RIPE NCC
RIPE NETWORK COORDINATION CENTRE

# RIPE NCC Routing Information Service (RIS)

## Overview

Colin Petrie | 14/12/2016 | RON++

# What is RIS?

# What is RIS?

- Worldwide network of BGP collectors
- Deployed at Internet Exchange Points
  - Including at AMS-IX and NL-IX
- Collects raw BGP data from peers
- Stores BGP routing table dumps
- 15+ years of history
- Used by network operators and researchers every day!

# Collector locations

# What is RIS?

- A huge archive of useful data about BGP routing activity
- A database where you can look up almost anything you want to know about routing
- We provide APIs to query all the data
- And of course, a nice shiny web interface!
  - RIPEstat
  - https://stat.ripe.net/

# Why RIS?

Why are we doing this?
A bit of history

# Why RIS?

- Original project was defined in RIPE-200
  - in 1999, when the BGP table was 64,000 routes!!
- Looking glasses are instantaneous
- Routing problems are also instantaneous
- BGP needs to be recorded, to track what is happening and what has happened.
- Also to provide statistics and reporting on routing table metrics

# Why the RIPE NCC RIS?

- RIPE NCC is a neutral body
- Experience running measurement platforms
  - Test Traffic Measurement project
  - RIPE Atlas
- Supporting our own members
  - who are mainly network operators
- Supporting the community
  - researchers
  - operators

# RIS data access

What can you get?
And how do you get it?

# Raw data!

- 15+ years of data available to download and analyse yourself :)
- https://www.ripe.net/analyse/internet-measurements/routing-information-service-ris/ris-raw-data
- Data stored in MRT (RFC6396) format
- Readable using BGPdump utility
  - Open source, maintained by RIPE NCC
  - https://bitbucket.org/ripencc/bgpdump

# Web interfaces and APIs

- Of course, if all we did was store the raw data, we'd just need a bunch of hard disks and an FTP server
- But you want to query all our lovely datasets!
- RIPEstat
  - https://stat.ripe.net/
- Our portal for everything you ever wanted to know!

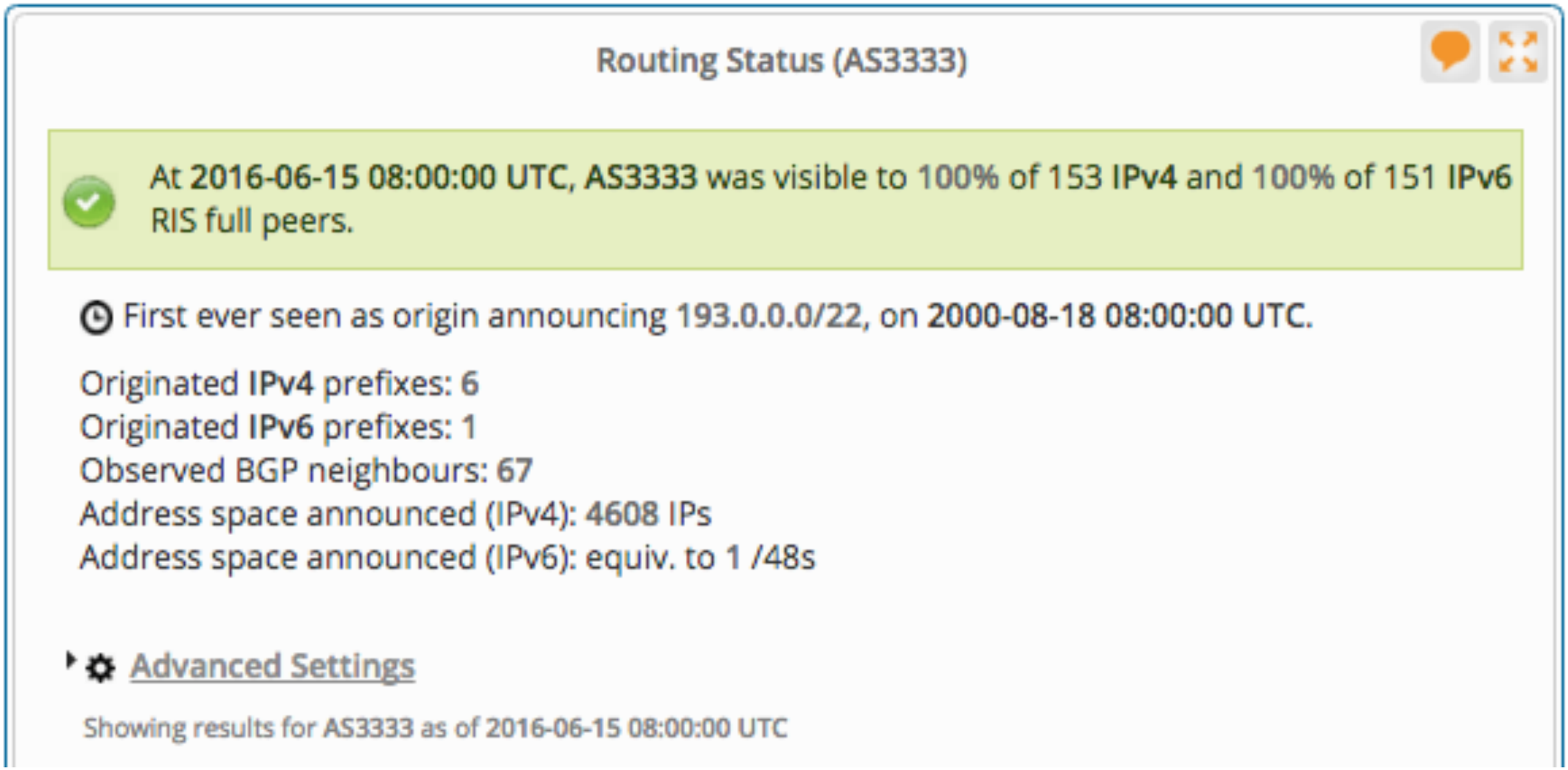
# RIPEstat

- RIPEstat is a web-based interface that provides everything you ever wanted to know about IP address space, Autonomous System Numbers (ASNs), and related information for hostnames and countries in one place.

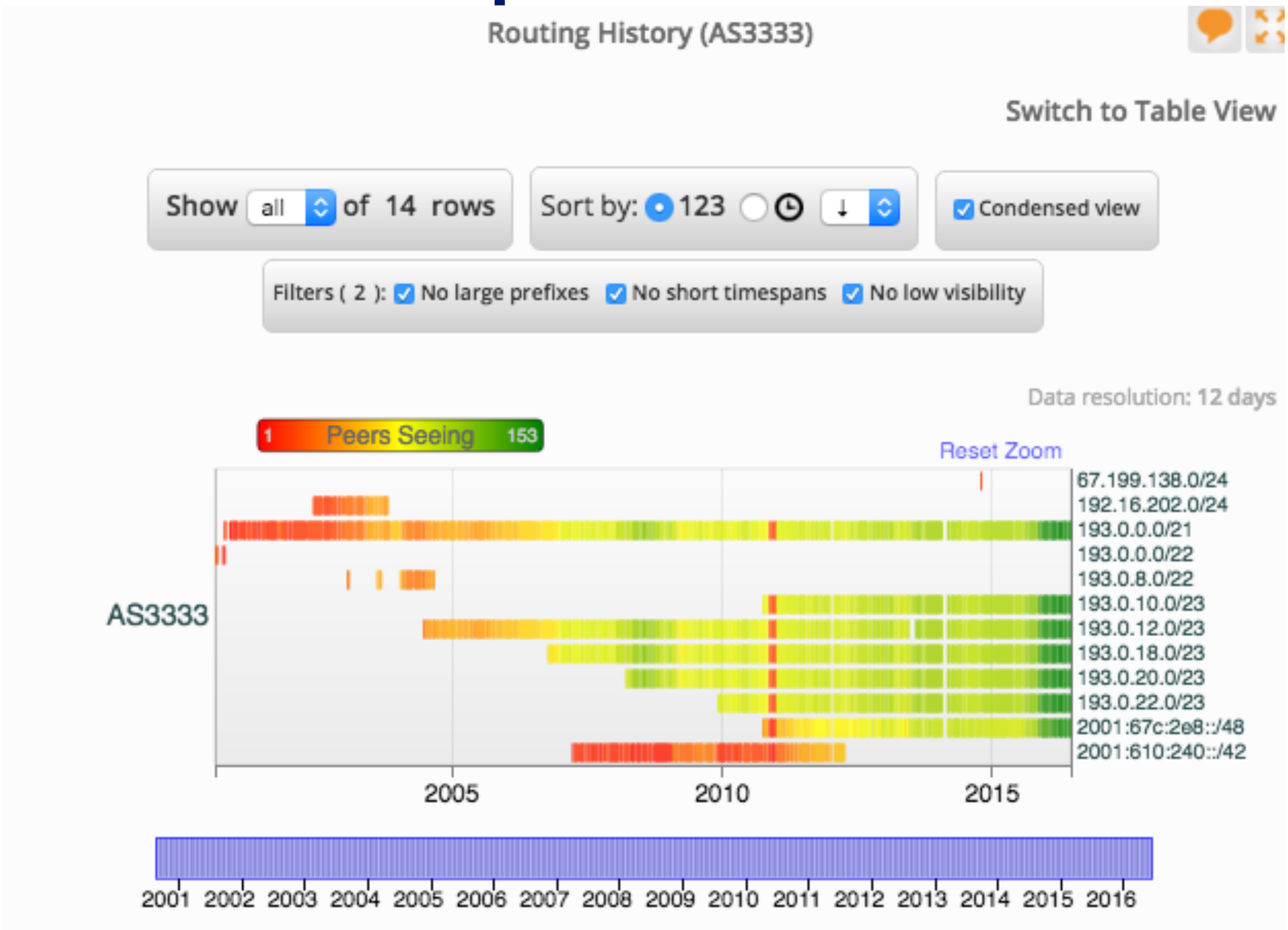
# RIPEstat

- What can you search for?
  - ASN (autonomous system number)
  - IPv4 address
  - IPv4 prefix
  - IPv6 address
  - IPv6 prefix
  - country code (ISO - e.g. NL, ES, US)

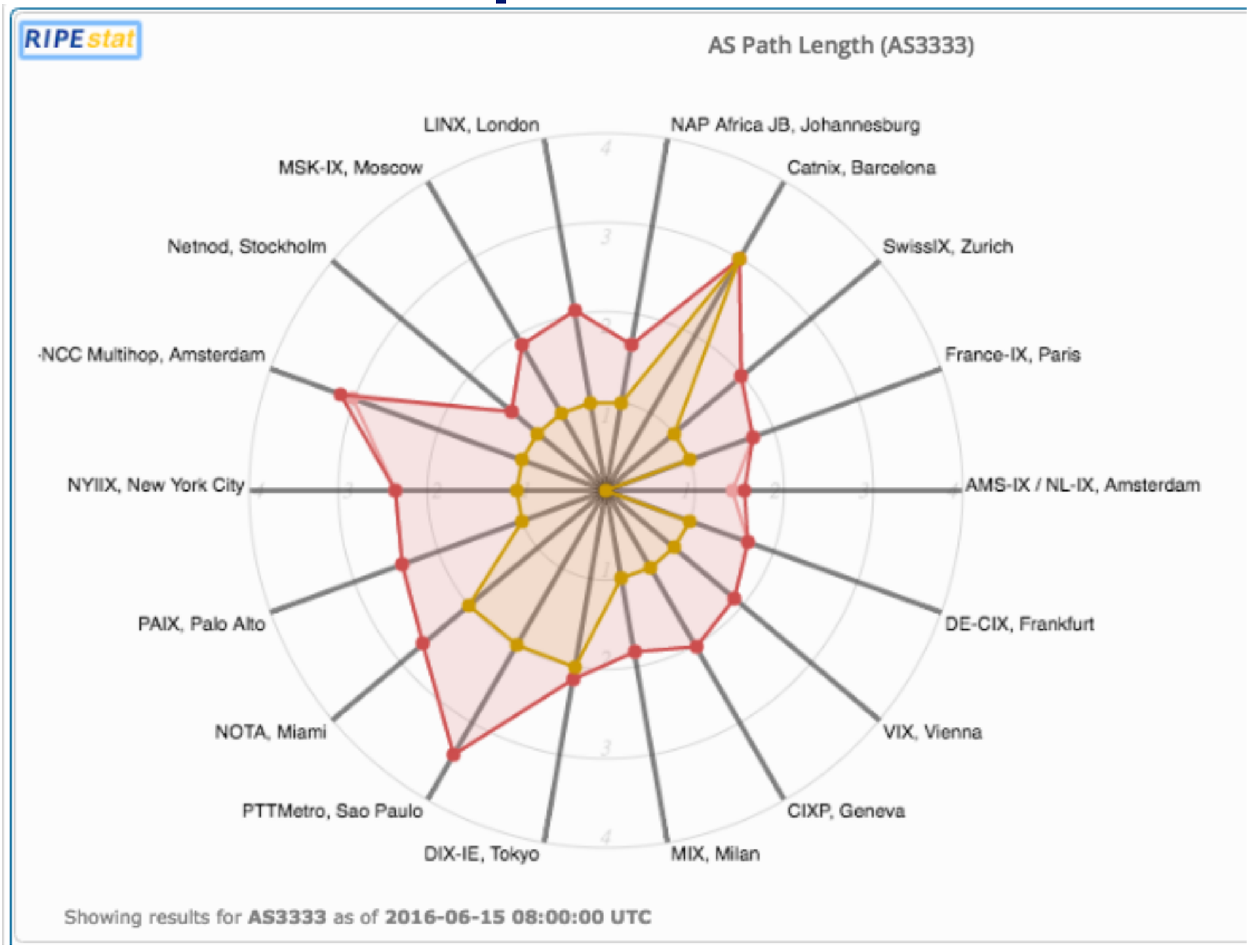
# RIPEstat examples



# RIPEstat examples



# RIPEstat examples



# RIPEstat examples

RIPEstat

Visibility (AS3333)

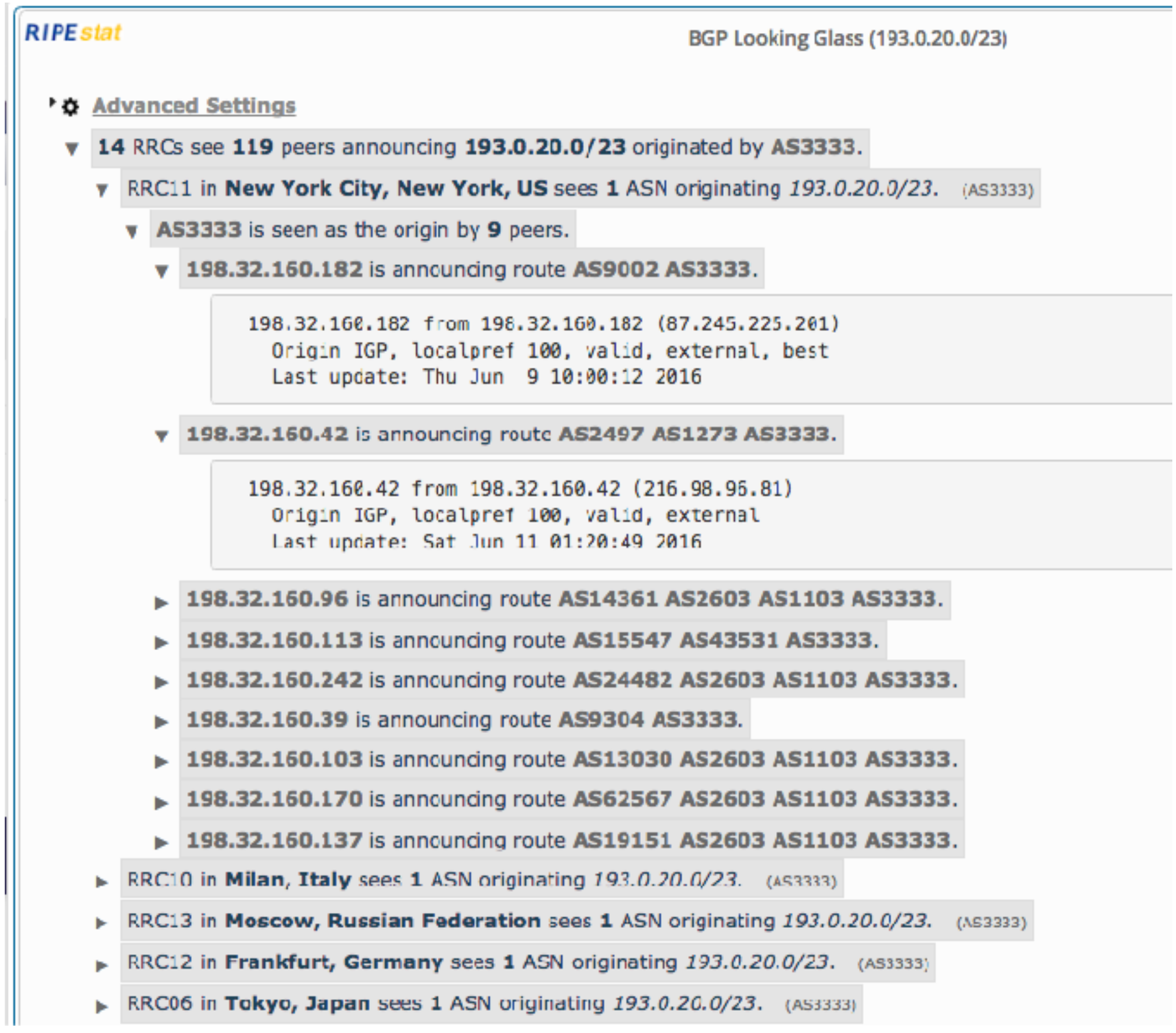
**AS3333** is visible by **100%** of 153 **IPv4** and **100%** of 151 **IPv6** RIS full peers.

**Visibility Location Details of AS3333**

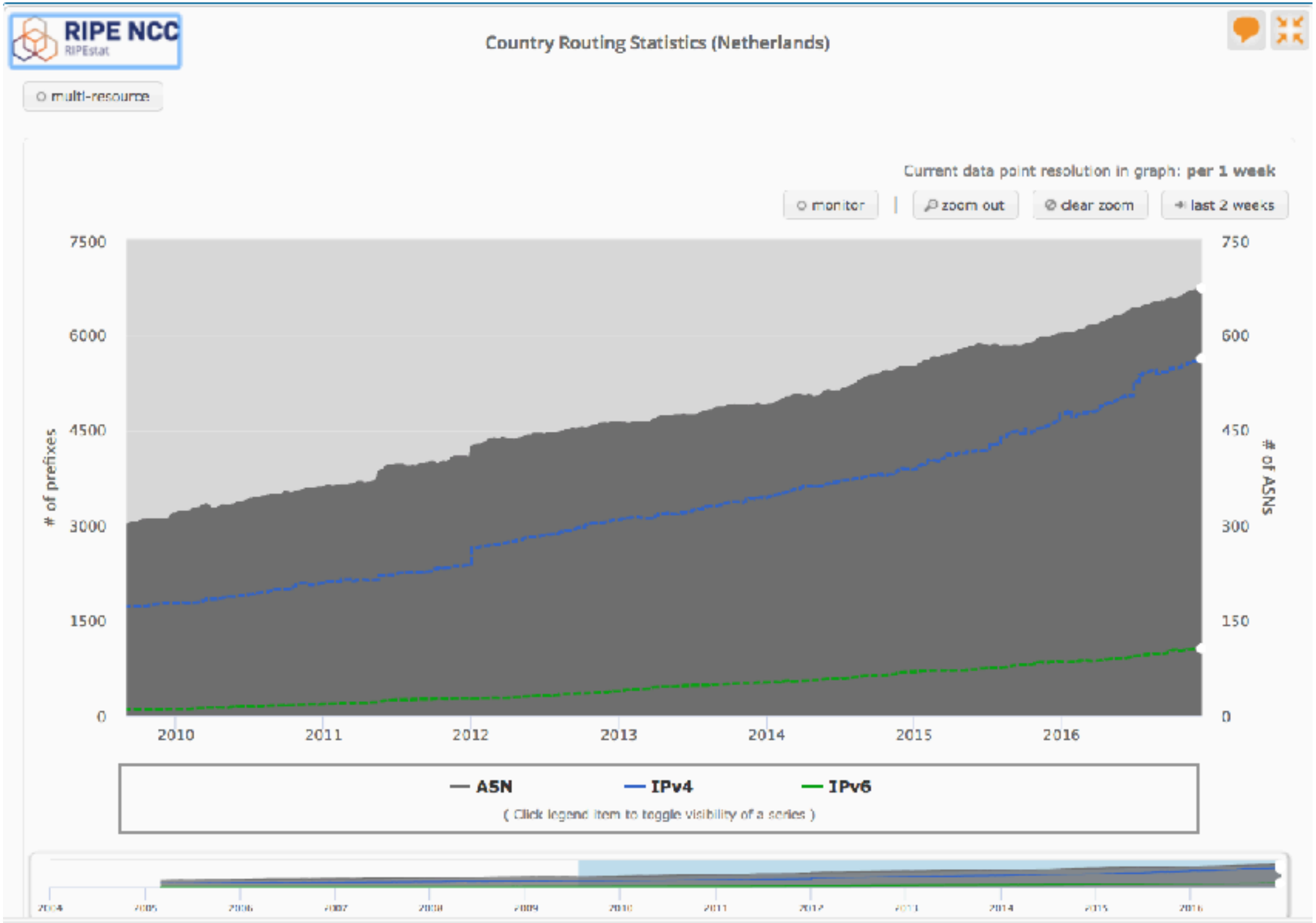
| RRC | IXP Location | Location | IPv4 peers seeing | IPv6 peers seeing | IPv4 Visibility | IPv6 Visibility |
|---|---|---|---|---|---|---|
| RRC00 | RIPE-NCC Multihop | Amsterdam, Netherlands | 13 of 13 | 10 of 10 | 100% | 100% |
| RRC01 | LINX | London, United Kingdom | 7 of 7 | 9 of 9 | 100% | 100% |
| RRC03 | AMS-IX / NL-IX | Amsterdam, Netherlands | 7 of 7 | 12 of 12 | 100% | 100% |
| RRC04 | CIXP | Geneva, Switzerland | 7 of 7 | 5 of 5 | 100% | 100% |
| RRC05 | VIX | Vienna, Austria | 4 of 4 | 6 of 6 | 100% | 100% |
| RRC06 | DIX-IE | Tokyo, Japan | 2 of 2 | 2 of 2 | 100% | 100% |
| RRC07 | Netnod | Stockholm, Sweden | 4 of 4 | 5 of 5 | 100% | 100% |
| RRC10 | MIX | Milan, Italy | 10 of 10 | 8 of 8 | 100% | 100% |
| RRC11 | NYIIX | New York City, US | 9 of 9 | 9 of 9 | 100% | 100% |
| RRC12 | DE-CIX | Frankfurt, Germany | 14 of 14 | 21 of 21 | 100% | 100% |
| RRC13 | MSK-IX | Moscow, Russian Federation | 11 of 11 | 4 of 4 | 100% | 100% |
| RRC14 | PAIX | Palo Alto, US | 7 of 7 | 8 of 8 | 100% | 100% |
| RRC15 | PTTMetro | Sao Paulo, Brazil | 14 of 14 | 10 of 10 | 100% | 100% |
| RRC16 | NOTA | Miami, US | 3 of 3 | 3 of 3 | 100% | 100% |
| RRC18 | Catnix | Barcelona, Spain | 1 of 1 | 1 of 1 | 100% | 100% |
| RRC19 | NAP Africa JB | Johannesburg, South Africa | 3 of 3 | 2 of 2 | 100% | 100% |
| RRC20 | SwissIX | Zurich, Switzerland | 18 of 18 | 18 of 18 | 100% | 100% |
| RRC21 | France-IX | Paris, France | 19 of 19 | 18 of 18 | 100% | 100% |

Showing results for **AS3333** as of **2016-06-15 08:00:00 UTC**

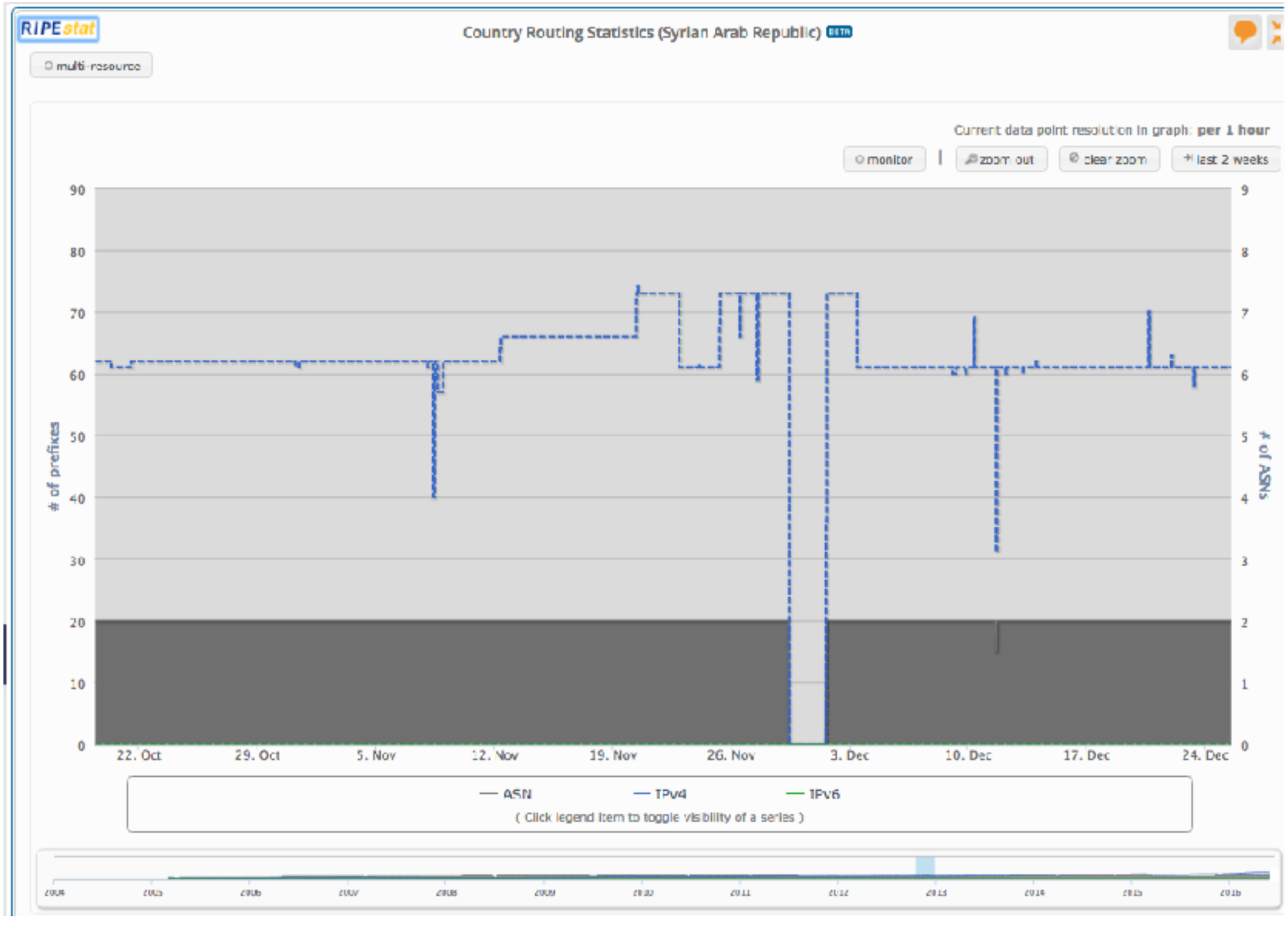
# RIPEstat examples



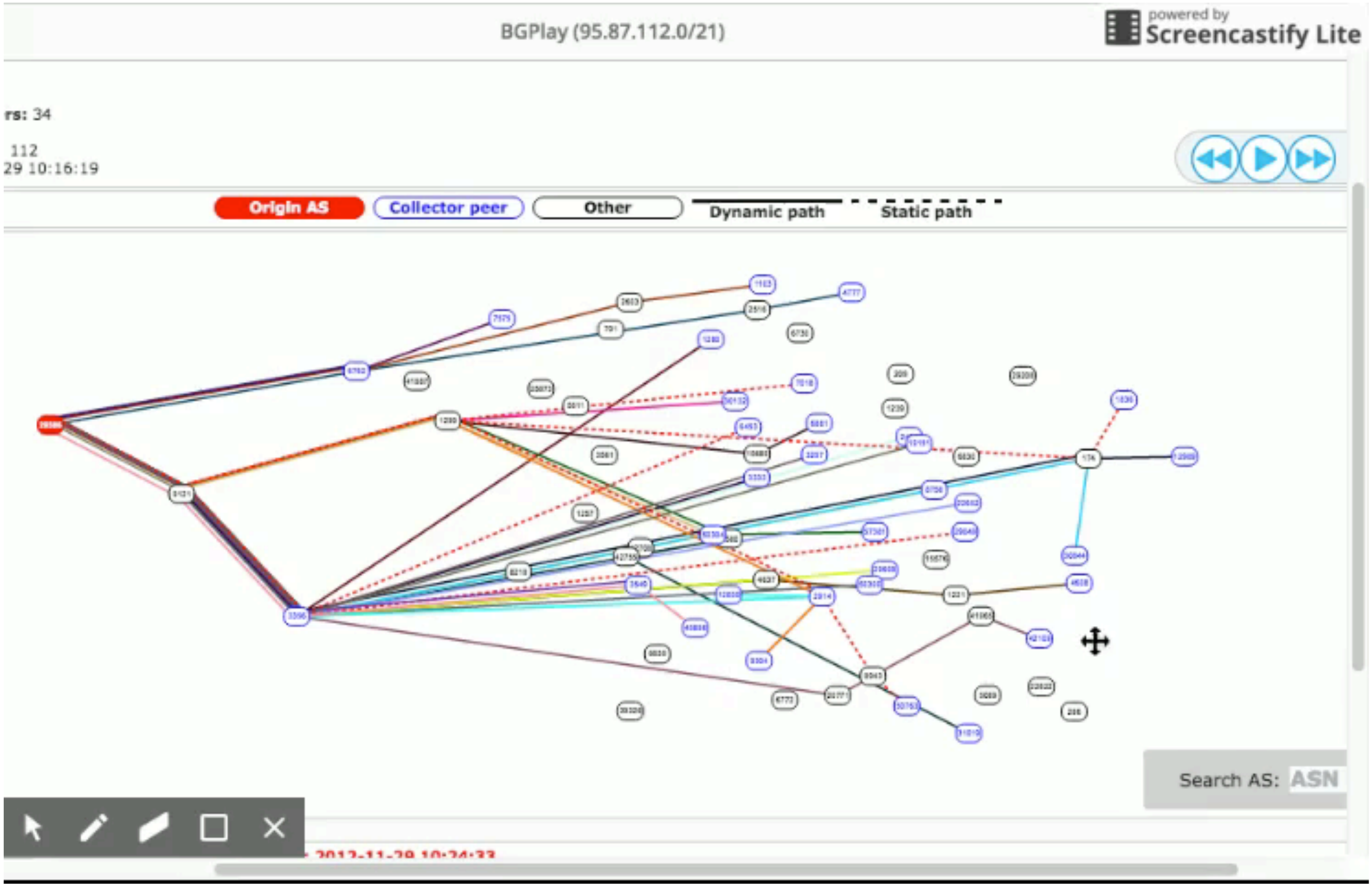
# RIPEstat examples



# RIPEstat examples



# RIPEstat examples



# Live stream - demo

- Prototype!!
- Let’s see if it works
- http://stream-dev.ris.ripe.net/demo

- Live stream enables new applications
  - BGP Hijack detection
  - real time anomaly analysis
  - live monitoring of your routes

# RIS data access

## How else can you get it

# RIPEstat Data API

- All these queries are available through an API
- Actually, all those shiny web interfaces use the API anyway
- You can use it too!! Write your own scripts etc
- https://stat.ripe.net/docs/data_api
- There are also some extra API calls which are not yet visualised

# RIPEstat Data API

- Remember this started because looking glasses are instantaneous?
- BGP State
  - https://stat.ripe.net/docs/data_api#BGPState
- This data call returns the state of BGP routes for a resource at a certain point in time, as observed by all the RIS collectors
- This is derived by applying a computation of state to the RIB dump (granularity=8h) that occurred exactly before that time, using the BGP updates observed between the RIB time and the query time.

# RIPEstat Data API - BGP State

- https://stat.ripe.net/data/bgp-state/data.json?resource=193.0.24.0/21&timestamp=2016-05-19T00:33:21
- Show me what this prefix looked like at exactly this time!

*State of AS2121 (RIPE meeting) prefix after we plugged in the router at RIPE72 in Copenhagen*

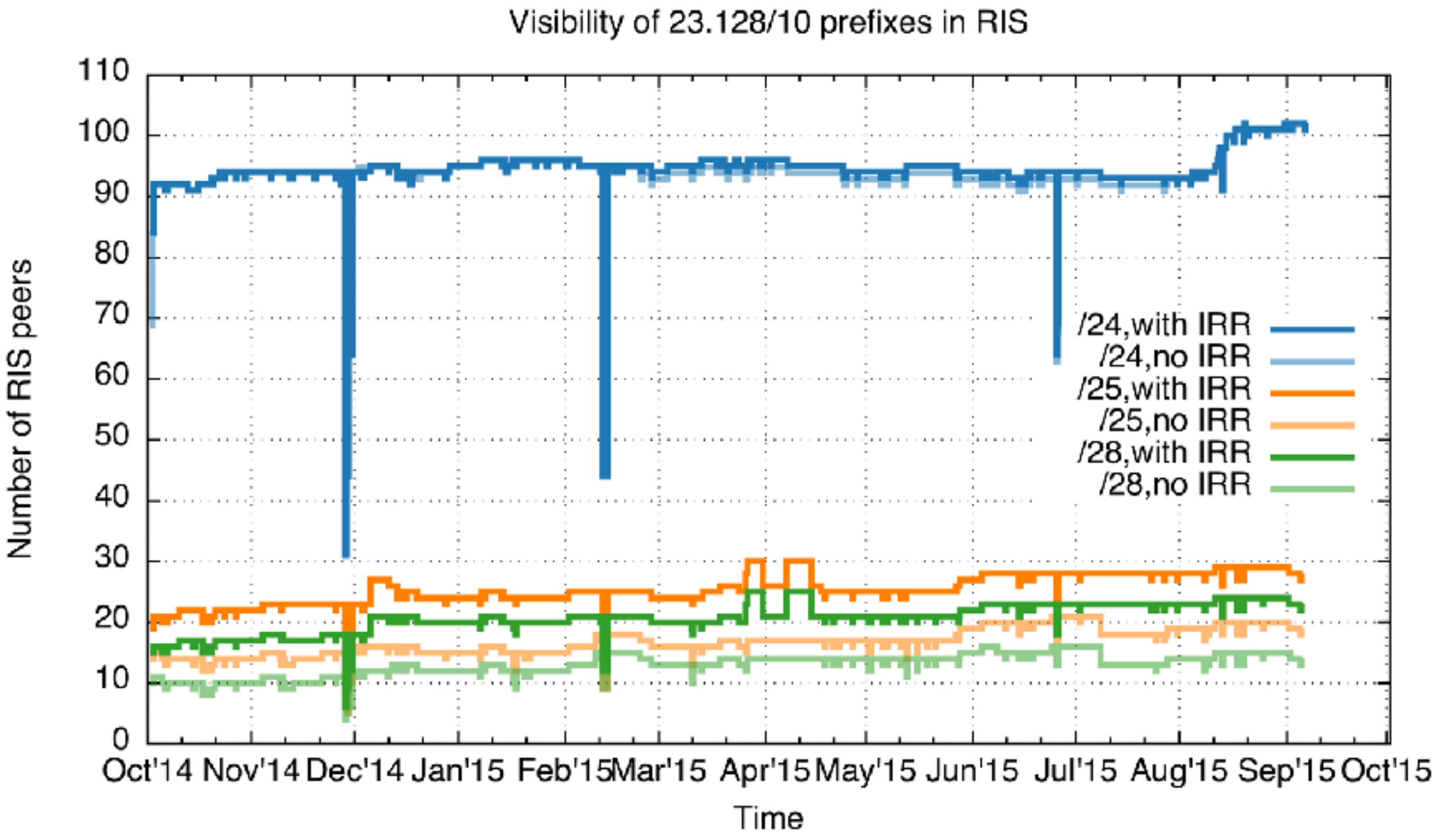

```
"data": {
    "bgp_state": [
        {
            "source_id": "00-12.0.1.63",
            "path": [
                7018,
                174,
                42525,
                2121
            ],
            "community": [
                "7018:5000",
                "7018:37232"
            ],
            "target_prefix": "193.0.24.0/21"
        },
        {
            "source_id": "00-146.228.1.3",
            "path": [
                1836,
                2852,
                21320,
                2603,
                42525,
                2121
            ],
            "community": [
                "1836:120",
                "1836:3100",
                "1836:3110",
                "2852:2852"
            ],
            "target_prefix": "193.0.24.0/21"
        },
        {
            "source_id": "00-176.12.110.8",
            "path": [
                50300,
                3356,
                42525,
                2121
            ],
            "community": [],
            "target_prefix": "193.0.24.0/21"
```

# What else can you do?

Lots of analysis that this data allows

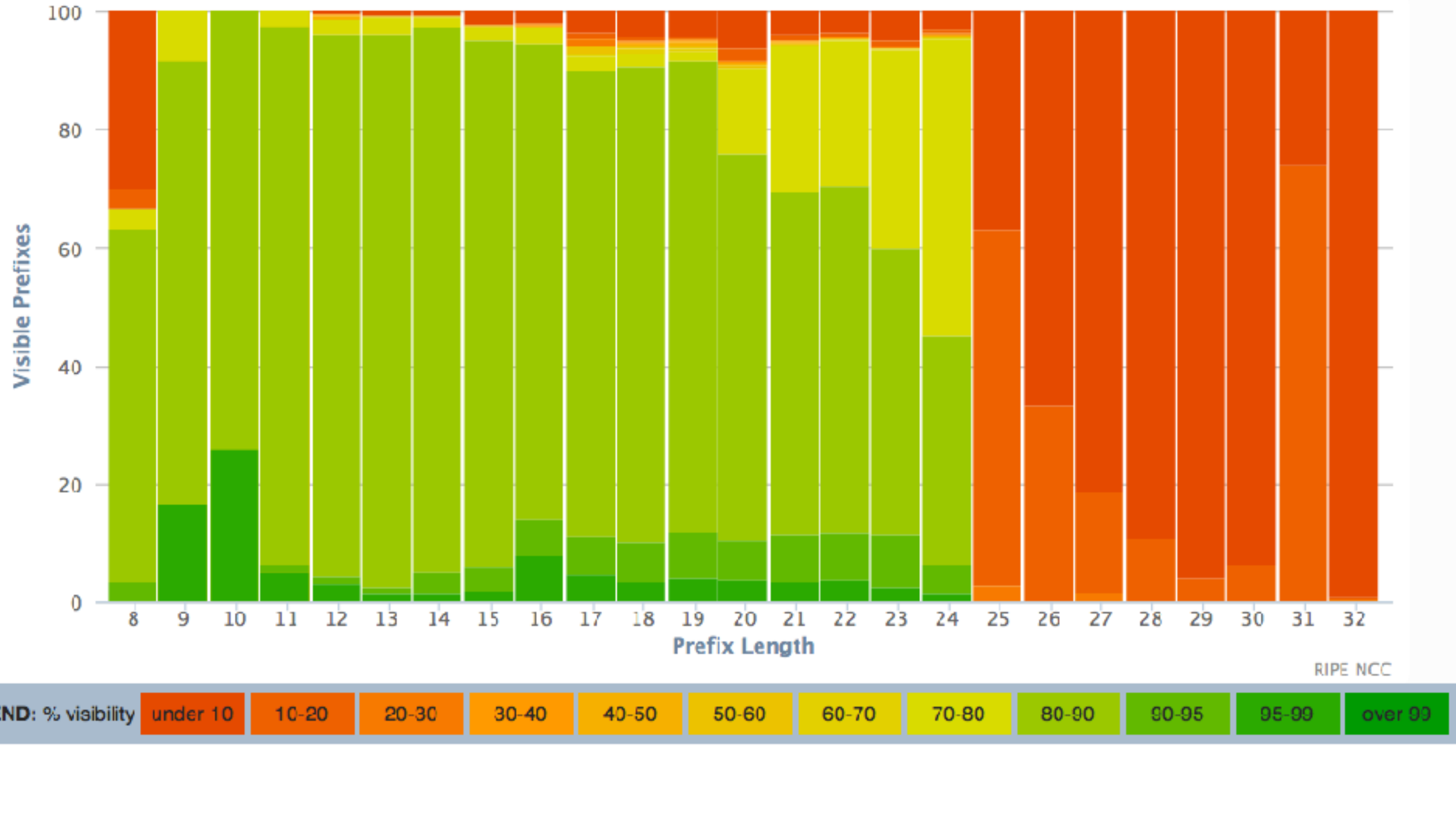
# Prefix reachability studies

- https://labs.ripe.net/Members/emileaben/has-the-routability-of-longer-than-24-prefixes-changed



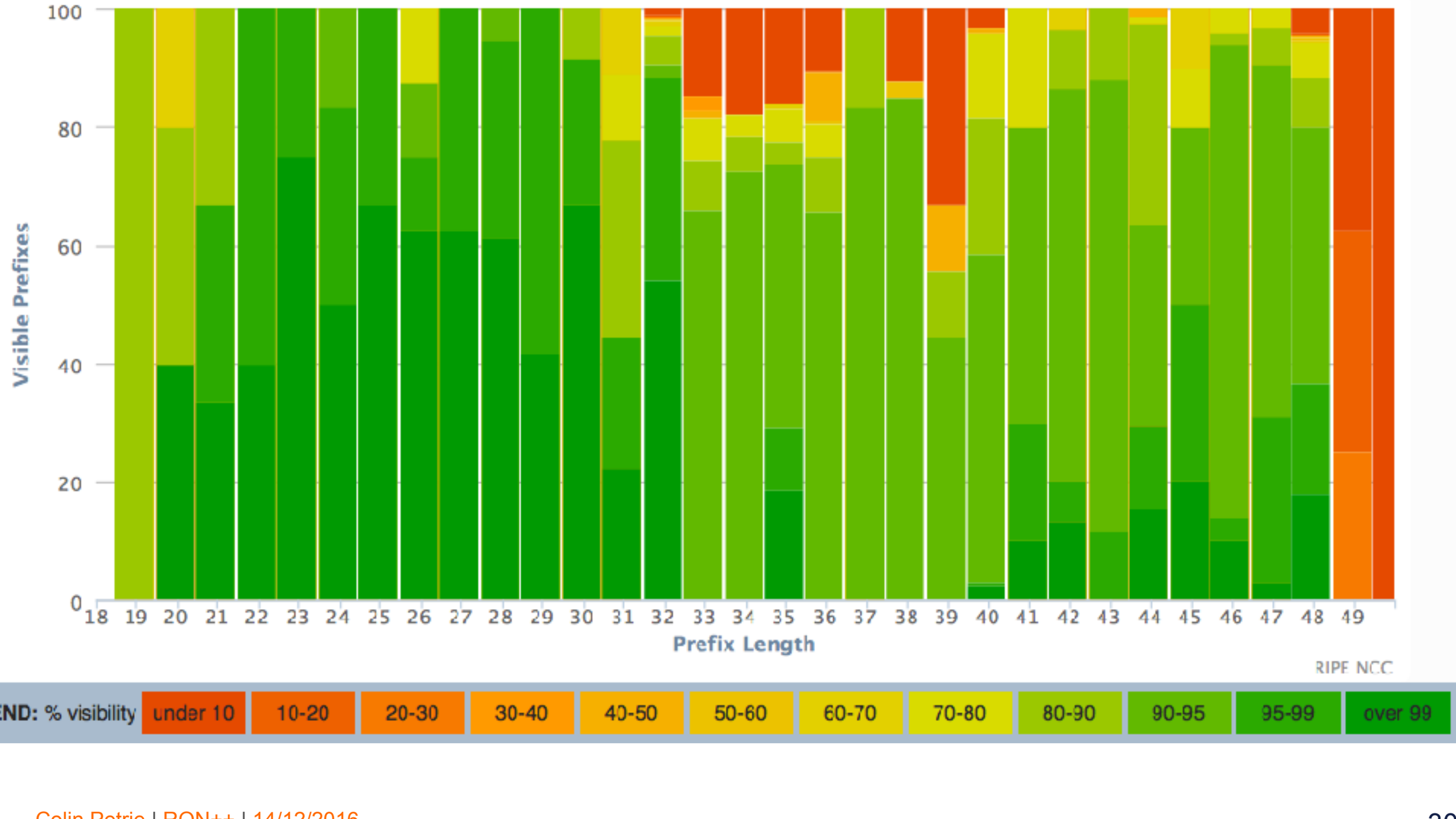
# Prefix length visibility

- https://labs.ripe.net/Members/dbayer/visibility-of-prefix-lengths



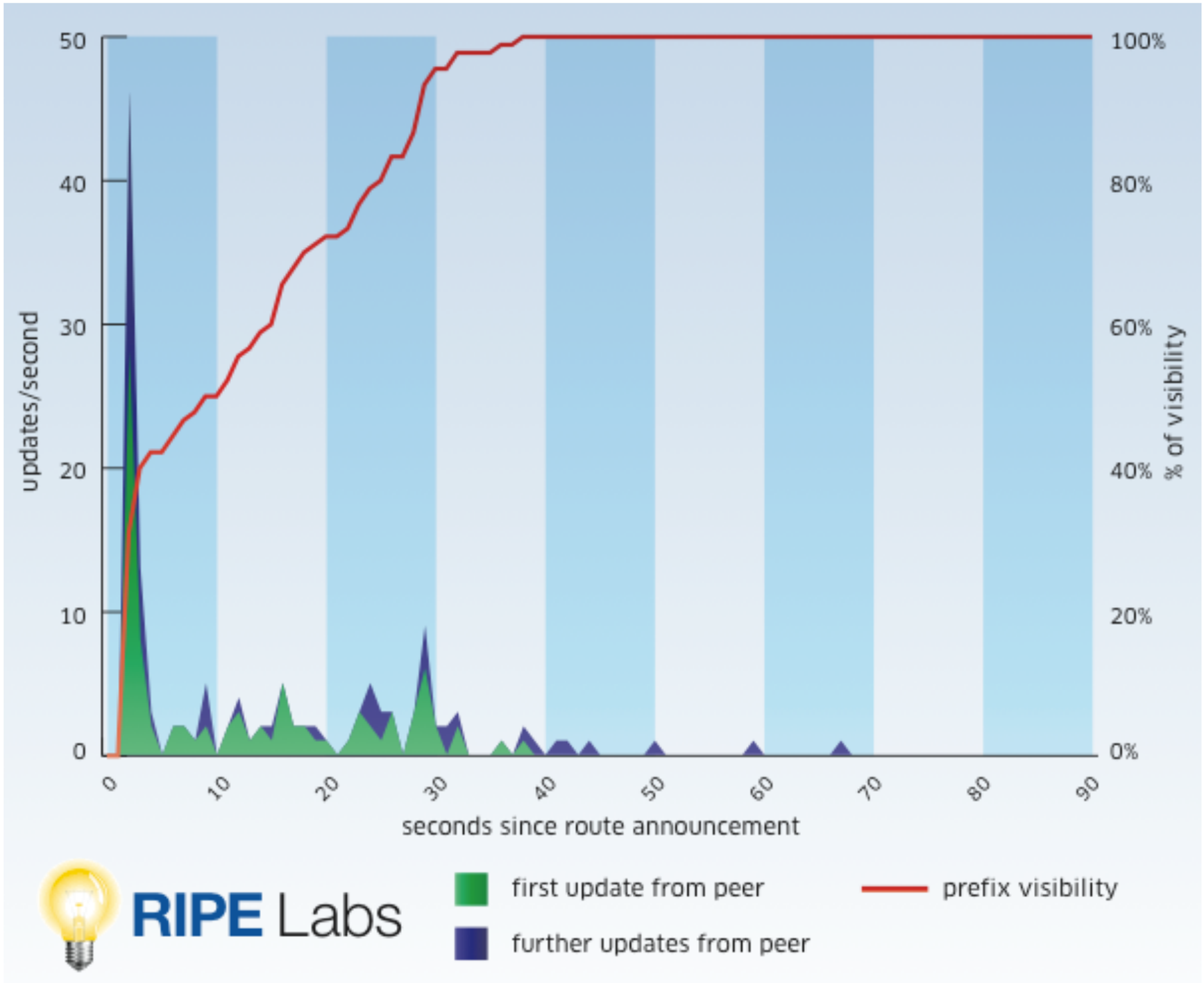
# Prefix length visibility

- https://labs.ripe.net/Members/dbayer/visibility-of-prefix-lengths



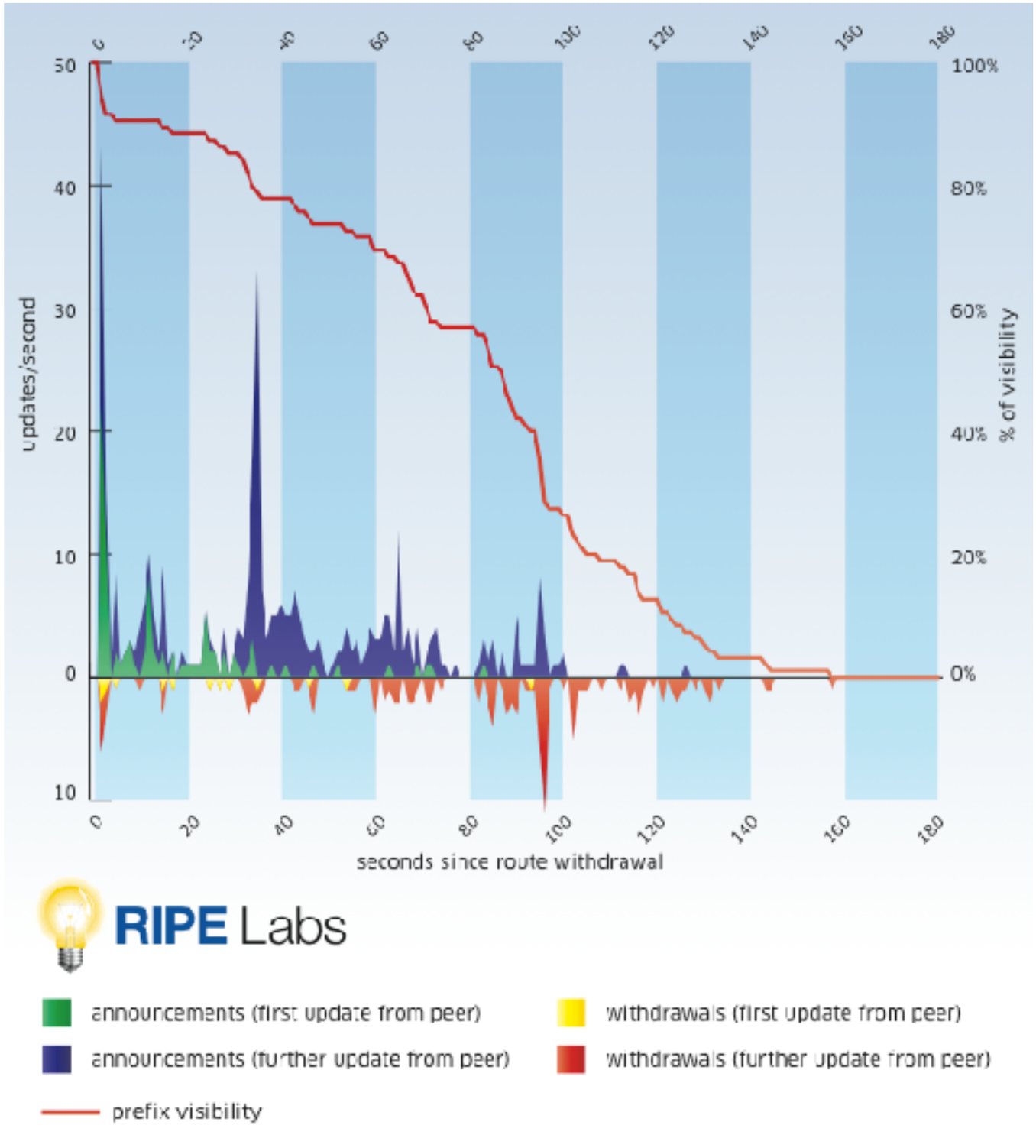
# BGP update propagation

- https://labs.ripe.net/Members/vastur/the-shape-of-a-bgp-update



# BGP update propagation

- https://labs.ripe.net/Members/vastur/the-shape-of-a-bgp-update



# How can you help?

# How can you help?

- Peer with us!!!
  - AS12654 @ AMS-IX, NL-IX
  - RRC03
  - http://www.ris.ripe.net/cgi-bin/peerreg.cgi
- Send us your routes
  - If you can send us your full BGP table, please do
  - If not, send us what you can!
  - We will be recording them forever ;-)

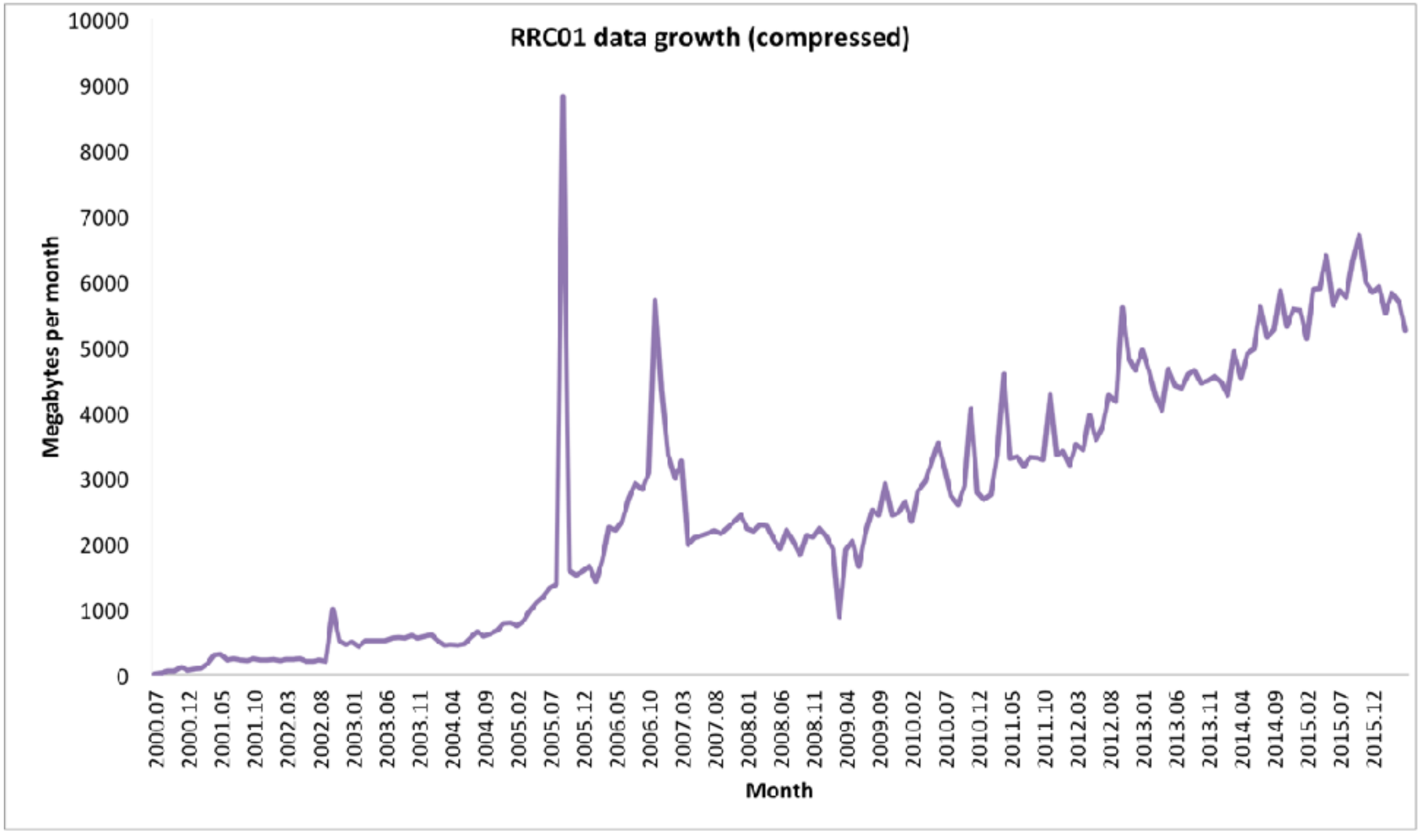
# RIS growth

Because the internet keeps growing

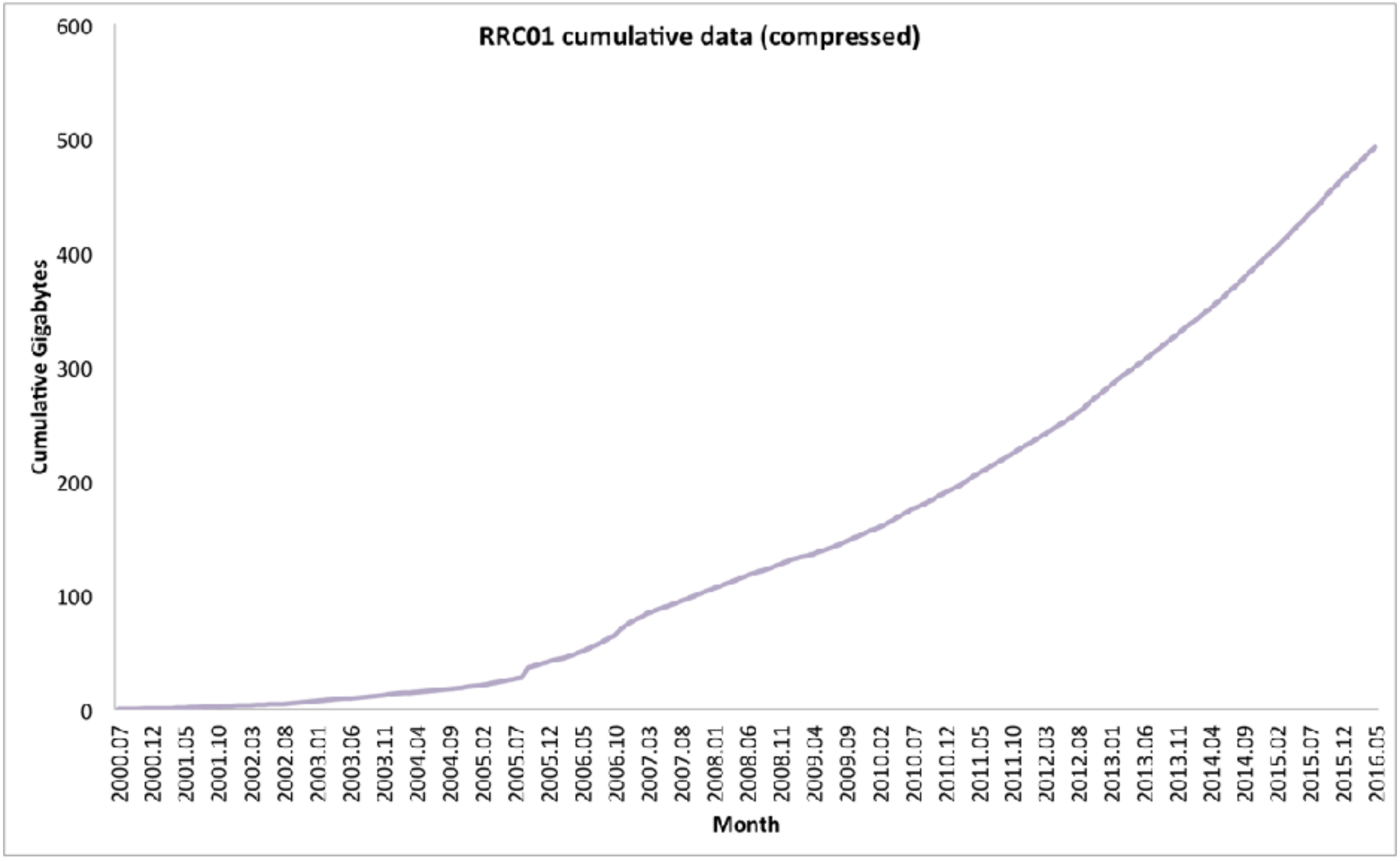
# Collector history

| Collector | Location | IXP | Deployed | Removed |
|---|---|---|---|---|
| RRC00 | Amsterdam | Multi-hop | 1999 | - |
| RRC01 | London | LINX | 2000 | - |
| RRC02 | Paris | SFINX | 2001 | 2008 |
| RRC03 | Amsterdam | AMS-IX | 2001 | - |
| RRC04 | Geneva | CIXP | 2001 | - |
| RRC05 | Vienna | VIX | 2001 | - |
| RRC06 | Tokyo | DIX-IE | 2001 | - |
| RRC07 | Stockholm | Netnod | 2002 | - |
| RRC08 | San Jose | MAE-West | 2002 | 2004 |
| RRC09 | Zurich | TIX | 2003 | 2004 |
| RRC10 | Milan | MIX | 2003 | - |
| RRC11 | New York | NYIIX | 2004 | - |
| RRC12 | Frankfurt | DE-CIX | 2004 | - |
| RRC13 | Moscow | MSK-IX | 2005 | - |
| RRC14 | Palo Alto | PAIX | 2005 | - |
| RRC15 | Sao Paulo | PTT-Metro SP | 2006 | - |
| RRC16 | Miami | NOTA | 2008 | - |
| RRC18 | Barcelona | CATNIX | 2015 | - |
| RRC19 | Johannesburg | NAPAfrica JB | 2016 | - |
| RRC20 | Zurich | SwissIX | 2015 | - |
| RRC21 | Paris | FranceIX | 2015 | - |

# RRC01 data production rate



# RRC01 cumulative data



# Data growth

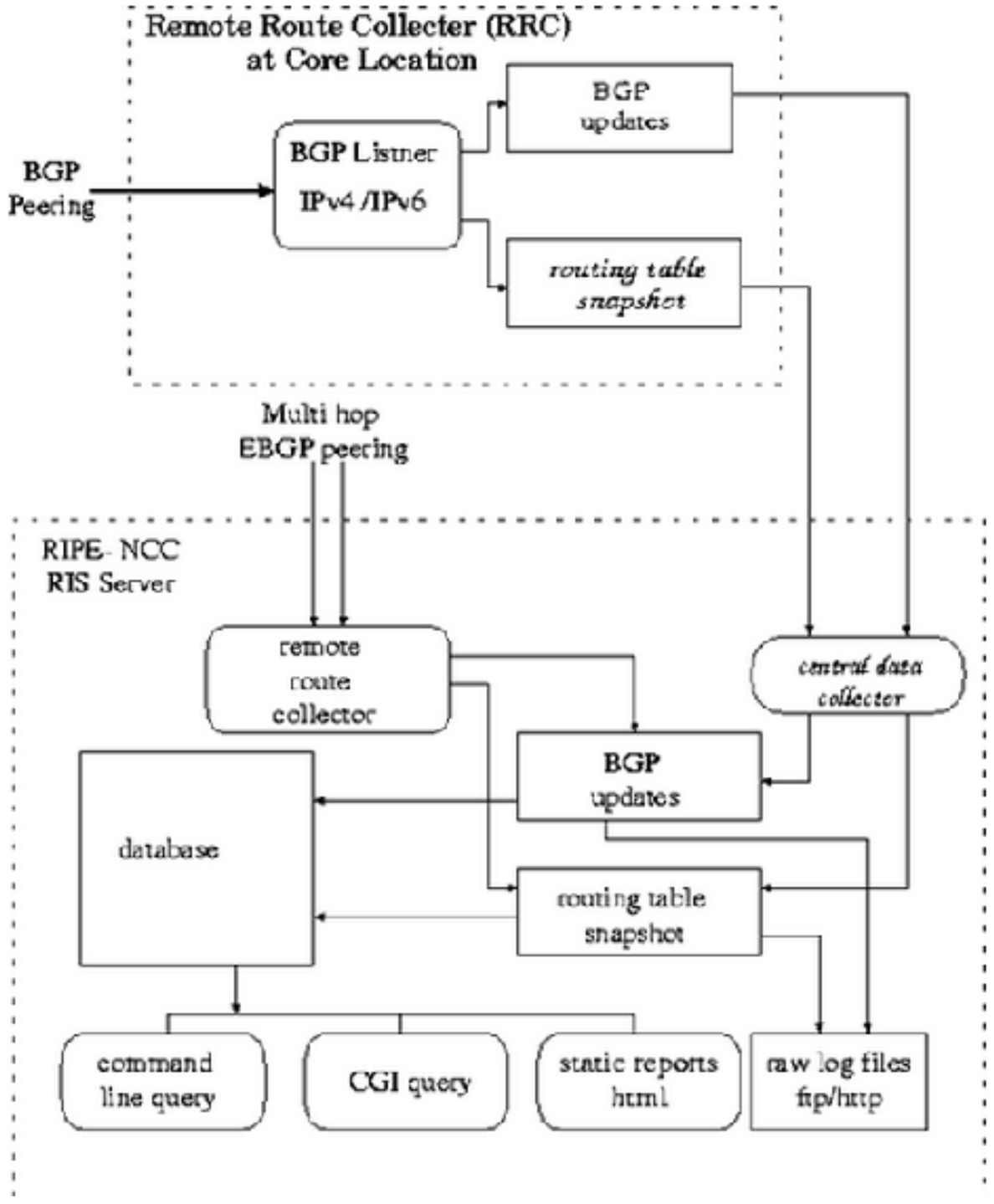
- More BGP routes
  - BGP table has grown from 60,000 to 600,000 routes
  - more BGP updates
  - larger RIB (table) dumps
- More RIS collectors
- More peers at each collector
- Non-linear growth curve ;)

# RIS Operations

As the system has evolved

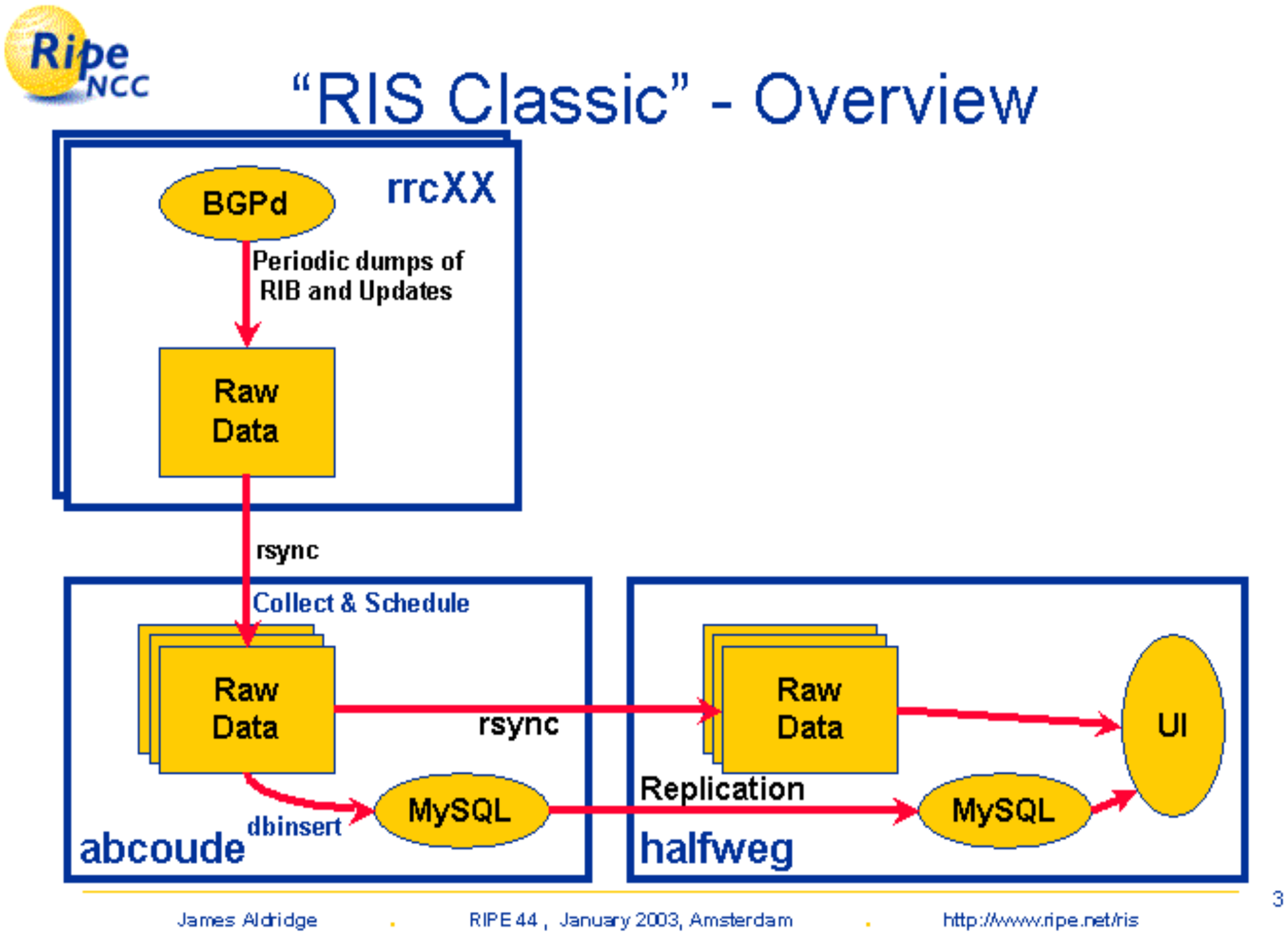
# Original architecture (1999)

- Diagram from RIPE-200 (original concept)
- Note ‘RIS Server’
  - singular!
- Also, the ‘database’
  - this becomes the hardest part!!




*Original RIS design (RIPE-200)*
*circa 1999*

# Classic architecture (~2003)



# Classic architecture

- MySQL database
  - many millions of BGP updates
  - hundreds of thousands of BGP routes, seen by hundreds of peers
  - route attributes (communities, AS_PATH, origin, med, etc)
  - ASN adjacencies
  - more/less specific matching
  - complex data schema

# Scaling MySQL

- Splitting and sharding
  - 8 MySQL servers
  - some collectors were so big they needed their own MySQL server!
- Data retention
  - database was only query-able for 3 months worth of data
  - the references grew too large, that every 3 months we basically had to drop all the data, and let it start again!!

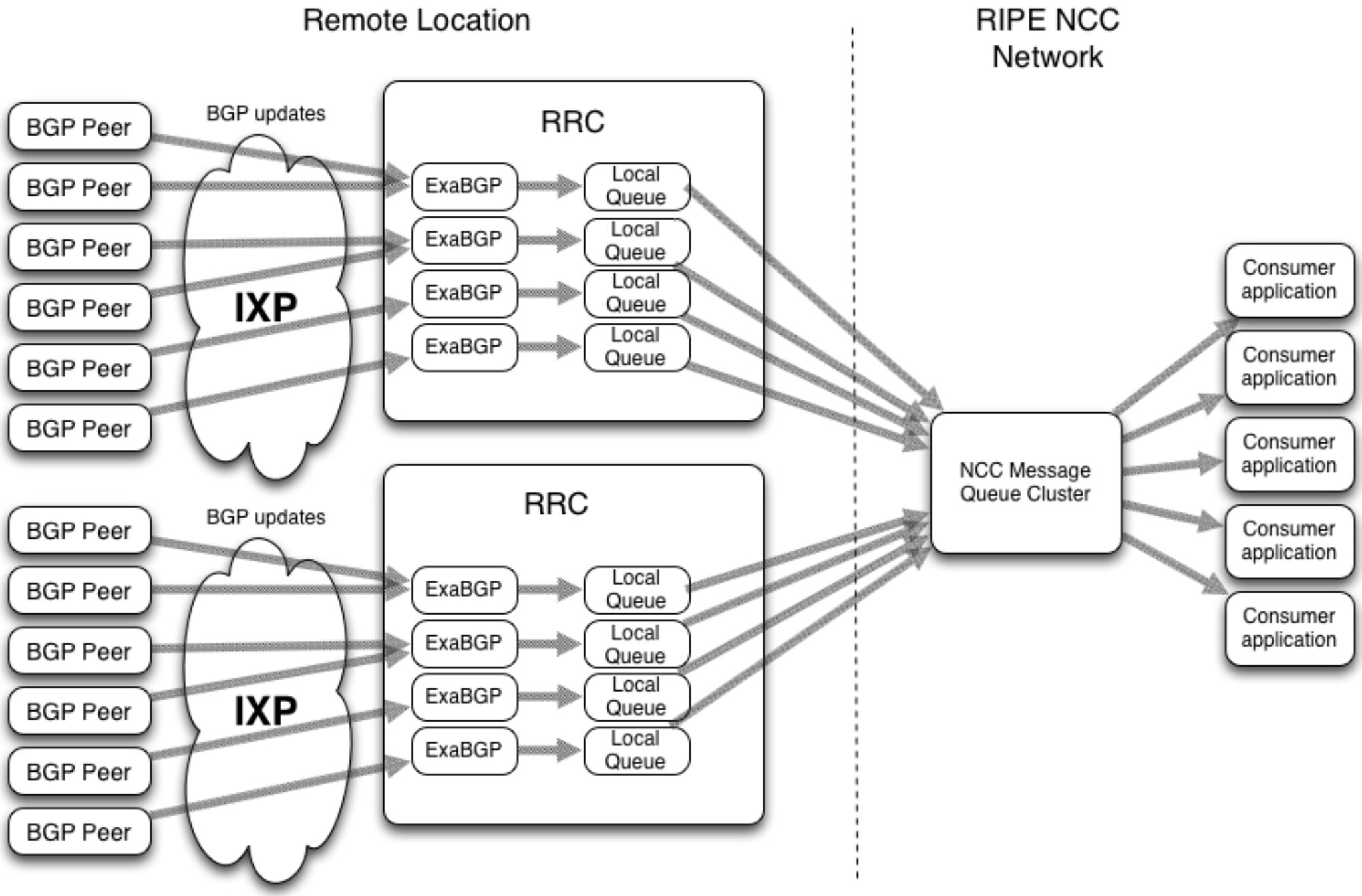
# Scaling the collectors

- Quagga used as BGP collector
- Single-threaded
  - Not as scalable on modern multi-core CPUs
- Locks updates during table-dump process
  - Requires that dump completes before the hold timer expires, or BGP session will drop
- Some data consistency issues
  - Sometimes updates are missing from the update dumps at the time of a table dump
  - This makes it difficult to accurately rebuild BGP state at a intermediate time, if updates are not reliable in-between

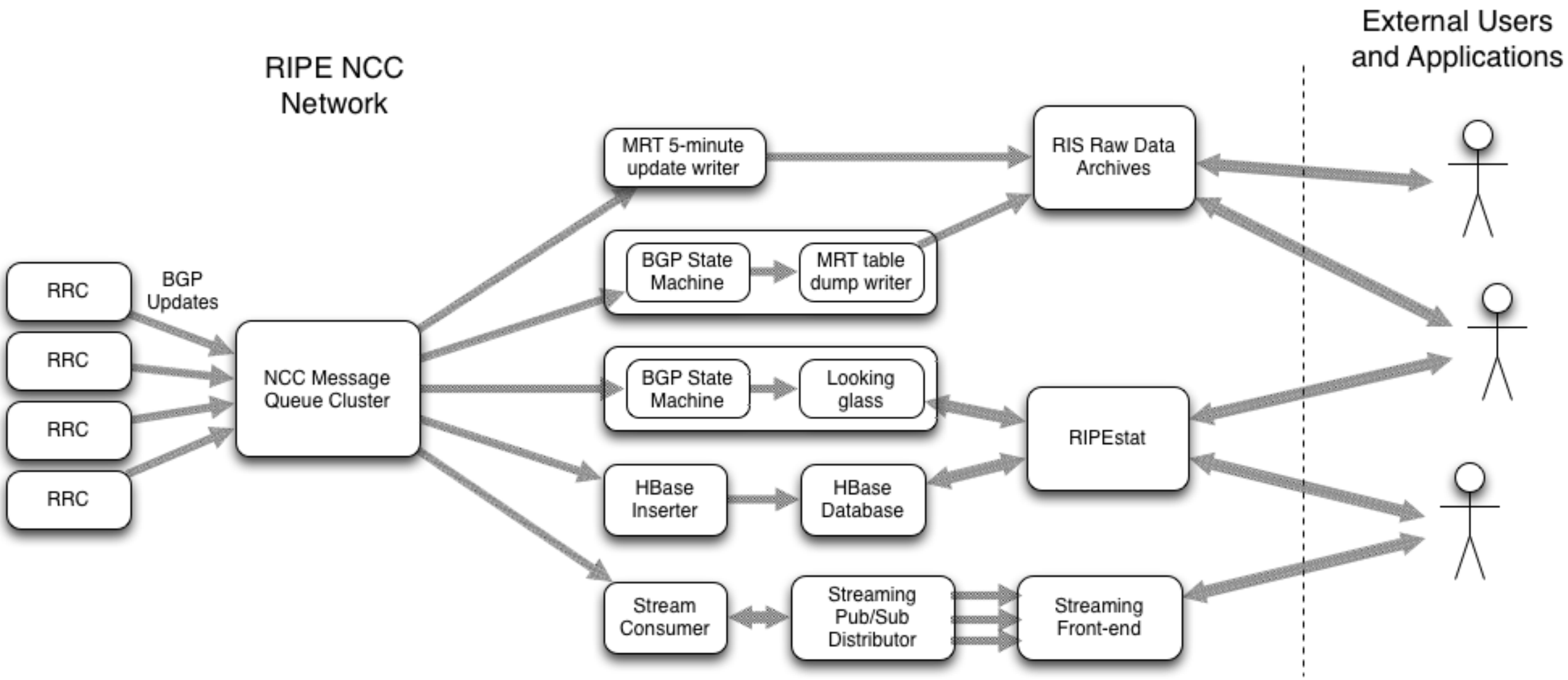
# RIS and Atlas Operations

Time for a redesign
(and this is the current design!)

# Data collection



# Back-end data distribution



# Data processing

- Apache Hadoop
  - An open-source software framework for distributed storage and distributed processing of very large data sets on computer clusters built from commodity hardware.
- “Big Data” storage and analytics
- Allows us to build a scalable storage and processing cluster
- Currently over 150 servers in the cluster!
  - Although the cluster is not only used for RIS!
  - Also used by RIPE Atlas and other projects

# Data processing - components

- HDFS
  - distributed, replicated, cluster filesystem
- YARN
  - compute resource manager and application scheduler
- Map/Reduce
  - massive batch job processing
- HBase
  - non-relational distributed database
  - large tables - billions of rows X millions of columns

# Data processing - components

- Spark
  - Cluster computing used for data stream processing
  - i.e. non-batch computing
- Azkaban
  - batch workflow job scheduler, dependency tracking etc
- Kafka
  - BGP/Atlas messaging bus

# Data processing - RIS

- Raw data inputs:
  - BGP updates events - everything must start from a BGP message!
  - BGP table dumps (which can also be derived from updates)
- Derived datasets
  - update-counts, first-last-seen, prefixes-transited-by-asn, peers-list, asn-stats, asn-adjacencies
  - country-code mapping
  - aggregated counts for historical overviews
  - distributed looking-glass processing

# Data processing - Atlas

- Raw data inputs:
  - Atlas measurement messages!
- Derived datasets
  - Mainly aggregated counters for different zoom levels
  - Probe traffic stats
  - Ping measurement stats (loss %, min, max, med etc)
  - DNS query stats (loss %, latency, etc)

# Questions

cpetrie@ripe.net