

Seeing DNS Differently: Interactive Exploration of a Year of Passive DNS

OARC 47 – 2026/08/09

Kim M (kim@acmesecurity.org)

One Year of Passive DNS: Too much to see?

DNS traffic generates too much data to inspect directly

One Year of Passive DNS: Too much to see?

DNS traffic generates too much data to inspect directly

The traditional approach of aggregation can hide important patterns

One Year of Passive DNS: Too much to see?

DNS traffic generates too much data to inspect directly

The traditional approach of aggregation can hide important patterns

Relevance of the aggregation depends on the user: analysts, engineers, and security teams may need different views of the same data.

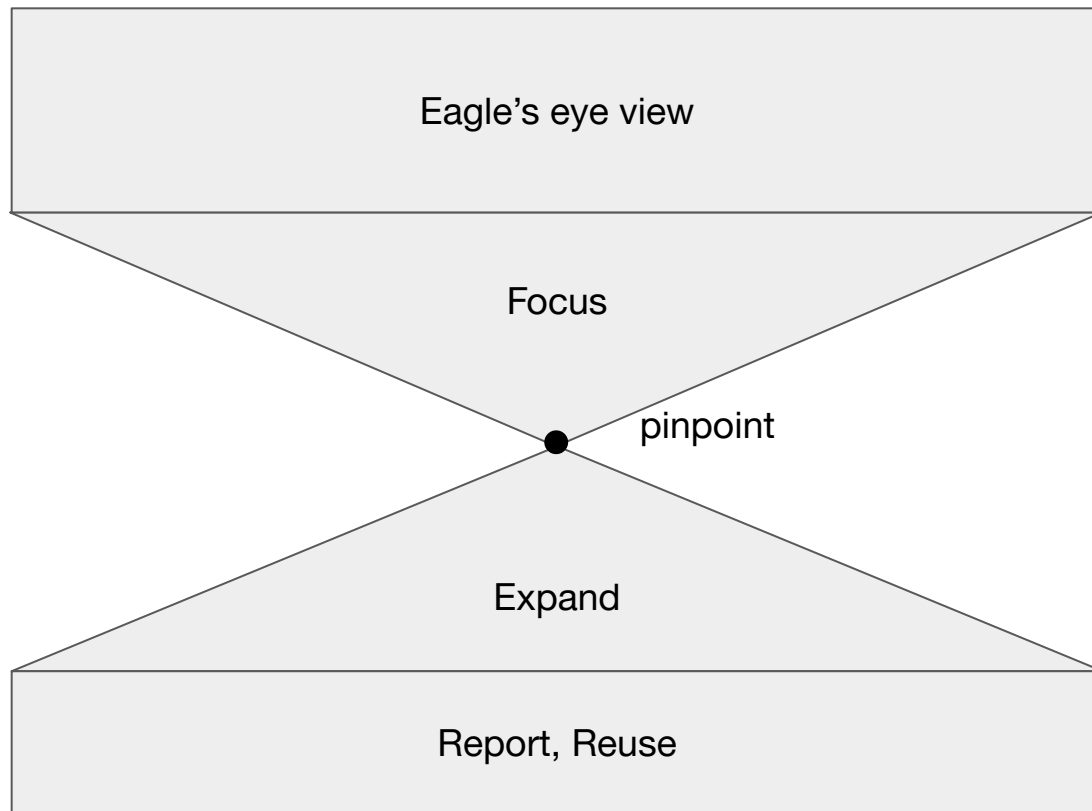
Why visualization by itself isn't enough

Visualization sparks curiosity and unexpected questions.

A fixed visualization creates friction:

Pursuing those questions often requires further development, of a visualization system, or work outside of it altogether.

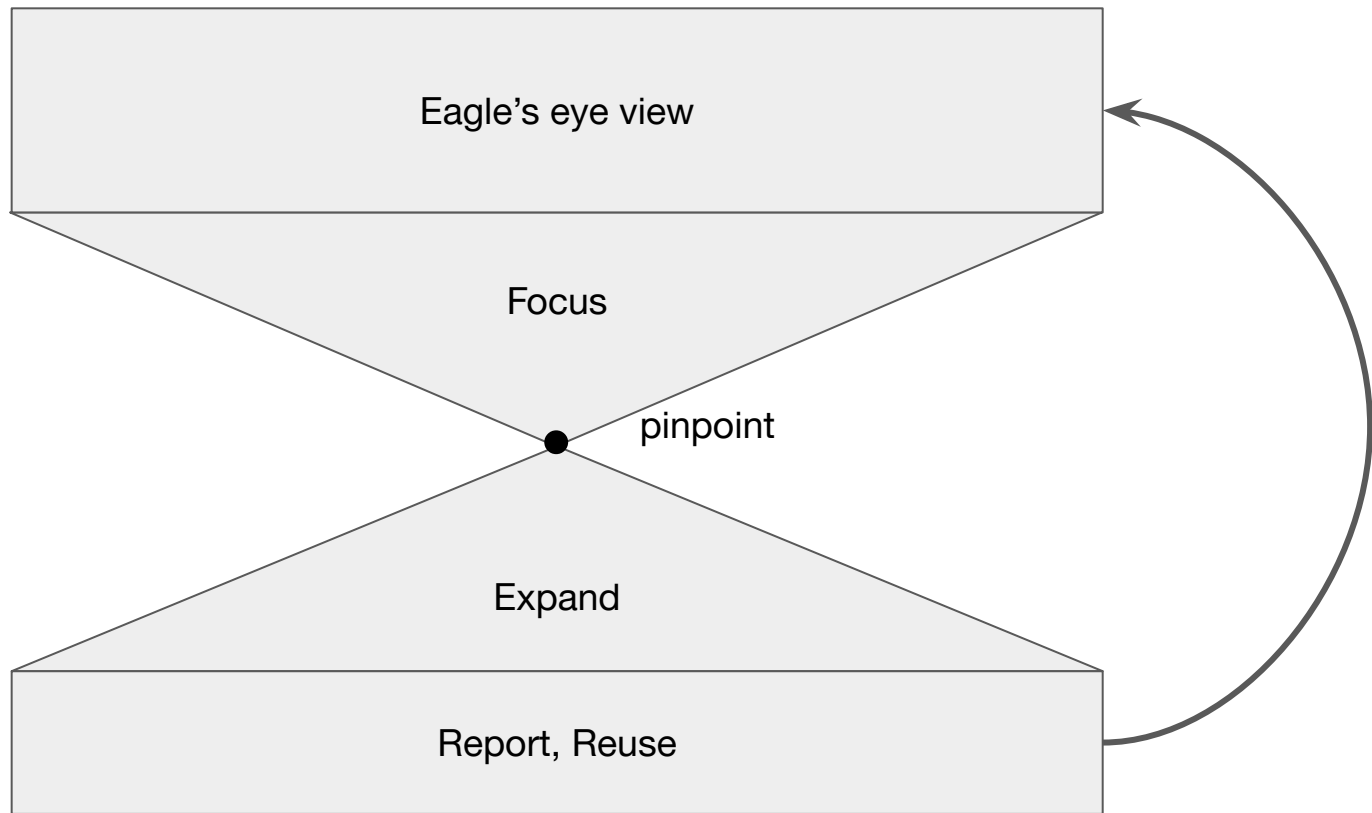
Our Approach: Visual Investigation



Our Approach: Visual Investigation

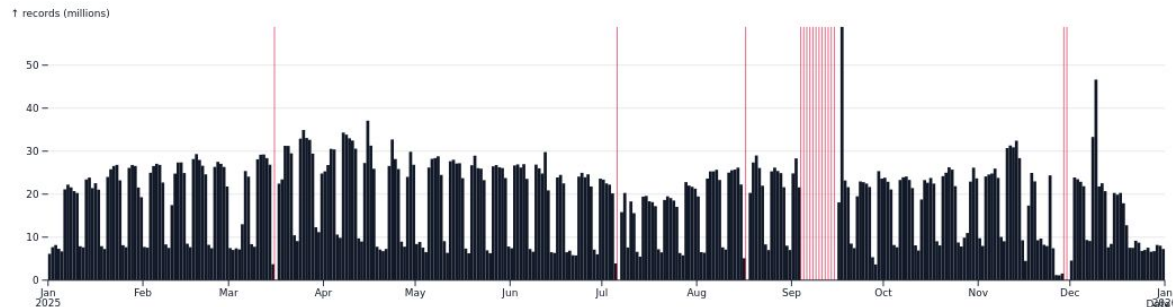
Eagle's eye view	Use main dashboards to detect overall patterns in data
Focus	Dynamically generate more detailed dashboards for a smaller time frame
Pinpoint	Directly investigate a behavior using pre existing components and dataloaders (and create your own!)
Report, Reuse	Report your findings visually and interactively, reuse the pipeline you created on other investigations

Our Approach: Visual Investigation

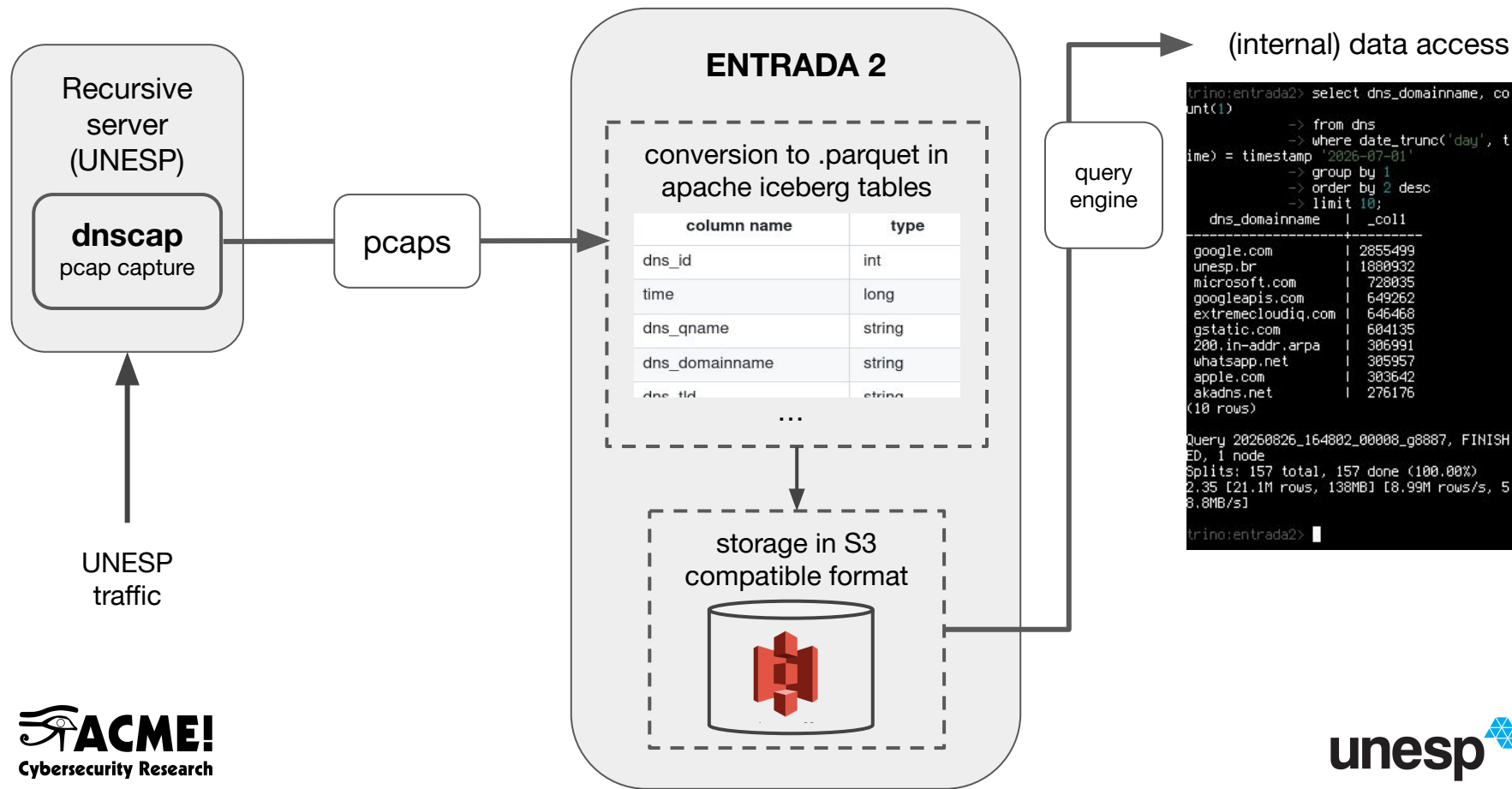


Dataset in this presentation

- Dataset from a primary recursive resolver in a multi campus university
- around 6.52 billion DNS records collected during 2025
- Coverage of 348/365 days (95.34%) (known collection instability)
- Contains Inbound and outgoing DNS queries
 - around 96.4% of inbound requests originates from the institutional AS
- Attributes: 46 fields covering DNS headers, EDNS, network information, geolocation, ASN, timing and DNS resource records (ENTRADA 2's data schema)



Data Collection



Architecture

Modularized queries through dataloaders

Dataloaders can be easily be used as adapters to modern big data architectures

This architecture allows for standardization of the data

Aggregation

targeting/removal of attributes within the data

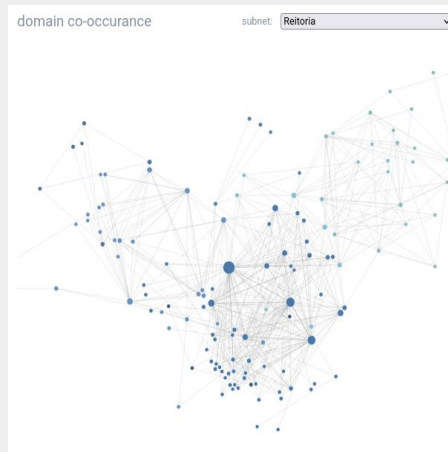
(what is relevant to that specific user)

Examples:

- Aggregation by service, subnet, range...
- Filtering CDNs, known infrastructure traffic
- Data enrichment, internal heuristics

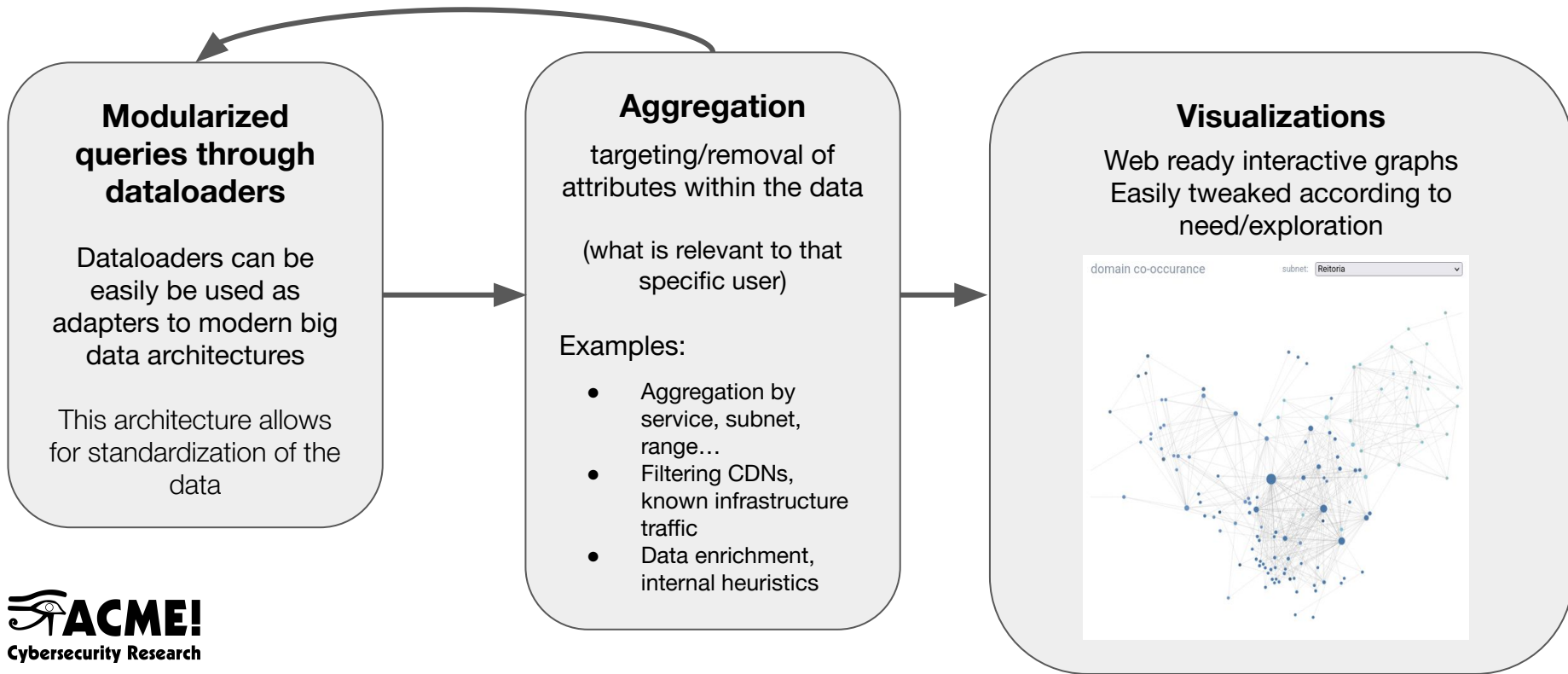
Visualizations

Web ready interactive graphs
Easily tweaked according to need/exploration



Architecture

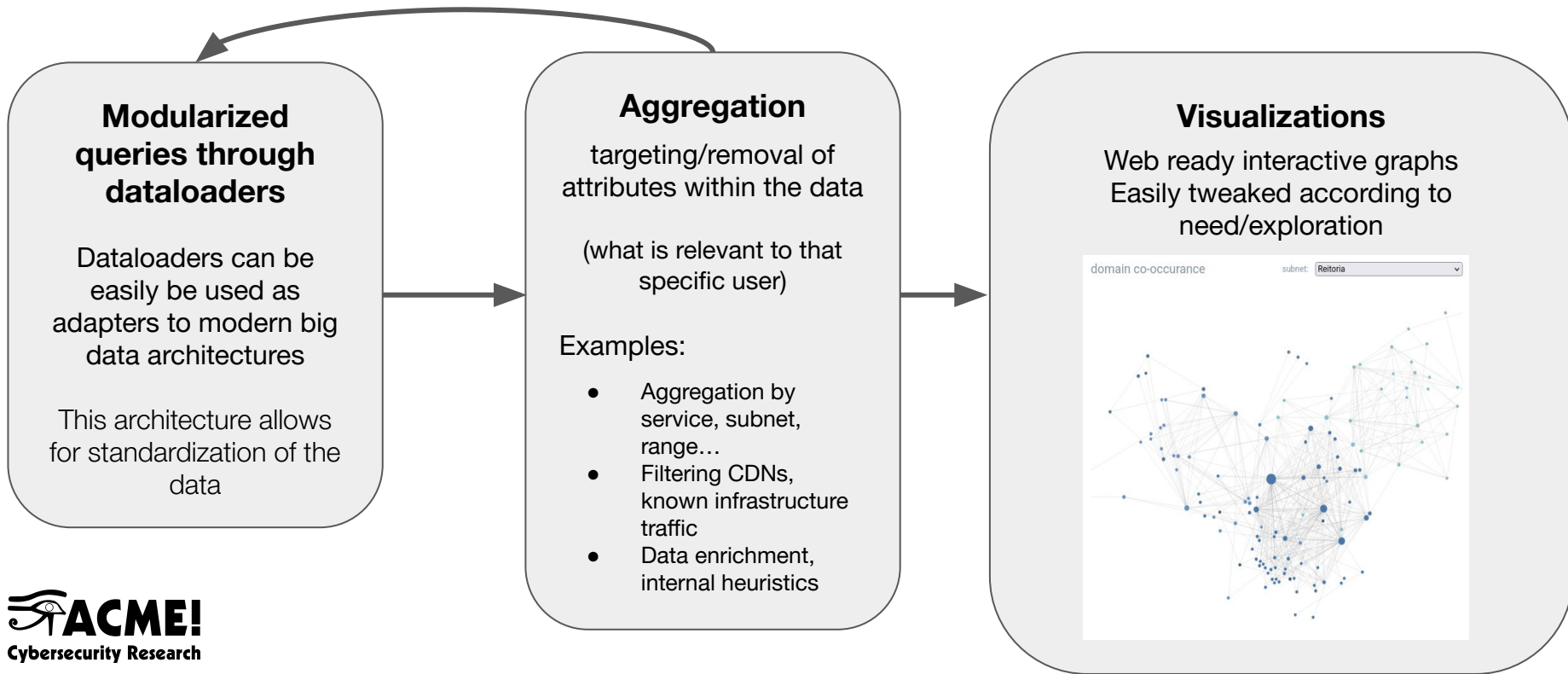
since query is modular....



Architecture

since query is modular....

aggregation rules can be “translated”
back into the query (performance gains!)

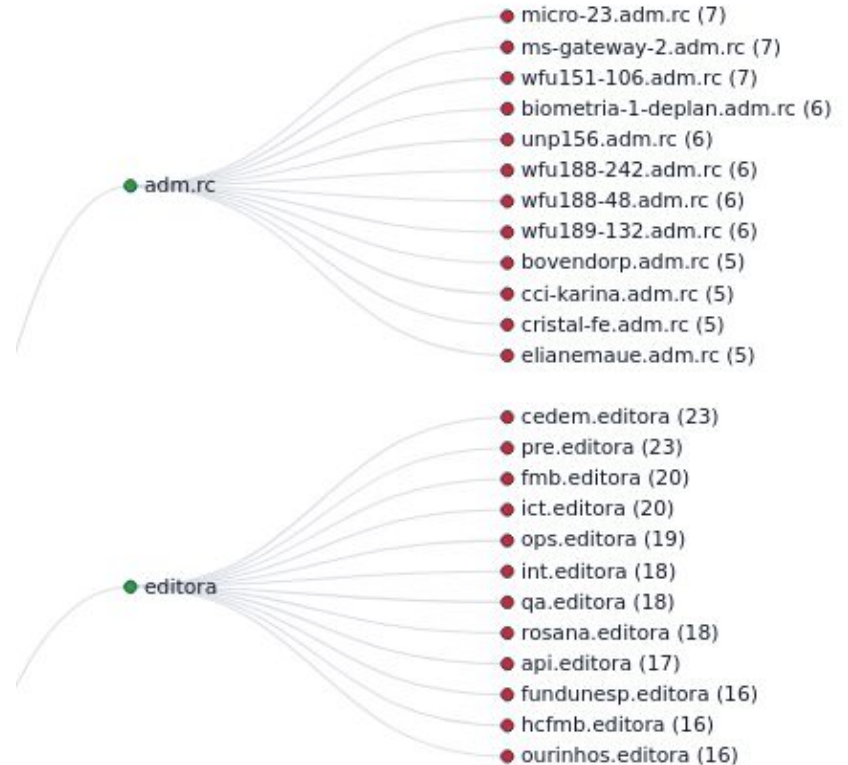


Demonstration

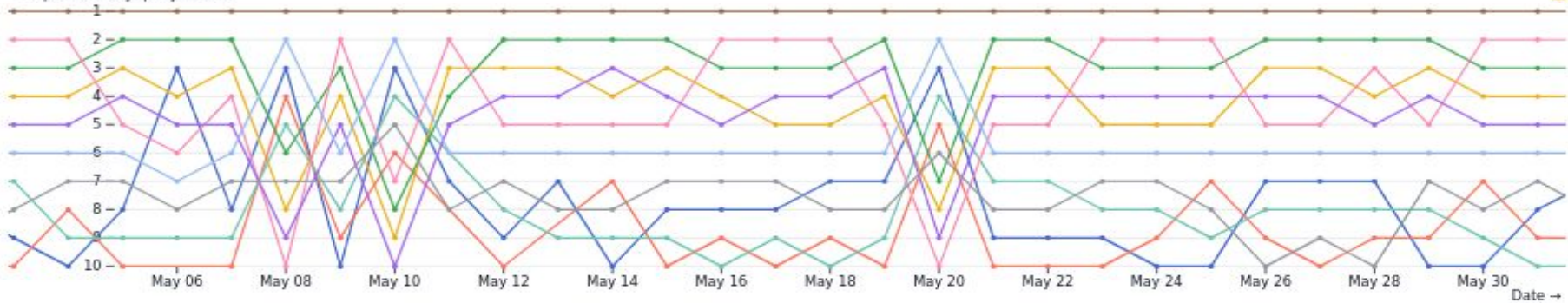
Demonstration

Results

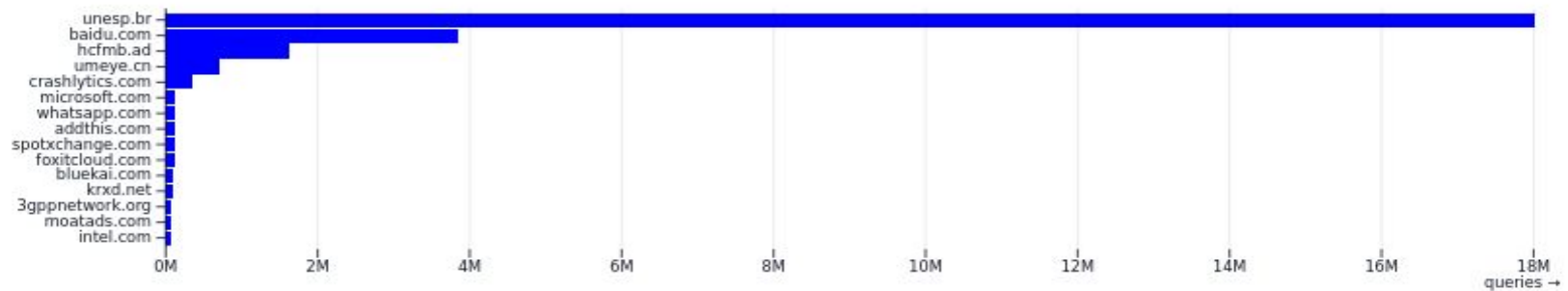
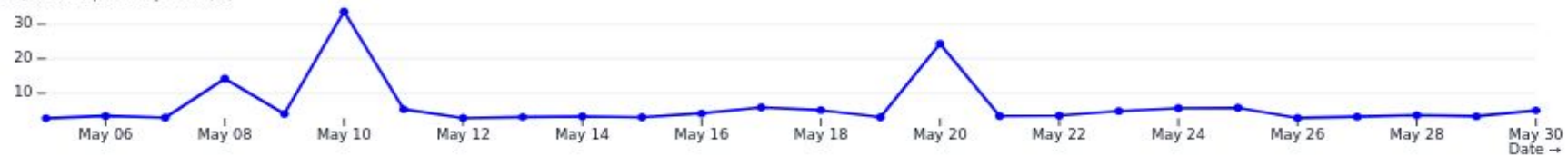
- On May 10, 2025, our recursive server received ~1.23 million queries, primarily from five ASNs geolocated in china, targeting unesp.br
- High QNAME diversity observed (630.436 individual qnames)
- 97.39% of queries were NXDOMAIN
- Actual institutional hostnames seemed to be incorporated into the NXDOMAIN queries, among combinations of common infrastructure terms



↓ top 10 asns by query volume



↑ distinct QNAMEs / parsed queries (%)



Results

A 28.5% overlap of FQDNs observed across the 5 ASNs, implicating coordination

Traffic from these five particular ASNs observed on three distinct dates:

day	active_asns	nxdomain	asn_pairs	mean_coincidence	max_coincidence	
1746403200000	1	113	0			
1746489600000	1	387	0			
1746662400000	5	355787	10	26.49	41.79	(08/05/2025)
1746835200000	5	1056731	10	9.82	13.19	(10/05/2025)
1747094400000	1	178	0			
1747180800000	1	110	0			
1747267200000	1	103	0			
1747353600000	1	125	0			
1747440000000	1	156	0			
1747526400000	1	184	0			
1747612800000	1	134	0			
1747699200000	5	569408	10	14.17	15.46	(20/05/2025)
1747785600000	1	118	0			

Conclusions Future Work

- Effective data exploration requires systems that ADAPT
 - “The tool should evolve with its users: new features and views can be driven by the questions they want to explore”
- Development of the tool is still ongoing
 - Integration into DSC datatool, other DNS measurement services
 - Support for additional data contexts (active DNS, RDAP records, etc)

Thank you!

OARC 47 – 2026/08/09

Kim M (kim@acmesecurity.org)