REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)* 01-Jul-2004	**2. REPORT TYPE** Quarterly	**3. DATES COVERED** *(From - To)* Apr 2003 - Jun 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 Apr 2003 - Jun 2003

5a. CONTRACT NUMBER
AO: L018/00

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

5c. PROGRAM ELEMENT NUMBER
0602301E

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
DI-MGMT-80227

6. AUTHOR(S)

Dr. K. C. Claffy

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-09-Qtr#8

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	8	**19b. TELEPHONE NUMBER** *(Include area code)* (858)534-8333

NMS Project Quarterly Report #Qtr8: 1-Apr-03 through 30-Jun-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 #Qtr8

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 April - June 2003.

2.0 Project Members

UCSD hours:
PI: 129.00
CAIDA Senior Staff: 387.00
CAIDA Staff: 481.60
Total Hours: **997.60**

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 at PAM2003 and WIDE meetings.**
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

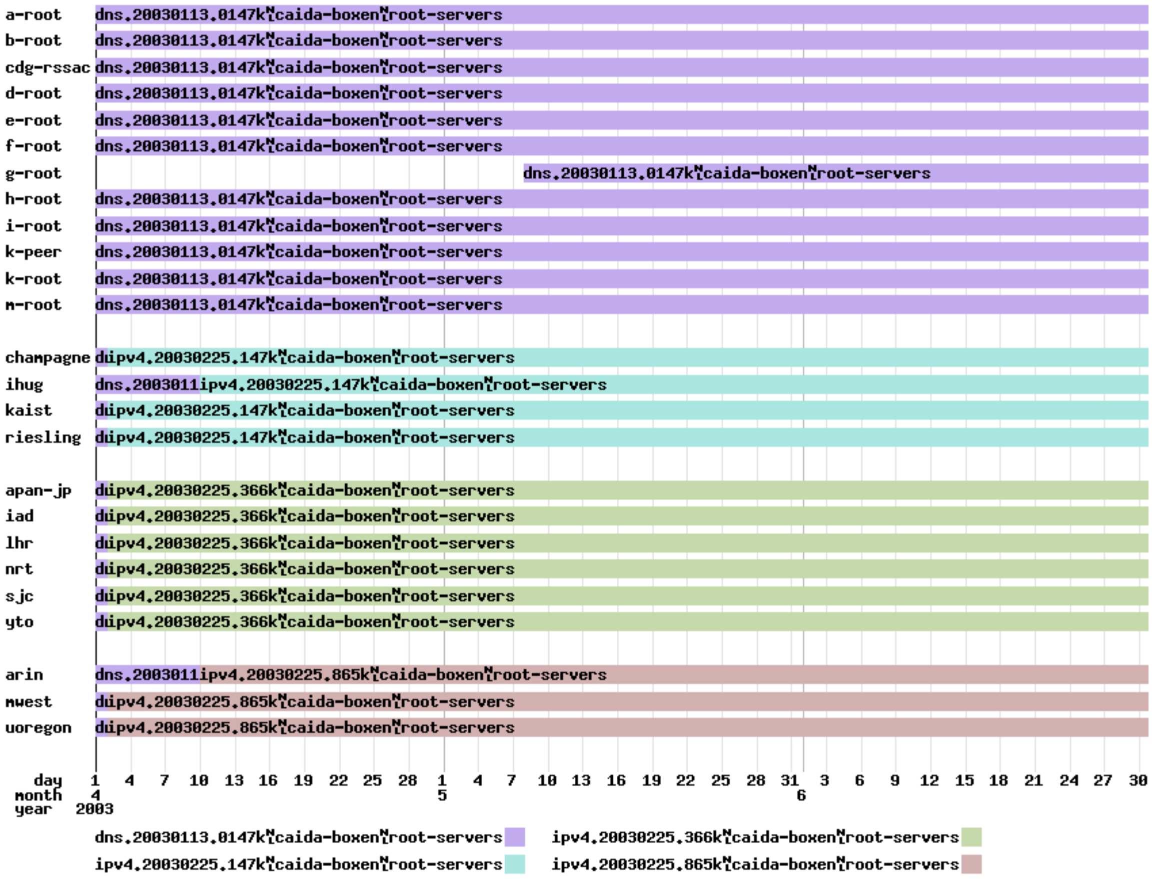
Approach

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 30-Jun-03 (25 monitors active):



Topology Analysis Results:

New Tools: Two new topology analysis tools were released this quarter: dnsstat, and iffinder.

dnsstat:

CAIDA released a new tool dnstat: http://www.caida.org/tools/utilities/dnsstat/

```
Usage:  ./crl_dnsstat [-C'<coral_command>']... [-p<len>] [-N<max>] [-anSDQhru] <source>...
```

-p<len>	aggregate hosts by CIDR prefix length <len> (default: 32)
-a	resolve IP addresses to hostnames (requires -p32)
-n	print DNS code numbers, not symbols
-S	ignore IP address of client (query Source)
-D	ignore IP address of server (query Destination)
-Q	don't count by query opcode/class/type
-h	print in more human-friendly format
-r	do not count msgs with RD set
-c	do not correct byte-swapped QDCOUNT
-dq	record requests only (default: requests and responses)
-dr	record responses only (default: requests and responses)
-b	write results in binary
-i	read binary input from dnsstat -b on stdin
-u	print contents of unusual msgs to stderr
-N<max>	keep data for at most <max> src/dst/proto entries
-s	print information on hash table usage

Improvements to software CoralReed include: possible values of <coral_command>, with default values in parentheses:

```
   version                          {config|f}=<filename>
   {iomode|m}=<mode_options>              source=<filename>[,<mode_options>]
   proto=[<subif>=]<proto>              allow=<subif>[=<proto>]
   deny=<subif>[=<proto>]               filter=<expr>
   p[ackets]=<N>  (0)                 d[uration]=<seconds>  (0)
   i[nterval]=<seconds>  (300.000000)     v[erbosity]=<level>  (1)
   norm[alize]={0|1}  (1)               comment=<string>
   e[rrfile]=<filename>                 sort  (do not sort)
   ignore_time_err  (abort on time error)
```

<mode_options> is a comma-separated list of any of:

```
   [first=]<N>     [!]last       [!]user       [!]ctrl
   [!]idle        [!]all        [!]echo       fw=<name>
   proto=[<subif>=]<proto>
   allow=<subif>[=<proto>]
   deny=<subif>[=<proto>]
   (default: first=554,rx)
```

iffinder:

CAIDA released a new tool “iffinder”: http://www.caida.org/tools/measurement/iffinder/. Iffinder discovers which interfaces (IP addresses) reside on the same router. Command line syntax offers the following parameters:

Usage: ./iffinder [options] <ip_file>	
Options (with defaults in brackets):	
-v<V>	verbosity <V> [1]

-w<W>	wait <W> seconds for response [10]
-c<C>	send at most <C> concurrent probes [30]
-p<P>	use dest ports in the range <P> to <P>+<C>*3-1 [33434]
-r<R>	send at most <R> probes per second [300]
-n<N>	probe each address in the list <N> times [1]
-N	don't probe newly discovered addresses
-R	use RECORD ROUTE option
-T	use RFC 1393 TRACEROUTE option (not widely supported)
-M	probe the /30-mate of every address

B. Workload / Performance Measurement Using Passive Monitors

OC48 Traces were successfully captured from the Metromedia Fiber Network (MFN) backbone in San Jose, CA and from the PAIX exchange point in San Jose, CA.

Analysis Results: Workload Characterization of an OC48 Link

OC48 Data from MFN, April – June 2003

We collected three traces from MFN during this period (April 24: 2 hours; May 7: 24 hours; May 8: 24 hours).

OC48 Data from PAIX, April – June 2003

We collected one trace from PAIX during this period (May 7: 1.25 hours).

C. Routing Measurement

No routing results to report during this reporting period.

Task 2. Archiving and Storage Task

Approach for Archiving skitter Data and Making Data Available to Researchers

Requestor	**Organization**	**Project**
Xiaoming Zhou	Delft University of Technology	Statistical analysis of Internet
Luciano Capitanio	Libera Universite maria Santissima Assunta, Italy	Internet visualization
Shi Zhou	Queen Mary, University of London	Internet topological properties
Feng Zhou	UC Berkeley	Overlay networks
Vukadinovic danica	ETHZ (Swiss Federal Institute of	AS and BGP

	Technology)	policy analysis
Prasanna Ganesan	Stanford University	Overlay networks
Petr Fiedler	Brno University of Technology – UAMT FEKT	Non-deterministic networks
Hao Chen	Simon Fraser University	Hybrid network model
Michael Collins	CERT and Carnegie Mellon University	DoS detection and filtering
Filipo Radicchi	Tor Vergata University	Network structure
Ferdo Ivanek	Stanford University	Internet graphs
Anukool Lakhina	Boston University	Topology studies
Denilson Martins	Federal University of Rio de Janeiro	IP traceback study
Juan Rodriguez Hervela	University Carlos III of Madrid, Spain	BGP behavior
David Ellis	Sussex University, UK	IP traceback
Edmund Hibbert	New jersey Institute of Technology	Sampling Bias
Ji Li	MIT	Region project
Apu Kapadia	UIUC	Mist routing infrastructure
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		18
Master's students using skitter data		37
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

CAIDA published three DNS analysis papers during this reporting period:

In "Wow, that's a lot of packets" (presented at PAM2003) we characterize DNS clients that send large numbers of queries to root DNS servers. Analysis of trace data from the two F root servers shows a number of interesting characteristics. Many root server clients send an excessive number of packets. We describe a few of the busiest sources in detail.

After classifying each query from the trace, we find that a very small percentage of the total traffic is legitimate.

In “Spectroscopy of DNS Update Traffic” we study attempts to dynamically update DNS records for private (RFC1918) addresses, by analyzing the frequency spectrum of updates observed at an authoritative name- server for these addresses. Using a discrete autocorrelation algorithm we found that updates series have periods of 60 or 75 minutes, which we identied as default settings of out- of-the-box Microsoft Windows 2000 and XP DNS software.

A variant of “Spectroscopy of DNS Update Traffic” was presented at the Workshop for Internet Applications (WIAPP). We developed a binary autocorrelation algorithm and discovered that updates come in infinite series with periods of 60 or 75 minutes. We identify both periods as default settings of out-of-the-box Microsoft Windows 2000 and XP DNS software. Identifying this common property of end-user environments helps to understand users' behavior on the Internet. To our knowledge this is the first study of the global impact of dynamic DNS.

6.0 Artifacts Developed During the Past Quarter

None

7.0 Issues

8.0 Near-term Plan

The following work is planned for 01-Jul-03 through 30-Sep-03:

General/Administrative Outreach and Reporting Plans

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

Task 3. Domain Name System (DNS) Infrastructure Model

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.

We plan to repeat the DNS caching nameserver experiments with the following changes:

- use more realistic TTLs for SLD zones/names
- implement multiple IP addresses per hostname
- use some CNAMEs, perhaps across domains
- force some percentage of 'lame delegations'?
- find another source for list of names to query
- address the "replaying trace too fast" problem that causes a big burst of queries before any replies come back.
- address the problem that root/TLD TTLs are longer than experiment duration.

We will also explore cases where a caching nameserver cannot fully communicate with a root nameserver since this seems to me to be the cause for most of the root server load.

9.0 Completed Travel

The following travel incurred expenses to this award and occurred during Year 2, Qtr 4, 1-Apr-03 through 30-Jun-03:

- Duane Wessels 4/6 – 4/8 Colorado to San Diego Passive and Active Measurement (PAM) workshop
- Margaret Murray 5/27 – 5/29 San Diego NMS meeting (parking, registration)
- Kc claffy 5/27 – 5/29 San Diego NMS meeting (registration)

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

- CAIDA presented a poster entitled “Modeling the Domain Name system (DNS) at the San Diego NMS PI meeting on May 27, 2003.
- Ken Keys refined printing within crl_delay for Nikhil Dave of SPAWAR to support a new print-all-packets option. Dr. Dave will be further testing/using this feature in the ForceNET Limited Objective Experiment (LOE).
- A rep from Sitara systems, Zaib Kaleem, saw crl_delay work in the ForceNET LOE and expressed interest in it for incorporation into some commercial products. CAIDA referred this request to Dr. William Decker at UCSD Technology Transfer (TIPS)
- Upon hearing that Microsoft had fixed problems with its implementation of a DNS caching server, CAIDA requested a copy of Microsoft’s W2.003K for

Duane Wessels to include in his DNS testing/modeling. See http://www.packet-pushers.net/dns/simulations/.

- Key Keys released a new version of crl_delay on May 29 to address requests from Nikhil Dave.
- CAIDA received a request from Denise Masi and Martin Fischer of MitreTek for packet interarrival times from real Internet links.
- Ken Keys advised Nikhil Dave on options for synchronizing NIC timestamps for a ship-to-shore monitoring project that will use crl_delay. Ken also added a flow duration measurement.

12.0 Publications and Presentations:

1. The following papers were published:
   a. Broido, E. Nemeth, and k. claffy, ``Spectroscopy of DNS Update Traffic,'', in WIAPP 2003, San Jose, CA, June 2003, WIAPP.
   b. Broido, E. Nemeth, and k. claffy, ``Spectroscopy of DNS Update Traffic,'', in ACM SIGMETRICS 2003. June 2003, ACM.
   c. D. Wessels, M. Fomenkov, and k. claffy, ``Wow, That's a lot of packets,'', in Passive and Active Network Measurement. Apr 2003, PAM.
   d. D. Moore, C. Shannon, G. Voelker, and S. Savage, ``Internet Quarantine: Requirements for Containing Self-Propagating Code,'', in INFOCOM 2003, San Francisco, CA, April 2003, INFOCOM 2003.
   e. T. Lindh and N. Brownlee, ``Integrating Active Methods and Flow Meters - an implementation using NeTraMet,'', in Passive and Active Network Measurement. Apr 2003, PAM.
   f. T. Lee, B. Huffaker, M. Fomenkov, and k. claffy, ``On the problem of optimization of DNS root servers' placement,'', in Passive and Active Network Measurement. Apr 2003, PAM.
   g. Y. Hyun, A. Broido, and k. claffy, ``On Third-party Addresses in Traceroute Paths,'', in Passive and Active Network Measurement. Apr 2003, PAM.
   h. M. Fomenkov, K. Keys, D. Moore, and k. claffy, ``Longitudinal study of Internet traffic from 1998-2001: a view from 20 high performance sites,'', Tech. rep., Cooperative Association for Internet Data Analysis (CAIDA), Apr 2003.

2. The following presentations were given:
   a. "priorities and challenges in Internet measurement simulation, and analysis" (LSN, Jun '03)
   b. "Traceroute and BGP AS Path Incongruities" (Internetworking 2003, Jun '03)
   c. "modeling the Domain Name System (DNS)" (NMS PI, May '03)
   d. "On Third-party Addresses in Traceroute Paths" (PAM, Apr '03)
   e. "Internet Quarantine: Requirements for Containing Self-Propagating Code" (INFOCOM, Apr '03)
   f. "Spectroscopy of Private DNS Update Sources" (WIAPP, Jun '03)
   g. "Atom-based Routing" (Internetworking 2003, Jun '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 Apr 2003 to 30 Jun 2003

Actual Costs Incurred: $ 74,238.00

Current Period:

UCSD
Labor Hours: 997.60 $ 43,504
ODC's: $ 4021
IDC's: $ 24,713
TOTAL: **$ 74,238**

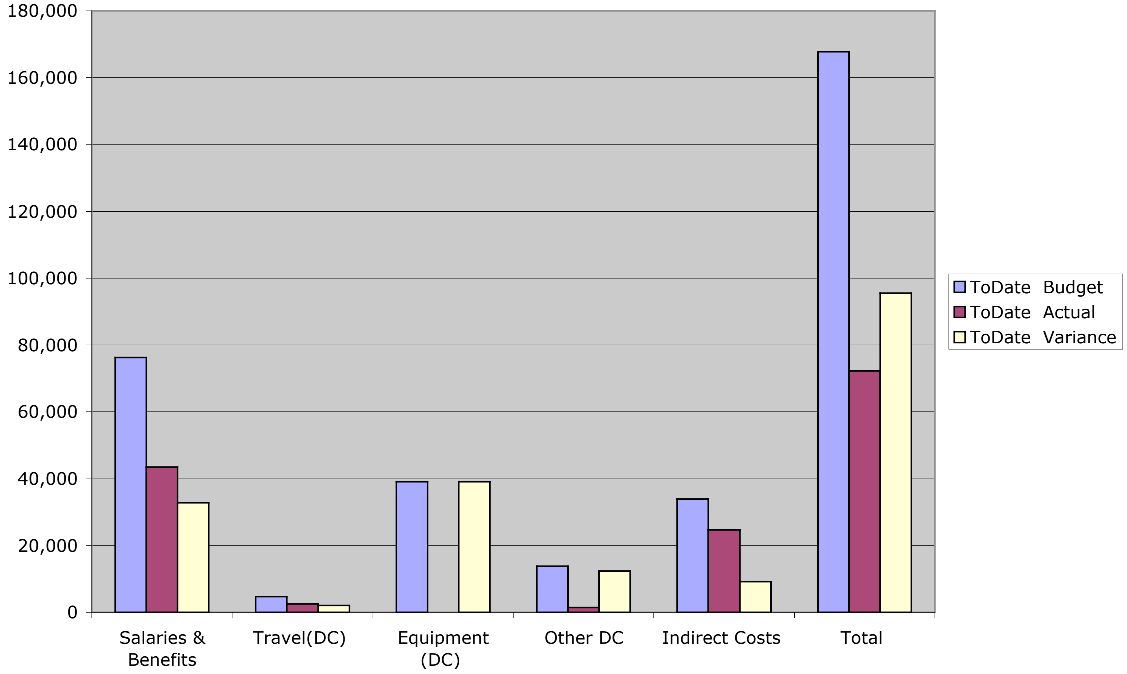
Cumulative to date:

Labor Hours: 21,192.78 $ 778,623
ODC's: $ 82,091
IDC's: $ 427,629
TOTAL: **$ 1,288,343**

Cost Curves for Apr - Jun 2003:

	ToDate Budget	**ToDate Actual**	**ToDate Variance**
Salaries & Benefits	76,306	43,504	32,802
Travel(DC)	4,678	2,566	2,112
Equipment (DC)	39,053	0	39,053
Other DC	13,822	1,455	12,367
Indirect Costs	33,866	24,713	9,153
Total	167,725	72,238	**95,487**

NMS Cost Curves Apr - Jun 2003


Salaries & Benefits
Travel(DC)
Equipment (DC)
Other DC
Indirect Costs
Total
ToDate Budget
ToDate Actual
ToDate Variance