AFRL-RI-RS-TR-2021-072

ADVANCING SCIENTIFIC STUDY OF INTERNET SECURITY AND TOPOLOGICAL STABILITY (ASSISTS)

UNIVERSITY OF CALIFORNIA SAN DIEGO - CENTER FOR APPLIED INTERNET DATA ANALYSIS (CAIDA)

APRIL 2021

FINAL TECHNICAL REPORT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

STINFO COPY

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE

■ **AIR FORCE MATERIEL COMMAND** ■ **UNITED STATES AIR FORCE** ■ **ROME, NY 13441**

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09. This report is available to the general public, including foreign nations. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RI-RS-TR-2021-072 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE CHIEF ENGINEER:

/ S /
FRANCES A. ROSE
Work Unit Manager

/ S /
GREGORY J. HADYNSKI
Assistant Technical Advisor,
Computing & Communications Division
Information Directorate

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
APRIL 2021	FINAL TECHNICAL REPORT	DEC 2017 – AUG 2020

4. TITLE AND SUBTITLE

ADVANCING SCIENTIFIC STUDY OF INTERNET SECURITY AND TOPOLOGICAL STABILITY (ASSISTS)

5a. CONTRACT NUMBER
FA8750-18-2-0049

5b. GRANT NUMBER
N/A

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Kimberly Claffy and Alberto Dainotti

5d. PROJECT NUMBER
DHSA

5e. TASK NUMBER
SS

5f. WORK UNIT NUMBER
IS

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of California San Diego
Center for Applied Internet Data Analysis (CAIDA)
900 Gilman Drive Dept 621
Lo Jolla CA 92093-0621

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory/RITGB
525 Brooks Road
Rome NY 13441-4505

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/RI

11. SPONSOR/MONITOR'S REPORT NUMBER
AFRL-RI-RS-TR-2021-072

12. DISTRIBUTION AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited. This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This project has two purposes: (1) providing data that supports identification and monitoring of network events, interdependencies and vulnerabilities; and (2) a data analysis system that improves our ability to identify, monitor, and mitigate infrastructure vulnerabilities that threaten the reliability of the nation's communication capabilities. As a Data Provider, we provided data sets that targeted security, stability, and resilience of communications infrastructure. As a Decision Analytics-as-a-Service Provider, we undertook a system integration effort to prototype a platform that enables experts to analyze diverse sets of live and historic macroscopic Internet data. By providing both data-preprocessing tools as well as the infrastructure for data sharing and analysis, we lowered the barrier for organizations to provide and consume relevant and sometimes sensitive data, and created an extensible environment for testing collaborative analysis and fusion of diverse Internet data in the context of cyber-security analytics.

15. SUBJECT TERMS

Real-time internet monitoring, vulnerability assessment, incident investigation, cybersecurity data sharing and analysis

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			**FRANCES A. ROSE**
U	U	U	UU	30	19b. TELEPHONE NUMBER (*Include area code*) **N/A**

TABLE OF CONTENTS

LIST OF TABLES

1. SUMMARY OF ACCOMPLISHMENTS

Research and development targeted at identifying and mitigating Internet security threats require current network data. To fulfill this need, researchers working for the Center for Applied Internet Data Analysis (CAIDA), a program at the San Diego Supercomputer Center (SDSC) which is based at the University of California, San Diego (UCSD), have been engaged in collecting and hosting Internet measurement data, and in developing analytic platforms enabling decision makers to better understand events related to security, stability, and resilience of critical communications infrastructure: episodes of disruption of Internet connectivity, Border Gateway Protocol (BGP) leaks and hijackings, denial of service attacks, scanning and other suspicious patterns of behavior. We curated and, as necessary, anonymized network data, and shared it with the vetted network and security researchers using the IMPACT portal and legal framework. We also prototyped an analytical online platform that fosters integration of disparate data sources, tools, services and expert knowledge with IMPACT's proven legal framework to support protected data sharing. By providing both data-preprocessing tools, as well as infrastructure for data sharing, analysis, and interactive visualization, this platform lowers the barrier for users wanting to provision/consume diverse sets of data, with the ultimate goal of creating a powerful marketplace of data, tools, and expertise.

The major challenges in our approach for Department of Homeland Security, Science and Technology Directorate (DHS S&T) Technical Topic Area (TTA)#1 were: development and maintenance of infrastructure to support sustainable collection, curation, and storage of large volumes of data, including data volume and quality, validation of data and inferences, general infrastructure scalability issues, and technical mechanisms for enabling privacy-respecting data sharing. To manage sustainability and reliability issues we embedded resiliency and redundancy into our system architecture, and distributed, fault-tolerant storage and analytics environments. To deal with storage scalability, we implemented expandable, distributed object-storage architecture, OpenStack SWIFT [23] system for many of our datasets. We adopted sophisticated pre-processing techniques to achieve data compression and minimal loss of data utility. We collaborated with the IMPACT legal team to refine the privacy-sensitive data-sharing framework that integrated proven disclosure control techniques to protect privacy without obliterating all utility in the data, with a policy approach that relies upon standard privacy principles and obligations of researchers and data providers.

The major technical challenges in our approach for TTA#2 related to the large integration efforts to combine various third-party open-source software frameworks with both our existing and newly developed components into a complex distributed platform. Another set of challenges arose in identifying and integrating an authentication and authorization framework that is reliable, configurable, and flexible enough to support both interactive user access to the platform, as well as programmatic access to the data and APIs.

2. INTRODUCTION

Large-scale Internet cyber-attacks and incidents – route hijacking, network outages, phishing campaigns, botnet activities, large-scale bug exploitation, etc. – represent a major threat to public safety and to both public and private strategic and financial assets. Mitigation and recovery, assessment of impacts and restoration costs, as well as prevention of further attacks of similar nature, are impeded by the fact that such events can remain unnoticed or hard to characterize, in terms of motivation, infrastructure used by the attacker, and scope. Because of their macroscopic nature, identifying such events and understanding their scope and dynamics requires three critical inputs: heterogeneous sources and types of data to cross-validate inferences; a system to enable close to real-time integration and interactive visualization of such data; and a team of experts with varied background and skills to soundly interpret fused data.

The Center for Applied Internet Data Analysis (CAIDA) at UC San Diego pursued these three inputs via participation in the IMPACT program across two TTAs: as a Data Provider (DP) - TTA#1, and as a Decision Analytics-as-a-Service Provider (DASP) - TTA#2.
As a Data Provider, we continued to provide data sets that had already proven relevant to security, stability, and resilience of networks. We also generated new data sets reflecting immediate threats, vulnerabilities, and hazards to the nation's critical communications infrastructures (e.g. detected outages, BGP hijacks, DoS attacks).
These datasets are now serving as inputs to HI-CUBE (*Hub for Internet Incidents Investigation)* the analytical platform that we developed for TTA#2, as well as available independently to researchers developing their own analytic capabilities.

As a Decision Analytics-as-a-Service Provider, we extended our existing system to support new analytic capabilities that integrate, correlate, and cross-validate multiple sources of measurement and meta-data to enable informed mitigation of and response to attacks and other disruptive events. Our system integration efforts resulted in a new platform – HI-CUBE (*Hub for Internet Incidents Investigation*) [1] – that allows experts to analyze diverse sets of live and historic macroscopic Internet data with interactive and visual tools in a trusted collaborative environment. HI-CUBE integrates four elements:

- web services and visual interfaces;
- data processing and analytics to support interactive querying and anomaly detection;
- new incident-based dataset creation and curation; and
- deployment and community outreach to support iterative improvements.

This platform offers a new model for enabling situational awareness of critical cyberinfrastructure, by flexibly supporting sharing and consumption of streamed and archived data sources targeting near-realtime and post-facto analysis of macroscopic cyber-security incidents. By providing both data-preprocessing tools, as well as the infrastructure for data sharing and analysis, HI-CUBE lowers the barrier for organizations to provision or consume data, and creates an extensible lab for testing cooperative analysis and fusion of diverse Internet data in the context of cyber-security analytics.

3. METHODS, ASSUMPTIONS, AND PROCEDURES

3.1 Supporting Cybersecurity Research through Network Data Collection and Curation (TTA#1)

A. Data Provider Tasks

As **Data Provider**, we collected, curated, and archived various Internet measurement data sets of relevance for cybersecurity research and development activities. To procure these data sets, CAIDA researchers used unique Internet measurement capabilities: distributed Archipelago measurement infrastructure tailored specifically for active probing experiments; and the UCSD Network Telescope that can observe manifestations of malicious activities in the Internet. We also resumed access to network monitors collecting samples of backbone Internet traffic and collected data as long as links were available – from March 2018 through January 2019.
We have almost completed development of a new system to capture bidirectional backbone traffic data on 100GB links, and laid foundation for testing and deployment.

The National Science Foundation (NSF) and Department of Homeland Security (DHS)-funded Archipelago (Ark) active measurement platform. Ark [2] is a platform designed, developed, and deployed by CAIDA for optimized, coordinated active network measurements. It is used for a variety of macroscopic Internet active measurement projects in support of empirical Internet research, including ongoing and ad-hoc measurements. Over 150 Ark monitors are deployed worldwide on six continents.
We measured IPv4 and IPv6 Internet topology from this Ark infrastructure, and mapped observed IP addresses to their DNS names in real-time. From this topology data we derived and heavily curated "Internet Topology Data Kits" [3], which included annotations of inferences relevant to analyzing the Internet as critical infrastructure, namely: router-level topology inferences, router-to-AS assignments, and geographic locations of each router. We also collected Prefix-Probing data, which consists of daily traceroutes to every announced BGP prefix from a subset of Ark monitors. Each monitor probes the entire set of targets independently and completes exactly one pass of the target set every calendar day (aligned on UTC boundaries). During the course of the project we added five new ITDK releases: 2017-08; 2018-03; 2019-01; 2019-04; 2020-01. We substantially improved ITDK production methodology utilizing traceroutes not only from our Archipelago measurement infrastructure but also from the RIPE Atlas measurement infrastructure [4].

The UCSD Network Telescope. The UCSD Network Telescope [5] consists of a large piece of globally announced IPv4 address space (/8 segment). This address space contains almost no legitimate hosts, so inbound traffic to non-existent machines is unsolicited, and anomalous in some way. Our network telescope contained approximately 1/256th of all public IPv4 addresses, so it receives roughly one out of every 256 packets sent by malicious software with an unbiased random number generator. (One-fourth of this address space was sold by its owner in July 2019, reducing the size of our "lens").
We collected pcap files (header and content) from the UCSD Network Telescope, instrumentation that monitors, strips the payload, and retains a sliding most recent one-month window of data on our machines, while archiving older data to an outside facility (NERSC).

In order to enable more efficient data storage, processing, and analysis, we post-processed these hourly pcap files using Corsaro software to extract the most important packet header fields and aggregate data into FlowTuple files. This is an ongoing dataset that started in October 2008 and is updated daily [6]. Another Telescope dataset that we indexed in IMPACT was *CAIDA UCSD Network Telescope Daily Randomly and Uniformly Spoofed Denial-of-Services* (RSDoS) *Attack Metadata* [7]. This dataset is derived from the unidirectional unsolicited IPv4 traffic reaching the UCSD Network Telescope. From the raw traffic data, we extract the backscatter (response) packets sent by victims of randomly and uniformly spoofed DoS attacks, summarize activity that relates to the same victim in an 'attack vector,' and produce a single CSV file of attack vectors per day.

Passive network monitors. From March 2018 till January 2019 we operated a passive network traffic monitor located at Equinix datacenter in New York, New York. That bidirectional link monitor consisted of a pair of physical machines. Both machines have a single Endace 6.2 DAG network monitoring card. A single DAG card is connected to a single direction of the bi-directional backbone link. Both machines had two Intel Dual-Core Xeon 3.00GHz CPUs, with 8 GB of memory and 1.3 TB of RAID5 data disk, running Linux 2.6.15 and DAG software version dag-2.5.7.1.2-.
Per our Research Agreement with the provider, after collecting a trace, we stripped the payload on the monitor before transferring files to CAIDA servers. Once on CAIDA servers, we compiled basic statistics, anonymized the raw traces using CryptoPAn prefix-preserving anonymization with the same key for all traces collected during a given calendar year, and packaged the data for release to vetted researchers. The resulting datasets [8] contain one-hour long samples of traffic collected monthly during a calendar year.
This dataset was our most popular source of data, but the link upgraded to 100GB monitors and our 10GB monitors aged out. In Year-2 of the project we worked on developing a new system to capture bidirectional backbone traffic data. We faced hardware availability and supply challenges due to Covid-19 closures that delayed the testing and deployment of the system. By the end of the project we have almost completed development of the system and laid foundation for testing and deployment.

B. Data Host Tasks

As a **Data Host**, we stored (including backups), managed, and served the available data to vetted security researchers. CAIDA personnel maintained numerous data-serving hardware platforms hosted in the machine room at the San Diego Supercomputer Center (SDSC). Our system administrators designed, configured and deployed these hosts to provide high availability for data collection, indexing, curation, and distribution. We expanded our hosting capabilities to support continuous growth of CAIDA datasets by deploying a new storage server (450 TB of total storage capacity). We are using the old data server for backup purposes.

One challenge inherent in Data Host activities was dealing with the huge data volumes. The data sets collected by CAIDA can grow to hundreds of terabytes, limiting the number of researchers who can use the data. The problem is especially acute for the UCSD Telescope data where during malicious activity outbreaks, data volumes can increase sharply, yet rapid analysis and

response are necessary. The speed, scope, and strength of today's automated malicious software require that relevant data are available in nearly real time.
As a Data Host, we adopted a distributed object storage platform Swift based on OpenStack for storing Telescope time series datasets. We migrated from Telescope data storage on local disk to a cloud-based system -- OpenStack SWIFT. All historical (starting 2008) Daily Randomly and Uniformly Spoofed Denial-of-Service *(*RSDoS) Attack Metadata, as well as historical and near-real-time aggregated Flow data are now stored in SWIFT. We also store at least one month of the most recently collected Near-Real-Time raw traffic traces data in SWIFT. Users can either download data from Swift or access these data via the native Swift API. We implemented the new VM-based analysis platform developed specifically for Telescope data users. Users can now log into a dedicated VM where they can access and process the Network Telescope data without contending for shared resources.

C. Generations of New Data Sets to Enable Effective Detection and Analysis of Incidents of Connectivity Disruptions

To support real-time network event identification and monitoring solutions to Cybersecurity Challenge Problems we generated new data sets that reflect immediate threats, vulnerabilities, and hazards to critical infrastructures, e.g., detected outages, BGP hijacks, DoS attacks, and other traffic anomalies, and meta-data to support analytics.
In addition to the two new ongoing UCSD Network Telescope datasets described above, other new data sets included: *AS Facilities Mapping* [9] -- the data set containing detailed information about individual interconnection facilities; *BGP Communities Dictionary* [10]– the data set containing geolocation information encoded by network operators into the Community attributes they set up for their networks; *Two years of RSDoS Attack Metadata* [11]– containing 'attack vectors' related to the same victim of randomly and uniformly spoofed DoS attacks as seen in the backscatter response.

We leveraged relevant data from other CAIDA research, including connectivity disruption logs; router-level and physical-level interconnection topology data; and crowd-sourced measurements of network vulnerabilities. These data sets now serve as inputs to the HI CUBE platform [1] (TTA#2), as well as being available independently to researchers developing their own analytic capabilities.

D. Contribution to IMPACT Team Activities and Project Development

We worked closely with IMPACT project team members to optimize IMPACT portal utility, convenience, and overall user experience. We participated in monthly project conference calls and in all PI meetings, updated IMPACT MOAs as needed to support new data; hosted project meetings, provided documentation, and outreach materials.
We cleaned and streamlined CAIDA publications catalogued in the IMPACT portal. We implemented procedures to provide access to all users whose requests were approved by IMPACT with the minimum delay.
We participated in marketing efforts, including meetings with US Cybercommand (May 2019), demo to the Internet Society (August 2019) and talks with them on providing data produced by

the Hijacking Observatory to support their Mutually Agreed Norms for Routing Security (MANRS) project. We presented IMPACT at multiple professional meetings, including the international workshop on Active Internet Measurements – Knowledge of Internet Structure: Measurement, Epistemology, and Technology (AIMS-KISMET) that we hosted on February 26-28th, 2020, at the San Diego Supercomputer Center, UCSD, La Jolla.

3.2 Developing HI-CUBE: Hub for Internet Incident Investigation (TTA#2)

Here we briefly summarize our approach to develop—as a **Decision Analytics-as-a-Service Provider**—new analytic capabilities that allow users to integrate, correlate, and cross-validate multiple sources of measurement and meta-data to enable informed mitigation of and response to attacks and other disruptive events. Our system integration efforts resulted in a new platform – HI-CUBE (*Hub for Internet Incidents Investigation*) [1] – that will allow experts to analyze diverse sets of live and historic macroscopic Internet data with interactive and visual tools in a trusted collaborative environment. Further details about our task execution and technical accomplishments are provided in Section 4.1.2.

A. Development of Web Services and Visual Interfaces

To prepare a web environment for collaborative investigation of incidents, we extended and refined the functionalities implemented in CAIDA's Charthouse [12]. We extended the authentication and authorization functionalities of the current Charthouse web application, including adding support for fine-grained data access control. We also developed a management interface for users, groups and shared data.

B. Design and Development of Software Infrastructure for Data Storage, Query, and Transformation

To upgrade the system to achieve an efficient, scalable, distributed infrastructure that allows flexible, complex and fast queries, we worked to replace our monolithic time series database (DBATS) with a distributed database for time-series analytics. We created a new analytics query engine based on Grafana [13], Telegraf [14], Apache Kafka [15] and InfluxDB [16], to which we added middleware to support automatic enrichment of time series with geographical and prefix-to-AS metadata.

B. Integration and Testing of HI-CUBE System in Operational Research Environments

To integrate the different components of the HI-CUBE system as well, as other data sources and deploy a prototype service, we acquired, deployed, and configured the hardware needed for hosting the HI-CUBE service, and migrated current databases, as well as integrated additional datasets. We also deployed and tuned the upgraded components of the infrastructure for big data analytics (query and analytics engine) based on Grafana [13], Telegraf [14], Apache Kafka [15], and InfluxDB [16].

C. Community Outreach and Service

We collected feedback presenting at Principle Investigator (PI) meetings and international workshops and we demoed our prototype to members of the community of cybersecurity researchers and analysts. We collaborated with other academic researchers and with the US Cybercommand.

4. RESULTS AND DISCUSSION

We successfully accomplished all the objectives of the project, met the milestones and produced all the deliverables.

4.1 Accomplishments Against Planned Objectives

4.1.1 TTA#1: Supporting Cybersecurity Research through Network Data Collection and Curation

Objective 1 (Data Provider): Collect, process and archive data to support cybersecurity research and development activities.

We collected, processed, and archived ~2.3 PB of new data -- adding more than 470,000 new files. By the end of the project we offered 23 datasets and three tools via IMPACT. Table 1 shows the current size of each group of datasets available via IMPACT at the end of report period. The Telescope data sizes are in Petabytes – 10^{15} bytes, while the Ark Topology and Traffic Trace data sizes are in Terabytes – 10^{12} bytes. The Telescope traffic data dominates the overall volume of data. Sizes of other data sets are negligible in comparison with these major categories.

Table 1: Sizes of CAIDA datasets indexed in the IMPACT portal at the end of the project

Data Collection	# of files	Size, compressed	Size, uncompressed
Ark IPv4 Routed /24	753555	12.9 TB	38.4 TB
Ark IPv4 Routed /24 DNS Names	4326	172.4 GB	648.7 GB
Ark IPv4 Prefix Probing	38805	4.0 TB	12.6 TB
Ark Internet Topology Data Kits	185	42.5 GB	312.8 GB
UCSD Network Telescope live data	720	62 TB	145 TB
UCSD Network Telescope archived data (at NERSC)	107583	2.9 PB	6.74 PB[4]
UCSD Network Telescope meta data	215067	332.5 TB	1.22 PB
UCSD Network Telescope Traffic Samples	76846	3.2TB	10.5TB
OC48 Internet Traces (anonymized)	220	79.9 GB	151.0 GB
Cumulative totals	**1197307**	**3.31 PB**	**8.17 PB**

Objective 2 (Data Host): Manage, maintain and share CAIDA data with vetted researchers

We provided access to our datasets to 161 users whose requests (161out of 324 requests) were pre-approved by IMPACT staff and approved by CAIDA. Users downloaded more than 16 TB of data during the report period. Table 2 shows number of data requests from IMPACT users and volumes of downloaded data. Note that the number of received requests is two times larger than the number of granted requests. Most of the rejections were due to the absence of a signed IMPACT restricted-data access Memorandum of Agreement (MOA).

Table 2: Number of data requests from IMPACT users and volumes of downloaded data during the report period

Dataset	Requests Received	Requests Granted[1]	Bytes Downloaded
Archival Telescope Data	59	49	375 GB
Near-real-time Telescope Data	47	17	N/A[2]
Network Telescope meta data	20	2	N/A[2]
RSDoS Attack Metadata	17	2	283 MB
Internet Topology with Ark	76	34	15.73 TB
DDoS	56	12	131.5 GB
Henya	3	3	N/A[3]
AS facilities Map	9	7	5.9 MB
Geolocated Router	16	16	3.9 MB
BGP Communities	12	12	10.6 MB
Border Mapping	3	1	0.37 MB
OC48 Traces	3	3	6.3 GB
AS rank	3	3	N/A[3]

[1] We granted all requests passed on to us by the IMPACT Coordinating Center (Blackfire).
[2] Near-Real Time Telescope data may be analyzed only on CAIDA machines and cannot be downloaded.
[3] CAIDA data analysis tool

Objective 3: Generate and index new datasets

By the end of the project, we offered 23 datasets and three tools via IMPACT. During the project period we indexed the following new datasets in IMPACT.

(1) **AS Facilities Mapping** (http://www.caida.org/data/as-facilities/) [9]
This dataset contains detailed information about individual interconnection facilities, including their names, operators, and geographic data; mapping of ASes to interconnection facilities; and types of peering interconnections used at those facilities. This data augments topological maps of AS connectivity with physical attributes and can be used to assess the resilience of interconnections in the event of natural disasters, identify facility or router outages, shed light on peering disputes, and help locate points of attacks, congestion, or instability on the Internet. Access is quasi-restricted https://www.impactcybertrust.org/dataset_view?idDataset=832

(2) **BGP Communities Dictionary** (http://www.caida.org/data/bgp-communities/) [10]
This BGP Community Dictionary Dataset focuses on Location-Encoding Ingress Communities which label the location where the prefix entered the network. The Dictionary represents our best effort to extract meaningful geolocation information encoded by network operators into the

Community attributes they set up for their networks. These data can be used as a source of meta-data to interpret/annotate other BGP data, including for inferring topological and geographic locations of disruptive events. Data is provided in a Tab Separated Values format. Each entry contains the BGP community, the geographic encoding, and the date when it was collected. Some records may also include additional (human-readable) location information, such as the city, facility, etc. The April 2018 release contained information about 7350 geolocated communities and 714 ASes in 88 countries. Access is quasi-restricted: https://www.impactcybertrust.org/dataset_view?idDataset=865

(3) **RSDoS Attack Metadata from 2015-2017** (http://www.caida.org/data/passive/rsdos-targets/) [11]

This dataset is aggregated from the unidirectional unsolicited IPv4 traffic reaching the UCSD Network Telescope from March 2015 to February 2017. From the raw traffic data, we extract the backscatter (response) packets sent by victims of randomly and uniformly spoofed DoS attacks, summarize activity that relates to the same victim in an 'attack vector,' and produce a single CSV file of attack vectors per day. The attack vector consists of a target IP address, statistical information about the attack, and geolocation and BGP routing metadata for that IP address. Possible uses of this data include: studying and modeling DoS attacks and characterizing victim populations.
Access is restricted: https://www.impactcybertrust.org/dataset_view?idDataset=915

(4) **CAIDA UCSD Network Telescope *Daily* RSDoS Attack Metadata [7]** (http://www.caida.org/data/passive/telescope-daily-rsdos.xml /)

This *ongoing* dataset is derived from the unidirectional unsolicited IPv4 traffic reaching the UCSD Network Telescope. From the raw traffic data, we extract the backscatter (response) packets sent by victims of randomly and uniformly spoofed DoS attacks, summarize activity that relates to the same victim in an attack vector, and produce a single CSV file of attack vectors per day. The attack vector consists of a target IP address, statistical information about the attack, and geolocation and BGP routing metadata for that IP address. Possible uses of this data include: studying and modeling DoS attacks and characterizing victim populations. This is an ongoing dataset that starts in October 2008 and is updated daily. This dataset contains coarse location information and is current data, therefore it must only be analyzed on CAIDA machines. https://www.impactcybertrust.org/dataset_view?idDataset=1159

(5) **CAIDA UCSD Network Telescope Aggregated Flow Datas**et [6] (http://www.caida.org/data/passive/telescope-flowtuple.xml)

This dataset consists of hourly files of unsolicited traffic captured by the UCSD Network Telescope traces and aggregated into the FlowTuple format. Raw data captured by the UCSD Network Telescope is stored in huge pcap files. In order to enable efficient storage, processing and analysis, we post-process raw data with our Corsaro tool to extract the most important packet header fields and aggregate data into FlowTuple files. This is an ongoing dataset that starts in October 2008 and is updated daily. This dataset contains IP addresses of the attacks victims, and therefore should only be analyzed on CAIDA machines.
https://www.impactcybertrust.org/dataset_view?idDataset=1160

(6) **Ark IPv4 Internet Topology Data Kits [3]**

(http://www.caida.org/data/internet-topology-data-kit/)
We added ITDK 2017-08, ITDK 2018-03, ITDK 2019-01, ITDK 2019-04, ITDK 2020-01 to our ongoing collection of this curated and aggregated data that started in 2010 and now includes 19 Kits. These ITDKs utilize traceroutes not only from our Archipelago measurement infrastructure but also from RIPE Atlas measurement infrastructure.
Access is restricted: https://www.impactcybertrust.org/dataset_view?idDataset=837

Objective 4: Contribute to IMPACT team activities and project development

CAIDA researchers have been active participants of the IMPACT project and made significant contributions to team activities aimed at improving the portal usability and user experience, as well as at outreach and marketing efforts. Selected highlights include:

- Improvements on the front end of CAIDA's data sharing infrastructure. As our experience with data sharing mechanisms grew, we improved data set representation and descriptions on CAIDA's web site, maintained a mailing list to keep the users of our data informed about changes and developments, and created a structure of responses to automate handling of user requests.
- Improvements on the back end of CAIDA's data sharing infrastructure. We maintained and grew our hardware infrastructure, adopted a distributed object storage platform Swift cluster based on OpenStack for storing Telescope time series datasets.
- Assisting developers with portal building and debugging. We made concerted efforts to work with Blackfire (IMPACT portal developer): participated in design discussions, tested and helped debug new deployments, made numerous suggestions aimed at improving portal operations, look and feel, and the overall user's experience.
- Co-organizing and hosting team meetings. We worked with IMPACT DHS Program Managers to organize project meetings and site visits at UCSD.
- Using portal data and procedures. CAIDA researchers not only contributed data to the portal, but also used data made available by other IMPACT team members to advance CAIDA's research programs. The benefits of these activities are two-fold: first, by applying for data as any outside user would do, we tested and evaluated usability factors, e.g., convenience and robustness, of portal operations. We proposed useful changes and modifications. Second, we demonstrated usefulness of IMPACT data offerings for cybersecurity research.
- Resource innovation. We worked with IMPACT team to measure and justify effectiveness of IMPACT resources and analyze the users' data and tools use and needs.
- Market innovation. We presented IMPACT at numerous professional meetings, participated in pitch development activities for potential industry and government entities.

4.1.2 TTA#2: Developing HI-CUBE: Hub for Internet Incident Investigation

Objective 2.1: Development of web services and visual interfaces.

Our objective in this task was to develop and deploy a web environment for collaborative investigation of incidents. We achieved this objective by first developing and deploying a proof of concept HI-CUBE [1] Web application based on the (previously funded by NSF and DHS) IODA project. We used this proof of concept to initially demo the HI-CUBE concept to other IMPACT performers, in order to receive early feedback and incorporate it in our development process. We then ported, upgraded and overhauled both the underlying ReactJS [17] web UI and Symfony (PHP) API applications [18] as part of building a HI-CUBE prototype, which we deployed at https://dev.hicube.caida.org. We integrated into the HI-CUBE prototype the interface of a BGP Observatory [19]: a system that monitors the Internet global routing plane 24/7 to detect malicious BGP activity and misconfiguration (https://dev.hicube.caida.org/feeds/hijacks/events).
We also developed a new experimental time-series visualization interface based on Grafana [13] and developed a process for rapidly deploying customized interfaces on a per data-provider basis. In parallel, we designed, developed and deployed an initial version of an authentication and authorization sub-system based on the OAuth open standard using the Globus [20] and Auth0 [21] *identity-as-a-service* providers. In 2020 we revised and extended this system, leveraging the Keycloak [22] open source framework to deploy an on-premises authentication and authorization service. This system also provides management interfaces for users, groups, and roles. We periodically demonstrated the new features and functionalities at PI meetings.

Objective 2.2: Design and development of software infrastructure for data storage, query, and transformation.

The main objective of this task was the design and development of HI-CUBE's efficient, scalable, distributed infrastructure for flexible, complex and fast large-data queries (as discussed in the "Capability Design Plan" and "Complete Design" deliverables).
A large part of our efforts in this task consisted in replacing our DBATS monolithic time series database (inherited from the IODA architecture) with an efficient distributed HTTP query server based on InfluxDB. To enable the transition of HI-CUBE to using a Big Data analytics query engine built on open source technology, including an InfluxDB backend, we developed a (dockerized) component responsible for automatically enriching time series with geographical and prefix-to-AS metadata (https://github.com/CAIDA/telegraf-friendlytagger). We then leveraged the open-source Grafana software framework to provide a flexible interface and query server that allows users to rapidly and interactively build queries across the various HI-CUBE data providers. We finally deployed, integrated, tested, and tuned the upgraded components of this infrastructure to enable large-scale data ingestion and analytics (the current deployment supports both the UCSD and Merit network telescope data sources).
In parallel, we designed and provisioned the HI-CUBE cloud storage system, centered around the OpenStack Swift [23] object storage. Throughout the project we continuously added (and shared) FlowTuple, BGP Hijacks Observatory, and OpenIntel DNS data [24]. At the time of writing, these three datasets consume more than 450 TB of storage. This infrastructure is used

both for programmatic access by other HI-CUBE components, as well as directly by HI-CUBE users—with access available both via the standard Swift and the widely used Amazon S3-compatible HTTP APIs. We also developed and deployed custom Swift middleware to provide IP-based ACLs, allowing us further control when handling sensitive datasets (https://github.com/CAIDA/hermes-iprange-acl). We performed a proof of concept analysis using the S3 compatible Swift API to process more than 5 TB of FlowTuple data using an Apache Spark cluster located at the University of Twente in The Netherlands (supporting another DHS S&T DDoS R&D effort). We also developed experimental middleware for Swift that dynamically removes sensitive fields from Avro data files depending on user permissions (https://github.com/CAIDA/hermes-avrofilter). Finally, we deployed a Globus endpoint on a virtual machine in our OpenStack cluster to provide HI-CUBE users access to data stored in Swift via the Globus data transfer platform.

Objective 2.3: Integration and testing of HI-CUBE system in operational research environments.

Our objective in this task was to integrate the different components and data sources of the HI-CUBE system. Executing this task required the purchase, configuration and deployment of various hardware components, including: *(i)* two SSD cluster machines for time series analytics that we added to our distributed time series collection platform (*TimeSeriesKafka*), *(ii)* a storage server for the OpenStack Swift storage cluster (including a new disk tray), and *(iii)* two web application servers.
We also deployed additional database services on the SSD server cluster as needed to support HI-CUBE data providers (e.g., an ElasticSearch cluster used by the BGP Hijacks Observatory).

As part of this task, we migrated existing time-series data feeds, such as data from both network telescope data providers (UCSD and Merit), into the new InfluxDB-based system. We then deployed, integrated, tested, and tuned the upgraded components of the distributed database system (including the query engine and the HTTP query server) based on InfluxDB to enable large-scale data ingestion and analytics (as described earlier when discussing technical achievements in Task 2.2).

We also invested significant effort performing the integration of HI-CUBE components and data sources. The majority of this integration work has already been discussed in the previous sections, but we also performed integration of two data sources represented by the CAIDA Network Telescope and BGP Observatory projects. We deployed and integrated our latest iteration of the real-time traffic analysis component of the STARDUST platform, which continuously captures and processes traffic reaching the UCSD network telescope. This upgrade reduces the effective latency of the telescope data available via HI-CUBE from several hours to approximately 1 minute. In parallel, in Year 2 of the project, we completed integration of the BGP Observatory data source, which included: designing and testing a new inference engine, debugging the event tagging sub-system, porting the Observatory UI to the HI-CUBE Web application framework, developing and deploying an ElasticSearch-centered pipeline, and updating the user interface.

As described in the previous section (Task 2.1), we deployed a proof of concept (alpha version) of the HI-CUBE web site in Year 1 of the project and later in Year 2 we deployed a prototype (beta) version accessible at https://dev.hicube.caida.org.

Objective 2.4: Community outreach and service.

We presented the HI-CUBE concept and demoed the first proof of concept and the subsequent prototype at the IMPACT PI meetings (2018, 2019), as well as the DHS S&T Cybersecurity and Innovation Showcase (2019). In addition, we submitted an abstract (on the topic of BGP Hijacking detection and mitigation) for consideration as a talk at the CISA Cybersecurity Summit (2019). We also presented status of the HI-CUBE project at the IMAPS 2018, WIE 2018 international workshops that we organized at UC San Diego.

In May 2019, we discussed the functionalities and methods of the BGP Hijacking Observatory in a two-day meeting with the US Cybercommand at the DreamPort facilities in Maryland (dreamport.tech) and collected feedback. In August 2019, we demoed and discussed the BGP Hijacking Observatory in a call with the Internet Society (ISOC) and discussed use by operators (including the possibility of a pilot evaluation of a preliminary deployment facilitated by ISOC), as well as by ISOC's observatory.

On September 9th-10th, 2019, we hosted the 2nd International Workshop on Darkspace and UnSolicited Traffic Analysis (DUST 2019) at the San Diego Supercomputer Center, UCSD, San Diego, California. The goal of the DUST workshop series is to bring together researchers, operators, and analysts interested in unsolicited traffic analysis, especially traffic destined to unassigned (dark) IP address space. During the workshop, we gave a demo of the HI-CUBE system and discussed potential collaborations and opportunities for data integration with the participants, who included industry and academia.

Finally, we assisted researchers from the University of Twente in The Netherlands in performing a proof of concept analysis of FlowTuple data stored on our Swift cluster from an Apache Spark cluster located at their university. We also created a Swift account for the group of Prof. Paul Barford, University of Wisconsin, Madison, to start integrating their NTP-based post-processed outage data into HI-CUBE. To support this project, we wrote and provided documentation on how to upload their data into HI-CUBE's cloud storage system.

Deliverables

Deliverable	Type	Date
TTA#1 Capability Design Plan	Report	Jan 2018
TTA#2 Capability Design Plan	Report	Jan 2018
TTA#1 Capability Complete Design Plan	Report	Feb 2018
TTA#2 Capability Complete Design Plan	Report	Feb 2018
Presentations at PI Meetings	Presentation	May 2018; Mar & Aug 2019; Mar 2020
Tri-annual Technical Status Reports	Report	May, Sep, 2018; Jan, May, Sep 2019 Jan, May, Sep 2020
Open source software package	Software	Sep 2020

4.3 Technology Transition and Transfer

We released as open-source the components of the HI-CUBE system:

- https://github.com/CAIDA/charthouse-ui — HI-CUBE Web Application
- https://github.com/CAIDA/charthouse-api — HI-CUBE HTTP API
- https://github.com/CAIDA/libtimeseries — Time-series abstraction library
- https://github.com/CAIDA/libipmeta — Library to support the execution of historical and realtime IP metadata lookups using commercial IP geolocation databases (NetAcuity, Maxmind) and CAIDA Prefix-To-AS databases.
- https://github.com/CAIDA/bgpstream — Framework for historical and realtime analysis of BGP data
- https://github.com/CAIDA/bgpview — Library for efficient reconstruction, transport and analysis of BGP routing tables and the identification of potential prefix hijacking events
- https://github.com/CAIDA/corsaro3 — Software suite for performing large-scale analysis of trace data
- https://github.com/CAIDA/telegraf-friendlytagger — A dockerized component responsible for automatically enriching time series with geographical and prefix-to-AS metadata
- https://github.com/CAIDA/hermes-iprange-acl — Swift middleware to provide IP-based ACLs
- https://github.com/CAIDA/hermes-avrofilter — Swift middleware to dynamically remove sensitive fields from Avro data files depending on user permissions.

5. CONCLUSIONS

The Department of Homeland Security (DHS) Science and Technology (S&T) Directorate supported this unique and ambitious data sharing program for more than 20 years and enabled substantial progress in terms of improving data sharing architectures, analysis, legal support, and transparency of processes. Much of the research and infrastructure enabled by the program would not exist otherwise.

Metrics for evaluation of the program have been a long-standing challenge, as the disincentives to share data in the first place still dwarf the effects of any other actions taken to expand use of the platform. We believe additional incentives to share data will require regulation, which we also believe is inevitable but must be undertaken carefully and with specific analysis questions driving the data sharing requirements. When the U.S. government gets to that point, we believe there will be many lessons to draw on from this project. We have been honored to work with DHS on the project and hope for further engagement as CISA expands its scope of responsibility in this area.

6. REFERENCES

[1] CAIDA UCSD: Hub for Internet Incidents Investigation, https://dev.hicube.caida.org
[2] CAIDA UCSD: Archipelago (Ark) Measurement Infrastructure, https://www.caida.org/projects/ark/
[3] CAIDA UCSD: The CAIDA UCSD Ark IPv4 Internet Topology Data Kits, https://www.caida.org/data/internet-topology-data-kits, DOI: 10.23721/107/1423888
[4] Ripe Network Coordination Center, RIPE Atlas, https://atlas.ripe.net/
[5] CAIDA: The UCSD Network Telescope, https://www.caida.org/projects/network_telescope/
[6] CAIDA UCSD: UCSD Network Telescope Aggregated Flow Dataset, https://www.caida.org/data/passive/telescope-flowtuple.xml, DOI: 10.23721/107/1503377
[7] CAIDA UCSD: UCSD Network Telescope Daily Randomly and Uniformly Spoofed Denial-of-Service (RSDoS) Attack Metadata - https://www.caida.org/data/passive/telescope-daily-rsdos.xml, DOI: 10.23721/107/1503376
[8] CAIDA UCSD: The CAIDA UCSD Anonymized Internet Traces, https://www.caida.org/data/passive/passive_dataset.xml
[9] CAIDA UCSD: The CAIDA UCSD AS Facilities Mapping, April 2017, https://www.caida.org/data/as-facilities/, DOI: 10.23721/107/1421848
[10] CAIDA UCSD: The CAIDA UCSD BGP Community Dictionary, https://www.caida.org/data/bgp-communities/, DOI: 10.23721/107/1435354
[11] CAIDA UCSD: Randomly and Uniformly Spoofed Denial-of-Service (RSDoS) Attack Metadata from 2015-2017, https://www.caida.org/data/passive/rsdos-targets/, DOI: 10.23721/107/1463169
[12] CAIDA UCSD Charthouse, https://charthouse.caida.org/
[13] Grafana, https://grafana.com/
[14] InfluxData: Telegraf, the open source server agent https://www.influxdata.com/time-series-platform/telegraf/
[15] Apache Kafka, https://kafka.apache.org/
[16] InfluxData: Influx DB https://www.influxdata.com/products/influxdb-overview/
[17] React JavaScript library, https://reactjs.org/
[18] Sympfony, https://symfony.com/doc/current/bundles/NelmioApiDocBundle/index.html
[19] CAIDA UCSD: BGP Hijacking Observatory, https://bgp.caida.org/
[20] Globus, https://www.globus.org/
[21] Auth0 Authentification Platform, https://auth0.com/
[22] Keyclock Open Source Identity and Access Management, https://www.keycloak.org/
[23] OpenStack SWIFT, https://docs.openstack.org/swift
[24] OpenIntel DNS data, https://openintel.nl/

APPENDIX A – PUBLICATIONS AND PRESENTATIONS

Publications

All of the following publications are also available on **https://www.caida.org/outreach/publications/papers** .

(1) https://arxiv.org/abs/1801.01085v2
Title: ARTEMIS: Neutralizing BGP Hijacking within a Minute (technical report)
Authors: Sermpezis, P., Kotronis, V., Gigis, P., Dimitropoulos, X., Cicalese, D., King, A., Dainotti, A. Date: Jan 2018
Venue: arXiv
Keywords: passive data analysis, routing, security, software/tools, topology

(2) http://www.caida.org/publications/papers/2018/survey_among_network_operators
Title: A Survey among Network Operators on BGP Prefix Hijacking
Authors: Sermpezis, P., Kotronis, V., Dainotti, A., Dimitropoulos, X. Date: Jan 2018
Venue: ACM SIGCOMM Computer Communication Review (CCR), v. 48 (1), pp. 64-69
Keywords: security, routing, passive data analysis, topology

(3) https://dl.acm.org/citation.cfm?doid=3234200.3234209
Title: On the Potential of BGP Flowspec for DDoS Mitigation at Two Sources: ISP and IXP
Authors: N. Hinze, M. Nawrocki, M. Jonker, A. Dainotti, T. Schmidt, and M. Wählisch
Date: August 2018
Venue: Proceeding of the ACM SIGCOMM 2018 Conference on Posters and Demos, pp. 57-59
Keywords: internet outages, measurement methodology, network telescope, security software/tools, topology

(4) http://www.caida.org/publications/papers/2018/inferring_carrier_grade_nat/
Title: Inferring Carrier-Grade NAT Deployment in the Wild
Authors: I. Livadariu, K. Benson, A. Elmokashfi, A. Dainotti, and A. Dhamdhere
Date: April 2018
Venue: INFOCOM
Keywords: IPv6, measurement methodology

(5) https://dl.acm.org/citation.cfm?id=3278571
Title: A First Joint Look at DoS Attacks and BGP Blackholing in the Wild
Authors: Jonker, Mattijs, Aiko Pras, Alberto Dainotti, and Anna Sperotto
Date: October 2018
Venue: Internet Measurement Conference
Keywords: Denial-of-Service; DDoS Mitigation; BGP; Blackholing

(6) https://arxiv.org/abs/1906.04426
Title: Chocolatine: Outage Detection for Internet Background Radiation
Authors: A. Guillot, R. Fontugne, P. Winter, P. Mérindol, A. King, A. Dainotti, and C. Pelsser
Date: June 2019
Venue: Proceedings of Network Traffic Measurement and Analysis Conference (TMA)
Keywords: internet outages, passive data analysis, security, network telescope

(7) https://tma.roc.cnam.fr/Proceedings/TMA_Paper_2.pdf
Title: Geo-Locating BGP prefixes
Authors: P. Winter, R. Padmanabhan, A. King, and A. Dainotti
Date: June 2019

Venue: Proceedings of Network Traffic Measurement and Analysis Conference (TMA)
Keywords: internet outages, measurement methodology, routing

(8) https://tma.roc.cnam.fr/Proceedings/TMA_Paper_4.pdf
Title: BGP hijacking classification
Authors: S. Cho, R. Fontugne, K. Cho, A. Dainotti, and P. Gill
Date: June 2019
Venue: Proceedings of Network Traffic Measurement and Analysis Conference (TMA)
Keywords: routing, security, topology

(9) https://ccronline.sigcomm.org/2019/ccr-july-2019/towards-passive-analysis-of-anycast-in-global-routing-unintended-impact-of-remote-peering/
Title: Towards Passive Analysis of Anycast in Global Routing: Unintended Impact of Remote Peering
Authors: R. Bian, S. Hao, H. Wang, A. Dhamdhere, A. Dainotti, and C. Cotton
Date: July 2019
Venue: ACM SIGCOMM Computer Communication Review, *49(3), p18*
Keywords: routing, dns, topology

(10)https://journals.sagepub.com/toc/jcrb/0/0
Title: At Home and Abroad: The Use of Denial-of-service Attacks during Elections in Nondemocratic Regimes.
Authors: P. Lutscher, N. Weidmann, M. Roberts, M. Jonker, A. King, and A. Dainotti
Date: July 2019
Venue: Journal of Conflict Resolution, DOI: 0022002719861676
Keywords: passive data analysis, security, network telescope

(11)http://www.caida.org/publications/papers/2019/toward_theory_harms_internet/
Title: Toward a Theory of Harms in the Internet Ecosystem.
Authors: D. Clark and k. claffy
Date: August 2019
Venue: CAIDA
Keywords: overview, policy

(12)http://www.caida.org/publications/papers/2019/residential_links_under_weather/
Title: Residential Links Under the Weather
Authors: R. Padmanabhan, A. Schulman, D. Levin, N. Spring
Date: August 2019
Venue: ACM SIGCOMM Conference Proceedings
Keywords: active data analysis, measurement methodology, routing

(13)https://dl.acm.org/doi/abs/10.1145/3355369.3355581
Title: Profiling BGP Serial Hijackers: Capturing Persistent Misbehavior in the Global Routing Table
Authors: Testart, C., Richter, P., King, A., Dainotti, A., & Clark, D.
Date: October 2019
Venue: Proceedings of the Internet Measurement Conference (IMC'19)
Keywords: routing, security, topology

(14)https://dl.acm.org/doi/10.1145/3355369.3355589
Title: Learning Regexes to Extract Router Names from Hostnames
Authors: Luckie, M., Huffaker, B., & claffy, K.
Date: October 2019

Venue: Proceedings of the Internet Measurement Conference (IMC'19)
Keywords: measurement methodology, software/tools, topology

(15)https://link.springer.com/chapter/10.1007%2F978-3-030-44081-7_5
Title: To Filter or Not to Filter: Measuring the Benefits of Registering in the RPKI Today
Authors: Testart, C., Richter, P., King, A., Dainotti, A., Clark, D.
Date: March 2020
Venue: Proceedings of the Passive and Active Measurement. PAM 2020
Keywords: Internet security, Routing, RPKI, BGP

(16)https://doi.org/10.1007/978-3-030-44081-7_13
Title: Unintended Consequences: Effects of Submarine Cable Deployment on Internet Routing.
Authors: Fanou R., Huffaker B., Mok R., Claffy K.C.
Date: March 2020
Venue: Proceedings of the Passive and Active Measurement. PAM 2020
Keywords: measurement methodology, BGP, topology

(17)https://dl.acm.org/doi/abs/10.1145/3402413.3402415
Title: RIPE IPmap active geolocation: mechanism and performance evaluation
Authors: Du, B., Candela, M., Huffaker B., Snoeren, A. C., claffy K.
Date: April 2020
Venue: ACM SIGCOMM Computer Communication Review 50.2 (2020): 3-10
Keywords: measurement methodology, routing, topology

(18)https://dl.acm.org/doi/10.1145/3392158
Title: vrfinder: Finding Outbound Addresses in Traceroute
Authors: Marder, A., Luckie, M., Huffaker, B., Claffy, KC.
Date: June 2020
Venue: SIGMETRICS/Performance Joint International Conference on Measurement and Modeling of Computer Systems, 2020
Keywords: Measurement methodology, routing, topology

(19)https://www.caida.org/publications/papers/2020/forgotten_side_dns/
Title: The Forgotten Side of DNS: Orphan and Abandoned Records.
Authors: Sommese, R., Jonker, M., van Rijswijk-Deij, R., Dainotti, A., & Sperotto, A
Date: June 2020
Venue: Workshop on Traffic Measurements for Cybersecurity, 2020
Keywords: dns, topology

(20)https://ieeexplore.ieee.org/document/9133283
Title: The DNS in IoT: Opportunities, Risks, and Challenges
Authors: Hesselman, C., Kaeo, M., Chapin, L., Claffy, K., Seiden, M., McPherson, D., .& Rasmussen, R.
Date: July 2020
Venue: IEEE Internet Computing, 2020
Keywords: dns, security

(21)https://www.jstor.org/stable/10.5325/jinfopoli.10.2020.0001
Title: Policy Challenges in Mapping Internet Interdomain Congestion.
Authors: claffy, KC., Clark, D. D., Bauer, S., & Dhamdhere, A
Date: July 2020
Venue: Journal of Information Policy, 10, pp. 1-44
Keywords: routing, topology

(22)https://www.sciencedirect.com/science/article/abs/pii/S1389128620311336
Title: Spoofed traffic inference at IXPs: Challenges, methods and analysis..
Authors: Müller, L., Luckie, M., Huffaker, B., claffy, KC & Barcellos, M
Date: December 2020
Venue: Computer Networks, p.107452
Keywords: routing, active analysis

Meetings and Presentations

1. *Meeting Name*: Workshop on Active Internet Measurements (AIMS)
Purpose: To promote discussion between academics, industry, policymakers, and funding agencies on active Internet measurement, exchange of research ideas and questions that have been answered, or could be answered, with proposed measurement infrastructures
Dates: March 13 - 15, 2018
Location: UC San Diego, La Jolla, CA
Attendees from the project: kc claffy, A. Dainotti, and M. Fomenkov, A. King
Presentations: A. Dainotti, ARTEMIS: "Neutralizing BGP Hijacking within a Minute"
http://www.caida.org/publications/presentations/2018/artemis_aims/

2. *Meeting Name:* DHS IMPACT PI Meeting
Purpose: Project status and progress reports, discussions between IMPACT PIs
Dates: May 7 - 9, 2018
Location: Boston, MA
Meeting Attendees: kc claffy, A. Dainotti, and M. Fomenkov
Presentations Made: kc claffy, " DHS IMPACT Project: CAIDA Update";
http://www.caida.org/publications/presentations/2018/impact_pi_may/
Dainotti, A "HICUBE: Hub for Internet Incident Investigation"
http://www.caida.org/publications/presentations/2018/hicube_impact_pi_may/

3. *Meeting Name:* CAIDA IMAPS Workshop
Purpose: Workshop on interdisciplinary research of political events that are visible through analysis of Internet measurement data
Dates: September 5 - 7, 2018
Location: La Jolla, CA
Meeting Attendees: kc claffy, A. Dainotti, A. King and M. Fomenkov
Presentations Made: A. King, "Realtime detection and analysis of Denial of Service attacks"

4. *Meeting Name:* DHS IMPACT PI Meeting
Purpose: To discuss Infrastructure, process and resources effectiveness and innovation.
Dates: March 4, 2019
Location: Menlo Park, CA
Meeting Attendees: kc claffy, A. Dainotti, A. King, E. Yulaeva
Presentations Made: A. Dainotti, *"*HI-CUBE: Hub for Internet Incident Investigation";
kc claffy, "DHS IMPACT Project: CAIDA Update"

5. *Meeting Name:* DHS 2019 S&T Cybersecurity and Innovation Showcase
Purpose: To introduce DHS-funded research projects to government, industry technology implementers, investors, angel funders, and other potential market transition partners.
Dates: March 18 - 20, 2019
Location: Washington DC
Meeting Attendees: kc claffy, A. Dainotti, Alistair King
Presentations Made: A. Dainotti, "HI-CUBE: Hub for Internet Incident Investigation";
kc claffy, "Advancing Scientific Study of Internet Security and Topological Stability (ASSISTS DP)"

6. *Meeting Name:* DHS IMPACT PI Meeting
Purpose: To discuss Infrastructure, process and resources effectiveness and innovation.
Dates: August 28-29, 2019
Location: SRI International, Washington DC
Meeting Attendees: kc claffy (remotely), A. Dainotti, E. Yulaeva
Presentations Made: A. Dainotti, "CAIDA's BGP (HIJACKING) Observatory";
kc claffy, "DHS IMPACT Project: CAIDA Update (TTA-1)"

7. *Meeting Name:* Visit at DreamPort / Collaboration on BGP hijacking with Geiger-Gremlin team at Cyber National Mission Force
Purpose: To discuss methods and infrastructure for BGP hijacking detection and mitigation.
Dates: May 2-3, 2019
Location: DreamPort, Columbia MD
Meeting Attendees: A. Dainotti
Presentations Made: A. Dainotti, demos and excerpts from "CAIDA's BGP Observatory", "Neutralizing BGP Hijacking with the ARTEMIS Open-source Tool", "Extracting Cybersecurity Metrics from Internet Mapping and Monitoring"

8. *Meeting Name:* CAIDA's BGP Hijacking Observatory
Purpose: To discuss with and demo to the ISOC MANRS project lead our current work to develop and deploy CAIDA's BGP Hijacking Observatory.
Dates: August 8, 2019
Location: Conference Call
Meeting Attendees: A. Dainotti
Presentations Made: A. Dainotti, demo and excerpts from "CAIDA's BGP Observatory"

9. *Meeting Name:* Workshop on Darkspace and UnSolicited Traffic Analysis (DUST)
Purpose: To bring together researchers, operators, and analysts interested in unsolicited traffic analysis, especially traffic destined to unassigned (dark) IP address space.
Dates: September 9-10, 2019
Location: SDSC, UCSD La Jolla, CA
Meeting Attendees: kc claffy, A. Dainotti, E. Yulaeva, A. King
Presentations Made: A. Dainotti,"HI-CUBE: Hub for Internet Incident Investigation";
A. King, "STARDUST – Sustainable Tools for Analysis and Research on Darknet UnSolocited Traffic"

10. *Meeting Name:* ACM Internet Measurement Conference 2019
Purpose: IMC is a three-day event focusing on Internet measurement and analysis.
Dates: October 21-23, 2019
Location: KIT Royal Tropical Institute, Amsterdam, Netherlands
Meeting Attendees: A. Dainotti, A. King
Presentations Made: C. Testart, P Richter, A. King, A. Dainotti, D. Clark "Profiling BGP Serial Hijackers: Capturing Persistent Misbehavior in the Global Routing Table"
M. Luckie, B. Huffaker, K. claffy "Learning to Extract Router Names from Hostnames"

11. *Meeting Name:* Workshop on Active Internet Measurements: Knowledge of Internet Structure: Measurement, Epistemology, and Technology (AIMS-KISMET)
Purpose: To bring together researchers, operators, and analysts to discuss possible scenarios for the future of the Internet.
Dates: February 26-28, 2020
Location: SDSC, UCSD La Jolla, CA
Meeting Attendees: kc claffy, A. Dainotti, E. Yulaeva, A. King, J. Polterock
Presentations Made: A. Dainotti,"*CAIDA's BGP (Hijacking) Observatory";*
kc claffy, "*CAIDA, and DHS IMPACT Program"*

12. Meeting Name: DHS IMPACT PI Meeting
Purpose: Future of IMPACT
Dates: March 23, 2020
Location: online
Meeting Attendees: A. Dainotti, kc claffy, E. Yulaeva, J. Polterock
Presentations Made: kc claffy "DHS IMPACT Project: CAIDA Update"
A. Dainotti "CAIDA's BGP (Hijacking) Observatory"

13. Meeting Name: Passive and Active Measurement Conference (PAM) 2020
Purpose: To bring together researchers and operators to discuss novel and emerging work in the area of network measurement and analysis.
Dates: March 30-31, 2020
Location: Online
Meeting Attendees: A. Dainotti,
*Presentations Made:*C. Testart, P. Richter, A. King, A. Dainotti, D. Clark "To Filter or not to Filter: Measuring the Benefits of Registering in the RPKI Today"

LIST OF ACRONYMS

ACL	Access Control List
AIMS-KISMET	Active Internet Measurements: Knowledge of Internet Structure: Measurement, Epistemology, and Technology
AS	Autonomous Systems
API	application programming interface
ARK	Archipelago
BGP	Border Gateway Protocol
CAIDA	Center for Applied Internet Data Analysis
CISA	Cybersecurity & Infrastructure Security Agency
CPU	Central Processing Unit
CSV	Comma-Separated Values
DAG	Directed Acyclic Graph
DASP	Decision Analytics-as-a-Service Provider
DBATS	Database of Aggregated Time Series
DDoS	Distributed Denial of Service
DHS	Department of Homeland Security
DNS	Domain Name System
DoS	Denial-of-Services
DP	Data Provider
DUST	Darkspace and UnSolicited Traffic Analysis
GB	Gigabytes
GHz	Gigahertz
HI-CUBE	Hub for Internet Incidents Investigation
HTTP	Hypertext Transfer Protocol
IMAPS	Internet Measurement And Political Science
IMPACT	Information Marketplace for Policy and Analysis of Cyber-risk & Trust
IODA	Internet Outage Detection and Analysis
IPv4	Internet Protocol version 4
IPv6	Internet Protocol version 6
ITDK	**Internet Topology Data Kit**
MANRS	Mutually Agreed Norms for Routing Security
MOA	Memorandum of Agreement
NERSC	National Energy Research Scientific Computing Center
NSF	National Science Foundation
NTP	Network Time Protocol
PB	Petabyte
PI	Primary Investigator
R&D	Research & Development
RAID	Redundant Array of Independent Disks
RSDoS	Randomly and Uniformly Spoofed Denial-of-Services
S&T	Science & Technology
SDSC	San Diego Supercomputer Center
SSD	Solid State Drives

STARDUST	Sustainable Tools for Analysis and Research on Darknet
UnSolicited Traffic	
TB	Terabytes
TTA	Technical Topic Area
UCSD	University of California, San Diego
UI	User Interface
US/U.S.	United States
UTC	Coordinated Universal Time
VM	Virtual Machine
WIE	Workshop on Internet Economics