Supporting Research and Development of Security Technologies through Network and Security Data Collection

Cooperative Agreement FA8750-12-2-0326, CAIDA, UCSD

HOSTING INFRASTRUCTURE DESCRIPTION

The following describes the data hosting infrastructure (Figure 1) deployed and maintained by CAIDA staff on the UC San Diego campus to host and serve datasets provisioned by CAIDA for the DHS PREDICT Project. The described infrastructure was not exclusively funded by the PREDICT Project.

1. Description.

To support the requirements that come with the roles of PREDICT Data Host (DH) and Data Provider (DP), the CAIDA group at UCSD maintains numerous computers mounted in the machine room at the San Diego Supercomputer Center. Our system administrators have designed, configured and deployed these hosts to provide high availability for data collection, indexing, curation, and distribution to researchers approved via the PREDICT vetting mechanism.

As a general strategy, we have chosen to deploy several hosts with moderately large (20-40TB), locally attached disk systems that make use of the ZFS file systems. These configurations optimize for cost of storage and availability for data consumers. We also run several systems that act as web servers for hosting project description pages as well as for data distribution to vetted account holders. We have begun making use of FreeBSD jails servers that mount exported file systems from the backend data servers.

Separately, as an experiment to compare cost of ownership, reliability, and performance, we also purchase cloud disk storage from our host department, SDSC (http://cloud.sdsc.edu/). We use this storage for archival of datasets and plan to also utilize the services for distribution of data.

Additionally, we make use of an Energy Research Computing Allocations Process (ERCAP) Allocation at the National Energy Research Scientific Computing Center (NERSC) facility, a division of the Lawrence Berkeley National Laboratory located in Berkeley, California. SDSC enjoys high bandwidth connectivity (10GB) with the NERSC.GOV domain allowing to conduct regular file transfers for archival of historical data.

2. System Inventory

1. Data Server (thoth.caida.org)
   OS: FreeBSD 8.2
   CPUs: 1 x 4 core Intel(R) Xeon(R) CPU E5620 @ 2.40GHz
   Memory: 6GB RAM
   Storage: 48 TB raw disk (38 TB after RAID 6 and zfs overhead)
   Description: This machine is our main data server.

2. Web Server (alcatraz.caida.org)
   OS: FreeBSD 8.2 (jail server)
   CPUs: 2 x 6 core Intel(R) Xeon(R) CPU X5670 @ 2.93GHz
   Software: Apache
   Memory: 48GB RAM
   Storage: minimum requirements for jailed operating system and mounted file systems.
   Description: The FreeBSD jail server that mounts the exported file systems from the thoth.caida.org data server to serve http://data.caida.org/. We run two redundant jail servers so that one can take over if the other fails.

3. Web Server (attica.caida.org)
   OS: FreeBSD 8.2 (jail server)
   CPUs: 2 x 6 core Intel(R) Xeon(R) CPU X5670 @ 2.93GHz
   Software: Apache
   Memory: 48GB RAM
   Storage: minimum requirements for jailed operating system and mounted file systems.
   Description: This machine is the backup jail server.

4. Web Server (indy.caida.org)
   OS: FreeBSD 7.3
   CPUs: 2 x 4 core Intel(R) Xeon(R) CPU X5667 @ 3.07GHz
   Software: Apache
   Memory: 32GB RAM
   Storage: 12 TB of raw disk (~9TB useable after RAID 5 and filesystem overhead)
   Description: This host is the main data server for the Archipelago measurement infrastructure. It is also the main distribution server for http://topo-data.caida.org.

5. Data Server (thor.caida.org)
   OS: FreeBSD 8.3
   CPUs: 2 x 6 core Intel(R) X5675 @ 3.07GHz
   Memory: 192GB RAM
   Storage: 102 TB raw disk (74 TB after RAID 6 and zfs overhead)
   Description: This host acts both as the primary data server and the primary analysis machine for the UCSD Network Telescope data.

6. Web Server (monsterzero.caida.org)
   OS: FreeBSD 6.2
   CPU: 1 x 2 core Intel(R) Pentium(R) D CPU 3.00GHz
   Software: Apache
   Memory: 1GB RAM
   Storage: 152GB local disk
   Description: This server hosts http://topo-data.caida.org/. It mounts all topology data over NFS from indy.caida.org.

7. Data and Compute Server (ogma.caida.org)
   OS: FreeBSD 8.3
   CPUs: 1 x 4 core Intel(R) Xeon(R) CPU W3530 @ 2.80GHz
   Memory: 6GB RAM
   Storage: 48 TB raw disk (38 TB after RAID 6 and zfs overhead)
   Description: This server hosts the UCSD Network Telescope data and acts as the main compute server for near realtime telescope data.

8. Web Server (cider.caida.org)
   OS: FreeBSD 5.4
   CPUs: 1 x 2 core Intel(R) Pentium(R) D CPU 3.00GHz
   Software: Apache
   Memory: 3GB RAM

Storage: 73GB local disk
Description: This server supports the web infrastructure that serves the dataset description pages, forms, and project web pages http://www.caida.org/data/ and http://www.caida.org/projects/predict/.

9. Data Archival and Storage (cloud.sdsc.edu)
OS: Linux (Rocks Clustering tool kit)
Software: OpenStack (Swift)
Storage: 45 TB
Description: We purchase cloud disk storage space from SDSC (http://cloud.sdsc.edu) for storing our system backups and archiving PREDICT data. 20 TB of this space originally came as a transitionary allocation from SDSC during graduate decommissioning of HPSS and SamQFS tape facilities. In the coming year, we will be charged for the entire 45 TB.

10. NERSC HPSS Tape Archive Allocation
OS: AIX
Software: HPSS Tape Services
Storage: 150 TB
Description: We enjoy a 150 TB allocation of HPSS tape resources at the NERSC facility where we archive our historical UCSD Network Telescope (darknet) data. As of the end of 2012, we have used approximately 105TB of this allocation.

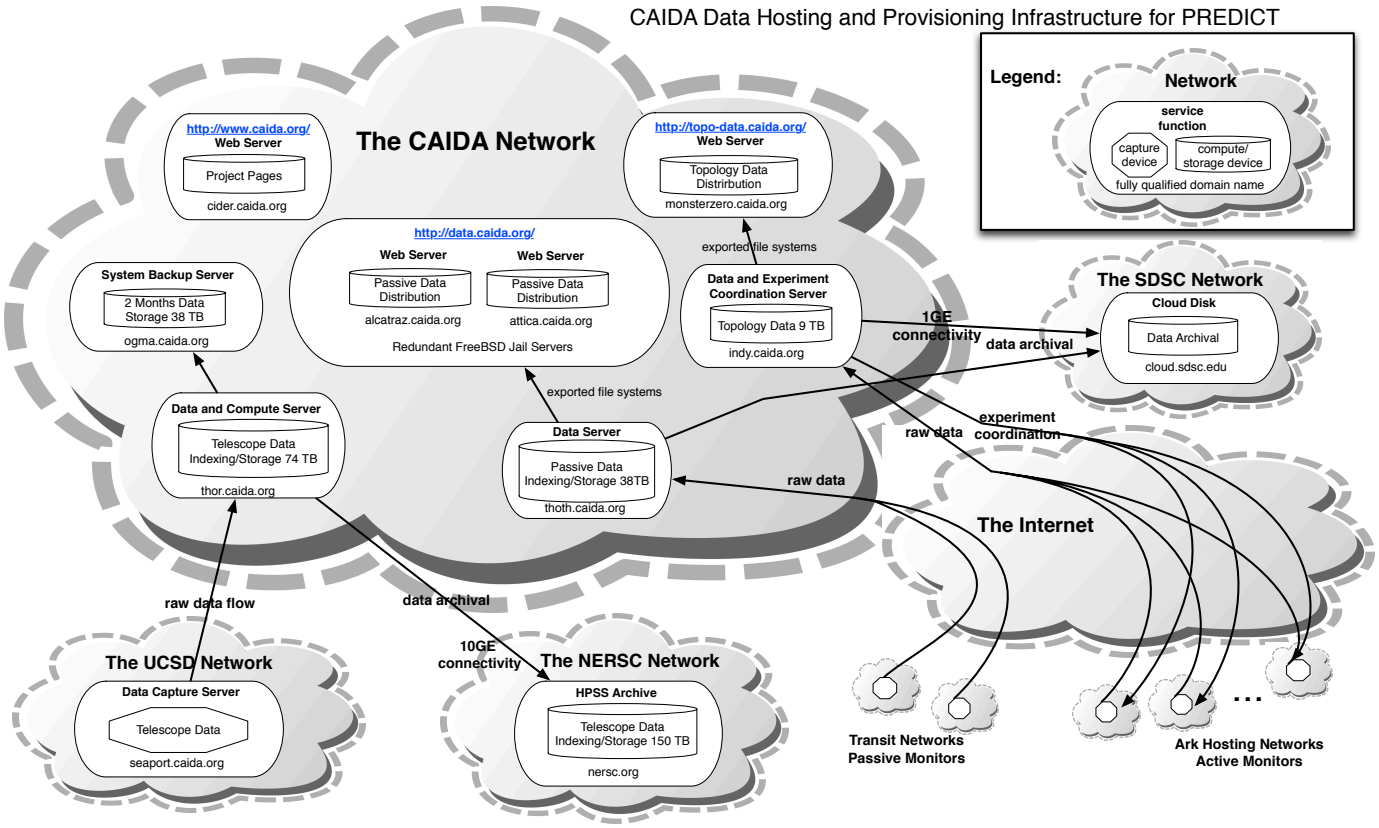

Figure 1. UCSD Data Hosting and Provisioning Infrastructure for PREDICT.