

Cartographic Capabilities for Critical Cyberinfrastructure (C4)

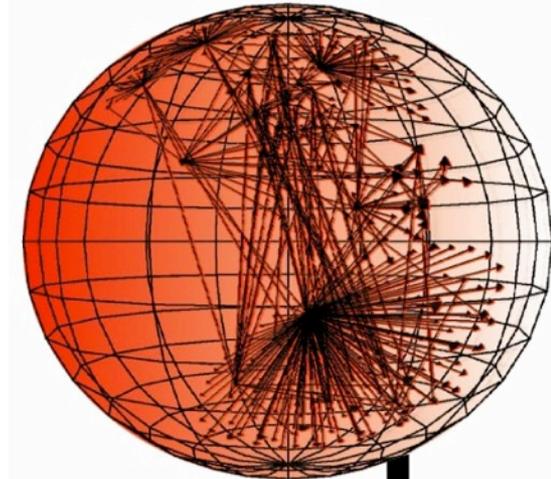
Cyber Security Division 2012 Principal Investigator Meeting

9-11 October 2012

Presenter: Bradley Huffaker
bhuffake@caida.org

PI: k claffy
kc@caida.org

Cooperative Association for Internet Data Analysis
San Diego Supercomputer Center
University of California at San Diego



caida

UC San Diego

Introduction: Team

- Cartographic Capabilities for Critical Cyber-infrastructure (C4) will address TTA #7 by improving our ability to **identify**, **monitor**, and **model** critical infrastructure.
 - **k claffy (PI)**, **founder and director** Cooperative Association for Internet Data Analysis (CAIDA)
 - **Young Hyun**, **architect and developer**, Archipelago and MIDAR
 - **Bradley Huffaker**, **researcher**, in Internet topology analysis and visualization.
 - **Amogh Dhamdhere**, **research scientist**, Internet topology and economic analysis
 - **Marina Fomenkov**, **project manager**, coordination, reporting, and data analysis
 - **Josh Polterock**, **technical manager**, oversee deployment data collection
 - **Ken Keys**, **programmer analyst**, software development, measurements, integration
 - **Daniel Andersen**, **system administrator**, administration and support of computer systems
 - **Alex Ma**, **web programmer**, developing and maintaining the CAIDA web site

Technical Approach

- Project Phases:
 - Phase I: Applied **Research** (18 months)
 - Phase II: **Development** (12 months)
 - Phase III: Technology **Demonstration** (6 months)
 - Testing and **Evaluation**
 - Internet Topology Data Kit (**ITDK**)

router
geolocation

router AS
assignment

hostnames

nodes

links

ITDK Datasets

Technical Approach: Phase I

Phase I: Applied Research: increase completeness, accuracy, and potency of Internet mapping at multiple granularities.

1. Improve completeness of macroscopic Internet maps
 - a) **Install** monitoring infrastructure in underserved regions
 - b) **Investigate** more efficient and scalable **probing techniques**
 - c) **Incorporate** additional IP address lists into alias resolution process
 - d) **Synthesize** comprehensive **topology** from all available sources

Ark monitor locations

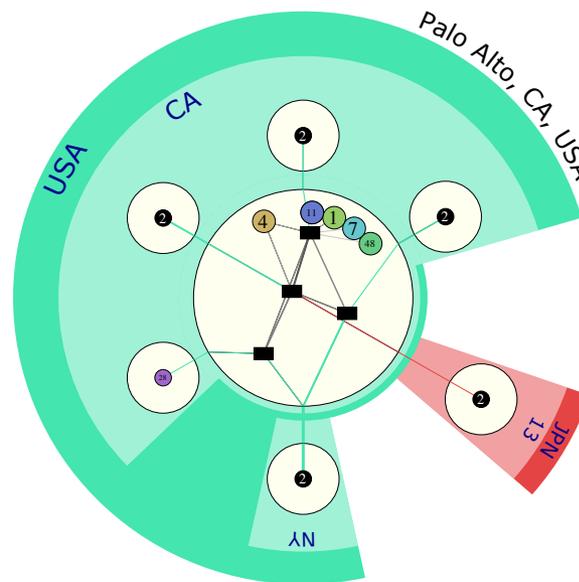


Technical Approach: Phase I

2. Increase accuracy of macroscopic Internet maps
 - a) investigate impact of **false link** inferences on different graphs
 - b) Investigate approaches to identifying **AS peering** links
 - c) Develop **validation support** for multiple levels of inference

3. Increase the **richness** of macroscopic Internet maps
 - a) Add annotations for intermediate (PoP/city-) **infrastructure** inferences
 - b) Add **economic AS** annotations
 - c) Extract **regional anomalies**

Inferred routers and ASes they connect to in IJ's Palo Alto PoP

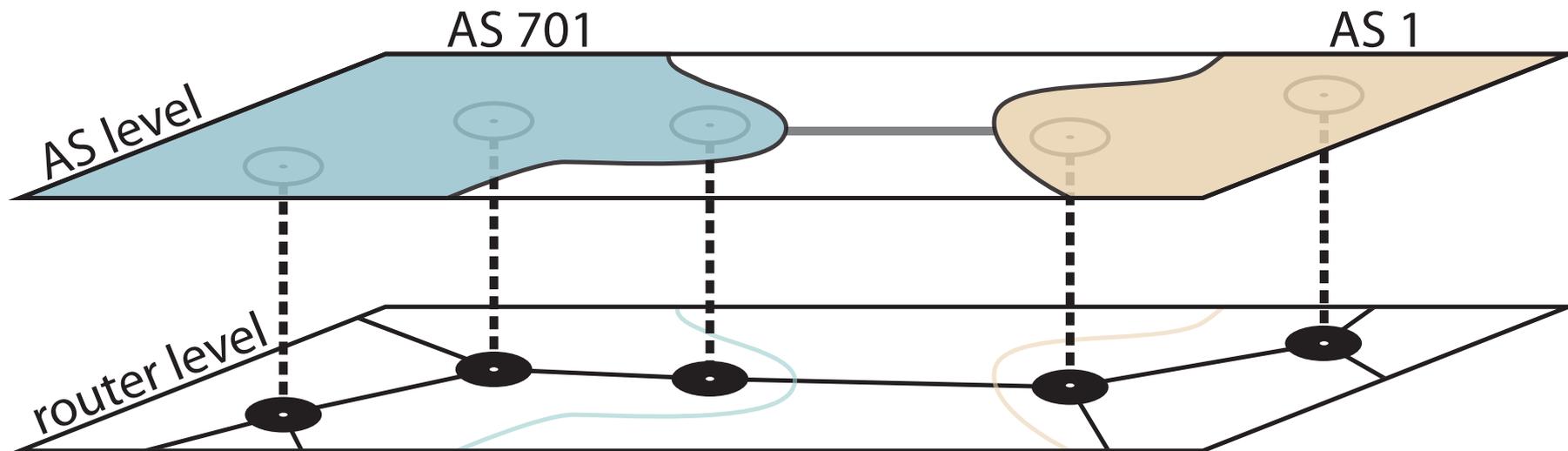


router	AS	AS Num.
■	②	2497
	④	4323
	①	1280
	④⑧	4826
	⑦	7575
	①①	11164
	②⑧	2828

Technical Approach: Phase II

Phase II: Development: Expand monitoring infrastructure and implement techniques investigated or designed in Phase I.

1. Create a **new** series of Internet Topology Data Kits (ITDK)
 - a) Install **new** and **upgrade** obsolete **Ark monitors**
 - b) Conduct **large scale alias resolution** probing runs every 3-6 months
 - c) **Analyze** data, derive topology graphs at **various levels of granularity**, update data and algorithms descriptions as necessary, share data

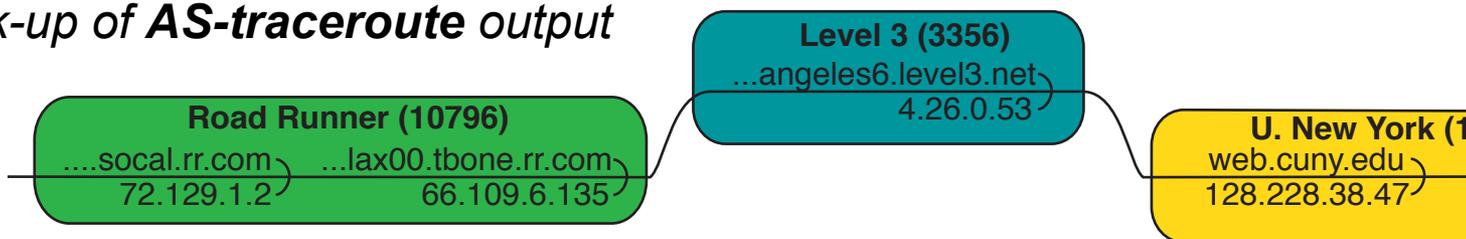


Technical Approach: Phase II

2. Develop interactive graphical interface to topology data
 - a) **Connectivity plus**: ownership structure, business relationships, geographic coverage, financial indicators
 - b) **Interactive** (GUI-based) corrections of (PoP,city,etc) inferences
 - c) **Develop validation support** for multiple levels of inference

3. Implement on-demand topology measurement tools
 - a) **Support** queries regarding **regional** performance dynamics
 - b) **Support** real-time **probing** destinations by **country**, **AS**, **prefix**, **org**
 - c) Provide a public **AS-traceroute** measurement tool

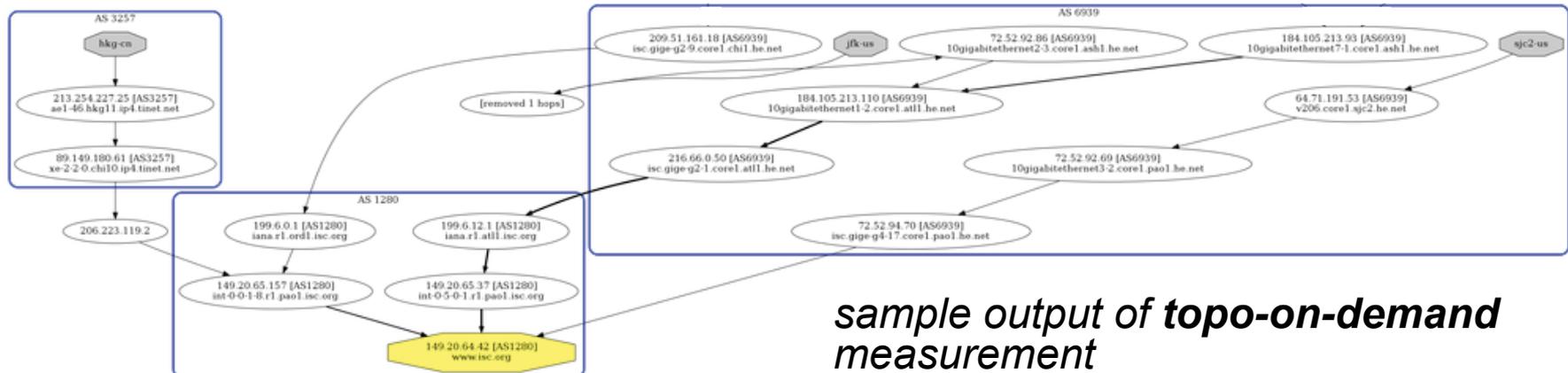
mock-up of AS-traceroute output



Technical Approach: Phase III

Phase III: Demonstration: Use developed technology to execute real-time delivery of rich cybersecurity-relevant knowledge to DHS.

1. Continue to **expand Ark** platform
 - a) Deploy remaining **monitors**
 - b) Re-assess optimal **configuration** for team probing
2. Demonstrate **capabilities** of the platform and technologies developed
3. Test “**topo-on-demand**” measurements in an operational environment
 - a) Support querying of **historical topology measurements**
 - b) Integrate diverse **sources** of data into coherent multi-level **map**

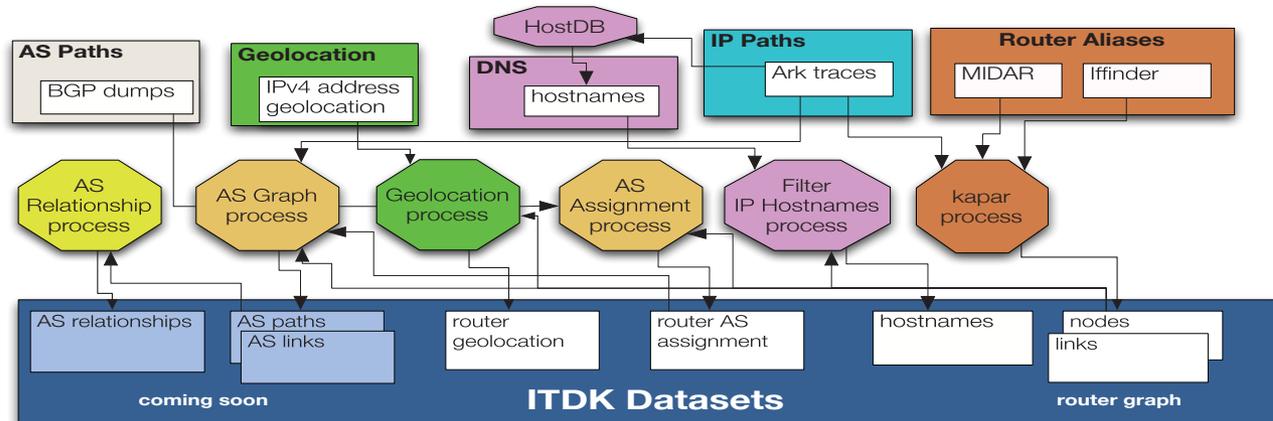


Technical Approach: Testing and Evaluation

Testing and Evaluation

1. **Large scale** measurement experiments on Ark
2. Four stages of measurement and analysis to **construct** a **router-level graph** (Estimation, Discover, Elimination, and Corroboration)
3. **Process** yields comprehensive multi-layer view of connectivity, **heavily curated** into an Internet Topology Data Kit (**ITDK**)
4. **Compare** and analyze standard **statistics** on resulting **topologies**
5. **Validate against** all available sources of **ground truth**
6. Rich user community: **825 user accounts** for topology data (sept12)

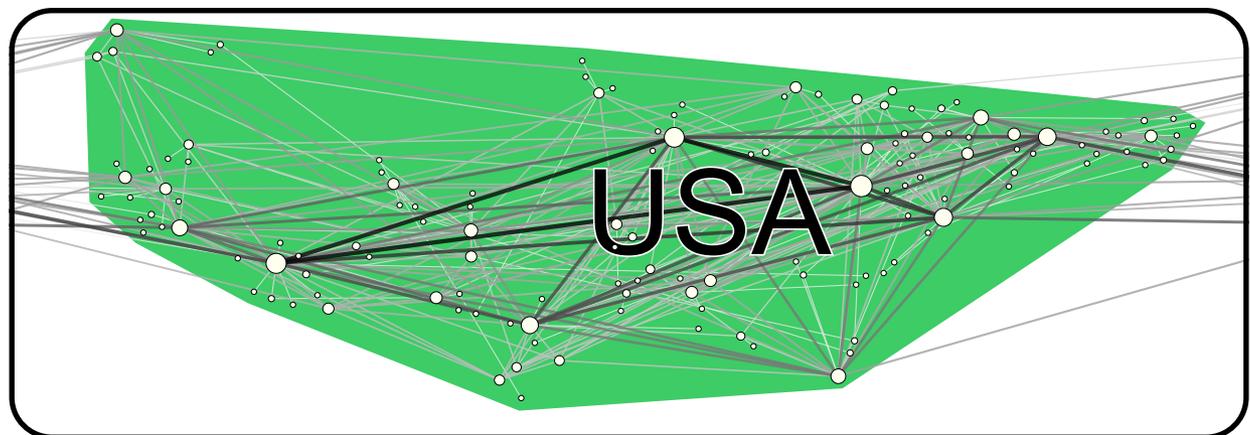
ITDK curation process



Milestones: Phase I (App.Res.)

- a) Upgrade alias resolution software to include additional IP address lists
- b) Beta-release of interactive PoP/city-level map validation functionality
- c) Test algorithms to classify AS peering links, including SFP links
- d) Deploy 8-10 Ark monitors (new locations or upgrades)
- e) Release economic annotation functionality for AS-level graphs
- f) Intermediate level infrastructural annotations
- g) Evaluate Ark-derived topology augmented w/ additional reachability data
- h) Extract data and analysis for specific regions of interest around the world

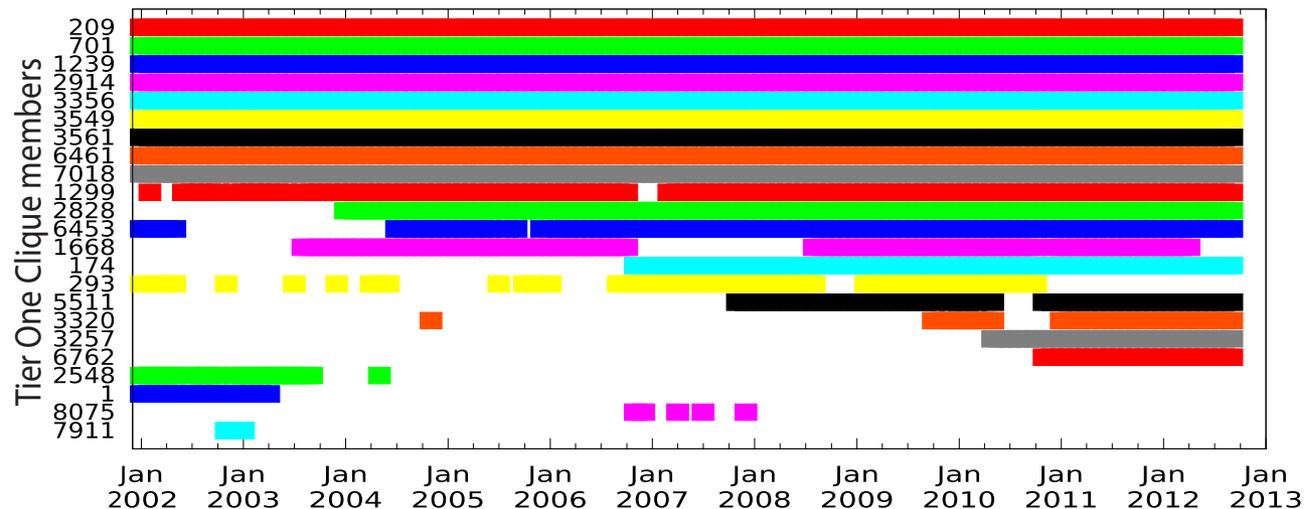
OX's (AS 2828)
USA PoP level graph



Milestones: Phase II (Dev.)

- a) Deploy 8-10 Ark monitors (new locations of upgrades)
- b) Create expanded ITDK (w/newly derived annotations)
- c) Develop query interface to historical & recent topology data
- d) Beta support for interactive user correction of AS meta-data
- e) Prototype AS-traceroute tool
- f) Prototype “topo-on-demand” measurements and interface

*Top Tier ASes
clique members
over time*



Milestones: Phase III (Demo)

- a) Test developed capabilities in realistic environment
- b) Deploy remaining monitors
- c) Upgrade AS Rank system

Organization Ranking by number of customers

AS Ranking | **Org Ranking** | Information for a single AS | Information for a single Org | Background | Data Sources | Help | [Org Ranking Help](#)

The top Organizations ranked by customer cone size are displayed below. Dataset: 2012.06.01

For information about a specific Org, enter its name:

Table shows 10 of 43174 Orgs. **sorted by** number of ASes in customer cone

Org rank	Org name	Num. ASes	customer cone						AS degree	Org degree
			Number of			Percentages of all				
			ASes	IPv4 prefixes	IPv4 addresses	ASes	IPv4 Prefixes	IPv4 Addresses		
1	Level 3 Communications	18	31,141	330,943	1,770,445,516	75%	79%	69%	4,375	3,872
2	TeliaNet Global Network	5	16,073	175,312	759,252,039	39%	41%	29%	742	662
3	Tinet SpA	2	14,595	179,793	794,176,773	35%	43%	31%	884	798
4	Cogent/PSI	3	14,060	164,075	667,755,727	34%	39%	26%	3,535	3,267
5	NTT America, Inc.	7	11,697	160,793	757,129,442	28%	38%	29%	809	703
6	TATA Communications	6	8,528	131,962	518,890,651	20%	31%	20%	842	777
7	Sprint	16	7,844	129,977	965,836,541	19%	31%	37%	1,023	906
8	TELECOM ITALIA SPARKLE S.p.A.	3	7,569	105,987	379,857,945	18%	25%	14%	262	236
9	Qwest Communications Company, LLC	11	7,253	122,871	759,408,992	17%	29%	29%	1,754	1,596
10	Init7 Global Backbone	424	6,744	72,220	395,631,197	16%	17%	15%	2,564	2,305

data sources

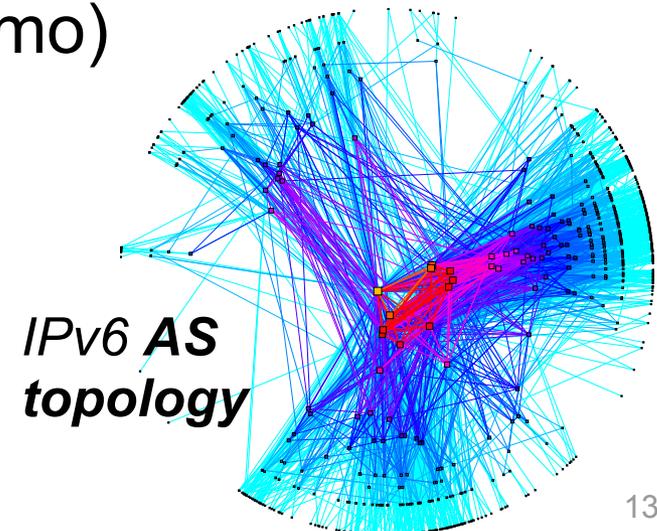
geolocation	database	2012.06.25	netacuity
organization	whois	0000.00.00	JPNIC, KRNIC, LACNIC
		2012.06.29	AFRINIC, APNIC, ARIN, LACNIC, RIPE
topology	BGP	2012.06.01, 2012.06.02, 2012.06.03, 2012.06.04, 2012.06.05	ripe rrc00, rrc01, rrc03, rrc04, rrc05, rrc06, rrc07, rrc10, rrc11, rrc12, rrc13, rrc14, rrc15
			routeviews eqix, isc, linx, routeviews2, saopaulo, sydney

Support for this work is provided by the U.S. Department of Homeland Security's Science and Technology Directorate (Project N66001-09-C-2029), the National Science Foundation Internet Laboratory for Empirical Network Science (Project CNS-0958547), and Cisco's University Research Program.

Cooperative Association for Internet Data Analysis | Based at the University of California's San Diego Supercomputer Center

Deliverables

- 1) raw topology data (all phases, continually)
- 2) support validation of (PoP/city) map inferences (15 mo)
- 3) annotated Internet topology map (18 mo)
- 4) Internet Topology Data Kits (18mo, 24mo, 30 mo)
- 5) query interface to topology database (30 mo)
- 6) *topo-on-demand* measurement functionality (30 mo)
- 7) upgraded AS Rank system (35 mo)
- 8) monthly and final reporting



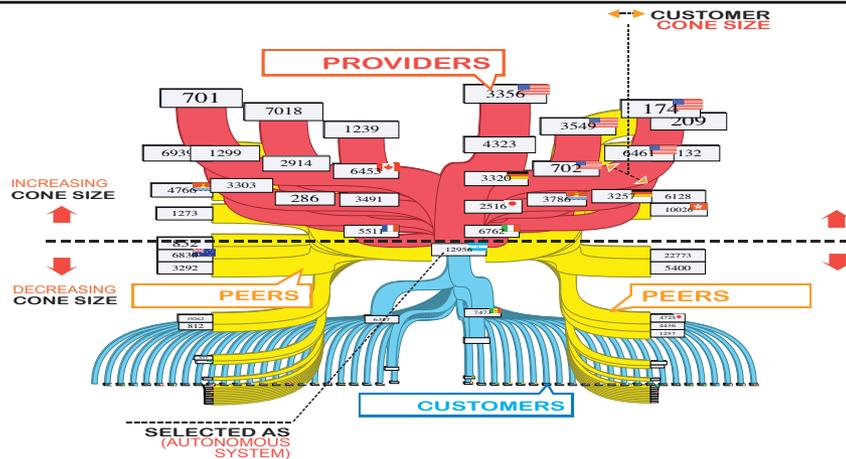
Technology Transition

- Use **PREDICT** to **share collected** topology data (7 Terabytes so far) w/academic researchers (90% of current users), CAIDA membership program to share data with industrial partners
- **Software** generally **open source** (UCSD or GPL)
- Participate in **advisory capacity** with **other agencies** (FCC, NSA, NIST) on how data can be used to support their missions
- Regular **workshops** and **engagement with operators** (NANOG, RIPE) to present results and solicit feedback

Quad Chart

BAA Number: Cybersecurity (11-02) TTA: #7
 Title: Cartographic Capabilities for Critical Cyberinfrastructure

Kimberly Claffy
 7 July 2011



Operational Capability

1. Expand measurement platform; ongoing Internet topology probing; periodic Internet Topology Data Kits (ITDK); interactive query interface to data; on-demand topology measurements; contribute resulting data to DHS S&T PREDICT.
2. Deploy 20+ additional monitors; collect, curate, analyze, and aggregate 2+TB/year topology data.
3. Estimated system maintenance cost: \$300K/year.
4. Results will improve our ability to identify, monitor, and model critical infrastructure, directly targeting goals of TTA #7.

Technical Approach

1. Develop measurement technologies and interdisciplinary analysis capabilities to execute timely delivery of richly annotated maps of critical Internet resources, both at physical and logical levels.
2. Increase completeness, accuracy and richness of Internet maps at multiple granularities; develop interactive interface to topology data; implement on-demand topology measurement capability.
3. Conduct ongoing global Internet topology measurements, provide annotated topology data, and support interactive AS ranking.
4. Current measurement platform, Ark, consists of 61 monitors across 29 countries and 6 continents. We have prototyped interactive validation functionality, and released preliminary annotated ITDKs.
5. Previous effort funded via BAA 07-09. Prototyping topo-on-demand capability and packaging analysis tools for release.

Schedule, Cost, Deliverables, & Contact Info:

Project starts 1 Oct 2012, Applied Research ends 31 Mar 2013.
 Development Phase ends 31 Mar 2014. Technology Demo Phase (optional) ends 30 Sept 2014 (Total 36 mo).

Deliverables: Internet topology datasets; topo-on-demand measurement tools; interactive validation and query interfaces; upgraded AS ranking service; monthly and final report.

POC: Sandra C, UCSD Contracts&Grants, 9500 Gilman Dr. MC 0934, La Jolla, CA 92093-0934 FAX 858-534-0280.