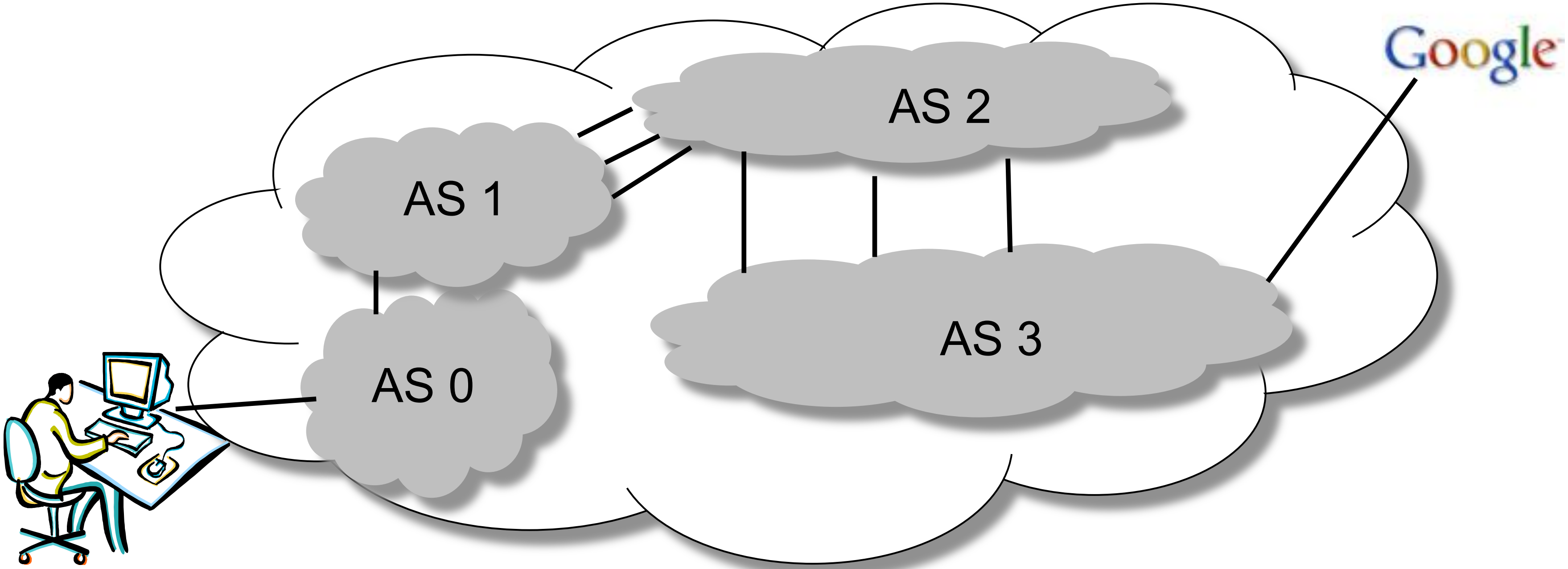
Internet measurements: topology discovery and dynamics

Renata Teixeira
MUSE Team
Inria Paris-Rocquencourt

Why measure the Internet topology?

- Network operators
  - Assist in network management, fault diagnosis
- Distributed services and applications
  - Select the best paths to use
- Researchers
  - Properties of Internet structure, dynamics
  - Economics of the Internet

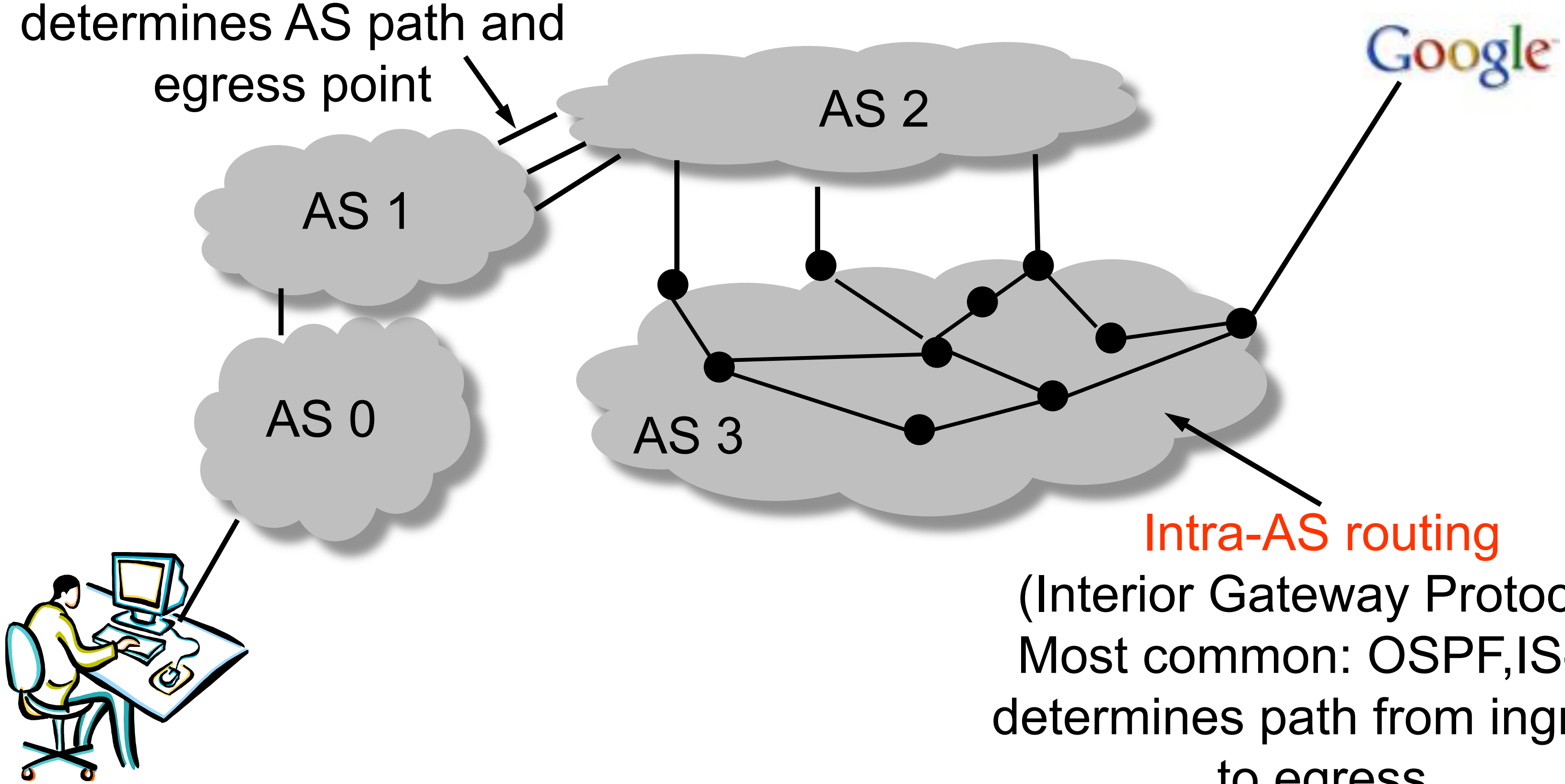
Internet: network of networks



- Internet = interconnection of Autonomous Systems (AS)
  - Distinct regions of administrative control
  - Routers/links managed by a single “institution”
  - Service provider, company, university, etc.

Hierarchical routing

Inter-AS routing
(Border Gateway Protocol)
determines AS path and
egress point



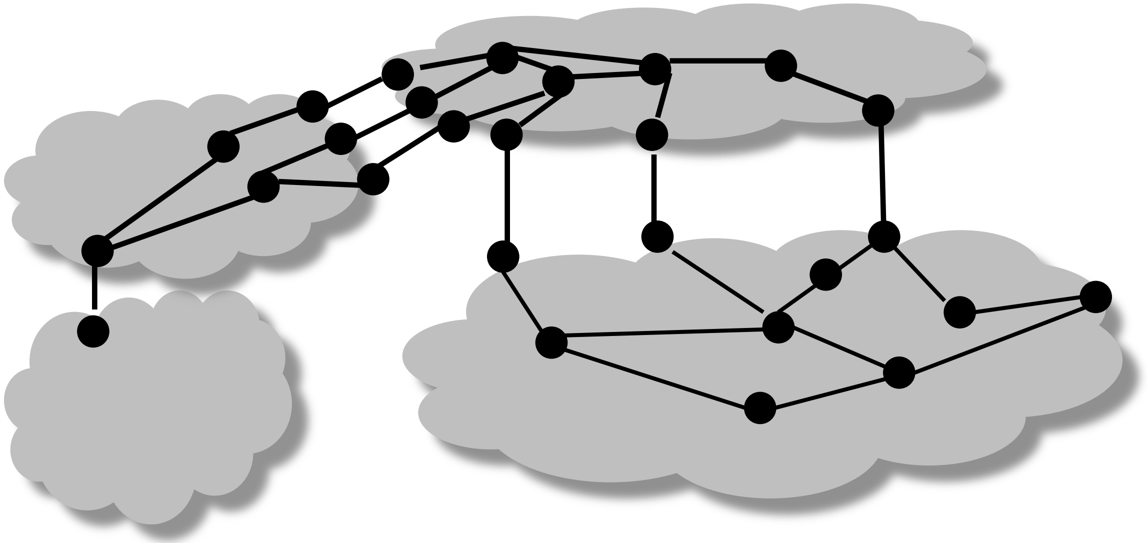
Intra-AS routing
(Interior Gateway Protocol)
Most common: OSPF,IS-IS
determines path from ingress
to egress

Outline

- Router-level topologies
  - Common network designs
  - Measuring with access to routers: OSPF/IS-IS monitors
  - Measuring without access to routers: Traceroute
- AS-level topology
  - Business relationships between ASes
  - BGP: Internet's inter-domain routing
  - Inferring AS topology from BGP and traceroute

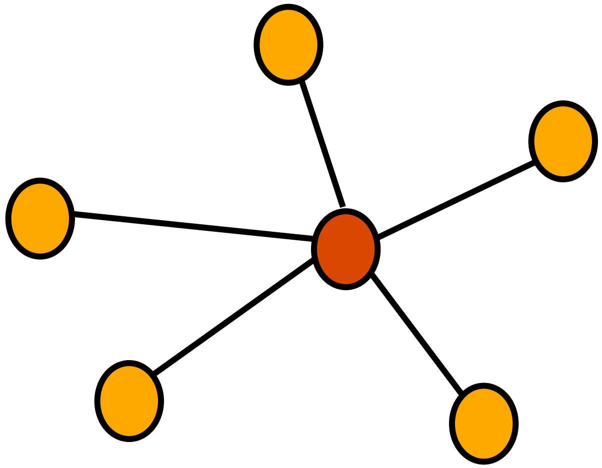
Router topology

- Node: router
- Edge: link

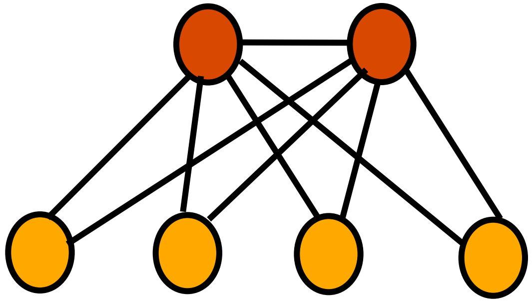


Hub-and-spoke topology

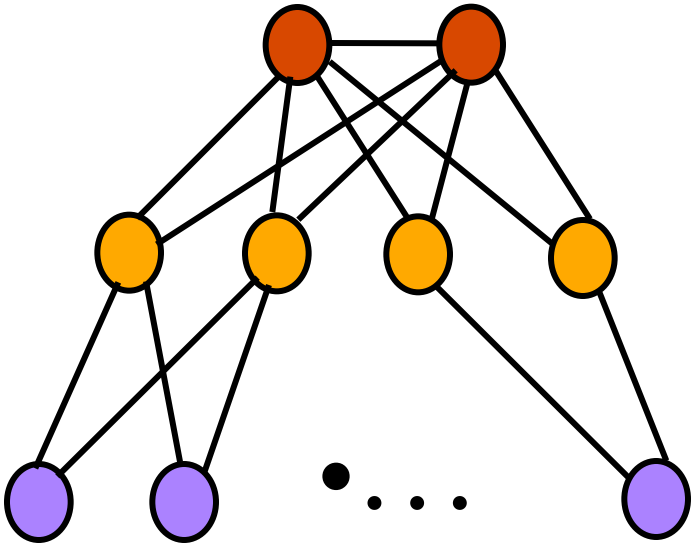
- Single hub node
  - Common in enterprise networks
  - Main location and satellite sites
  - Simple design and trivial routing
- Problems
  - Single point of failure
  - Bandwidth limitations
  - High delay between sites
  - Costs to backhaul to hub



Simple alternatives to hub-and-spoke

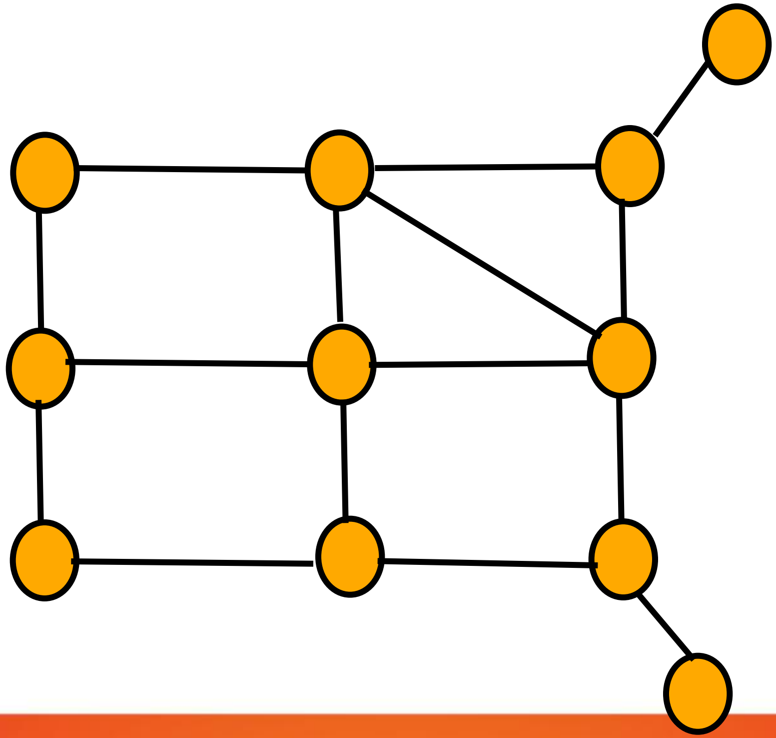


- Dual hub-and-spoke
  - Higher reliability
  - Higher cost
  - Good building block
- Levels of hierarchy
  - Reduce backhaul cost
  - Aggregate the bandwidth
  - Shorter site-to-site delay



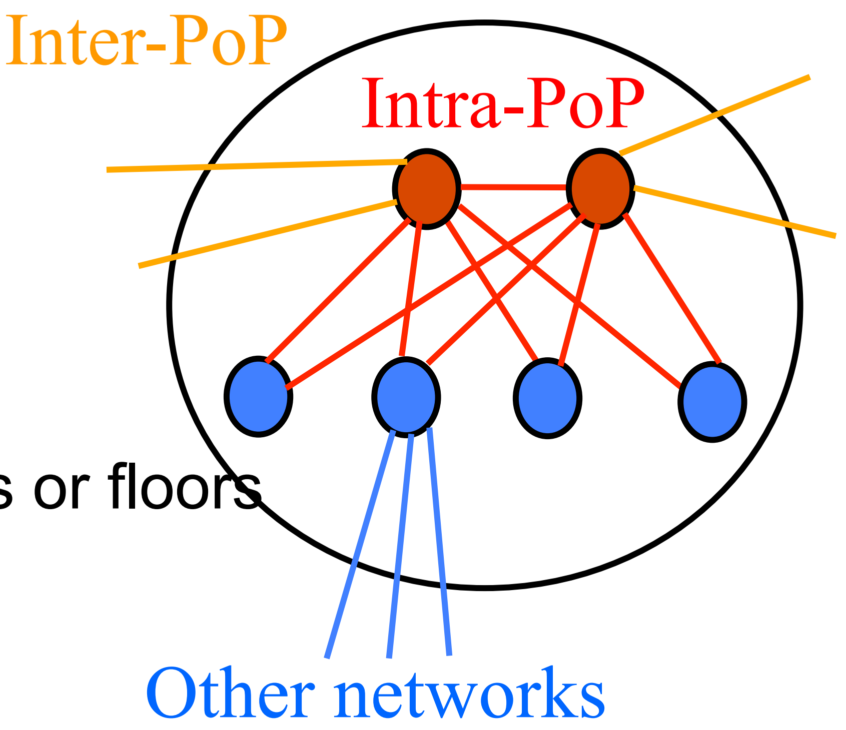
Backbone networks

- Multiple Points-of-Presence (PoPs)
- Lots of communication between PoPs
- Accommodate traffic demands and limit delay



Points-of-Presence (PoPs)

- Inter-PoP links
  - Long distances
  - High bandwidth
- Intra-PoP links
  - Short cables between racks or floors
  - Aggregated bandwidth
- Links to other networks
  - Wide range of media and bandwidth



Measuring router topology

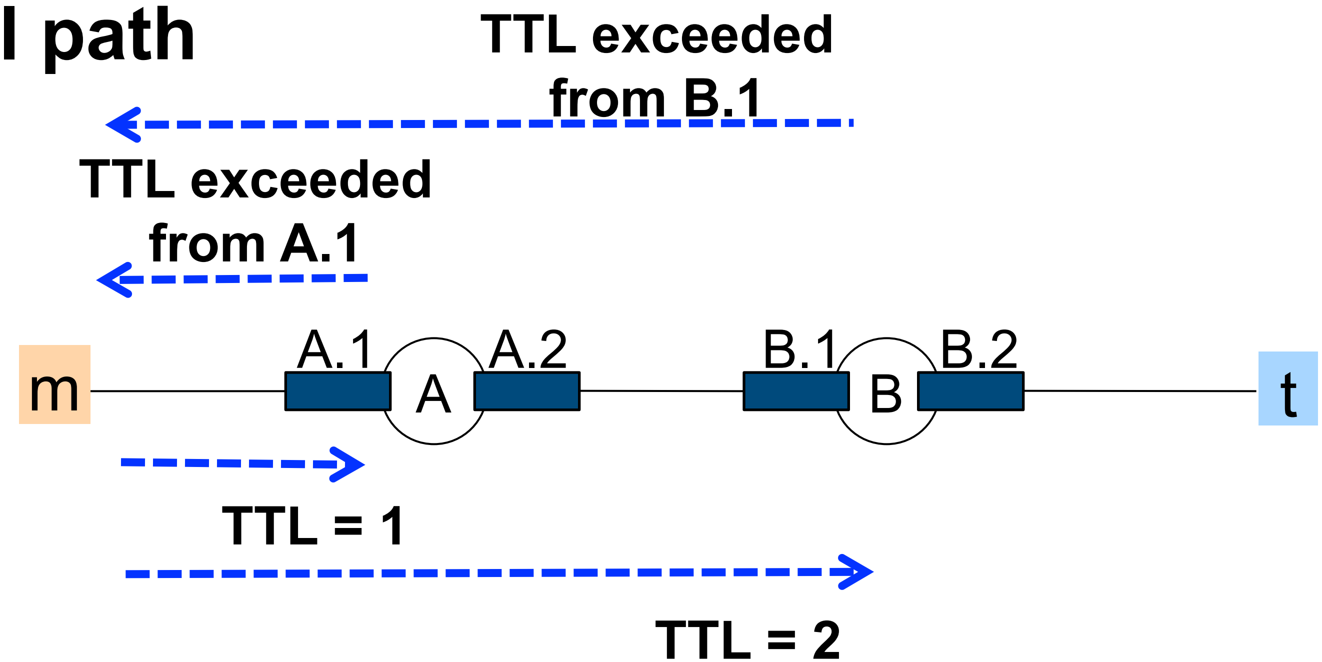
- With access to routers
  - Topology of one network
  - Routing monitors (OSPF or IS-IS)
- No access to routers
  - Multi-AS topology or from end-hosts
  - Monitors issue active probes: traceroute

Router topology from routing messages

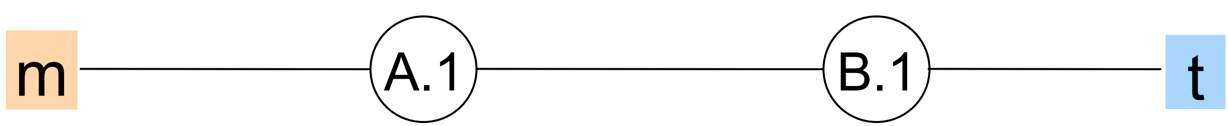
- Routing protocols flood state of each link
  - Periodically refresh link state
  - Report any changes: link down, up, cost change
- Monitor listens to link-state messages
  - Acts as a regular router
    - AT&T's OSPFmon or Sprint's PyRT for IS-IS
- Combining link states gives the topology
  - Easy to maintain, messages report any changes

Inferring a path without access to routers: traceroute

Actual path



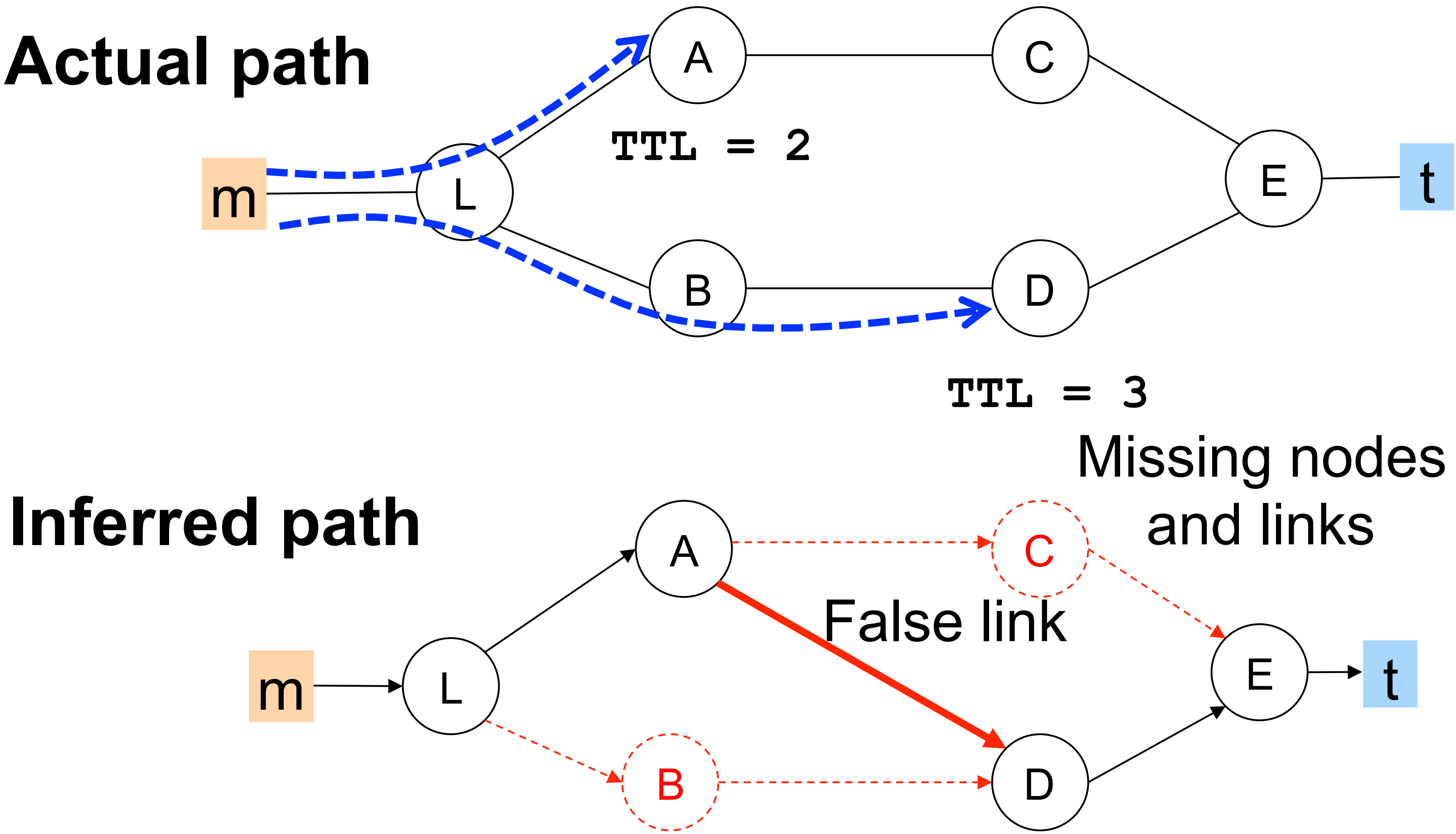
Inferred path



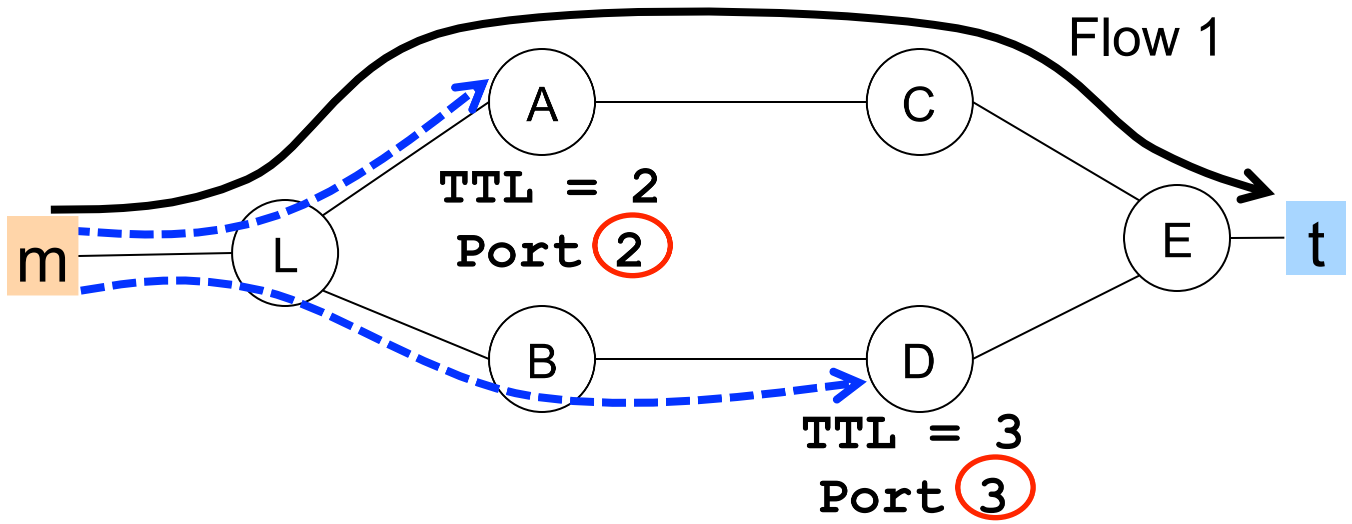
A traceroute path can be incomplete

- Load balancing is widely used
  - Traceroute only probes one path
- Sometimes taceroute has no answer (stars)
  - ICMP rate limiting
  - Anonymous routers
- Tunnelling (e.g., MPLS) may hide routers
  - Routers inside the tunnel may not decrement TTL

Traceroute under load balancing



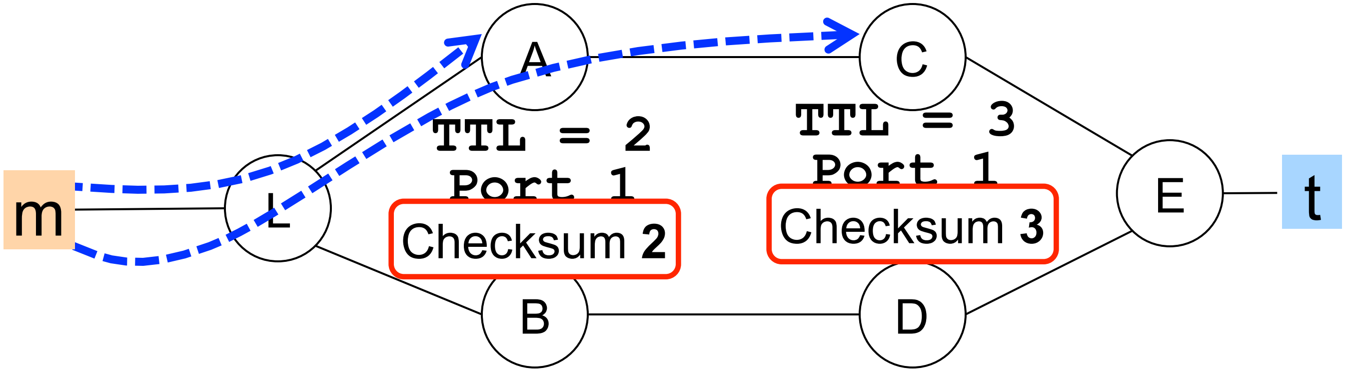
Errors happen even under per-flow load balancing



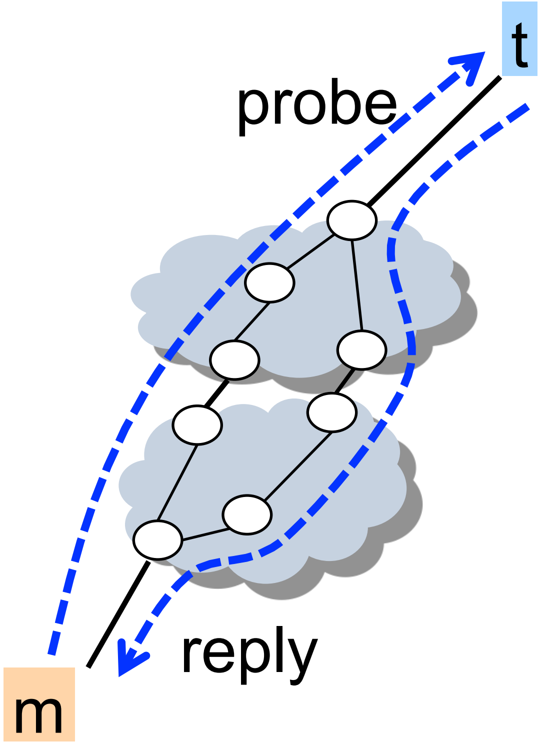
- Traceroute uses the destination port as identifier
  - Needs to match probe to response
  - Response only has the header of the issued probe

Paris traceroute

- Solves the problem with per-flow load balancing
  - Probes to a destination belong to same flow
- Changes the location of the probe identifier
  - Use the UDP checksum

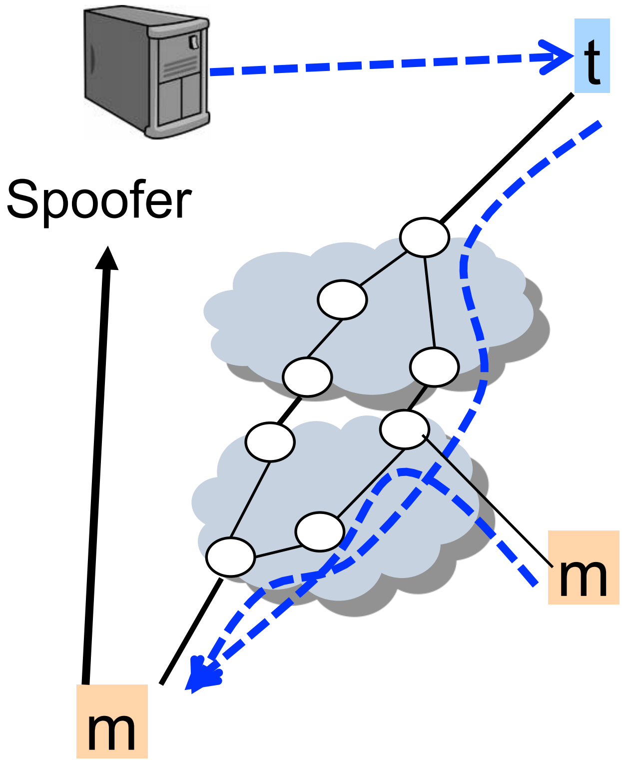


Traceroute measures the forward path



- Paths can be asymmetric
  - Load balancing
  - Hot-potato routing

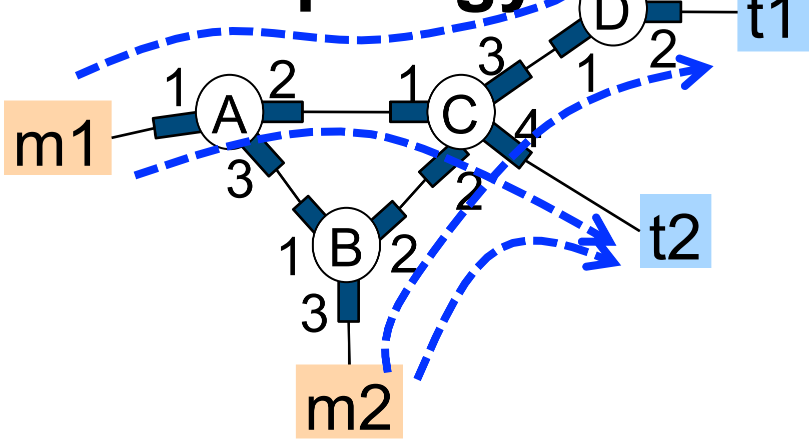
Reverse traceroute



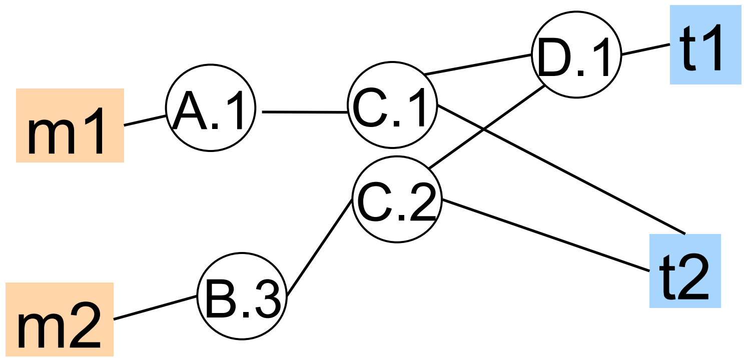
- IP options work on forward and reverse path
  - Record Route (RR) option: 9 hops
- Leverage multiple monitors
  - Get baseline paths
  - Assume destination-based routing
- Spoofing to select best monitor
  - Spoofer sends spoofed probe with source address of the monitor

Topology from traceroutes

Actual topology



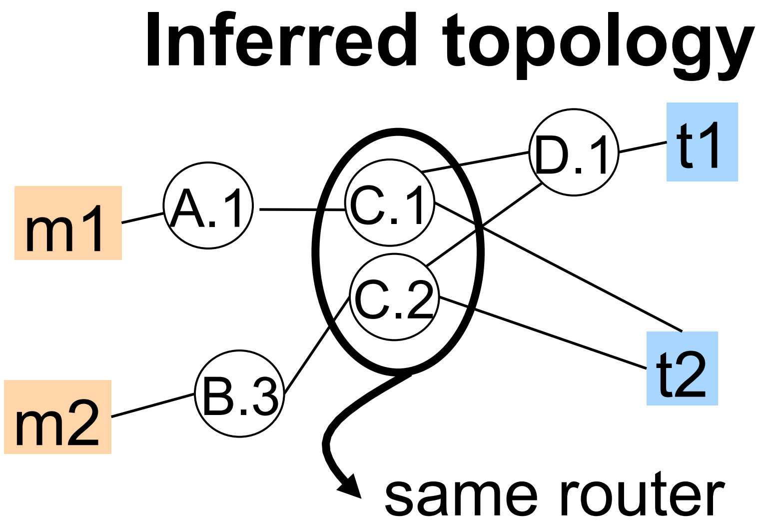
Inferred topology



- Inferred nodes = interfaces, not routers
- Coverage depends on monitors and targets
  - Misses links and routers
  - Some links and routers appear multiple times

Alias resolution: Map interfaces to routers

- Direct probing
  - Probe an interface, may receive response from another
  - Responses from the same router will have close IP identifiers and same TTL
- Record-route IP option
  - Records up to nine IP addresses of routers in the path
- CAIDA's MIDAR tool
  - Large scale alias resolution



Large-scale topology: coverage

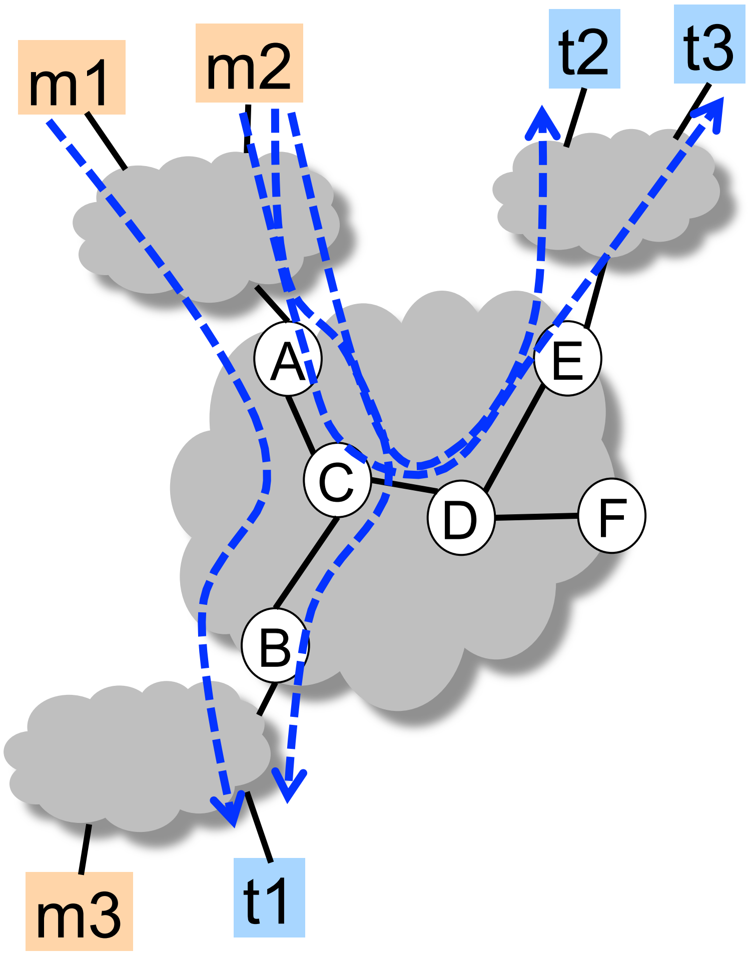
- Few monitors, lots of destinations
  - Deploying monitors is hard
  - Can probe any destination connected to the Internet
- Example: CAIDA's Ark
  - Monitors: 94
  - Destinations: All routed /24 IPv4 prefixes (9.5 million)
  - Optimization: Group of monitors split destination list
    - Measures full destination list in 2/3 days

Increasing the number of monitors

- Peer-to-peer monitoring software
  - E.g.: Dimes (~400); EdgeScope (~900K)
  - Advantage: Easy deployment
  - Problem: little control
- Low cost monitors
  - E.g.: Ark's Raspberry Pi monitor, RIPE Atlas
  - Advantage: more control
  - Problem: Need more user engagement

Inferring topology of one AS

- Rocketfuel topologies
  - Only one traceroute that enter in one ingress and leave via the same egress
  - Alias resolution with IPID
  - DNS names to map routers to PoPs
- Topology errors
  - Missed links: lack of vantage points, incomplete traceroutes
  - Added links: incorrect alias resolution, adding reverse links

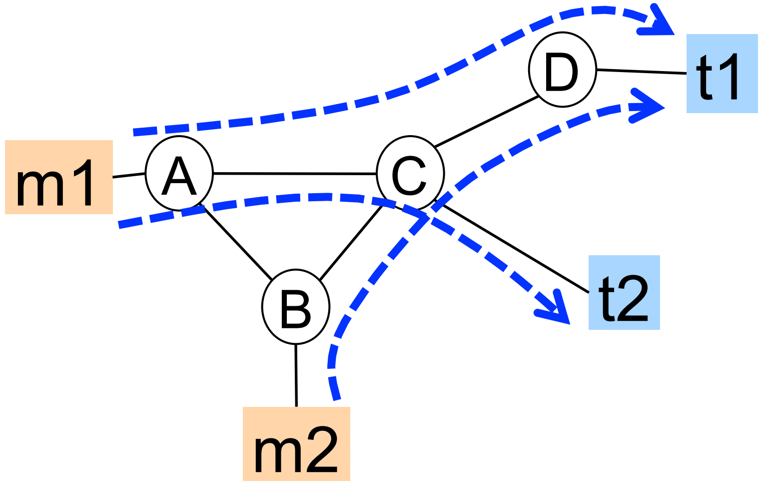


Measuring topology dynamics

- Probing a large topology takes time
  - E.g., probing 1200 targets from PlanetLab nodes takes 5 minutes on average (using 30 threads)
  - Probing more targets covers more links
  - But, getting a topology snapshot takes longer
- Snapshot may be inaccurate
  - Paths may change during snapshot
- Hard to get up-to-date topology
  - To know that a path changed, need to re-probe

Faster topology snapshots with tree assumption

- Probing redundancy
  - Intra-monitor
  - Inter-monitor
- Doubletree
  - Assume tree structure
  - Combines backward and forward probing to eliminate redundancy
- Topology errors
  - Load balancing and traffic engineering violate tree assumption



Tracking large number of paths with multi-path detection

- Observation: Internet paths are mostly stable
  - Repeatedly probing paths waste probes
- Dtrack: Probe according to path stability
  - Change detection: lightweight probing for speed
    - Allocates more probes to unstable paths
  - Path remapping: accuracy with Paris traceroute
    - Local remapping

Summary: Router-level topologies

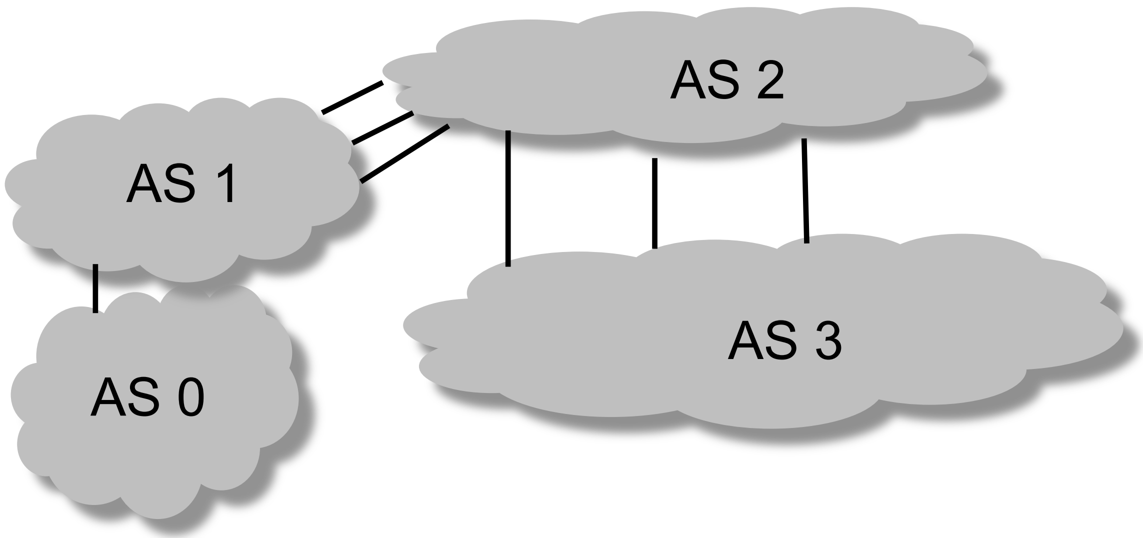
- With access to routers
  - Topology of one AS
  - Observe routing messages
- Without access to routers
  - Traceroute + alias resolution
  - Challenges
    - Incomplete traceroutes
    - Cover all routers and links in Internet
    - Probe fast enough to observe fine-grained dynamics

Outline

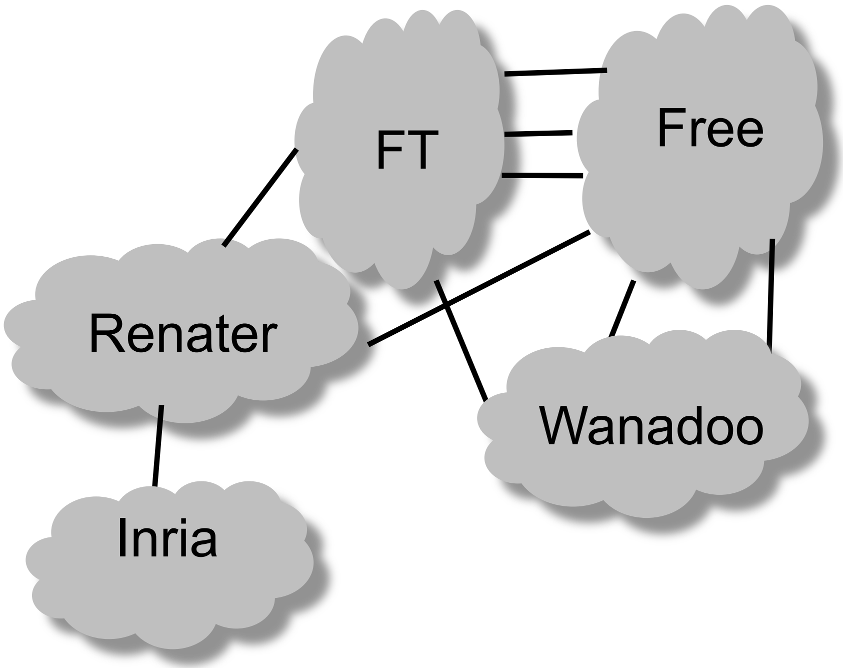
- Router-level topologies
  - Common network designs
  - Measuring with access to routers: OSPF/IS-IS monitors
  - Measuring without access to routers: Traceroute
- AS-level topology
  - Business relationships between ASes
  - BGP: Internet's inter-domain routing
  - Inferring AS topology from BGP

AS topology

- Node: AS
- Edge: relationship between ASes

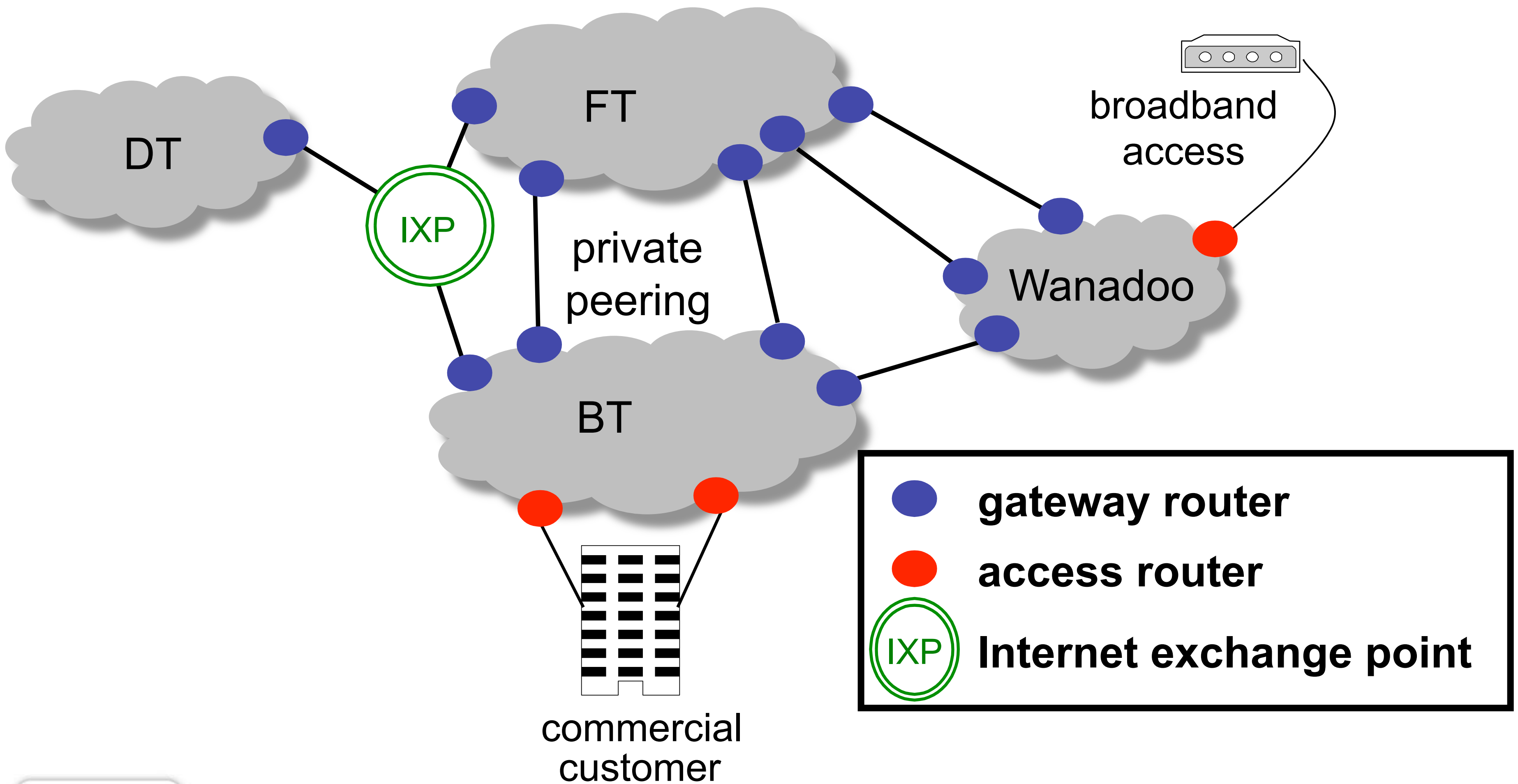


Hierarchy of ASes



- Large, tier-1 provider with a nationwide backbone
  - At the "core" of the Internet, don't have providers
- Medium-sized regional provider with smaller backbone
- Small network run by a single company or university

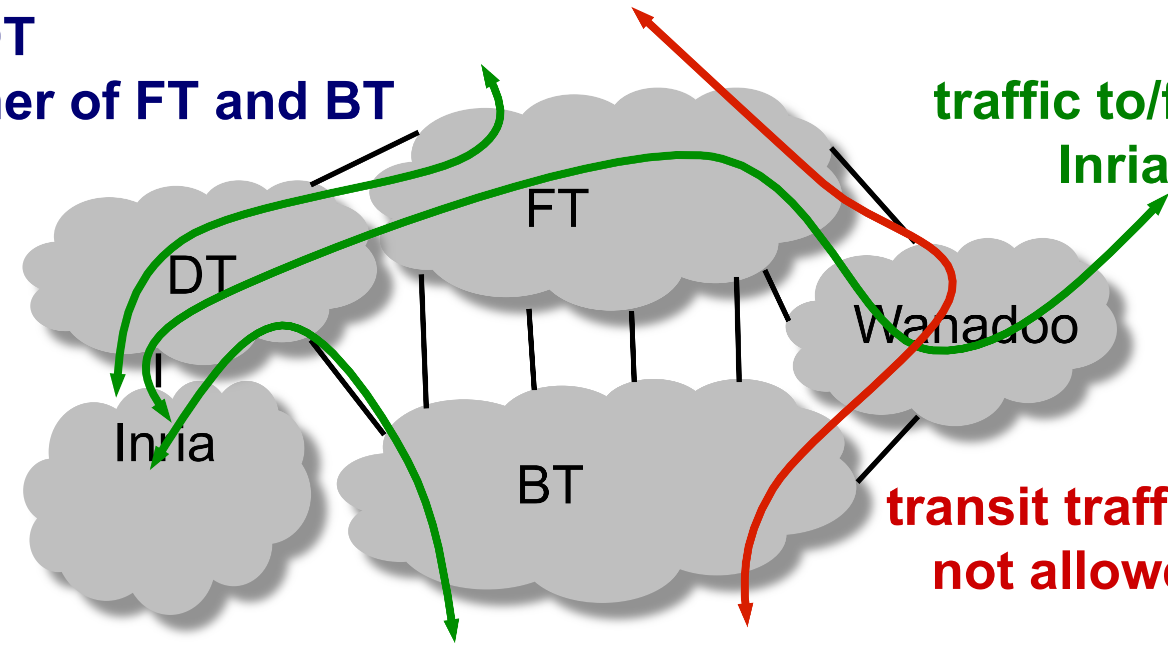
Connections between networks



Customer-provider relationship

- Customer needs to be reachable from everyone
  - Provider exports routes learned from customer to everyone
- Customer does not want to provide transit service
  - Customer does not export from one provider to another

Inria is customer of DT
Wanadoo is a customer of FT and BT



traffic to/from Inria

transit traffic is not allowed

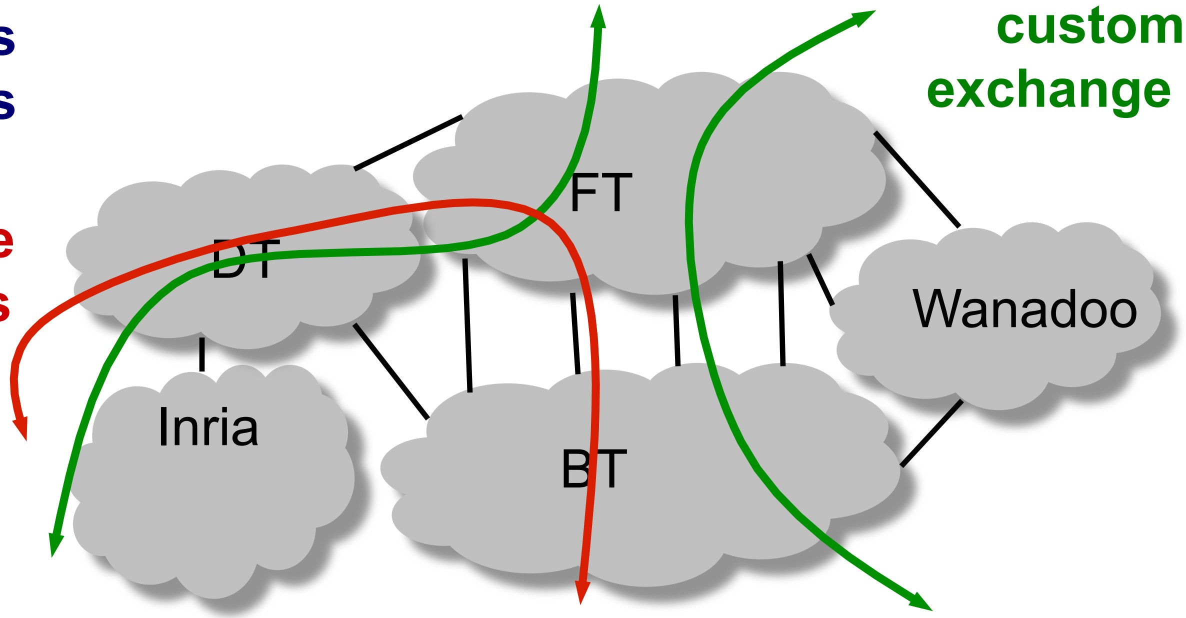
Peer-peer relationship

- Peers exchange traffic between customers
  - AS exports only customer routes to a peer
  - AS exports a peer's routes only to its customers

FT and BT are peers
FT and DT are peers

FT doesn't provide transit for its peers

customers exchange traffic

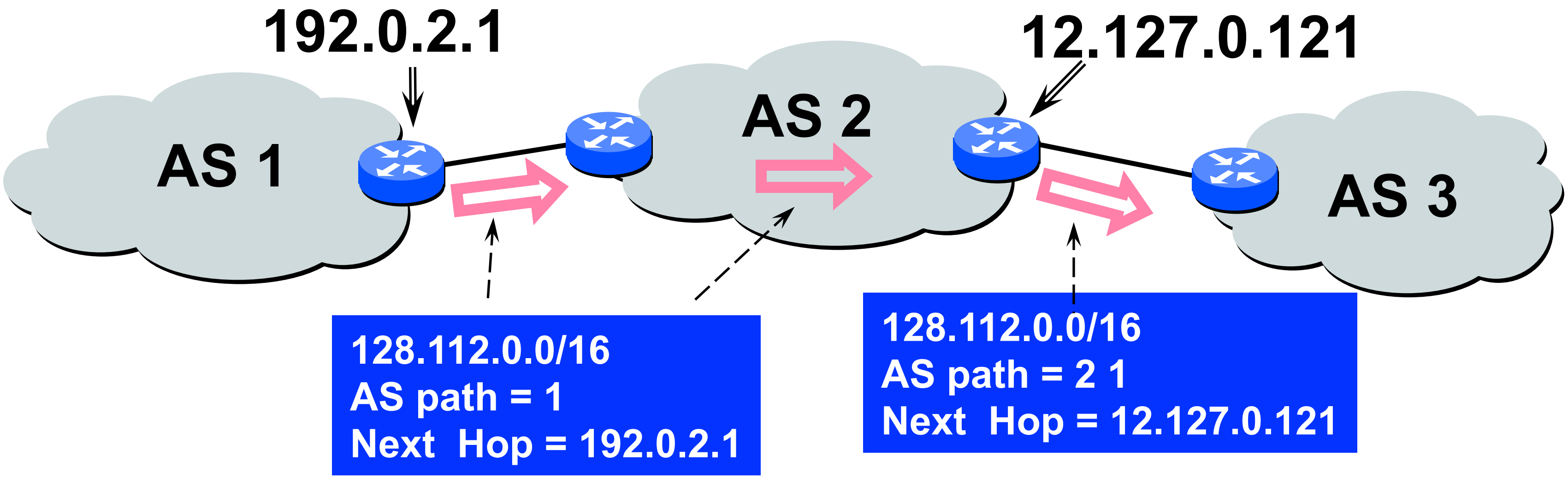


Border Gateway Protocol (BGP)

- Inter-domain routing protocol for the Internet
  - Prefix-based path-vector protocol
  - Policy-based routing based on AS Paths
  - Evolved during the past 20 years

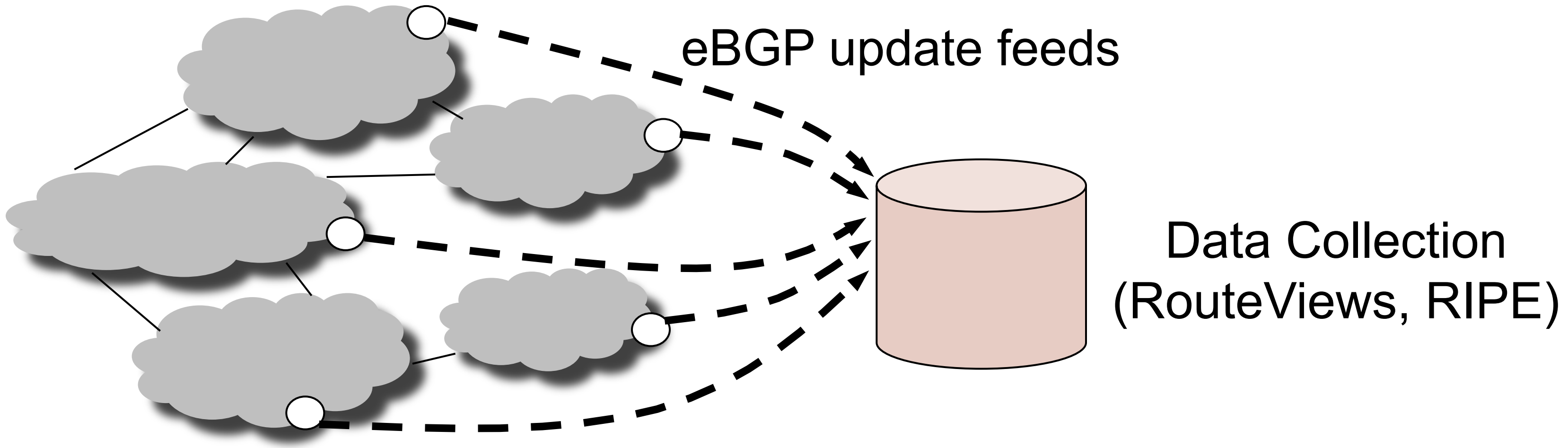
BGP route

- Destination prefix (e.g,. 128.112.0.0/16)
- Route attributes, including
  - AS path (e.g., “2 1”)
  - Next-hop IP address (e.g., 12.127.0.121)



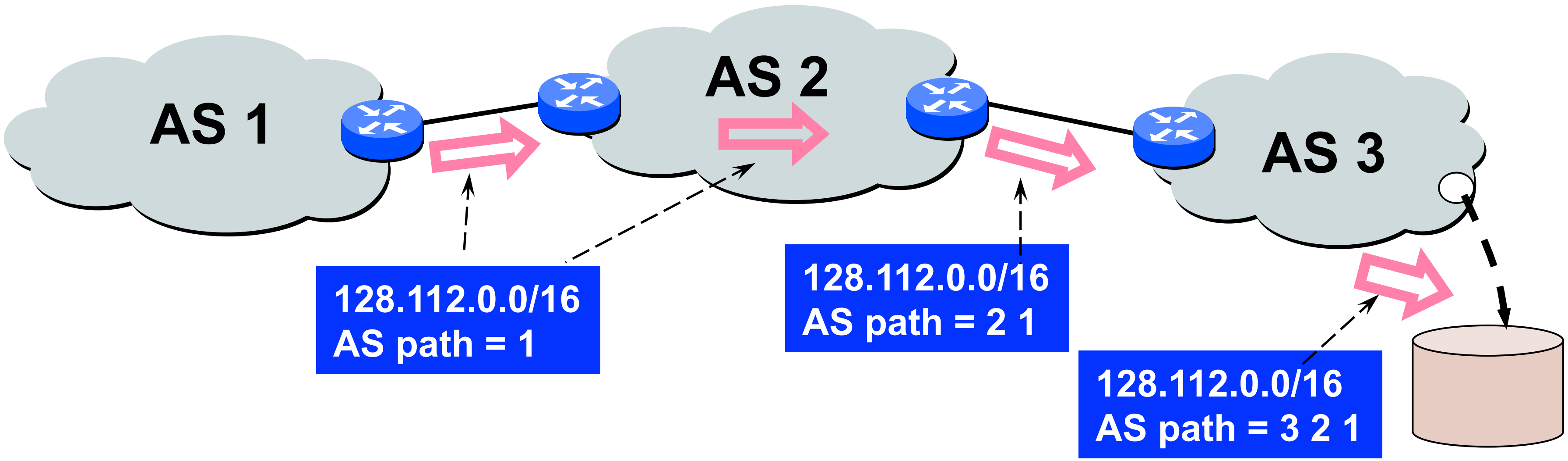
Passive BGP measurements

- Passive measurements: public BGP data
  - RouteViews, RIPE RIS



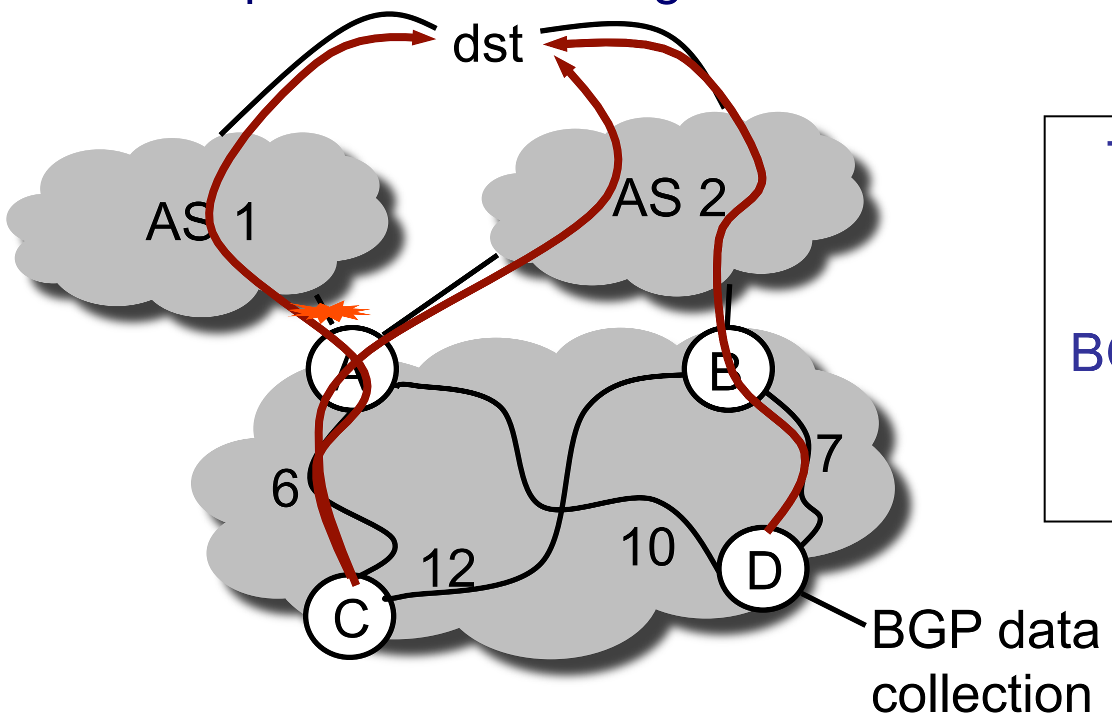
AS topology from BGP data

- Example: AS path = 3 2 1
  - Nodes: 1, 2, 3
  - Edges: (1,2), (2,3)



Problem: Each router’s view is unique

Myth: The BGP updates from a single router accurately represent the AS.

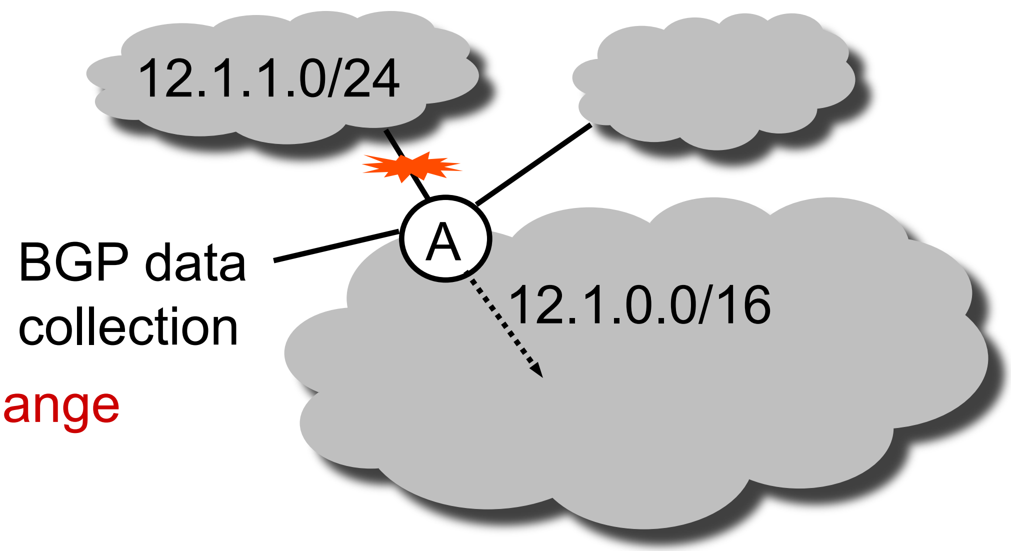


The measurement system needs to capture the BGP routing changes from all border routers

No change

Problem: Route aggregation hides information

Myth:BGP data from a router accurately represents changes on that router.



No change

The measurement system needs to know
all routes the router knows.

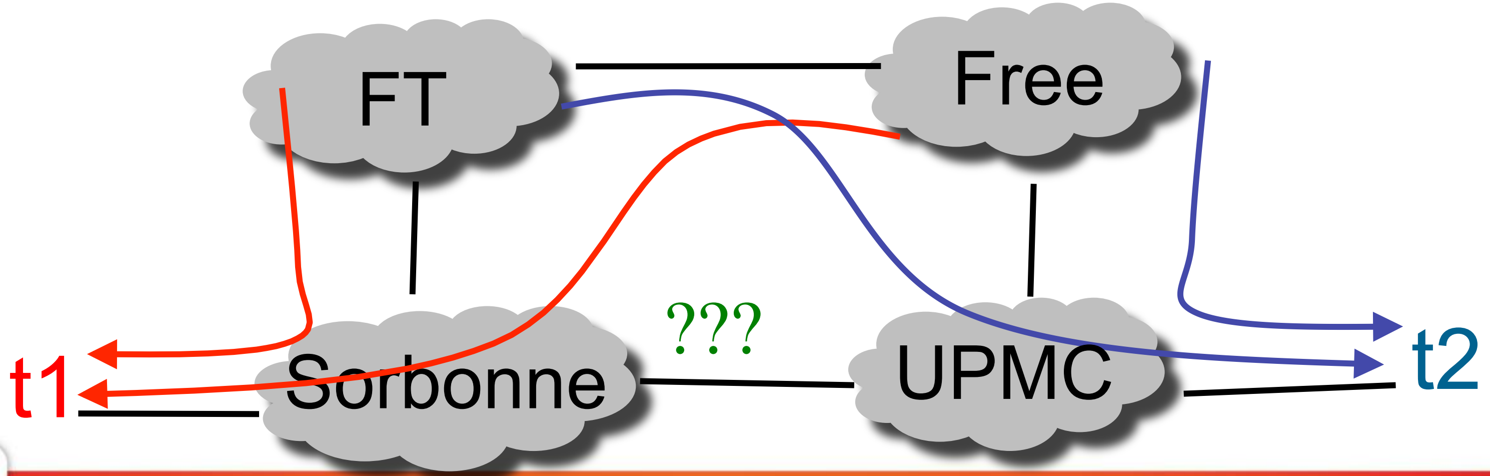
Using traceroutes to improve AS topologies

1	169.229.62.1	AS25	Berkeley
2	169.229.59.225	AS25	
3	128.32.255.169	AS25	
4	128.32.0.249	AS25	
5	128.32.0.66	AS11423	Calren
6	209.247.159.109	AS3356	Level3
7	*	AS3356	
8	64.159.1.46	AS3356	
9	209.247.9.170	AS3356	
10	66.185.138.33	AS1668	AOL
11	*	AS1668	
12	66.185.136.17	AS1668	
13	64.236.16.52	AS5662	CNN

- IP to AS mapping
  - Internet registries: Whois
  - Origin AS of BGP prefix

Challenges of Inter-AS Mapping

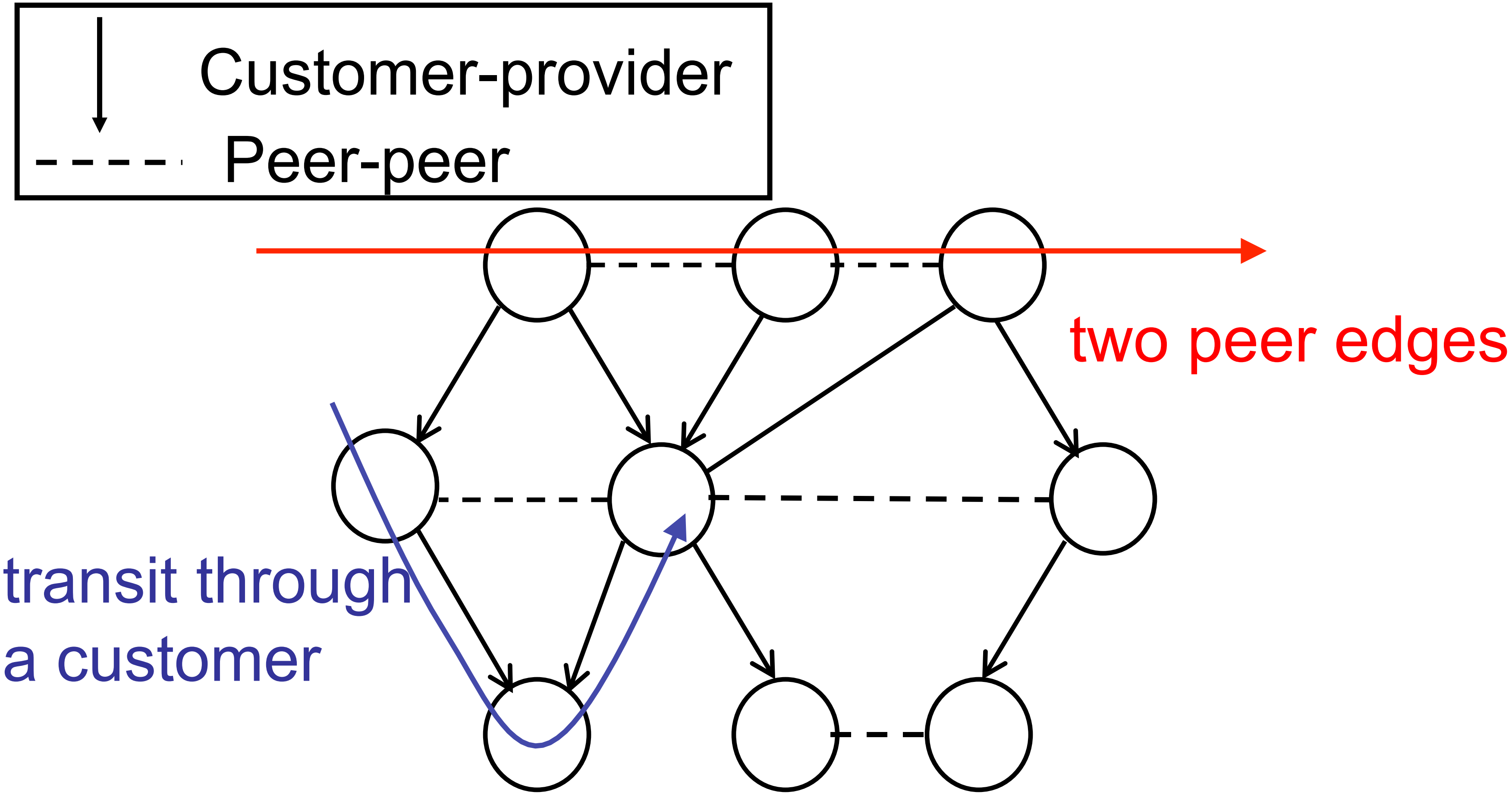
- Mapping traceroute hops to ASes is hard
  - Need an accurate registry of IP address ownership
  - Whois data are notoriously out of date
- Collecting diverse interdomain data is hard
  - Especially hard to see peer-peer edges



Inferring AS Relationships

- Key idea
  - The business relationships determine the routing policies
  - The routing policies determine the paths that are chosen
  - So, look at the chosen paths and infer the policies
- Example: AS path "1 7018 88" implies
  - AS 7018 allows AS 1 to reach AS 88
  - Each "triple" tells something about transit service
- Collect and analyze AS path data
  - Identify which ASes can transit through the other
  - … and which other ASes they are able to reach this way

Paths you should never see ("Invalid")



Challenges of relationship inference

- Incomplete measurement data
  - Hard to get a complete view of the AS graph
  - Especially hard to see peer-peer edges low in hierarchy
- Real relationships are sometime more complex
  - Peer is one part of the world, customer in another
  - Other kinds of relationships (e.g., backup and sibling)
  - Special relationships for certain destination prefixes
- EdgeScope: more complete AS topologies
  - Traceroutes from Bittorrent clients + sophisticated heuristics

Summary: AS-level topologies

- Sources of AS paths
  - Public BGP repositories
  - Traceroutes + IP-AS mapping
- Challenges
  - Can’t always model one AS as a node
  - Hard to observe links closer to the edge

REFERENCES

Router-level topology from inside

- IS-IS monitoring
  - R. Mortier, “Python Routeing Toolkit (`PyRT')”, https://research.sprintlabs.com/pyrt/
- OSPF monitoring
  - A. Shaikh and A. Greenberg, “OSPF Monitoring: Architecture, Design and Deployment Experience”, NSDI 2004
- Commercial products
  - Packet Design: http://www.packetdesign.com/

Traceroute

- Original traceroute tool
  - V. Jacobson, traceroute, February, 1989.
- Tracing accurate paths under load-balancing
  - B. Augustin *et al.*, “Avoiding traceroute anomalies with Paris traceroute”, IMC, 2006.
  - D. Veitch, B. Augustin, R. Teixeira, and T. Friedman, " Failure Control in Multipath Route Tracing", in Proc. of IEEE Infocom, April 2009.
- Reverse traceroute
  - E. Katz-Bassett, H. Madhyastha, V. Adhikari, C. Scott, J. Sherry, P. van Wesep, A. Krishnamurthy, T. Anderson, “Reverse traceroute”, NSDI, 2010.

Router-level topology with traceroute

- Use of record route to obtain more accurate topologies
  - R. Sherwood, A. Bender, N. Spring, “DisCarte: A Disjunctive Internet Cartographer”, SIGCOMM, 2008.
- Large-scale alias resolution
  - K. Keys, Y. Hyun, M. Luckie, and k. claffy, "Internet-Scale IPv4 Alias Resolution with MIDAR", IEEE/ACM Transactions on Networking, vol. 21, no. 2, pp. 383--399, Apr 2013.

Optimizing router-level topology discovery

- Reducing overhead to trace topology of a network and alias resolution with direct probing
  - N. Spring, R. Mahajan, and D. Wetherall, “Measuring ISP Topologies with Rocketfuel”, SIGCOMM 2002.
- Reducing overhead to take a topology snapshot
  - B. Donnet, P. Raoult, T. Friedman, and M. Crovella, “Efficient Algorithms for Large-Scale Topology Discovery”, SIGMETRICS, 2005.
- Tracking topology changes
  - I. Cunha, R. Teixeira, D. Veitch, and C. Diot, "Predicting and Tracking Internet Path Changes, in Proc. of ACM SIGCOMM, August 2011.

Macroscopic topology measurement systems

- CAIDA’s Ark
  - http://www.caida.org/projects/ark/
- Dimes
  - http://www.netdimes.org
- iPlane
  - http://iplane.cs.washington.edu/
- Northwestern’s EdgeScope
  - http://aqualab.cs.northwestern.edu/projects/86-edgescope-sharing-the-view-from-a-distributed-internet-telescope

BGP monitors

- RouteViews
  - http://www.routeviews.org/
- RIPE-RIS
  - http://www.ripe.net/data-tools/stats/ris/routing-information-service
- Cyclops: Aggregates data from multiple monitors
  - http://cyclops.cs.ucla.edu/

AS-level topologies

- Obtaining AS paths from traceroutes
  - Z. M. Mao, J. Rexford, J. Wang, R. H. Katz, “Towards an Accurate AS-Level Traceroute Tool”, SIGCOMM 2003.
- More complete AS-level topology
  - K. Chen, D. R. Choffnes, R. Potharaju, Y. Chen, F. E. Bustamante, D. Pei, Y. Zhao, “Where the Sidewalk Ends: Extending the Internet AS Graph Using Traceroutes From P2P Users”, CoNEXT, 2009.
- More accurate model of the AS topology
  - W. Mühlbauer, A. Feldmann, O. Maennel, M. Roughan, and S. Uhlig, “Building an AS-topology model that captures route diversity” ACM SIGCOMM 2006.

AS relationship inference

- L. Subramanian, S. Agarwal, J. Rexford, and R. H. Katz, "Characterizing the Internet hierarchy from multiple vantage points," IEEE INFOCOM, 2002
- M. Luckie, B. Huffaker, k. claffy, A. Dhamdhere, and V. Giotsas, "AS Relationships, Customer Cones, and Validation", IMC, 2013.

AS-level topologies: Be aware

- R. V. Oliveira, D. Pei, Walter Willinger, B. Zhang, L. Zhang, “The (in)completeness of the observed internet AS-level structure”, IEEE/ACM Trans. Netw. 18(1), 2010.
- M. Roughan, W. Willinger, O. Maennel, D. Perouli, R. Bush “10 Lessons from 10 Years of Measuring and Modeling the Internet's Autonomous Systems”, IEEE JSAC 29(9), 2011.