High Performance Computing at Sandia National Labs

R. Michael Cahoon, *John P. Noe*, and *Walter H. Vandevender*[1],
Scientific Computing Systems, Sandia National Laboratories

ABSTRACT: *Sandia's High Performance Computing Environment requires a hierarchy of resources ranging from desktop, to department, to centralized, and finally to very high-end corporate resources capable of teraflop performance linked via high-capacity Asynchronous Transfer Mode (ATM) networks. The mission of the Scientific Computing Systems Department is to provide the support infrastructure for an integrated corporate scientific computing environment that will meet Sandia's needs in high-performance and midrange computing, network storage, operational support tools, and systems management. This paper describes current efforts at SNL/NM to expand and modernize centralized computing resources in support of this mission.*

1 Introduction

Sandia National Laboratories, a US Department of Energy laboratory, works with the US Defense Department, the National Science Foundation, NASA, and industry to develop High Performance Computing (HPC) technologies and apply them to nationally important problems. Its mission includes national security, industrial competitiveness, energy resources, and environmental quality. At Sandia, computers are being used to design and optimize materials ranging from catalysts to opto-electronics, and simulations are replacing tests and experiments that are environmentally unacceptable or prohibitively expensive.

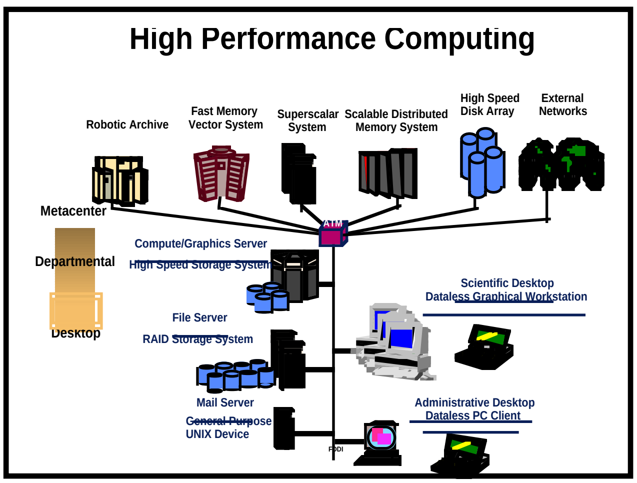

Figure 1: A Fully Integrated Environment.

Sandia uses a variety of computing resources to address the wide-range of computational problems investigated at the laboratories. As shown in Figure 1, our network presents a single system image reaching from the desktop to very high-end machines. The mission of the Scientific Computing Systems department is to provide this integration over a wide-range of platforms.

2 Hardware Infrastructure

Until recently Sandia's major computing resources at the high-end were composed of: 1) an Intel Paragon with 3,744 processors and 38 Gbytes (gigabytes) of memory; 2) one 8-processor Cray YMP and one 2-processor Cray Y-MP, each with 512 MBytes (Megabytes) memory and a 2048 MByte solid state disk and 3) two central file storage systems composed of two Storage Tek tape robots, each with 2 TBytes (terabytes) capacity and two Convex C-series front-end systems each with 100 GByte disk cache. Separate Cray and mass-storage systems were provided for classified and unclassified computing while the Paragon is transitioned between environments on a scheduled basis. Other resources include two nCube 2s, each with 1,024 processors and 4 Gbytes of memory and a variety of high-end workstation systems.

2.1 Mid-range transition

Beginning in 1994, Sandia began a transition to more cost effective platforms that provide departmental level computing. A major benefit of this strategy is that compute servers are more 'local' to the users and machines are tailored to a specific class of user and/or application.

[1] This work was supported by the United States Department of Energy under Contract DE-AC04-94AL85000

By the end of FY95 the two Cray YMP's will be replaced with: 1) a 10 processor SP2 with 128 MByte/node, 2) an 8 processor SP2 with 512 MByte/node, 3) a 4 processor SGI Power Challenge, 4) 3 Cray J90 series for unclassified computing (one J916/8-2048 and two J916/8-1024) and 5) a Cray J916/4-1024 for classified computing. Also, the Convex front-ends to the mass-storage systems will be replaced with SGI Power Challenge systems with updated Titan Unitree software.

2.2 Teraflop System

Sandia, the Department of Energy (DoE), and Intel recently signed a multi-year contract to deploy a 1 teraflop system by December 1996 and to develop a machine capable of 10 teraflops. The initial system will consist of over 4000 compute nodes, 268 GBytes of memory, and two 1 TByte disk systems. Each compute node will consist of two Intel P6 processors, memory scalable from 64 Mbytes to 256 Mbytes, and industry standard PCI bus.

2.3 High Performance Storage Systems

Sandia, in partnership with other DoE labs, NASA laboratories, and IBM is developing software systems to manage the next generation of mass-storage technology. Sandia's High Performance Storage Facility (figure 2), HPSS, is expected to be on-line during the fall of 1995. The initial system will consist of an IBM 3494 tape archive (27 TBytes capacity without compression) and 8 IBM 3590 tape drives. The system is controlled by 4 IBM RS6000/5xx workstations with OC3c (155 Mbit/sec) interfaces. As a network attached device, the HPSS system will be connected to the Paragon with OC12 (622 Mbit/sec) and HiPPI interfaces.

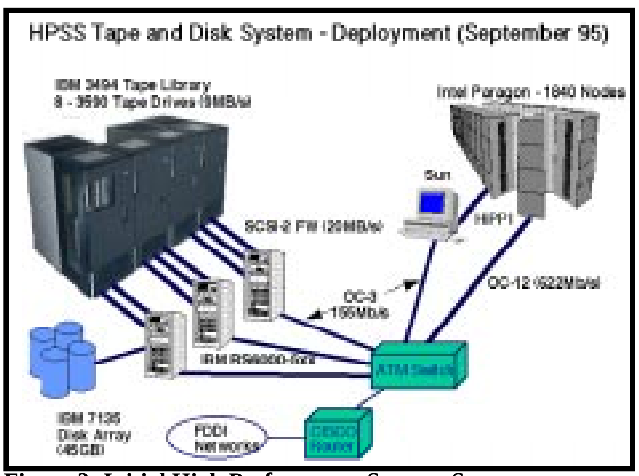

Figure 2: Initial High Performance Storage System

3 Networking Infrastructure

Sandia embraced asynchronous transfer mode (ATM) as a networking technology several years ago and has been very active in deploying and promoting this technology for both the wide-area and local-area networks. As the only distributed DoE laboratory, with facilities in Albuquerque, New Mexico, and Livermore, California, Sandia requires long-distance, high-speed secure communications. The need to conduct large-scale simulations over very long distances has driven work in high-speed encryption/decryption. Moreover, Sandia is a member of AT&T's experimental university network and the Bay Area Gigabit Testbed, and has partnered with communications companies to help form the National Information Infrastructure Testbed.

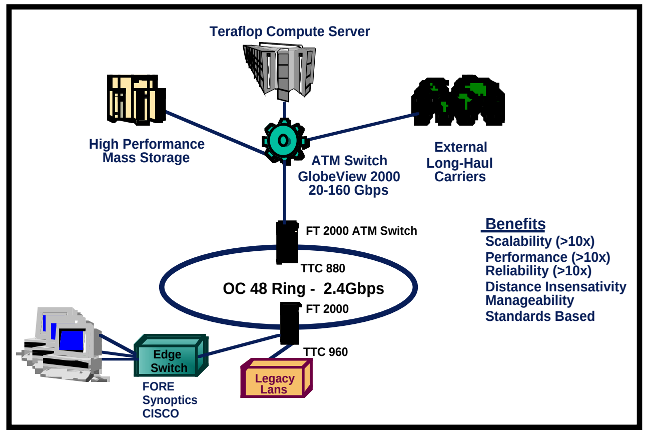

Figure 3: Wide- Local-Area Networks

Figure 3 shows a high-level view of the networks at Sandia, New Mexico. The central facility, which will house the Teraflop, HPSS systems, and other major equipment, is connected to an AT&T GlobeView high capacity ATM switch, national networks, and Sandia's own internal networks. These internal networks are connected through technical control centers (TCC) via single-mode fiber and synchronous optical network (SONET) switches. The entire Sandia campus, which spans about 7 miles, is connected via the TCCs. Each TCC provides connections to the backbone and a variety of ATM edge switches. The edge switches, from a number of different vendors, provide connectivity to devices with ATM interfaces and legacy FDDI/ethernet local area networks.

4 SecureNet Network

The SecureNet computing network provides connectivity among various participating DoE sites to enable the sharing of computer resources and classified data in order to facilitate research and development efforts. Using Motorola NES encryption units attached to the DOE Energy Sciences Net (ESNet), this network is undergoing the final stages of DOE approval for classified use, with a full demonstration of functionality before October 1, 1995. Figure 4 shows SecureNet access into the Sandia National Laboratories/New Mexico participating network.

TIE-In™ (The Technology Information Environment for Industry) provides unprecedented access to a variety of resources at the laboratories. On-line solutions are provided in the form of packaged technologies, which employ graphical user interfaces, design guides and assistants, help, tutorials, and manuals to produce "smart front-ends". TIE-In provides a central point of access to a variety of technical solutions, which are integrated and connected to supporting computational and physical resources. The goal is to provide easy-to-use, self-sufficient environments for utilizing laboratories' technologies to obtain practical and valuable - yet affordable - solutions.

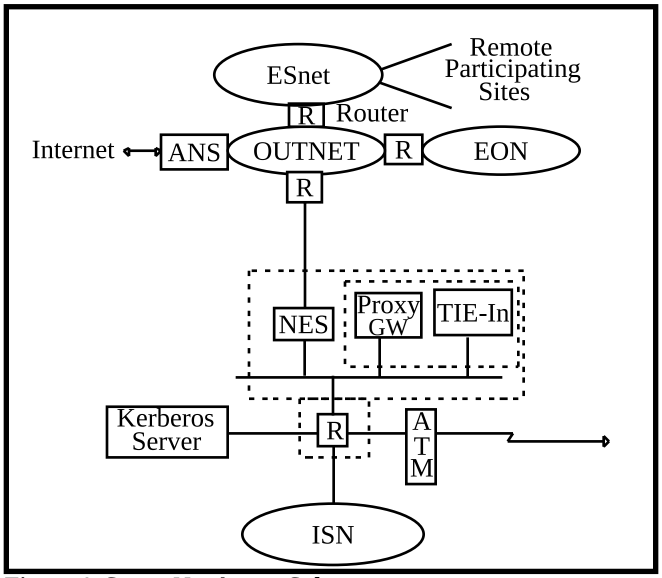

Figure 4: SecureNet Access Subnet

5 Distributed Computing Environment

As the Laboratories move from a facilities-based test environment to one with much greater dependence upon computational simulation, the need for an integrated computer environment that provides a single system image is most important. This single system image will provide: 1) a user-friendly environment where one-time authentication propagates throughout the campus, 2) a robust network queuing system where jobs are routed to the appropriate HPC resource, 3) a mechanism that enables critical applications to obtain shared resources, 4) a network infrastructure that enables compute, visualization and storage servers to scale as the complexity of the applications scale and 4) security measures designed to provide a high degree of interoperability while meeting DoE requirements.

The department will focus on developing the Distributed Computing Environment (DCE) services shown in Figure 5. This will enable applications to achieve the desired level of interoperability.

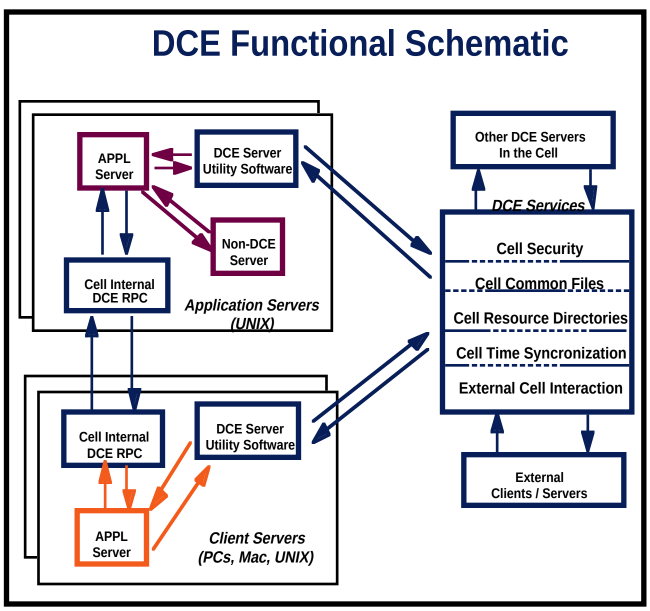

Figure 5: OSF DCE Environment

5.1 Network Queuing Environment

The current batch queuing environment is currently based upon Kerberos Version 5 and NQS. Sandia is investigating a variety of alternative solutions. The CraySoft NQE product and the Production Control System (PCS) from LLNL are candidates.

5.2 Single Login

Sandia has employed the Kerberos authentication/verification system for several years. This system provides a rudimentary mechanism for multi-system access. However, Kerberos was an afterthought applied to our then-current network design. With DCE, Kerberos is integrated into all aspects of DCE services that require authentication. We expect to be able to provide access to all services for every customer via a single DCE login associated with that customer. This will greatly simplify customer access to Sandia DCE computing resources and improve security by requiring customers to retain only one password per network (unclassified and classified network will remain completely separate.)

6 Applications

Sandia scientists are devoting considerable effort to develop highly scalable solutions to a variety of nationally important problems and techniques.

6.1 Shock Physics

Using C++ on the Intel Paragon, Sandia has performed the largest known weapons safety problem using a code (CTH) that models the multiphase (solid-liquid-vapor), strength, and fracture properties of materials. These 3-D problems are characterized by large material deformations: penetration, perforation, fragmentation, high explosive initiation and detonation, and very high-speed impact.

6.2 Electronic Structures

Researchers are developing a variety of quantum-mechanical methods to computationally predict the properties of materials. Applications range from semiconductors to catalysts. The goal is to identify materials with application-specific properties by using computer simulations rather than time-consuming and expensive trial-and-error experiments.

The parallel version of BandPW, an electronic structure model that solves Schrodinger's equations using plane-wave basis, is ideally suited for semiconductors and opto-electronics. Quest has been developed at Sandia to compute the electronic structure of 'large' systems, both molecules (anticancer molecules, catalysts, proteins, and polymers) and periodic crystals (surface catalysts and structural metals).

6.3 Atomistic Methods

The effort to develop scalable molecular dynamics algorithms for simulating systems of millions of atoms has lead to partnerships with Wright-Patterson Air Force Base and with Dupont and Bristol-Meyers Squibb. The Sandia approach models interactions between atoms via force potentials. Typical parallel implementations assign either a region of space or a set of atoms to each processor--so-called spatial or atomistic decompositions. A recently developed force decomposition algorithm has proven to be several times faster for irregular systems of molecules (for example, clusters, surfaces).

6.4 Other Applications

Other on-going areas of research including enabling technologies (operating systems, linear-equation solvers) low-density flows, electrical defibrillation, seismic data processing, medical imaging are beyond the scope of this paper. More information can be obtained from Sandia's World Wide Web server,

http://www.sandia.gov