Statement of Work
for
Protected Repository for the Defense of Infrastructure Against Cyber Threats (PREDICT)
The University of California, San Diego (UCSD)

1.0 Objective

The objective of the Protected Repository for the Defense of Infrastructure against Cyber Threats (PREDICT) is design, develop, and implement a large-scale, privacy-protected, dataset repository of real network and system traffic for use by the cyber security research community, both in the U.S. and international. The PREDICT will accelerate design, production, and evaluation of next-generation cyber security solutions, including commercial products. The University of California, San Diego (UCSD) will support PREDICT objectives as defined by this cooperative agreement.

2.0 Scope

The scope of the work to be performed by UCSD is as both a Data Host (DH) and a Data Provider (DP). As a data provider, UCSD asserts ownership or a right to control and disclose to researchers the data UCSD provides. UCSD will provide data, data collection, data curation, data catalog support, data use restrictions, privacy and anonymization of data, Internal Review Boards or ethics review support, legal support, data risk analysis, and data evolution support. As a data host, UCSD will maintain computing infrastructure to store data and host data received from external sources, within resources. UCSD will, also, host its own datasets and external data, and provide legal support as needed for data hosting tasks. UCSD will also provide PREDICT project support for Principal Investigator meetings, metrics development, PREDICT outreach and the PREDICT Application Review Board. Overall, dataset(s) provided by UCSD for publication and hosting must be compliant with any law or regulation that is pertinent to the dataset content, to include Department of Homeland Security (DHS) privacy policies, and full compliance with the PREDICT legal framework (which includes international dissemination), described below.

3.0 Background

UCSD has been a data host and data provider for DHS's PREDICT project during its initial phase. The primary goal of PREDICT is to bridge the gap between producers of security-relevant network operations data and technology developers and evaluators who can leverage this data to accelerate the design, production, and evaluation of next-generation cyber security solutions. Central to PREDICT management is the PREDICT Coordination Center (PCC). The PCC facilitates the release of datasets by data hosts to approved researchers, subject to the terms and conditions set forth by DHS, the PCC, data providers, and data hosts. In support of these activities, the PCC develops, hosts and maintains a web portal (http://www.predict.org) that advertises the datasets available from the PREDICT project and automates the generation of appropriate agreements for and between PREDICT entities. The primary agreements managed by the PCC are the following:

a. Memorandum of Agreement (MOA) between the PCC and Researcher
b. MOA between the PCC and Data Host.
c. MOA between PCC and Data Provider.
d. Data Use Agreement between the Data Provider and the Researcher.

Other PREDICT program entities and external interfaces are as follows:

a. PREDICT Coordinating Center (PCC) - The PCC manages the PREDICT data catalog and operations, processes researchers' applications for PREDICT data, and handles administrative matters.

b. Researcher - a person who requests PREDICT datasets in an individual capacity and who has been identified by a Referring Organization as someone who has a legitimate need for the data.

c. Principal Investigator – a researcher identified in a contract with the Government funded to perform PREDICT tasks.

d. Referring Organization - an entity that identifies a Principal Researcher as someone who is affiliated or aligned with the Referring Organization and who has a legitimate need for PREDICT datasets.

e. Research Organization - an organization that desires to have research conducted on its behalf and designates individuals as Data Custodians to request and be responsible for PREDICT datasets.

f. Application Review Board (ARB) - an entity that reviews and approves or rejects applications for Data from Researchers. Data Hosts and Data Providers may be required to participate in this forum.

The relationship between PREDICT entities and the workflow of PREDICT information and agreements is depicted in Figure 1, below.

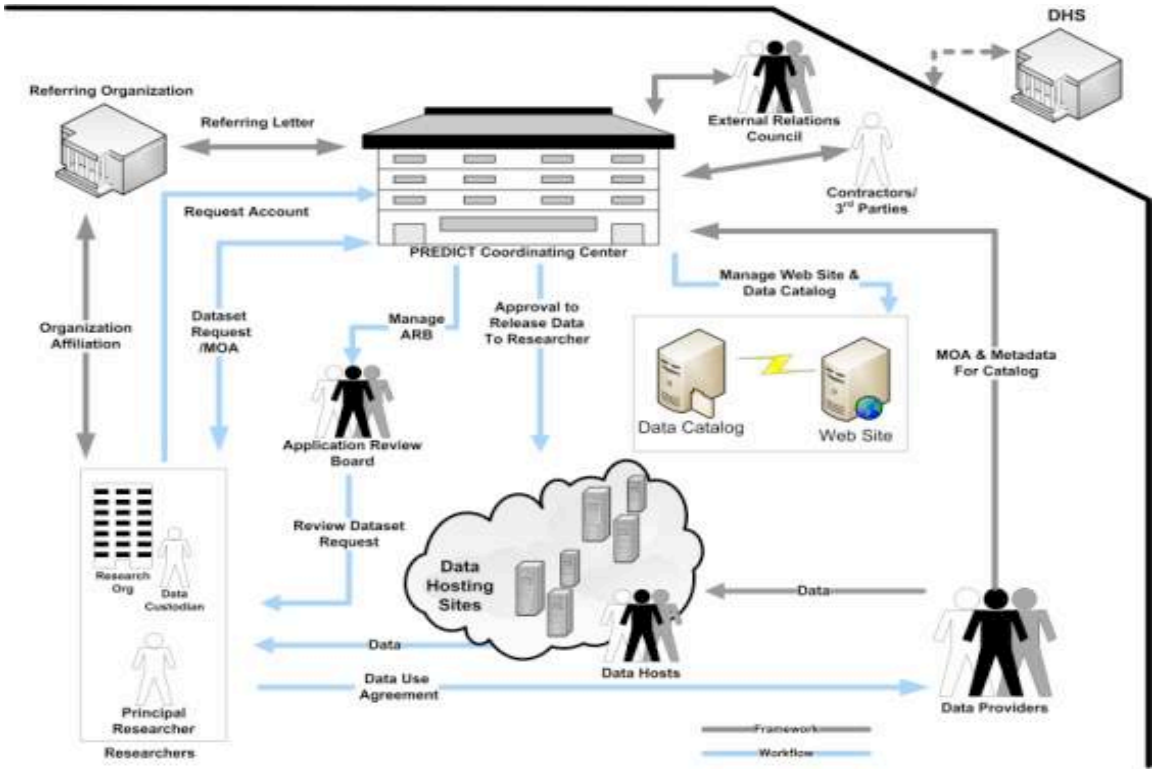

Figure 1. PREDICT Work Flow

4.0 Technical Requirements.

The contractor, to include subcontractors, shall accomplish PREDICT data providing, data hosting, and project support as follows:

4.1 Providing Data

The contractor shall implement and document data providers processes and relationships depicted in Figure 1 and described below, to include adherence to the PREDICT legal framework. Specifically, the contractor shall support PREDICT by providing multiple internet traffic datasets including:

a. Internet topology Data
   (1) Internet Topology Measured from Ark Platform will be provided including:
      i. IPv4 Routed /24 Topology (forward IPv4 paths, reply Time-to-Live (TTL), Round-Trip-Time (RTT), and ICMP responses)
      ii. IPv4 Routed /24 DNS Names (fully-qualified domain names for IP addresses in the IPv4 Routed /24 dataset)
      iii. IPv6 Topology (IP paths, RTT, TTL, and ICMP for IPv6)
   (2) Active Internet Topology Measurements with Skitter will be provided. This 4-TB archive contains forward IP paths and RTTs to hundreds of thousands of IPv4 destination addresses collected in 1998-2008 using the skitter probing tool from 24 monitors on 4 continents. These legacy measurements can be used to study the historical development of macroscopic connectivity and performance of the Internet.
   (3) Internet Topology Data Kits (ITDK) - Router-level topology data, router- to-AS assignments, geographic location of each router, and DNS lookups of all observed

IP addresses will be provided.

These Internet topology data sets can be used for modeling and simulation of malware propagation and containment measures, infrastructure stability vulnerability assessments, longitudinal studies of Internet topology evolution, Internet address mapping and inferences.

b. Blackhole Address Space Data – The UCSD Real-time Network Telescope Data will be provided in PCAP format. Telescope data is available in near real-time (1-hour delayed), covering a sliding most recent two-month window, to strategic projects that cannot use a static archived snapshot of data. Static archived snapshots will also be provided. This blackhole address space data supports study of the origin and characteristics of Internet pollution, evaluating various malware collection approaches, developing efficient mitigation strategies, and monitoring Internet censorship or outage events on a global scale.
c. IP Packet Header Data – The data to be provided is OC48 Peering Point IP Packet Headers. This data set consists of three packet traces captured at an OC48 Peering Point in 2002-2003. This legacy data contains packet headers in PCAP format and supports research on Internet traffic and classification, including analysis of security-related events.

4.1.1 Data Collection

Provide a collection mechanism for any dataset provided to the PREDICT project. Deliver a document describing the physical, logical and functional configuration of the collection mechanism to include and any planned evolution of the collection mechanism.

4.1.2 Data Curation

Document and implement a process for managing the format and organization of any dataset provided to the PREDICT project for dissemination via the PREDICT legal framework.

4.1.3 PREDICT Data Catalog Support

Support the maintenance of the PREDICT data catalog hosted by the PCC consistent with the data collection and data curation tasks.

4.1.4 Data Use Restrictions

Describe any restrictions on the use of any datasets provided to PREDICT and identify any issues that would prevent the dissemination of any dataset(s) internationally.

4.1.5 Privacy and Anonymization

Describe all efforts and methods employed to ensure that the data is legally collected and that the data to be provided is compliant with privacy laws. Describe any alteration of the data (e.g., anonymization) and appropriate disclosure control processes. If none of the aforementioned activities are needed, explicitly state non applicability.

4.1.6 Institutional Review Boards or Ethics Reviews

Describe and implement any Institutional Review Board (IRB), or ethics review processes related to dataset requests, required to release the data they available via the PREDICT project. The description shall include nominal timelines, the issues the IRB, or ethics review, would have to consider, and the expected frequency of IRB interaction or ethics reviews.

4.1.7 Legal Support
UCSD shall review and negotiate the applicable legal documents required to support the PREDICT legal framework and efforts of the PREDICT Coordinating Center (PCC). Describe how UCSD will meet PREDICT goals for the execution of PREDICT legal documents including Data Provider Memoranda of Agreement (MOAs) and MOAs with the PCC. As each of the aforementioned documents are living, these documents will require revision without any predictability, discuss the availability of decision making legal support and any other required entity needed to execute such documents. UCSD shall respond to these documents consistent with UCSD policy and within seven (7) business days.

4.1.8 Data Risk Analysis
Provide a risk analysis of any dataset provided that addresses any federal, state, local, and international laws that are relevant to the collection and dissemination of the dataset, as well as any ethical issues.

4.1.9 Data Evolution
If applicable, describe how data collections are planned to evolve as devices, architectures, and protocols evolve.

4.2 Data Hosting
Implement and document the PREDICT data hosting processes and relationships Figure 1 and below. Support adherence to the PREDICT legal framework and make data available to approved researchers in accordance with PREDICT policies. Data hosting tasks are below.

4.2.1 Data Hosting Processes
Provide a data hosting infrastructure to support the PREDICT project. Describe any expansion plans for hosts and bandwidth to be needed as a result of traffic growth. Describe scenarios and processes for the dissemination of data that may occur via media. Provide a system description document for data hosting infrastructure to be employed in the performance of this SOW. The description will include hardware, software, logical configuration, and mirroring or redundancy equivalence.

4.2.2 External Data
Host data from external sources as a result of coordination with data sources outside of the PREDICT program providing data for use via PREDICT. Indicate the availability of infrastructure for hosting external data. Describe a plan for curating data from external sources, if willing to host external data.

4.2.3 Legal Support
UCSD shall review and negotiate the applicable legal documents required to support the PREDICT legal framework and efforts of the PREDICT Coordinating Center (PCC). Describe how they will meet PREDICT goals for the execution of PREDICT legal documents including Data Host Memoranda of Agreement (MOAs) and MOAs with the PCC. As each of the aforementioned documents are living, these documents will require revision without any predictability. Discuss the availability of decision making legal support and any other required entity needed to execute such documents. UCSD shall respond to these documents consistent with UCSD policy and within a seven (7) business days.

4.3 Project Support

Implement and document PREDICT project support, as follows:

4.3.1 Meeting Support

Designate a principal investigator (PI). The PI shall support and attend PI Meetings not to exceed three (3) times per year. Provide status briefings and participation in PREDICT planning and outreach. Provide an on-site venue for hosting the PI Meeting, up to three (3) times per year. Participate in program meetings, including teleconferences, program reviews and other technical interchange meetings.

4.3.2 PREDICT Outreach

Describe and implement a plan to publicize the availability of the data provided via PREDICT, and support outreach by supporting Government approved workshops, conferences and other technical forums.

4.3.3 Metrics

Propose, provide and monitor metrics to describe the utility of the datasets provided and also the growth and management of the data hosted as well as data provided. Collaborate and cooperate with PREDICT participants on the establishment and monitoring of project-level metrics.

4.3.4 PREDICT Application Review Board (ARB) Support

Support the PREDICT ARB activities whenever datasets provided are on the agenda to screen researcher organizations, referring organizations and researchers for legitimacy of purpose and intent of PREDICT data use.

5. Deliverables

The Contractor performs fundamental research on a reasonable efforts basis and in accordance with UC policy. The contractor shall provide the following deliverables:

5.1 Project Management Plan

The contractor shall provide a project management plan that documents the artifacts and addresses all of the technical requirements in Section 4.0. Delivery shall occur within forty-five (45) days of award, with updates to be delivered annually thereafter.

5.2 Hosting Infrastructure Description Document

The contractor shall use provide a system description document for data hosting infrastructure to be employed in the performance of the SOW. The description will include hardware, software, logical configuration, and mirroring or redundancy equivalence. Delivery shall occur within forty-five (45) days of award, with updates to be delivered annually thereafter.

5.3 Quarterly Technical Status Report

The contractor shall provide a quarterly technical report that includes the following:

a. Task progress
b. Project plans for the next quarter
c. Issues, concerns or suggestions
d. Datasets added and size
e. Dataset Collection/Size Storage Update
f. Hosting capacity

g. Hosting capacity used
h. Changes to hosting infrastructure
i. Monthly changes in metrics for the quarter
j. PREDICT, policy, ethics or IRB issues
k. Dataset requests
l. Dataset distribution
m. Travel and presentations
n. Papers published
o. Experiments
p. Listing of external papers giving attribution to the contractor data provided by PREDICT
q. Additional highlights (e.g. press releases)

5.4 Monthly Financial Status Report

The contractor shall provide a monthly financial report that includes the following:

a. Amount Obligated
b. Amount Expended
c. Amount Invoiced for the month
d. Projected monthly expenditures
e. Expenses incurred but not invoiced

5.5 Briefings and Research Papers. The contractor shall deliver all presentations and/or research papers supported by this SOW and grant to the Department of Homeland Security (DHS), and the Government of the United States of America or those acting on its behalf, a nonexclusive, irrevocable, worldwide license to use the presentation or other material delivered and to reproduce, distribute copies to the public, and perform publicly and display publicly such materials by or on behalf of the DHS and the United States Government.

5.6 Final Report. The contractor shall deliver a final report that summarizes all activities during the period of performance.