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 :)  
- 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

Routing Status (195.137.160.0/24)

At 2017-10-03 16:00:00 UTC, 195.137.160.0/24 was 100% visible (by 158 of 158 RIS full peers).

First ever seen announced by AS35594, on 2005-10-04 16:00:00 UTC.

Originated by: AS6697 (valid route object in RIPE)

No less-specific covering prefixes.

Advanced Settings

Showing results for 195.137.160.0/24 as of 2017-10-03 16:00:00 UTC
RIPEstat Examples
RIPEstat Examples: Selectel case
RIPEstat Examples: Selectel case
RIPEstat Examples

BGP Looking Glass (195.137.160.0/24)

- 14 RRCs see 115 peers announcing 195.137.160.0/24 originated by AS6697.

- RRC11 in New York City, New York, US sees 1 ASN originating 195.137.160.0/24. (AS6697)
- RRC10 in Milan, Italy sees 1 ASN originating 195.137.160.0/24. (AS6697)
- RRC00 in Amsterdam, Netherlands sees 1 ASN originating 195.137.160.0/24. (AS6697)
- RRC01 in London, United Kingdom sees 1 ASN originating 195.137.160.0/24. (AS6697)
- RRC15 in Sao Paulo, Brazil sees 1 ASN originating 195.137.160.0/24. (AS6697)

AS6697 is seen as the origin by 14 peers.

- 187.16.217.48 is announcing route AS16735 AS41095 AS6697.
  187.16.217.48 from 187.16.217.48 (200.225.196.252)
  Origin IGP, localpref 100, valid, external, best
  Community: 16735:5 16735:6101
  Last update: Mon Sep 25 13:17:37 2017

  187.16.220.193 from 187.16.220.193 (177.129.136.254)
  Origin IGP, localpref 100, valid, external
  Last update: Tue Oct 3 04:19:37 2017
### 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

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

Country Routing Statistics (Belarus)

- Current data point resolution in graph: per 1 week

- # of prefixes
- # of ASNs

- ASN
- IPv4
- IPv6

(Click legend item to toggle visibility of a series)

Showing results for Belarus from 2004-01-01 00:00:00 UTC to 2017-10-03 19:34:59 UTC
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](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](https://labs.ripe.net/Members/vastur/the-shape-of-a-bgp-update)
RIS growth

Because the internet keeps growing
## Collector history

<table>
<thead>
<tr>
<th>Collector</th>
<th>Location</th>
<th>IXP</th>
<th>Deployed</th>
<th>Removed</th>
</tr>
</thead>
<tbody>
<tr>
<td>RRC00</td>
<td>Amsterdam</td>
<td>Multi-hop</td>
<td>1999</td>
<td>-</td>
</tr>
<tr>
<td>RRC01</td>
<td>London</td>
<td>LINX</td>
<td>2000</td>
<td>-</td>
</tr>
<tr>
<td>RRC02</td>
<td>Paris</td>
<td>SFINX</td>
<td>2001</td>
<td>2008</td>
</tr>
<tr>
<td>RRC03</td>
<td>Amsterdam</td>
<td>AMS-IX</td>
<td>2001</td>
<td>-</td>
</tr>
<tr>
<td>RRC04</td>
<td>Geneva</td>
<td>CIXP</td>
<td>2001</td>
<td>-</td>
</tr>
<tr>
<td>RRC05</td>
<td>Vienna</td>
<td>VIX</td>
<td>2001</td>
<td>-</td>
</tr>
<tr>
<td>RRC06</td>
<td>Tokyo</td>
<td>DIX-IE</td>
<td>2001</td>
<td>-</td>
</tr>
<tr>
<td>RRC07</td>
<td>Stockholm</td>
<td>Netnod</td>
<td>2002</td>
<td>-</td>
</tr>
<tr>
<td>RRC08</td>
<td>San Jose</td>
<td>MAE-West</td>
<td>2002</td>
<td>2004</td>
</tr>
<tr>
<td>RRC09</td>
<td>Zurich</td>
<td>TIX</td>
<td>2003</td>
<td>2004</td>
</tr>
<tr>
<td>RRC10</td>
<td>Milan</td>
<td>MIX</td>
<td>2003</td>
<td>-</td>
</tr>
<tr>
<td>RRC11</td>
<td>New York</td>
<td>NYIIX</td>
<td>2004</td>
<td>-</td>
</tr>
<tr>
<td>RRC12</td>
<td>Frankfurt</td>
<td>DE-CIX</td>
<td>2004</td>
<td>-</td>
</tr>
<tr>
<td>RRC13</td>
<td>Moscow</td>
<td>MSK-IX</td>
<td>2005</td>
<td>-</td>
</tr>
<tr>
<td>RRC14</td>
<td>Palo Alto</td>
<td>PAIX</td>
<td>2005</td>
<td>-</td>
</tr>
<tr>
<td>RRC15</td>
<td>Sao Paulo</td>
<td>PTT-Metro SP</td>
<td>2006</td>
<td>-</td>
</tr>
<tr>
<td>RRC16</td>
<td>Miami</td>
<td>NOTA</td>
<td>2008</td>
<td>-</td>
</tr>
<tr>
<td>RRC18</td>
<td>Barcelona</td>
<td>CATNIX</td>
<td>2015</td>
<td>-</td>
</tr>
<tr>
<td>RRC17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RRC19</td>
<td>Johannesburg</td>
<td>NAPAfrica JB</td>
<td>2016</td>
<td>-</td>
</tr>
<tr>
<td>RRC20</td>
<td>Zurich</td>
<td>SwissIX</td>
<td>2015</td>
<td>-</td>
</tr>
<tr>
<td>RRC21</td>
<td>Paris</td>
<td>FranceIX</td>
<td>2015</td>
<td>-</td>
</tr>
<tr>
<td>RRC22</td>
<td>Bucharest</td>
<td>InterLAN</td>
<td>coming soon</td>
<td>-</td>
</tr>
<tr>
<td>RRC23</td>
<td>Singapore</td>
<td>Equinix SG</td>
<td>coming soon</td>
<td>-</td>
</tr>
</tbody>
</table>
Size of compressed raw BGP updates per month

<table>
<thead>
<tr>
<th>Year</th>
<th>Month</th>
<th>Gigabytes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>01</td>
<td>0</td>
</tr>
<tr>
<td>2001</td>
<td>06</td>
<td>0</td>
</tr>
<tr>
<td>2002</td>
<td>01</td>
<td>0</td>
</tr>
<tr>
<td>2002</td>
<td>06</td>
<td>0</td>
</tr>
<tr>
<td>2003</td>
<td>01</td>
<td>0</td>
</tr>
<tr>
<td>2003</td>
<td>06</td>
<td>0</td>
</tr>
<tr>
<td>2004</td>
<td>01</td>
<td>0</td>
</tr>
<tr>
<td>2004</td>
<td>06</td>
<td>0</td>
</tr>
<tr>
<td>2005</td>
<td>01</td>
<td>0</td>
</tr>
<tr>
<td>2005</td>
<td>06</td>
<td>0</td>
</tr>
<tr>
<td>2006</td>
<td>01</td>
<td>0</td>
</tr>
<tr>
<td>2006</td>
<td>06</td>
<td>0</td>
</tr>
<tr>
<td>2007</td>
<td>01</td>
<td>0</td>
</tr>
<tr>
<td>2007</td>
<td>06</td>
<td>0</td>
</tr>
<tr>
<td>2008</td>
<td>01</td>
<td>0</td>
</tr>
<tr>
<td>2008</td>
<td>06</td>
<td>0</td>
</tr>
<tr>
<td>2009</td>
<td>01</td>
<td>0</td>
</tr>
<tr>
<td>2009</td>
<td>06</td>
<td>0</td>
</tr>
<tr>
<td>2010</td>
<td>01</td>
<td>0</td>
</tr>
<tr>
<td>2010</td>
<td>06</td>
<td>0</td>
</tr>
<tr>
<td>2011</td>
<td>01</td>
<td>0</td>
</tr>
<tr>
<td>2011</td>
<td>06</td>
<td>0</td>
</tr>
<tr>
<td>2012</td>
<td>01</td>
<td>0</td>
</tr>
<tr>
<td>2012</td>
<td>06</td>
<td>0</td>
</tr>
<tr>
<td>2013</td>
<td>01</td>
<td>0</td>
</tr>
<tr>
<td>2013</td>
<td>06</td>
<td>0</td>
</tr>
<tr>
<td>2014</td>
<td>01</td>
<td>0</td>
</tr>
<tr>
<td>2014</td>
<td>06</td>
<td>0</td>
</tr>
<tr>
<td>2015</td>
<td>01</td>
<td>0</td>
</tr>
<tr>
<td>2015</td>
<td>06</td>
<td>0</td>
</tr>
<tr>
<td>2016</td>
<td>01</td>
<td>0</td>
</tr>
<tr>
<td>2016</td>
<td>06</td>
<td>0</td>
</tr>
<tr>
<td>2017</td>
<td>01</td>
<td>0</td>
</tr>
<tr>
<td>2017</td>
<td>06</td>
<td>0</td>
</tr>
<tr>
<td>2018</td>
<td>01</td>
<td>0</td>
</tr>
<tr>
<td>2018</td>
<td>06</td>
<td>0</td>
</tr>
<tr>
<td>2019</td>
<td>01</td>
<td>0</td>
</tr>
<tr>
<td>2019</td>
<td>06</td>
<td>0</td>
</tr>
</tbody>
</table>

Oleg Muravskiy | ENOG 14 | October 2017
Size of compressed BGP dumps per month
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 ;)

Oleg Muravskiy | ENOG 14 | October 2017
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

- BGPd
  - Periodic dumps of RIB and Updates
  - rsync

- Raw Data
  - MySQL
  - dbinsert

- abcoude
  - Collect & Schedule

- halfweg
  - Replication

- MySQL
  - rsync

- UI

“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

- BGPd
- rrcXX
- Raw Data
- MySQL
- dbinsert
- replications
- UI
- abcoude
- halfweg

James Aldridge, RIPE 44, January 2003, Amsterdam
http://www.ripe.net/ris
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 an 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 ;-)
Interested in hosting an RRC?
Questions

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