Detecting BGP Anomalies

Geoff Huston
APNIC Labs

BGP Anomaly Detection

- The objective of an BGP anomaly detector is to process BGP updates and automatically detect “anomalies” where the prefix or the path does not appear to be aligned to “normal” routing behaviours

- The challenge is to train an automated system to generate a useful model of “normal” and “anomalous” classification of BGP updates

Approaches to Anomaly Detection

I - Rule-based systems

- Have network operators describe their routing policies
- Trigger notification on detected exceptions

Approaches to Anomaly Detection

I - Rule-based systems

- Have network operators describe their routing policies
- Trigger notification on detected exceptions

For example:

AS131072

Originates:

- 192.0.2.0/24
- 2001:DB8::/32

- Next-Hop AS
  - AS4608
- Downstream AS:
  - Nil

Approaches to Anomaly Detection

I - Rule-based systems

- Have network operators describe their routing policies
- Trigger notification on detected exceptions

For example:

AS131072

Originates:

- 192.0.2.0/24
- 2001:DB8::/32

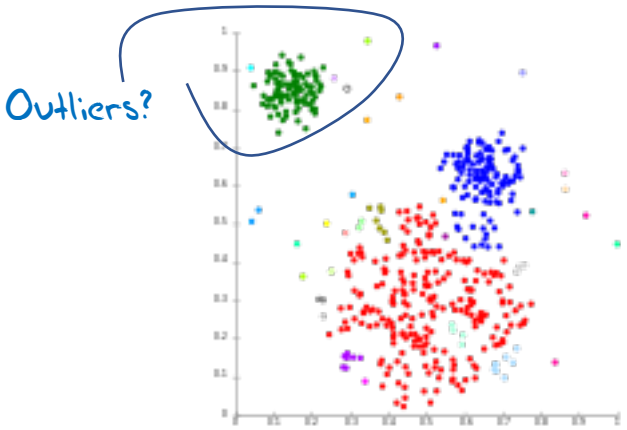
- Next-Hop AS
  - AS4608
- Downstream AS:
  - Nil

```
BGP4MP|1566349266|A|192.0.2.0/24|3556 4608 131072|IGP
BGP4MP|1566349266|A|192.0.2.0/24|3556 4823 4777 131072|IGP
```

Approaches to Anomaly Detection

II – Machine Based Learning

- Feed updates into a parameter generator
- Perform n-dimensional cluster analysis on the data set
- Identify outliers as potential anomalies



Approaches to Anomaly Detection

III – Heuristics

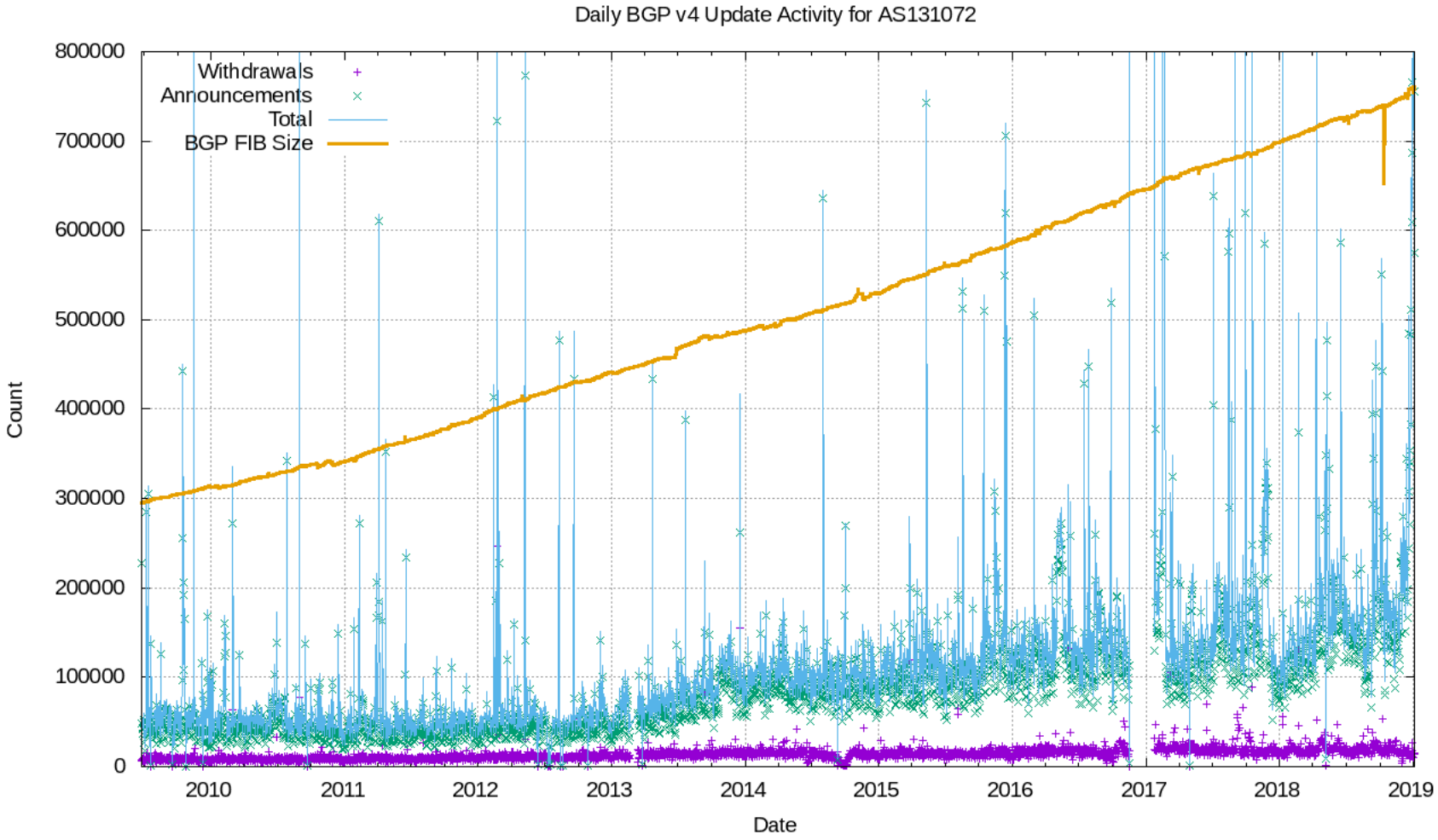
- Feed updates into an analyser
- Generate a n-dimensional ‘score’ for the update
- Use thresholds to pick out out candidate anomalies

- Which is the focus of this project…

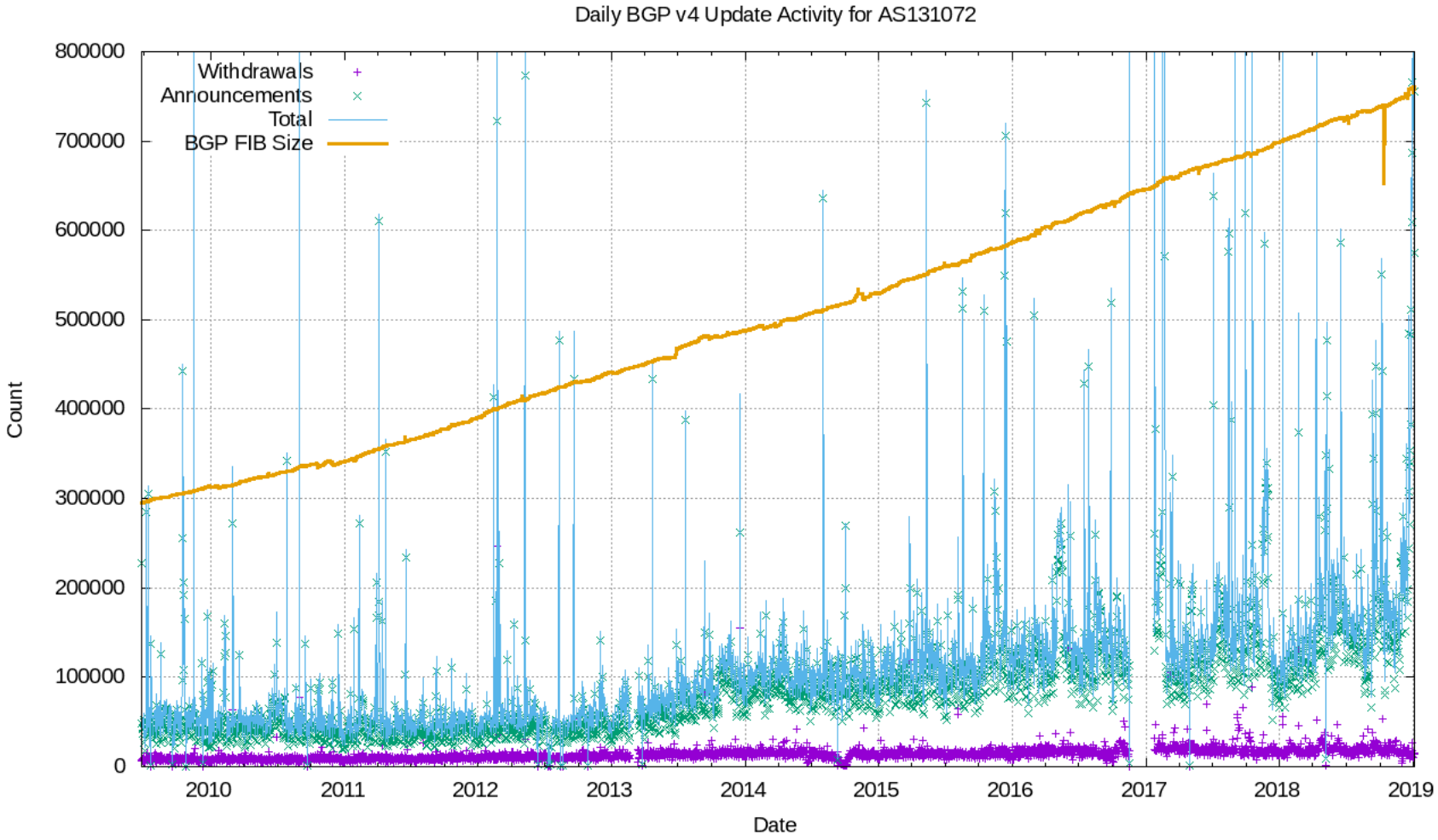
BGP is a Chatty Protocol

- It’s a distance vector protocol
- Which means that it converges through exhaustion via progressive refinement, not through direct computation (as is the case with SPF protocols)
- Which also implies that there are transient states that are not stable
- Which implies that when we look for anomalies in BGP updates there is a huge amount of BGP chaff to work through!

BGP Update Profile



BGP Update Profile



Over 10 years the number of prefixes has more than doubled

Number of prefix updates was stable for 4 years, then has been rising slowly

Number of prefix withdrawals has been steady

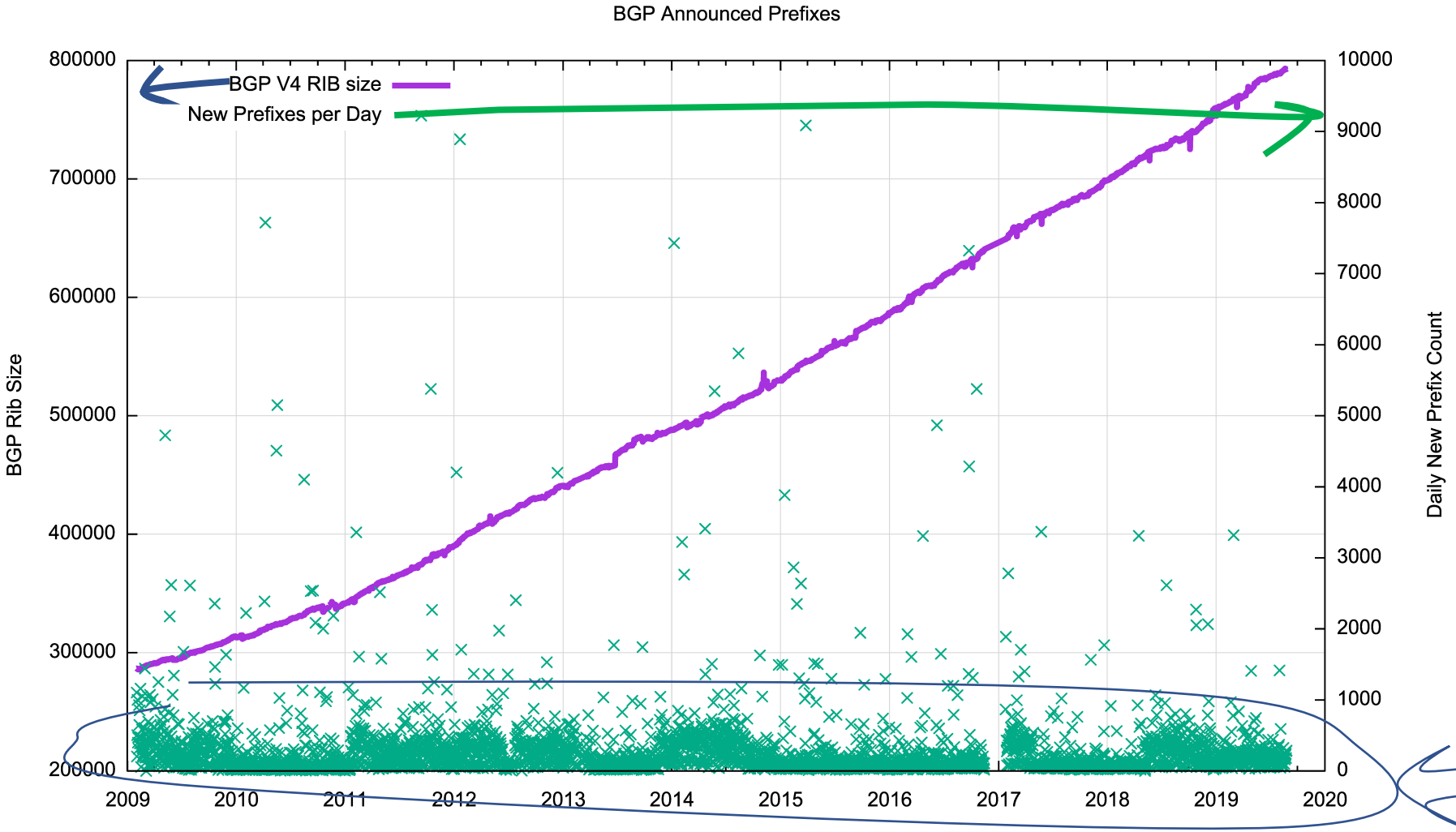
BGP Update Profile

- A conventional default-free IPv4 eBGP session at the edge of the Internet will process some 150,000 – 200,000 prefix announcements per day
- And some 10,000 prefix withdrawals per day
- This is a relatively stable profile for eBGP update activity

BGP Update Profile

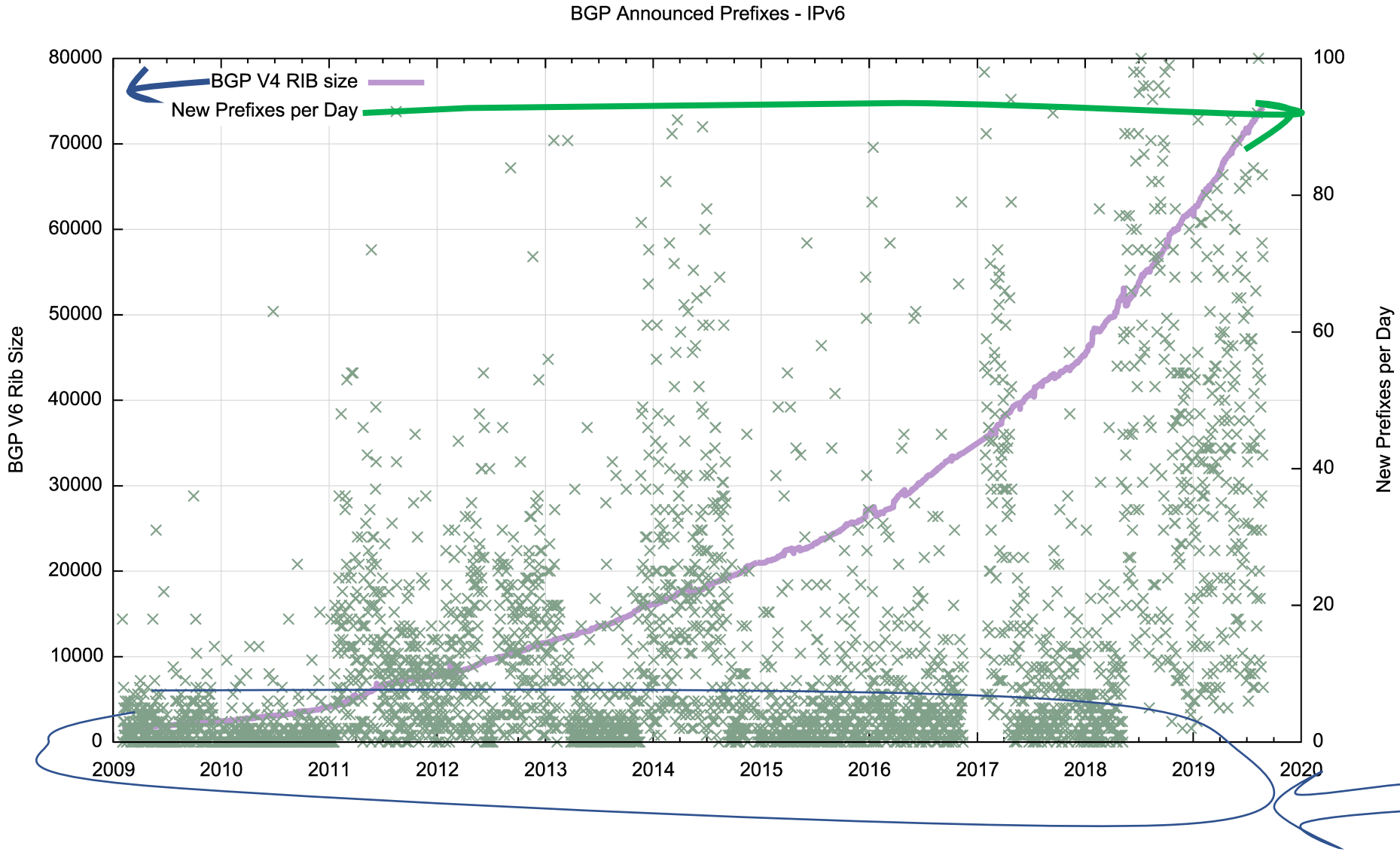
- It is useful to understand how much of this protocol traffic is a by product of the operation of the protocol, how much is realignment of the network topology and how much is “new” reachability information
- Let’s count the daily number of prefixes in the eBGP RIB and the daily count of previously unseen prefixes

BGP Prefix Updates - IPv4



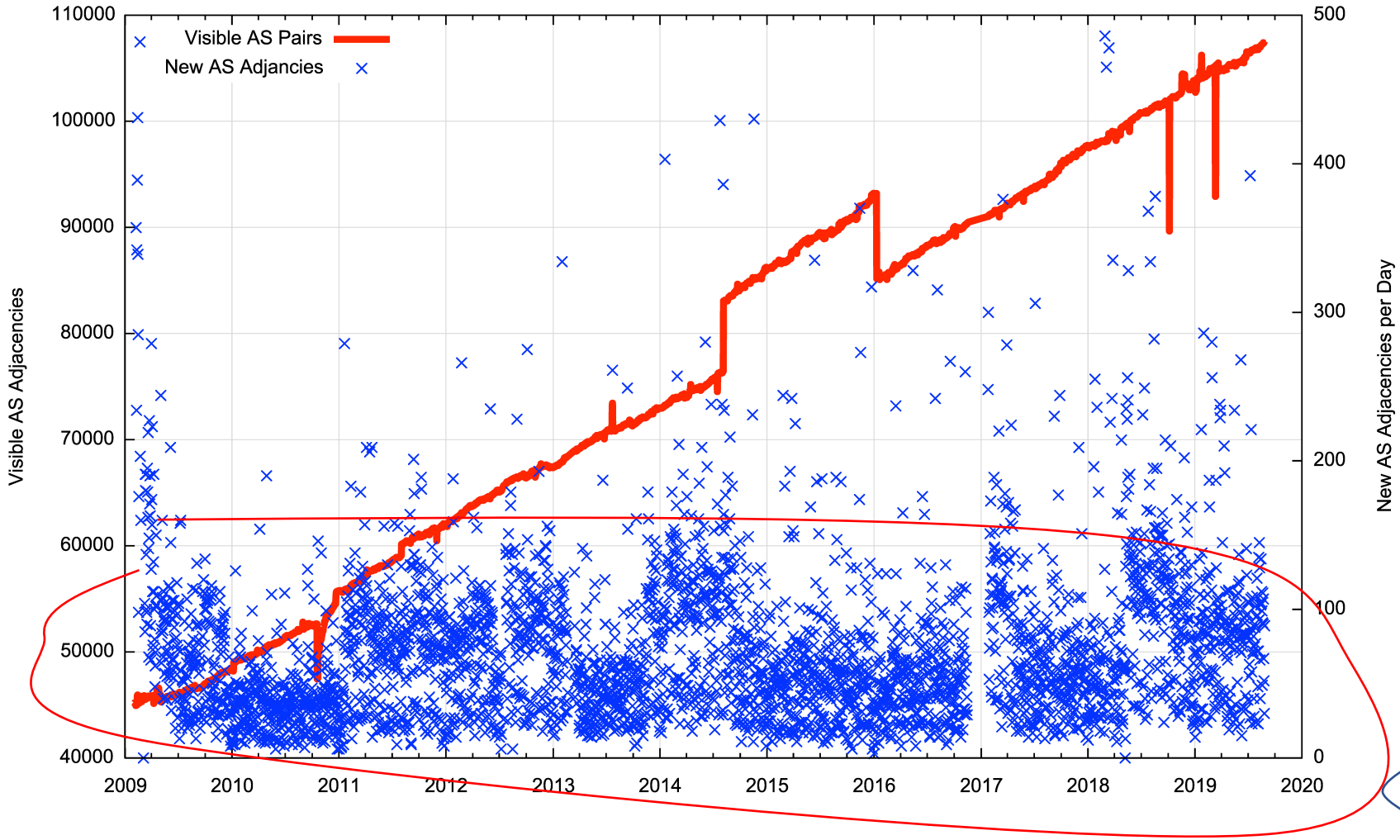
There are some 200-300 ”new” prefixes per day in the IPv4 BGP RIB

BGP Prefix Updates - IPv6



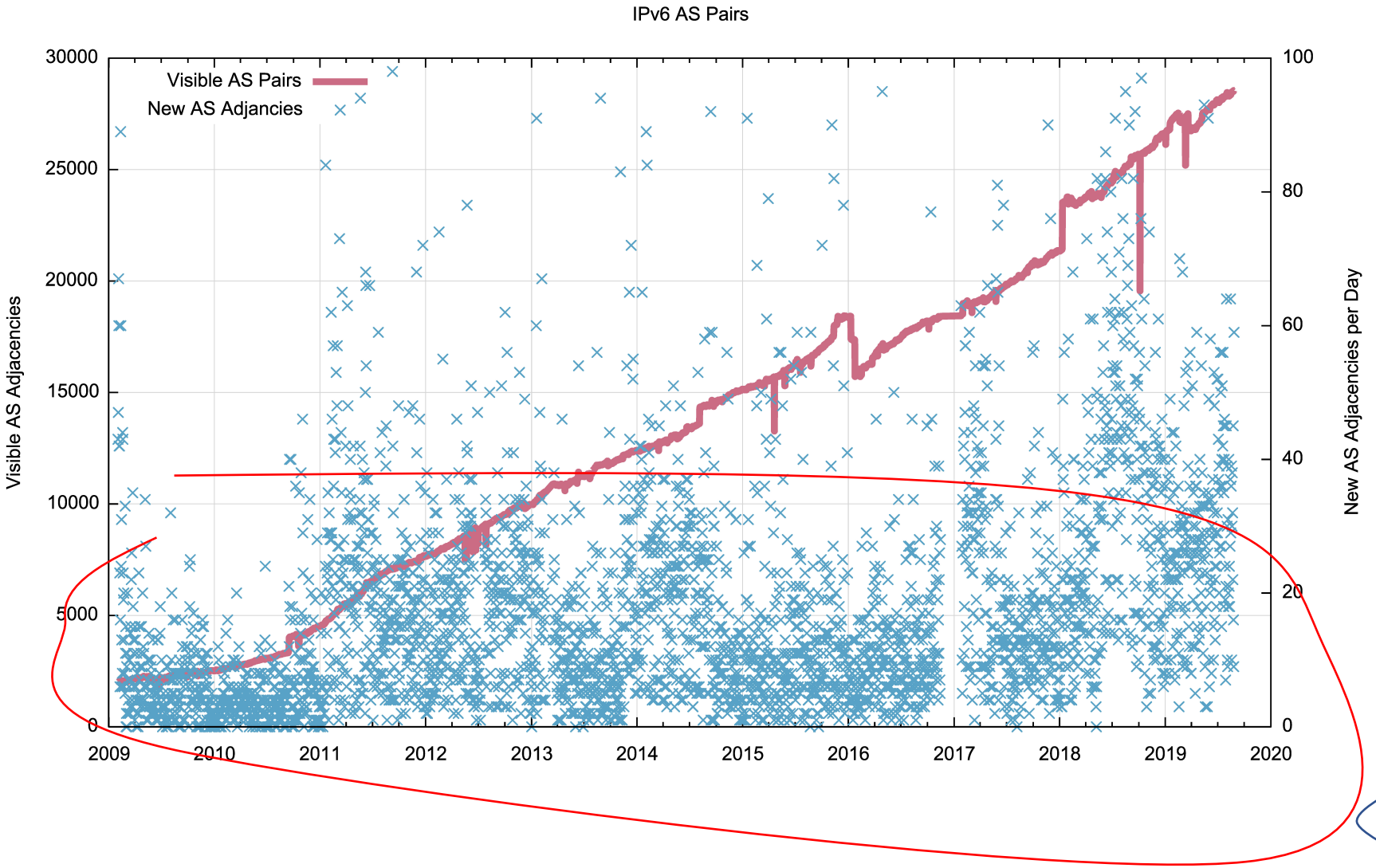
There were some 5-10 ”new” prefixes per day in the IPv6 BGP RIB

BGP AS Adjacency Updates - IPv4



There are up to 100 ”new” AS Adjacencies per day in the IPv4 BGP RIB

BGP AS Adjacency Updates - V6



There are up to 30 ”new” AS Adjacencies per day in the IPv4 BGP RIB

BGP Updates and Information Content

- BGP updates tend to repeat previous information most of the time
  - 150,000 updates per day, but only 200 – 300 previously unseen prefixes and 100 previously unseen AS adjacencies per day
- The “new” information content volume in BGP updates is relatively small, and is scale free

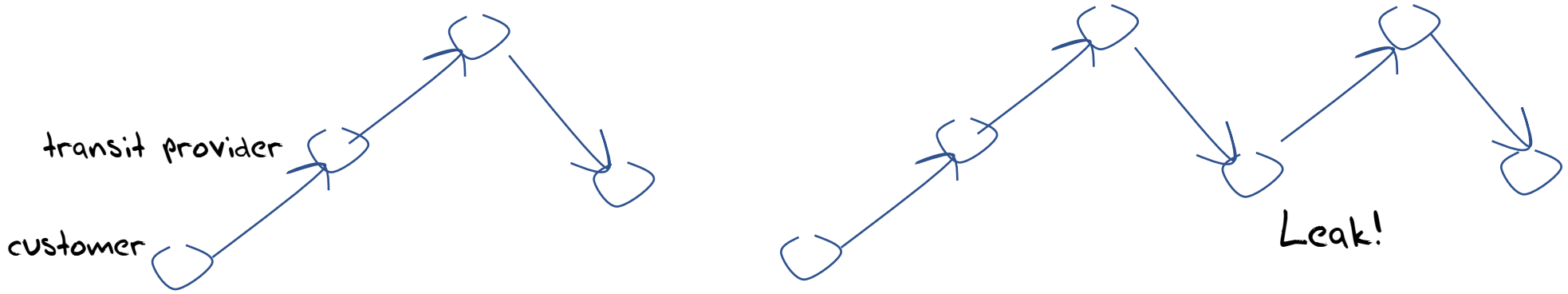
  (The rate of growth is not directly related to the size of the network)
- If we use “new” information as a trigger point to look for unusual BGP activity we might have a useful way to filter BGP updates

BGP Update Processing

- For each Prefix, load the prefix into an aggregate and more specific context tree
- For each AS Path, analyse the AS ordered adjacencies and infer the provider / peer / customer relationship
- Geolocate the prefix and the originating AS
- Check the ROA status
- Check the IRR material for this prefix
- Now apply an anomaly “interest” level calculation on announcement and withdrawal

What is "interesting" in BGP Updates?

- AS paths that contain “valleys”
  - The AS at the trough of the valley is leaking routes from one upstream provider to another



What is "interesting" in BGP Updates?

- AS paths that contain “valleys”
- AS paths that contain ASes in “unusual” places
  - The AS path is synthetic and is not a BGP generated ‘history’ of the propagation of the update
  - It could be a poison AS to deflect traffic for TE
  - Or an effort to create a false ‘best path’ to redirect traffic

What is "interesting" in BGP Updates?

- AS paths that contain “valleys”
- AS paths that contain ASes in “unusual” places
- Previously unseen prefixes
  - Address hijack?
  - Bogon?

What is "interesting" in BGP Updates?

- AS paths that contain “valleys”
- AS paths that contain ASes in “unusual” places
- Previously unseen prefixes
- Previously unseen ASes
  - Fake AS Path?
  - AS hijack?

What is "interesting" in BGP Updates?

- AS paths that contain “valleys”
- AS paths that contain ASes in “unusual” places
- Previously unseen prefixes
- Previously unseen ASes
- Prefixes that are more specifics or aggregates where the origin ASs are different
  - “Hole punching”
  - Or route hijack by more specifics

What is "interesting" in BGP Updates?

- AS paths that contain “valleys”
- AS paths that contain ASes in “unusual” places
- Previously unseen prefixes
- Previously unseen ASes
- Prefixes that are more specifics or aggregates where the origin ASs are different
- Prefixes where the geolocation of the more specifics are different
  - “Hole punching”
  - Or route hijack by more specifics

What is "interesting" in BGP Updates?

- AS paths that contain “valleys”
- AS paths that contain ASes in “unusual” places
- Previously unseen prefixes
- Previously unseen ASes
- Prefixes that are more specifics or aggregates where the origin ASs are different
- Prefixes where the geolocation of the more specifics are different
- Prefixes where the geolocation of the prefix and the Originating AS are different
  - route hijack?

What is "interesting" in BGP Updates?

- AS paths that contain “valleys”
- AS paths that contain ASes in “unusual” places
- Previously unseen prefixes
- Previously unseen ASes
- Prefixes that are more specifics or aggregates where the origin ASs are different
- Prefixes where the geolocation of the more specifics are different
- Prefixes where the geolocation of the prefix and the Originating AS are different
- Short lived prefixes
  - After “a period” its no longer an anomaly but part of the set of BGP ground truths

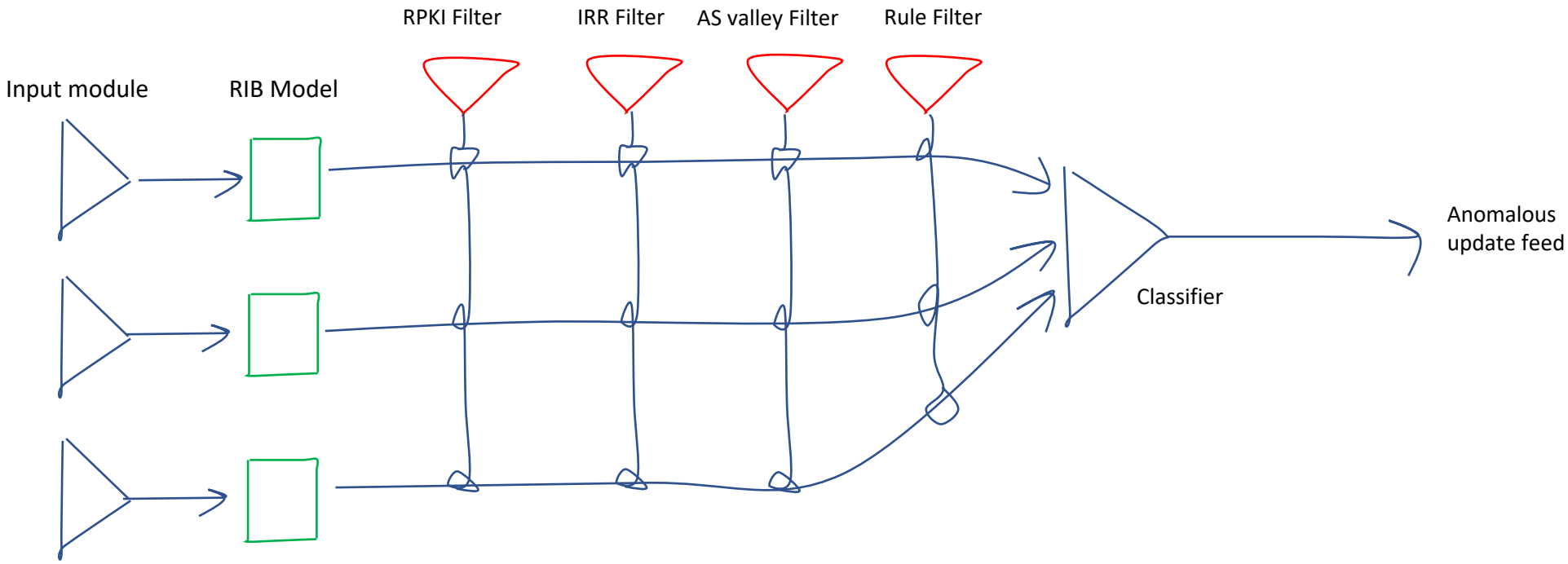
Interest and Intent Filtering

- RPKI can be interpreted as a strong statement of routing intent
- IRR data can also be interpreted as intent, although without the same level of clarity of intent
- More specific announcement floods are “interesting”

This Project

- Yet another BGP Anomaly detector
- Why another?
  - Open source code base (C)
  - Generic design that can cope with feeds from one or more BGP speakers
  - Intended to use plug in filter sets, to allow both specific rule applications and more general anomaly detection

Overall Architecture



Reporting

- How should the tool report?
  - JSON feed
  - Web Archive
  - Linked into RIPEStat
  - Other report formats?

Current Status

Flikr: P.A.H. http://bit.ly/320TEF4

Interested in this work?

You can play too:

- Pass an eBGP feed to a detector
- Take a copy of the code and apply it to your own BGP feeds?
- Subscribing to a BGP anomaly feed service using your rule set
- Interested in subscribing to a general BGP anomaly feed

APNIC's Role

- We share an interest in a secure and stable routing system
- We'd like to help operators by informing them of the status of routing stability
- We are interested in trying to measure the incidence of BGP anomalies over time to inform the community about the severity and incidence of these anomalies

Thanks!