COMPUTER & COMPUTATIONAL SCIENCES

SC 2003 Panel

Battle of the Network Stars!

Wu Feng

feng@lanl.gov

Research & Development in Advanced Network Technology (RADIANT)

Computer & Computational Sciences Division

Los Alamos National Laboratory

NNSA

Los Alamos NATIONAL LABORATORY

Panelists

- Brad Booth, Intel Corp.
- Jeff Chase, Duke University
- Dhabaleswar K. (DK) Panda, The Ohio State University
- Fabrizio Petrini, Los Alamos National Laboratory
- Jim Pinkerton, Microsoft Corp.
- Anthony Skjellum, MPI Software Technology

Setting the Stage …

- In the tradition of the ABC TV show that pitted Hollywood stars from different networks against each other, this panel reprises the show where the network stars in this case are InfiniBand, Myrinet, Quadrics, SCI, and 10-Gigabit Ethernet.

- The panel will address two major sets of questions:
  1. Which interconnect is the best for high-performance computing and why?
  2. What are the future trends in high-performance networking, and what are the implications of these trends?

Where Are We Now (Arguably) ?

	10GigE	Infiniband	Myrinet	SCI	Quadrics
Network Environment	Any: LAN, SAN, MAN, WAN	SAN	SAN	SAN	SAN
Scalability	IP-routed → "Infinite" # of nodes		Source-routed → ?		Source-routed → ?
Cost Per Port	$$$$ (2004: $$) (2005: $)	$$	$$	$$$	$$$
Performance MPI-to-MPI	916 MB/s 10.0 us (at socket level)	843 MB/s 6.0 us	250 MB/s 6.3 us	230 MB/s 3.7 us	908 MB/s 2.4 us
Protocols	Native TCP/IP TCP offload RDMA over TCP/IP	RDMA	RDMA	RDMA (or RSM: Remote Shared Memory)	RDMA
Total Cost of Ownership	$	$$$	$$$$	$$$	$$$$$

What will be the metrics of tomorrow?

Specific Questions

- "Head-to-Head" Battle of Network Interconnects
  - Each panelist was invited onto the panel due to their expertise with specific network interconnects. Why is "your" solution the better one?
  - Given that we, as a community, focus on the typical quantitative measure of latency and throughput for network interconnects, what other ways should we be evaluating interconnects?
  - Will the "status quo" in networking continue? That is, Ethernet as a commodity interconnect that often doubles as a cheap commodity solution for clusters with Infiniband, Quadrics, and Myrinet "relegated" to high-end and more costly clusters?
  - With 10-Gigabit Ethernet processors on the horizon, will RDMA/TCP/10GigE be sufficient in matching the performance of Quadrics, Infiniband, and Myrinet? And if so, what does this foretell of the future of these latter interconnects?
  - What assumptions must interconnects make about the underlying arhictecture (or what assumptions would they like to make)? PCI-X? PCI Express? Intel's "Communications Streaming Architecture" or a network co-process?

Specific Questions

- Future Trends and Implications
  - Moore's Law forecasts the doubling of processor speeds every 18 months. Arguably network speeds have been doubling every 12 months on average. Will there come a time where we focus more on "supernetworking"?
  - "Sockets or Bust?": Does the interface to the network have to be a sockets interface? Or will application programmers be willing to rely on "friendlier but inefficient" MPI software to hide such details?
  - In five years, how will today's interconnects evolve and/or compete in high-performance computing?
  - Infiniband started out as a high-performance I/O technology but has evolved into a general network interconnect for high-performance clusters. Will it replace Myrinet or Quadrics as the costlier high-performance interconnect for high-end clusters?
  - What implications, if any, are there for direct-access file systems (DAFS)?