HI^3 HUB for Internet Incident Investigation

Alberto Dainotti
alberto@caida.org

Center for Applied Internet Data Analysis
University of California, San Diego

caida
www.caida.org

TALK OUTLINE

- **Context**: Large scale Internet Incidents
- **Need**: Combine interactive tools / live data / expertise in one place
- HI^3 : definition and **approach**
- Hypothetical **examples** (focus on data types and diverse providers)
- High-level **diagram**
- **Development** Roadmap
- A demo of **functionalities** using the current proof of concept
- **Competition**/Synergies

LARGE-SCALE INCIDENTS

a threat to private and national assets

- **large-scale Internet incidents** *(hijacks, outages, spam and fishing campaigns, botnet activities, scanning, large-scale bug exploitation)* are a major threat to public safety and to both public and private strategic and financial assets
- *Often:*
  - ***unnoticed***
  - ***hard to understand*** (dynamics, motivation, infrastructure used, source, target)
    - hard to mitigate, prevent, etc.
    - hard to assess the damage
    - hard to assess restoration

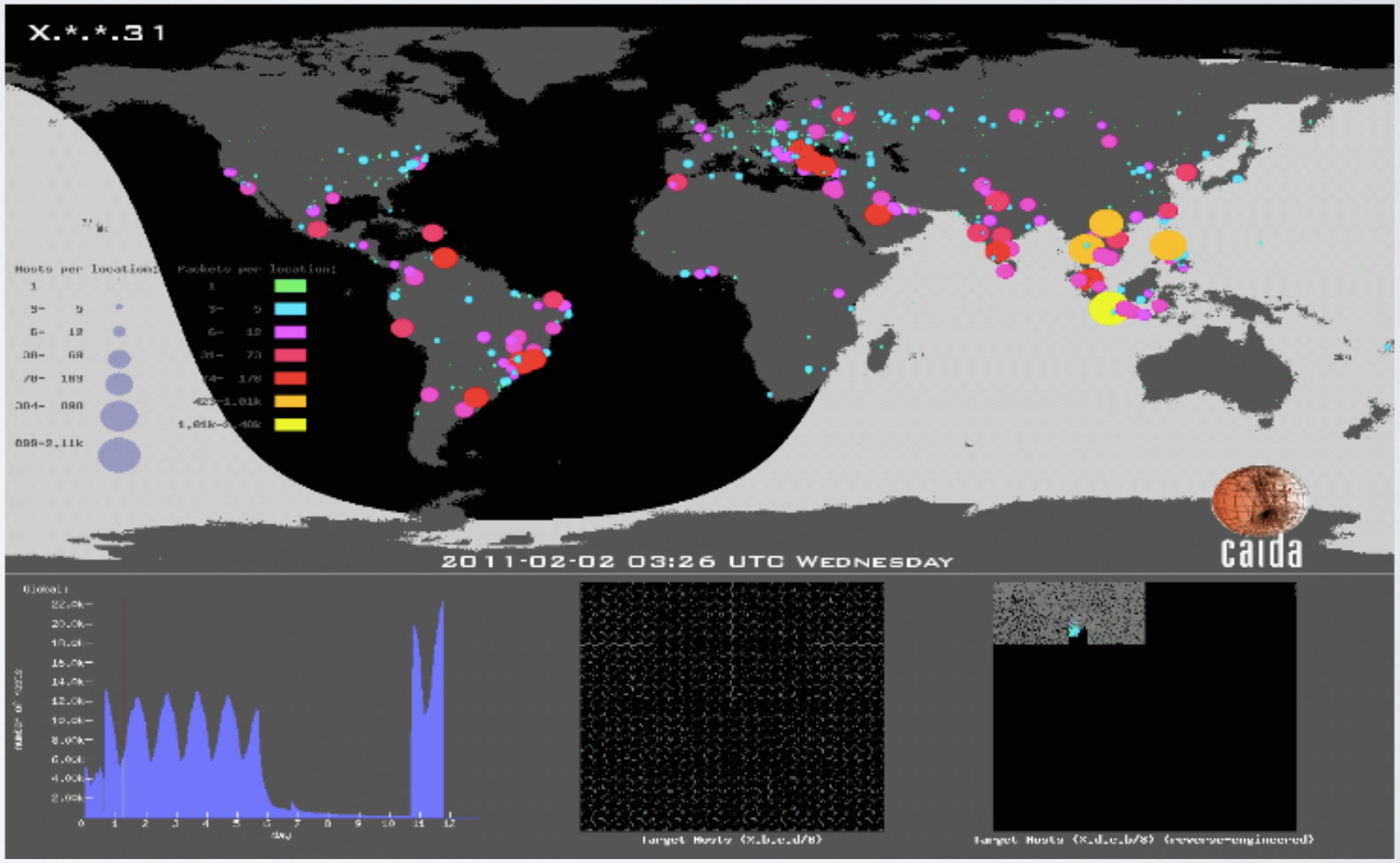
UNDER THE RADAR

the "sipscan" was massive and unnoticed

- *February 2011 - 3M hosts **covertly scanning** the whole IPv4 Internet in 12 days*

- *Following months - Massive **exploitation of VoIP infrastructure***

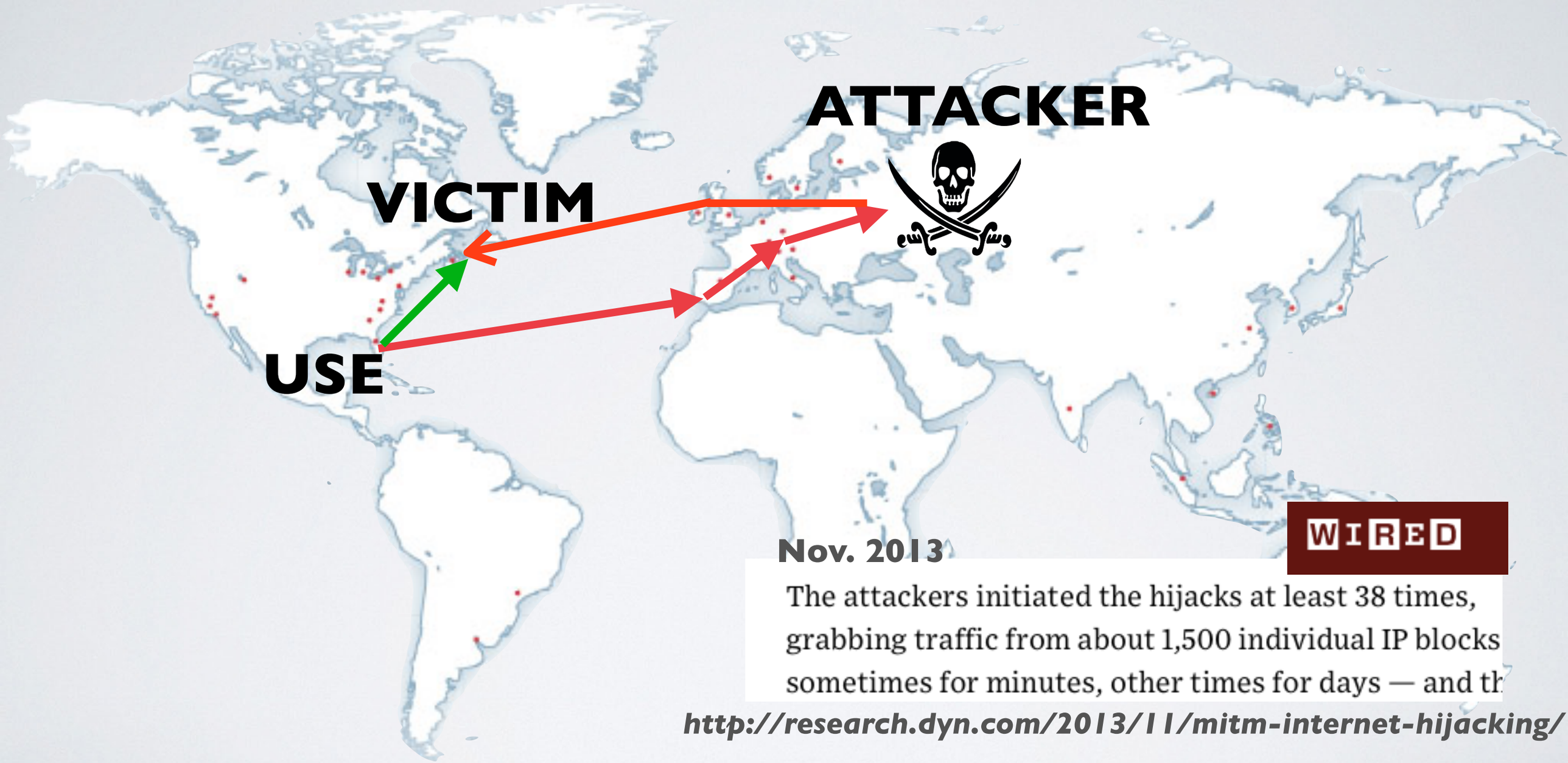
- *VoIP Fraud costs $40 billion per year*



video: http://www.caida.org/publications/papers/2015/analysis_slash_zero/supplemental/sipscan.composite.mp4

UNDER THE RADAR

BGP mitm attacks constantly go unnoticed



Nov. 2013

WIRED

The attackers initiated the hijacks at least 38 times, grabbing traffic from about 1,500 individual IP blocks sometimes for minutes, other times for days — and th

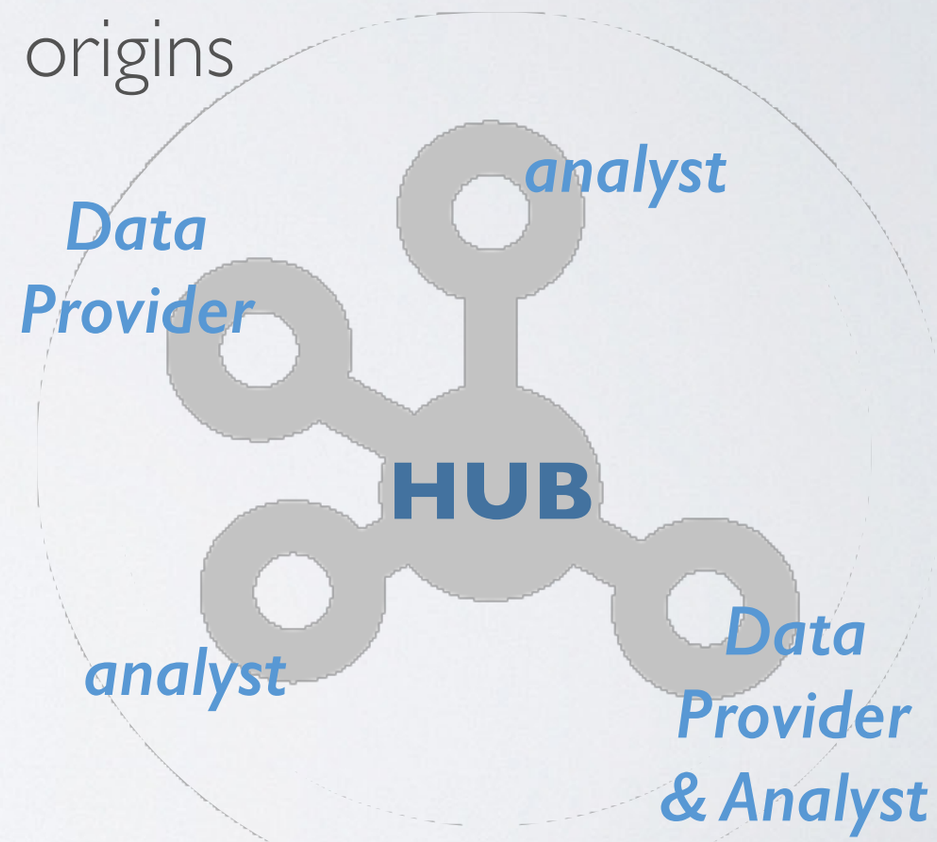
http://research.dyn.com/2013/11/mitm-internet-hijacking/

WE NEED

many features in one "place"

Effective (timely) analysis of these events requires

- ability to act **fast** with **agility**
- **combination** of data of different types and origins
  - sharing **live data**
  - on-the-fly data extraction/aggregation
  - interactive visual tools for quick exploration
- **teamwork** of heterogeneous expertise
- a **trusted** environment



HUB for Internet Incident Investigation

towards a virtual situation room

- A web-based collaborative environment
- with trusted groups of vetted experts (through a legal framework)
- producing analyses with interactive and visual tools
- based on diverse sets of streamed (and historical) data

THE HI^3 APPROACH

three key elements

1. Combination and correlation of **diverse** Internet cyber-security data streams around a set of common dimensions: *time and* ***Internet Coordinates***

2. Data analytics in the form of interactive **exploratory data analysis** and configurable event **detection**

3. Trusted **realtime collaborative** environment

THE HI^3 APPROACH

1. **Combine/correlate diverse Internet cyber-security data:** center data *organization, processing, querying and visualization* around a set of common dimensions: ***time and Internet Coordinates***

- **IP addresses** (IPs, /24s, /16s, …)
- **BGP**: prefixes, AS numbers, siblings, AS-relationships, AS customer cones
- **Topologies**: AS graph, router-level topology, physical (links, facilities, …)
- **Geopolitical layer**: geographic coordinates, administrative/political boundaries (country, region, county, province, city, zip code, building, etc.)
- **DNS**: records, active and passive databases
- **Whois** and routing registries

THE HI^3 APPROACH

2. **Data analytics in the form of exploratory data analysis and configurable event detection**

- interactive **navigation** through many (tens of millions) data streams
- interactive + live data **visualization** interfaces (e.g., hundreds of time series in a single graph)
- users can **apply functions to the data** and observe the results immediately applied to the current visualization ("Internet Matlab" analogy)
- configurable **automated detection of anomalies and dashboards**

THE HI^3 APPROACH

3. Trusted collaborative environment

- potentially leveraging IMPACT legal framework and vetting
- users can create **trusted groups**
- **realtime collaboration** (as in Google Docs)
- users can save **personalized** organizations of data, bookmarks to dashboards and live graphs, …
- **open access to public data** creates the opportunity to attract both additional insights into the large pool of data available as well as new users that might join restricted groups or form other collaborations

HI³ HYPOTHETICAL EXAMPLE #1

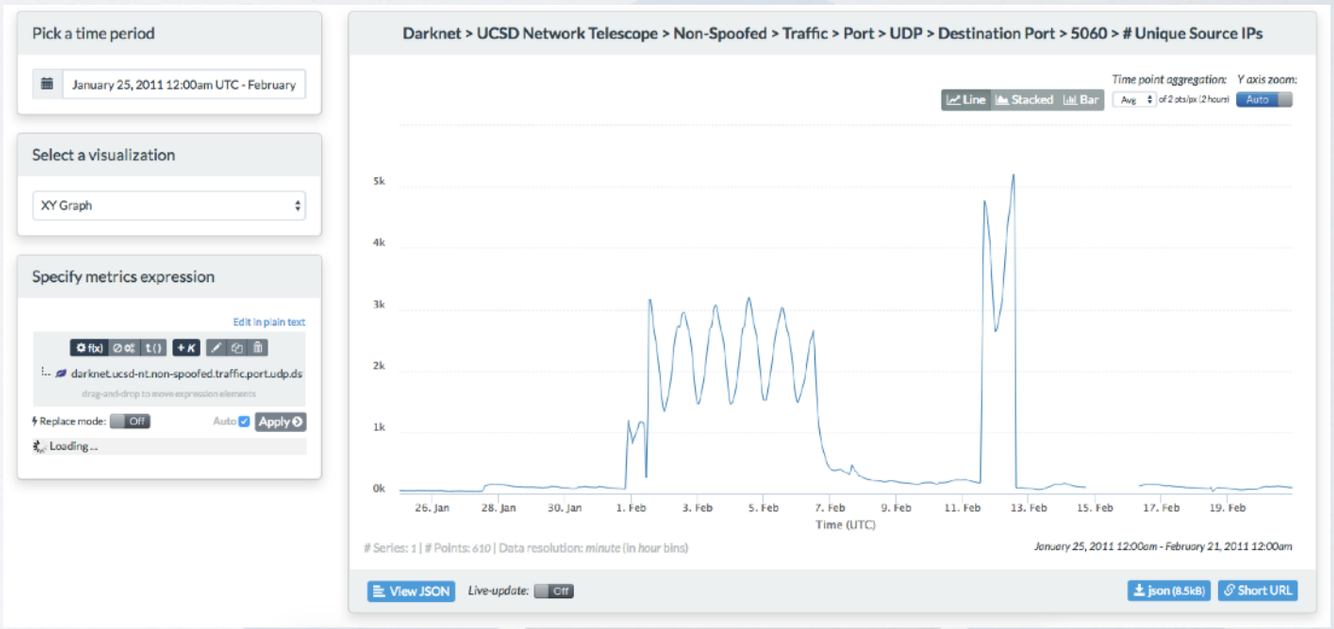
Investigating BGP hijacking

- CAIDA's **BGP hijacking** monitoring system identifies a set of potential **attacks** involving certain ASes
- Farsight's **spam telemetry data** shows an increase in spam sent by these ASes
- Visual exploration and correlation of data helps to **identify patterns**

HI³ HYPOTHETICAL EXAMPLE #2

Discovering a Botnet Scan (sipscan)

- Alert on large increase of (pre-filtered) **traffic** reaching UCSD's **network telescope** towards port UDP 5060
- Analyst sends a request for a Spark query to CAIDA returning **set of IP addresses**/prefixes/ASes
- Data correlates with Symantec's **malware database** allowing us to identify the source as the "Sality" Botnet



HI³ HYPOTHETICAL EXAMPLE #3

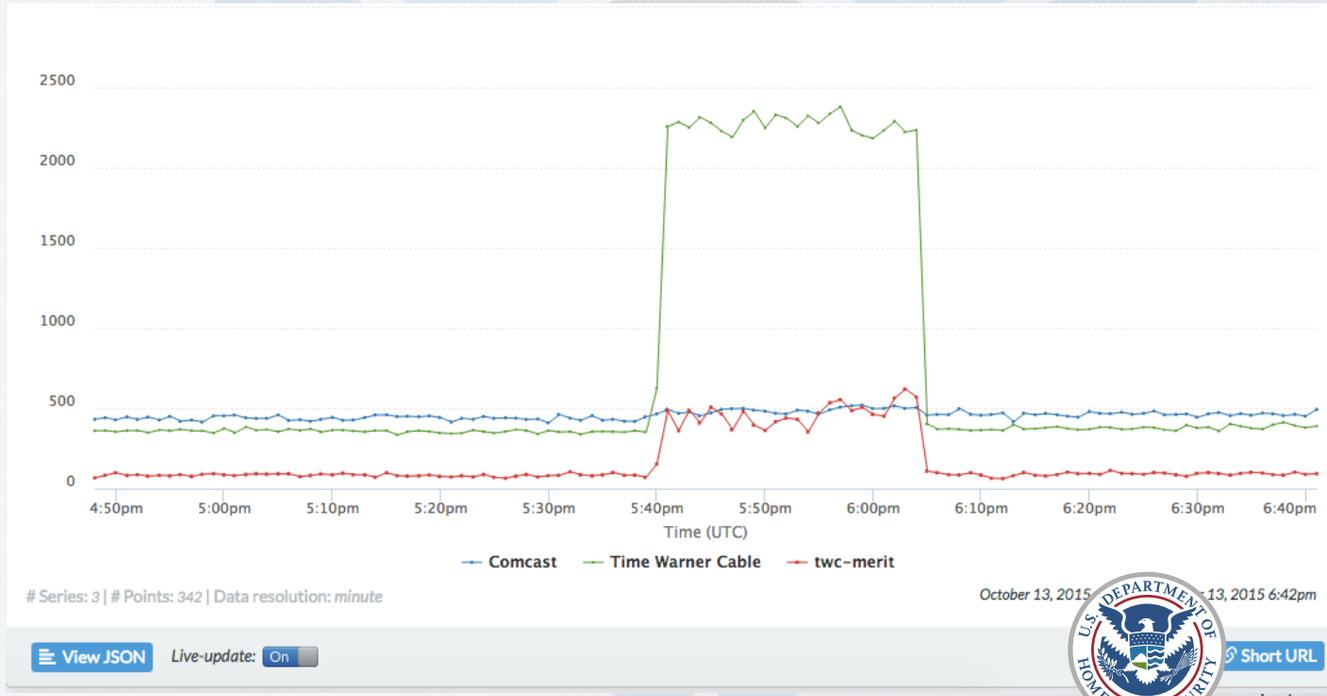
Investigating a new massive bug-exploitation

- 13th Oct 2015 - from 5.40pm UTC for about **25 minutes a large spike in source IPs from all continents**

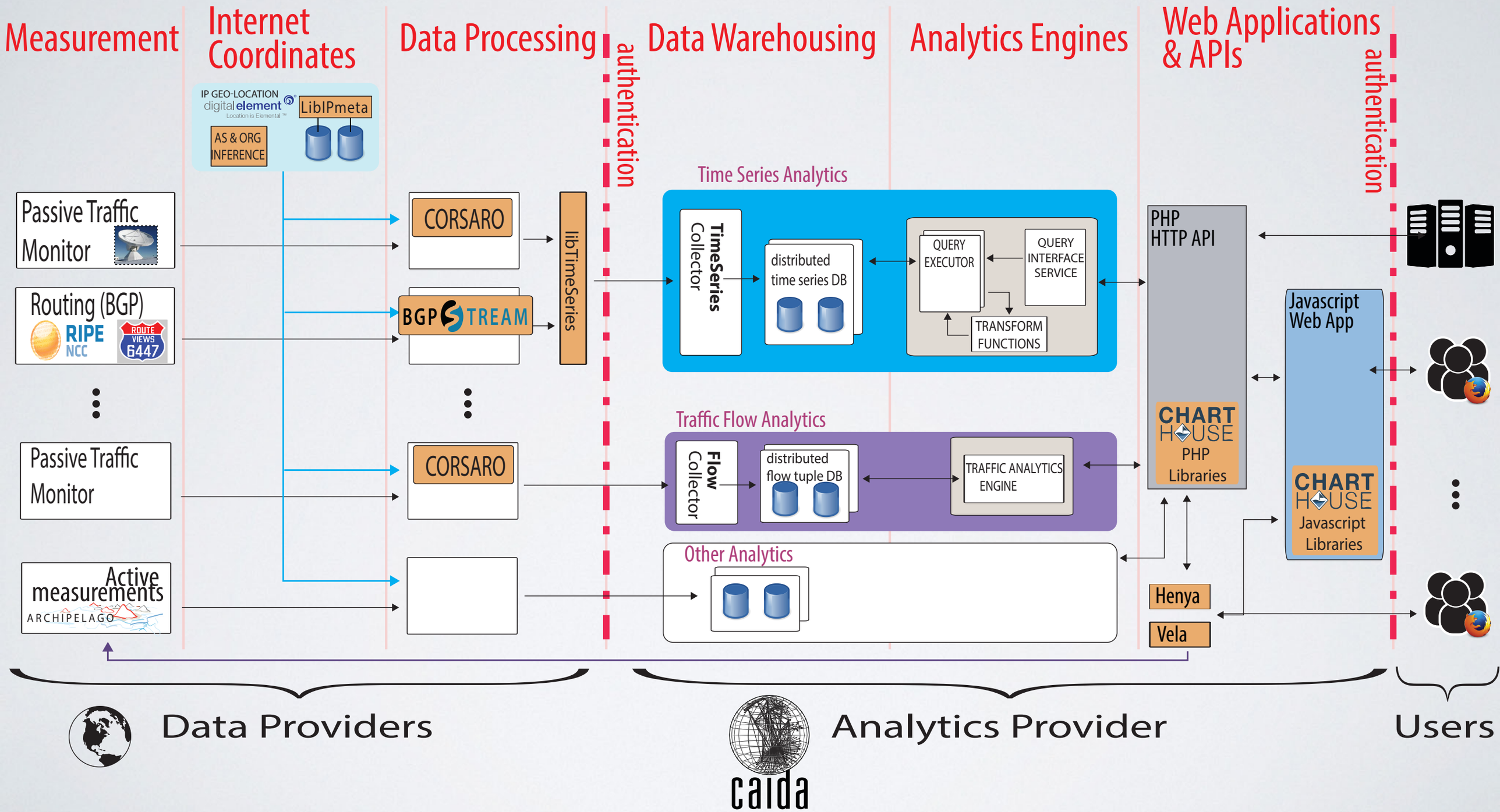
- but among top 5 **US providers** we see it coming **only from TWC**!

- backscatter from UPnP machines that received spoofed packets to their **UDP port 1900 (UPnP)** originating from UDP port 80

- It is unclear why the packets would be spoofed, but it could be an exploitation attempt that injects and runs code on the remote machine



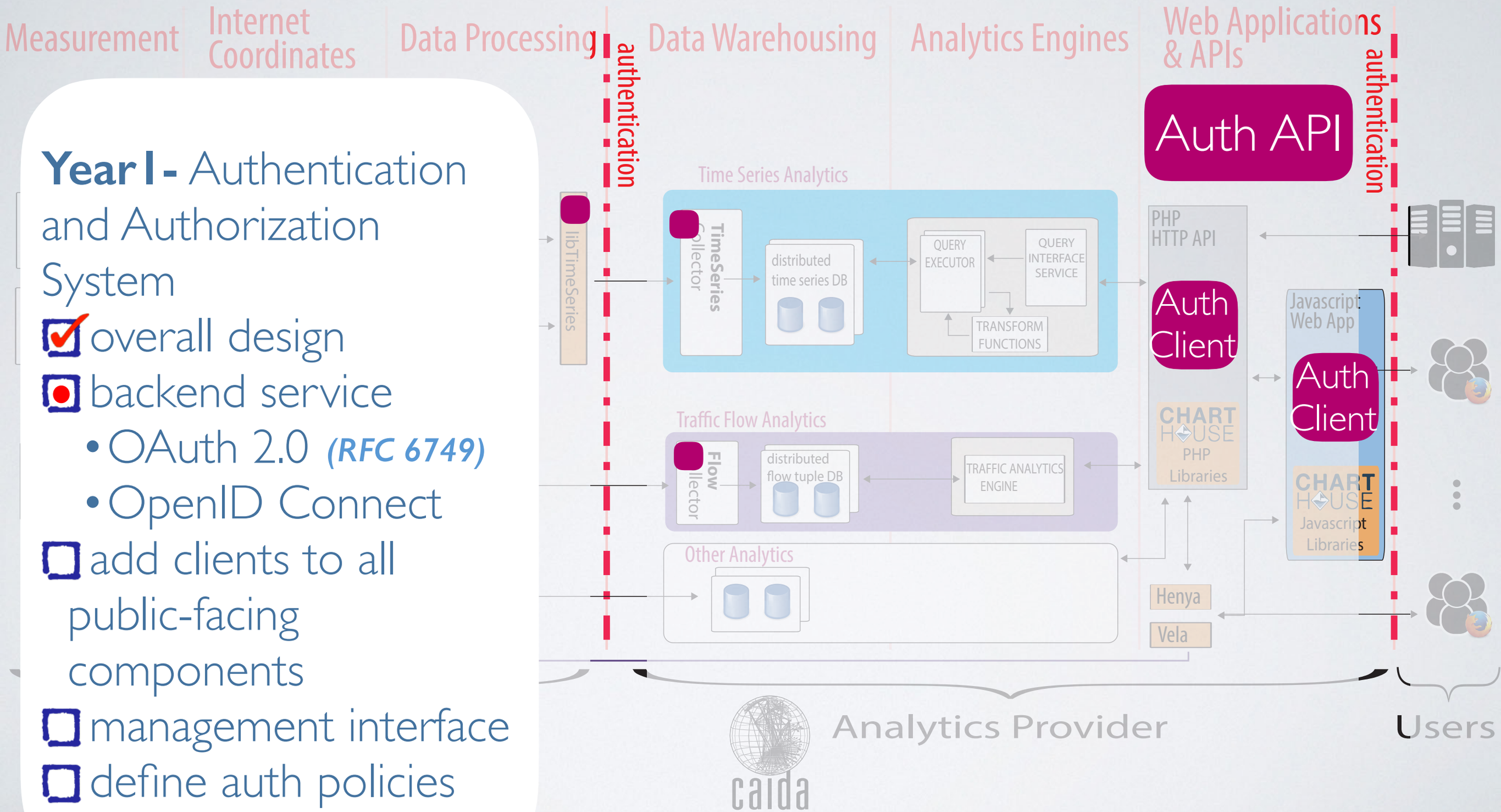
HI³ ECOSYSTEM



HI³ DEVEL. ROADMAP

Year I - Authentication and Authorization System

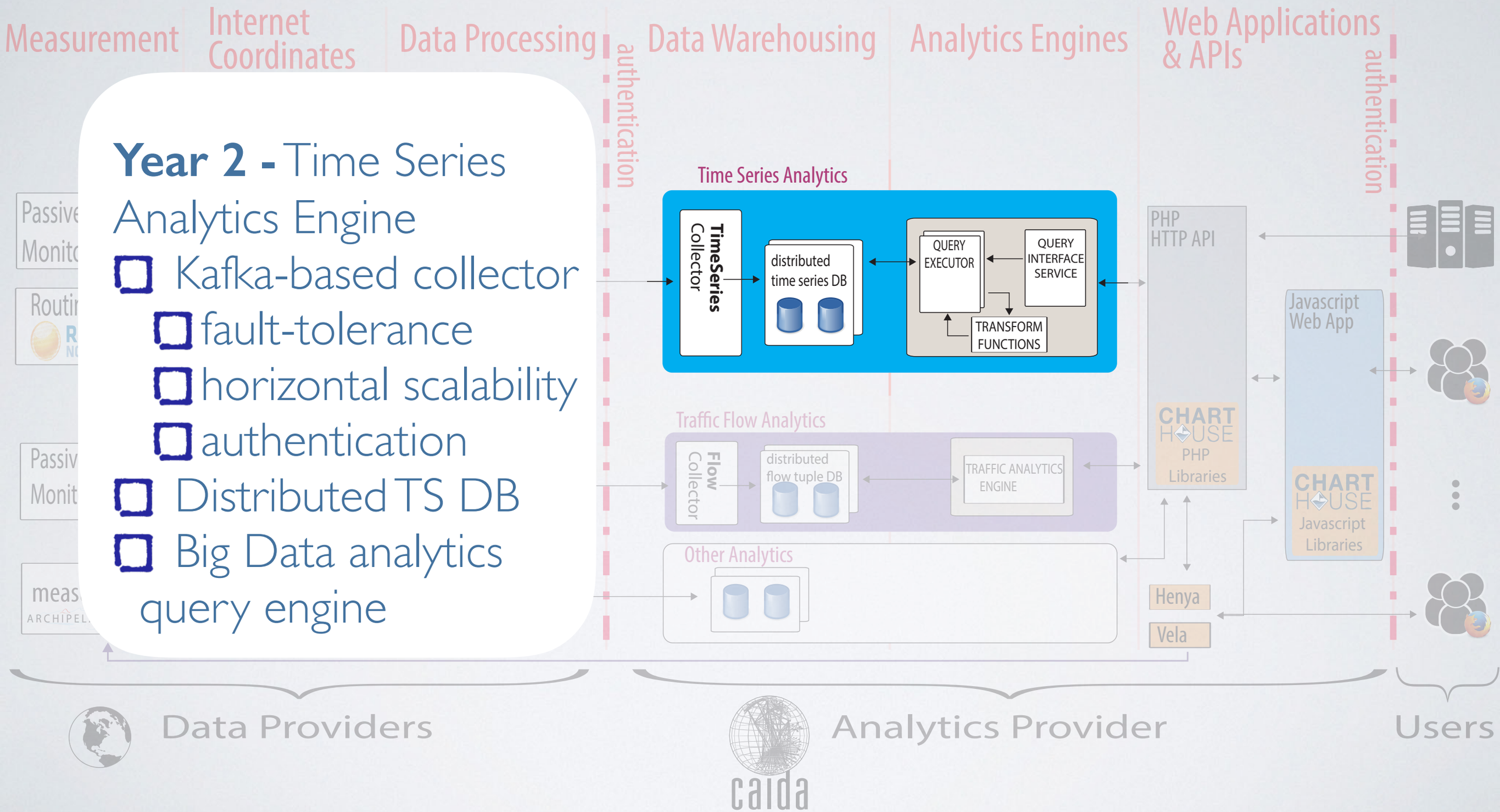
- ☑ overall design
- ◉ backend service
  - OAuth 2.0 ***(RFC 6749)***
  - OpenID Connect
- ☐ add clients to all public-facing components
- ☐ management interface
- ☐ define auth policies



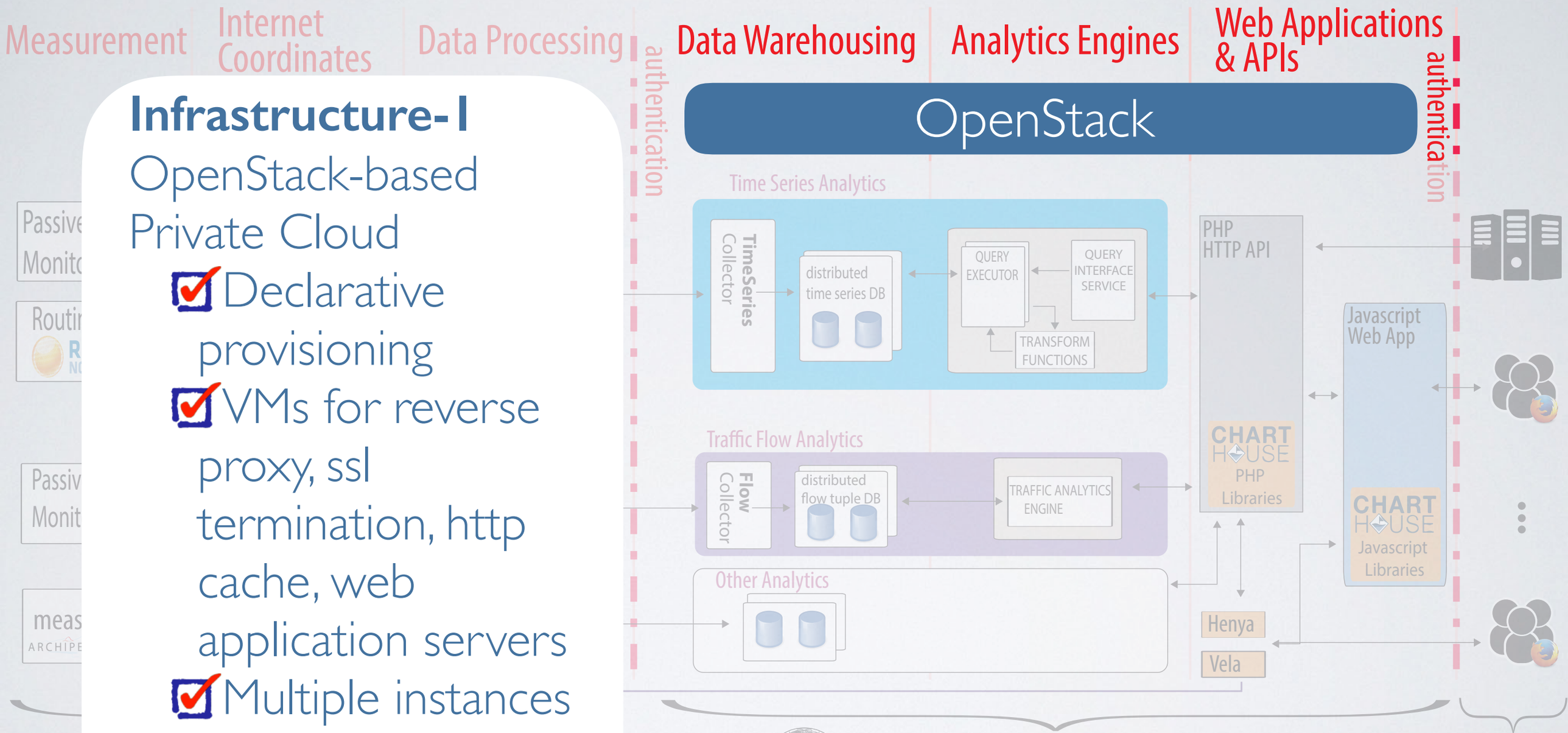
HI³ DEVEL. ROADMAP

Year 2 - Time Series Analytics Engine

- Kafka-based collector
  - fault-tolerance
  - horizontal scalability
  - authentication
- Distributed TS DB
- Big Data analytics query engine



HI^3 DEVEL. ROADMAP



Infrastructure-I

OpenStack-based Private Cloud

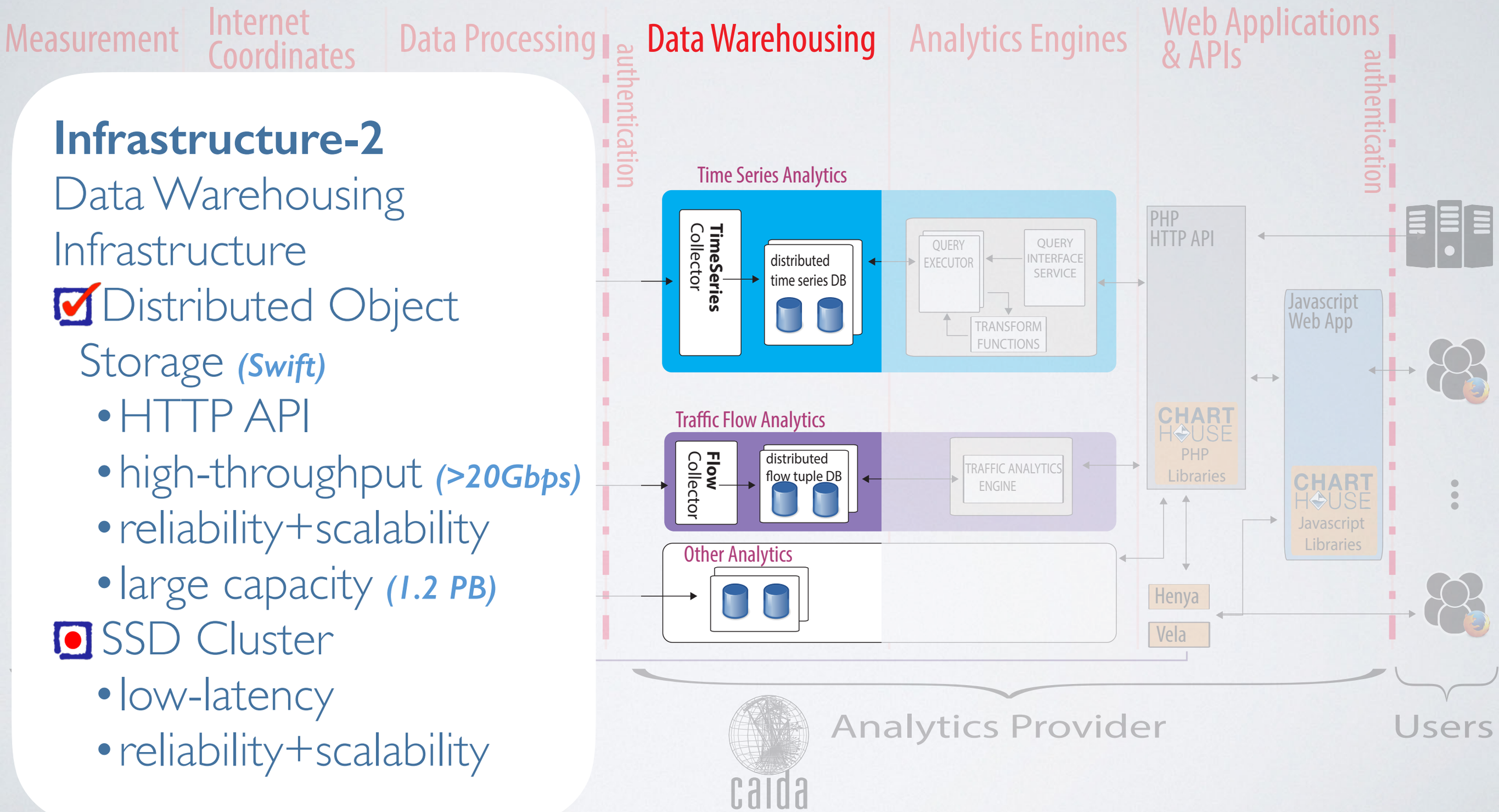
- ☑ Declarative provisioning
- ☑ VMs for reverse proxy, ssl termination, http cache, web application servers
- ☑ Multiple instances for devel and testing
- ☐ Data analytics engines

HI³ DEVEL. ROADMAP

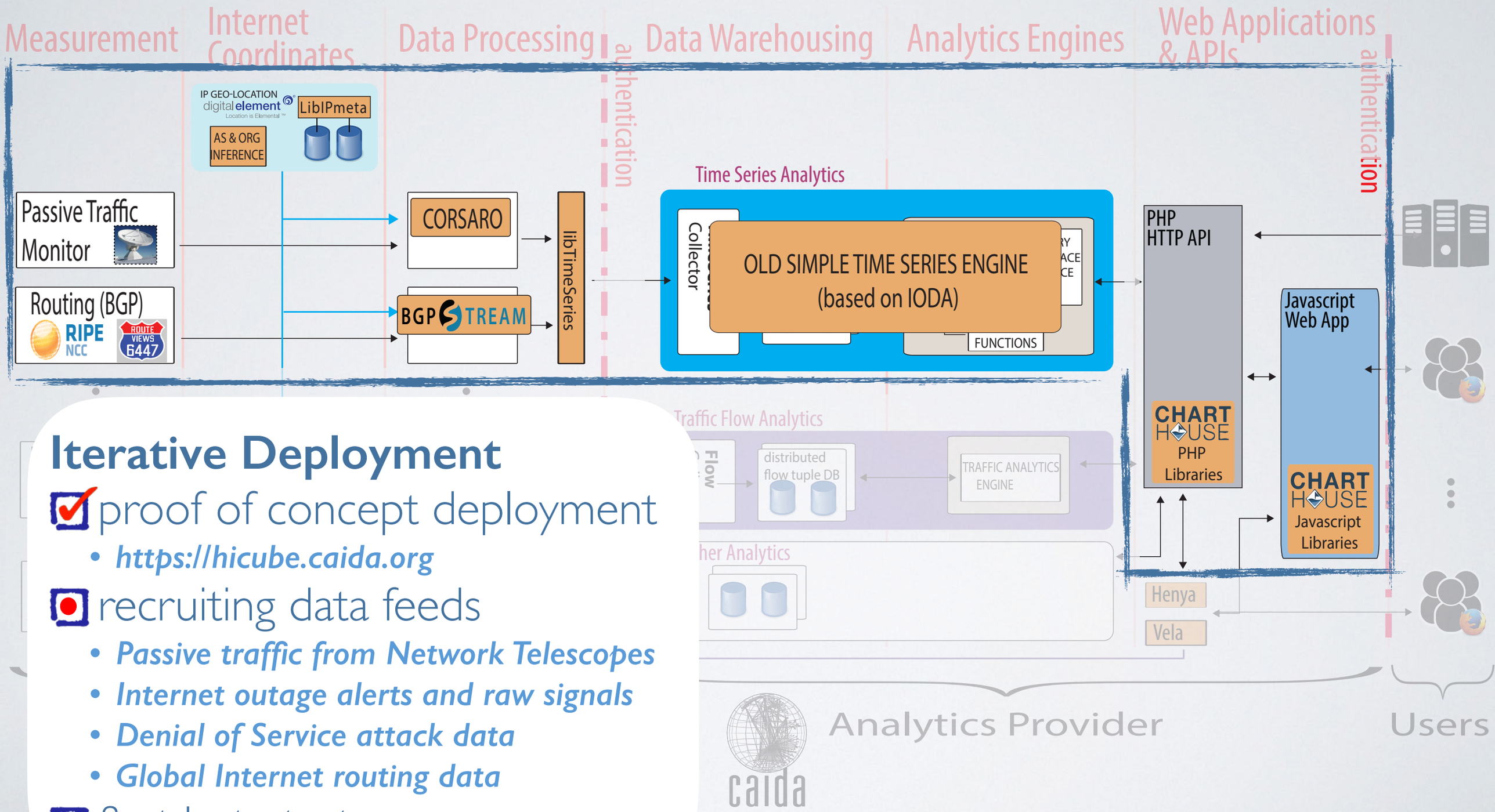
Infrastructure-2

Data Warehousing Infrastructure

- ☑ Distributed Object Storage *(Swift)*
  - HTTP API
  - high-throughput *(>20Gbps)*
  - reliability+scalability
  - large capacity *(1.2 PB)*
- ● SSD Cluster
  - low-latency
  - reliability+scalability



HI³ DEVEL. ROADMAP



Iterative Deployment

- ☑ proof of concept deployment
  - *https://hicube.caida.org*
- ◉ recruiting data feeds
  - *Passive traffic from Network Telescopes*
  - *Internet outage alerts and raw signals*
  - *Denial of Service attack data*
  - *Global Internet routing data*
- ☐ first beta-testers soon

HI^3 "POC" DEMO

HI³ HYPOTHETICAL EXAMPLE #3

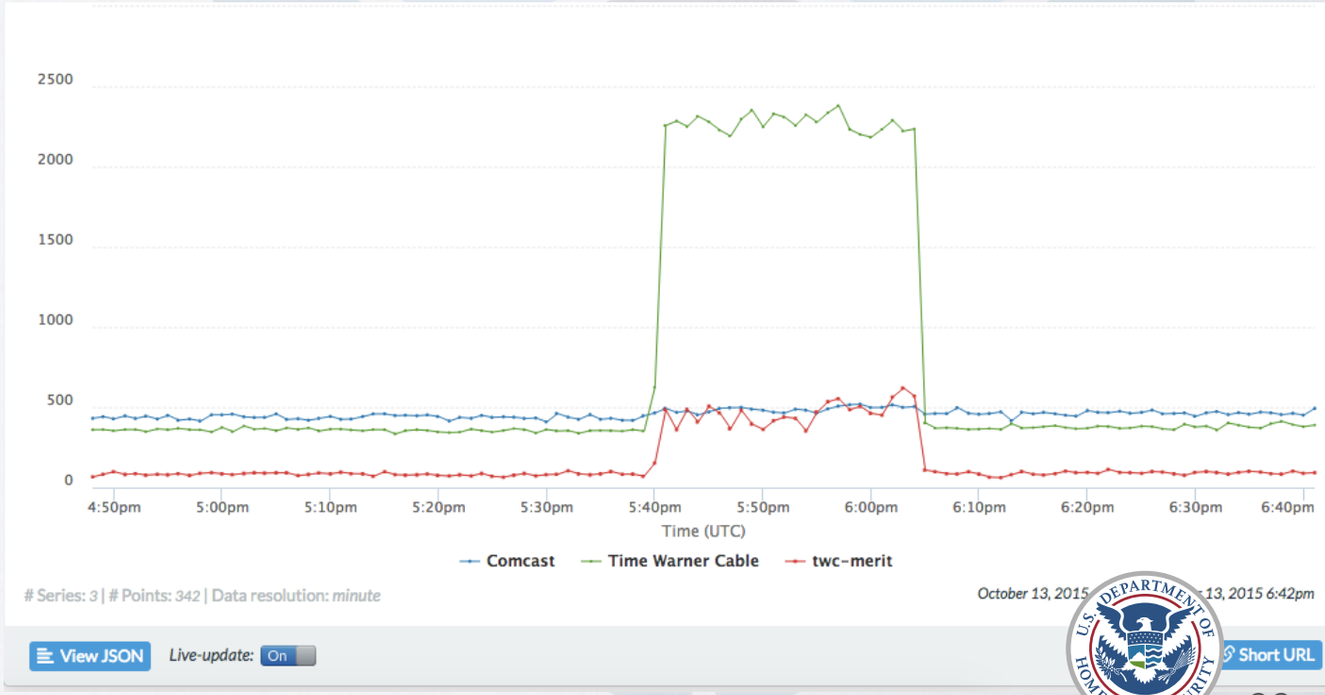
Investigating a new massive bug-exploitation

- 13th Oct 2015 - from 5.40pm UTC for about **25 minutes a large spike in source IPs from all continents**

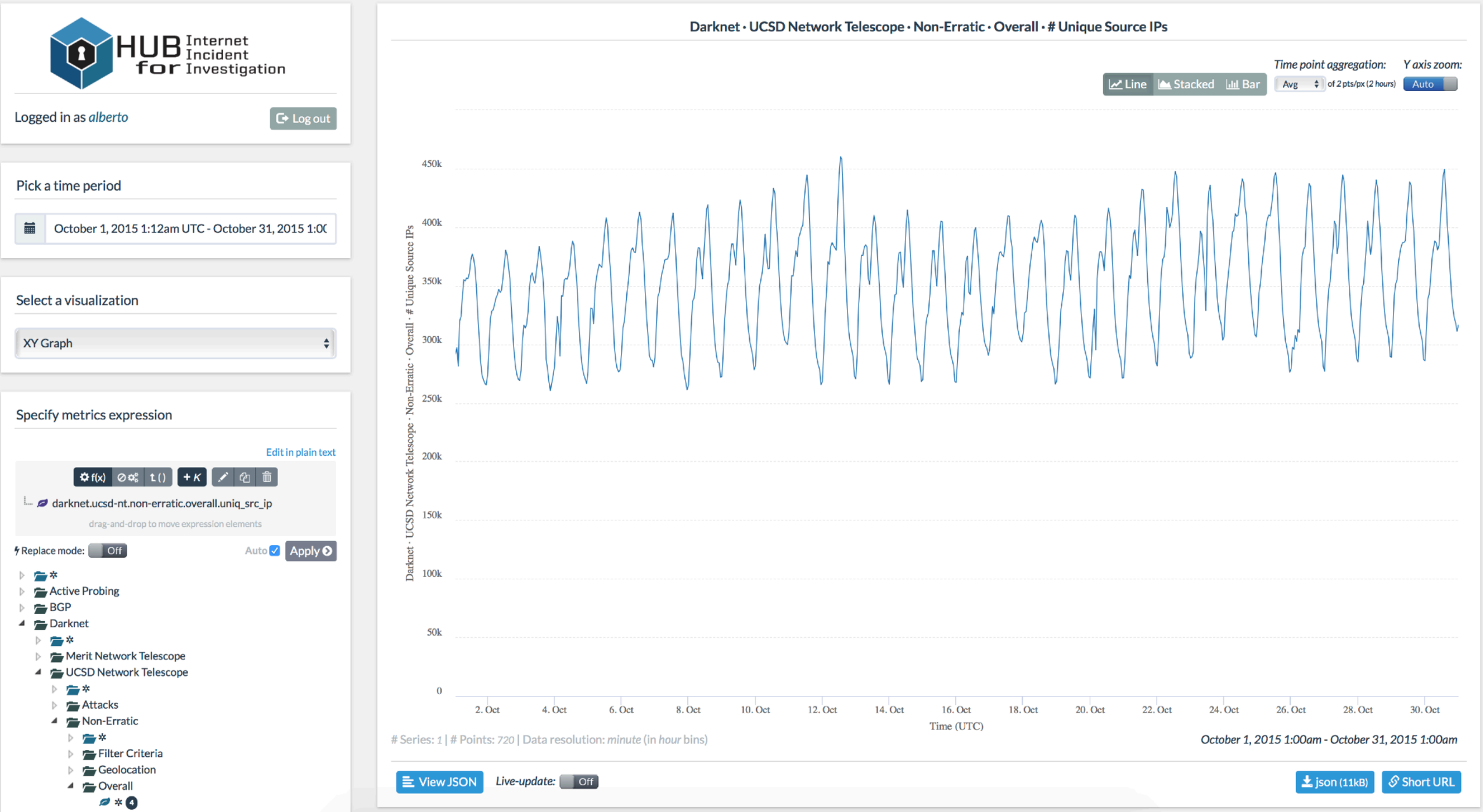
- but among top 5 **US providers** we see it coming **only from TWC**!

- backscatter from UPnP machines that received spoofed packets to their **UDP port 1900 (UPnP)** originating from UDP port 80

- It is unclear why the packets would be spoofed, but it could be an exploitation attempt that injects and runs code on the remote machine

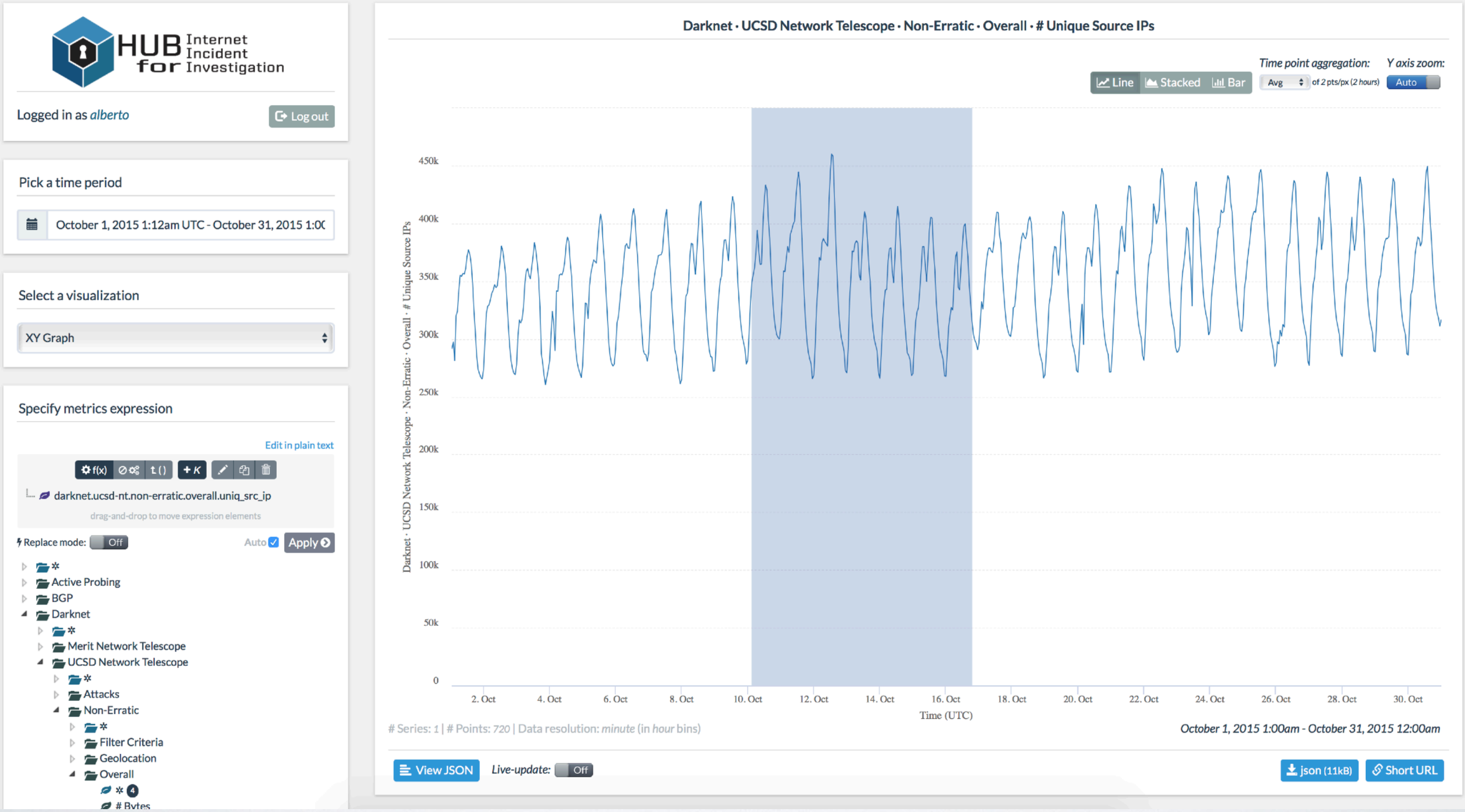


UNIQ SRC IP IN TELESCOPE IN OCT 2015



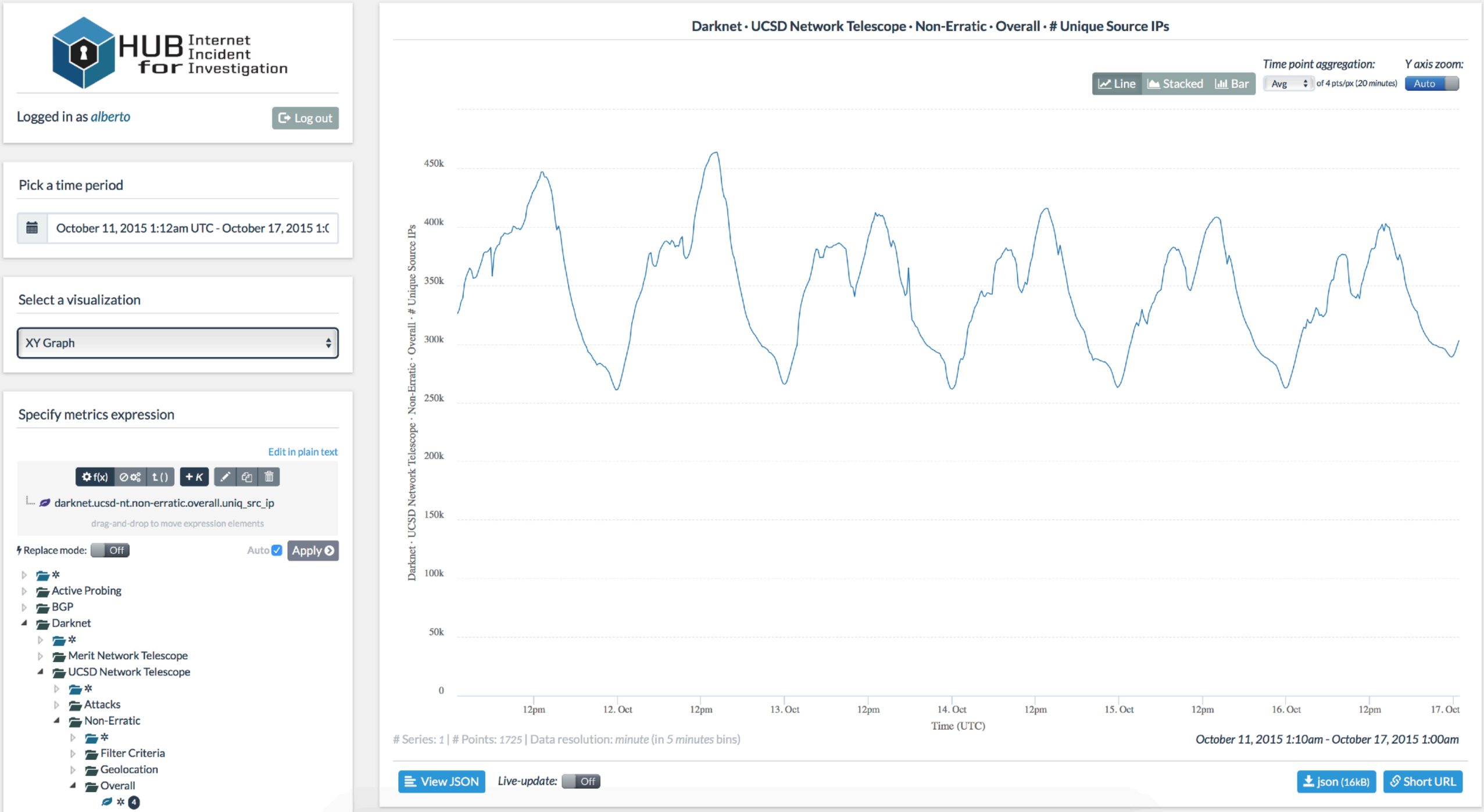
https://hicube.caida.org/@81

INTERACTIVE ZOOM



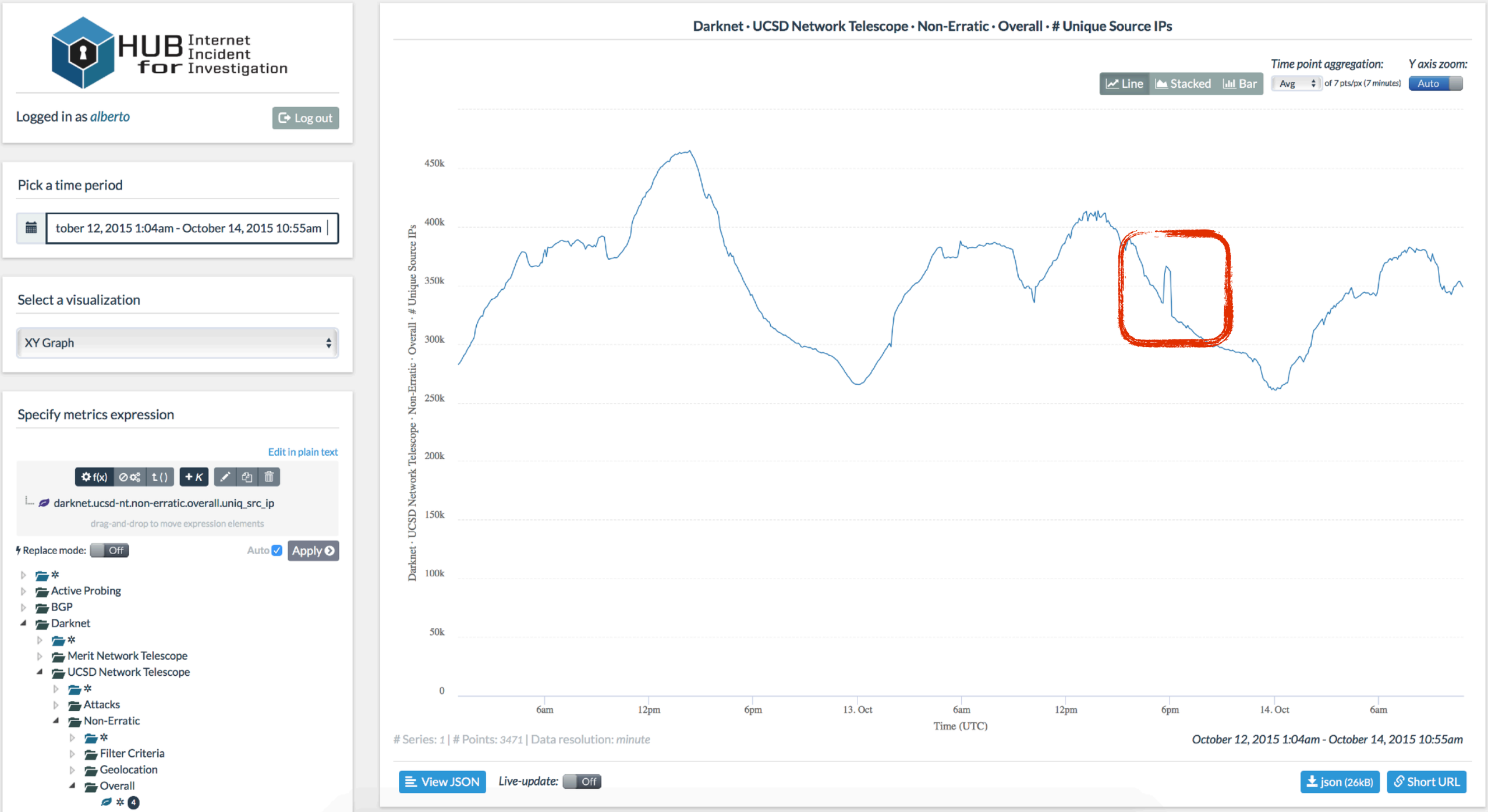
https://hicube.caida.org/@81

WEEK OF OCT 13



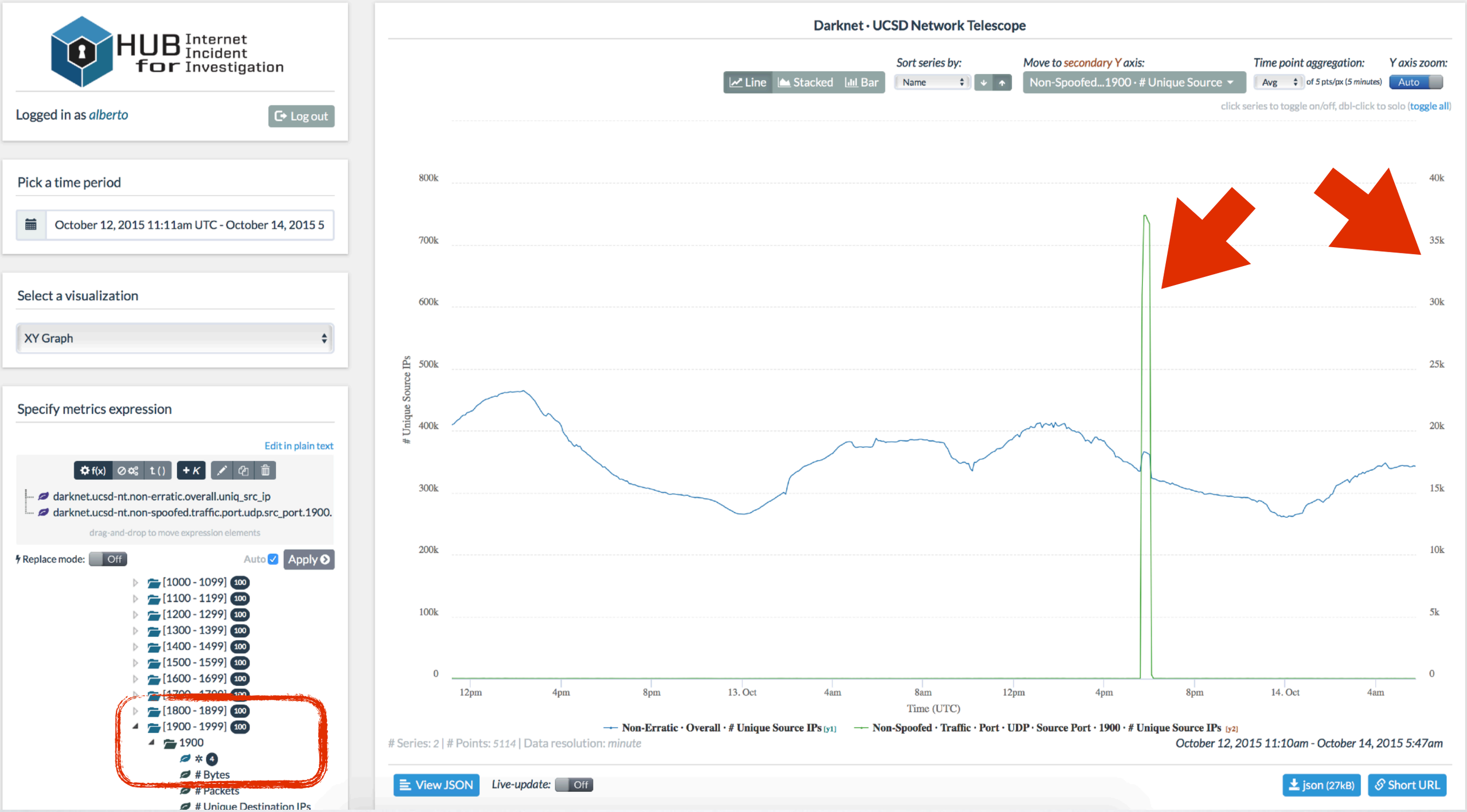
https://hicube.caida.org/@Yy

ZOOM AROUND OCT 13



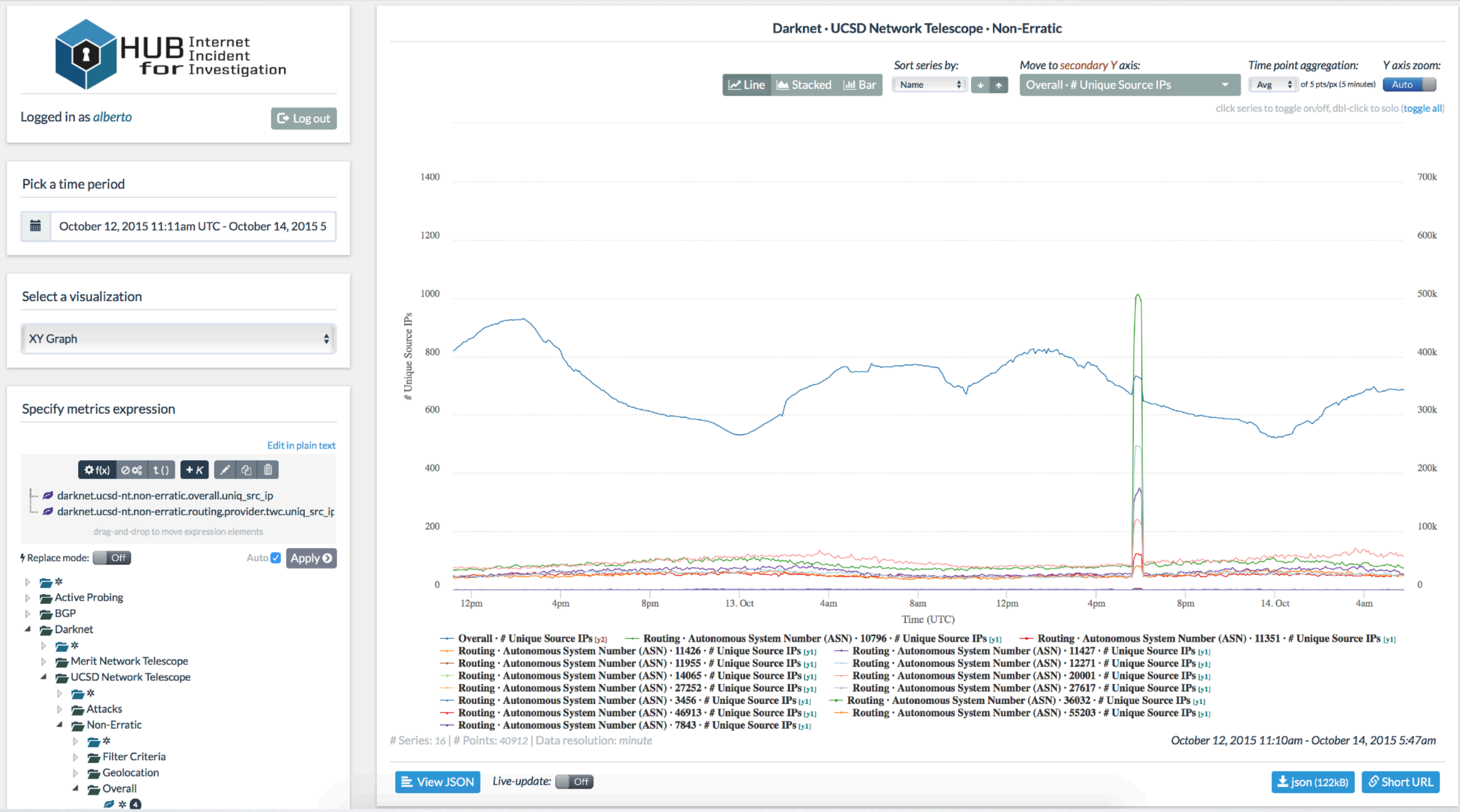
https://hicube.caida.org/@i3

ADDED UDP SRC PORT 1900



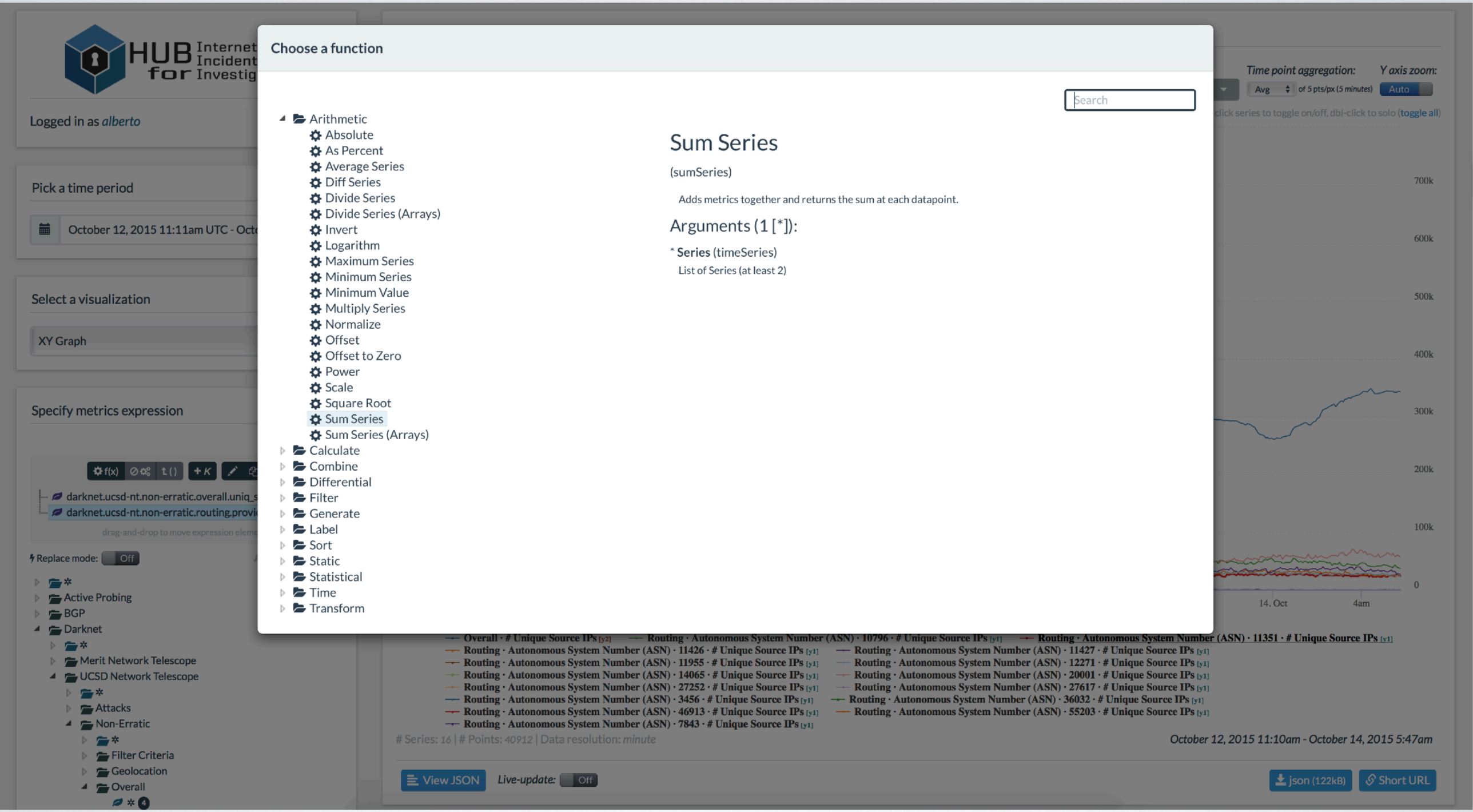
https://hicube.caida.org/@Khe

ADDED TWC-ONLY TRAFFIC

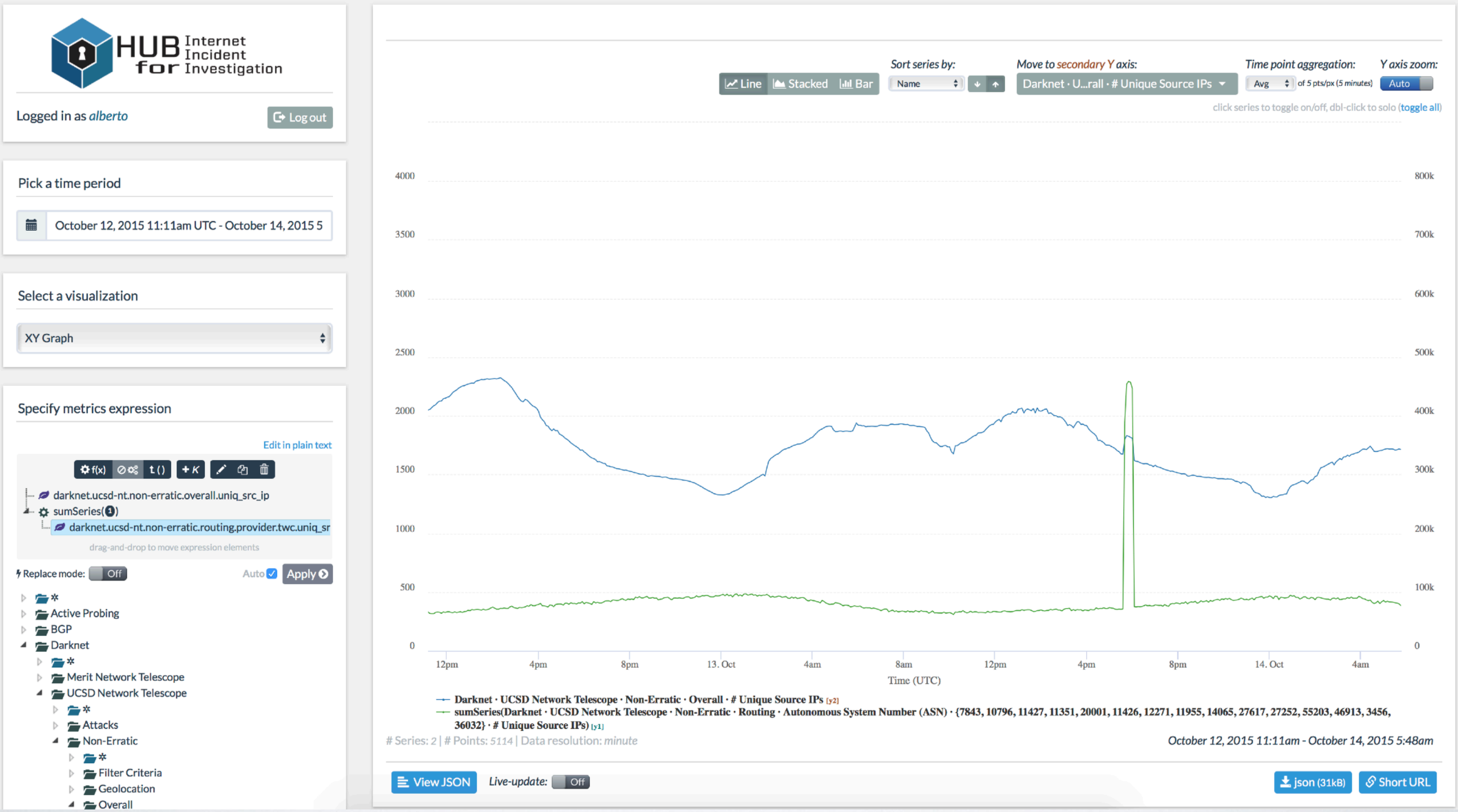


https://hicube.caida.org/@eT

SUM OF ALL TWC SIBLING AS

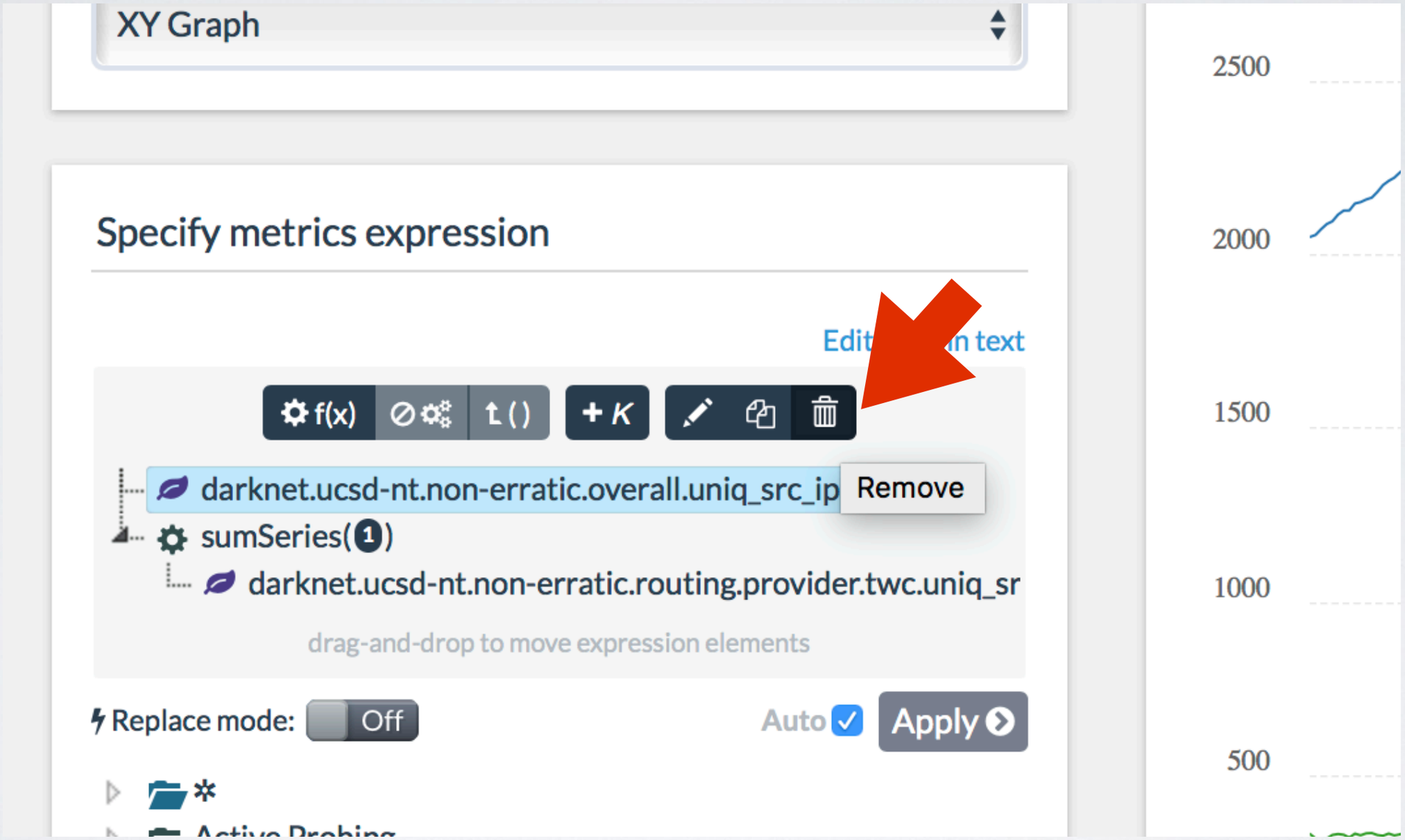


SUM OF ALL TWC SIBLING AS

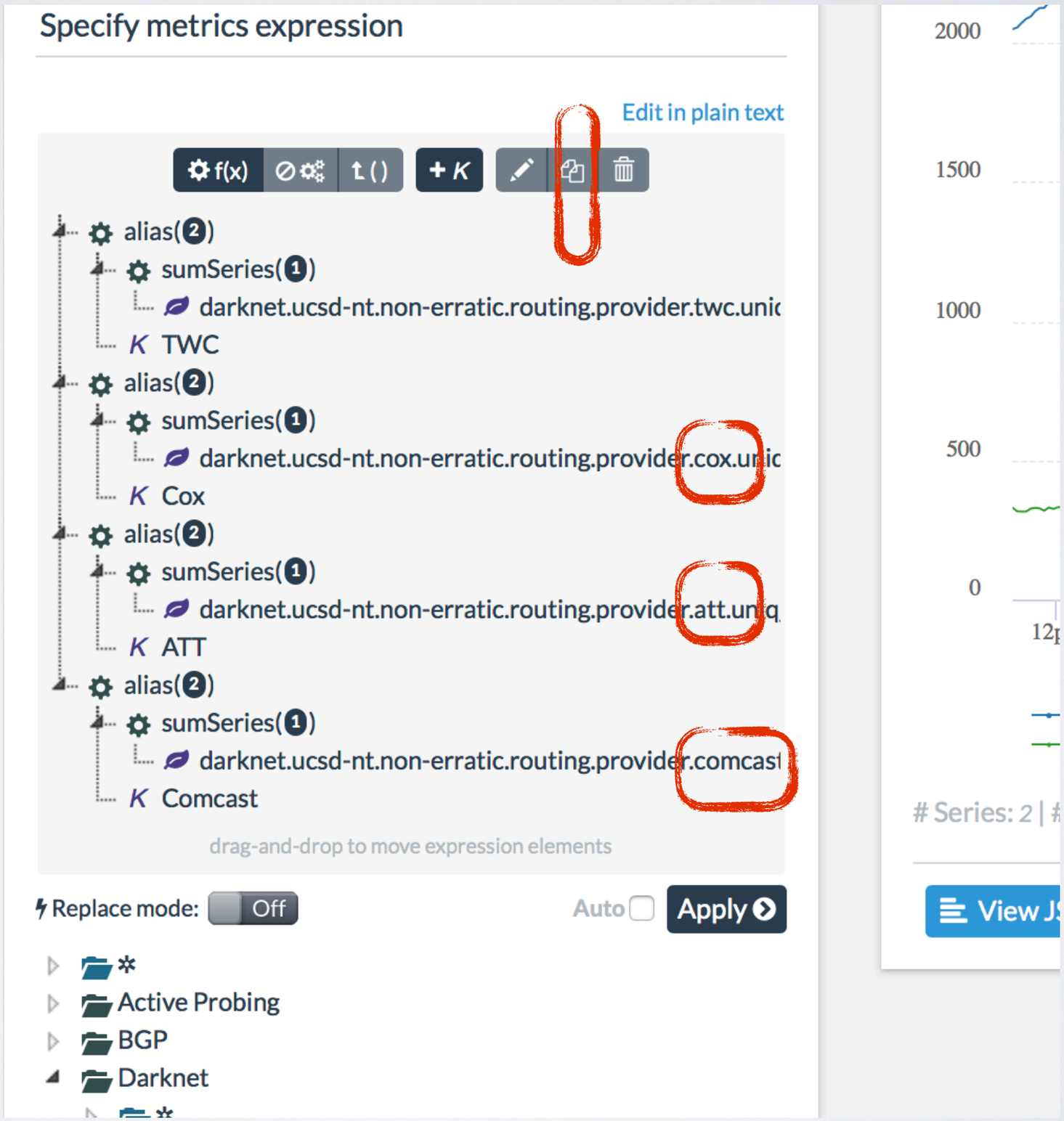


https://hicube.caida.org/@i0

REMOVING OVERALL AND ADDING COX, COMCAST, AT&T

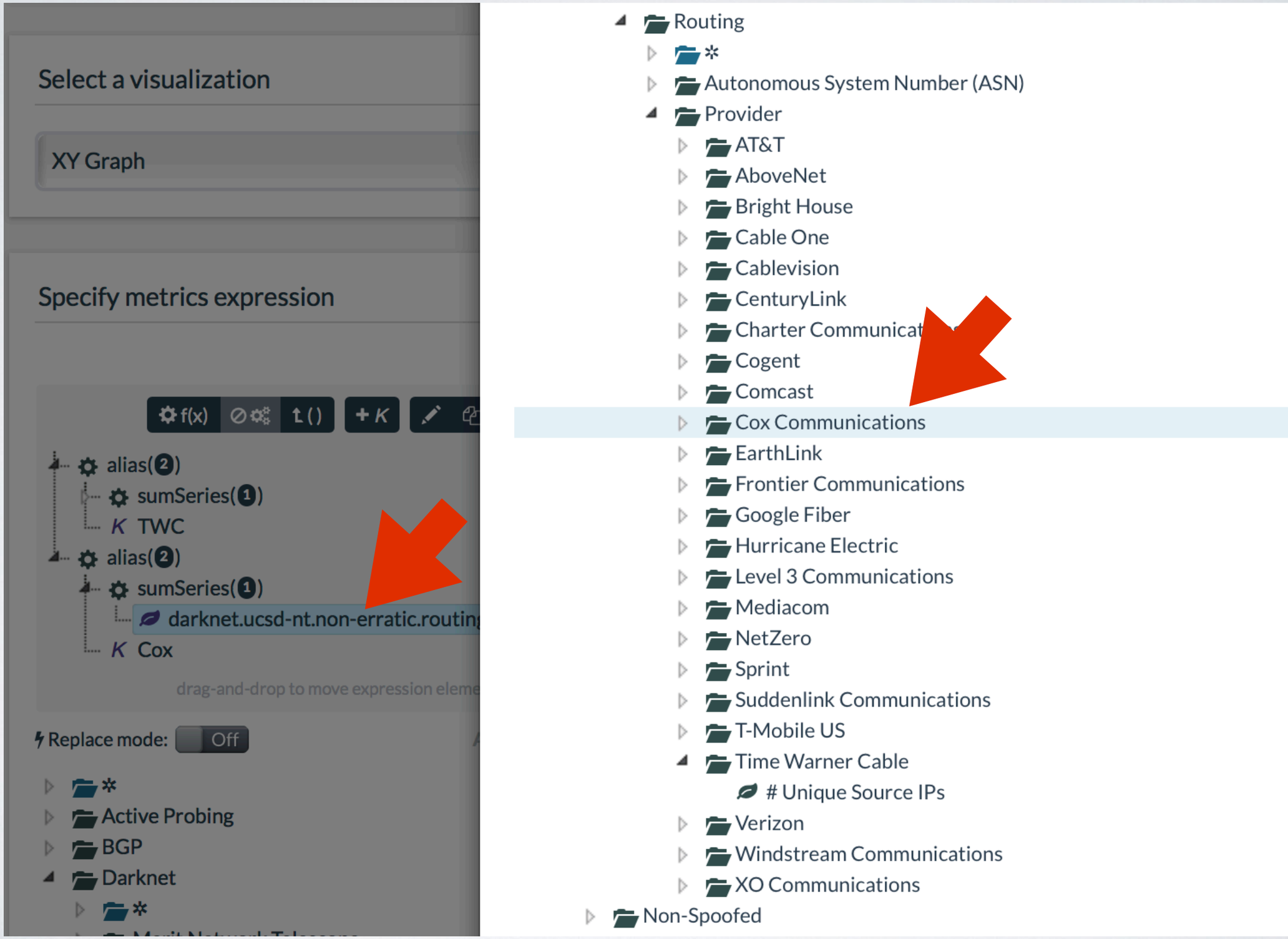


ADDING COX, COMCAST, AT&T (CLONING AND EDITING METRIC EXPRESSIONS)



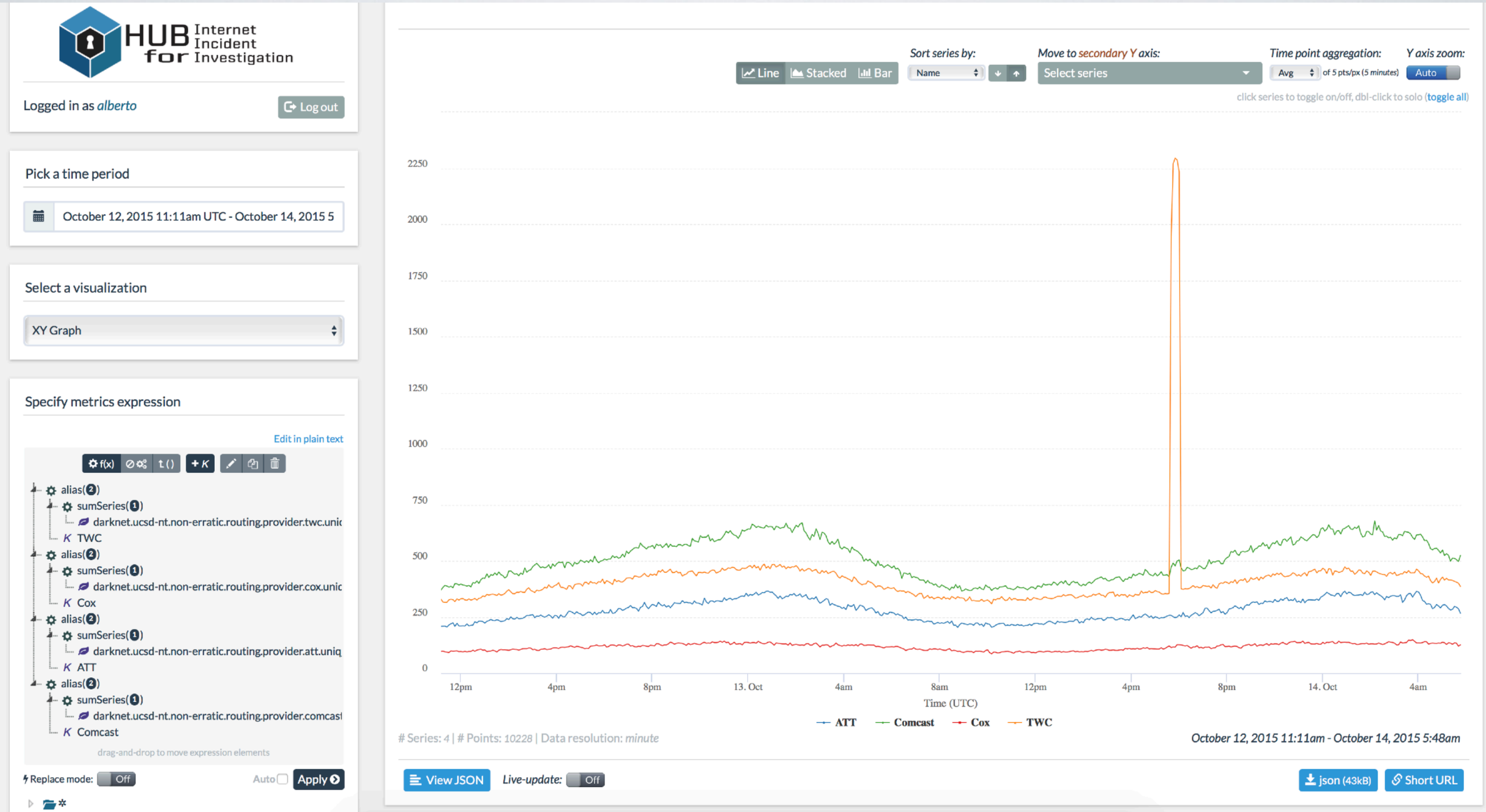
https://hicube.caida.org/@Pg

REMOVING OVERALL AND ADDING COX, COMCAST, AT&T



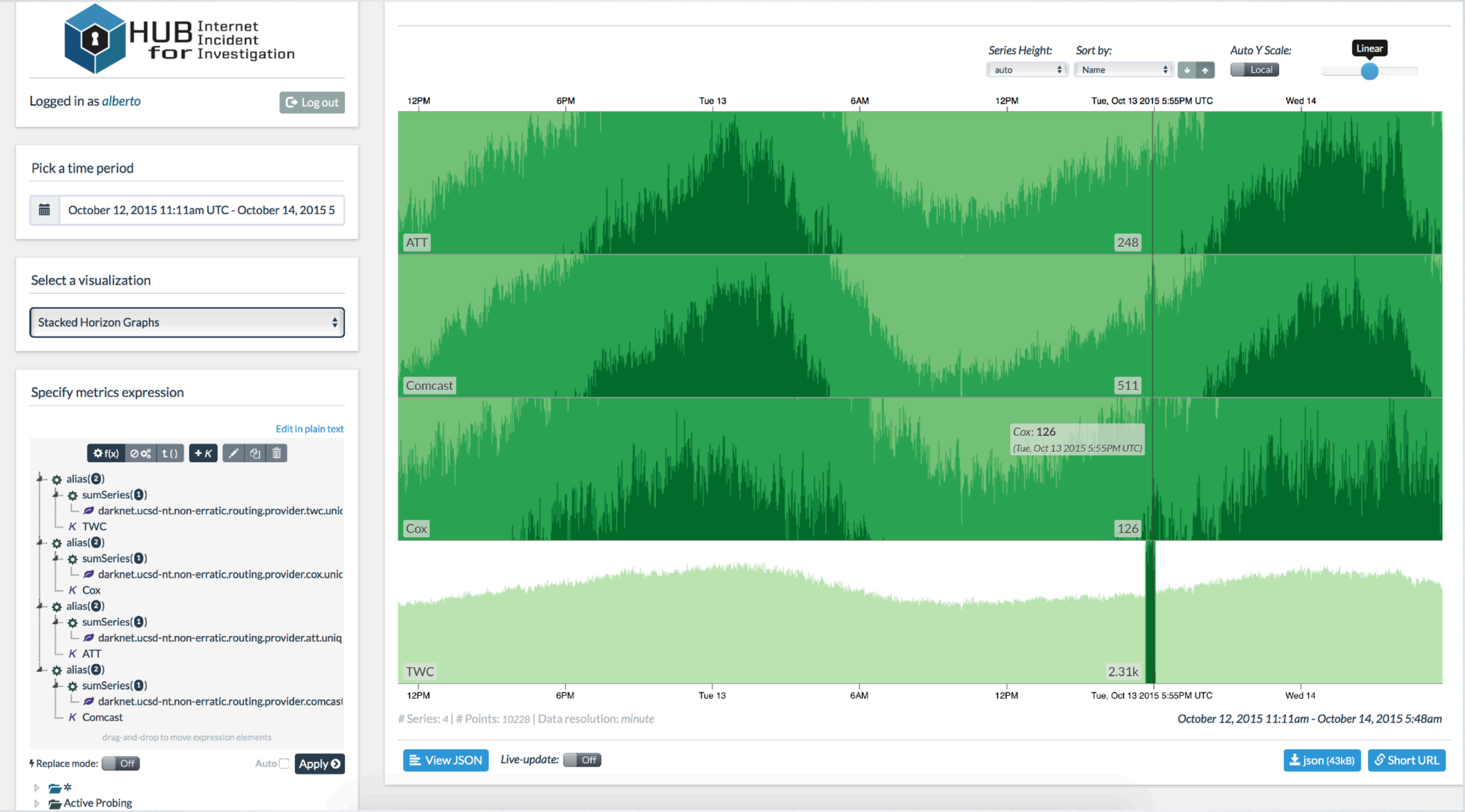
https://hicube.caida.org/@i0

TWC, COX, COMCAST, AT&T: PEAK ONLY IN TIME WARNER CABLE



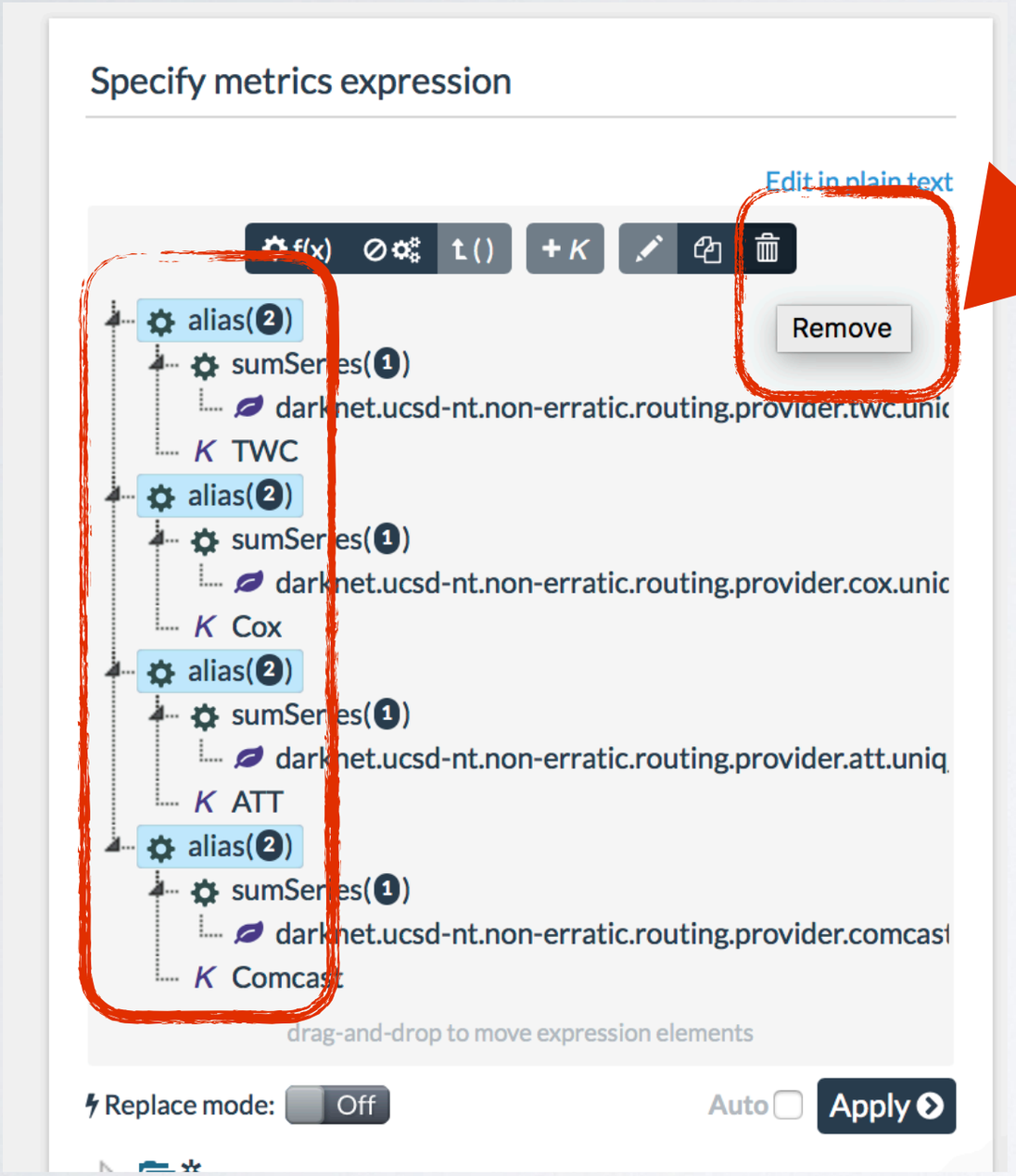
https://hicube.caida.org/@Pg

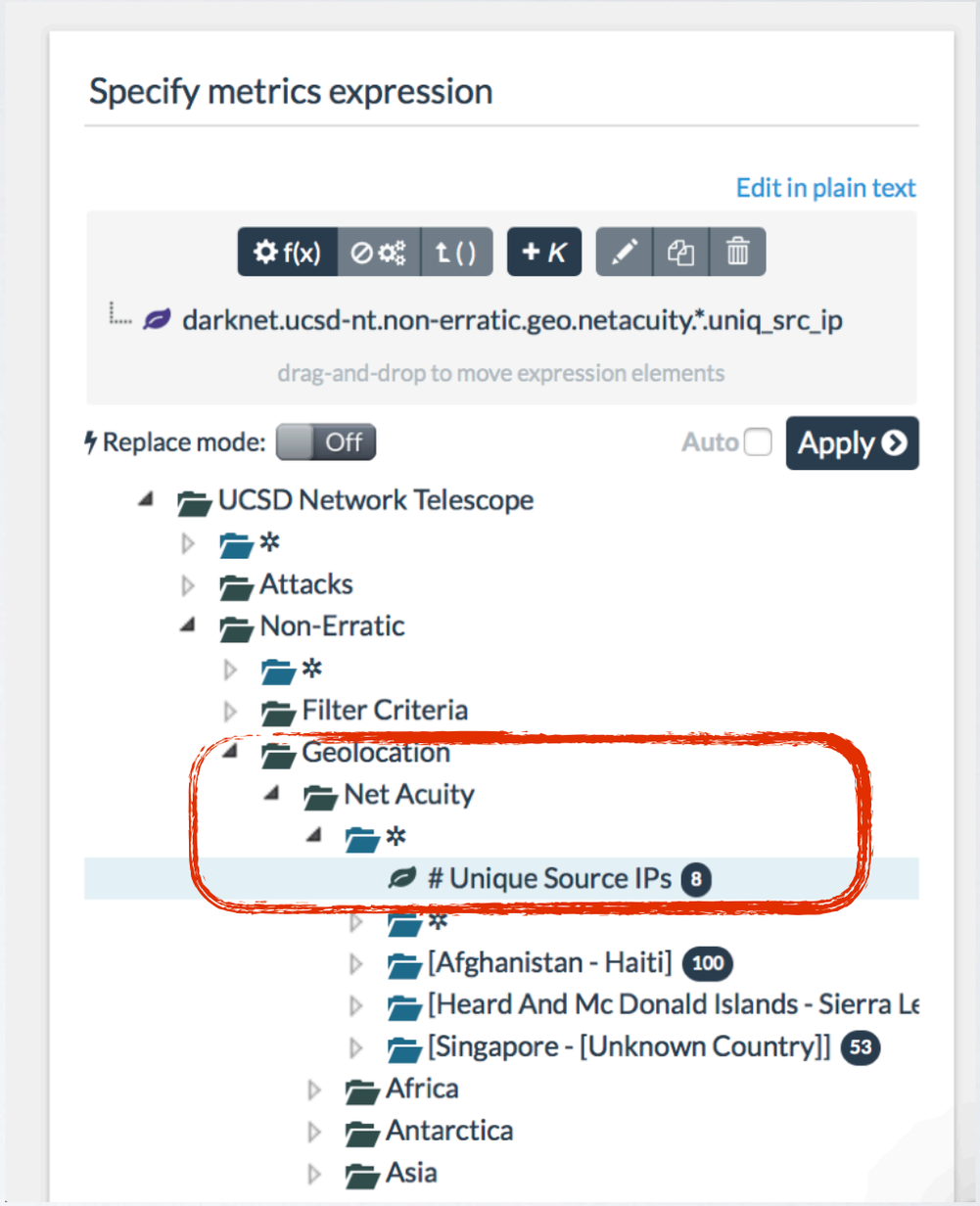
TWC, COX, COMCAST, AT&T (STACKED HORIZON CHARTS)



https://hicube.caida.org/@om

SWITCHING TO WORLD VIEW





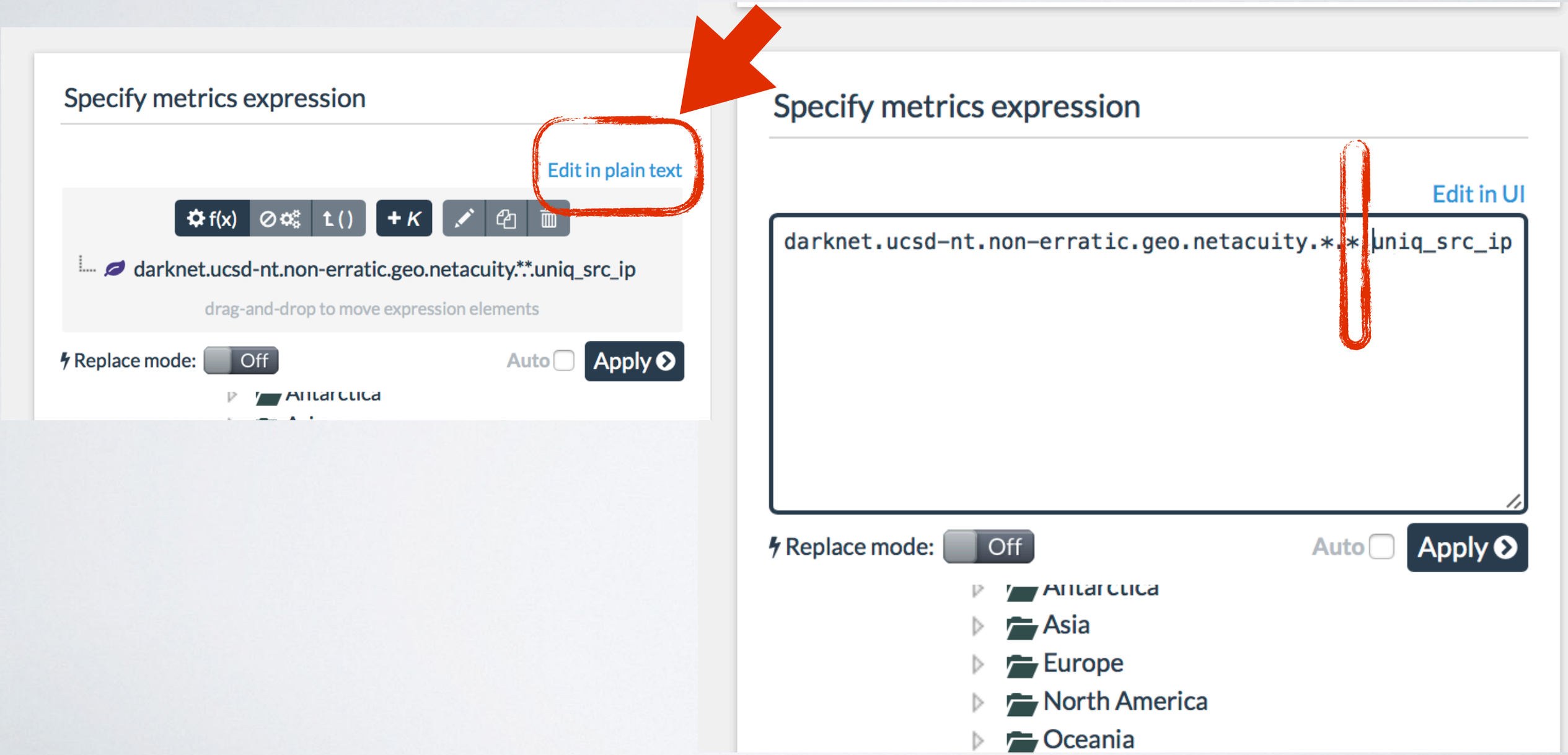
SWITCHING TO WORLD VIEW

https://hicube.caida.org/@6l

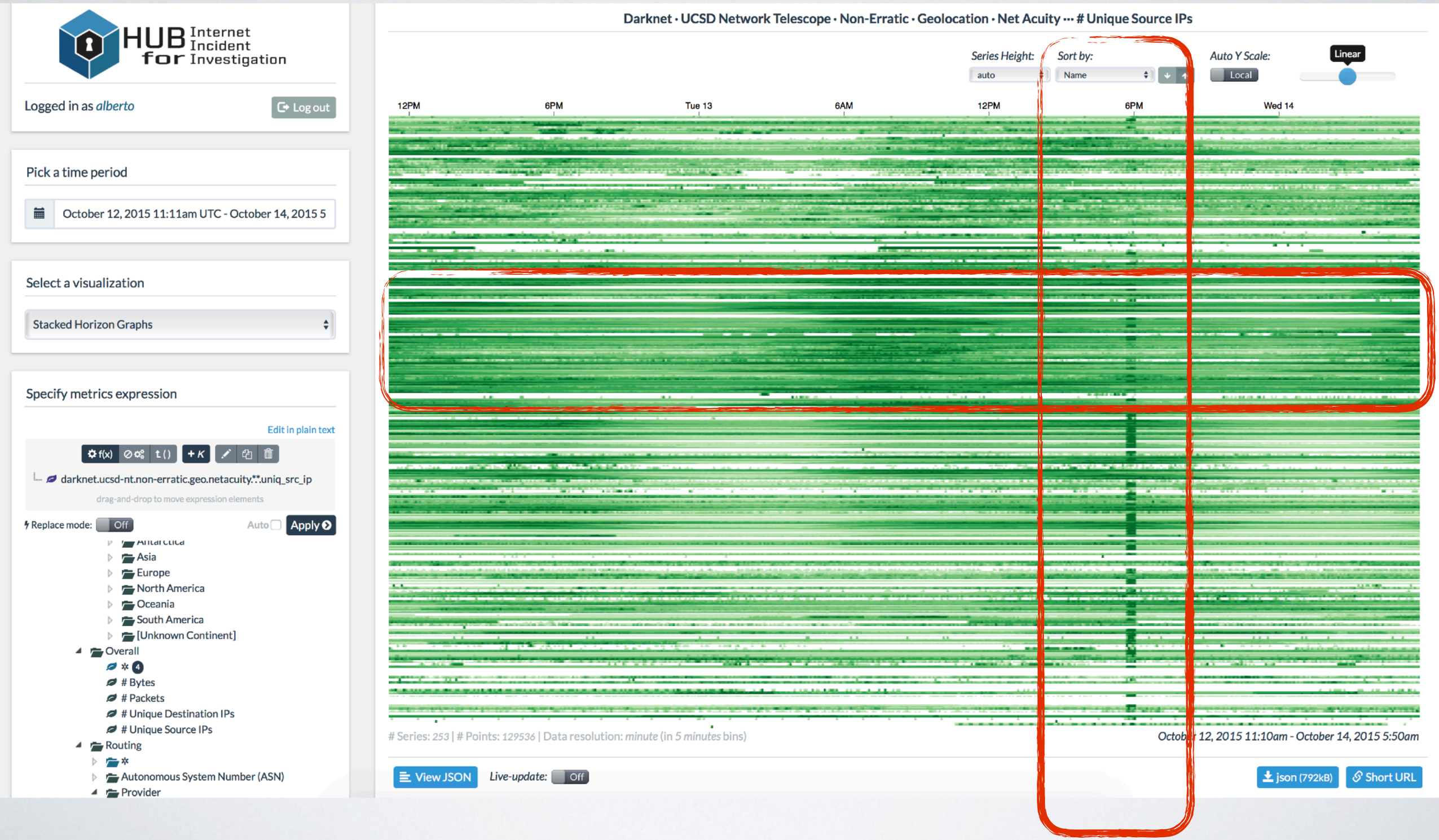
FROM EVERY CONTINENT; NOT MUCH FROM ASIA THOUGH

https://hicube.caida.org/@6l

EXPRESSIONS CAN BE EDITED MANUALLY: ADD * TO SWITCH FROM CONTINENTS TO COUNTRIES

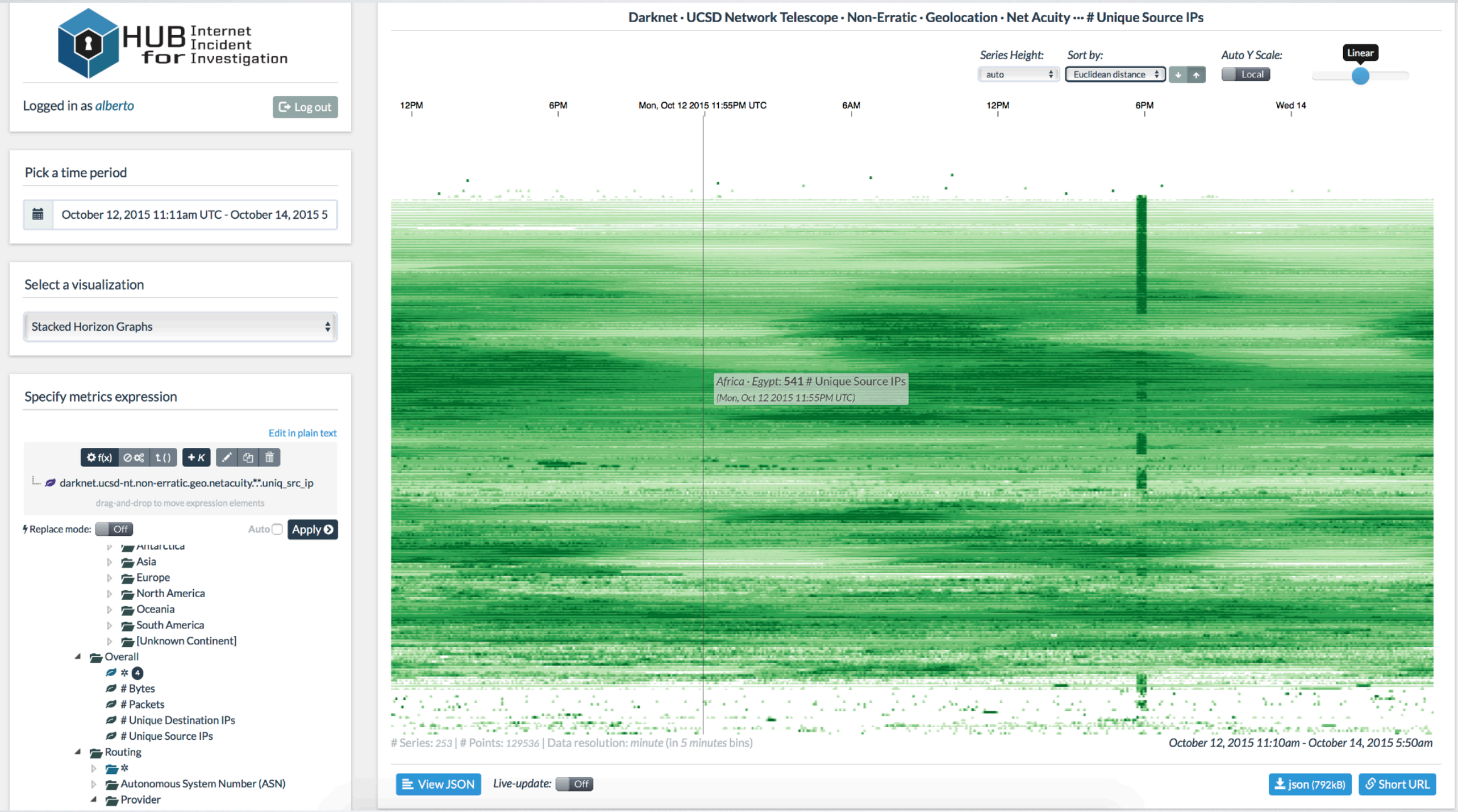


VIEW PER COUNTRY: SHOWS SOME ASIAN COUNTRIES SHOW THE PEAK



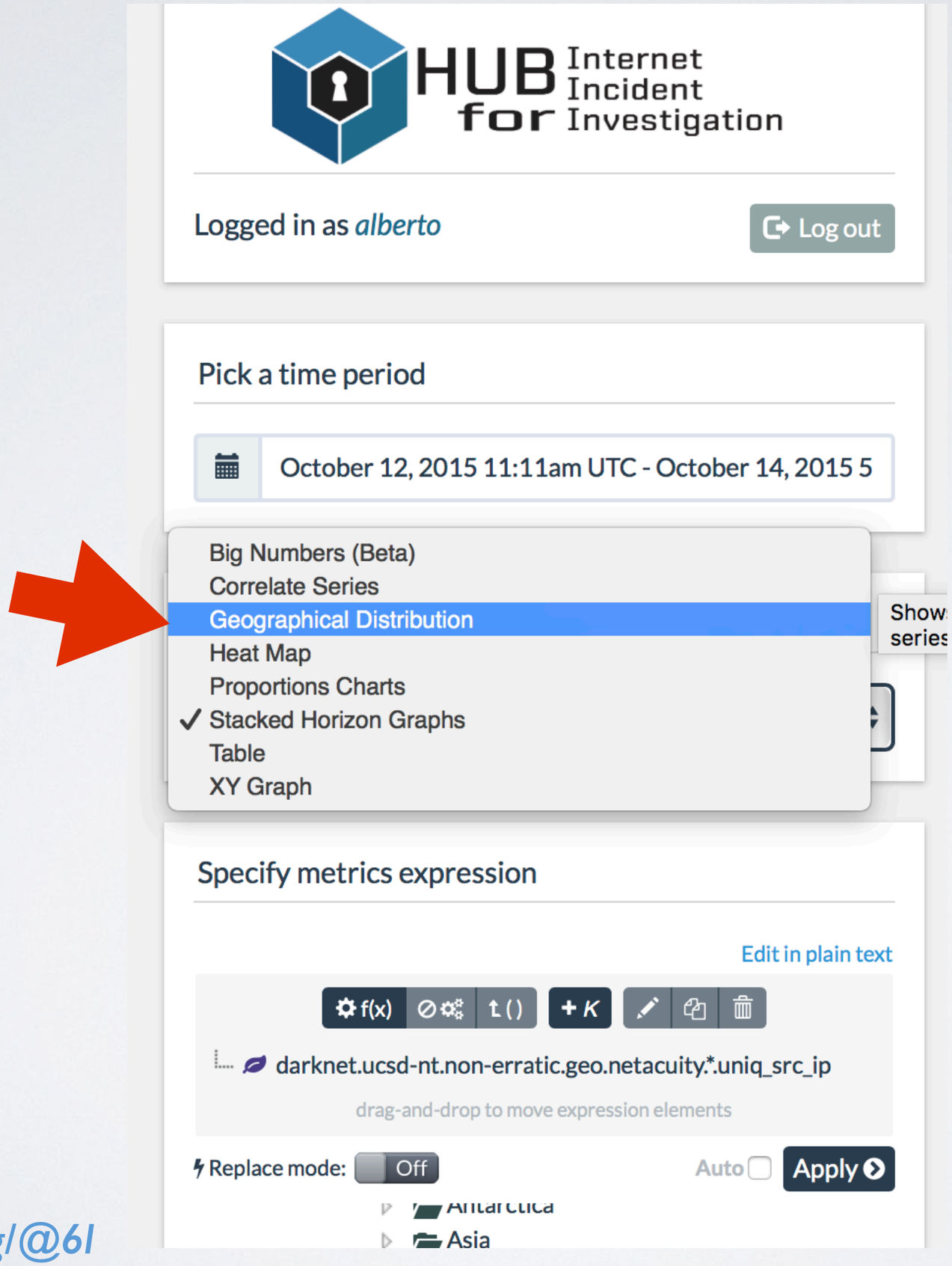
https://hicube.caida.org/@8m0

SORTING BY SIMILARITY



https://hicube.caida.org/@7mm

SWITCHING TO GEOMAP VIEW



https://hicube.caida.org/@6l

SWITCHING TO GEOMAP VIEW: ANIMATION SHOWS CHINA IS NOT PARTICIPATING

HUB for Internet Incident Investigation

Logged in as *alberto* Log out

Pick a time period

October 13, 2015 12:05am UTC - October 13, 2015 1

Select a visualization

Geographical Distribution

Specify metrics expression

Edit in plain text

normalize(1)

darknet.ucsd-nt.non-erratic.geo.netacuity.**.uniq_src_ip

drag-and-drop to move expression elements

Replace mode: Off Auto Apply

- Antarctica
- Asia
- Europe
- North America
- Oceania
- South America
- [Unknown Continent]
- Overall
  - *
  - # Bytes
  - # Packets
  - # Unique Destination IPs
  - # Unique Source IPs
- Routing
  - *
  - Autonomous System Number (ASN)

normalize(Darknet · UCSD Network Telescope · Non-Erratic · Geolocation · Net Acuity ··· # Unique Source IPs)

Exp base 6

Color scale: Fixed

Range: 0.00 – 1.00

0.800 1.00

Arctic Ocean · Greenland · Iceland · Canada · United States · Russia · Kazakhstan · Mongolia · China · India · Brazil · Argentina · Australia · Madagascar · South Africa · Indonesia · Philippines · Papua New Guinea · New Zealand · North Pacific Ocean · North Atlantic Ocean · South Pacific Ocean · South Atlantic Ocean · Indian Ocean

Leaflet | Map data © OpenStreetMap contributors

8 minutes /frame 2 fps

3AM 6AM 9AM 12PM 3PM 6PM 9PM

Series: 253 | # Points: 164680 | Data resolution: *minute* (in *2 minutes* bins)

October 13, 2015 5:45pm - October 13, 2015 6:00pm

View JSON Live-update: Off

json (2.4MB) Short URL

HI^3 BENEFITS

- Enhances our **ability to detect and understand** large-scale incidents
  - *live* streams of data
  - multi-source data is *combined* (same interface, tools, taxonomy, …)
  - dashboards and alert creation/subscription services
  - *collaborative* analyses
  - *crowdsourcing* analysis of public data
- **Extensible**
- **Increases the utility of the data/tools** and enables their in-context-evaluation
- **Lowers the barrier** for
  - stakeholders to provision/consume the data
  - more experts to engage/emerge and play with data

COMPETITION

each, focused on a single class of events

- typically focused on one class of events / one data source type
  - don't allow access to raw data
  - don't allow to "play" w/ the data
- single-user focused

COMPETITIONLEMENTARITY

it's more about complementarity

- **External synergy**
  - DHS IMPACT data providers
    - *in:* can feed HI-CUBE
    - *out:* HI-CUBE generates new raw data requests
  - THREAT/INTELLIGENCE platforms (e.g., ThreatConnect)
    - *in:* previously encoded threats may inspire analyses in HI-CUBE
    - *out:* they would use results from analyses made in HI-CUBE

- **Integration synergy**
  - integration of tools and datasets

HUB for Internet Incident Investigation

THANKS