CS 204: BGP

Jiasi Chen

Lectures: MWF 12:10-1pm

Humanities and Social Sciences 1403

http://www.cs.ucr.edu/~jiasi/teaching/cs204_spring17/

Overview

- AS relationships
- Inter-AS routing
  - BGP
  - Example
- Paper discussion

Q: How to “glue together” the “network of networks”?

Where did YouTube go?

Source: https://www.cnet.com/news/how-pakistan-knocked-youtube-offline-and-how-to-make-sure-it-never-happens-again/

Where did YouTube go?

- In 2008, Pakistani government decided to block YouTube
- Pakistan Telecom (PT) began advertising a route to YouTube
  - Advertised 256 addresses within YouTube's IP block
  - Actually led to a "black hole"
- A Hong Kong-based telecom company picked up the advertisement
- Announcement spread to other major ISPs within 10s of seconds
- YouTube countermeasures
  - Advertise 64 addresses within YouTube's IP block
  - More specific rule should override general one
- Full recovery after ~2 hours after PT stopped advertising the route

Source: https://www.cnet.com/news/how-pakistan-knocked-youtube-offline-and-how-to-make-sure-it-never-happens-again/

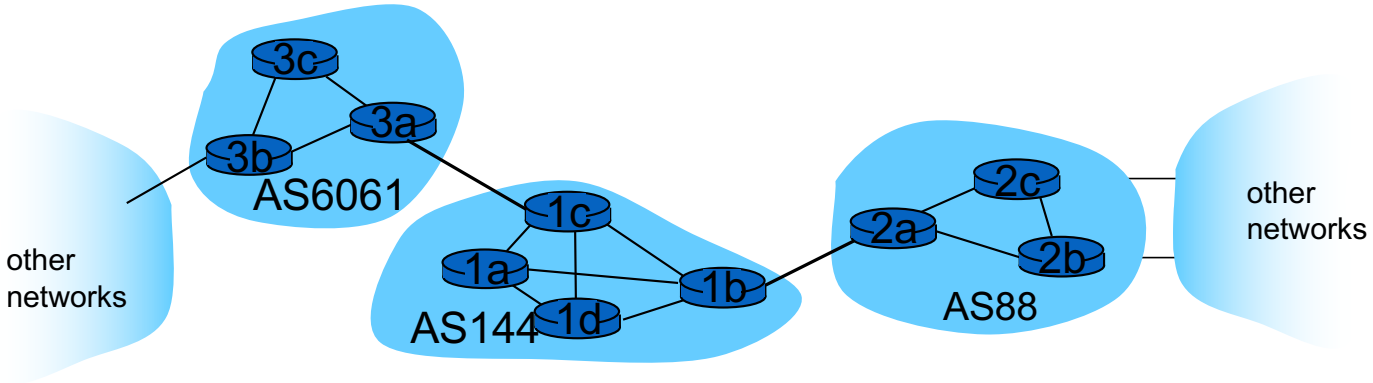
Overview

- AS relationships
- Inter-AS routing
  - BGP
  - Example
- Paper discussion

Q: How to “glue together” the “network of networks”?

Autonomous Systems

- Autonomous system (AS)
  - Unit of routing policy
  - ~50k ASes in use
  - E.g., UCR has AS#6061, AT&T has AS#144, Princeton has AS#88

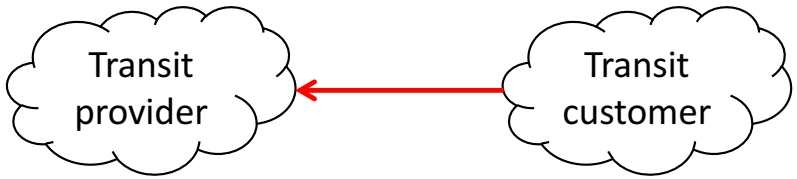


Peering and Transit ISPs

- Peering

- Traffic flows are bi-directional
- ISPs jointly pay for equipment costs

- Transit



- Traffic flows are bi-directional
- Arrow = Payment from customer to provider for upstream and downstream traffic

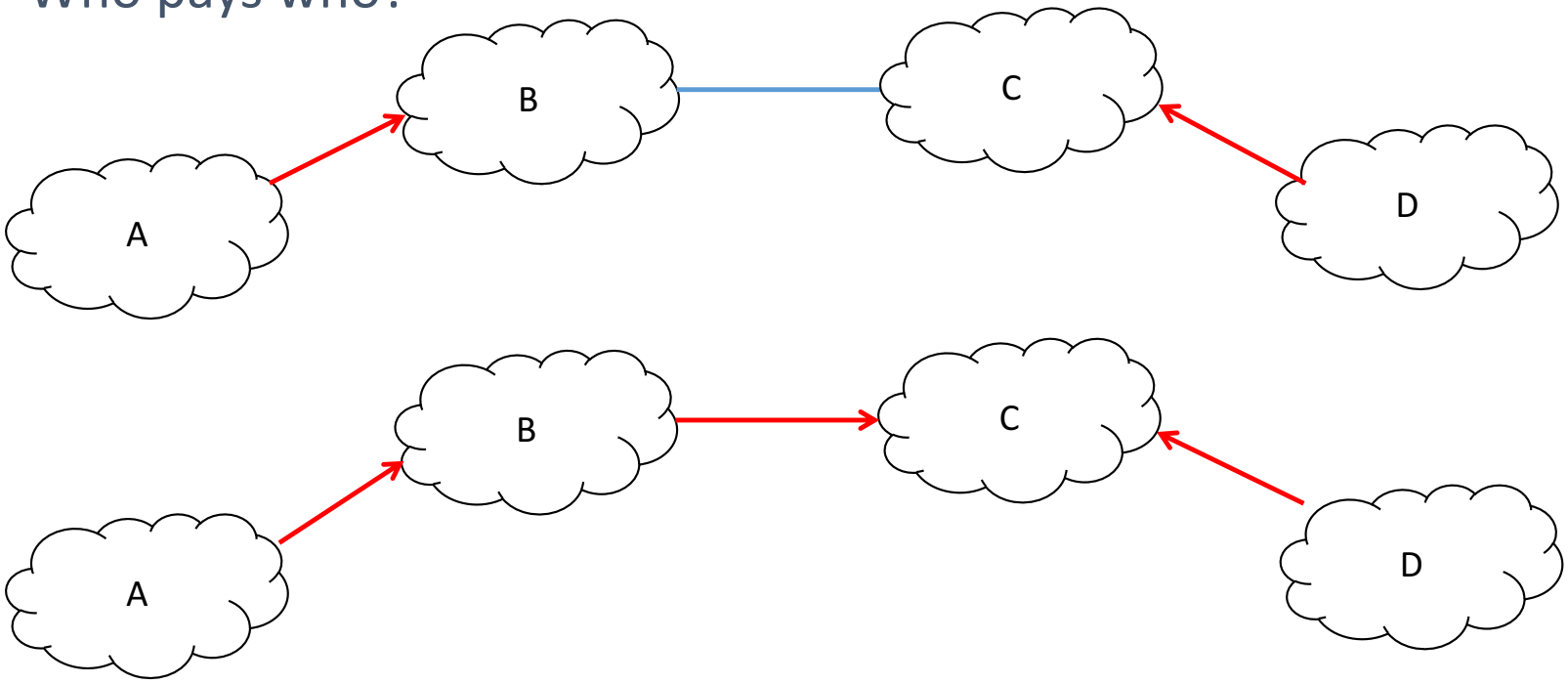
Pricing Contract

- Pricing contracts are typically not public information
- $/Mbps/month for all traffic to all destinations

- Variations
  - Paid peering
    - E.g. Netflix paid Comcast for direct peering
  - Backplane peering
    - Charge small ISPs for access to ISP's peers
  - Regional pricing
    - Pay to access different geographical regions, own customers vs external ISPs

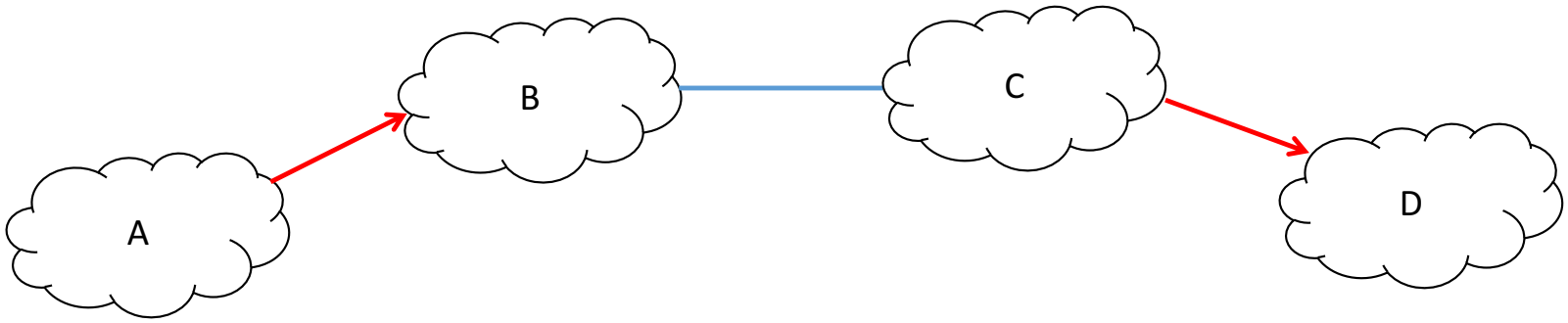
Examples

- Does traffic flow between A and D?
- Who pays who?

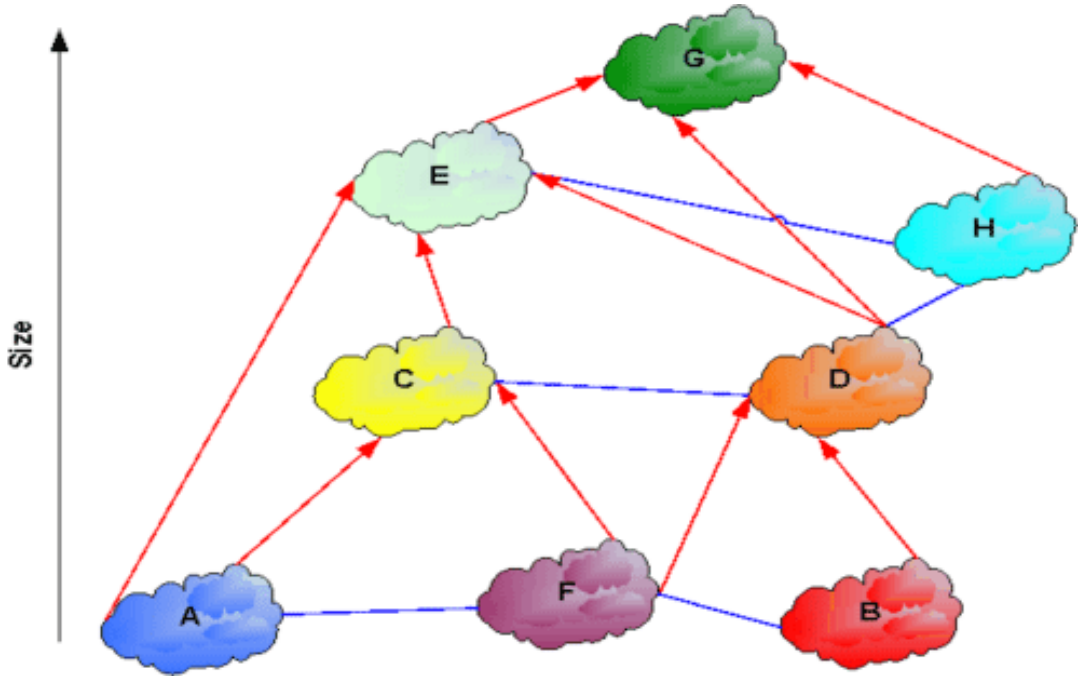


Examples (2)

Peering connection only open to customers



Q: Will C announce B to D?



Source: https://arstechnica.com/features/2008/09/peering-and-transit/2/

Who can network G see?

- Network G can see all the networks because networks E, D and H buy transit from it.

Can A see B through F?

- Network A can see network F and its customers directly, but not network B through network F.

Can C see B through D or F?

- Network C can see Network B through its peer D, but not via its transit customer F.

Will traffic from C to H go through E or D or both?

- Traffic from C to H will go through E, but not through D.

Internet Exchange Points (IXP)

- When two networks peer, it attracts other networks to peer there too
- Transit providers
- Direct connection between ISPs still preferred
- Run as non-profits (Europe) or private business (USA)
  - Provide network equipment, switches, etc.
  - Monthly fee to join the IXP

Overview

- AS relationships
- Inter-AS routing
  - BGP
  - Example
- Paper discussion

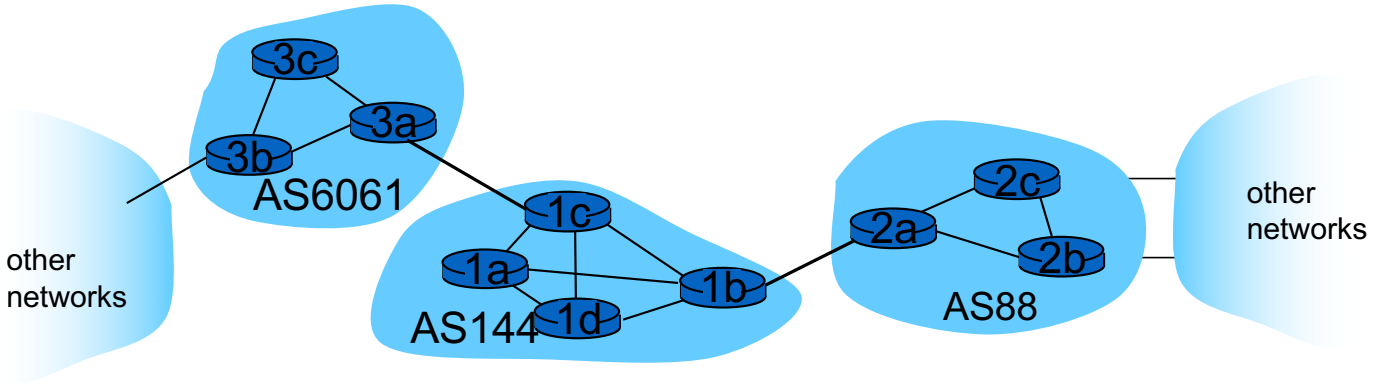
Q: How to “glue together” the “network of networks”?

Review of Routing

- Inter-AS routing
  - BGP
- Intra-AS routing
  - RIP
  - OSPF

Link-state?

Distance vector?



Why different Intra-, Inter-AS routing ?

policy:

- inter-AS: admin wants control over how its traffic routed, who routes through its net.
- intra-AS: single admin, so no policy decisions needed

scale:

- hierarchical routing saves table size, reduced update traffic

performance:

- intra-AS: can focus on performance
- inter-AS: policy may dominate over performance

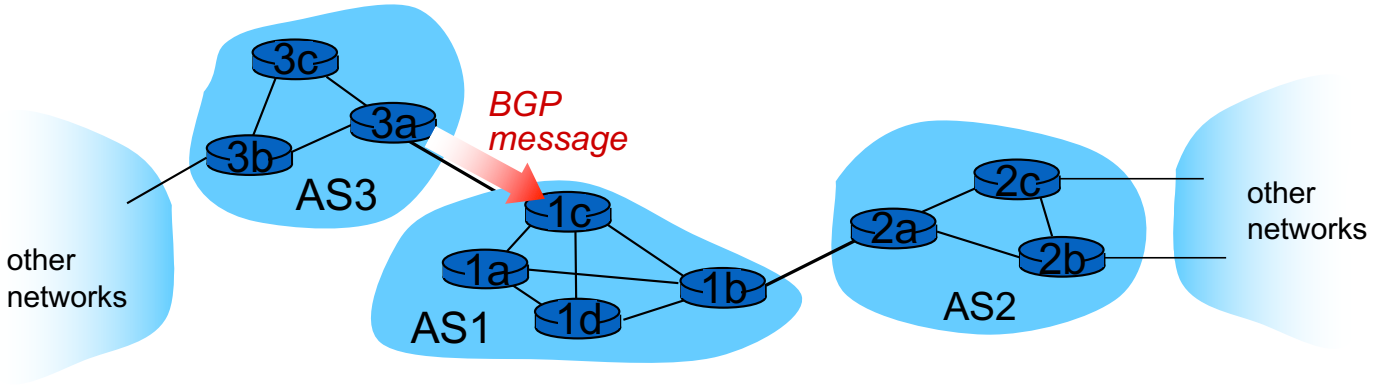
Internet inter-AS routing: BGP

- BGP (Border Gateway Protocol): *the* de facto inter-domain routing protocol
  - "glue that holds the Internet together"
- BGP provides each AS a means to:
  - eBGP: obtain subnet reachability information from neighboring ASs.
  - iBGP: propagate reachability information to all AS-internal routers.
  - determine "good" routes to other networks based on reachability information and policy.
- allows subnet to advertise its existence to rest of Internet: *"I am here"*

BGP basics

- BGP session: two BGP routers ("peers") exchange BGP messages:
  - advertising *paths* to different destination network prefixes ("path vector" protocol)
  - exchanged over semi-permanent TCP connections

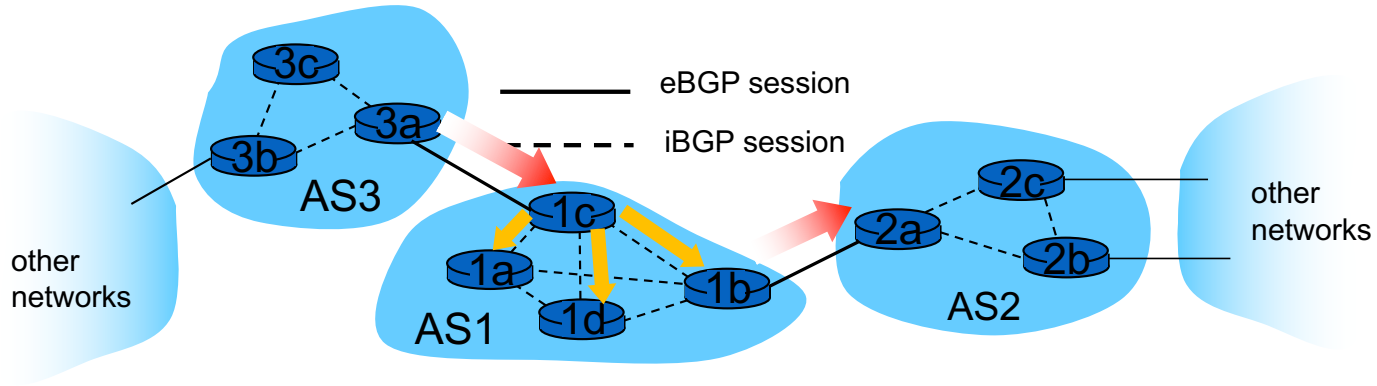
- when AS3 advertises a prefix to AS1:
  - AS3 *promises* it will forward datagrams towards that prefix
  - AS3 can aggregate prefixes in its advertisement



Path attributes and BGP routes

- advertised prefix includes BGP attributes
  - prefix + attributes = "route"
- two important attributes:
  - AS-PATH: contains ASs through which prefix advertisement has passed
  - NEXT-HOP: indicates specific internal-AS router to next-hop AS
- Example
  - Prefix: 138.16.64/22
  - AS-PATH: AS3 AS15 …
  - NEXT-HOP: 201.44.13.125

BGP basics: distributing path information



BGP messages

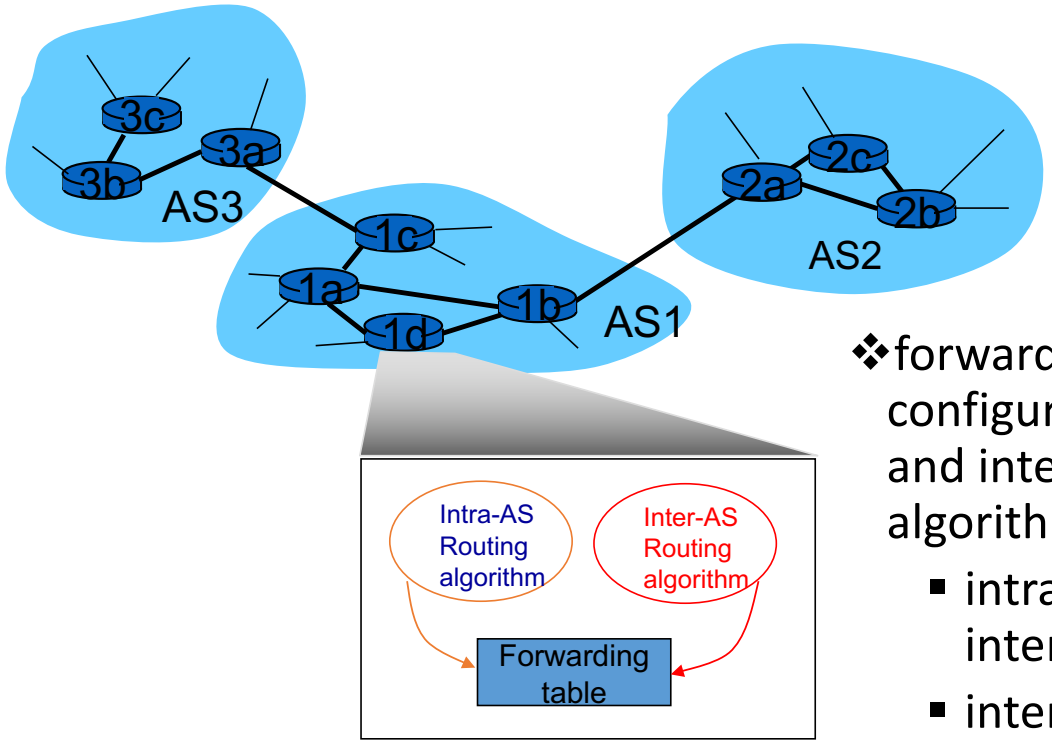
- BGP messages exchanged between peers over TCP connection
- BGP messages:
  - OPEN: opens TCP connection to peer and authenticates sender
  - UPDATE: advertises new path (or withdraws old)
  - KEEPALIVE: keeps connection alive in absence of UPDATES; also ACKs OPEN request
  - NOTIFICATION: reports errors in previous msg; also used to close connection

Overview

- AS relationships
- Inter-AS routing
  - BGP
  - Example
- Paper discussion

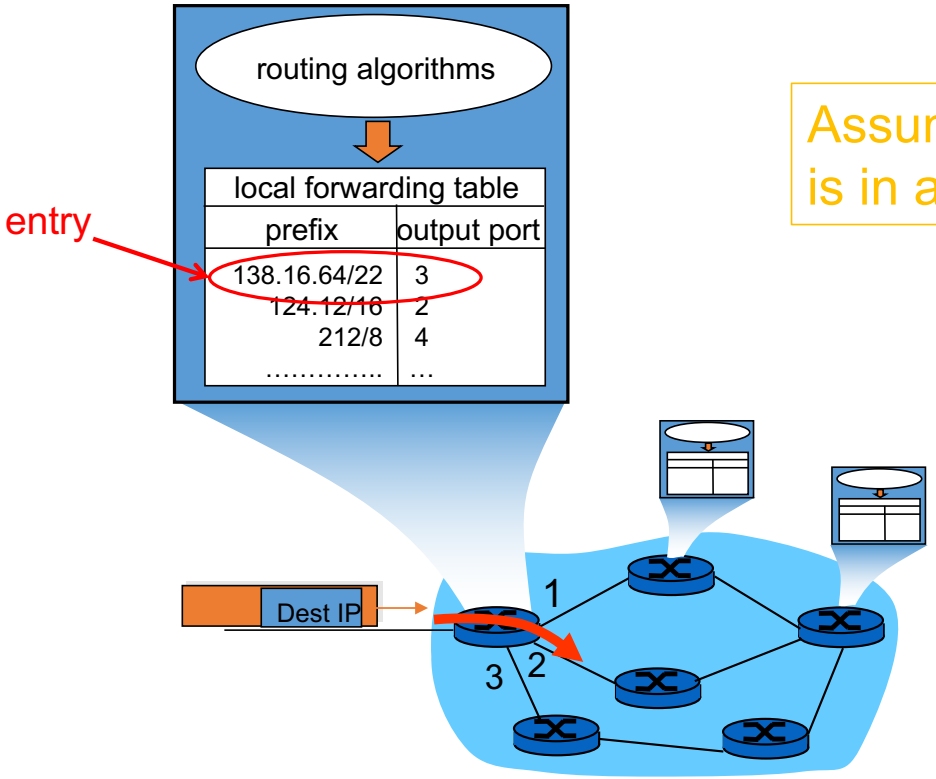
Q: How to “glue together” the “network of networks”?

Interconnected ASes



- forwarding table configured by both intra- and inter-AS routing algorithm
  - intra-AS sets entries for internal dests
  - inter-AS & intra-AS sets entries for external dests

How does entry get in forwarding table?



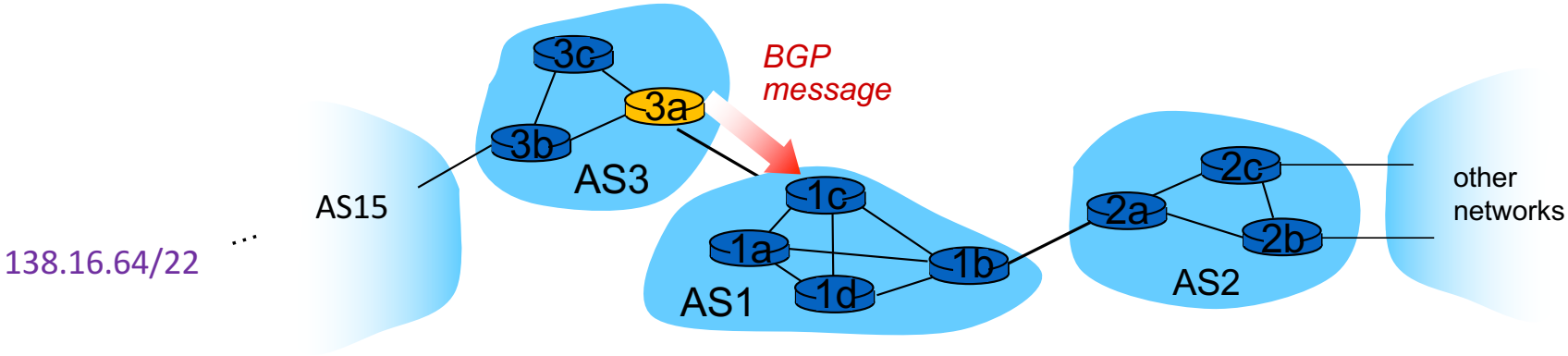
Assume prefix 138.16.64/22 is in another AS.

How does entry get in forwarding table?

High-level overview

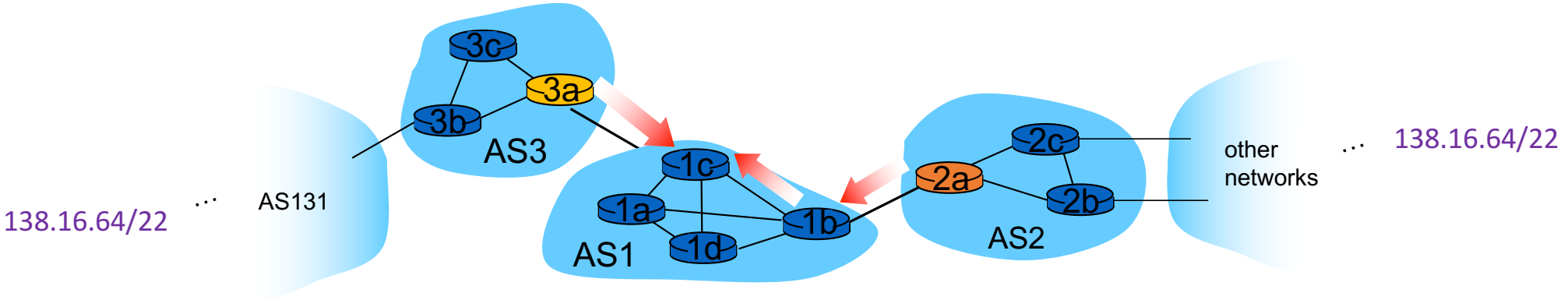
1. Router becomes aware of prefix
2. Router determines output port for prefix
3. Router enters prefix-port in forwarding table

Router becomes aware of prefix



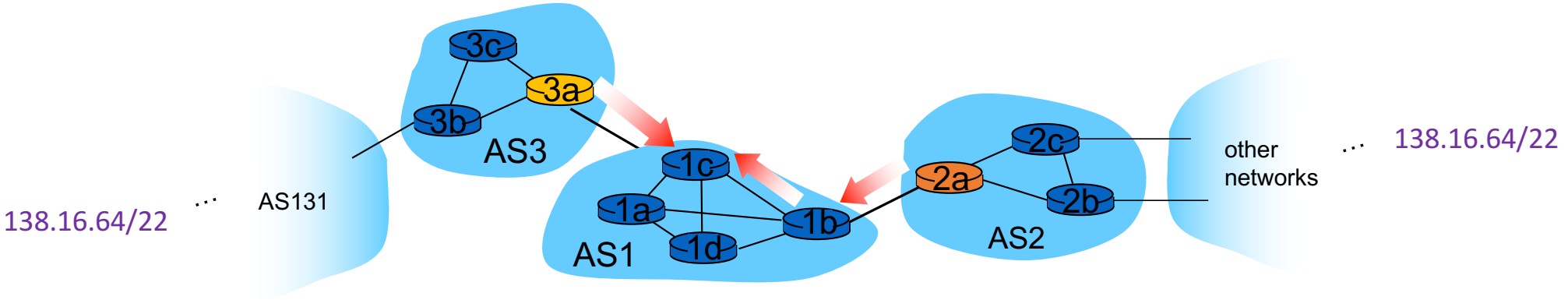
- BGP message contains "routes"
- "route" is a prefix and attributes: AS-PATH, NEXT-HOP,…
- Example
  - Prefix: 138.16.64/22
  - AS-PATH: AS3 AS15 …
  - NEXT-HOP: 201.44.13.125

Router may receive multiple routes



- Router may receive multiple routes for same prefix
- Which route to pick?
  1. local preference value attribute: policy decision
  2. shortest AS-PATH
  3. closest NEXT-HOP router: hot potato routing
  4. additional criteria

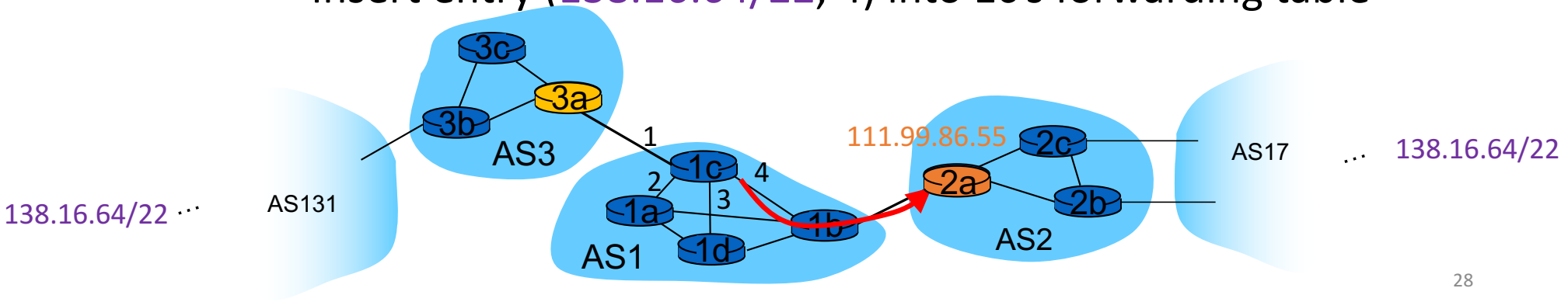
2. Shortest AS Path



- AS3 AS131 AS201 to 138.16.64/22
- AS2 AS17 to 138.16.64/22

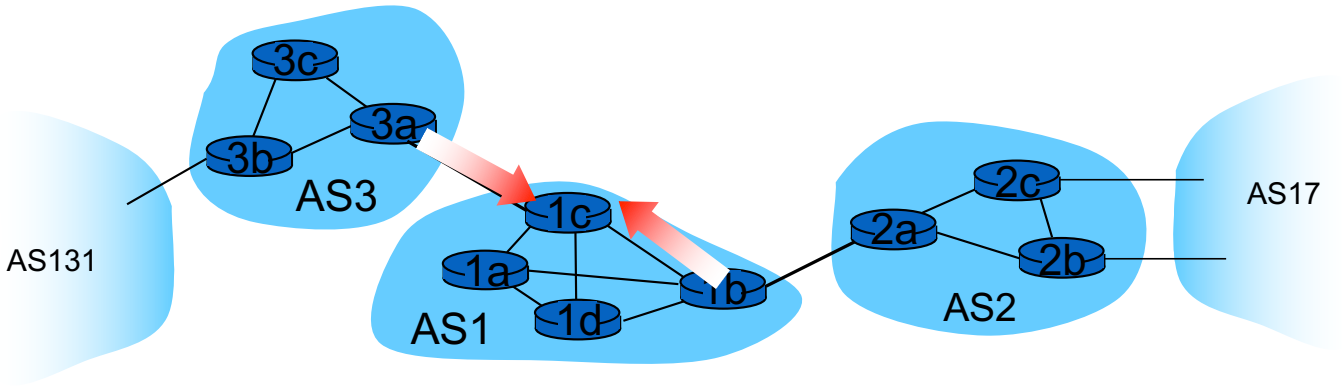
Use intra-domain routing

- Use selected route's NEXT-HOP attribute
  - NEXT-HOP = IP address of the router interface that begins the AS PATH
- Example:
  - AS-PATH: AS2 AS17 …; NEXT-HOP: 111.99.86.55
- Router uses OSPF to find shortest path from 1c to 111.99.86.55
- Insert entry (138.16.64/22, 4) into 1c's forwarding table

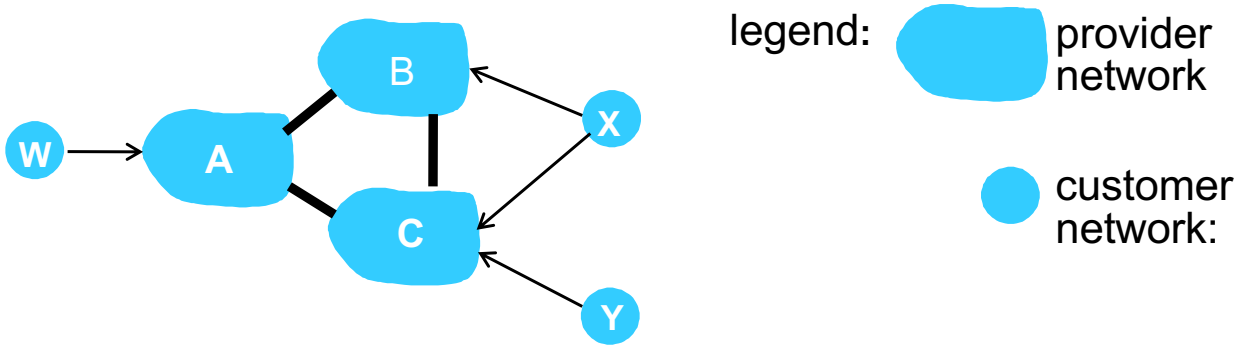


3. Closest NEXT-HOP Router

- Suppose there two or more best inter-routes.
- Then choose route with closest NEXT-HOP
  - Use OSPF to determine which gateway is closest
  - Q: From 1c, chose AS3 AS131 or AS2 AS17?
  - A: route AS3 AS131 since it is closer

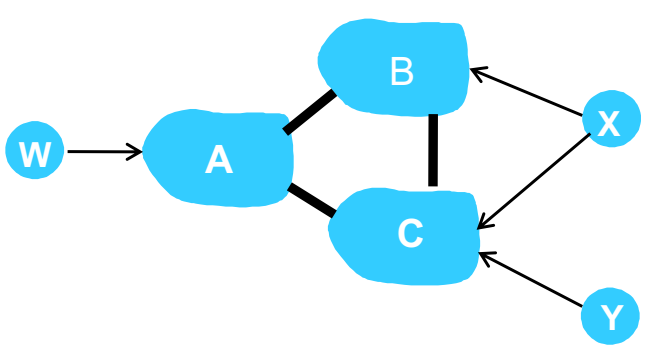


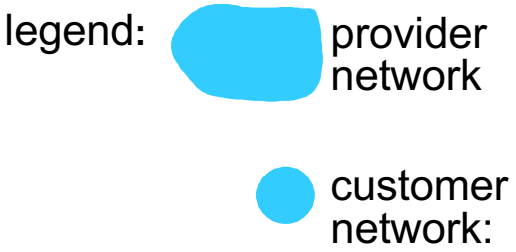
1. Policy decision



- A,B,C are *provider networks*
- X,W,Y are customer (of provider networks)
- X is *dual-homed:* attached to two networks

1. Policy decision





- A advertises path AW to B
- B advertises path BAW to X
- Q: Should B advertise path BAW to C?
  - No way! B gets no "revenue" for routing CBAW since neither W nor C are B's customers
  - B wants to force C to route to w via A
  - B wants to route *only* to/from its customers!

How does entry get in forwarding table?

Summary

1. Router becomes aware of prefix
   - via BGP route advertisements from other routers
2. Filter the route based on policy ($$$)
3. Determine router output port for prefix
   - Use BGP route selection to find best inter-AS route
   - Use OSPF to find best intra-AS route leading to best inter-AS route
   - Router identifies router port for that best route
4. Enter prefix-port entry in forwarding table

In Practice

- Suppose you want to map the Internet…

IP address	AS#
173.246.82.76/30	40191
14.163.30.128/30	45899
14.163.30.144/30	45899
197.178.128.0/17	33771
199.241.113.0/24	30047
24.72.183.0/24	47887
24.72.167.0/24	47887
62.24.127.0/24	12455
103.237.33.0/24	58558
103.254.169.0/24	59149
142.234.125.0/24	15003
194.135.25.0/24	2118
208.81.73.0/24	27621
66.193.109.0/24	4323
110.78.161.0/24	23456
131.60.200.0/23	391
23.122.48.0/23	7018
177.154.78.0/23	28349
69.32.246.0/23	27365
177.8.252.0/23	262707
197.42.86.0/23	8452

AS#	AS#	Relationship
4323	12122	p2p
23151	19406	unknown
1103	21345	unknown
12714	20562	p2p
9298	4808	c2p
3267	196695	unknown
24626	43265	p2c
37432	33763	unknown
3741	50683	unknown
11030	45896	unknown
54345	33660	c2p
28310	52768	p2c
209	62507	p2c
47541	12637	unknown
23624	4713	c2p
59467	42632	c2p
9505	7529	unknown
20505	25180	p2p
62480	33659	c2p
56730	41678	unknown
29636	34407	unknown

IP Address

- Princeton IP ranges
  - 128.112.0.0/16
  - 140.180.0.0/16
  - 204.153.48.0/23
  - 66.180.177.0/24
  - 192.12.53.0/24
- UCR IP ranges
  - 138.23.0.0/16
  - 192.31.146.0/24
  - 192.31.148.0/24
  - 192.35.223.0/24

Q: What is the difference between IP address blocks and AS#?

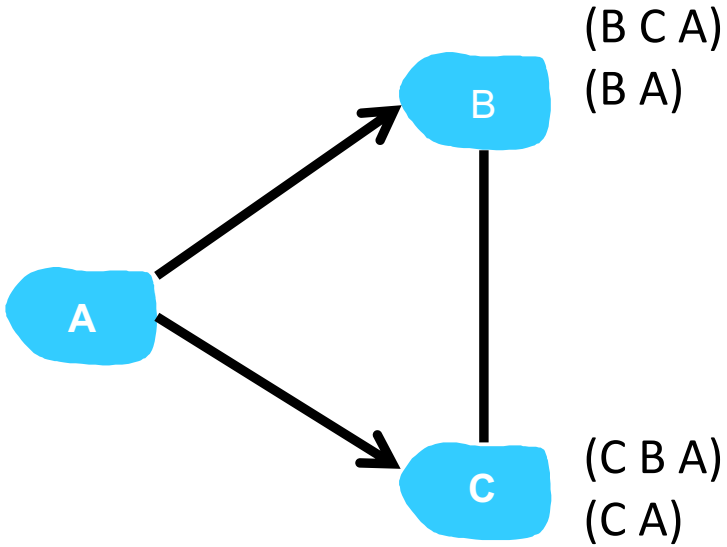
Overview

- AS relationships
- Inter-AS routing
  - BGP
  - Example
- Paper discussion

Q: How to “glue together” the “network of networks”?

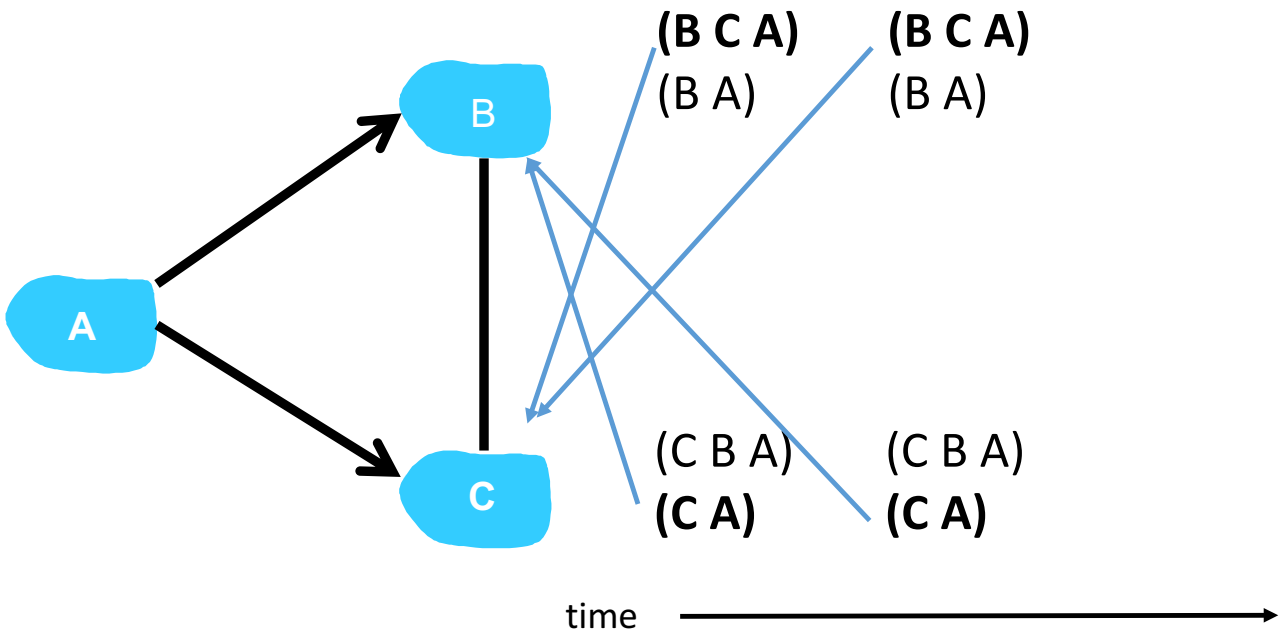
Stable BGP Routing

- Trying to get to destination A
- Routes listed in order of preference



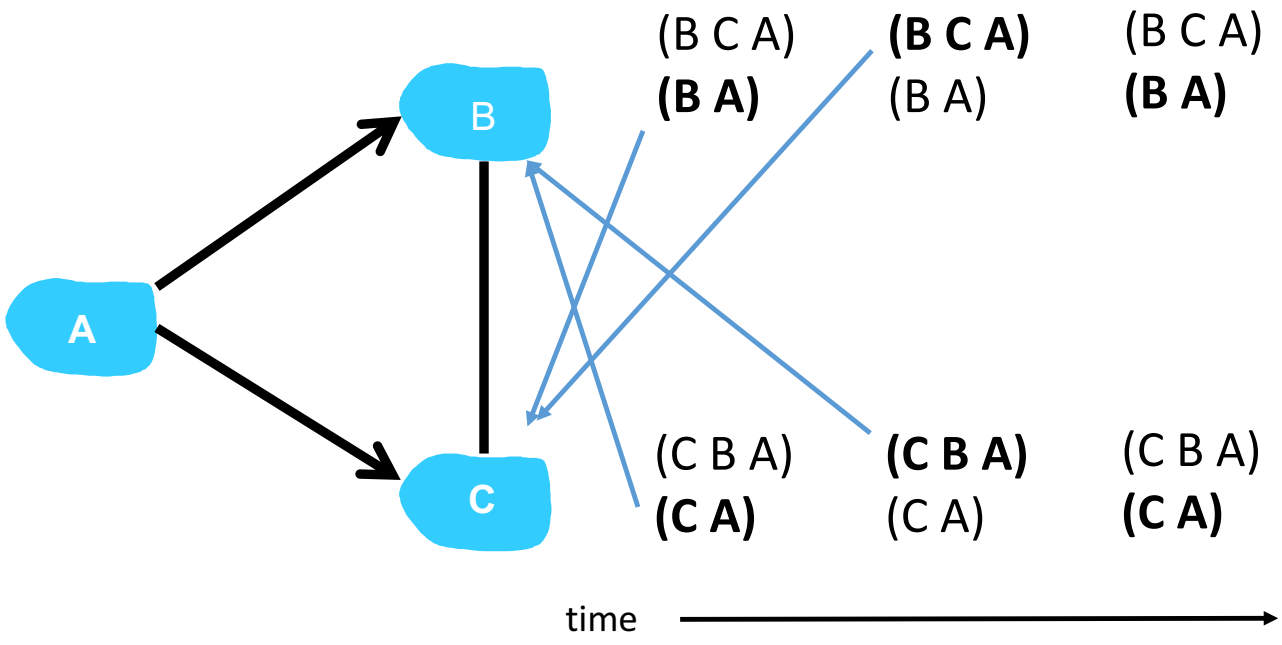
Stable BGP routing

- Suppose we start off with a certain initial configuration



Stable BGP routing

- Suppose we start off with the second choice options…



Paper Discussion

- What are implicit and explicit policies?
- What are some of the underlying assumptions of the model?
- Do you think it is feasible to have a centralized route registry?

Sources

- *Computer Networking: A Top-Down Approach*, Kurose & Ross
- Lixin Gao and Jennifer Rexford, “Stable Internet Routing Without Global Coordination,” *IEEE Trans. Networking*, 2001.