Practical Searches for Stability in iBGP

Ashley Flavel  Matthew Roughan  Nigel Bean  Aman Shaikh
Routing Protocol

• A distributed mechanism to determine the best route to a destination.

• Routers propagate their own best route to their neighbors.
  – Best route may not be the shortest-path

• Non-shortest path routing protocols can oscillate!
  – Routers persistently altering their decision in response to one-another.
  – Nemesis to stability
    • Waldo’s nemesis is Odlaw
What is Routing Oscillation?

Preference to destination
What is Routing Oscillation?
What is Routing Oscillation?
What is Routing Oscillation?

Preference to destination

Preference to destination

Preference to destination

Preference to destination
What is Routing Oscillation?

Preference to destination
What is Routing Oscillation?
What is Routing Oscillation?
What is Routing Oscillation?

Preference to destination
What is Routing Oscillation?
What is Routing Oscillation?
Motivation

• Why is oscillation bad?
  – Network reliability is impossible
  – Network predictability is impossible
  – Debugging a network is impossible
  – Degrades router performance

• Routing oscillation should never occur!
  – Full control over network
Motivation

• Oscillation does occur in practice!
  – How do we detect oscillations?

• What are the causes of oscillation inside an AS?
  – MED oscillation
    • Filtering or resetting this attribute prevents this form of oscillation
  – iBGP topology oscillation
    • Caused by the interaction between iBGP and the IGP.
iBGP Topology Oscillation

- iBGP and IGP run on different topologies!
  - iBGP is overlaid on physical topology
  - iBGP used to determine route to external destination
  - IGP used to determine route to internal destination
  - They interact
    - Hot-potato routing
    - Lead to oscillation
BGP Decision Process

- Each router’s decision is independent and based on its set of available routes
- Set of available routes depends on iBGP topology

1. Highest Local-Preference
2. Shortest AS Path
3. Best Origin
4. Lowest MED
5. Shortest IGP distance to egress router (hot-potato)
6. Tie-break
Our Approach

• We create an abstract model of the interaction between iBGP and the IGP.

• Prove oscillatory properties of the abstract model
  – Localize where oscillation can occur.

• Where a configuration is not oscillatory we mathematically prove this is the case.
iBGP Route Reflection

- Used in preference to full-mesh of iBGP sessions in large networks.
  - Due to scalability concerns
- Two classes of routers
  - Clients
  - Route-reflectors
- Multi-level hierarchy possible
  - We focus our attention to two levels
iBGP Route Reflection

- Clients propagate *externally learned* routes to parents
iBGP Route Reflection

- Clients propagate *externally learned* routes to parents
iBGP Route Reflection

- Clients propagate **externally learned** routes to parents

![Diagram of iBGP Route Reflection]

- Route-reflector (RR)
- Client
- Client-parent
- iBGP session
- RR-RR
- iBGP session
iBGP Route Reflection

- Clients propagate externally learned routes to parents
iBGP Route Reflection

- Route-reflectors propagate client routes to all other iBGP neighbors.
iBGP Route Reflection

- Route-reflectors propagate client routes to all other iBGP neighbors.
iBGP Route Reflection

- Route-reflectors propagate client routes to all other iBGP neighbors.
iBGP Route Reflection

- Route-reflectors reflect routes learned from other route-reflectors to clients
iBGP Route Reflection

- Route-reflectors reflect routes learned from other route-reflectors to clients
Reliance Graph

• What happens when routers’ decisions oscillate?
  – Multiple routers persistently alter their decision in response to the decision of others

• If a router can learn of its best route from another router, then it is reliant on it

• The reliance graph captures all possible reliances.
Reliance Graph Basics

• A reliance graph is on a per egress instance basis
  – Set of routers which have a direct egress
  – Equally attractive over AS-wide decision steps

• A directed edge in the reliance graph exists if a router’s decision is reliant on another.

• Oscillation requires a cycle in the reliance graph
Reliance Graph

- All iBGP sessions can be bi-directional edges in reliance graph.
- We can prune most of these edges!
  - iBGP pruning
    - Some neighbors will never propagate a route
    - Based on route-reflection
  - IGP pruning
    - Some neighbors will never propagate a best route
    - Based on IGP distances
iBGP Pruning

- Routers in the egress instance
  - Will select its own route
  - Not reliant on any other router
iBGP Pruning

- Routers in the egress instance
  - Will select its own route
  - Not reliant on any other router

- Clients can only select a route learned from their parent.
  - Do not propagate any information to their parents
iBGP Pruning

• A route-reflector can only be reliant on another route-reflector if that route-reflector has a client in the egress instance.
iBGP Pruning

- A route-reflector can only be reliant on another route-reflector if that route-reflector has a client in the egress instance.
- We can prune further using IGP distances
- Route-reflectors only reliance on their ‘best’ client
Co-reliance groups

- We partition the reliance graph into strongly connected components
  - Termed co-reliance groups
- Oscillation can only occur within a non-singleton co-reliance group.
- The only possible location for a non-singleton co-reliance group
  - route-reflectors with clients in the egress instance

![Diagram showing co-reliance groups and their connections](image-url)
Will a co-reliance group oscillate?

• Maybe.
  – Not everything in a striped sweater is Waldo
• Analyze each co-reliance group independently
  – Oscillation can only occur within a co-reliance group.
Co-reliance group algebra

- A co-reliance group only contains route-reflectors with clients in the egress instance.
- Each route-reflector has egresses which it can learn via:
  - a client (direct route), and
  - another route-reflector (indirect)
    - Irrelevant which indirect route is chosen.
- Each router in co-reliance group only has two route choices!
  - Direct (d) or Indirect (i)
- **Labels:** direct (d), indirect (i)
- **Preference Relationship:** $i > d$
- **Outbound Arc Labels:** i if node is d, nothing if node is i
Example

- **Labels**: direct (d), indirect (i)
- **Preference Relationship**: i > d
- **Outbound Arc Labels**: i if node is d, nothing if node is i
Example

- **Labels:** direct (d), indirect (i)
- **Preference Relationship:** i > d
- **Outbound Arc Labels:** i if node is d, nothing if node is i
Example

• **Labels:** direct (d), indirect (i)
• **Preference Relationship:** i > d
• **Outbound Arc Labels:** i if node is d, nothing if node is i
Example

- **Labels**: direct (d), indirect (i)
- **Preference Relationship**: i > d
- **Outbound Arc Labels**: i if node is d, nothing if node is i
- **One stable solution**
Example

- Alternative stable solution
- If all possible message orderings result in the group settling to a solution - the system is stable.

```
Example

- Alternative stable solution
- If all possible message orderings result in the group settling to a solution - the system is stable.

```

```
- Alternative stable solution
- If all possible message orderings result in the group settling to a solution - the system is stable.

```

```
- Alternative stable solution
- If all possible message orderings result in the group settling to a solution - the system is stable.

```
Three Node Co-reliance Group

- Will this co-reliance group oscillate?
State Machine

Diagram of a state machine with states labeled as ddd, idi, ddi, iiid, did, ddi, and iii.
How many reliance graphs?

• So far we have focused on a single egress instance
•Analyzing all egress instances results in a combinatorial explosion.
•Some border routers never used in combination.
  – Filters on border routers
•We can further scope the problem
  – Prioritize the checking of current egress instances
Practical Analysis

• Based on topology and routing state of a Tier 2 AS
  – Approx 500 routers.
  – 954 egress instances found (maximum number of egress routers = 17).
  – Power set of these egresses raised the number of egress instances to 204,621

• Results
  – Stable
  – Under 15 minutes to run

• Further optimizations possible
  – Can run in parallel
  – Can reduce co-reliance groups to equivalent forms
  – Can keep a library of co-reliance groups
What does this mean?

• Proof a large practical iBGP configuration’s stability.
  – Not a simulation.

• Can determine the oscillatory properties of
  – configuration changes prior to their implementation
  – the current network state
  – the network as it evolves (e.g. due to failures)

• Guarantee the stability of a configuration.

• Pinpoint the routers responsible for oscillatory modes.
Conclusion and Future Work

• We can determine the oscillatory properties of a network configuration.
  – Fast
  – Scalable for realistic networks
  – It is a proof of stability or you find where the oscillation is.

• Future Work
  – How do you fix oscillation?
Where’s Waldo?