designing a Global Internet Measurement Infrastructure

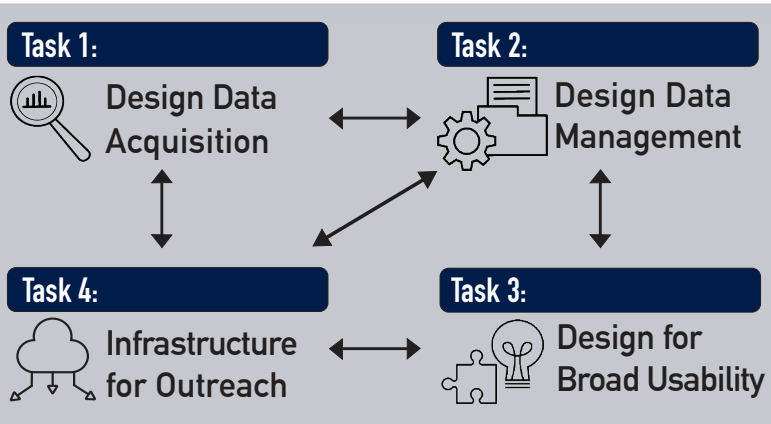
GMI

kc claffy (CAIDA/UCSD), David Clark (CSAIL/MIT), Steve Huter (NSRC/UO)

OVERVIEW

The goal of this project is to design a new generation of measurement infrastructure for the Internet, which will support collection, curation, archiving, and expanded sharing of data needed to advance critical scientific research on the security, stability, and resilience of Internet infrastructure. While the Internet has become critical infrastructure permeating all aspects of modern society, its security and trustworthy character are subject to constant threats and attacks. The security of the Internet infrastructure is a high priority for the security research community, but that community is greatly hindered by a lack of relevant data. Our project attempts to improve this situation by gathering data that can inform the development of long-lasting advances in security.

Our focus is on the Internet as a data transport service, and the vulnerabilities specific to that layer: attacks on the Internet global routing system (the Border Gateway Protocol or BGP) that deflect traffic to bogus destinations (a persistent problem), abuses of the Domain Name System (a widespread, pernicious problem), attacks on the key management system (the Certificate Authorities) that underpin identity and authentication on the Internet, and spoofing of Internet addresses in order to disrupt regions of the Internet with untraceable traffic (Denial of Service attacks).



These challenges are *foundational*: they affect the reliable operation of every application that operates over the Internet. Lack of attention to security in early design decisions, persistent disagreement about the best path to improvement, and lowest-cost operational practices surrounding these layers have allowed malicious actors to execute and scale harmful behavior including interception and disruption of traffic, denial of service and phishing attacks, and distribution and execution of malware.

PROJECT STRUCTURE

We organize this project in four tasks.

1. Our first task is to design, prototype, test and evaluate a new highly distributed **network measurement platform** capable of capturing four types of security-relevant data (topology, routing, unsolicited traffic, DNS), as well as hosting vetted experiments. This task will require consideration of both hardware and virtualized software deployments, in a modular architecture that allows hosting sites to opt in to measurements consistent with their evolving policies.
2. Our second task covers many facets of **data management**: curation, including anonymization, post-processing and analytics. We will design capabilities to make the datasets easy to discover, use, and share, including consistent APIs, meta-data, and efficient dissemination approaches.
3. Our third task focuses on community-oriented infrastructure that will enable use of the data for a broad set of **cybersecurity research** and beyond. This task will tackle issues with sensitive data that raises privacy or corporate concerns. One goal is to bridge the gap between emerging data disclosure control technologies and network and security practitioners. We will explore the relevance of computer science advances such as differential privacy and secure multi-party computation, to cybersecurity research priorities. To foster broad use of data, we will build on successful data-sharing agreements, demonstrate their utility with commercial case studies, and socialize these among our partners and the larger community.
4. Task four will include workshops, and curriculum development to support **STEM/cybersecurity workforce training**. To scale methods for training students and researchers how to analyze Internet measurement data, we will develop a Network Infrastructure Data Science course, including modules on responsible (ethical, privacy-respecting) use of data and analytics.

EVALUATING OUR DESIGN

As part of this design phase, we need a subset of the Internet where we can work with operators, deploy and evaluate prototypes, and develop practices for community access to our data. Operators of 10 R&E networks have committed to work with us to design and prototype a measurement platform on their networks. These R&E networks encompass campus networks as well as regional, national and transnational networks focused on the needs of the R&E community.

These needs include high performance processing of scientific data, student privacy, protection of unreleased research, and sustainability.

A key goal of this project is to allow vetted researchers to deploy their own measurement experiments on the resulting distributed platform. This objective will require mechanisms and policy to ensure that experiments are consistent with policies of the different parties hosting the devices and will not cause harm to the Internet or the reputation of the hosting networks.

PROJECT MANAGEMENT & PARTNERS

The lead organization for the 3-year Design Phase of this project is the Center for Applied Internet Data Analysis (CAIDA) at UC San Diego. CAIDA is recognized by the international Internet research community as a unique source of Internet data. CAIDA has developed and sustained programs to share data and knowledge, including sensitive knowledge about critical infrastructure, for two decades. CAIDA's web site lists over 1900 publications by non-CAIDA authors that made use of CAIDA data.

CAIDA brings years of experience with legal and ethical frameworks for data sharing, as well as experience in data collection, curation, handling, transformation, and anonymization to protect privacy, and organizational processes and accounting infrastructure to handle many classes of data, different modes and levels of access, and computational resources to support Internet science.

Our team includes two community-focused organizations with a 20-year history of collecting and sharing the most security-relevant data about the global routing system: NSRC/UO (Route Views) and RIPE (the EU's IP address registry, which operates the Routing Information Service). In addition to deep technical expertise, NSRC will help with engagement, strategic coordination, and project management.

kc claffy, the founder and director of CAIDA, and David Clark (MIT) are principal investigators of this project. Clark plays a central role in this proposed work, with five decades of experience with Internet architecture, security, measurement, and policy guidance.

Our critical assets also include partners with commercial network engineering expertise, data science techniques, and technology to manage data integrity, availability, and privacy. To reach the implementation phase, this project will require a rich structure of working groups and shared intellectual property to enable sustainable and secure infrastructure operations and use.

POLICY ENGAGEMENT

As we concentrate on the design of infrastructure and methods to gather and work with data, we will also consider how to translate technical understanding into actionable knowledge of direct use to society. For example, one premise of our work is that we cannot overcome today's security threats to network infrastructure by responding to individual events. Rather we must shift the landscape by understanding the operation of these Internet systems and behavior patterns of malicious actors, with the goal of finding operational practices that will thwart attackers. This is a fundamentally interdisciplinary challenge, with economic and policy implications. An essential goal of this project is to provide a missing piece to tackle this challenge: objective, third-party data-driven analysis that can inform the technological as well as governance architectures of Internet infrastructure.

BROADER IMPACTS

This design project will enable application of data intensive methods to study the global Internet infrastructure, supporting scientific and engineering advances to navigate current and future Internet-related harms. The project will contribute to a broad range of disciplines that now depend on data about the Internet, including network science, socioeconomic studies, international relations, and political science, and will play a key role in raising the next generation of U.S. leaders in information technology. The resulting capabilities to support data acquisition, curation, and sharing will have an inherent equalizing effect on the research community. Moreover, if successful, this cyberinfrastructure will increase the trustworthiness of the Internet for U.S. citizens and serve as a model for rest of the world.

This material is based on research sponsored by the National Science Foundation (NSF) grant OAC-2131987.

UC San Diego

UNIVERSITY OF OREGON