MARTIN MARIETTA ENERGY SYSTEMS LIBRARIES
3 4456 0148809 4

ORNL/TM-10172

ornl

OAK RIDGE NATIONAL LABORATORY

MARTIN MARIETTA

ADVANCED DISTRIBUTED PROCESSING WITH FOCUS AND PC/FOCUS: PLANNING CONSIDERATIONS AND PHASED IMPLEMENTATION

Cathrine E. Snyder
Donald R. Hanlon, Jr.
Linda J. Bettis
Joanne O'Malley

OAK RIDGE NATIONAL LABORATORY
CENTRAL RESEARCH LIBRARY
CIRCULATION SECTION
4500N ROOM 175
LIBRARY LOAN COPY
DO NOT TRANSFER TO ANOTHER PERSON
If you wish someone else to see this report, send in name with report and the library will arrange a loan.

OPERATED BY
MARTIN MARIETTA ENERGY SYSTEMS, INC.
FOR THE UNITED STATES
DEPARTMENT OF ENERGY

Printed in the United States of America. Available from
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Road, Springfield, Virginia 22161
NTIS price codes—Printed Copy: A03 Microfiche A01

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.

ORNL/TM-10172

Energy Division

ADVANCED DISTRIBUTED PROCESSING WITH FOCUS AND PC/FOCUS: PLANNING CONSIDERATIONS AND PHASED IMPLEMENTATION

Cathrine E. Snyder

Donald R. Hanlon, Jr.
University of Tennessee

Linda J. Bettis
The Davis Group, Inc. Washington, D.C.

Joanne O'Malley
Navy Regional Data Automation Center, Washington, D.C.

Date of Issue: November 1986

Prepared by
Oak Ridge National Laboratory
Oak Ridge, Tennessee 37831
operated by
Martin Marietta Energy Systems, Inc.
for the
U.S. Department of Energy
under Contract No. DE-AC05-84OR21400

MARTIN MARIETTA ENERGY SYSTEMS LIBRARIES

3 4456 0148809 4

CONTENTS

ABSTRACT v

1. INTRODUCTION 1

2. PLANNING 2

2.1 CENTRAL PROCESSING 2
2.2 LOCAL CONFIGURATION 3
2.3 REMOTE COMMUNICATIONS 4
2.4 LOCAL APPLICATION PROCESSING IN A SHARED DATA BASE ENVIRONMENT 6
2.5 DISTRIBUTED DATA BASE 7

3. PHASED IMPLEMENTATION: INTRODUCTION INTO THE USER ENVIRONMENT 9

3.1 LOCAL MICROPROCESSOR SOFTWARE 10
3.2 LASER PRINTER 11
3.3 LOCAL AREA NETWORK 12
3.4 DISK SERVER 12
3.5 SPLIT DATA BASE 14
3.6 FULL FILE SERVER FUNCTIONALITY 14

4. CONCLUSION 17

5. REFERENCES 18

FIGURES

Fig. 1 Phase 3 system configuration 13

Fig. 2 Phase 6 system configuration 16

ABSTRACT

A distributed data system that included shared data resident in a mainframe/PC local area network environment was successfully introduced to nontechnical end-users. The system provided the end-user with the control of responsiveness and reliability inherent in local application processing, along with the back-up support and archival storage/retrieval provided by a mainframe computer. A semaphore-driven locking scheme ensured referential data integrity across the shared data system, and HASP protocols provided fast and reliable file transfer capability. The project demonstrates that distributed methods can be applied to information systems to increase their efficiency and productivity. When the resources are on hand for such a system, planning and a phased approach to the appropriate implementation can bring success even in an end-user environment.

1. INTRODUCTION

As microprocessors become commonplace on the desktops of users who routinely access a mainframe data base in their daily work, many opportunities present themselves for taking advantage of the local processing power of the personal computer. However, managers have often provided their data processing workers with microcomputers solely for the purpose of doing word processing and spreadsheets, while restricting the microcomputer's use in data processing to accessing the mainframe as a dumb terminal. In today's fast-growing data processing environments, central mainframe access can be slow and unreliable. Why neglect the microcomputer's capability to be a part of a distributed system that reduces the users' dependence on the mainframe environment?

The fact is that although distributed data processing with microcomputers can be the solution to the lack of responsiveness and reliability often experienced by the users of mainframe computers, such systems have rarely been installed in purely end-user situations. One roadblock to the implementation of a distributed system in a micro/mainframe environment is lack of technical expertise on the part of the users, and a lack of time in the daily routine for users to gain that expertise. However, more effective data processing with enthusiastic user acceptance can be attained if distributed technology is brought to the end-user through careful planning and a phased implementation that does not interrupt his or her busy schedule.

A recent prototype project executed for the Navy Regional Data Automation Center (NARDAC) In Washington, D.C. illustrates the good results obtainable when introducing distributed processing with a local logic component into the end-user environment. The system was assembled at the Office of the Navy Comptroller Financial Control Division (NCB-3), which is responsible for Navy budget execution tracking. The users' previous environment was a central processor operated by NARDAC, with a data base utilizing the fourth generation language FOCUS residing on an Amdahl V7/A under MVS/TSO. The user accessed the data base and performed application processing in terminal emulation mode from IBM XTs and used Data 100 workstations for high speed remote printing.

The project goal was to set up a prototype of a distributed environment that simulated the users' central data base access. State-of-the-art technology in data processing and communications was to be used, and new developments in hardware and software were implemented throughout the project. The micro/mainframe system that evolved has improved data processing reliability and responsiveness as perceived by the user; its successful introduction into a nontechnical environment was made possible by rigorous attention to planning and phased implementation.

2. PLANNING

At NCB-3, the success of the distributed system implementation was due in large part to the considerable time spent in planning. Although planning is an essential high-level activity, it is often bypassed when it comes to microcomputer acquisition. This can be a critical error.[1] End-users, especially, need the security of knowing that their changing environment is based on an orderly process in which they participate and which they accept.

A good approach to planning the acquisition of a distributive processing system is to outline the configuration of the new system based on a thorough understanding of the current data processing environment and on the functional aspects of how the users operate their applications. The plan should also include a comprehensive implementation procedure.

A modular approach to the planning process facilitates discussions with both users and their managers, and is easily accomplished when planning a micro/mainframe system. In the modular approach to the project discussed here, the authors considered separately five elements of distributed data processing: central processing, local configuration, remote communications, local applications processing, and distribution of the data base.

2.1 CENTRAL PROCESSING

The first module of the planning effort addressed two characteristics of central mainframe systems that generally cause the greatest end-user frustration: slow response time and lack of reliability. Such frustration could be eliminated by removing the mainframe from the users' environment and relying on a local processor, such as a minicomputer, for all application work.

However, some of the good characteristics of central processing are difficult to provide at the local level. The central facility at NARDAC Washington, like most other data processing centers, has procedures for backup and recovery, done automatically at the central site. This service would be difficult and impractical to duplicate in an end-user environment. Another important mainframe service is large and accessible storage capacity. The archiving, tracking, and

retrieving of history files are not activities that suit end-users, but like most users who process large amounts of data, the NCB-3 users needed this capability.

The plan for NCB-3's distributed system configuration called for the central processor to be a vital part of the final system, providing at the very least system back-up and archiving. The implementation plan also envisioned a phased approach to application processing, where only part of the data base and its applications would be brought to the local level in a prototype situation. Only after the prototype had been successfully implemented would total application processing be attempted at the local level, retaining the mainframe back-up and archiving capabilities.

2.2 LOCAL CONFIGURATION

After considering the role of the central processor, the next logical step in planning for a distributed system is to configure the local components. We considered three options for the NCB-3 local environment:

1. Continue to use the IBM PC XTs as terminals, or dumb workstations, and connect them to a local mini- or other micro-computer.

2. Use the XTs singly as stand-alone micros, bring down data base segments one at a time to each XT, process on the micro, and send the updates back to the mainframe.

3. Connect the existing PCs in a local area network (LAN) and let them work together.

The option of using the IBM PC XTs as terminals did not meet the project goal of using new technology, nor would it exploit the full potential of the XTs. The second option, using the PCs as stand-alone micros, could take advantage of the processing power inherent in the XTs but did not provide the shared, multi-user data base access required by the NCB-3 staff. Therefore, the third option, connecting the XTs in a LAN, was chosen for the NCB-3 environment. This allows the micros to have shared access to the data base, to share peripherals such as a laser printer, and to communicate with each other[2].

Once it was decided that the distributed system called for a LAN, a specific product had to be chosen from several available on the market. The recent emergence of standards in

local area networking allows the selection process to focus on performance, functionality, and cost[3] as opposed to a specific brand of hardware. Many articles on the attributes of various LANs are available in the open literature[4,5] to help the planner in making a choice; the important thing to remember is that the product(s) should be chosen based on the users' business needs and office requirements.

The planners for NCB-3 had three functional criteria for their LAN choice: device support (IBM PC XTs, file sharer, printer, and external communications via modem); applications support (the network's ability to support use of PC/FOCUS and to transfer data to and from the Amdahl mainframe data base); and the immediate availability of the LAN product. Several LANs fit these criteria, but the planners at NCB-3 had two other criteria they wished to meet. First, the LAN should support a nonproprietary server in order to allow the evaluation of the IBM AT, which had then just come on the market. Second, the LAN should be compatible with the phased introduction of the sophisticated file server software marketed by Novell in order to free the end-user from the technical task related to space allocations on a shared disk. The 3-COM LAN met all these criteria. Because it was also well received in the literature surveys[6], it was chosen for the NCB-3 project.

2.3 REMOTE COMMUNICATIONS

A plan for distributive processing must next specify the method of transferring data between the mainframe and the local environment. Ideally, this transfer should be accomplished at high speed without errors. There are other desirable features, such as a query system for file transfer status, restart capability, and character blocking for efficient transfer of large files. Practically, however, there may be constraints placed by the central host and line equipment available.

The search for a robust file transfer protocol for the NCB-3 distributed processing system uncovered many issues. Very early it was found that when data files are transferred from a hierarchical data base, file expansion occurs. A 3.2 megabyte FOCUS file could expand into 18 megabytes of "flat," transferrable file. Using the asynchronous lines (1200 BPS) available to connect the NCB-3 user with the mainframe at NARDAC Washington, 18 megabytes of data would take about two days to transfer under ideal conditions. Considering retransmittals necessary because of noisy lines, line hits, or bad data, it might take considerably longer to transfer such a file. Clearly, this was not acceptable. In planning for a

better solution, we considered two ways to handle communications from the mainframe to the local environment with FOCUS:

1. Use PC/FOCUS standard software with its PC-HOLD command to create the flat files, and the PC/FOCUS transfer command, XFER, to send them by asynchronous file transfer, but limit the number of records in the downloaded segment.

2. Use methods other than PC/FOCUS software for more robust file transfer.

While using PC/FOCUS Release 1.0 with 1200 BPS asynchronous communications, it was possible to transfer 658 records of 150-200 characters each in about fifteen minutes. The later PC/FOCUS release 1.5 has enhancements such as error correction and dynamic blocking that speed up file transfer, given good asynchronous lines. However, with any asynchronous transmission, there is a limitation in line speed; it requires much valuable time when users download a sizable segment of the data base. Another problem is that most asynchronous transmission protocols have no restart capability. If a user nears the end of a large transfer and loses the line, he would have to start again from the beginning (a nerve-wracking process that destroys productivity).

Another PC/FOCUS capability considered for the project was the interface to IRMA, the bisynchronous 3270 package. However, the NCB-3 site did not have 3270 communications to the data center. Because IRMA is also limited in line speed (9600 BPS), we chose not to pursue obtaining 3270 service.

Our second option was to evaluate communications capabilities other than FOCUS. There are asynchronous file transfer products, such as Smartcom and Crosstalk, that provide better error correction (based on the XMODEM protocol) than did PC/FOCUS Release 1.0. Kermit, produced by Columbia University, also has restart capability and is quite popular. But all of these, being asynchronous protocols, have the same outstanding limitation as FOCUS software—they are slow.

We therefore considered various synchronous communications protocols. The 2780/3780 remote job entry products, such as those produced by AST Research and Winterhalter, provide some beneficial features: data can be blocked at 512 characters; there are automatic retransmissions of data and data compression capabilities; and a line speed of 19.2 KBPS can be attained. Certainly, this provides a more robust file transfer, but there is a serious drawback in that these

2780/3780 protocols were designed for workstations. They "think" in terms of cardreaders and printers, providing an 80-character record size for cardreaders and a 144-character record size for printers. Odd-sized or lengthy records in FOCUS would be difficult to transfer. An additional limitation of these protocols is that they have no restart capability.

A better choice for on-line use is the Houston Automatic Spooling Program (HASP), marketed by Barr and Persyst. This protocol is similar to the 2780 protocols but has better compression capability and supports faster line speeds (up to 56 KBPS). Also, the HASP protocol can deal with the larger records one expects with FOCUS. Because of multileaving features, it also allows such capabilities as querying the host system in the middle of a file transfer to ask, "Where are we?"

For the users of our distributed system, we planned for the most practical remote communications solution that would allow data transfer without disrupting their customary workday. We decided on a two-phased implementation, first making use of the facilities immediately available—the 1200 BPS communication lines and the PC/FOCUS asynchronous software. We solved the two-day transfer problem by downloading the data base in "segments"; these segments of data were arranged within a certain size range that would transfer along 1200 BPS lines in 15 to 20 minutes. This worked well because users rarely needed more than several segments at one time. In other situations, it may be necessary to plan for a faster file transfer immediately. We, too, planned to have a higher performance file transfer as soon as possible in order to download the entire data base, and required the NCB-3 users to work at 1200 BPS for only four months before we shifted to the HASP products.

2.4 LOCAL APPLICATION PROCESSING IN A SHARED DATA BASE ENVIRONMENT

When there is a requirement for simultaneous multiuser access to the data base, maintaining data integrity within the system is a concern that must be addressed. In the mainframe FOCUS environment, data integrity is provided by the FOCUS Simultaneous Usage (SU) facility. However, once shared data base access is established on the LAN, PC software is required to handle the update of each individual record. Without this software, data integrity is threatened. For example, let's say that two users retrieve the same segment of the data base

early in the same day and process some of the data. Whoever finishes processing first and restores the modified data will have his updates wiped out when the second person later restores the same segment.

Two methods can be considered for preventing such losses and maintaining data integrity in a shared data base environment:

1. Use PC data base software that furnishes data integrity protection.

2. Engineer custom software for multi-user access to the data.

At the time of the NCB-3 planning effort, IBI was working to develop FOCUS software that would operate on the local level as an SU facility, but it would not be ready within the time frame of our project.* In fact, many PC data base management systems that might be used in a LAN environment do not furnish data integrity protection in a multi-user situation.

However, the second method for preserving data integrity, engineering a limited locking capability with custom software, will be satisfactory in many situations. For the NCB-3 locally shared data base, semaphores and a pair of programs (lock and unlock) written in C provided the capability of preserving data integrity when several workers shared the data base. When the NCB-3 user downloads a group of records from the server to his workstation, a lock is set on that group. No one else can get access to that data while it is being processed. When the worker is finished, the group of records is automatically unlocked after it is successfully restored to the server.

However, engineered locking is not the ultimate solution for maintaining data integrity in a shared data base on a LAN because it is possible to access the data base directly without going through the menus that call up the engineered programs. Data integrity would then be compromised despite the semaphore lock. But the use of semaphores allowed us to set up a practical local processing situation, and the technique should be appropriate for most LAN users.

---

*This PC software is now available from IBI for the IBM PC Network and Nestar Systems Inc. Plan Series network.

2.5 DISTRIBUTED DATA BASE

The data protection just discussed occurs on the local level, but in a true distributed environment one also has to plan for a broader issue inherent in the fact that the data base resides in multiple locations. It is essential to plan for update coordination between the possible places where the real data could be found. In our distributed system, the data were moved from the mainframe down to the LAN, and from the LAN server to the workstations. In the case where the database software is designed to be distributed, it may not be necessary to move the data in this way. There seem to be three ways to determine where the data is located at any given time:

1. Download the entire data base when any portion of it is needed on the network.

2. Use software, perhaps data base software, to keep track of where the data reside at any given time in a method transparent to the user.

3. Download a group of records from the data base and have the user manually keep track of where the data reside.

In the NCB-3 prototype, which included the use of live data on the LAN, the data could be found in three different places: on the mainframe, on the server, or on the individual workstations. Because we had already decided to start by using asynchronous communications, the first option of downloading the entire data base was discarded temporarily because of the lengthy file transfer time.

The second option, software that supports a true distributed data base environment, is not available at the present time. However, several software vendors are taking advantage of recently established LAN standards to develop more sophisticated software[7]. In a method transparent to the user, this future software will be able to locate any data, regardless of its physical location within the system, and to automatically update changes in all instances of the data each time they are uploaded.

It may be some time before a true distributed data base program is available to the end user, but this should not be a deterrent to setting up micro/mainframe distributed systems today. System designers should instead be aware of the need to remind the users to keep track of where the "real" data

base exists at any point in time. The NCB-3 users manually kept track of the location of their data without problems despite the fact that they were accustomed to updating data on the mainframe without taking the LAN into consideration.

3. PHASED IMPLEMENTATION: INTRODUCTION INTO THE USER ENVIRONMENT

Although a good planning effort is essential, the method of introducing new technology into the end-user environment is also important, both to user acceptance and to creating improved data processing. The ADP professional needs to consider the problem of teaching the users new technology, and the challenge of evaluating new technologies for their use in production environments. There are basically two ways of approaching system installation: (1) implement all new technology at once, or (2) prepare a phased implementation.

The first option has the virtues of simplicity and speed (i.e., bring the technology in, set it up, train the users, and leave). However, this procedure seldom leaves the user happy or ready to tackle the new system. The proliferation of information centers to assist end-users attests to the fact that hasty implementation of new technology is wide-spread and not entirely satisfactory. End-users who do not have access to such information centers will need some other method of getting started in order to prevent frustration when required to use technology they do not have time to master[8]. Modern training methods (such as computer-aided instruction, interactive video, and self-help menus) have been developed, but they are not always available to small groups, and do not in and of themselves guarantee successful introduction of new systems to the nontechnical user.

We recommend phased implementation as a method that allows the introduction of new techniques into daily routine without user frustration. Frustration can be avoided when appropriate training is included with each phase, supported by stepwise production of clear documentation. The documentation can be created and updated as the phases progress, taking advantage of user input along the way to ensure a responsive product.

Phased implementation allows the users to practice what they learn before they tackle the next complication, so that the hours spent in training can be productive. It also lends itself well to a prototype situation, where planners, system designers, and users seek a good understanding of how the new

system will function[9]. Adjustments and enhancements can easily be made in ways that will be cost-effective and efficient. Phased implementation should be tailored to the specific end-user environment, but the method used at the NCB-3 site will serve as an appropriate example of the approach.

We chose to implement distributed technology at NCB-3 in six distinct phases: local microprocessor software, laser printer, local area network, disk server, split data base, and the addition of a file server to provide full LAN functionality. These or similar phases would be appropriate in many situations. The important thing is to follow a rigorous planning effort with a phased implementation that is flexible, responding to lessons learned about user capabilities and system functions.

3.1 LOCAL MICROPROCESSOR SOFTWARE

In the first phase of system installation, we limited the changes for the users to bringing some of their current mainframe FOCUS application processing down into a stand-alone microprocessor (IBM PC XT) using PC/FOCUS. The object of this phase was to see how the PC data base software would perform and how it would be accepted by the users. Because of the differences between using PC/FOCUS on a micro and the processing with FOCUS software on the mainframe, there were some training issues involved. In fact, the amount of training required for the users was greater than anticipated. We had assumed that because the users were already working with XTs, they would be familiar with DOS. This proved to be an incorrect assumption because the users had used the XTs only as dumb workstations and had not learned to use DOS. This confirmed that flexibility in training has to be a part of the implementation process and that one should not assume anything about a user's familiarity with equipment and methods. The first sections of the user manual were developed and included tailored instructions for DOS.

Although the NCB-3 prototype was chiefly intended to take advantage of the XTs and to improve systems reliability and responsiveness, an unforeseen benefit of using the PC data base software was immediately realized. The user was accustomed to having asynchronous TTY communications with the mainframe data base, which meant being restricted to editing one line at a time. When the data base was moved down to PC/FOCUS, full-screen capability was available for data entry in the form of the PC/FOCUS utility, FIDEL. Originally we had

intended to do no programming at all and to bring down the applications exactly as they existed on the mainframe; however, FIDEL was such an obvious advantage to the users that we decided to rewrite the main menus and the update process to allow its use. This was no great complication, taking only a few days of FOCUS programmer time. The actual downloading of data in the chosen segment size—approximately 500-600 records at 132 characters per record (about 66 kilobytes)—took about 15-20 minutes with asynchronous 1200 baud communications. The users did not complain about this, perhaps because they were warned ahead of time that downloading FOCUS files was going to take time. With some personal training and help, our users were now enthusiastic about their new environment and were ready for the next phase to begin.

3.2 LASER PRINTER

In the second phase, the FOCUS application system reporting functions were made available at the PC/FOCUS level by making a laser printer available to the local PCs. This involved making some adjustments in the users' FOCUS reporting programs having to do with height and width of the printout. There were also some adjustments that had to be made in the PC environment because of memory limitations, but we were able to split the FOCUS programs in a satisfactory manner.

There are exciting capabilities inherent with laser printers, such as a variety of font types and options for orienting the report on the paper. However, FOCUS and many other PC software packages do not supply print drivers to take advantage of the font and orienting options. This problem can be solved by using additional software. For the NCB-3 project, we used a package called PRINTMAN, a menu-driven printer utility. Using PRINTMAN, it is possible to choose fonts by creating copy files, which go to the printer and enable it to set the correct escape sequence. However, with software such as PRINTMAN one cannot drive the laser printer to mix fonts in a single FOCUS report. Through PRINTMAN, we provided the users with a default option of orienting the reports in landscape mode because of the width of their reports. We updated the user documentation to include laser printer operations and report production base on the PRINTMAN functions.

The laser printer was also beneficial because the users had been producing some very large reports, pasting them together, and then reducing them on a copier. With the laser

printer, they are able to reduce the size of the type and directly print most of their reports on 8 1/2" x 11" paper.

Once the users became capable of using the laser printer for reports produced with their local processing, they discovered another unforeseen benefit of PC data base software: PC/FOCUS included a facility for ad hoc reporting called TABLETALK. This software had not been available for use with their TTY connection to the mainframe. The NCB-3 users get congressional requests for single pieces of information, to which they had been taking two or more days to reply. With their local processing, TABLETALK, and the laser printer, reporting is so much easier and faster that the user can now reply to such requests within half a day.

We found that the user training with TABLETALK and the laser printer was very simple, requiring only about an hour of demonstration before the user was ready to begin. However, more software needs to be developed so that these users and others like them can avail themselves of all the features of laser printer technology.

3.3 LOCAL AREA NETWORK

The purpose of Phase 3 was to set up a functional local area network. Setting up a computer network calls for a new organizational component: someone must be in charge of the network.

However, user training and demonstration of the network capabilities at NCB-3 were simplified because we had opted not to train the users in the network administrator functions. In most end-user situations, the network administration function is perhaps best filled temporarily by an ADP professional. This allows the users to become acclimatized to their new environment in a stress-free manner. As use of the LAN proceeds, it will become apparent who among the users displays the curiosity and inclination to be given authority as system administrator. The training of the user responsible for the LAN will then be much easier because he has had a chance to become an experienced user first.

We found that simply sharing a printer is a good way to introduce end-users to the advantages of sharing resources on the LAN. The first LAN configuration at NCB-3 therefore

consisted of three inter-connected XTs, an AT that was brought in to act as a server, and a laser printer (Fig. 1). The users readily used the laser printer, providing a good test situation for the network and the updated user documentation.

3.4 DISK SERVER

The introduction of file sharing by implementing the disk server functionality of the LAN software into the local area network is the logical next phase. By "disk server" we mean programs that provide disk sharing for the users, displaying the files on the shared storage device and transferring them at the users' command.

To implement this phase, the NCB-3 users downloaded from the mainframe the same files they had been working with on the isolated PCs to the shared disk on the AT. By itself, the disk server provides no locking system that will preserve data integrity. Therefore, the custom-engineered record-locking scheme was installed with the disk server to prevent users from downloading the same data at the same time. In addition to new system menus to guide them through this downloading process, the users were provided with updated documentation. The engineered locking scheme* was successful in preserving data integrity.

3.5 SPLIT DATA BASE

The distributed system prototype is more likely to be successful if the next phase comprises a gradual increase in the amount of mainframe data distributed, and hence in the amount of local application processing. Careful analysis of the users' work and data will reveal how best to go about this in different prototypes.

At NCB-3, the data were split according to the functions of an application group, that is, a group whose duties required the users to access only certain distinct parts of the data base. Each application group could then easily download its split portion of the data in its entirety to the LAN and do all the application processing on the local level.

---

* The C code written for the locking of FOCUS data files is available upon request from the authors.

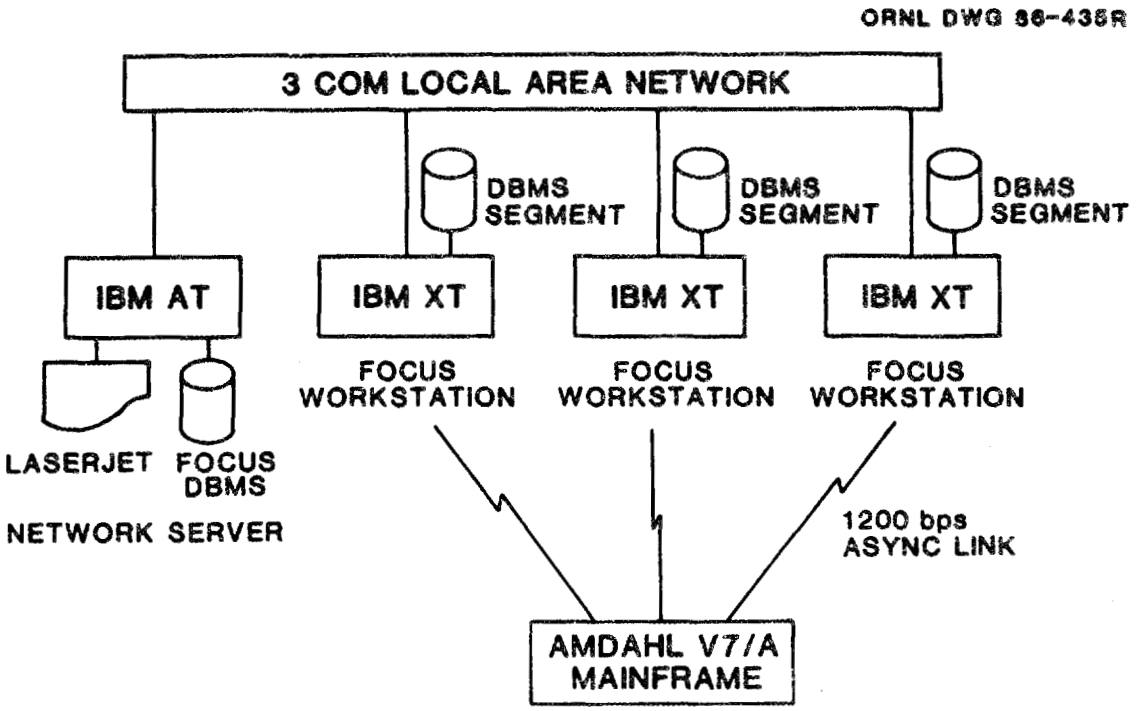

Fig. 1. Phase 3 configuration of the NCB-3 prototype.

A group that was accustomed to downloading 18 MB of data might only need 3 MB, a reasonable amount of data to download with the asynchronous communication being used in the early stages of the project. System menus and user documentation were again updated to reflect these changes.

3.6 FULL FILE SERVER FUNCTIONALITY

True file server functionality can only be realized by implementing file server software. We define file server functionality as follows:

* Disk allocation defined at the directory and subdirectory level;

* Data protection provided at the file level through lockout on files in use to prevent multiple user updates from destroying data integrity;

* User access rights defined on a file-by-file basis; and

* Necessary disk space dynamically reallocated as files grow in size.

Although LAN manufacturers do not necessarily include file servers in their systems, Novell Inc. markets file server software compatible with MS DOS 3.1 that can be used with Novell LANs or installed on top of other vendor's LAN hardware[10]. Using only a disk server (i.e., lacking file server capability), every computer in the system must keep a file allocation table (FAT), which keeps track of the areas on the disk that are available and those that have been accessed by other computers. With many users on the system, this becomes complicated, and one device on the system must be made responsible for updating all of the FATs. Errors and discrepancies are common, and there is a definite threat to data integrity.

MS-DOS 3.1 provides the Redirector, a program that reroutes requests from local devices and sends them to remote devices, allowing the file server to automatically keep a FAT for the whole system. All requests for information by the other computers can be sent via MS DOS 3.1's Redirector to the file server, which then accesses the disk, taking into account all other active or pending requests and dynamically allocating storage space. Compared to a disk server, this system

is far more orderly and controlled, and it is less prone to failure[10].

Beyond preserving data integrity, fully functional file servers can provide other valuable services such as automatic backup and recovery. They also allow users to move to different workstations without physically moving any secondary storage devices, and allow workstations to operate without using disks. A high-level file server can store, retrieve, or update the data base by individual files, or individual pages within the files, or at the byte level where the update and retrieve fields can vary in length[11].

At NCB-3, the Novell software performed as expected. Because the users had been using our engineered record--locking, the introduction of the file server did not affect them directly. The network administrator role was, however, now simplified considerably. The system menu remained the same, but the user documentation was updated to reflect network command changes. Documentation for network administration is being developed for later inclusion with the user documentation. The configuration of the mature system as it now exists (Phase 6) is illustrated in Fig. 2.

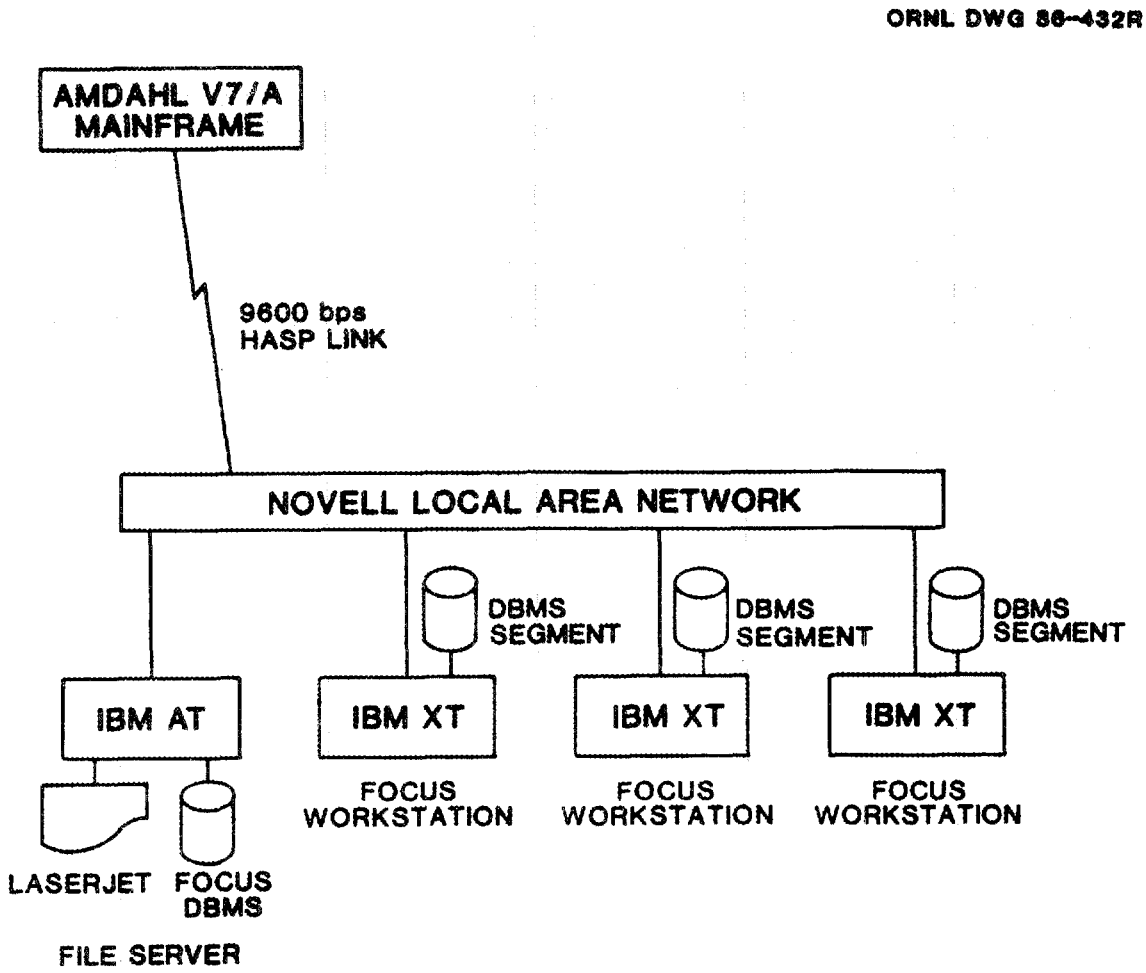

Fig. 2. Phase 6 configuration of the NCB-3 prototype.

4. CONCLUSION

True distributed data base management systems, able to automatically maintain and locate data anywhere on the network, are still commercially unavailable. Planners and implementors of distributed systems must therefore carefully consider and introduce methods for preserving data integrity and for ascertaining location of data at any given time. Nevertheless, end-users can be successfully introduced to advanced distributed processing systems that take advantage of the personal computers in their work environment. Modular planning and phased implementation are the keys to such success.

5. REFERENCES

1. Lynn Bateman, "Lack of Planning Hampers Micros' Effectiveness," Government Computer News, 4(18), 72 (September 27, 1985).

2. William H. Lowe, Jr., "Local Area Networks: the Exploration Has Just Begun," The Office, 100(5), 32 (November 1984).

3. Michael Durr, "Expanding Capabilities in High Performance Networking," pp. 151-154 in COMPCON Spring 86, Thirty-First IEEE Computer Society International Conference, March 3-6, 1986, Cathedral Hill Hotel, San Francisco, California, published by IEEE Computer Society Press, 1730 Massachusetts Avenue, N.W., Washington, D.C. 20036-1903, 1986, IEEE Computer Society Order Number 692, IEEE Catalog Number 86CH2285-5.

4. Jonathan Sachs, "Six Leading LANs," PC World, 3(2), 112 (February 1985).

5. John Dickinson, "LAN Speed Trials," PC Magazine, 4(3), 168-175 (February 5, 1985).

6. The Editors of PC Magazine, "Battle of the Network Stars," PC Magazine, 4(3), 180-186 (February 5, 1985).

7. Gregory Ennis, "PC Network Services for Distributed System Design," pp. 155-160 in COMPCON Spring 86, Thirty-First IEEE Computer Society International Conference, March 3-6, 1986, Cathedral Hill Hotel, San Francisco, California, published by IEEE Computer Society Press, 1730 Massachusetts Avenue, N.W., Washington, D.C. 20036-1903, 1986, IEEE Computer Society Order Number 692, IEEE Catalog Number 86CH2285-5.

8. "Need for Microcomputer Training Shown in Study of Major Corporations," Journal of Accountancy, 159(4), 31 (April 1985).

9. Dorine Andrews and Naomi Leventhal, "Prototypes Best for Flight-Testing Ideas," Government Computer News, 4(23), 57 (December 6, 1985).

10. Mike Hurwicz, "MS-DOS 3.1 Makes it Easy to Use IBM PCs on a Network," Data Communications, 14(12), 230 (November 1985).

11. Liba Svobodova, "File Servers for Network-Based Distributed Systems," Computing Surveys, 16(4), 354 (December 1984).

INTERNAL DISTRIBUTION

1. R. W. Barnes
2. M. E. Boling
3. J. Christian
4. N. E. Clapp
5. F. Clark
6. L. A. Clinard
7. F. Coleman
8. J. Craven
9. G. A. Dailey
10. J. M. Fielden
11. W. Fulkerson
12. N. B. Gove
13. M. H. Graves
14. R. K. Gryder
15. D. R. Hanlon
16. J. Hardee
17. T. Hayes
18. A. Huntley
19. A. S. Loebl
20. J. Longmire
21. F. C. Maienschein
22. K. C. Miller
23. H. Pfuderer
24. B. T. Rhyne
25. J. A. Russell
26. K. E. Shaffer
27. J. D. Shelton
28-40. C. E. Snyder
41. T. D. Wallace
42. W. B. Wood
43. R. E. Ziegler
44. Central Research Library
45. Document Reference Section
46-48. Laboratory Records, ORNL-RC
49. Laboratory Records—Record Copy
50. ORNL Patent Section

EXTERNAL DISTRIBUTION

51. A. R. Allen, Financial Systems Specialist, General Electric Company, Fairfield, CT 06341

52. R. Baeslach, Ernst L. Whinney, Room 606, 1821 Jefferson Davis Highway, Arlington, VA 22202

53. C. Baker, Crystal Mall Two, Room 606, Jefferson Davis Highway, Arlington, VA 22202

54. R. S. Barnfknecht, Pacific Bell, Room 2S100T, 2600 Cominio Ramon, San Ramon, CA 94583

55. K. Bell, CIGNA Corporation, 1600 Arch Street, Philadelphia, PA 19103

56-62. L. J. Bettis, Davis Group Inc., Suite 405, 1701 K Street, NW, Washington, DC 20036

63. E. Buckler, Navy Regional Data Automation Center, Code 42, Bldg. 143, Washington Navy Yard, Washington, DC 20374

64. M. Busch, FMP Management Information Systems, Crystal Mall Two, Room 600, Washington, DC 20362

65. J. G. Carbonell, Associate Professor of Computer Science, Carnegie-Mellon University, Pittsburgh, PA 15213

66. P. T. Drisko, Information Services, Bechtel Petroleum, Inc., Fifty Beale Street, P.O. Box 3965, San Francisco, CA 94119

67. A. J. Crosby, C.P. Systems, 105 Cecil Street, #22-03/04 The Octagon, Singapore, China 0106

68. S. M. Gillis, Dean of Graduate School, Duke University, Durham, NC 27706

69. K. Hall, Navy Regional Data Automation Center, Code 44, Bldg. 143, Washington Navy Yard, Washington, DC 20374

70. F. R. Kalhammer, Vice President and Head of the Energy Management and Utilization Division, Electric Power Research Institute, P.O. Box 10412, Palo Alto, CA 94303

71. R. E. Kasperson, Professor of Government and Geography, Graduate School of Geography, Clark University, Worcester, MA 01610

72. V. King, Logical Data Solutions, 6825 Capri Place, Bethesda, MA 20817

73. M. Lessen, Consulting Engineer, 12 Country Club Drive, Rochester, NY 14618

74. W. Levander, Navy Regional Data Automation Center, Code 4422, Bldg. 143, Washington Vay Yard, Washington, DC 20374

75. T. P. Lo, Rockwell International Corporation, P.O. Box 10462, Dallas, TX 75207

76. J. O'Malley, Navy Regional Data Automation Center, Code 44, Bldg. 143, Washington Navy Yard, Washington, DC 20374

77. T. Padgett, Information Builders, Inc., 4151 Ashford-Dunwoody Road, Suite 625, Atlanta, GA 30319

78. P. Quigley, Manitoba Data Services, 215 Garry Street, Tenth Floor, Winnipeg, Manitoba R3C 3P3

79. L. L. Radcliffe, Research & Waste Management Division, U.S. Department of Energy, Oak Ridge Operations, P.O. Box E, Oak Ridge, TN 37831

80. W. K. Reising, Arthur Adnerson and Company, 69 West Washington Street, Chicago, IL 60602

81. M. A. Simms, Prophet Systems Group, Eight Copples Lane, Wallingford, PA 19086

82. R. Vertner, City of Indianapolis, 1948 City County Building, Indianapolis, IN 46204

83. C. O. Wilson, Owens-Corning Fiberglass, Route 16, Technical Center, Granville, OH 43202

84. D. Wszolek, Information Builders, Inc., 1250 Broadway, New York, NY 10001

85. A. Wurglitz, 12005 Sun Rise Valley Dr., Reston, VA 22091

86. Office of Assistant Manager for Energy Research and Development, Department of Energy, Oak Ridge Operations Office, P.O. Box E, Oak Ridge, TN 37831

87-116. Technical Information Center, U.S. Department of Energy, Oak Ridge Operations, P.O. Box 62, Oak Ridge, TN 37831