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	Jan 2003 - Mar 2003

4. TITLE AND SUBTITLE

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

Progress Status and Management Report
for the period Jan 2003 - Mar 2003

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-08-Qtr#7

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	12	19b. TELEPHONE NUMBER *(Include area code)* (858)534-8333

NMS Project Quarterly Report #Qtr7: 1-Jan-03 through 31-Mar-03

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 #Qtr7

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 January – March 2003.

2.0 Project Members

UCSD hours:
PI: 137.60
CAIDA Senior Staff: 344
CAIDA Staff: n
Total Hours: **1100.80**

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

Status	Task	Notes
Cancelled	**Task 1 Year 3**	**Monitoring and archiving work now funded by NCS is reported here for continuity**
Cancelled	**Task 2 Year 3**	
	Task 3 Year 3 Milestones:	
Complete	Report on DNS damage; suggest protection strategies.	Reports given Nov 02 at NANOG and NMS PI Meeting
Ongoing	Report on validity of BIND name server affinity	**Additional experiments run and analysis published**
Complete	Provide initial model of DNS behavior for NMS Integration Prototype	DNS Query Workload for simulations derived from real DNS root server traffic
Begun	Document, package, and distribute passive tools (CoralReef and/or NeTraMet) and methods for their use to monitor the DNS infrastructure	**Updates to custom CoralReef application crl_delay released to Nikhil Dave**

5.0 Major Accomplishments and Results to Date

Task 1. Monitoring Task

A. Topology Measurement Using Active Probes

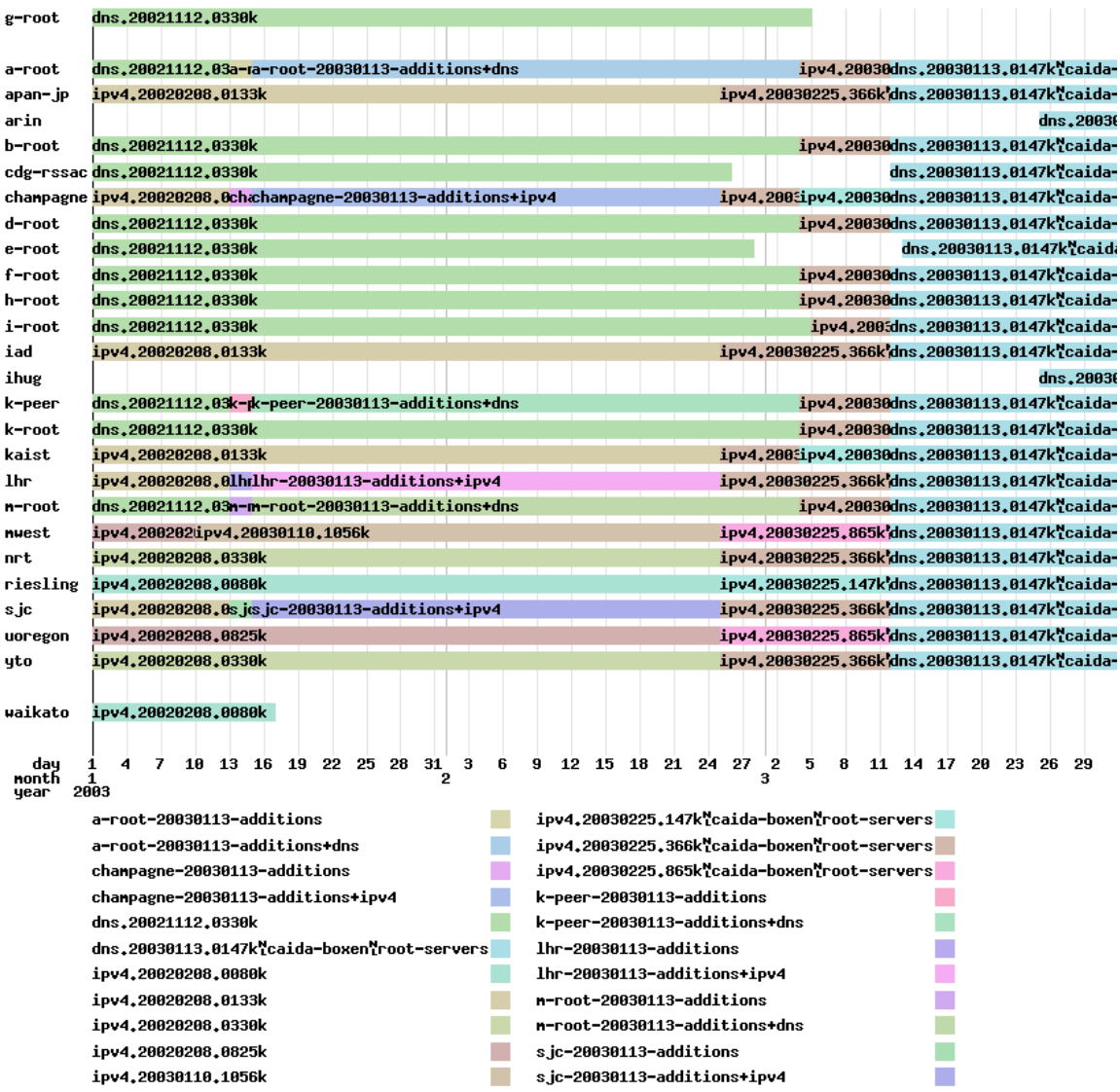
Approach

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-03 (26 monitors active):



B. Workload / Performance Measurement Using Passive Monitors

A one-hour OC48 trace was successfully captured from the Metromedia Fiber Network (MFN) backbone in San Jose, CA.

C. Routing Measurement

CAIDA released a set of router recommendations (See: http://www.caida.org/analysis/measurement/recommendations/routers.xml in response to a request from DARPA. These recommendations consider vendor/provider support in

providing information needed to support realistic Internet modeling and simulation.

Task 2, Archiving and Storage Task

Approach for Archiving skitter Data and Making Data Available to Researchers

Requestor	**Organization**	**Project**
Mark Handley	International Computer Science Institute	Evaluate alternative routing schemes for BGP
Aditya Namjoshi	University of Kentucky	BGP policies
Dr. William Semancik	DoD Laboratory for Telecommunications Sciences	Cyber defense mechanisms
Dawn Song	Carnegie Mellon U	IP traceback and DdoS defense
Adrian Perrig	Carnegie Mellon U	IP traceback and performance
Ma Tianbai	Chinese University of Hong Kong	Massive graphs and etworking
Kuai Xu	University of Minnesota	Exchange points
Gengxin Zhang	Queen Mary College, University of London	Simulations for Saveguard project
Ken Hui	Chinese University of Hong Kong	Topology studies
Dhiman Barman	Boston University	RTT distributions
David Fuhrmann	National Communications System (NCS)	Topology for Internet health
Derrick Kong	BBN (for DoD Laboratory for Telecommunications Sciences	Cyber defense mechanisms
Chou Kang-Hsien	National Chiao-Tung University, Distributed system and network Security lab	DDoS simulations
Christopher Kruegel	UC Santa Barbara	Intrusion detection
Kun Zhang	Georgia Institute of Technology	Exchange points
Karine Barzilai-Nahon	Tel-Aviv University	Internet access
Jing Huang	Institute of Computing technology, China	Correlate Internet metrics
Z. Morley Mao	UC Berkeley	AS mapping techniques
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		15
Master's students using skitter data		10

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.

Task 3. Domain Name System (DNS) Infrastructure Model

The Domain Name System (DNS) is a fundamental component of the modern Internet, providing a critical link between human users and Internet routing infrastructure by mapping host names to IP addresses. The DNS utilizes a hierarchical name space divided into zones that are distributed among the name servers. This hierarchy is manifested in the familiar "dots" structure. Each zone has one or more authoritative name servers. These are dedicated servers, whose job is to answer queries for names within their zone(s).

In order to reach a machine with the name "not.invisible.net', one must send a query to the DNS server responsible for machines and/or sub-domains in the domain *.invisible.net. The authoritative machine for *.invisible.net will be looked up by sending a query to the server authoritatively responsible for *.net. Such a server is called a global top-level domain (gTLD) server. Information on the appropriate gTLD can be obtained from one of the root servers. Currently there are 13 gTLD servers and 13 root servers.

CAIDA conducted a series of experiments comparing performance of different implementations of DNS caching nameserver software. Results will appear soon at: http://www.caida.org/outreach/papers/2003/dnspackets/ and in a poster "Modeling the Domain Name system (DNS).

The DNS Query Workload used in the simulation experiments consists of 7,507,544 hostnames derived from 24 hours of IRCache logs at a root name server. Filters on this real data removed invalid data (e.g. using IP address for hostname). Then we extract unique Second Level Domain (SLD) zones and valid unique Top Level Domain (TLD) zones. We keep hostnames with invalid TLDs to model error handling. This workload is played back in the simulator lab as fast as possible.

The first round of experiments compared three caching name server implementations:

a. BIND 9.2.2
b. DJBDNS 1.05 (with 100M cache)
c. Windows 2000 V5.0.49664

In Experiment 1, we simulated no delay and no query loss. Experiment 2 simulated no delay, but 10% loss. Experiment 3 also simulated no loss, but with linearly increasing delay.

6.0 Artifacts Developed During the Past Quarter

None.

7.0 Issues

None.

8.0 Near-term Plan

The following work is planned for 01-Apr-03 through 30-Jun-03:

General/Administrative Outreach and Reporting Plans

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

Task 3. DNS Analysis

Overall goals: Build a model of DNS behavior. Investigate whether the current design will scale to serve continued IP address space growth. Conduct controlled experiments to identify parameters crucial to proper DNS operation.

- Report on DNS damage from non-caching DNS clients or ill-formed (illegal) queries. Suggest strategies for protecting the DNS.
- Report on the validity of BIND name server affinity algorithm.
- Provide initial model of DNS behavior for NMS Integration Prototype
- Document, package, and distribute passive tools (CoralReef and/or NeTraMet) and methods for their use to monitor the DNS infrastructure.

9.0 Completed Travel

The following travel incurred expenses to this award and occurred during Year 2, Qtr 3, 1-Jan-03 through 31-Mar-03:

- Nevil Brownlee 3/14 – 3/20 San Francisco IETF56
- Andre Broido 3/28 – 3/31 San Francisco HSN 2003
- Duane Wessels 3/20 – 3/21 Colorado to San Diego – WIDE meeting
- Brad Huffaker 3/16 – 3/21 San Francisco, IETF

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

10.0 Equipment Purchases and Description

No equipment was purchased during this quarter.

11.0 Significant Events

- Ken Keys sent Nikhil Dave a revised version of crl_delay that tracks the moving window size that a given existing tcp connection is using - i.e., how many un-acked packets that the connection will allow the sender outstanding at a given packet release time. Nikhil is interested in knowing this parameter as a function of time to supplement the great info already coming from crl_delay.
- CAIDA provided some packet level traces to ISI for use in their RAMP tool to produce *ns* models. (<= 1 hour of packet trace with anonymized IP addresses from either UCSD, SDNAP, Auckland or some other campus level traffic monitor.)
- CAIDA provided flow summary data to GT to be used to model the background traffic not modeled by RAMP. (flow size distribution over a 24 hour period, measured in packets, bytes, and duration, from either UCSD, SDNAP Auckland, or some other campus level traffic monitor, averaged over each 5 minutes and including src-port and dst-port.)
- CAIDA also provided a data set from 2001 showing numbers of attacks as seen reflected in backscatter data from UCSD's network telescope.
- CAIDA submitted information about five CAIDA tools to the NMS Model Inventory. NMS can take partial not complete credit for *CoralReef, NeTraMet, dnstat* and *iffinder*. (CoralReef and NeTraMet are both huge projects going on for years, funded by several different grants. *dnstop* was funded by WIDE.
- CAIDA began negotiations with John Todd of the National Communications System to fund a mechanism to allow NCS to fund CAIDA activities in support of NCS goals.

12.0 Publications and Presentations:

1. The following papers were published:
    a. Y. Hyun, A. Broido, and k. claffy, ``Traceroute and BGP AS Path Incongruities,", Tech. rep., Cooperative Association for Internet Data Analysis (CAIDA), Mar 2003.
    b. A. Broido, R. King, E. Nemeth, and k. claffy, ``Radon Spectroscopy of Inter-Packet Delay,", in IEEE HSN 2003. March 2003, IEEE.
    c. D. Moore, V. Paxson, S. Savage, C. Shannon, S. Staniford, and N. Weaver, ``The Spread of the Sapphire/Slammer Worm,", Tech. rep., CAIDA, ICSI, Silicon Defense, UC Berkeley EECS and UC San Diego CSE, Jan 2003.
    d. R. Beverly and k. claffy, ``Wide-Area IP Multicast Traffic Characterization,", IEEE Network, vol. Jan/Feb 2003, Jan 2003.

2. The following presentations were given:

a. "bandwidth estimation: measurement methodologies and applications" (DOE ESnet, Feb '03)
b. "Understanding Global Internet Health" (UC Regents, Feb '03)
c. "Understanding Global Internet Health" (Jan '03)

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 Jan 2003 to 31 Mar 2003

Actual Costs Incurred: $ 1,216,105

Current Period:

UCSD
Labor Hours: 1100.80 $ 37,323
ODC's: $ 523
IDC's: $ 19,680
TOTAL: **$ 57,526**

Cumulative to date:

Labor Hours: 20,195.18 $ 735,119
ODC's: $ 78,070
IDC's: $ 402,916
TOTAL: **$ 1,216,105**

This revision of last quarter's cost curves reflects budgeting against the actual funds received instead of the budget plan for the total awarded amount.

Cost Curves for Jan - Mar 2003:

	ToDate Budget	**ToDate Actual**	**ToDate Variance**
Salaries & Benefits	113,629	37,323	76,306

Benefits			
Travel(DC)	4,753	75	4,678
Equipment (DC)	39,053	0	39,053
Other DC	14,270	448	13,822
Indirect Costs	53,546	19,680	33,866
Total	225,251	57,526	**167,725**