# a Abstract

**Title**: IODA-NP: Multi-source Realtime Detection of Macroscopic Internet Connectivity Disruption

**Total Anticipated Costs: $3M**

As most of the world's communication infrastructure converges on use of the Internet, governments around the world are increasingly realizing they must understand how to provide the level of reliability and resilience we have come to take for granted from the public switched telephony network, now on its way out in the U.S. But in spite of the societal and economic impact of Internet connectivity disruptions, we lack any rigorous understanding of Internet outages or even sufficient tools for their systematic and timely identification. This gap is rooted in persistent challenges in Internet measurement and monitoring, such as the lack of measurement-aware design of Internet protocols, inherent complexity of (noisy) network measurement data, and the lack of systematic ground-truth data, combined with the highly diverse and continuously evolving nature of Internet infrastructure. Challenges in Internet connectivity outage identification are exacerbated by the heterogeneity of disruption events and their characteristics: many endemic or external factors can trigger an outage, ranging from human error such as misconfiguration, government-mandated shutdowns, and cyber-attacks, to cable cuts, network failures, natural disasters, power outages, etc.

In the last 5 years, the Center for Applied Internet Data Analysis (CAIDA) at UC San Diego (Offerer) leveraged its two decades of experience in Internet measurement and data analysis, to tackle this research problem. Specifically, we focused on the detection and characterization of **large-scale Internet outages, *i.e.,* connectivity disruptions significantly affecting large geographic regions or specific Internet operators**. Our work proceeded in parallel along two dimensions: we proposed and combined innovative methodologies for outage detection and we developed software frameworks as building blocks to implement and test our approaches. As a result, early this year we presented a prototype of our *"IODA" (Internet Outage Detection and Analysis)* system to detect events of connectivity disruption in near-realtime.

In response to CSD BAA, HSHQDC-17-R-B0002, via strategically planned participation in the PARIDINE program TTA#1, we propose to conduct applied research and development in order to *(i)* define **a rigorous framework, *IODA-NP (Next Phase)*, for large-scale outage detection based on IODA methodologies and software building blocks**, *(ii)* perform methodological and technological improvements to our approaches and systems in order to develop and deploy this framework as a **near-realtime capability monitoring the Internet 24/7**, and *(iii)* **systematically identify and quantitatively evaluate its capabilities and limitations in the real world**. In addition, we propose an optional Pilot Task in which we will integrate, deploy and test our solution into operation in cooperation with an entity identified by the DHS. Our project includes applied research, software development, new data analytics, systems integration, operations and maintenance, validation, and event analysis and reporting.

# b Performance Goals

The Regents of the University of California; University of California, San Diego on behalf of the San Diego Supercomputer Center's Center for Applied Internet Data Analysis (CAIDA), offer this technical proposal, which includes the following goals: (1) a measurement and data analysis system, based on a rigorously defined framework, which will improve our ability to identify and monitor episodes of macroscopic Internet connectivity disruption that threaten the security and reliability of the nation's communication capabilities; (2) a Web interface to access dashboards and alerts and visually inspect specific events (software-as-a-service) as well as programmatic APIs to access both live streams and historical data of the alerts generated by the system; (3) periodic reports about the events detected by the system and analyzing its performance, limitations and potential improvements; (4) (*optional*) deployment and testing of our solution into operation in cooperation with an entity identified by the DHS.

The proposed work targets objectives outlined in TTA#1 of PARIDINE's solicitation: Definition, Identification, and Production of Network / Internet Disruptive Events (NIDEs). We will develop and evaluate a methodology and build, integrate, and operate multiple open-source software tools for near-realtime detection of large-scale Internet outages affecting large geographic regions or specific Internet operators. Our detection, reporting, and analysis system will help identify different types of events of connectivity disruption with diverse potential impact on communications, other critical infrastructures, public safety, financial services, etc. To achieve the four mentioned goals, we propose to leverage the results of technologies and infrastructure funded by the Department of Homeland Security and the National Science Foundation, in particular the methodologies and the software components developed in CAIDA's *IODA (Internet Outage Detection and Analysis)* project. We will combine outage inference approaches operating both at the Internet control plane (BGP) and data plane, using both passive (Internet Background Radiation) and active approaches (continuous probing). We will define specific metrics and provide definitions of events for each inference approach and propose an applied methodology to fuse these inferences in order to (i) increase their accuracy and (ii) yield a better understanding of the characteristics of the disruptive event detected (e.g., a cable cut or a natural disaster will likely affect the control plane, whereas network congestion will appear as a data-plane anomaly only in most situations). In addition to APIs enabling access to alerts generated by our service, to enhance user capabilities to understand, quantify and categorize each event, we will provide visual interfaces for the inspection and correlation of current and historical data.

Our proposed work in response to PARIDINE's call will result in ***IODA-NP (IODA Next Phase)***, a rigorously defined and evaluated (on real events) framework and a usable standalone product (provided as software-as-a-service) as well as methods and software modules that in the future could be integrated into broader solutions (e.g., for risk assessment, root cause analysis, or event prediction). The resulting technologies will support situational awareness and decision analytic needs of NCCIC and other government agencies. We perform this applied research and development on a reasonable efforts basis.

# c Detailed Technical Approach

**Introduction**

The focus of our proposed work is the identification and characterization of macroscopic events of connectivity disruption that affect the Internet, potentially disrupting other critical infrastructure components that now rely on Internet communications, e.g., power grid, VoIP enhanced 911, financial institutions. These types of events cover the disruptive event scenarios exemplified in sections 2.7.1, 2.7.2, 2.7.4, 2.7.5, 2.7.6 of the PARIDINE BAA. CAIDA pioneered methodologies to characterize large-scale Internet outages in 2011 and 2012, caused by natural disasters and human intervention In 2012 we received funding from the NSF to start a project called IODA (Internet Outage Detection and Analysis) to develop a proof-of-concept system implementing our methodologies [1]. To support this project, we developed several IODA software components (e.g., [2]) thanks to this and other funding from NSF and DHS S&T. We completed this project in early 2017, resulting in mature methodologies to extract "liveness" signals from various types of Internet measurement data, a set of open source software components and visual interfaces, and a demo dashboard. In response to BAA HSHQDC-17-R-B0002, we propose to take IODA to the *next phase* (IODA-NP) and produce a framework and a tool addressing the requirements of the PARIDINE solicitation, significantly advancing the current state of the art in near-realtime global Internet monitoring and specifically in the detection and classification of disruptive Internet events such as connectivity outages.

**Existing Work: IODA**

In IODA we developed measurement and data analysis methodologies based on three distinct types of sources of Internet measurement data:

1. *Internet Background Radiation (IBR)* is one-way unsolicited traffic generated by millions of Internet hosts worldwide, due to misconfiguration, malware propagation, scanning, etc. IODA is the only outage inference system using IBR, thanks to a methodology that we demonstrated is capable of detecting outages caused by state censorship [3], natural disasters [4], and border router misconfiguration in Autonomous Systems (ASes) [5]. From IBR, we filter out spoofed traffic and bursty traffic components (e.g., due to scanning from large botnets) and extract a "liveness signal" based on the number of distinct source IP addresses observed from a given geographic region or AS. We collect IBR traffic through the UCSD Network Telescope, an almost entirely unutilized /8 IPv4 address block, estimated to observe 1/256th of all the IBR generated in the Internet. As of March 2017, the telescope captures more than 1TB of compressed traffic per day. We continuously process this traffic using our Corsaro open-source software platform.
2. *BGP routing information*. For this data source, we leverage the collection infrastructure operated by the RouteViews and RIPE RIS projects. We infer the state of the routing tables exported by hundreds of operational routers by processing BGP updates and RIB dumps and we extract information about which network blocks (BGP prefixes) appear reachable on the Internet control plane from most of these vantage points. Different from other organizations occasionally reporting BGP-visible connectivity disruption (*e.g.,* Renesys/Dyn), our approach counts visible /24 blocks instead of prefixes, more meaningfully quantifying which fraction of the address space normally announced by an AS or from a region is reachable at a certain point in time. We continuously process data from more than 300 operational BGP routers using our BGPStream open source software framework.

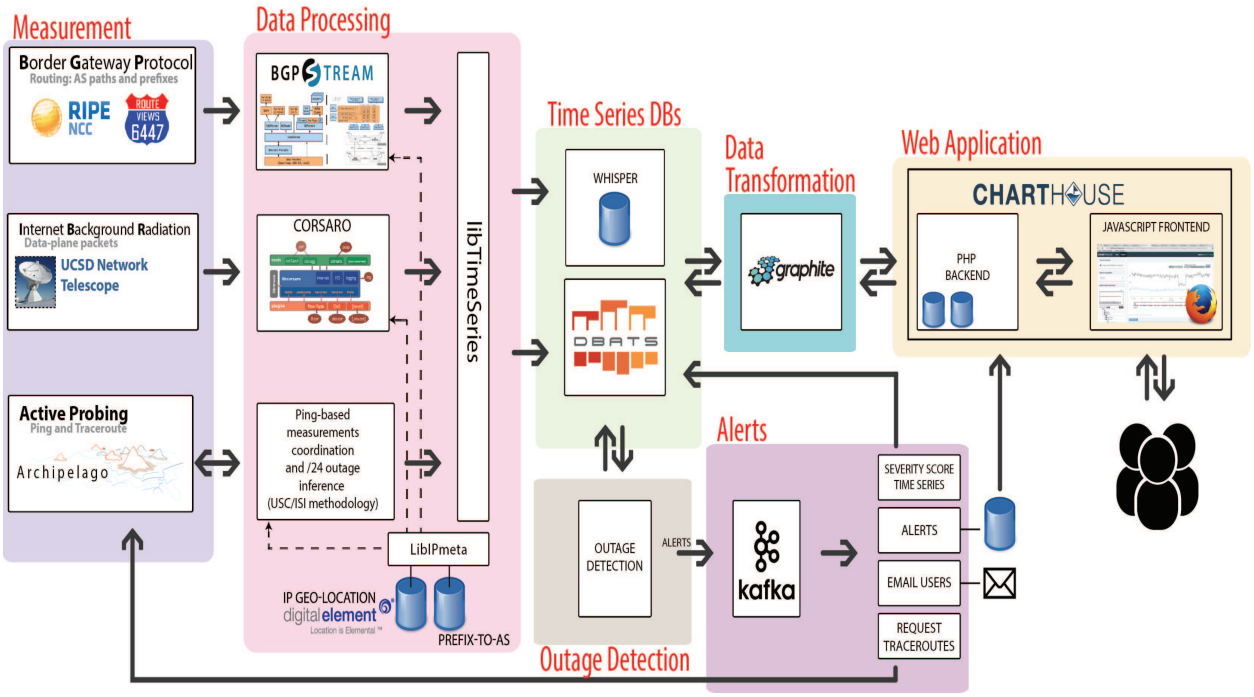


Figure 1: Overview diagram of the IODA/IODA-NP architecture [1].

3. *Active probing*. We periodically probe approximately 3.5 M /24 network blocks worldwide and adaptively send more probes upon lack of response using the *Trinocular* methodology developed by ISI/USC [6]. We run our measurements from a few dozen CAIDA Archipelago (Ark) nodes distributed worldwide and from a central node at UC San Diego.

Figure 1 outlines the high-level overview of the current IODA architecture. Raw measurement data from each source (*Measurement* block in the diagram) is processed and augmented through IP geo-location and IP-to-AS lookups (*Data Processing* block), in order to extract a time series "liveness signal" for each observed region and AS, according to each data source. We continuously (*e.g.,* every minute, in the case of IBR) add data points to these time series stored in CAIDA's DBATS open source high performance time series database (*Time Series DB*). We implemented simple preliminary outage detection triggered by drops of the signals below thresholds that we arbitrarily selected, based on brief experimental observation of the effect of the thresholds on the frequency of alerts generated (*Outage Detection*). Measuring the normalized offsets from the threshold we obtain an estimate of the significance of the detected event. If the event is detected through more than one data source (e.g., BGP and IBR) we multiply the respective offsets to combine inferences and obtain an overall confidence level. The alerts are fed into an Apache Kafka queue (*Alerts*) and then stored as time series data points. In the same block we started implementing functionalities to trigger additional Ark measurements (*e.g.,* traceroutes) upon event detection. We developed a demo dashboard *Web Application*, showing alerts and per data source signals for events detected by the system, implemented in PHP/Javascript and accessing the time series databases through a Python backend (*Data Transformation*) based on the Graphite open source software.

**Proposed Work: IODA-NP**

Five years of research and development in IODA make us uniquely positioned to deliver to the PARIDINE program an effective tool for the detection of events of connectivity disruption. The initial IODA project yielded two significant achievements. *(i)* Our measurement and inference approaches proved useful in highlighting and characterizing disconnection events. We showed ev-

idence of outages captured by IODA that affected entire countries (*e.g.,* in Iraq because of state censorship, in North Korea with speculations about cyberwarfare, in Turkey because of a power outage), smaller regions (the impact of Hurricane Sandy on the U.S. east coast, cable cuts in the San Francisco Bay Area), and operators (Time Warner Cable outages in the U.S. in 2014 and 2015). We also demonstrated that combining the three data sources can improve coverage, increase the confidence level of inferences, and help classify events and reveal their root cause. However, our current approaches lack formal metric-based definitions of targeted events and quantitative objectives in terms of accuracy and coverage. We have developed only a naive anomaly detection approach with arbitrary thresholds and we have not systematically and rigorously validated its inferences. In other words, we still need to investigate, exploit, and engineer into a rigorous framework the full potential of our methods. *(ii)* We have developed powerful software frameworks and modules that allow us to cope with the vast complexity and heterogeneity of the data continuously analyzed in IODA and to demonstrate the practical feasibility of our approach. However, our system is still at a proof-of-concept stage: it is fragile and susceptible to overload, which is problematic for a system expected to monitor critical infrastructure 24/7; it also produces inferences with a delay of two hours and coarse time granularity in some cases.

Our proposed applied research and development of the IODA-NP framework will advance along two dimensions: ***Methodologies***, to formally define and evaluate our framework and improve methods in terms of scope of detected events and detection accuracy; ***System performance***, to reduce latency, generate output with finer time granularities, and achieve sufficient reliability to support 24/7 monitoring. The evidence-based iterative development approach described in TTA#1 of the PARIDINE call perfectly matches the path we envision to fill these gaps. Our proposed work (in the first two years) is organized in four tasks that directly map to the four goals of TTA#1: Definitions and Metrics, Development of the Framework, Report and Analysis of Detected Events and Validation, Development of APIs; we propose a fifth task to coordinate our Iterative Development process. Finally, our optional Pilot tasks will allow us to deploy, integrate and evaluate IODA-NP into an operational environment based on the needs of a partner entity identified by the DHS.

**Technical Challenges**

The major technical challenges in our approach are: development and maintenance of infrastructure to support sustainable collection, processing, and storage of large volumes of diverse data, including data volume and quality, validation of data and inferences, and general infrastructure scalability issues. Further, the highly interactive nature of IODA presents a major technical challenge in terms of the design, implementation and optimization of several distributed processing systems in order to minimize end-to-end query latency. Other challenges include the integration of diverse external data sets of widely varying formats and quality; integration of software components, both internally and externally developed; and supporting systems that must operate continuously (24/7) to provide real-time data collection, processing and visualization of data. Many of these challenges we have successfully overcome with the IODA project, but we anticipate similar challenges in IODA-NP.

**Strategies for Mitigating Risks**

To support risks, we budgeted for dedicated system administration resources to manage inevitable reliability issues with COTS hardware and solutions. We are relying on open source software components so that we can modify code as needed to scale to our needs. We have embedded resiliency and redundancy into our system architecture, and distributed, fault-tolerant storage and analytics environments. To deal with storage scalability, we will use an expandable, distributed

object-storage architecture for archival and distribution of large datasets, and we will implement and execute sophisticated pre-processing techniques to achieve data compression and minimal loss of data utility. Fortunately we have demonstrated successful approaches to all of the challenges described above, in the current IODA platform, although each one takes sometimes significant time and effort to overcome. We have budgeted for sufficient time to iterate our development throughout the project as well as get repeated feedback from collaborators and colleagues, *e.g.,* by presenting current state of technology at each PI meeting. For all software and systems development, we will follow an iterative implementation approach where we will rapidly develop a working system, and then iteratively deploy, evaluate, refine, and adapt components based on results of benchmarking and user testing.

Another priority is validation: given the prevalence of potentially false positives in detection of connectivity disruptions, we integrate diverse data sources for cross-validation of inferences. We will confirm, the detected events as well as analyze undetected events possibly disclosed by other sources, comparing our inferences with information reported by public media, nonprofit organizations, and operators' mailing lists (e.g., *outages* mailing list, NANOG, RIPE). We also will leverage our trust relationships with operators to acquire ground truth against which to validate our inferences. In addition, we will collaborate with other PARIDINE performers working in the same space to cross-validate our results.

# d *

References


[1] A. Dainotti and K. Claffy, "Detection and analysis of large-scale Internet infrastructure outages (IODA)." `http://www.caida.org/projects/ioda/`, 2012.

[2] C. Orsini, A. King, D. Giordano, V. Giotsas, and A. Dainotti, "BGPStream: a software framework for live and historical BGP data analysis," in *Internet Measurement Conference (IMC)*, Nov 2016.

[3] A. Dainotti, C. Squarcella, E. Aben, K. C. Claffy, M. Chiesa, M. Russo, and A. Pescapé, "Analysis of country-wide Internet outages caused by censorship," in *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement*, IMC '11, pp. 1–18, ACM, 2011.

[4] A. Dainotti, R. Amman, E. Aben, and K. C. Claffy, "Extracting benefit from harm: using malware pollution to analyze the impact of political and geophysical events on the internet," *SIGCOMM Comput. Commun. Rev.*, vol. 42, pp. 31–39, Jan. 2012.

[5] K. Benson, A. Dainotti, k. claffy, and E. Aben, "Gaining Insight into AS-level Outages through Analysis of Internet Background Radiation," in *Traffic Monitoring and Analysis Workshop (TMA)*, Apr 2013.

[6] L. Quan, J. Heidemann, and Y. Pradkin, "Trinocular: Understanding Internet Reliability Through Adaptive Probing," pp. 255–266, 2013.


# e Testing and Evaluation

**Proposed use for DHS/S&T.** The system will allow DHS components to improve the ability to identify, monitor, and model critical cyber-infrastructure, and to assess the impact of disruptions in terms of number and type of networks affected, and how long before connectivity is restored.

**Operational Utility Assessment Plan.** The initial testing environment for the first two years of the project will be PARIDINE PI meetings and CAIDA workshops series (AIMS, WIE, DUST, IMAPS). To measure, assess and evaluate the current and projected operational value of this technology, we propose to visit partner locations including the NCCIC and FCC to provide hands-on tutorials and documentation to explain how to effectively use the IODA platform. The iterative process that is essential to our software development methodology will allow us to integrate feedback from DHS components on how to improve operational utility into subsequent versions of our platform. We will seek also seek testing and evaluation from network operators (through NANOG and IETF interactions). Their interest in an ability to use the tool for monitoring and detection of outages, hijacks, and other disruptive events will form the basis of our utility assessment.

**Metrics of utility and management of data and/or tools** We will use traditional quantitative metrics including number of users, activity on project mailing lists from external participants, usage statistics of the IODA platform. We will also solicit feedback at workshops and via periodic surveys of users.