University of Glasgow

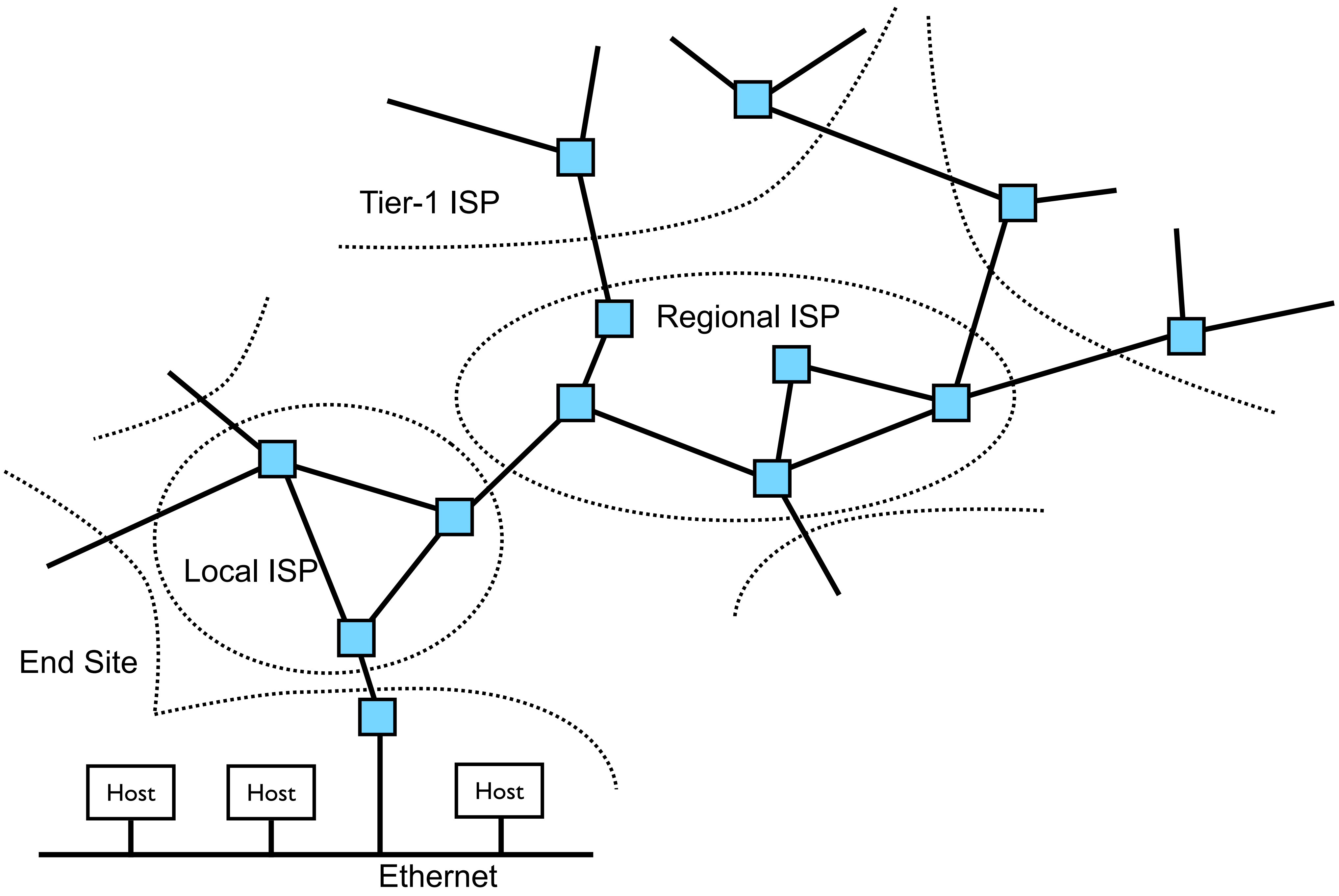
Network Layer (4): Interdomain Routing

Networked Systems 3
Lecture 11

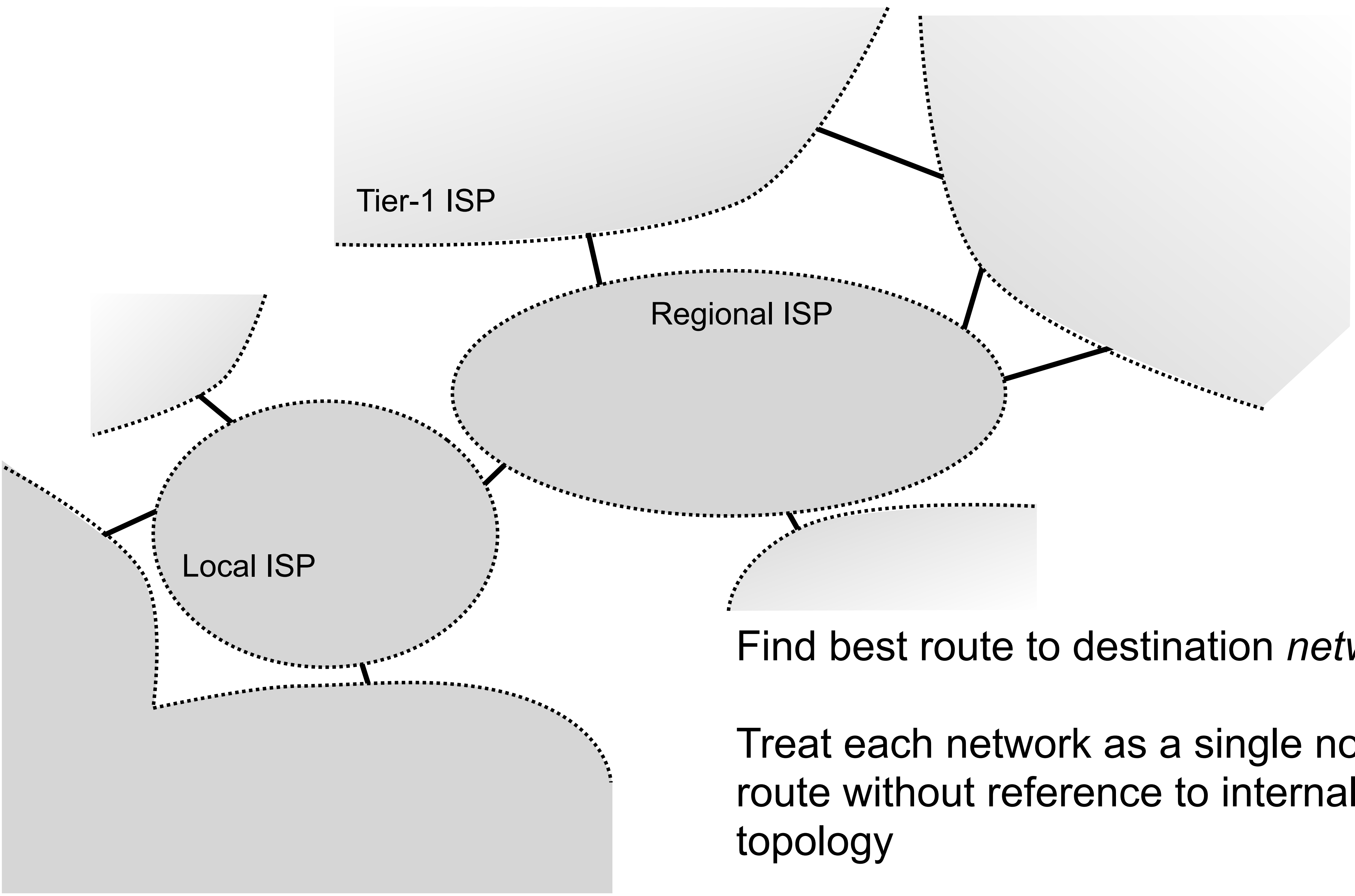
Lecture Outline

- Interdomain routing
  - Autonomous systems and the Internet AS-level topology
  - BGP and Internet routing

Interdomain Unicast Routing



Interdomain Unicast Routing

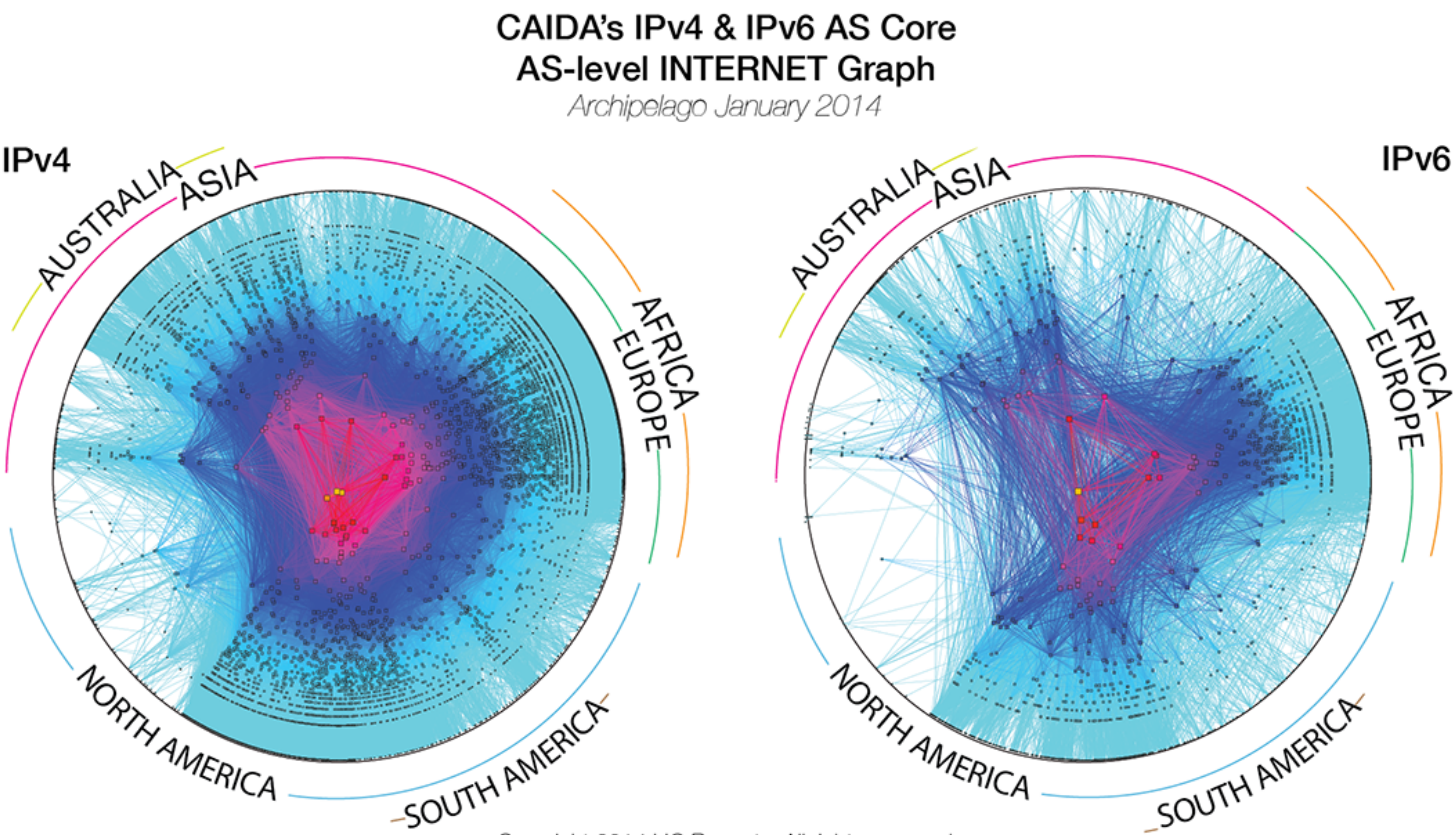


Find best route to destination *network*

Treat each network as a single node, and route without reference to internal network topology

Autonomous Systems

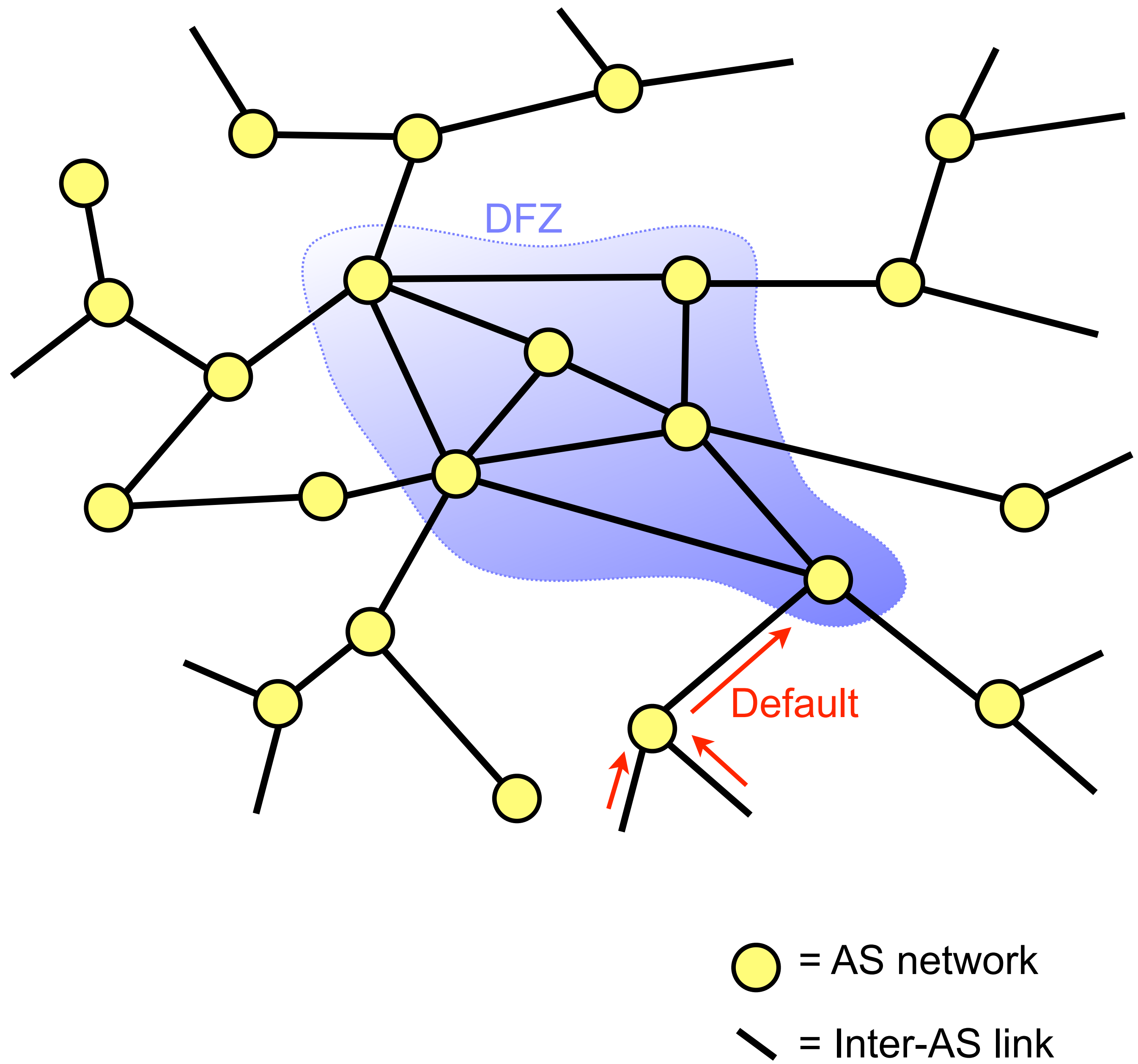
- Network comprised of autonomous systems (ASes)
  - Each AS is an independently administered network
    - An Internet service provider, or other organisation, that operates a network and wants to participate in interdomain routing
    - Some organisations operate more than one AS
      - For ease of administration; due to company mergers; etc.
    - Each AS is identified by a unique number, allocated by the RIR
      - ~50,000 AS numbers allocated: http://bgp.potaroo.net/cidr/autnums.html (Jan 2015)
- Routing problem is finding best AS-level path from source AS to destination AS
  - Treat each AS as a node on the routing graph (the "AS topology graph")
  - Treat connections between ASes as edges in the graph



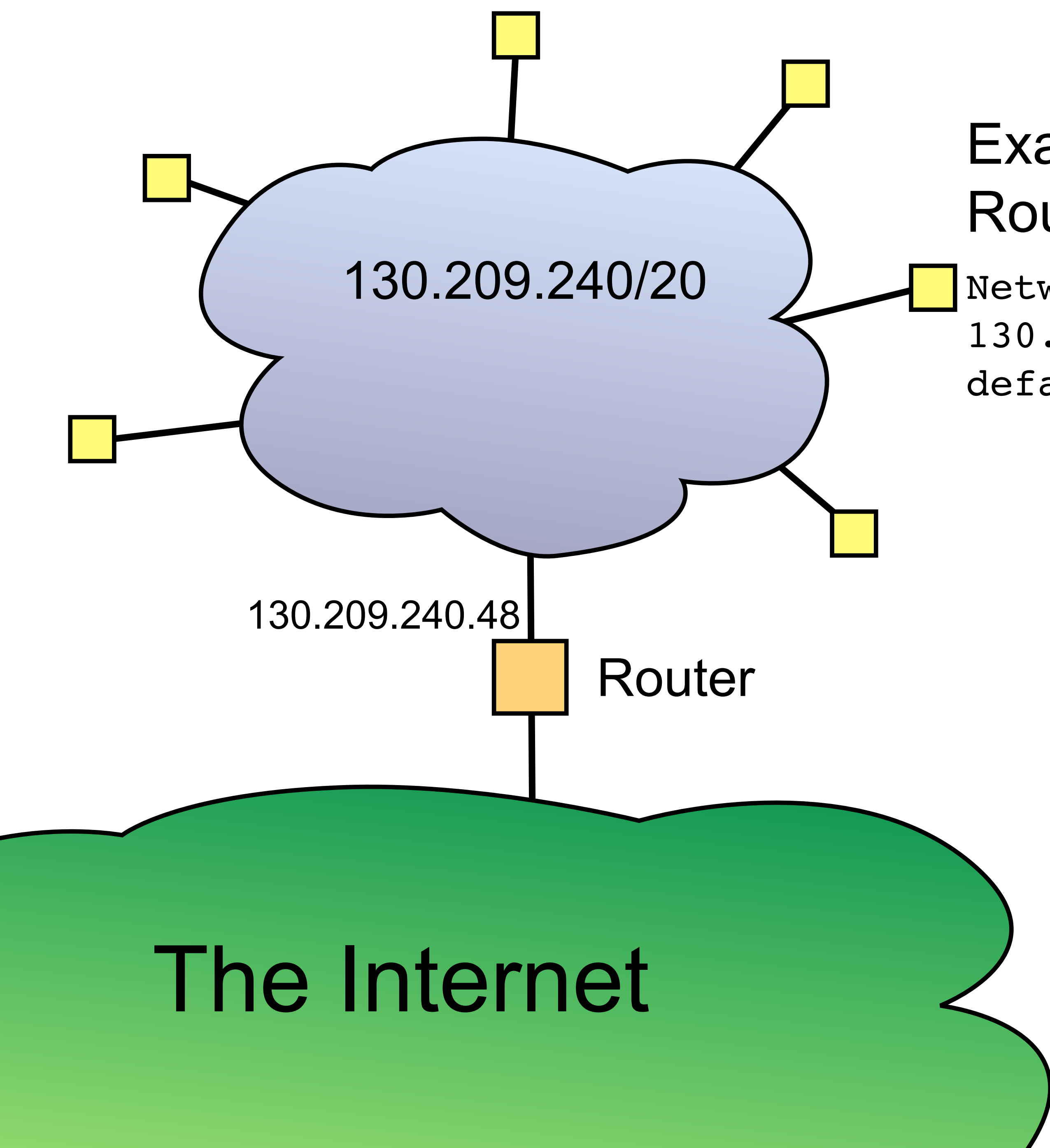
http://www.caida.org/research/topology/as_core_network/2014/

Default Routes and the DFZ

- The AS-level topology:
  - Well connected core networks
  - Sparsely connected edges, getting service from the core networks
- Edge networks can use a *default route* to the core
- Core networks need a full routing table
  - The *default free zone* (DFZ)



Routing at the Edge



Example:
Routing table for hosts in Glasgow SoCS

Network:	Netmask:	Gateway:
130.209.240.0	255.255.240.0	eth0
default	0.0.0.0	130.209.240.48

Routing in the DFZ

- Core networks are well-connected, must know about every other network
  - The *default free zone* where there is no default route
  - Route based on policy, not necessarily shortest path
    - Use AS x in preference to AS y
    - Use AS x only to reach addresses in this range
    - Use the path that crosses the fewest number of ASes
    - Avoid ASes located in that country
  - Requires complete AS-level topology information

Routing Policy

- Interdomain routing is between competitors
  - ASes are network operators and businesses that compete for customers
  - Implication: an AS is unlikely to trust its neighbours
- Routing must consider policy
  - Policy restrictions on who can determine your topology
  - Policy restrictions on which route data can follow
  - Prefer control over routing, even if that means data doesn't necessarily follow the best (shortest) path – the shortest path might pass through a competitor's network, or a country you politically disagree with, or over an expensive link…

Border Gateway Protocol

- Interdomain routing in the Internet uses the Border Gateway Protocol (BGP)
  - External BGP (eBGP) used to exchange routing information between ASes
    - Neighbouring ASes configure an eBGP session to exchange routes
    - Runs over a TCP connection between routers; exchanges knowledge of the AS graph topology
    - Used to derive "best" route to each destination; installed in routers to control forwarding
  - Internal BGP (iBGP) propagates routing information to routers within an AS
    - The intra-domain routing protocol handles routing within the AS
    - iBGP distributes information on how to reach external destinations

Routing Information Exchanged in eBGP

- eBGP routers advertise lists of IP address ranges ("prefixes") and their associated AS-level paths
- Combined to form a routing table

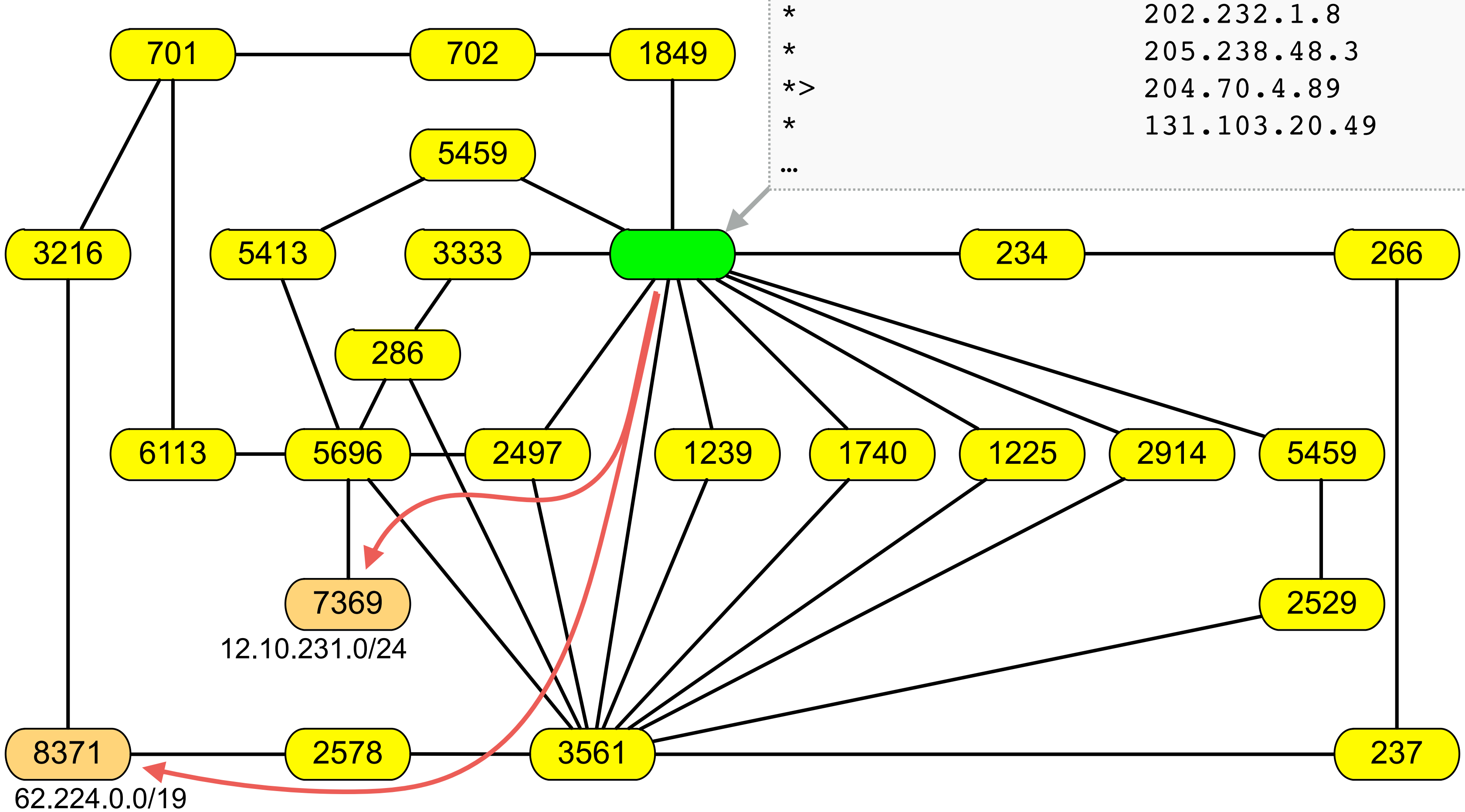
```
   Prefix            Next Hop           AS Path
…
*  12.10.231.0/24    194.68.130.254     5459 5413 5696 7369 i
*                    158.43.133.48      1849 702 701 6113 5696 7369 i
*                    193.0.0.242        3333 286 5696 7369 i
*                    204.212.44.128     234 266 237 3561 5696 7369 i
*>                   202.232.1.8        2497 5696 7369 i
*                    204.70.4.89        3561 5696 7369 i
*                    131.103.20.49      1225 3561 5696 7369 i
*  62.224.0.0/19     134.24.127.3       1740 3561 2578 8371 i
*                    194.68.130.254     5459 2529 3561 2578 8371 i
*                    158.43.133.48      1849 702 701 3216 3216 3216 8371 8371 i
*                    193.0.0.242        3333 286 3561 2578 8371 i
*                    144.228.240.93     1239 3561 2578 8371 i
*                    204.212.44.128     234 266 237 3561 2578 8371 i
*                    202.232.1.8        2497 3561 2578 8371 i
*                    205.238.48.3       2914 3561 2578 8371 i
*>                   204.70.4.89        3561 2578 8371 i
*                    131.103.20.49      1225 3561 2578 8371 i
…
```

Hosts with IP addresses in the range 12.10.231.0 - 12.10.231.255 are in AS 7369. That AS is best reached via AS 2497 and then AS 5696. Packets destined for those addresses should be sent to address 202.232.1.8 next, from where they will be forwarded.

AS Topology Graph

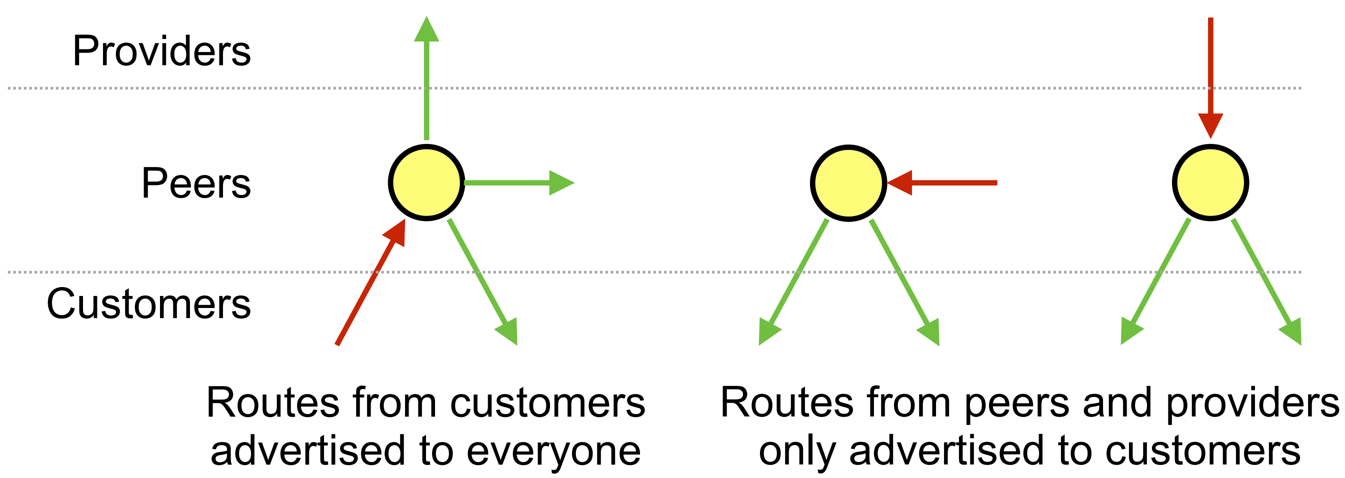
An example fragment of the AS topology graph:

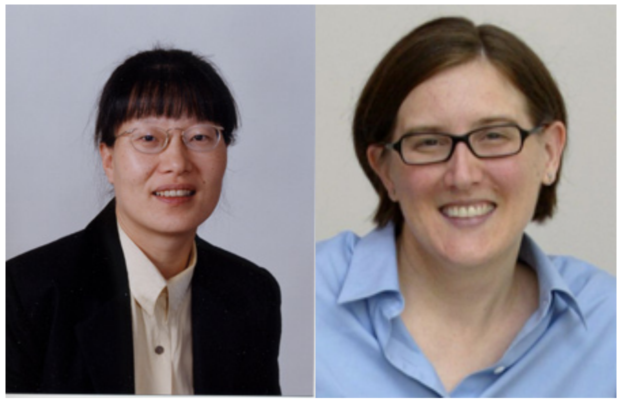
```
   Prefix            Next Hop          AS Path
…
*  12.10.231.0/24    194.68.130.254    5459 5413 5696 7369 i
*                    158.43.133.48     1849 702 701 6113 5696 7369 i
*                    193.0.0.242       3333 286 5696 7369 i
*                    204.212.44.128    234 266 237 3561 5696 7369 i
*>                   202.232.1.8       2497 5696 7369 i
*                    204.70.4.89       3561 5696 7369 i
*                    131.103.20.49     1225 3561 5696 7369 i
*  62.224.0.0/19     134.24.127.3      1740 3561 2578 8371 i
*                    194.68.130.254    5459 2529 3561 2578 8371 i
*                    158.43.133.48     1849 702 701 3216 3216 3216 8371 8371 i
*                    193.0.0.242       3333 286 3561 2578 8371 i
*                    144.228.240.93    1239 3561 2578 8371 i
*                    204.212.44.128    234 266 237 3561 2578 8371 i
*                    202.232.1.8       2497 3561 2578 8371 i
*                    205.238.48.3      2914 3561 2578 8371 i
*>                   204.70.4.89       3561 2578 8371 i
*                    131.103.20.49     1225 3561 2578 8371 i
…
```



Routing Policy in eBGP

- Each AS chooses what routes to advertise to its neighbours
- Doesn't need to advertise everything it receives
    - Usual to drop some routes from the advertisement – depends on the chosen routing policy
    - Common approach: the Gao-Rexford rules:





Lixin Gao Jennifer Rexford

Ensures the AS graph is a valley-free DAG (recommended, but not required, policy)

BGP Routing Decision Process

- BGP routers receive path vectors from neighbouring ASes giving possible routes to prefixes
  - Filtered based on the policy of each AS in the path from the source
- BGP decision process is complex and policy-driven
  - Choose what route to install for destination prefix in forwarding table based on multiple criteria – policy, shortest path, etc.
  - BGP doesn't always find a route, even if one exists, as may be prohibited by policy
  - Routes are often not the shortest AS path
  - Mapping business goals to BGP policies is a poorly documented process, with many operational secrets

Table 2: Simplified BGP decision process [6, 24]. This table was also provided with the survey.

#	Criteria
1	Highest LocalPref
2	Lowest AS Path Length
3	Lowest origin type
4	Lowest MED
5	eBGP-learned over iBGP-learned
6	Lowest IGP cost to border router (hot-potato routing)
7	If both paths are external, prefer the path that was received first (i.e., the oldest path) [6]
8	Lowest router ID (to break ties)

Source: Phillipa Gill, Michael Schapira, and Sharon Goldberg, "A Survey of Interdomain Routing Policies", ACM CCR, V44, N1, January 2014, p29-34

Summary

- The interdomain routing problem
  - Autonomous systems
  - Routing on the AS graph
  - Trust and policy constraints
- Interdomain routing in the Internet
  - BGP