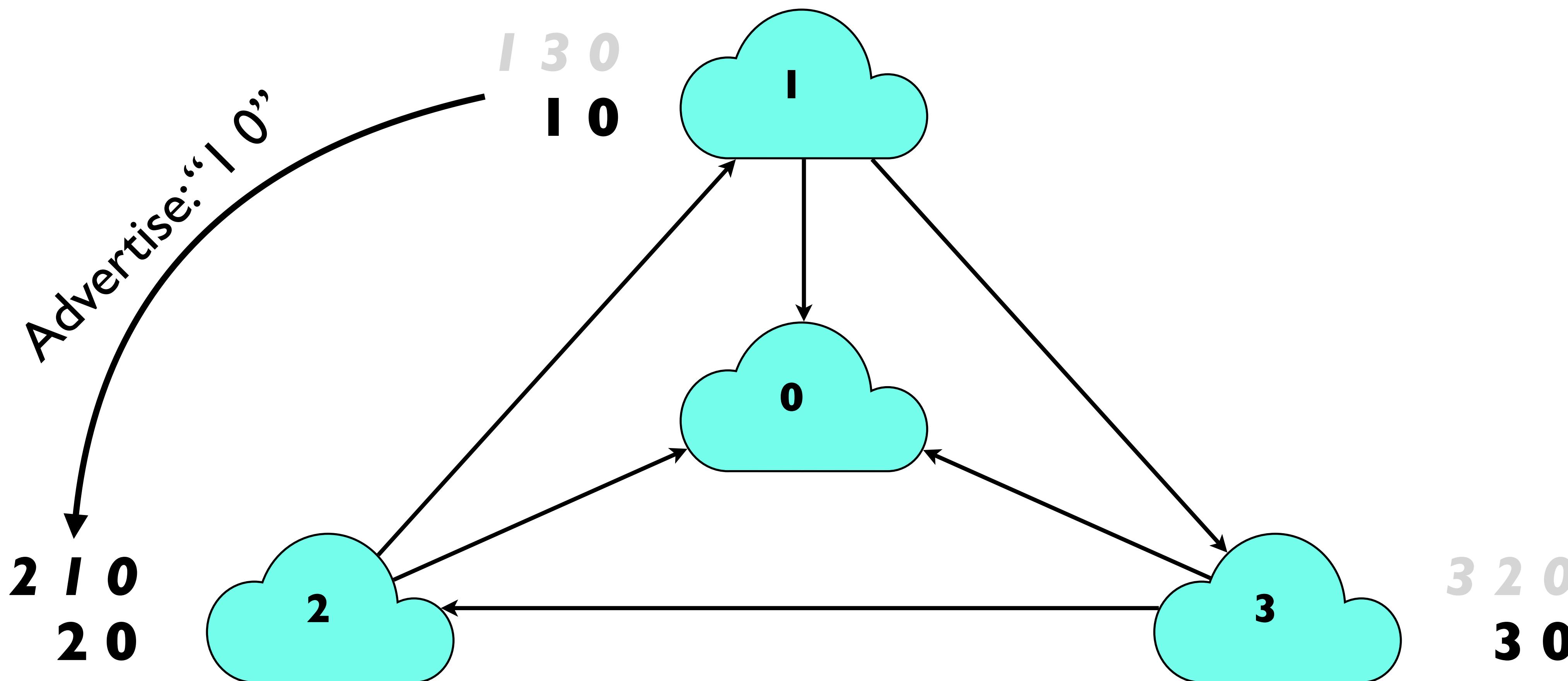
Example of Policy Oscillation

Initially: Nodes 1, 2, 3 know only shortest path to 0

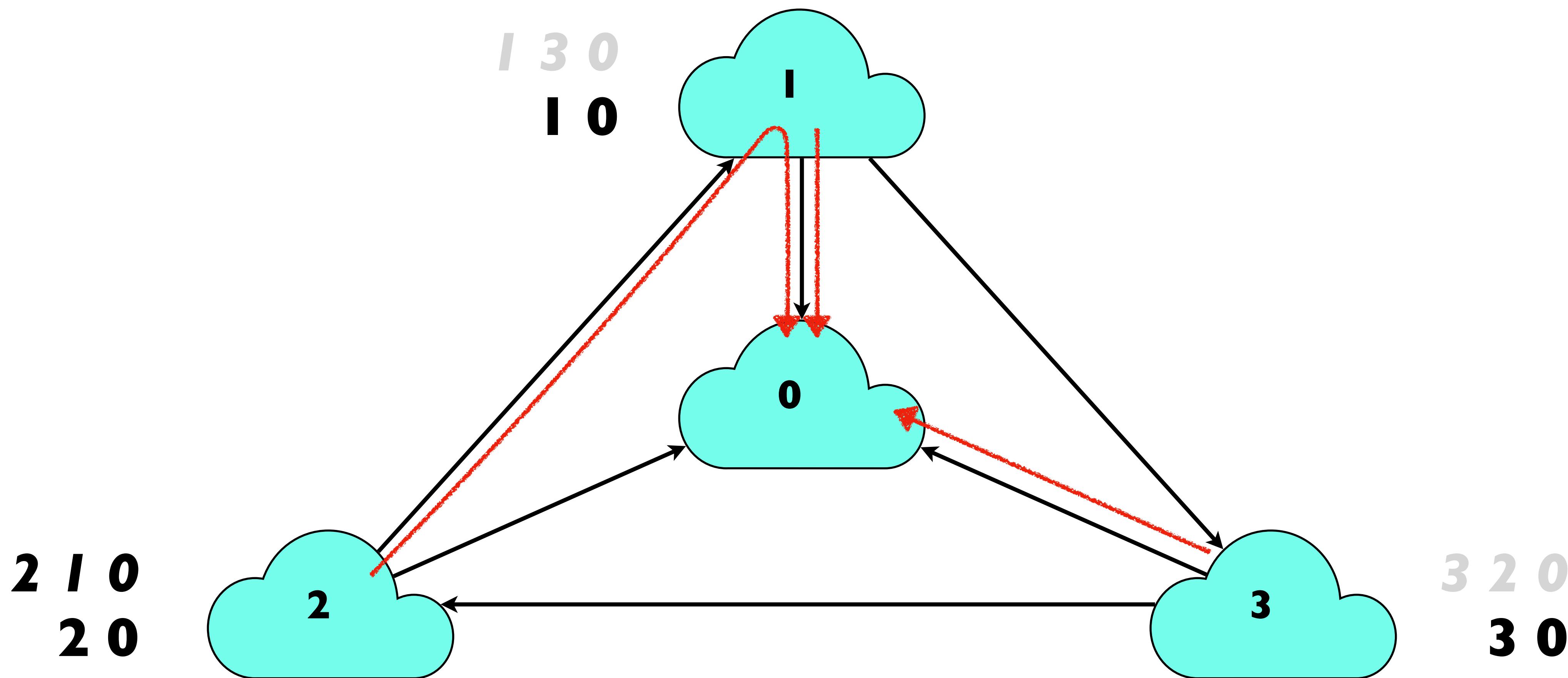


Example of Policy Oscillation

I **advertises** its path “I 0” to 2

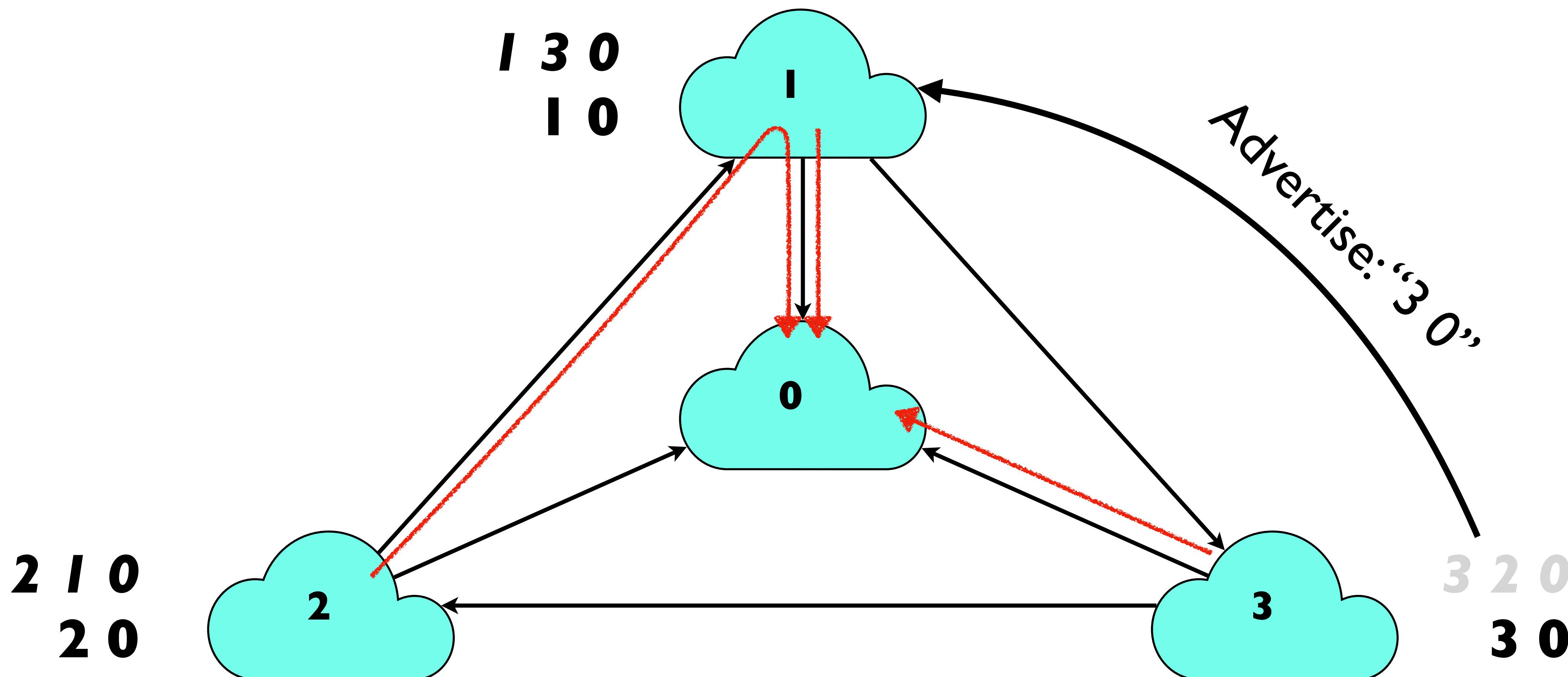


Example of Policy Oscillation

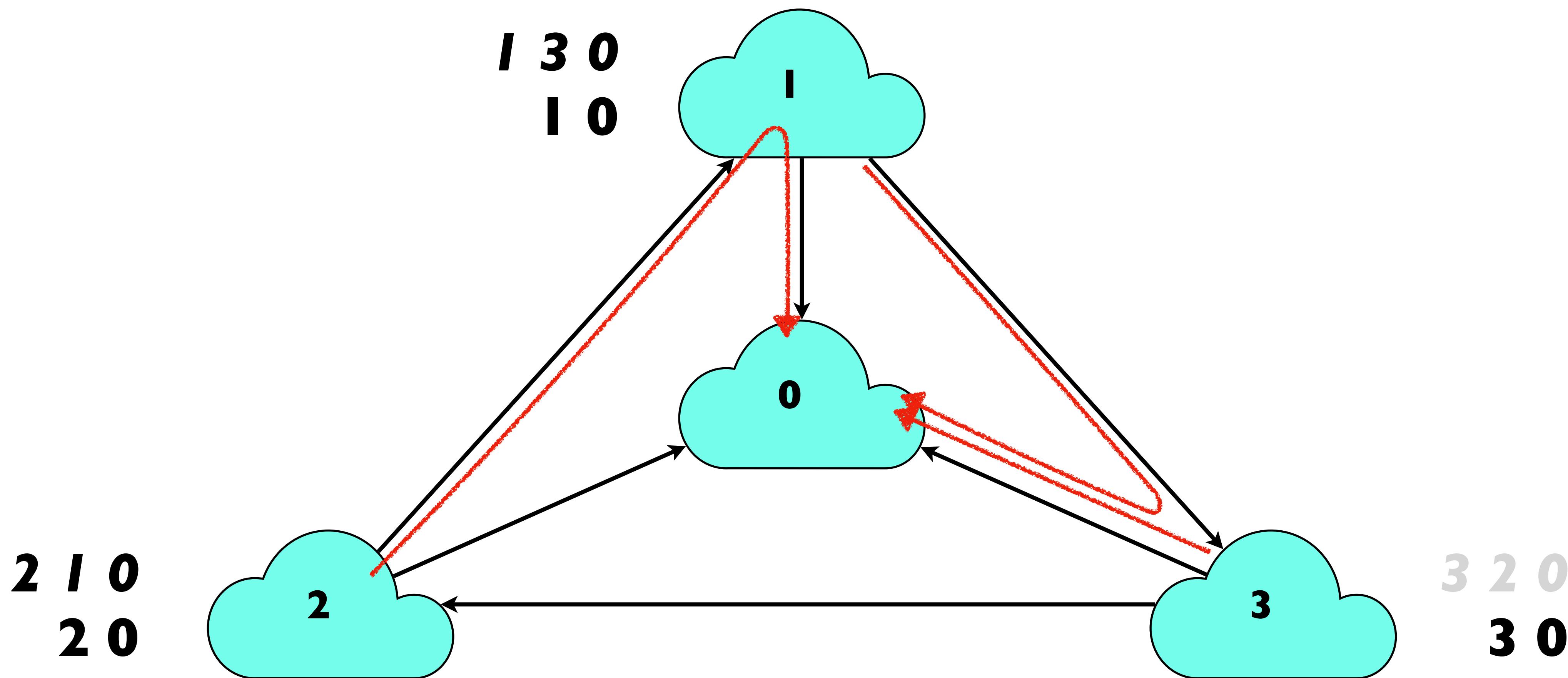


Example of Policy Oscillation

3 advertises its path “3 0” to I

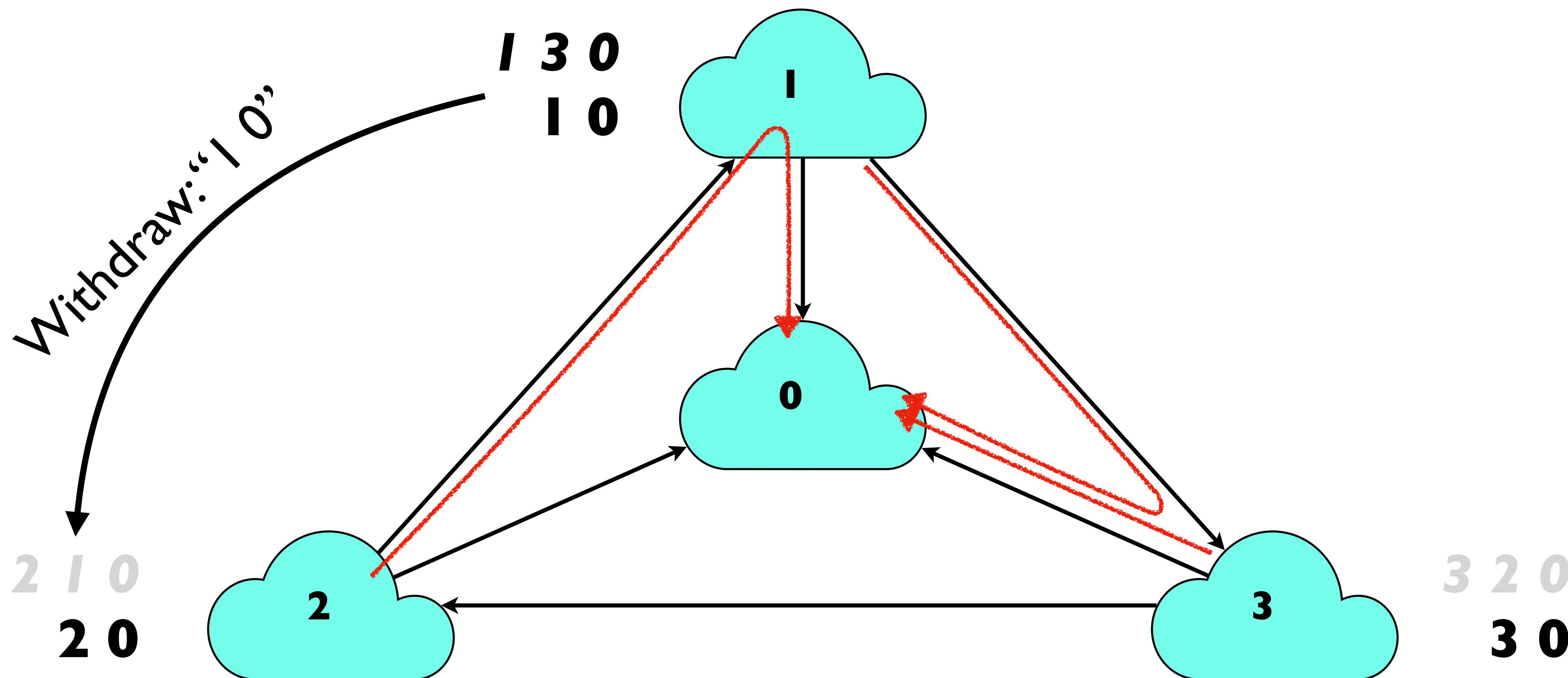


Example of Policy Oscillation

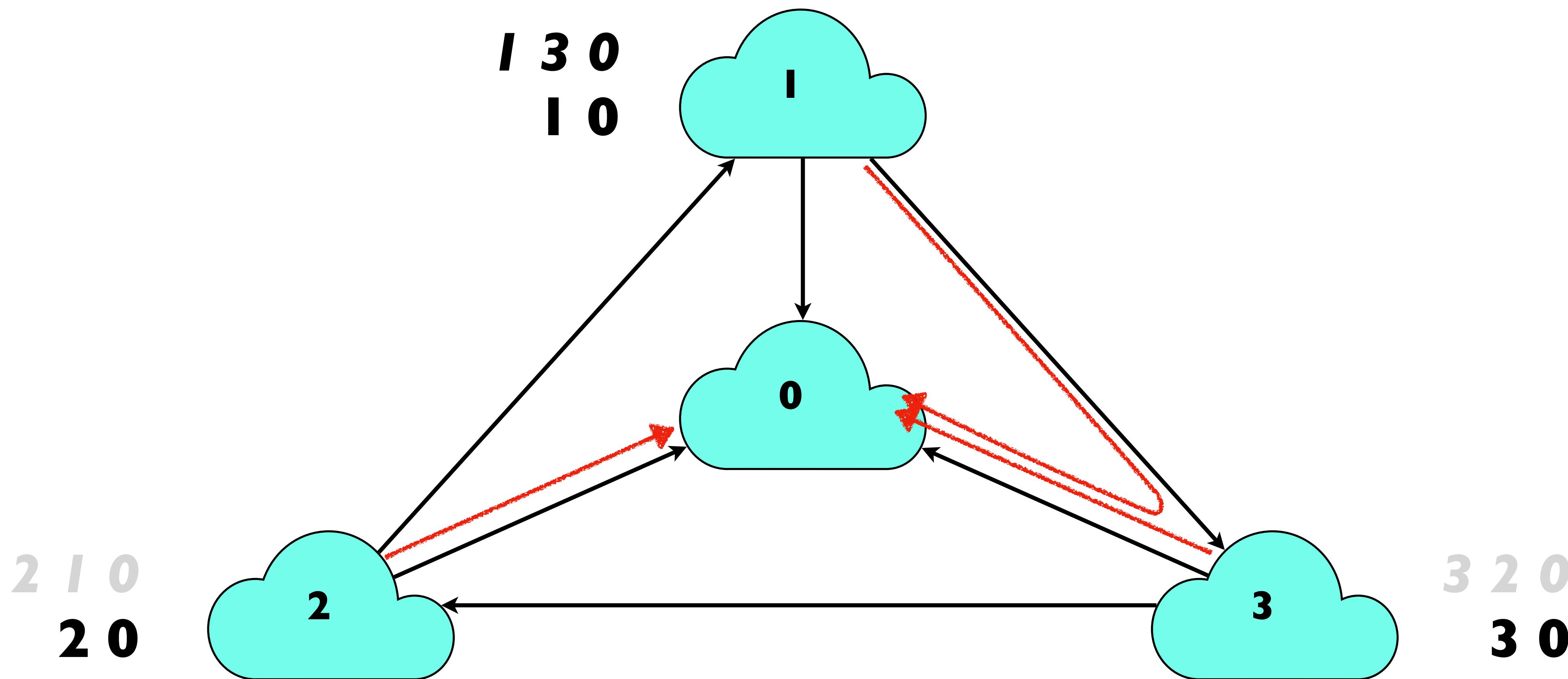


Example of Policy Oscillation

I withdraws its path “I 0” from 2

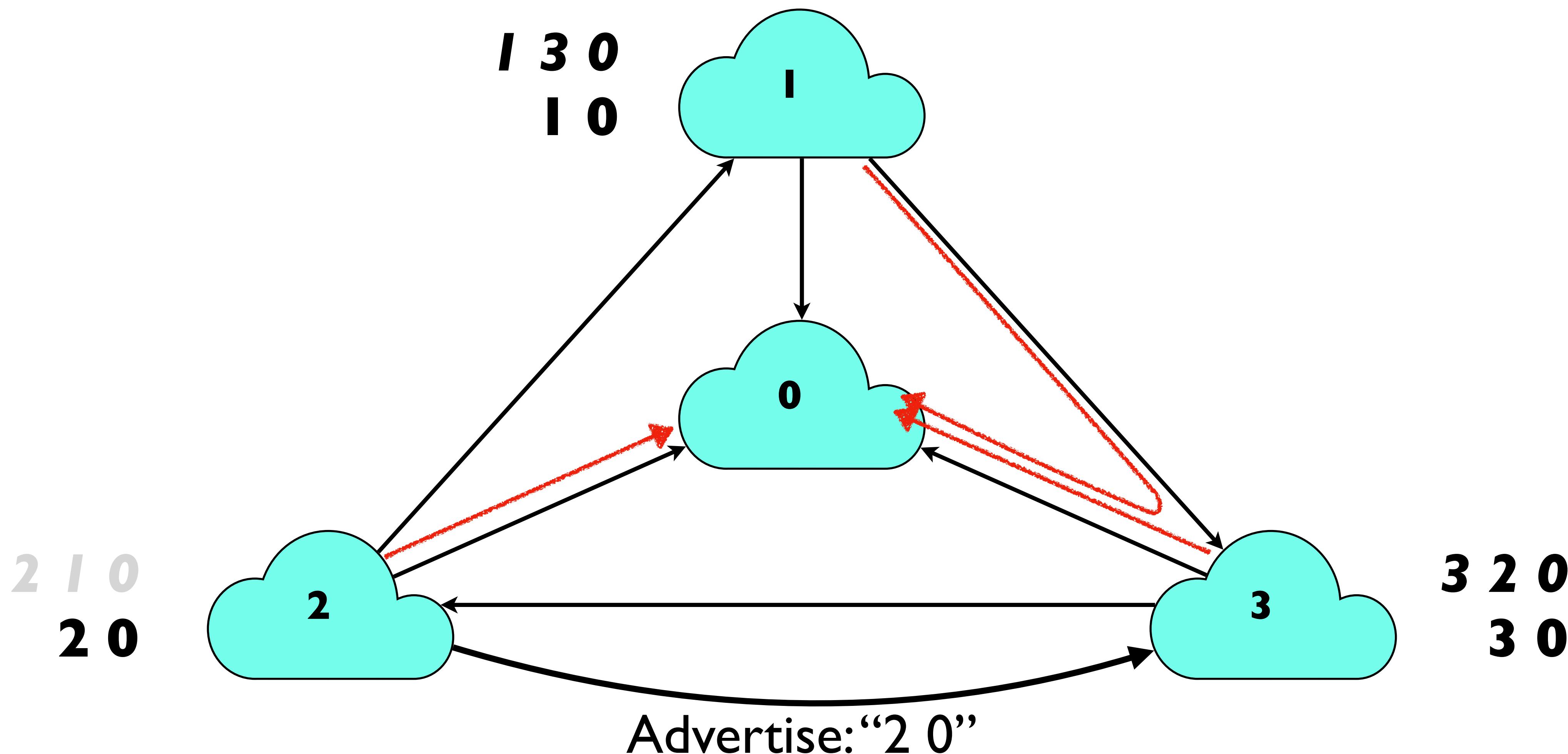


Example of Policy Oscillation

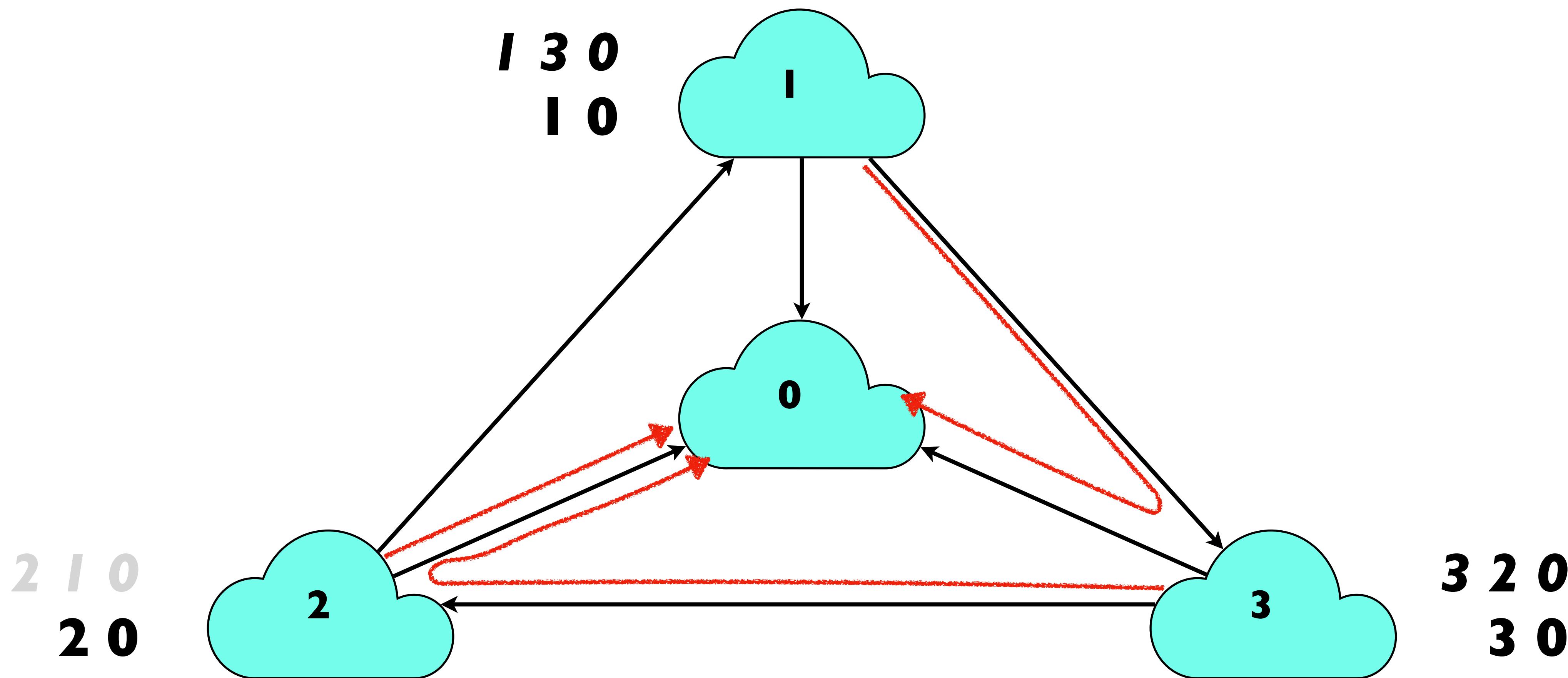


Example of Policy Oscillation

2 advertises its path “2 0” to 3

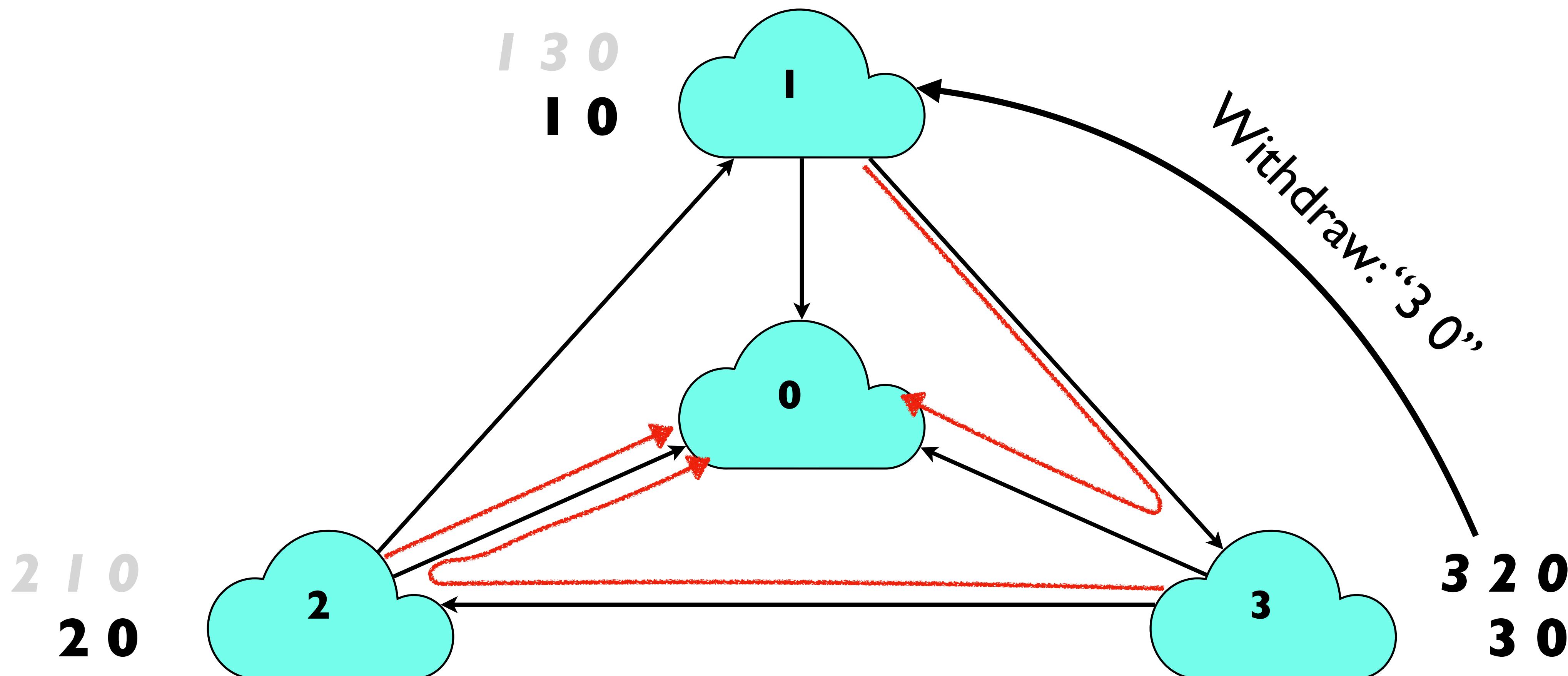


Example of Policy Oscillation

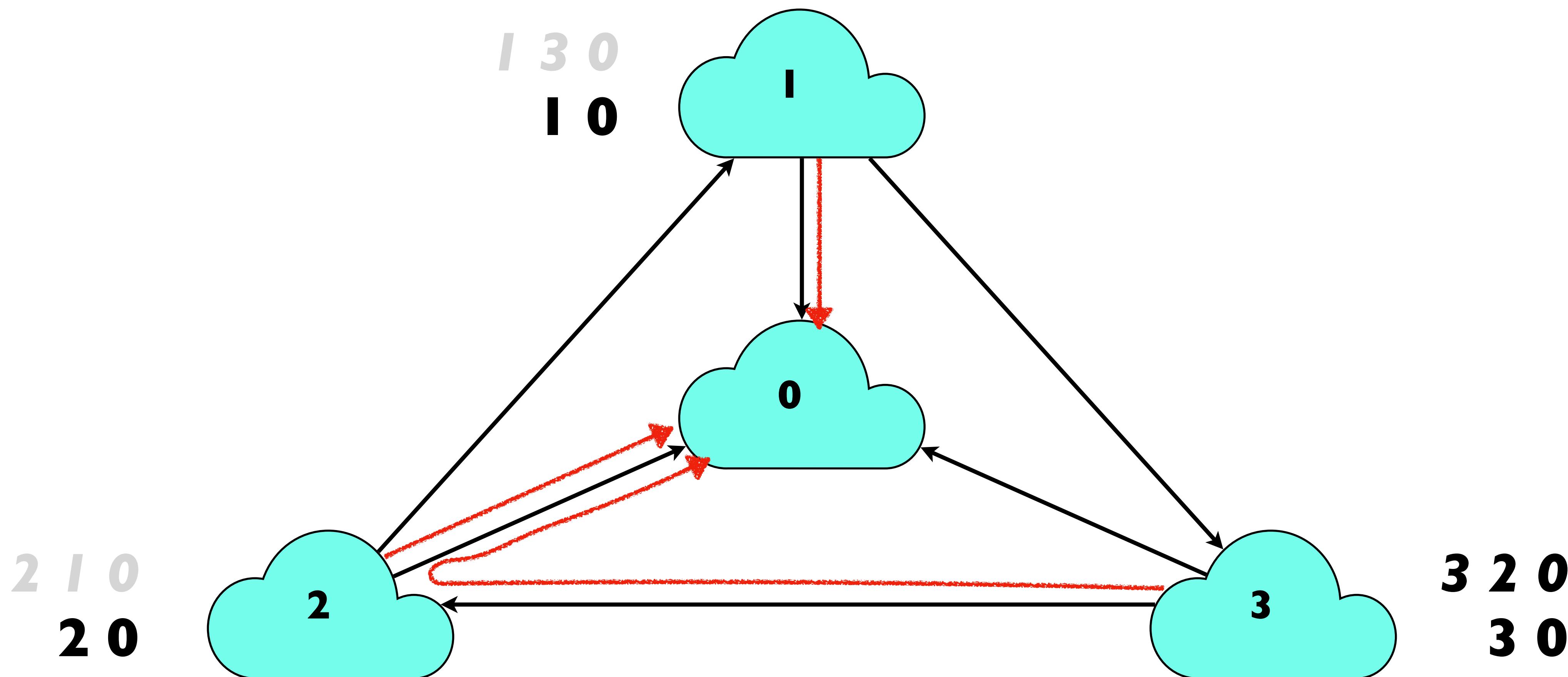


Example of Policy Oscillation

3 withdraws its path “3 0” from I

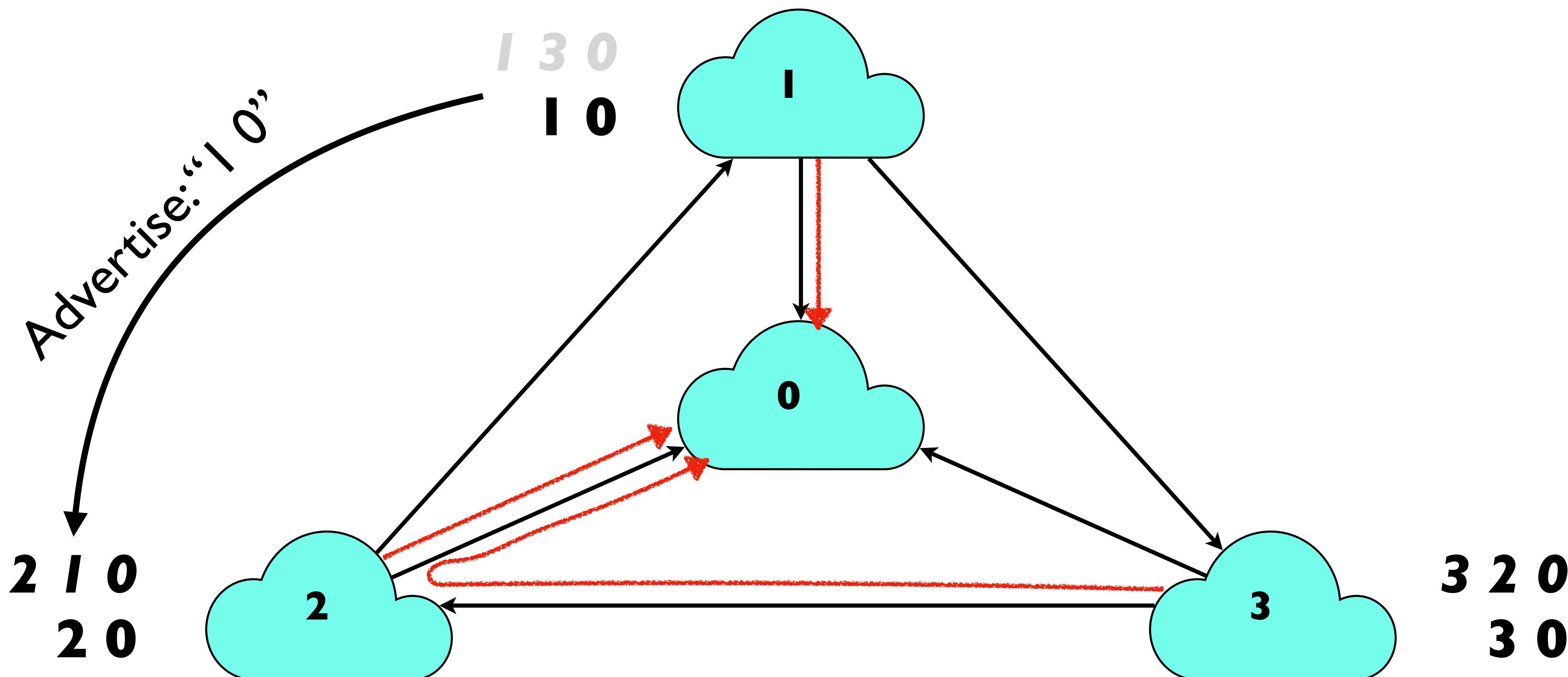


Example of Policy Oscillation

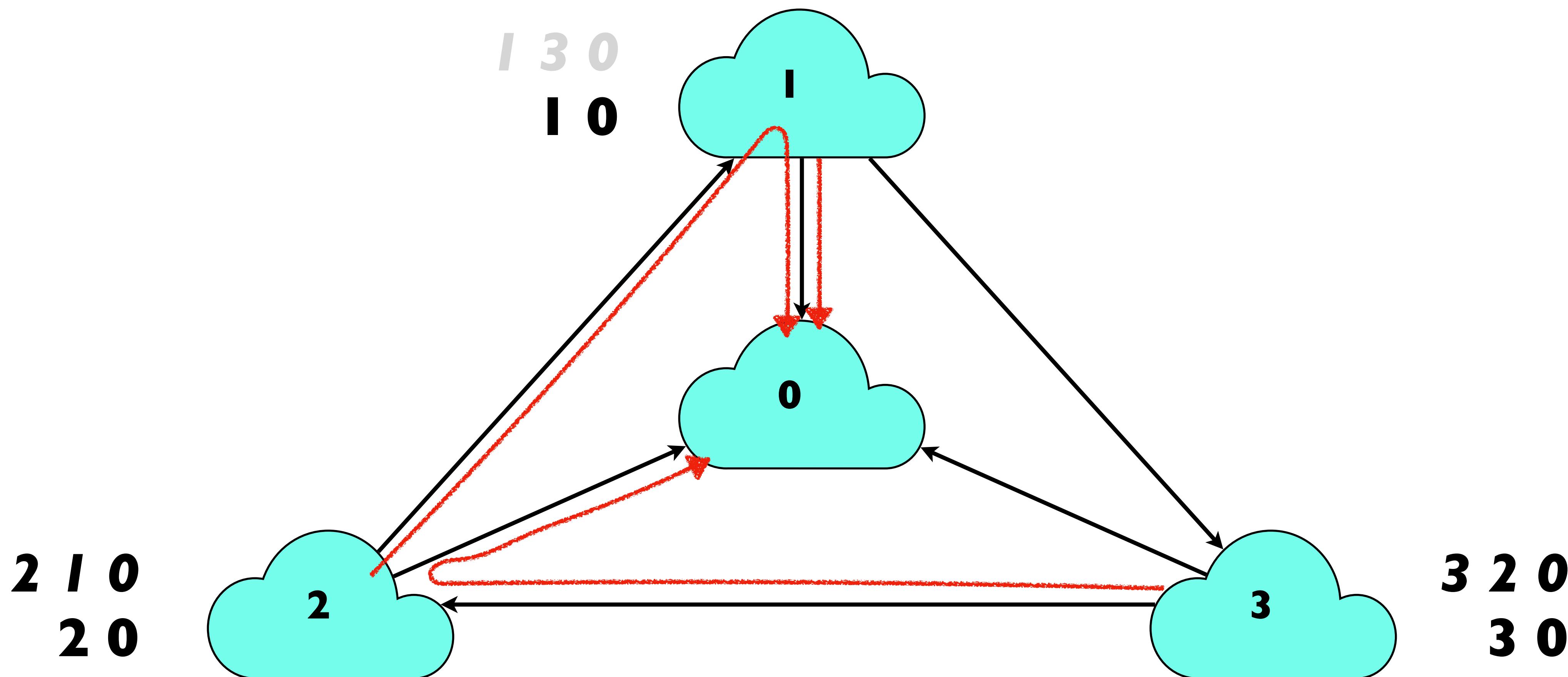


Example of Policy Oscillation

I **advertises** its path “I 0” to 2

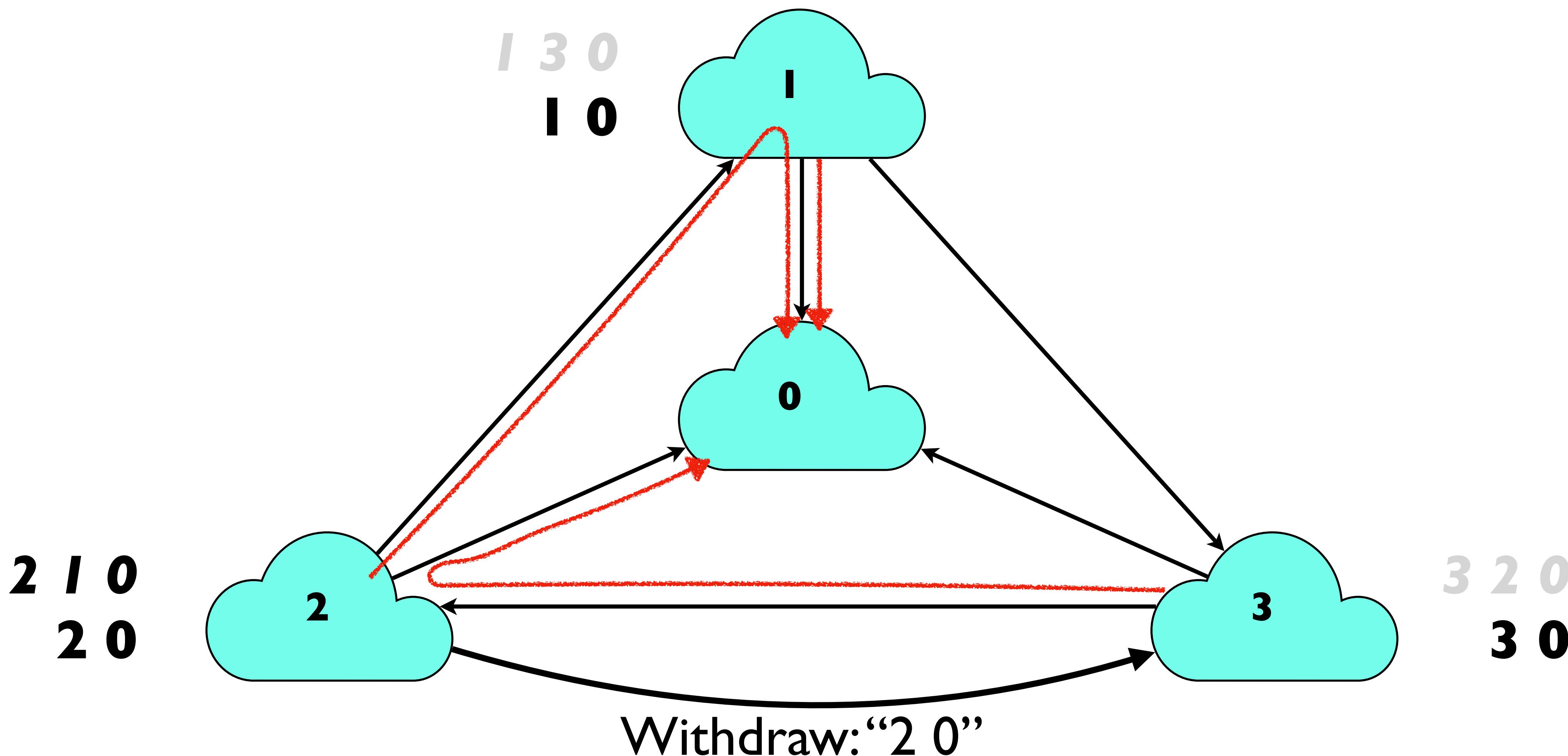


Example of Policy Oscillation



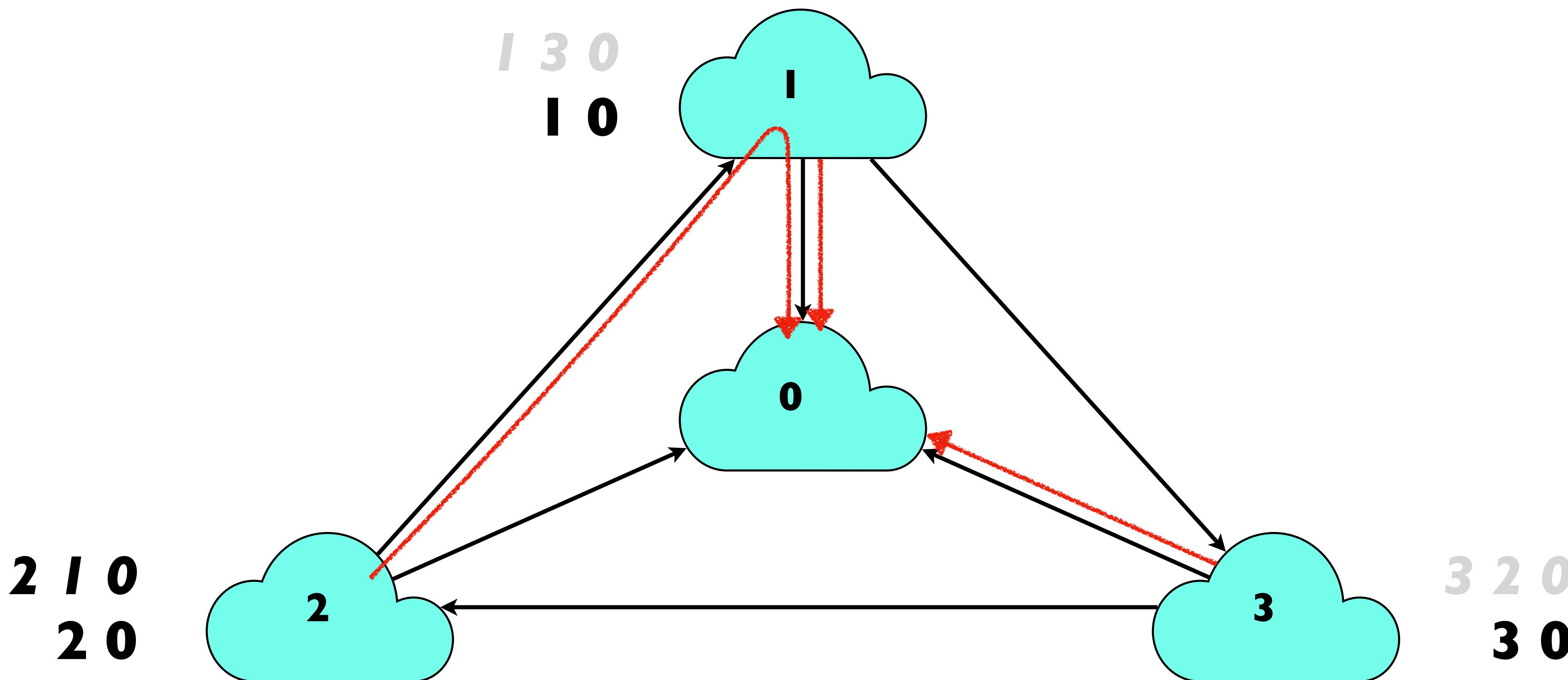
Example of Policy Oscillation

2 withdraws its path “2 0” from 3



Example of Policy Oscillation

We are back where we started!



Convergence

- **Result:** If all AS policies follow “Gao-Rexford” rules, BGP is guaranteed to converge (safety)
- For arbitrary policies, BGP may fail to converge!

Convergence

- **Result:** If all AS policies follow “Gao-Rexford” rules, BGP is guaranteed to converge (safety)
- For arbitrary policies, BGP may fail to converge!
- Why should this trouble us?

Performance Non-issues

Performance Non-issues

- **Internal routing (non)**

- Domains typically use “hot potato” routing
- Not always optimal, but economically expedient

Performance Non-issues

- **Internal routing (non)**

- Domains typically use “hot potato” routing
- Not always optimal, but economically expedient

- **Policy not about performance (non)**

- So policy-chosen paths aren’t shortest

Performance Non-issues

- **Internal routing (non)**

- Domains typically use “hot potato” routing
- Not always optimal, but economically expedient

- **Policy not about performance (non)**

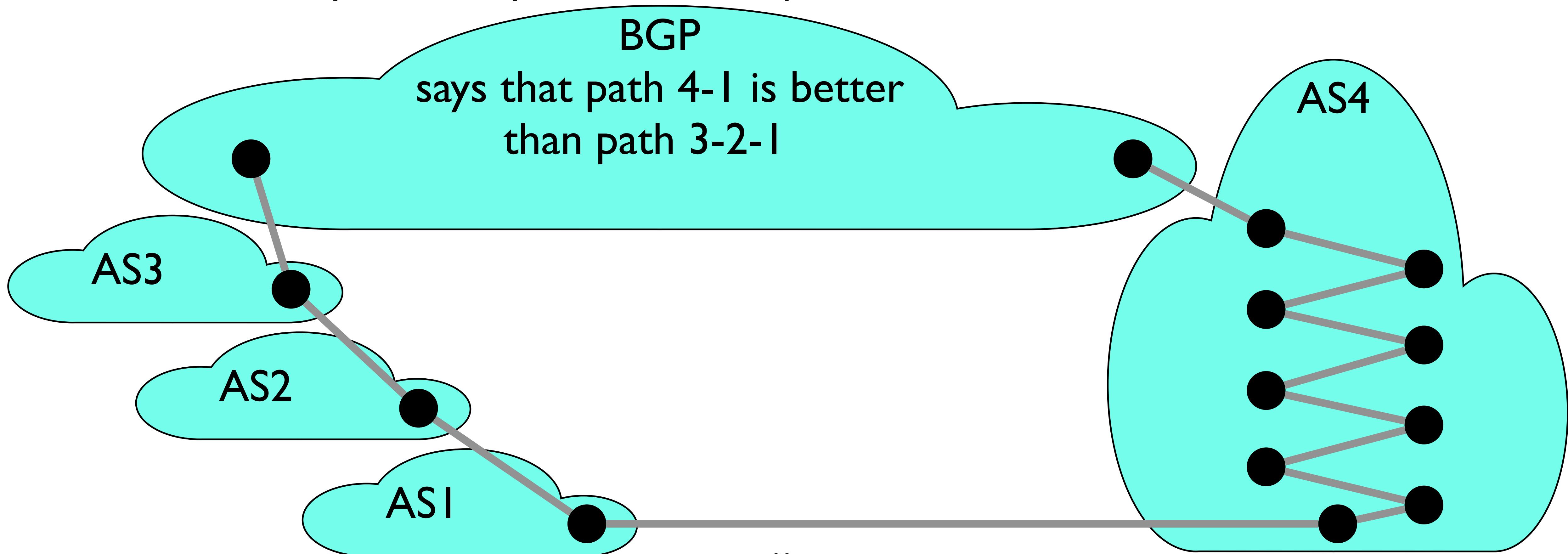
- So policy-chosen paths aren’t shortest

- **AS path length can be misleading (non)**

- 20% of paths inflated by at least 5 router hops

Performance Example

- **AS path length can be misleading**
 - An AS may have many router-level hops



Real Performance Issue: Slow Convergence

Real Performance Issue: Slow Convergence

- **BGP Outages are biggest source of Internet problems**

Real Performance Issue: Slow Convergence

- **BGP Outages are biggest source of Internet problems**
- **Labovitz et. al. SIGCOMM'97**
 - 10% of route available less than 95% of the time

Real Performance Issue: Slow Convergence

- **BGP Outages are biggest source of Internet problems**
- **Labovitz et. al. SIGCOMM'97**
 - 10% of route available less than 95% of the time
- **Labovitz et. al. SIGCOMM'00**
 - 40% of path outages take 30+ minutes to repair

Real Performance Issue: Slow Convergence

- **BGP Outages are biggest source of Internet problems**
- **Labovitz et. al. SIGCOMM'97**
 - 10% of route available less than 95% of the time
- **Labovitz et. al. SIGCOMM'00**
 - 40% of path outages take 30+ minutes to repair
- **But most popular paths are very stable**

BGP Misconfigurations

BGP Misconfigurations

- **BGP Protocol is both bloated and underspecified**

- Lots of attributes
- Lots of leeway in how to set and interpret attributes
- Necessary to allow autonomy, diverse policies
- But also gives operators plenty of rope

BGP Misconfigurations

- **BGP Protocol is both bloated and underspecified**
 - Lots of attributes
 - Lots of leeway in how to set and interpret attributes
 - Necessary to allow autonomy, diverse policies
 - But also gives operators plenty of rope
- **Much of this configuration is manual and *ad hoc***

BGP Misconfigurations

- **BGP Protocol is both bloated and underspecified**

- Lots of attributes
- Lots of leeway in how to set and interpret attributes
- Necessary to allow autonomy, diverse policies
- But also gives operators plenty of rope

- **Much of this configuration is manual and *ad hoc***

- **And the core abstraction is fundamentally flawed**

- Disjoint per-router configuration to affect AS-wide policy
- Now strong industry interest in changing this!

BGP: How did we get here?

BGP: How did we get here?

- **BGP was designed for a different time**

- Before commercial ISPs and their needs
- Before address aggregation
- Before multi-homing

- **1989: BGP-1 [RFC 1105]**
 - Replacement for EGP (1984, RFC 904)
- **1990: BGP-2 [RFC 1163]**
- **1991: BGP-3 [RFC 1267]**
- **1995: BGP-4 [RFC 1771]**
 - Support for Classless Interdomain Routing (CIDR)

BGP: How did we get here?

- **BGP was designed for a different time**
 - Before commercial ISPs and their needs
 - Before address aggregation
 - Before multi-homing
- **We don't get a second chance: ‘clean slate’ designs virtually impossible to deploy**

BGP: How did we get here?

- **BGP was designed for a different time**
 - Before commercial ISPs and their needs
 - Before address aggregation
 - Before multi-homing
- **We don't get a second chance: ‘clean slate’ designs virtually impossible to deploy**
- **Thought experiment: how would you design a policy driven interdomain routing solution? How would you deploy it?**

Next time, on CPSC 433/533

- **Wrap up the network layer!**

- The IPv4 header
- Other exciting details