Stable Internet Routing Without Global Coordination

Jennifer Rexford

Internet and Networking Systems

AT&T Labs - Research; Florham Park, NJ

http://www.research.att.com/~jrex

Joint work with Lixin Gao (Smith College)

http://www.research.att.com/~jrex/papers/sigmetrics00.ps

Outline

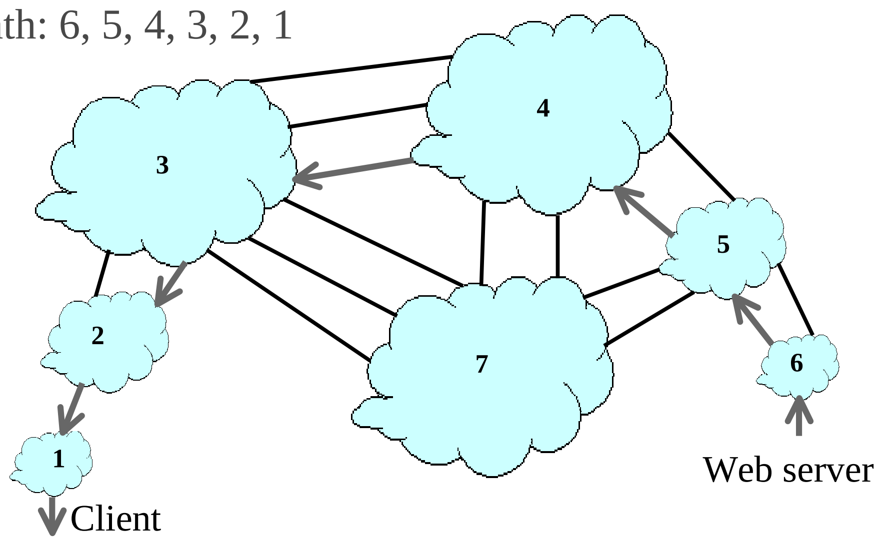
- Background
  - Internet architecture
  - Border Gateway Protocol (BGP)
  - BGP convergence problems
- Ensuring BGP convergence
  - Commercial relationships between ASes
  - Export policies and hierarchical AS graph
  - Guidelines for how each AS ranks routes
- Extensions and ongoing work

Internet Architecture

- Divided into Autonomous Systems
  – Distinct regions of administrative control (~6000-7000)
  – Set of routers and links managed by a single institution
  – Service provider, company, university, …
- Hierarchy of Autonomous Systems
  – Large, tier-1 provider with a nationwide backbone
  – Medium-sized regional provider with smaller backbone
  – Small network run by a single company or university
- Interaction between Autonomous Systems
  – Internal topology is not shared between ASes
  – … but, neighboring ASes interact to coordinate routing

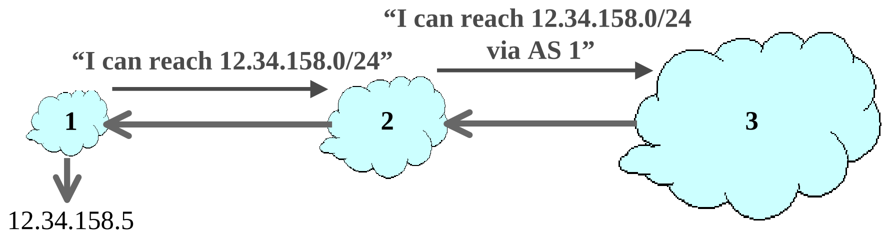
Autonomous Systems (ASes)

Path: 6, 5, 4, 3, 2, 1



Interdomain Routing (Between ASes)

- ASes exchange info about who they can reach
- Local policies for path selection (which to use?)
- Local policies for route propagation (who to tell?)
- Policies configured by the AS's network operator



Border Gateway Protocol

- Exchanging route advertisements
  - Pair of routers speak BGP over a TCP connection
  - Advertise best route for a prefix to neighboring ASes
  - Withdraw a route when it is no longer available
- Processing route advertisements
  - Import policies (manipulate incoming advertisements)
  - Decision process (select best route to given prefix)
  - Export policies (manipulate outgoing advertisement)

Route Advertisement

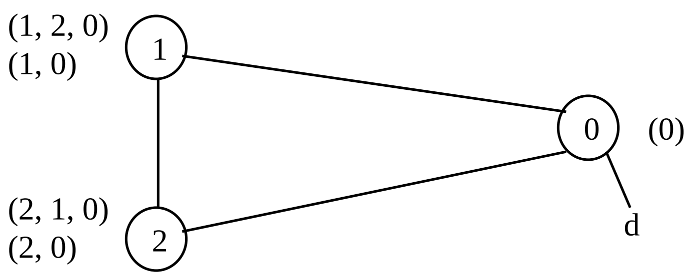
- IP prefix
  - 12.34.158.0/24
- AS path
  - (6, 5, 4, 3, 2)
- Other attributes
  - Multiple exit discriminator
  - Next-hop IP address
  - Community set
  - Local preference

BGP Decision Process

- Highest local preference
  - Set by import policies upon receiving advertisement
- Shortest AS path
  - Included in the route advertisement
- Smallest multiple exit discriminator
  - Included in the route advertisement
- Smallest internal path cost
  - Based on intradomain routing protocol (e.g., OSPF)
- Smallest next-hop router id
  - Final tie-break

Stable and Unstable States

- Stable state:
  AS1: (1, 0)
  AS2: (2, 1, 0)
- Unstable state:
  AS1: (1, 0)
  AS2: (2, 0)



Internet Routing Registry

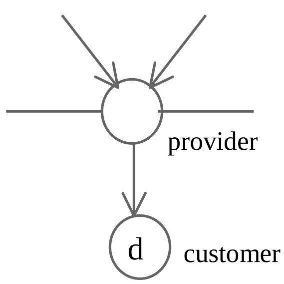
- Strawman solution to convergence problem
  - Create a global Internet routing registry
  - Require each AS to publish its BGP policies
  - Check for conflicting BGP policies
- Problems
  - ASes may be unwilling to reveal BGP policies
  - Keeping the registry up-to-date would be difficult
  - Checking for convergence problems is NP-complete
  - Link/router failure may result in an unstable system

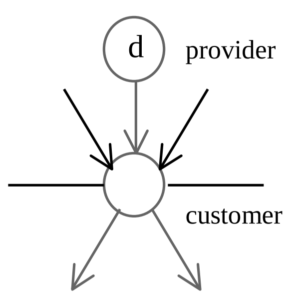
Our Approach

- Goals
  - Allow complex local policies for each AS
  - Do not force ASes to divulge their policies
  - Do not change the BGP protocol & decision process
  - Ensure convergence under failures and policy changes
- Exploit commercial relationships between ASes
  - Export policies based on AS relationships
  - Hierarchical structure of the AS graph
  - Guidelines for import policies based on relationships

Customer-Provider Relationship

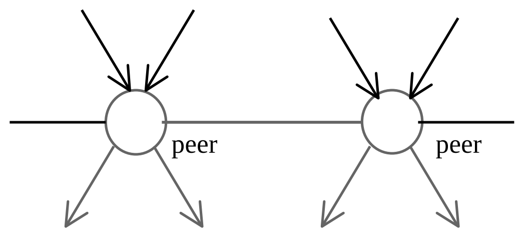
- Customer pays provider for access to the Internet
- AS exports customer's routes to *all* neighbors
- AS exports provider's routes *only* to its customers





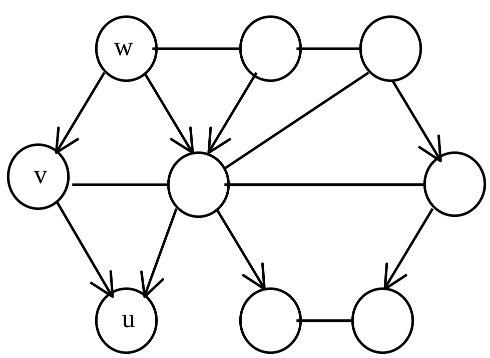
Private Peering

- ◆ Peers exchange traffic between their customers
- ◆ Free of charge (assumption of even traffic load)
- ◆ AS exports a peer's routes *only* to its customers



Hierarchical AS Graph

- ◆ If u is a customer of v, and v is a customer of w, then w is not a customer of u
- ◆ Provider-customer graph is a directed, acyclic graph



Our Import Policy Guidelines

- Classify routes based on next-hop AS
  - Customer routes, peer routes, and provider routes
- Assign local preference based on classification
  - Prefer *customer* routes over peer and provider routes
- Allow *any* local preference within a class
  - Do not impose ranking between customer routes, or between peer and provider routes
- Consistent with traffic engineering practices
  - Customers pay for service, and providers are paid
  - Peer relationship contingent on balanced traffic load

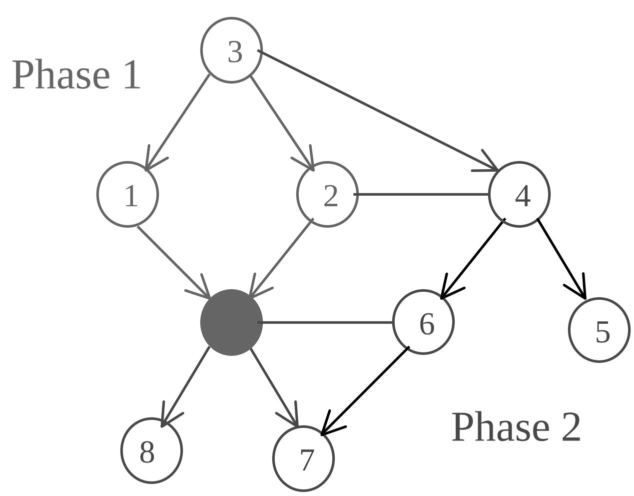
Solving the Convergence Problem

- Assumptions
  - Acyclic customer-provider graph
  - Export policies based on AS relationships
  - Import policies that favor customer routes
- Result
  - Guaranteed convergence of the routing protocol
  - Holds under link/router failures and policy changes
- Sketch of proof
  - Activation sequence that leads to a stable state
  - Any “fair” activation sequence includes this sequence

Proof Details

- Phase 1: selecting customer routes
  - Activate ASes in an order that obeys the partial order defined by the *customer-provider* relationships
  - Decision of one AS cannot cause an earlier AS to change its mind
- Phase 2: selecting peer and provider routes
  - Activate remaining ASes in an order that obeys the partial order of the *provider-customer* relationships
  - Decision of phase-2 AS cannot affect a phase 1 AS
  - Decision of one phase-2 AS cannot cause an earlier phase-2 AS to change its mind

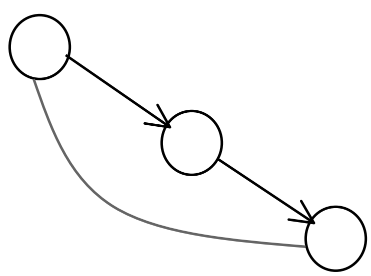
Example Activation Sequence



More Flexible Import Policies

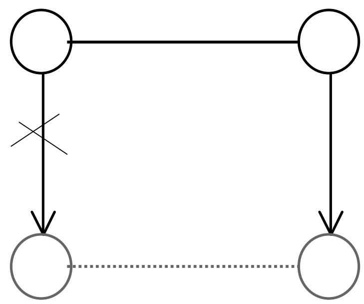
- Allowing more flexibility in assigning local pref
  - Allow peer route with same local pref as customer route
  - Never choose a peer route over a shorter customer route
  - Greater flexibility for balancing traffic load
- Stricter AS graph assumptions
  - Hierarchical provider-customer relationship (as before)
  - No private peering with (direct or indirect) providers (assume that peers are roughly the same size)
- Guaranteed convergence, even under failures

Violation of Stricter AS Graph Assumption



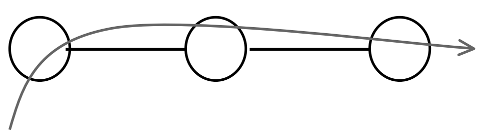
Backup Relationships

- Allow two ASes to provide backup service
- Liberal export policies for backup relationship
- Assign lowest preference to backup routes
- Guaranteed BGP convergence, even under failures



Other Extensions

- Complex AS relationships
  – AS pair with different relationship for different prefixes
  – AS pair with both a backup and a peer relationships
  – AS providing transit service between two peer ASes



- Stability under changing AS relationships
  – Customer-provider to/from peer-peer
  – Customer-provider to/from provider-customer

Conclusions

- Avoiding convergence problems
  - Hierarchical AS relationships
  - Export policies based on relationships
  - Guidelines for import policies based on relationships
- Salient features
  - No global coordination (locally implementable)
  - No changes to BGP protocol or decision process
  - Guaranteed convergence, even under failures
  - Guidelines consistent with financial incentives

Ongoing Work

- Inferring AS relationships
  - AS relationships correspond to export policies
  - Export policies translate into routes in routing tables
  - Routing table entries can be used to infer relationships
  - Paper by Lixin Gao in *Global Internet Symposium 2000*
- Systematic study of backup relationships
  - Multi-homed backup and peer backup relationships
  - Choose between backup routes based on AS relationships and number of backup links in path
  - Joint work with Lixin Gao and Tim Griffin

Ongoing Work (Continued)

- Traffic engineering for ISP networks
  - Interdomain traffic model (ingress to set of egress links)
  - Derivation of traffic demands from edge measurement
  - Model of path selection in intradomain routing (OSPF)
  - Snapshot of current topology and route configuration
  - Environment for experimenting with routing changes
  - Joint work with A. Greenberg, A. Feldmann, C. Lund, N. Reingold, and F. True (SIGCOMM'00, …)