

TTT/CR/22 (do not delete this) **Technology Transfer for
Accelerating Web Access at NSSL in Support of
Hurricane Imagery**

HPCC FY 2004 Final Report

May 25, 2005

Principal Investigator: **Alex Hsia**
Line Organization: OAR
Routing Code: R/OM62
Address: NOAA-Boulder NOC
325 Broadway St
MS R/OM62
Boulder, CO 80305
Phone: (303) 497-6351
Fax: (303) 497-6951
E-mail Address: Alex.Hsia@noaa.gov

Jerry.Janssen@noaa.gov
Gary.Skaggs@noaa.gov
Thomas.Renkevans@noaa.gov

Proposal Theme: **HPCC Technology Transfer**

Funding Summary: FY 2004 \$ 20,000

Technology Transfer for Accelerating Web Access at NSSL in Support of Hurricane Imagery

HPCC Final Report for FY 2004

Prepared by: Alex Hsia

Performance:

The main deliverable of an operational 3-way load balancing of GOES and SSD satellite data web sites was provided for the 2004 hurricane season. This 3-way load balancing consisted of the main origin sites for GOES <<http://www.goes.noaa.gov>> and SSD <<http://www.ssd.noaa.gov>> in Camp Springs, Maryland and the mirror sites in Boulder, Colorado and the National Severe Storms Laboratory (NSSL) in Norman, Oklahoma. However, the funding for the load balancer ended up coming from another source, so the HPCC funds that were dedicated for this purpose were spent to help the Pacific Marine Environmental Laboratory (PMEL) in Seattle, Washington, acquire a load balancer.

In particular, the addition of the NSSL site as a mirror to the GOES and SSD distributed web farms through the use of a Foundry ServerIron XL load balancer allowed NOAA to successfully provide the public with satellite imagery during the active hurricane season in 2004. During peak utilization the NSSL squid farm was providing approximately 135 Mbps of data to the public with about 14-15 million hits per day.

The system documentation is included in this report.

Questions regarding this project can be directed to any of the investigators listed on the title page, and Michael Knezevich <Michael.T.Knezevich@noaa.gov> at PMEL.

Project Summary:

The Satellite Services Division (SSD) of NESDIS (co-located with NCEP in Camp Springs, MD) processes GOES satellite data into forecast products utilized by NWS, emergency managers, researchers and the general public. The basic product suite of GOES imagery can be accessed at <http://www.goes.noaa.gov> and at <http://www.ssd.noaa.gov>. Currently, these are among the most frequently accessed sites in NOAA, particularly during hurricane season when public interest in satellite imagery of these weather phenomena is at its peak.

This proposal utilized investigations and implementations supported by the HPCC Program to extend a technology solution elsewhere. Specifically, the investigation funded in FY02 titled "Accelerating Retrieval of Web Content" <<http://nwweb.nwn.noaa.gov/hpcc/nwg/fy02-funded/we/ngi-nw-04.html>> led to a prototype implementation in Boulder, which was extended to NSSL.

The process of getting the GOES and SSD sites served out of NSSL included the following:

- Replicating the data from the origin server to the mirror site
- Configuring Squid servers to serve as a proxy cache for the websites
- Acquiring and configuring a load balancer for the farm of Squid servers
- Configuring the DNS servers to globally load balance the hostname

Development of the data content transfer and updates via rsync was accomplished by Ed Ladd (RS Information Systems and NESDIS/SSD) and Doug Kennedy (NSSL). Additionally Apache virtual host configuration at NSSL was accomplished by Doug Kennedy.

The current Norman-NOC web server is a Sun Enterprise 220R server, with 2-450 Mhz processors, 2 GB RAM, and 72 GB of on-board disk space. This server houses the repository of data that gets synced from the origin sites such as the GOES and SSD websites and provides the web frontend for the farm of Squid proxy servers that are behind the Foundry ServerIron XL load balancer.

A Foundry ServerIron XL was selected as the load balancer for the Squid cluster due to its performance, and success in load balancing the Boulder webfarm. The Foundry ServerIron XL is capable of providing 1 million concurrent sessions and 19,000 connections per second. As mentioned earlier, the ServerIron XL was purchased through other means so the \$14,000 in this proposal that was dedicated towards that capital expense went to help PMEL purchase a ServerIron GT EGx4.

The DNS resolution for the distributed web farms is currently accomplished through a set of two Cisco Distributed Directors (DD) that are located at the Boulder NOC and the SSMC NOC. The DDs were configured to distribute the global load evenly amongst the three sites: NESDIS, Boulder and NSSL.

One anomaly was discovered during operation during the last hurricane season which led to a configuration standard for those sites that are mirror sites for more than one website. The problem is due to the fact that we are distributing the load based on a portion metric which sends a fixed percentage of DNS requests to a certain site. If a certain site, such as Boulder and NSSL, is a mirror for multiple sites, such as NHC and GOES, and they use the same IP address for the virtual hosts, then the DD may not send off requests in the desired portion. This problem arises from the fact that the DD maintains counters based on IP address so sites such as Boulder and NSSL would not get as much traffic as the origin site for the GOES/SSD websites. The solution is to use a dedicated IP address for each set of mirrored websites where independent portion control is desired. GOES and SSD are considered common websites and thus can use a single IP address.

After the load balancer is put into operation and a farm of web servers or Squid proxies are placed behind it, the bottleneck now becomes one of bandwidth as the system can scale very easily by adding additional nodes. The addition of more nodes can be rather quick with automated cloning of systems with software like SystemImager <<http://www.systemimager.org>>. Care must be taken to ensure that the network path between the load balancer and the WAN links is over network segments that have adequate capacity and that adequate there is adequate WAN capacity to support the public demand.

Expenditure Summary:

<u>Category</u>	<u>Detailed Description</u>	<u>Amount</u>	<u>Matching</u>
Personnel Compensation	Contract Personnel		
	NOAA-Boulder NOC	\$ 2,000	\$ 5,000
	NESDIS/SSD	\$ 1,000	\$ 1,000
	NSSL	\$ 1,000	\$ 1,000
Capital Expenses	Layer 7 switch	\$ 14,000	\$ 21,000
Training/Travel	NOAA-Boulder NOC	\$ 1,000	\$ 1,000
	NSSL	\$ 1,000	\$ 2,000
Total Spent:		\$ 20,000	\$ 31,000

Future Direction:

Now that the GOES and SSD websites have been successfully mirrored to the Boulder and NSSL web farms additional capacity at other major NOAA campuses should be explored. If there were to be an outage at any of the current three mirror sites during a major weather event then the remaining two sites may become overwhelmed leading to a possible cascade failure.

The National Hurricane Center (NHC) website is currently mirrored at four major NOAA campuses: SSMC, Boulder, Seattle and Norman. The GOES and SSD websites should be mirrored to those same sites.

The load balancing hardware was purchased for PMEL as part of this project so the next steps are to mirror the content from NESDIS to PMEL and build up their webfarm to handle the load. The funds required to serve the data out of PMEL may be cost prohibitive, but it provides NOAA an alternative to outsourcing the content to a

commercial provider.