REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
01-Jul-2004	Quarterly	Jul 2002 - Sep 2002

4. TITLE AND SUBTITLE

Macroscopic Internet Data Collection and Analysis in Support of the NMS Community:

Progress Status and Management Report
for the period Jul 2002 - Sep 2002

5a. CONTRACT NUMBER
AO: L018/00

5b. GRANT NUMBER
N66001-01-1-8909

5c. PROGRAM ELEMENT NUMBER
0602301E

6. AUTHOR(S)

Dr. K. C. Claffy

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
DI-MGMT-80227

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Cooperative Association for Internet Data Analysis (CAIDA)
University of California, San Diego (UCSD)
9500 Gilman Drive MS-0934
La Jolla, CA 92093-0934

8. PERFORMING ORGANIZATION REPORT NUMBER
CAIDA-NMS-06-Qtr#5

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
SPAWARSYSCEN SAN DIEGO
CODE D212
BLDG A33 ROOM 1602W
53560 HULL STREET
SAN DIEGO CA 92152-5001

10. SPONSOR/MONITOR'S ACRONYM(S)
DARPA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This UCSD/CAIDA project focuses on advancing the capacity to monitor, depict, and predict traffic behavior on current and advanced networks, through developing and deploying tools to better engineer and operate networks and to identify traffic anomalies in real time. CAIDA will concentrate efforts in the development of tools to automate the discovery and visualization of Internet topology and peering relationships, monitor and analyze Internet traffic behavior on high speed links, detect and control resource use (security), and provide for storage and analysis of data collected in aforementioned efforts.

15. SUBJECT TERMS
network monitoring, network topology, network performance, network workload characterization, routing, CoralReef, skitter, OC48, Gigabit Ethernet, Domain Name Service (DNS), AS, Denial-of-Service, Code Red, Nimda

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			K.C. Claffy, Ph.D.
U	U	U	UU	11	19b. TELEPHONE NUMBER *(Include area code)* (858)534-8333

NMS Project Quarterly Report #Qtr5: 1-Jul-02 through 30-Sep-02

SUBMITTED TO Receiving Officer
SPAWARSYSCEN - SAN DIEGO
e-mail address: nms@spawar.navy.mil

Nikhil Dave and Steve Fujii
Technical Representatives {nik, fujii}@spawar.navy.mil

Tiffany Townsend
Contract Specialist, Code 2211
SPAWARSYSCEN SAN DIEGO CONTRACTS D212
PHONE 619-553-5472
FAX 619-553-4464
tiffany.townsend@spawar.navy.mil

SUBMITTED BY
University of California, San Diego (UCSD)
9500 Gilman Drive
La Jolla, CA 92093-0505

Principal Investigator
Dr. Kimberly Claffy
PHONE 858-534-8333
FAX 858-822-0861
kc@caida.org

Contract/Financial Contact
Pamela J. Alexander
PHONE 858-534-0240
FAX 858-534-0280
pjalexander@ucsd.edu

Quarterly Status Report #Qtr5

Macroscopic Internet Data Collection and Analysis in Support of the NMS Community

1.0 Purpose of Report

This status report is the quarterly cooperative agreement report that summarizes the effort expended by the UCSD's Cooperative Association for Internet Data Analysis (CAIDA)

program in support of SPAWARSYSCEN-SAN DIEGO and DARPA on Agreement N66001-01-1-8909 during July - September 2002.

2.0 Project Members

UCSD hours:
PI: 120.40
CAIDA Senior Staff: 731
CAIDA Staff: 1229.80
Total Hours: **2,081.20**

3.0 Project Description

This UCSD/CAIDA project focuses on advancing the capacity to monitor, depict, and predict traffic behavior on current and advanced networks, through developing and deploying tools to better engineer and operate networks and to identify traffic anomalies in real time. CAIDA will concentrate efforts in the development of tools to automate the discovery and visualization of Internet topology and peering relationships, monitor and analyze Internet traffic behavior on high speed links, detect and control resource use (security), and provide for storage and analysis of data collected in aforementioned efforts.

4.0 Performance Against Plan

(Please note: Changes since the last reporting period are in boldface type, and links have been updated with new content.)

Status	**Task 1 Year 2 Milestones:**	**Notes**
Progress	Add 5 additional skitter source sites	**Done**
Ongoing	Add 5 workload monitor sites	Added web page for accessing NeTraMet measurements
Complete	Develop comprehensive website(s) for public availability of data	• root/gTLD DNS performance plots • skitter daily summaries • NMS project progress • CoralReef analysis of SDNAP

Status	**Task 2 Year 2 Milestones:**	**Notes**
Ongoing	Establish archive and interactive database for community access to skitter, mantra, routing, and	• research community collaborators • skitter daily summaries

	CoralReef data.	• Real-time workload characterization of SDNAP
Ongoing	Solicit community feedback regarding needed data types, formats, and dataset sizes.	**Discussions occurred at USENIX, RIPE and NCS.**
Ongoing	Work with the NMS community to design common experiments	**Ken Keys worked with Nikhil Dave to resolve a problem with crl_delay.**

5.0 Major Accomplishments and Results to Date

Task 1. Monitoring Task

A. Topology Measurement Using Active Probes

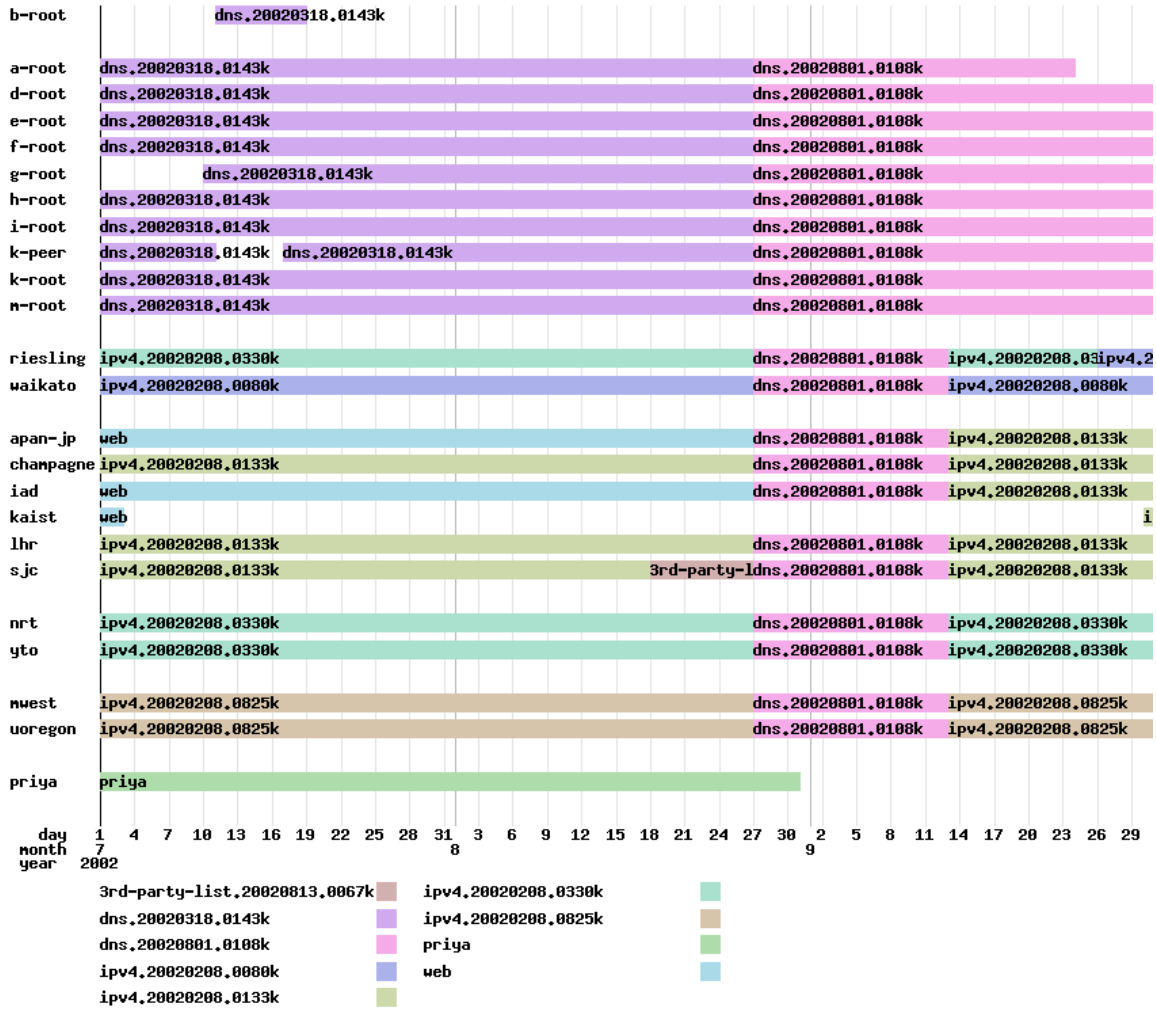
Approach

skitter is a CAIDA tool that measures both the forward path and round trip time (RTT) to a set of destination hosts by sending probe packets through the network. It does not require any configuration or cooperation from the remote sites on its target list. In order to reveal global IP topology, CAIDA's Macroscopic Topology Measurement and Mapping project builds software and infrastructure to:

- Collect forward path (layer 3) and RTT data
- Acquire infrastructure-wide global connectivity information
- Analyze the visibility and frequency of IP routing changes
- Visualize network-wide IP connectivity

An essential design goal of skitter is to execute its pervasive measurement while placing minimal load on the infrastructure and upon final destination hosts. To achieve this goal, skitter packets are small (52 bytes in length), and we restrict the frequency of probing to 1 packet every 2 minutes per destination and 300 packets per second to all destinations. To improve the accuracy of its round trip time calculations, CAIDA added a kernel module to the FreeBSD operating system platform used by its skitter monitors. Kernel timestamping does not solve the synchronization issue required for one-way measurements, but reduces variance caused by multitasking processing when taking round trip measurements. This feature helps to capture performance variations across the infrastructure more effectively. By comparing data from various sources, we can identify points of congestion and performance degradation or areas for potential improvements in the infrastructure.

skitter Monitor Status as of 31-Mar-02 (24 monitors active):



A. Topology Analysis Results:

CAIDA published "Distance Metrics in the Internet" in which we compared four Internet distance metrics and analyzed the predictive power of these metrics in selecting, from a given source, the lowest latency destination from among a candidate set. The four metrics are: IP path length; autonomous system (AS) path length; great circle geographic distance; and previously measured round trip time (RTT). We describe general properties of these four metrics and, using an unprecedented volume of real Internet macroscopic topology and RTT data, compare their correlation with actual RTT to the destination. The new methodology we propose for testing different metrics is suitable for evaluating new distance estimation techniques as they become available. This paper is publicly available at http://www.caida.org/outreach/papers/2002/Distance/

B. Workload / Performance Measurement Using Passive Monitors

One OC48 trace was successfully captured from the Metromedia Fiber Network (MFN) backbone in San Jose, CA. We collected one 7 hour trace on August 14, 2002.

During August, a project occurred to collect simultaneous measurements for data analysis correlation. Measurements included iffinder on the SJC skitter box; an OC-12 DAG trace at the UCSD/SDSC external gateway; and tcpdumps on all DNS root skitter boxes and on the F-root logger machine.

DNS Analysis

CAIDA began working with F-root operators to analyze a serious macroscopic DNS performance issue: RFC1918 (black-hole) DNS updates. RFC1918 private address space traffic by definition should remain local, yet root nameservers must waste significant time and resources processing these illegitimate DNS queries. CAIDA devised log file analysis techniques to identify the OS sending the illegitimate DNS queries, including code for analyzing unique DNS query hosts and domain names in PTR updates.
See http://www.caida.org/outreach/papers/2001/Rssac2001a/ and
http://www.caida.org/outreach/papers/2001/DNSPerfMeas/ for publications.

C. Routing Measurement

Generating a Map of Exchange Point Usage:

Skitter paths are represented by the IP addresses of the nodes that the packet visits in order to get to its destination. Since destinations may cross domains, the IP addresses can be mapped to Autonomous System numbers (ASNs), as is done currently in skitter dailies. However, skitter dailies are innately inaccurate when IP addresses in the paths belong to Exchange Points (XPs) where multiple ASes may come to peer. We are refining our methodology for generating daily skitter-based AS charts to account for these methodological idiosycracies in xchange point usage.

Our new methodology starts by modifying the prefix-to-ASN mapping that we gather from Routeviews (http://www.routeviews.org). We identify the prefixes that belong to an XP and map prefix-to-XP name. IP addresses in the skitter paths to the longest match can map to either an ASN or an XP. We then analyze XP usage. For example, we can infer that if AS1 sends a packet to XP_PAIX and XP_PAIX sends that packet to AS2, then AS1 is peering with AS2 at XP_PAIX. Once we have identified the paths using ASNs and XP names, we can break down the paths into FROM -> TO tuples and use Otter to plot those tuples. Finally, we use NetGeo to determine the geographical location of the exchange points. We recognize that the methologies are still primitive in terms of the empirical complexity of the Internet, and yet also the most advanced `state of the art' at this time.

Task 2, Archiving and Storage Task

Approach for Archiving skitter Data and Making Data Available to Researchers

Requestor	**Organization**	**Project**
Thinh Nguyen	UC Berkeley	Routing for video streaming
Weixiong Rao	Shanghai Jiaotong University, China	P2P routing
Abhinav Kothari	K.K. WAGH C.O.E Hirabhai Haridas Vidyanagari Amrutdham, Panchavati, INDIA	Geographic host mapping
Stefan Bender	Saarland University, Saarbrucken, Germany	Host distance estimation
Samson Lee	University of Technology, Sydney, Australia	Policy based network management
Phil Lowden	North Carolina State Univerisity	Compare cost-based vs. policy-based routing.
Other collaborative projects at: http://www.caida.org/projects/nms/reports/skitter_comuse.xml		
Previous collaborative projects at: http://www.caida.org/projects/nms/reports/prev_skitter_comuse.xml		
PhD students using skitter data		7
Master's students using skitter data		5
About publicly available skitter data: http://www.caida.org/cgi-bin/skitter_summary/main.pl		

Approach for Archiving CoralReef Data

1. CAIDA maintains a SDNAP report generator, publishing workload characterization results at http://www.caida.org/dynamic/analysis/workload/sdnap/0_0_/. Results are updated every 5 minutes.
2. CAIDA archives CoralReef data for special purpose studies as needed, but must limit data collection to available disk space.

6.0 Artifacts Developed During the Past Quarter

None

7.0 Issues

We continue to operate at a significant deficit based on notification of Year 2 funds. We received $352,875 in April and a $150,000 conadd in June. Discussions with SPAWAR

and DARPA revealed that the DARPA funds used to fund Task 1 were allocated only for Year 1, CAIDA is unable to complete Year 2, Task 2 or the proposed conadd work without Task 1. Current discussions concern possible changes to the SOW. In August, CAIDA prepared a new SOW for Yr 3 (Oct 1 – Sep 30, 2003) to work within a budget amount of $351,316.

8.0 Near-term Plan

The following work is planned for 01-Oct-02 through 31-Dec-02:

General/Administrative Outreach and Reporting Plans

- Submit Quarterly Report to SPAWAR covering progress, status and management.

Task 1. Monitoring Task Plans

- **A. Topology Measurement**

  - CAIDA will continue to collect and analyze data for its macroscopic topology project.

- **B. Workload Measurement**

  - CAIDA will continue to analyze traces gathered from OC48 links at Metromedia Fiber Network (MFN) in San Jose. We are trying to find other locations for OC48 traffic taps, but need additional funding for that.

  - Refinement of the CoralReef software suite will continue, (http://www.caida.org/tools/measurement/coralreef/) .

- **C. Routing Measurement**

CAIDA will continue to refine methodology and results from ongoing routing studies.

Task 2, Archiving and Storage Task Plans

- We will continue to collect and analyze data collected from skitter sources deployed in the field.

- We will continue to make skitter topology and performance data available to researchers via password authentication for use in their research and monitor results. See: http://www.caida.org/projects/nms/reports/skitter_comuse.xml

- We will continue briefings to the Internet community on the purpose and results of skitter active monitoring and will solicit their feedback.

- We will refine the collection and archiving of skitter data

9.0 Completed Travel

The following travel incurred expenses to this award and occurred during Year 2, Qtr 1, 1-Jul-02 through 30-Sep-02:

- David Moore 8/7 – 8/8 San Francisco – Usenix Security Conference

Other related travel occurred but was not charged to this award.

This quarter we moved $2,396 of international travel expenses off of this grant. Nevil Brownlee's trip 4/27/02 – 6/21/02 to attend a RIPE meeting involved travel from Auckland, New Zealand to Boston to Amsterdam, but we inadvertently did not request or receive the required prior permission.

10.0 Equipment Purchases and Description

Three FreeBSD systems and one workstation monitor were purchased.

11.0 Significant Events

- Kc claffy pursued discussions with John Todd of the National Communications System (NCS), an agency wanting to fund tools and data for modeling Internet health. CAIDA was asked to investigate whether NCS might be able to fund Option 1 and/or Option 2 of this effort.
- Ken Keys communicated with Lucent developers concerning issues he discovered while testing the Lucent Optistar gigE interface.
- Ken Keys rewrote crl_oneway (crl_delay) to track TCP ACK delays and statistics in addition to one way delay statistics. Ken also corresponded with Ron Nolte of SAIC regarding design refinement options.
- Kc claffy provided 3, 6, and 12 month milestones (in five possible research areas: I. Passive monitoring, II. Infrastructure protection (homeland security), III. Active monitoring; IV. Routing/toplogy analysis and modeling: V. Internet Spectroscopy) to David Nicol (for presentation to Sri Kumar) for meeting NMS program performance and scalability goals.
- CAIDA provided source code for its walrus 3D directed graph visualizer to Dr. John Poindexter, Director of the IAO office at DARPA, who is interested in using walrus in IAO's Total Info Awareness Program,
- Kc claffy corresponded with Sri Kumar to answer his questions about CAIDA's Walrus 3D image illustrating Code-Red worm infections within prefix 24.0.0.0/8 on July 19, 2001.
- CAIDA provided Anup Ghosh of DARPA with a walrus 3D visualization of the CodeRed worm for use in a DARPA presentation.

12.0 Publications and Presentations:

1. The following papers were published:
   - B. Huffaker, M. Fomenkov, D. Plummer, D. Moore, and k. claffy, ``Distance Metrics in the Internet,", in IEEE International Telecommunications Symposium (ITS), Brazil, Sept 2002, IEEE.
   - N. Brownlee and k. claffy, ``Understanding Internet Traffic Streams: Dragonflies and Tortoises,", IEEE Communications, Jul 2002.
2. The following presentations were given:
   a. Kc claffy presented "Internet measurements: myths about Internet Data" presented July 2002 at SDSC. http://www.caida.org/outreach/presentations/2003/lsn20030610/
   b. David Moore presented "Network Telescopes: Observing Small or Distant Security Events" August 8, 2002 at the USENIX Security conference in San Francisco. http://www.caida.org/outreach/presentations/2002/usenix_sec/
   c. Kc claffy presented "measurement and analysis of the root DNS system: update" to the DNS working group at the 43rd RIPE meeting September 9-13, 2002 in Rhodes, Greece. http://www.caida.org/outreach/presentations/dns0209/
   d. Brad Huffaker and Marina Fomenkov presented "Distance Metrics in the Internet" September 8-12, 2002 in Natal, Brazil. http://www.caida.org/outreach/presentations/2002/Distance/
3. CAIDA prepared and submitted a Technical Report, Financial Report and Quad Chart for the IPTO 2002 Project Summary Collection.
4. CAIDA prepared a CAIDA NMS project summary slide in support of an interim Director's review of NMS.

13.0 FINANCIAL INFORMATION:

Contract #: N66001-01-1-8909

Contract Period of Performance: 5 Jun 2001 to 5 Jun 2004

Ceiling Value: $ 1,726,160

Current Obligated Funds: $1,726,160

Reporting Period: 1 Jul 2002 to 30 Sep 2002

Actual Costs Incurred: $ 1,089,947

Current Period:

UCSD
Labor Hours: 2,081.20 $ 85,513
ODC's: $ 11,534

IDC's: $ 46,440
TOTAL: **$ 143,487**

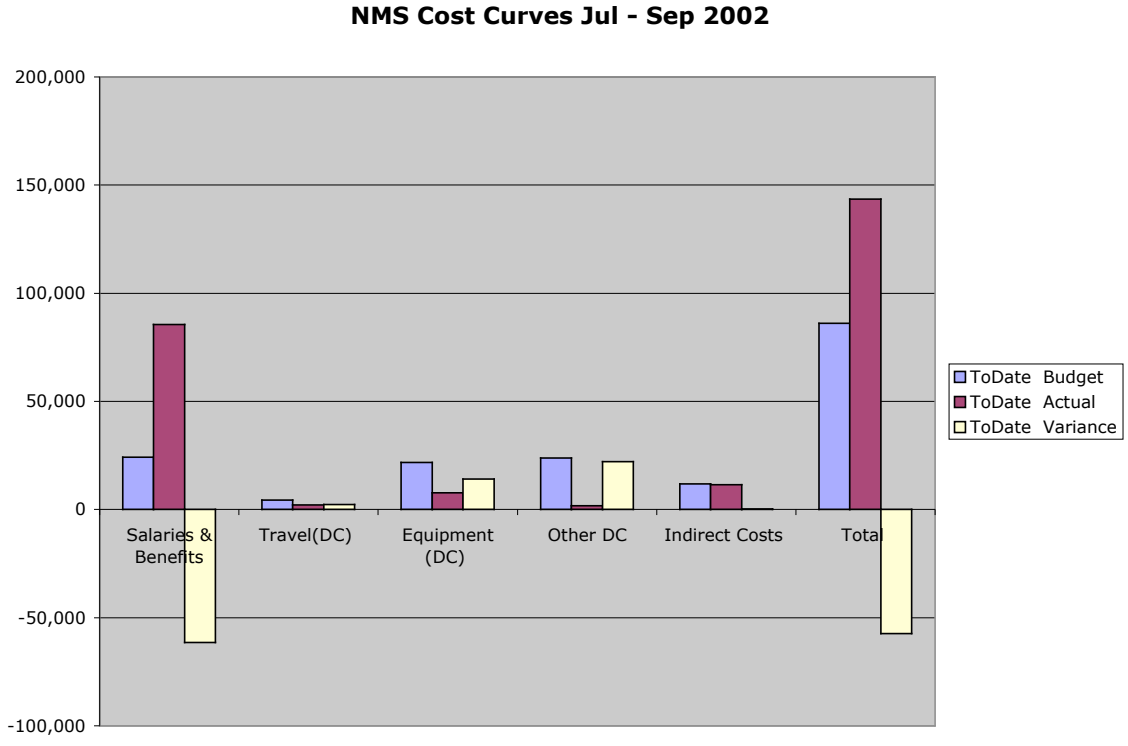
Cumulative to date:

Labor Hours: 18,191.97 $ 664,699
ODC's: $ 65,491
IDC's: $ 359,757
TOTAL: **$ 1,089,947**

Cost Curves for Jul - Sep 2002:

	ToDate Budget	**ToDate Actual**	**ToDate Variance**
Salaries & Benefits	24,107	85,513	-61,406
Travel(DC)	4,450	2,114	2,336
Equipment (DC)	21,791	7,738	14,053
Other DC	23,875	1,682	22,193
Indirect Costs	11,831	11,534	297
Total	86,054	143,487	**-57,433**

See Section 7.0 Issues for an explanation of the discrepancy between spending level and budget.


NMS Cost Curves Jul - Sep 2002
200,000
150,000
100,000
50,000
0
-50,000
-100,000
Salaries & Benefits
Travel(DC)
Equipment (DC)
Other DC
Indirect Costs
Total
ToDate Budget
ToDate Actual
ToDate Variance