Conf-910323--11

LBL-30411

Lawrence Berkeley Laboratory

UNIVERSITY OF CALIFORNIA

Received by OSTI

MAY 28 1991

Information and Computing Sciences Division

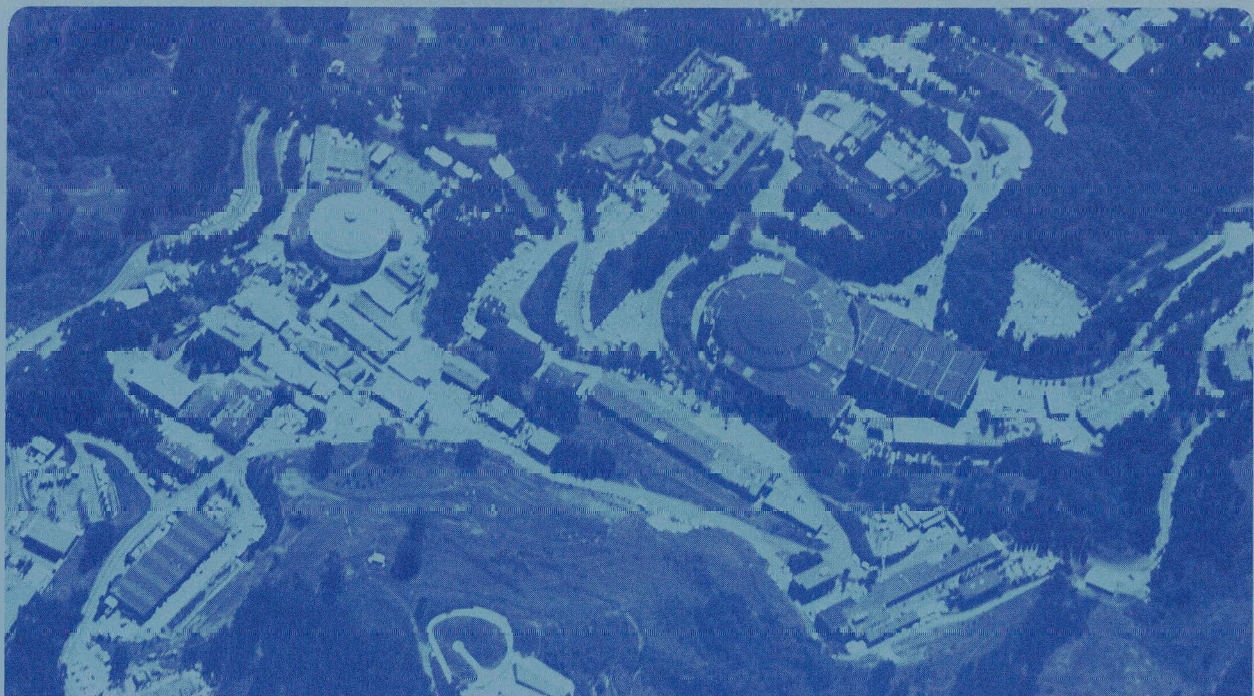
Presented at the Conference on Computing in High Energy Physics, Tsukuba, Japan, March 11–15, 1991, and to be published in the Proceedings

The Role of the <UNIX®> Central Computing Facility in a Multi-Purpose National Laboratory

C.A. Eades

March 1991

DO NOT MICROFILM COVER



Prepared for the U.S. Department of Energy under Contract Number DE-AC03-76SF00098

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. Neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial products process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California and shall not be used for advertising or product endorsement purposes.

Lawrence Berkeley Laboratory is an equal opportunity employer.

LBL--30411

DE91 012361

The Role of the UNIX™ Central Computing Facility in a Multi-Purpose National Laboratory*

Craig A. Eades
Information and Computing Sciences Division
Lawrence Berkeley Laboratory
One Cyclotron Road
Berkeley, CA 94720

March 5, 1991

Prepared for the 1991 Conference on Computing in High Energy Physics
Tsukuta, Japan, March 11–15, 1991

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

*This work was supported by the Director, Office of Energy Research, Office of High Energy Physics and Nuclear Physics, of the U.S. Department of Energy under Contract DE-AC03-76SF00098

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

The Role of the <UNIX™> Central Computing Facility in a Multi-Purpose National Laboratory

Craig A. Eades

Information and Computing Services
Lawrence Berkeley Laboratory
Berkeley, California, 94720

ABSTRACT

In the last ten years we have seen a significant change in the role of the central computing facility. This has been brought about by the new technology and the evolving needs of the research community. In this paper we discuss the current efforts of the central computing facility's UNIX Group at Lawrence Berkeley Laboratory to address these changes in technology and the needs of its research groups. This paper is divided into three general areas. In the first, management, we discuss the costs of system management in a distributed computing environment; current computing issues, including the impact of workstation technology on the laboratory infrastructure, the advantages and disadvantages of diskless workstations in the laboratory environment as well as the need for high-reliability file services it implies. In the second, mass storage, we discuss the need for and an implementation of mass storage service to the laboratory community. We will also discuss issues such as network access and integration strategies. In the third, we will discuss other types of services provided by the central facility, including common software export via high availability Network File Services (NFS™) servers, distributed high-quality output devices and the role of the central facility in providing computer security.

INTRODUCTION

Ten years ago, the central computing facility at many institutions was a monolithic entity providing cpu cycles and disk space to a captive audience. The cost of acquiring and maintaining a computer was so high that users often had no choice but to utilize the central facility, paying whatever price it charged. A typical user viewed the central facility as a monopoly provider of service, its computing policies as capricious and excessive. It arbitrarily imposed, for instance, file naming requirements, login name restrictions, scheduling of downtime, charging policies and so forth.

Today, it is not uncommon for research groups to possess a computing capability equal to that available from the central facility. The advent of inexpensive computing has allowed research groups to acquire and maintain their own computing systems. Departmental computing provides local control of a resource that is becoming an ever greater part of the jobs performed at research facilities. Since policies and procedures can be set locally, they can be as lax or as restrictive as the philosophy of the departmental group could support.

The development of greater local computing capability will have a dramatic impact on the way central facilities must conduct their business and how research groups will satisfy their computational needs. The central facility will have to find new ways to support research efforts. It will move from being a commodity provider of generic cpu cycles to a service based provider of skills and expertise.

Why UNIX?

UNIX is often the operating system of choice, not necessarily because it is the best, but because it offers a rich computing environment at a very low cost. The UNIX operating system itself is quite simple and easy to understand. Source code for the system is readily available (and quite inexpensive for educational and research institutions), so adding new hardware or extending the operating system is possible.

The UNIX operating system is a major reason for the growth in the use of TCP/IP network protocols. Almost all new networking research is done first on UNIX platforms. Workstations running UNIX and equipped with TCP/IP software have become common throughout the research community providing electronic mail access and file and data sharing among users at distant sites.

Because UNIX is easy to port to a new processor, hardware vendors often select it as their operating system, allowing manufacturers to concentrate on hardware development instead of developing an operating system. The selection of UNIX or a derivative as the operating system to run on the current RISC processors is an example of this choice.

Portability of applications is another powerful motive for choosing UNIX. Institutions have become tired of developing software for proprietary operating systems, only to become trapped in a particular architecture, dependent on a particular vendor. Although there are exceptions, UNIX in most cases provides languages and tools to develop portable applications. By designing future applications with portability in mind, developers can extend their software investment into new architectures at little cost.

For the immediate future, we can expect that the fastest and the most inexpensive workstation processors will be supporting UNIX. Its simple structure lends itself to efficient implementation on a wide variety of architectures and its mature support environment as well as the growing body of software vendors who support it, will provide a very strong motive for hardware vendors to supply UNIX for their processors.

Distributed Computing Management

Regardless of who manages a workstation, or where it is located, there are certain institutional requirements which must be met. The workstation cannot adversely affect other systems; it must not be run in a manner which will not present security risks, and it must protect corporate data residing on such a machine.

A distinction should be made between personal computers (such as MACINTOSH computers) and more complex scientific workstations. A personal computer is designed to be minimally configurable. It is packaged for the novice to install. Networking provided by the vendor is not complex, is usually homogeneous, and so is easy to install and maintain. The user doesn't have to understand its relationship to other computers in a network environment.

Although the equipment has become small and simple, the operating system has not. It contains all of the capabilities of a larger more physically complex system, and presents all of the management requirements.

There are some fixed costs for workstation management. These costs may be hidden, perhaps by assigning the task of management as an additional duty to a researcher or graduate student, or may be explicit, by either hiring or contracting this service. Hiding management costs by using "free" labor can in fact, drive up the cost of research and can be more expensive as "free" management climbs the learning curve.

Because of their very nature, scientific workstations are usually connected to a network. Failure to provide sound management services can have a significant impact, not only on the viability and functionality of the associated system, but of other systems as well. A workstation, improperly configured, can start a broadcast storm which will be deleterious to the performance of most other machines on that net.

Failing to install the operating system properly can lead to (among other things) significant breaches in the security of the system. Many operating systems are shipped from the vendor with security holes known throughout the hacker community. Unless a knowledgeable individual is installing the system, these problems may be left uncorrected, leaving the system vulnerable to penetration by others.

Although file system maintenance and backup is one of the most basic of system management functions, it is surprising how often it is neglected. Failure to observe basic management procedures can result in the loss of data.

Diskless workstations, where the operating system and its associated files are managed remotely seem to be a good alternative.

Three types of server/client relationships can exist in a distributed computing environment.

1. Diskless workstations are easiest to administer. Clients load all file systems across a network and there is no local data. Their server must be available for them to operate. Diskless workstations require all the resources that diskfull workstations require, but all file systems, including the root and swap file systems, are accessed across a network.

2. Dataless clients have a local root and swap partition, but all other files are accessed across the network. User file systems will exist on a remote server, creating a dependency on that server. This type of client is more difficult to

administer since the operating system must be upgraded or installed on an individual basis by the user or manager.

3. Datafull clients have local root and swap and user file systems. This places the least load on the network but is the most difficult to administer.

At LBL, as at other institutions, there is debate about the utility of diskless workstations. Given their lower acquisition cost and ease of maintenance & installation, they are often an attractive choice for research groups with limited equipment budgets. However, two aspects of diskless workstation performance raise concern: (1) their ability to perform well without local storage, and (2) their impact on the network infrastructure of the laboratory.

In discussing the impact that diskless workstations have on a network, we need to consider how that load is generated. The load occurs when diskless workstations boot and they are loading their kernel across the network, and when performing a page fault either through image activation or program swap. Minimizing any of these factors will mitigate any negative impact that the workstation would have on a network.

Recent Investigation into the way that workstations are used at LBL has given us some insight into the load that diskless workstations place on the network. Most workstations are used for text editing, for sending and receiving mail, and for communication with other computer systems. Often they perform a single function, such as text editing, for long periods of time.

Our experience is that diskless workstations place a load of less than 1 packet per second on the network on average; this is a load that most high speed, well managed networks can tolerate.

When considering performance we must realize that diskless workstations are usually memory constrained. Increasing main memory size will often bring significant increases in performance and reduce the swap rate. This is especially true in older workstations which often run large windowing systems and thus are memory starved at 8MB. If diskless performance is not satisfactory, further performance improvements in performance can be realized by placing the temporary file system in memory[1].

However, there is a class of users who benefit from the addition of local storage. These users have an extremely high swap rate usually found where large compilations are done frequently. In our environment, the number of individuals that perform this type of work is limited.

Diskless workstations are an extremely viable option especially given the rising costs of labor, and the marginal performance penalty imposed by running

[1] J. Polk, Sun Microsystem, "NFS Config", Presented Oct 1990 at the Conference on Large Systems Administration, Colorado Springs, Co.

a workstation diskless. Future advanced in networking technology will further mitigate these negatives.

The Role of the Central Facility

Even though there is a trend toward departmental computing, there remain many areas of service and technology which a central facility can more effectively provide. It should be the job of the central facility to investigate these areas of opportunity and to provide service where it can do so more efficiently than can the community.

The central facility is responsible for institutional planning. No research group can really have a perspective which encompasses the entire institution. Issues exist which, while perhaps not significant to the individual are important to the institution as a whole. Examples of this are networking infrastructure, computer security, and corporate data conservation.

The central facility can take advantage of economies of scale and provide inexpensive management services are to the laboratory community. At LBL, our experience with various UNIX server/client configurations leads us to believe that management of a file server and 5 diskless client workstations requires approximately one quarter of one full time employee (FTE). As more diskless workstations are added, economies of scale are generated, reducing the per workstation cost. Included in this cost are all aspects of systems and network configuration, insuring file system integrity and routine system management tasks such as backup and operating system update. It it is possible to manage such a system with less of a personnel investment than one-quarter FTE; however, such systems tend to be managed less well and do not provide a rich software environment .At our institution, the cost of this support must be borne by the research group; however, it is fixed cost that can be planned for and because it is a cost based service, it provides a tool for both the central facility and the community to predict the minimum management costs. Because of this we have a clearer picture of one of the component costs of research.

At LBL, the community is able to select from a wide range and configuration of services; they can select a single service such as account management, or subscribe to a program of full support (as described briefly above) with various levels of support in between.

With partial management, groups may reduce the rate they pay the central facility for management by performing some aspects of management themselves. For instance, the Scientific Visualization Group at LBL provides all services for its computer systems except operating system installation and file system backup. This gives the group local control of their systems for policy decisions and software installation and provides a cost effective way to reduce system management costs.

Managing the multitude

The ideal scientific workstation might be described as one with infinite cpu, unfailing infinite storage, infinite network bandwidth and infinite mean time between failure (MTBF). Diskless workstations come closest to providing these characteristics. The small number of moving parts (some without any moving parts) provide high reliability and office environment "friendliness". Because most workstations are memory-constrained, and because of the way that most workstations are used, the lack of a local disk does not have a negative impact on performance; instead it provides significant gains in ease of administration.

A diskless workstation provides many important benefits. Because there is no local operating system managers can update many workstations at the same time. The low purchase price eases acquisition costs, allowing further penetration of workstation technologies into the laboratory. The use of network services such as Sun's NIS (Formerly Yellow Pages) allows users to access many machines and to have an environment familiar to them.

To support diskless workstations, and diskfull workstations mounting common file systems, high reliability file servers are required. Although fault-tolerant systems have been available for some time, and disk technology can provide on-line recovery via RAID schemes, fault-tolerant servers employing these features that are capable of supporting diskless workstations are not available. Vendors supporting diskless workstations and performing work in multi-processing do not indicate that this will be a priority in the near future. However, high availability of file systems is an institutional need.

[AT LBL we have found that one problem with central management of systems owned by other groups is the users; perception that their problems are not considered important by the support group. This leads to a feeling of abandonment and powerlessness by the users. To address this problem, we assign a coordinator to each machine, whether it is a server or a workstation and regardless of its class. The coordinator is the responsible advocate for that machine; he is directly responsible to the supported group and responsible to central facility management. It is important to note that the coordinator is not required to perform any system functions on the machines he coordinates; instead he insures that activities necessary to the maintenance of the machine are performed.]---needs to move somewhere else, or deleted.

Mass Storage

Two types of mass storage are necessary in a distributed computing environment: high availability storage for critical software , and an archival storage facility which can store large files at a very low cost. A successful mass storage facility will provide both of these to its clients in a seamless manner.

The LBL Mass Storage System

Present computing environments such as LBL are very large and complex, containing a diverse collection of hardware and operating systems. A mass storage system needs to be able to serve this heterogeneous environment.

NFS, Sun's Network File system and FTP, the industry standard file transfer protocol are being coupled with advances in optical and magnetic storage and robot technology to provide a low-cost mass storage system.

The architecture of the LBL mass storage system is taken from the IEEE Mass Storage Reference Model[2] . The mass storage system will be a UNIX-based, large-scale, distributed file and storage management system. The system will provide arbitrarily large amounts of storage capacity and fast access to stored data with full location transparency. All servers and daemons can be run on a single machine, or multiple machines, and will accept client requests from a variety of methods including FTP and NFS.

Using NFS, clients will see the mass storage system as an infinite file system which can be mounted locally. Programs which currently read and write to local files systems will be able to read and write to files residing on the mass storage system transparently.

Using FTP, clients without NFS access can replicate and restore files from the mass storage system. Policy agents can be written on each platform to automatically control file migration to the mass storage system.

On the mass storage system itself, files will migrate from the caching store (to consist of high speed magnetic disk), to low cost/Mbyte backing stores, (magneto-optical disk). Based on a least-recently-used algorithm, these files will then migrate to archival storage media (high density tape). On reference, the file will migrate (fault) back to the cache for access.

The mass storage system will also grow as technology in the area of storage media and as workstation and networking technology progress providing larger lower cost file space and improved access times.

Since the IEEE standard supports the simplest randomly addressable file abstraction, (a bit file), more complex file structures can be created if needed.

Although the mass storage system will contain many gigabytes of storage, current file size limitations of 2GB in most versions of UNIX will impose some restrictions which may affect applications using large data sets. This must be considered a temporary problem as the increasing size of storage devices drives the requirement for greater file size.

---

[2]IEEE Technical Committee on Mass Storage Systems and Technology "Mass Storage System Reference Model: Version 4", Draft standard of the Institute of Electrical and Electronics Engineers, S. Coleman, S. Miller, ed.

A more complete description of the LBL mass storage system is contained in "Building a Mass Storage Server for Physics Applications", given at this conference.

High Availability Storage

In addition to the mass storage system, another type of storage is desirable for files which must be frequently accessed or which need to be available on fast storage media directly connected to the network.

Among these are codes which consume a large amount of space, change frequently and are difficult to build. An example is the X11 windowing software. Since the mass storage system will consist of tiers of storage with files migrating from fast cache storage to higher density lower cost storage using a least recently used algorithm, it may not be an appropriate place for files which change often, or are accessed frequently. Constant access to a large file will keep it in the faster cache, while the fault process for files stored in the slower storage media may be intolerable to certain file usage.

The consequences may be severe if the server exporting a windowing system to a large number of clients fails. In this case, dependency on the imported files is so great that no other work can be done while the server is unavailable.

Although fault tolerance is not now available at a reasonable cost, single board dedicated computers and mass storage devices with very high MTBF can provide very high availability file servers.

We have dedicated a low-cost (<$5K) workstation with 4GB of disk storage to this task. It performs no function other than serving files across the network. It does not run any network services other than those necessary to perform this task. There are no user accounts, and the operating system is kept as stable as possible. The inherent reliability of single board computers and the high MTBF has provided a platform with very high availability. At LBL over 100 workstations receive their windowing and font software in this fashion with very satisfactory results.

Other Central Facility Services

There are many other areas in which the central facility can provide service. Some of these services provided by the LBL central facility are discussed below.

Network backups

The central facility provides backup service for both UNIX-based workstations and personal computers. Both backup services run unattended and utilize high capacity 8mm tape drives.

A single server with five 8mm tape drives performs nightly incremental and weekly full backups for 45GB of disk on 35 UNIX-based servers. Two print servers provide backup service for over 250 Macintosh computers.

Central Mail Gateway

The central facility maintains a mail server which provides a well known address to which any user can have mail sent. The gateway provides appropriate routing and delivery of electronic mail from heterogeneous networks.

Security

The central facility at LBL provides two security services. The first, a laboratory wide User Identification (UID) server improves the security of files exported via NFS. The second, an extensible network security probing system called NetSweep allows the central facility to quickly check all local hosts connected to LBLnet for known security problems.

The uid and Group Identification (GID) 2-tuple mapping is the basic method of establishing identification and access rights for users in UNIX systems. In an environment where cross mounting of file systems is practiced, it is important to preserve a unique relationship between individuals and uids.

The UID server centrally administers the LBL uid/gid domain. Electronic mail is sent to the server containing the users name, login id and group affiliation. If the user has already been issued a uid the server returns the previously assigned uid. Otherwise a new uid is assigned and the system manager is notified by return mail. In practice the process takes minutes.

The second facility, NetSweep, allows the central facility to quickly install security tests and check all of the local hosts connected to LBLnet. NetSweep is run periodically to check all the computers on the laboratory's network for known security violations.

Software

Distributed software is an area which the central facility can provide invaluable support for research groups. There is an abundance of public domain software which can be extremely useful to research groups. Examples of this are GNUEmacs and GAWK, available from the Free Software Foundation. However, an individual research group may not have the necessary disk space or the requisite software environment for compilation. At LBL, we export X11 software to the laboratory community. As this is the interface software for a large part of the laboratory community, it is important that it have high availability. AT LBL, a server is dedicated to the export of software to the laboratory. This server has no users and is designed to have a very high availability.

Visualization

The central facility supports development efforts in new data analysis and display technology. One example of this is the use of a data flow visualization system based on the notion of visual programming, such as Stardent Computer's Application Visualization System (AVS). The acquisition of graphic engines, software and high-quality continuous-tone color output devices by the central facility has given researchers in Particle Physics, Earth Sciences, Mathematics, Research Medicine, The Human Genome Center and Nuclear Science novel new ways to prepare images from scientific data.

PROGRAM DEVELOPMENT

There is a certain class of software which has broad applicability to the laboratory population. This software includes graphics libraries, interfaces such as X-11, and large data-bases.

The central facility can acquire this software and make it available. At the user level, software is acquired, maintained and distributed in an up-to-date fashion.

PROJECT LEADERSHIP

Because of its unique perspective in matters relating to computing at the laboratory, the central facility can undertake a leadership role in the acquisition of new computing technologies. At Lawrence Berkeley Laboratory the central facility participates in many Beta test programs with vendors. Participation in these programs provides two main benefits. First, it has a direct influence on emerging technology. Beta sites have direct contact with the group implementing the product. By using this link, a great deal of influence over the final product can be generated. The end result is a supported product which meets the needs of the laboratory more closely.

Second is the advance access and integration of new hardware into the computing environment prior to its release to the general public. Such advance access provides a mechanism to ease the implementation process for new hardware. For workstations, a Beta test program can significantly increase the life span of such equipment, bringing, instead of a 24 month utility, as much as 30 months' lifetime.

Distributed Printing

The distributed printing service[3] at LBL consists of multiple-vendor printers and print servers connected to the LBLnet and other LocalTalk networks throughout the laboratory. The printers are attached to UNIX based print servers running derivatives of the 4.X BSD operating system. Printer access via

---

[3] W. Johnston, D. Hall, "Unix Based Distributed Printing in a Diverse Environment", Advanced Development Projects, LBLID-21440,Lawrence Berkeley Laboratory, June, 1986

the print servers is through direct ethernet connections, through AppleTalk connections via Kinetics FastPath Gateways and through RS232 serial connections. The print servers are located within the central computing facility. Users can access the Distributed Printing System from the LBL VMS cluster and from any LBL UNIX computer connected to LBLnet. Since, at the printer level, the network is LocalTalk, Macintosh computers at the laboratory connected to LocalTalk can also use of any of the Postscript printers.

Six Sun workstations act as print servers for the entire laboratory. Local workstations or servers never process any print jobs. Instead, the print job is forwarded to a print server. Each print server contains all of the software necessary to process the print job and direct it to the appropriate printer. A user can select a local printer or a printer near the office of an associate. This strategy frees the local system from the task of processing a job from text into another printer language and also frees user from maintaining the complex and large font and software trees necessary.

This service now supports over 150 printers ranging from 4 page per minute 300 dpi HP and Apple laserwriters to 12 page per minute 600 dpi laserprinters to high quality continuous color output devices. Any of these devices can be accessed by any other system at the institution.

As this distributed printing service became more complex and diverse, so do the possible modes of failure. Even small hardware and software failures can cause unpredictable results, unnecessary delays and frustration to the user community. To support this very large distributed printing plant, a Distributed Printing Monitor[4] was developed. The monitor utilizes the network to check each component of Distributed Printing to the lowest level possible if trouble is detected, mail is sent to the operations staff for immediate action.

CONCLUSION

The central facility can capitalize on its expertise in computing to provide solutions which will have an institution-wide impact. It can capitalize on its links with vendors to direct development of hardware and software in a direction advantageous to its community. It can capitalize on its ability to reliability store large amounts of data. It can capitalize on its connectivity to other research institutions to assist in the flow of technology and information into the institution. It can capitalize on its expertise to develop new software methodologies which will help researchers better understand their data.

AT LBL the central facility has responded to these changes in technology by examining its role and providing service where that service can be unique, qualitatively different or more efficiently provided, than from its community.

[4]R. Rendler "A Distributed Printing Monitor", LBLID-27492, Computing Services Department, Information and Computing Sciences Division, Lawrence Berkeley Laboratory, May 1989

Far from being outmoded, we believe the need for a central facility will continue as dependence on workstation technology increases at our institutions. Policy does not need to be made to protect the income or the traditional areas of service of the central facility. An active, responsive facility will generate more than enough work to support its existence.