EECS 489
Computer Networks

Winter 2025

Mosharaf Chowdhury

Material with thanks to Aditya Akella, Sugih Jamin, Philip Levis, Sylvia Ratnasamy, Peter Steenkiste, and many other colleagues.

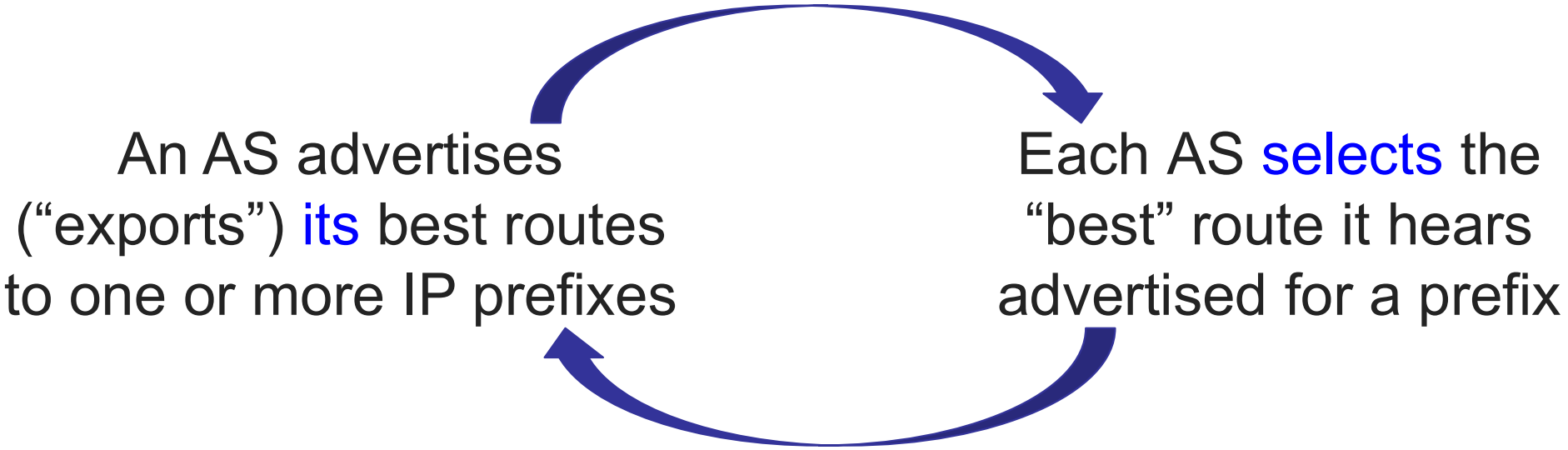
Agenda

- BGP basics
- BGP policies and how they are implemented
- BGP protocol details
- BGP issues in practice

Inter-domain 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
- BGP (Border Gateway Protocol) is the Inter-domain routing protocol
  - Implemented by AS border routers

BGP: Basic idea



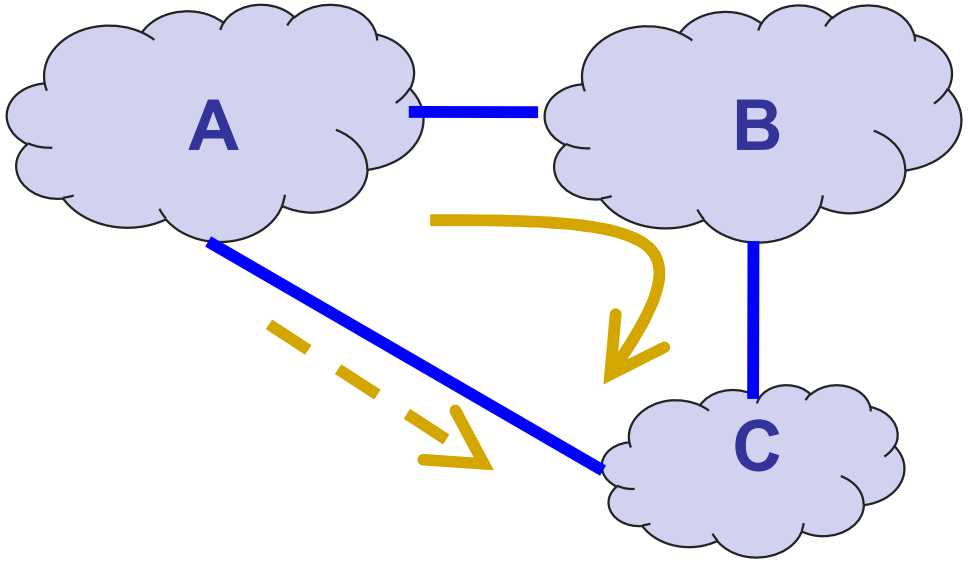
You’ve heard this story before!

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!

BGP & DV differences: (1) Not picking shortest-path routes

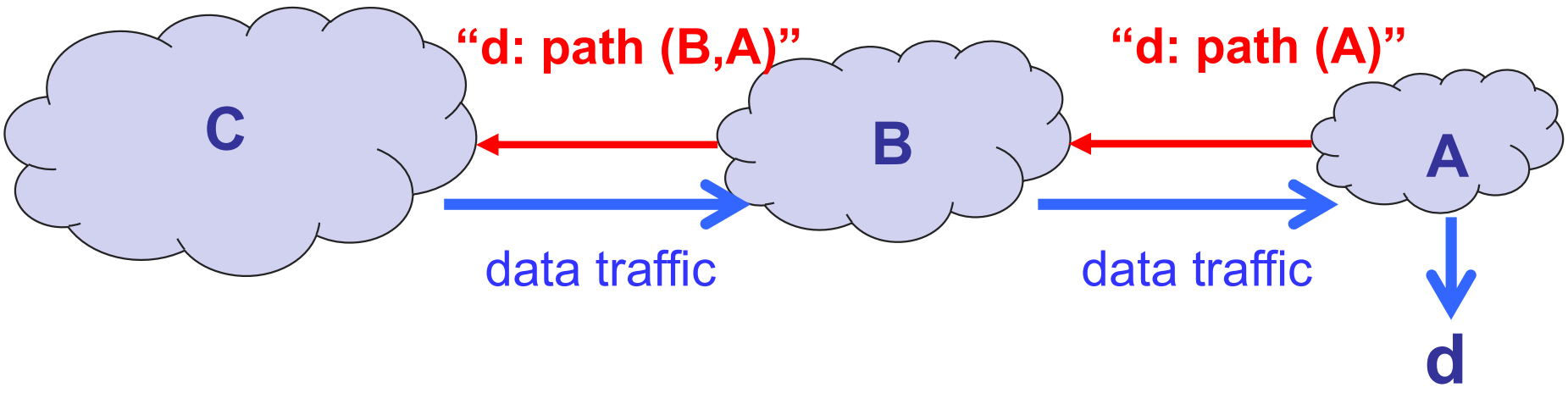
- BGP selects the best route based on policy, not shortest distance (i.e., least-cost)
- AS A may prefer “A,B,C” over “A,C”



- How do we avoid loops?

BGP & DV differences: (2) Path-Vector routing

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

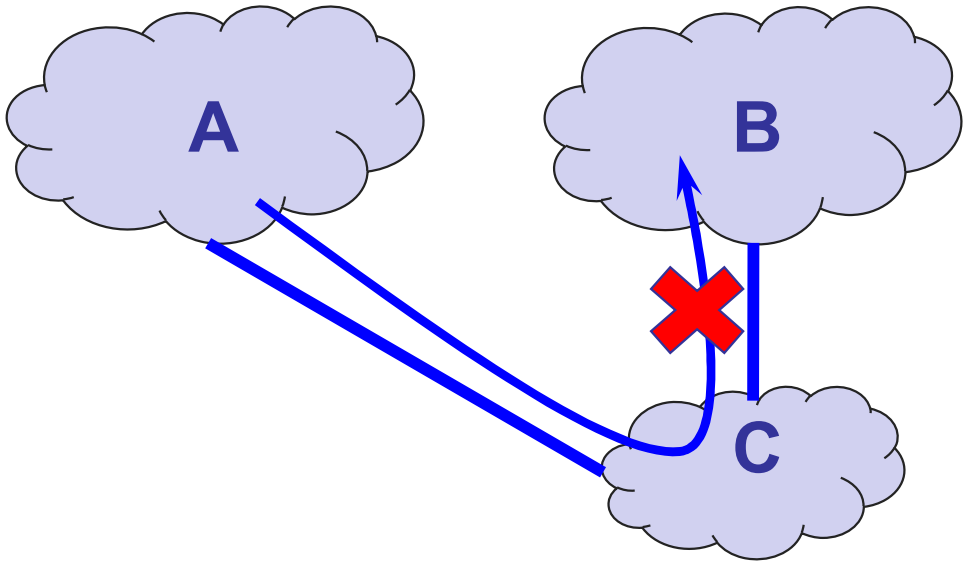


BGP & DV differences: (2) Path-Vector routing

- Key idea: advertise the entire path
  - Distance vector: send distance metric per destination
  - Path vector: send the entire path for each destination
- Benefits
  - Loop avoidance is straightforward (simply discard paths with loops)
  - Flexible and expressive policies based on entire path

BGP & DV differences: (3)
Selective route advertisement

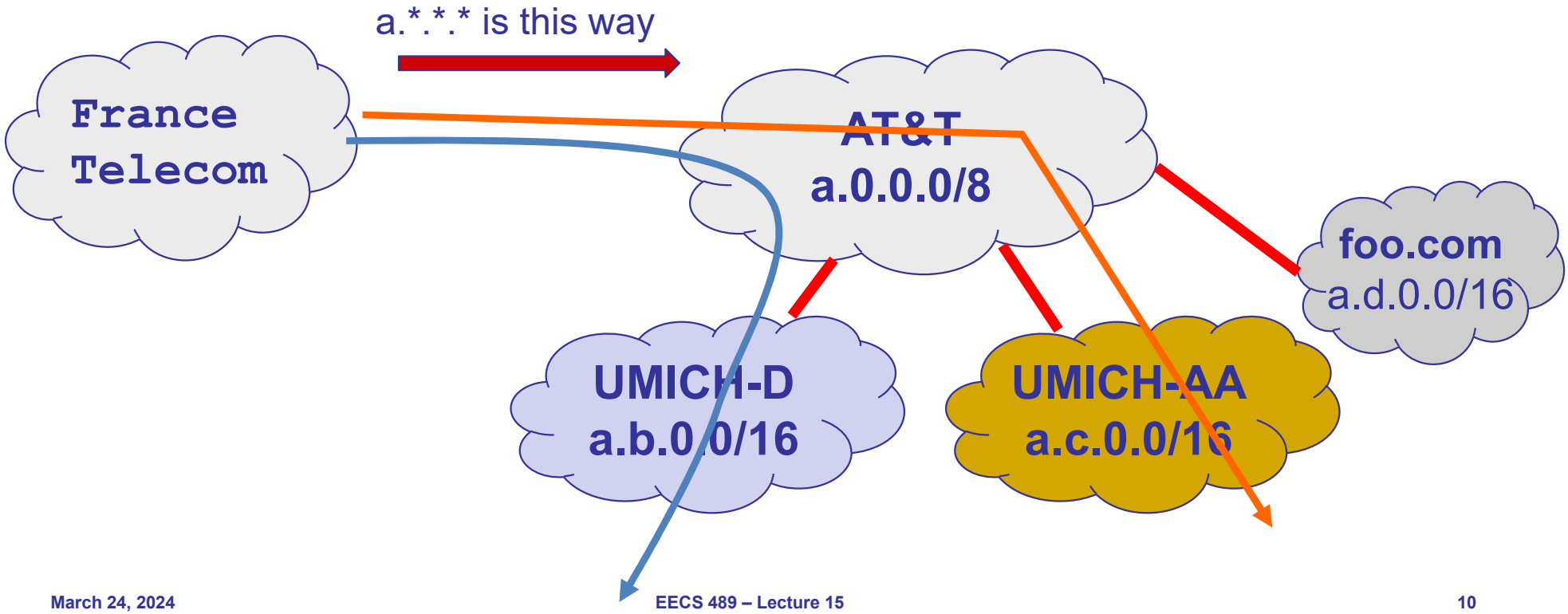
- For policy reasons, an AS may choose not to advertise a route to a destination
- Hence, reachability is not guaranteed even if graph is physically connected



AS-C does not want to carry traffic to AS-B

BGP & DV differences: (4) BGP may aggregate routes

- For scalability, BGP may aggregate routes for different prefixes

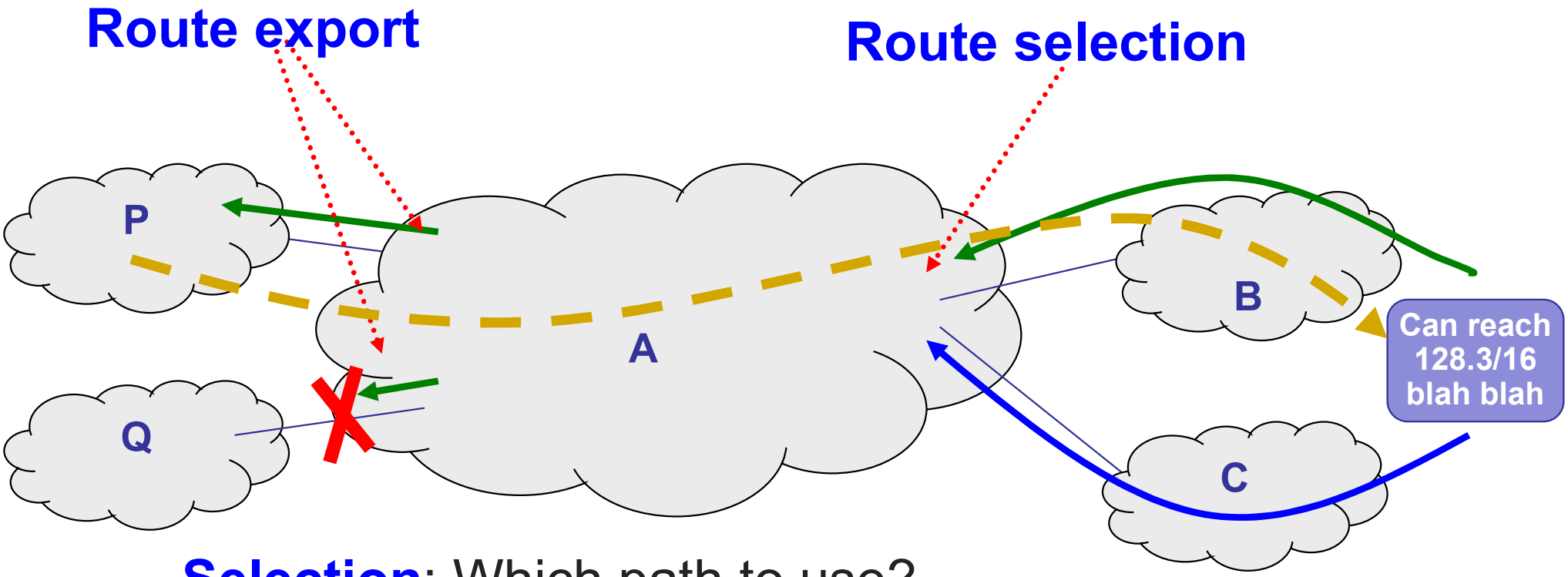


Topology & policy shaped by inter-AS business relationship

- 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

BGP POLICIES

Policy dictates 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

Typical selection policies

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

Typical export policy

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

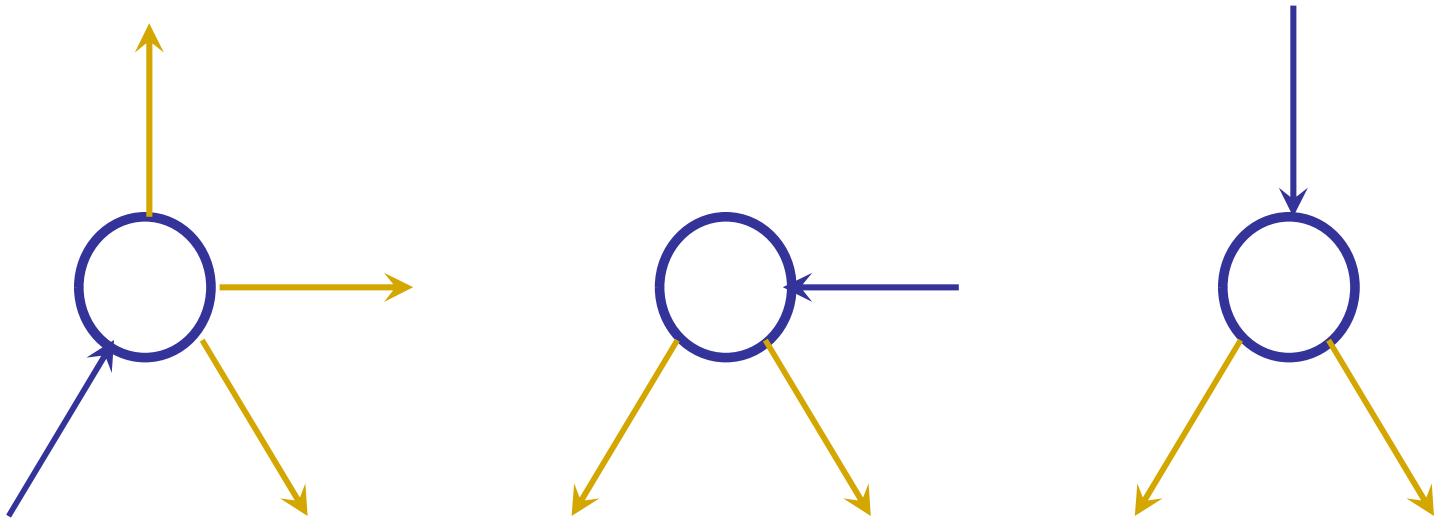
We'll refer to these as the "Gao-Rexford" rules (capture common – but not required! – practice)

Gao-Rexford

Providers

Peers

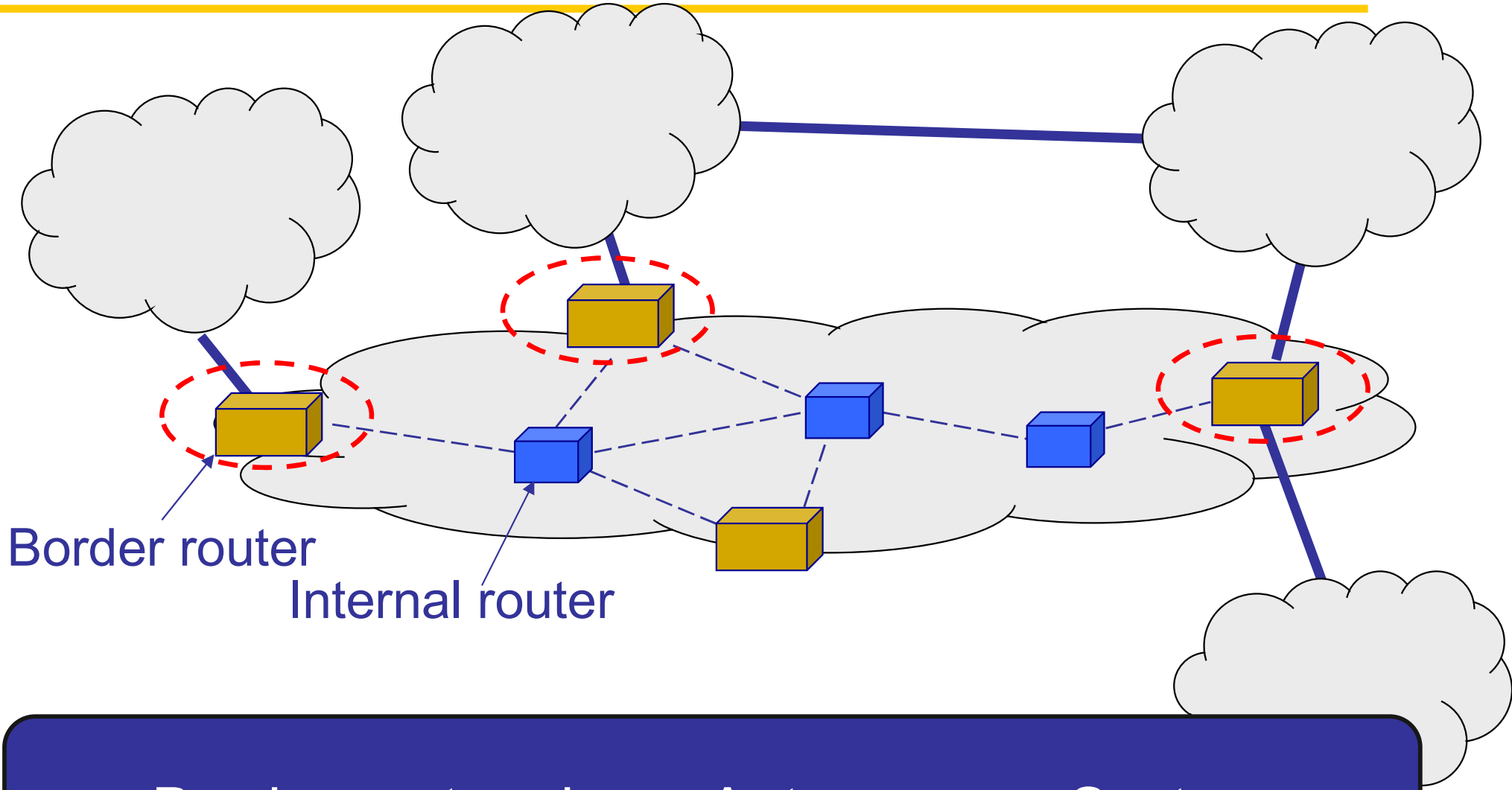
Customers



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

BGP PROTOCOL DETAILS

Who speaks BGP?

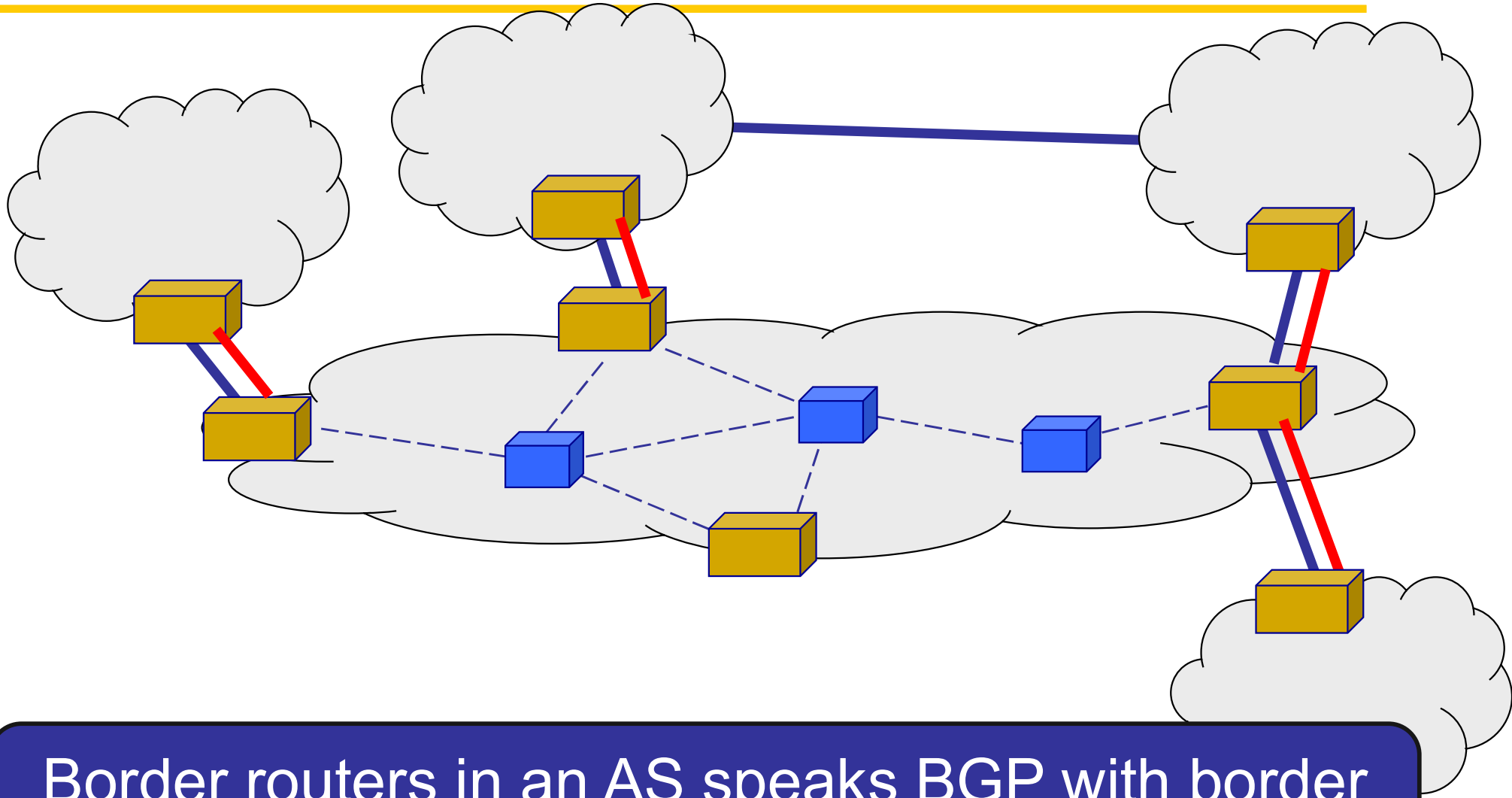


Border routers in an Autonomous System

What does “speak BGP” mean?

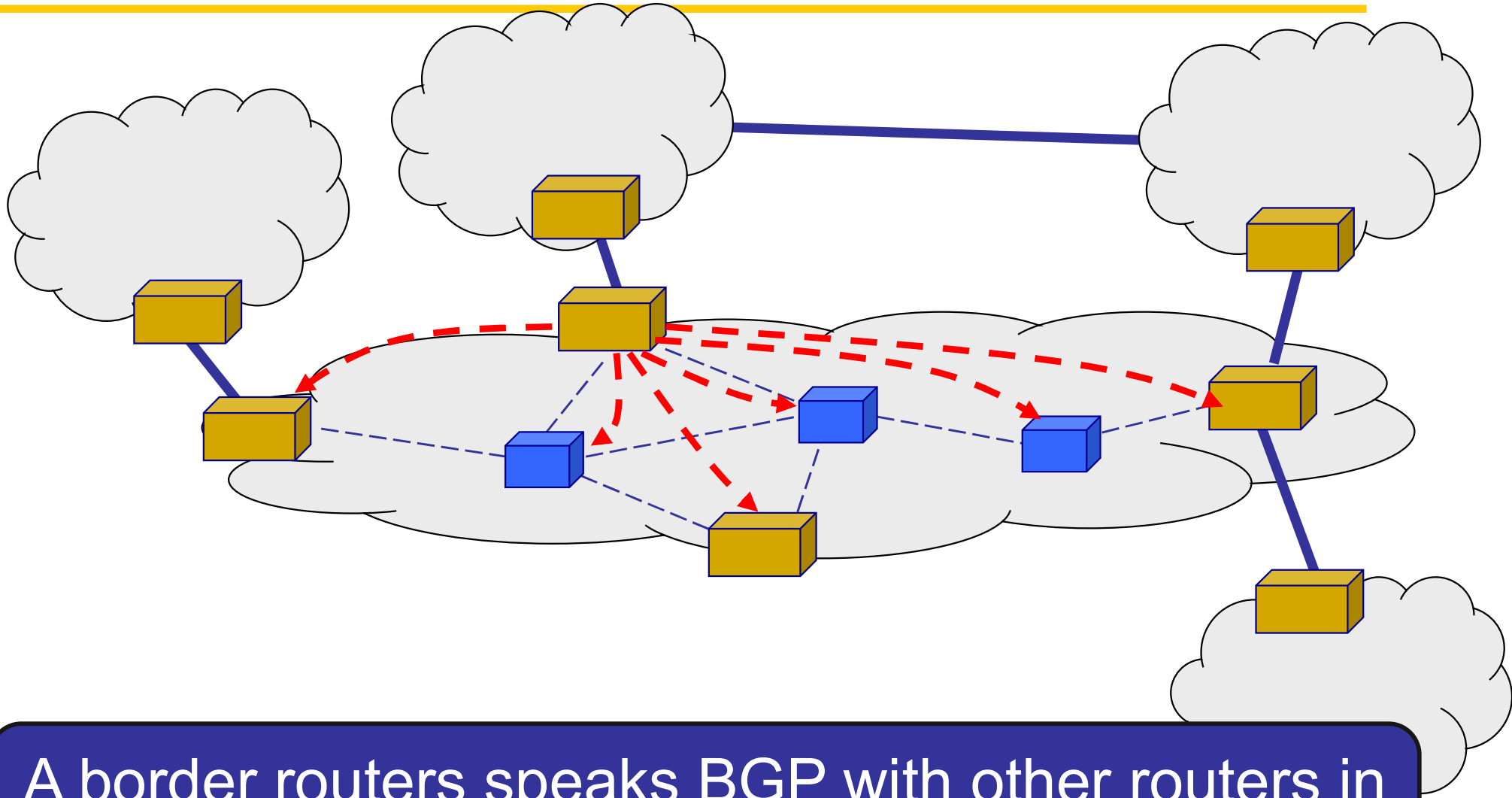
- Implement the BGP protocol standard
  - Read more here: http://tools.ietf.org/html/rfc4271
- Specifies what messages to exchange with other BGP “speakers”
  - Message types (e.g., route advertisements, updates)
  - Message syntax
- How to process these messages
  - E.g., “when you receive a BGP update,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

BGP sessions: External



Border routers in an AS speaks BGP with border routers in other ASes using eBGP sessions

BGP sessions: Internal



A border routers speaks BGP with other routers in the same AS using iBGP sessions

eBGP, iBGP, and 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” = Intra-domain routing protocol
  - Provide internal reachability
  - E.g., OSPF, RIP

eBGP, iBGP, and IGP together

- Learn routes to external destination using eBGP
- Distribute externally learned routes internally using iBGP
- Travel shortest path to egress using IGP

Basic messages in BGP

- Open
  - Establishes BGP session (BGP uses TCP)
- Notification
  - Report unusual conditions
- Update
  - Inform neighbor of new routes
  - Inform neighbor of old routes that become inactive
- Keep-alive
  - Inform neighbor that connection is still viable

Route updates

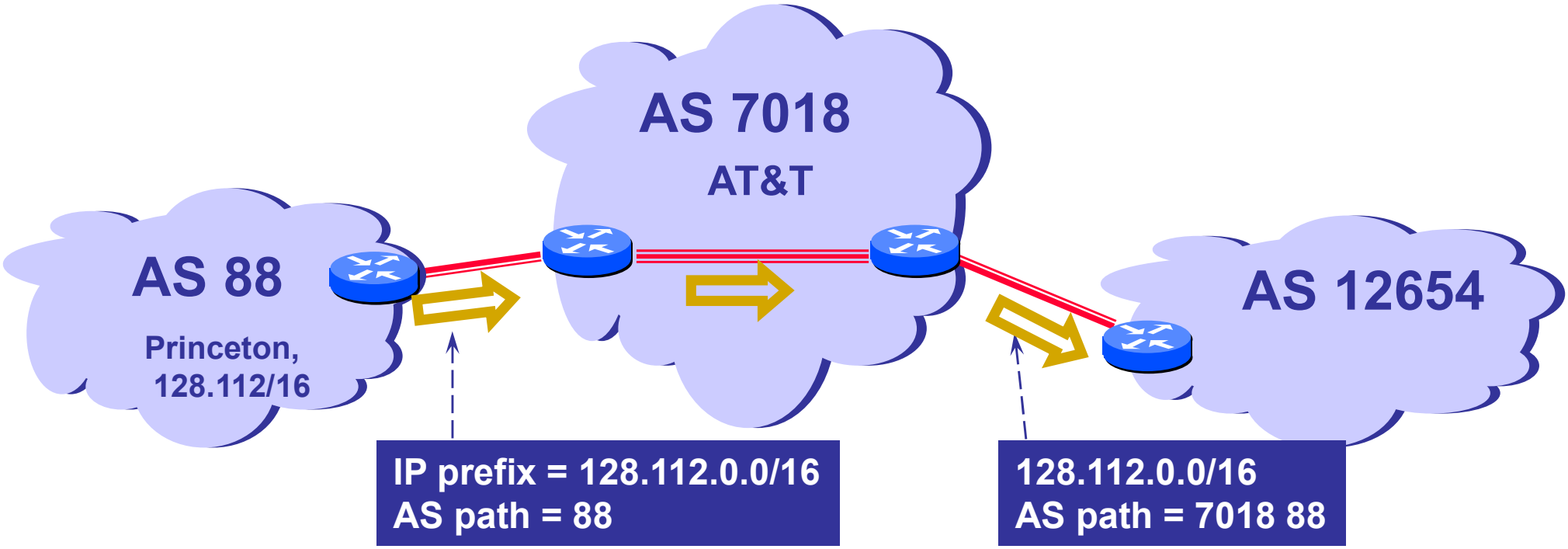
- Format <IP prefix: route attributes>
  - Attributes describe properties of the route
- Two kinds of updates
  - Announcements: new routes or changes to existing routes
  - Withdrawal: remove routes that no longer exist

Route attributes

- 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

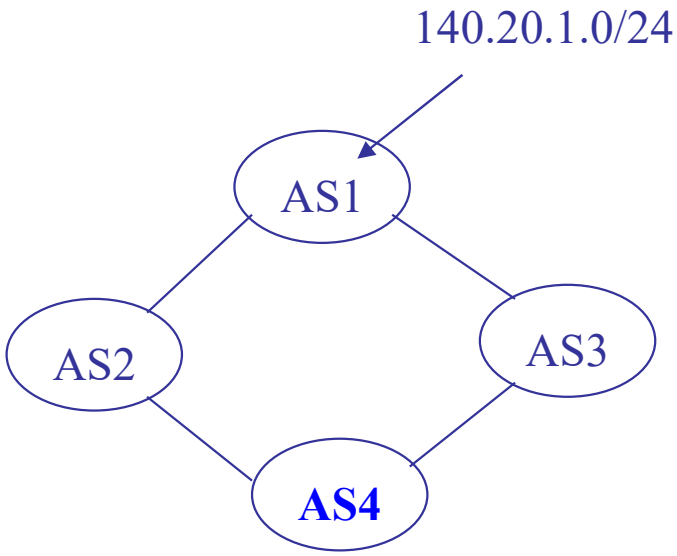
Attributes: (1) ASPATH

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



Attributes: (2) LOCAL PREF

- Local preference in choosing between different AS paths
  - Local to an AS; carried only in iBGP messages
- The higher the value the more preferred

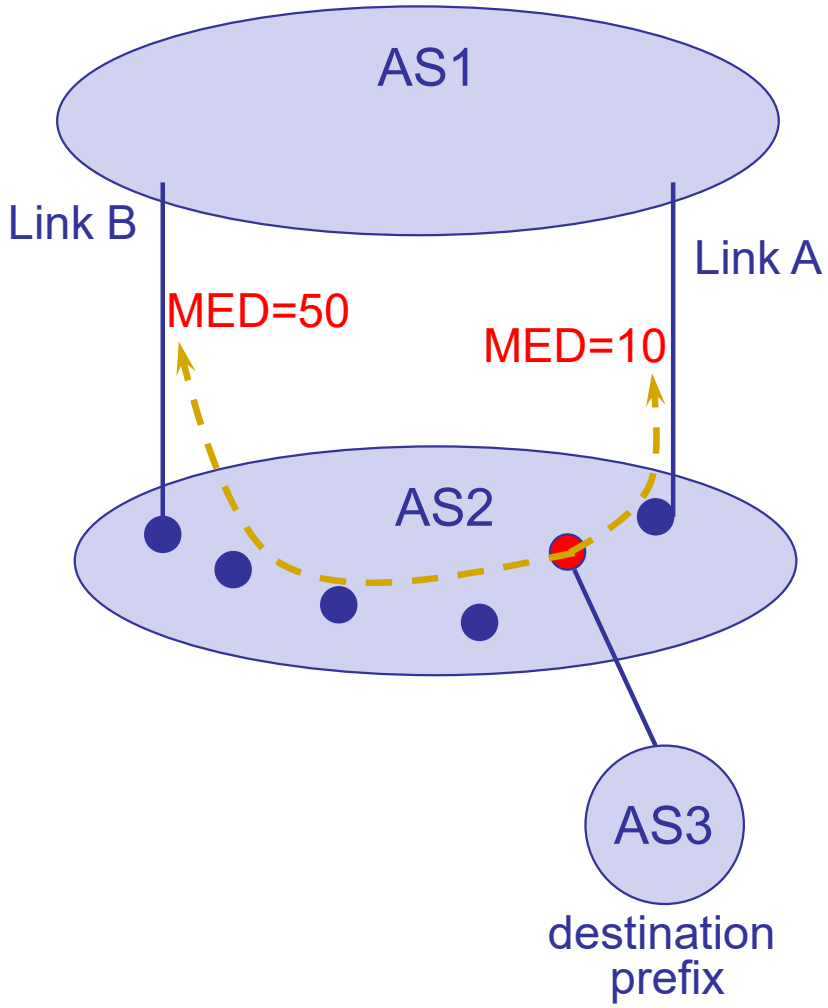


BGP table at AS4:

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

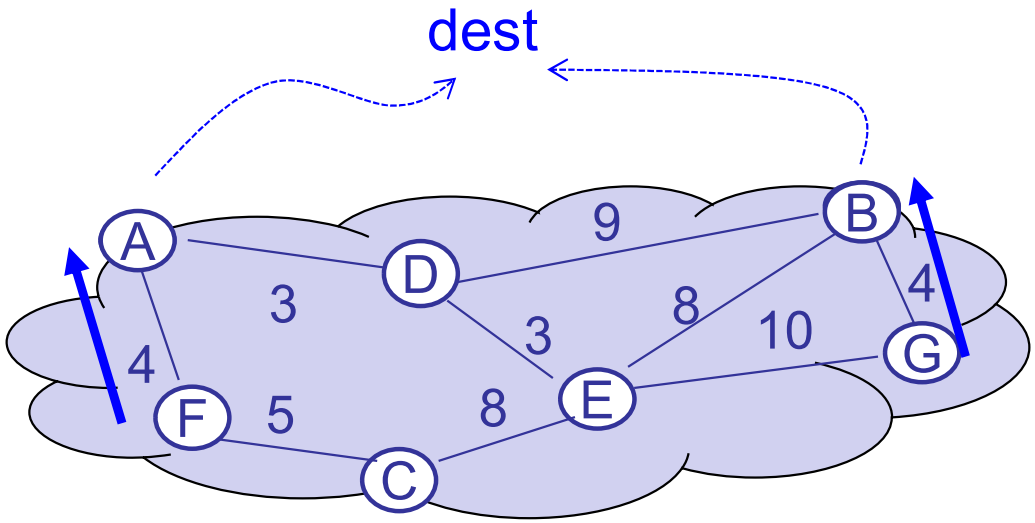
Attributes: (3) MED

- Multi-exit discriminator is used when ASes are interconnected via 2 or more links; it specifies how close a prefix is to the link it is announced on
- Lower is better
- AS that announces a prefix sets MED
- AS receiving the 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 intra-domain protocol

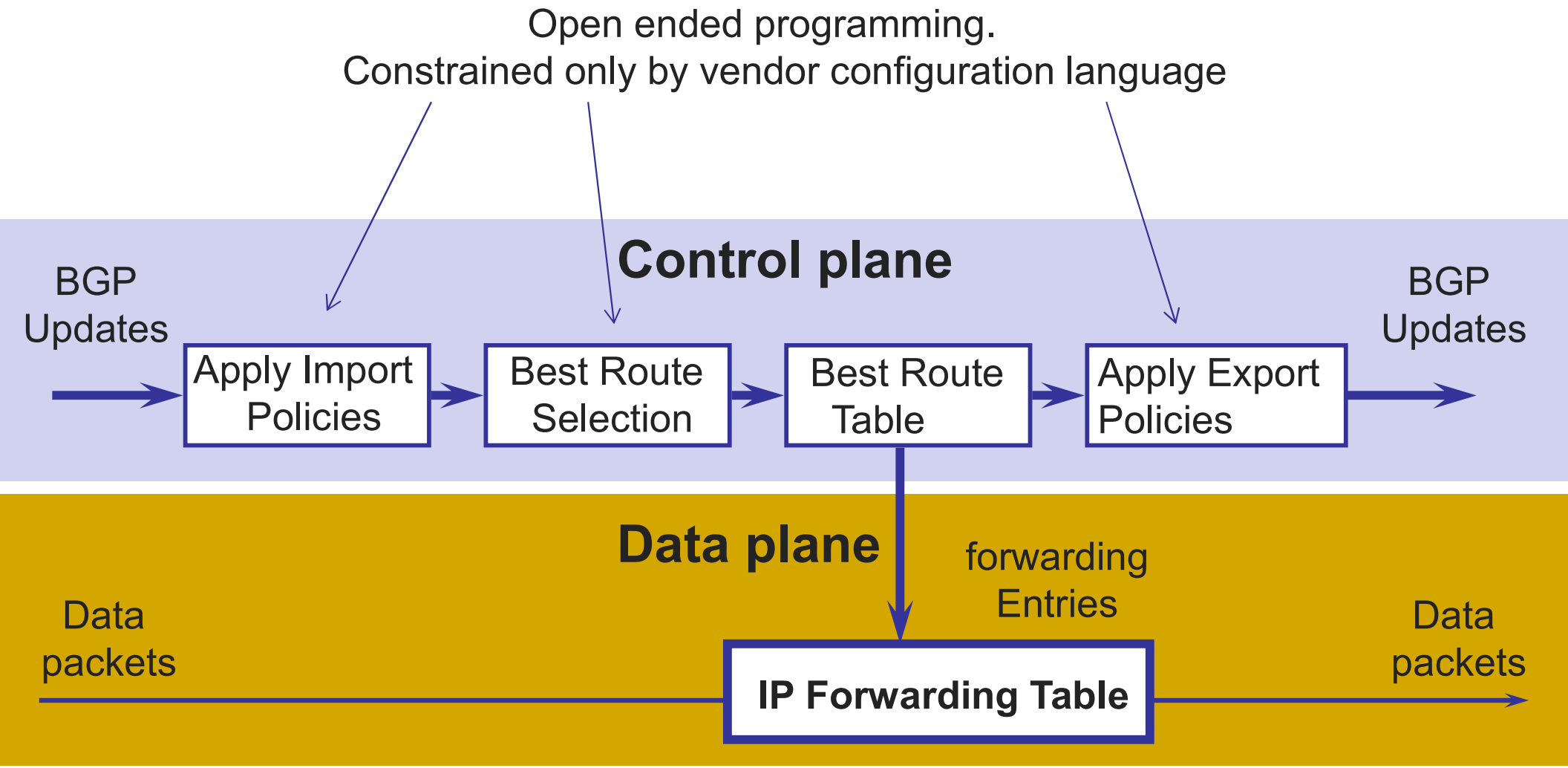


Using attributes

- 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



5-MINUTE BREAK!

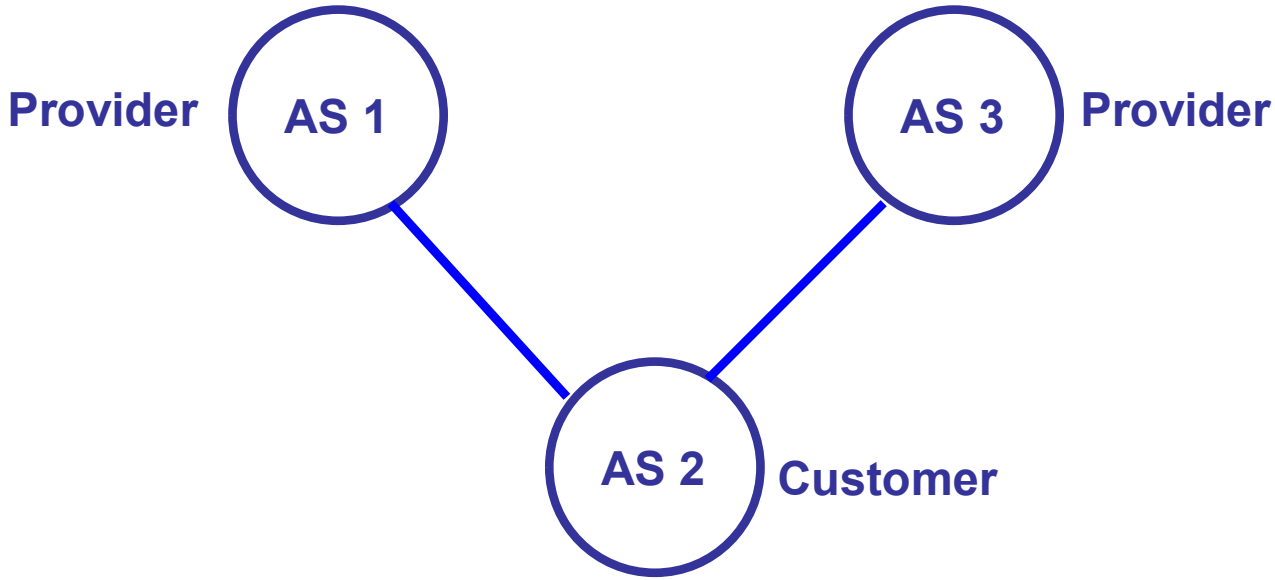
BGP ISSUES IN PRACTICE

Issues with BGP

- Reachability
- Security
- Convergence
- Performance
- Anomalies

Reachability

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



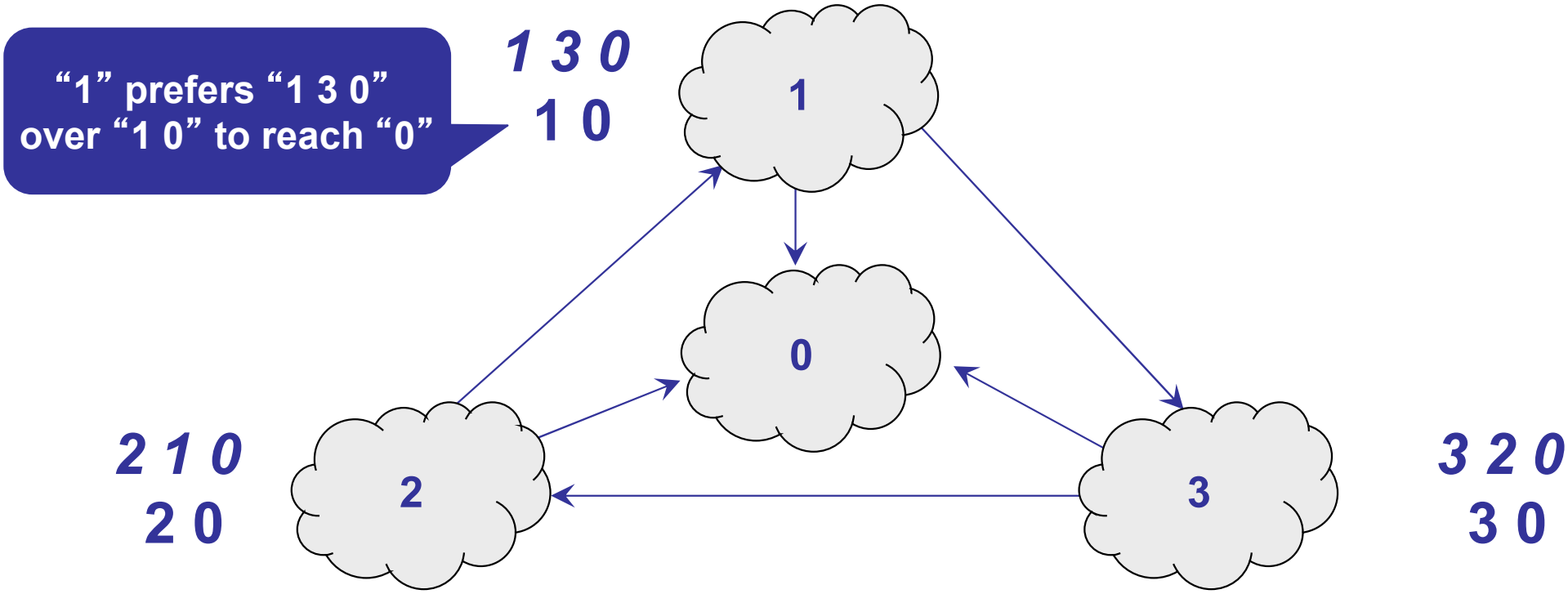
Security

- An AS can claim to serve a prefix that they do not have a route to (blackholing)
  - Problem not specific to policy or path vector
  - Important because of AS autonomy
  - Fixable: make ASes "prove" they have a path
- AS may forward packets along a route different from what is advertised
  - Tell customers about fictitious short path…
  - Much harder to fix!
  - More: http://queue.acm.org/detail.cfm?id=2668966

Convergence

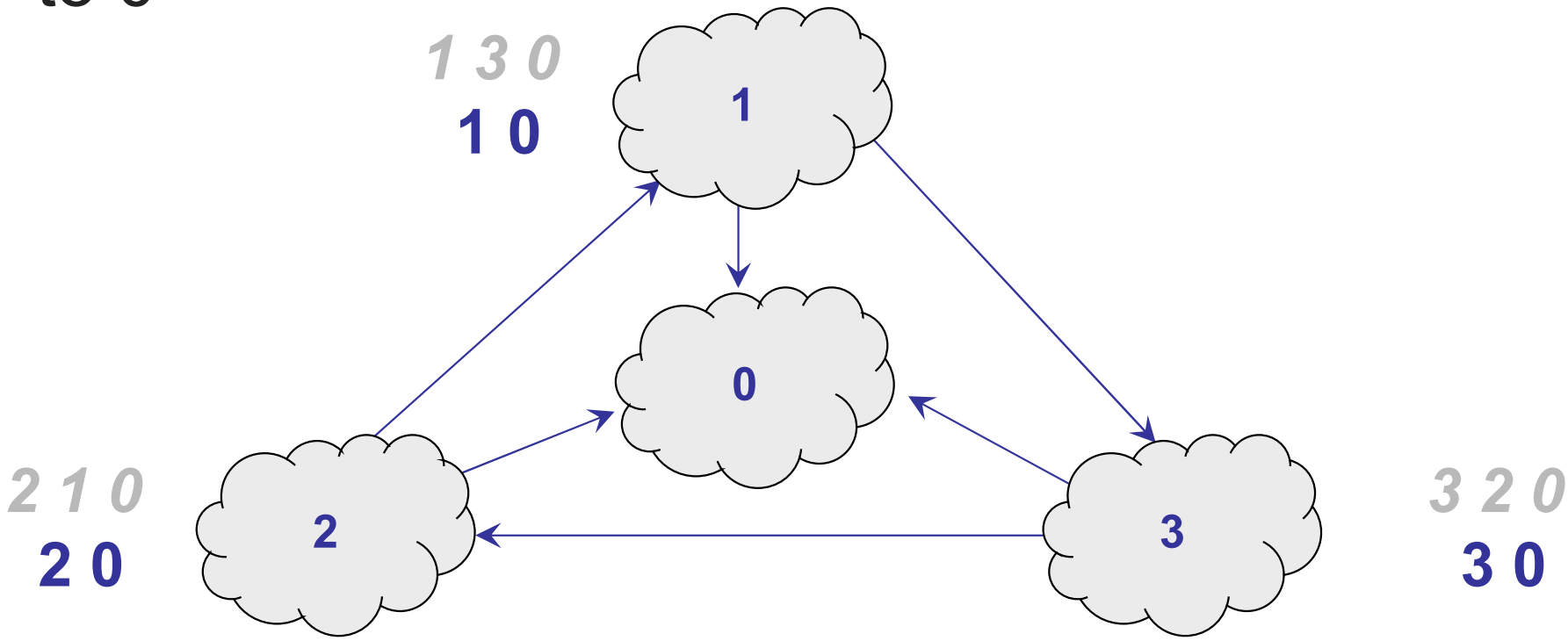
- If all AS policies follow “Gao-Rexford” rules, BGP is guaranteed to converge
- For arbitrary policies, BGP may fail to converge!

Example of policy oscillation



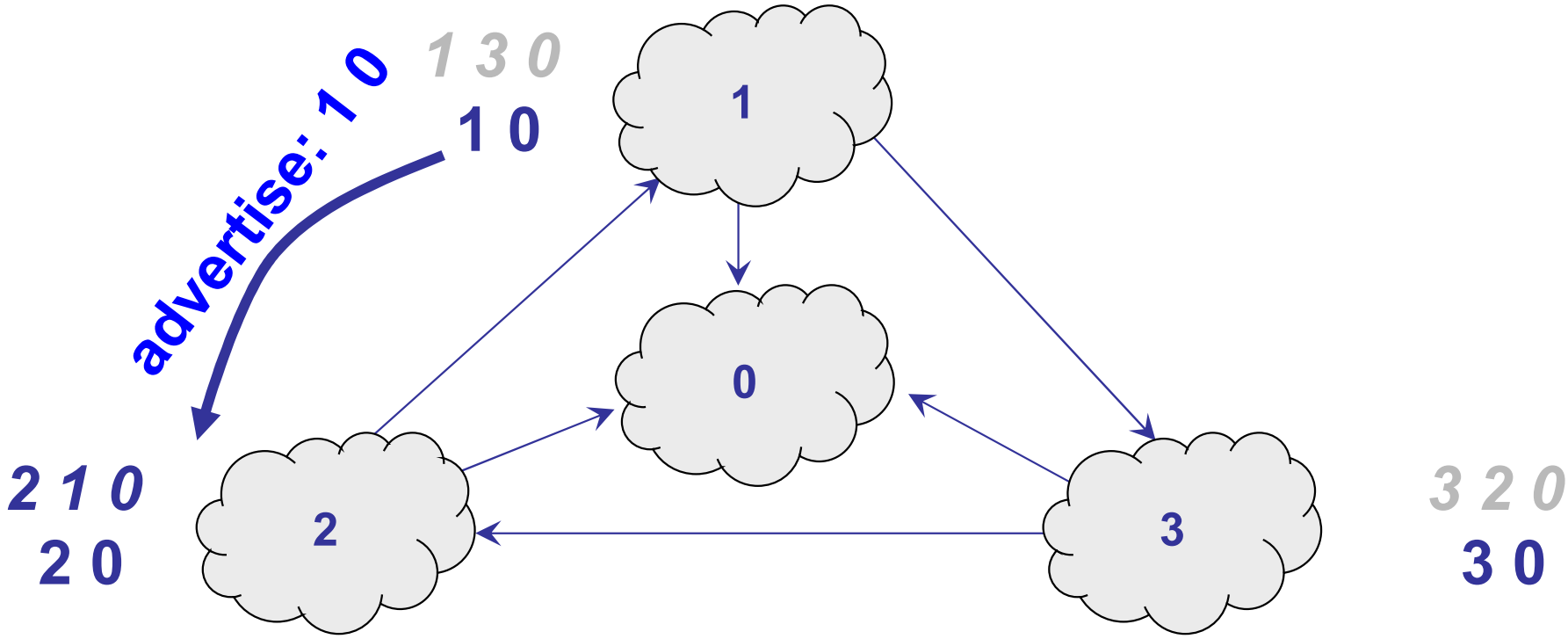
Step-by-step of policy oscillation

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

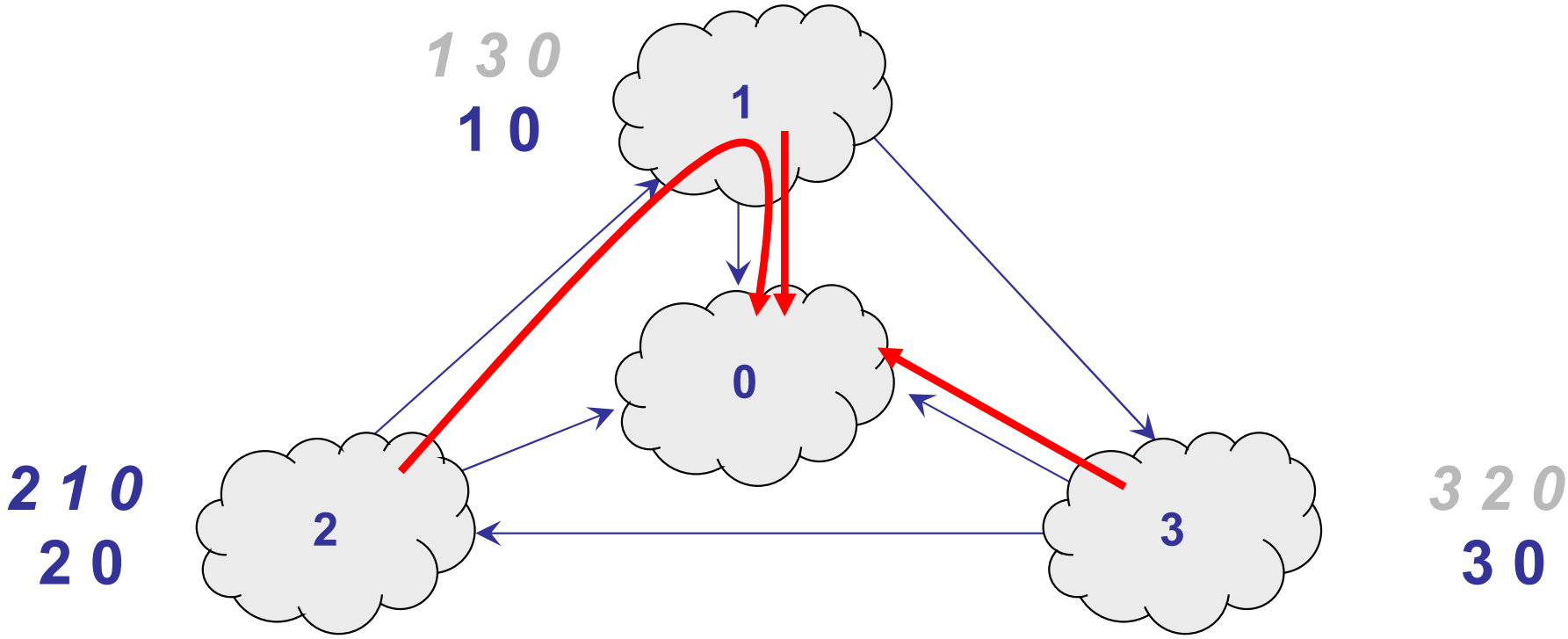


Step-by-step of policy oscillation

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

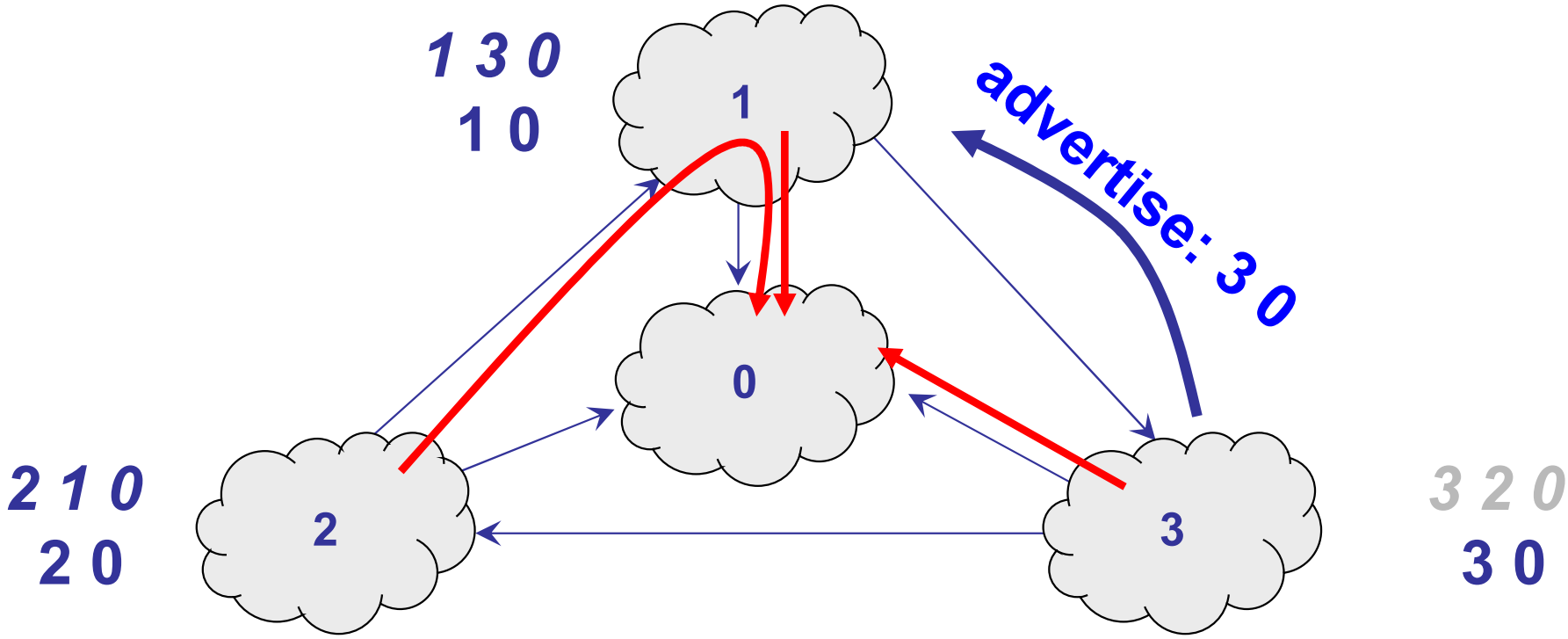


Step-by-step of policy oscillation

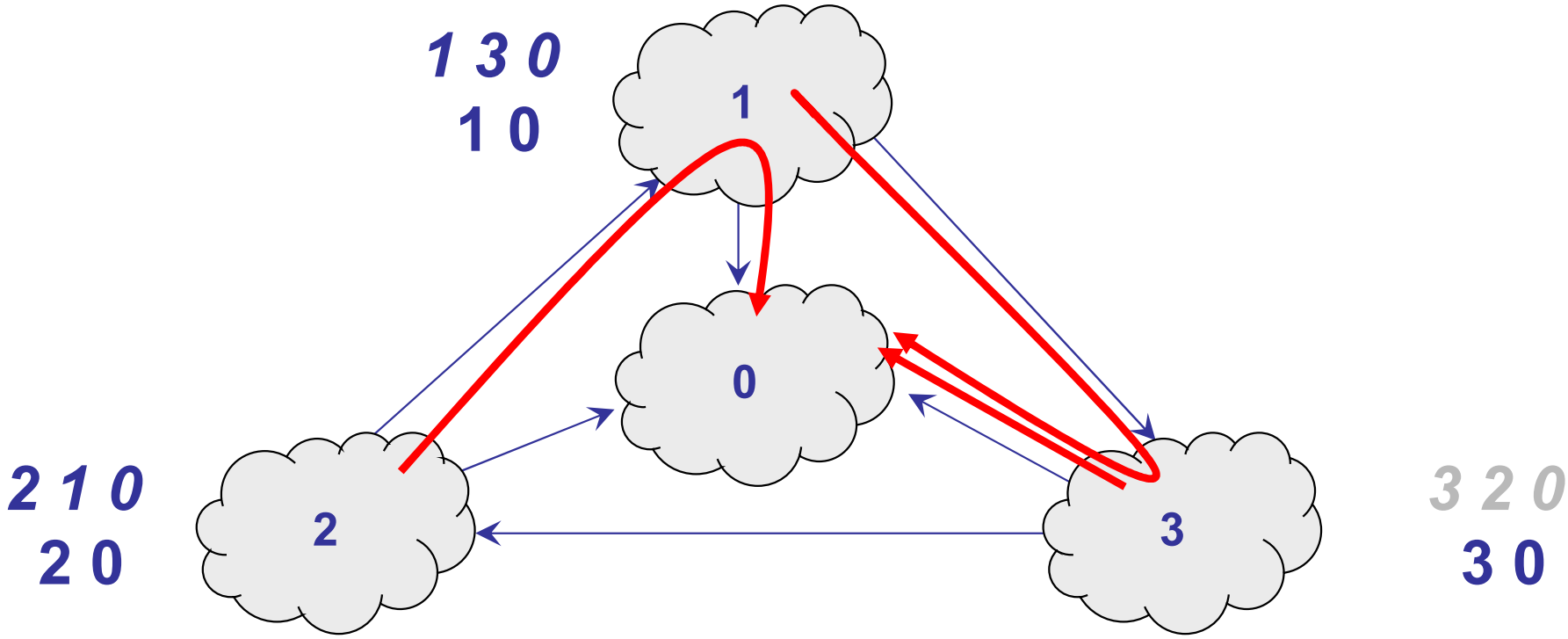


Step-by-step of policy oscillation

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

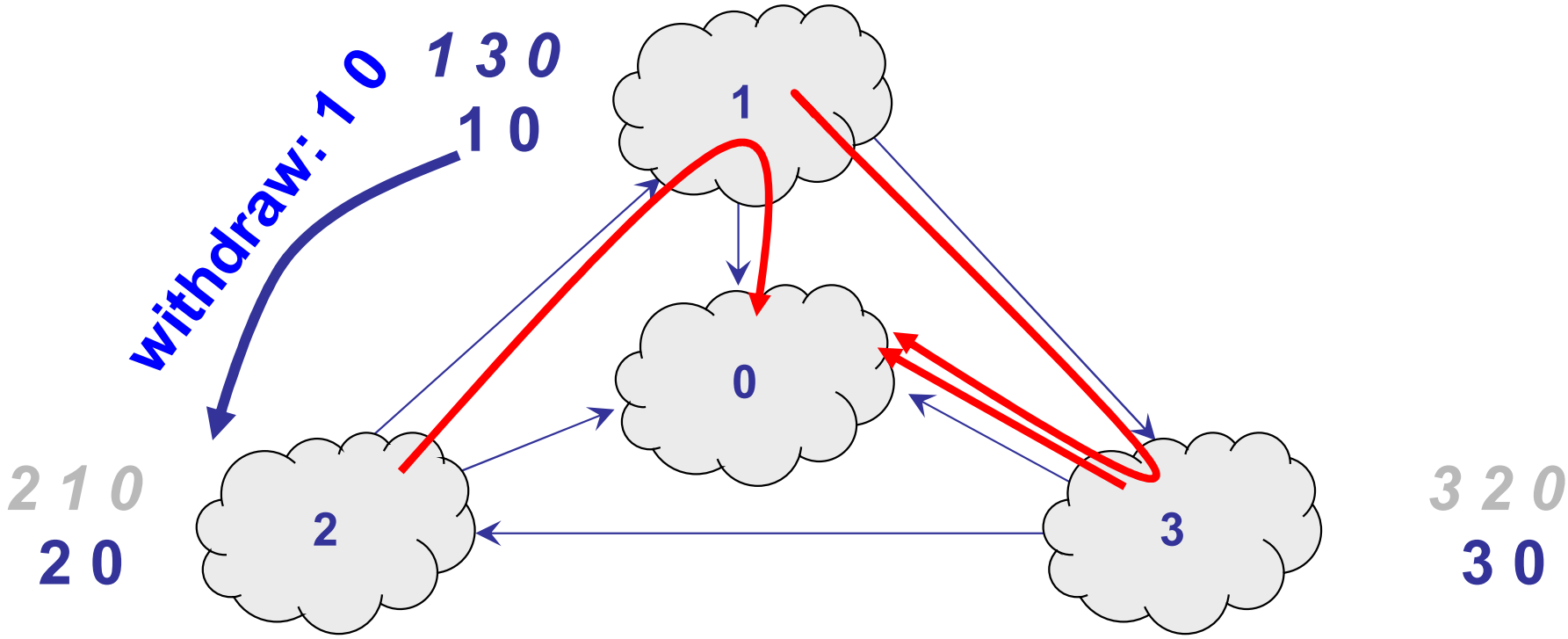


Step-by-step of policy oscillation

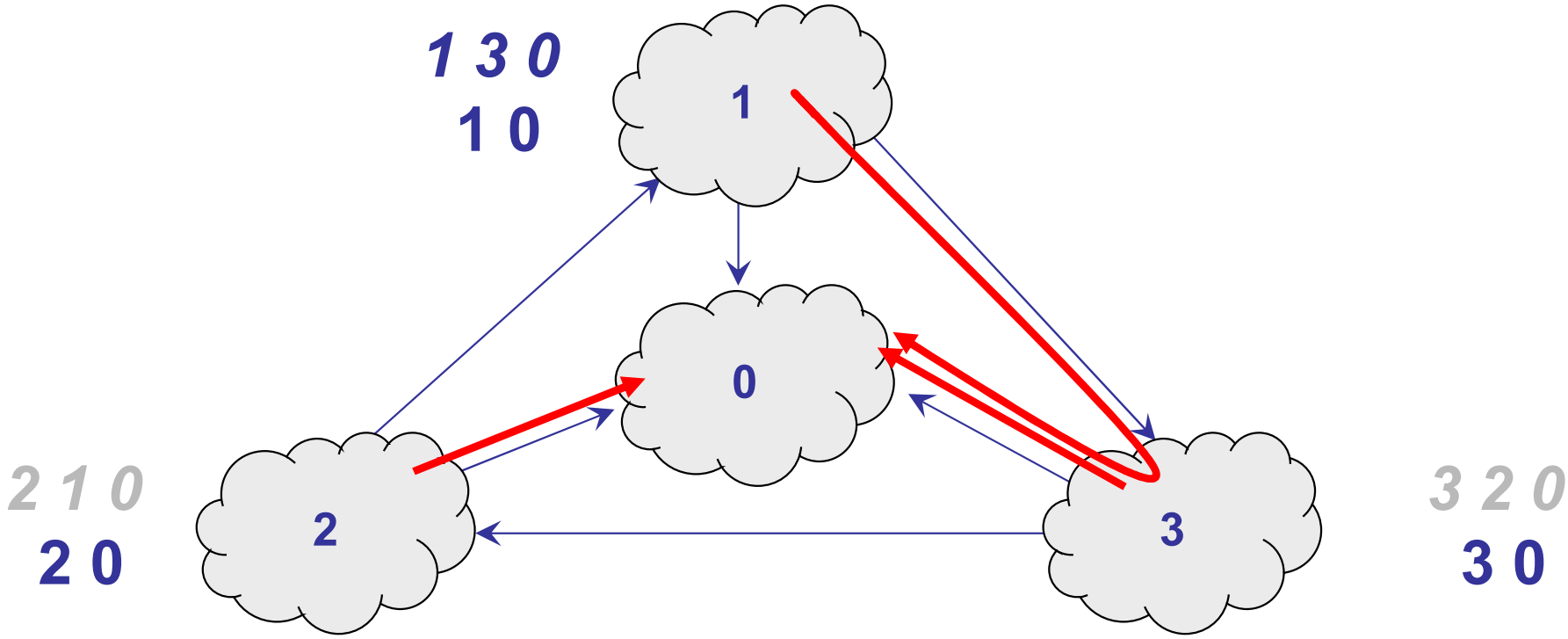


Step-by-step of policy oscillation

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

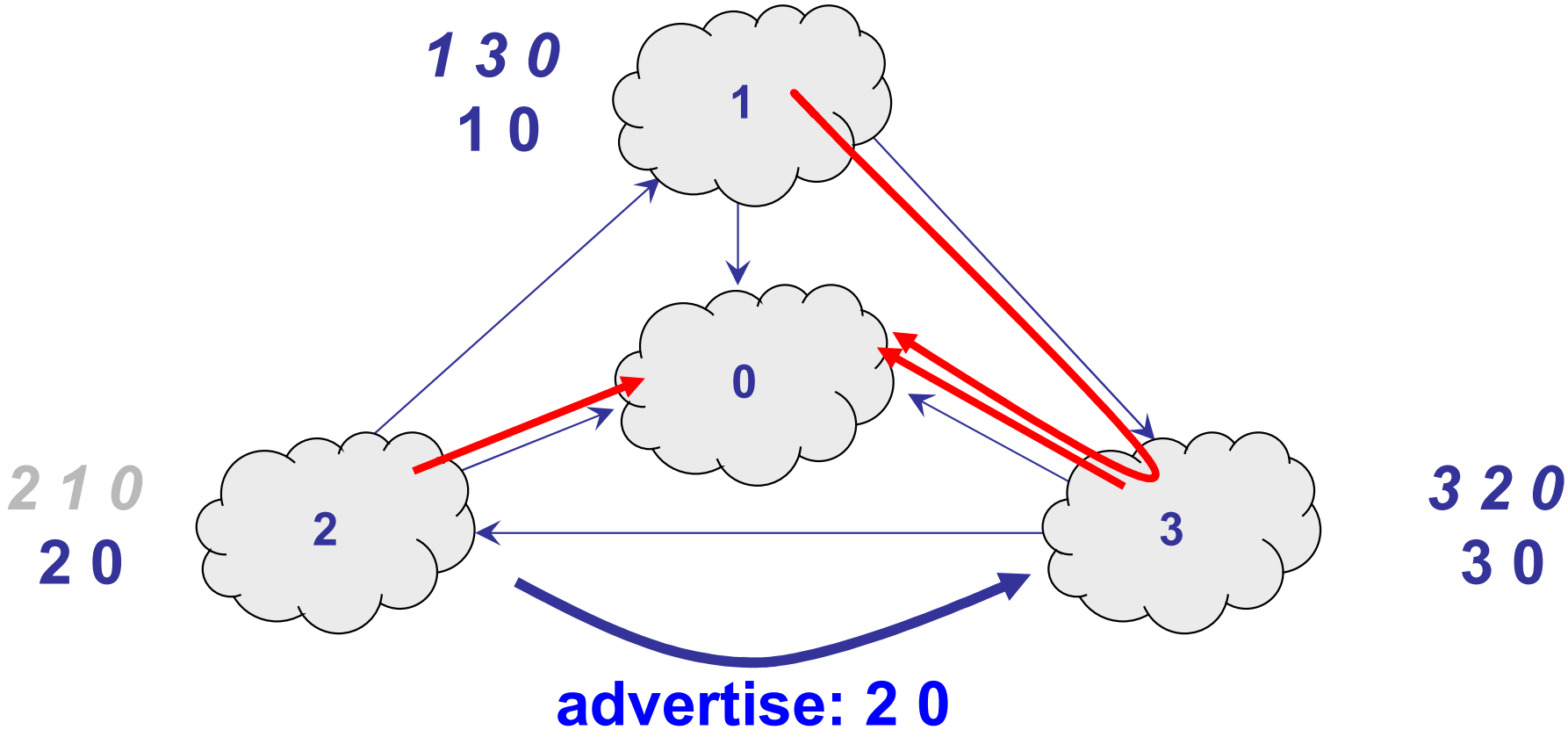


Step-by-step of policy oscillation

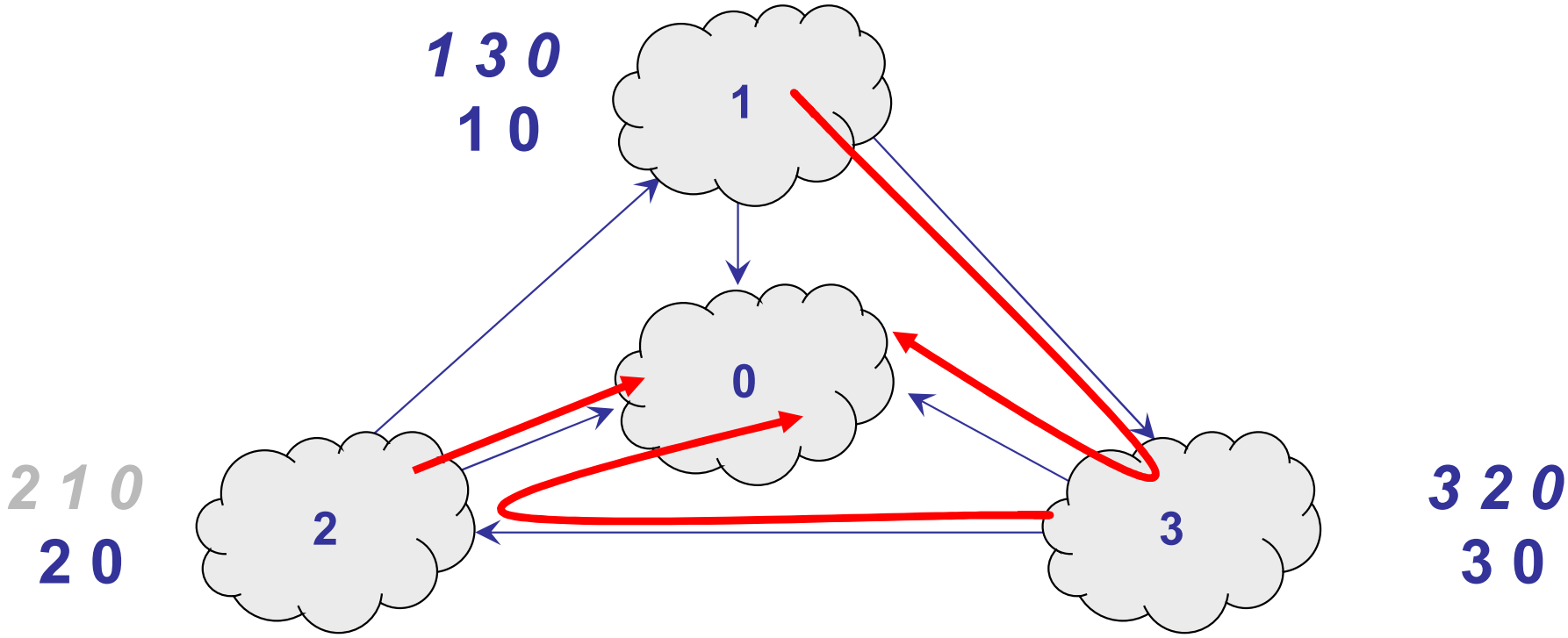


Step-by-step of policy oscillation

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

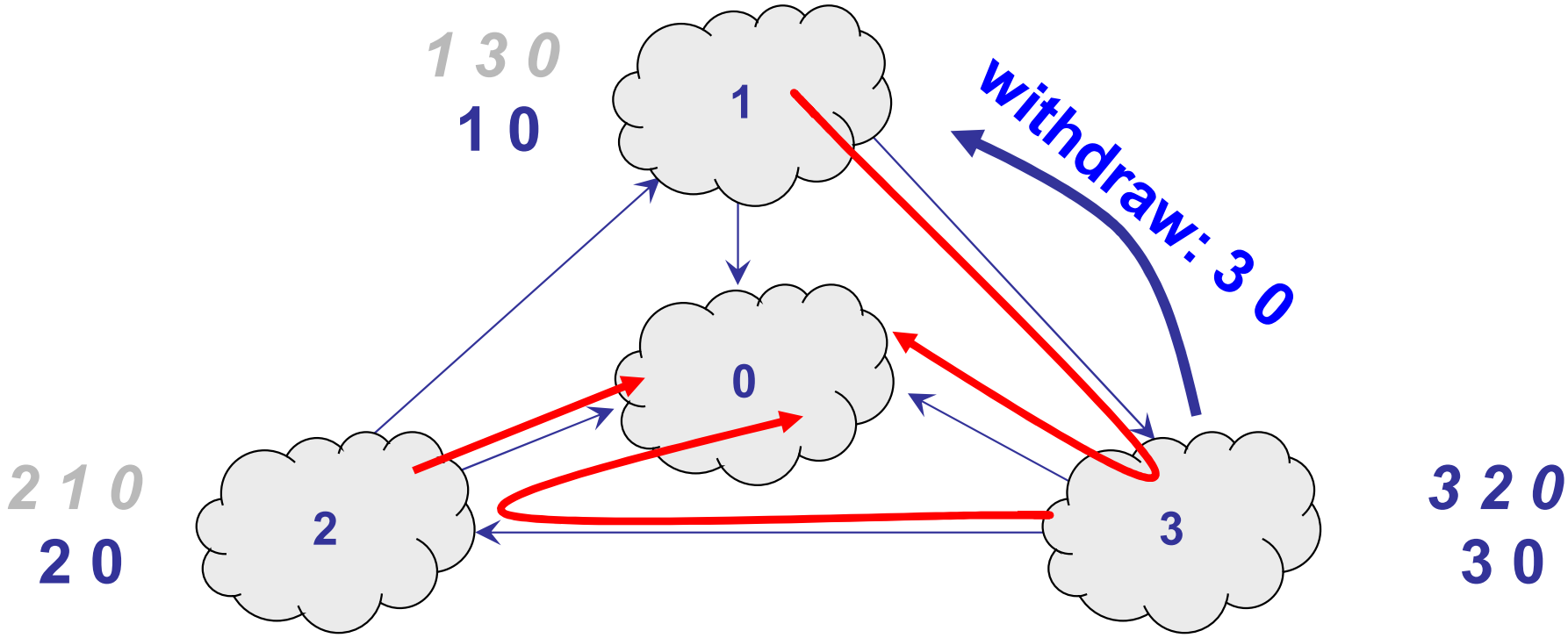


Step-by-step of policy oscillation

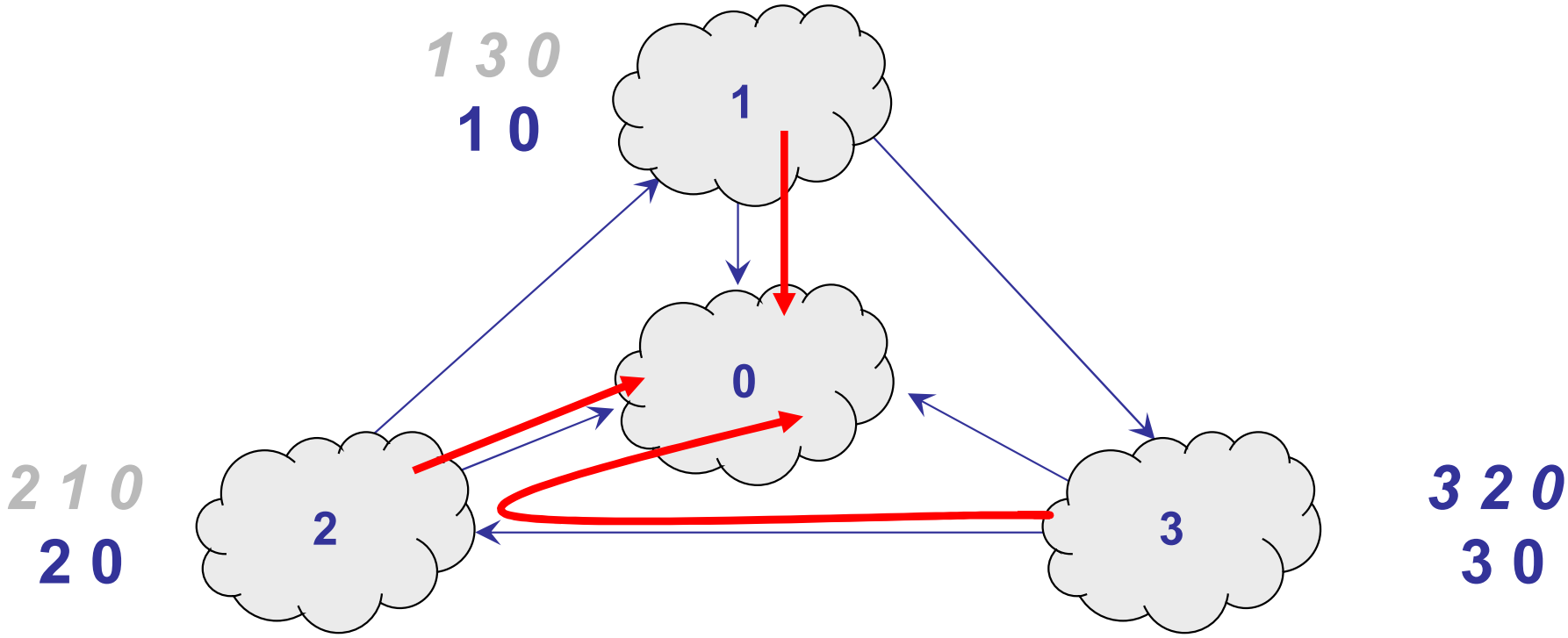


Step-by-step of policy oscillation

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

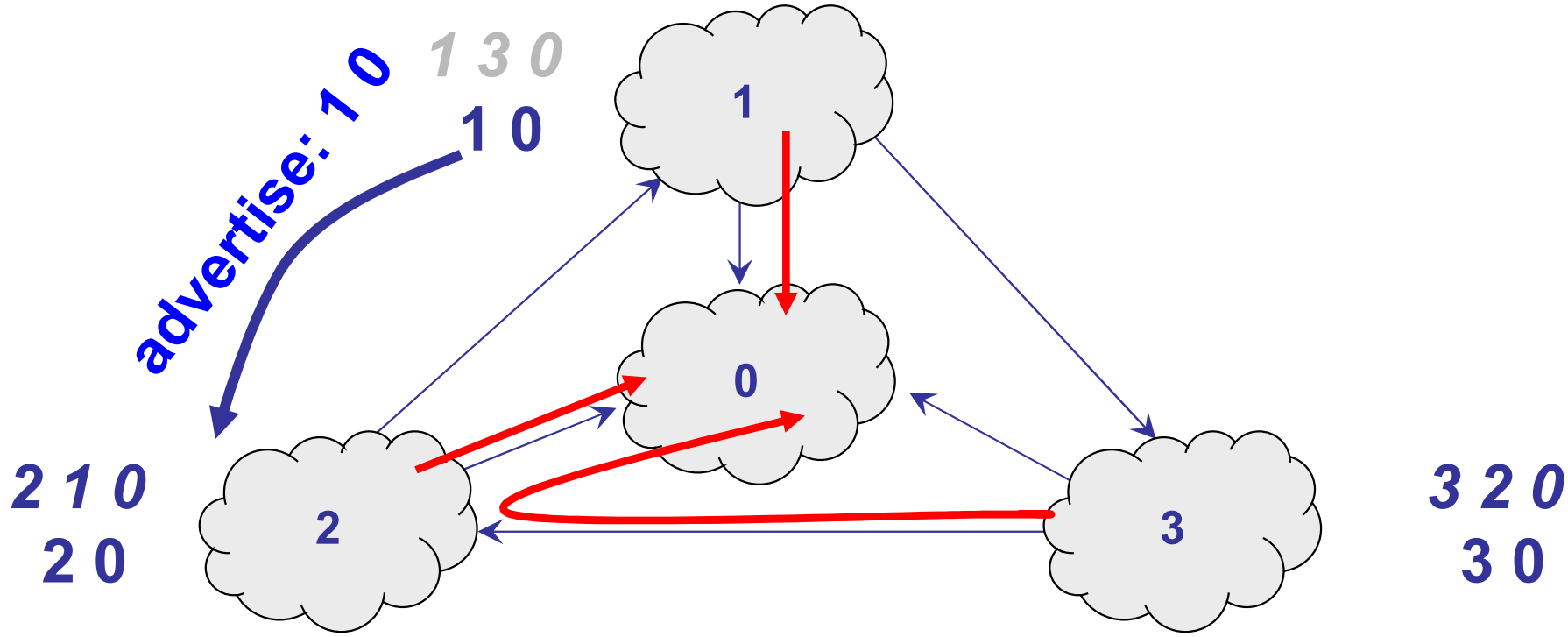


Step-by-step of policy oscillation

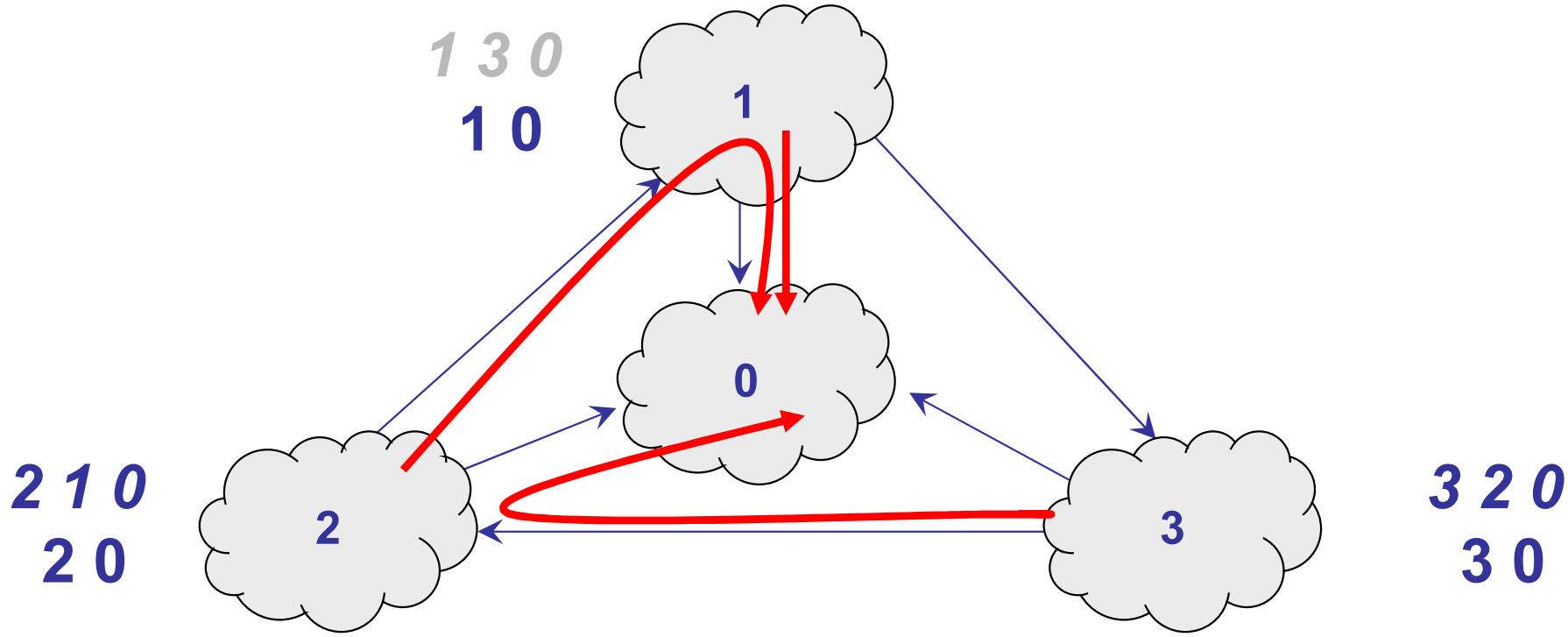


Step-by-step of policy oscillation

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

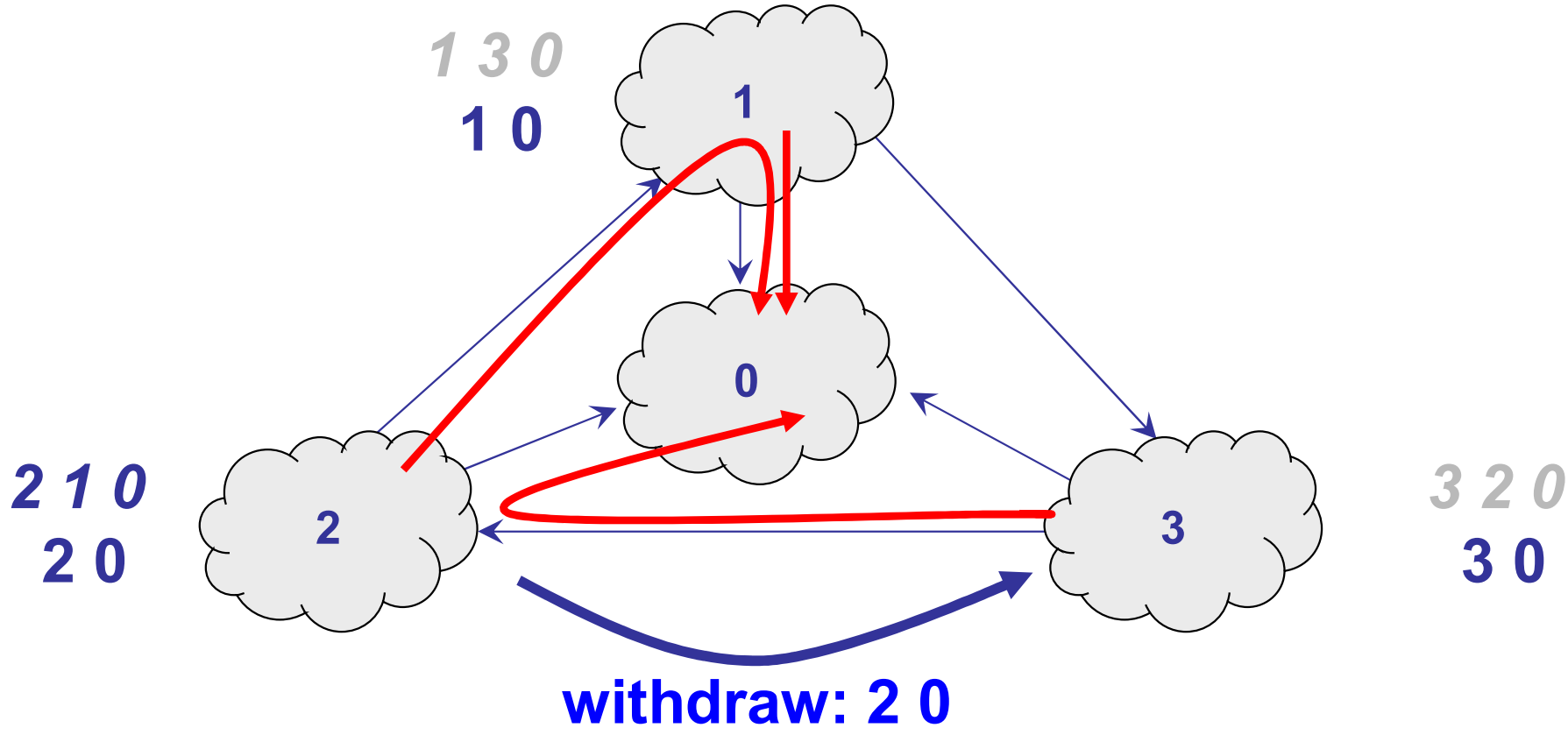


Step-by-step of policy oscillation

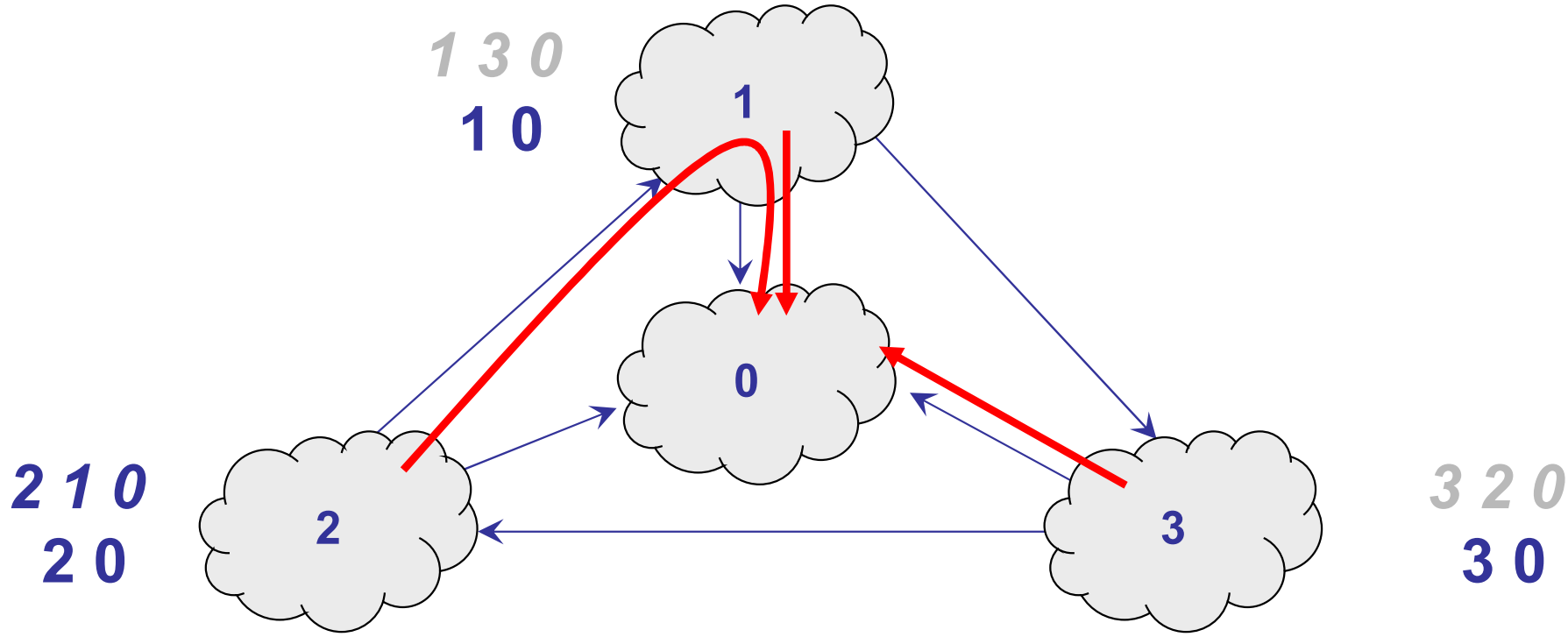


Step-by-step of policy oscillation

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



We're back to where we started



Convergence

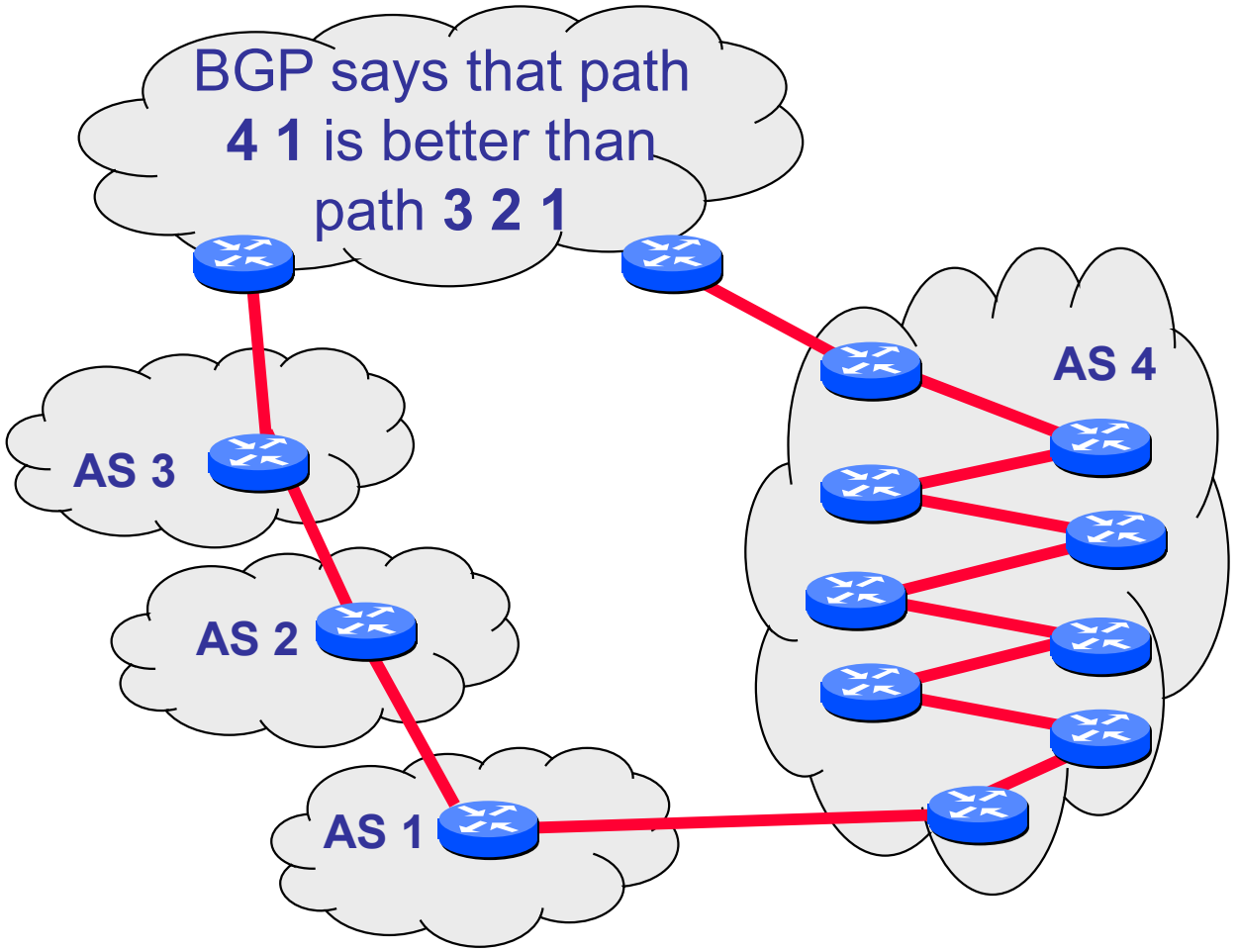
- If all AS policies follow “Gao-Rexford” rules, BGP is guaranteed to converge
- For arbitrary policies, BGP may fail to converge!

Performance nonissues

- Internal routing
  - Domains typically use “hot potato” routing
  - Not always optimal, but economically expedient
- Policy is not always about performance
  - Policy-driven paths aren’t the shortest
- AS path length can be misleading
  - 20% of paths inflated by at least 5 router hops

AS path length can be misleading

- An AS may have many router-level hops



Real performance issue: Slow convergence

- BGP outages are biggest source of Internet problems
- Most popular paths are very stable
- Outages and other issues are very common
  - Check out https://radar.cloudflare.com/routing

BGP misconfigurations

- BGP protocol is bloated yet 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
- Configuration is mostly manual and ad hoc
  - Disjoint per-router configuration to effect AS-wide policy

Summary

- Network layer deals with data plane (forwarding) and control plane (routing)
- Control plane deals with intra-domain routing (LS and DV) and inter-domain routing (BGP)

- Next class: SDN