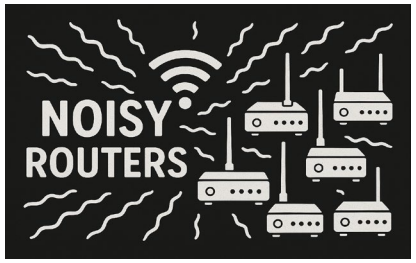
Noisy Routers: Investigating the Make-Up of Route Collector Data

Ebrima Jaw[†], **Joint work with**: Thomas Krenc[‡§], Moritz Müller[†*], kc claffy[§], Lambert Nieuwenhuis[†], Cristian Hesselman[†*]

[†]University of Twente, Enschede, The Netherlands
[§]CAIDA, UC San Diego La Jolla, CA, USA
[*]SIDN Labs, Arnhem, The Netherlands
[‡]IIJ Research Laboratory, Japan

MAPRG, IETF-126, Vienna, Austria

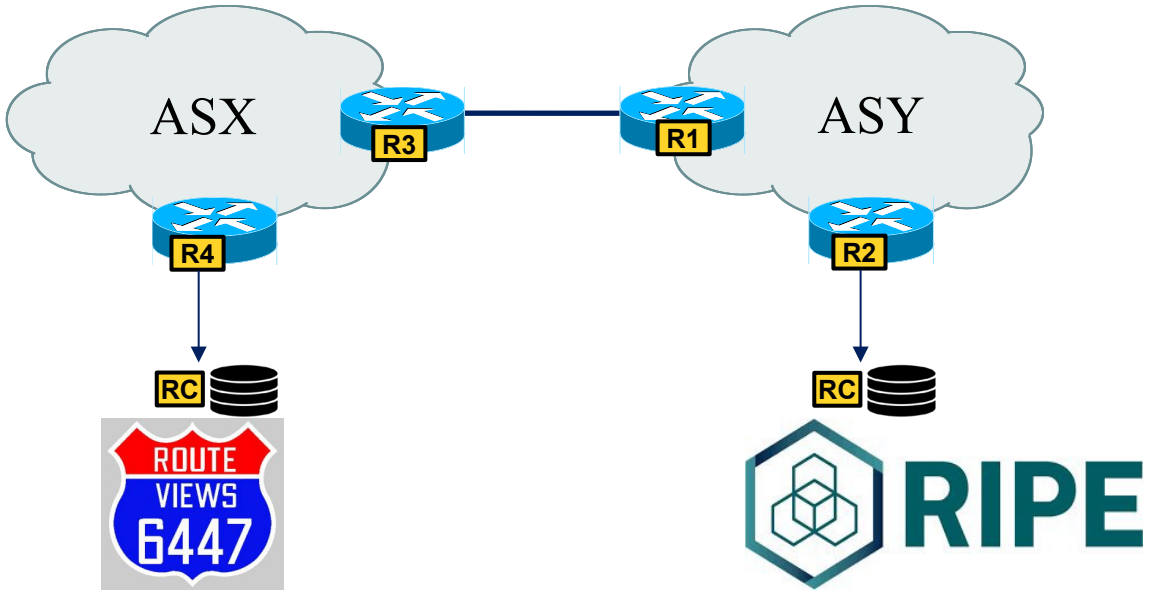
23.07.2026



Route Collectors: Overview & Significance

A subset of ASes (**peers**) share their routing data with **RouteViews** and **RIPE RIS** in **MRT** format

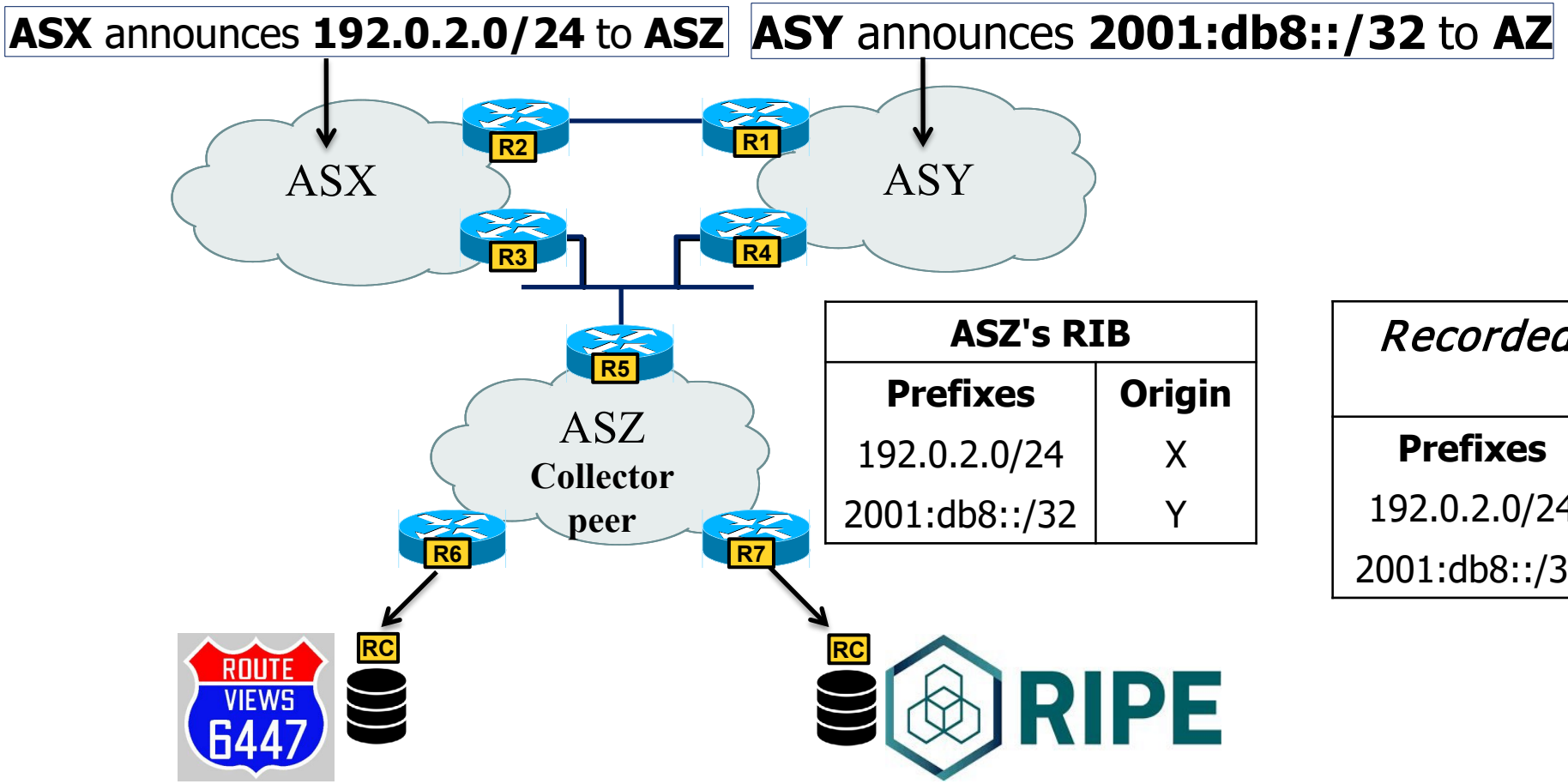
The collected data support both **operational** and **research** activities:



- *Operators:* Network monitoring, troubleshooting, *&* traffic engineering
- *Researchers:* Internet topology analysis and routing security research

Expected Normal BGP Announcement Operation

Typically, **routers** should only trigger updates to reflect **configuration changes**:
Newly announce, **withdrawn prefixes** or **link failures**

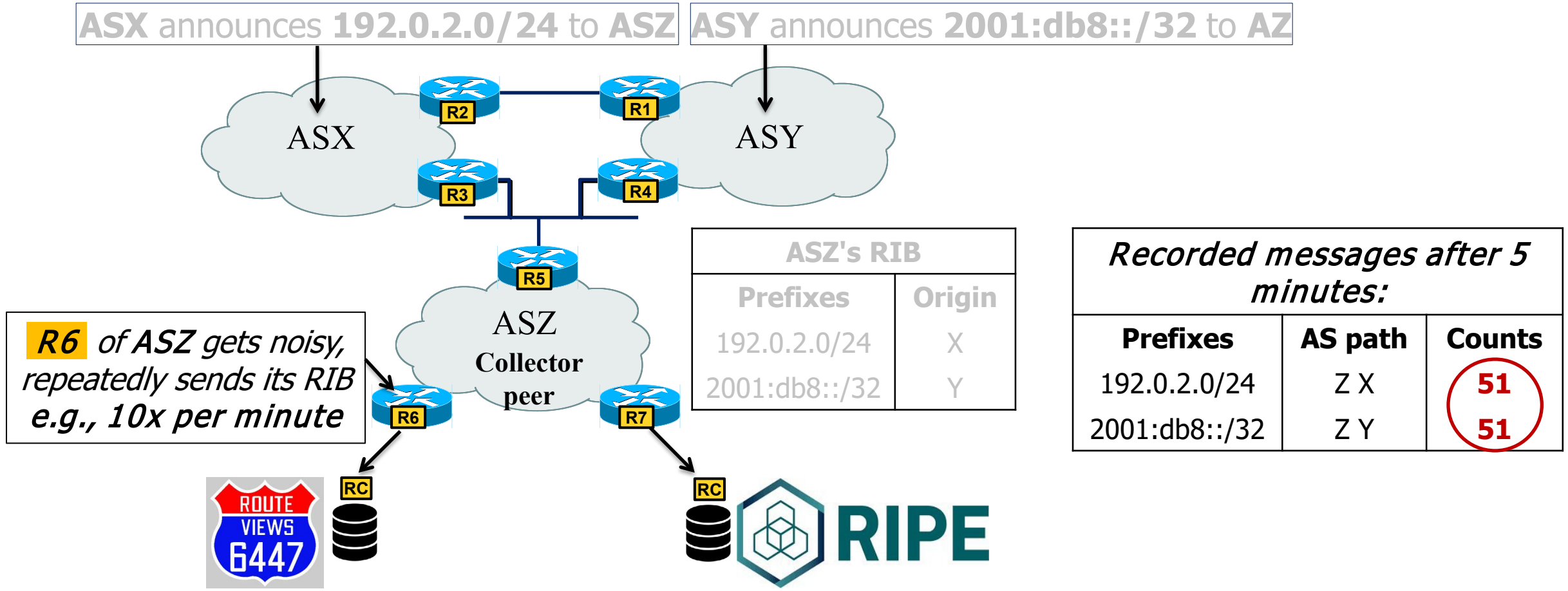


ASZ's RIB	
Prefixes	**Origin**
192.0.2.0/24	X
2001:db8::/32	Y

Recorded messages after 5 minutes:		
Prefixes	**AS path**	**Counts**
192.0.2.0/24	Z X	1
2001:db8::/32	Z Y	1

Potential Noisy Routers: Continuous RIB Re-announcements

A **buggy** or **misconfigured** router can generate **significant unnecessary announcements**



ASZ's RIB	
Prefixes	**Origin**
192.0.2.0/24	X
2001:db8::/32	Y

Recorded messages after 5 minutes:

Prefixes	AS path	Counts
192.0.2.0/24	Z X	**51**
2001:db8::/32	Z Y	**51**

Our Motivation

Data management challenges for route collector projects

Re: Noisy prefixes in BGP

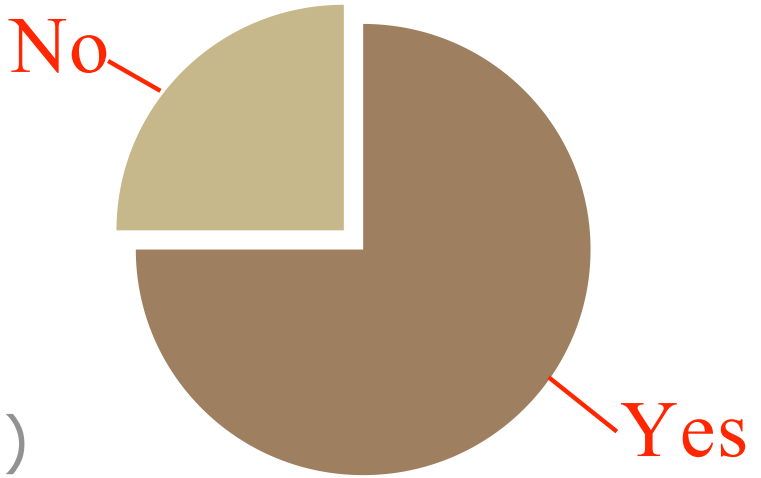
From: [redacted] <pfs () routeviews org>
Date: Mon, 10 Feb 2025 09:42:09 +1000

Route Collection at the RIPE NCC - Where are we and where should we go?

[redacted]

- Increased storage and computational overhead for researchers

Do you find the data from RIS and RouteViews expensive to process in terms of computational resources?



Survey conducted among BGP researchers (Alfroy et al.)

Our Main Goal:
Quantify and Characterize the Composition of RouteViews MRT Files

Our Definition of Noise in BGP Data & Its Impact

- Repeated announcements for the same prefix with unchanged attributes within a short time window (e.g., a minute).
- We consider these announcements as **noise** because they contribute little or no new routing information.
- Increases router processing load
- Unnecessarily inflate MRT file sizes

Datasets and Preprocessing

- We extracted **daily update share per collector peer** over 13 years

Start Date: 2012-01-01		**End Date**: 2025-02-28
Collectors	Peer ASNs	Update Counts (A/W)
36	649	2.6 Trillion

Datasets and Preprocessing

- We extracted daily update share per collector peer over 13 years

Start Date: 2012-01-01		**End Date**: 2025-02-28
Collectors	Peers ASNs	Update Counts (A/W)
36	649	2.6 Trillion

- We also used two months of recent RouteViews data (**83.17 billion updates**) to **quantify** and **characterize** the composition of MRT files.

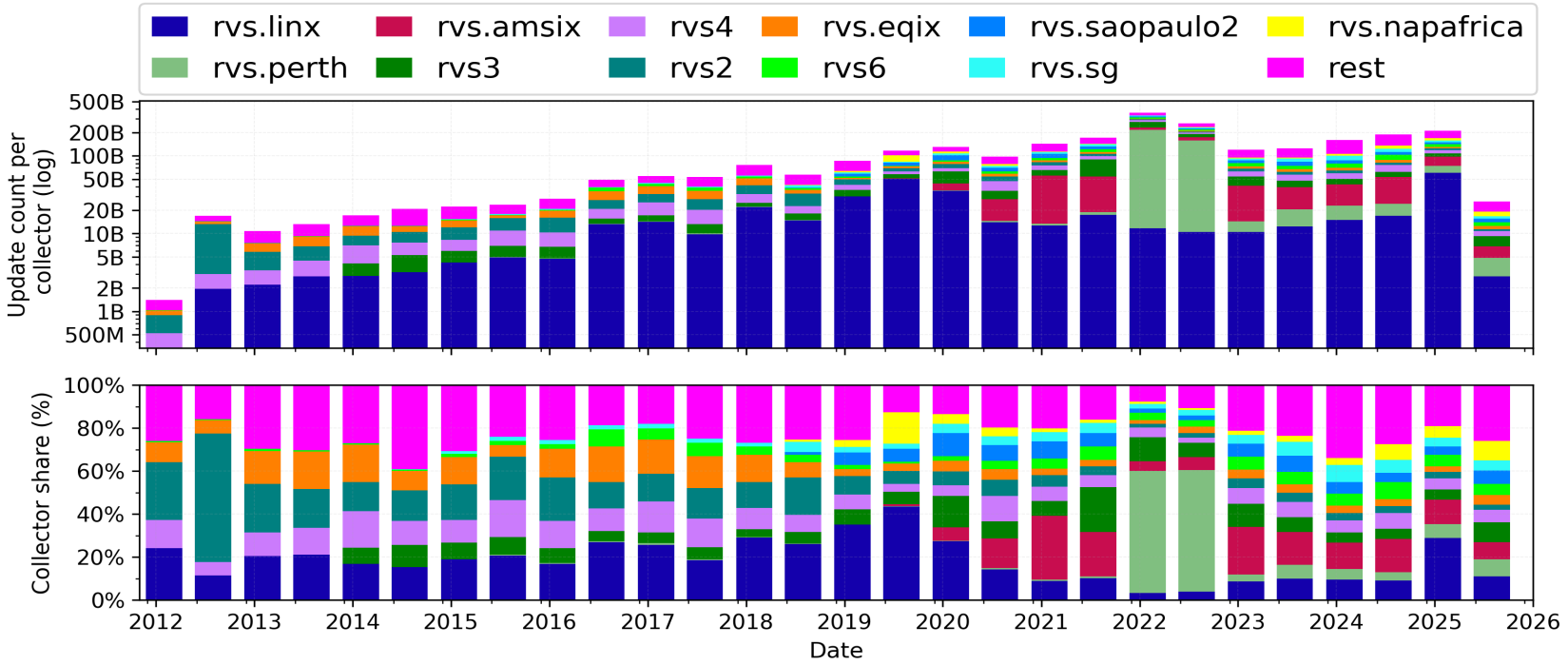
	December 2021		December 2024	
	IPv4	IPv6	IPv4	IPv6
MRT files	104.15K (86.84K)		98.26K (93.47K)	
Sessions	505	401	682	592
Peer ASNs	305	230	321	278
Announcements	38.67B	11.16B	11.79B	18.52B
Unique prefixes	1.21M	307.23K	1.22M	272.87K

Goal #1:
Measure update distribution across RouteViews collectors

Methodology: Quantifying Update Share Across Collectors

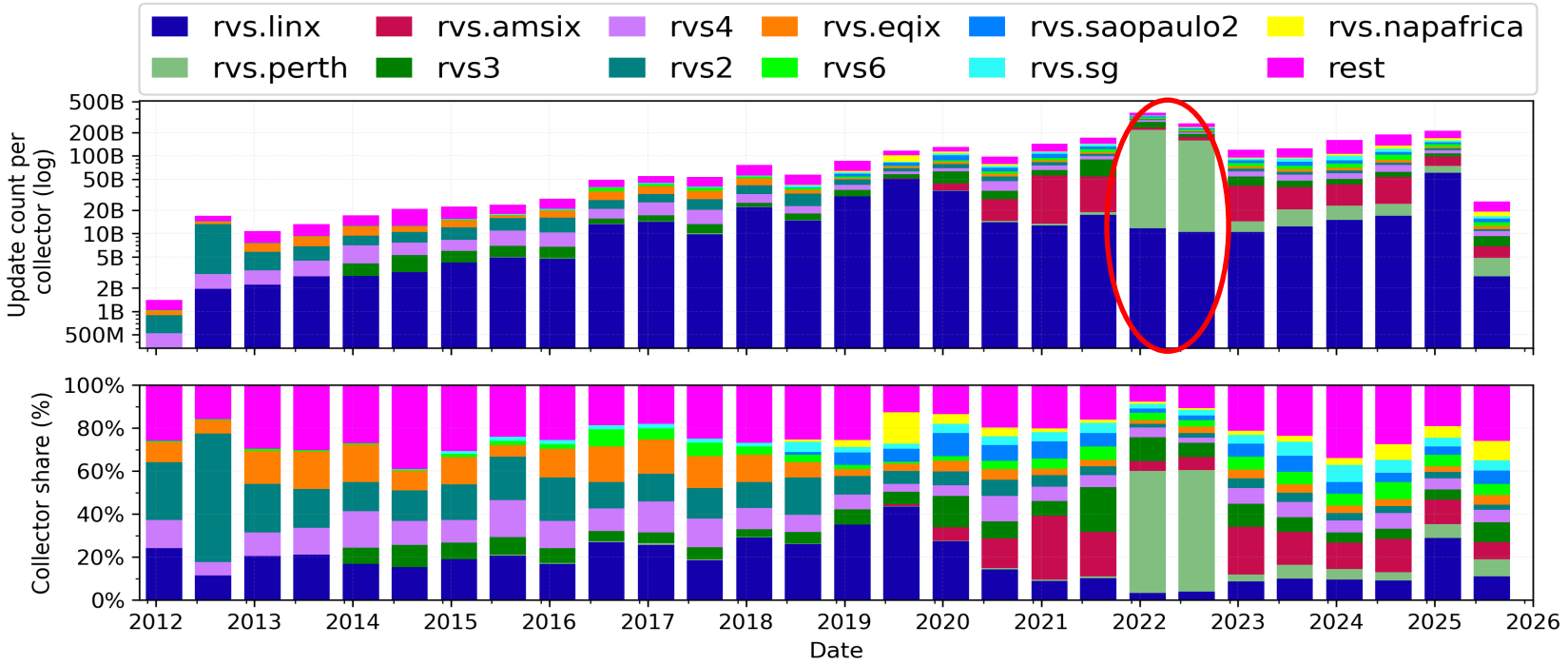
- Daily update counts per collector
- We **aggregate daily totals** into **six-month** intervals
- We compute each collector's **average update volume** per interval
- Rank the collectors by **average update volume**
- We select the **top 11 collectors** and **aggregate** all others as "**rest**"

Results #1 Update Share Across Collectors



Of the 36 collectors, the top 5 collectors accounted for **54% of updates** for the 13 years.

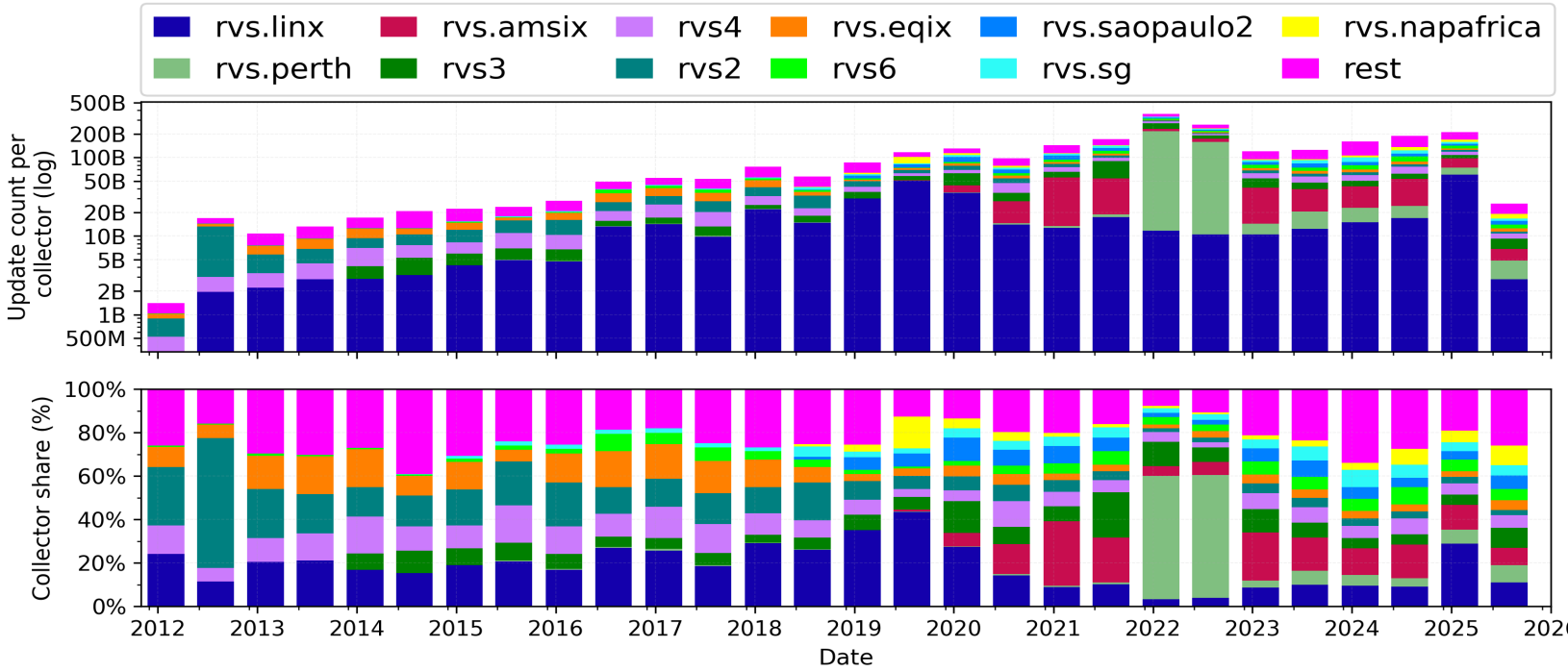
Results #1 Update Share Across Collectors



Of the 36 collectors, the top 5 collectors accounted for **54% of updates** for the 13 years.

Perth and **Linx** together accounted for 30.4% of all updates across the 13 years

Results #1 Update Share Across Collectors

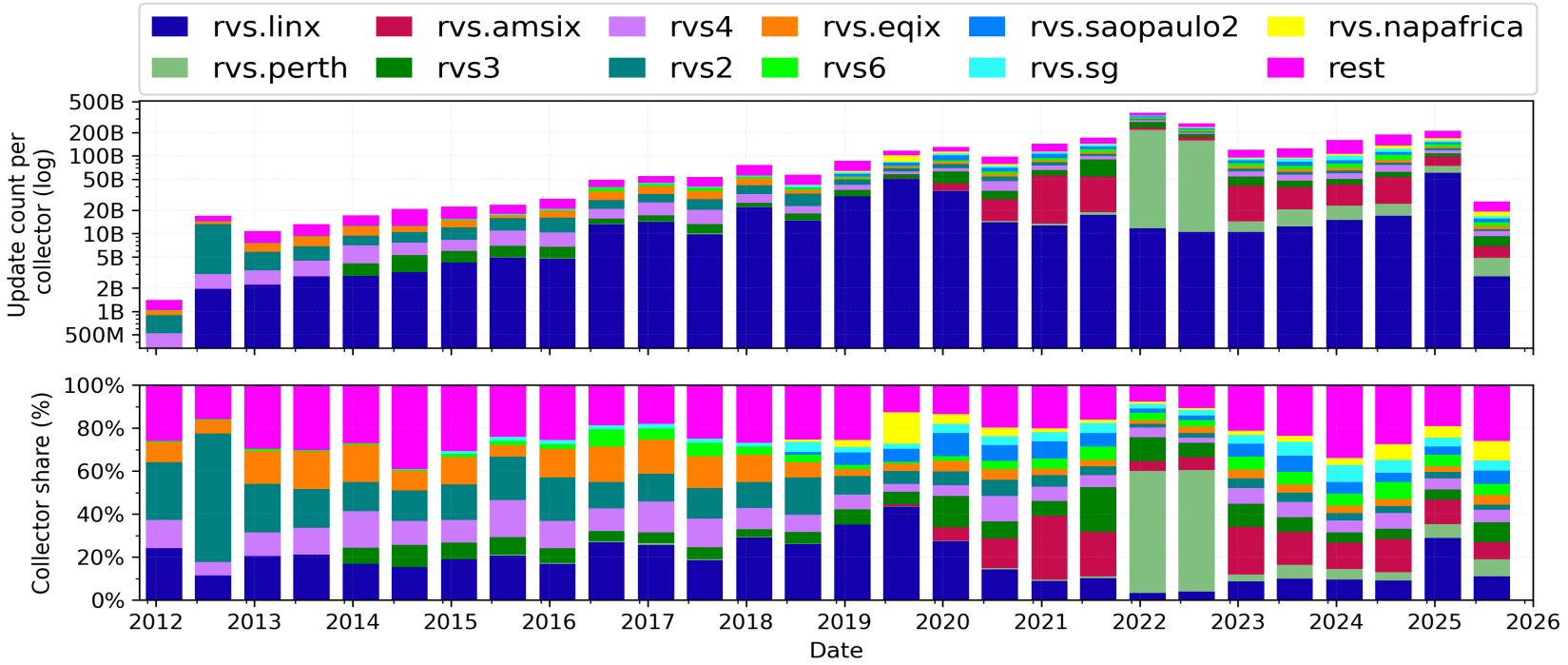


Of the 36 collectors, the top 5 collectors accounted for **54% of updates** for the 13 years

Perth and **Linx** together accounted for 30.4% of all updates across the 13 years

Mostly **repetitive announcements, less withdrawals**

Results #1 Update Share Across Collectors



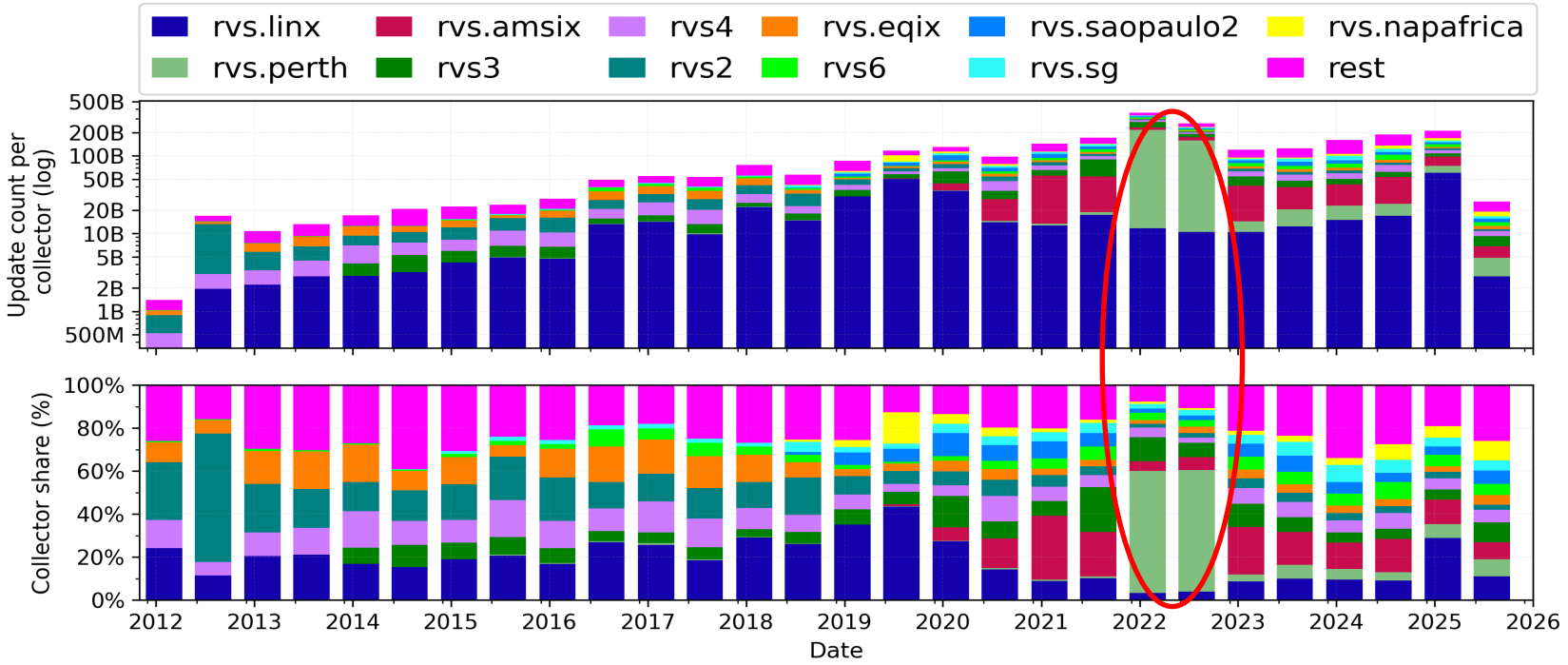
Of the 36 collectors, the top 5 collectors accounted for **54% of updates** for the 13 years.

Perth and **Linx** together accounted for 30.4% of all updates across the 13 years

Mostly **repetitive announcements**, **less withdrawals**

More than **one-third** of the collectors were each responsible for fewer than **1%** of updates.

Results #1 Update Share Across Collectors



Of the 36 collectors, the top 5 collectors accounted for **54% of updates** for the 13 years.

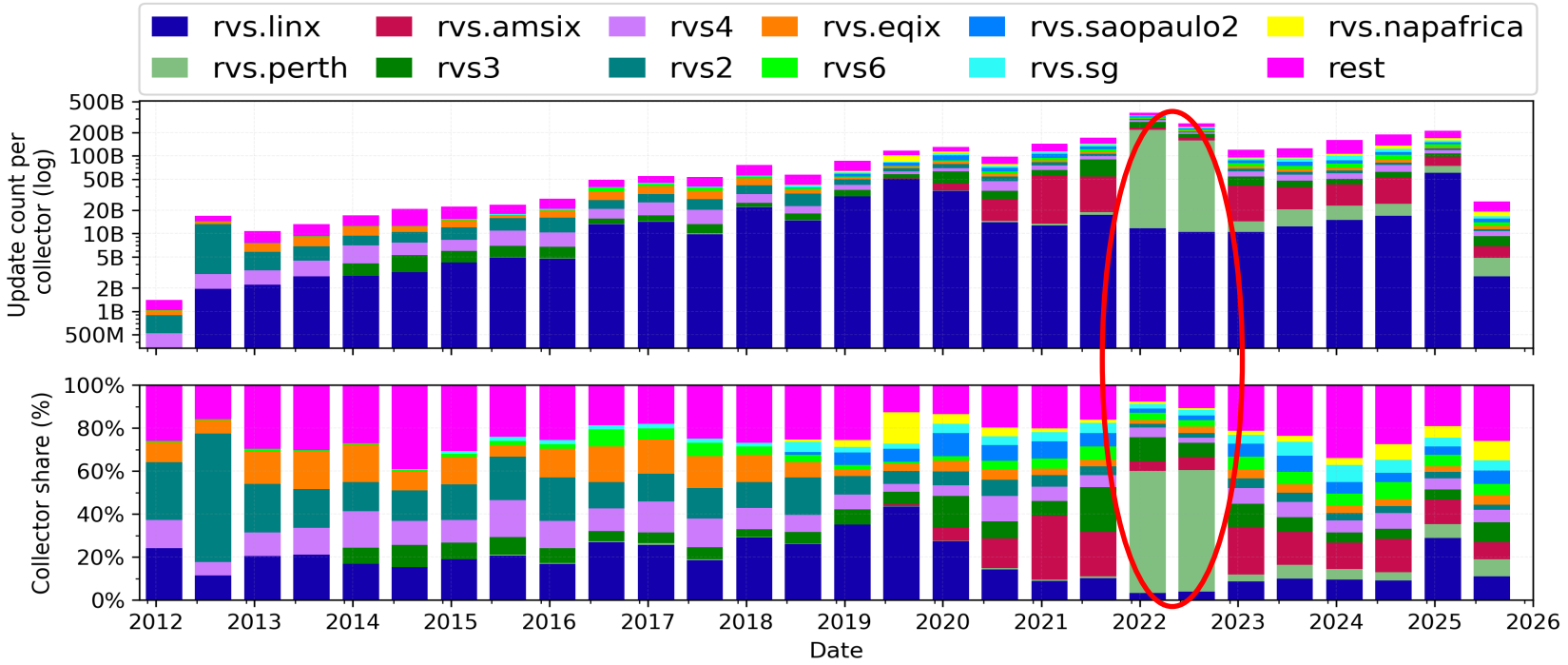
Perth and **Linx** together accounted for 30.4% of all updates across the 13 years

Mostly **repetitive announcements**, **less withdrawals**

More than **one-third** of the collectors were each responsible for fewer than 1% of updates.

Perth recorded the largest routing event ever with **74B** updates in Nov 2021 (**3%** of all updates across collectors)

Results #1 Update Share Across Collectors



Of the 36 collectors, the top 5 collectors accounted for **54% of updates** for the 13 years.

Perth and **Linx** together accounted for 30.4% of all updates across the 13 years

Mostly **repetitive announcements**, **less withdrawals**

More than one-third of the collectors were each responsible for fewer than 1% of updates.

Perth recorded the largest routing event ever with **74.3B** updates in Nov 2021 (**3%** of all updates across collectors)

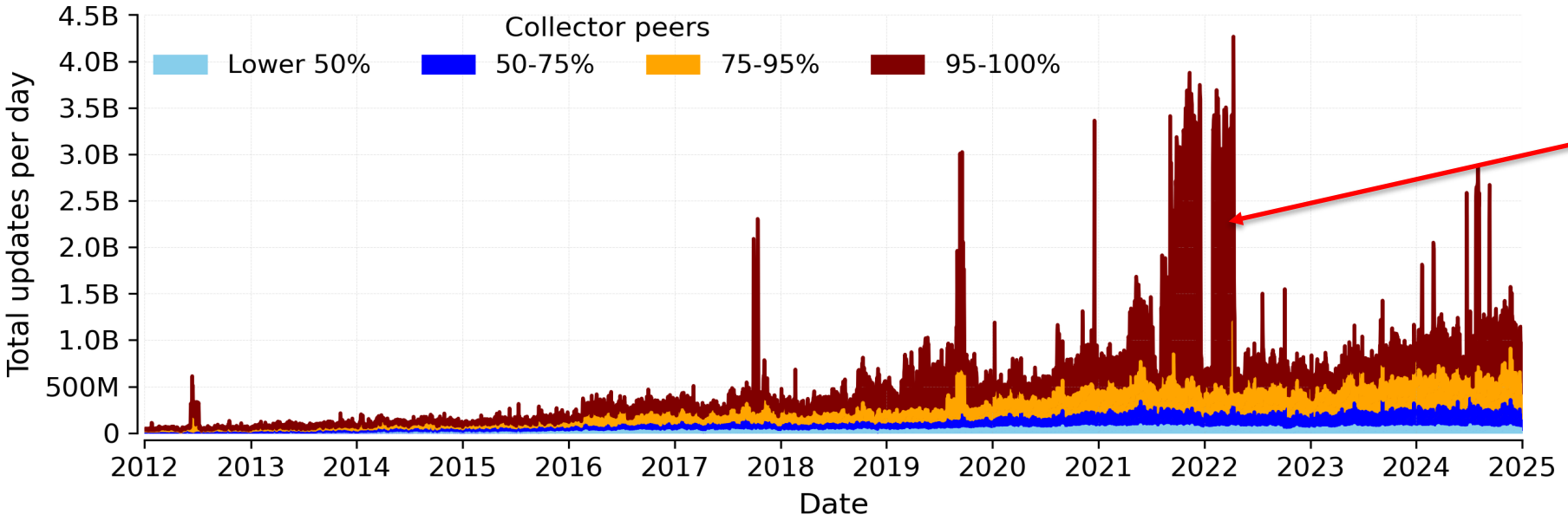
Key Takeaways: Update distribution across collectors are **heavily skewed**, & not correlated to **peer counts**

Goal #2:
Evaluate whether high-frequency updates are concentrated among a few peers

Methodology: Collector–Peer Contribution Analysis

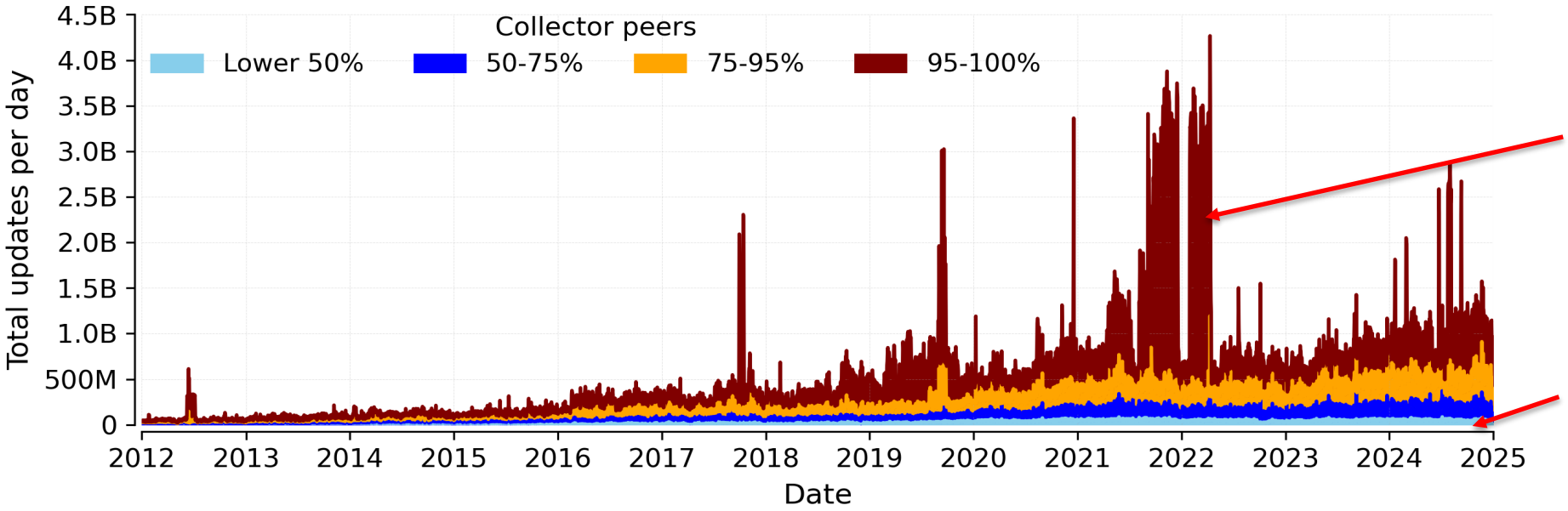
- We construct daily update time series for each collector–peer pair
- We rank collector–peer pairs by daily update count for comparability
- We assigned the ranked peers into daily quantiles (95–100%, and lower 50%)
- We quantify the **top contributors** and aggregate the **least noisy pairs** as "others"

Results #2: Collector Peer Update Contributions



Of the **2.6T** updates, the top 5% collector-peers accounted for **1.5T** (**55.86%**) of total updates

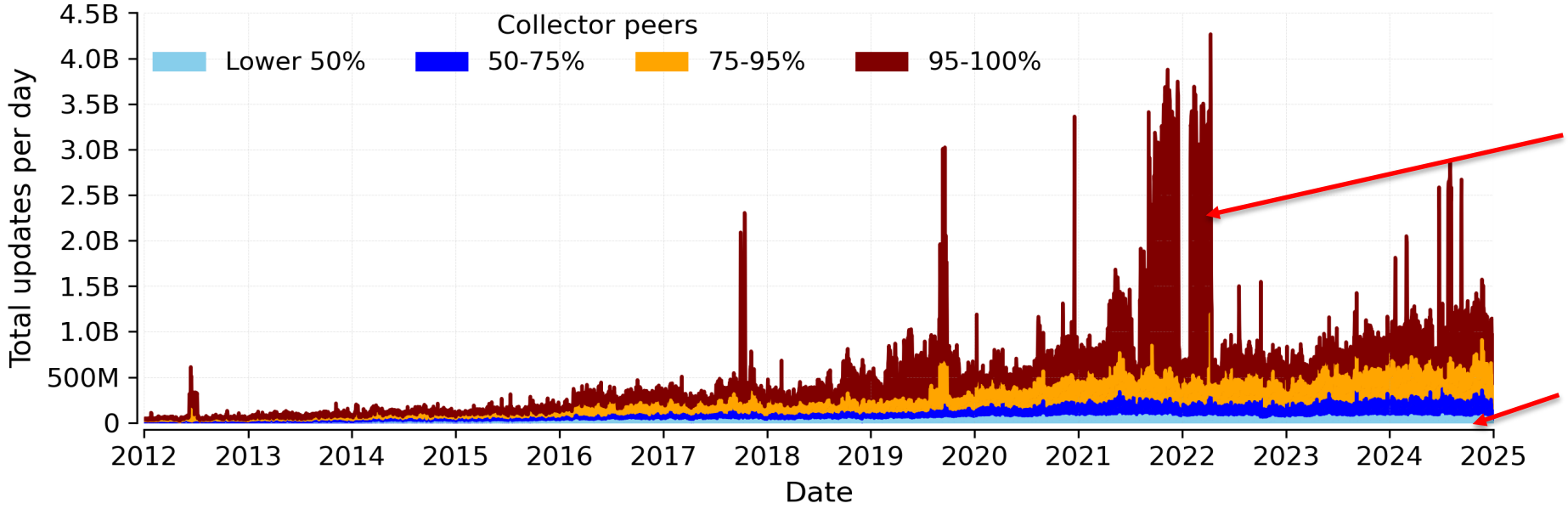
Results #2: Collector Peer Update Contributions



Of the 2.6T updates, the top 5% collector-peers accounted for 1.5T (55.86%) of total updates.

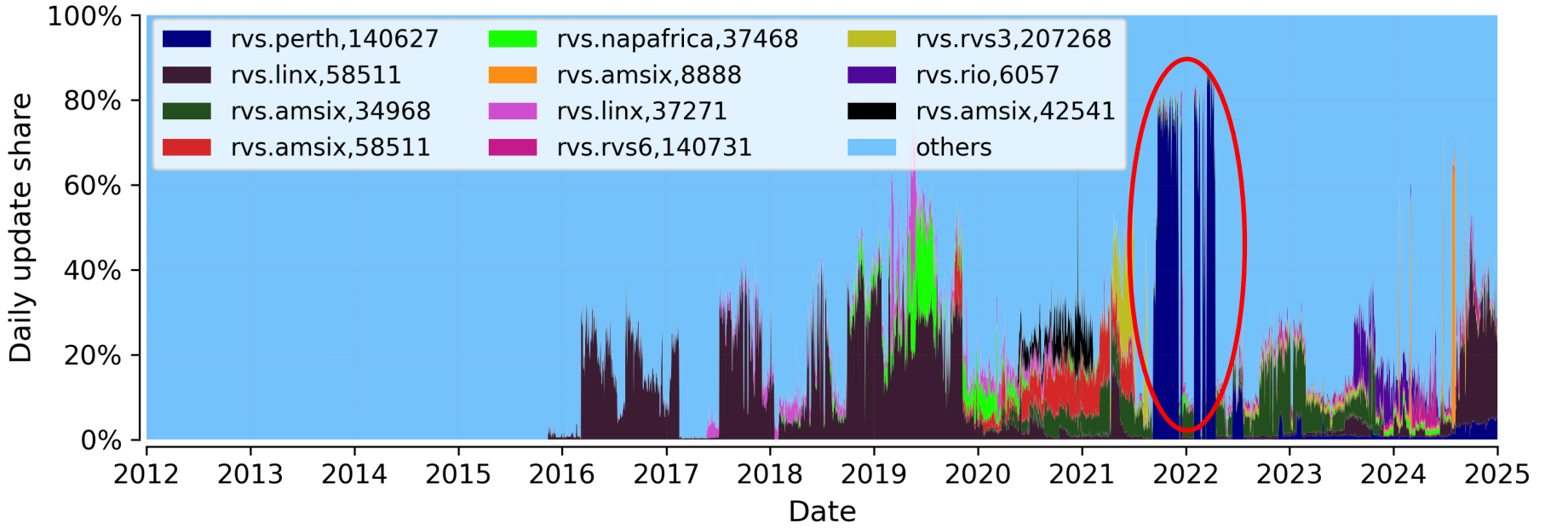
In contrast, the **lower 50%** of peers generated **~137B** (**5.17%**) of the total updates

Results #2: Collector Peer Update Contributions



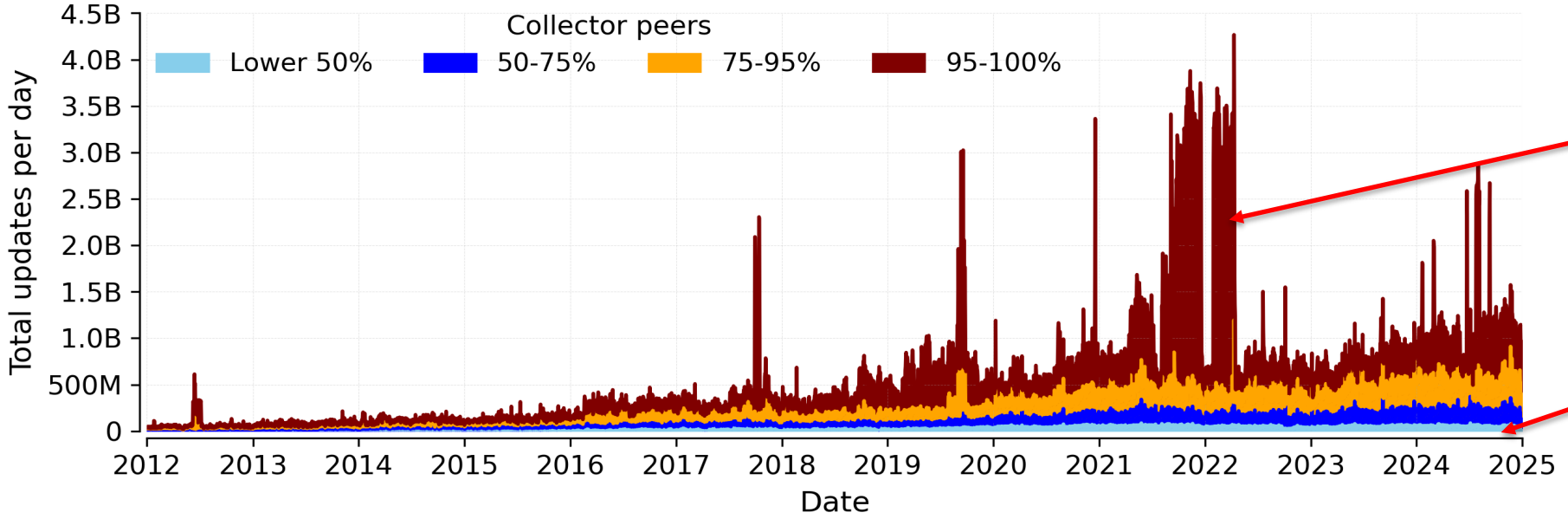
Of the 2.6T updates, the top 5% collector-peers accounted for 1.5T (55.86%) of total updates.

In contrast, the lower 50% of peers generated 136.5B (5.17%) of total updates



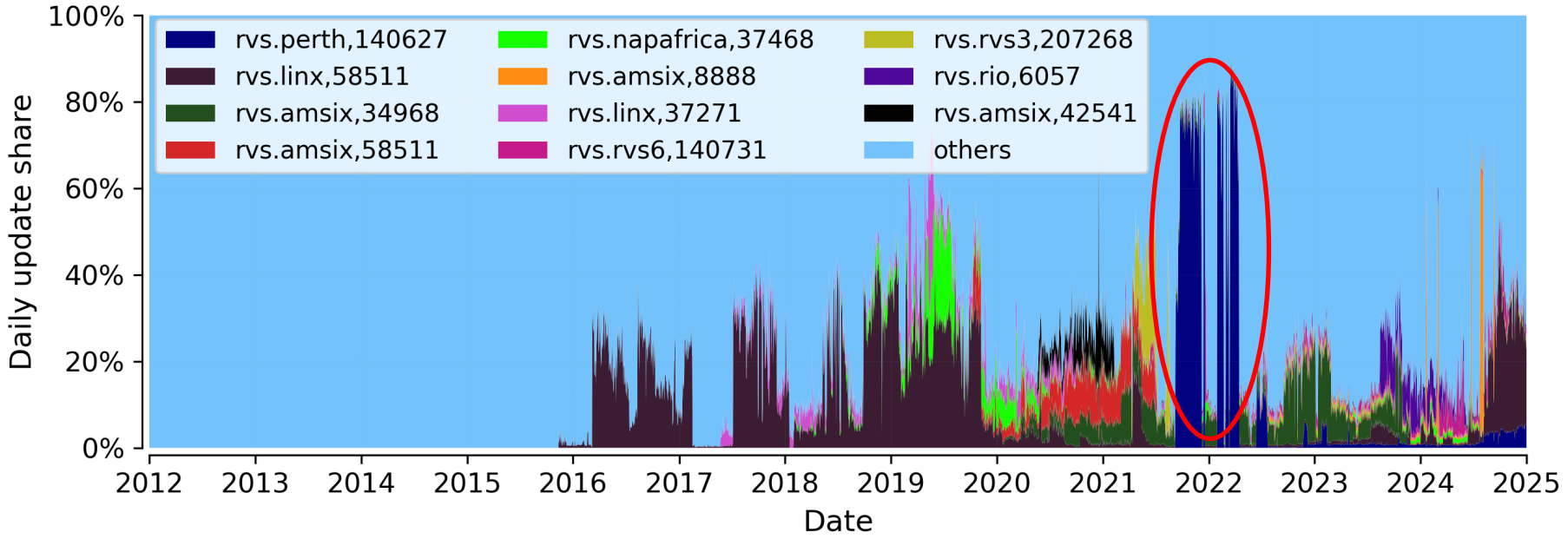
AS140627 alone contributed **14%** of all updates, and up to **88%** on some days.

Results #2: Collector Peer Update Contributions



Of the 2.6T updates, the top 5% collector-peers accounted for 1.5T (55.86%) of total updates.

In contrast, the lower 50% of peers generated 136.5B (5.17%) of total updates



AS140627 alone contributed 14% of all updates, and up to 88% on some days.

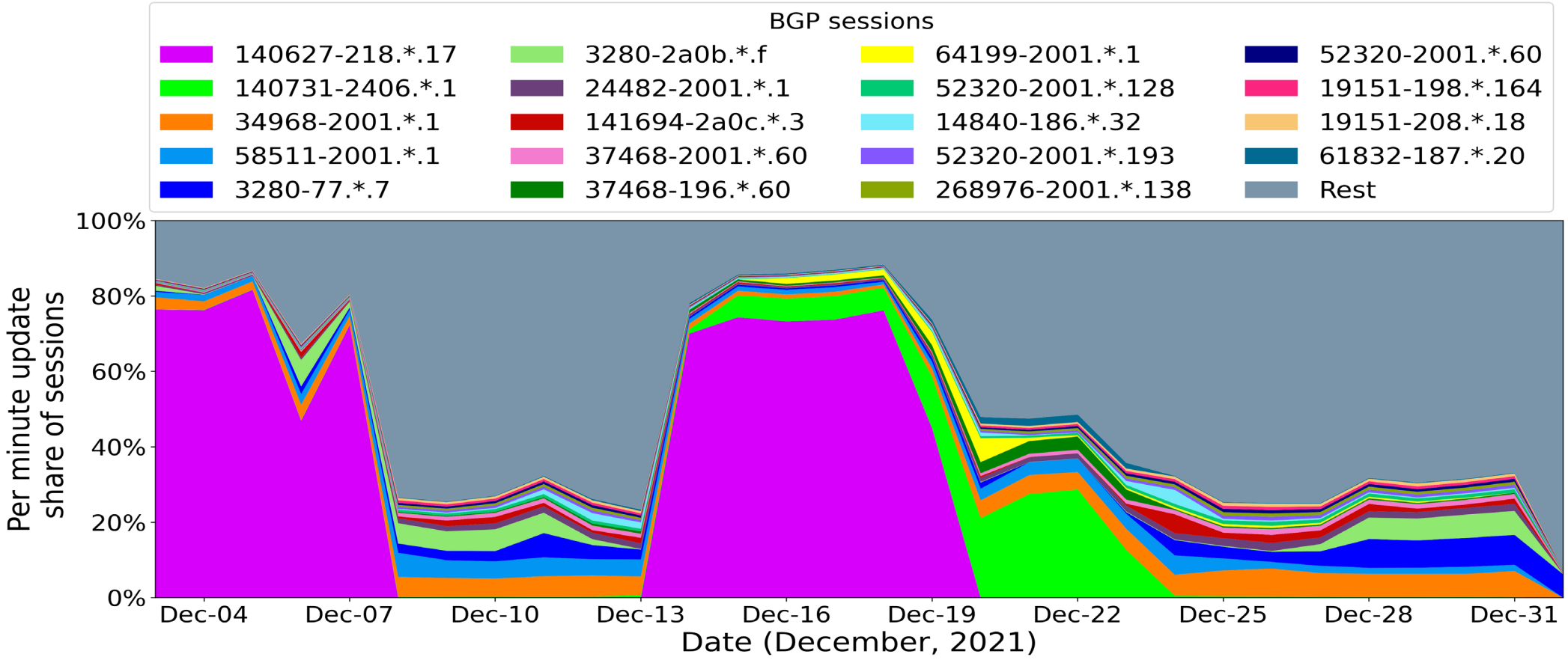
Key Takeaway: Noise is both *prevalent* and *unpredictable*, & it is driven by few col. peers.

Goal #3:
Measure how unevenly updates are distributed across *sessions*, *prefixes*, and *AS paths*

Methodology: Statistical Analysis of Update Variability

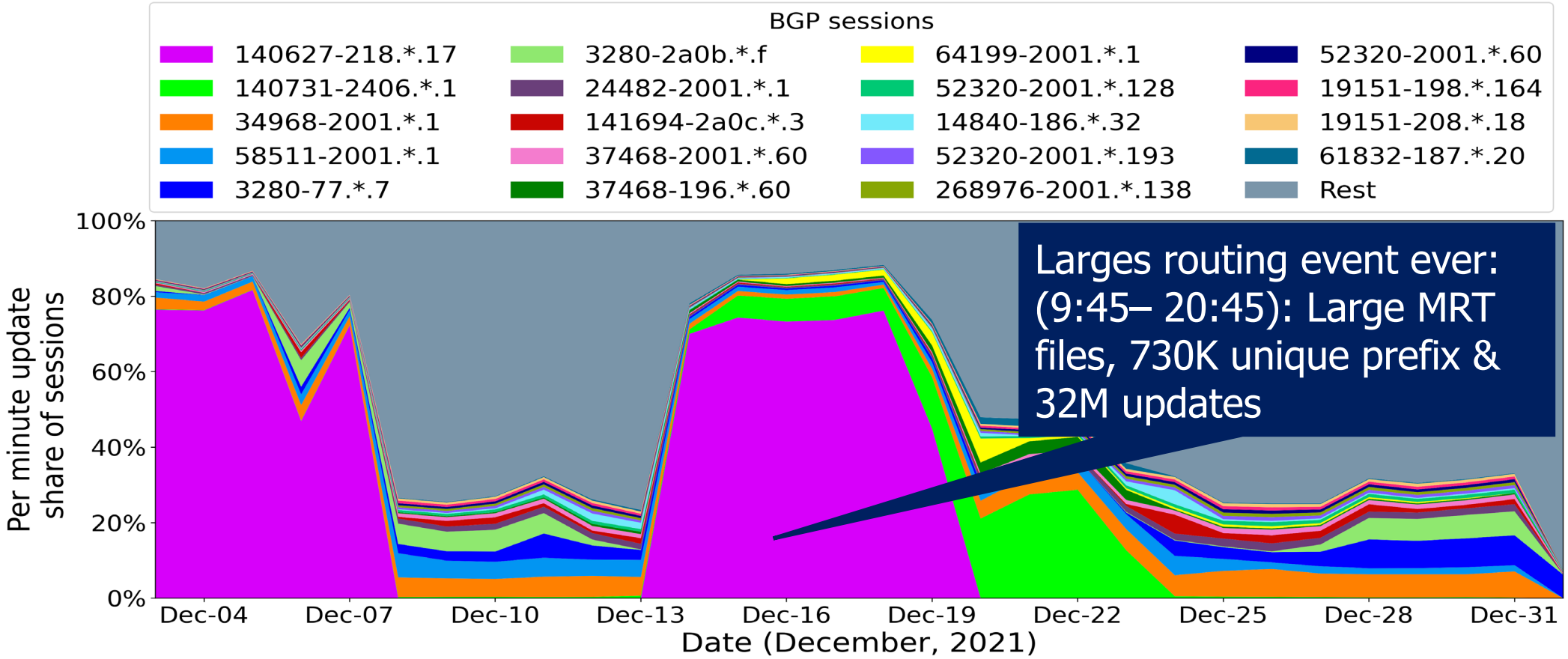
- We compute per-minute update share for each BGP session over the two months.
- We applied statistical methods to quantify update concentration and variability
- **Concentration** captures dominant sources; **variability**: fluctuating update volumes
- We apply the same methods for the **prefixes** and **AS paths** analysis.

Results #3: Update Distributions Across Sessions



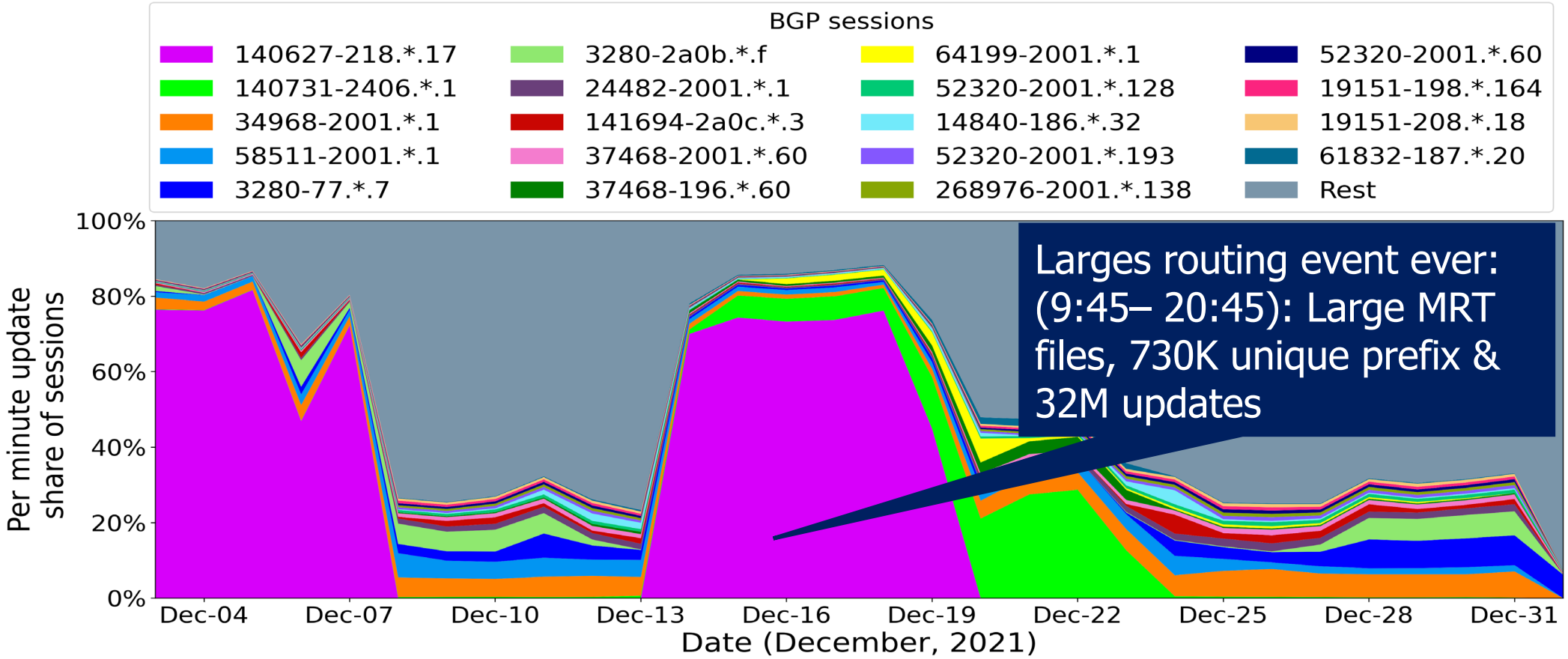
Only **19** (**2%**) of sessions accounted for **71.32%** of updates for Dec 2021, while **98%** of the sessions contributed only **28.68%** of updates.

Results #3: Update Distributions Across Sessions



Only 19 (2%) of sessions accounted for 71.32% of updates for Dec 2021, while 98% of the sessions contributed only 28.68% of updates.

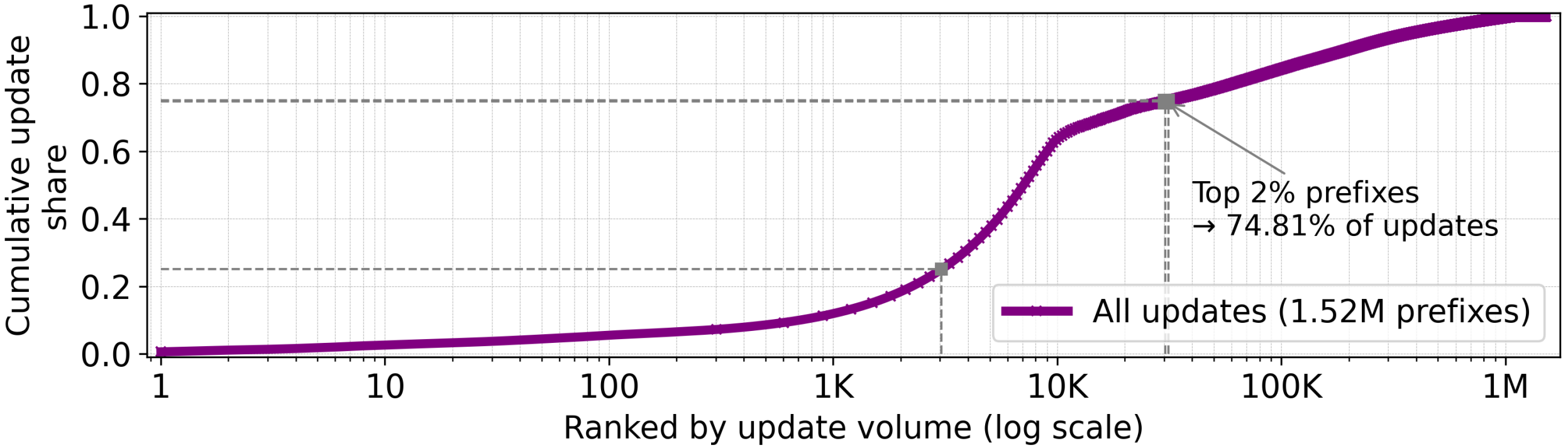
Results #3: Update Distributions Across Sessions



Only 19 (2%) of sessions accounted 71.32% of updates for Dec 2021, while 98% of the sessions contributed only 28.68% of updates.

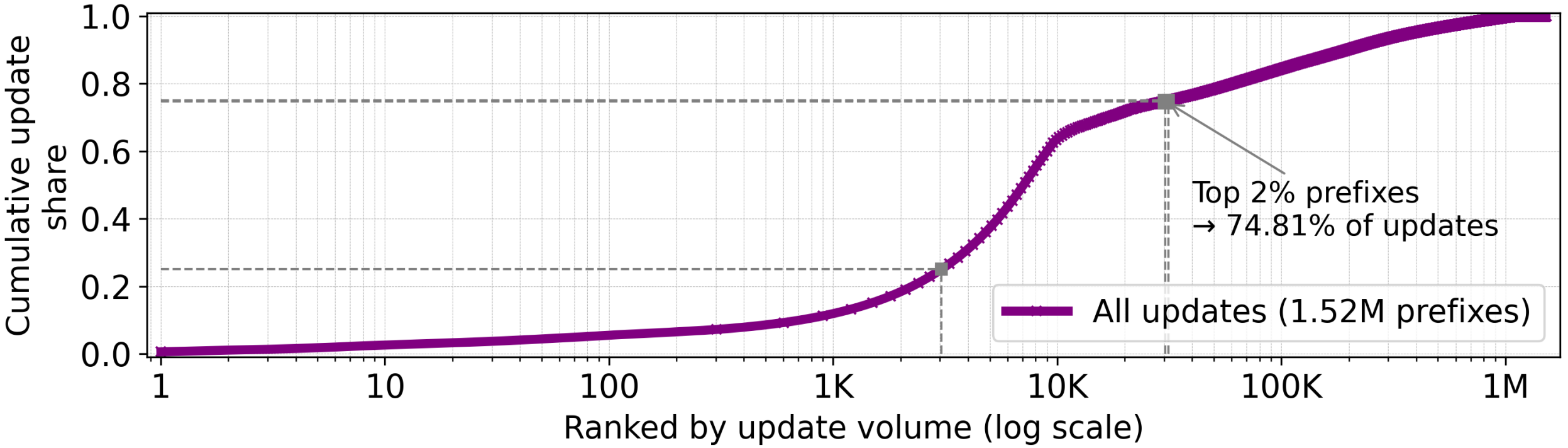
Takeaway: Small subset of sessions consistently generates most of the repeated updates (*redundant*)

Results #4: Update Distribution Across Prefixes



Only **2%** of prefixes appeared in over **two-thirds** of the total updates in Dec 2021

Results #4: Update Distribution Across Prefixes



Only **2%** of prefixes appeared in over **two-thirds** of the total updates in Dec 2021

~75% of prefixes appeared in only **5% of updates**, indicating relatively stable behavior.

Key Takeaways: A small fraction of **unstable prefixes** contributed a disproportionate share of updates

Key Takeaways and Operator Feedback

- Noise in BGP data is both **prevalent** and **unpredictable**, driven by a few peers

Key Takeaways and Operator Feedback

- Noise in BGP data is both **prevalent** and **unpredictable**, driven by a few peers
- A Few of the **sessions**, **prefixes** & **AS paths** contributed large redundant updates

Key Takeaways and Operator Feedback

- Noise in BGP data is both **prevalent** and **erratic**, driven by a few collector peers
- A Few of the **sessions**, **prefixes** & **AS paths** contributed large redundant updates
- **Prefixes** are **not universally noisy**, often confined to a few collectors (e.g., Perth)

Key Takeaways and Operator Feedback

- Noise in BGP data is both **prevalent** and **erratic**, driven by a few collector peers
- A Few of the **sessions**, **prefixes** & **AS paths** contributed large redundant updates
- **Prefixes** are **not universally noisy**, often confined to few collectors (e.g., Perth)
- **Noise** is largely due to **faulty** or **misconfigured** routers, not inherently from **origin ASes**

Key Takeaways and Operator Feedback

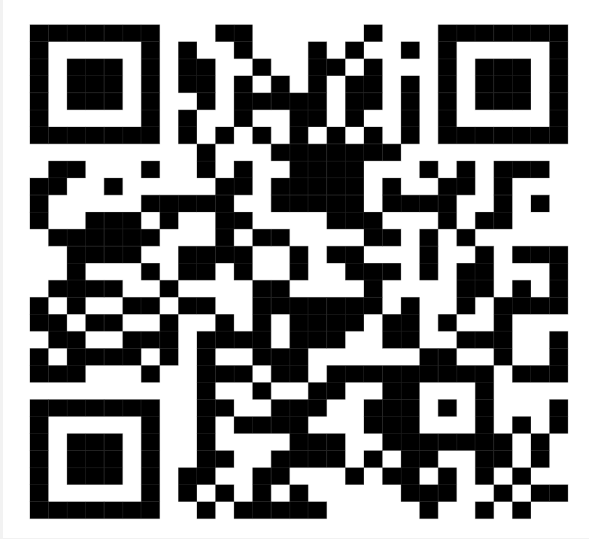
- Noise in BGP data is both **prevalent** and **erratic**, driven by a few collector peers
- A Few of the **sessions**, **prefixes** & **AS paths** contributed large redundant updates
- **Prefixes** are **not universally noisy**, often confined to few collectors (e.g., Perth)
- **Noise** is largely due to **faulty** or **misconfigured** routers, not inherently from origin **ASes**
- Redundant updates **unnecessarily inflate MRT** files without introducing **new routing state**.

Key Takeaways and Operator Feedback

- Noise in BGP data is both **prevalent** and **erratic**, driven by a few collector peers
- A Few of the **sessions**, **prefixes** & **AS paths** contributed large redundant updates
- **Prefixes** are **not universally noisy**, often confined to few collectors (e.g., Perth)
- **Noise** is largely due to **faulty** or **misconfigured** routers, not inherently from origin **ASes**
- Redundant updates **unnecessarily inflate MRT** files without introducing **new routing state**.

- **Call for Operator Feedback:**
  - Seeking operator insights into the root causes of observed announcement patterns

THANK YOU!



Please read our full
paper for more details

e.jaw@utwente.nl

UNIVERSITY
OF TWENTE.