



**RIPE NCC**  
RIPE NETWORK COORDINATION CENTRE

# Developing the RIPE Routing Information System (RIS)

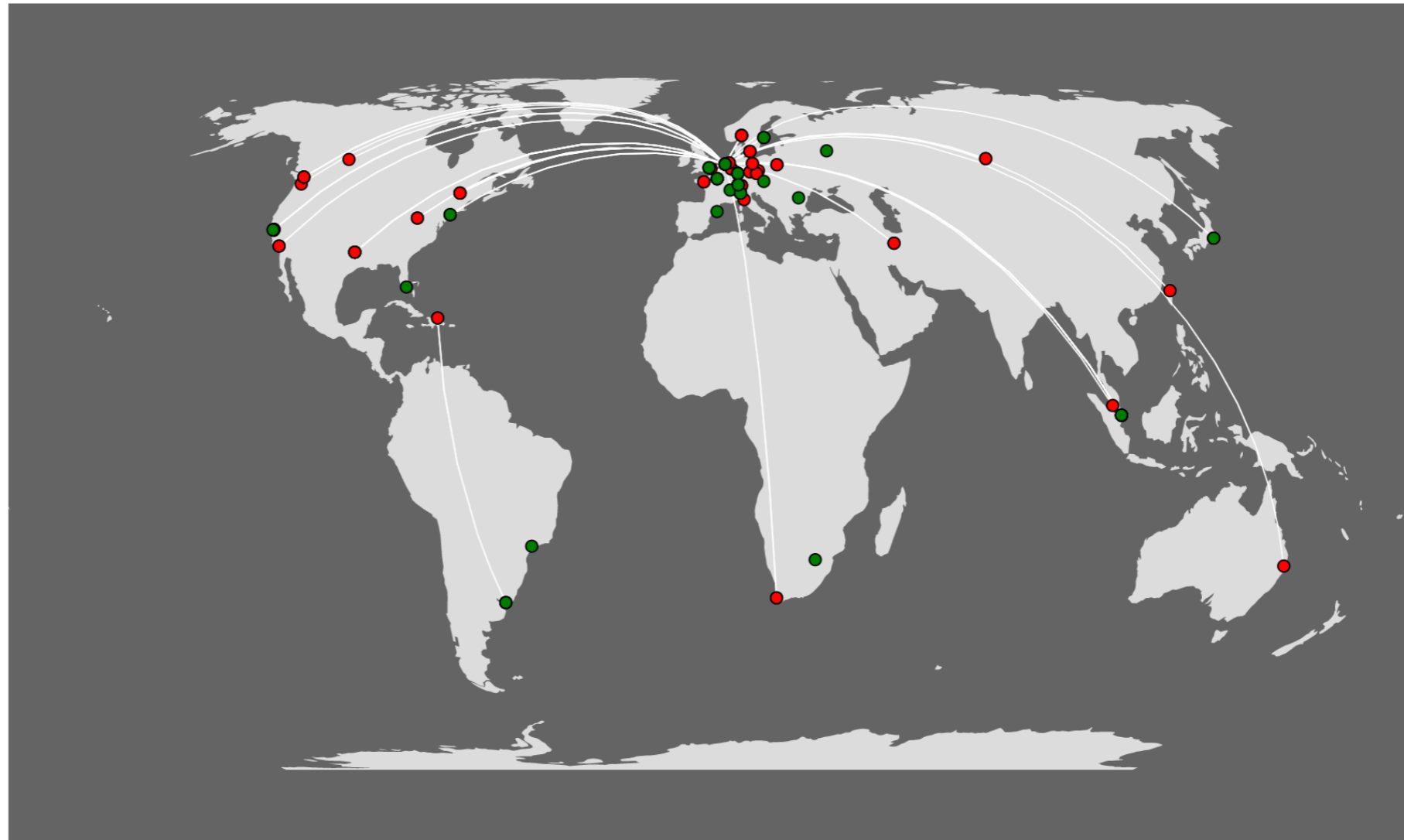
## New Adventures in RIS



# What is RIS?

- Collects BGP data from a diverse set of networks on the Internet
- [ris.ripe.net](https://ris.ripe.net)
- Many tools and analyses of Internet interdomain behaviour depend on RIS
  - RIPEstat , RISlive
  - Internet outages analyses on RIPE Labs
  - RPKI stats

# 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
```

```
465M   rrc00/2019.10/bview.20191001.0000.gz
198M   rrc01/2019.10/bview.20191001.0000.gz
172M   rrc03/2019.10/bview.20191001.0000.gz
29M    rrc04/2019.10/bview.20191001.0000.gz
42M    rrc05/2019.10/bview.20191001.0000.gz
19M    rrc06/2019.10/bview.20191001.0000.gz
44M    rrc07/2019.10/bview.20191001.0000.gz
119M   rrc10/2019.10/bview.20191001.0000.gz
47M    rrc11/2019.10/bview.20191001.0000.gz
202M   rrc12/2019.10/bview.20191001.0000.gz
69M    rrc13/2019.10/bview.20191001.0000.gz
51M    rrc14/2019.10/bview.20191001.0000.gz
169M   rrc15/2019.10/bview.20191001.0000.gz
24M    rrc16/2019.10/bview.20191001.0000.gz
14M    rrc18/2019.10/bview.20191001.0000.gz
52M    rrc19/2019.10/bview.20191001.0000.gz
188M   rrc20/2019.10/bview.20191001.0000.gz
148M   rrc21/2019.10/bview.20191001.0000.gz
68K    rrc22/2019.10/bview.20191001.0000.gz
25M    rrc23/2019.10/bview.20191001.0000.gz
22M    rrc24/2019.10/bview.20191001.0000.gz
2.1G   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 strategy: Add route collectors at IXPs
- Do we need other strategies for better diversity (= less data processing, more signal)



# 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](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.<sup>[2]</sup>

## 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.<sup>[9]</sup>

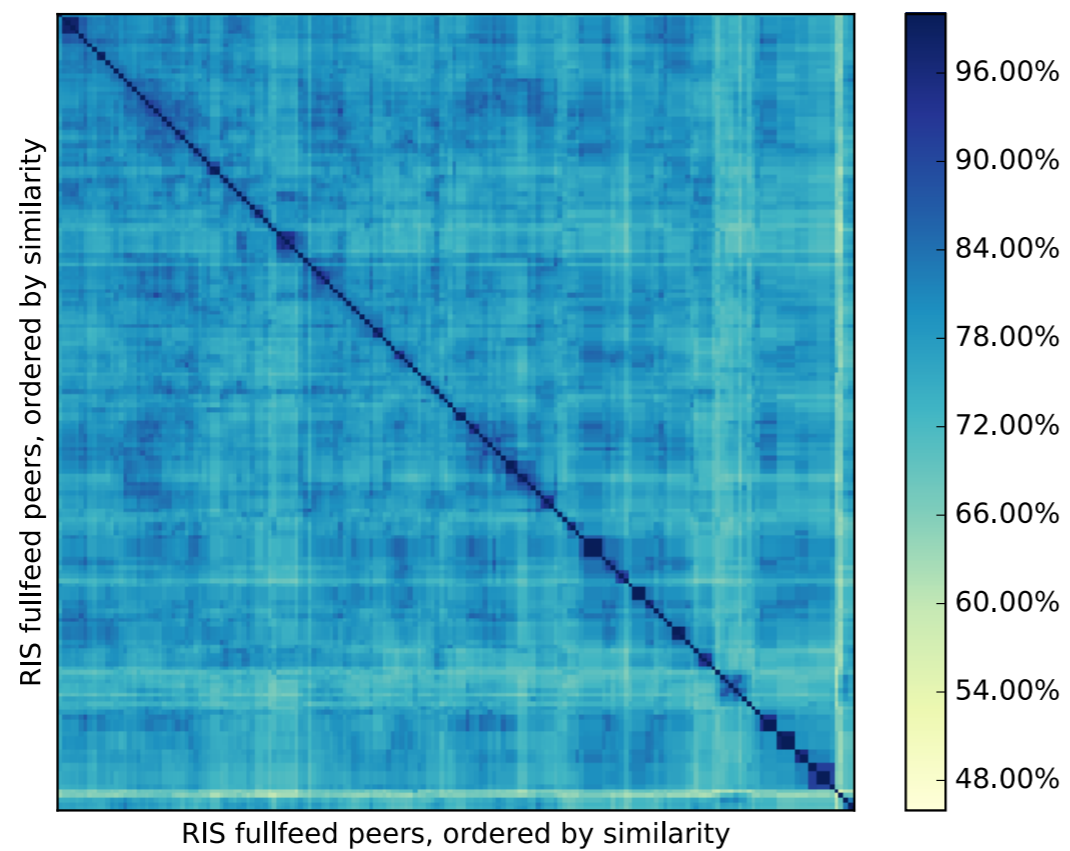
### Power

Convenience sampling is characterized with insufficient **power** to identify differences of population subgroups.<sup>[10]</sup>

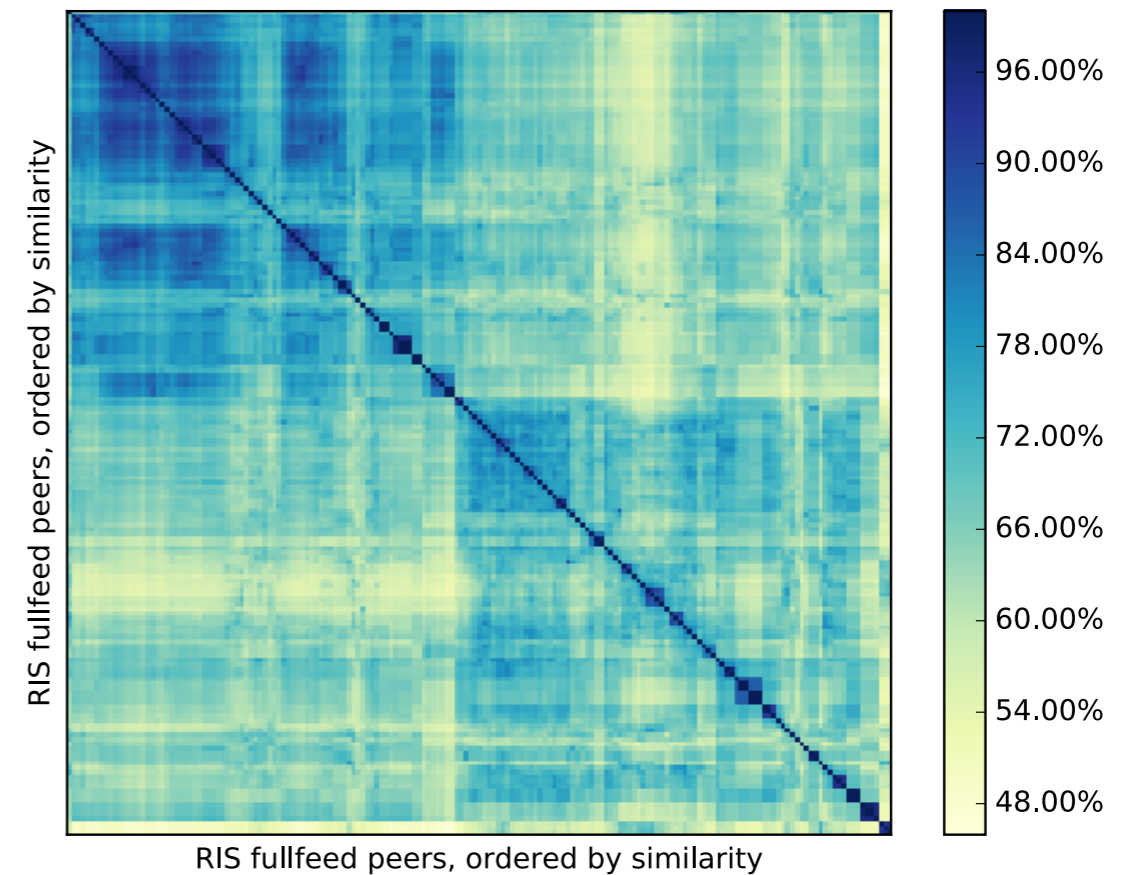
# Diversity in RIS



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
- Globally visible events we see (but also with much less data)
- 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?
  - NOGs?

# 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?

<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



[emile.aben@ripe.net](mailto:emile.aben@ripe.net)  
[@meileaben](https://twitter.com/meileaben)