

---

# China Clipper Project: An Overview and Wide Area TCP Results

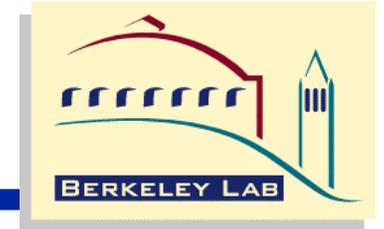
**Brian L. Tierney (bltierney@lbl.gov)**

*Future Technologies Group*

**Lawrence Berkeley National Laboratory**

# Clipper Project

---



- **Goals**
  - **Develop technologies required for distributed data-intensive applications**
  - **Apply to high energy physics (HEP) data analysis**
- **Participants**
  - **Argonne National Laboratory**
  - **Lawrence Berkeley National Laboratory**
  - **Stanford Linear Accelerator Center (SLAC)**

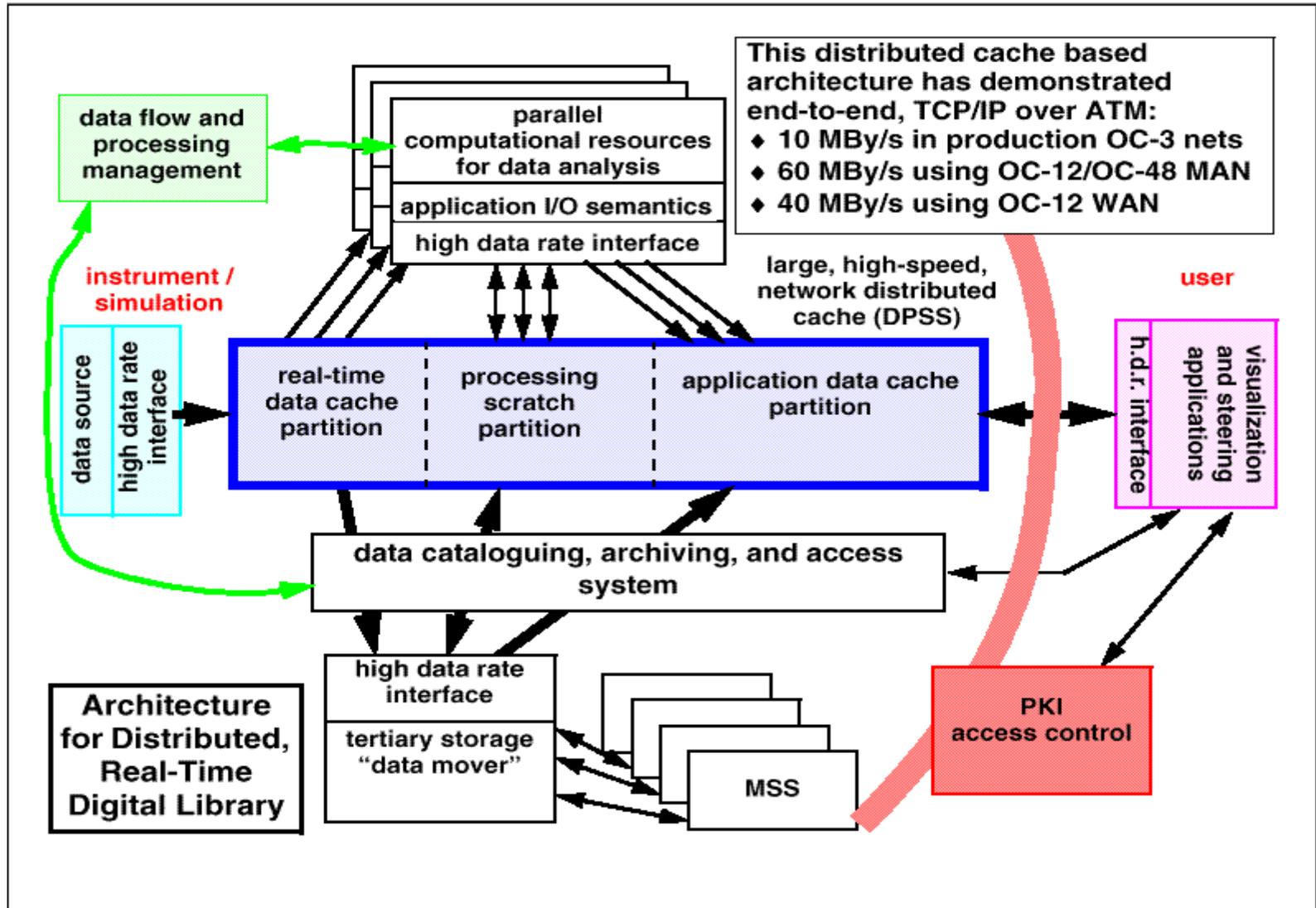
# Clipper Technologies

---



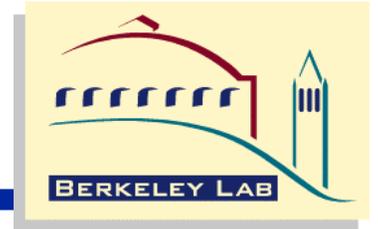
- **Distributed Parallel File System**
  - High-speed, low-cost data cache
- **Globus**
  - End-to-end resource management
- **ESnet and NTON**
  - OC12 networks
- **HPSS and Objectivity**
  - Data archives

# Clipper Data Architecture



# Key Features of the Architecture

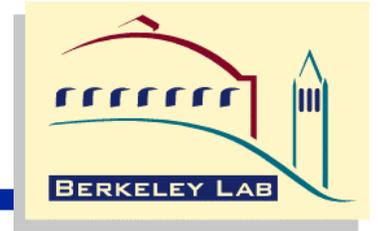
---



- **Very high-speed cache that is distributed, scaleable, and dynamically configurable**
- **Common, low-level, high data rate interface that supports various application I/O semantics**
- **High-speed tertiary storage interface**
- **Data cataloguing and access system**
- **Data Model:**
  - **data sources deposit data in cache, and consumers take data from the cache, usually writing processed data back to the cache**

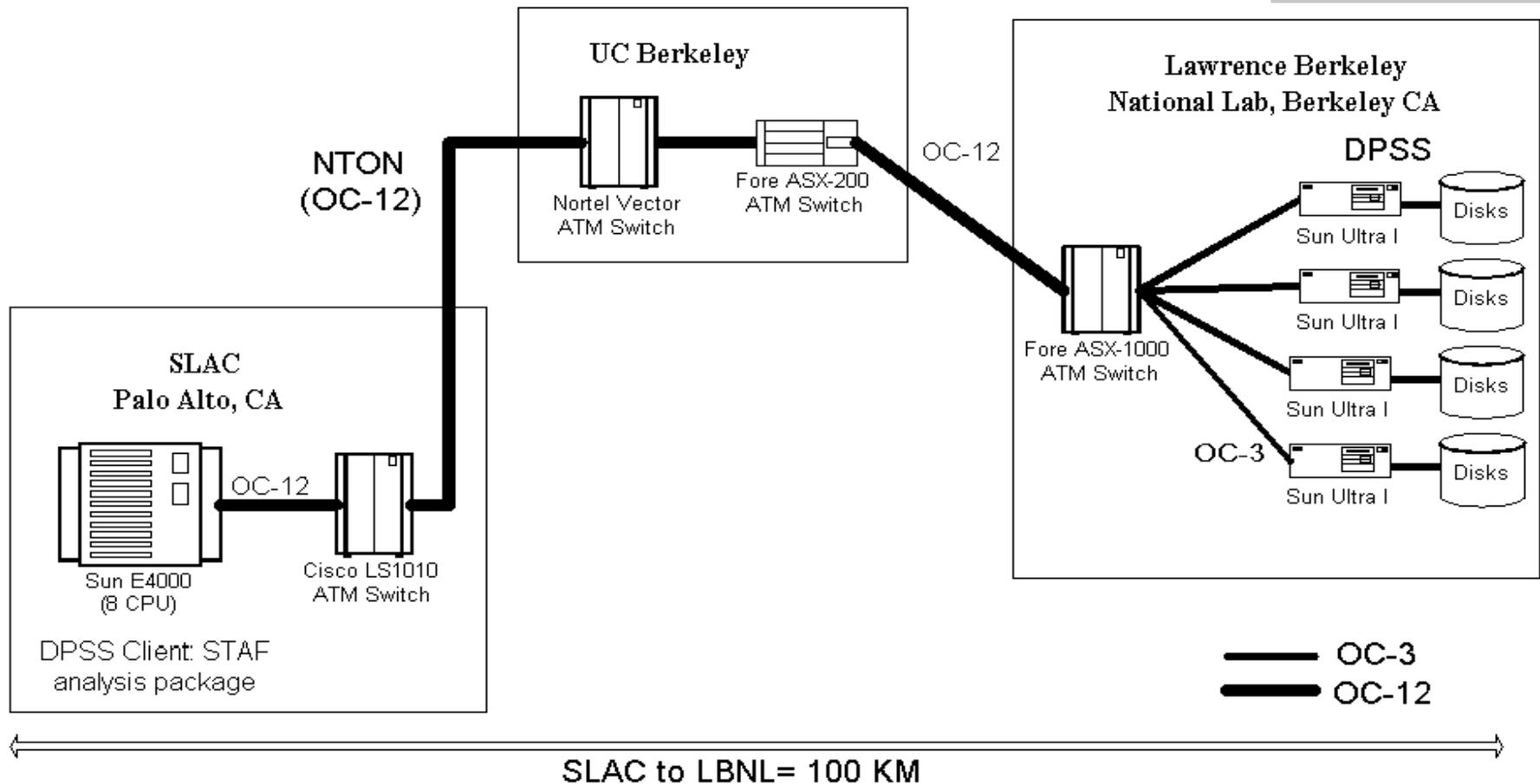
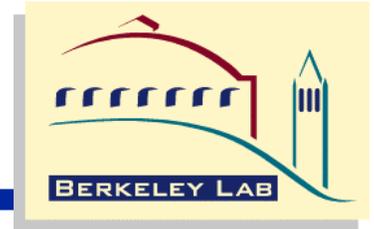
# Clipper Status (March 99)

---



- Initial data transfer experiments conducted
  - LBNL-SLAC: 57 MB/s for HEP application
  - ANL-LBNL: 35 MB/s for HEP application
- Prototype advance reservation capabilities demonstrated in Globus
- Work on network reservations, Objectivity, etc., proceeding

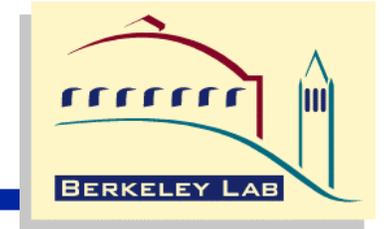
# LBNL / SLAC HENP Application Experiment



Achieved 57 MBytes/sec (450 Mbits/sec) of user data delivered to the application

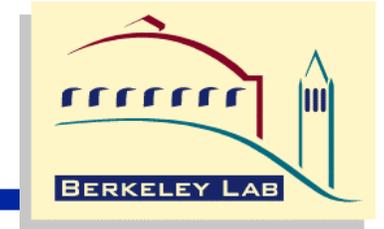
# LBL/SLAC Performance Results

---



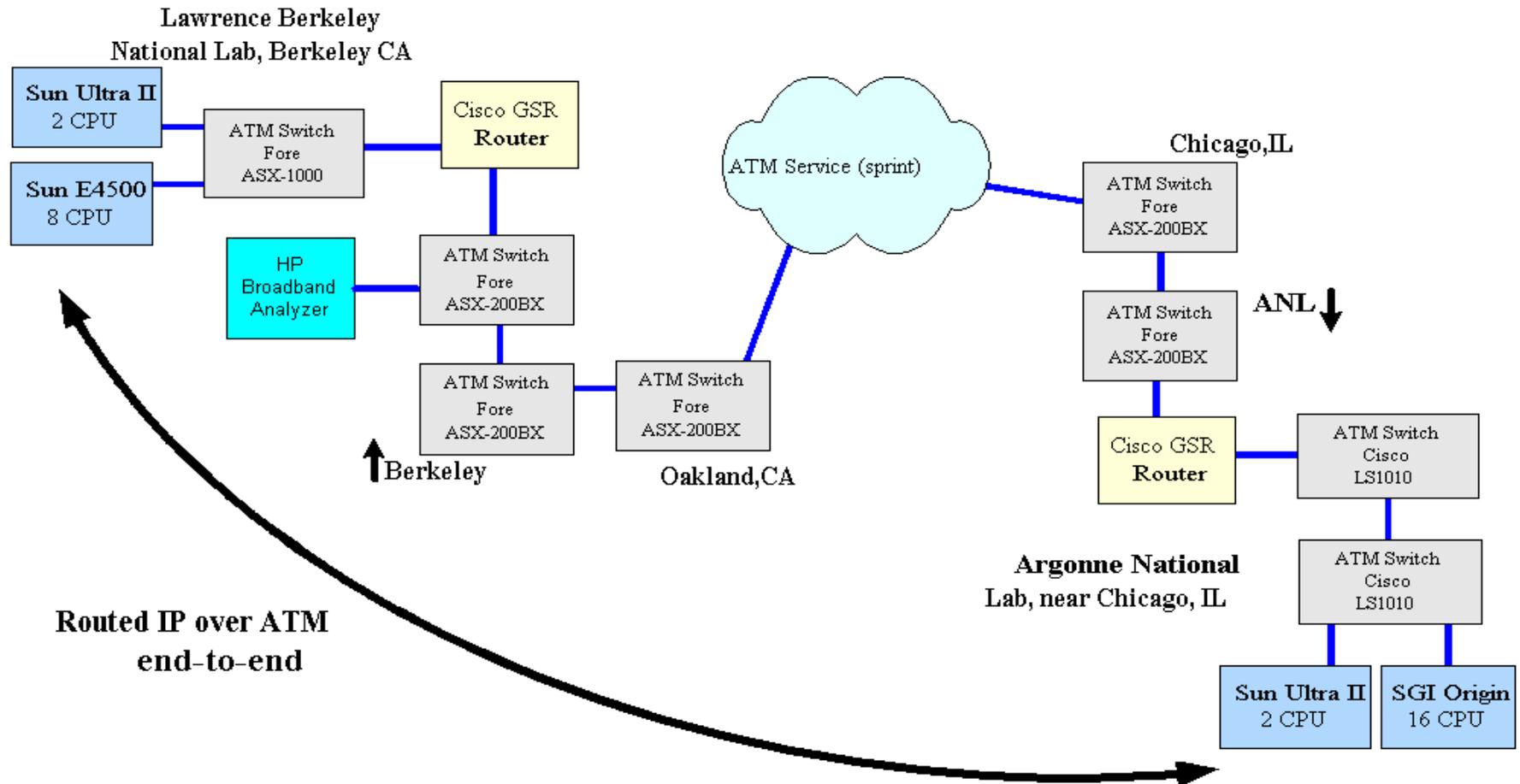
- Experiments conducted over NTON
  - Application network was IP over OC-12 (622 Mbit/sec) ATM.
- An application (STAF: Physics Analysis package) running on a Sun Enterprise-4000 SMP at SLAC (Palo Alto) read data from four distributed disk servers at LBNL (Berkeley), parsed the XDR records and placed the data into the application memory

# Results



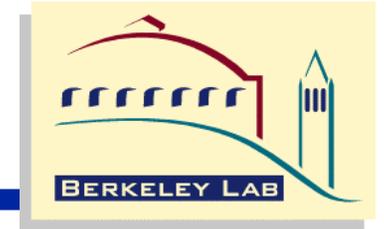
- **Each DPSS server transfer rate is 14.25 MBytes/sec**
- **OC-12 receiver was able read data from 4 servers in parallel at 57 Mbytes/sec**
  - **this is the rate of data delivered from datasets in a distributed cache to the remote application memory, ready for analysis algorithms to commence operation.**
- **This is equivalent to 4.5 TeraBytes/day!**
- **Latency for a single 64 KByte data block is 25 ms, so pipelining is very important**

# ANL/LBNL TCP Experiments



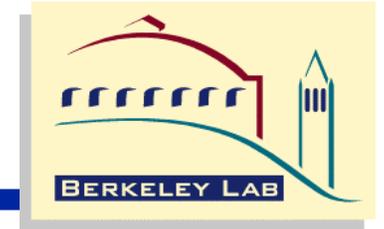
2/9/99 - RLN

# TCP Testing



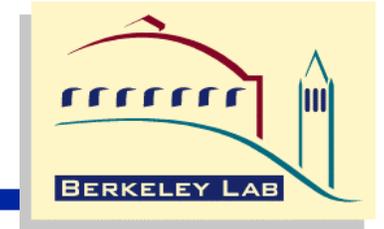
- Tested TCP throughput using `ttcp` with 4 MB send and receive buffers on a “private” network: no other traffic during these tests
  - Results: Throughput = 150 to 300 Mbps
- Installed the SACK patch (RFC 2018) from Sun
  - Results: Throughput = 300 to 480 Mbps
- Variance due to cell loss: GSR routers reported 3-4 packet losses during typical 3-5 minute test
- Replaced port in Oakland ATM switch to try to correct cell loss problem
  - Results: 340 to 480 Mbps: less packet loss now, resulting in smaller range of throughput

# HP ATM Tester



- **Added add an HP Broadband Analyzer at LBNL, and setup a loopback in the GSR at ANL**
  - HP Analyzer does “GCRA (Generic Cell Rate Algorithm) compliance testing” (traffic shaping)
- **Discovered a bug in the GSR policing code**
  - GSR has PCR (Peak Cell Rate), SCR (sustained) and MBS (Max burst size) "equivalents" that are setable. SCR was being ignored.
  - The GSR was honoring the PCR and not the SCR. This was tested by issuing pings FROM the GSR to the HP
- **After fixing this bug: achieved 572 Mbps (ATM rate) (max ATM over OC-12 = 600 Mbps)**

# Summary of TCP Performance



- **Early informal testing**

Test	Throughput Range (Mbits/sec)
TCP: Local LBL loop through GSR	400 to 513 Mbps
TCP: LBNL to ANL (no SACK)	150 to 300 Mbps
TCP: LBNL to ANL (with SACK)	300 to 480 Mbps

- **Current Results (10 GB transfer, shared link)**

Test	Min	Max	Average	Std
ANL to LBNL	278	393	346	36.69
LBNL to ANL	285	387	352	23.59

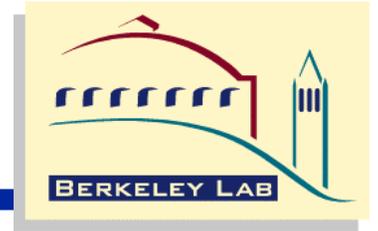
# Why the Large Variance in Throughput?

---



- Most TCP traces show 1-3 “glitches” during a 10 GB transfer (see traces on following slides)
  - GSR router at ANL reports CRC errors on input
  - very hard to determine source of these errors

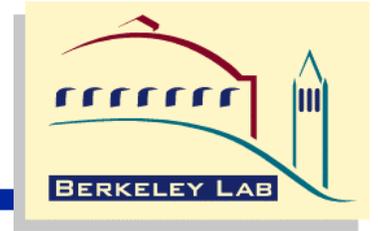
# Issues



- **Still very hard to find problems**
  - **ATM switches still do not accurately report cell loss (it was a LOT of work to track down the bad ATM switch card)**
  - **Can not see into ISP ATM cloud**
  - **Often not getting the service you are paying for**

# For More Information

---



- **Clipper project**
  - <http://www-didc.lbl.gov/Clipper>
- **TCP/IP results**
  - **Contact Brian Tierney, [bltierney@lbl.gov](mailto:bltierney@lbl.gov)**
- **INFOCOM 99 Gigabit Network Workshop Talk on LBNL-ANL results:**
  - <http://www-didc.lbl.gov/publications.html>