CleanBGP: Verifying the Consistency of BGP Data

Ashley Flavel
Olaf Maennel
Belinda Chiera
Matthew Roughan
Nigel Bean
Why is BGP Data Important?
Why is BGP Data Important?

- Raw BGP data
- Updates
- Router Decisions
- Stability or Instability of Network
- Future Router Requirements
- Anomaly Detection
- What-if analysis
- Traffic Matrix Estimation
- Oscillation Detection
- Bogon Prefixes
- Link Failures
- Predicting impact of policy changes
- Predicting impact of topology changes
- Link Utilization
Why is BGP Data Important?

- Raw BGP data
  - Update Rate Statistics
  - Router Decisions
  - Network Topology

- BGP updates
  - Oscillation Detection
  - Bogon Prefixes
  - Link Failures
  - Predicting impact of policy changes
  - Predicting impact of topology changes
  - Link Utilization
  - Future Network Design

- Route Monitor
  - Periodic table dump
Why is **Accurate** BGP Data Important?
How Can We Check the Accuracy?
CleanBGP
Data Sources

• Tables
  – Current route of monitored router to all possible destinations (prefixes)
  – Periodically written to disk
    • RIPE (8 hours)
    • RouteViews (2 hours)

• Updates
  – BGP is incremental protocol
    • No periodic retransmission of routes
    • Generally small fraction of routes in table updated in a short interval
      – Except when a BGP session is first being established
The Border Gateway Protocol
BGP Session Failures

[Diagram of a network with nodes labeled 0 to 6, showing BGP session failures with numbers in square brackets at each node.]

TRC Mathematical Modelling
BGP Session Failures

---

0
1
2
3
4
5
6

[1 2 0]
[0]
[2 0]
[3 0]
[4 1 0]
[5 3 0]
[6 4 1 0]
BGP Session Failures

Diagram showing a network with nodes labeled 0, 1, 2, 3, 4, 5, 6. Nodes 0 and 3 are highlighted with red circles. Connections between nodes are labeled with numbers: [1 2 0], [0], [2 0], [3 0], [4 2 0], [5 3 0], [6 4 1 0], and [6 4 1 0].
BGP Session Failures
BGP Session Failures
Data Consistency

- The BGP table is the construction of the last update for each prefix.
  - A table at t1 plus updates in the interval [t1,t2] is equivalent to the table at t2.
- In the recorded data this is not always the case!
Measurement Artifact 1
Measurement Artifact 1
Measurement Artifact 1

- Monitoring Session Reset
  - During downtime, no updates recorded
  - After session reset all routes currently in the table are re-advertised
Measurement Artifact 1

- **Monitoring Session Reset**
  - During downtime, no updates recorded
  - After session reset all routes currently in the table are re-advertised
Measurement Artifact 2
Measurement Artifact 2

- Update Re-ordering
  - ‘Almost simultaneous’ updates either
    - recorded in incorrect order; or
Measurement Artifact 2

• Update Re-ordering
  – ‘Almost simultaneous’ updates either
    • recorded in incorrect order; or
    • applied to table in the incorrect order
      – Serious consequences when software router used as operational router
      – Invalid state!

![Diagram showing time t1, t2, and recorded data vs. actual activity]
Other Measurement Artifacts

- **Missing Updates**
  - Hardware issues prevent all updates being written to data warehouse

- **Incomplete Table**
  - The table is not written completely to data warehouse
Evidence of Measurement Artifact

• **What do we see in the data?**
  – Constructed table differences
  – Almost simultaneous updates
  – No routing activity for extended period
  – Burst of routing announcements

• **State Information**
  – Some data sources have session UP/DOWN meta-data.
  – Oldest prefix in table
    • During a session re-establishment ALL prefixes are re-announced.
    • When a session reset definitely did not occur
    • When a session reset may have occurred

• **Predict the cause of an inconsistency based on evidence**
Detection of Measurement Artifact

- Inconsistency in Constructed and Recorded Table
  - A session reset may not cause an inconsistency!
    - No withdrawals may occur during downtime
    - Still an artifact
      - Re-establishment phase updates

- Sliding window on update timeseries
  - Threshold of duplicates or unique prefixes
  - Downtime
    - Hold-time a good threshold when `keep-alives’ recorded
Localization of Measurement Artifact

- Update timeseries split into bins
- Find group of suspicious bins around detected time
  - Include single ‘normal’ bins
  - Detected time one bin either side of group
  - Captures multiple resets in one interval
- A bin is suspicious if
  - No updates
  - Large number of unique prefixes
  - Large number of duplicates
- Conservative detection/localization provides confidence in data!
Cleaning Data

• Exclusion
  – Exclude the data affected from further analysis
  – Recommended

• Estimation
  – What actually happened?
    • Remove duplicates during measurement artifact interval
    • Place updates where appropriate
      – Table provides some help here
      – Mark the updates which we introduce/remove
What Did We Find?

• Analyzed several RIPE monitors for several months
  – Inconsistent data in about 5% of tables
  – 81% of inconsistencies caused by re-ordered updates!
  – Session resets contributed 10% of inconsistencies
    • Much more frequent detection when no inconsistency
    • State information for validation
  – Almost an hour on Jan 21, 2007 where no updates are recorded
    • Not caused by a session reset
Summary

• Important to validate your data!
• Cross-checking provides an increased level of confidence in data
• Developing a tool based on these results
  – Including automatic threshold setting
Summary

• Important to validate your data!
• Cross-checking provides an increased level of confidence in data
• Developing a tool based on these results
  – Including automatic threshold setting
• I’m looking for a job 😊