Some Foundational Problems in Interdomain Routing

Nick Feamster, Hari Balakrishnan
M.I.T. Computer Science and Artificial Intelligence Laboratory

Jennifer Rexford
AT&T Labs -- Research
The state of interdomain routing research

- Many years of work, some successes
- Drilling down into esoteric, BGP-specific arcana
- Unfortunately, not a lot of yield in improvements

Maybe problems are intrinsic to interdomain routing?

Questions:
- What makes interdomain routing so difficult?
- Which problems are intrinsic vs. bad design?
Interdomain Routing Has a Lot to Do

- Stability, fast convergence, loop freedom, security...

- Reachability to global destinations
  - Millions of destinations
  - Thousands of independent networks

- Contractual agreements
  - Transit
  - Peering
  - Partial/paid peering or transit

- Traffic engineering
  - Load balance (inbound and outbound)
  - Backup
Interdomain Routing Model

Local Policy:
- Ranking function to select the best route
- Export rules determine route visibility

Intra-AS Route Dissemination:
- Routers within an AS exchange best routes

![Diagram showing interdomain routing with labels for eBGP and iBGP, routers exchanging routes, and export policy determining route visibility.]

Policy determines ranking over possible routes.

Export policy determines availability of routes to other ASes.

Routers inside an AS exchange best routes.
Open Issues Discussed in the Paper

- **Policy-Induced Problems**
  - Policy disputes between ASes (i.e., oscillation)
    - Intrinsic tension between expressiveness and stability
  - Policy enforcement
  - Secure route advertisement

- **Scalability-Induced Problems**
  - Network partitions, forwarding loops, and oscillations
    - Bad design choices for route dissemination and computation
  - Inability to determine cause of update (slow convergence, difficulty of root-cause analysis)
  - Prefix aggregation hides reachability information
Tension between expressiveness and stability

**Requirement:** Stable path assignment

**Problem:** Local rankings of each AS may conflict. [GW01]
Tension between expressiveness and stability

**Requirement:** Stable path assignment

**Problem:** Local rankings of each AS may conflict. [GW01]

For any possible path assignment to 0, some node will always want to switch to a better path.
Tension between expressiveness and stability

**Requirement:** Stable path assignment

**Problem:** Local rankings of each AS may conflict. [GW01]

For any possible path assignment to 0, some node will always want to switch to a better path.
Requirement: Stable path assignment

Problem: Local rankings of each AS may conflict. [GW01]

For any possible path assignment to 0, some node will always want to switch to a better path.
Tension between expressiveness and stability

**Requirement:** Stable path assignment

**Problem:** Local rankings of each AS may conflict. [GW01]

For any possible path assignment to 0, some node will always want to switch to a better path.
Tension between expressiveness and stability

**Requirement:** Stable path assignment

**Problem:** Local rankings of each AS may conflict. [GW01]

For any possible path assignment to 0, some node will always want to switch to a better path.
Tension between expressiveness and stability

What’s known:

• BGP may be unstable.

• Restricting policies can guarantee stability.
  
  ▶ No "dispute wheel" implies stability. [GW01]
  ▶ Restricting rankings and export policies according to "peering" and "customer-provider" relationships implies stability. [GR01]

\[
\begin{array}{ccc}
1 & 3 & 0 \\
10 & & \\
\end{array}
\quad \quad
\quad
\begin{array}{ccc}
3 & 2 & 0 \\
30 & & \\
\end{array}
\quad \quad
\begin{array}{ccc}
0 & & \\
& & \\
\end{array}
\quad \quad
\begin{array}{ccc}
2 & 1 & 0 \\
20 & & \\
\end{array}
\quad \quad
\quad
\begin{array}{ccc}
& & \\
& & \\
\end{array}
\]

Peer

Peer

Peer

AS 2 does not advertise "2 0" to its peer.

What’s the problem?

• Conditions on export restrict relationships.
• It makes no sense to restrict these relationships.
AS 2 does not advertise "2 0" to its peer.

**Restricting available routes won’t work:**

- Stable if 2 does not advertise "2 0" to 3
- But what if 3 pays 2 to see "2 0"?

**Note:** These contracts exist today!

**Assuming no restrictions on available routes:**

- What types of rankings will guarantee stability?
Don’t embed common practices into tomorrow’s protocols!

Research directions:

• Determining restrictions on rankings, given no restrictions on topology or contracts
  ▶ Are these rankings too restrictive for traffic engineering, etc.?

• A separate protocol on a slower timescale for contracts and negotiation?
Scalability-Induced Problem: Partitions

**Requirement:** Many routers and sessions

**Today:** Route reflectors reduce the number of sessions.

**Problem:** Route reflection does not distribute all routes.
Scalability-Induced Problem: Partitions

**Requirement:** Many routers and sessions

**Today:** Route reflectors reduce the number of sessions.

**Problem:** Route reflection does not distribute all routes.

How to guarantee path visibility?
Scalability-Induced Problem: Partitions

**Requirement:** Many routers and sessions

**Today:** Route reflectors reduce the number of sessions.

**Problem:** Route reflection does not distribute all routes.

How to guarantee path visibility?

All top-level routers must be fully meshed.
Loops and Oscillations

What’s known:

- iBGP/IGP interaction causes loops and oscillations.
What’s known:
• iBGP/IGP interaction causes loops and oscillations.
Loops and Oscillations

What’s known:

- iBGP/IGP interaction causes loops and oscillations.
- Restrictions guarantee stable, loop-free paths.
  - Every shortest path in an AS must be a valid "signaling path". [GW01]
  - Route reflectors should be close to clients. [GW02]
What’s the problem?

- Can’t have redundant, topologically diverse RRs
  - Route reflectors must be placed based on IGP topology.
  
This artifact results from bad design.

Research directions:

- Weaker conditions for forwarding correctness?
- Why not tunnel from ingress to egress?
Possible Solution: Routing Control Platform

- Compute consistent routes using complete state.
- Control routing protocol interactions.

Correct dissemination and computation of routes.
Moving forward, we can:
- Continue fixing BGP
- Propose an entirely new protocol

In either case, we should try to distinguish intrinsic tradeoffs from design choices.

Intrinsic problems:
- Tension between expressiveness and stability
- Inability to determine the cause of an update
- Aggregation of prefixes vs. control of traffic

Design choices:
- iBGP loops, partitions, and oscillations
- multiple-exit discriminator (MED) attribute
Policy Disputes

Even given stable inputs, BGP may not converge.

For any possible path assignment to 0, some node will always want to switch to a better path.
Policy Disputes

Even given stable inputs, BGP may not converge.

For any possible path assignment to 0, some node will always want to switch to a better path.
• Every relationship is "provider-customer" or "peering" (no cycles)

• **Ranking** rules:
  ▶ Prefer route through a customer over one through a peer.

• **Export** rules:
  ▶ Export all routes to customers.
  ▶ Export only customer routes to peers and providers.

AS 2 does not advertise "2 0" to its peer.
Do these constraints close the book on BGP convergence? 

We think not.

- Customer-provider/peer relationships are global.
- Export policies are contractual (they involve money).

AS 3 might pay AS 2 to advertise "2 0".
Scalability Makes Interdomain Routing Difficult

**Requirement:** Many routers and sessions
**Problem:** Route reflection does not distribute all routes.
**Result:** Partitions, oscillations, and forwarding loops.

**Requirement:** Many destinations.
**Problem:** Aggregation hides information about reachability to destinations.
**Result:** Coarse information about end-host reachability.

**Requirement:** Many ASes.
**Problem:** AS path hides router-level path information.
**Result:** Slow convergence, inability to pinpoint failures.