RIPE NCC
RIPE NETWORK COORDINATION CENTRE

RIPE NCC Routing Information Service (RIS)

2017 Update

Oleg Muravskiy | ENOG 17 | October 2017

What is RIS?

What is RIS?

- Routing Information Service
- Worldwide network of BGP collectors
- Deployed at Internet Exchange Points
- Collects raw BGP data from peers
- Stores BGP updates and routing table dumps
- 15+ years of history
- Used by network operators and researchers every day

Collector locations

- 18 collectors and growing
- 600+ peers
- 150+ full-feed peers

Why RIS?

Why are we doing this?
A bit of history

Why RIS?

- Original project was defined in RIPE-200 in 1999:

  “In other words, it can be regarded as one integrated Looking-Glass for the entire Internet that includes history information”

- Looking glasses are instantaneous
- Routing problems are also instantaneous
- BGP history is 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!

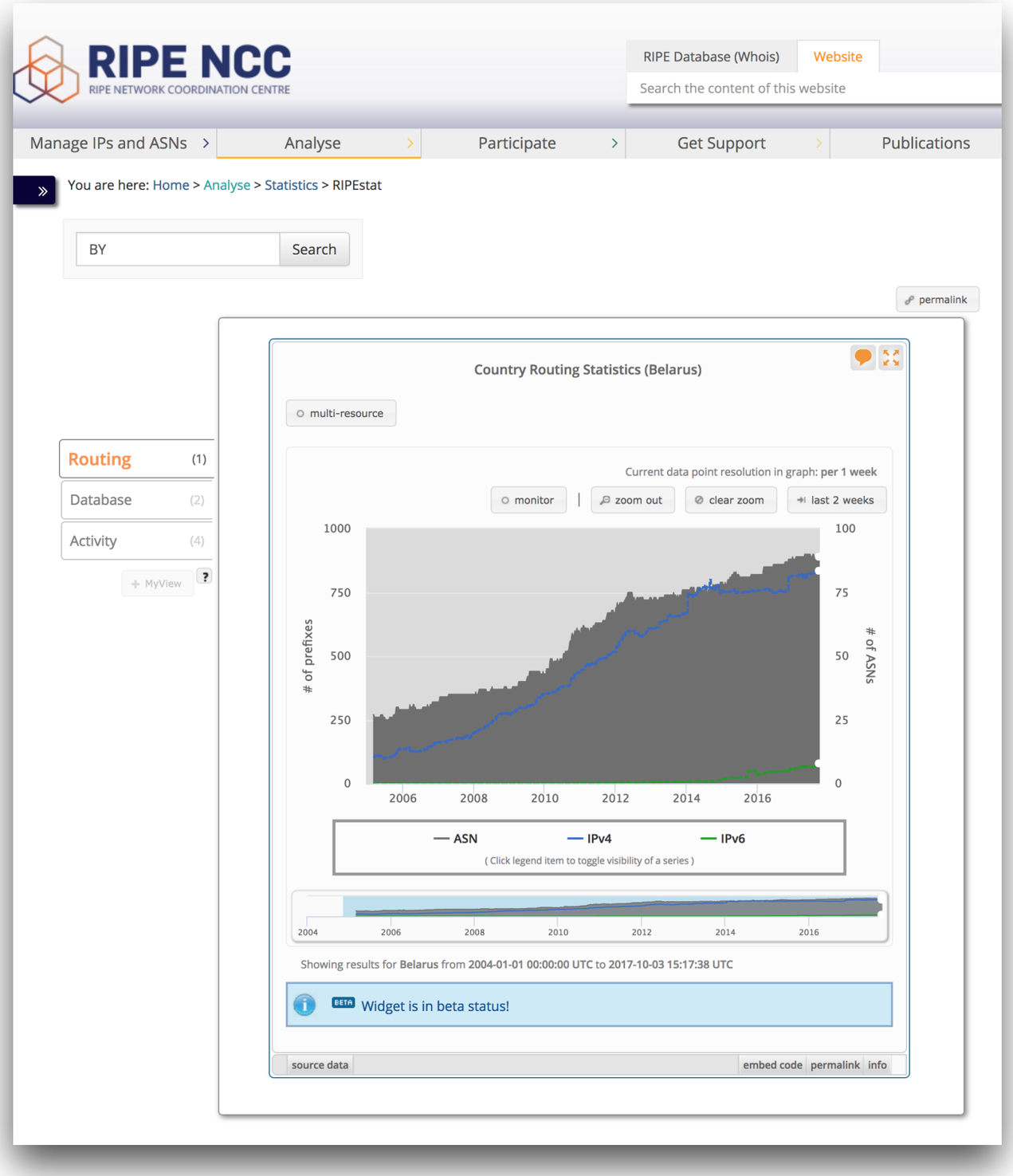
- 15+ years of raw data (5.8 TB) available to download and analyse yourself :)
  - https://www.ripe.net/analyse/internet-measurements/routing-information-service-ris/ris-raw-data
- Data stored in MRT (RFC6396) format
- Readable using BGPdump utility
  - open source, maintained by RIPE NCC
  - https://bitbucket.org/ripencc/bgpdump
- …and by other tools

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!

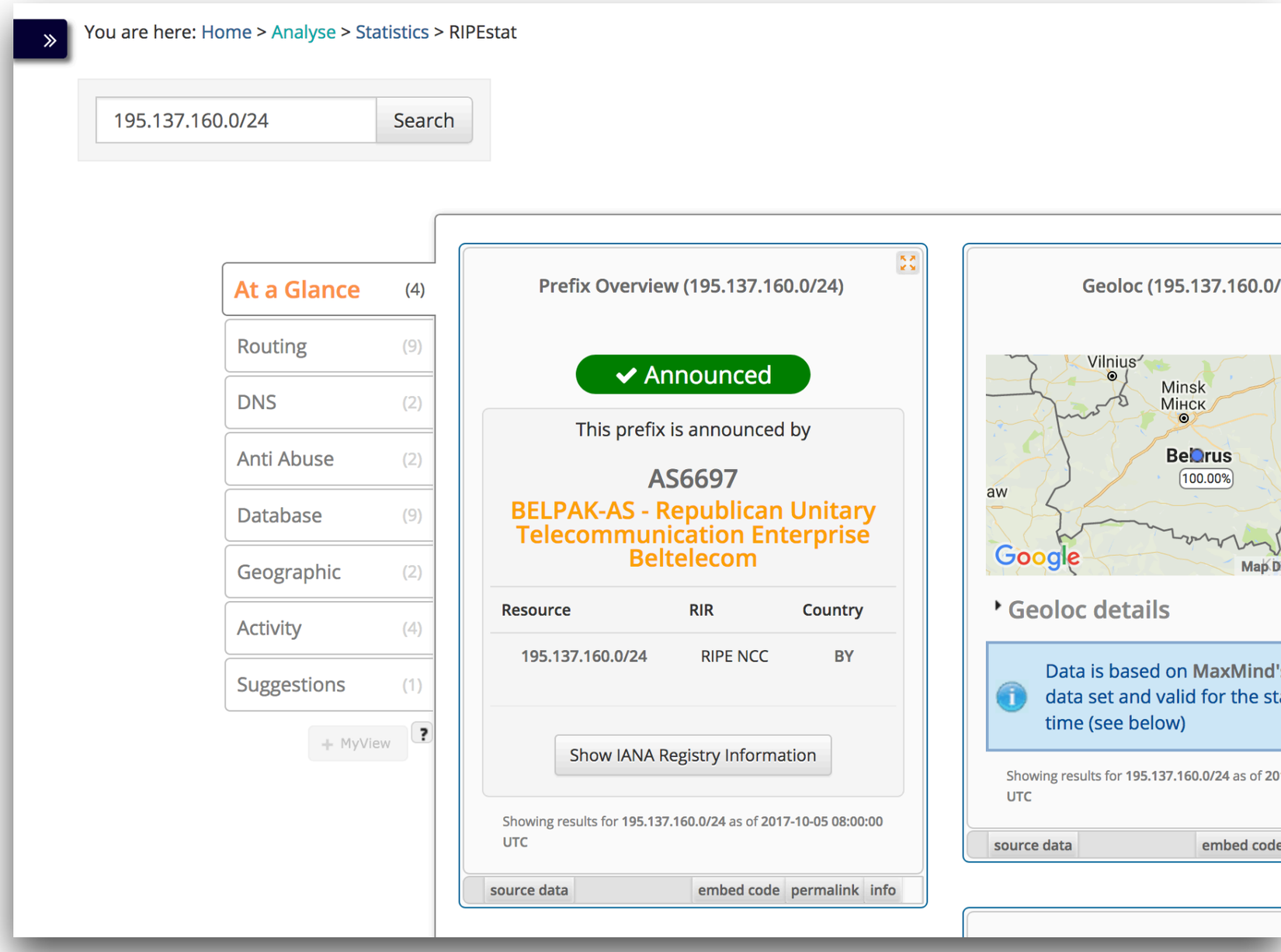
RIPEstat

- RIPEstat (stat.ripe.net) 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

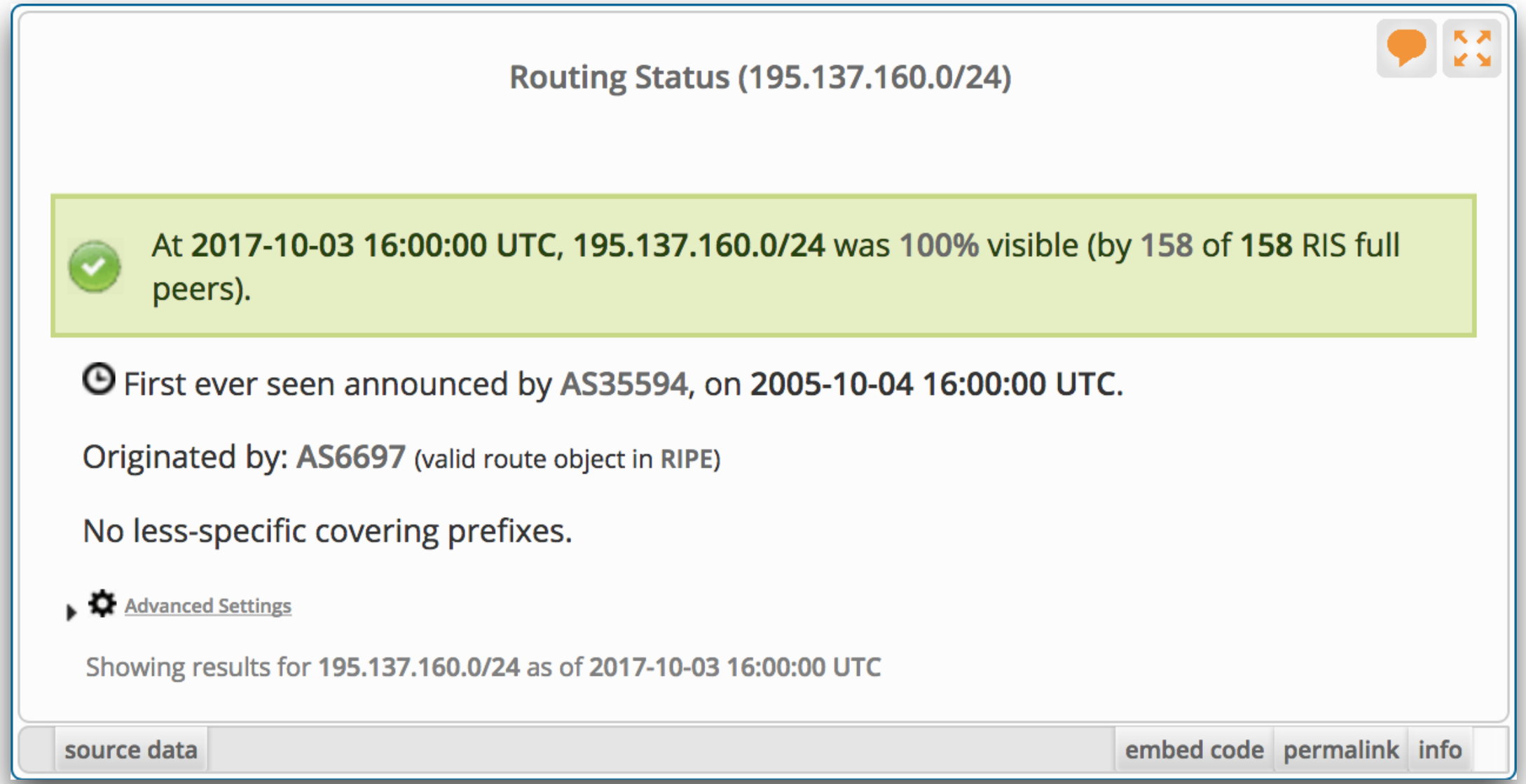


RIPEstat

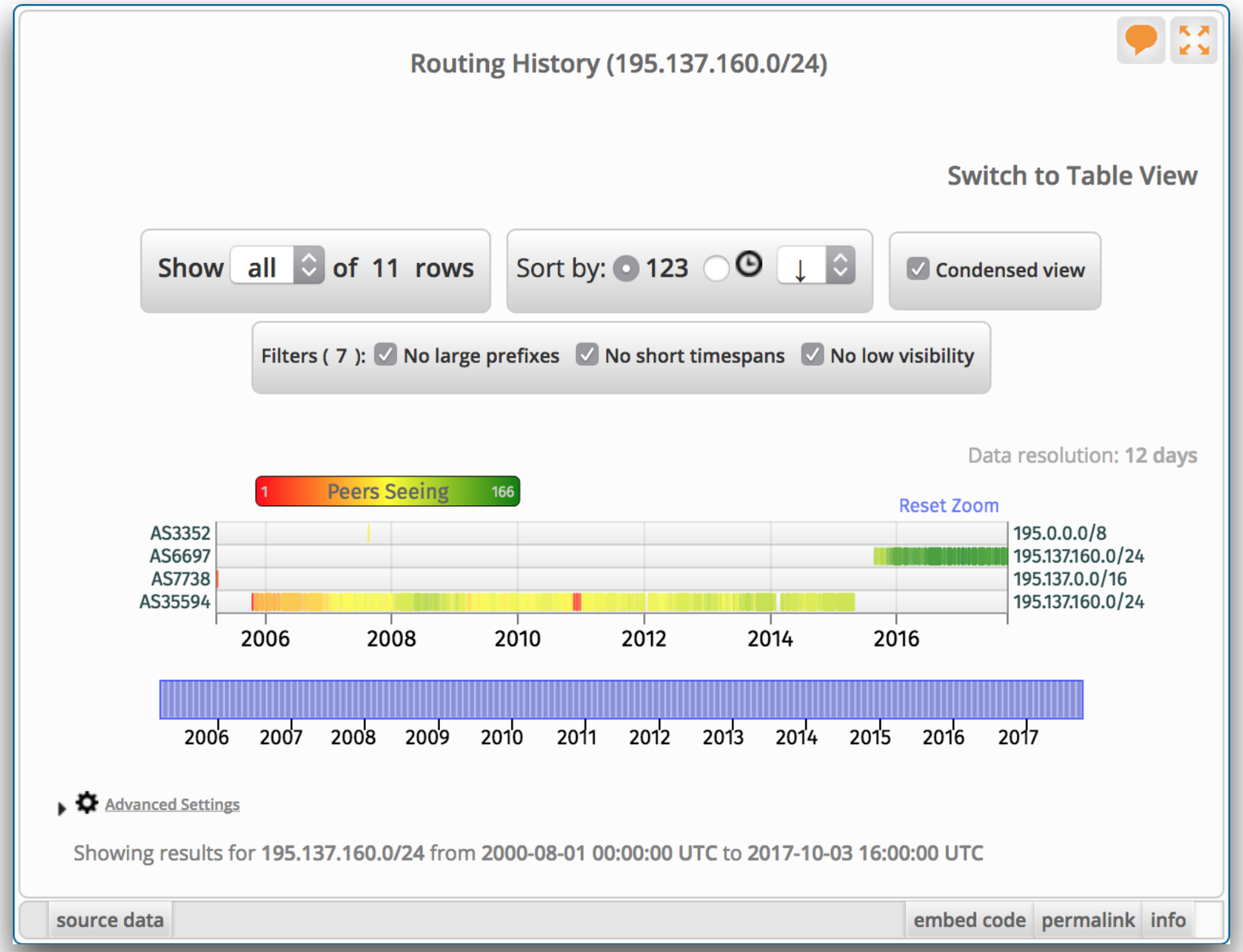
- What can you search for?
  - ASN
  - IPv4 address
  - IPv4 prefix
  - IPv6 address
  - IPv6 prefix
  - country



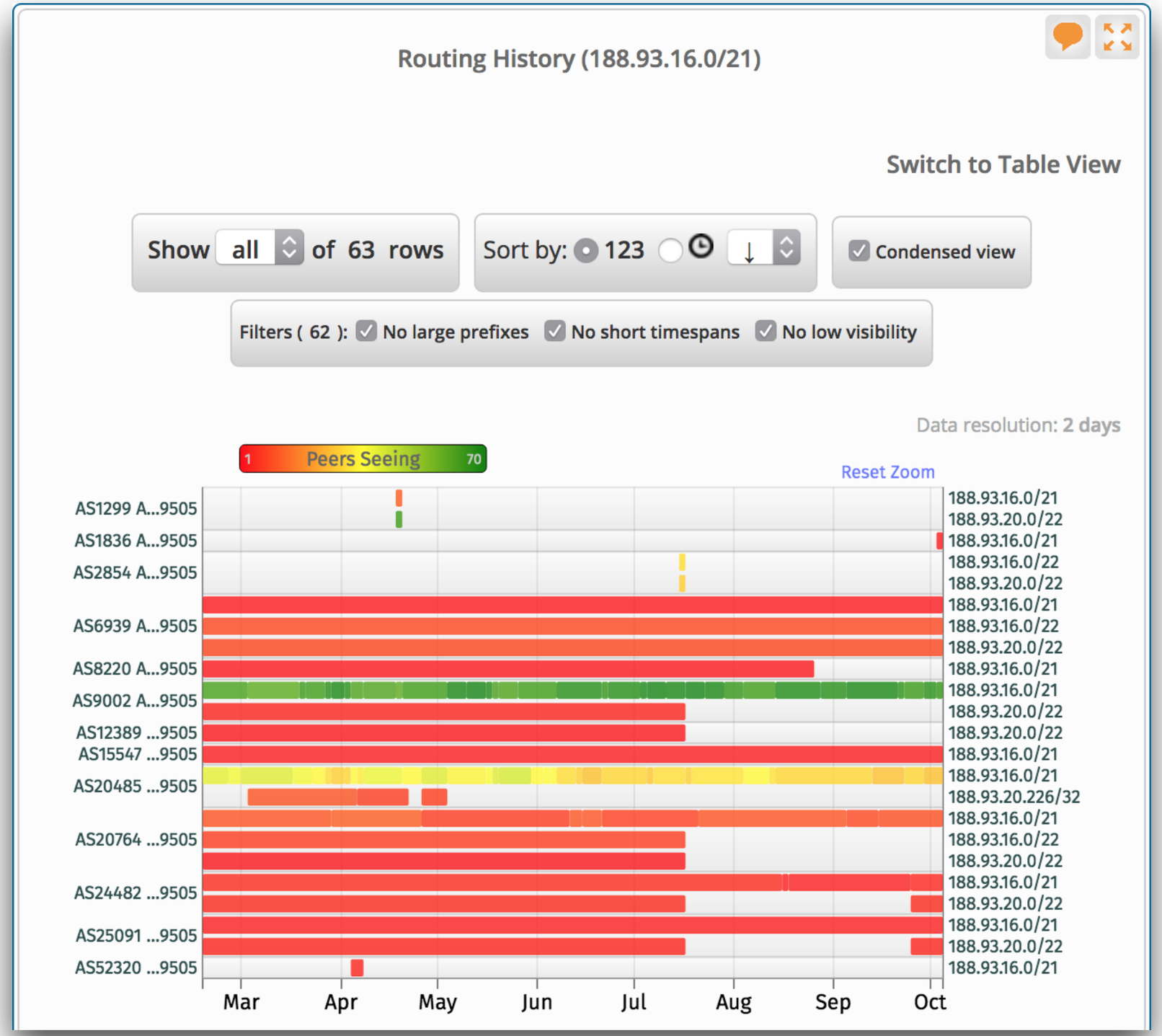
RIPEstat Examples



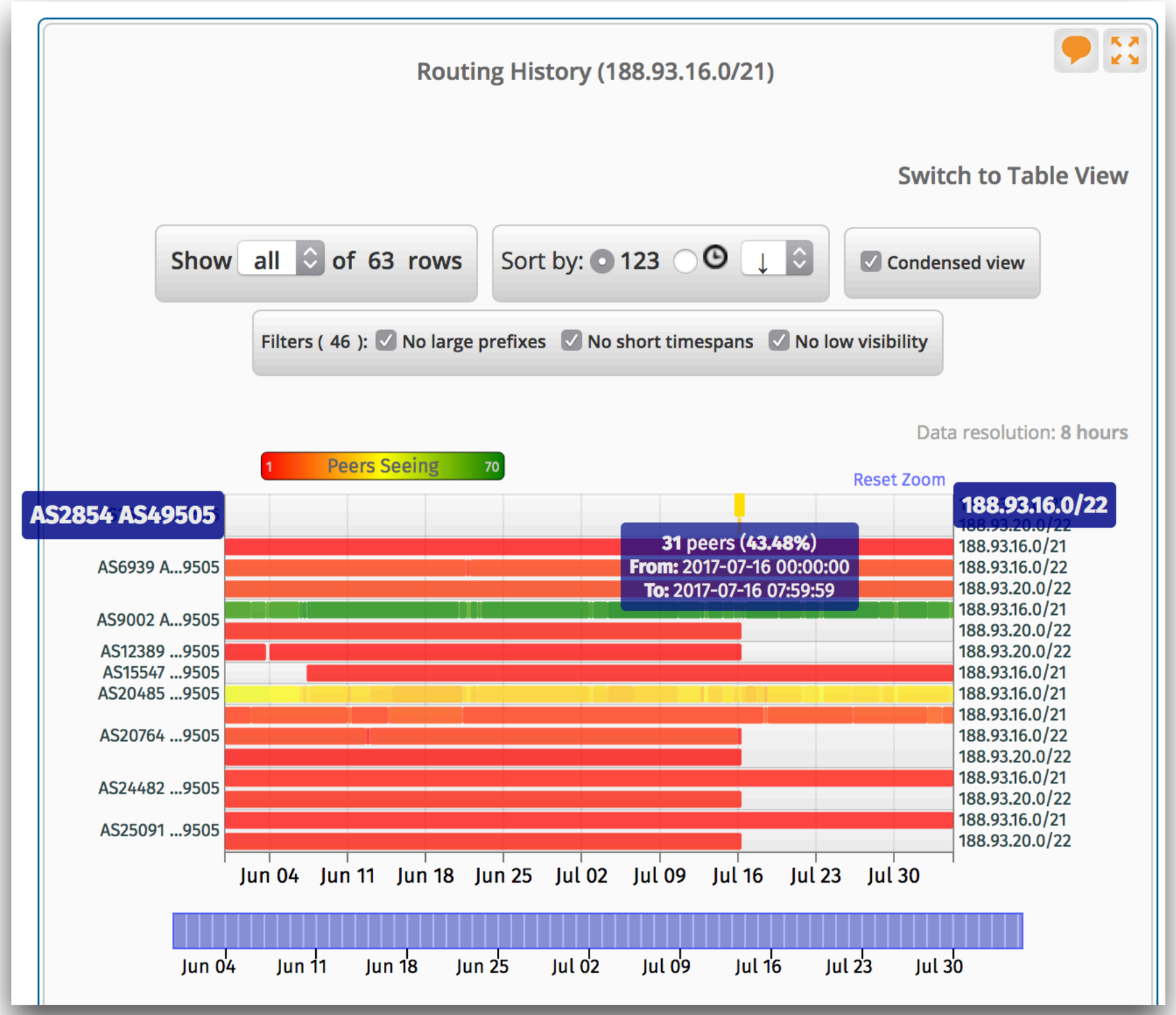
RIPEstat Examples



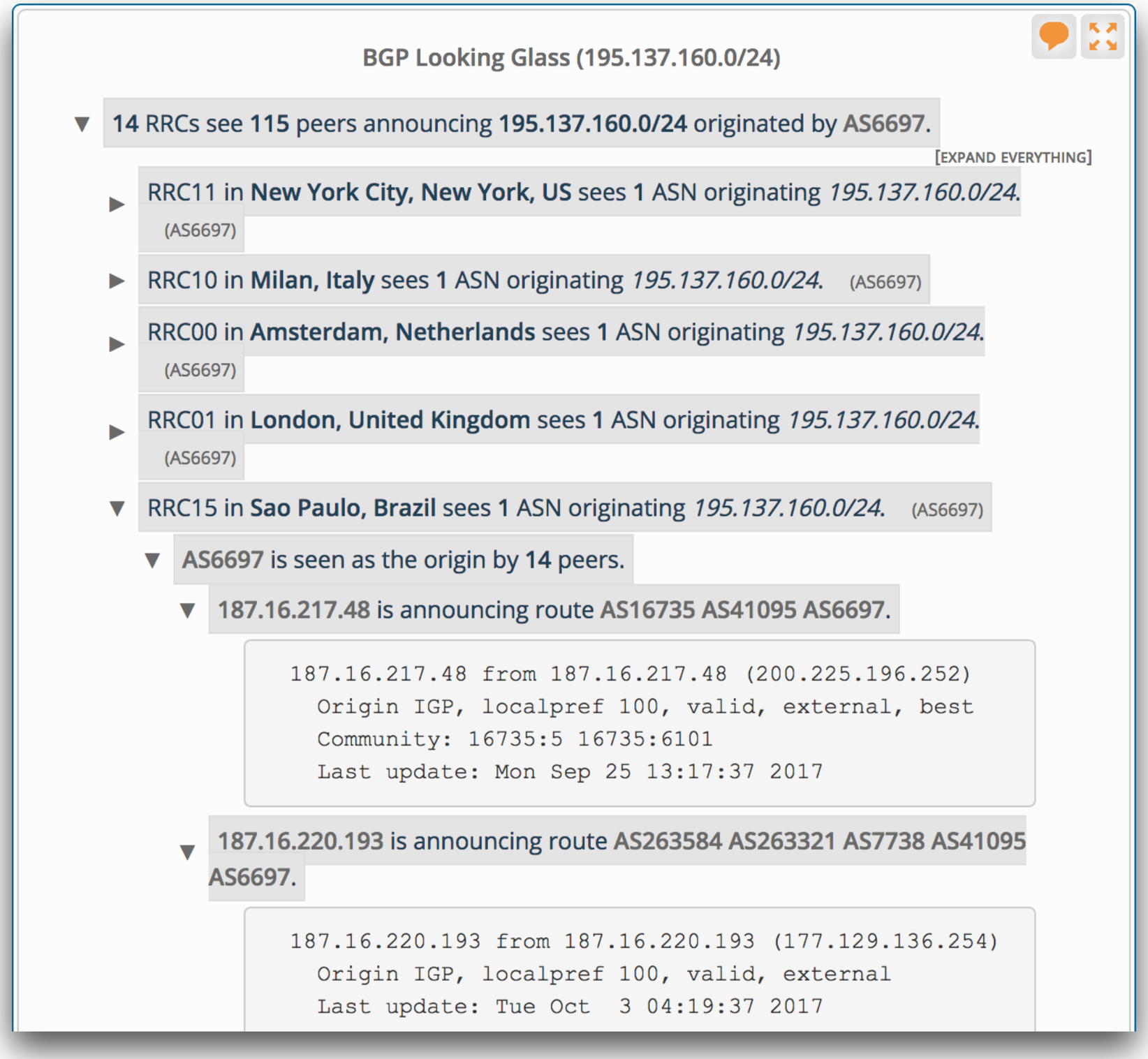
RIPEstat Examples: Selectel case



RIPEstat Examples: Selectel case



RIPEstat Examples



RIPEstat Examples

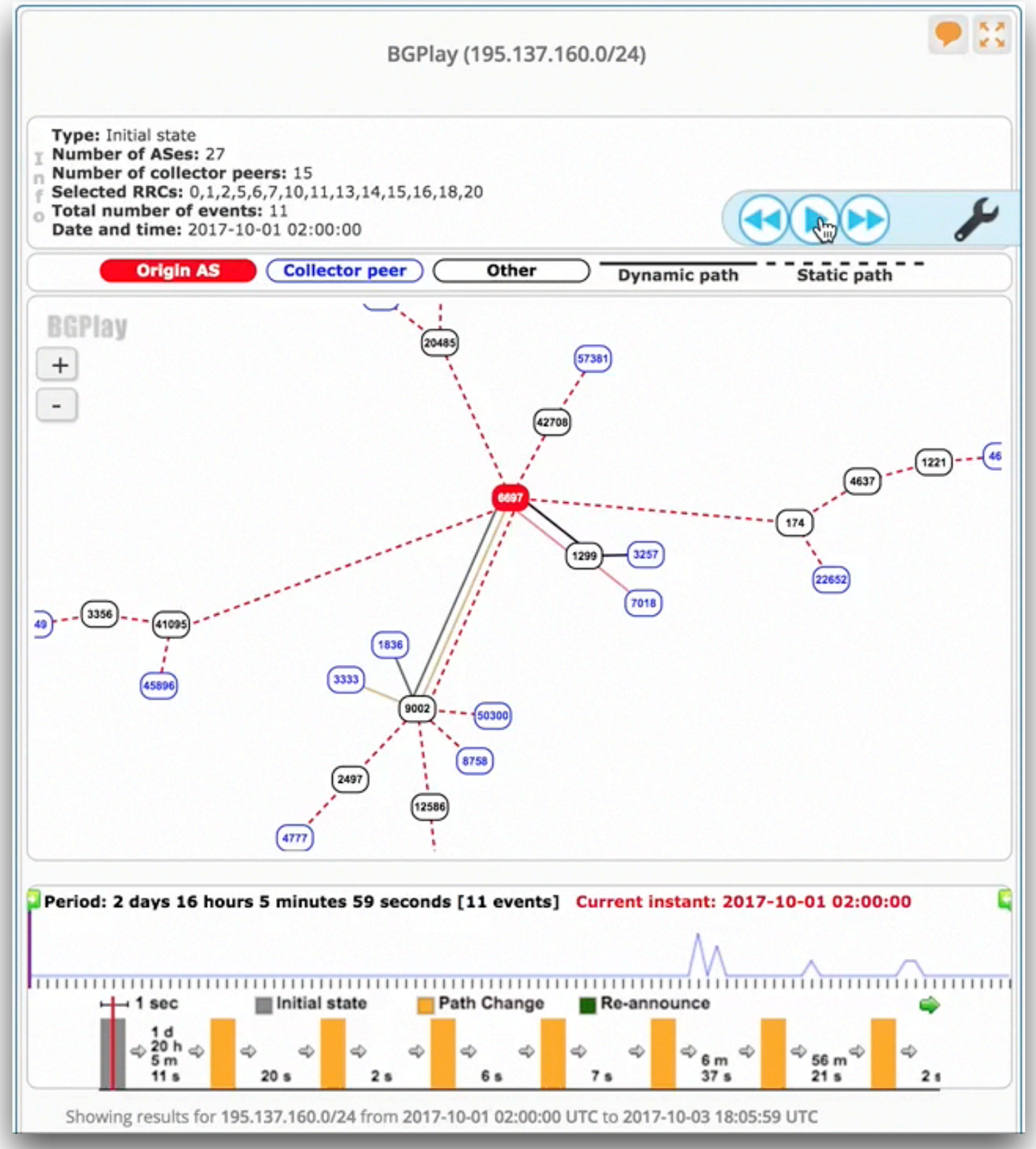
Visibility (195.137.160.0/24)

195.137.160.0/24 is visible by 100% of 158 IPv4 RIS full peers.

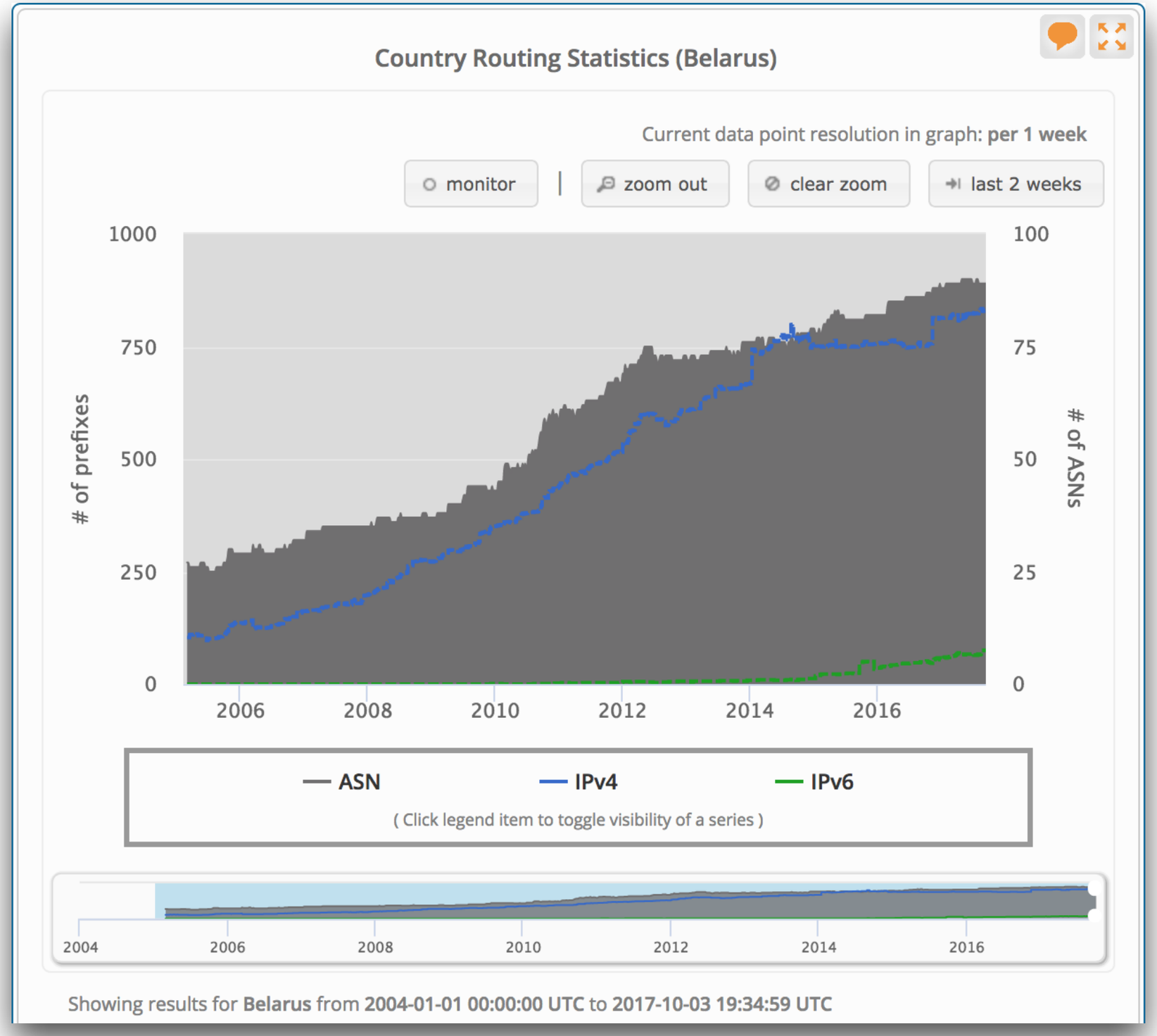
Visibility Location Details of 195.137.160.0/24

RRC	IXP Location	Location	IPv4 peers seeing	IPv4 Visibility
RRC00	RIPE-NCC Multihop	Amsterdam, Netherlands	15 of 15	100%
RRC01	LINX	London, United Kingdom	7 of 7	100%
RRC03	AMS-IX / NL-IX	Amsterdam, Netherlands	7 of 7	100%
RRC04	CIXP	Geneva, Switzerland	8 of 8	100%
RRC05	VIX	Vienna, Austria	3 of 3	100%
RRC06	DIX-IE	Tokyo, Japan	2 of 2	100%
RRC07	Netnod	Stockholm, Sweden	5 of 5	100%
RRC10	MIX	Milan, Italy	9 of 9	100%
RRC11	NYIIX	New York City, US	9 of 9	100%
RRC12	DE-CIX	Frankfurt, Germany	12 of 12	100%
RRC13	MSK-IX	Moscow, Russian Federation	11 of 11	100%
RRC14	PAIX	Palo Alto, US	7 of 7	100%
RRC15	PTTMetro	Sao Paulo, Brazil	14 of 14	100%

RIPEstat Examples



RIPEstat Examples



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 – Selectel case

- Show me who and how was announcing this prefix at that time
- https://stat.ripe.net/data/bgp-state/data.json?resource=188.93.16.2&timestamp=2017-07-16T00:51:23Z


```
"data": {
    "bgp_state": [
        {
            "source_id": "00-178.255.145.243",
            "path": [50304, 42708, 2854, 49505],
            "community": [],
            "target_prefix": "188.93.16.0/22"
        },
        {
            "source_id": "00-193.0.0.56",
            "path": [3333, 1136, 24785, 24785, 24785, 24785, 20562, 2854, 49505],
            "community": [],
            "target_prefix": "188.93.16.0/22"
        },
        ...
],
"query_time": "2017-07-16T00:51:23",
"resource": "188.93.16.0/22",
"nr_routes": 44
```

Live stream demo

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

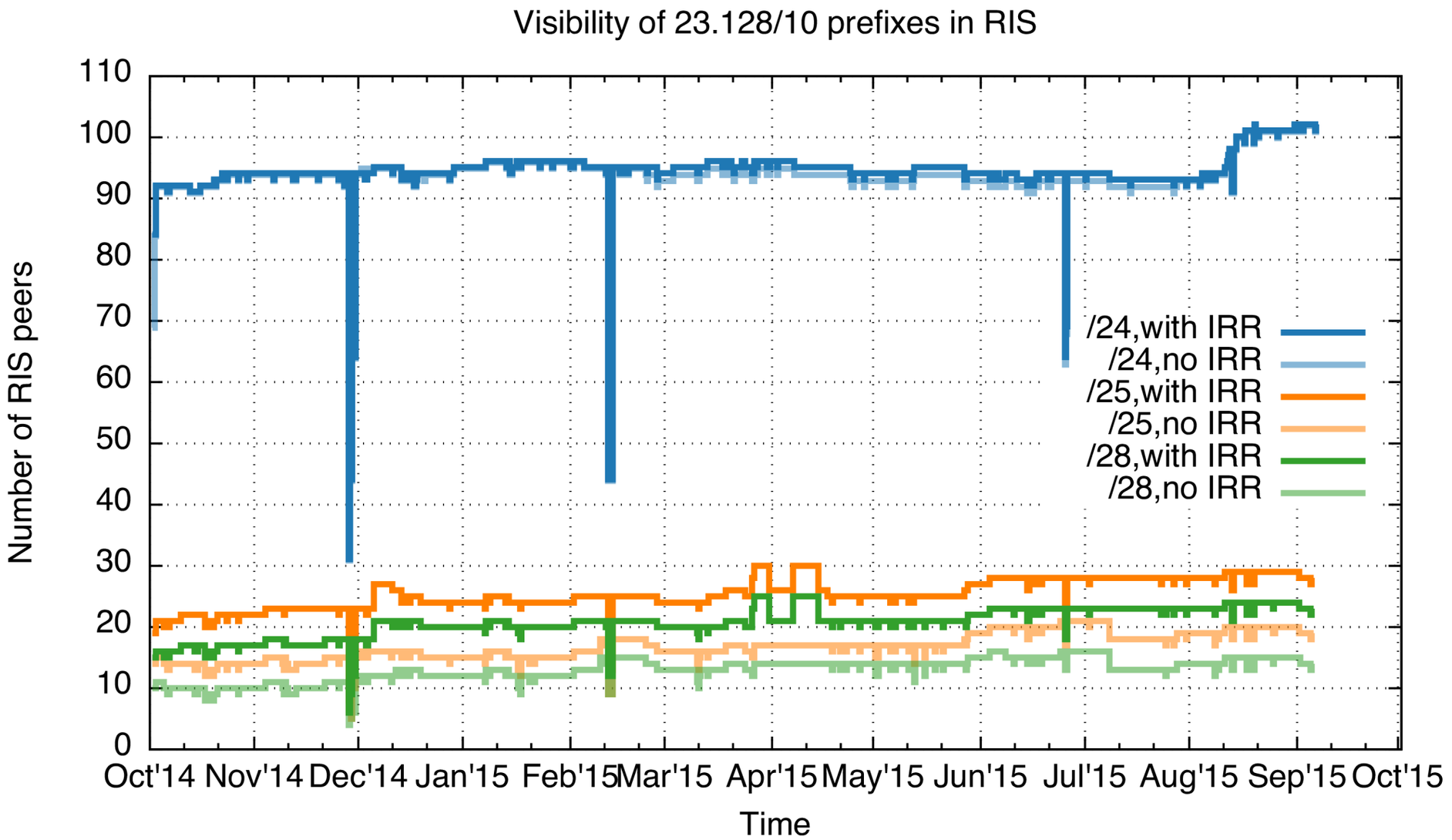
- Live stream enables new applications
  - BGP hijack detection
  - real time anomaly analysis
  - live monitoring of your routes

What else can you do?

Lots of analysis that this data allows

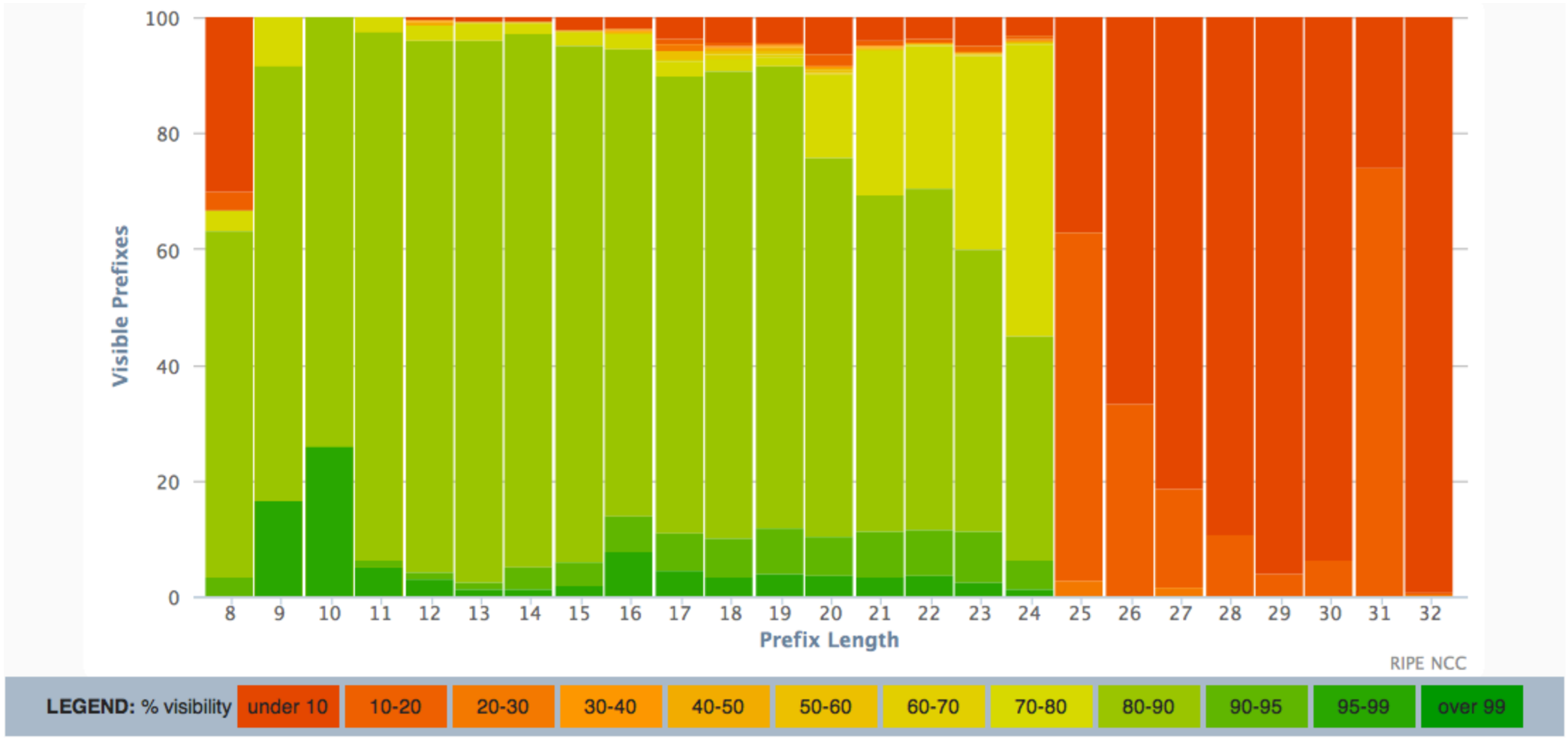
Prefix reachability studies

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



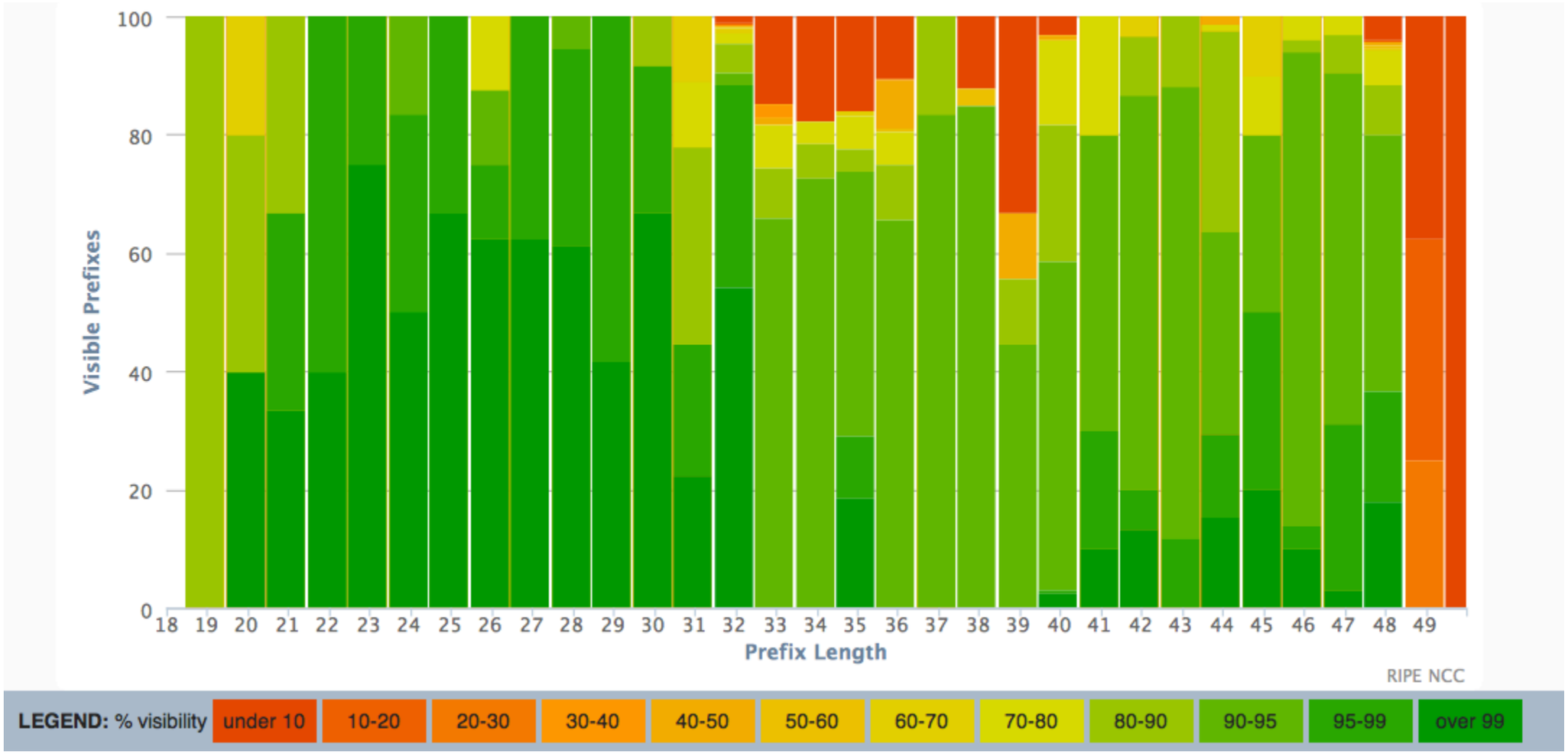
Prefix length visibility in IPv4

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



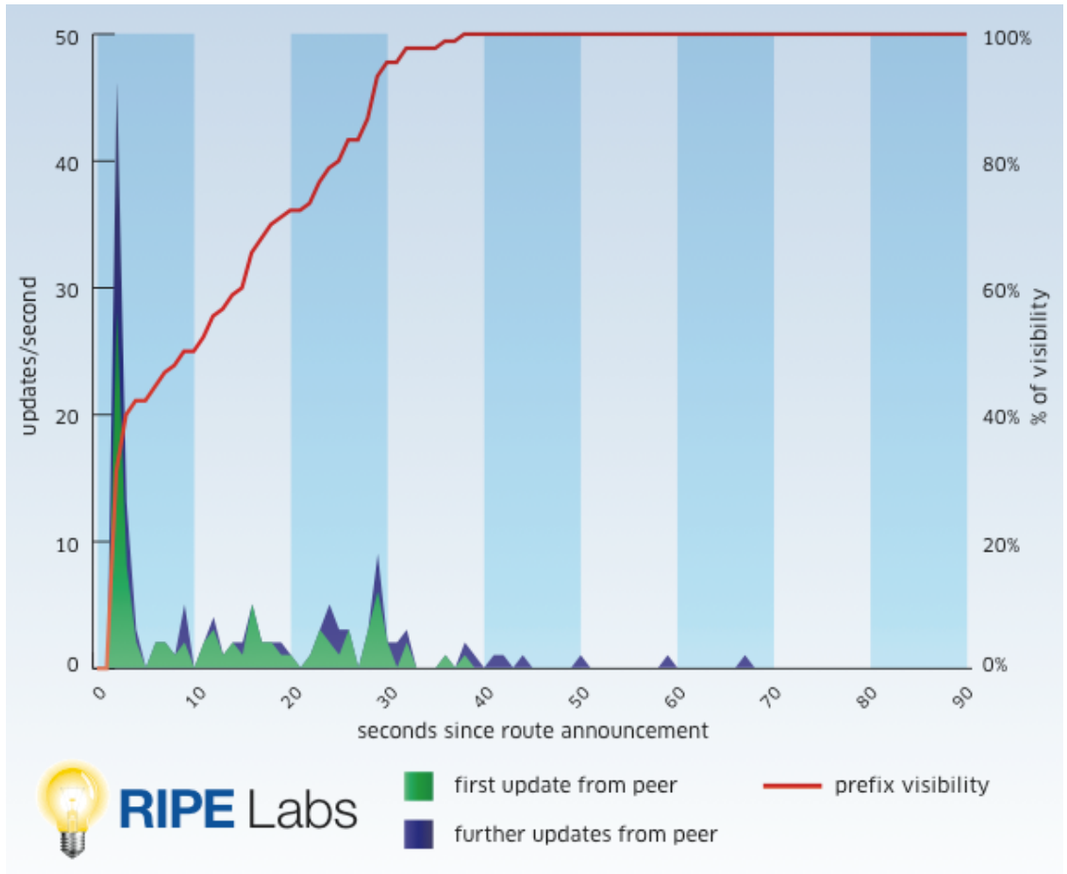
Prefix length visibility in IPv6

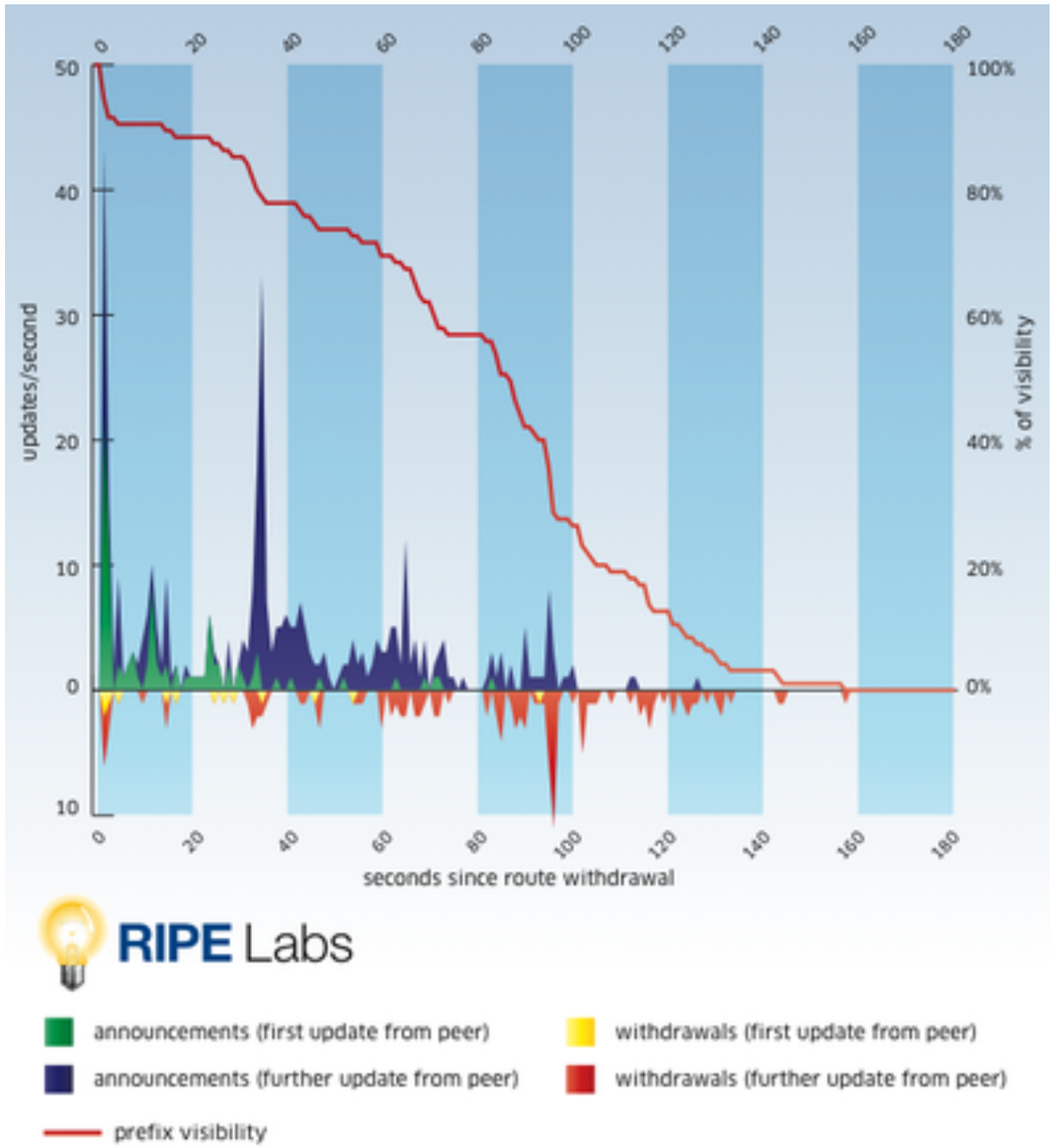
- https://labs.ripe.net/Members/dbayer/visibility-of-prefix-lengths



BGP update propagation

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





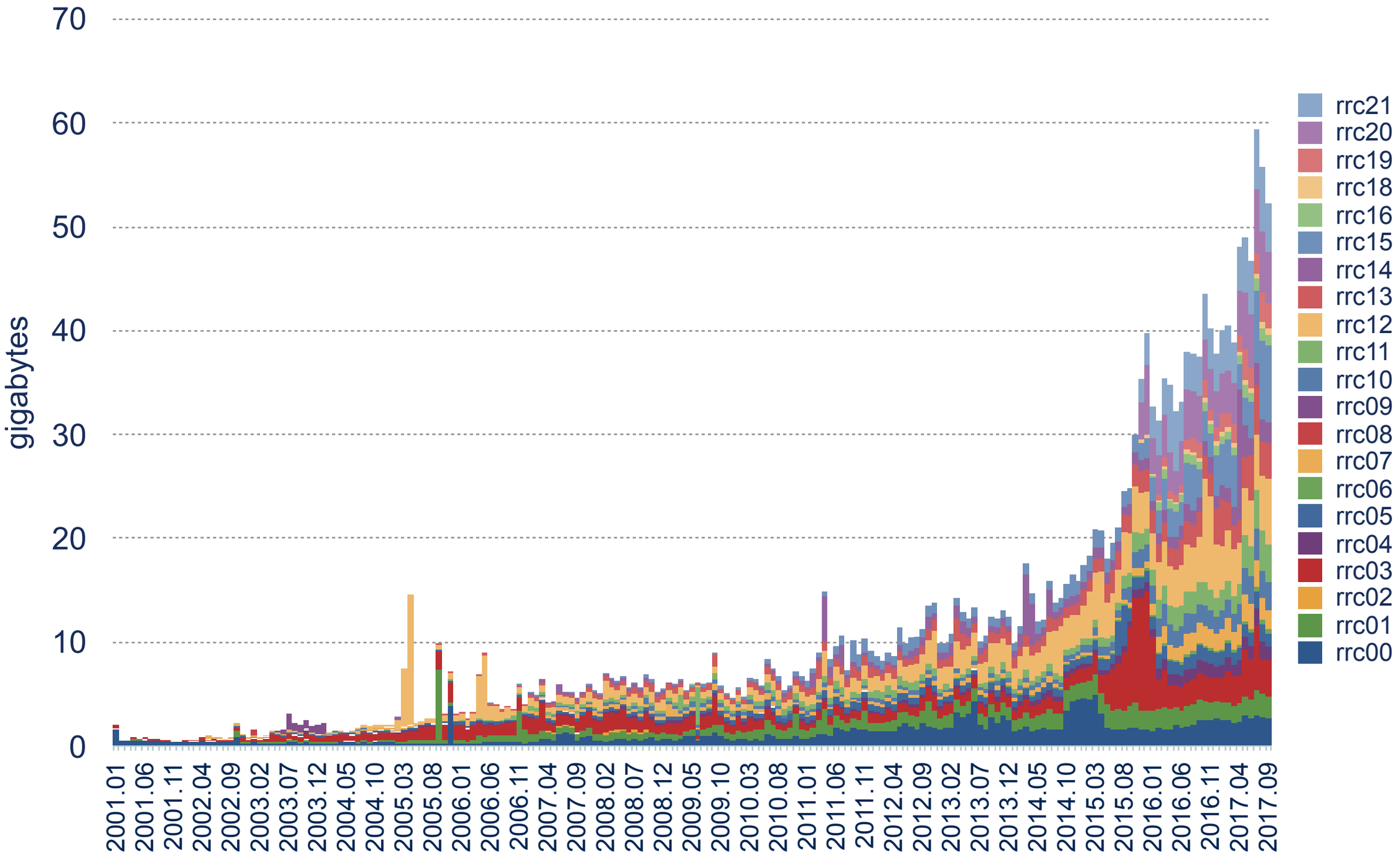
RIS growth

Because the internet keeps growing

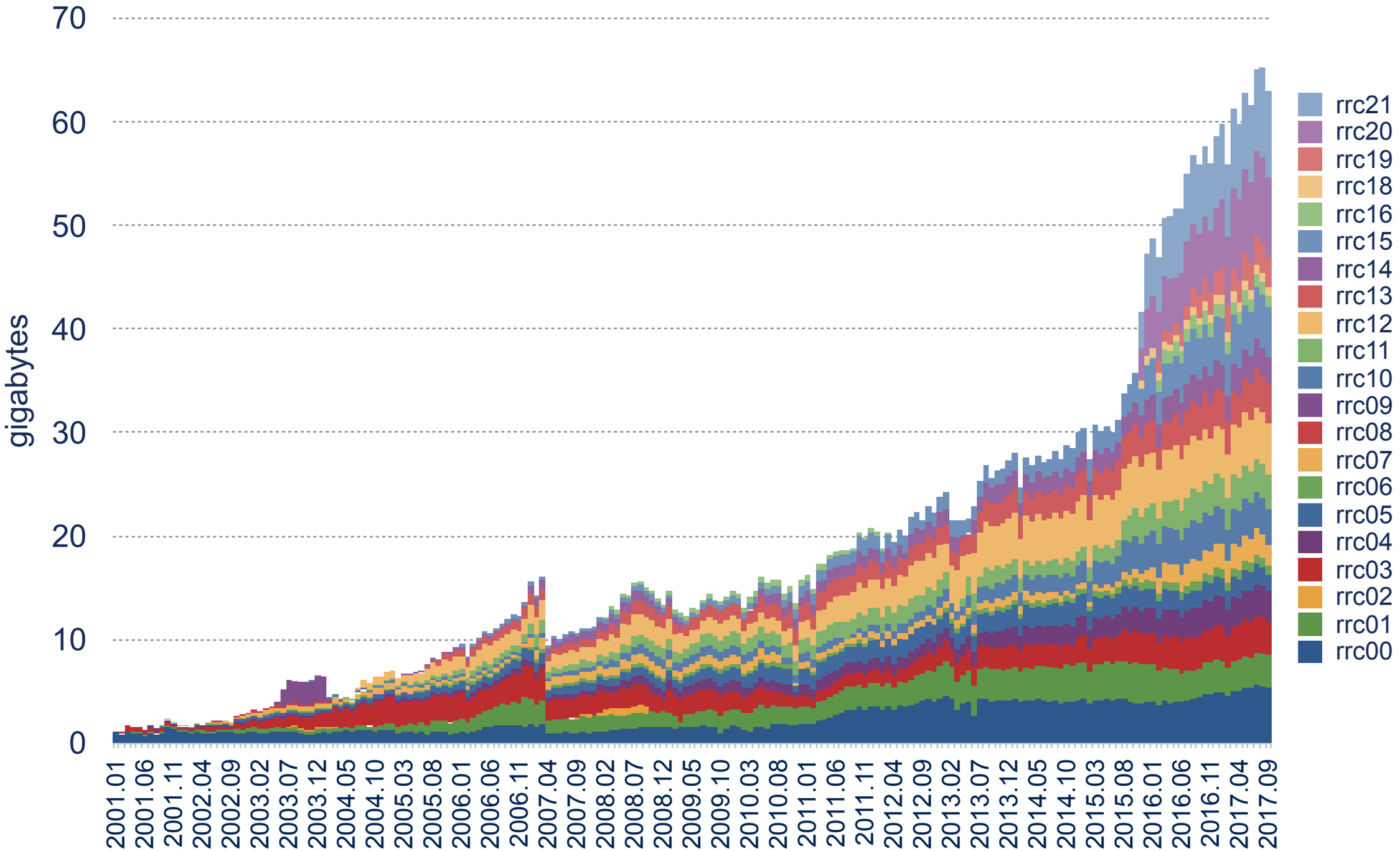
Collector history

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	-
RRC17				
RRC19	Johannesburg	NAPAfrica JB	2016	-
RRC20	Zurich	SwissIX	2015	-
RRC21	Paris	FranceIX	2015	-
RRC22	Bucharest	InterLAN	coming soon	
RRC23	Singapore	Equinix SG	coming soon	

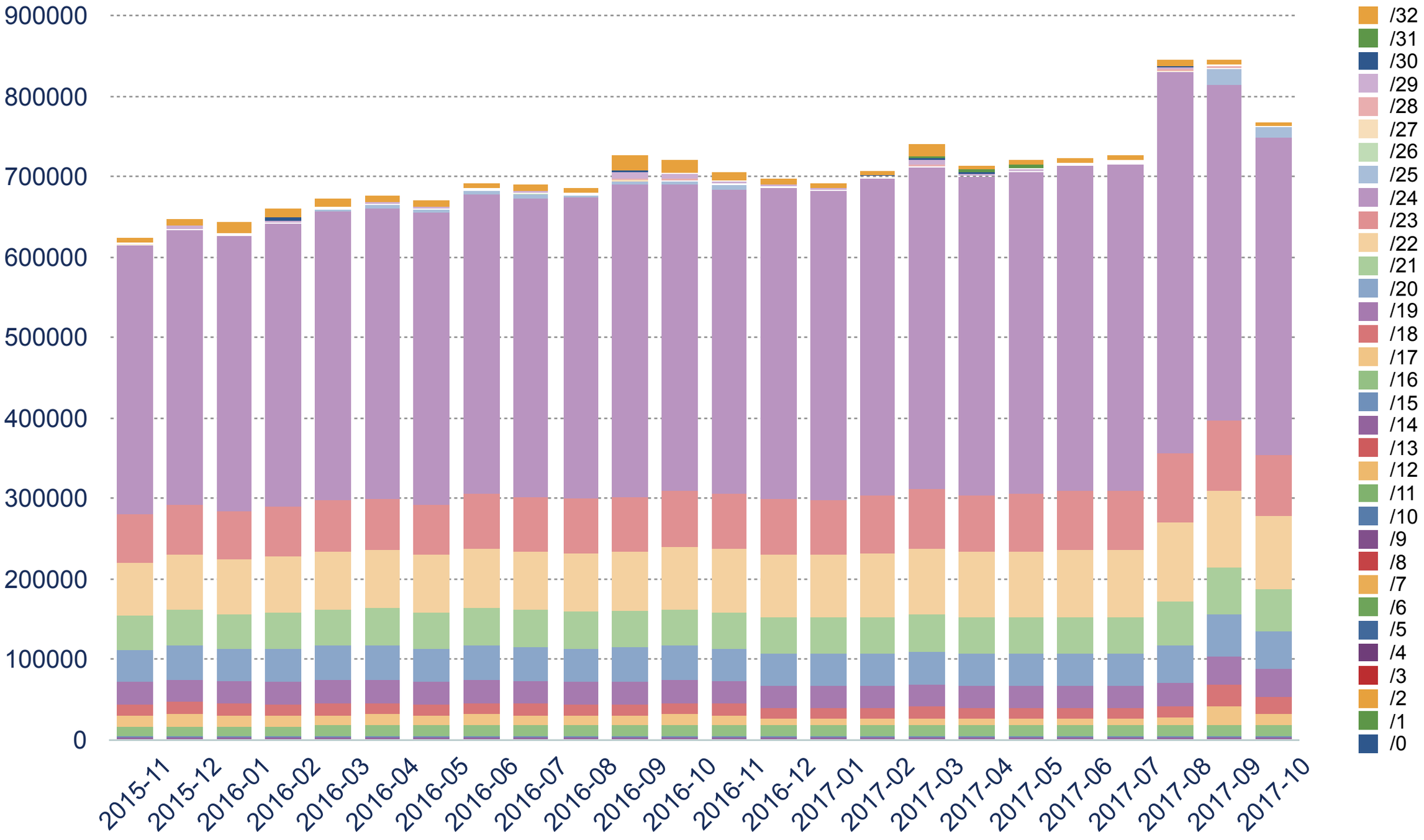
Size of compressed raw BGP updates per month



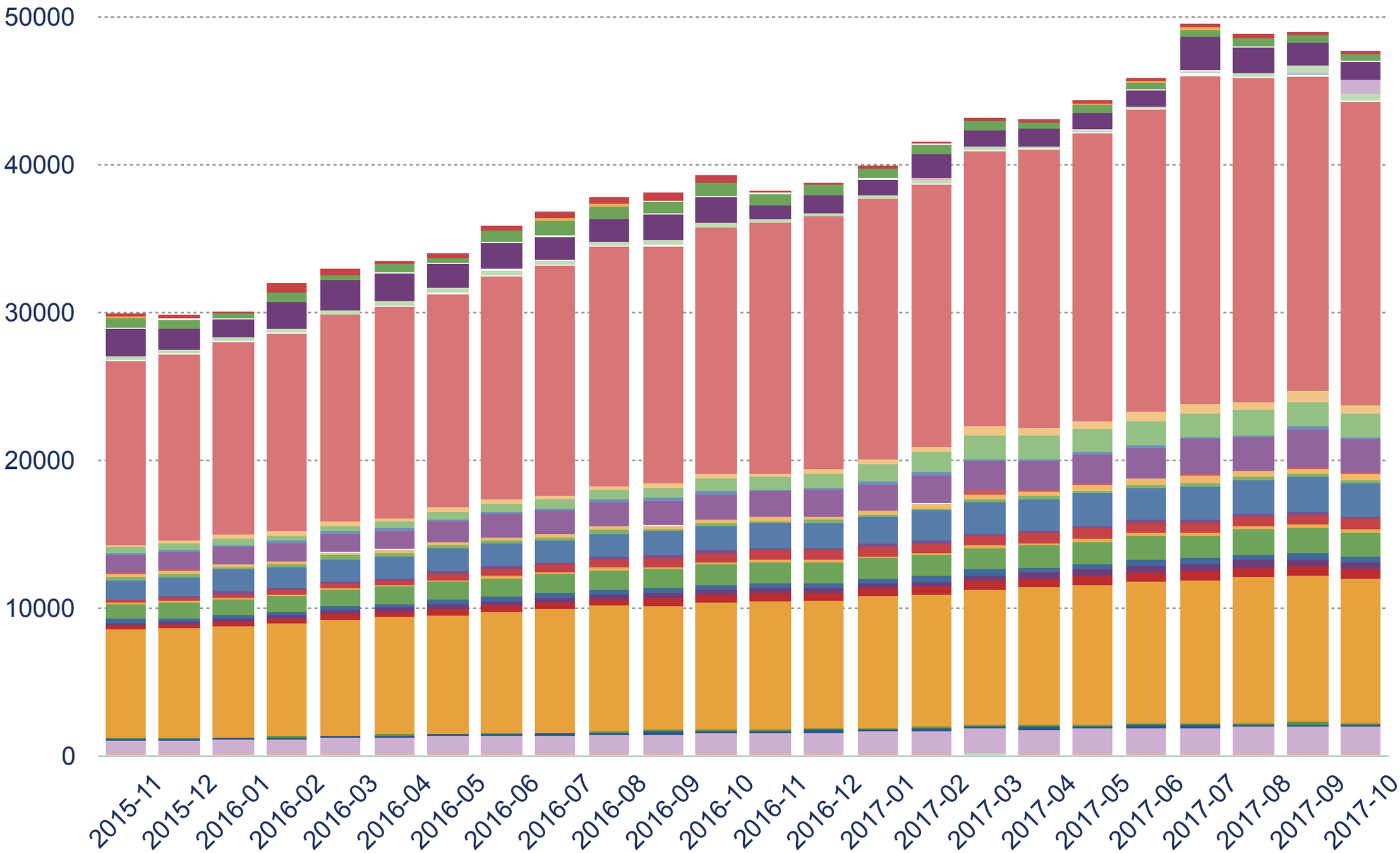
Size of compressed BGP dumps per month



Number of IPv4 prefixes seen



Number of IPv6 prefixes seen



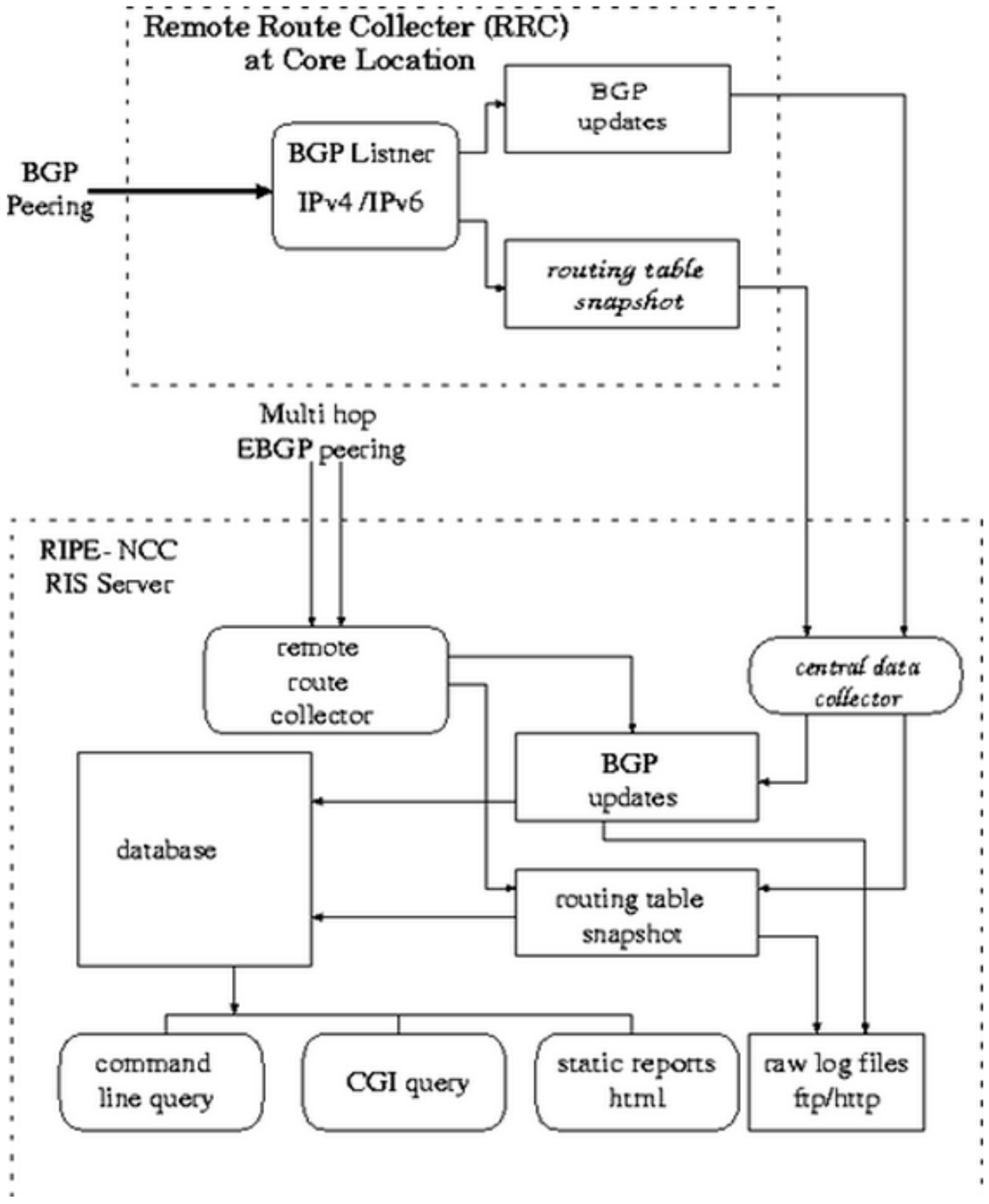
Data growth

- 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

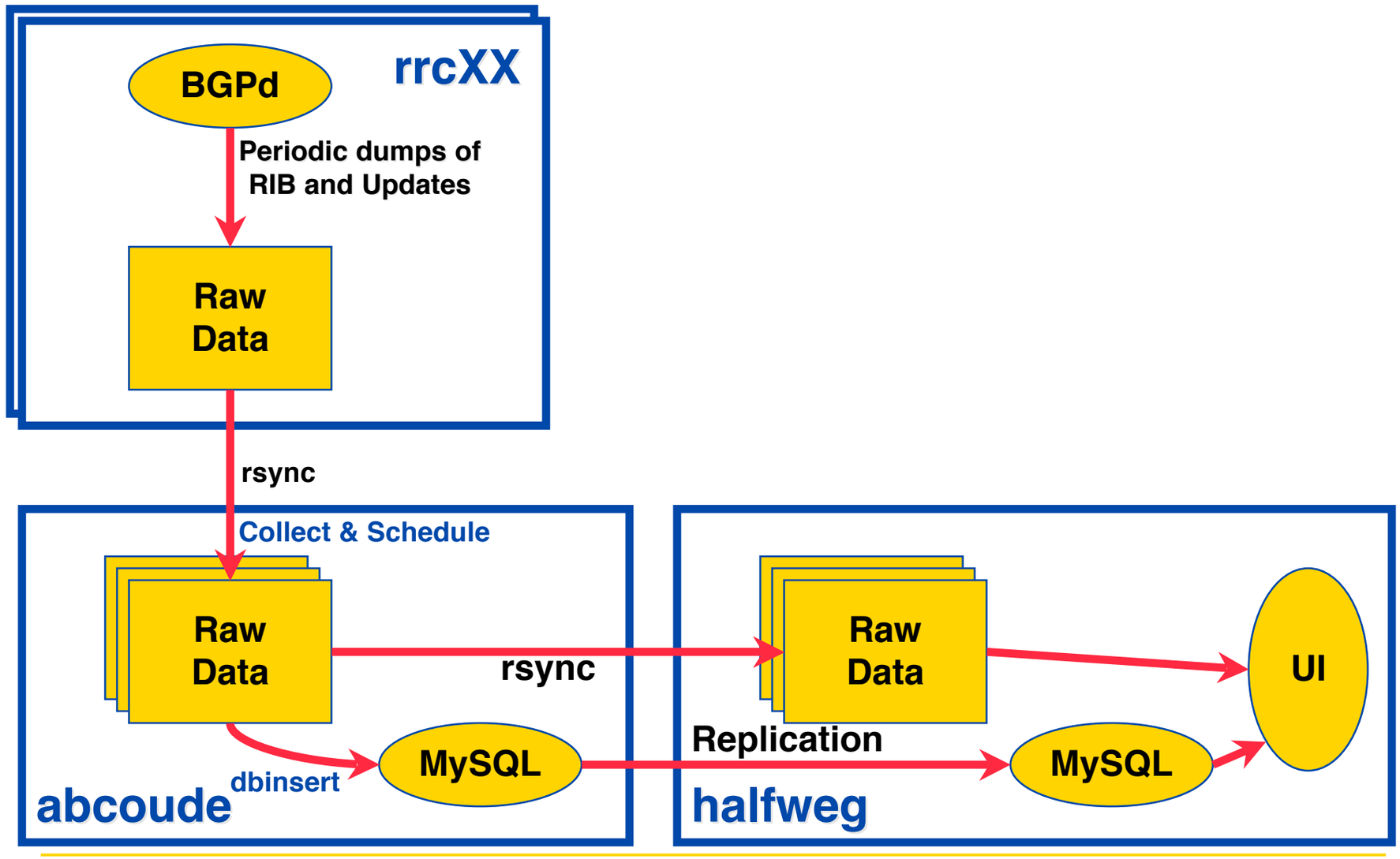
Original architecture (1999)



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

“Classic” architecture (2003, 9 collectors)

“RIS Classic” - Overview



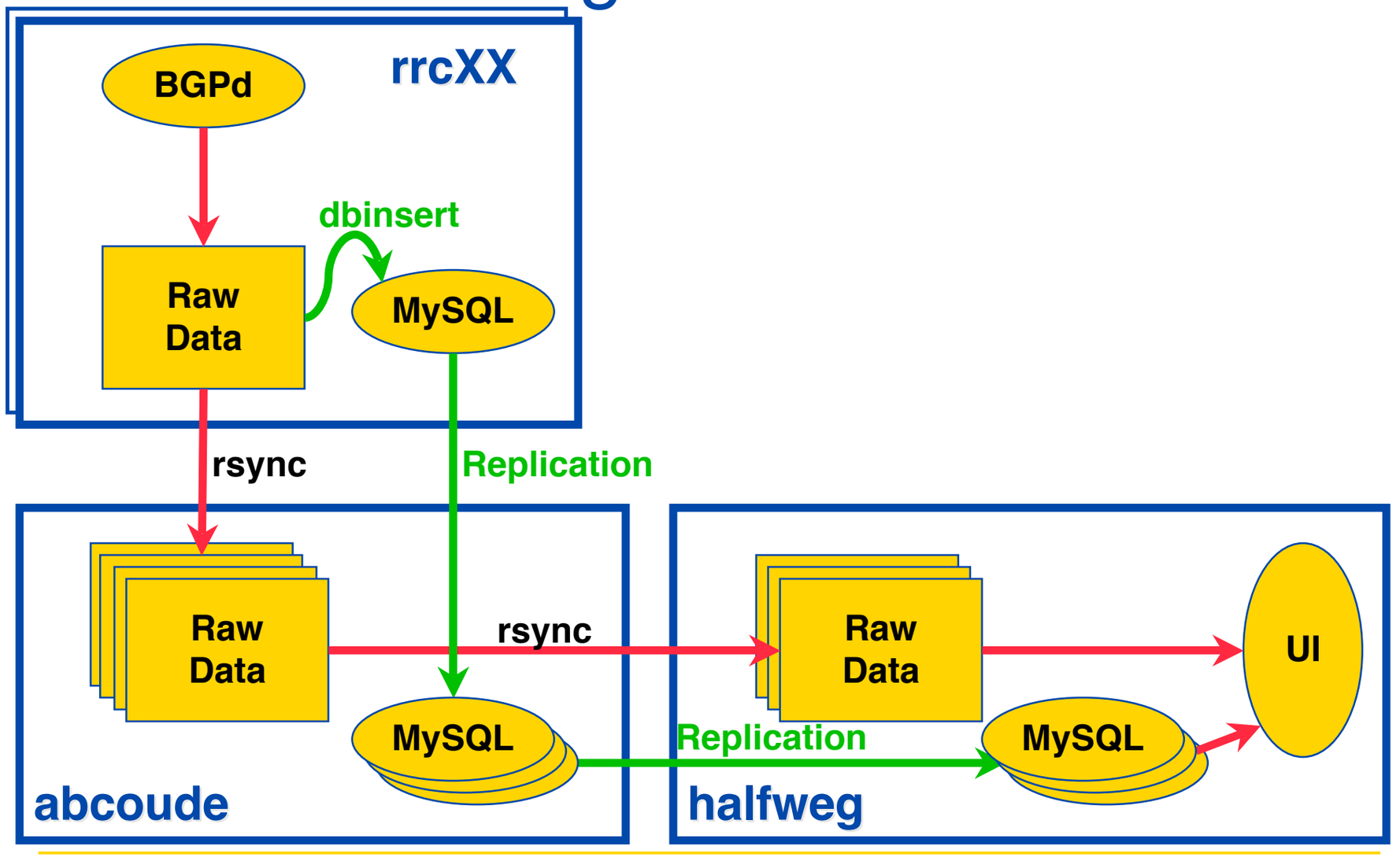
“Classic” architecture (2003, 9 collectors)

Problems

- Database insertion of data from 9 route collectors on a single central machine is slow
  - Little headroom to allow for abnormal cases
  - Can sometimes take more than 24 hours to insert a single day’s data
  - Little capacity to add more RRCs or full BGP feeds
- Limited attributes are stored in the database:
  - Only first 255 characters of AS Path stored
  - Other BGP attributes (communities, MEDs, etc.) ignored

“RISng” architecture (2003, 9 collectors)

RISng - Overview



Scaling the Database

- 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
- Time for Big Data!

Big 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
- Allows us to build a scalable storage and processing cluster
  - Attributes and aggregations for all historical data are available
- Currently over 150 servers in the cluster
  - Although the cluster is not only used for RIS
  - Also used by RIPE Atlas and other projects

Big Data processing – components

- HDFS
  - distributed, replicated, cluster filesystem
- HBase
  - non-relational distributed database
  - large tables – billions of rows × millions of columns
- Map/Reduce
  - massive batch job processing
- Azkaban
  - “crontab” for Hadoop, with dependency tracking and more
- Kafka
  - Message Queue and stream processing

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

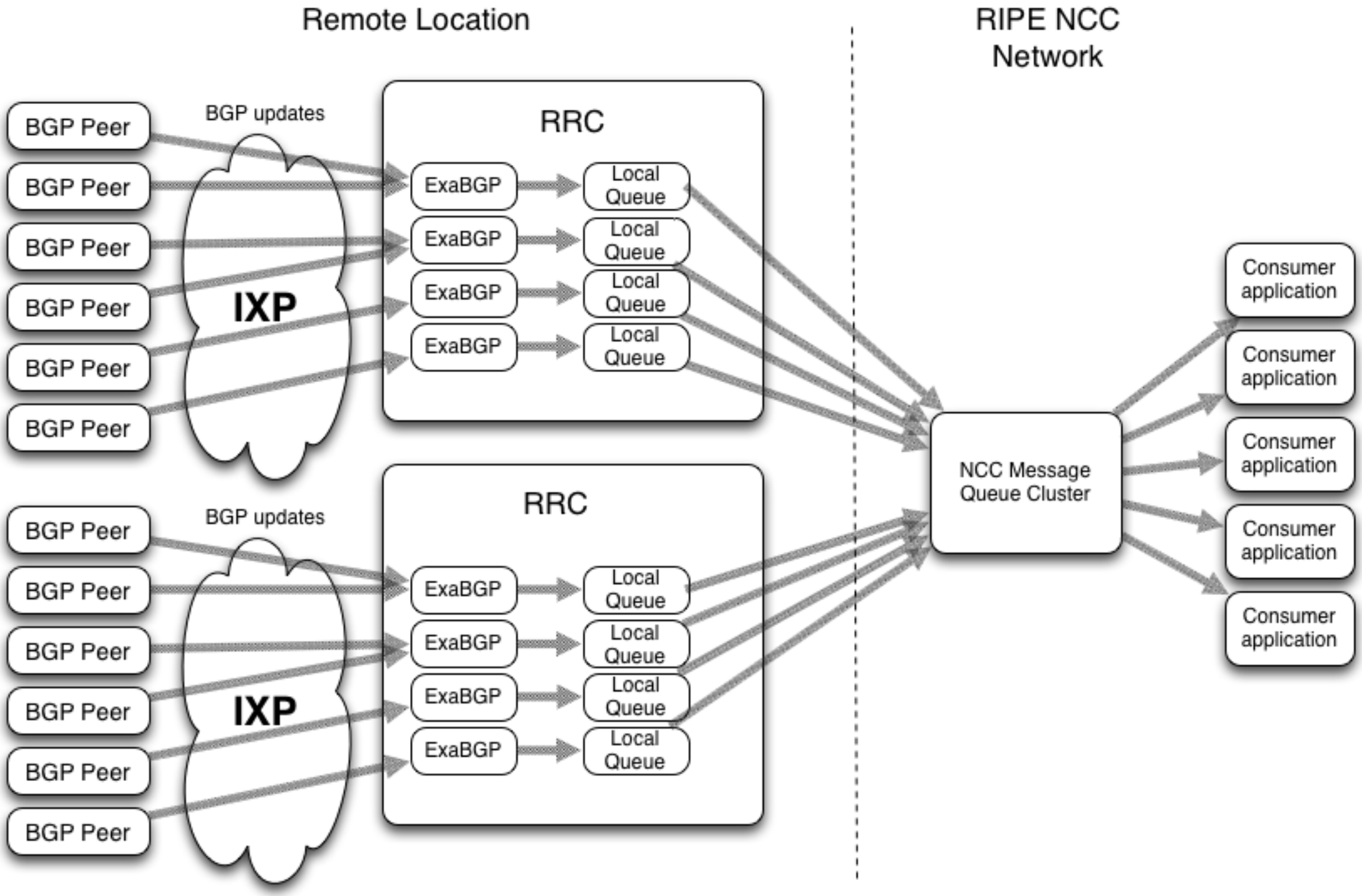
RIS operations

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

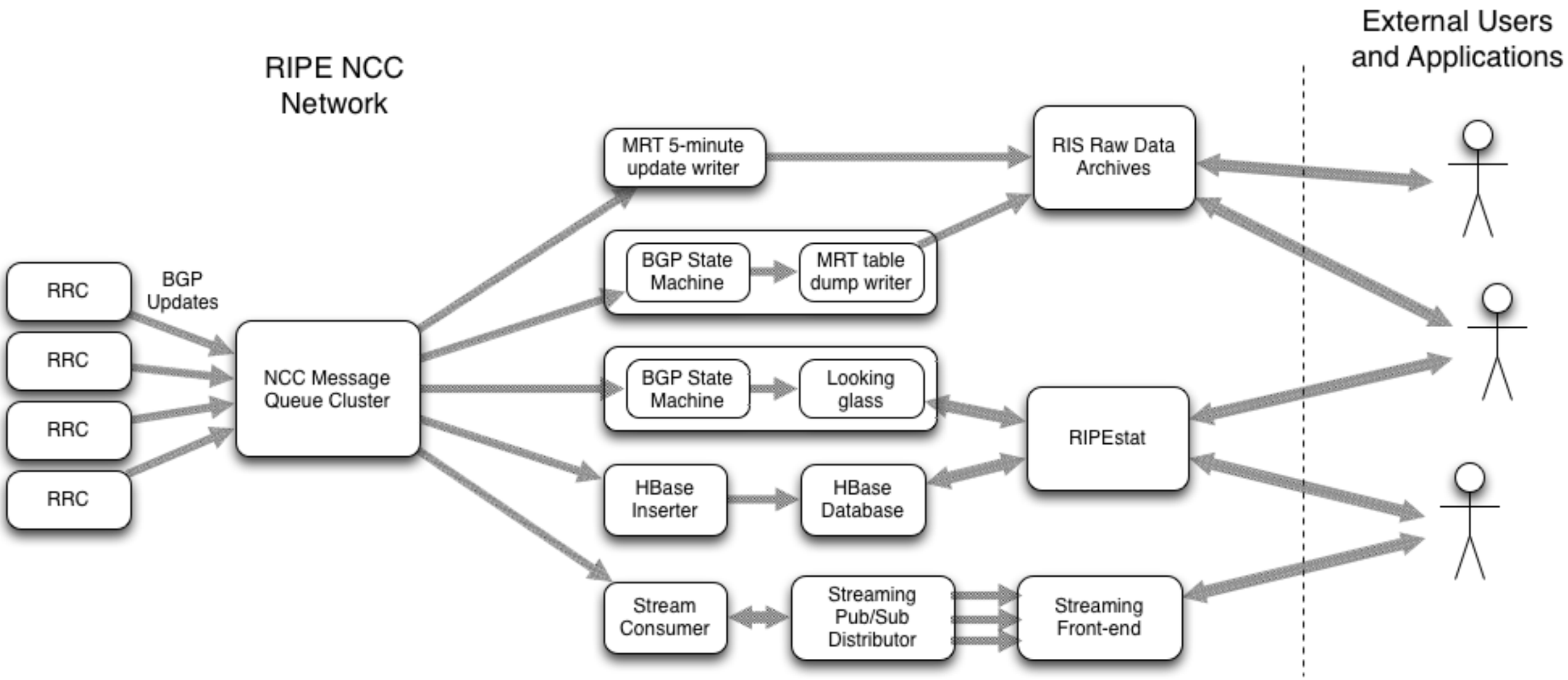
New architecture

- Multiple BGP daemons (ExaBGP) – at least 1 per core
  - lightweight daemon
  - finally could saturate RRC server
- Message Queue
  - RabbitMQ ➠ Kafka
- Stream processing
  - raw updates files
  - Looking Glass
- Batch processing
  - aggregations

Data collection



Back-end data distribution



How can you help?

How can you help?

- Peer with us!!!
  - AS12654
  - 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 ;-)

CIXP, Geneva (RRC04)

DIX-IE, Tokyo (RRC06)

Netnod, Stockholm (RRC07)

MIX, Milan (RRC10)

NYIIX, New York City (RRC11)

MSK-IX, Moscow (RRC13)

PAIX, Palo Alto (RRC14)

PTTMetro, Sao Paulo (RRC15)

NOTA, Miami (RRC16)

Catnix, Barcelona (RRC18)

NAP Africa JB, Johannesburg (RRC19)

SwissIX, Zurich (RRC20)

France-IX, Paris (RRC21)

Interested in hosting an RRC?

Questions

ris@ripe.net
https://ripe.net/ris
https://stat.ripe.net