CS4450

Computer Networks:
Architecture and Protocols

Lecture 15

Border-Gateway Protocol

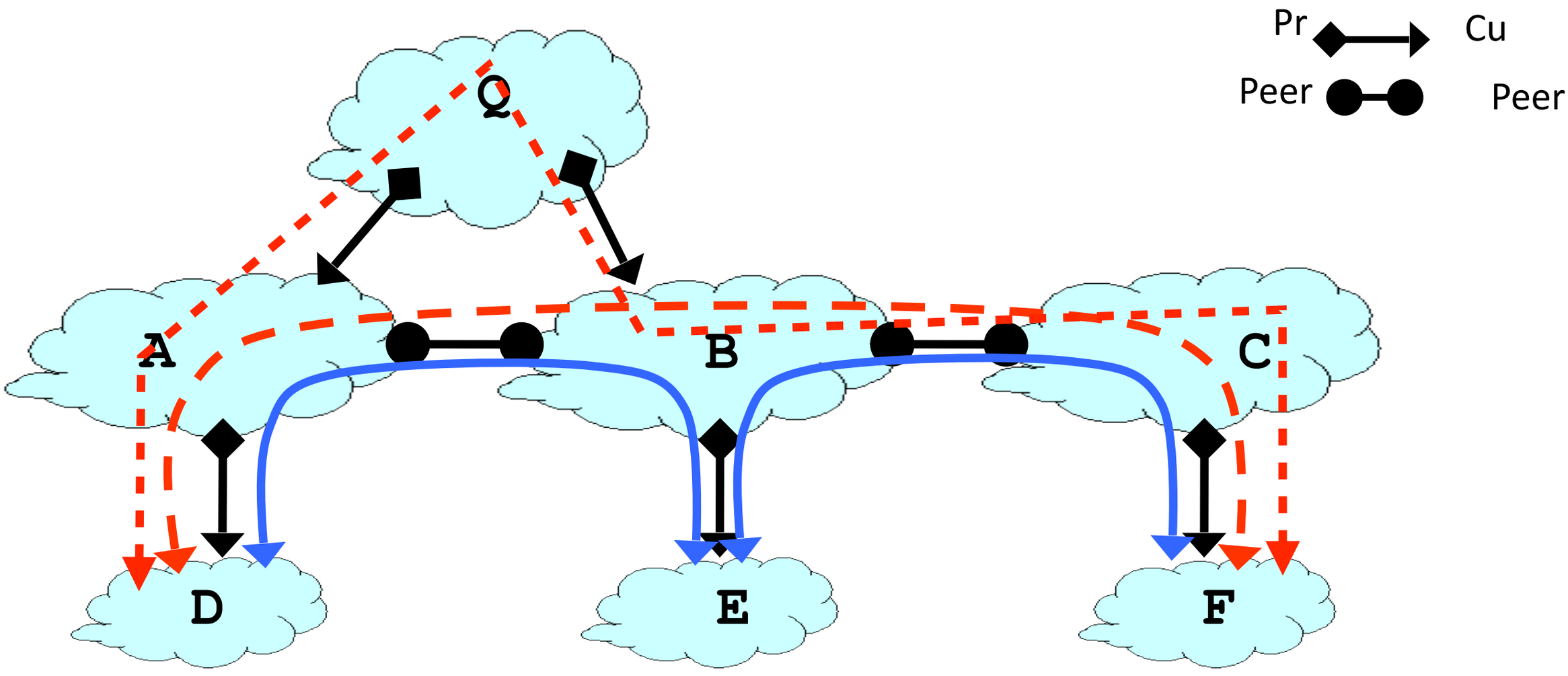
Rachit Agarwal

Goals for Today's Lecture

- Wrap up Inter-domain routing (Border-Gateway Protocol (BGP))
    - Driven by "business goals", rather than "performance goals"
    - We will focus on a synchronous version:
        - One node in the network acts at a time
        - In practice, BGP implementations are asynchronous

Recap from last lecture

Recap: Inter-domain Routing Follows the Money



traffic allowed traffic not allowed

- ASes provide “transit” between their customers
- Peers do not provide transit between other peers

Recap: Administrative Structure Shapes Inter-domain Routing

- ASes want freedom to pick routes based on policy
  - *"My traffic can't be carried over my competitor's network!"*
  - *"I don't want to carry A's traffic through my network!"*
  - Cannot be expressed as Internet-wide "least cost"

- ASes want autonomy
  - Want to choose their own internal routing protocol
  - Want to choose their own policy

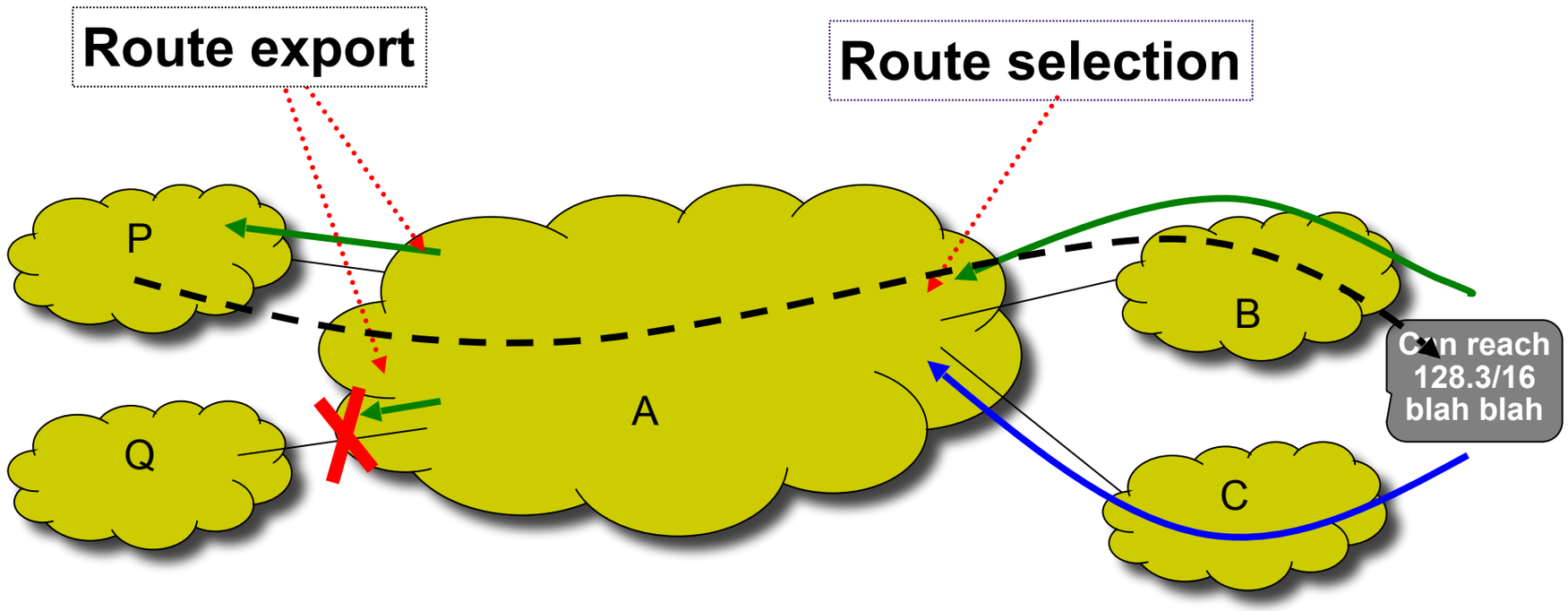
- ASes want privacy
  - Choice of network topology, routing policies, etc.

Recap: BGP is Inspired by Distance Vector

- Per-destination route advertisements
- No global sharing of network topology
- Iterative and distributed convergence on paths
- But, four key differences
  - BGP does not pick shortest paths
  - Each node announces one or multiple PATHs per destination
  - Selective Route advertisement: not all paths are announced
  - BGP may aggregate paths
    - may announce one path for multiple destinations

Recap: Policy:

Imposed in how routes are selected and exported



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

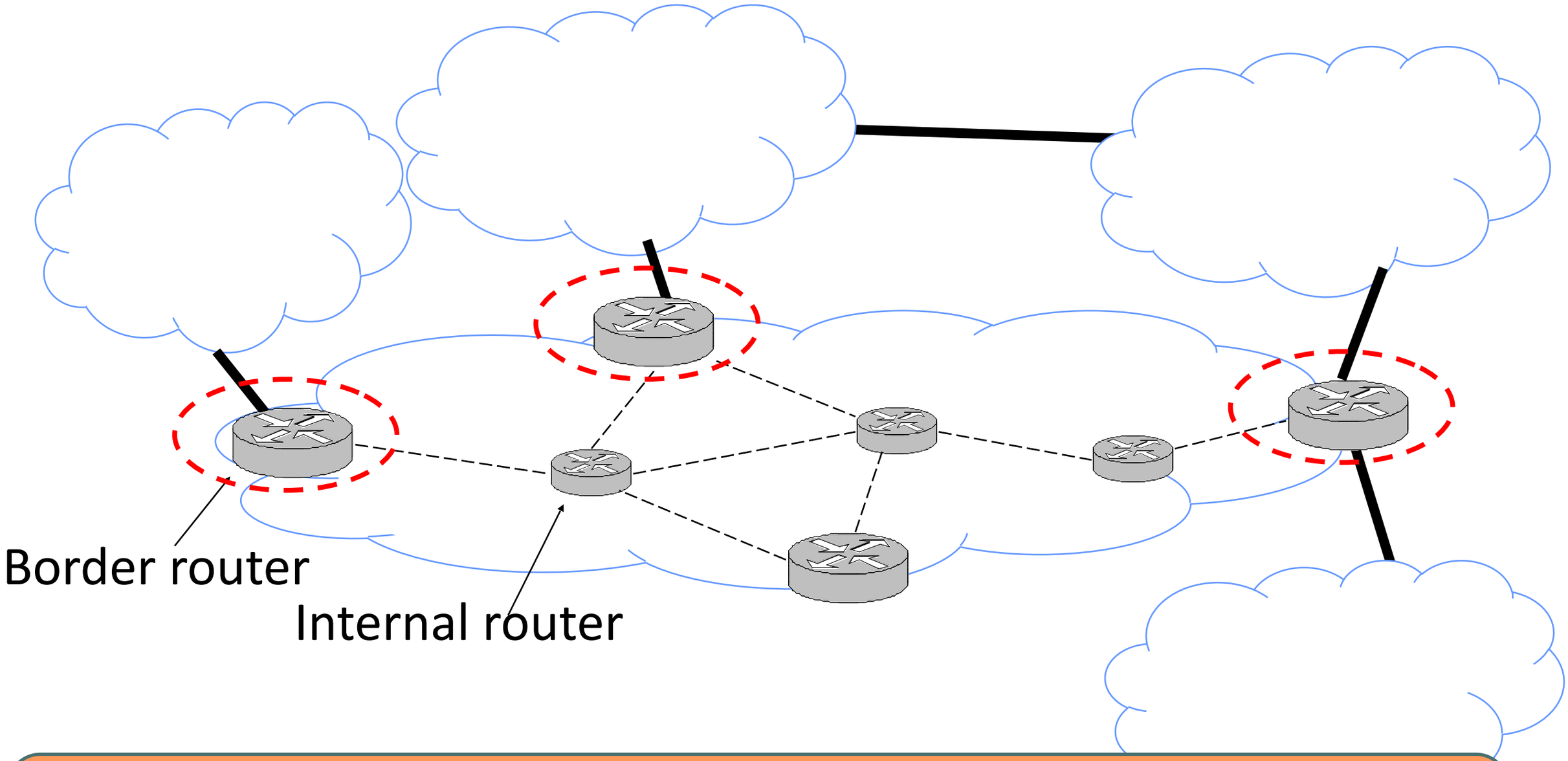
Recap: Typical Export Policy

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

Known as the “Gao-Rexford” rules
Capture common (but not required!) practice

BGP protocol details

Who speaks BGP?



Border routers at an Autonomous System

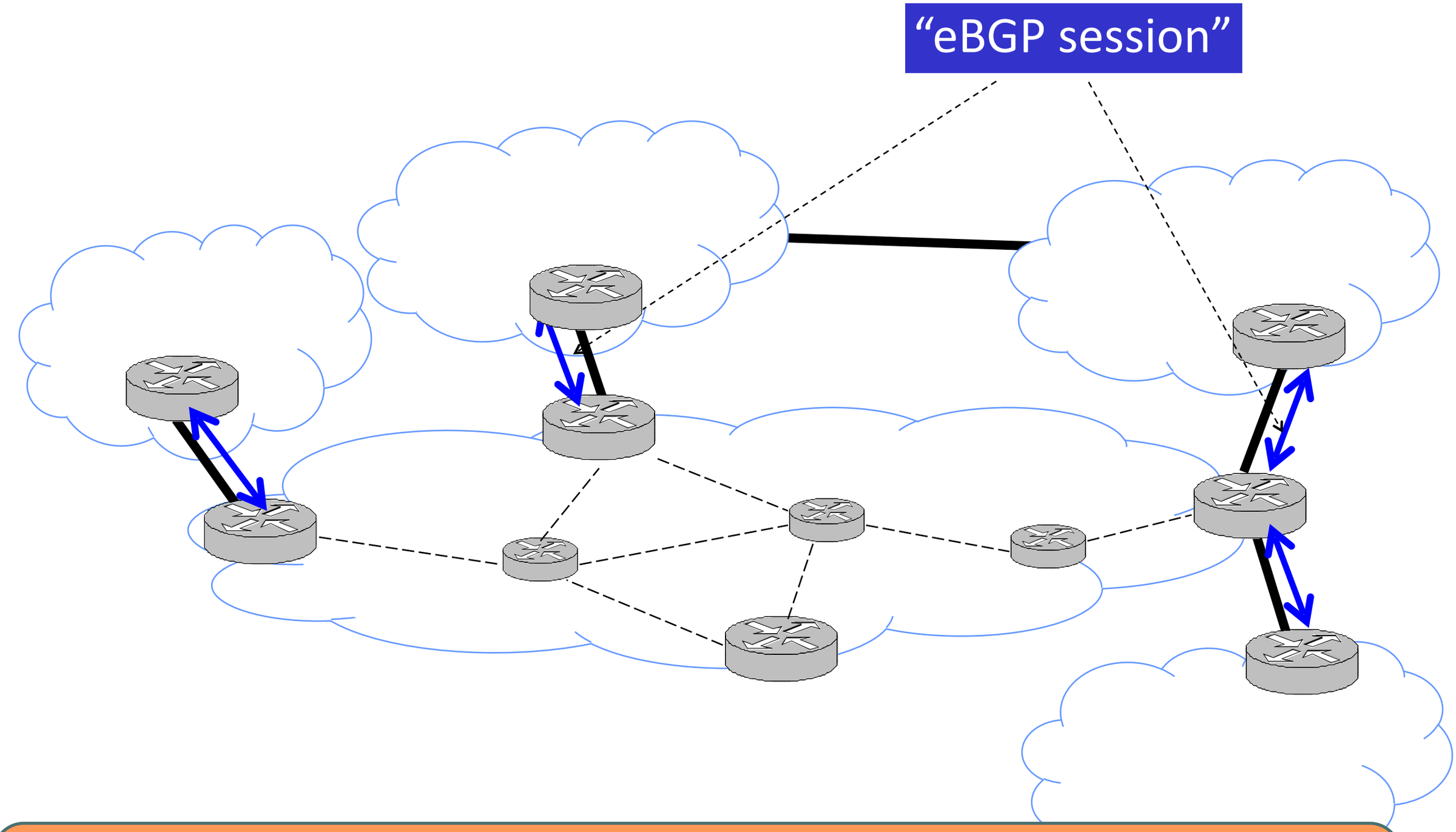
What Does “speak BGP” Mean?

- Implement the BGP Protocol Standard
  - Internet Engineering Task Force (IETF) RFC 4271

- Specifies what messages to exchange with other BGP “speakers”
  - Message types (e.g. route advertisements, updates)
  - Message syntax

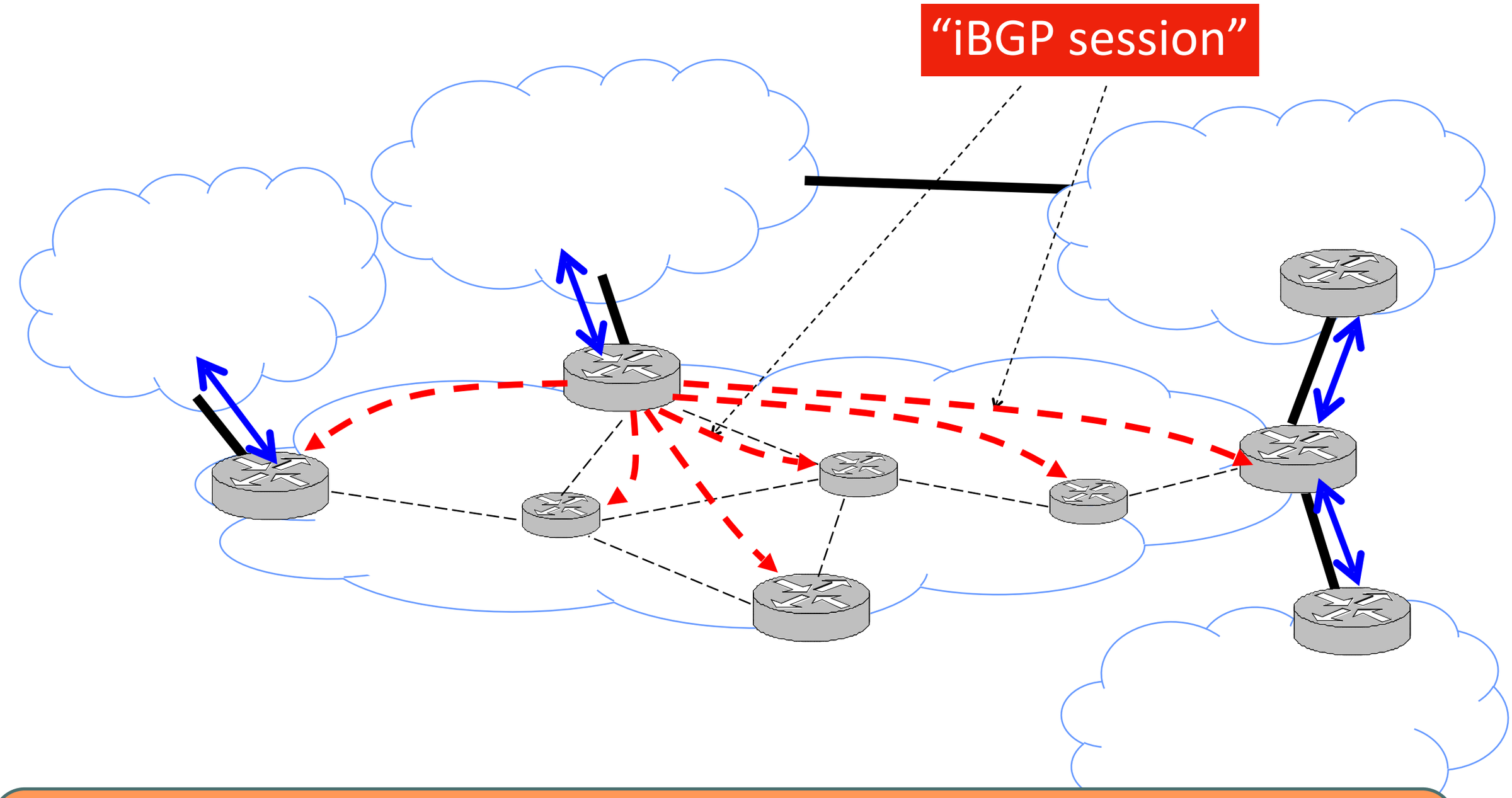
- Specifies how to process these messages
  - When you receive a BGP update, do x
  - Follows BGP state machine in the protocol spec and policy decisions, etc.

BGP Sessions



A border router speaks BGP with
border routers in other ASes

BGP Sessions



A border router speaks BGP with other (interior and border) routers in its own AS

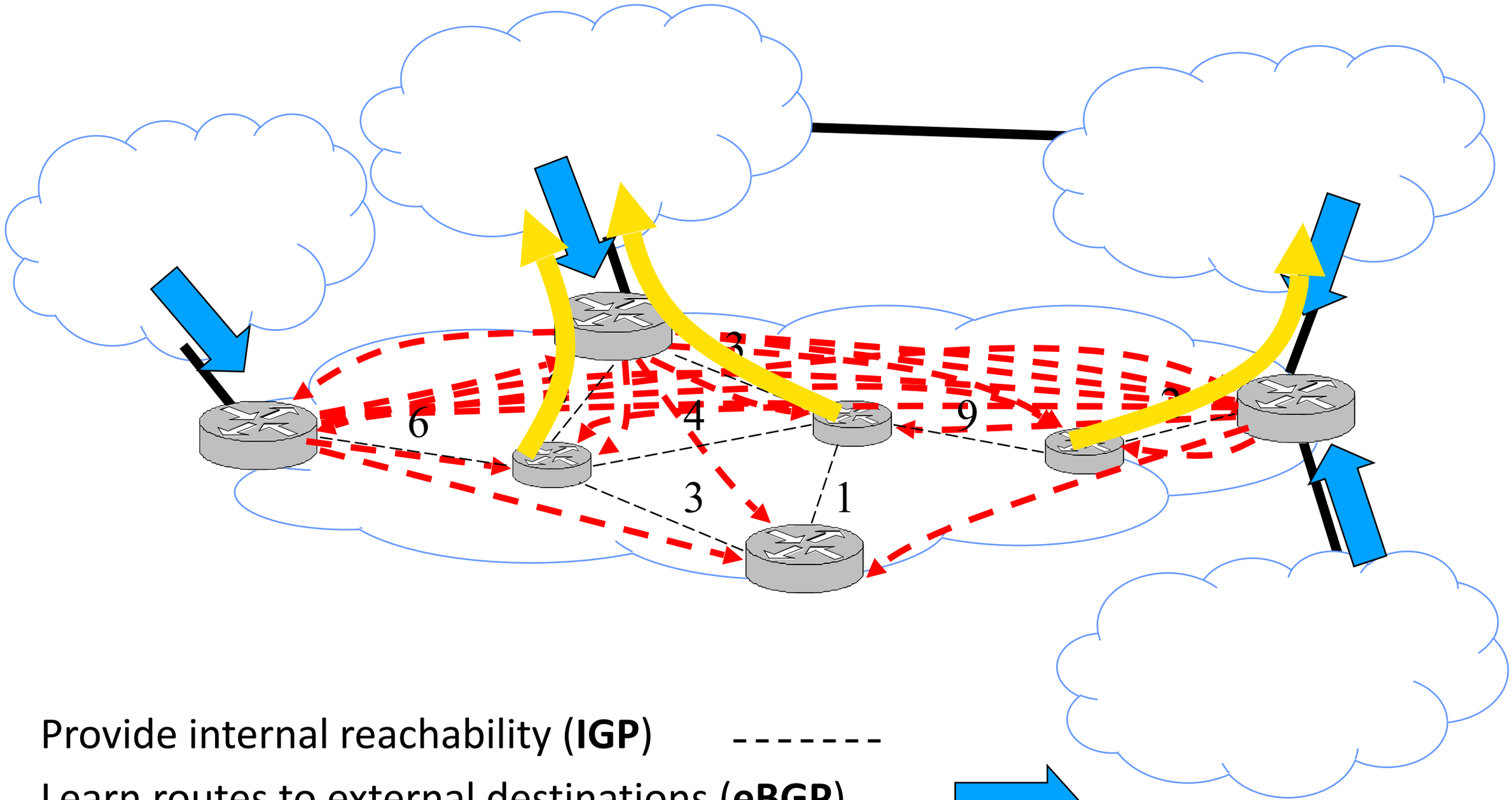
eBGP, iBGP, IGP

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

- iBGP: BGP sessions between border routers and other routers within the same AS
  - Distribute externally learned routes internally

- IGP: Interior Gateway Protocol = Intradomain routing protocol
  - Provides internal reachability
  - e.g. OSPF, RIP

Putting the Pieces Together



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

- **Open**
  - Establishes BGP session

- **Update**
  - Inform neighbor of new routes
  - Inform neighbor of old routes that become inactive

- **Keepalive**
  - Inform neighbor that connection is still viable

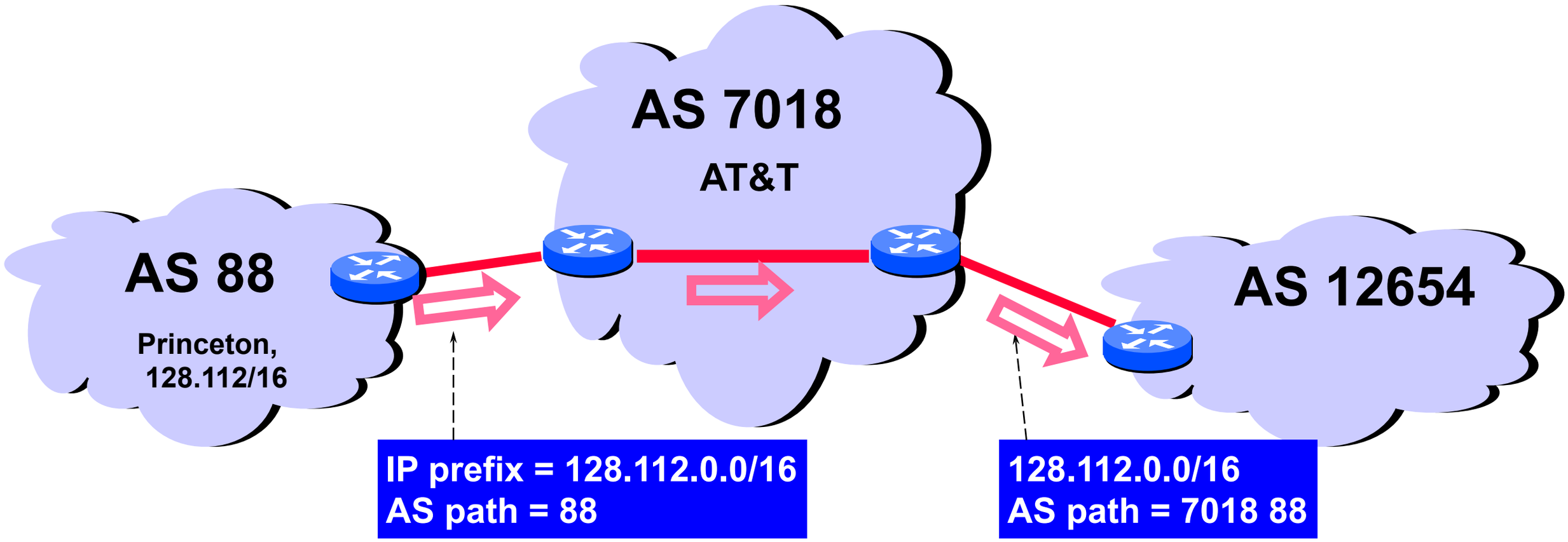
Route Updates

- Format: *<IP prefix: route attributes>*
- Two kinds of updates:
  - Announcements: new routes or changes to existing routes
  - Withdrawals: remove routes that no longer exist

- Route Attributes
  - Describe routes, used in 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
  - Many standardized attributes in BGP

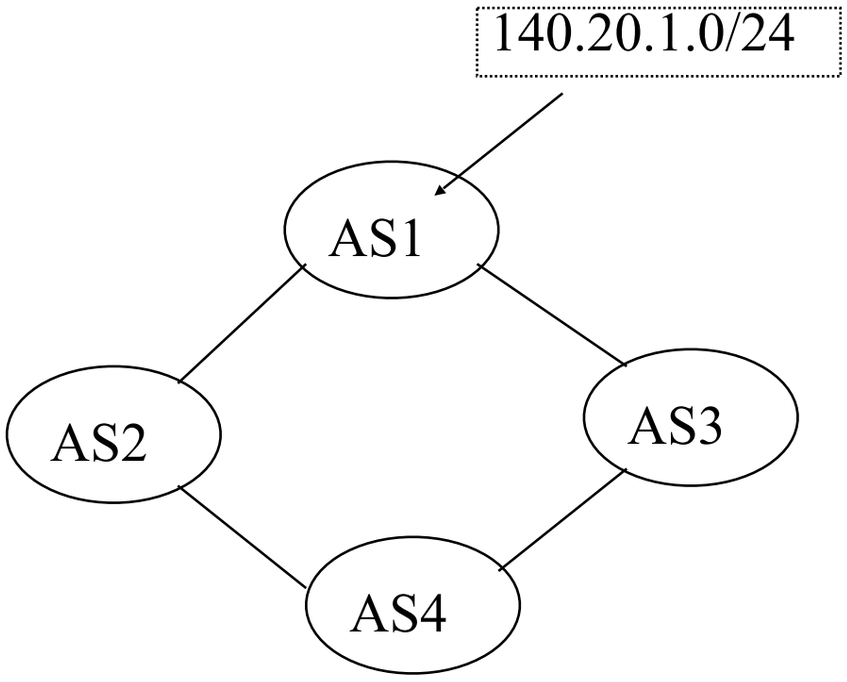
Route Attributes (1): ASPATH

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



Route 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

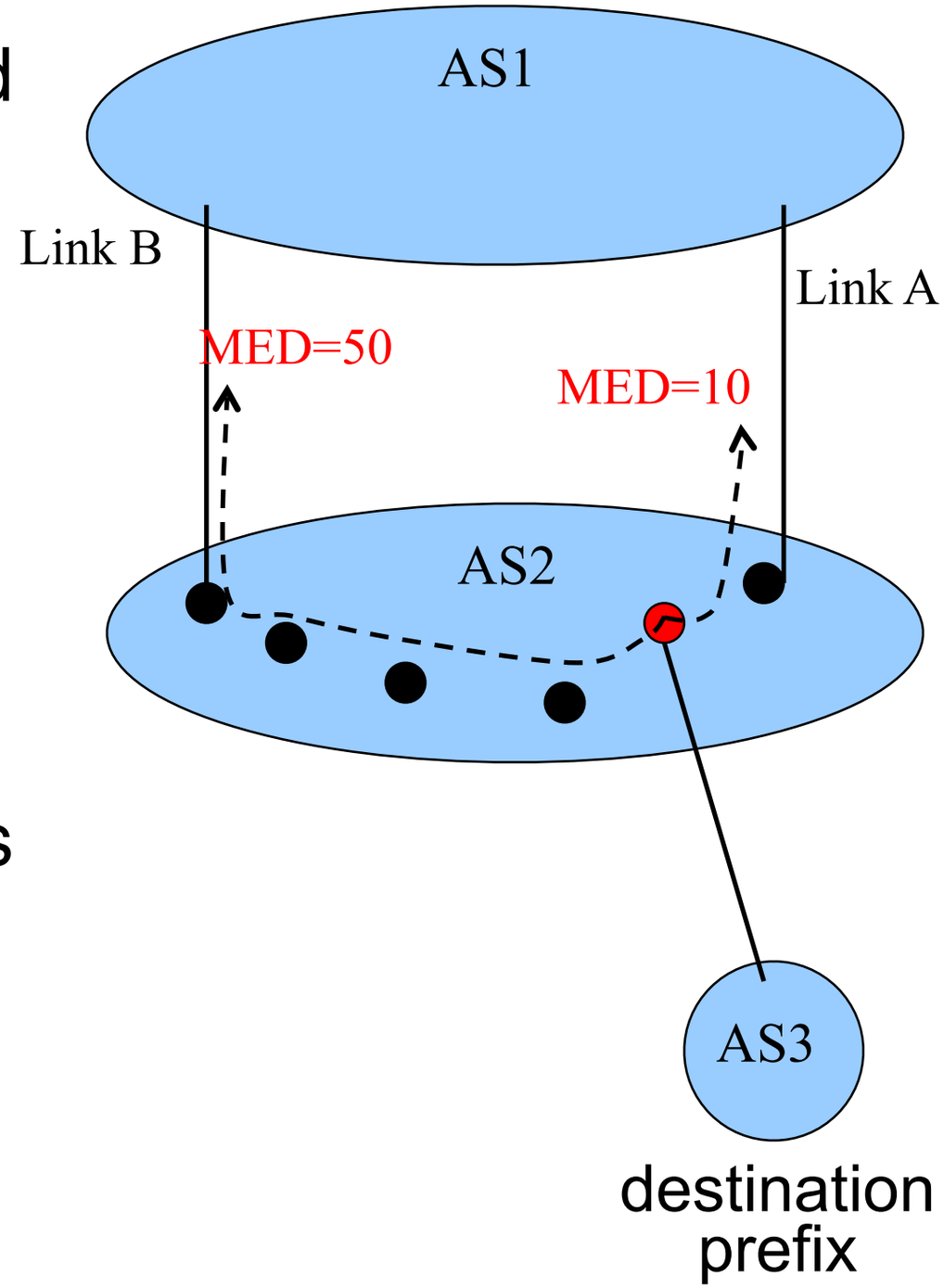


BGP table at AS4:

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

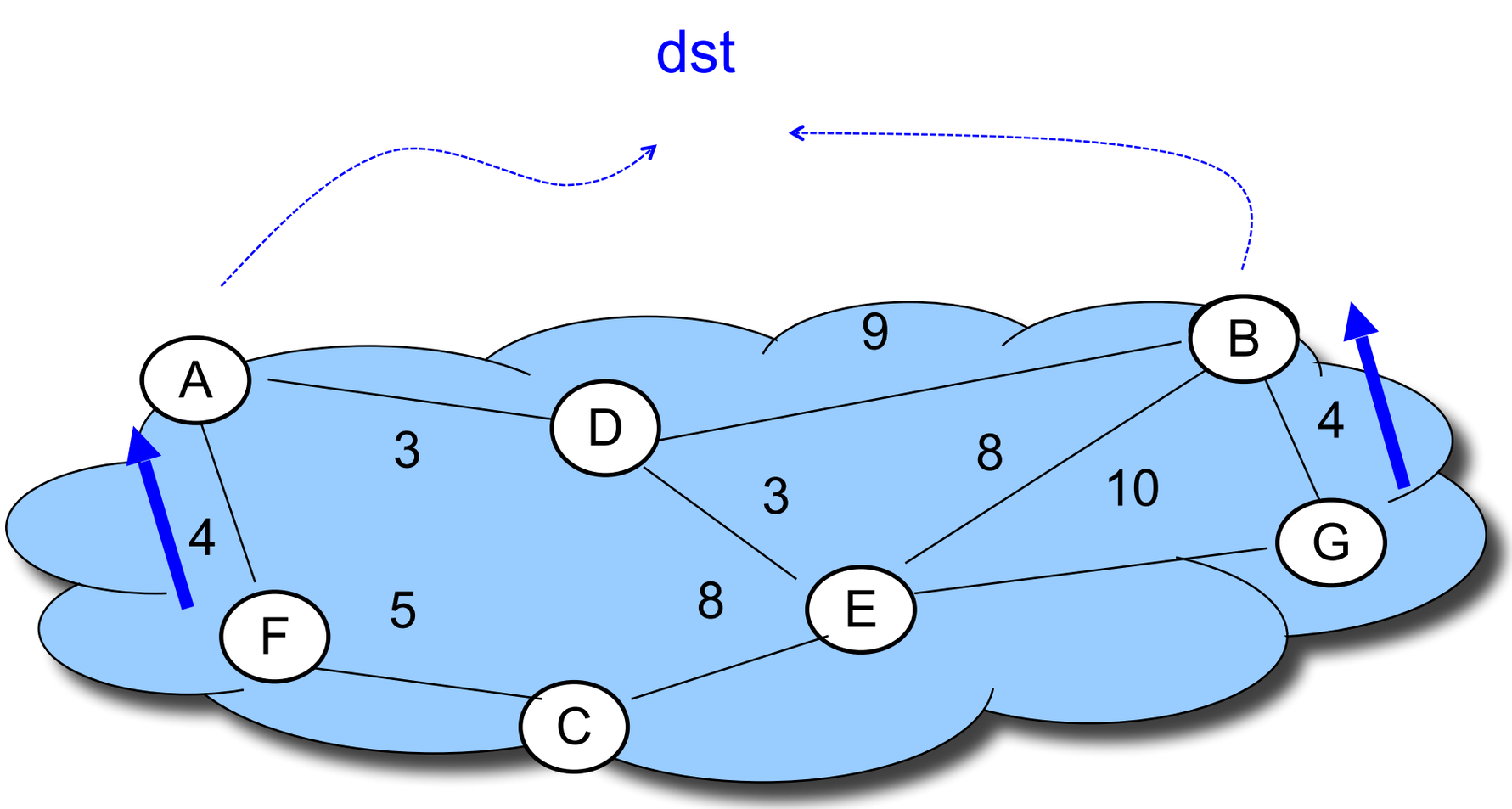
Route Attributes (3) : MED

- "Multi-Exit Discriminator"
- Used when ASes are interconnected via two or more links
  - Specifies 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



Route Attributes (4): IGP Cost

- Used for hot-potato routing
  - Each router selects the closest egress point based on the path cost in intra-domain protocol



Using Attributes

- Rules for route selection in priority order

1. Make or save money (send to customer > peer > provider)
2. Maximize performance (smallest AS path length)
3. Minimize use of my network bandwidth (“hot potato”)
4. …

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	iBGP path	Lowest IGP cost to next hop (egress router)
6	Router ID	Smallest next-hop router's IP address as tie-breaker

BGP Update Processing

Open ended programming.
Constrained only by vendor configuration language

Control plane

BGP Updates

Apply Import Policies

Best Route Selection

Best Route Table

Apply Export Policies

BGP Updates

Data plane

forwarding Entries

Data packets

IP Forwarding Table

Data packets

BGP Issues

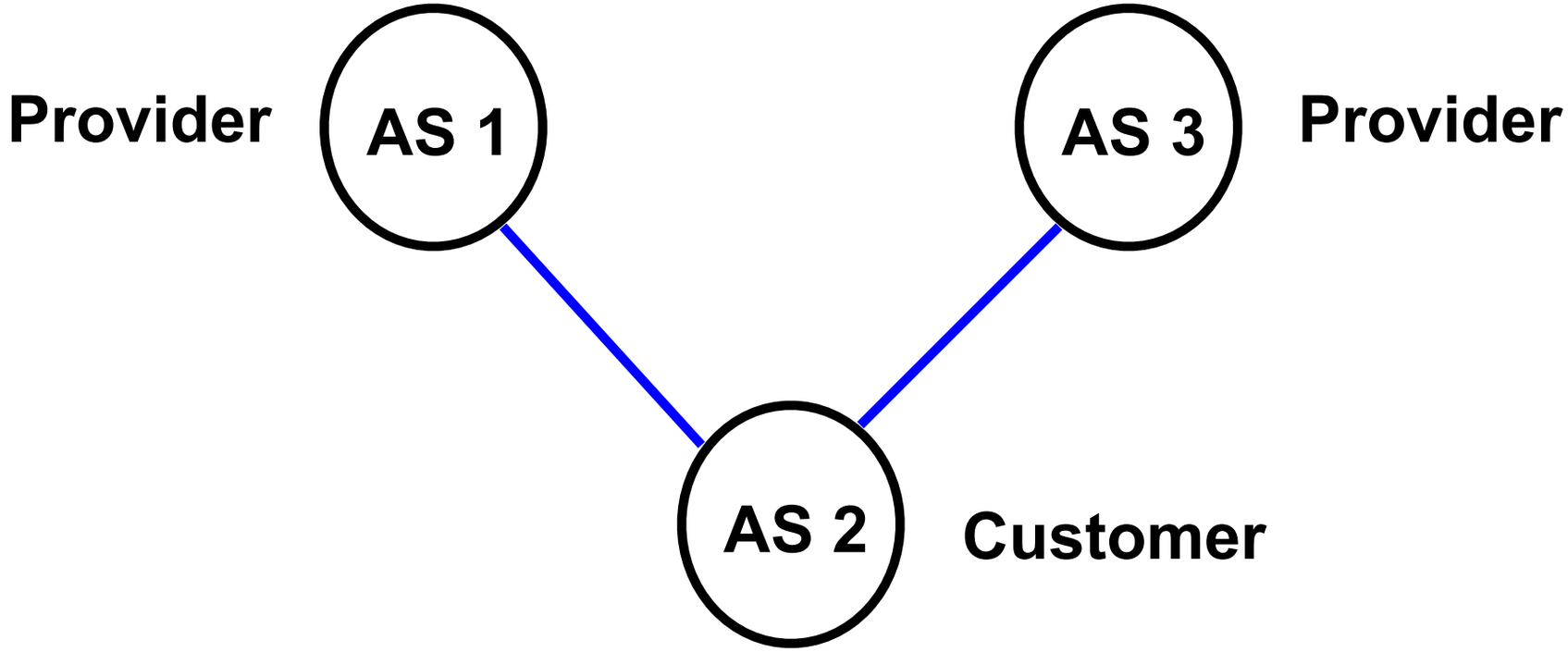
BGP: Issues

- Reachability
- Security
- Convergence
- Performance
- Anomalies

Reachability

- In normal routing, if graph is connected then reachability is assured

- With policy routing, this doesn't always hold



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*

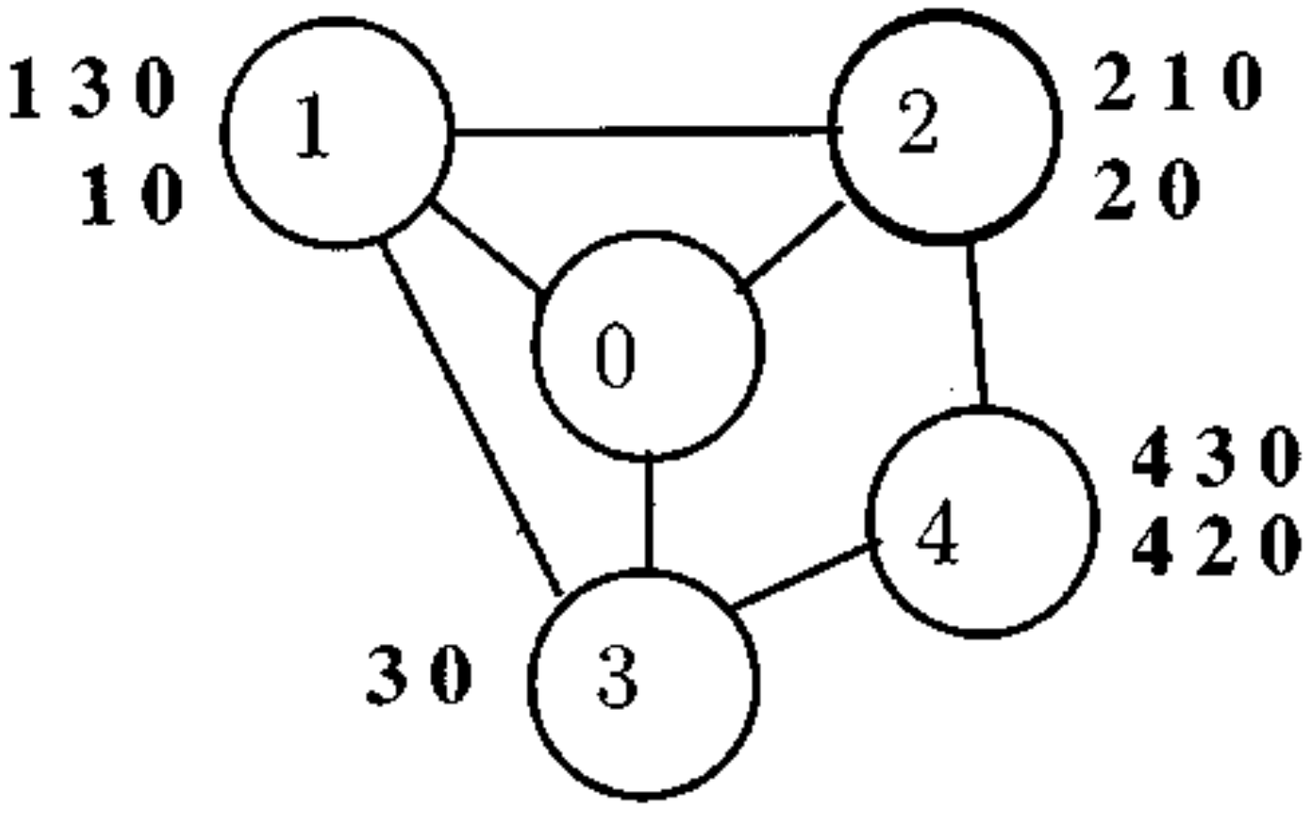
- But…
  - AS may forward packets along a route different from what is advertised
    - Tell customers about a fictitious short path…
    - Much harder to fix!

Convergence

- If all AS policies follow Gao-Rexford rules,
  - Then BGP is guaranteed to converge (safety)

- For arbitrary policies, BGP may fail to converge!

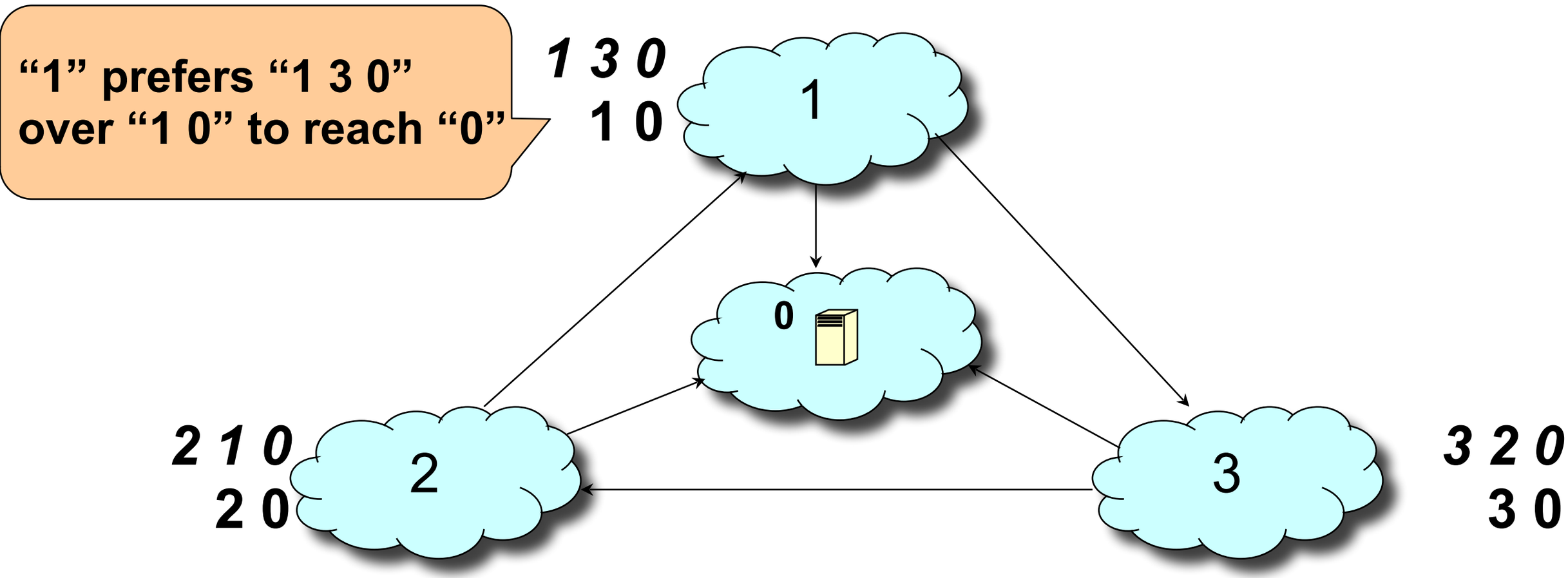
BGP Example (All good)



GOOD GADGET

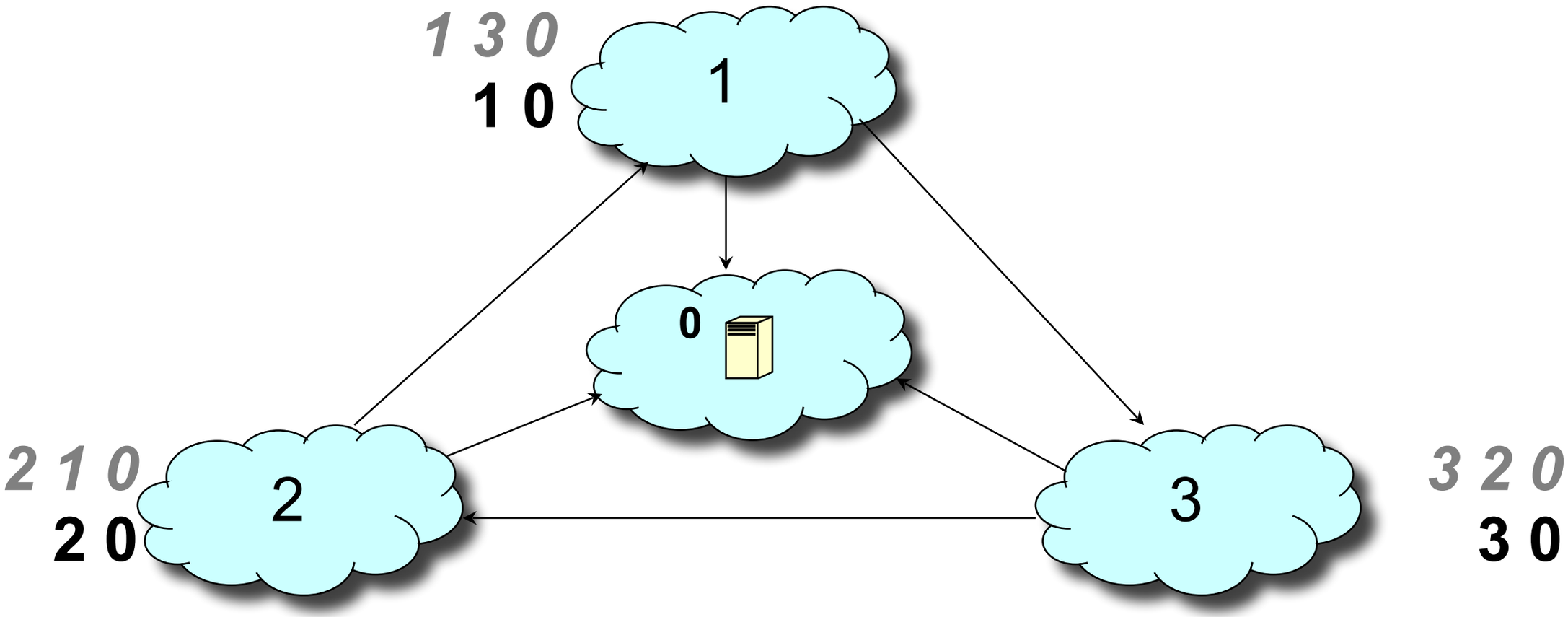
	1	2	3	4
R1	10	20	**30**	-
R2	10	20	**30**	430
R3	130	20	30	430

Example of Policy Oscillation



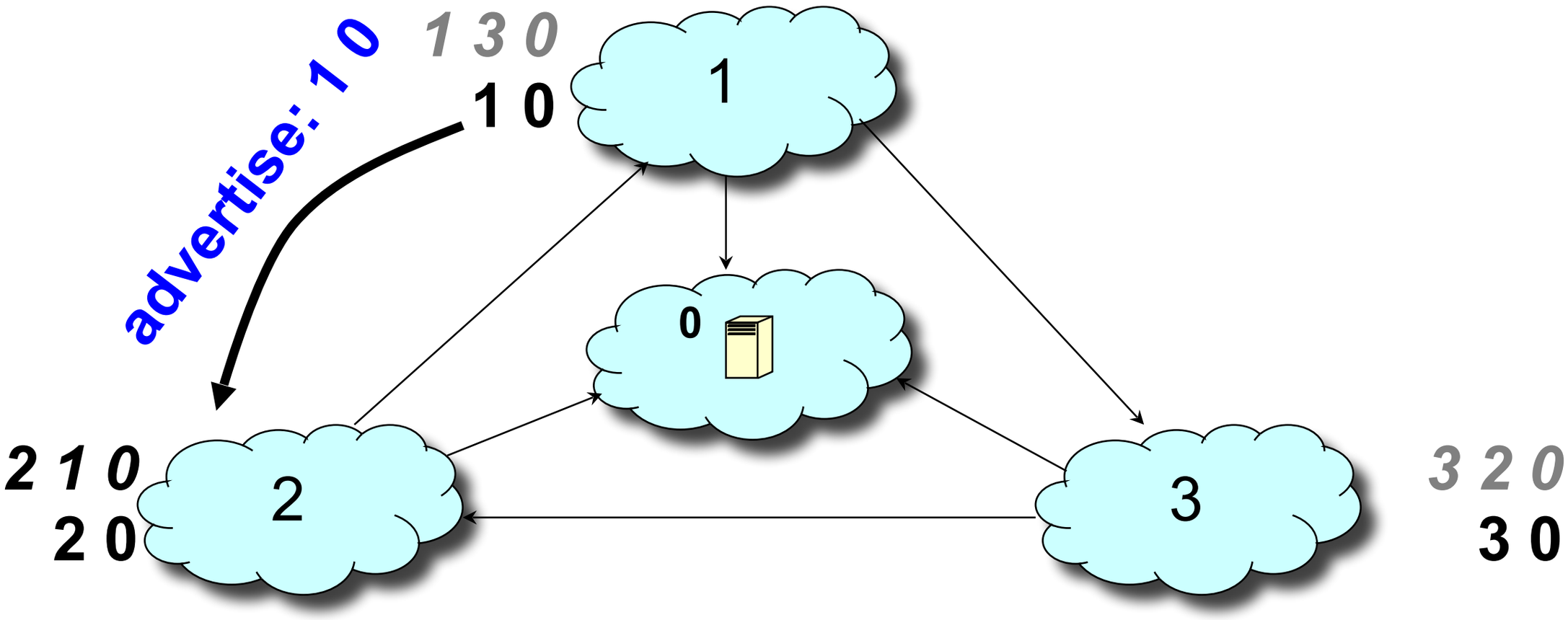
Step-by-step Policy Oscillation

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

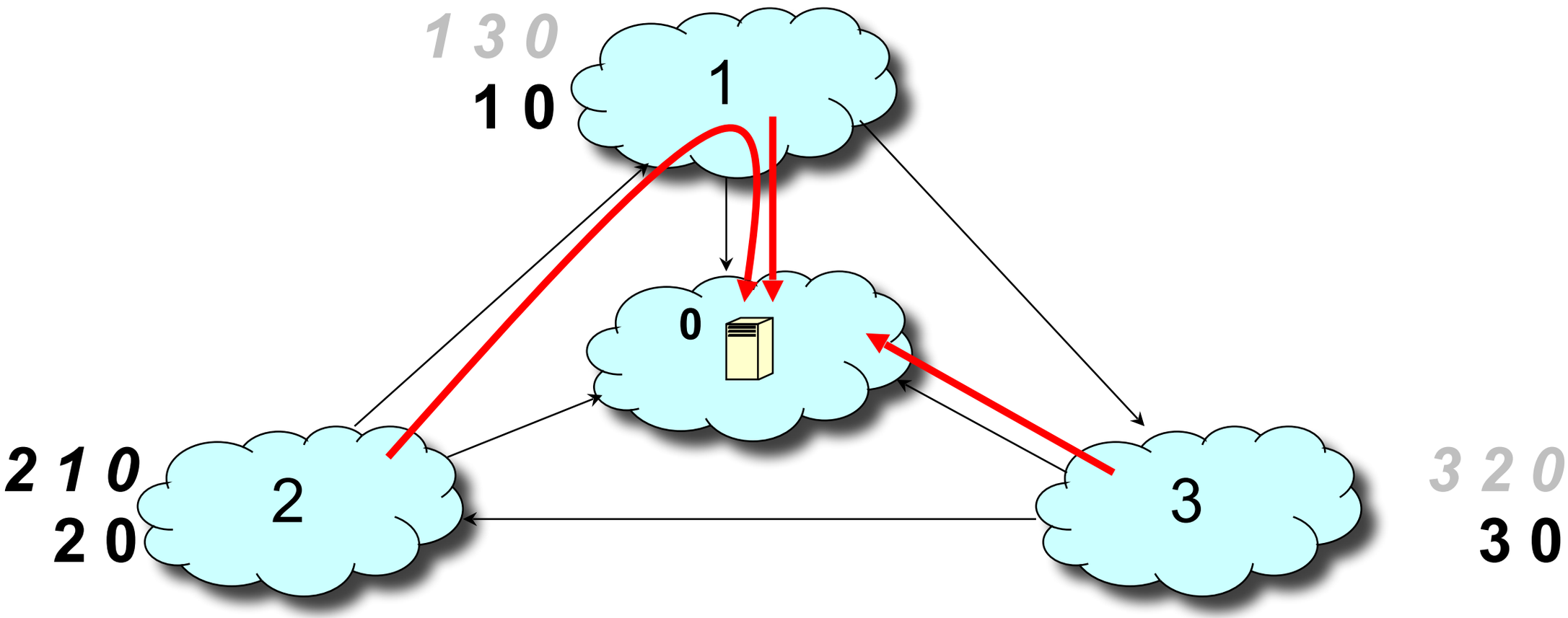


Step-by-step Policy Oscillation

1 **advertises** its path 1 0 to 2

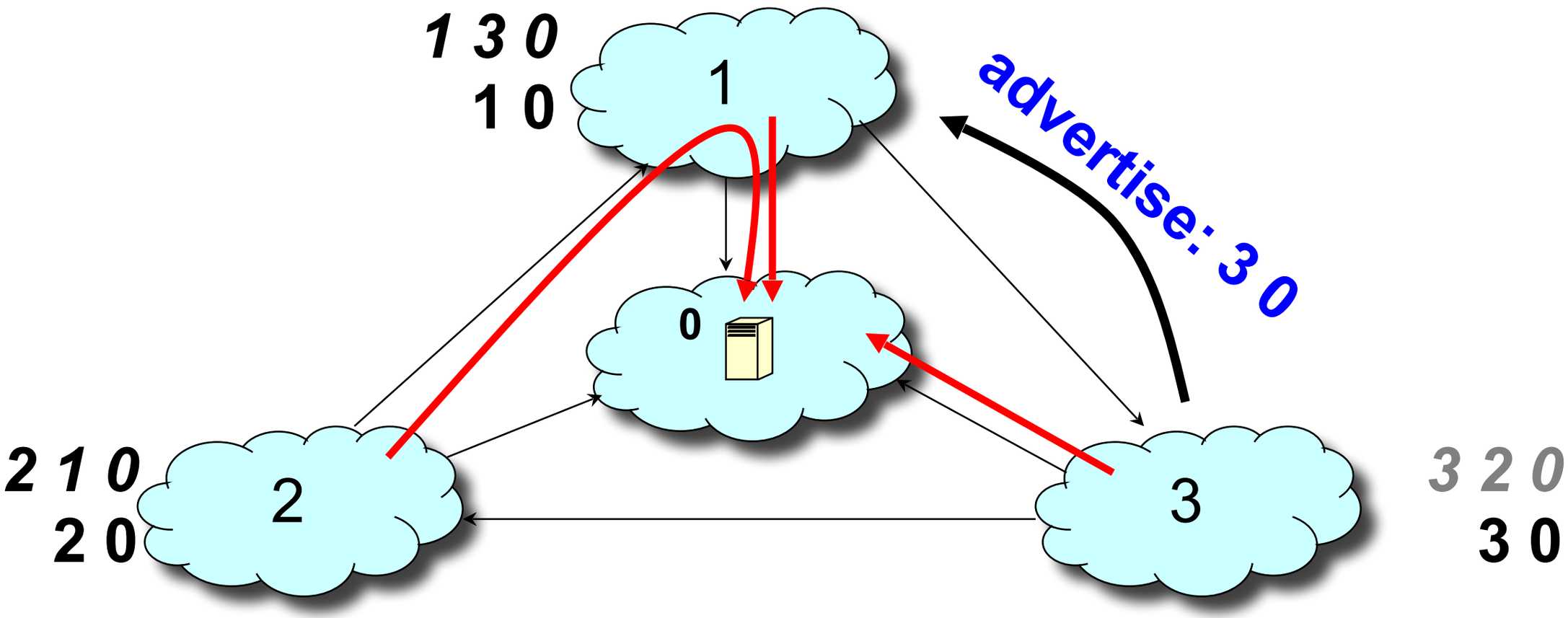


Step-by-step Policy Oscillation

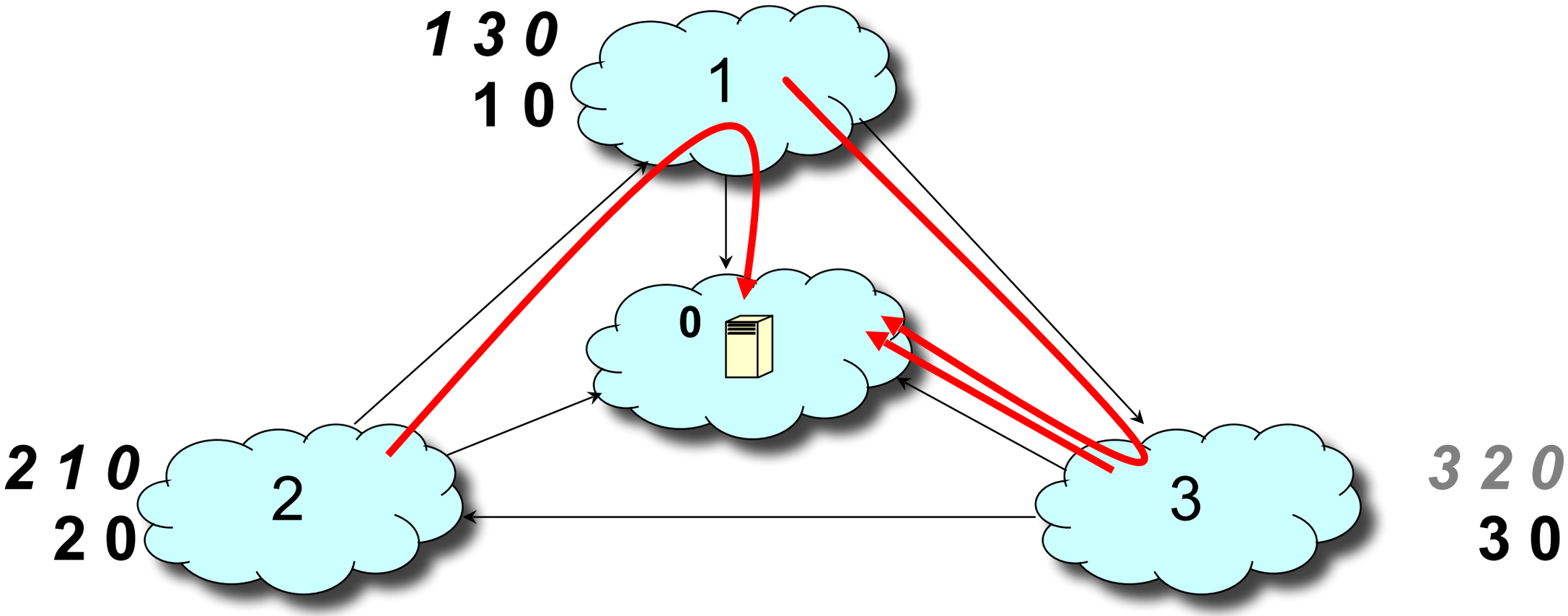


Step-by-step Policy Oscillation

3 **advertises** its path 3 0 to 1

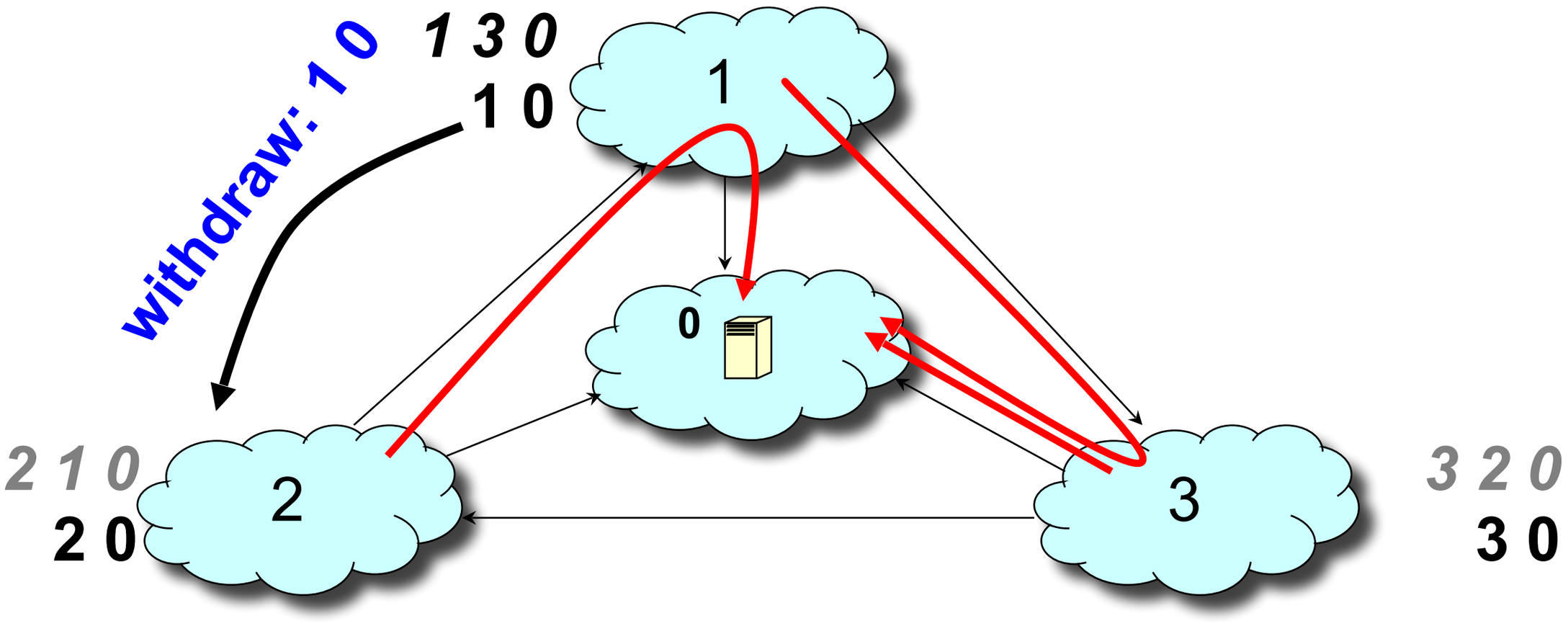


Step-by-step Policy Oscillation

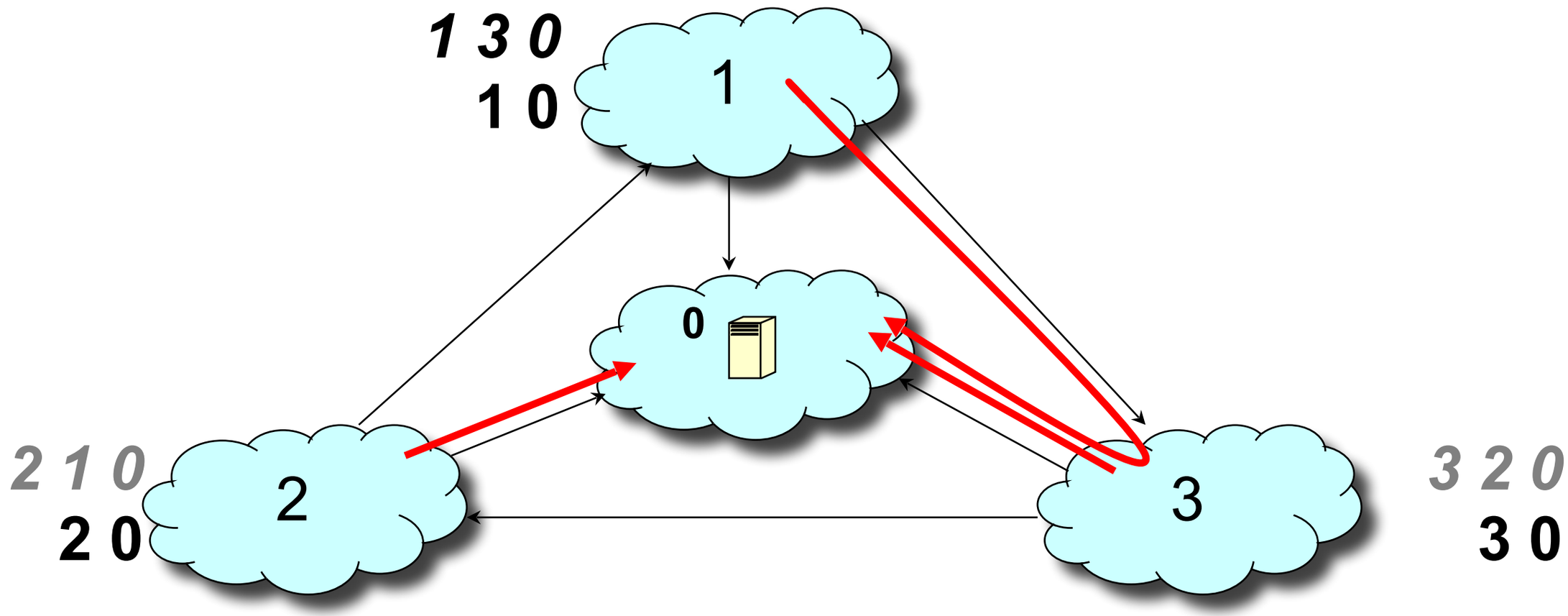


Step-by-step Policy Oscillation

1 **withdraws** its path 1 0 from 2

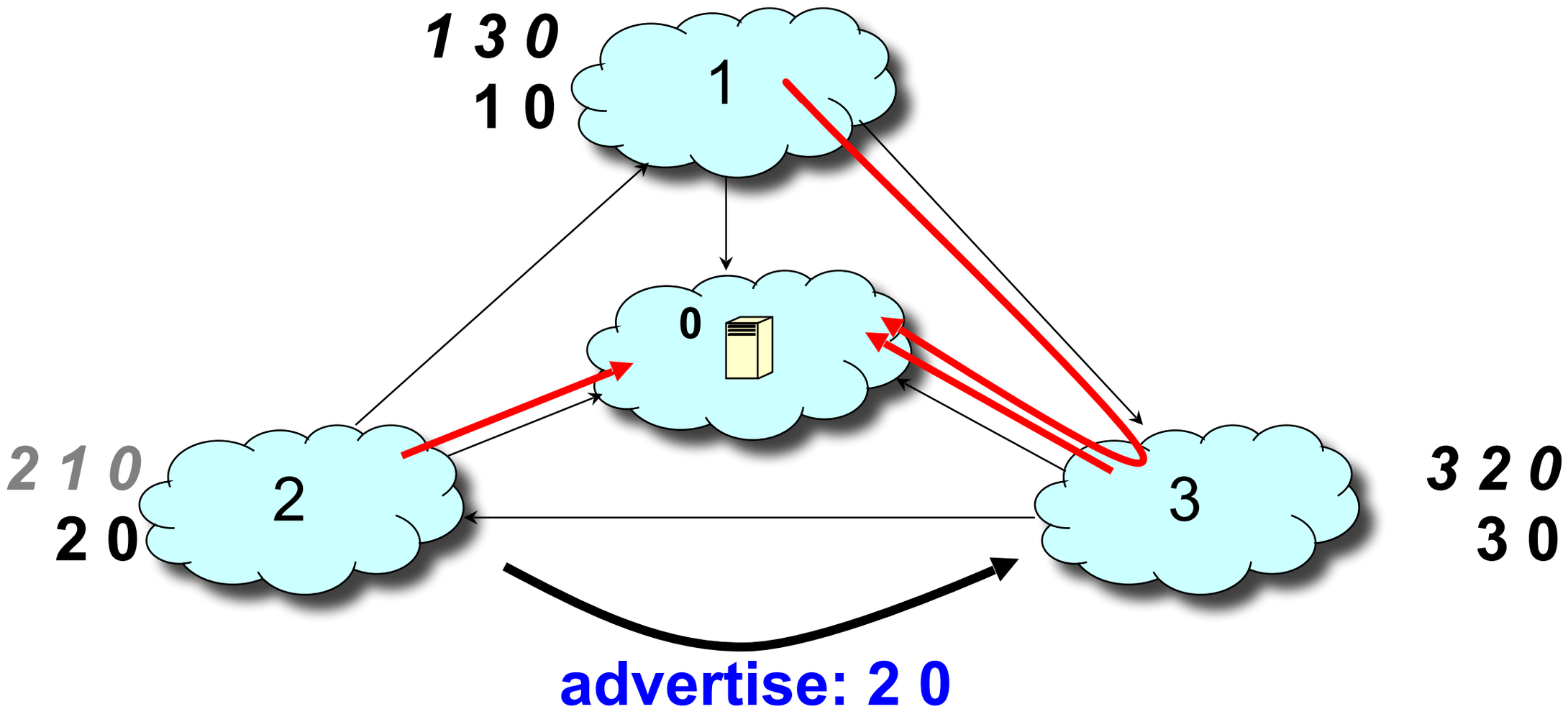


Step-by-step Policy Oscillation

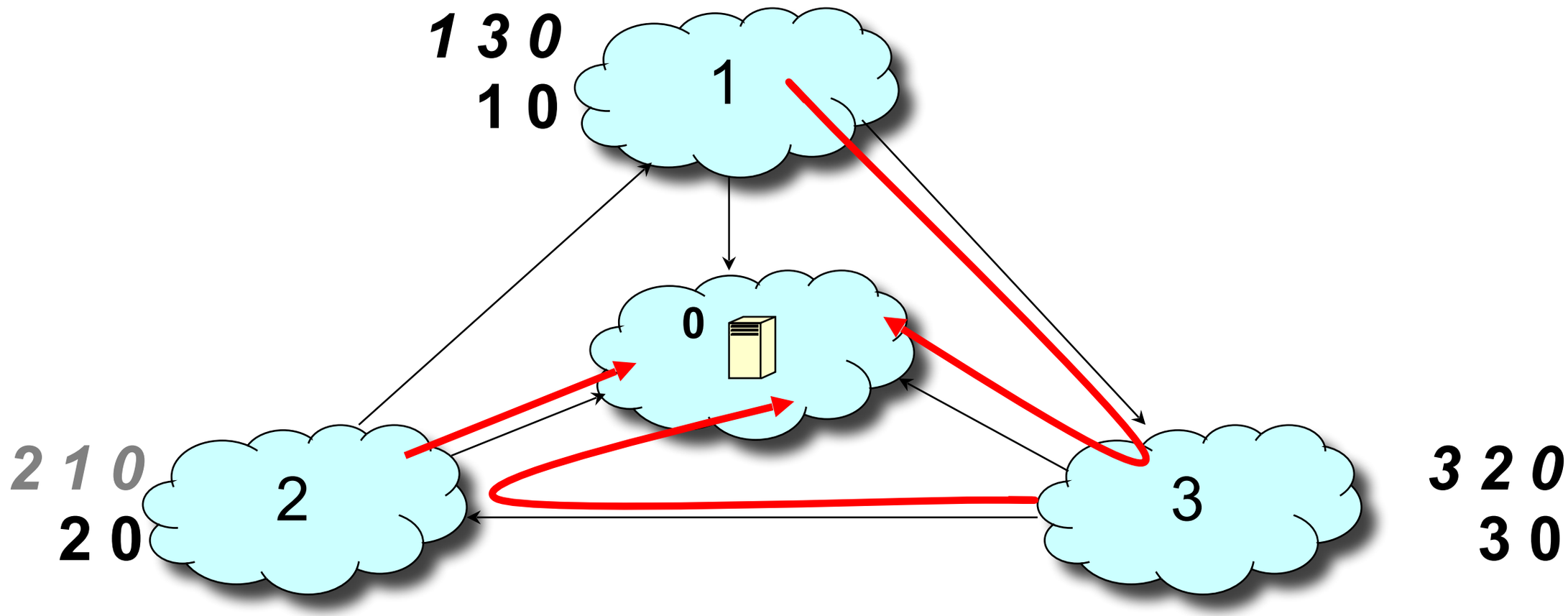


Step-by-step Policy Oscillation

2 **advertises** its path 2 0 to 3

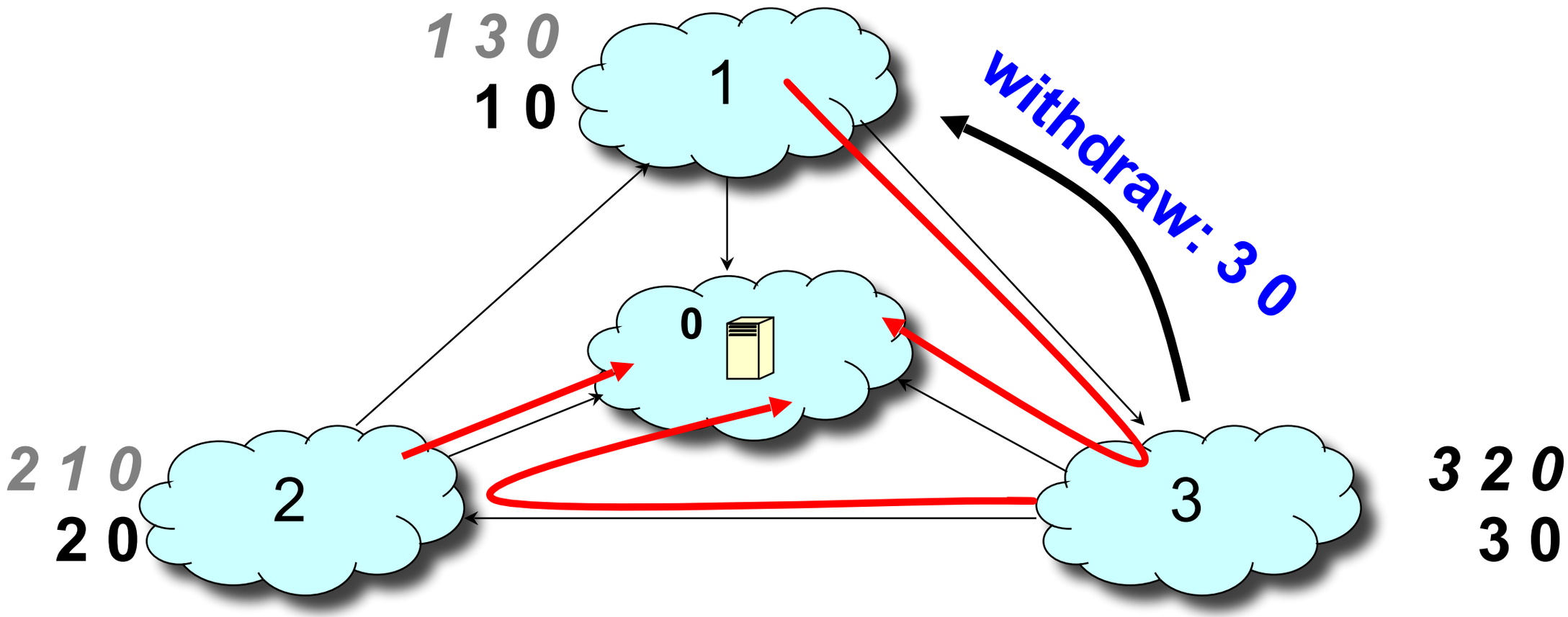


Step-by-step Policy Oscillation

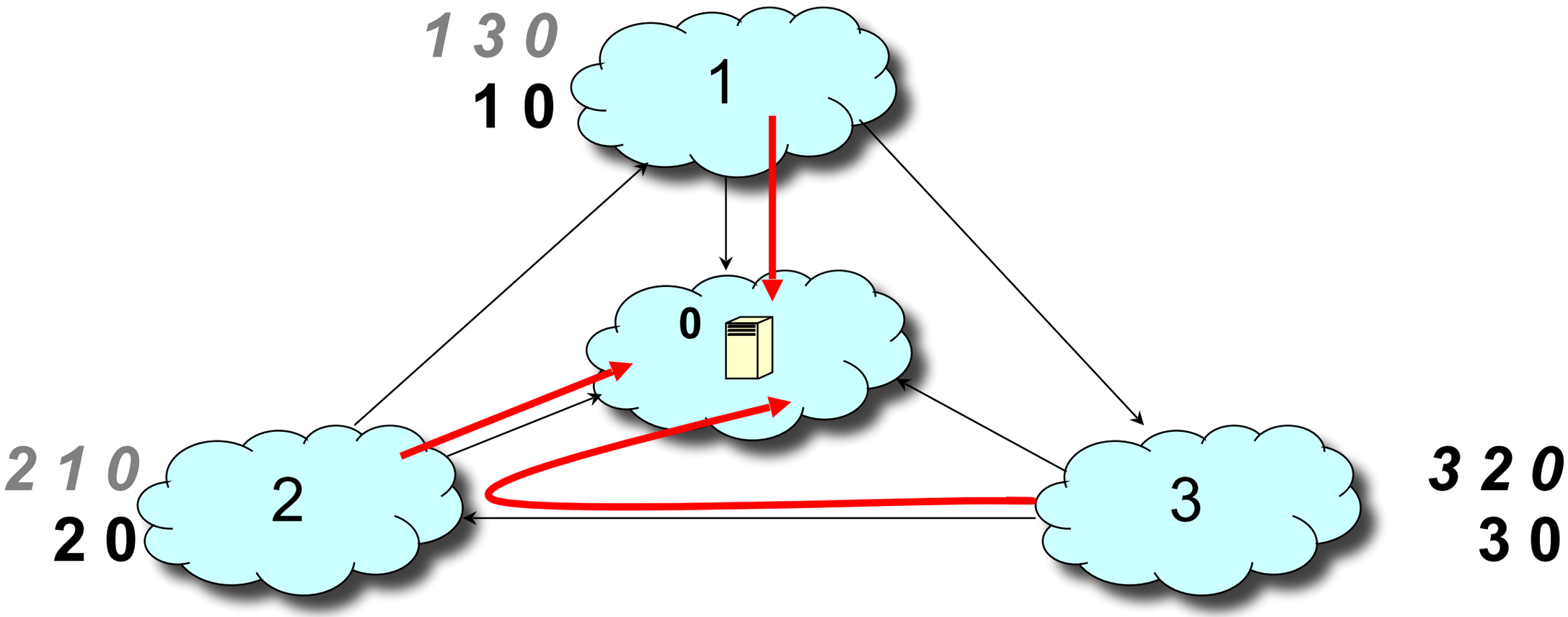


Step-by-step Policy Oscillation

3 **withdraws** its path 3 0 from 1

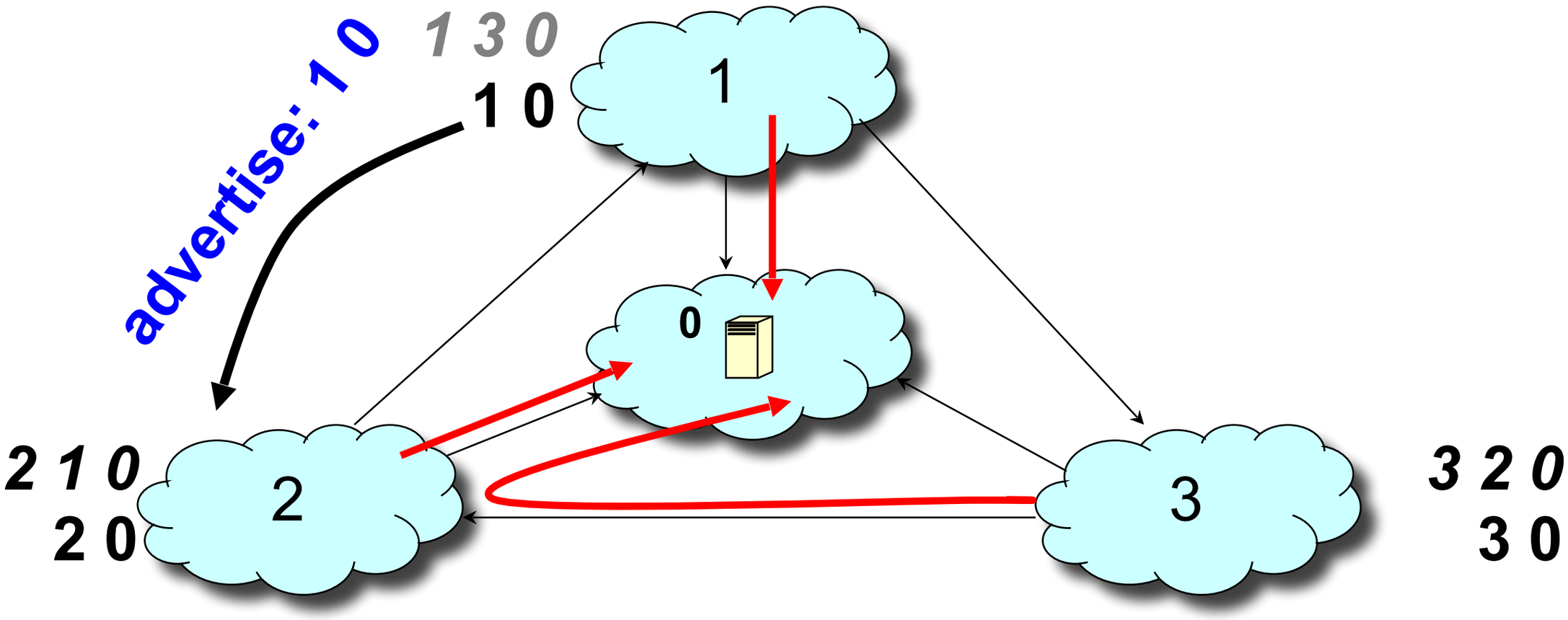


Step-by-step Policy Oscillation

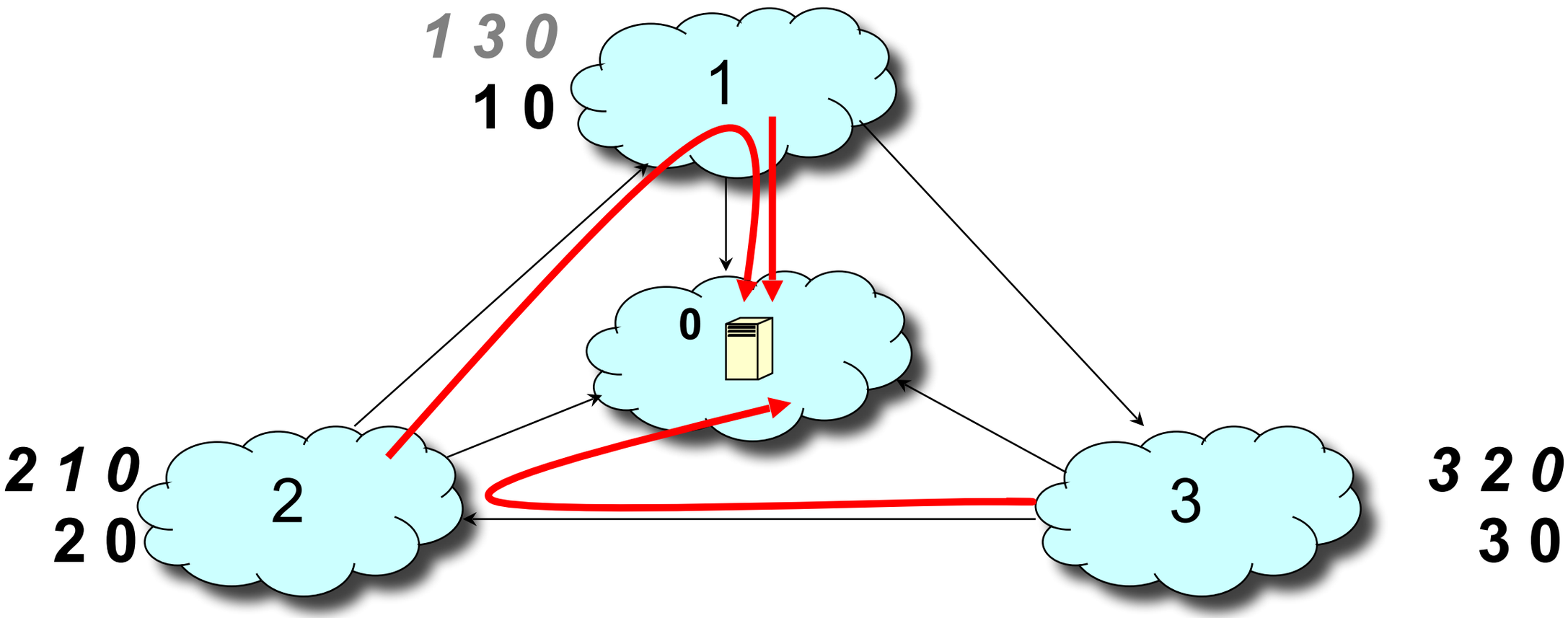


Step-by-step Policy Oscillation

1 **advertises** its path 1 0 to 2

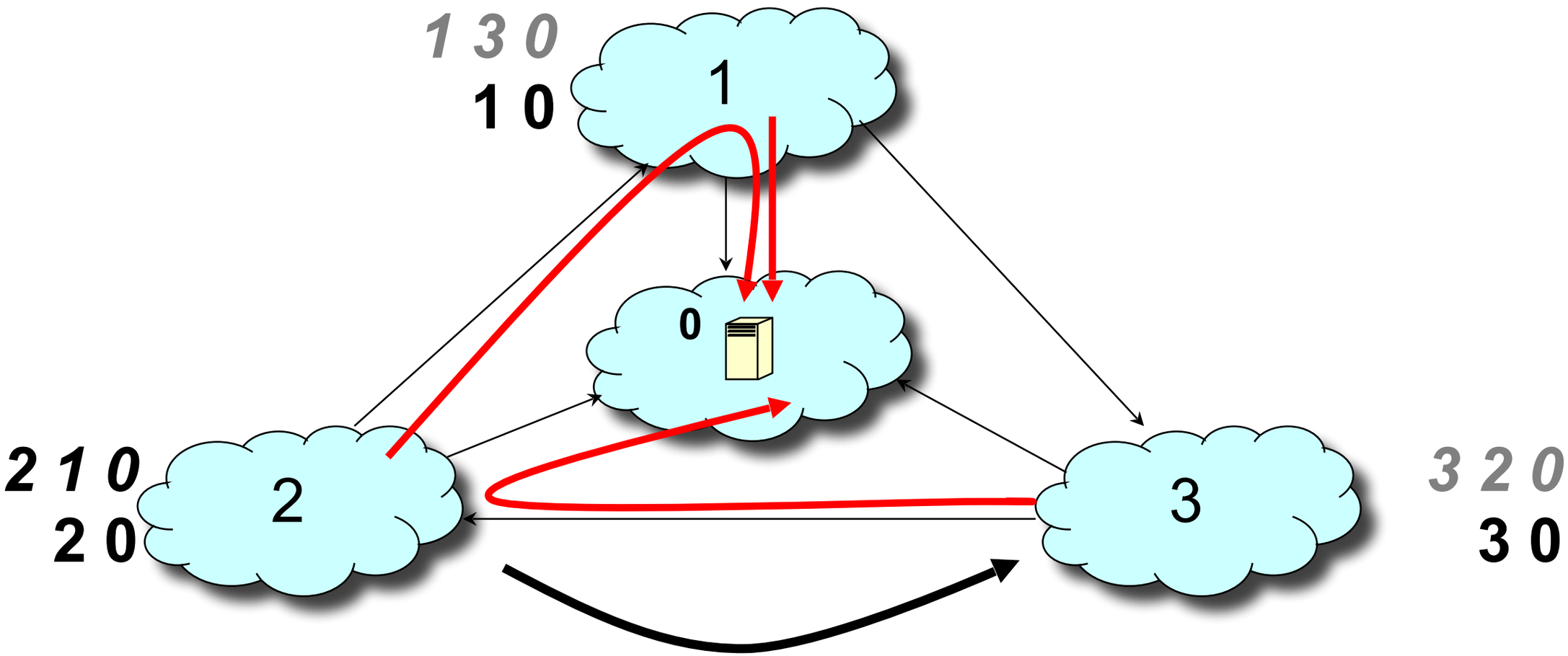


Step-by-step Policy Oscillation

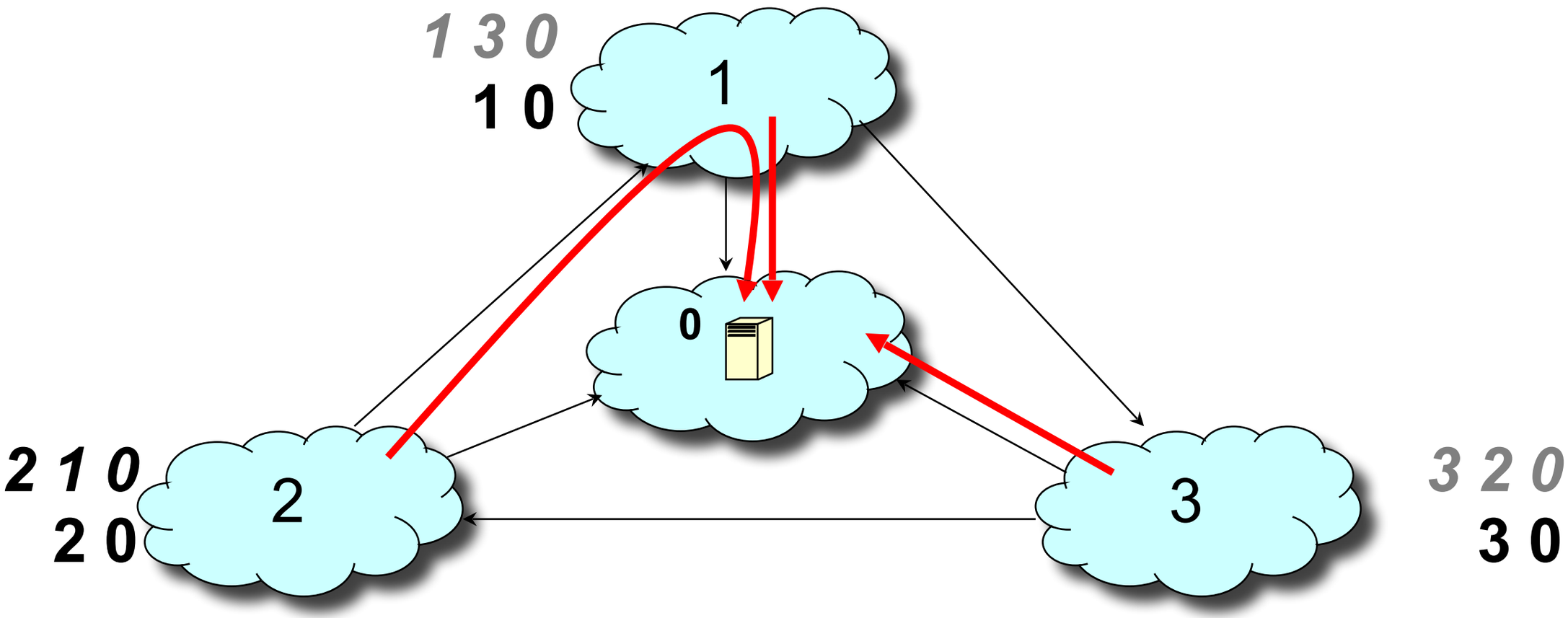


Step-by-step Policy Oscillation

2 **withdraws** its path 2 0 from 3

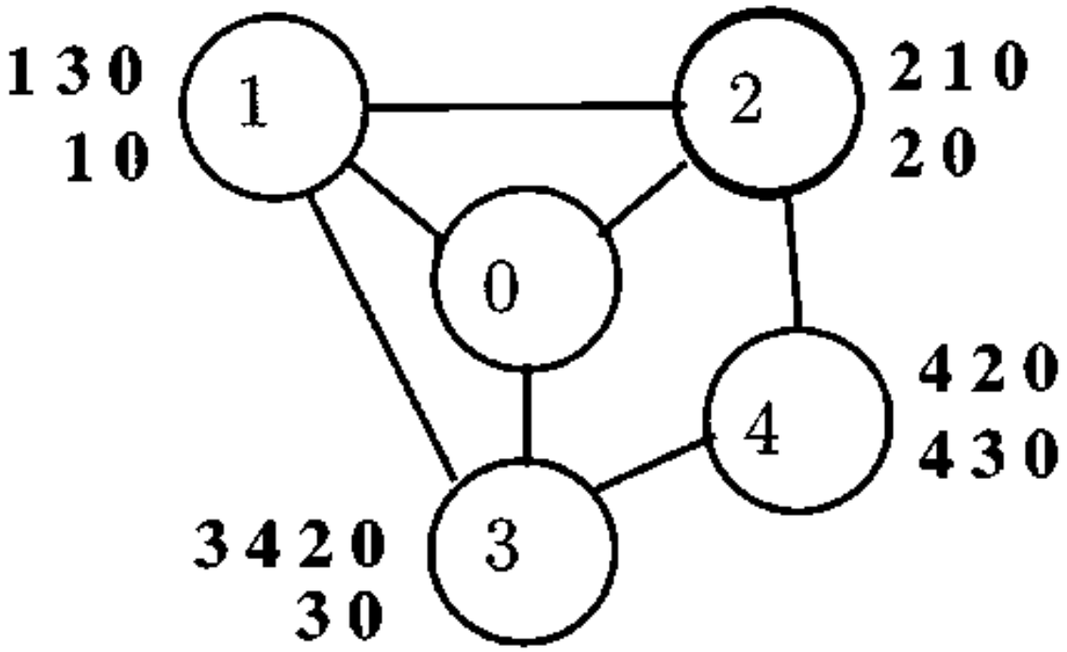


Step-by-step Policy Oscillation



We are back to where we started!

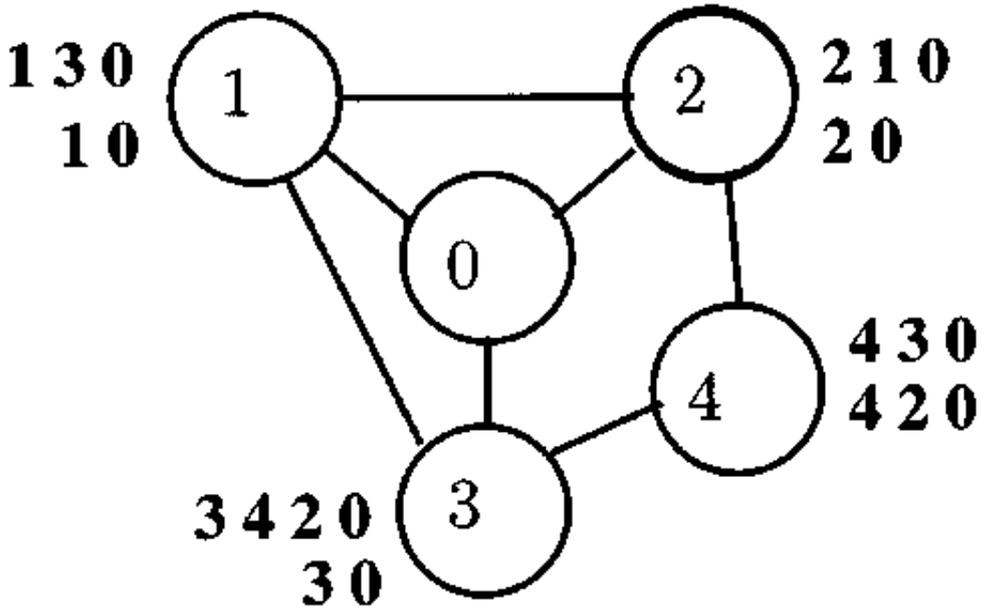
BGP Example (Persistent Loops)



BAD GADGET

	1	2	3	4
R1	10	**20**	30	-
R2	10	20	30	**420**
R3	**10**	20	3420	420
R4	10	**210**	3420	420
R5	10	210	3420	-
R6	10	210	**30**	-
R7	**130**	210	30	-
R8	130	**20**	30	-
R9	130	20	30	**420**
R10	130	20	**3420**	420
R11	10	20	3420	420

BGP Example (Bad bad bad)



NAUGHTY GADGET

	1	2	3	4
R1	10	20	**30**	-
R2	10	20	**30**	430
R3	130	20	30	430

	1	2	3	4
R1	10	**20**	30	-
R2	10	20	30	**420**
R3	**10**	20	3420	420
R4	10	**210**	3420	420
R5	10	210	3420	**-**
R6	10	210	**30**	-
R7	**130**	210	30	-
R8	130	**20**	30	-
R9	130	20	30	**420**
R10	130	20	**3420**	420
R11	10	20	3420	420

Convergence

- If all AS policies follow Gao-Rexford rules,
  - Then BGP is guaranteed to converge (safety)

- For arbitrary policies, BGP may fail to converge!

- Why should this trouble us?

Performance Non-Issues

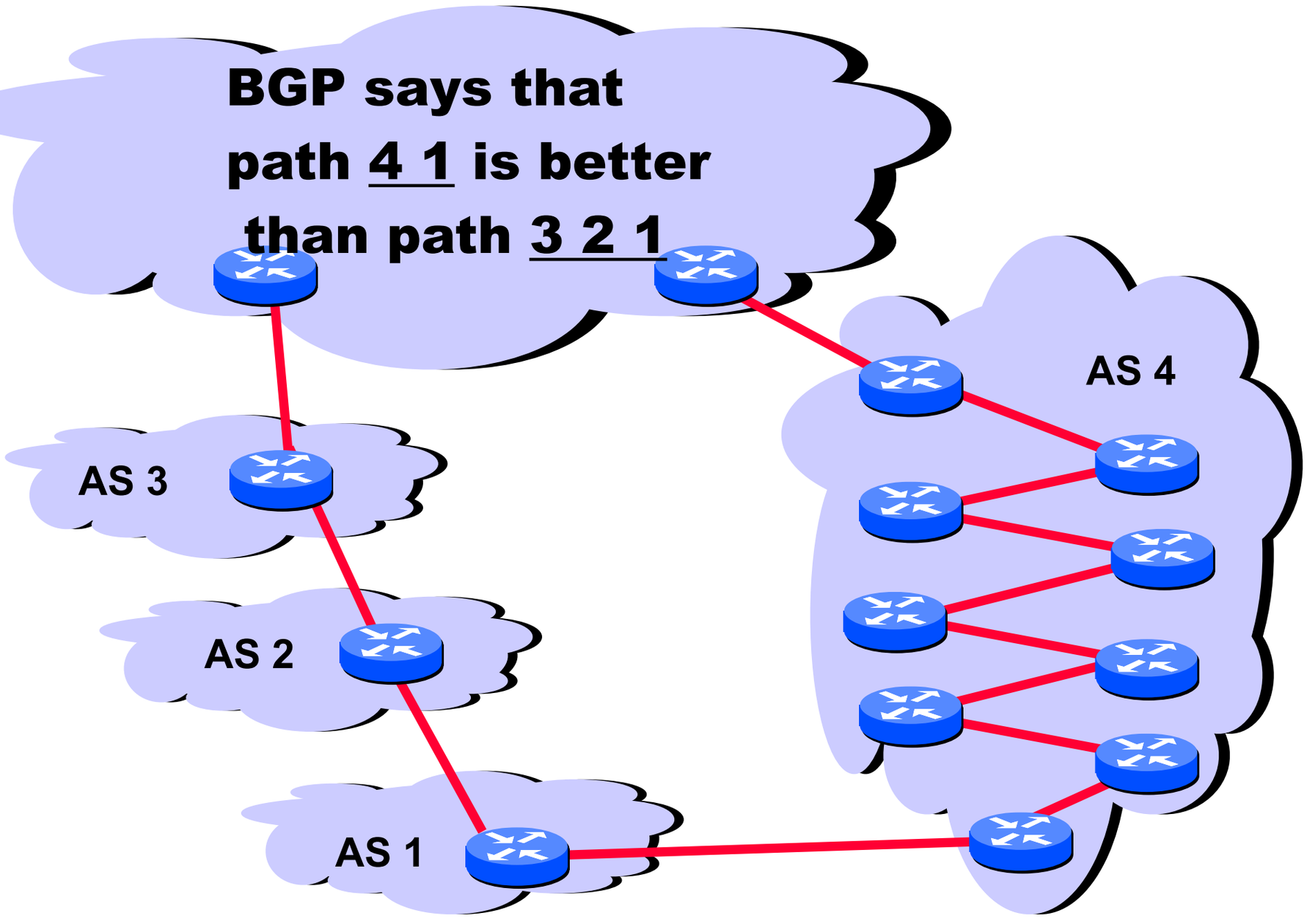
- Internal Routing
  - Domains typically use “hot potato” routing
  - Not always optimal, but economically expedient

- Policy not about performance
  - So policy-chosen paths aren’t shortest

- AS path length can be misleading
  - 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



Performance: Real Issue

Slow Convergence

- BGP outages are biggest source of Internet problems

- Labovitz et al. *SIGCOMM'97*
  - 10% of routes available less than 95% of the time
  - Less than 35% of routes available 99.99% of the time

- Labovitz et al. *SIGCOMM 2000*
  - 40% of path outages take 30+ minutes to repair

- But most popular paths are very stable