



RIPE NCC
RIPE NETWORK COORDINATION CENTRE

Developing the RIPE Routing Information System (RIS)

New Adventures in RIS



What is RIS?

- Collects BGP data from a diverse set of networks on the Internet
- ris.ripe.net
- Many interfaces
 - RIPEstat , RIS Live, raw data
 - Internet outages analyses on RIPE Labs
 - RPKI stats
 - How are you or your members using it?

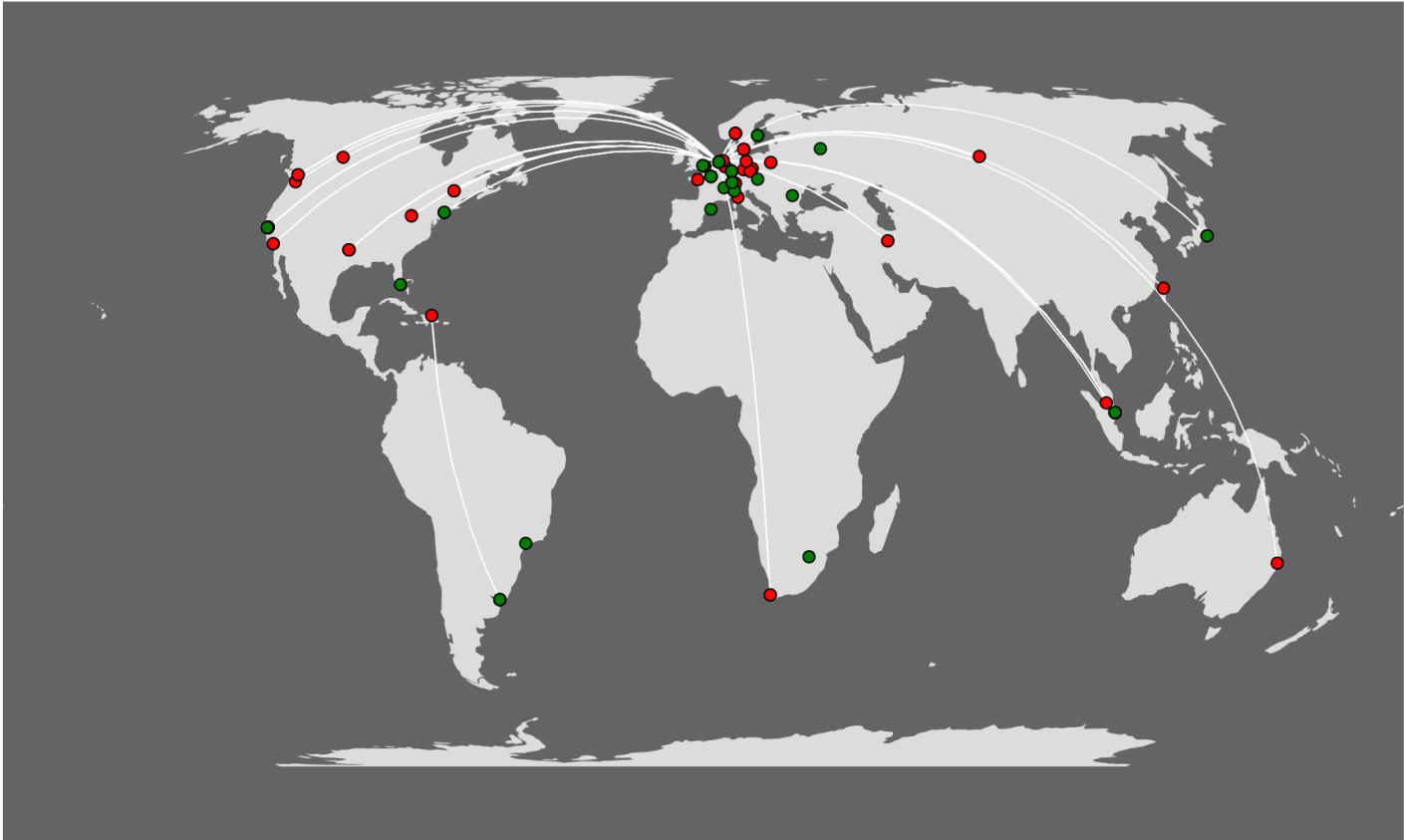
What Do We Want RIS To Be?



- A sustainable service that helps create a safe and stable Internet
- To collect and re-distribute the most relevant BGP routing information and strive for shortest possible time to insight



What is RIS? (in pictures)



RIS is Growing!



```
199M   rrc00/2019.01/bview.20190101.0000.gz
103M   rrc01/2019.01/bview.20190101.0000.gz
102M   rrc03/2019.01/bview.20190101.0000.gz
32M    rrc04/2019.01/bview.20190101.0000.gz
32M    rrc05/2019.01/bview.20190101.0000.gz
16M    rrc06/2019.01/bview.20190101.0000.gz
32M    rrc07/2019.01/bview.20190101.0000.gz
82M    rrc10/2019.01/bview.20190101.0000.gz
48M    rrc11/2019.01/bview.20190101.0000.gz
120M   rrc12/2019.01/bview.20190101.0000.gz
55M    rrc13/2019.01/bview.20190101.0000.gz
45M    rrc14/2019.01/bview.20190101.0000.gz
121M   rrc15/2019.01/bview.20190101.0000.gz
29M    rrc16/2019.01/bview.20190101.0000.gz
15M    rrc18/2019.01/bview.20190101.0000.gz
37M    rrc19/2019.01/bview.20190101.0000.gz
148M   rrc20/2019.01/bview.20190101.0000.gz
110M   rrc21/2019.01/bview.20190101.0000.gz
4.0K   rrc22/2019.01/bview.20190101.0000.gz
22M    rrc23/2019.01/bview.20190101.0000.gz
1.4G   total
```

```
586M   rrc00/2020.01/bview.20200101.0000.gz
214M   rrc01/2020.01/bview.20200101.0000.gz
217M   rrc03/2020.01/bview.20200101.0000.gz
30M    rrc04/2020.01/bview.20200101.0000.gz
45M    rrc05/2020.01/bview.20200101.0000.gz
20M    rrc06/2020.01/bview.20200101.0000.gz
46M    rrc07/2020.01/bview.20200101.0000.gz
105M   rrc10/2020.01/bview.20200101.0000.gz
57M    rrc11/2020.01/bview.20200101.0000.gz
264M   rrc12/2020.01/bview.20200101.0000.gz
70M    rrc13/2020.01/bview.20200101.0000.gz
53M    rrc14/2020.01/bview.20200101.0000.gz
147M   rrc15/2020.01/bview.20200101.0000.gz
22M    rrc16/2020.01/bview.20200101.0000.gz
14M    rrc18/2020.01/bview.20200101.0000.gz
57M    rrc19/2020.01/bview.20200101.0000.gz
205M   rrc20/2020.01/bview.20200101.0000.gz
175M   rrc21/2020.01/bview.20200101.0000.gz
21M    rrc22/2020.01/bview.20200101.0000.gz
22M    rrc23/2020.01/bview.20200101.0000.gz
16M    rrc24/2020.01/bview.20200101.0000.gz
2.4G   total
```

Downside: Analysis takes twice as long



Redundancy

- Do we have redundancies in the data?
- Is RIS diverse?
 - What does this mean for BGPlay, RIS Live?
- Current expansion:
 - Add route collectors (RRCs) at IXPs
 - Add peers at multihop or 'local' RRCs
- Can we think of strategies for better diversity
 - Less data processing, more signal, **shorter time to insight**



Diversity and Bias

- Is RIS (or any route collector project) representative of the Internet?
- The way we “sample the Internet” suggests it is biased
- Value for RIS peers:
 - For the good of the Internet
 - “I look better in Internet rankings”
- We observe the “clue core”
- Are we in a “Filter Bubble”?



Convenience Sampling

- https://en.wikipedia.org/wiki/Convenience_sampling

Convenience sampling

From Wikipedia, the free encyclopedia

Convenience sampling (also known as **grab sampling**, **accidental sampling**, or **opportunity sampling**) is a type of **non-probability sampling** that involves the **sample** being drawn from that part of the population that is close to hand. This type of sampling is most useful for **pilot testing**.

Advantages [\[edit \]](#)

Convenience sampling can be used by almost anyone and has been around for generations. One of the reasons that it is most often used is due to the numerous advantages it provides. This method is extremely speedy, easy, readily available, and cost effective, causing it to be an attractive option to most researchers.^[2]

Disadvantages [\[edit \]](#)

Even though convenience sampling can be easy to obtain, its disadvantages usually outweigh the advantages. This sampling technique may be more appropriate for one type of study and less for another.

Bias

The results of the convenience sampling cannot be generalized to the target **population** because of the potential **bias** of the sampling technique due to under-representation of subgroups in the sample in comparison to the **population** of interest. The **bias** of the sample cannot be measured. Therefore, inferences based on the convenience sampling should be made only about the sample itself.^[9]

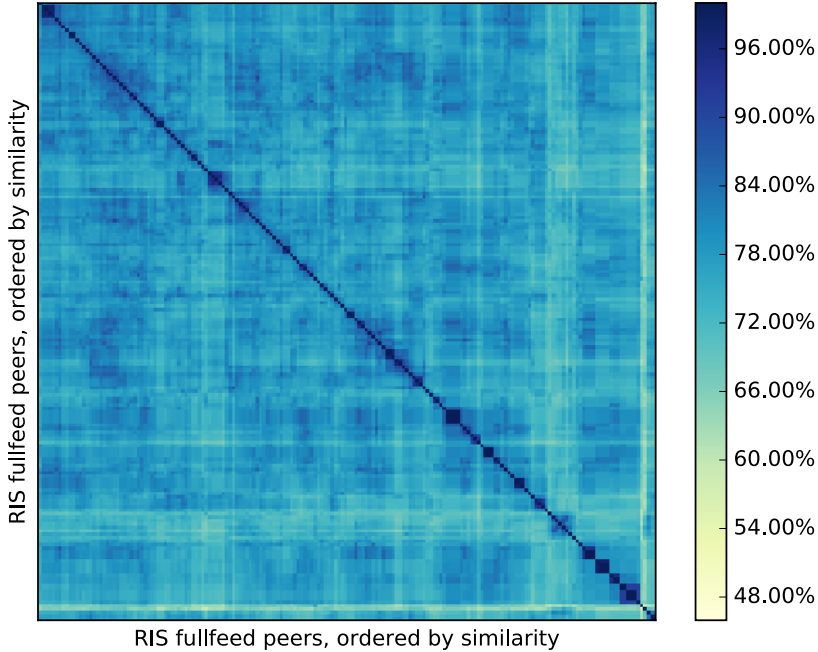
Power

Convenience sampling is characterized with insufficient **power** to identify differences of population subgroups.^[10]

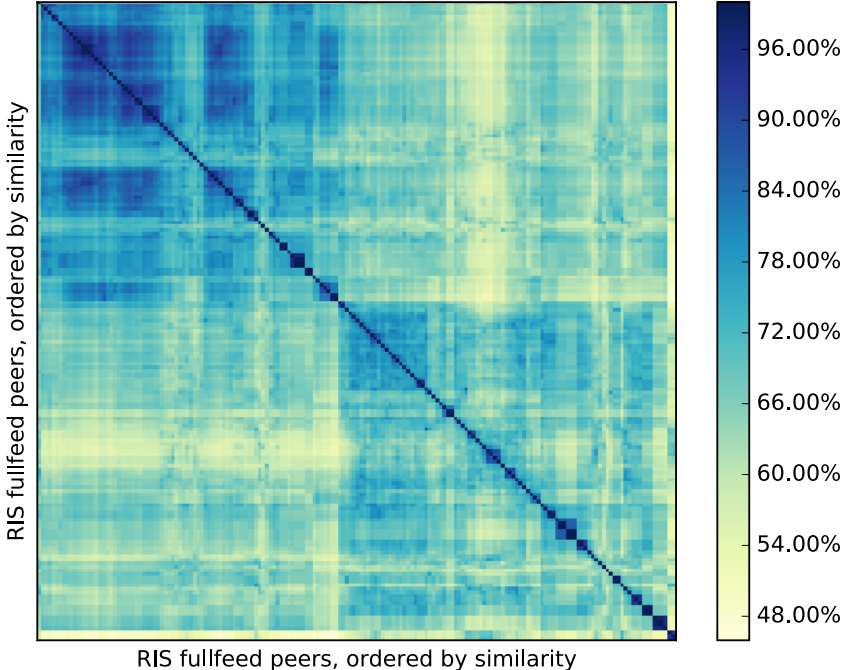


Measuring Diversity in RIS

Similarity matrix for RIS peers IPv4



Similarity matrix for RIS peers IPv6



adapted from:

<https://labs.ripe.net/Members/emileaben/how-diverse-is-ris>



Example: BGP Hijacks

- By making RIS more diverse, we'll be able to see hijacks that currently fly under the radar
- If we cover enough ASNs that are central to a certain region/country we can quantify the effect of a hijack on that region/country
- Detecting local (scoped) events needs diversity



How?

- Technical
 - BMP / ADD_PATH
 - Focus on multi-hop collectors (regional ones?)
- Incentives
 - Value for peers: Peer-centric interfaces/analysis?
 - T-shirts?
- Targeting?
 - Network Operator Groups?



Conclusion

- We can't answer the "is this representative"-question
- We can assess when peers add to our diversity
- ***Do we want to move there and how?***

<https://seclists.org/nanog/2019/Aug/369>

Also, one of the strengths to the 'monitoring as a service' folks is their number of collection points and breadth of ASN to which they interconnect those points/ RISLive, I think, reports out from ~37 or so RIPE probes, how do we (the internet) get more deployed (or better interconnection to the current sets)? and maybe even more importantly... what's the right spread/location/interconnectivity map for these probes?



Questions To You

- What would a peering strategy for RIS look like?
- What data ingress (peers) to focus on?
 - Tier1-like networks (central for whole Internet)
 - Locally influential networks (those central to transit for a given region/country)
 - Very local, high interconnect density (IXPs!)
 - Route-servers?
 - ‘Local’ or multihop RRCs
 - Only local tables?



Questions To You

- Better meta-data = more insight
- What do we know about RIS peers?
 - Some stuff we can try infer (is it a full feed?)
- What can and should we expose?
 - Helps with insight
 - Geographical location



Questions



emile.aben@ripe.net
@meileaben



```
# Hegemony and RIS peers. Address family: IPv4
# fullfeed inference: 731733 (>90% of 95pctile of nr pfxes)
asn      hgm%    in_ris  fullfeed_in_ris  asn_name
1299     12.5    True    False            TELIANET Telia Company AB, EU
3356     11.2    False   False            LEVEL3 - Level 3 Parent, LLC, US
6939     6.9     True    False            HURRICANE - Hurricane Electric LLC, US
174      6.9     True    True             COGENT-174 - Cogent Communications, US
2914     6.0     True    True             NTT-COMMUNICATIONS-2914 - NTT America, Inc., US
4134     5.2     False   False            CHINANET-BACKBONE No.31,Jin-rong Street, CN
7018     4.3     True    True             ATT-INTERNET4 - AT&T Services, Inc., US
209      4.2     False   False            CENTURYLINK-US-LEGACY-QWEST - Qwest Communications Company, LLC, US
3257     3.3     False   False            GTT-BACKBONE GTT Communications Inc., DE
721      3.0     False   False            DNIC-ASBLK-00721-00726 - DoD Network Information Center, US
701      2.9     False   False            UUNET - MCI Communications Services, Inc. d/b/a Verizon Business, US
4837     2.8     False   False            CHINA169-BACKBONE CHINA UNICOM China169 Backbone, CN
7922     2.6     False   False            COMCAST-7922 - Comcast Cable Communications, LLC, US
6762     2.1     True    True             SEABONE-NET TELECOM ITALIA SPARKLE S.p.A., IT
58453    2.1     False   False            CMI-INT-HK Level 30, Tower 1, HK
5511     2.0     True    False            Opentransit Orange S.A., FR
9808     2.0     False   False            CMNET-GD Guangdong Mobile Communication Co.Ltd., CN
6453     2.0     True    True             AS6453 - TATA COMMUNICATIONS (AMERICA) INC, US
12956    2.0     True    False            Telefonica Telefonica International Wholesale Services, SL, ES
4766     1.8     False   False            KIXS-AS-KR Korea Telecom, KR
6461     1.7     True    False            ZAYO-6461 - Zayo Bandwidth, US
17676    1.7     False   False            GIGAINFRA Softbank BB Corp., JP
3320     1.6     False   False            DTAG Deutsche Telekom AG, DE
16509    1.6     False   False            AMAZON-02 - Amazon.com, Inc., US
7843     1.5     False   False            TWCABLE-BACKBONE - Time Warner Cable Internet LLC, US
27064    1.3     False   False            DNIC-ASBLK-27032-27159 - DoD Network Information Center, US
4713     1.2     False   False            OCN NTT Communications Corporation, JP
8075     1.1     True    False            MICROSOFT-CORP-MSN-AS-BLOCK - Microsoft Corporation, US
6167     1.1     False   False            CELLCO-PART - Cellco Partnership DBA Verizon Wireless, US
6830     1.1     True    False            LGI-UPC Liberty Global Operations B.V., AT
367      1.1     False   False            DNIC-ASBLK-00306-00371 - DoD Network Information Center, US
1239     1.0     False   False            SPRINTLINK - Sprint, US
```