Understanding Operational Routing
(part II)

Geoffrey Xie
Naval Postgraduate School

Joint work with
Franck Le* (formerly CMU student, now at IBM)
Hui Zhang (CMU)

July 6, 2011
Route Aggregation

**Unallocated Child Prefix:** e.g., 10.1.33.0/24

<table>
<thead>
<tr>
<th>Destination</th>
<th>Next-hop</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.1.1.0/24</td>
<td>19.1.1.2</td>
</tr>
<tr>
<td>10.1.2.0/24</td>
<td>19.1.1.2</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>10.1.32.0/24</td>
<td>19.1.1.2</td>
</tr>
</tbody>
</table>

**Aggregate Route**

<table>
<thead>
<tr>
<th>Destination</th>
<th>Next-hop</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.1.0.0/16</td>
<td>19.1.1.2</td>
</tr>
</tbody>
</table>

**Route Aggregation** reduces size of routing tables and contains propagation of route flaps.
Challenges for Understanding RA

• White-box
  – ISPs do not disclose details of incidents nor configurations
  – Difficult to assess prevalence of anomalies and to what extent RA contributes to them

• Black-box
  – Can expose anomalies but difficult to identify root causes
  – Studies discovered loops, and conjectured RA to be responsible for them
Proposed Approach

• Study RA behavior as implemented in today’s routers
  – No precise specification of RA behaviors
  – Many questions left unanswered

• Propose a model to
  – analyze and reason about RA,
  – identify routing abnormalities that can be root-caused to route aggregation, and
  – design and prove more formal and comprehensive RA guidelines to ensure safe routing
Main Results on RA

• A set of experiments to characterize RA behaviors
  – Behaviors vary significantly across protocols and vendors

• A canonical router model with two new router primitives
  – New model captures diversity of RA behaviors

• A comprehensive analysis of RA
  – Expose four new types of routing anomaly (e.g., route loss)
  – Identify the root causes for each anomaly
  – Explain shortcoming of current vendor guidelines
  – Present sufficient conditions for the RA to guarantee routing safety

• A clean-slate proposal for RA
Part 2 Outline

• Background
• Experiments
• Model
• Anomalies & Root Causes
• Analysis of Existing Solutions
• New Safety Conditions
• Clean-Slate Approach
Persistent Forwarding Loop

<table>
<thead>
<tr>
<th>Destination</th>
<th>Next-hop</th>
</tr>
</thead>
<tbody>
<tr>
<td>128.2.0.0/16</td>
<td>19.2.2.2</td>
</tr>
<tr>
<td>128.2.1.0/24</td>
<td>10.1.2.1</td>
</tr>
<tr>
<td>128.2.2.0/24</td>
<td>10.1.2.1</td>
</tr>
<tr>
<td>128.2.5.0/24</td>
<td>10.1.2.1</td>
</tr>
</tbody>
</table>

Internet Service Provider
19.2.2.1
128.2.0.0/16
0.0.0.0/0

Enterprise Network
128.2.0.0/16
19.2.2.2
128.2.1.0/24
128.2.2.0/24
128.2.5.0/24
Persistent Forwarding Loop

<table>
<thead>
<tr>
<th>Destination</th>
<th>Next-hop</th>
</tr>
</thead>
<tbody>
<tr>
<td>128.2.0.0/16</td>
<td>19.2.2.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Destination</th>
<th>Next-hop</th>
</tr>
</thead>
<tbody>
<tr>
<td>128.2.1.0/24</td>
<td>10.1.2.1</td>
</tr>
<tr>
<td>128.2.2.0/24</td>
<td>10.1.2.1</td>
</tr>
<tr>
<td>128.2.5.0/24</td>
<td>10.1.2.1</td>
</tr>
</tbody>
</table>

Enterprise Network
128.2.0.0/16
Blackhole

10.1.1.10

10.1.0.0/23

10.0.0.0/23

10.0.0.0/23

10.1.0.0/24

X

10.1.1.0/24
Blackhole

10.1.0.0/24
10.1.1.0/24
10.1.0.0/23
10.1.0.0/23
10.1.1.10
10.1.0.0/24
10.1.1.0/24
Outline

• Background
• Experiments
• Model
• Anomalies & Root Causes
• Analysis of Existing Solutions
• New Safety Conditions
• Clean-Slate Approach
Experiment Setup

Input Child routes  Aggregation Router  Output Aggregate route
## RA Behaviors

<table>
<thead>
<tr>
<th>Vendor</th>
<th>Routing Protocol</th>
<th>Modes of RA</th>
<th>Metric</th>
<th>Sink Route</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Auto</td>
<td>Manual</td>
<td></td>
</tr>
<tr>
<td>Cisco</td>
<td>EIGRP</td>
<td>Y</td>
<td>Interface</td>
<td>min()</td>
</tr>
<tr>
<td>Cisco</td>
<td>RIP</td>
<td>Y</td>
<td>Interface</td>
<td>min()</td>
</tr>
<tr>
<td>Cisco</td>
<td>OSPF</td>
<td>N</td>
<td>Area</td>
<td>min(), max()</td>
</tr>
<tr>
<td>Cisco</td>
<td>BGP</td>
<td>Y</td>
<td>AS</td>
<td>Origin/AS-SET</td>
</tr>
<tr>
<td>Juniper</td>
<td>*</td>
<td>N</td>
<td>Creation per router; advmt per instance</td>
<td>Customizable</td>
</tr>
</tbody>
</table>
Outline

• Background
• Experiments
• Model
• Anomalies & Root Causes
• Analysis of Existing Solutions
• New Safety Conditions
• Clean-Slate Approach
Model

• Motivation
  – Ad-hoc nature of current design & implementation
  – Unified model to reason about RA and identify root-cause of anomalies

• Insight
  – Diversity of behaviors can be captured by two primitives
  – Outcomes depend on where those two primitives are applied (RIB, FIB, etc.)
RA Primitives

• **add-sink()**
  - Creates* sink route
  - Assigns specific AD value
  * Sink route may not get installed in FIB

• **adv-aggr()**
  - Cisco only; handles advertisement of aggregate routes
  - Performed for each protocol instance and per routing process

**Location where the primitive examines for presence of child routes differs depending on implementation**

**JUNOS implementations rely on export policies to announce aggregate routes**
Canonical Router Model (Cisco IOS)
Canonical Router Model (Juniper JUNOS)

Model allows us to infer whether a route is installed in FIB and how it may be advertised by a router.
Unexpected Route Loss

Route Protocol
Route Selection Alg.

Local RIB

EIGRP RIBin

Routing Protocol
Route Selection

Table:

<table>
<thead>
<tr>
<th>Destination</th>
<th>route (AD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>192.168.1.0/24</td>
<td>r1 (90)</td>
</tr>
<tr>
<td>192.168.0.0/16</td>
<td>r2 (90) → r3 (5)</td>
</tr>
<tr>
<td>192.0.0.0/8</td>
<td>r4 (5)</td>
</tr>
</tbody>
</table>

Diagram:

- r1: 192.168.1.0/24
- r2: 192.168.0.0/16
- r3: 192.168.0.0/16
- r4: 192.0.0.0/8

Add-Sink
Adv-Aggr

**Unexpected Route Loss**

- **Local RIB**
- **EIGRP RIBin**
- **Routing Protocol**
- **Route Selection**
- **FIB**
- **Add-sink()**
- **EIGRP RIBout**
- **Adv-aggr()**

- **r₁**: 192.168.1.0/24
- **r₂**: 192.168.0.0/16
- **r₃**: 192.168.0.0/16
- **r₄**: 192.0.0.0/8

- **l₁**: rr
- **l₂**: rr
- **l₃**: rr
- **l₄**: rr

- **No RA on l₄**
- **Two routes r₁ and r₂ received to different prefixes**
- **One would expect both routes to be advertised on l₄**
Outline

• Background
• Experiments
• Model
• Anomalies & Root Causes
• Analysis of Existing Solutions
• New Safety Conditions
• Clean-Slate Approach
Anomalies

• In-depth analysis based on model
  1. Identify root causes of known anomalies
  2. Disclose four new anomalies

• Key findings
  1. Sink route creation does not guarantee installation in FIB
  2. Sink routes do not mean anomaly-free routing (e.g., route loss)
Root Causes of Known Anomalies

• Strict Monotonicity: A route’s preference should strictly decrease when propagated

• Forwarding loops for unallocated child prefixes:
  – Lack of sink routes
  – Router advertises aggregate routes with “holes”
  – Propagation of non-existent routes, which should be least preferred
    ➔ Violation of SM

• Blackholes for some child prefixes:
  – Interactions between multiple routers
  – More difficult to identify
New Anomalies

• Previously known:
  – Forwarding loops (unallocated prefixes)
  – Blackholes

• New discoveries:
  – Route Loss
  – Forwarding loops (allocated prefixes)
  – Route Oscillations
  – Perpetual Count-to-Infinity
Perpetual Count to Infinity

Advertisement cycle continues until metric reaches max value; route is discarded, sink route is installed again, ...

\[ \rightarrow \text{Perpetual count to infinity} \]
Outline

• Background
• Experiments
• Model
• Anomalies & Root Causes
• Analysis of Existing Solutions
• New Safety Conditions
• Clean-Slate Approach
Inadequacy of Vendor Guidelines

• Installation of sink routes

• Topology specific guidelines
Vendor Guidelines: Sink Routes

“A router that generates an aggregate route for multiple, more-specific routes must discard packets that match the aggregate route, but not any of the more-specific routes. In other words, the “next hop” for the aggregate route should be the null destination. This is necessary to prevent forwarding loops when some addresses covered by the aggregate are not reachable.”

• Correctness criterion, but how to meet it?

• Creation of sink route does not guarantee installation in FIB

• Also, Installation of sink route in FIB can create blackholes
Vendor Guidelines: Topology Specific

**Approach 1**
- Add links

**Approach 2**
- Only choke point can act as aggregation router
Difficulty of Anomaly Detection

Theorem:

The problem of determining whether the collection of route aggregation configurations in a network is vulnerable to persistent forwarding loops is NP-hard.
Outline

• Background
• Experiments
• Model
• Anomalies & Root Causes
• **Analysis of Existing Solutions**
• New Safety Conditions
• Clean-Slate Approach
New Safety Conditions

• Desired Properties
  1. Convergence
  2. Loop-free packet forwarding paths
  3. No blackholes
  4. No unexpected route losses

• Challenges
  – network-wide coordination
  – stringent requirement on router
Sufficient Conditions

1. Convergence and loop-free paths
   - *For an aggregate prefix, a sink route is added and always possesses the lowest unique AD value*

2. Blackhole Prevention
   - A router advertises an aggregate route only if a set of routes that fully covers the address space of the aggregate route is present in the FIB

3. Prevention of Route Loss
   - *add-sink() and adv-aggr() should not create a new route in the presence of another route in the router FIB advertising the same prefix as the aggregate.*
Outline

• Background
• Experiments
• Model
• Anomalies & Root Causes
• Analysis of Existing Solutions
• New Safety Conditions
• Clean-Slate Approach
Motivation

• RA reduces FIB sizes, by

  – relying on the hierarchical IP addressing

  – allowing routers to treat multiple destination networks collectively as a single logical destination
Semantics of Route

• A route implies all destinations included in the advertised prefix to be reachable

• Route aggregation violates this assumption

→ Forwarding loops, blackholes, etc.
Proposed Solution

• Novel concept of *negative route*

• It allows a router to indicate subsets of prefixes for which it does not have a route

• Difference from existing invalid routes:

  Negative routes are stored in RIB and FIB. It eliminates all anomalies due to unreachable sub-prefixes in aggregate route.
RA Conclusions

• RA is essential for Internet routing scalability

• Current design is prone to anomalies
  – Forwarding loops, perpetual count to infinity, unexpected route loss, blackholes, etc.

• We proposed the first analytical model to reason about RA
  – Needs more attention from community
A Final Remark

Theory / model → Design → Practice

guide → guide
More details of the RA analysis are in


Thank you!
Related Work


[WSR 05] R. White, D. Slice, and A. Retana.


Backup Slides
add-sink() primitive

Primalve 1 add-sink($E$, $A$)

**Input:** (1) $E$ - routes from the router FIB or routes from a protocol specific Route Information Base (RIB); (2) $A$ - all aggregate routes configured on the router

1. $S = \{\}$
2. Remove existing sink routes from $E$
3. for all $a \in A$ do
4. \hspace{1em} if there exists a child route of $a$ in $E$ then
5. \hspace{2em} Set AD value and Null next hop for $a$;
6. \hspace{2em} Add $a$ to $S$;
7. \hspace{1em} end if
8. end for
9. Present $S$ to the route selection procedure
**Primitive 2** \texttt{adv-aggr}(E, A)

**Input:**
1. \(E\) - routes from a protocol specific RIBOut;
2. \(A\) - all aggregate routes configured on given interface

1. Remove sink routes from \(E\)
2. \texttt{for all} \(a \in A\) \texttt{do}
3. \texttt{if} there exists a child route of \(a\) in \(E\) \texttt{then}
4. Remove all child routes of \(a\) in \(E\);
5. \(a.m = \text{metric}(a, E)\);
6. Add \(a\) to \(E\);
7. \texttt{end if}
8. \texttt{end for}
9. Advertise \(E\) on the interface