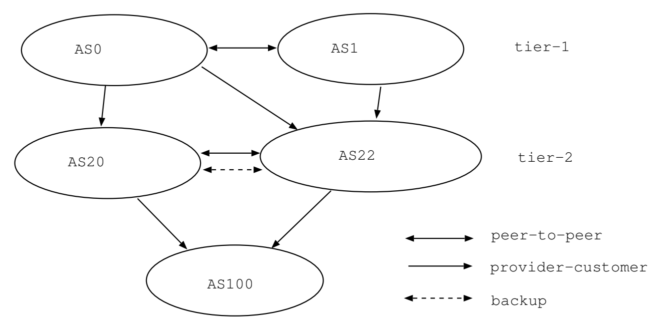
eecs489 COMPUTER NETWORKS

Lecture 18:
Policy Routing in BGP

BGP Policy Routing

Commercial relationship between ASs:

- peering: peers agree to exchange traffic for free
  - AT&T peers with Sprint
- customer-provider: customer pays provider for access
  - UM is a customer of Merit
  - Merit is a customer of AT&T, NTT, Internet2, NLR
- backup

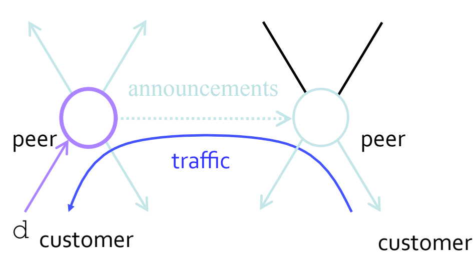


Peering Relationship

Peers exchange traffic between customers

- AS exports only customer routes to a peer
- AS exports a peer's routes only to its customers
- often the relationship is settlement-free (i.e., no money exchanged)

Traffic to/from the peer and its customers



[Rexford]

Customer-Provider Relationship

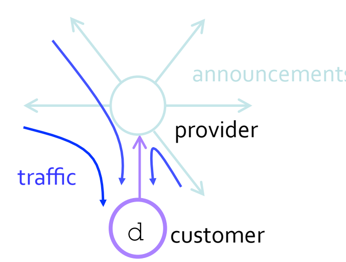
Customer needs to be reachable from everyone

- provider tells all its neighbors how to reach the customer

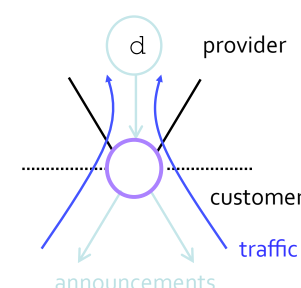
Customer does not want to provide transit service

- customer does not let its providers route through it

Traffic to the customer



Traffic from the customer

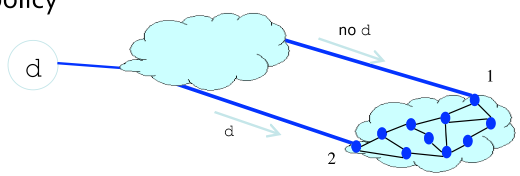


[Rexford]

BGP Policy Routing

An AS's export policy (which routes it will advertise):

- to a customer: all routes
- to a peer or service provider:
  - routes to all its own APs and to its customers' APs,
  - but not to APs learned from other providers or peers
- internal routing of an AS is effected by its neighbors' route export policy

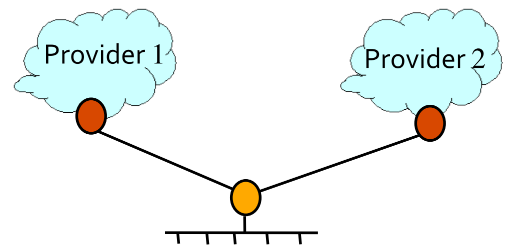


all traffic to d must be routed to egress point 2

Multi-Homing: ≥ 2 Providers

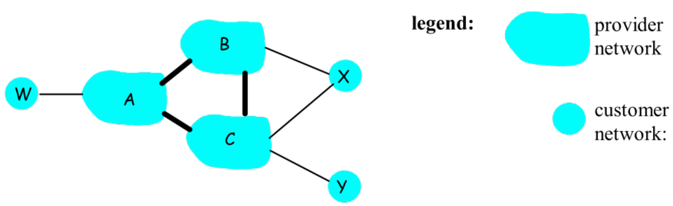
Motivations for multi-homing

- extra reliability, survive single ISP failure
- financial leverage through competition
- better performance by selecting better path
- gaming the 95^{th}-percentile billing model



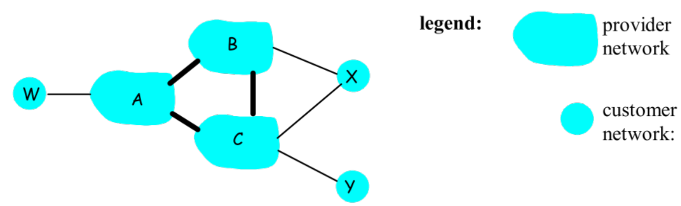
[Rexford]

BGP Routing Policy Example



A,B,C are provider networks
X,W,Y are customers (of provider networks)
X is dual-homed: attached to two networks
X does not want to carry traffic from B to C
.. so X will not advertise to B a route to C

BGP Routing Policy Example



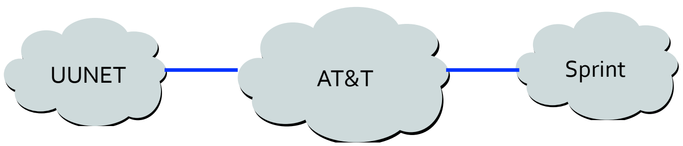
A advertises to B the path AW
B advertises to X the path BAW
B does not advertise to C the path BAW

- 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!

BGP Policy Tools

Export policies: how an AS sets attributes for routes it advertises

- always prepends itself to the AS-PATH
- multiple-exit discriminator (MED): an AS can tell a neighbor its preferred ingress point
- discard some route announcements
  - limit propagation of routing information
  - example: don't announce routes from one peer to another



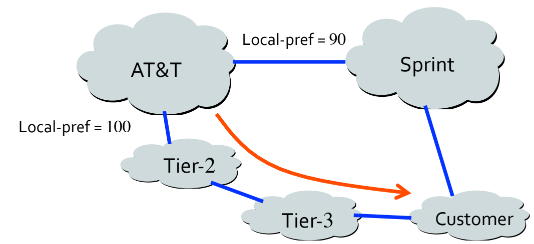
BGP Policy Tools

An AS may learn more than one route to some APs

Each AS applies its own local preference to choose route

Import policies: which of the advertised routes to use

- always checks AS-PATH against routing loop
- local preference: an AS can specify its preferred egress point to reach another AS, in spite of AS path length
- example: prefer customer over peer



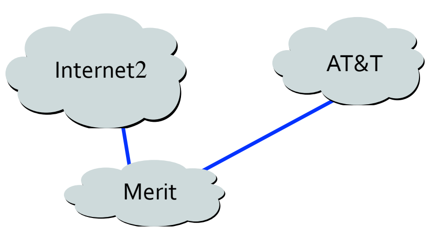
Import Policy: Filtering

Discard some route announcements

- detect configuration mistakes and attacks

Examples: filter customer's advertised APs

- discard route if AP not owned by the customer
- discard route that contains other large ISP in AS path

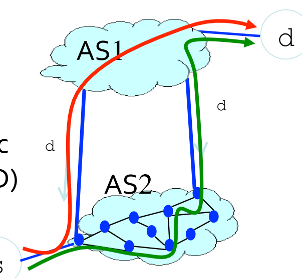


[Rexford]

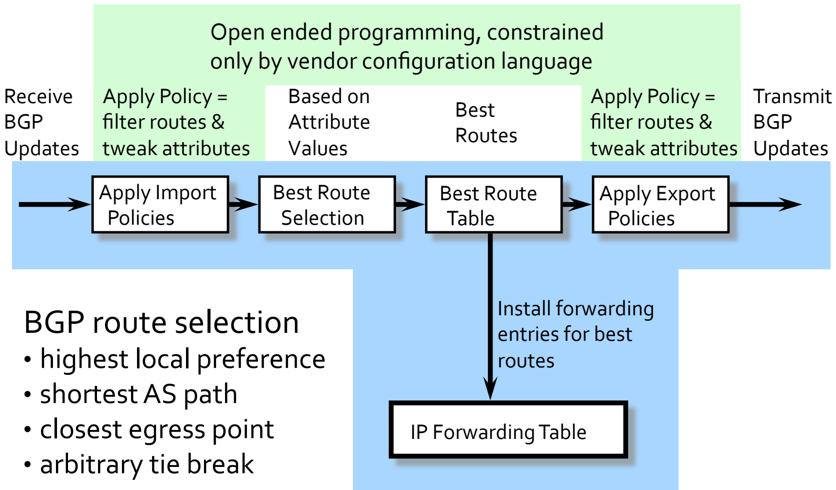
BGP Policy Tools

Export policies: how an AS sets attributes for routes it advertises, to influence the way its neighboring ASs behave

- AS prepending: artificially inflate the AS path length (by repeating the AS number) to convince neighbors to use a different AS
- cold-potato routing: AS1 prefers ingress closest to destination prefix
- hot-potato routing: AS2 prefers egress (NEXT-HOP) closest to traffic source (ignoring the other AS's MED)



BGP Policy: Implementation



BGP route selection
- highest local preference
- shortest AS path
- closest egress point
- arbitrary tie break

[Rexford]

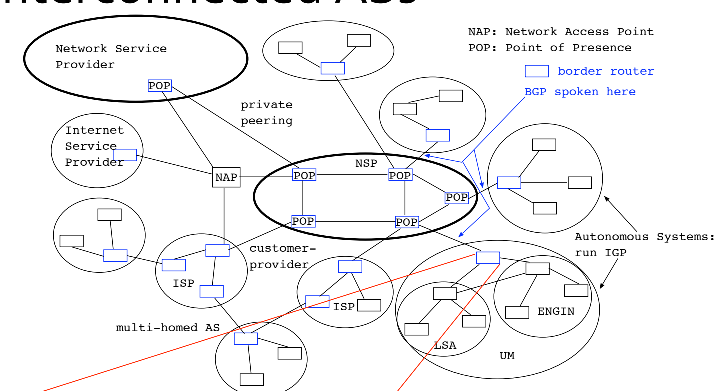
Why Separate Inter-AS Routing ?

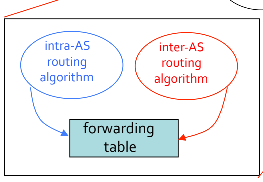
Scale:
hierarchical routing saves table size, reduced update traffic

Policy:
Intra-AS: single admin, so no policy decisions needed
Inter-AS: admin wants control over how its traffic is routed and who routes through its network, i.e., policy driven

Performance:
Intra-AS: can focus on performance
Inter-AS: policy may dominate over performance

Interconnected ASs



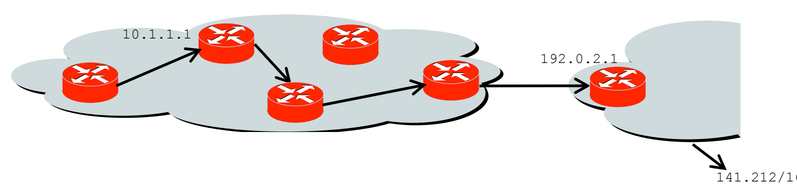


Forwarding table is configured by both intra- and inter-AS routing algorithms
- intra-AS sets entries for internal destinations
- inter-AS & intra-AS set entries for external destinations

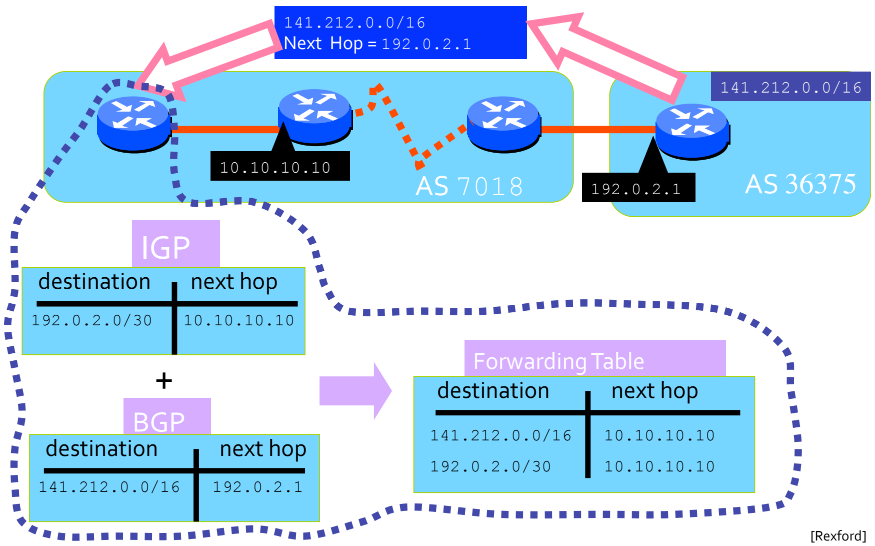
Joining BGP and IGP Information

Border Gateway Protocol (BGP)
- announces reachability to external destinations
- maps a destination prefix to an egress point
  - `141.212.0.0/16` reached via `192.0.2.1`

Interior Gateway Protocol (IGP)
- used to compute paths within the AS
- maps an egress point to an outgoing link
  - `192.0.2.1` reached via `10.1.1.1`



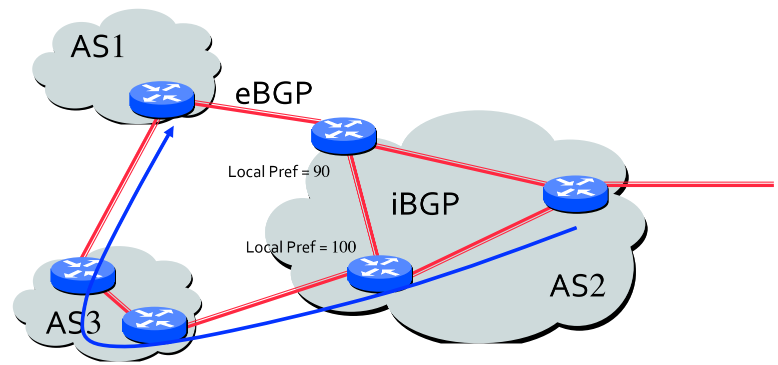
Joining BGP with IGP Information



[Rexford]

An AS is not a Single Node

Multiple routers in an AS
- normally, external routes are not propagated within an AS
- internal BGP (iBGP) allows two border routers of an AS to distribute BGP information within the AS
- sets up iBGP sessions between internal routers



[Rexford]

Causes of BGP Routing Changes

Topology changes
- equipments going up or down
- deployment of new routers or sessions

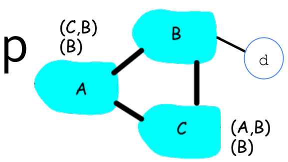
BGP session failures
- due to equipment failures, maintenance, etc.
- or, due to congestion on the physical path

Changes in routing policy
- changes in preferences in the routes
- changes in whether the route is exported

Persistent protocol oscillation
- conflicts between policies of different ASs

[Rexford]

BGP Routing Policy Loop

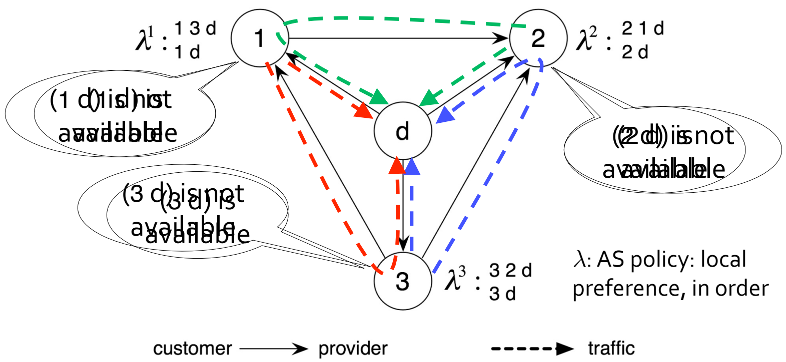


A favors *C* for *B*
C favors *A*

Current approach to prevent BGP policy loops:
- ISPs register their policies with the Internet Routing Registry (IRR)
- policy specified in a standard language
- conflicts can be statically checked
- (policy loop is different from routing loop and is independent of the use of path vector)

Problems:
- policies must be revealed and updated
- static checking for convergence is NP-hard
- possible for BGP not to converge under router/link failure or policy changes

BGP Is Not Guaranteed to Converge



Example known as a "dispute wheel"

Conclusions

BGP is addressing a hard problem

- routing protocol operating at a global scale, with tens of thousands of independent networks, that each has its own policy goals, and all want fast convergence

Key features of BGP

- prefix-based path-vector protocol
- incremental updates (announcements and withdrawals)
- policies applied at import and export of routes
- interaction with the IGP to compute forwarding tables
- internal BGP to distribute information within an AS

[Rexford]

BGP Converges Slowly

Path vector avoids counting-to-infinity

- but ASs must still explore many alternate paths to find the highest-ranked path available

Fortunately, in practice

- most popular destinations have very stable BGP routes
- and most instability lies in a few unpopular destinations

Still, lower BGP convergence delay is a goal

- can be tens of seconds to tens of minutes
- high for important real-time (audio/video) applications
- or even just interactive application, like Web browsing