



August 24, 2012

The Archipelago (Ark) Project<sup>1</sup> is an active measurement infrastructure for Internet topology discovery operated by CAIDA. The primary goals are to

- reduce the effort needed to develop and deploy sophisticated large-scale measurements,
- provide a step toward a community-oriented measurement infrastructure by allowing collaborators to run their vetted measurement tasks on a security-hardened distributed platform, and
- collect and distribute datasets that support a variety of research, engineering, and operational interests that promote a robust, scalable global Internet infrastructure.

Ark is tailored specifically for active network measurement. This allows Ark to be simpler than other general-purpose distributed experimental platforms and to concentrate on providing facilities that directly address the needs of networking research. For example, we provide a facility for communication and coordination that makes it easier to write distributed measurements that must work together to achieve a goal. Our aim is to lower the barrier to bringing novel and interesting measurement techniques to life.

Organizations that wish to host an Ark monitor agree to our Memorandum of Cooperation (MoC)<sup>2</sup> and donate rack space, power, and cooling for a server dedicated to Ark measurement. Each Ark measurement node broadens the view of the global Internet for the network research community and increases the completeness and accuracy of data representing the topological structure of the Internet. Our main goal in deploying Ark monitors is to achieve a geographically and topologically diverse view of the Internet. As of August 2012, the Ark infrastructure consists of 61 monitors, 28 of which are IPv6-capable, distributed over 32 countries worldwide. Geographically,<sup>3</sup> monitor distribution is concentrated in North America (20 monitors), Europe (18 monitors), and the Asia-Pacific Rim (12 monitors). The remaining monitors are located in South America, Africa, and elsewhere.

An important focus of Ark is large-scale traceroute-based topology measurements from multiple vantage points to the global IPv4 and IPv6 Internet. The results of these measurements, which have been ongoing since September 2007, are made available to the research community in several datasets. The primary dataset, the “IPv4 Routed /24 Topology Dataset”<sup>4</sup>, contains 7 TB of data as of August 2012. In addition, we provide the results of DNS name lookups for all IP addresses observed in traceroute paths<sup>5</sup>; Autonomous System (AS) links derived from IP paths<sup>6</sup>; and “Macroscopic Internet Topology Data Kits (ITDKs)”<sup>7</sup> that provide router-level topology derived from IP paths using multiple alias-resolution techniques.

Network researchers use CAIDA topology data to conceive, develop, and test their models and methods. The range of scientific experiments has successfully demonstrated our vision of a metaphorical distributed measurement “operating system” to support empirical Internet science.

1 <http://www.caida.org/project/ark/>

2 <http://www.caida.org/projects/ark/moc/ark-moc-v1.xml>

3 <http://www.caida.org/data/monitors/monitor-map-ark.xml>

4 [http://www.caida.org/data/active/ipv4\\_routed\\_24\\_topology\\_dataset.xml](http://www.caida.org/data/active/ipv4_routed_24_topology_dataset.xml)

5 [http://www.caida.org/data/active/ipv4\\_dnsnames\\_dataset.xml](http://www.caida.org/data/active/ipv4_dnsnames_dataset.xml)

6 [http://www.caida.org/data/active/ipv4\\_routed\\_topology\\_aslinks\\_dataset.xml](http://www.caida.org/data/active/ipv4_routed_topology_aslinks_dataset.xml)

7 <http://www.caida.org/data/active/internet-topology-data-kit/>

Sincerely,

Josh Polterock,

Manager of Scientific Projects  
The Archipelago Team

Mail : Josh Polterock  
CAIDA/SDSC/UC San Diego, MC-0505  
9500 Gilman Drive  
La Jolla, CA 92093-0505  
Phone : (858) 822-0957  
Email : [ark-info@caida.org](mailto:ark-info@caida.org)

## **About CAIDA**

The Cooperative Association for Internet Data Analysis (CAIDA), based at the University of California's San Diego Supercomputer Center in San Diego, US, is a collaborative undertaking among organizations in the commercial, government, and research sectors aimed at promoting greater cooperation in the engineering and maintenance of a robust, scalable global Internet infrastructure.

CAIDA investigates practical and theoretical aspects of the Internet in order to:

- provide macroscopic insights into Internet infrastructure, behavior, usage, and evolution,
- foster a collaborative environment in which data can be acquired, analyzed, and (as appropriate) shared,
- improve the integrity of the field of Internet science, and
- inform science, technology, and communications public policies.