EN.601.414/614 Computer Networks

BGP

Xin Jin

Fall 2020 (TuTh 1:30-2:45pm on Zoom)

JOHNS HOPKINS UNIVERSITY

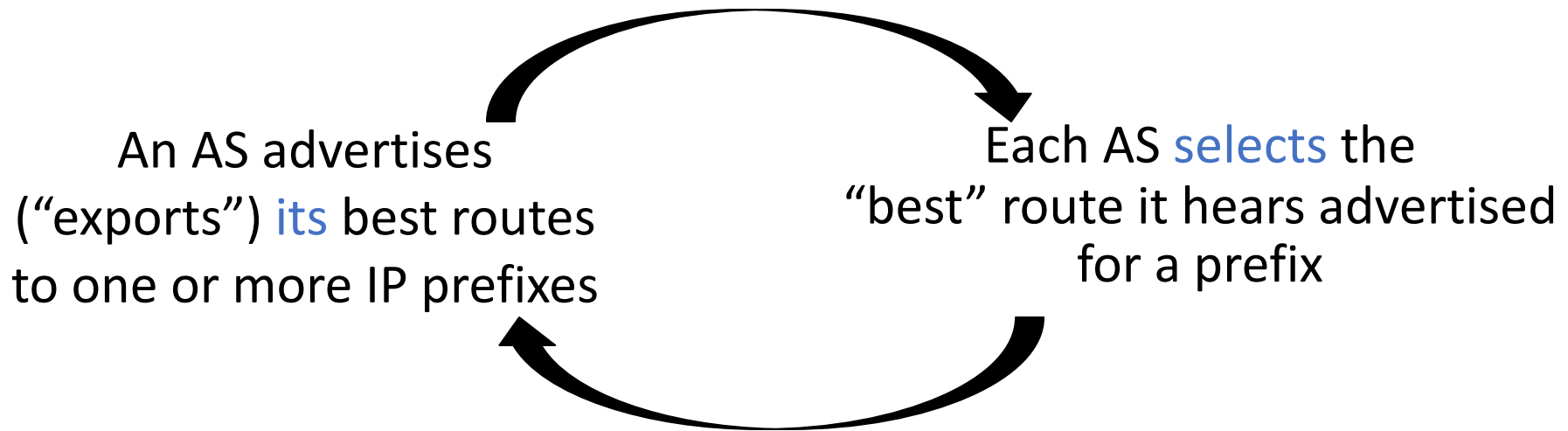https://github.com/xinjin/course-net

Agenda

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

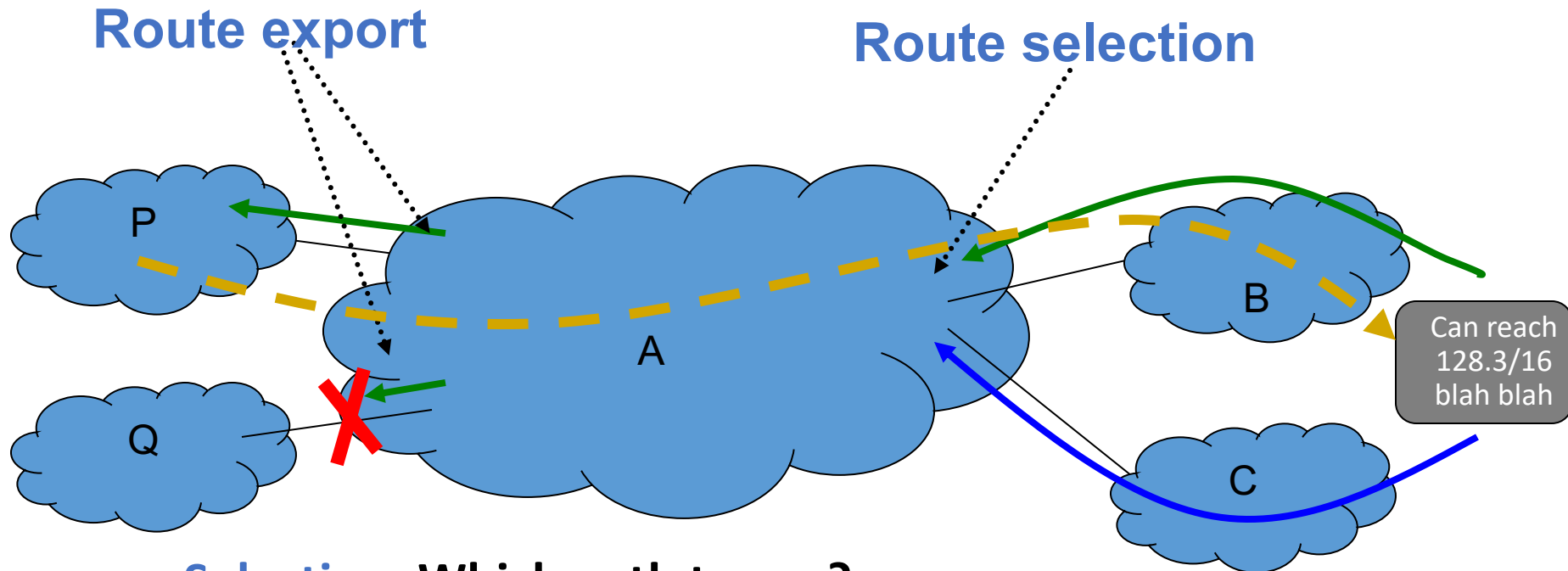# Inter-domain routing: Setup

- **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

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 policies

Policy dictates how routes are "selected" and "exported"

Route export

Route selection

P

A

Q

B

C

Can reach 128.3/16 blah blah

- **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

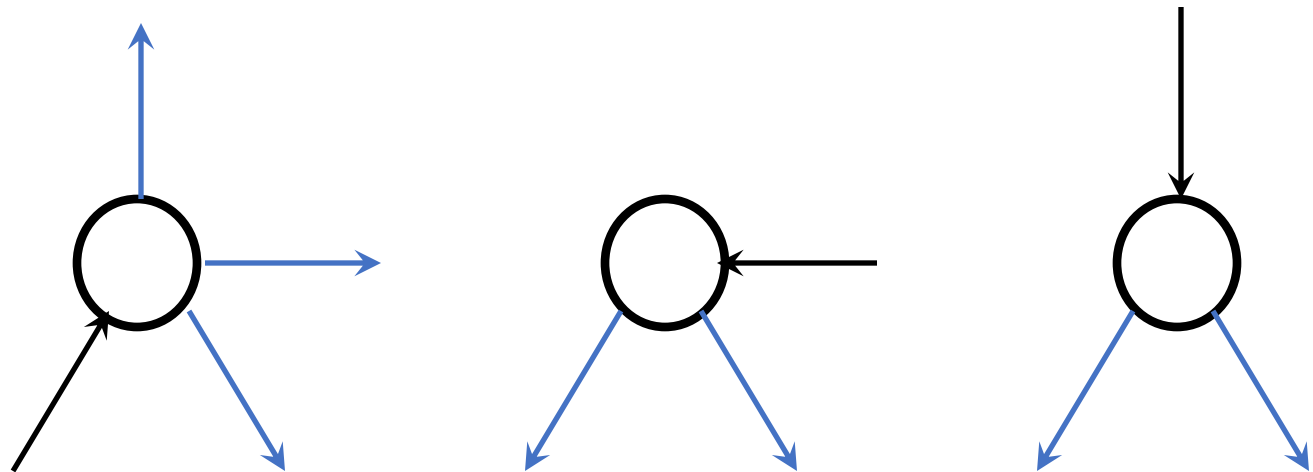We'll refer to these as the "Gao-Rexford" rules (capture common – but not required! – practice)

Gao-Rexford

Providers

Peers

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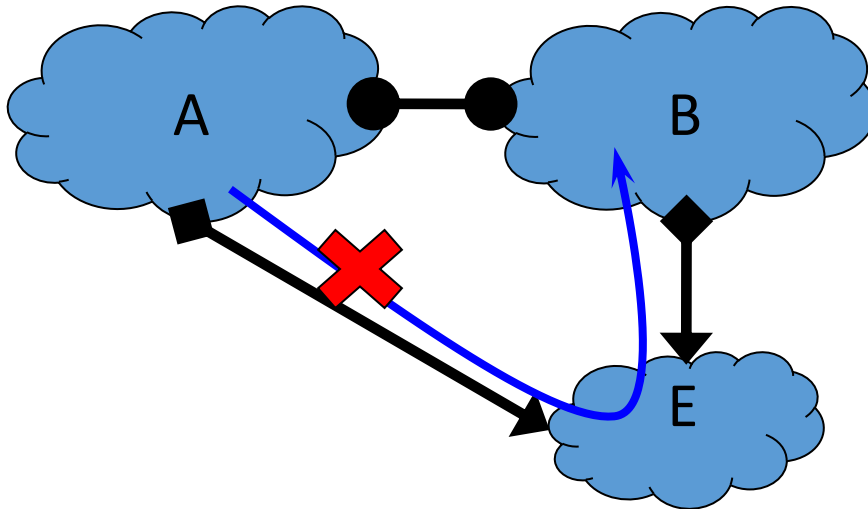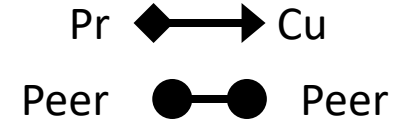
With Gao-Rexford, the AS policy graph is a
DAG (directed acyclic graph) and routes are "valley free"

Valley-Free Routing

- **Number links as (+1, 0, -1) for customer-to-provider, peer and provider-to-customer**
- **In any path should only see sequence of +1, followed by at most one 0, followed by sequence of -1**
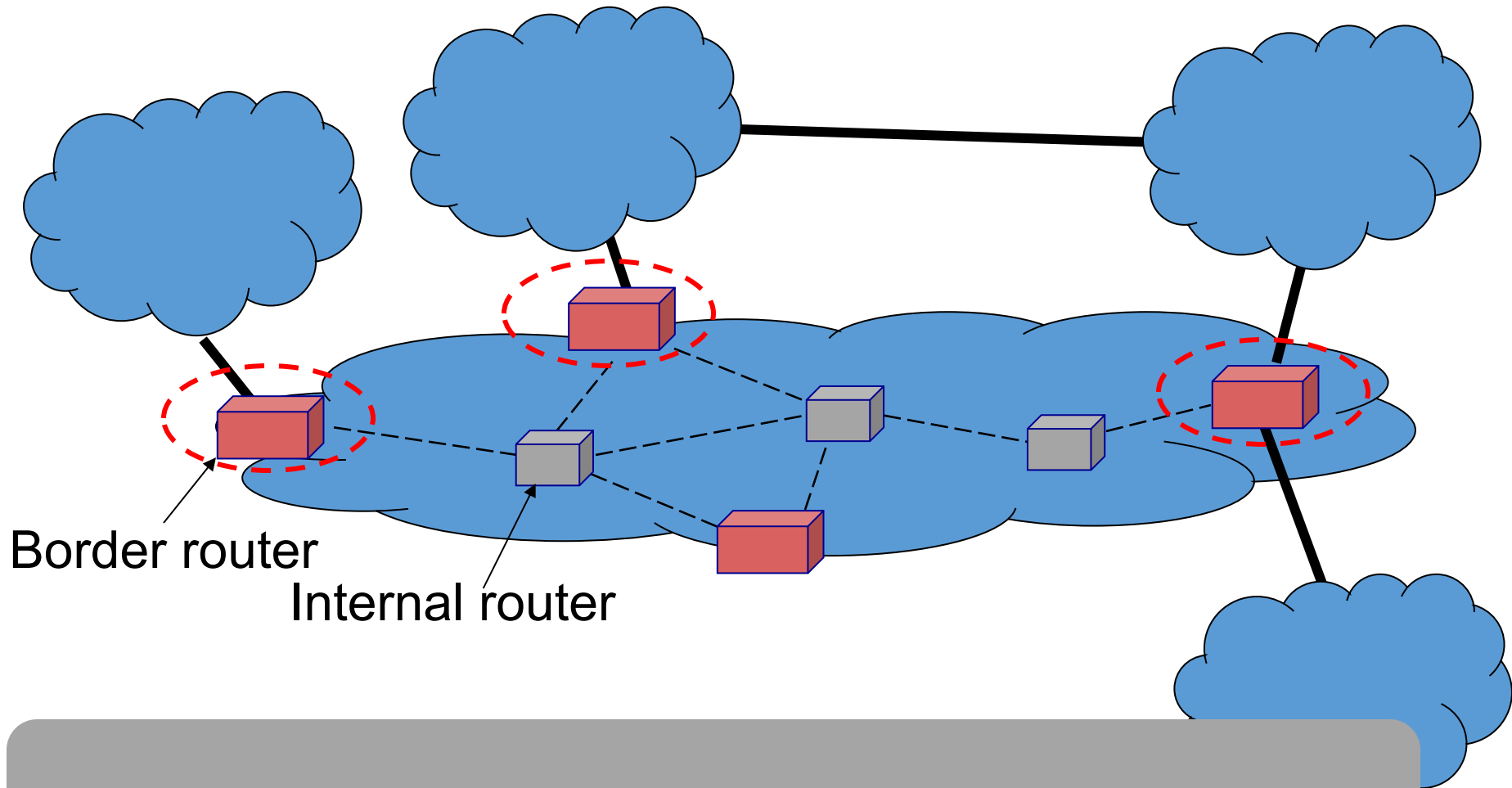
Example: Valley-Free Routing

Pr ◆━━▶ Cu

Peer ●━━● Peer

- The path is (-1, +1). It is not valley-free.

BGP Protocol details

Who speaks BGP?

Border router

Internal router

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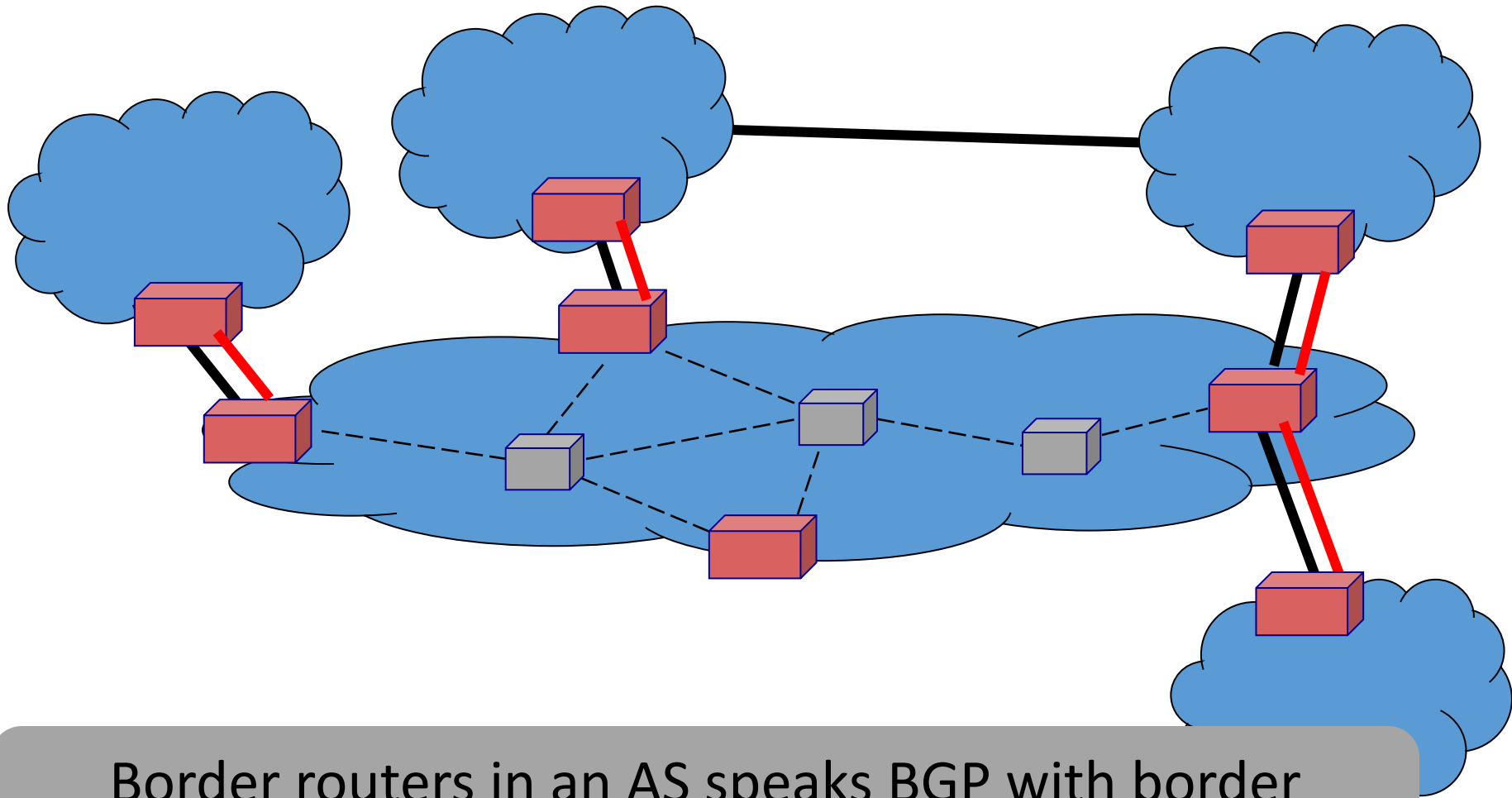
Border routers in an AS speaks BGP with border routers in other ASes using eBGP sessions

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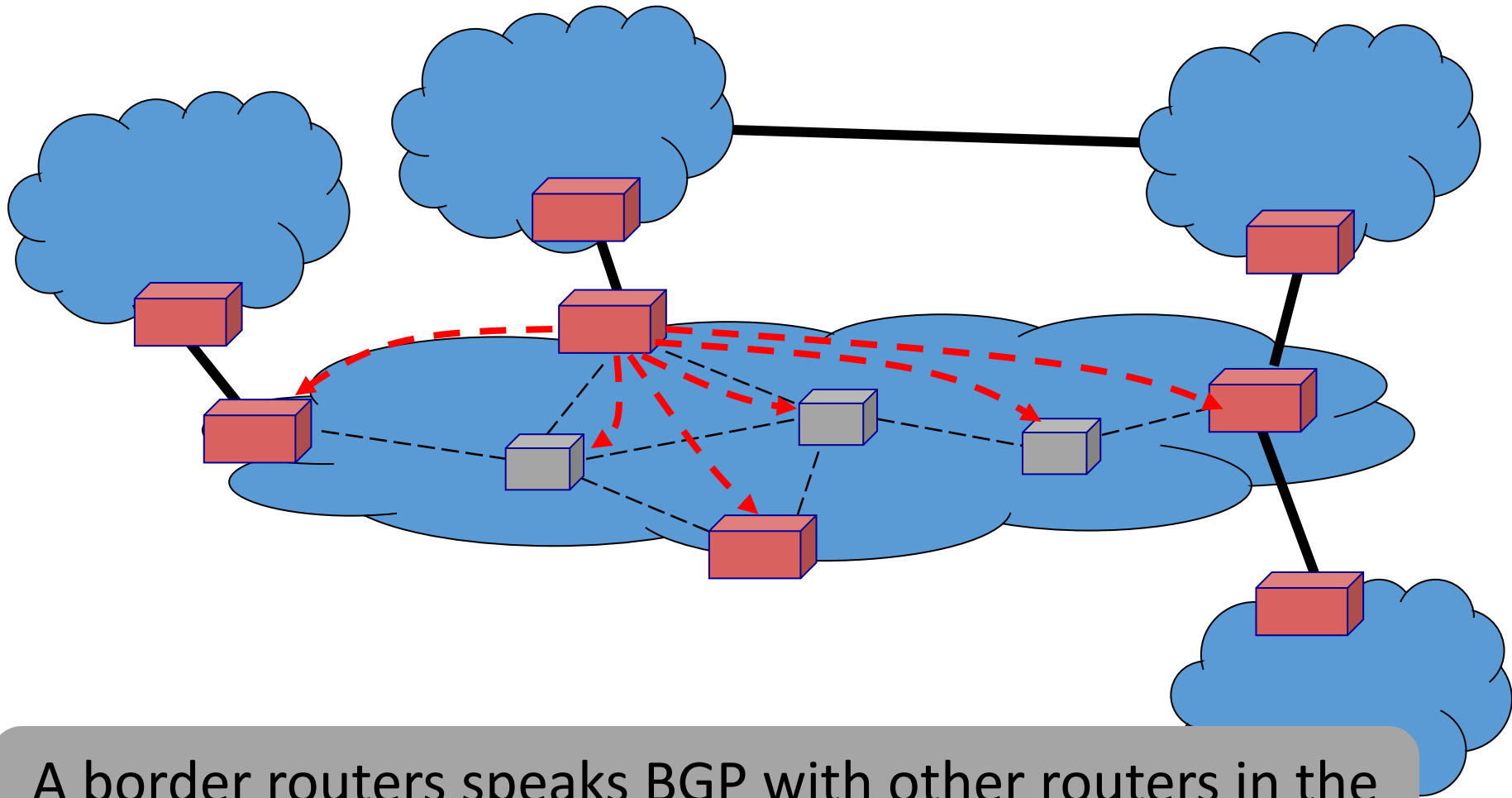
A border routers speaks BGP with other routers in the same AS using iBGP sessions

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
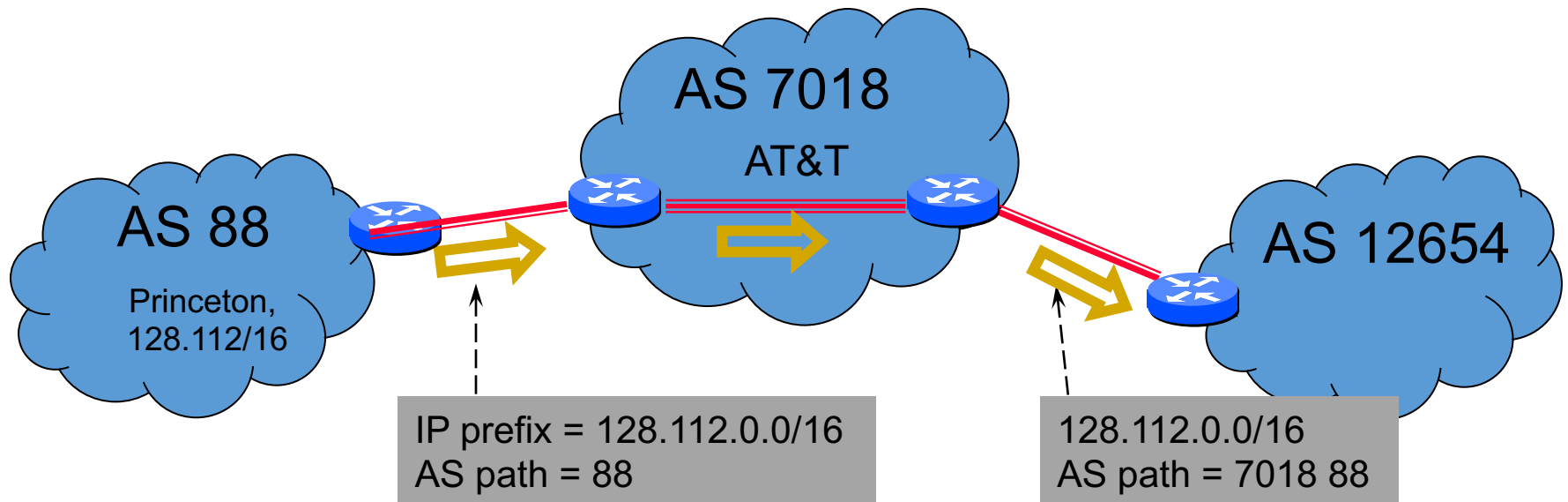
Route updates

- **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)**

AS 7018

AT&T

AS 88

Princeton,
128.112/16

AS 12654

IP prefix = 128.112.0.0/16
AS path = 88

128.112.0.0/16
AS path = 7018 88

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**
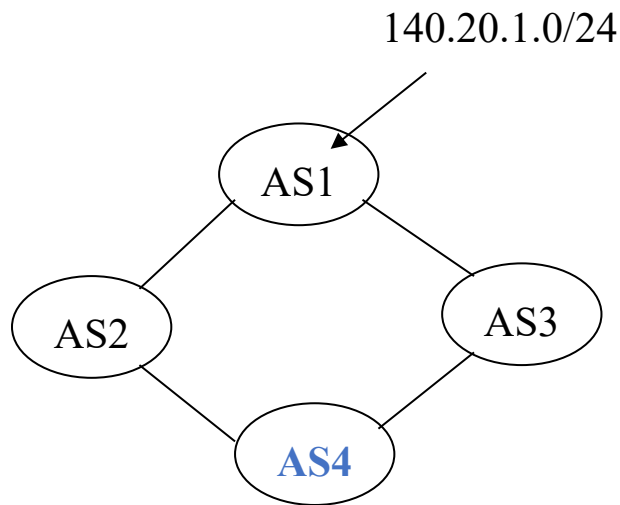
140.20.1.0/24

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**

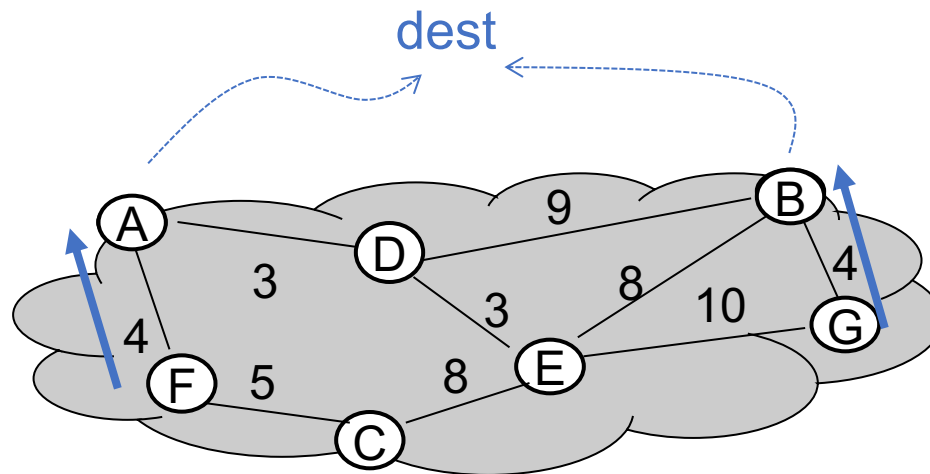- **AS receiving the prefix (optionally!) uses MED to select link**

AS1

Link B

Link A

MED=50

MED=10

AS2

AS3

destination prefix
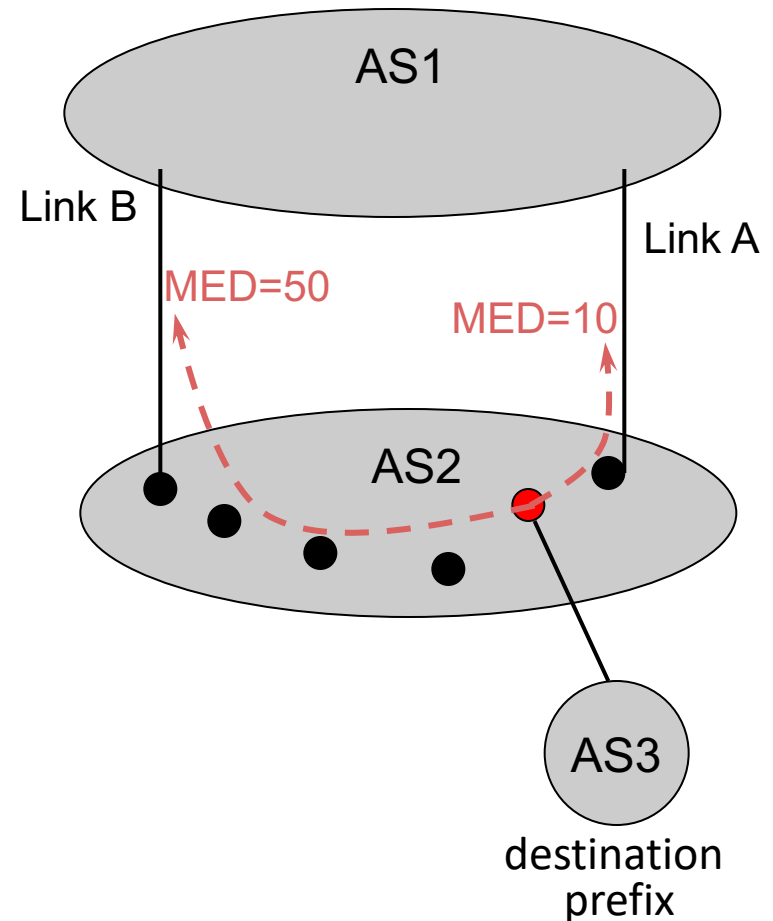
Attributes: (4) IGP cost

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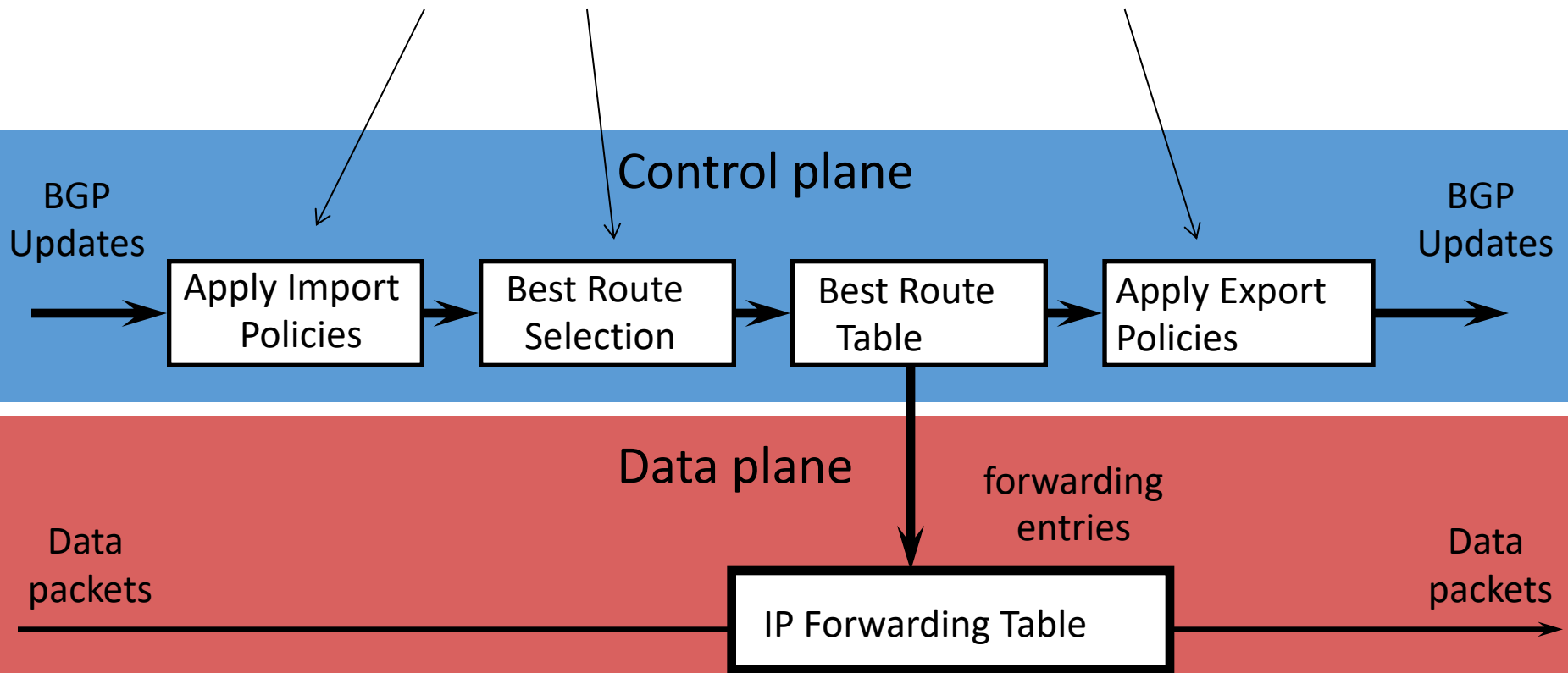
Using attributes

- **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

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
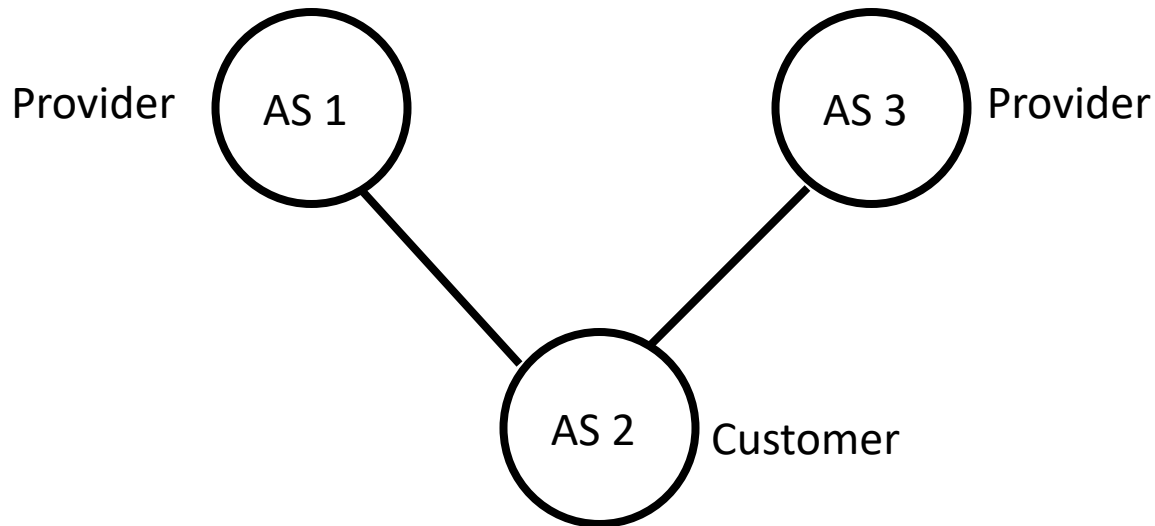entries

Data
packets

IP Forwarding Table

Data
packets

BGP issues 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**

Provider (AS 1)   (AS 3) Provider

(AS 2) Customer

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**
  - ➢A set of rules that decide the preferences of routes, e.g., prefer a route via a customer over a route via a provider or peer

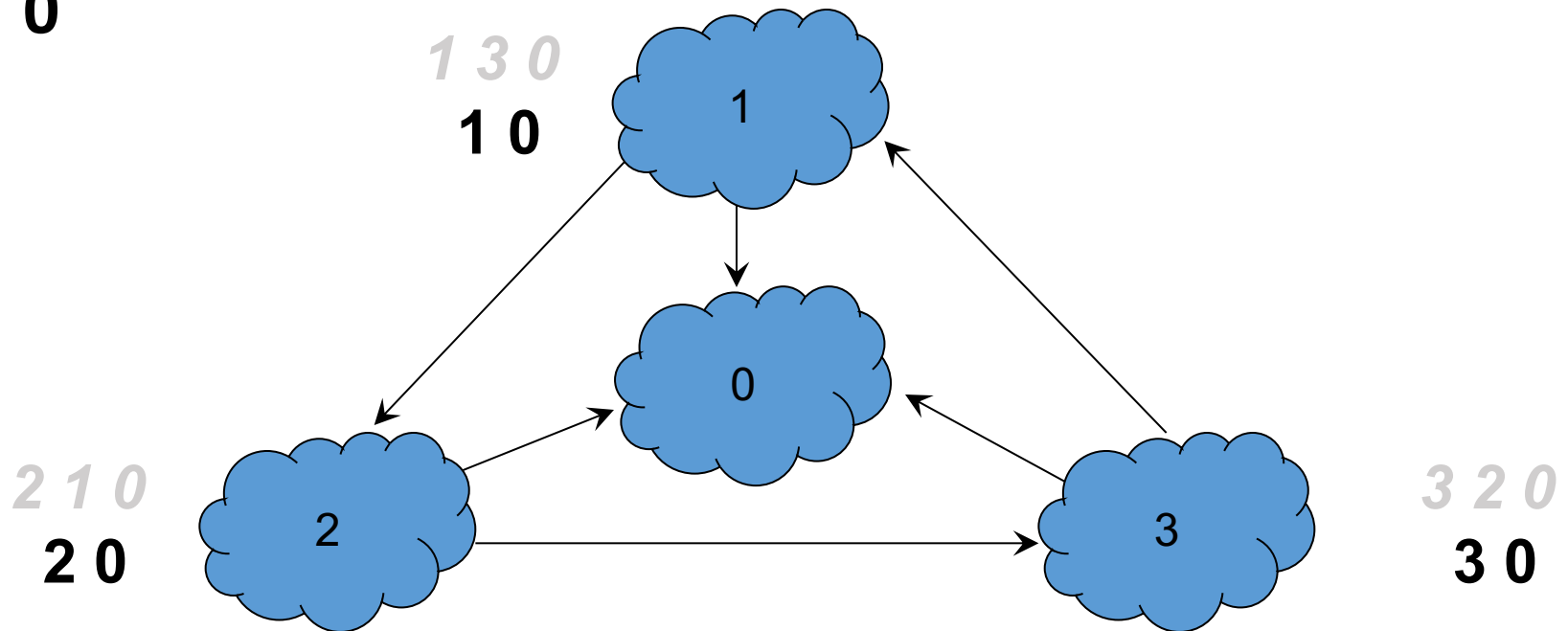- **For arbitrary policies, BGP may fail to converge!**
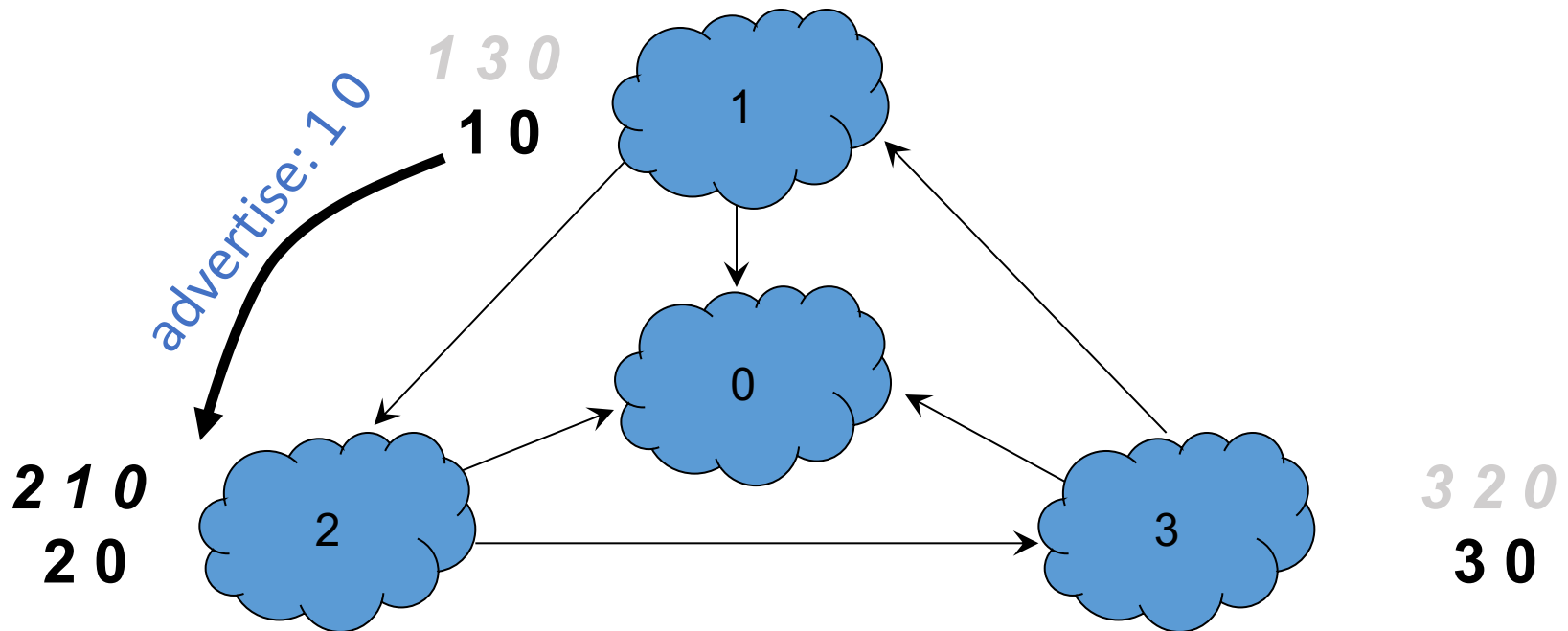
Example of policy oscillation
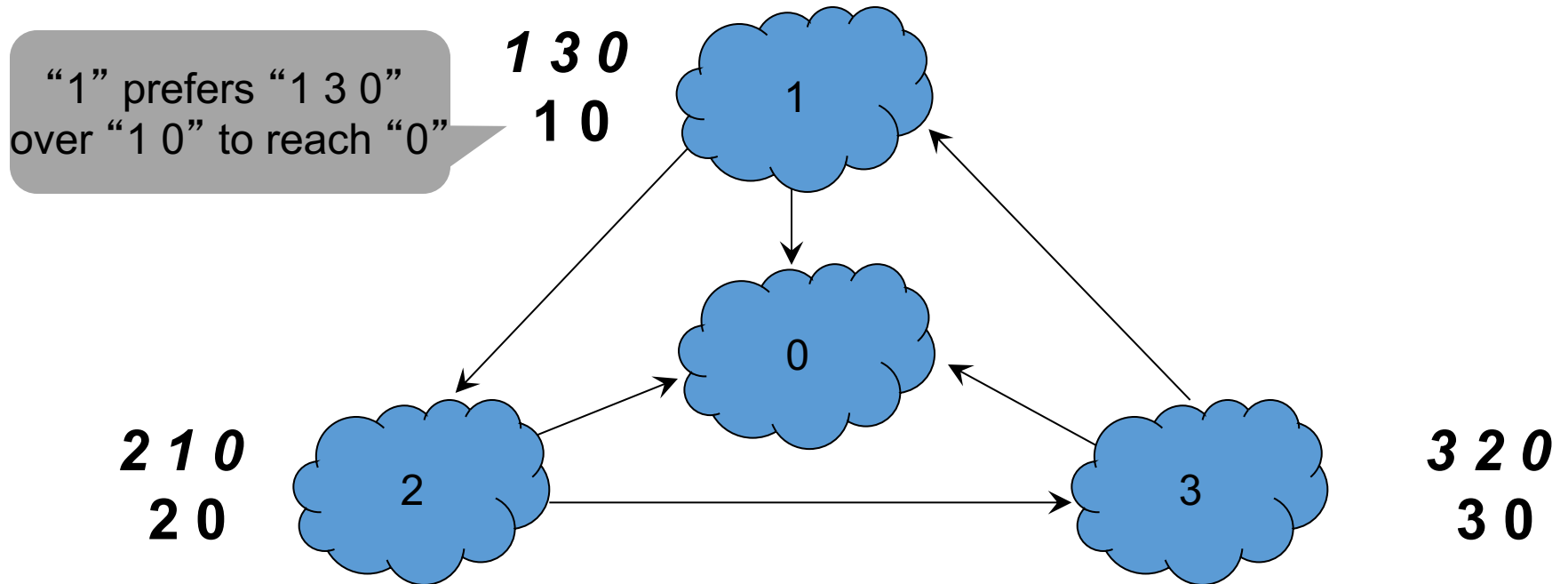
Step-by-step of policy oscillation

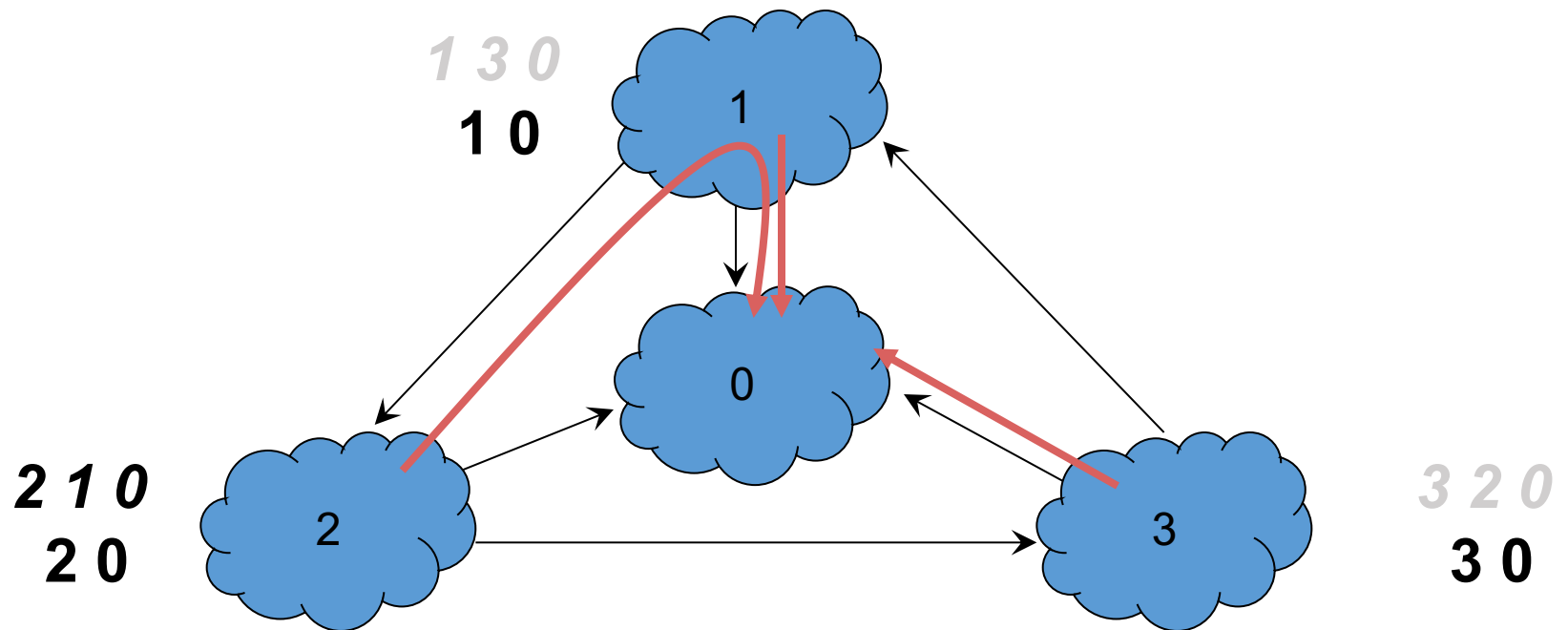- **Initially: nodes 1, 2, 3 know only shortest path to 0**

1 3 0

1 0

1

0

2 1 0

2 0

2

3 2 0

3 0

3

Step-by-step of policy oscillation

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

Step-by-step of policy oscillation

1 3 0
1 0

2 1 0
2 0

0

3 2 0
3 0

Step-by-step of policy oscillation

- **3 advertises its path 3 0 to 1**
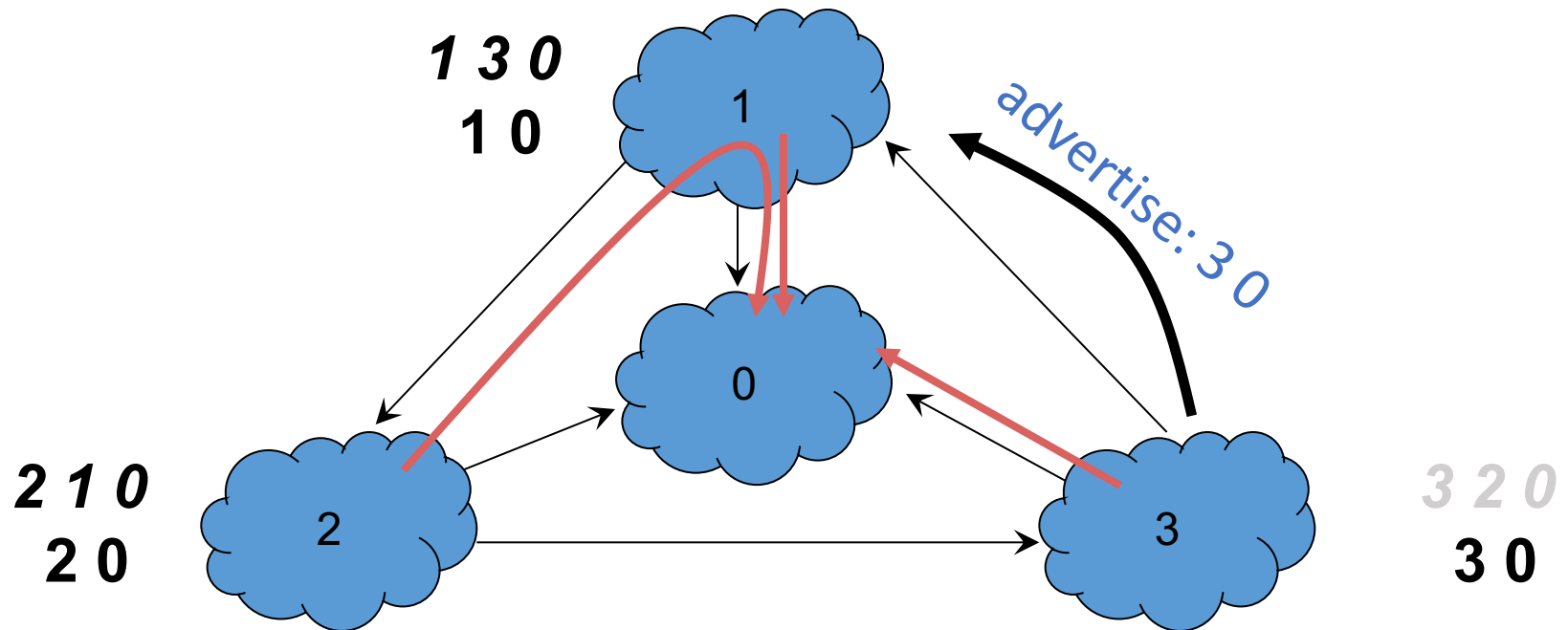
Step-by-step of policy oscillation
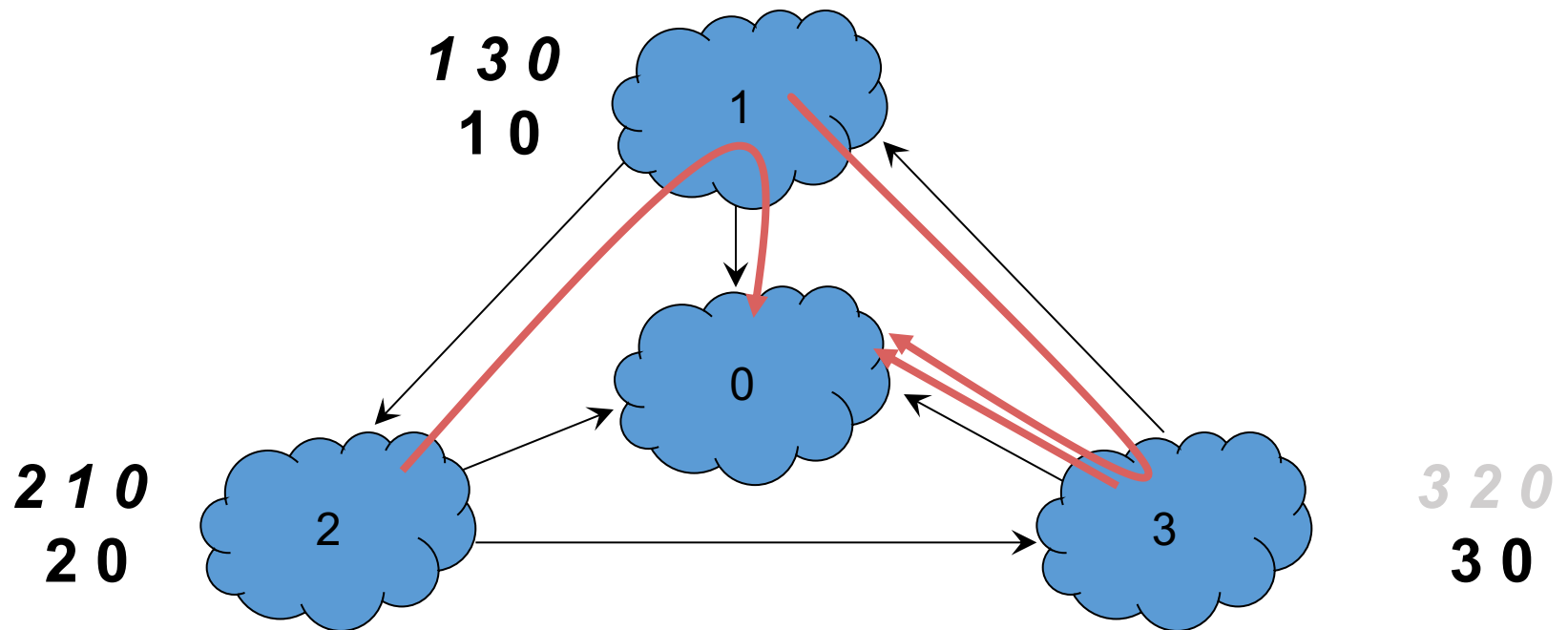
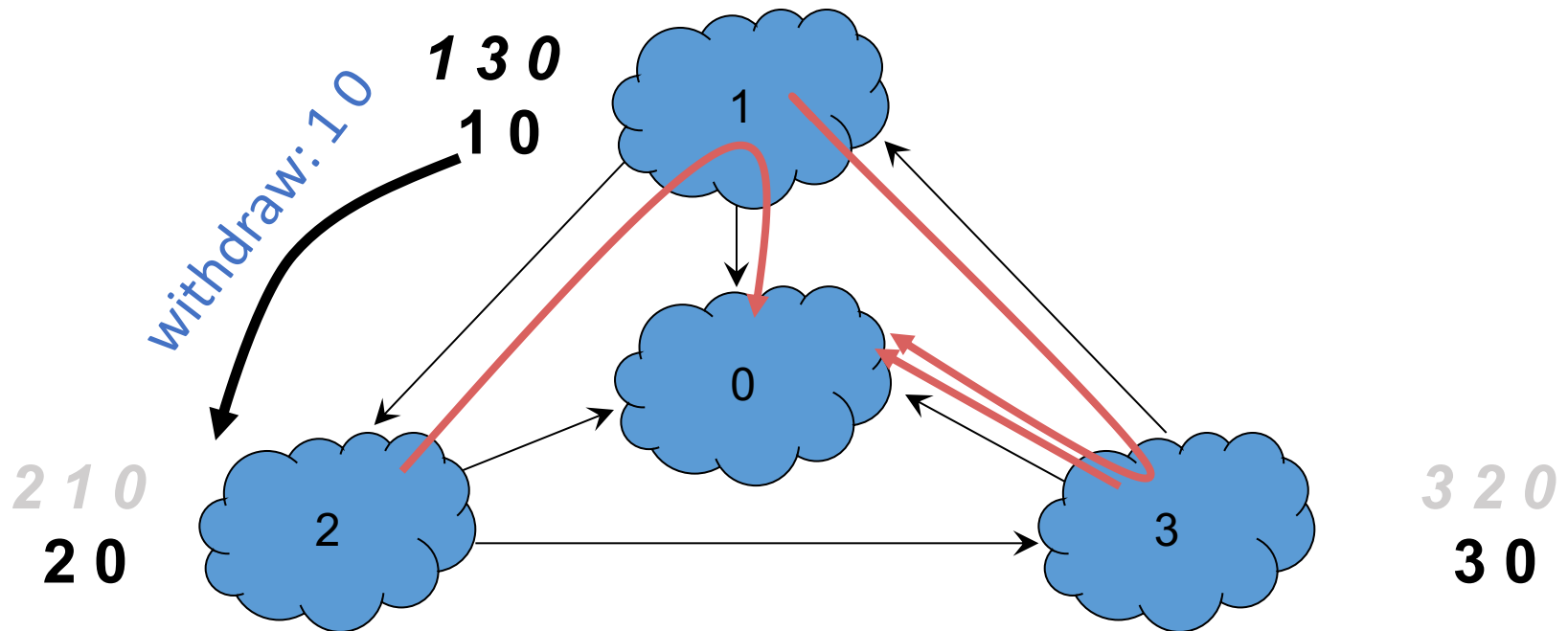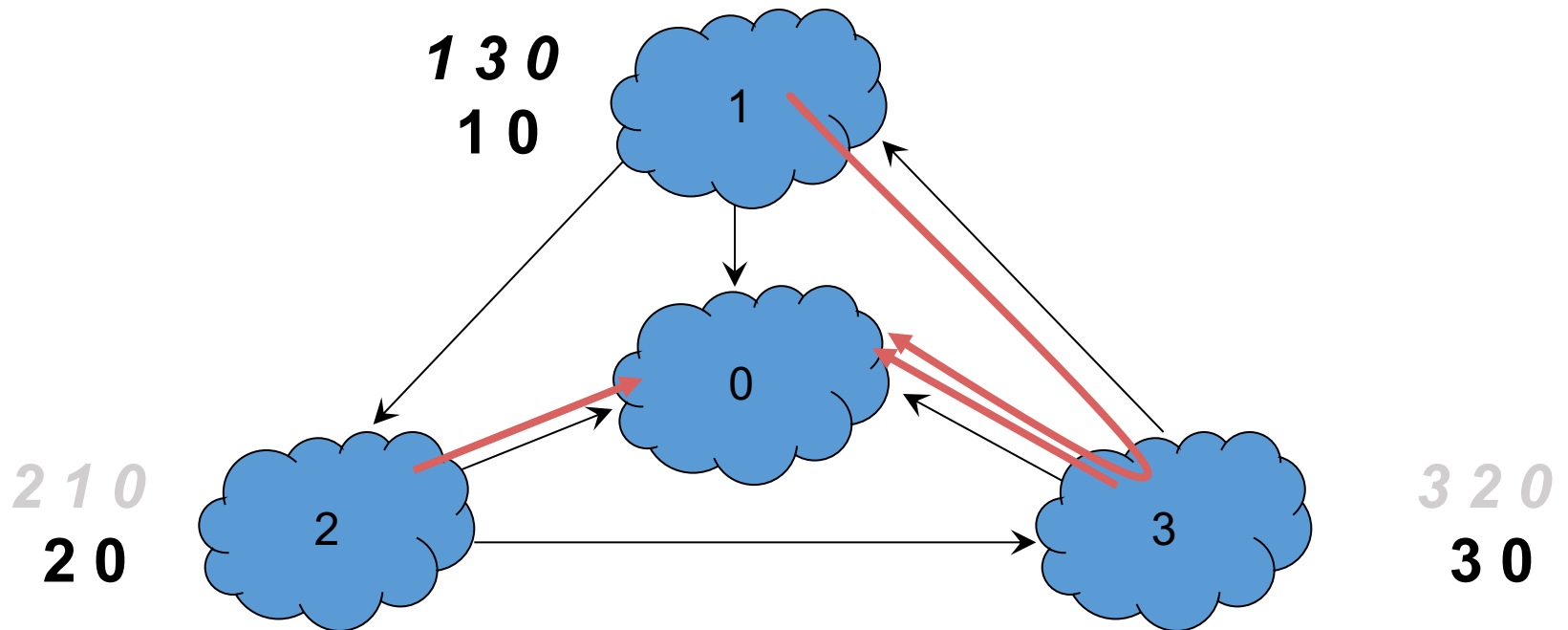# Step-by-step of policy oscillation
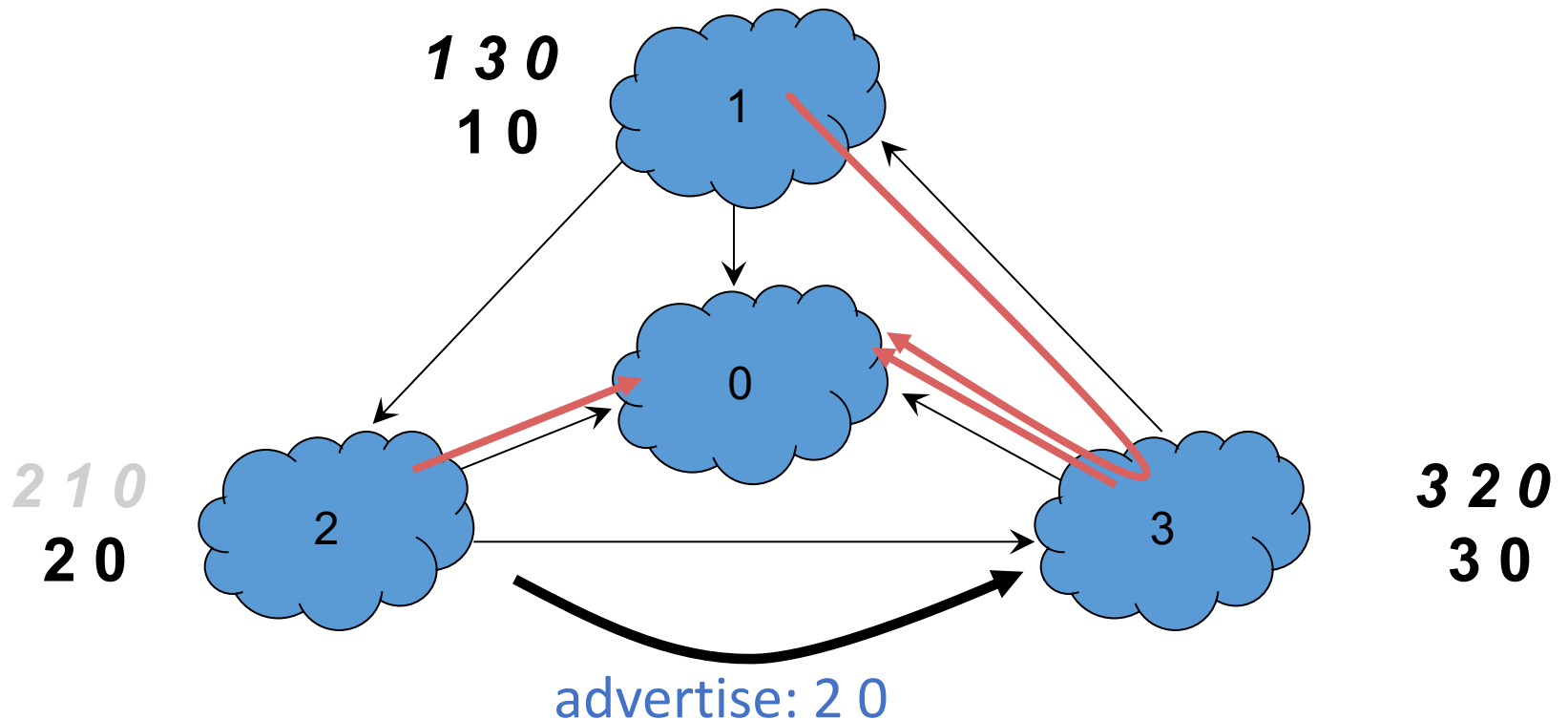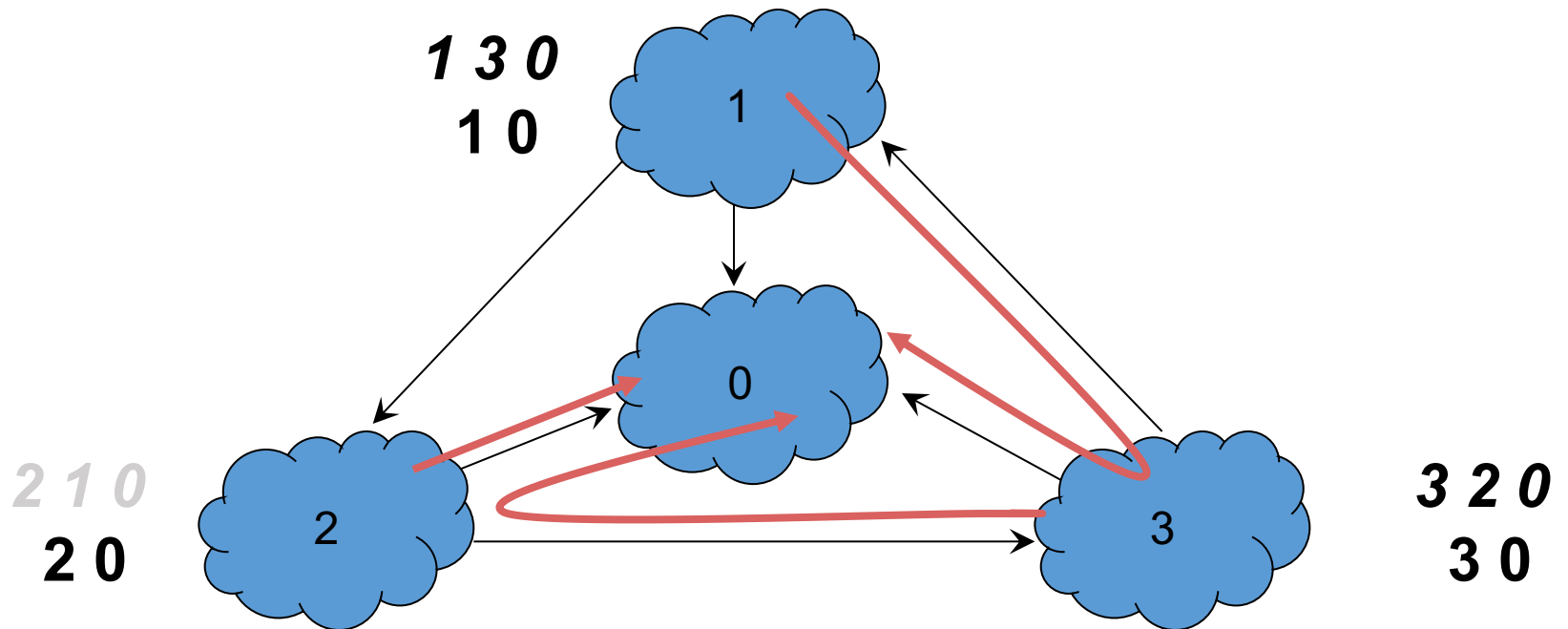
- **1 withdraws its path 1 0 from 2**

1 3 0
1 0

withdraw: 1 0

2 1 0
2 0

0

3 2 0
3 0

39

Step-by-step of policy oscillation

1 3 0
1 0

2 1 0
2 0

3 2 0
3 0

Step-by-step of policy oscillation

• **2 advertises its path 2 0 to 3**

advertise: 2 0

Step-by-step of policy oscillation

1 3 0
1 0

2 1 0
2 0

3 2 0
3 0

Step-by-step of policy oscillation

• **3 withdraws its path 3 0 from 1**
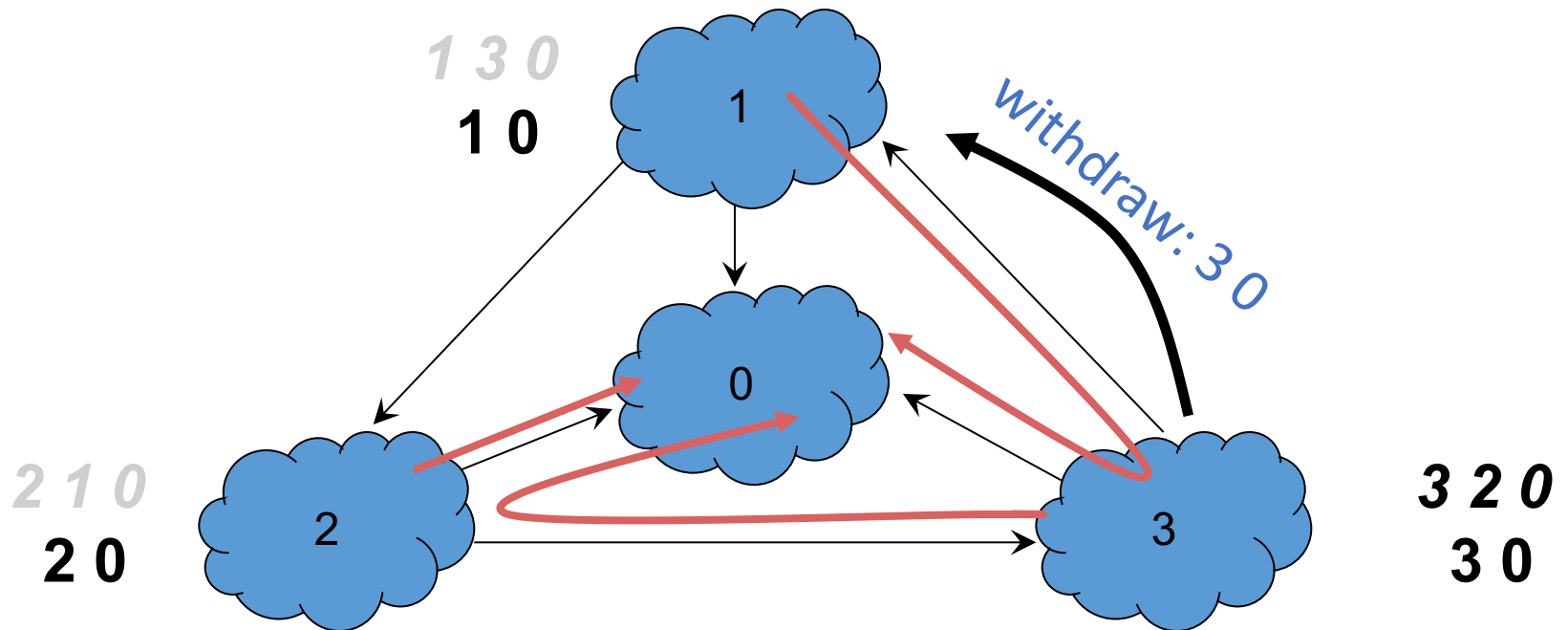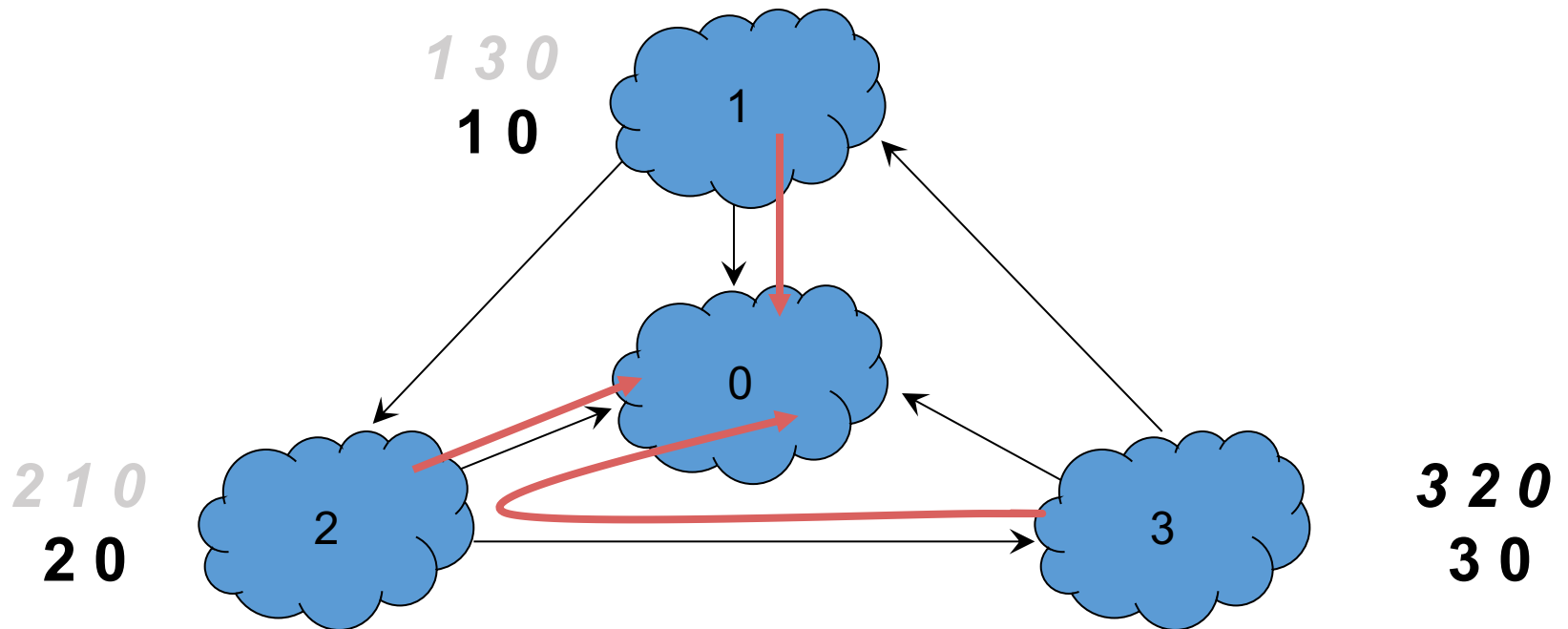
43

Step-by-step of policy oscillation

1 3 0
1 0

2 1 0
2 0

0

1

2

3

3 2 0
3 0

Step-by-step of policy oscillation

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

Step-by-step of policy oscillation

1 3 0

1 0

2 1 0
2 0

3 2 0
3 0
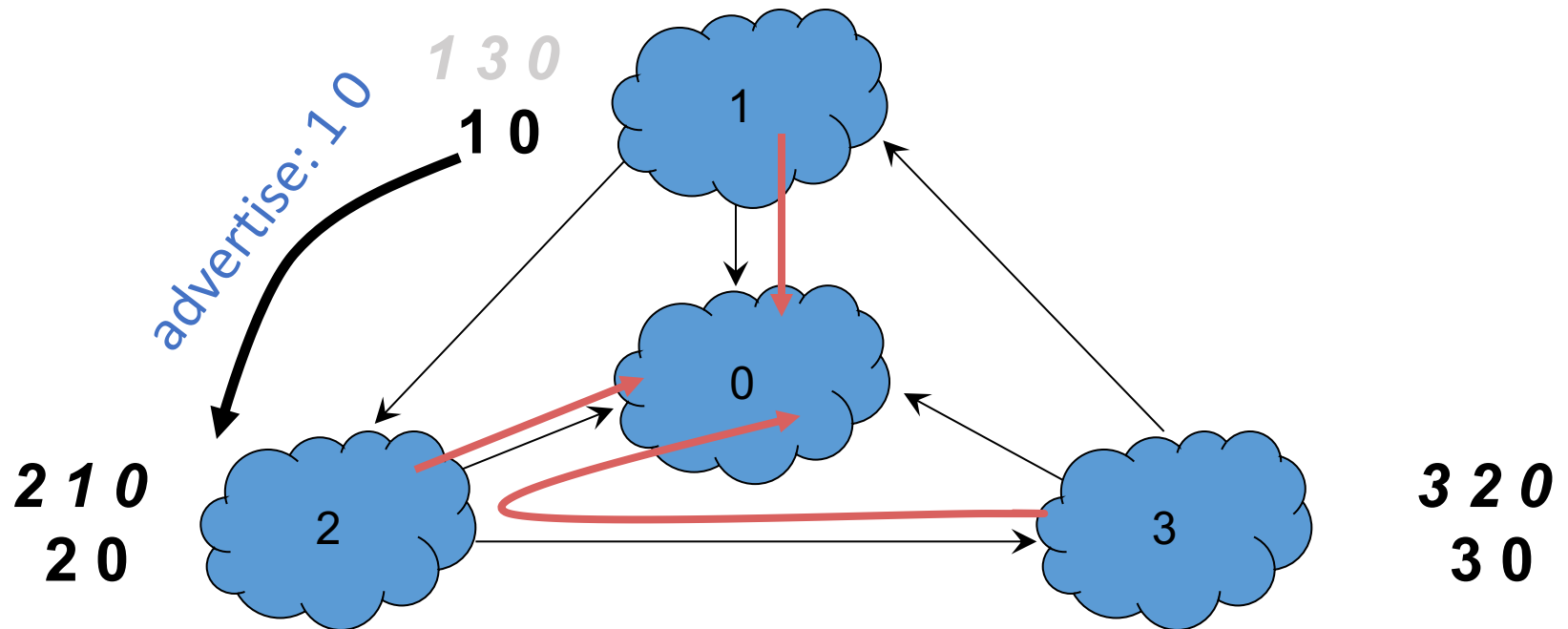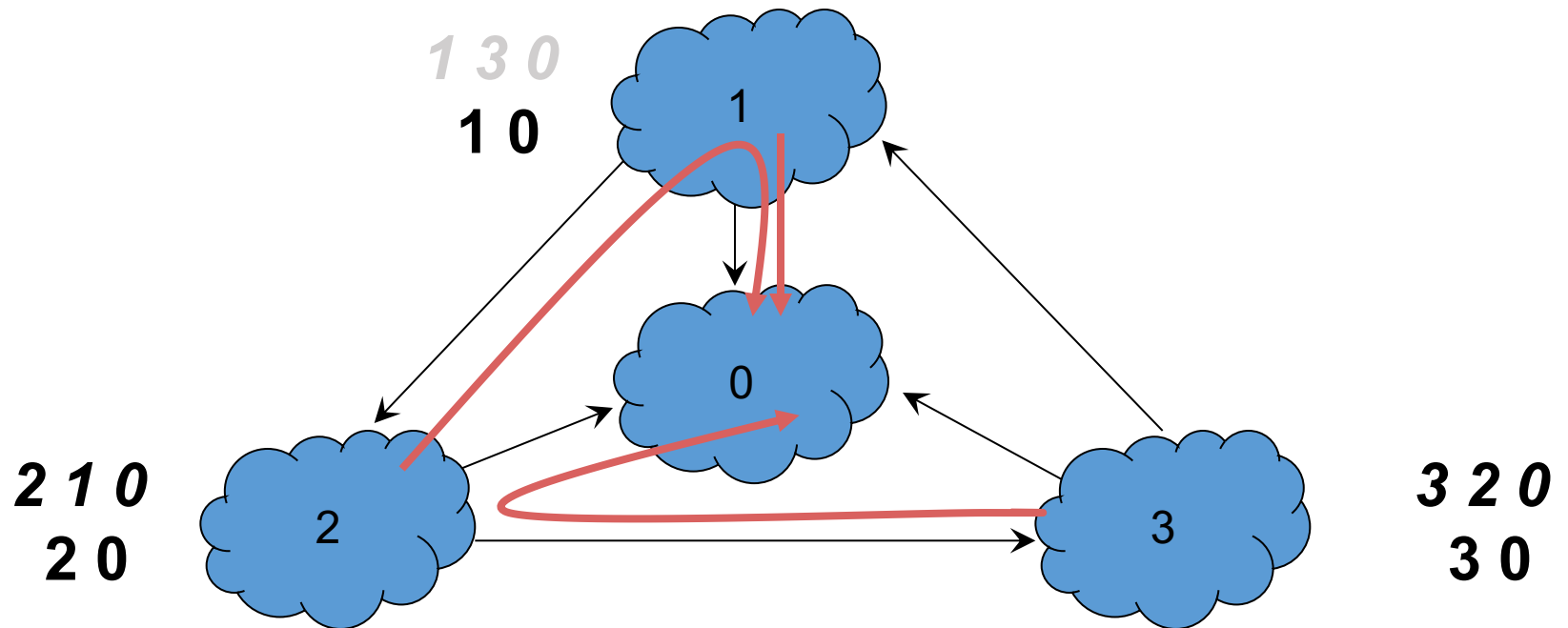
Step-by-step of policy oscillation

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

1 3 0
1 0

2 1 0
2 0

3 2 0
3 0

withdraw: 2 0
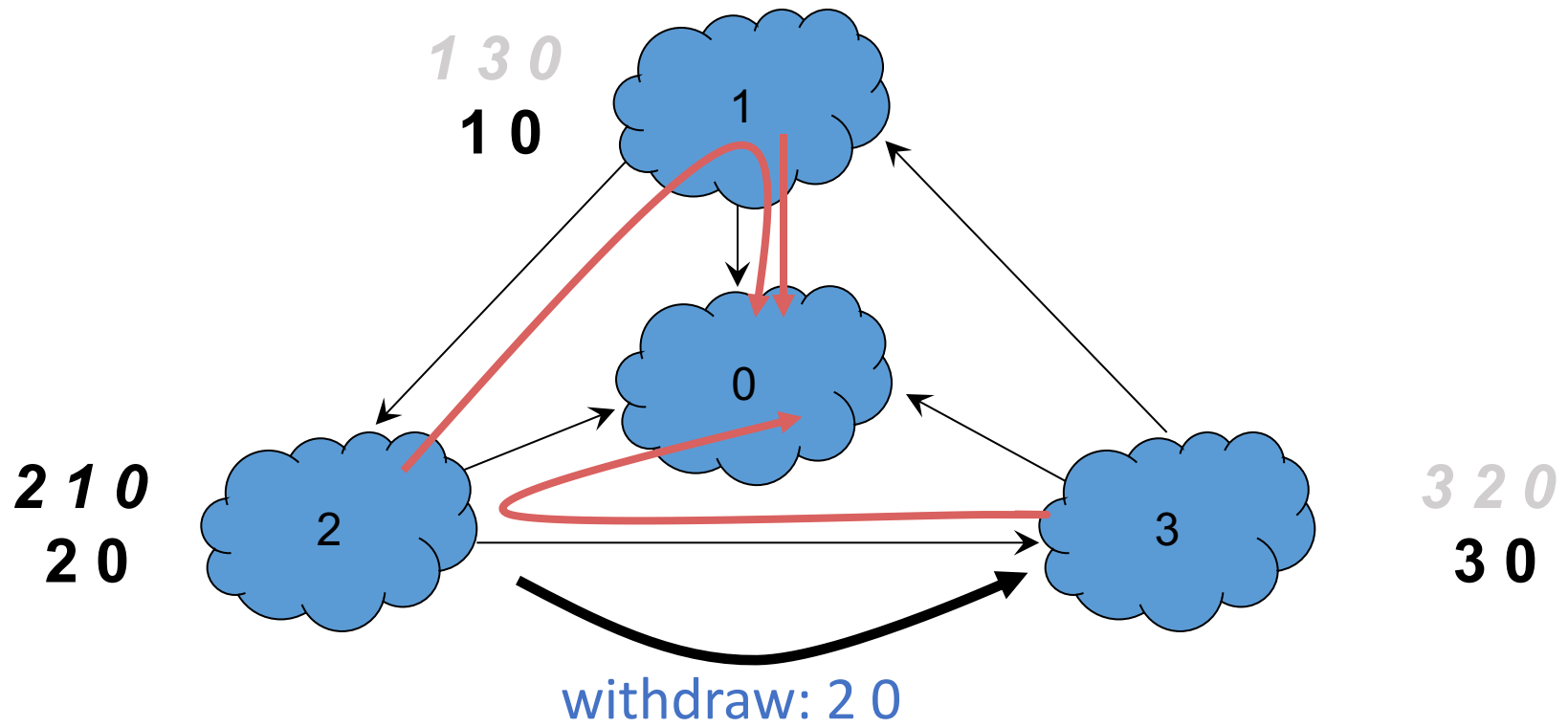
We're back to where we started
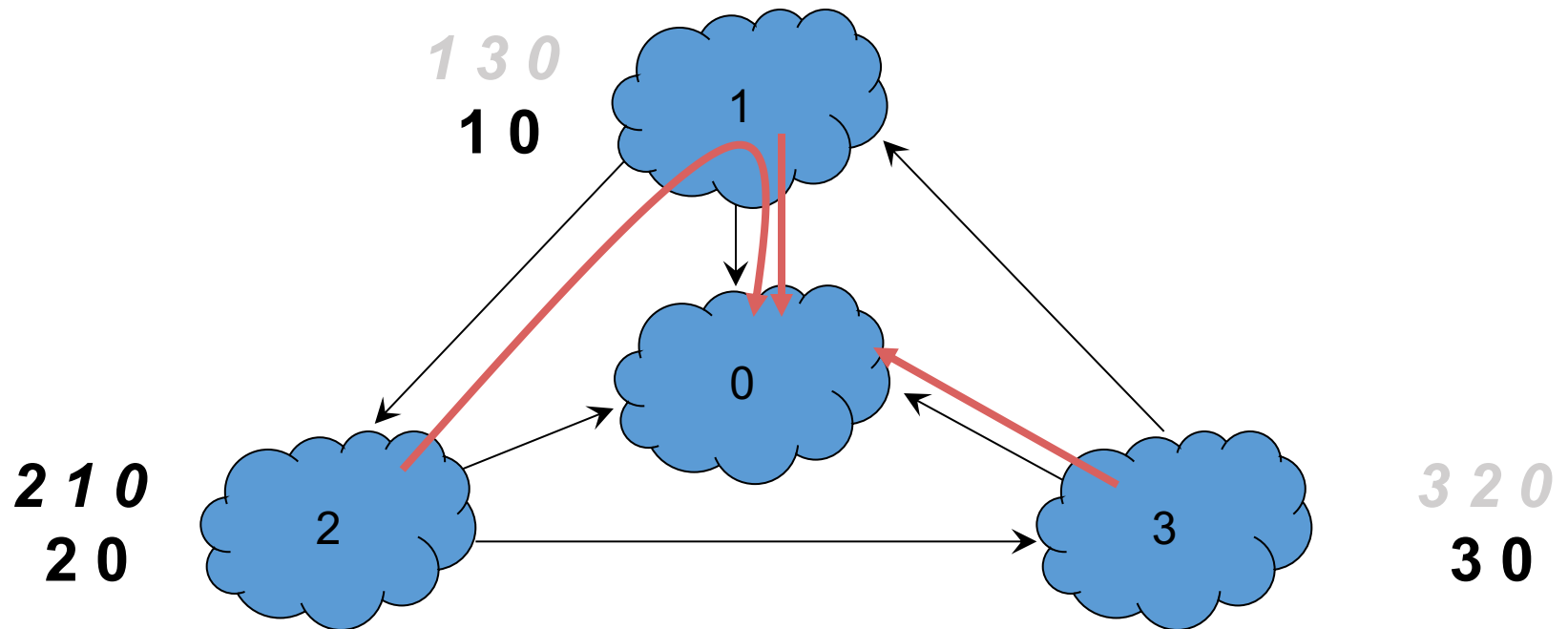
Convergence

- **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**

BGP says that path **4 1** is better than path **3 2 1**

AS 3

AS 2

AS 1

AS 4
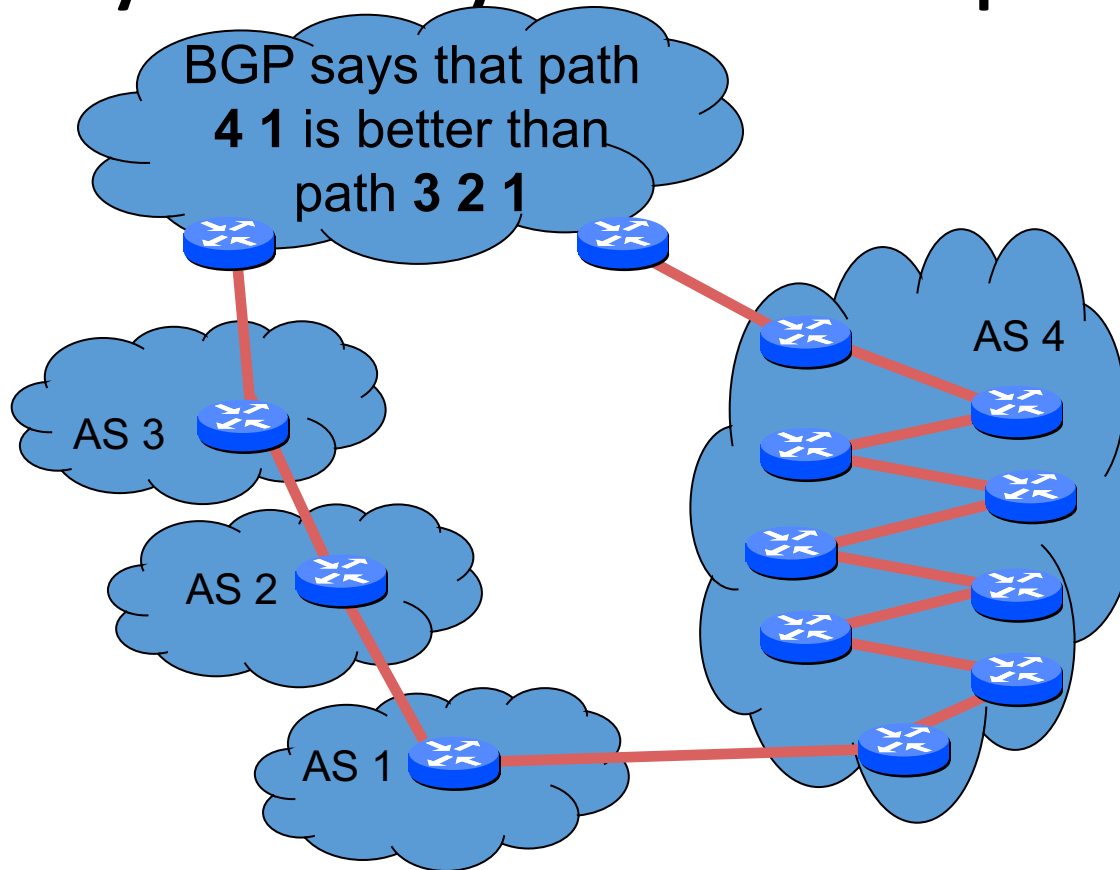
Real performance issue: Slow convergence

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

- **Most popular paths are very stable**

- **Outages are still very common**
  - ➢Check out https://bgpstream.com/

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

Group Discussion

- **Topic: BGP**
  - ➢What are the issues with BGP? Can you propose some ideas to fix them?

- **Discuss in groups, and each group chooses a leader to summarize the discussion**
  - ➢**Everyone should speak.**
  - ➢**Turn on your audio and video. Do not mute.**

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 lecture: Programmable Networks**

Thanks!
Q&A