Network Layer:
Global Internet Routing (Interdomain, BGP); DHCP; Forwarding

Qiao Xiang, Congming Gao, **Qiang Su**

https://sngroup.org.cn/courses/cnns-xmuf25/index.shtml

11/27/2024

This deck of slides are heavily based on CPSC 433/533 at Yale University, by courtesy of Dr. Y. Richard Yang.

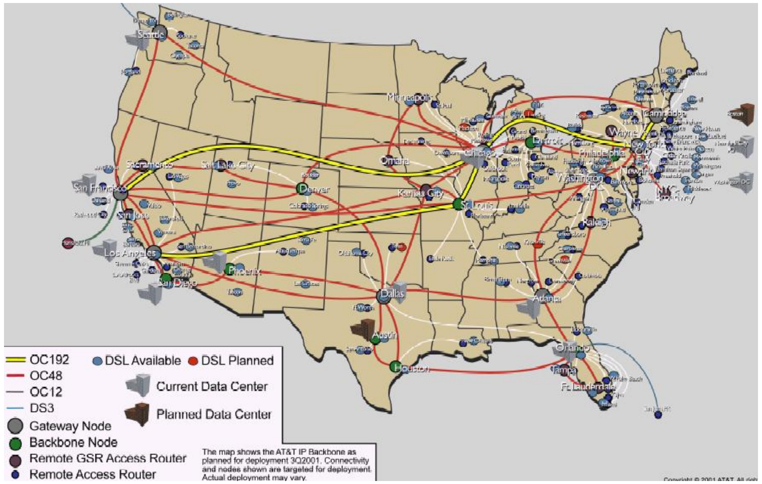
Outline

- Admin and recap
- Network control plane
  - Routing
    - Routing computation
      - Basic routing computation protocols
        - Distance vector protocols (distributed computing)
        - Link state protocols (distributed state synchronization)
      - Global Internet routing

Recap: Routing Context

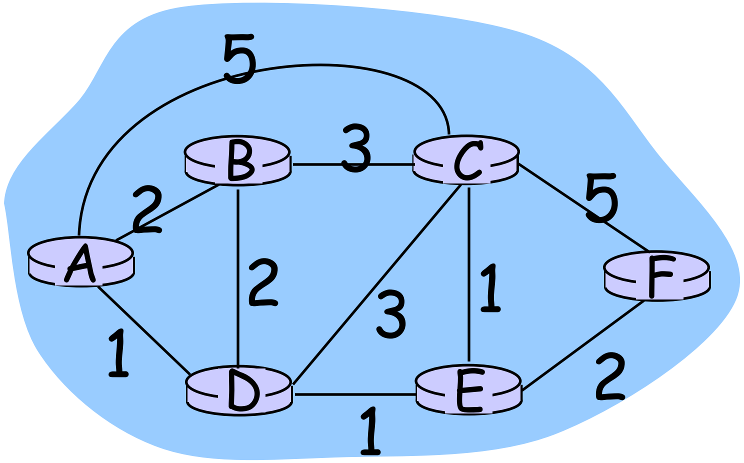
Routing

Goal: determine "good" paths (sequences of routers) thru networks from source to dest.



Often depends on a graph abstraction:

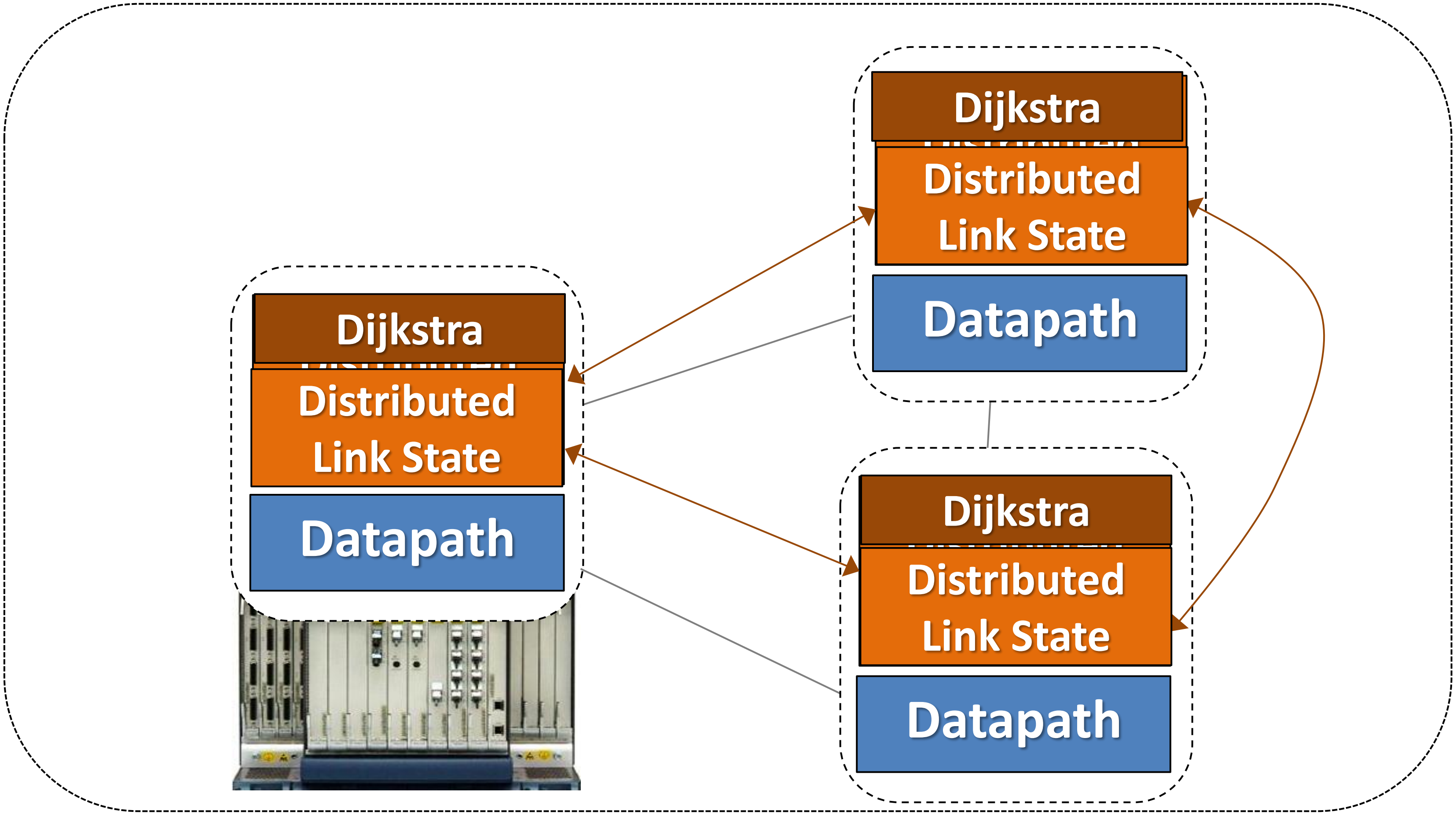
- graph nodes are routers
- graph edges are physical links
  - links have properties: delay, capacity, $ cost, policy



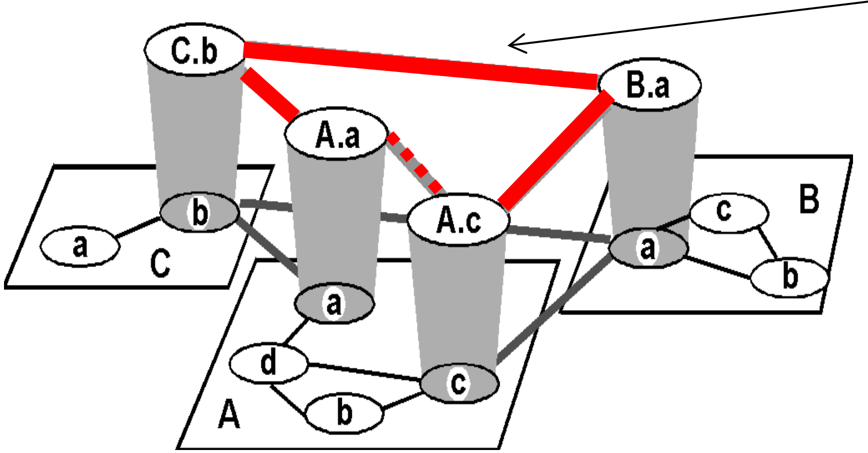
Recap: Link-State Routing

- Basic idea: Not distributed computing, only distributed state distribution
- Net topology, link costs are distributed to all nodes
  - all nodes have same info
  - Each node computes its shortest paths from itself to all other nodes
    - standard Dijkstra's algorithm as path compute alg
    - Allows multiple same-cost paths
    - Multiple cost metrics per link (for type of service routing)
- Most commonly used routing protocol (e.g., OSPF/ISIS) by most networks in Internet

Roadmap: Routing Computation Architecture Spectrum



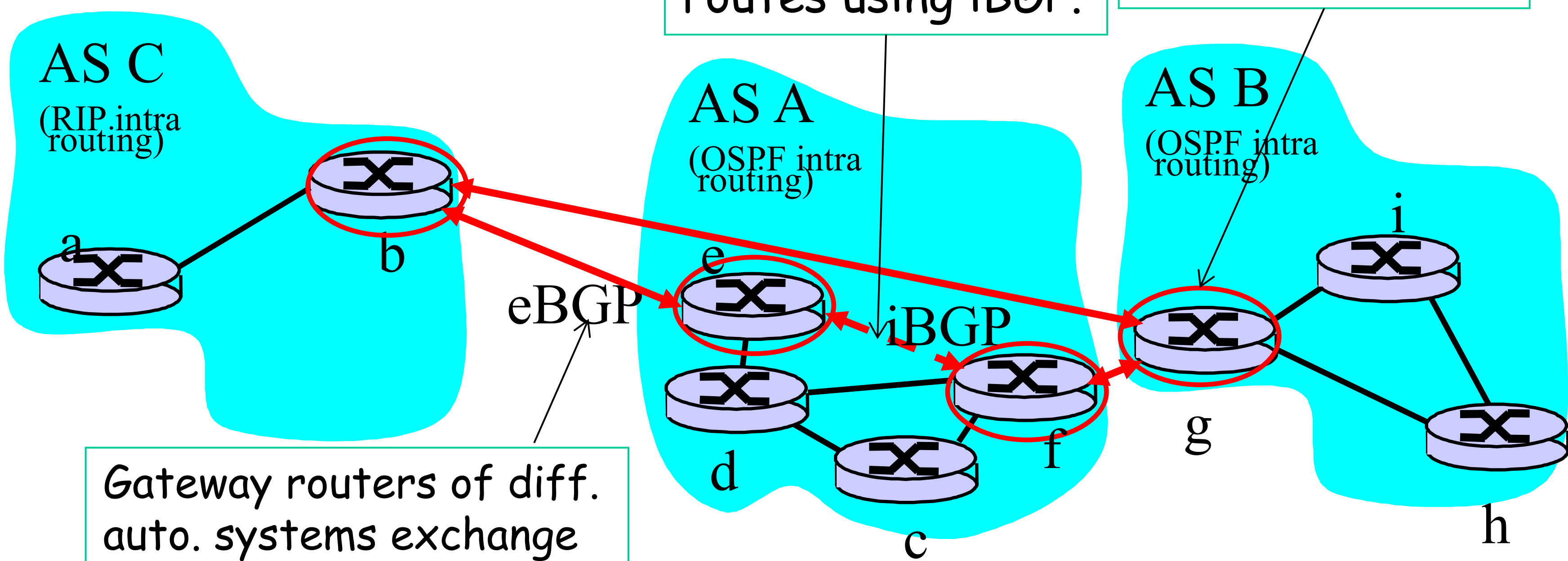
Recap: Routing with Autonomous Systems



Inter-AS routers form an overlay

Gateway routers participate in intradomain to learn internal routes.

Gateway routers of same AS share learned external routes using iBGP.



Gateway routers of diff. auto. systems exchange routes using eBGP

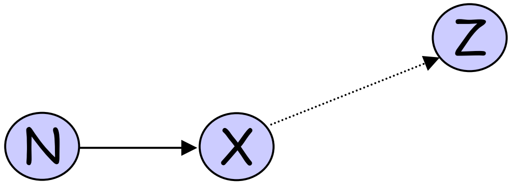
Recap: BGP Basic Operations

- BGP is **a Path Vector** protocol
  - similar to Distance Vector protocol
  - a border gateway sends to a neighbor *entire path* (i.e., *a sequence of ASNs*) to a destination, e.g.,
    - gateway X sends to neighbor N its path to dest. Z:

$$\text{path}(X,Z) = X, Y1, Y2, Y3, \ldots, Z$$

  - if N selects path(X, Z) advertised by X, then:

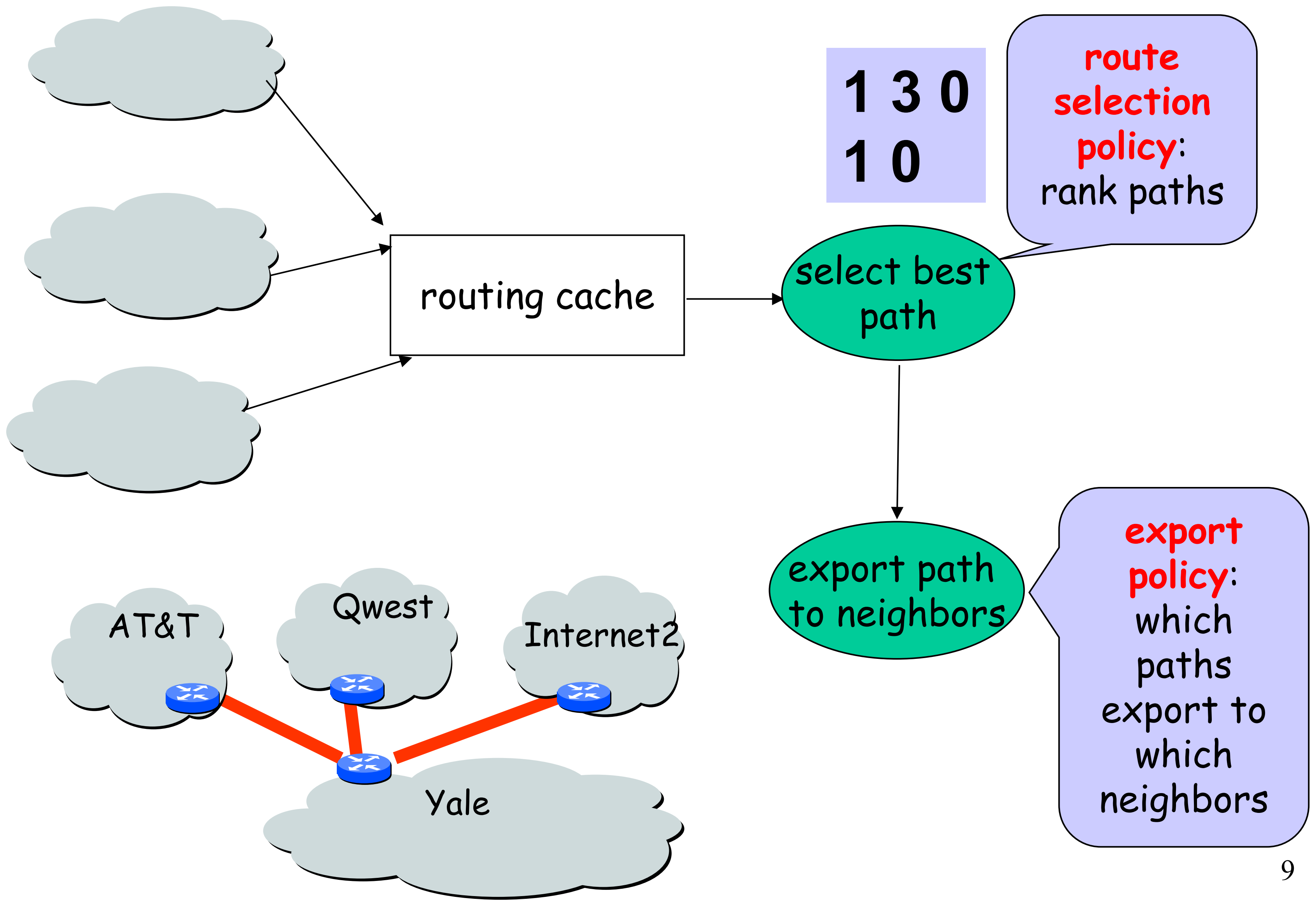
$$\text{path}(N,Z) = N, \text{path}(X,Z)$$



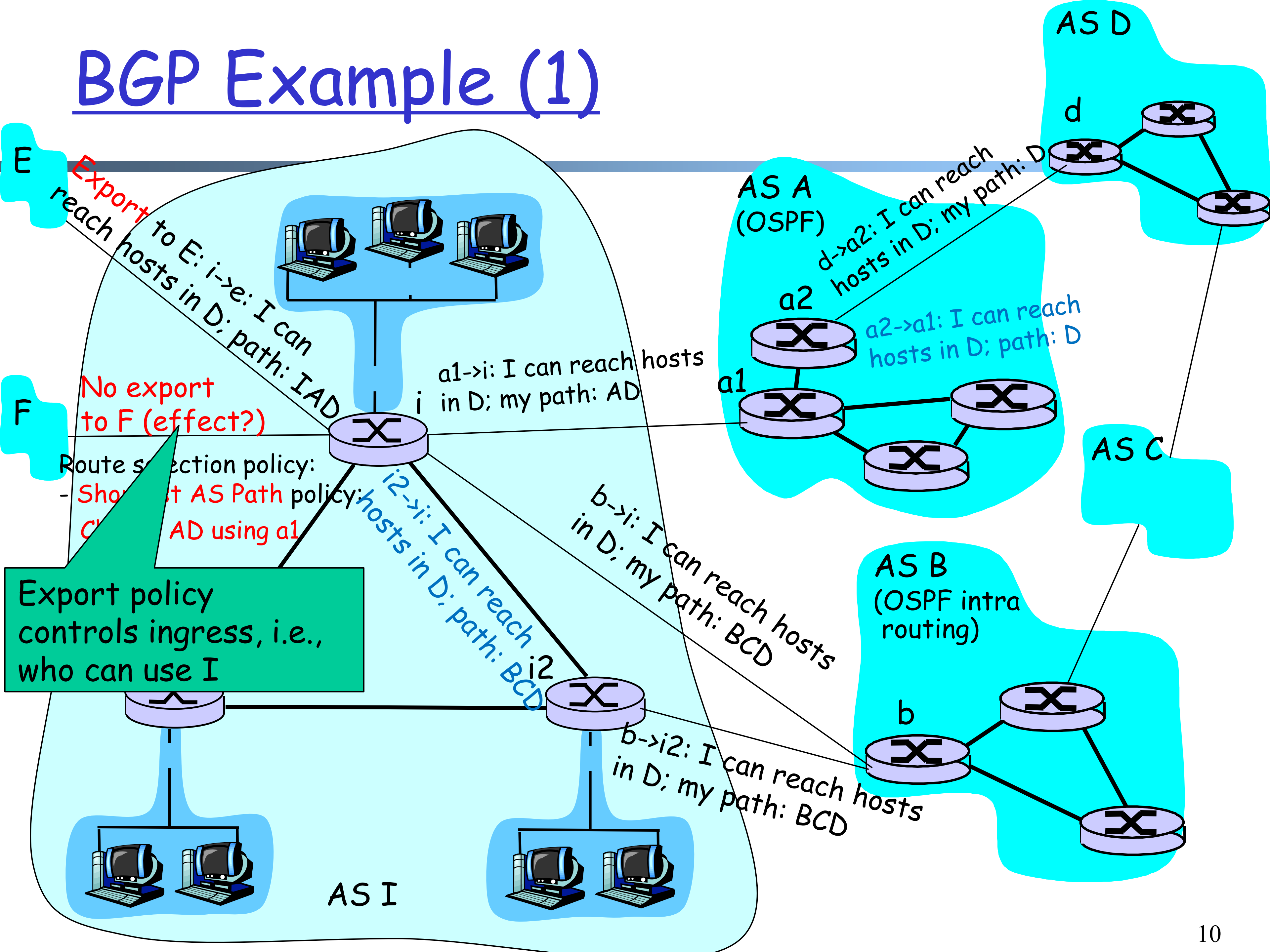
Outline

- Admin and recap
- Network control plane
  - Routing
    - Link weights assignment
    - Routing computation
      - Basic routing computation protocols
      - Global Internet routing
        - Basic architecture
        - *BGP (Border Gateway Protocol): The de facto Inter-domain routing standard*
          - Basic operations
          - *BGP as a policy routing framework (control interdomain routes)*

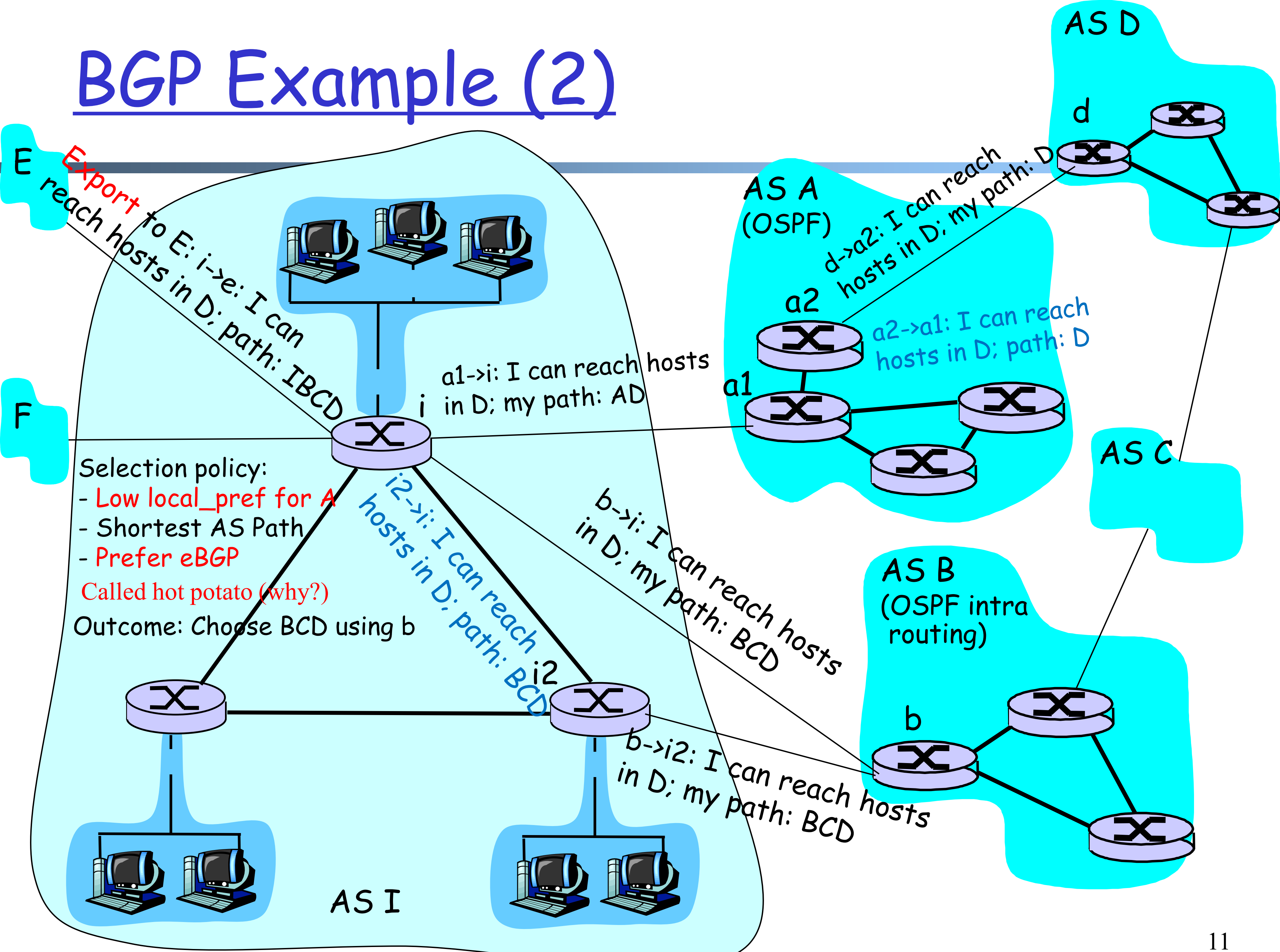
BGP Policy Routing Framework: Decision Components



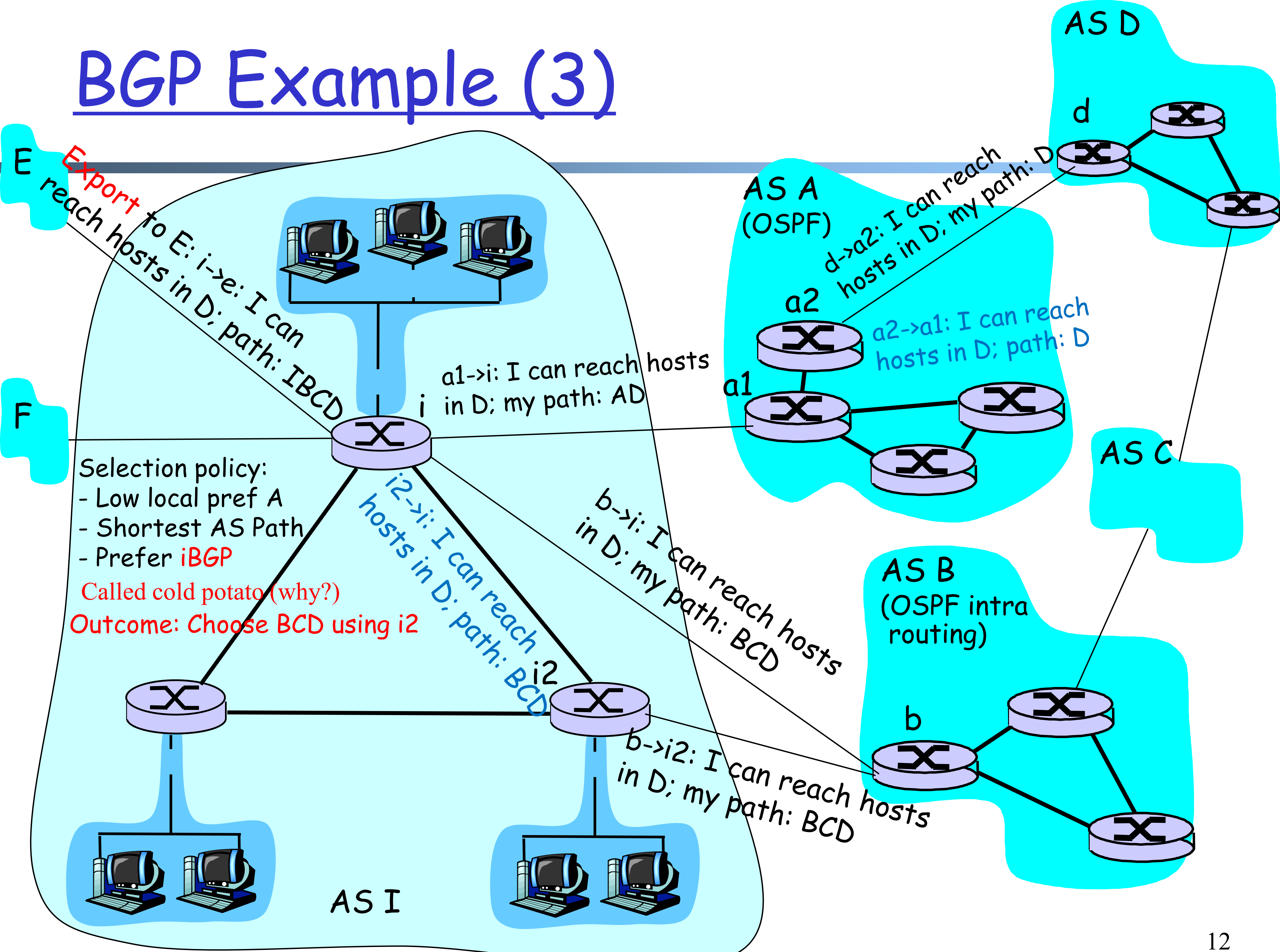
BGP Example (1)



BGP Example (2)



BGP Example (3)



Observing BGP Paths

- Using one of the looking glass servers:
  http://www.bgp4.as/looking-glasses

  https://www.gin.ntt.net/looking-glass/

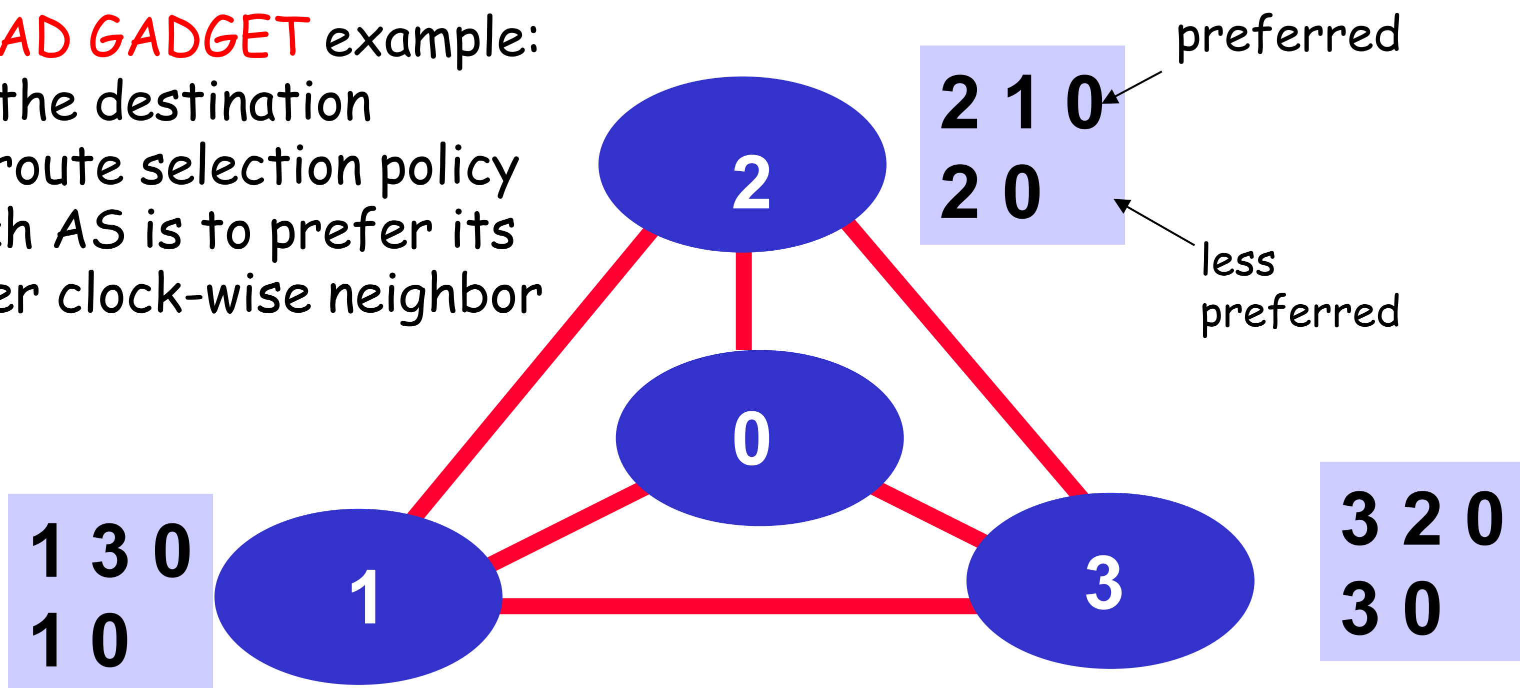
Outline

- Admin and recap
- Network control plane
  - Routing
    - Link weights assignment
    - Routing computation
      - Basic routing computation protocols
      - Global Internet routing
        - Basic architecture
        - *BGP (Border Gateway Protocol): The de facto Inter-domain routing standard*
          - Basic operations
          - BGP as a policy routing framework (control interdomain routes)
          - *Policy/interdomain routing analysis*

Motivation: Policy Routing Stability

- A policy routing system can be considered as a system to aggregate local preferences, but aggregation may not be always successful.

The BAD GADGET example:
- 0 is the destination
- the route selection policy of each AS is to prefer its counter clock-wise neighbor



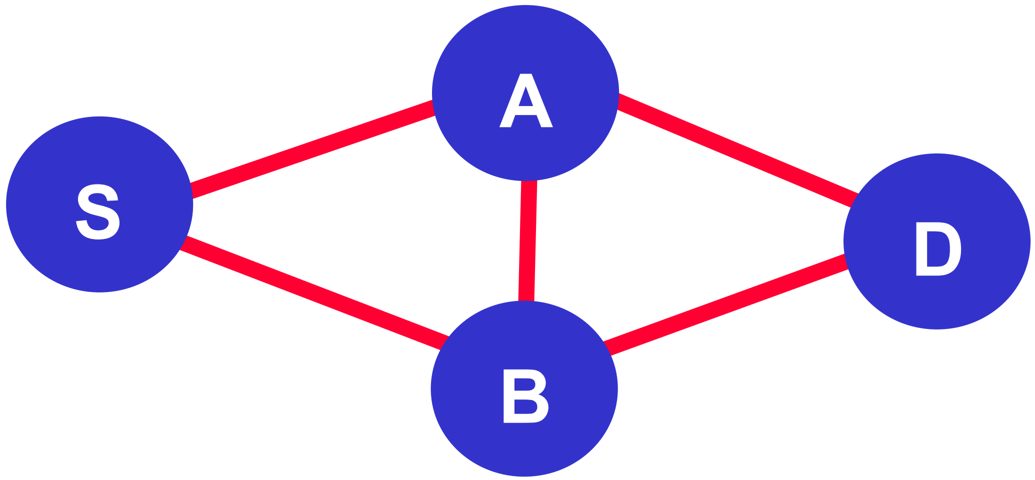
Policy (preferences) aggregation fails: routing instability !

General Framework of Preference Aggregation

- Also called Social Choice
  - Given individual preferences, define a framework to aggregate individual preferences:
    - A set of choices: a, b, c, …
    - A set of voters 1, 2, …
      - Each voter has a preference (ranking) of all choices, e.g.,
        - » voter 1: a > b > c
        - » voter 2: a > c > b
        - » voter 3: a > c > b
    - A well-specified aggregation rule (protocol) computes an aggregation of ranking, e.g.,
      - Society (network): a > b > c

For more details: see Semih Salihoglu (2007). Interdomain Routing as a Social Choice.

Example: Aggregation of Global Preference



- Choices (for S->D route): SAD, SBD, SABD, SBAD
- Voters S, A, B, D
- Each voter has a preference, e.g.,
  - S: SAD > SBD > SABD > SBAD
  - ...

Arrow's Impossibility Theorem

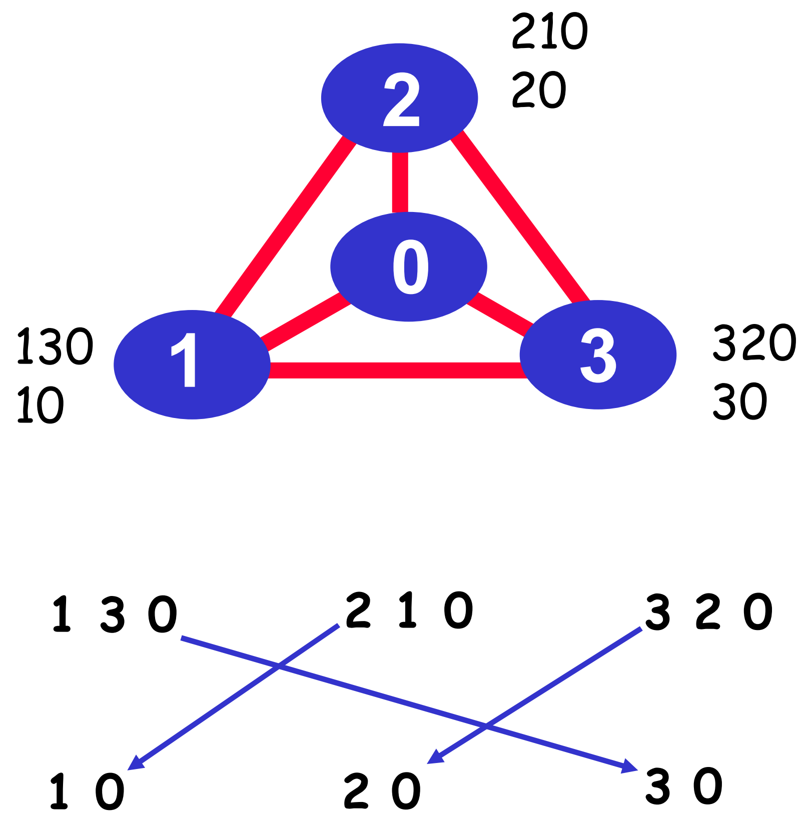
- Axioms:
  - Transitivity
    - if a > b & b > c, then a > c
  - Unanimity:
    - If all participants prefer a over b (a > b) => a > b
  - Independence of irrelevant alternatives (IIA)
    - Social ranking of a and b depends only on the relative ranking of a and b among all participants
- Result:
  - Arrow's Theorem: Any constitution that respects transitivity, unanimity and IIA is a dictatorship.

Outline

- Admin and recap
- Network control plane
  - Routing
    - Link weights assignment
    - Routing computation
      - Basic routing computation protocols
      - Global Internet routing
        - Basic architecture
        - BGP (Border Gateway Protocol): The de facto Inter-domain routing standard
          - Basic operations
          - BGP as a policy routing framework (control interdomain routes)
          - Policy/interdomain routing analysis
            - Global preference aggregation and Arrow's Theorem [Optional]
            - *Local preference aggregation*

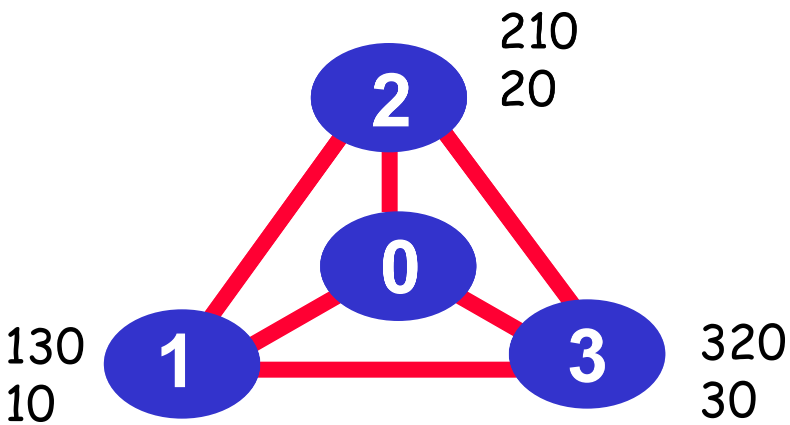
BGP w/ Local Preference

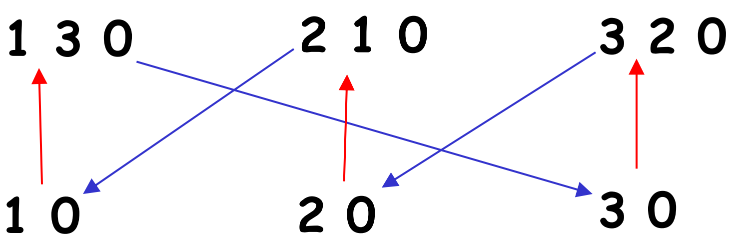
- BGP preferences are typically local (only on paths start from itself)
- Hence the preferences have dependency (priority)
  - The "closer" a node to the destination, the more "powerful" it may be



Complete Dependency: P-Graph

- Complete dependency can be captured by a structure called P-graph
- Nodes in P-graph are feasible paths
- Edges represent priority (low to high)
  - A directed edge from path N_1P_1 to P_1
    - intuition: to let N_1 choose N_1P_1, P_1 must be chosen and exported to N_1
  - A directed edge from a lower ranked path to a higher ranked path
    - intuition: the higher ranked path should be considered first

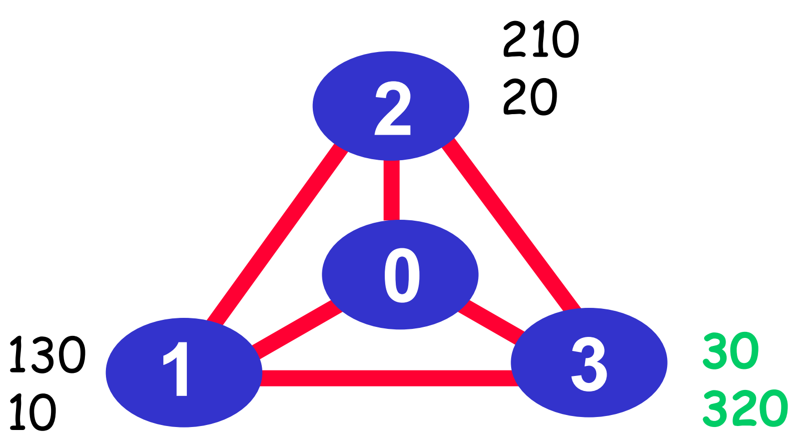


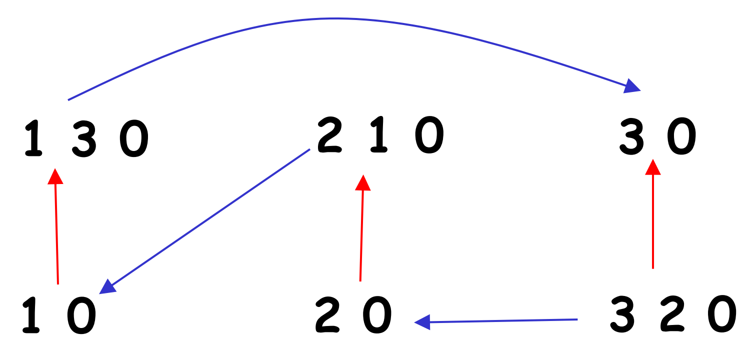


Any observation on the P-graph?

P-Graph and BGP Convergence

- If the P-graph of the networks has no loop, then policy routing converges.
  - intuition: choose the path node from the partial order graph with no out-going edge to non-fixed path nodes, fix the path node, eliminate all no longer feasible; continue
- Example: suppose we swap the order of 30 and 320



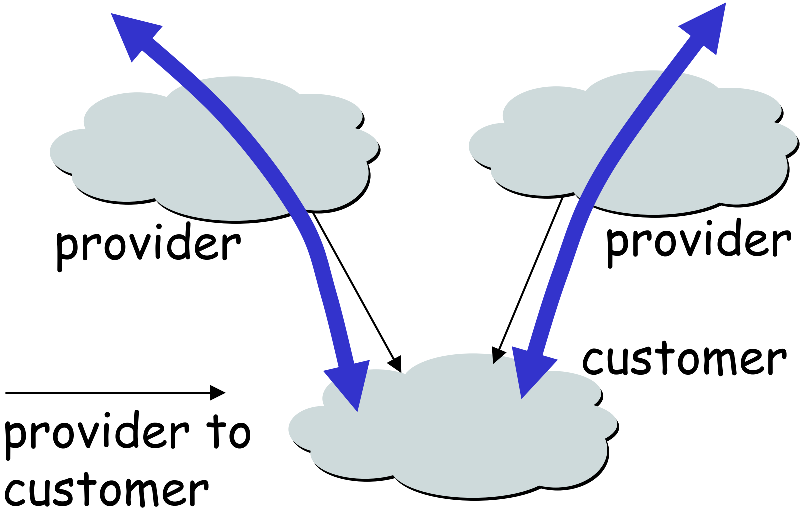


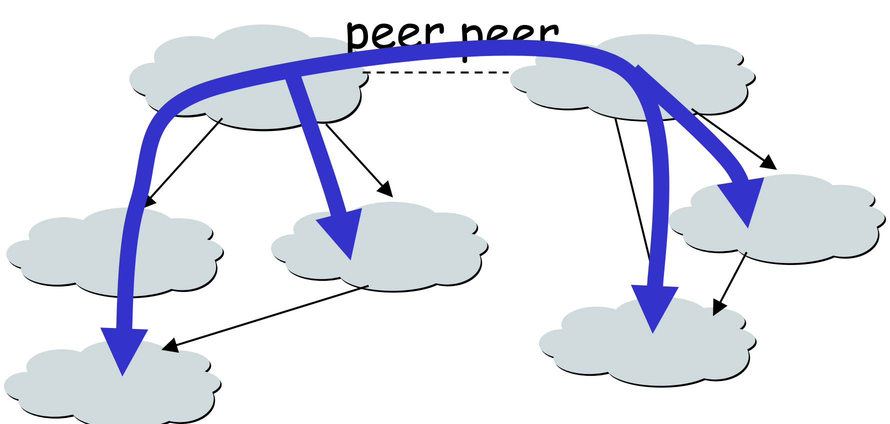
Outline

- Admin and recap
- Network control plane
  - Routing
    - Link weights assignment
    - Routing computation
      - Basic routing computation protocols
      - Global Internet routing
        - Basic architecture
        - BGP (Border Gateway Protocol): The de facto Inter-domain routing standard
          - Basic operations
          - BGP as a policy routing framework (control interdomain routes)
          - Policy/interdomain routing analysis
            - Global preference aggregation and Arrow's Theorem
            - Local preference aggregation
            - ***Economics and interdomain routing patterns***

Internet Economy: Two Types of Business Relationship

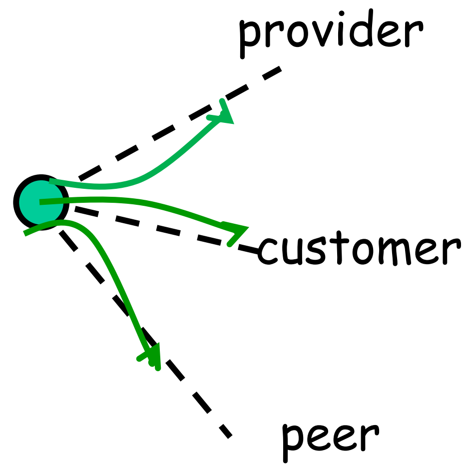
- *Customer provider relationship*
  - a provider is an AS that connects the customer to the rest of the Internet
  - customer pays the provider for the transit service
  - e.g., XMU is a customer of CERNET and China Telecom
- *Peer-to-peer relationship*
  - mutually agree to exchange traffic between their respective **customers** only
  - there is no payment between peers





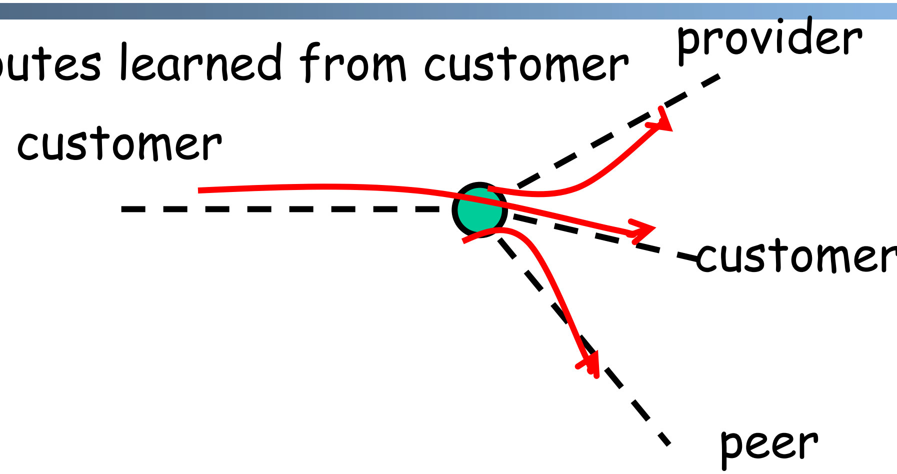
Route Selection Policies and Economics

- Route selection (ranking) policy:
  - the typical route selection policy is to prefer customers over peers/providers to reach a destination, i.e., Customer > pEer/Provider



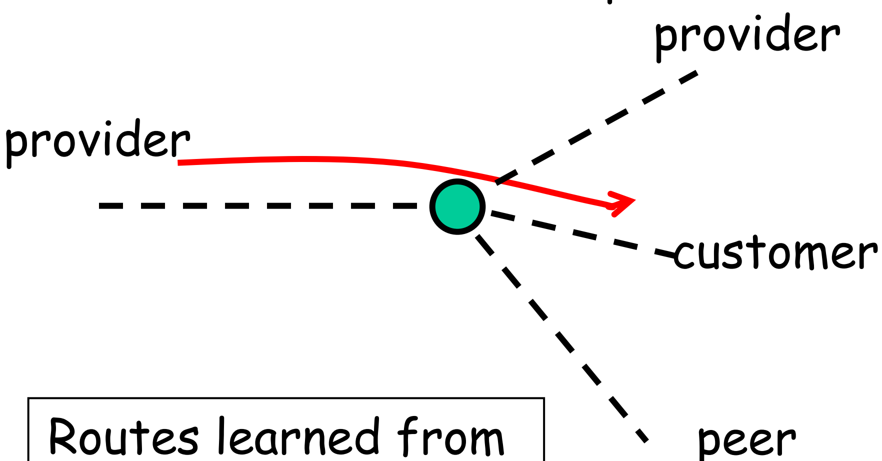
Export Policies and Economics

case 1: routes learned from customer



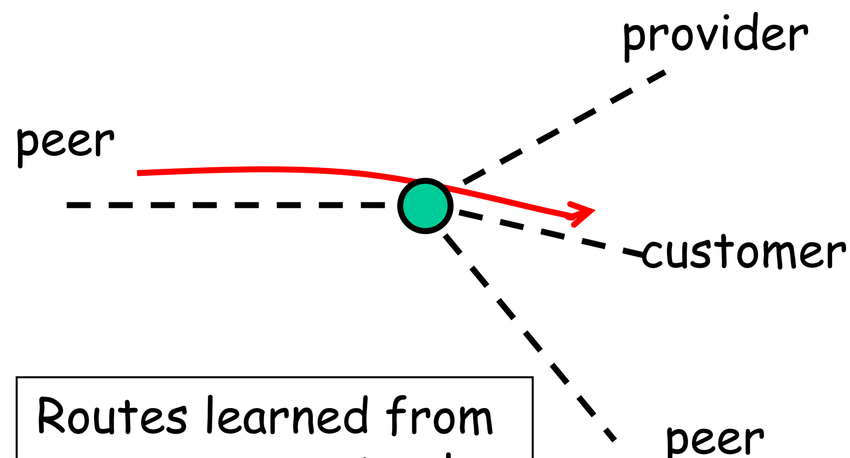
Routes learned from a customer are sent to all other neighbors

case 2: routes learned from provider



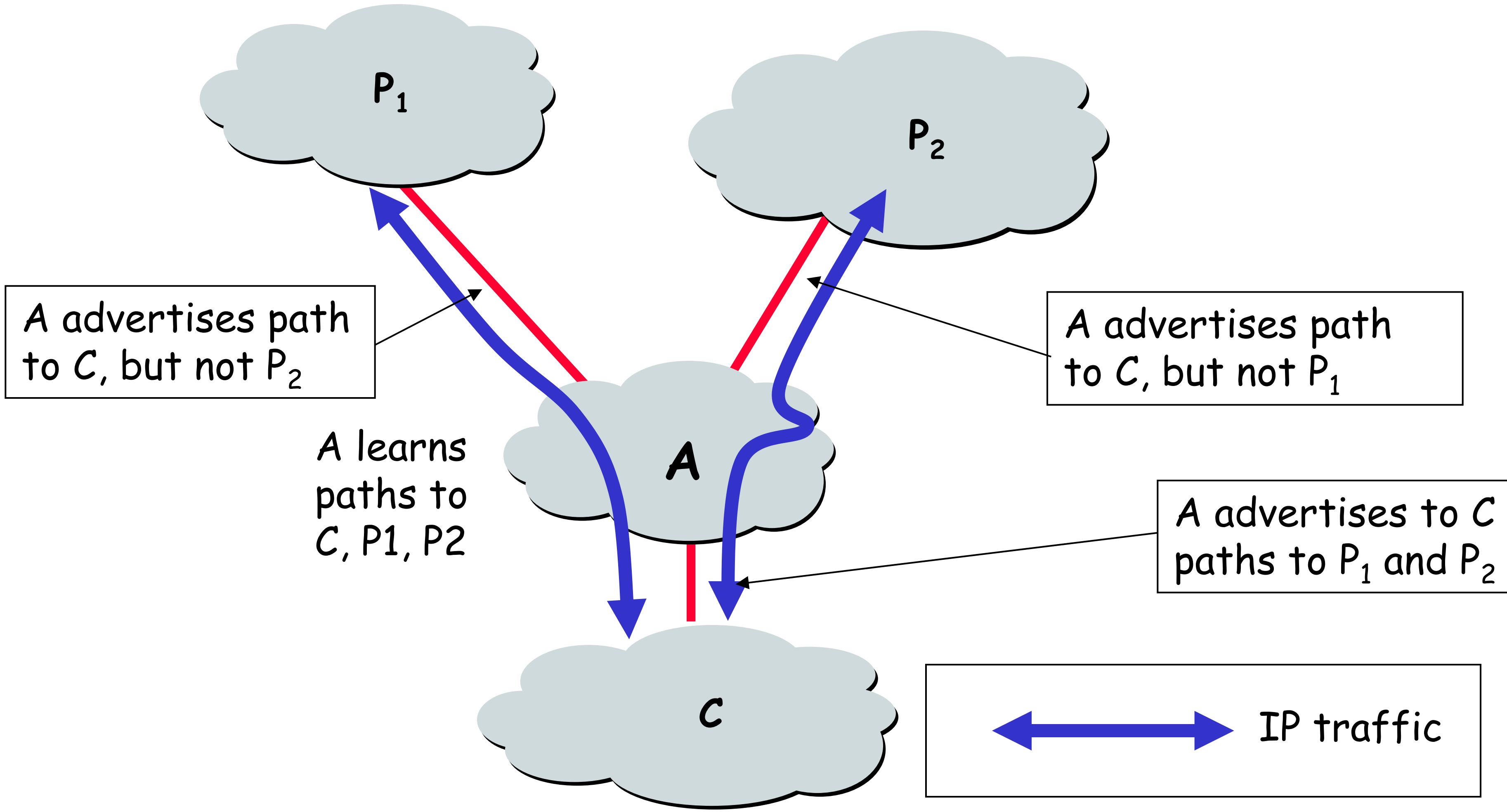
Routes learned from a provider are sent only to customers

case 3: routes learned from peer



Routes learned from a peer are sent only to customers

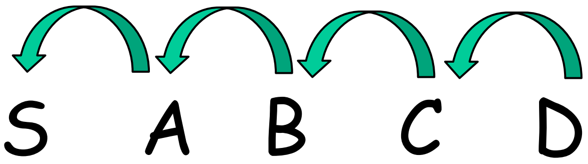
Example: Typical Export -> No-Valley Routing



Suppose P_1 and P_2 are providers of A; A is a provider of C

Typical Export Policies Route Patterns

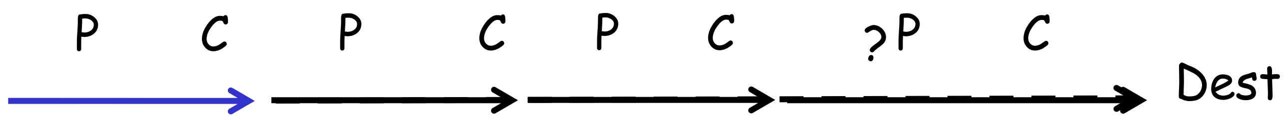
- Assume a BGP path SABCD to destination AS D. Consider the business relationship between each pair:



- Three types of business relationships:
  - PC (provider-customer)
  - CP (customer-provider)
  - PP (peer-peer)

Typical Export Policies Route Patterns

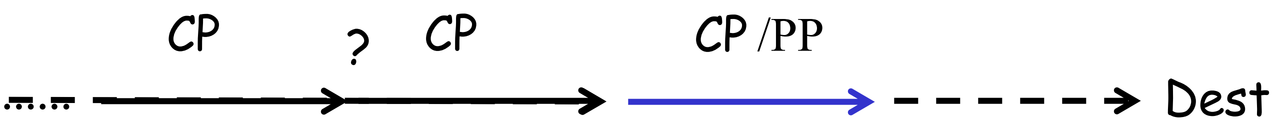
- Invariant 1 of valid BGP routes (with labels representing business relationship)



Reasoning: only route learned from customer is sent to provider; thus after a PC, it is always PC to the destination

Typical Export Policies Route Patterns

- Invariant 2 of valid BGP routes (with labels representing business relationship)



Reasoning: routes learned from peer or provider are sent to only customers; thus all relationship before is CP.

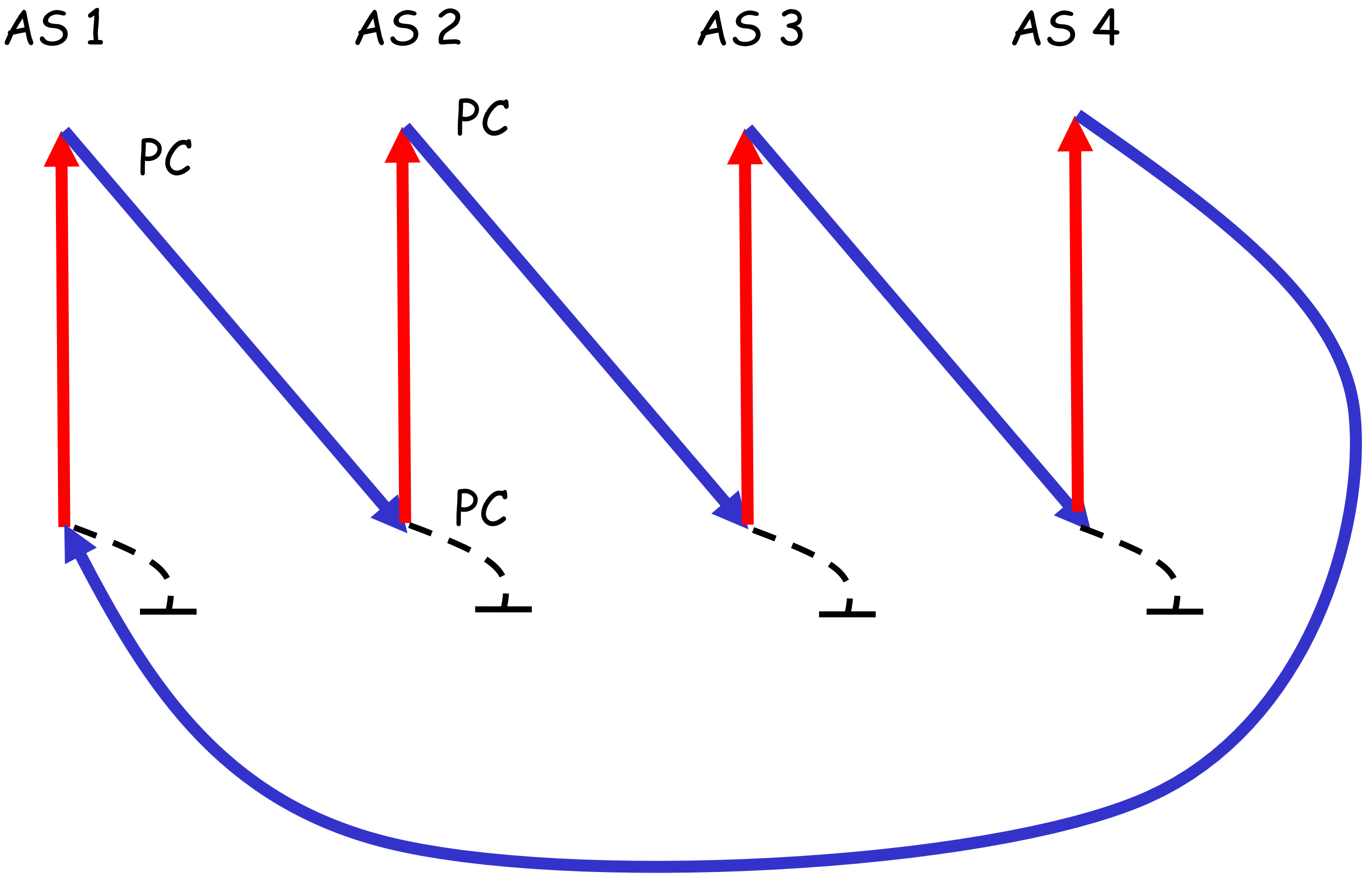
Stability of BGP Policy Routing

- Suppose
  1. there is no loop formed by provider-customer relationship in the Internet
  2. each AS uses typical route selection policy:
     C > E/P
  3. each AS uses the typical export policies

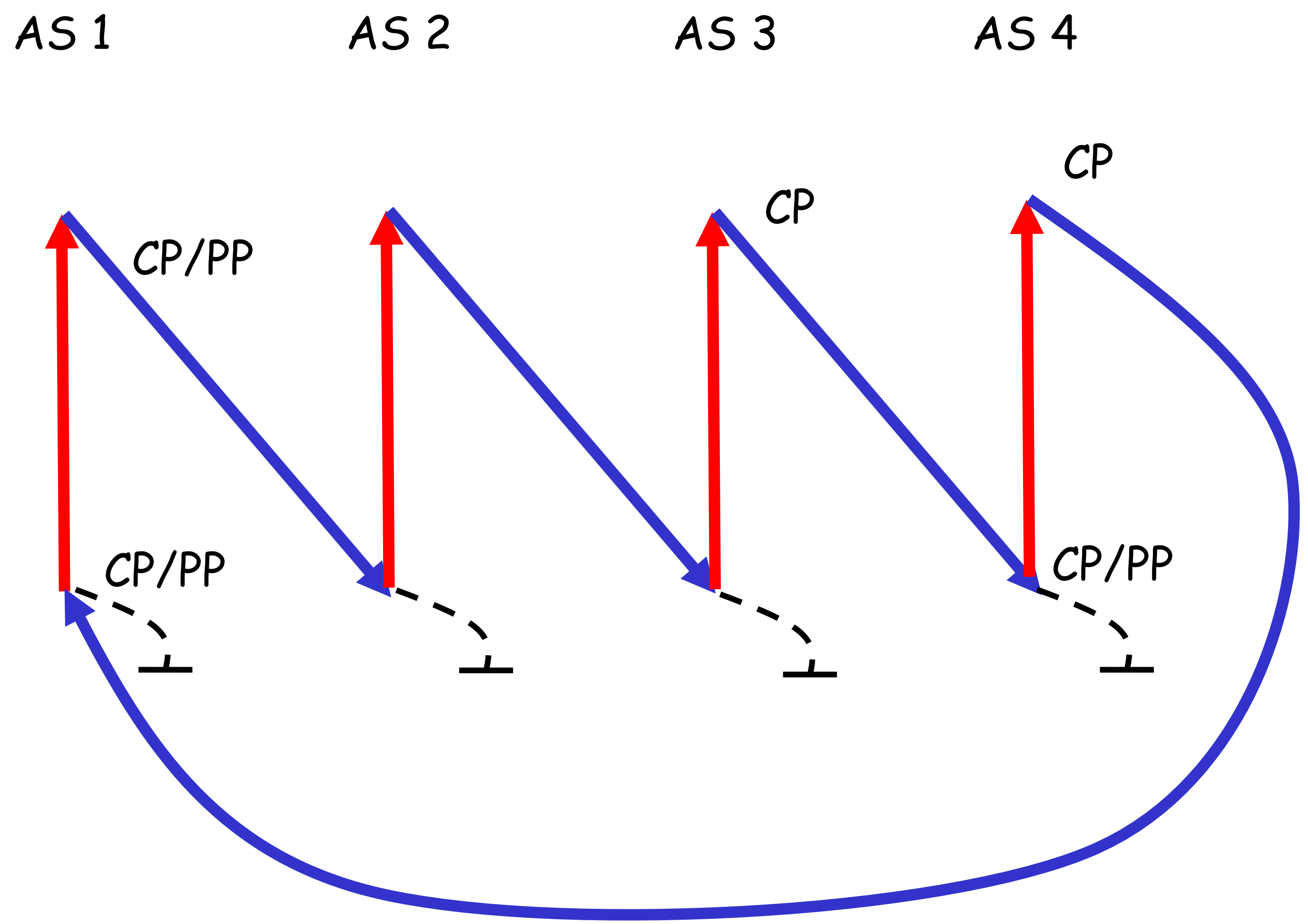
- Then policy routing always converges (i.e., is stable).

Case 1: A Link is PC

Proof by contradiction. Assume a loop in P-graph. Consider a fixed link. in the loop



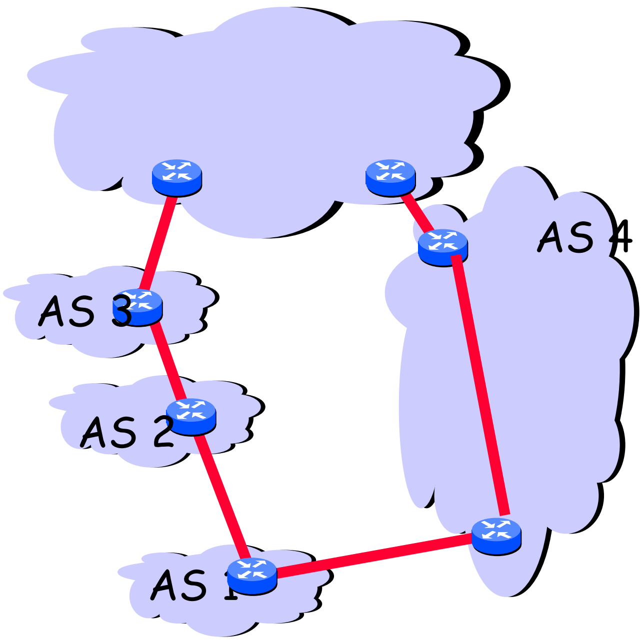
Case 2: Link is CP/PP



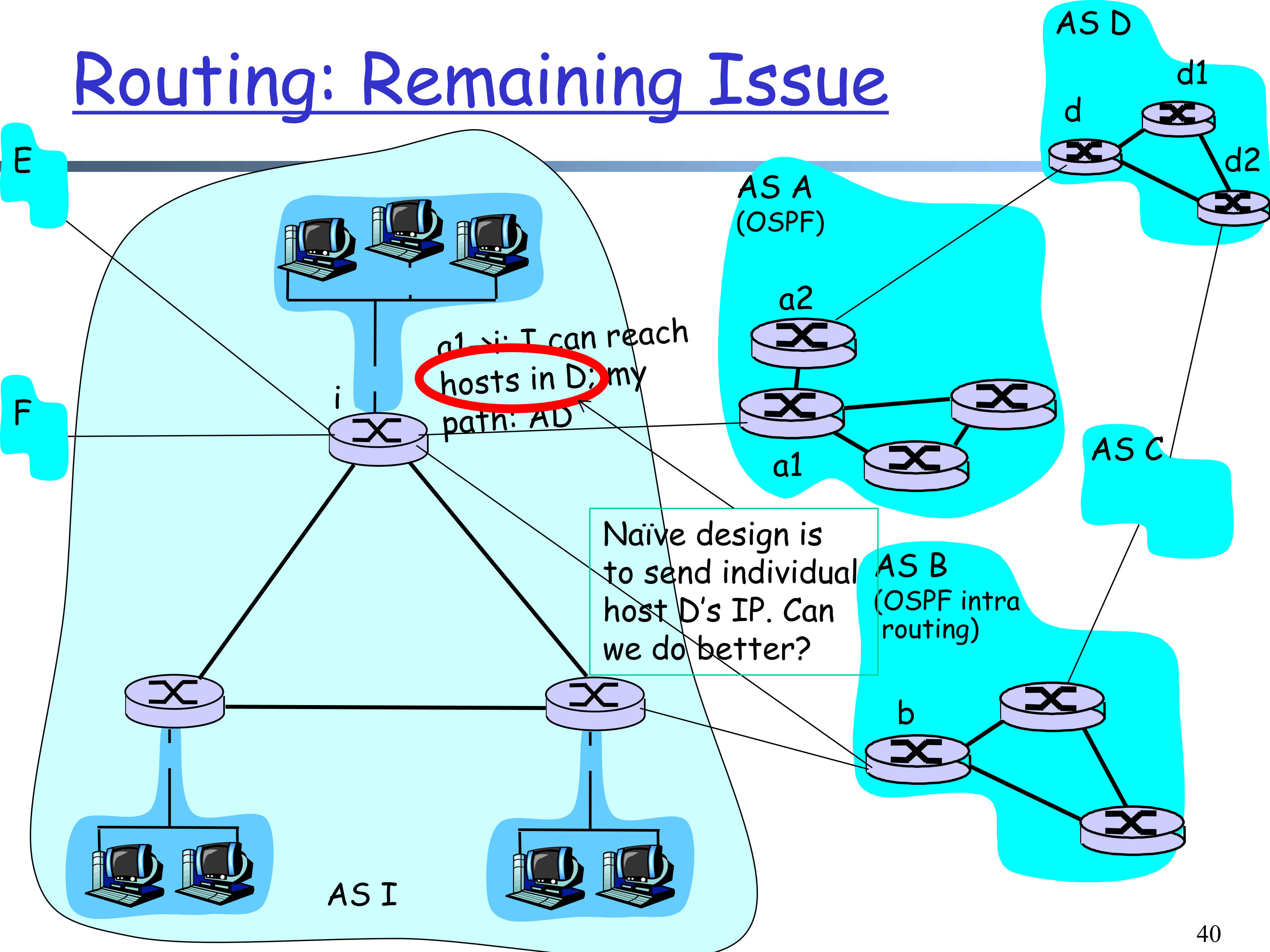
Summary: BGP Policy Routing

- Advantage
  - satisfies current demand

- Issues
  - policy dispute can lead to instability
    - current Internet economy provides a stability framework, but if the framework changes, we may see instability
  - Hierarchical routing can be inefficient



Routing: Remaining Issue



Outline

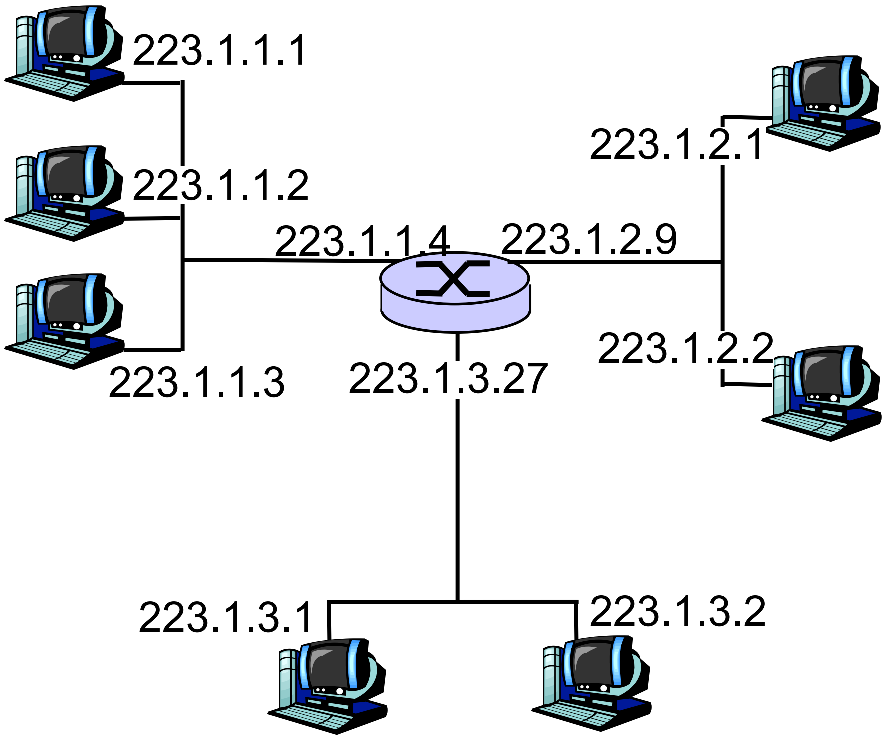
- Admin and recap
- Network control plane
  - Routing
    - Link weights assignment
    - Routing computation
      - Basic routing computation protocols
      - Global Internet routing
        - Basic architecture
        - BGP (Border Gateway Protocol): The de facto Inter-domain routing standard
          - Basic operations
          - BGP as a policy routing framework (control interdomain routes)
          - Policy/interdomain routing analysis
            - Global preference aggregation and Arrow's Theorem
            - Local preference aggregation
            - Economics and interdomain routing patterns
          - ***IP addresses for Interdomain routing***

IP Addressing Scheme: Requirements

- Uniqueness: We need an address to uniquely identify each destination
- Aggregability : Routing scalability needs flexibility in aggregation of destination addresses
  - we want to aggregate as a large set of destinations as possible in BGP announcements
- Current: the unit of routing in the Internet is a classless interdomain routing (CIDR) address

IP Address: Uniqueness

- **IPv4 address:** A 32-bit unique identifier for an *interface*
- *interface:*
  - routers typically have multiple interfaces
  - host may have multiple interfaces




```
%/sbin/ifconfig -a
```

223.1.3.2 = 11011111 00000001 00000011 00000010

223	1	3	2

e.g., /etc/sysconfig/network-scripts/ifcfg-enp0s25
%ifup

Classless InterDomain Routing (CIDR) Address: Aggregation

- A CIDR address partitions an IP address into two parts
  - A prefix representing the network portion, and the rest (host part)
  - address format: a.b.c.d/x, where x is # bits in network portion of address

network part | host part

11001000 00010111 00010000 00000000

200.23.16.0/23

Some systems use mask (1's to indicate network bits), instead of the /x format

CIDR Aggregation in BGP