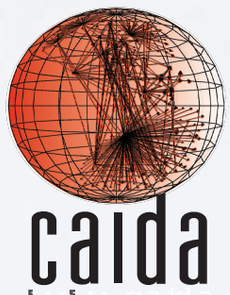
HI-Cube / HI³
Hub for Internet Incidents Investigation

Alberto Dainotti

alberto@caida.org



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

LARGE-SCALE INCIDENTS

a threat to private and national assets

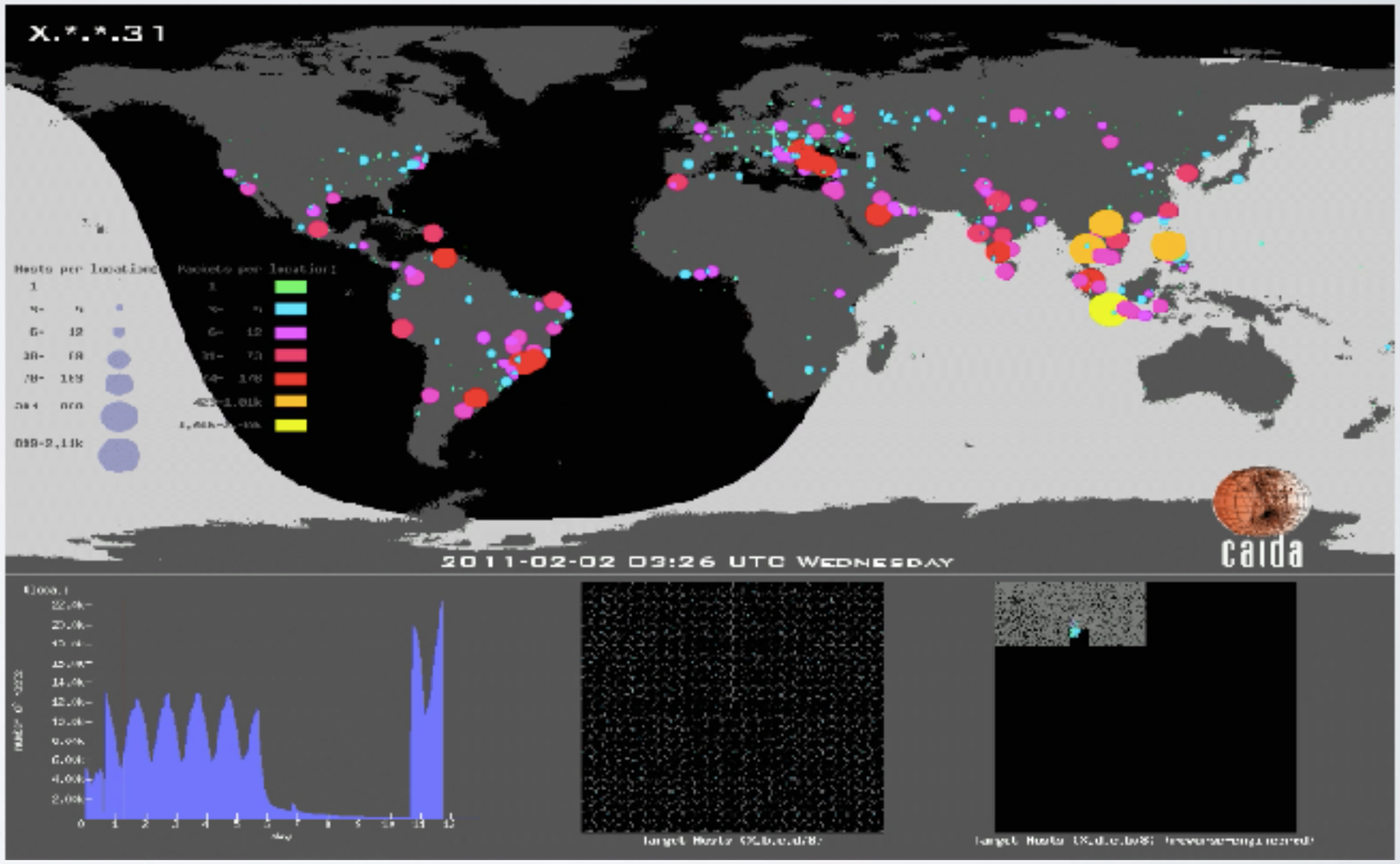
- **large-scale Internet incidents** *(hijacks, outages, spam and fishing campaigns, botnet activities, scanning, large-scale bug exploitation)* are a major threat to public safety and to both public and private strategic and financial assets
- *Often:*
  - ***unnoticed***
  - ***hard to understand*** (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

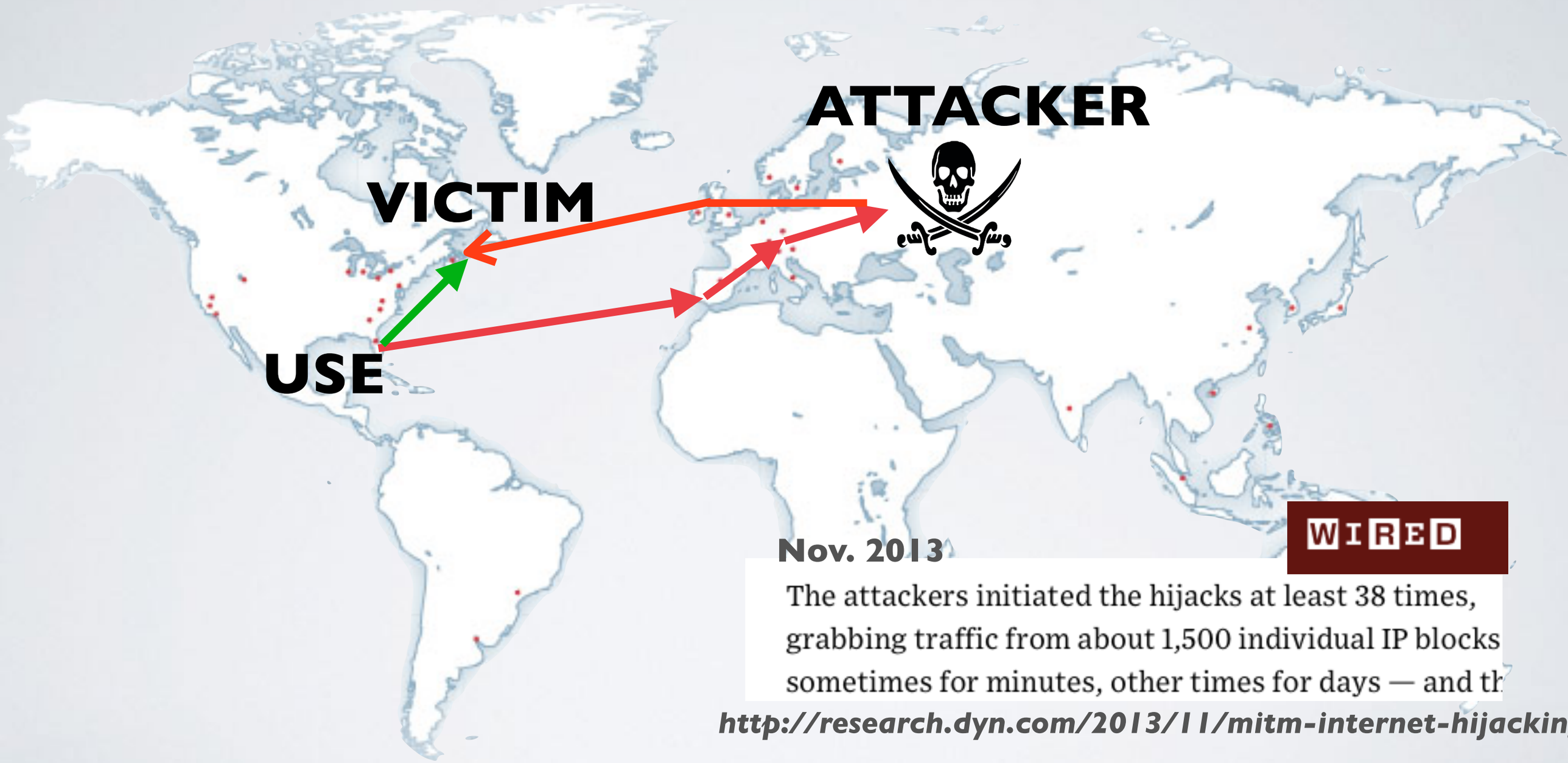
- *February 2011*
- *3M hosts covertly scanning the whole IPv4 Internet in 12 days*
- *Massive exploitation of VoIP infrastructure in the following months*

- *VoIP Fraud costs $40 billion per year*



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-hijackin

WE NEED

many features in one "place"

- Effective analysis of these events requires
  - data **extraction/aggregation**
  - **combination** of **data** of different type and origin
  - data and tool **sharing**,
  - **teamwork** of heterogeneous expertise
  - ability to act **fast** with **agility**
  - a **trusted** environment

THE HI^3 VISION

towards a distributed virtual situation room

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

THE HI^3 APPROACH

- Combination and correlation of diverse Internet cyber-security data
  - centering data *organization, processing, querying and visualization* around a set of common dimensions: time and **Internet Coordinates**

- Data analytics in the form of exploratory data analysis and event detection
  - interactive **navigation** through tens of millions of data streams
  - interactive + live data **visualization** interfaces (hundreds of time series and their 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

- Trusted collaborative environment
  - 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

THE HI^3 APPROACH

Internet Coordinates

- Primitives and Taxonomies for ***Internet Geography***:
  - *IP addresses* and their *aggregations* (IPs, /24s, prefixes, ASes, Siblings)
  - *geopolitical layer*: *geographic coordinates, administrative/political* (country, region, county, province, city, zip code, building, etc.)
  - DNS: *records, passive DNS and active DNS databases*
  - BGP: *prefixes, AS numbers*, …
  - BGP econ/etc: *siblings, AS-relationships*, AS customer cones, CDNs
  - *Whois* and routing registries
  - Internet *census*
  - Internet *cybercensus*: profiling of hosts and networks
  - Topologies: *AS graph, router-level* topology, *physical* (links, facilities, …)

BOOTSTRAP

building on top of existing platforms

- Our ***Web interface*** for monitoring the Internet 24/7 to detect large-scale internet outages: ***visualization tools*** for exploration, correlation, rapid-prototyping, dashboards — *ioda.caida.org*

- ***Infrastructure*** for managing millions of streams of (archived) ***time series***

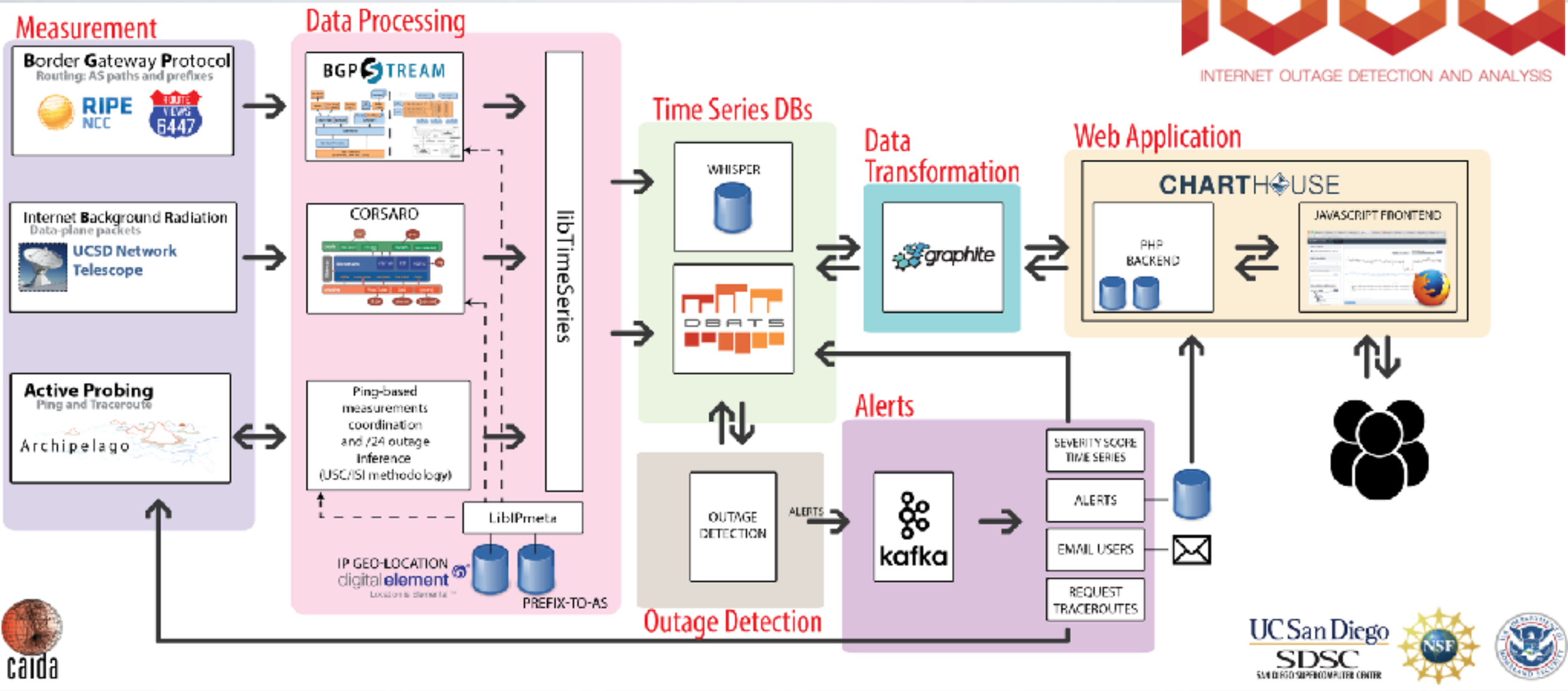
- ***Software components*** and ***data*** for ***Internet Geography***

- ***Legal framework*** from IMPACT

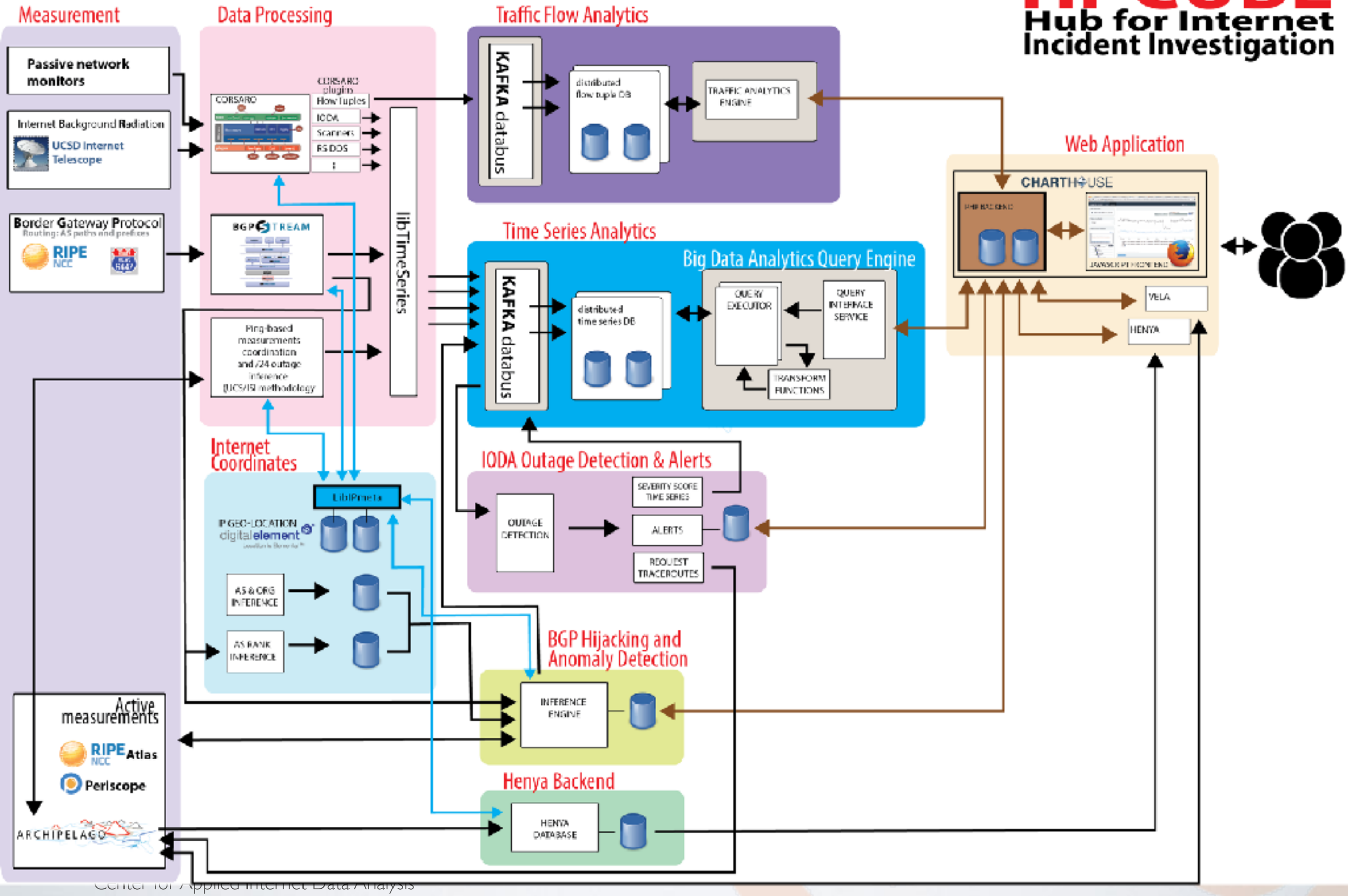
INFRASTRUCTURE DEMO

IODA

INTERNET OUTAGE DETECTION AND ANALYSIS



HI-CUBE
Hub for Internet Incident Investigation



INTERNET OUTAGE DETECTION AND ANALYSIS

REALTIME
processing

Multi-User
& Collaborative

AGGREGATE
by time/internet coordinates

SCALABILITY

LARGE +
DIVERSE
data

\+

DATA FEEDS

=

HI-CUBE
Hub for Internet Incident Investigation

TRANSFORM /
CORRELATE

MORE ANALITCS

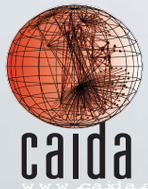
MORE FUNCTIONALITIES

VISUALIZE

IMPACT legal framework
+ DHS vetting

Center for Applied Internet Data Analysis
University of California San Diego

THANKS



Center for Applied Internet Data Analysis
University of California San Diego