Project Status Report

Reporting period: 04/01/2023 - 09/30/2023
Project title:
Mid-Scale RI-1 Design Project (M1:DP):
Designing a Global Measurement Infrastructure to Improve Internet Security (GMI3S)
OAC-2131987

Principal Investigator: kc Claffy, Bradley Huffaker (UCSD), David Clark (MIT)
Project Manager: Elena Yulaeva
Lead Institution: CAIDA, UCSD
Other Institutions: NSRC (U Oregon), MIT
Cognizant PO:

Table of Contents

1. Summary of project status

A brief summary of the project's overall status on technical progress, cost and performance.

Award Duration	Start date: 10/01/2021	Planned close out: **09/30/2025***
Project Finish Date	Planned Early Finish:	Estimated Early Finish:
Project %-complete	50%	

2. Near-Term Milestones

Include milestones with the scheduled dates or actual/forecast dates that are in current and the next reporting period, and milestones (with past scheduled dates) that are delayed to future reporting period. (**Completed deliverables have bold font dates**.)

WBS	Subsystem	Milestone	Scheduled Date	Actual date (A) /Forecast Date (F)
1.1	1.1.3	**Monitors Requirements documented, hardware and software needs assessed**	9/30/2023	**9/30/2023(A)**
	1.1.3.3	**Telescope data monitoring needs compiled**	05/31/2023	**05/31/2023(A)**
	1.1.3.4	**Two-way traffic data monitoring needs compiled**	05/31/2023	**05/31/2023(A)**
	1.1.3.5	**BGP data monitoring needs compiled**	09/30/2023	**09/30/2023(A)**
	1.1.3.6	**Active measurements needs compiled**	03/31/2023	**05/31/2023(A)**
	1.1.3.7	DNS monitoring needs compiled	04/30/2023	03/31/2024(F)
	1.1.4.1	**Monitor specification report draft posted for internal feedback**	03/31/2023	**08/15/2023(A)**
	1.1.4.2	Draft (2) for community feedback	05/31/2023	10/30/2023(F)
	1.1.4.3	Post final report for public comments	09/30/2023	03/31/2024(F)
	1.1.5	Monitor software Prototyped	03/31/2024	03/31/2024(F)
	1.1.5.5	DNS data monitoring software prototyped	03/31/2024	03/31/2024(F)
	1.1.6	Monitor deployment prototyped	03/31/2024	03/31/2024(F)
	1.1.6.2	**Telescope data monitoring deployment**	07/01/2023	**07/01/2023(A)**
	1.1.6.3	Two-way traffic monitor deployment	1/31/2023	12/31/2023(F)
	1.1.6.6	**DNS data monitoring deployment**	05/31/2023	**05/31/2023(A)**
	1.1.6.8	**10 nodes of multiple measurements deployed**	09/30/2023	**09/30/2023(A)**
	1.1.6.9	Additional 10 nodes of multiple measurements deployed	03/31/2024	03/31/2024(F)
	1.1.7.2	Third-party experiment deployed	01/31/2024	01/31/2024(F)
	1.1.7.3	Evaluation report of Data Acquisition Component created	03/31/2024	03/31/2024(F)
	1.1.9	**Integrated report "Internet infrastructure security vulnerabilities" and "Data Needs" published**	2/28/2023	**5/01/2023 (A)**

1.2	**1.2.1.1**	**Data storage hardware requirement documented, draft posted for stakeholders' review**	3/31/2023	**9/30/2023(A)**
	1.2.1.2	Community feedback incorporated into the data storage requirements document	3/31/2023	11/30/2023(F)
	1.2.2	Data storage systems specifications published	7/01/2023	3/31/2024(F)
	1.2.2.2	Data storage systems specifications documented, **draft posted for stakeholders' review**	2/28/2023	01/01/2024(F)
	1.2.2.3	Community feedback incorporated into document	07/01/2023	02/28/2024(F)
	1.2.3.4-2	**YR2 Data and metadata standards specifications published**	9/30/2023	**09/30/2023(A)**
	1.2.4.2	**Report on pcap real-time analysis tools**	7/01/2023	**7/01/2023(A)**
	1.2.4.3	Specification of tools for data curation and documentation, report created, annually updated	03/31/2024	03/31/2024(F)
	1.2.5.2.1	Incorporate datasets used by IYP	7/01/2023	03/31/2024(F)
	1.2.5.4.2	**Update of data and metadata API**	09/30/2023	**09/30/2023(A)**
	1.2.8	**Approaches to data dissemination designed, documented**	09/30/2023	**09/30/2023(A)**
	1.2.8.1	**Report on the latest big data storage/management technologies**	3/31/2023	**08/31/2023(A)**
1.3	1.3.1	Data discovery tools prototyped	07/01/2023	**09/30/2023(A)**
	1.3.1.4	**Report on other non-CAIDA datasets and tools integration**	05/31/2023	**5/31/2023(A)**
	1.3.1.5	Report on integration of automated meta-data/data citation creation into catalog.	07/01/2023	03/31/2024(F)
	1.3.2	Software for disclosure control developed	03/31/2024	03/31/2024(F)
	1.3.2.2	Workshop (disclosure control) conducted	03/31/2024	03/31/2024(F)
	1.3.2.5	At least two practices to enable legit research access to various data types prototyped and evaluated	03/31/2024	03/31/2024
	1.3.2.1	**Report on the gaps between privacy-preservation techniques and network and**	12/31/2022	**07/31/2023(A)**

		security research needs		
	1.3.2.3	**Gaps that privacy techniques can support identified, report created and shared**	1/31/2023	**09/30/2023(A)**
	1.3.2.4	**Create taxonomy of data**	12/31/2022	**09/30/2023(A)**
	1.3.3	Report on Policy tools	03/31/2024	03/31/2024
	1.3.3.6	Report on other countries' approaches to policy tools for disclosure control	3/31/2023	**09/30/2023(A)**
	1.3.3.7	Agreements with at least two commercial data providers put in place (DomainTools, Iconectiv, IPinfo)	7/31/2023	**07/31/2023(A)**
	1.3.4	Case studies on Extensibility		
	1.3.4.2	State of Internet report created and shared (1.3.4.2 Conduct meetings and compile data to create community-authored "State of the Internet report")	9/30/2023	9/30/2024 (F)
	1.3.4.3	"State of the DDoS attacks" report	9/30/2023	12/31/2023 (F)
	1.3.4.5	Appropriate external datasets and tools, and organizations to include in the extensibility case studies identified, documented	7/31/2023	12/31/2023 (F)
	1.3.4.6	Case studies on extensibility of policy framework conducted, documented and shared	03/31/2024	03/31/2024
1.4	**1.4.3.2**	**Online course on NIDS developed**	09/30/2023	**04/30/2023(A)**
	1.4.3.3	Video tutorials on nodes deployment and management created	3/31/2023	02/28/2024(F)
	1.4.3.4	**Quarterly calls conducted, minutes shared**	09/30/2023	**09/30/2023(A)**
	1.4.3.5	**Project Presentations**	09/30/2023	**09/30/2023(A)**

3. Technical progress highlights

Highlights of progress in this period, by near-term tasks, including the following select achievements:

- Published the combined version of "Internet infrastructure security vulnerabilities" (Milestone 1.1.1) and "Data Needs" (Milestone 1.1.2) reports at the project site https://gmi3s.caida.org/outcomes/documents/vulnerabilities-harms-dataneeds_v2.4.pdf

- Attended two FCC workshops on BGP routing security and subsequently submitted as a https://catalog.caida.org/paper/2023_internet_science_moonshot advocating for a novel set of operational standards in routing security.
- Developed first draft of Monitor Specification Report describing monitor requirements (Milestone 1.1.4.1) as well as hardware and software needs (Milestone 1.1.3)
- Advanced BGP2Go development, including a new peer statistics dashboard (http://nids.caida.org:45000/cgi-bin/peerstats.sh) and slide deck describing data growth challenges https://gmi3s.caida.org/reports/tkrenc-bgpcollectors.pdf
- Submitted paper on design of next-generation BGP monitoring architecture to HotNets'23 https://catalog.caida.org/paper/2023_internet_science_moonshot

- Led multi-stakeholder industry work party (part of ICANN Security Advisory Committee) to resolve a long-standing DNS security issue that involves recommendations regarding data to analyze mitigation effectiveness (Milestone 1.1.3.7)
- Developed slide deck describing proposed design for active measurement architecture, presented to 4 IXPs who have offered to help us deploy it (two R&E IXPs, two commercial)
- Initiated new industry partnership with Gcore to deploy active measurement monitors on their commercial network (Milestone 1.1.6.5)
- Designed and developed a variety of new software for active measurement platform to support coordination among nodes, and to optimize performance and sustainability of low-resource nodes, run active measurements on the nodes using Python wrappers around *scamper* (Milestone 1.1.6.5)
- Designed and implemented new software framework to support active measurements from clouds (Milestone 1.1.6.5)
- Simplified our Annotated Schema design draft (now v1.1) based on feedback from the community, and updated AS Rank's data schema to align with these updates.
- Designed and prototyped new methods for visualization of Internet-scale topology data (Milestone 1.2.5.4)
- Established working group to integrate Internet data sets into UC San Diego Data Science Machine Learning Platform (DSMLP) (Milestone 1.2.8)
- Created data-sharing agreement to support DNS TLD Zone file data sharing; engage with ccTLD operator from Portugal to test out feasibility of current agreement. (Milestone 1.3.3.1)
- Completed initial draft of Differential Privacy Gap Analysis Report (Milestone 1.3.2.3)
- Completed a case study in differential privacy (Milestone 1.3.2.1)
- Summarized progress on working with industry data and achieved next step in sharing UCSD data with industry for commercial use (Milestone 1.3.3.1)
- Summarized latest EU Digital Services Act and how other countries are approaching this issue (Milestone 1.3.3.6)
- Conducted an Active Internet Measurements workshop which led to development of Active Measurement section of Monitor Specific Report (described above) (Milestone 1.4.1)

https://www.caida.org/funding/msri-gmi3s/reports/Active_measurements_needs.pdf

- Designed a mechanism to directly measure traffic through individual IXP members.
- Held second Strategic Advisory Council meeting (Milestone 1.5.2.3), which led to decision to establish Research Advisory Council composed of independent academic security researchers.
- Conducted various other meetings with stakeholders to get feedback on design (Milestone 1.4.1)
- RouteViews team compiled a draft architecture strategy which we are now decomposing into tasks for the revised PEP.
- Explored the current RouteViews data setup in BigQuery, and provided insights into the challenges and potential improvements in handling BGP data in Google BigQuery (Milestone 1.2.3.2)
- Enhanced the BGP2GO prototype by mapping every single BGP-related identifier to a specific MRT update file in RouteViews and updating the UI to support Keycloak authentication.
- Compiled the first draft of the report on the latest big Internet measurement data storage and management technologies, and posted it for internal review.
- Supported 12 publications in top networking/measurement venues (Milestone 1.4.5)
- We took over the FANTAIL development tasks. Vetted researchers can request access to the beta version of the system for evaluation. These activities were co-funded by NSF CNS-1925729 which expired August 3, 2023, but the lead developer abruptly retired mid-project to take care of his aging parents.
- We manually integrated a handful of external datasets including Peeringdb, NetAcuity, MaxMind, RV BGP data, and ICANN CZDS zone files into the GMI catalog.
- We designed and partially implemented a new Resource Access Authorization Portal that incorporates Keycloak and provides a comprehensive set of functionalities for managing user access to GMI resources and securing sensitive information.
- We continued our monthly DDoS meetings. We focused on the comparison between different vantage points and their specific (biased) views on long-term attack trends. We started analysis of various industry reports focusing on aspects like report accessibility, structure, language, measurement methods, definitions of attacks, and trends across reports. The main goal of this analysis is to compare views from different data sources, including academic as well as industry data.
- We undertook small group GMI-BGP meetings with European collaborators (Cristel Pelsser and her team) to discuss various dimensions of optimization of BGP infrastructure and data curation.
- We held a workshop 1-5 May 2023, focused on BGP and active measurements.
- We continued weekly calls with the GMI policy/economics working group to stay aware of data needs being articulated by policy analysts and Internet economists.
- We held our second Strategic Advisory Committee meeting on April 21, 2023.
- The infrastructure development activities in this project contributed to five accepted IMC papers last year (2022), and four this year (2023).

4. List of Acronyms

AIMS	Active Internet Measurements

AMPRNet	AMateur Packet Radio Network (Network 44)
API	Application Programming Interface
Ark	CAIDA Archipelago Measurement Infrastructure
AS	Autonomous System
AUA	Acceptable Use Agreement
AUP	Acceptable Use Policy
AY	At Year
BCP	Baseline Change Proposal
BGP	Border Gateway Protocol
BMP	BGP Monitoring Protocol
CAIDA	The Center for Applied Internet Data Analysis
CDN	Content Delivery Network
CDR	Conceptual Design Report
CENIC	California Educational & Research Network Information Center
CFR	Code of Federal Regulations
CI	Cyber Infrastructure
CISE	Computer and Information Science Engineering
CNS	Computer and Networked Systems
CMP	Configuration Management Plan
CPI	Cost Performance Index
CSAIL	Computer Science & Artificial Intelligence Laboratory
CY	Calendar Year
DDOS	Distributed Denial of Service
DHS	Department of Homeland Security
DNS	Domain Name System
DPM	Deputy PM
DPU	Data Processing Unit

DREN	Defence Research and Engineering Network
DZDB	DNS Zone Database
EAC	Estimate at Completion
eMMC	embedded Multi Media Card
ES&H	Environment, Safety and Health
EVMS	Earned Value Management System
FCC	Federal Communication Commisions
FFRDC	Federally Funded Research and Development Center
FTE	Full Time Equivalent Employee
FY	Fiscal Year
GE	Gigabit Ethernet
GMI3S	Global Measurement Infrastructure to Improve Internet Security
ICANN	Internet Corporation for Assigned Names and Numbers
IHR	Internet JHealth Report
IIJ	Internet Initiative Japan
IMC	Internet Measurement Conference
IOT	Internet of Things
ISP	Internet Service Provider
IYP	Internet Yellow Pages (IIJ)
KINDNS	Knowledge-Sharing and Instantiating Norms for DNS and Naming Security
L2	Level 2
LFO	Large Facilities Office
LoC	Letter of Collaboration
MANRS	Mutually Agreed Norms for Routing Security
MIT	Massachusetts Institute of Technology
MREFC	Major Research Equipment and Facilities Construction
NIC	Network Interface Card

NIDS	Network Infrastructure Data Science
NIS	Network and Information Security (EU)
NIST	National Institute of Standards and Technology
NITRD	Network and Information Technology Research and Development
NLP	Natural Language Processing
NOG	Internet Network Operators' Group
NSF	National Science Foundation
NSRC	Network Startup Resource Center
NTIA	National Telecommunications and Information Administration
OAC	Office of Advance Cyberinfrastructure
OMB	Office of Management and Budget
PI	Principal Investigator
PII	Personally Identifiable Information
PEP	Project Execution Plan
PM	Project Manager
PMCS	Project Management Control System
PO	Program Officer 5
PRP/NRP	Pacific Research Platform/National Research Platform

QA	Quality Assurance
QC	Quality Control
R&D	Research and Development
RIPE-NCC	Regional Internet Registry for Europe Network Coordination Centre
RIS	Routing Information System
RLS	Resource Loaded Schedule
ROA	Routing Origin Authorization
RPKI	Resource Public Key Infrastructure

RV	Route Views
SD card	Secure Digital card
SDSC	San Diego Supercomputer Center
SOW	Statement of Work
SPI	Schedule Performance Index
SSAC	Security and Stability Advisory Committee
TLD	Top-level domain
TPC	Total Project Cost
UO	University of Oregon
U.S.	United States
USC ISI	USC Information Science Institute
VP	Vantage Point
WBS	Work Breakdown Structure