# The Next Generation of BGP Data Collection Platforms

**Thomas Holterbach**
University of Strasbourg

RIPE 90
May 15, 2025
Lisbon

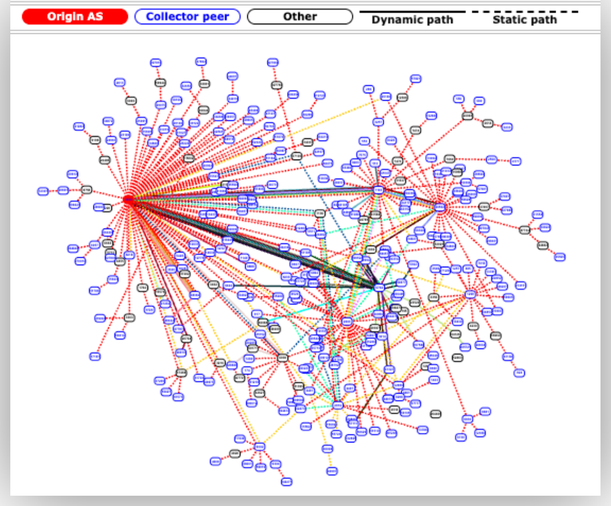
*Joint work with:*
Thomas Alfroy
Cristel Pelsser

# The BGP data collected by platforms like RIS, RouteViews and PCH is essential for understanding and monitoring Internet routing

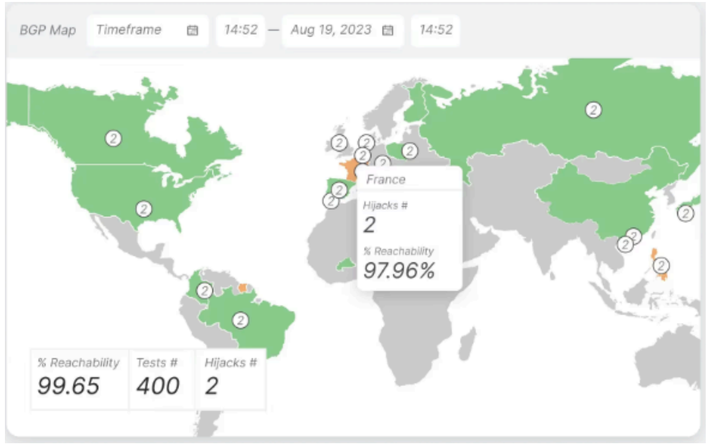
arrangements between ASes.

| AS Rank ▲ | AS Number ▽ | Organization | | cone size (ASes) ▽ |
|---|---|---|---|---|
| 1 | 3356 | Level 3 Parent, LLC | | 53982 |
| 2 | 1299 | Arelion Sweden AB | | 41919 |
| 3 | 174 | Cogent Communications | | 37559 |
| 4 | 3257 | GTT Communications Inc. | | 36334 |
| 5 | 2914 | NTT America, Inc. | | 26976 |
| 6 | 6939 | Hurricane Electric LLC | | 21697 |
| 7 | 6762 | Telecom Italia S.p.A. | | 20271 |
| 8 | 6453 | TATA COMMUNICATIONS (AMERICA) INC | | 19334 |
| 9 | 6461 | Zayo Bandwidth | | 18116 |
| 10 | 3491 | PCCW Global, Inc. | | 11530 |

CAIDA's AS rank




BGPlay




Catchpoint's BGP map

# Routers that share their BGP routes are like satellites observing the Earth's surface

Each router (or "*vantage point*") provides a partial view

But together they allow monitoring at larger scale

bgproutes.io

Our new next-gen BGP route collection platform

https://bgproutes.io

You can peer with us!

# bgproutes.io

Our new next-gen BGP route collection platform

We just launched
our new website!

https://bgproutes.io

You can peer with us!

bgproutes.io

# Our new next-gen BGP route collection platform

- Collection: High coverage
- Storage: Low data management cost
- Distribution: Fast API with high granularity

https://bgproutes.io

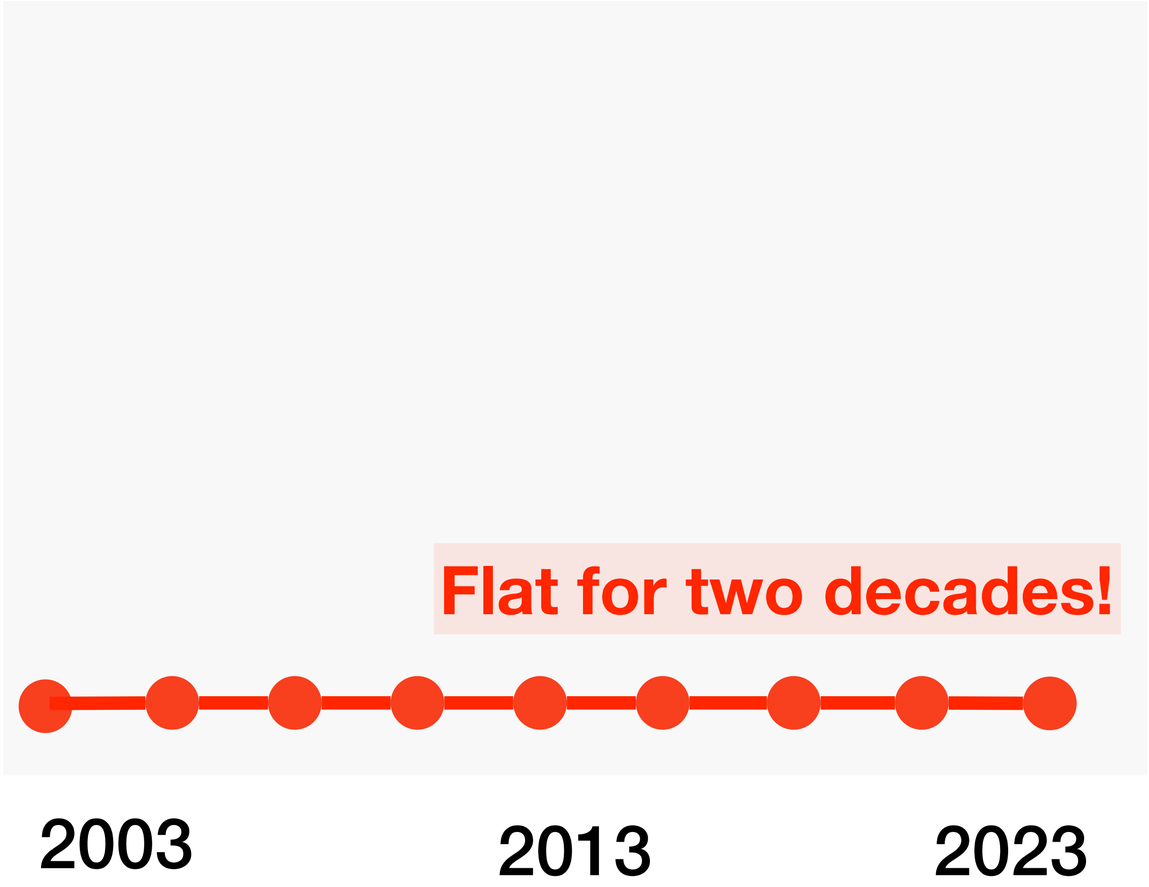
You can peer with us!

# 1.2%

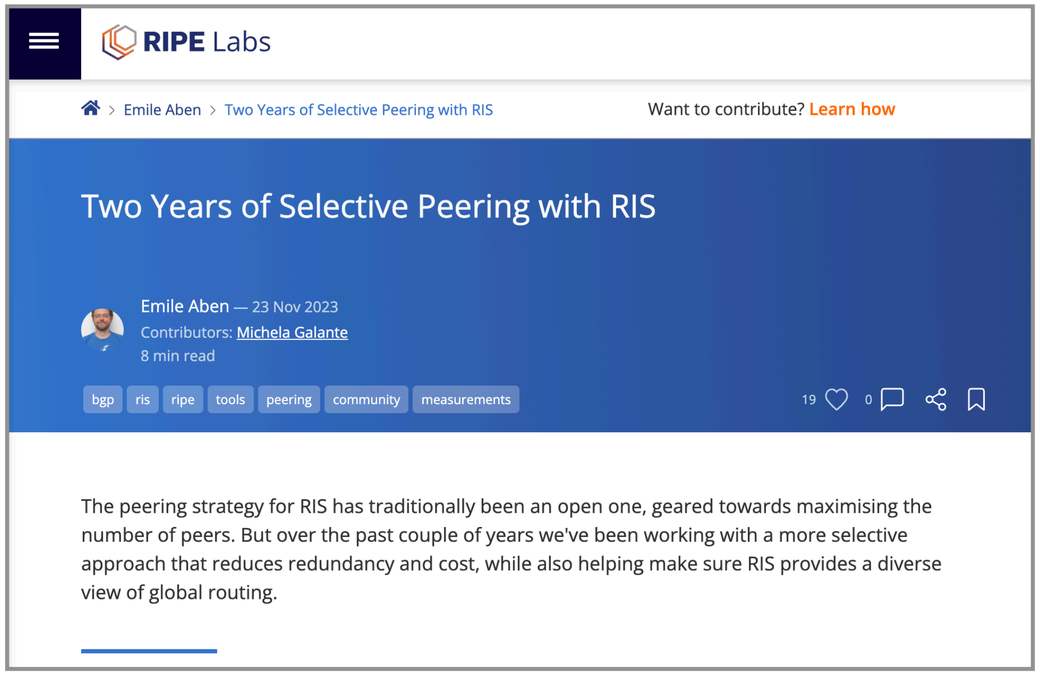
% of ASes sharing their BGP data (RIS+RouteViews) — **1.2%**

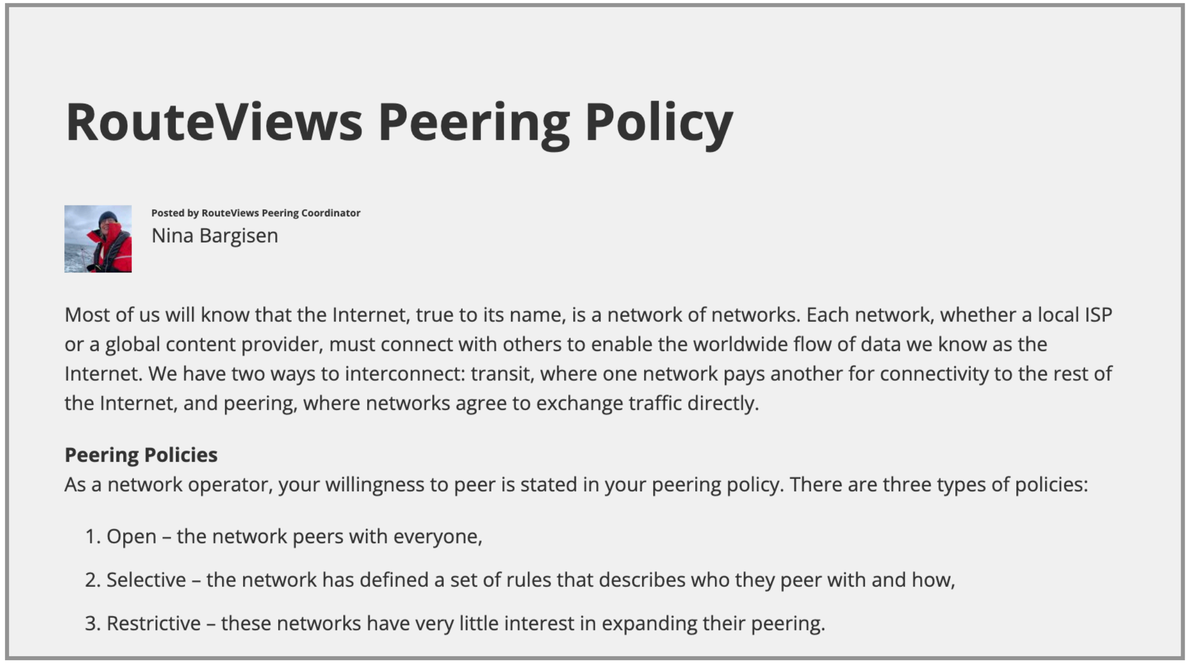
% of ASes sharing their BGP data (RIS+RouteViews)

**1.2%**



# And it is unlikely to improve anytime soon:
# RIS and RouteViews have recently adopted a selective a peering policy



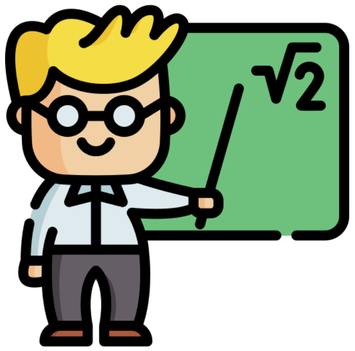



*But is a 1.2% coverage really bad?*

*But is a 1.2% coverage really bad?*

*It is hard to know without ground truth*

But it likely depends on the use case



# We measured the impact of a 1.2% coverage on three use cases using simulations*

**Use case #1:** Peer-to-peer link observation

**Use case #2:** Forged-origin hijack detection

*Our simulations use c-bgp on topologies with 6k ASes

# We measured the impact of a 1.2% coverage on three use cases using simulations*

**Use case #1:** Peer-to-peer link observation
We miss 84% of the links

**Use case #2:** Forged-origin hijack detection

*Our simulations use c-bgp on topologies with 6k ASes

# We measured the impact of a 1.2% coverage on three use cases using simulations*

**Use case #1:** Peer-to-peer link observation
We miss 84% of the links

**Use case #2:** Forged-origin hijack detection
We miss 24% of the Type-1 hijacks

*Our simulations use c-bgp on topologies with 6k ASes

# We measured the impact of a 1.2% coverage on three use cases using simulations*

**Use case #1:** Peer-to-peer link observation
We miss 84% of the links

**Use case #2:** Forged-origin hijack detection
We miss 24% of the Type-1 hijacks

For more details and use cases
see our SIGCOMM'24 paper

*Our simulations use c-bgp
on topologies with 6k ASes

bgproutes.io

Our new next-gen BGP route collection platform

- Collection: High coverage
- Storage: Low data management cost
- Distribution: Fast API with high granularity

https://bgproutes.io

# bgproutes.io

Our new next-gen BGP route collection platform

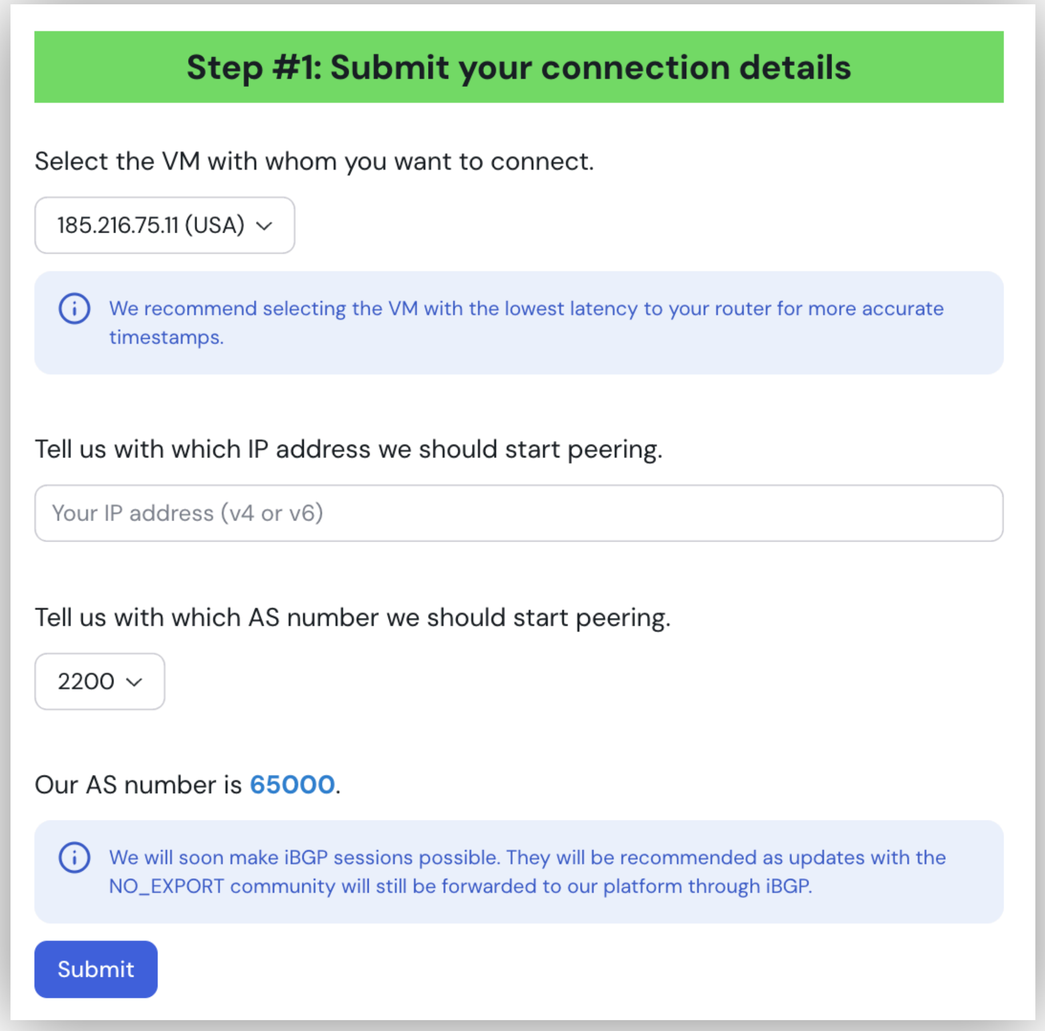
- Collection: High coverage
- Storage: Low data management cost
- Distribution: Fast API with high granularity

https://bgproutes.io

# **bgproutes.io** simplifies, automates and opens BGP data contribution to every ASN

Network operators can authenticate using peeringDB

Network operators just have to fill a form to start peering with bgproutes.io




bgproutes.io/contribute

# Our goal is not to compete against existing platforms

## Each existing platform has its strengths and weaknesses

Peering

| | RIPE RIS | RouteViews | PCH | CGTF RIS |
|---|---|---|---|---|
| Physical peering | ✓ | ✓ | ✓ | ? |
| Remote peering | ✓ | ✓ | ✗ | ? |
| Open peering policy | ✗ | ✗ | ✓ | ✗ |

# Our goal is not to compete against existing platforms

## Each existing platform has its strengths and weaknesses

| | | RIPE RIS | RouteViews | PCH | CGTF RIS |
|---|---|---|---|---|---|
| Peering | Physical peering | ✓ | ✓ | ✓ | ? |
| | Remote peering | ✓ | ✓ | ✗ | ? |
| | Open peering policy | ✗ | ✗ | ✓ | ✗ |
| Data | API to query historical data | ✓ | ✗ | ✗ | ✗ |
| | Real time stream | ✓ | ✓ | ✗ | ✗ |
| | Keep all data forever | ✓ | ✓ | ✓ | ✓ |

# Our goal is not to compete against existing platforms

## Each existing platform has its strengths and weaknesses

| | | RIPE RIS | RouteViews | PCH | CGTF RIS | bgproutes.io |
|---|---|---|---|---|---|---|
| Peering | Physical peering | ✓ | ✓ | ✓ | ? | ✗ |
| | Remote peering | ✓ | ✓ | ✗ | ? | ✓ |
| | Open peering policy | ✗ | ✗ | ✓ | ✗ | ✓ |
| Data | API to query historical data | ✓ | ✗ | ✗ | ✗ | ✓ |
| | Real time stream | ✓ | ✓ | ✗ | ✗ | ✓ |
| | Keep all data forever | ✓ | ✓ | ✓ | ✓ | ? |

# **bgproutes.io** embraces all existing platforms by centralizing their data in one unified repository

**bgproutes.io** already stores data for more than 5000 vantage points

## Data providers

| Name | Number of vantage points |
|---|---|
| bgroutes.io | **v4:** 32 **v6:** 2 |
| RIPE RIS | **v4:** 833 **v6:** 686 |
| RouteViews | **v4:** 735 **v6:** 695 |
| PCH | **v4:** 2129 **v6:** 390 |
| CGTF RIS Soon | **v4:** 0 **v6:** 0 |

bgproutes.io/vantage_points

# bgproutes.io

## Our new next-gen BGP route collection platform

- Collection: High coverage
- Storage: Low data management cost
- Distribution: Fast API with high granularity

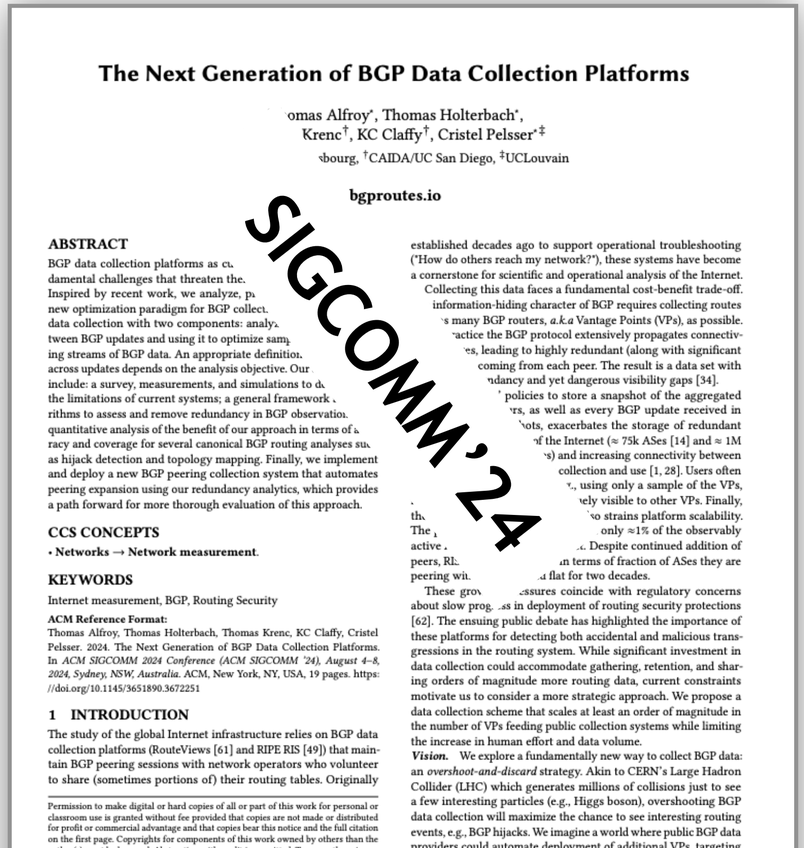
https://bgproutes.io

# **bgproutes.io** lowers data management costs thanks to new BGP compression algorithms

## **Lossless** compression
e.g., for recent data

Under submission

## **Lossy** compression
e.g., for old data



The Next Generation of BGP Data Collection Platforms

omas Alfroy*, Thomas Holterbach*,
Krenc†, KC Claffy†, Cristel Pelsser*‡
bourg, †CAIDA/UC San Diego, ‡UCLouvain

bgproutes.io

SIGCOMM'24

ABSTRACT

BGP data collection platforms as c
damental challenges that threaten the
Inspired by recent work, we analyze, p
new optimization paradigm for BGP collec
data collection with two components: analy
tween BGP updates and using it to optimize sam
ing streams of BGP data. An appropriate definition
across updates depends on the analysis objective. Our
include: a survey, measurements, and simulations to d
the limitations of current systems; a general framework
rithms to assess and remove redundancy in BGP observatio
quantitative analysis of the benefit of our approach in terms of
racy and coverage for several canonical BGP routing analyses su
as hijack detection and topology mapping. Finally, we implement
and deploy a new BGP peering collection system that automates
peering expansion using our redundancy analytics, which provides
a path forward for more thorough evaluation of this approach.

CCS CONCEPTS

• Networks → Network measurement.

KEYWORDS

Internet measurement, BGP, Routing Security

ACM Reference Format:
Thomas Alfroy, Thomas Holterbach, Thomas Krenc, KC Claffy, Cristel Pelsser. 2024. The Next Generation of BGP Data Collection Platforms. In *ACM SIGCOMM 2024 Conference (ACM SIGCOMM '24), August 4–8, 2024, Sydney, NSW, Australia.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3651890.3672251

1 INTRODUCTION

The study of the global Internet infrastructure relies on BGP data collection platforms (RouteViews [61] and RIPE RIS [49]) that maintain BGP peering sessions with network operators who volunteer to share (sometimes portions of) their routing tables. Originally established decades ago to support operational troubleshooting ("How do others reach my network?"), these systems have become a cornerstone for scientific and operational analysis of the Internet.

Collecting this data faces a fundamental cost-benefit trade-off. information-hiding character of BGP requires collecting routes s many BGP routers, *a.k.a* Vantage Points (VPs), as possible. actice the BGP protocol extensively propagates connectivs, leading to highly redundant (along with significant coming from each peer. The result is a data set with ndancy and yet dangerous visibility gaps [34].

policies to store a snapshot of the aggregated rs, as well as every BGP update received in ots, exacerbates the storage of redundant f the Internet (≈ 75k ASes [14] and ≈ 1M s) and increasing connectivity between collection and use [1, 28]. Users often , using only a sample of the VPs, ely visible to other VPs. Finally, th o strains platform scalability. The only ≈1% of the observably active . Despite continued addition of peers, RI n terms of fraction of ASes they are peering wi flat for two decades.

These gro ssures coincide with regulatory concerns about slow prog ss in deployment of routing security protections [62]. The ensuing public debate has highlighted the importance of these platforms for detecting both accidental and malicious transgressions in the routing system. While significant investment in data collection could accommodate gathering, retention, and sharing orders of magnitude more routing data, current constraints motivate us to consider a more strategic approach. We propose a data collection scheme that scales at least an order of magnitude in the number of VPs feeding public collection systems while limiting the increase in human effort and data volume.

***Vision.*** We explore a fundamentally new way to collect BGP data: an *overshoot-and-discard* strategy. Akin to CERN's Large Hadron Collider (LHC) which generates millions of collisions just to see a few interesting particles (e.g., Higgs boson), overshooting BGP data collection will maximize the chance to see interesting routing events, e.g., BGP hijacks. We imagine a world where public BGP data providers could automate deployment of additional VPs, targeting

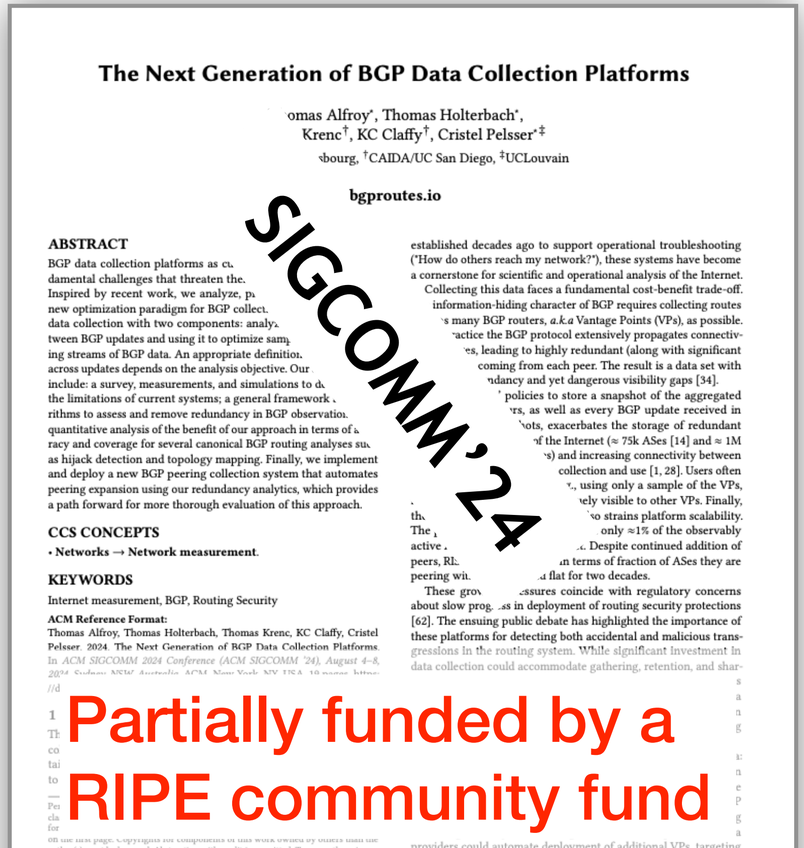
Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the

# **bgproutes.io** lowers data management costs thanks to new BGP compression algorithms

## **Lossless** compression
e.g., for recent data

Under submission

## **Lossy** compression
e.g., for old data



The Next Generation of BGP Data Collection Platforms

omas Alfroy*, Thomas Holterbach*,
Krenc†, KC Claffy†, Cristel Pelsser*‡
bourg, †CAIDA/UC San Diego, ‡UCLouvain

bgproutes.io

ABSTRACT

BGP data collection platforms as c
damental challenges that threaten the
Inspired by recent work, we analyze, p
new optimization paradigm for BGP collec
data collection with two components: analy
tween BGP updates and using it to optimize sam
ing streams of BGP data. An appropriate definition
across updates depends on the analysis objective. Our
include: a survey, measurements, and simulations to d
the limitations of current systems; a general framework
rithms to assess and remove redundancy in BGP observatio
quantitative analysis of the benefit of our approach in terms of
racy and coverage for several canonical BGP routing analyses su
as hijack detection and topology mapping. Finally, we implement
and deploy a new BGP peering collection system that automates
peering expansion using our redundancy analytics, which provides
a path forward for more thorough evaluation of this approach.

CCS CONCEPTS

• Networks → Network measurement.

KEYWORDS

Internet measurement, BGP, Routing Security

ACM Reference Format:
Thomas Alfroy, Thomas Holterbach, Thomas Krenc, KC Claffy, Cristel Pelsser. 2024. The Next Generation of BGP Data Collection Platforms. In *ACM SIGCOMM 2024 Conference (ACM SIGCOMM '24), August 4–8, 2024, Sydney, NSW, Australia.* ACM, New York, NY, USA, 19 pages. https://d

established decades ago to support operational troubleshooting ("How do others reach my network?"), these systems have become a cornerstone for scientific and operational analysis of the Internet.
Collecting this data faces a fundamental cost-benefit trade-off. information-hiding character of BGP requires collecting routes s many BGP routers, *a.k.a* Vantage Points (VPs), as possible. actice the BGP protocol extensively propagates connectiv- es, leading to highly redundant (along with significant coming from each peer. The result is a data set with ndancy and yet dangerous visibility gaps [34].
policies to store a snapshot of the aggregated rs, as well as every BGP update received in ots, exacerbates the storage of redundant f the Internet (≈ 75k ASes [14] and ≈ 1M s) and increasing connectivity between collection and use [1, 28]. Users often ., using only a sample of the VPs, ely visible to other VPs. Finally, th o strains platform scalability. The only ≈1% of the observably active . Despite continued addition of peers, RI n terms of fraction of ASes they are peering wi flat for two decades.
These gro ssures coincide with regulatory concerns about slow prog ss in deployment of routing security protections [62]. The ensuing public debate has highlighted the importance of these platforms for detecting both accidental and malicious transgressions in the routing system. While significant investment in data collection could accommodate gathering, retention, and shar-

SIGCOMM'24

Partially funded by a RIPE community fund

# For now, **bgproutes.io** only uses lossless compression

## Lossless compression
e.g., for recent data

Under submission

## Lossy compression
e.g., for old data

The Next Generation of BGP Data Collection Platforms

Thomas Alfroy*, Thomas Holterbach*,
Thomas Krenc†, KC Claffy†, Cristel Pelsser*‡

*University of Strasbourg, †CAIDA/UC San Diego, ‡UCLouvain

bgproutes.io

SIGCOMM '24

bgproutes.io

Our new next-gen BGP route collection platform

- Collection: High coverage
- Storage: Low data management cost
- Distribution: Fast API with high granularity

https://bgproutes.io

## Collecting more data is useful

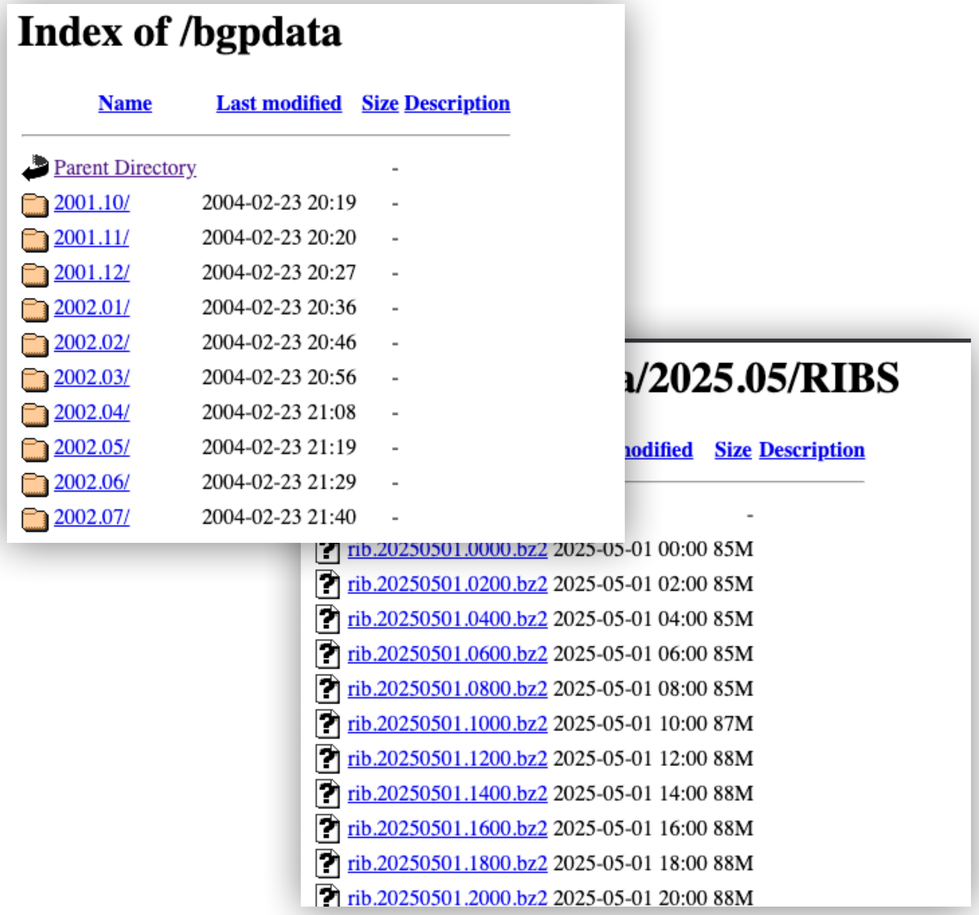
## But not when users cannot effectively process it

# Collecting more data is useful

## But not when users cannot effectively process it

Current MRT archives consist of 1000+ compressed files for updates and RIB dumps

The BGP data is intermingled across all these files




RouteViews MRT archive

# Imagine an network operator trying to retrieve all BGP updates for which their ASN appear in the AS path on May 14, 2025

# Imagine an network operator trying to retrieve all BGP updates for which their ASN appear in the AS path on May 14, 2025

RouteViews MRT archive (10+TBs)

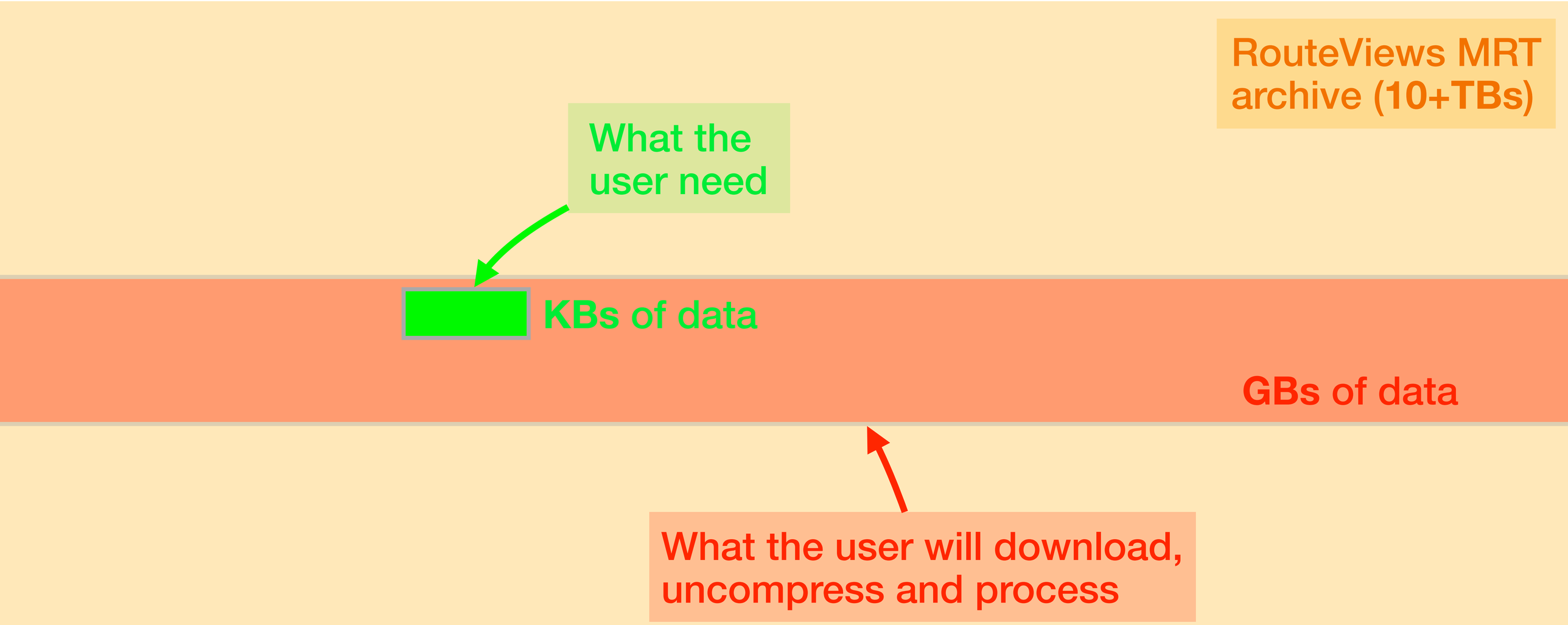
# Imagine an network operator trying to retrieve all BGP updates for which their ASN appear in the AS path on May 14, 2025

RouteViews MRT archive (10+TBs)

What the user need

KBs of data

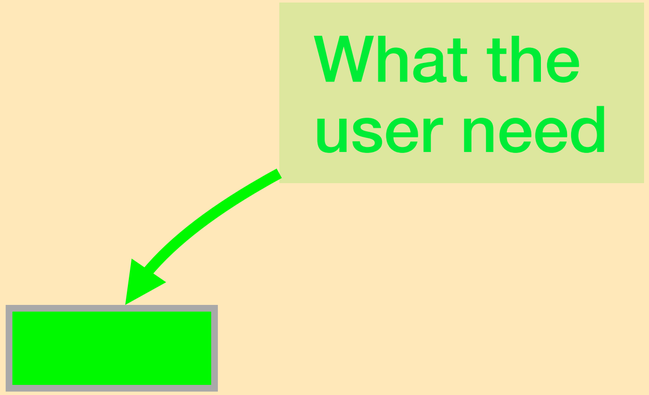
# Imagine an network operator trying to retrieve all BGP updates for which their ASN appear in the AS path on May 14, 2025



# Today, researchers often regretfully resort to sampling

## Results of our survey in our SIGCOMM'24 paper

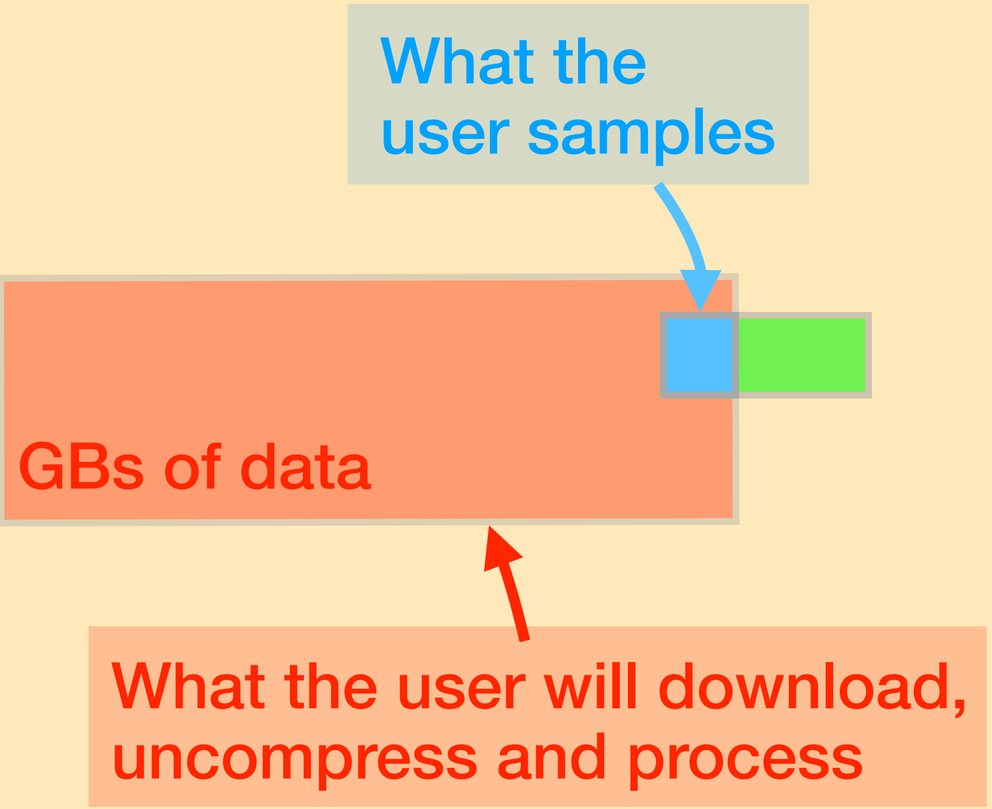
RouteViews MRT archive (10+TBs)



# Today, researchers often regretfully resort to sampling

## Results of our survey in our SIGCOMM'24 paper

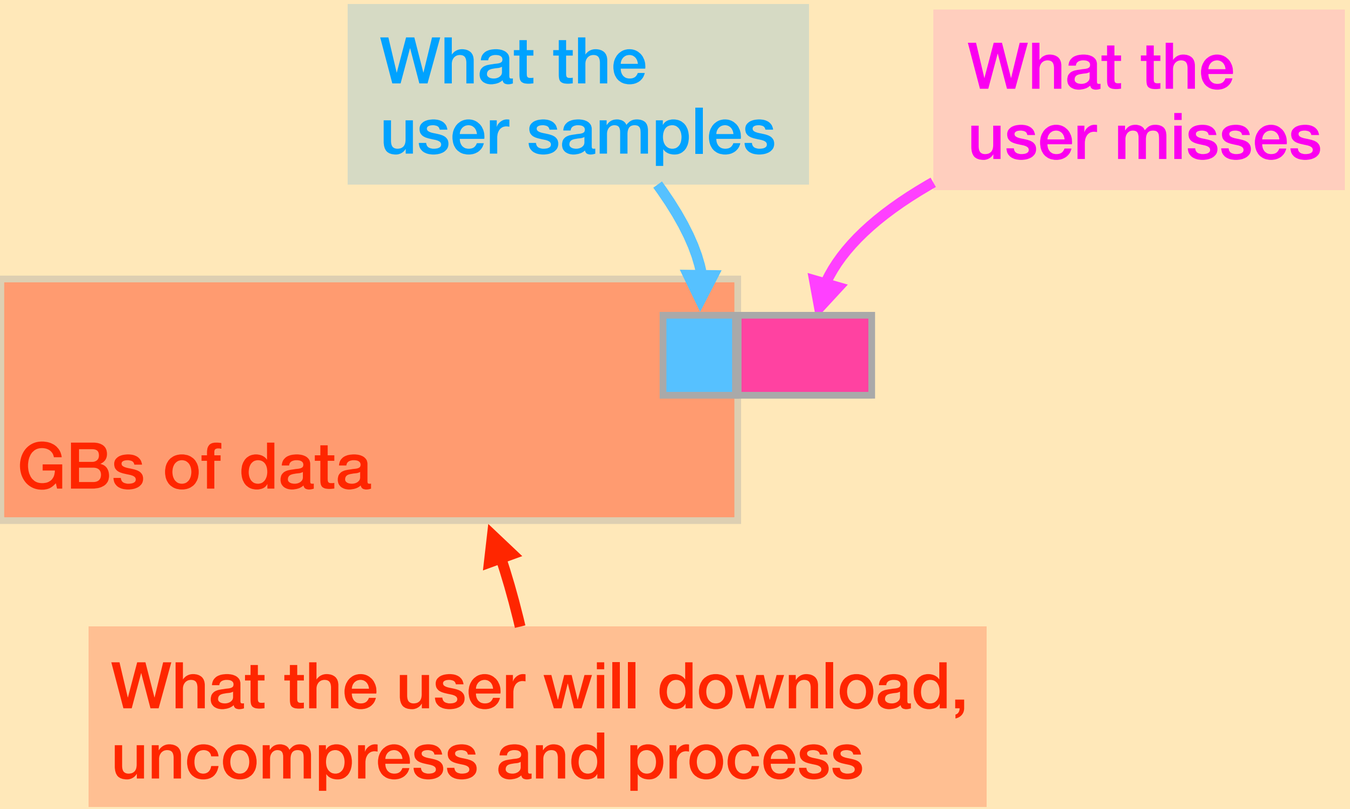
RouteViews MRT archive (10+TBs)



# Today, researchers often regretfully resort to sampling

## Results of our survey in our SIGCOMM'24 paper

RouteViews MRT archive (10+TBs)



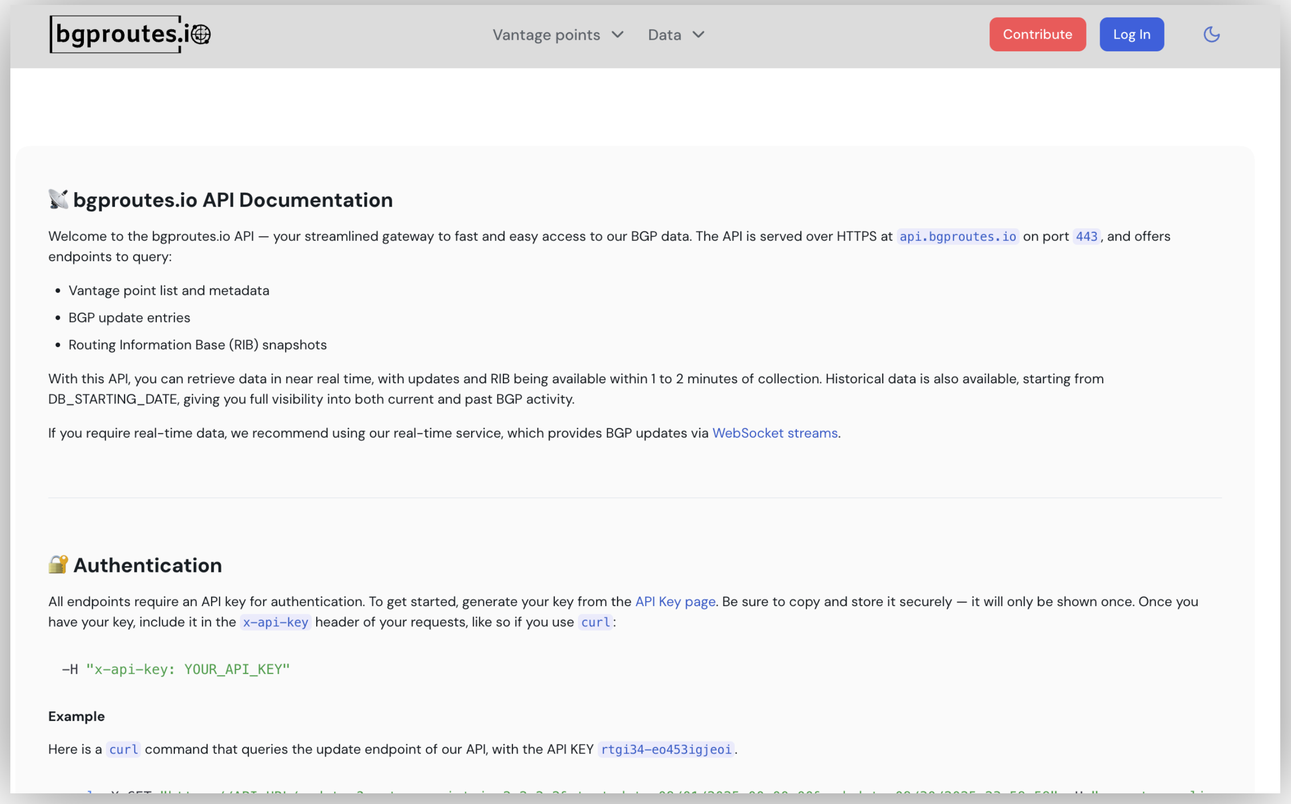
# bgproutes.io comes with a simple and fast API that provides high-granularity access to BGP data

The API offers three endpoints:

vantage_points

updates

rib




bgproutes.io/data_api

# Imagine an network operator trying to retrieve all BGP updates for which their ASN appear in the AS path on May 14, 2025

```
# Let’s use our Python client to retrieve the data
from pybgproutesapi import vantage_points, updates
```

# Imagine an network operator trying to retrieve all BGP updates for which their ASN appear in the AS path on May 14, 2025

```
# Let’s use our Python client to retrieve the data
from pybgproutesapi import vantage_points, updates

# Let’s retrieve the vantage points we want to use
vps = vantage_points(source=[‘bgproutes.io', ‘ris’])
```

# Imagine an network operator trying to retrieve all BGP updates for which their ASN appear in the AS path on May 14, 2025

```
# Let’s use our Python client to retrieve the data
from pybgproutesapi import vantage_points, updates

# Let’s retrieve the vantage points we want to use
vps = vantage_points(source=[‘bgproutes.io', ‘ris’])

# Finally, we retrieve the updates and print them for every VP
for vp in vps:
    vp_upd = updates(
        vp_ip=vp,
        start_date=“2025-05-14T00:00:00”,
        end_date=“2025-05-15T00:00:00”,
        aspath_regexp=“^3333 | 3333 | 3333$“

    print (vp_upd)
```

# bgproutes.io

## Our new next-gen BGP route collection platform

- Collection: High coverage
- Storage: Low data management cost
- Distribution: Fast API with high granularity

https://bgproutes.io

You can peer with us!