LEGIBILITY NOTICE

A major purpose of the Technical Information Center is to provide the broadest dissemination possible of information contained in DOE's Research and Development Reports to business, industry, the academic community, and federal, state and local governments.

Although a small portion of this report is not reproducible, it is being made available to expedite the availability of information on the research discussed herein.

Received by Com

FEB 09 1990

Los Alamos National Laboratory is operated by the University of California for the United States Department of Energy under contract W-7405-ENG-36

LA-UR--90-32

DE90 007887

TITLE RE-ENGINEERING THE LOS ALAMOS COMMON FILE SYSTEM

AUTHOR(S) Ronald D. Christman
Danny P. Cook
Christina W. MercierSUBMITTED TO Tenth IEEE Symposium on Mass Storage Systems
Monterey, California
May 7-10, 1990**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.

By accepting this article for publication, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes.

The U.S. Agency further states that it requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy.

Los Alamos National Laboratory
Los Alamos, New Mexico 87545

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

RE-ENGINEERING THE LOS ALAMOS COMMON FILE SYSTEM

**Ronald D. Christman
Danny P. Cook
Christina W. Mercier**

**Los Alamos National Laboratory
Los Alamos, New Mexico**

ABSTRACT

The Los Alamos National Laboratory's Common File System is being substantially upgraded so it will continue to meet the future data storage needs of the Laboratory and other sites. The nature of computing is changing with the advent of workstations, different work loads, and standard approaches to networking and operating systems. We reviewed the pros and cons of the Common File System and numerous existing and planned systems. Our conclusion was that the Common File System was a valuable resource with many strengths that should be preserved for the future.

We have arrived at a set of recommendations for changes and additions that we feel will allow the Common File System to continue to meet data storage needs in both the short and long term. The recommendations for 1990 - 1992 support the transition to Unix on supercomputers and support high performance networking. The recommendations for 1992 - 1995 support the transition to a Unix type file system that provides a significant level of file location transparency. For 1995 - 2000 the recommendations are necessarily less precise but support the notion of standard security, transparent file systems and faster networking.

INTRODUCTION

The computing environment at Los Alamos National Laboratory is expected to change dramatically in the 1990's. A study was done to determine the approach the Common File System (CFS) should take for solving the Laboratory's data storage needs in both the short and long term. The study involved determining the current strengths and weakness of CFS, researching the agreed upon scenario for the Laboratory's computing environment of the 1990's, investigating other mass storage systems in use and under development, and finally making recommendations.

METHOD OF OBTAINING INPUT AND SUGGESTIONS

Input on user satisfaction for the CFS was obtained from the following sources: (1) a survey sent to all users, (2) a newsletter article asking for suggestions, (3) interviews with user consultants who had a lot of experience and contact with CFS users, (4) interviews with users that do a large amount of computing and make extensive use of CFS resources, (5) interviews with Los Alamos experts in operating systems, networks and computer security, (6) discussions with inventors, users and manufacturers of state of the art file systems, and (7) requests for comments by distributing a report and giving review presentations.

FUTURE COMPUTING ENVIRONMENT AT LOS ALAMOS

The early 1990's will be a period of transition at the Laboratory as computer users move from using conventional terminals to accessing supercomputers and other Integrated Computing Network (ICN) facilities from scientific workstations³. The ICN is a large heterogeneous network of supercomputers, mainframes, special purpose research computers, workstations, and personal computers. In addition to computing power, the ICN provides a centralized storage system (CFS), a graphics and output system (PAGES), and a job scheduling and routing service (FOCUS). The Unix operating system will become the Laboratory standard operating system. There will be a demand for network transparent window systems and graphical interfaces to ICN resources. High-speed networking which accommodates voice and video will evolve over the decade. Computing power and storage requirements will continue to increase at both the Central Computing Facility and at the user's site in order to support the changing computing environment.

The trend of the new development work in the ICN will be to give the workstation users a more transparent view of the ICN resources. Networking will move toward standard interfaces and protocols. High-speed file transfers will become available when the Los Alamos Multiple Crossbar Network⁵ (MCN) or an equivalent high-speed network comes on line. The Los Alamos ICN will continue to be a multipartition security environment. Our major customers will continue their emphasis on developing and using large codes that consume a significant amount of ICN resources. New applications based on workstations will increase productivity resulting in more and different use of ICN resources.

THE COMMON FILE SYSTEM

The Common File System¹ (CFS) is the centralized file storage system used to support the storage needs of users on the ICN. CFS provides an essentially unlimited file store which is used by the other computers in the network for active and archival files. CTSS (currently the major Los Alamos supercomputer production operating system) does not provide permanent disk storage for user files. Users are responsible for storing their permanent files in the CFS. CFS uses a hierarchy of storage devices consisting of IBM 3380 disks and IBM 3480 cartridge tapes and determines where to store a file initially on the basis of the file size. Migration software is run on a periodic basis to move files up and down the storage hierarchy based on user specified activity level, file size, frequency of access, storage availability, and the storage capabilities of the various storage devices. The CFS software operates on IBM computers under the MVS/SP or MVS/XA operating system. It is a multitasking system written primarily in PL/I. CFS has been in operation at Los Alamos since mid 1979.

The CFS capabilities are provided through a standard set of requests⁶ that allow users to store and retrieve files, and to build and manipulate tree structured directories that assist in the management of their files. The user request interface is through an application level utility that is available on all network machines. Some of the more common requests are CREATE a root directory, ADD subdirectories, SAVE files, GET files, DELETE files, and LIST information about files and directories.

The CFS software has protection mechanisms that enable it to function in a security environment of classified and unclassified data along with cleared and uncleared users. These facilities have allowed Los Alamos to run in a multipartition mode consisting of four different network partitions including a completely Open network and two Secure networks restricted to cleared users. The CFS is connected to each of these networks and is able to service the file storage requirements of each in a secure fashion.

CFS Usage Statistics and Growth

Storage and data transmission requirements for the Los Alamos computing network have increased dramatically since CFS was first made operational in mid 1979. These requirements have closely correlated with the growth of the raw computing capacity available on the network.

At the end of 1980, after one year of use, CFS contained 155,000 files representing 1.46 trillion bits of data. In an average week, 50,000 file transfers occurred with an aggregate throughput of 2 trillion bits. As of Fall 1989, CFS stored 4 million files of which 31% were on disk and the remaining 69% were on tape. This represented a total data store of 100 trillion bits of which 97% were on tape and 3% were on disk. The growth of CFS for 1989 is estimated at 45 to 50 trillion bits. The peak weekly growth is 1.5 trillion bits. In late 1989 on a typical day 28,000 file transmissions are done with an aggregate throughput of approximately 800 gigabits.

EVALUATION OF CURRENT CFS CAPABILITIES (PROS, CONS)

Input from users of the Common File System provided a summary of positive and negative aspects of CFS. Different types of users often have conflicting opinions about aspects of CFS. Workstation users tended to want more flexible features while the users of supercomputers want more performance oriented features.

Performance, reliability, availability, and data integrity are considered very good qualities of CFS. New storage devices are easy to add as required. The CFS interfaces have good functionality and are reasonably easy to use. The security rules have been implemented well and allow the operation of a single file server to support all computing partitions. Files can be accessed from any machine in the network as long as security rules are followed. The demand for increased storage and faster file transmission has been met by CFS.

On the negative side, CFS does not support wild cards and subtree operations. Standard networking protocols and a Unix or SUN Network File System⁷ (NFS) type interface are not supported. File sharing needs to be enhanced by allowing group access to files. The CFS LIST command does not have options such as an alphabetically ordered listing and does not support the Unix 'ls' command format. File locking, automatic file backup, and file versions are not supported. Tools to allow users to automatically manage their storage and files are not adequate. Supercomputer users want faster file transmission rates for their massive data transfer requirements. The CFS file name size is not long enough for Unix users and a convenient way for workstation users to transmit an entire directory to CFS is not available.

OTHER MSS SYSTEMS REVIEWED

We reviewed several existing and planned systems and evaluated them for possible use at Los Alamos.⁸ Information was obtained from papers, manuals and private communications. The major systems we reviewed were Carnegie Mellon's Andrew File System (AFS), University of Michigan's Institutional File System (IFS), Sun's Network File System (NFS), NASA Ames Research Center's Mass Storage System II (MSS II), Lawrence Livermore National Laboratory's MSS (LLNL MSS), University of California at Berkeley's Sprite File System, and Control Data Corporation's CYBER Storage Server. We also discussed hardware and data storage plans with IBM and Cray Research Inc. Ten other systems were also reviewed.

Many of the systems reviewed (e.g. IFS, MSS II, LLNL MSS) are in development or in preliminary usage so that it is impossible to determine their ultimate performance and capacity. Some systems (i.e. NFS⁷, AFS⁸, Sprite⁹) are not designed to support the more than 8400 CFS users and a yearly growth rate of 50 trillion bits. The Unix based file systems are attractive for their transparent user interface (i.e. NFS, MSS II, Sprite, CYBER Storage Server), however, the Unix security implementation is not acceptable for our multipartition security environment. The Unix file system as a mass storage platform⁷ has serious problems that require kernel modifications to solve. For example, NASA Ames⁷ had to modify Amdahl's UTS operating system to allow for a large number of files, large file sizes, as well as incorporate performance enhancements. NASA Ames is working toward standardization of their Unix modifications. If this standardization is achieved, the Unix file system will become a more desirable platform for a mass storage implementation.

CONCLUSION

A new or enhanced data storage system must provide at least the functionality of the current CFS. It must provide the same or a greater level of responsiveness, reliability, security and expandability that has typified the CFS. It must provide enhanced service to users at high technology workstations and those using the Unix based supercomputers in the ICN. It must continue to provide good service to users who use conventional terminals.

The available and planned mass storage systems we reviewed do not provide all the capability, reliability, performance and security controls that exist in CFS. Some are stronger in some areas than CFS but no other system provides as good an overall solution for the Los Alamos environment as the CFS. However, CFS needs many long overdue enhancements to improve the user and networking interfaces. These enhancements need to maintain the benefits of CFS such as performance, reliability and security but replace or improve weak areas such as the network interface and the user interface for Unix users.

GENERAL RECOMMENDATIONS

The recommendations are divided into short term (1990-1992), mid term (1992-1995), and long term (1995-2000). The changes and additions specified will allow CFS to continue to meet the data storage needs of Los Alamos National Laboratory. We have also recommended that a significant effort be maintained to track hardware and software developments in mass storage technology to determine if at some point they may enhance or replace CFS. Because of the large amount of software changes needed, we have also recommended that the latest software methodology tools and techniques be used for software development in order to continue producing a quality product.

SHORT TERM RECOMMENDATIONS (1990 - 1992)

The short term recommendations support the transition to Unix as the standard operating system at the Laboratory and high performance networking.

Implement a Unix File Transfer Program (FTP) Interface and Standard Network Protocols in the CFS. - Develop a FTP interface to the CFS. This will provide a Unix like interface to CFS that supports full file movement. Restructure CFS so that a commercial TCP/IP package can be installed with minimum modifications. Implement other features such as increasing the maximum file size and adding wild card capability.

High Performance Networking - Implement the High Performance Parallel Interface (HPPI) channel to increase data transmission rates through the network and connect to the Crosspoint Star⁵ (CP*) experimental testbed. Implement a disk striping system to speed up disk access for massive files. Install an automatic tape library to reduce operator tape handling overhead.

CFS Graphic Interface for Workstations - Add features to better support graphical interfaces to CFS for workstation users.

Security - Implement intrusion detection tools, upgrade acquired software to Los Alamos standards, and support security changes needed to support Unix on the supercomputers.

MID TERM RECOMMENDATIONS (1992 - 1995)

The mid term recommendations support the transition to a Unix type file system that has a significant level of file location transparency.

Unix File and Path Names - Support file and path names consistent with the Unix length and character set standards.

Transparent Interface to CFS - Implement file block caching to support byte level access to CFS files. Support a Unix interface to CFS that is transparent in the sense that the users and application programs view and use CFS files in the same way as local files. Provide users with a locking mechanism to control concurrent access to the same data, user controlled file archiving, and backup.

Transition to a Standard Mass Storage System - Restructure CFS to support appropriate standards that may have evolved from standardization efforts such as the IEEE Reference Model for Mass Storage Systems¹¹.

Security - Participate in the development of a security model for the workstation environment.

Production Version of the Multiple Crossbar Network (MCN) or an Equivalent High-Speed Network - Support new requirements needed for the production version of MCN or an equivalent high-speed network implementation.

LONG TERM RECOMMENDATIONS (1995 - 2000)

The long term recommendations are more speculative but support the continuing transition to standard network and file system components used in a distributed environment.

Standard File Systems - Implement a high performance network disk system with the goal of moving all rotating storage devices off the supercomputers. Develop a standard file system interface for supercomputer solid state disk. Support operating system controlled file migration between client computers and network storage.

New File System Technologies - Implement a transparent, possibly distributed, file system with a global file space.

Standard Security - Implement within the MCN a standard and approved security control system.

Faster Networking - Implement the successor to the MCN.

Multiple Networking Interfaces and New Devices - Support Mass Storage Controllers with super fast input/output capability and push for File Transport protocols and network protocols implemented in hardware.

REFERENCES

1. Collins, Maurice W, Catherine W. Mexal, "The Los Alamos Common File System," *DIGEST OF PAPERS, Ninth IEEE Symposium on Mass Storage Systems*, October 31 - November 3, 1988.
2. Christman, Ronald D., Danny P. Cook, Christina W. Mercier, "Data Storage for the 1990s," (internal publication LA-UR-89-3697) Los Alamos National Laboratory, October 17, 1989.
3. Phillips, Dick, Chair of the User Interface Team, et al "The Los Alamos Computing Environment for the 1990s," (internal publication) Los Alamos National Laboratory, March 10, 1989.
4. Blaylock, John, Chair of the UNICOS Workplan Group, et al "UNICOS Migration Workplan," (internal publication) Los Alamos National Laboratory, October 1989.
5. Hoebelheinrich, Randy, Richard Thomsen, "Multiple Crossbar Network: A Switched High-Speed Local Network", 14th Conference on Local Computer Networks, Minneapolis, Minnesota, Oct 1989.
6. "Common File System - CFS Interface Reference," (internal publication CIC#895) Los Alamos National Laboratory, December 1987.
7. Sandberg, Russel "The Sun Network File System: Design, Implementation, and Experience" Printed by Sun Microsystems Inc., 1987.
8. Howard, John, Michael Kazar, et al, "Scale and Performance in a Distributed File System", *ACM Transactions on Computer Systems*, Vol.6, No. 1, Feb 1988.
9. Nelson, Mike, Brent Welch, John Ousterhout, "Caching in the Sprite Network File System", *ACM Transactions on Computer Systems*, Vol 16, No. 1, Feb 1988.
10. Henderson, Robert L., Alan Poston, "MSS II and RASH, A Mainframe UNIX Based Mass Storage System with a Rapid Access Storage Hierarchy File Management System", *USENIX*, Winter 1989.
11. Miller, S. W. "IEEE Reference Model for Mass Storage Systems," *Advances in Computers*, Vol.27, 1988, pp 157 - 210.