THESIS REPORT

Master's Degree

Computer Network Analysis through Design of Experimentation

by H. Kim
Advisor: G.M. Zhang

M.S. 93-22

INSTITUTE FOR SYSTEMS RESEARCH

Sponsored by
the National Science Foundation
Engineering Research Center Program,
the University of Maryland,
Harvard University,
and Industry

ABSTRACT

Title of Thesis: Computer Network Analysis through Design of Experimentation

Name of degree candidate: Hanju Kim

Degree and Year: Master of Science, 1993

Thesis directed by: Dr. Guangming Zhang
Assistant Professor
Institute for Systems Research &
Department of Mechanical Engineering

Modern society is rapidly evolving into an information society. Today's technologies that allow the transformation, communication, and storage of information have produced fundamental changes in the manner in which businesses are conducted.

This thesis research consists of the analysis of the Local Area Network (LAN) with an emphasis on performance. The LAN in the College of Behavioral and Social Sciences at the University of Maryland at College Park is investigated for the purpose of this research. A performance analysis is carried out for the LAN environment in the Lefrak Hall. Workstation speed, server speed, server memory, and the number of workstations on LAN are taken as four important system parameters and a 2^4 full factorial design is performed on the LAN in the Lefrak Hall. The responding time during a 19 MB file transmission is defined as the LAN operation efficiency. An empirical model is established to describe the effects of those four system parameters on the communication between the server and the workstation.

Significant findings, which lead to the recommendations for improving the performance of the LAN at the present time and within the current working environment, are:

1) Server memory is not a key factor to improve the communication efficiency.
2) The other three systems parameters have significant effects with the workstation speed being most important.
3) The two-factor interaction between workstation speed and server speed indicates that the gain from increasing either workstation or server speed will be double-counted in improving the communication efficiency.

With such findings, it is recommended that each department within the College of Behavioral and Social Sciences should upgrade existing workstations to higher speed. Also, instead of adding extra memory to the file server, the speed of the file server (processing power) should be increased. It is also recommended that new network technology such as switching hubs be used to increase the bandwidth of the network for further improving network efficiency.

Computer Network Analysis through Design of Experimentation

by

Hanju Kim

Thesis submitted to the Faculty of the Graduate School of
The University of Maryland in partial fulfillment
of the requirements for the degree of
Master of Science
1993

Advisory Committee:

Assistant Professor Guangming Zhang, Chairman/Advisor
Assistant Professor Linda Milor
Associate Professor Adrianos Papamarcou

Acknowledgment

I gratefully acknowledge the guidance, encouragement, and support of Dr. Guangming Zhang, in completing this thesis.

I would like to thank Dr. Linda Milor and Dr. Adrianos Papamarcou for being members on my thesis advisory committee. I also would like to thank everyone in the Office of Academic Computing Services at the University of Maryland, especially Matthew Farrelly, Marie Gates, and Matt Piermarini, for their support. A special thanks goes to Song Zhao for her moral support as well as for providing the resources necessary to complete this thesis.

Last, but not least, I would like to thank my parents, my sister, my brother-in-law, and my cousin for standing behind me every step of my academic career.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction to Computer Network

With a rapid advancement toward computerization, more and more organization opt to store information on computers. Nowadays, universities, government agencies, and private firms heavily rely on computer networks for information exchange and management. As the information stored in computer is accumulated, the way of accessing the information becomes increasingly vital. There is always the need for developing new and innovative computer networks to access and share information.

1.1 Review of Computer Networks

Recall the first appearance of telephone networks in the 1950's. By using a modem, digital signals are transported over an analog medium (telephone line). These networks transported data over a distance expeditiously and economically. However, the telephone network has disadvantages such as low fixed bandwidth of 4kHz and a level of performance designed primarily to meet the needs of voice communication. As a collection of nodes interconnected by transmission facilities, the data transmission network provides point-to-point communication only. Such networks are also referred to as private line networks or as networks providing private line services. Figure 1.1 illustrates the physical facility typical of voice and data transmission. As illustrated, the transmission channel could not carry voice and data interchangeably - a fact known as functional separation. Although it was possible to regenerate data signals as frequently as desired, selective regeneration of data signal was impossible in functionally separate voice (analog) and data (digital) transmission environments. Therefore, an integrated transmission environment requires interchangeability to transport voice in the analog format and data in the digital format. In fact, the ability to regenerate data signals was the key element in improving transmission performance in the 1950's. In order to meet communication needs by a single transmission network, it would require physical transmission channels between each and every end point. For n end points, the solution would require n(n-1)/2 transmission channels to provide a full interconnection between all end points. This solution is clearly undesirable, especially for a large number of end points.

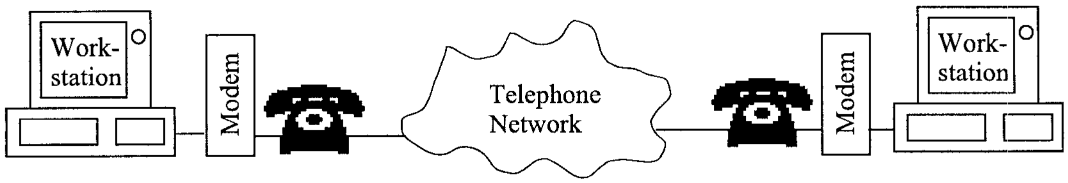


Figure 1.1 Initial Network Approach

During the 1960's, a private network approach (end-to-end) was developed. As shown in Figure 1.2, the network was composed of concentrators using a star topology. A star topology was where all the processing was concentrated in the central host processor and the front end was a transmission control unit responsible for only elementary Data Link Control (DLC) functions. One of the disadvantages under this scheme was a lack of line sharing or terminal sharing. A given line, and all terminals on it, were part of the access path to only one application program. Therefore, two different applications would require two terminals and two lines. In 1970s, an improvement was made. The front end became a programmable communication controller doing DLC and more. This design allowed terminals to share a line to separate applications located as the same host. However, this private network approach also resulted in a fragmented approach to solving the needs of data communications users. In addition, it was potentially expensive for small customer groups due to lack of sufficient sharing. Lack of interconnectability among private networks was another factor preventing from speeding up data communication.

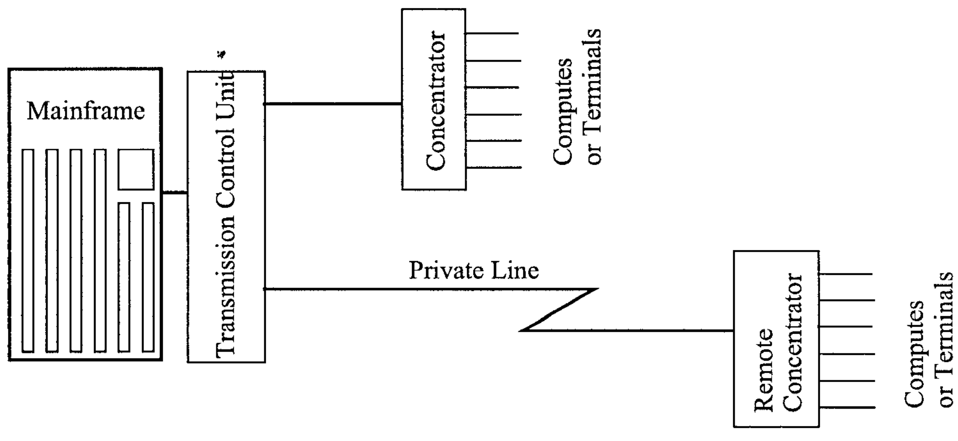


Figure 1.2 Intermediate Network Approach

The preceding considerations suggested the need for adopting a shared network approach, rather than a fragmented approach (previous approach) to data communications. In essence, a computer network is not merely a collection of terminals, communication lines, and computers, but an organized structure to which computers and terminals are attached and through which they communicate. Such a network should have the ability to connect all devices that meet a defined set of access and protocol requirements. This approach called the shared network approach is illustrated in Figure 1.3. In a shared network approach, communications between two end points demand much more than a simple transport connection.

When two people interact over a telephone, a voice-band (4 kHz) connection would be a necessary requirement. However, before they can meaningfully converse with each other, they must speak the same language. Even within the same language words have multiple meanings. A new challenge in information and data communication is both the accuracy and efficiency communications. So they must have some means of resolving ambiguities that may arise. As in human communication, such transport requirements are basic to all computer communication functions. An efficient and accurate transport network is thus a necessary part of computer communication.

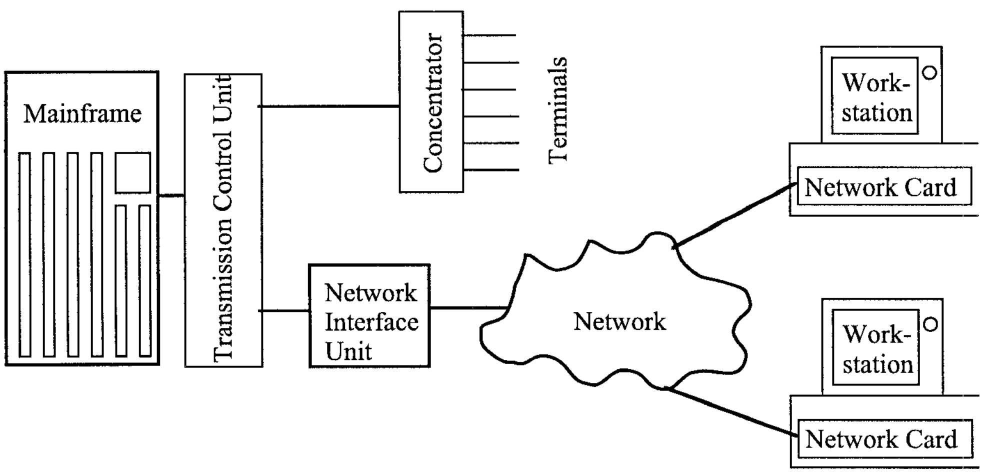


Figure 1.3 Shared Network Approach

1.2 Need for Local Area Network (LAN)

Modern society is rapidly evolving into an information society. Today's technology that allows the transformation, communication, and storage of information has produced fundamental changes in the manner in which businesses are conducted, government agencies are operated, educational institutions are functioned. Compared to what existed just a few decades ago, this change is dramatic and dynamic. With access to the most current information from different locations, geographically separated individuals may have the most current information through computer networks. Such accessibility results in considerable savings of time and an increase in productivity. At present, data is the most rapidly growing part of the realm of the information world due to the rapidly increasing need for information . Besides sharing information, most organizations want to share high-cost equipment such as laser printers, color printers, high-speed modems, back-up units, CD-ROMs, etc. The Local Area Network (LAN) is a computer communication network which allows workstations to share information, peripherals, etc. within a limited (local) area. All of a LAN's functionality is provided by a high capacity computer called "file server". The LAN also provides a connection to external networks such as Internet or other LANs to have even more resources available to LAN users. A LAN's capability to provide the above services makes a LAN a necessary requirement for many organizations.

1.3 Scope of LAN Analysis

In this thesis research, emphasis is given to Local Area Networks. To demonstrate the importance of performing a computer network analysis to improve the efficiency of the LAN, the local area network that exists in the College of Behavioral and Social Sciences in the University of Maryland at College Park will be investigated. The College of Behavioral and Social Sciences has several buildings, therefore its computer network consists of several LANs connected together. Analysis of LANs in this thesis work is based on the LAN environment that exists in Lefrak Hall. The analysis focuses on important issues that the LANs in Lefrak Hall face today. These issues include (1) users request faster computers to increase their productivity, (2) the need to make the decision of how to upgrade the file server, and (3) the need to improve the overall response time of the LANs. It is hoped that results from this thesis research will be recommended to the College of Behavioral and Social Sciences for improving their LAN performance and providing guideline for them to follow whenever they want to upgrade their LANs.

1.4 Thesis Organization

In addition to this chapter, there are five other chapters in this thesis. In chapter 2, the concept of a LAN is discussed with a focus on the definition of a LAN and its constituent elements as well as LAN applications. In chapter 3, the process of analyzing network system performance is discussed with a focus on the LANs in Lefrak Hall. Results from the performance analysis process provide possible solutions to resolve the current performance problems. In chapter 4, LAN performance criteria are established, throughput is identified, and data is collected. With the data collected, results are analyzed, and the recommendations are made for the College of Behavioral and Social Sciences to consider. In chapter 5, new network technology that is currently being developed is discussed. Chapter 6 summarizes the thesis research and gives suggestions for future work in this research area.

Chapter 2 Local Area Network

This chapter defines what a Local Area Network(LAN) is as well as terms that are associated with Local Area Networks. As described briefly in the previous chapter, a LAN is a computer communication network which allows workstations to share information, resources, and peripherals within a limited (local) area. The LAN also provides connections to external network such as Internet or other LANs so as to provide LAN users with more resources as illustrated in Figure 2.2.

2.1 Definition of Local Area Network

A local area network (LAN) is a system by which microcomputers can share information and resources within a limited (local) area. A LAN requires that individual workstations be physically connected by cabling (usually coaxial or twisted pair). In addition, some network software resides on a disk (hard or floppy disk) to permit the sharing of peripherals, data, and application programs. The primary motivation for most organizations to install a local area network is to take advantage of an existing investment in microcomputers, peripherals, and software. By installing a LAN, an organization can exploit existing equipment to share hardware and software resources. Each microcomputer attached to the network retains its ability to work as an independent personal computer and run existing software. However, it also becomes a network workstation capable of accessing information and software that is located on the network file server. When a user issues a command, network software intercepts the command and routes it to either the network file server or to its own workstation. Users at the workstation can process any data or run any application on the LAN as if the files are on their own workstation.

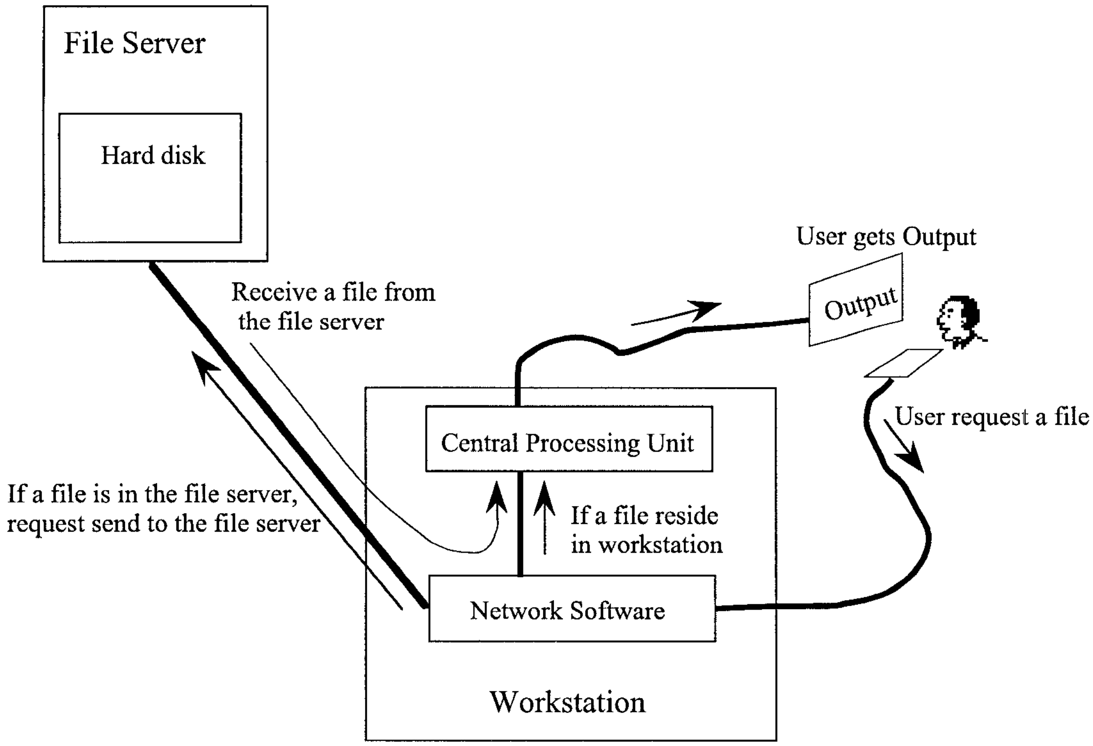


Figure 2.1 Routing of User's Request

2.2 Local Area Network Site

A Local Area Network Site is a relatively well-defined and restricted geographic area which is usually a single building . LANs intended for use on a university campus must be capable of spanning a maximum distance of several kilometers; LAN extensions within an office complex is unlikely to involve distances much greater than 1 kilometer. A local area network may provide facilities for wholly internal communications within a site, but may also permit access to external networks. A LAN that allows users to access an external network must have the capability to interface to that external network, via an appropriate device such as a "gateway" facility. The LAN must also provide a means by which the internal user can address the external network and the subscribers connected to the external network. For example, a LAN user at the Lefrak Hall can send electronic mail to a LAN user at the A.V. Williams building, since both of these LANs are interconnected through the devices such as a gateway.

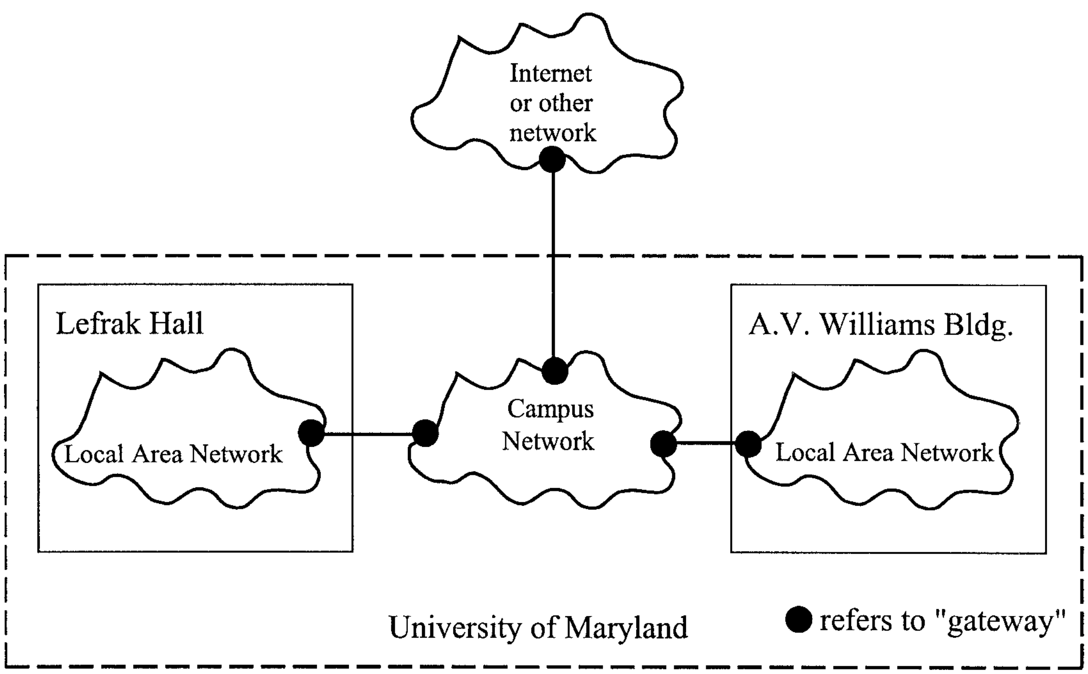


Figure 2.2 LAN-to-LAN and LAN-to-External Network

2.3 Applications for LANs

Traditionally, voice communication has dominated internal communications and resource sharing. However, the LAN environment provides many valuable services from which users can benefit such as information distribution and management, resource sharing, messages transactions, and provide connection to external networks for variety resources. For example, a faculty member could store a data file to be used in conjunction with an assignment to be done by all class members.

2.3.1 Information Distribution and Management

An example of one of the benefits of a LAN provided to a user is Information Distribution and Management. Information Distribution and Management involves the storage and management of information (in one or more locations) which is retrieved or distributed to multiple locations on the same LAN. The information can consist of common program files or users' data files. LANs provide a common storage area by which users within the same group can immediately access the up-to-date information relevant to, say, group projects. For example, an Economics department at a university could have access to data or economic indicators. By virtue of a LAN, these data can be stored in a location accessible to all the researchers in the department. LANs also provide a private storage area so that when users connect to LAN with their unique ID, they have a private storage area that no other users have access to, except for the LAN manager. This feature allows LAN users to use any workstation that is on the LAN and still have full access to their files. For example, if one of economics professors' computer is down, then he/she can use any networked computer to connect to the LAN using his/her own network ID to have all the access to his/her files on the network.

Beside getting access to their public and private information, users may use application software that is on the LAN. Considering the myriad of software applications on the market today, a LAN manager can easily provide the most popular software packages such as word processing and spread sheet programs. However, these software packages need to be upgraded frequently and upgrading software on each workstation is a very time consuming process. Therefore, a LAN can store these applications on the file server which allows the users to run the most recent applications directly from the LAN.

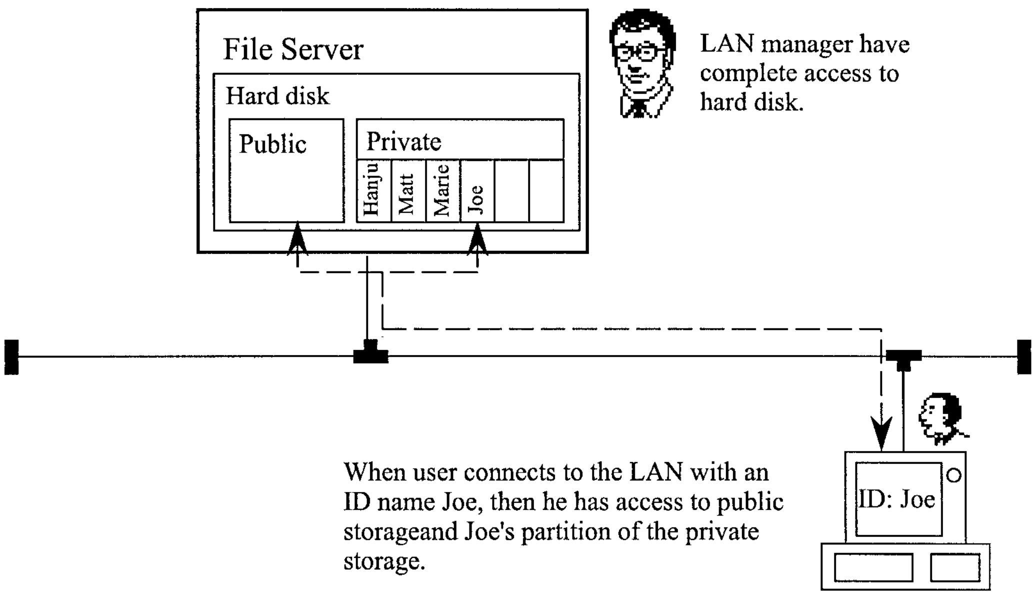


Figure 2.3 Accessing Information on the File Server

2.3.2 Resource Sharing

By virtue of having a LAN, a group of networked users can have access to expensive equipment such as high speed line-printers and color laser printers, an expense that would not be justified for individual users. Facility sharing often provides a primary financial justification for acquiring a LAN. The relevant equipment currently includes high quality laser printers, CD-ROMs, tape backup units, hard disks, etc. The users of such equipment may be located physically close to the equipment (in the case of a laser printer) or some distance away (in the case of a hard disk or tape backup unit). As previously mentioned, LAN users can take advantage of any networked device as shown in Figure 2.4.

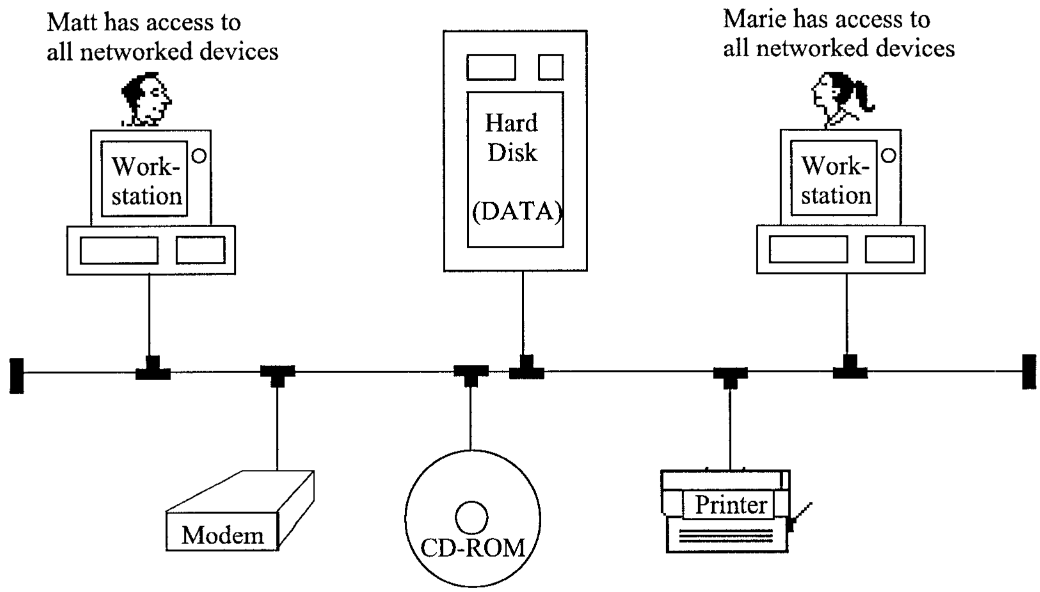


Figure 2.4 Resource Sharing

2.3.3 Messages and Transactions

Messages involve two-way communication between two terminals, which may require momentary storage. For example, an electronic mail (e-mail) message will be intermediately stored, then forwarded to the addressee. Most message-type applications involve one-way communication. Although a message may stimulate a response, there is a permissible time delay of seconds, minutes, or even hours before the response is expected or required. In some cases, a local network will allow multiple internal locations to access an external network. In these cases, regardless of the network protocol type involved, the internal network must be capable of interfacing in a transparent way to the external network, see Figure 2.2. As electronic mail gains popularity within computer users, access to external networks has become a strong justification for acquiring a local area network for many organizations.

2.4 LAN Elements

Complex systems such as today's LAN consist of many elements that contribute to their performance. Therefore, it is important to understand some of the main features of a LAN. There are three main components that make up a LAN: the Network Topology (Cabling), the File Server, and the Workstation. In this chapter, these three components will be introduced and their contribution to LAN performance will be defined. To make this discussion more concrete, the Novell NetWare network operating system environment will be used to discuss issues of LAN performance.

2.4.1 Network Media

Every LAN must have cabling to link individual workstations together with file servers and other network peripherals. Ethernet is the most popular network topology and there are several different types of cabling available for Ethernet, with considerable range in cost and in capacity.

Twisted pair cable is a point-to-point cabling scheme, and its topology represents a star (See Figure 2.5). A star topology permits centralized diagnostics of all network functions. All messages come through the central repeater hub, so it is easy to analyze all station messages and reveal any error messages and so forth. Twisted pair cable consists of four strands of 22-26 AWG wire (standard telephone wire) and may be run for 100m between the Ethernet controller in computer and repeater hub. With this cabling scheme, the transceiver is physically located on the Ethernet card. The twisted pair cabling scheme is gaining popularity, since it is relatively inexpensive and more manageable/reliable than coaxial cable. For example, if the cable that links a workstation with its repeater hub is damaged in any way, then only that workstation will be affected.

Whereas in a similar circumstance, a workstation connected with coaxial cable will cause the entire LAN to go down.

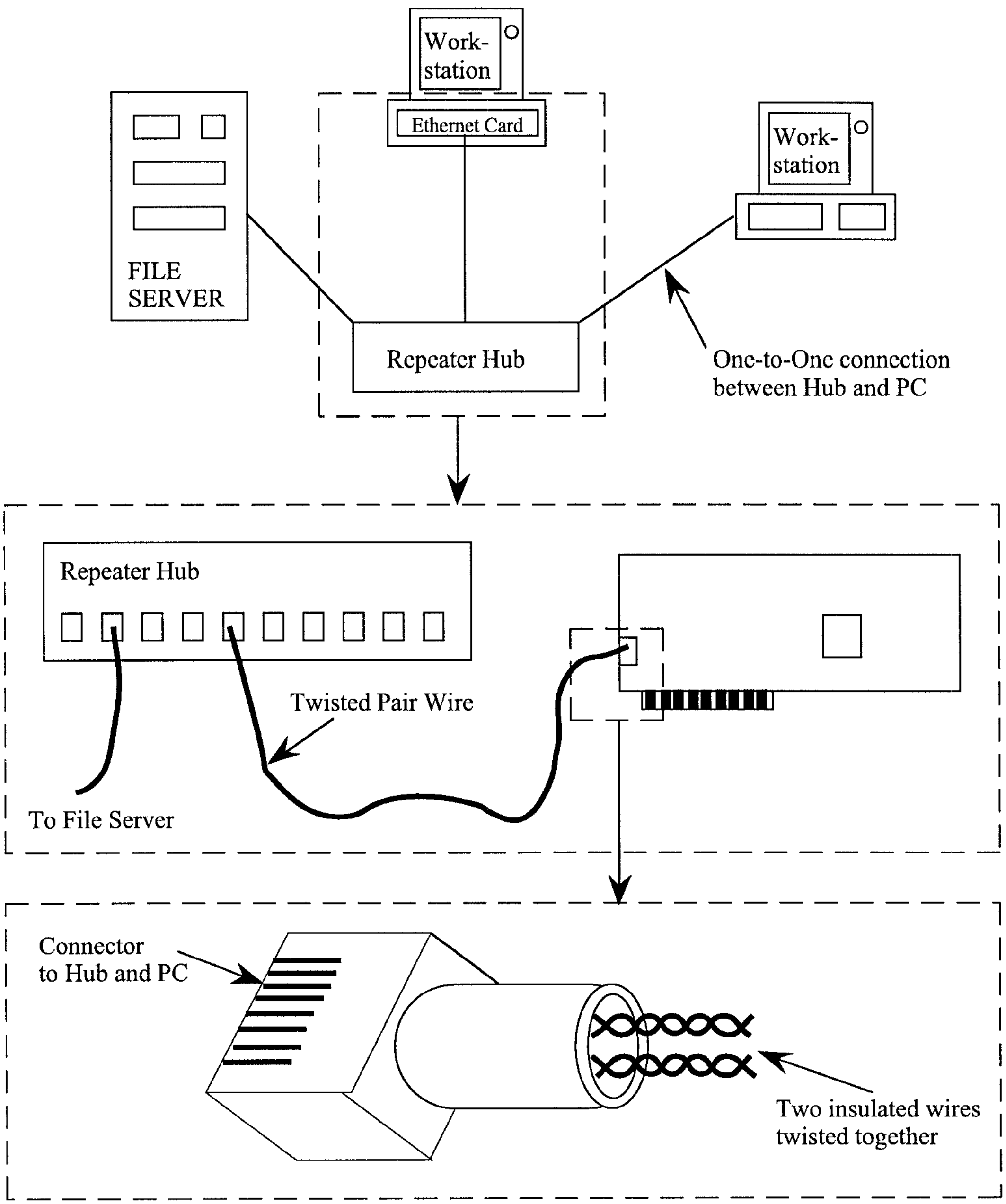


Figure 2.5 Twisted Pair Cabling

There are two wiring schemes for coaxial cabling: thick coaxial cable and thin coaxial cable. The main difference between thick and thin coaxial cable is that the thin coaxial network interface card does not require an external transceiver, but the thick coaxial network interface card does. Since most LANs only utilize thin coaxial cable, only thin coaxial cable is discussed here. In thin coaxial cabling, the transceiver is physically located on the Ethernet controller card in workstations. Workstations are attached to this cable through the use of BNC connectors and T connectors. Each cable segment endpoint contains a male BNC connector. These connectors attach to a T connector. Thin coaxial cabling scheme represents a bus topology (Figure 2.6). Thin coaxial cable, 10Base2, is composed of a copper conductor surrounded by insulation. Also, the inner insulation is wrapped by aluminum or copper and is protected by outer insulation. Coaxial cable has been the medium of choice of many of the major LANs, due to its simplicity. However, recently it is losing its popularity to twisted pair cabling because it is difficult to manage and maintain network security. The coaxial cabling scheme is a long extension of coaxial cable with workstations, file server, and other network peripherals connected along. When there is a break in the cable extension, every connected workstation and device will lose its connection from the file server. An additional drawback is that it is hard to diagnose where the breakage is, especially when the cable extension is long.

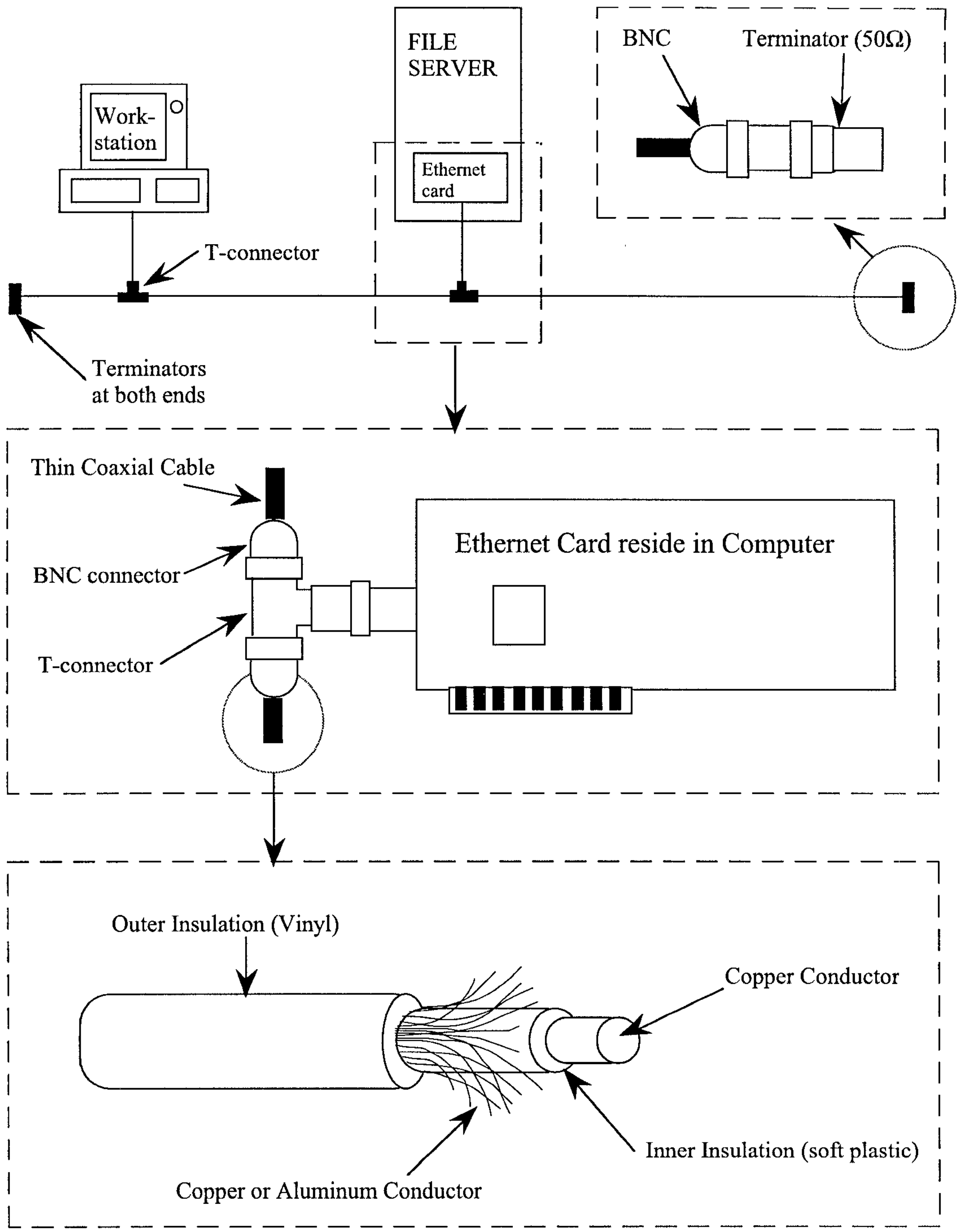


Figure 2.6 Coaxial Cabling

Yet another cabling option is fiber optic cable. Fiber optics in a LAN provide number of advantages over coaxial and twist pair cabling. Its data transmission rates far exceed coax and twisted pair, it is immune to electromagnetic or radio-frequency interference, and is capable of sending signals several miles without loss of signal. Due to its characteristics, it is being widely used in connecting LANs (between buildings). The cable is made of pure glass drawn into very thin fibers to form a core and the fibers are surrounded by cladding, a layer of glass with a lower refractive index than the glass in the core. A fiber-optic network uses a laser or light emitting diode (LED) to send a signal through the core portion of the cable. At the receiving end of the cable, the message is translated back into a digital or analog signal by a photo diode. While fiber optic cabling is a high performance cabling scheme, it is still costly compared to twisted pair or coaxial cabling (average Ethernet card for coaxial and twisted pair costs approximately $150 while fiber optic Ethernet card cost over $500). Fiber optic cabling is usually used to connect hubs or LANs in different buildings, and it is not widely used to connect workstations due to high cost and complex installation. Fiber optic cabling topology is similar to twist pair cabling: there is a one-to-one connection between hub and workstations (it maintains the advantage over coaxial cabling scheme in the sense that in the case of a cable breakage, only the affected workstation goes down). Typically fiber optic cabling requires 2 cables for each networked computer; one for transmitting and other for receiving, see Figure 2.7.

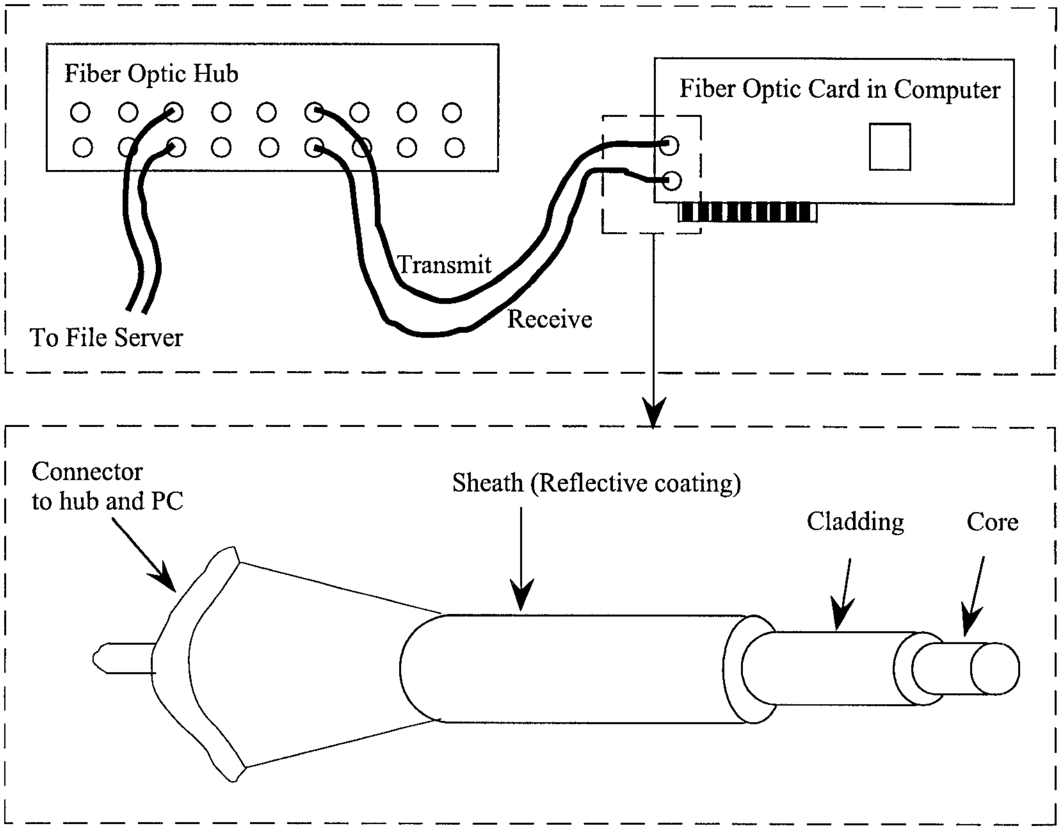


Figure 2.7 Fiber Optic Cabling

2.4.2 File Server

A file server is a microcomputer with a large-capacity hard disk that serves up files upon request from workstations that are connected to the network. It should be also noted that sending files to the workstation is not the only function of the file server. The file server also provide other services such the following:

1. Printing Services

User at the workstation can print to printers that are connected to the LAN. Printers are connected to the LAN in the same way as workstations, using one of the cabling schemes discussed previously. Each printer has a unique name so that users will be able to distinguish which printer they are printing their files to. When users print their file to one of networked printers, file server intercepts the file, check the printer name, and route a print job to that printer.

2. Device Sharing

Device sharing works the same way as printing. Network devices include big hard disk, CD-ROM drive, Modem, etc. Each of these devices also have a unique name and users will be able to access them as if they are connect directly to their workstations. When users request to use such devices, file server routes the request to that device, establishing a connection between the workstation and the device.

3. Electronic Messaging.

Electronic messaging, also known as an electronic mail, is another service that a file server provide to LAN users. Each user have an unique identification name (user ID) that they use to connect to the file server. This user ID is used to send electronic mail to other users on the LAN. When one user send an electronic mail to another user on the LAN, the file server receives an electronic mail and check its destination (receiving user ID) and route them to the user with that ID.

As more users depend on the LAN, it is very important that the file server be as reliable as possible, or a fault tolerant system. A fault-tolerant file server can be best described as a file server that can be added to a basic system to decrease the amount of downtime that can be expected. Fault-tolerant add-ins can increase server reliability significantly, by eliminating the cause of failure or providing a method of dealing with system failures without bringing down the server. The following are some of the possible approaches to making a fault-tolerant file server.

1. The use of an uninterrupted power supply (UPS).

A UPS will prevent surges, blow-outs, or complete losses of power from causing damage to the server hardware. LAN analysts need to make sure that the UPS provides sufficient KvH to provide enough power for the server and any external devices like hard disks and monitors.

2. The use of a backup system

Although backup is an ordinary procedure for maintaining LAN servers, the LAN manager needs to make sure that the backup device has sufficient capacity to handle the server in a single pass. Also, the LAN manager should have backup policies and make sure that the data is protected adequately.

3. Use hard disk mirroring

The LAN manager should consider hard disk mirroring, which allows the file server to duplicate network files to another hard disk (requires more than one hard disk) in case of hard disk failure. In case of hard disk failure, the other disk has an exact duplicate of the files so that users will not lose any data.

The Central Processing Unit (CPU) is a key component in the server and other workstations. The Novell network operating system, due to the nature of the program, requires INTEL processors better than a 80286 processor. The most popular operating system from Novell is NetWare 3.11, which requires a CPU equal or better than the INTEL386 (386, 486, and Pentium). Each processor can process information at different rates, known as the clock speed: the clock speed varies from 25Mhz to 66Mhz for 386, 486, and Pentium processors. Even if 386 and 486 process information at the same rate (clock speed), the 486 processor will process information more quickly due to its more efficient design - more internal processing units called transistors are built into 486 than

386.

The BUS is a data path between the CPU and other devices such as the network card and hard disk. There are several bus architectures that are available for the INTEL-based processor (386, 486, and Pentium): ISA, Microchannel, EISA, VESA Local Bus and PCI Local Bus. ISA stands for Industry Standard Architecture and is an old bus technology that is still the most popular in personal computers. Microchannel is a IBM standard that is used in IBM's higher-end personal computers such as PS/2 lines. EISA stands for Extended Industry Standard Architecture and co-developed by several computer companies (excluding IBM) to compete with IBM's Microchannel bus architecture. Peripheral Components Interconnect (PCI) and Video Enhance Standard Architecture (VESA) local bus technology are the newest additions to the bus architecture and they are currently used primarily for graphics applications at this time. Although the ISA architecture can be used in a file server, EISA and Microchannel bus architectures are more popular when it comes to file servers, since EISA and Microchannel provide faster data transfer between the bus and the CPU. The LAN designer should decide on the bus architecture (preferably, EISA or Microchannel) before selecting a file server for a LAN.

There are many different network interface cards for a Novell file server. However, this choice depends on the file server's bus architecture. For Microchannel based file servers, a network interface card must be Microchannel compatible, however, EISA based file servers can accommodate both ISA and EISA network interface cards. Yet, an ISA network interface card in an EISA based file server will not take advantage of the faster data transfer of the EISA bus architecture. There are some network cards, such as 3Com's EtherLink III which utilize parallel tasking(sends and receive data at the same time), allowing faster file transfer. However, both the workstation and the file server should have a parallel tasking capability card to take advantage of such a feature.

Memory size is also important in the Novell NetWare LAN environment because in addition to keeping the LAN's operating system in memory, it also maintains the File Allocation Table (FAT). FAT is a table where the server keeps track of files. The total memory requirements for Novell NetWare 3.11 should be calculated as follows[1].

1. Calculate the memory requirement for each volume. A volume consists of a partition of a physical hard disk. The more disk space used in the file server, the more memory is required since the FAT is stored in memory. Also, support for several operating systems (Macintosh, OS/2 and UNIX NFS) on hard disk requires additional memory. It should also be noted that the file server can have several hard disks and these hard disks can be named as a single volume (SYS) or several volumes (SYS, SYS1, DATA, and etc.).

   For each volume:

   Memory = X * Volume Size (in MB) ÷ Block Size (default is 4)

   X = 0.023 if volume has only DOS support

   X = 0.032 if volume has support for additional operating systems

   X values are predetermined by the Novell and using a X value lower than specified above may cause file server to malfunction (for example, users may lose file or file server may go down).

2. Add the memory requirements for all volumes. A file server can have several volumes and they should be added to compute the total memory requirement.

   $\text{Total Volume Memory} = \text{Memory}_{\text{Volume SYS}} + \text{Memory}_{\text{Volume SYS1}} + \ldots$

1NOVELL NetWare Version 3.11 Installation Guide, Novell, Inc.

3. Add 2MB for the LAN operating system and round the figures to the next higher MB. This gives the total memory requirements for the file server. However, if the figure is less than 4MB, then the total memory requirements should be 4MB, since 4MB is the minimum requirement for the Novell NetWare 3.11 operating system.

For example, if a file server has two 500MB hard disks with 2 separate volumes: one with both DOS and Macintosh operating systems (SYS Volume) and the other with only a DOS operating system (DATA Volume):

$Memory_{SYS} = 0.032 \times 500 \div 4$,then $Memory_{SYS}$ is 4MB

$Memory_{DATA} = 0.023 \times 500 \div 4$, then $Memory_{DATA}$ is 2.87MB

Total Volume Memory = $Memory_{SYS} + Memory_{DATA}$ = 4MB + 2.87MB

Adding 2MB for the operating system: 4MB + 2.86MB + 2MB = 8.87MB

Rounding to the next higher MB: 9MB

Since the calculated total memory is more than 4MB, in this example, the file server has a 9MB total memory requirement.

2.4.3 Workstations

There are several different types of workstations that can be connected to the Novell file server: workstations differ by the types of operating systems they are running. The most popular operating systems are DOS, Macintosh, OS/2, and UNIX; however, the vast majority of LANs are designed for IBM PCs or Apple Macintosh computers. In this section, a few workstation elements that contribute to network performance will be discussed.

The Central Processing Unit (CPU) is another key component of workstations and file servers. The most popular workstations in the Novell LAN are IBM DOS and Apple Macintosh computers. IBM DOS computers have CPUs made by INTEL and Macintosh computers use Motorola's CPUs. As discussed in the previous section on the file server, each processor can process information at different rates and some have a more efficient design than others.

There are several bus architectures that are available for INTEL based processors (386, 486, and Pentium): ISA, Microchannel(MC), EISA, VESA Local Bus and PCI Local Bus - see section 4.2.2. There are two main bus architectures for Apple Macintosh computers: Nu-bus and processor direct. Nu-bus is similar to the bus architecture found in INTEL based computers, while processor direct is a direct connection to the processor which allows data to be fed into the CPU directly. Processor direct architecture may communicate faster with the CPU. The difference in bus architecture is dependent upon which CPU is being used for that workstation - if a faster CPU is used, then a faster bus architecture is used so that the CPU does not have to wait long to get data from a bus.

There are many different network interface cards for workstations that vary by the type of bus architecture that is being used for that workstation. INTEL based computers use same network interface cards as the Novell file server. Recently, network interface units for portable computers (laptops or computers without expansion slots) were developed: a network interface unit is an external device that connects to the parallel port of a computer. Network interface cards for Apple Macintosh computers also vary by the type of bus architecture. Moreover, due to the high utilization of SCSI in Macintosh computers, there are many SCSI network interface units connected to a LAN.

Memory size in a workstation does not contribute to network performance directly, however, more memory is required in order to connect workstations to a LAN. This may cause problems for some DOS applications, since DOS has a memory limitation of 640KB, thus, the use of some type of memory managers (such as QEMM or HIMEM.SYS) is essential. For instance, normally DOS based computers require system files in order to attach to a LAN. Typical system files are shown in the table below.

File Name	File Size
LSL.COM	15.2KB
3C503.COM	3.9KB
IPXODI.COM	19.2KB
NETX.COM	47.5KB

Table 2.1 Network Software File Size

From Table 2.1, the DOS computer requires at least 85.8KB out of 640KB for network software. That means that a program that requires more than 554.2KB (640KB - 85.8KB) to run will not be executed. It then makes sense that program managers that allow DOS to use memory more than 640KB is essential in such cases. For Apple Macintosh computers, more memory must be added to computer without any modifications to operating system.

2.4.4 Workstation Load in LAN

It has been previously shown in other studies that the number of workstations on a LAN determines the performance of the LAN[2]. From Figure 2.8, the file server's throughput (megabits/second) increases as the number of workstations increase; however we can see that the throughput declines as the number of workstations exceeds thirty.

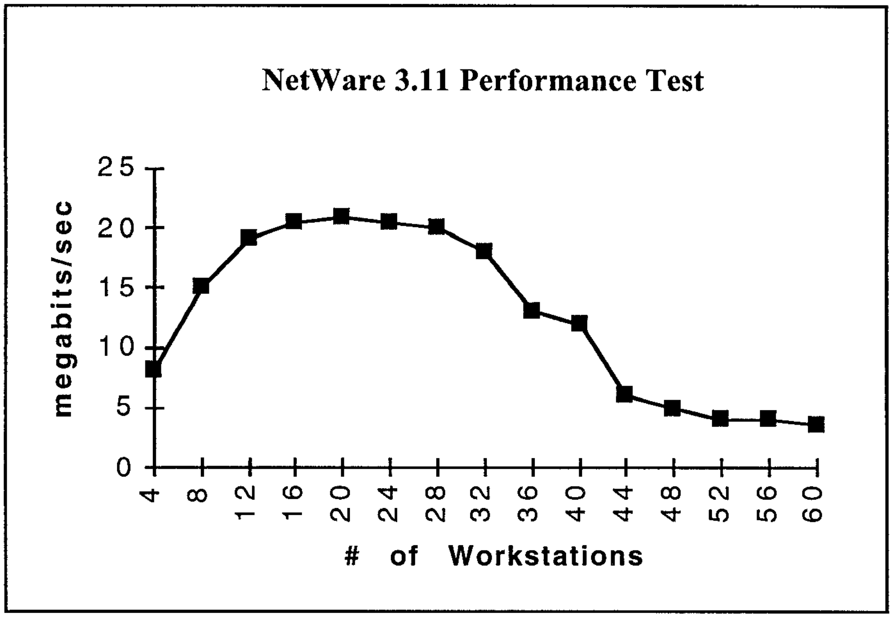


Figure 2.8 File Server Throughput vs. Number of Workstations

---

[2]ZD Labs, "NetWare 3.12 performs well", PCWeek, October, 1993, pp. 115

Chapter 3 Performance Analysis Process

During the life cycle of the LAN, LAN managers may require to solve some performance problems that may arise. LAN managers should consider performance analysis of a LAN to come up with possible solutions to problems. Performance analysis process is used to identify current LAN's performance parameters such as server speed and server memory size and come up with possible solutions to current LAN's performance problems such as slow response time. Regardless of the problem, the LAN manager must follow a certain procedure to ensure that all the necessary factors are considered. A LAN manager must define the performance problem, then the primary approach to solving the problem must be derived. Once the approach to the problem is derived, each criteria and throughput must be identified to come up with a mathematical model for the problem. Once this is accomplished, the LAN manager can collect data, apply the model, analyze the results, then make decisions of how best to solve the problem. LAN manager can follow the performance analysis process which consists of the eight major steps as shown in Figure 3.1.

In this chapter, performance analysis process is discussed and the LAN at the college of Behavioral and Social Sciences is analyzed using this process. Throughout the definition process, the problem is transformed from a physical status to preliminary mathematical function(s) whose inputs, outputs, constraints, and objectives are identified. The performance problems are defined and modeled in many different functions. The purpose of each function is to quantify a corresponding performance measurement. The performance problems are not just limited to addressing technical issues such as the technical capability of the LAN, but they also address management issues related to resources, user satisfaction, business policy, schedule and cost control. In many cases, the nature of the problem may appear to be obvious, but in all actuality the precise definition of the problem may be the most difficult part of the entire process. Unless the performance problem is clearly defined, it is doubtful whether an analysis of any type will be meaningful.

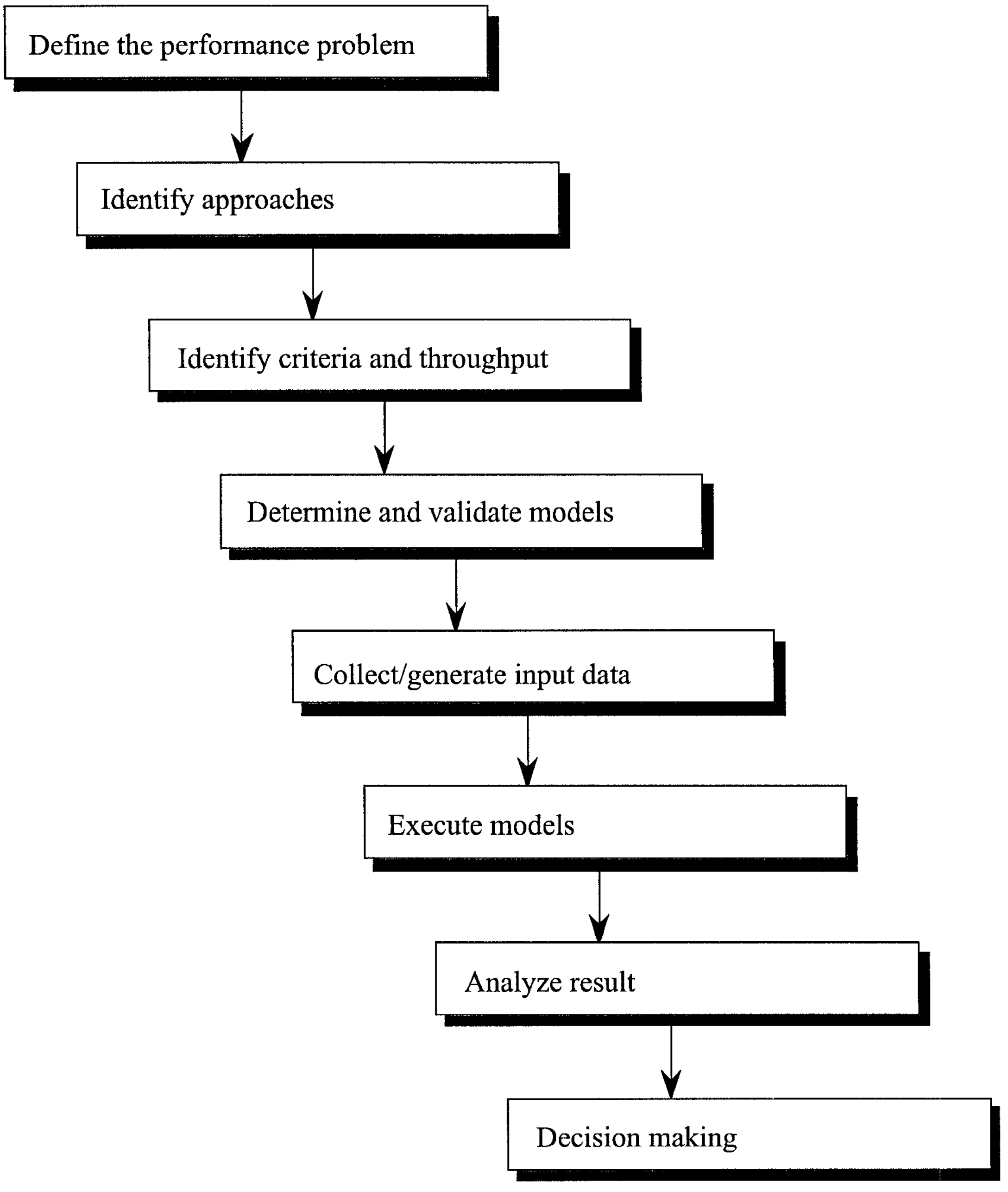


Figure 3.1 Performance Analysis Process

3.1 Define the Performance Problem

The first step the LAN manager must apply is to examine the LAN requirements in order to define the performance problem so that it can be studied in an efficient, controlled, and timely manner. There are several parameters which shall be considered to define the performance problem.

1. Performance and physical parameters: definition of the operating characteristics or functions of the LAN.

   The LAN at the college of Behavioral and Social Sciences in Lefrak Hall provides network services to faculty, staff, and students. While faculty members and staff have their own workstations, students rely on the public computer labs operated by the Office of Academic Computer Services (OACS - Computer Support Unit for the college of Behavioral and Social Sciences). The file server is called "lefrak" and has following capacity.

   IBM PS/2 Model 95XP: 486DX
   33Mhz clock speed
   3 Gigabytes of disk space with Name Space Support
   26Mb RAM

   The PS/2 Model has 26Mb RAM, which is the minimum memory requirement (see chapter 2.4.2) to run Novell NetWare 3.11 with 3 gigabytes hard disk while supporting the Name Space. Although, the file server uses 486 processor, the server runs at only 33Mhz, since 33Mhz was the fastest server at the time of purchase 3 years ago.

2. Use requirement: anticipated use of the LAN and its elements (e.g., hours of operation per day, on-off sequences, operational cycles per months).

   The file server (denoted as the Lefrak file server) must be on-line 24 hours a day and 365 days a year. Since the file server allows users to dial into the file server from home, the file server is being used at all times. Faculty members and graduates students use the file server as a common storage area to share information for their research projects. Word processing and electronic mail are applications programs that are used most on the file server[3]. Due to the high usage of electronic mail, users at the college of Behavioral and Social Sciences are heavily rely on the file server to communicate with each other.

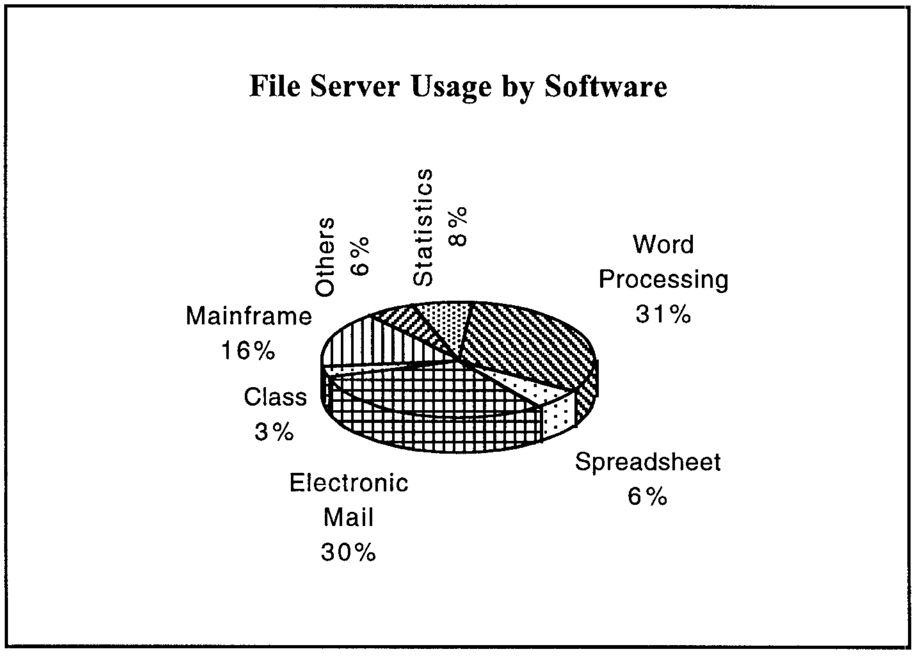


Figure 3.2 File Server Usage by Software

---

[3]File Server Usage Database record is provided by the Office of Academic Computing Services. Database is collected during the 92 fiscal year.

3. <u>Effectiveness factors</u>: LAN requirements specified as figures of merit

The availability in terms of Mean Time Between Failure (MTBF) and Mean Time To Repair (MTTR):

$$\text{Availability} = \frac{\text{MTBF}}{\text{MTBF} + \text{MTTR}}$$

The availability of the LAN at Lefrak Hall is calculated as follows[4]:

$$\text{Availability} = \frac{\text{MTBF}}{\text{MTBF} + \text{MTTR}} = \frac{60(\text{days})}{60(\text{days}) + 0.833(\text{day})} = 98.6\%$$

There was a total of 6 server failures during the 1992 fiscal year and their mean time to repair was 2 hours (which translates to 0.833 of day) for the LAN at the College of Behavioral and Social Sciences, which translates to 98.6 % availability. This indicates that the file server is very reliable.

Potential LAN purchasers seek higher business efficiency and lower operating costs with the constraints of keeping down capital expenditure and avoiding disturbances in the network site. Here are some of performance problems that the LAN at the College of Behavioral and Social Sciences are facing.

• Users have increased over the years, however, the file server have not been upgraded for 3 years. The average number of users on the file server during office hours has reached

---

[4]MTBF and MTTR data is provided by the OACS. The data is calculated during the 92 fiscal year.

the maximum number of users supported by the network operating system[5] (250 concurrent users at any given time). It is necessary to look into how we can maximize the server performance with the limited cost allocated for network equipment, see Figure 3.4.

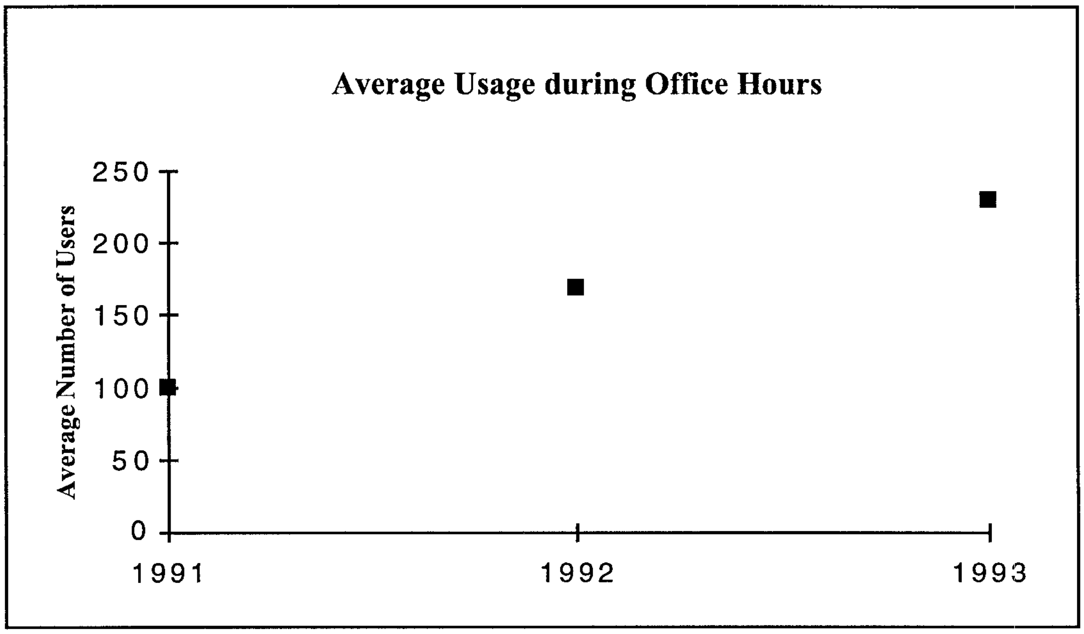


Figure 3.3 Average Number of Users on the File Server

• Users from various department requesting solutions to problems such as the sluggishness of the LAN, in other words, it takes a long time to access application programs over the LAN. Each department should upgrade their own hardware and the Office of Academic Computing Services does not have enough money allocated to help every department that reports such problems[6]. From Figure 3.4, $30,000 is allocated for

[5]Concurrent number of users on the file server is obtained by the server statistics database.
[6]Money allocated for the 93 fiscal year

the purchase and upgrade of the network work equipment, which includes network interface card, network cables, network hard disk, etc. Although, $37,000 is allocated for the computer equipment, it is only used to purchase or upgrade workstations.

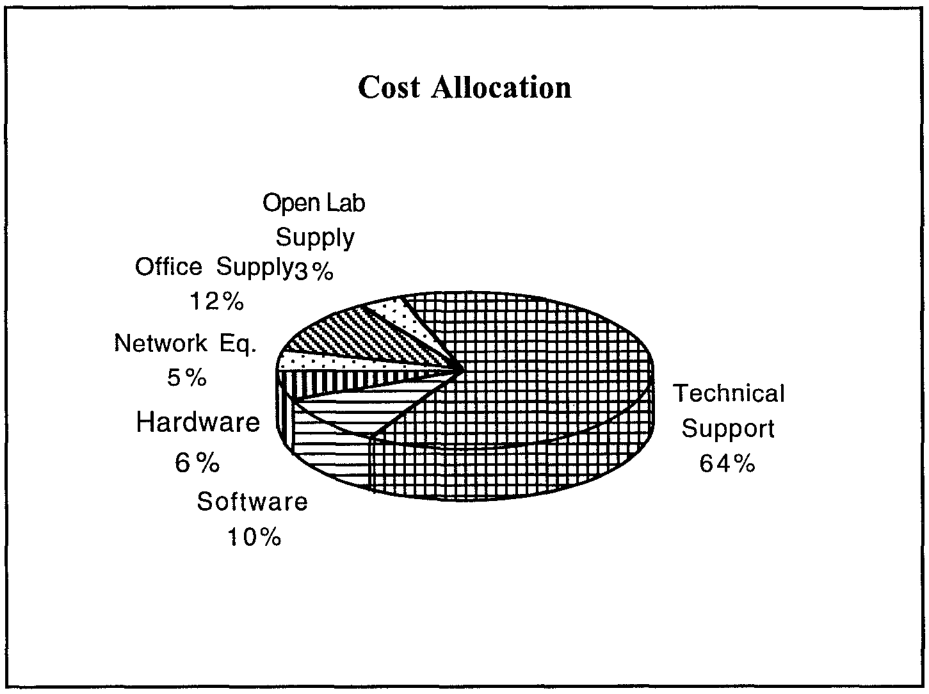


Figure 3.4 Cost Allocation

• User would like to use networked peripherals as easily as if they are attached to their workstations. Users would prefer Graphical User Interface to select networked devices.

The Graphical User Interface (GUI) is a rapidly developing computing environment. GUI replaces text-based commands with graphical representations, so that users don't have to memorize commands. In the LAN environment, where there are many network printers, hard disks, and other devices, GUI can be a pleasant interface to network users. For example, if a user wants to use a network printer in the Economics department, the actual command would resemble the following:

g:\system\capture /s=econ /q=apple_laser /nt /nb /ti=8

g:\system\capture is a network program to initiate the connection between the workstation and a network printer.

/s=econ means that the printer is located in the Economics department (Economics file server)

/q=apple_laser means that user want to use a printer named as "apple_laser". This unique name is predetermined when the printer is connected to the LAN.

/nt tells the printer not to use tab

/nb tells the printer not to print a banner (cover page that gets printed before every print job with user's name or ID.

/ti is a time-out value that server waits for the amount of time (in seconds) specified before sending a print job to the printer (8 seconds is a default)

Above is an example of one command that is used in Novell's LAN to connect to a printer. Novice users may not know the command, and it may be hard for even experienced users to remember the optional parameter. This text command can be replaced by a GUI such as following.

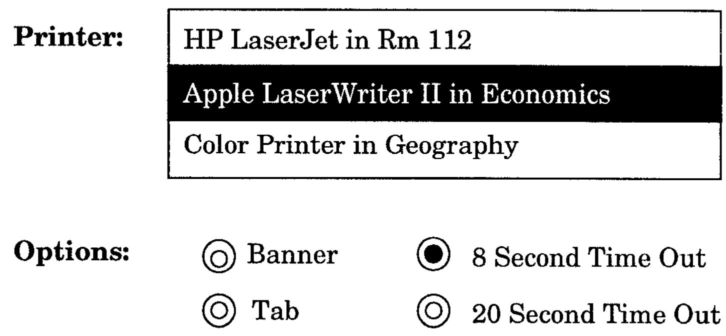


Figure 3.5 Graphical User Interface

By replacing a command line with GUI, users do not have to memorize commands or optional parameters. Such an implementation makes the LAN easy to use and results in high user satisfaction.

3.2 Identify Approaches

With performance problems defined, the primary approach to solving the problem must be derived. At this point the LAN manager can identify all feasible approaches that will lead to secondary solutions to the performance problem. All feasible candidates must be considered initially, but it should be noted that the more alternatives considered, the more complicated the performance analysis process becomes. Thus it is necessary to clearly distinguish between all possible approaches to ensure against inadvertent omissions or redundant submissions, and then to eliminate those candidates which are clearly unattractive or undesirable. For a complex system such as today's LAN, it may be hard to come up with any feasible approaches. Therefore, the LAN manager may seek possible approaches from outside the organization. The LAN manager may consider approaches that have been used by other organizations and apply them to customize their organization.

There are several feasible approaches that are identified in the previous step.

1. Increase Workstation Speed

   This approach is expected to contribute to the total performance of the LAN, since the faster workstation will be able to process information sent by the file server; however, it is costly to upgrade hundreds of workstations (There are approximately 586[7]-INTEL286 based workstation on the College of Behavioral and Social Sciences' LAN)

2. Increase Server Speed

   This approach is also expected to contribute the total performance of the LAN due to its faster processing power to process request from the workstation. This approach is quite acceptable in a sense that upgrading a CPU (motherboard) is fractional compare to upgrading several hundreds of workstations.

3. Increase Server Memory

   As discussed in chapter 2.4.2, memory size is also important in the Novell NetWare LAN environment, since the server maintains its own File Allocation Table (FAT). Any extra memory put into the file server should be used as a cache memory (server guesses which file will be requested next from the workstation, then reads the file into a cache memory to send a file quickly to the workstation). It is expected that the total performance of the LAN is due to the faster responses to workstations' request for file transfers.

4. Limit the Amount of Workstation on LAN

   In chapter 2.4.4 (Figure 2.8), the files server's throughput (megabits/second) increases as the number of workstations increase, however we can see that the throughput declines as the number of workstations exceeds thirty. Limiting thirty workstation to

---

[7]From the Inventory Data Sheet of the Office of Academic Computing Services (as of Sept. 1993)

the file server is not feasible considering the current situation of the college of Behavioral and Social Sciences' LAN (there are around 230 users connected to the file server during the business hours). However, it would be interesting to see if the number of workstations on the LAN make any contribution to the total performance in real working environment. The real working environment refers to the case where most people run word-processing and e-mail programs rather than constantly transferring files between the file server and workstation.

3.3 Identify Criteria and Throughput

The actual criteria selected for the evaluation process may vary considerably. Certain criteria may not be suitable for evaluating certain problems. The nature of the problem, the level of detail, and the level of accuracy required for analysis generally set the conditions for criteria selection. At the system level, parameters of primary concern include LAN performance capability (productivity, response time, utilization, etc.) and operational availability. The selected criteria are evaluated through measurements taken of criterion variables and calculations of their associated risk factors. It is hard to quantify performance capabilities such as productivity. Also, LAN managers may have several criteria since each department of an organization may have different priorities. The LAN manager must consider such a priority when selecting criteria and throughput.

While different organization may have different priorities on their LAN, OACS's main concerns are whether to increase the server memory, upgrade to faster speed workstations, or upgrade to faster server. These factors can be analyzed to see their contribution rate to the total performance of the LAN in terms of seconds. Each criterion is put to test to find their contribution to the performance of the LAN. Each criterion and throughput is discussed in chapter 4.

3.4 Determine and Validate Models

In this step, the LAN manager must transform the performance problem from a physical status into a symbolic (mathematical) model. Models have proven to be quite useful in the performance analysis process because they not only enable the analyst to simulate a real-world situation for evaluation, but they can often be manipulated more easily then a full-scale system. Several approaches can then be evaluated and any unforeseen modifications that the LAN may require can be observed and implemented. This aspect of modeling could inadvertently save time and money that would have inevitably been spent making modifications to the real LAN. A LAN manager must always remember that models are only approximations of a real situation. In addition to the obvious limitations of a modeling environment, there are other events that can occur in a real-world situation which cannot be predicted. Due to these uncertainties, it is virtually impossible to simulate a LAN's working environment completely, unless the testing is done on the real setting. Therefore, all assumptions must be clearly stated in order to correctly evaluate certain criteria variables which are dependent upon these assumptions and judge the merit (worth) of their evaluations. To ensure that a model which has been constructed to represent one of the LAN performance measurements is indeed the most accurate representation, its validity must be tested. Many organizations already have a heavy investment in LAN equipment (workstations) during the acquisition of LAN, so that it may be possible to consider using one department to postulate the model. Although each department may have different priorities, such a base modeling can be applied to the whole organization. For example, the Center for Substance Abuse at the University of Maryland has a LAN with 18 workstations. During the process of upgrading workstations, we were able to experience the obvious performance increase in LAN's response time. Although a test was not done in the most accurate manner, WordPerfect 5.1 (word processing program used widely throughout the

University of Maryland) opened 2 seconds faster with a new 486 computer than with a 386 workstation. At that time, we are not able to find out how much the total performance is improved over the previous, however, it is a definite criterion to be considered. The effect of using faster computer is calculated in the next chapter.

3.5 Collect/Generate Input Data

One of the most important steps in the performance analysis process is to specify the requirements for appropriate input data. The right type of data must be collected in a timely manner and presented in the proper format. To measure the performance of the LAN, 8 files were chosen totaling 1,940,386 bytes of information. This ensures that a fair comparison is made for the server memory size criterion. If one file is chosen, then it would be easier for the file server to guess where to read first, however, transferring 8 files somewhat eliminates an obvious advantage of a cache memory. This is a more realistic condition that may occur during the normal use of the LAN. Also, the data has to be large enough so that it will give sufficient time to fully interact with the LAN environment.

3.6 Execute Model

When data is gathered, the model is exercised to quantify the performance criteria. It is also important to calculate the standard error of the output to eliminate the uncertainty that LAN manager might have on certain key input parameters. By executing the model, the throughput is collected. The throughput is collected such that the computer displays the time when a task is started and displays the time when a task is completed. This method of collecting (instead of using a stopwatch) throughput will eliminate human response time of 0.53 seconds[8]: The mean response time is when a

---

[8]Mark S. Sanders and Ernest J. McCormick, <u>Human Factors In Engineering and</u>

person is anticipating to response, but mean response time increases to 0.73 seconds when the response was due to the surprise. Such a response time may lead to the false result and one must pay extra attention, especially when there is not much difference between the results.

3.7 Analyze Results

The execution of the model brings about results which must be analyzed. If various alternatives are considered as possible solutions to a performance problem, the results must be compared. At this point, the LAN manager is expected to tabulate or state the results in a form (mathematical model) which can be easily understood and prepare a list of recommendations. Once the result is presented as a mathematical model, it is clear to see how much each criteria has effect on the total performance of the LAN. It is easy to see estimate the result if there is any input change, therefore, it will be helpful during the decision process.

3.8 Decision Process

Finally, what remains is the decision-making process. The LAN manager presents recommendations to a superior, who then draws conclusions and assigns appropriate actions. Generally, this process is less complicated if the results have clarity and distinction. A corrective-action-loop may sometimes take place within the decision process. It occurs when the authority is dissatisfied with the performance of the LAN. Based on the recommendations, a corrective action is initialed that will require further analysis and perhaps another decision. This process generally continues until the performance of the LAN is considered to be acceptable.

Design, Sixth Edition, 1987. McGraw-Hill, Chapter 8.

3.9 Performance Analysis

We have seen the importance of performance analysis in analyzing a LAN. Using the performance measurements outlined above, the LAN manager can accurately quantify the various performance aspects of a LAN. The process detailed in this chapter provides a method for performing a complete performance analysis. Although the design of experiment in this thesis work was done after the installation of the LAN, it would be advised that such a technique is used from the beginning of the LAN acquisition. When the performance analysis process is used throughout the LAN acquisition, the LAN managers can have confidence that the LAN will be delivered meeting requirements, on schedule, and within the cost allocated. In the next chapter, the LAN is analyzed using performance analysis process.

Chapter 4 Quantitative Analysis of LANs

In a previous chapter, we defined performance analysis process. A LAN manager needs to use techniques that can help him/her decide which approach is suitable for his/her particular environment. This chapter shows a statistical technique that can help the LAN manager analyze the LAN's performance.

4.1 Statistical Quality Control in LAN Design/Improvement

Statistical Quality Control (SQC) is a collection of techniques, controls, and processes to be used for the improvement of the quality design of a system. Such a technique can also be applied to the design and improvement of a LAN. This can be accomplished through the collection of data related to the implementation and operations that are predetermined by the design of the LAN. By applying SQC properly on a LAN specification, considerable faith can be placed in the quality of the desired LAN performance. The desired performance is achieved by fulfilling the LAN requirements that are based on research regarding the user's needs. Therefore, to produce a high performance LAN, there must be teamwork (designers, managers, and users) throughout all the stages of the design. This shows that SQC requires an inter-human understanding and commitment through experimentation in LAN design and improvement.

4.2 High Performance LAN through Experimentation

Statistics is concerned with the design of experiments, the analysis of data, and the drawing of conclusions (or inferences) based on the observed data. The observed data that is obtained is used to troubleshoot the current LAN for deficiencies in quality design and to develop better LANs. This can be done by extracting information from the observed data. Valid and meaningful data are available from either passive observation of the design process or from purposeful experiments. It has become clearly evident that LAN managers need to push the quality issue during the planning and execution of experiments, so that quality becomes an integral part of every aspect of the LAN life cycle. High performance LAN design through experimentation can be achieved through six stages of experimentation, where each stage depends on the information obtained from the previous stage. These stages are described in the next 6 sub-sections.

4.2.1 Process Analysis

In the process analysis stage, real problems are translated into the design of experiments problems. This involves five activities.

The five activities are as follows:

1. Choosing the factors to study.

   As discussed in the previous chapter, some of LAN elements contribute directly to the LAN's performance. For example, the file server has the highest I/O throughput (megabits are transferred in one second) when there was 24 workstations on the LAN rather than 4 workstations on the LAN (discussed in the chapter 2.4). While different organizations may have different priorities on their LAN, OACS's main concerns are whether to increase the server memory, upgrade to a faster workstations, or upgrade to faster server. These factors can be analyzed to see their contribution rate to the total performance of the LAN.

2. Choosing the responses to measure.

   The LAN performance can be translated into how fast users can retrieve a particular file or how fast users can run an application programs that reside on the server. Such a response can be measure by seconds. For example, in OACS's LAN environment, it takes as much as 10 seconds to load a WordPerfect (word processing software) program and most other network resident programs. Therefore, the response time is measured in seconds.

3. Picking the general form of the mathematical model that you think will best describe the process.

Mathematical model such as below will be used to describe the performance of the LAN in OACS. For three variables denoted as A, B, and C, one will have a mathematical model as follows:

$$Y_i = \text{Mean} + \frac{A}{2}X_1 + \frac{B}{2}X_2 + \frac{C}{2}X_3 + \frac{AB}{2}X_1X_2 + \frac{AC}{2}X_1X_3 + \frac{BC}{2}X_2X_3 + \frac{ABC}{2}X_1X_2X_3$$

$$\begin{aligned}
\text{For } y_i &= y_1, & X_1 &= -1, & X_2 &= -1, & X_3 &= -1 \\
y_i &= y_2, & X_1 &= +1, & X_2 &= -1, & X_3 &= -1 \\
y_i &= y_3, & X_1 &= -1, & X_2 &= +1, & X_3 &= -1 \\
y_i &= y_4, & X_1 &= +1, & X_2 &= +1, & X_3 &= -1 \\
y_i &= y_5, & X_1 &= -1, & X_2 &= -1, & X_3 &= +1 \\
y_i &= y_6, & X_1 &= +1, & X_2 &= -1, & X_3 &= +1 \\
y_i &= y_7, & X_1 &= -1, & X_2 &= +1, & X_3 &= +1 \\
y_i &= y_8, & X_1 &= +1, & X_2 &= +1, & X_3 &= +1
\end{aligned}$$

Such a mathematical model estimates the total response time (Y) of the LAN which depends on the individual factors (A, B, C) and interactive terms between factors (AB, AC, BC, ABC). This model can be also be expanded to accommodate more than 3 factors. The development of such a mathematical model is described in later sections.

4. Deciding whether the experiment needs to be run in groups of runs (blocks). The experiment will be run in groups of 5. When the each test conditions is set up, 5 runs are made for each test setting to find the averages and their estimated errors.

5. Deciding whether to aggregate raw responses and how to aggregate them. Responses are the time it takes to transfer files from the file server to a workstation. Response is measured such that computer keeps the start and ending time to transfer files. Ending time is subtracted from the start time and the difference is the total response time to be recorded for analysis of data.

4.2.2 Performing the Runs

Having chosen the design, the next activity is to actually manipulate the LAN to implement the design. A useful tool at this stage is a work sheet. The work sheet indicates all the factor settings for each run, in units that are specified by the design. It also includes space for recording the values of all the responses for each run. It should clearly indicate the units of measurement and the number of decimals necessary for each response. The work sheet indicates the order in which the runs of the experiment should be performed, see Table 4.4.

4.2.3 Analyzing the Data

When the runs are complete, it is time to explore the data. This exploration results in a mathematical model that describes the relationship between the factors (LAN elements) and the responses (LAN response time). We will be able to come up with specified mathematical model (derived from the general form of the mathematical model) - some factors will have more effect on performance than others and some may not have any effects at all. Building this model involves using and understanding various graphs and analytical tools, see Section 4.4.

4.2.4 Conclusions and Recommendation

Armed with a mathematical model of the process, we can apply what we have learned to the real world. Using the model, we can conclude which approach (factor) made the biggest contribution or which approach (factor) did not make any contribution to the performance of the LAN. Then, we will be able to decide whether to spend money to upgrade a server, workstation, or other to achieve the higher LAN performance.

In many cases, results and conclusions from an experiment bring up new questions for further research. It is possible that the predictions made before the experiment can be quite far from the experimental result which can also lead us to focus on different aspect of the LAN elements. At any rate, the more we know about a process, the better the experiment will be. It makes sense to view an investigation as a sequence of experiments that give you a clearer understanding of the process. Figure 4.1 illustrates the stages of quality design process through experimentation using SQC.

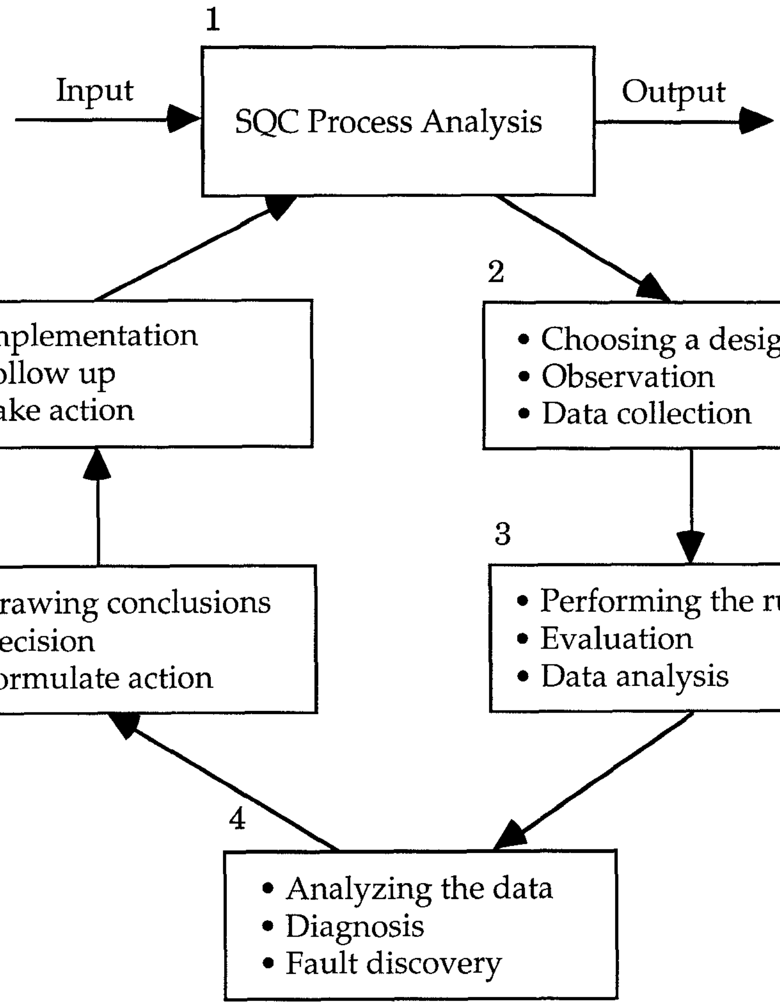


Figure 4.1 Quality Design Process Stages

4.3 Principles and Method in the Design of the Experiment

Randomization is a principle that should be applied for the purpose of randomly distributing unknowns: this will guarantee inferential validity in the face of an unspecified disturbance. Blocking is another principle which states that experiments should be made close together in time and space to assure that they have the same basis. Experiments made close together in time or space are often more similar than those made further apart, and hence can often provide a basis for blocking.

Factorial design (specifically, two-level factorial design), based on the principle of the orthogonal array, is a useful experiment to measure the effects of one or more variables on a response. A two-level factorial design which consists of four variables is denoted as 2^4 factorial design. A complete 2^3 factorial design contains $2^4 = 16$ unique test conditions. Also, the two-level factorial design is important for a number of reasons:

1. Two-level factorial designs require relatively few runs per factor studied; although they are unable to fully explore a wide region in the factor space, they can indicate major trends and so determine a promising direction for further experimentation.
2. When a more thorough local exploration is needed, two-level factorial design can be suitably augmented to form composite designs.
3. The interpretation of observations produced by the two-level factorial design can proceed largely by using common sense and elementary arithmetic.

The number of tests required by a full 2^k factorial design increases geometrically as k is increased. However, as k gets bigger, the desired result can often be obtained by performing only a fraction of the full factorial design. Such a technique is called the fractional factorial design.

4.4 Factorial Design in the LAN Environment

The LAN performance yield is related to several factors: workstation speed, server speed, server memory size, and number of workstations connected in the Novell LAN environment. In order to achieve a the highest performance yield possible, it is important to know how these four factors are related to overall network performance.

However, there are many more factors that contribute to the LAN performance: packet burst mode, server BUS architecture, interface card, etc.

File servers and workstations running at clock speeds ranging from 33Mhz and 66Mhz are considered for the experiment. A file server with a 500MB hard disk running the NetWare 3.11 operating system requires 6MB (see section 4.2). Since the server maintains its own File Allocation Table (FAT), a table that keeps track of where files are located in memory, as well as LAN operating systems, a total server memory ranging from 6MB to 8MB is used. This is done to determine the effect of having more memory than the minimum requirement. The number of workstation connected to the file server is varied from 2 to 8, to see whether the server idle time has any effect on the LAN performance. Workstations, except the two workstations in testing, are IBM ValuePoint computers running at 33Mhz. The testing environment was constructed using a coaxial cabling scheme, see Table 4.2. The following hardware is used for the experiment.

File Server:	CPU: 486DX/2 running at 66Mhz
	BUS: EISA
	Network Card: 3Com 3c579 (EISA based)
	Hard disk: 500MB SCSI hard disk (12ms)
	Total memory: 6MB (1-4MB module and 1-2MB module)
	8MB (2-4MB module)

File Server:	CPU: 486DX running at 33Mhz
	BUS: EISA
	Network Card: 3Com 3c579 (EISA based)
	Hard disk: 500MB SCSI hard disk (12ms)

Total memory: 6MB (1-4MB module and 1-2MB module)

8MB (2-4MB module)

Workstation: CPU: 486DX/2 running at 66Mhz

BUS: ISA

Network Card: 3Com 3c509 (ISA based)

Workstation: CPU: 486DX running at 33Mhz

BUS: ISA

Network Card: 3Com 3c509 (ISA based)

Two levels were chosen for each variable. The high and low levels of the four variables are listed in Table 4.1. Workstation and server speed are chosen because 66Mhz is the fastest PC on the market and 33Mhz is the most popular speed for PCs. 6MB of server memory is a minimum requirement for the tested server configuration and 8MB of server memory gives 2MB extra cache memory (the server reads data into cache memory before the workstation requests it, for quick response). Two workstations on a LAN would be the smallest LAN (one workstation would not require a LAN) and 8 workstations on a LAN can represent a small working group.

Variable	Condition	Low Level	High Level
Workstation speed	Mhz	33	66
Server speed	Mhz	33	66
Server memory	MB	6	8
Number of workstations on LAN	workstations	2	8

Table 4.1 Variable Levels for the LAN Experiment

With the four variables shown in Table 4.1, 16 different tests (2^4 factorial design) can be performed. The test combination (test 1 through test 16) can be found on Table 4.2 and Table 4.3.

Test Number	Workstation speed	Server speed	Server memory	Number of workstations
1	33	33	6	2
2	66	33	6	2
3	33	66	6	2
4	66	66	6	2
5	33	33	8	2
6	66	33	8	2
7	33	66	8	2
8	66	66	8	2
9	33	33	6	8
10	66	33	6	8
11	33	66	6	8
12	66	66	6	8
13	33	33	8	8
14	66	33	8	8
15	33	66	8	8
16	66	66	8	8

Table 4.2 Data from a Factorial Design (original units of the variable)

The table below (Table 4.3) shows the coded units of the variables. The variables are coded such that the high level will be denoted by +1, and the low level by -1. By doing this, regardless of the physical conditions represented by the two levels, the basic design of any two-level factorial becomes a simple tabulation of a systematic arrangement of plus and minus "ones". Low level (-1) and high level (+1) are chosen in the following manner:

Workstation Speed and Server Speed:

Average: $\frac{33+66}{2}=49.5$ Average of the difference: $\frac{66-33}{2}=16.5$

Low level: $\frac{33-49.5}{16.5}=-1$ High level: $\frac{66-49.5}{16.5}=+1$

Server Memory:

Average: $\frac{6+8}{2}=7$ Average of the difference: $\frac{8-6}{2}=1$

Low level: $\frac{6-7}{1}=-1$ High level: $\frac{8-7}{1}=+1$

Number of Workstations:

Average: $\frac{2+8}{2}=5$ Average of the difference: $\frac{8-2}{2}=3$

Low level: $\frac{2-5}{3}=-1$ High level: $\frac{8-5}{3}=+1$

The best way to write down the coded units (-1,+1) for the sixteen test conditions is as follows:

1. For the variable workstation speed, write down a column of -1, +1, -1, +1, -1, +1,

-1, +1, -1, +1, -1, +1, -1, +1, -1, +1. The signs alternate each time (i.e., $2^0 = 1$, one alternation every time). The column length is 2^k, where k is the number of variable.

2. For the variable server speed, write down a column of -1, -1, +1, +1, -1, -1, +1, + 1, -1, -1, +1, + 1, -1, -1, +1, + 1. The signs alternate in pairs (i.e., $2^1 = 2$, alternate in pairs).

3. For the variable server memory size, write down a column of -1, -1, -1, -1, +l +1, +1, +1, -1, -1, -1, -1, +l +1, +1, +1. The signs alternate in groups of four (i.e., $2^2 = 4$, alternate in groups of four).

4. For the variable number of workstations, write down a column of -1, -1, -1, -1, -1, -1, -1, -1, +1, +1, +1, +1, +1, +1, +1, +1. The signs alternate in groups of eight (i.e., $2^3 = 8$, alternate in groups of eight).

Test Number	Workstation speed	Server speed	Server memory	Number of workstations
1	-	-	-	-
2	+	-	-	-
3	-	+	-	-
4	+	+	-	-
5	-	-	+	-
6	+	-	+	-
7	-	+	+	-
8	+	+	+	-
9	-	-	-	+
10	+	-	-	+
11	-	+	-	+
12	+	+	-	+
13	-	-	+	+
14	+	-	+	+
15	-	+	+	+
16	+	+	+	+

Table 4.3 Data from a Factorial Design (coded units of variables)

4.4.1 Collection of Data

The experiment was set up to collect data from sixteen different tests as in Table 4.2 and 4.3. Then each test was run 5 times and its result are averaged and a standard deviation is calculated, see Table 4.4 below. The test result represents the number of seconds it took to transfer 1,940,386 bytes of information (consisting of 8 files) between the file server and the workstation.

The results from a 2^4 factorial design are shown in Table 4.4, and the graphical representation of the test is given in Figure 4.3. If only the three variables are considered, the 2^3 factorial design can be represented geometrically as a cube on the three mutually perpendicular coordinate axes X_1, X_2, and X_3, as shown in Figure 4.2. The numbers encircled at the eight corner points of the cube represent the corresponding test numbers in standard order. However, with the four variables, two of the 2^3 cubes will represent the 2^4 factorial design: one cube for low level of the fourth variable and another cube for high level of fourth variable, see Figure 4.3.

Test	1	2	3	4	5	Avg	Std	s^2
1	48	48	47	47	47	47.4	.55	.3
2	42	43	41	42	42	42.0	.71	.5
3	46	47	46	46	47	46.4	.55	.3
4	40	41	39	40	39	39.8	.84	.7
5	47	48	47	47	47	47.2	.45	.2
6	42	43	42	42	42	42.2	.45	.2
7	47	46	46	46	47	46.4	.55	.3
8	40	41	40	40	39	40.0	.55	.3
9	46	46	46	47	46	46.2	.45	.2
10	42	41	40	42	41	41.2	.84	.7
11	46	46	44	45	46	45.4	.89	.8
12	39	39	40	39	39	39.2	.45	.2
13	46	47	47	46	46	46.4	.55	.3
14	42	41	40	41	41	41.0	.71	.5
15	45	46	45	45	45	45.2	.45	.2
16	40	39	38	39	39	39	.71	.5

Table 4.4 Data Sheet (units are in seconds)

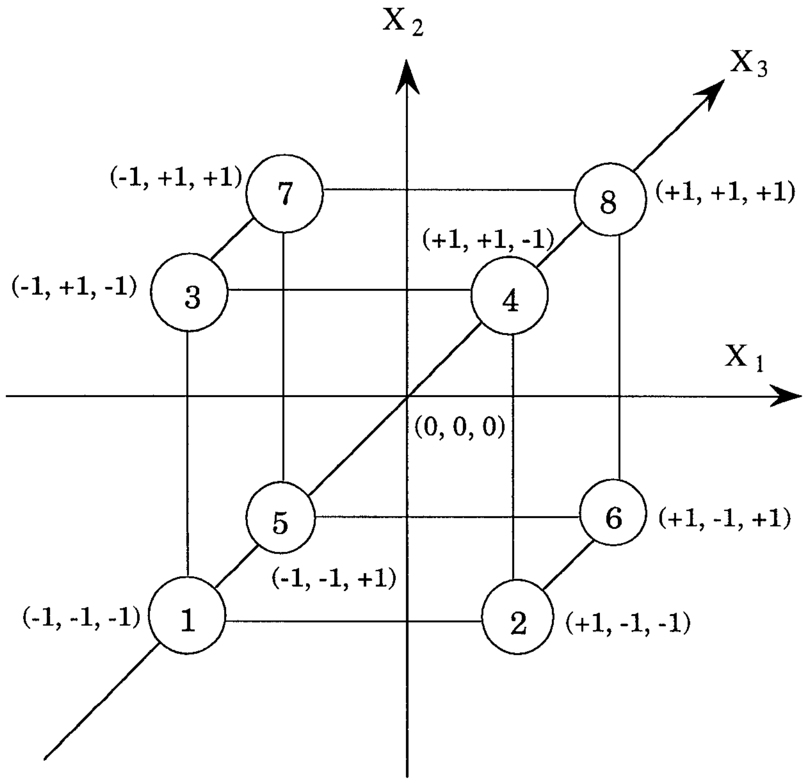


Figure 4.2 Graphical Representation of the 2^3 Factorial Design

With the four variables, two of the 2^3 cubes will represent the 2^4 factorial design: one cube for low level of the fourth variable and another cube for high level of fourth variable, see Figure 4.3.

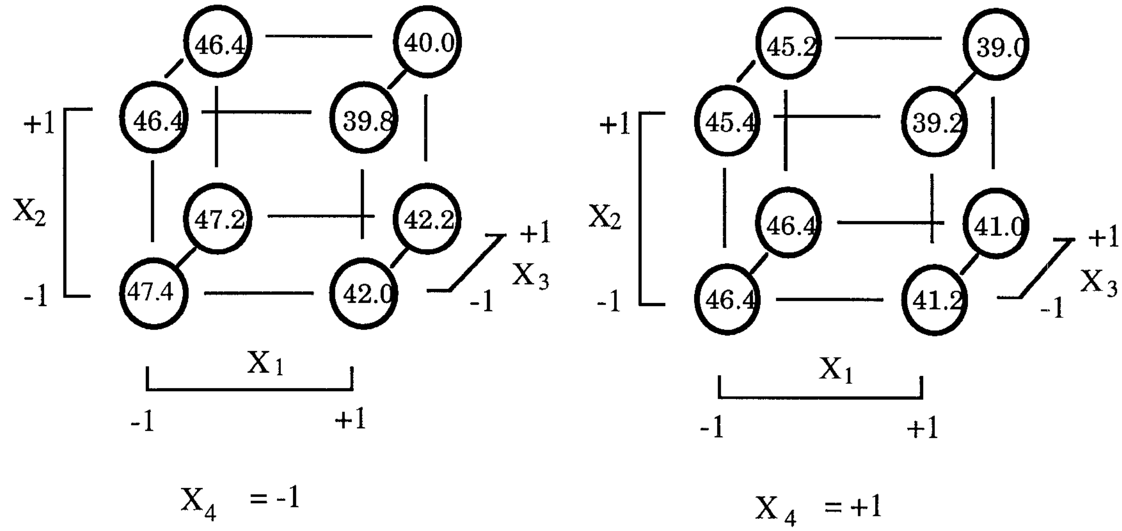


Figure 4.3 Graphical Representation of the 2^4 Factorial Design

Table 4.5 below show the average, main effects, two-factor interaction effects, three-factors interactions effects, and four-factors interactions effects. Effects which are underlined indicates that they made a significant contribution to the LAN's performance.

The pooled estimate of run variance is

$$s^2 = \frac{s_1^2 + s_2^2 + \ldots + s_{16}^2}{16}$$

The pooled estimate of run variance s^2 can be calculated from the last column in Table 4.4.

$$s^2_{pooled} = \frac{.3+.5+.3+.7+.2+.2+.3+.3+.2+.7+.8+.2+.3+.5+.2+.5}{16} = \frac{6.2}{16} = .38$$

Using this pooled estimate of run variance, we are able to estimate the variation associated with effects. Since each main effect and interaction is a statistic of the form $\bar{y}_{+} - \bar{y}_{-}$, the total runs made in conducting a two-level factorial design was 80. As a result, the variance of each effect is given by

$$\text{Variance(effect)} = \text{Var}(\bar{y}_{+} - \bar{y}_{-})$$

$$s^2_{effect} = \frac{4}{N} s^2_{pooled} = .019$$

N = the total number of runs (80)

$$s_{effect} = .14$$

Using this information, we are able to identify those effects which are not significant from a statistical point of view, as seen in Table 4.5. From Table 4.5, the server memory (M) has no significant main effect or any interactive effects. This suggests that X3 (server memory) does not have any effect on LAN performance. Therefore, the 2^4 factorial design can be treated as 2^3 factorial design. From the graphical representation of the 2^4 factorial design shown in Figure 4.3, the cube on the left (X4=-1) represents the result when 2 workstations are on the LAN, while the cube on the right (X4=+1) represents the result when 8 workstations are on the LAN. By combining the front 4 encircled numbers with the back 4 encircled numbers, X3 (server memory) can be eliminated, as shown in Figure 4.4.

Effect	Effect ± standard error
Average	43.44 ± .07 (.14 / 2)
Main effects	
Workstation Speed, W	-5.77 ± .14
Server Speed, S	-1.53 ± .14
Server Memory, M	-0.03 ± .14
# of Workstation on LAN, N	-0.98 ± .14
Two-factors interactions	
WS	-0.57 ± .14
WM	0.03 ± .14
WN	-0.08 ± .14
SM	-0.03 ± .14
SN	-0.03 ± .14
MN	-0.08 ± .14
Three-factor interaction	
WSM	0.03 ± .14
WSN	0.08 ± .14
SMN	-0.08 ± .14
WMN	-0.13 ± .14
Four-factor interaction	
WSMN	0.075 ± .14

Table 4.5 Effects and Standard Errors for 2^4 Factorial Design

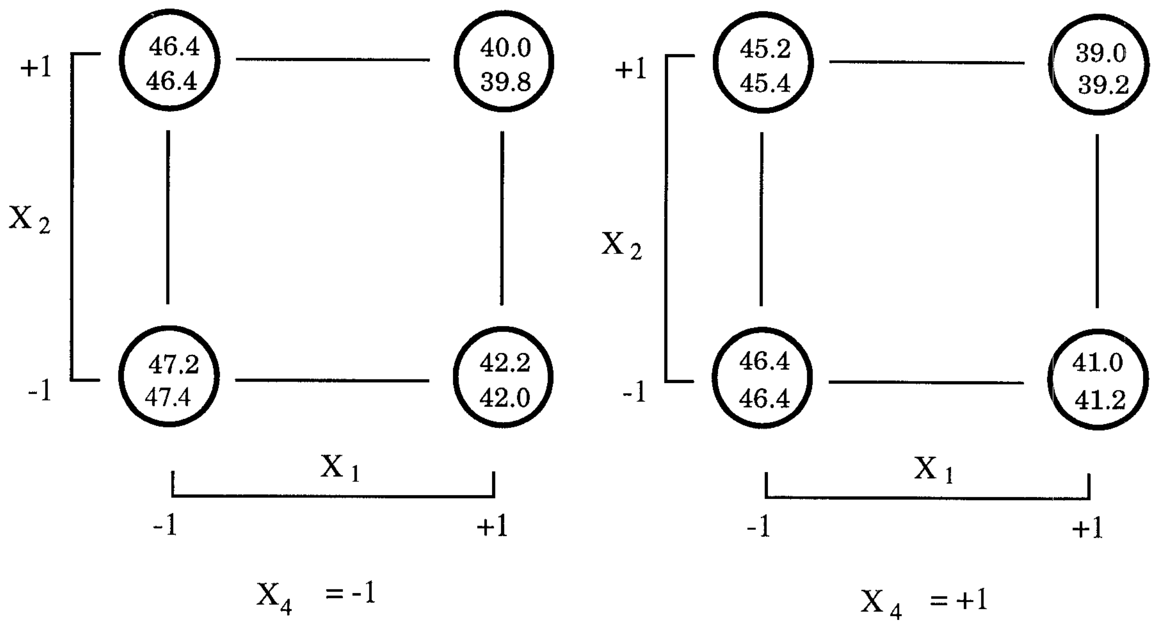


Figure 4.4 Elimination of X3 Variable

Once, X3 (server memory) is eliminated, the graphical representation of 2^3 factorial design can be drawn by combining two two-way diagrams, as shown in the Figure 4.3. The left two-way diagram goes to the front of the cube and the right two-way diagram goes to the back of the cube, making the graphical representation of the 2^3 factorial design, see the Figure 4.5.

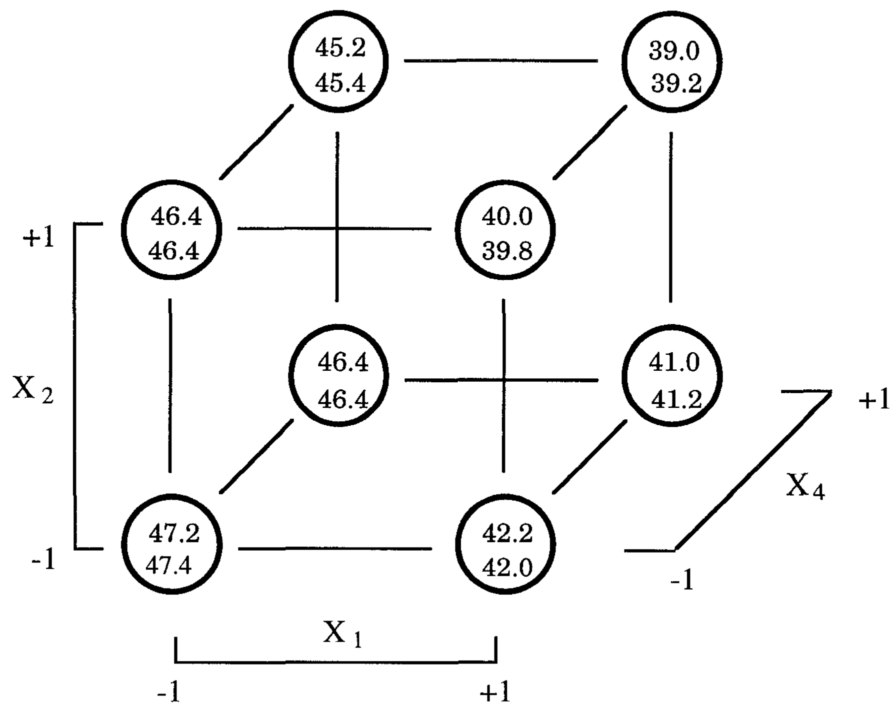


Figure 4.5 The 2^3 Factorial Design with Experimental Result

4.4.2 2^3 Factorial Design

With only three variables remaining, the 2^4 factorial design experiment becomes a 2^3 factorial design (8 different tests). Table 4.6 shows variable levels of only 3 variables without the variable 4 (the server memory size).

Variables	Condition	Low Level	High Level
Workstation speed	Mhz	33	66
Server speed	Mhz	33	66
Number of workstations on LAN	workstations	2	8

Table 4.6 Variable Levels for the LAN Experiment

With two levels for each of the three factors (Table 4.6), it is possible to prepare eight tests with the conditions shown in Table 4.7.

Test Number	Workstation Speed (Mhz)	Server Speed (Mhz)	Workstation Number (#)
1	33	33	2
2	66	33	2
3	33	66	2
4	66	66	2
5	33	33	8
6	66	33	8
7	33	66	8
8	66	66	8

Table 4.7 Data from the 2^3 Factorial Design (original units of variable)

With the 16 tests performed previously (Table 4.4), it is feasible to combine tests 1 through 4 with tests 4 through 8, since they have same test conditions (workstation speed, server speed, and number of workstations). In a similar manner, tests 9 through 13 can be combined with tests 14 through 18. As a result, each test now has 10 values instead of 5 as shown in the Table 4.8.

Test	1	2	3	4	5	6	7	8	9	10	Avg	Std	s^2
1	48	48	47	47	47	47	48	47	47	47	47.3	.48	.23
2	42	43	41	42	42	42	43	42	42	42	42.1	.57	.32
3	46	47	46	46	47	47	46	46	46	47	46.4	.52	.27
4	40	41	39	40	39	40	41	40	40	39	39.9	.74	.54
5	46	46	46	47	46	46	47	47	46	46	46.3	.48	.23
6	42	41	40	42	41	42	41	40	41	41	41.1	.74	.54
7	46	46	44	45	46	45	46	45	45	45	45.3	.67	.46
8	39	39	40	39	39	40	39	38	39	39	39.1	.57	.32

Table 4.8 Data Sheet for 2^3 (units are seconds)

Now, using Tables 4.8 and 4.9, the geometric representation of the 2^3 factorial design can be set up as shown in Figure 4.6.

Test Number	Server Architecture	Workstation CPU	Workstation Architecture
1	-	-	-
2	+	-	-
3	-	+	-
4	+	+	-
5	-	-	+
6	+	-	+
7	-	+	+
8	+	+	+

Table 4.9 Data from the 2^3 Factorial Design (coded units of variables)

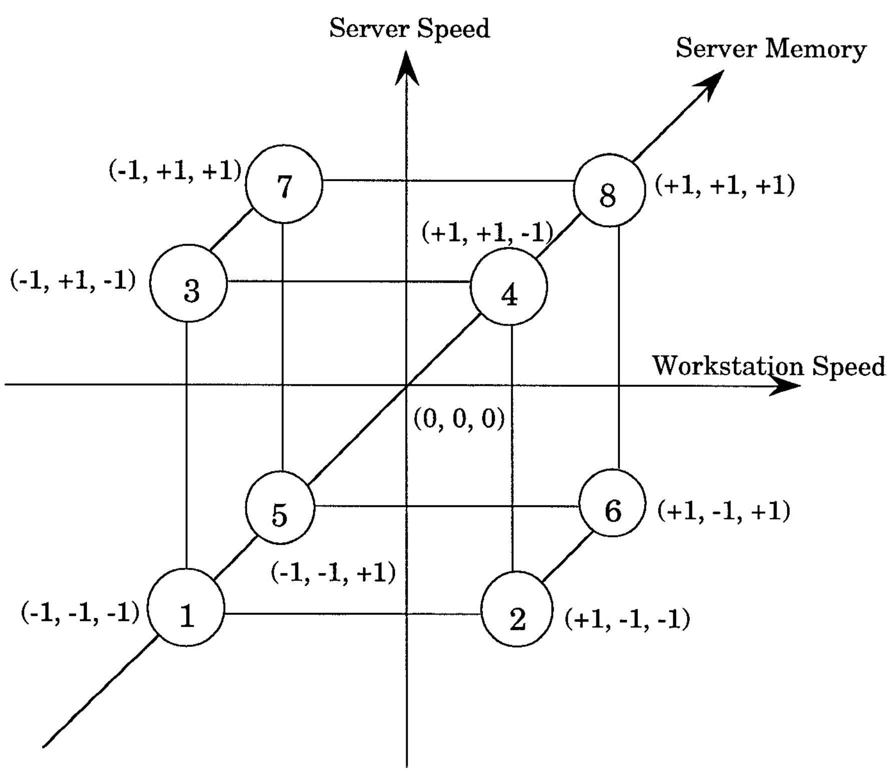


Figure 4.6 Graphical Representation of a 2^3 Design for LAN

Figures 4.6 and 4.7 illustrate the graphical representation of the factorial design of network performance. To draw a graphical representation, include the three variables as three mutually perpendicular coordinate axes: workstation speed, server speed, and the number of workstations on the LAN; then the 2^3 factorial design can be represented geometrically as a cube.

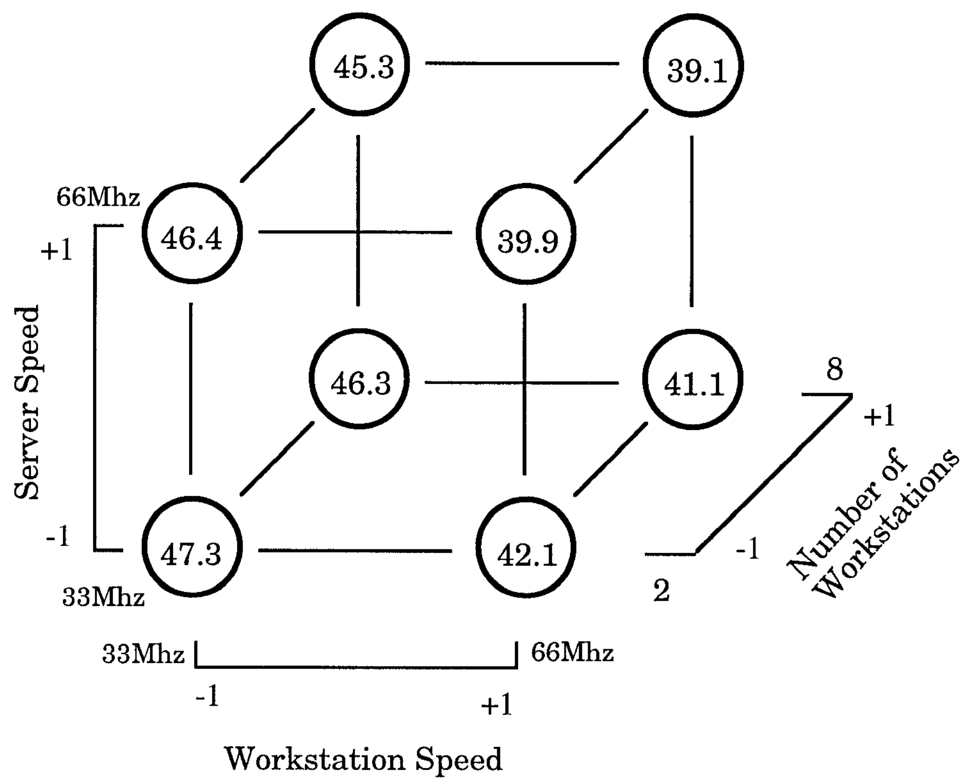


Figure 4.7 The Cube (2^3 Factorial Design for LAN)

4.4.2.1 Calculation of the Main Effects

We group the eight yield into four groups; for example, (y_2, y_1), (y_4, y_3), (y_6, y_5), and (y_8, y_7). Each of the four groups represents a pair of tests which were performed with the same server speed and number of workstations on the LAN, but with different workstation speed - ranging from 33Mhz to 66Mhz. Therefore, the difference in the yield between y_2 and y_1 represents the effect of the workstation speed on the result when 2 workstations are connected to a 33Mhz file server.

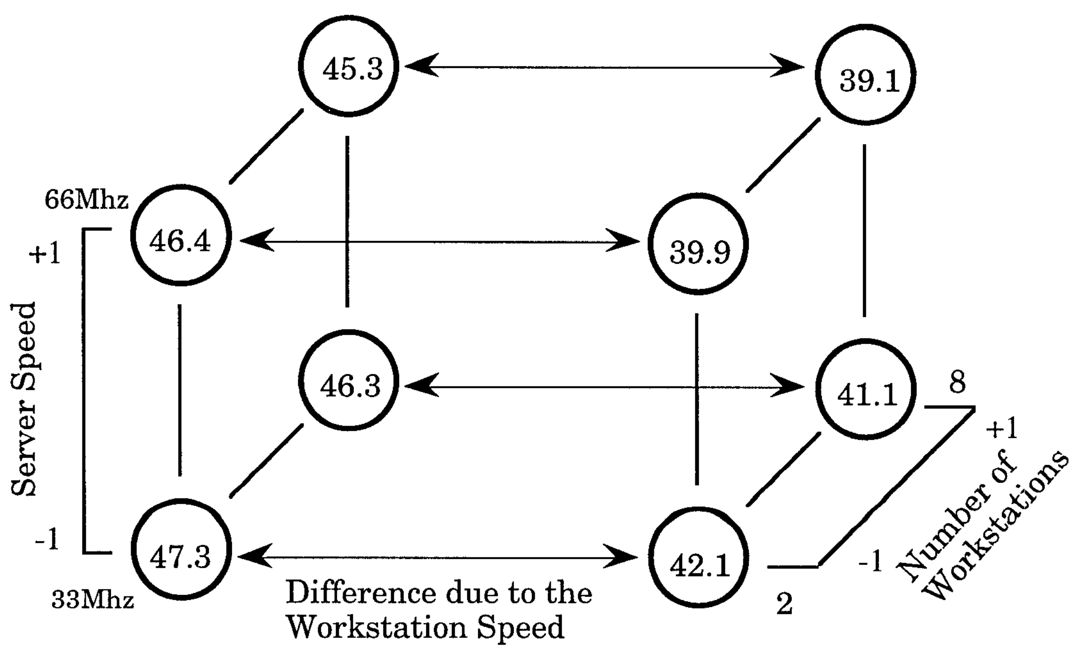


Figure 4.8 Main Effect of the Workstation Speed

The four groups give four differences in the result, which are due to workstation speed variation from high(66Mhz) to low(33Mhz), see Figure 4.8 and Table 4.10. The main effect of the workstation speed shows the average effect of the workstation speed over all conditions of the other variables.

Condition at which comparison is made		
Individual Contrasts	Server Speed	# of Workstations
$y_2 - y_1 = 42.1 - 47.3 = - 5.2$	33	2
$y_4 - y_3 = 39.9 - 46.4 = - 6.5$	66	2
$y_6 - y_5 = 41.1 - 46.3 = - 5.2$	33	8
$y_8 - y_7 = 39.1 - 45.3 = - 6.2$	66	8
Main Effect of Workstation Speed: ((-5.2) + (-6.5) + (-5.2) + (-6.2)) / 4 = - 5.8		

Table 4.10 Main Effect of the Workstation Speed

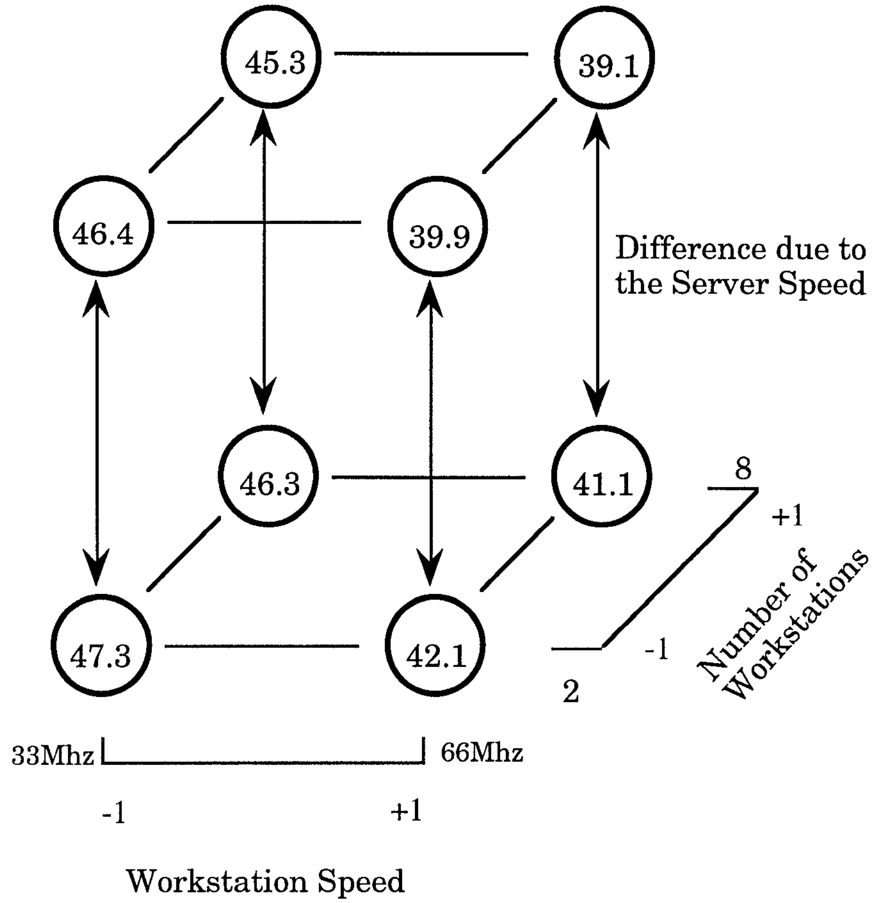


Figure 4.9 Main Effect of the Server Speed

From the figure above (Figure 4.9), the four groups give four differences in the result, which are due to server speed variation from high(66Mhz) to low(33Mhz), see Table 4.11. The Figure 4.9 illustrates the difference in yield due to the difference in the server speed. The main effect of the server speed shows the average effect of the workstation speed over all conditions of the other variables.

Condition at which comparison is made		
Individual Contrasts	Workstation Speed	# of Workstation
$y_3 - y_1 = 46.4 - 47.3 = -0.9$	33	2
$y_4 - y_2 = 39.9 - 42.1 = -2.2$	66	2
$y_7 - y_5 = 45.3 - 46.3 = -1.0$	33	8
$y_8 - y_6 = 39.1 - 41.1 = -2.0$	66	8
Main Effect of Server Speed: ((-0.9) + (-2.2) + (-1.0) + (-2.0)) / 4 = -1.5		

Table 4.11 Main Effect of the Server Speed

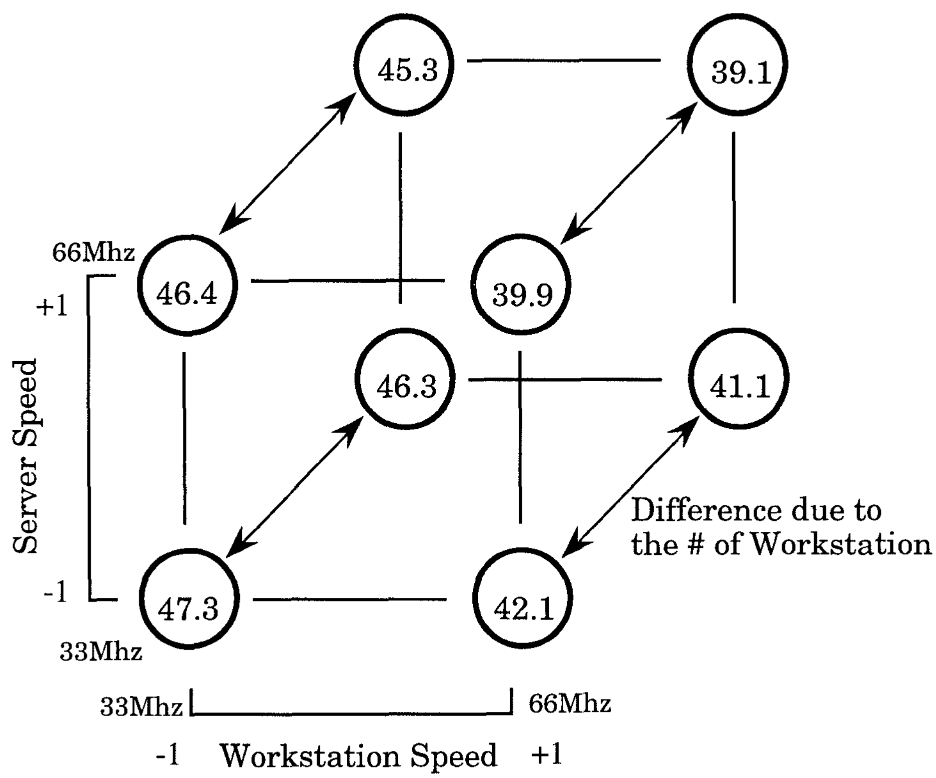


Figure 4.10 Main Effect of the # of Workstations

From Figure 4.10, the four groups give four differences in the result, which are due to the number of workstations on the LAN ranging from high (8 workstations) to low (2 workstations), see Table 4.12. Figure 4.10 above illustrates the yield difference due to the number of workstations on the LAN. The main effect of the server speed measures the average effect of the workstation speed over all conditions of the other variables as calculated on the Table 4.12 below.

Condition at which the comparison is made		
Individual Contrasts	Workstation Speed	Server Speed
$y_5 - y_1 = 46.3 - 47.3 = -1.0$	33	33
$y_6 - y_2 = 41.4 - 42.1 = -1.0$	66	33
$y_7 - y_3 = 45.3 - 46.4 = -1.1$	33	66
$y_8 - y_4 = 39.1 - 39.9 = -0.8$	66	66
Main Effect of the Number of Workstations $((-1.0) + (-1.0) + (-1.1) + (-0.8)) / 4 = -1.0$		

Table 4.12 Main Effect of the Number of Workstations

The main effect for each of the three variables can be seen as the difference between the two averages such that the main effect = $\bar{y}_+ - \bar{y}_-$, where $\bar{y}_+$ is the average response for the high level (+1) of the variables and $\bar{y}_-$ is the average response for the low level (-1). Thus, the main effect can also be calculated as seen below. The result shown below has the same effect as in Table 4.10.

$$Workstation\_Speed = \frac{y_2+y_4+y_6+y_8}{4} - \frac{y_1+y_3+y_5+y_7}{4}$$

$$Workstation\_Speed = \frac{42.1+39.9+41.1+39.1}{4} - \frac{47.3+46.4+46.3+45.3}{4}$$

$$Workstation\_Speed = 40.55 - 46.32 = -5.77$$

The same calculation can be done for the main effect of the workstation speed and the server speed.

Server Speed Effect:

$$Server\_Speed = \frac{y_3+y_4+y_7+y_8}{4} - \frac{y_1+y_2+y_5+y_6}{4}$$

$$Server\_Speed = \frac{46,4+39,9+45.3+39.1}{4} - \frac{47.3+42.1+46.3+41.1}{4}$$

$$Server\_Speed = 42.67 - 44.2 = -1.53$$

Number of the Workstations on the LAN effect:

$$\#\_Workstation = \frac{y_5+y_6+y_7+y_8}{4} - \frac{y_1+y_2+y_3+y_4}{4}$$

$$\#\_Workstation = \frac{46.3+41.1+45.3+39.1}{4} - \frac{47.3+42.1+46.4+39.9}{4}$$

$$\#\_Workstation = 42.95 - 43.93 = -0.98$$

4.4.2.2 Calculation of the Interaction Effect

To calculate the interaction effect, set up a matrix table as in Table 4.13. To simplify, workstation speed is denoted by W, server speed as S, and number of workstations by N. The interaction effects are designated by workstation speed and server speed (WS), workstation speed and number of workstations (WN), server speed and number of workstations (SN), and workstation speed, server speed, and number of workstations (WSN).

Test	W	S	N	WS	WN	SN	WSN	Y
1	-	-	-	+	+	+	-	47.3
2	+	-	-	-	-	+	+	42.1
3	-	+	-	-	+	-	+	46.4
4	+	+	-	+	-	-	-	39.9
5	-	-	+	+	-	-	+	46.3
6	+	-	+	-	+	-	-	41.1
7	-	+	+	-	-	+	-	45.3
8	+	+	+	+	+	+	+	39.1

Table 4.13 Design Matrix with Interaction Effects

By examining Table 4.10, it is evident that the workstation speed effect on the result is almost same, although a different number of workstations are attached to the file server . When 2 workstations are connected to the file server, the average workstation speed effect is:

$$((-5.2) + (-6.5))/2) = -5.85$$

When there are 8 workstations on the LAN, the average workstation speed effect is:

$$((-5.2) + (-6.2))/2) = -5.7$$

A measure of non-linearity or interaction between the workstation speed and the number of workstations on the LAN is supplied by the difference between the average workstation speed effect with 2 workstations on the LAN and the average workstation speed with 8 workstations on the LAN. Then, half of this difference is called the workstation speed by the number of workstations on the LAN interaction or the WN interaction. Therefore, from the average server architecture effect the WN interaction is:

$$\text{WN interaction} = ((-5.85) - (-5.7))/2 = -0.075$$

The WN interaction may also be thought of as one-half the difference in the average number of workstations effect at the two levels of workstation speed. In the 33Mhz workstations, the average number of workstations effect is (Table 4.12):

$$((-1.0) + (-1.1))/2 = -1.05$$

With 66Mhz workstations, the average workstation architecture effect is :

$$((-1.0) + (-0.8))/2 = -0.9$$

Therefore, the WN interaction is:

$$((-1.05) - (-0.9)) / 2 = -0.075$$

A geometrical interpretation of the WN interaction can also be made by pushing the top surface of the Figure 4.9 down to the bottom surface and averaging the two corner values at each of the four corners, see Figure 4.10.

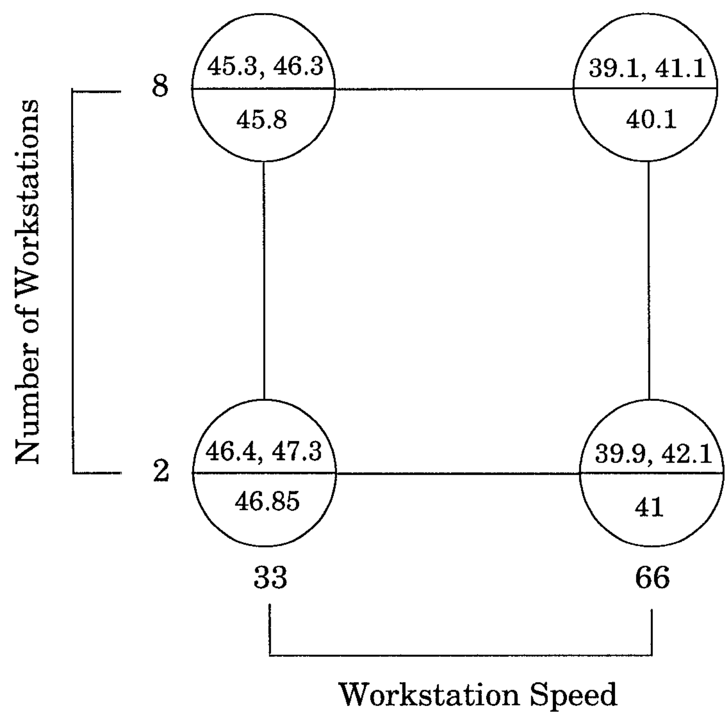


Figure 4.11 WN Interaction on Two-Dimension

Using Figure 4.11, the WN interaction can be drawn as below (Figure 4.12).

Using Figure 4.12, it is easy to see that the difference between the two increments is -0.15 (5.7-5.85= -0.15). Therefore, the WN interaction effect is:

$$WN = \frac{1}{2}x(-0.15) = -0.075$$

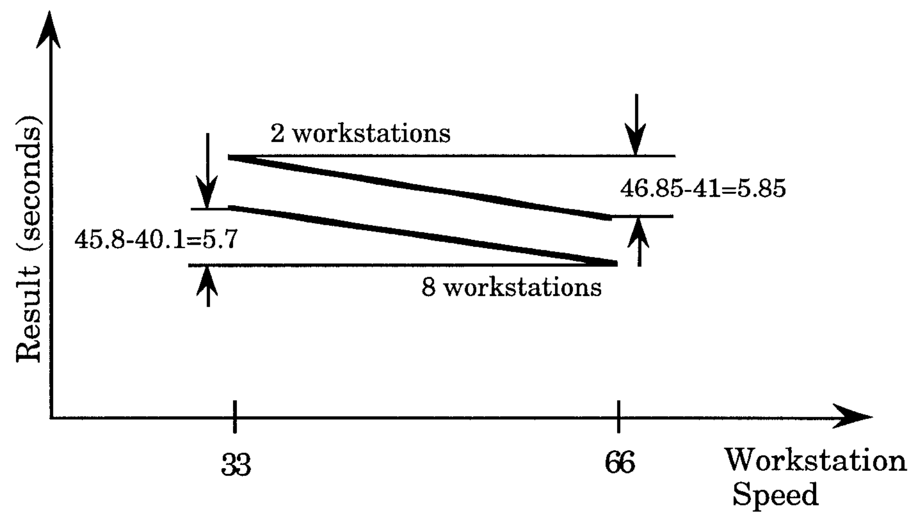


Figure 4.12 WN Interaction Graph

The WS interaction effects can also be calculated in the same manner and by a geometrical interpretation similar to figure above. A geometrical interpretation of the WS interaction could also be made by pulling the back surface of Figure 4.10 to the front surface and averaging the two corner values at each of the four corners, see Figure 4.13.

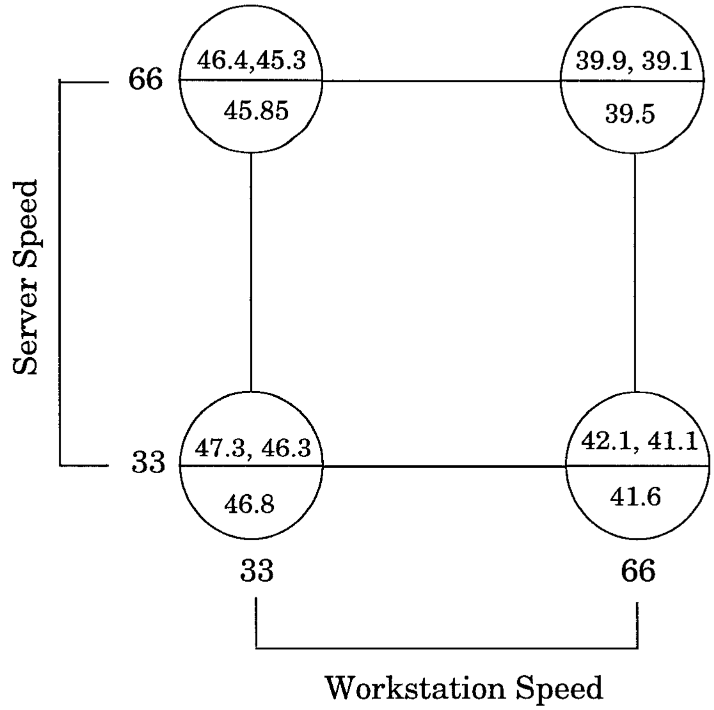


Figure 4.13 WS Interaction on Two-Dimension

Using Figure 4.13, WS interaction can be drawn as below (Figure 4.14). In Figure 4.13, it is easy to see that the difference between the two increments is -1.15 (5.2-6.35= -1.15). Therefore, the WS interactions become:

$$WS = \frac{1}{2} x(-1.15) = -0.575$$

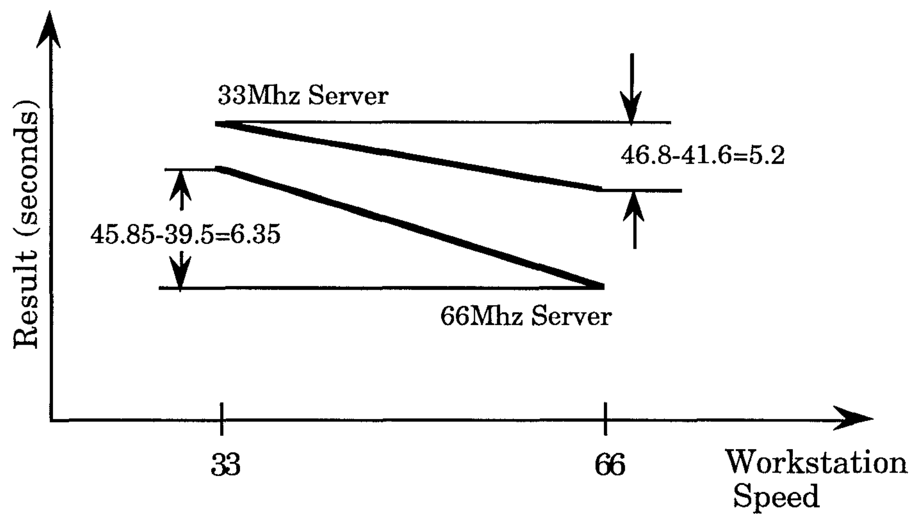


Figure 4.14 WS Interaction Graph

The SN interaction effects can also be calculated in the same manner and by geometrical interpretation similar to the figure above. Figure 4.15 is generated by combining the left surface with the right surface (see Figure 4.8) and averaging the two corner values at each of the four corners.

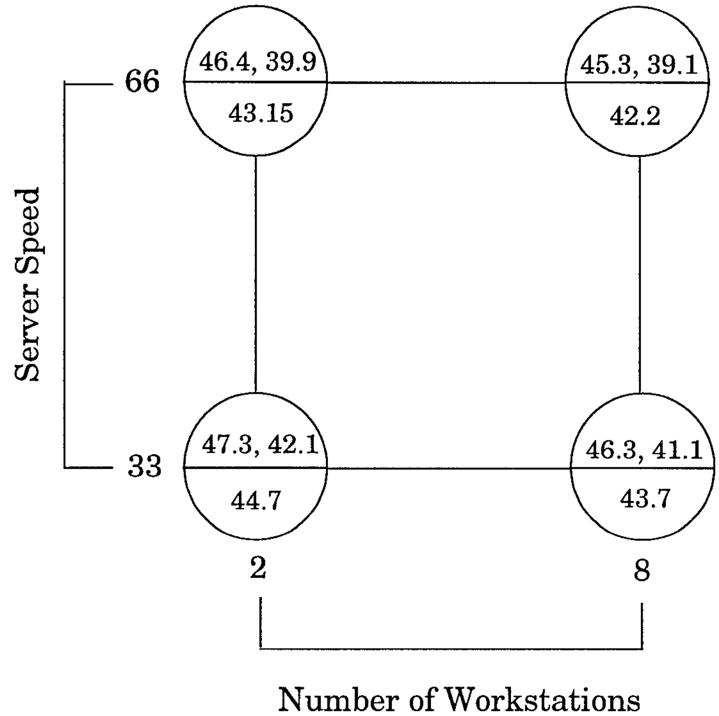


Figure 4.15 SN Interaction on Two-Dimension

Using Figure 4.15, the SN interaction can be drawn as below (Figure 4.16). In Figure 4.16, it is easy to see that the difference between the two increments is -0.05 (0.95-1.0= -0.05). Therefore, the SN interactions become:

$$SN = \frac{1}{2} x(-0.05) = -0.025$$

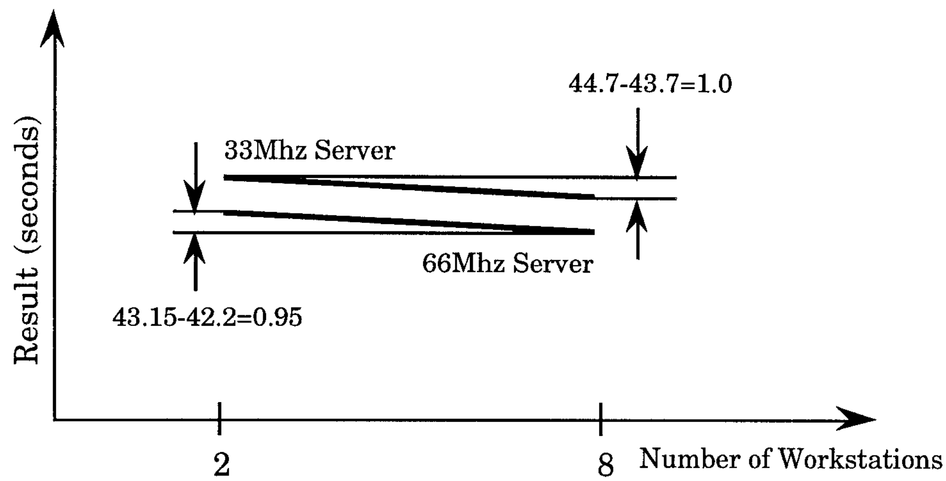


Figure 4.16 SN Interaction Graph

The three-variable interaction can be calculated by considering the workstation speed by server speed interaction WS. Two measures of the WS interaction are available from the experiment performed here. The two measures are taken from the number of workstations on the LAN. The WS interaction with 8 workstations on the LAN (+) is :

$$WS = \frac{(y_4 - y_3) - (y_2 - y_1)}{2} = \frac{(39.9 - 46.4) - (42.1 - 47.3)}{2}$$

$$WS = \frac{((-6.5) - (-5.2))}{2} = -0.65$$

The WS interaction with 2 workstations on the LAN (-) is:

$$WS = \frac{(y_8 - y_7) - (y_6 - y_5)}{2} = \frac{(39.1 - 45.3) - (41.1 - 46.3)}{2}$$

$$WS = \frac{((-6.2) - (-5.2))}{2} = -0.5$$

Half the difference of the two variable interaction defines the three variable interaction of workstation speed, server speed, and the number of workstations on LAN, WSN.

$$WSN = \frac{(-0.5 - (-0.65))}{2} = 0.075$$

To describe the three-factor interaction effect graphically, Figure 4.17 is used. The cube (Figure 4.7) is divided into 2 two-way diagram, one at the low level number of workstations on the LAN (2 workstations) and the other at the high level number of workstations on the LAN (8 workstations). These two two-way diagrams are used to calculate two-factor interactions of workstation speed and server speed, then the average difference of the two is contrasted for workstation speed, at the fast server speed and low server speed. Finally, the three factor interaction (WSN) is calculated by finding the average difference in the two local two-factor interactions calculated previously. The graphical representation of the WSN interaction is also shown in Figure 4.18.

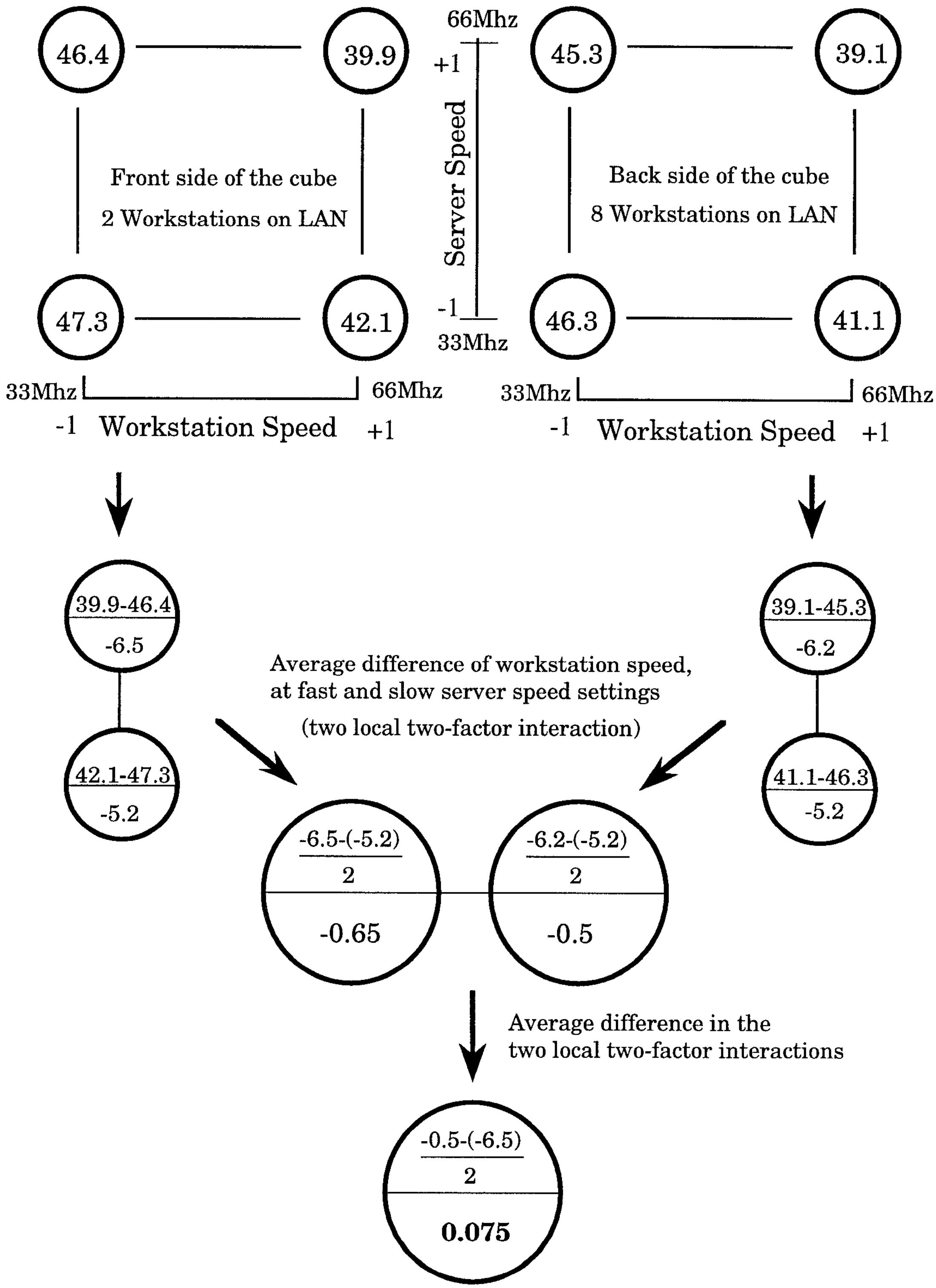
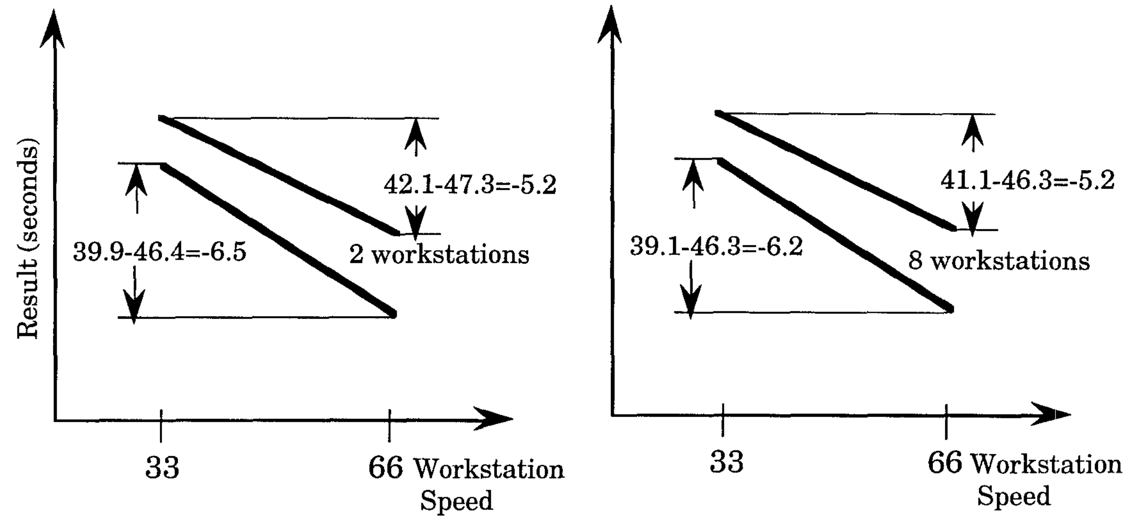


Figure 4.17 Three-Factor Interaction

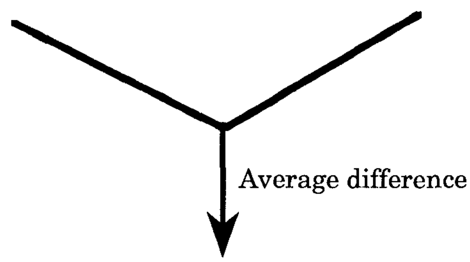


$$WS = \frac{((-6.5)-(-5.2))}{2} = -0.65$$

with 2 workstations on LAN

$$WS = \frac{((-6.2)-(-5.2))}{2} = -0.5$$

with 8 workstations on LAN



$$WSN = \frac{(-0.5-(-0.65))}{2} = 0.075$$

Figure 4.18 WSN Interaction Graph

4.4.2.3 Generalized Method for the Calculation of Effects

With the understanding of the average effects and interactions discussed previously, a generalized method for the calculation of effects is discussed in this section. LAN analysts who consider more than three variables should find this method simple to calculate. The matrix shown below (Table 4.14) is derived from Table 4.13. Table 4.14 has an extra row with a divider, derived from the geometrical representation. For example, column W (either -1 or +1) is multiplied by column Y, followed by a summing of all the numbers, and finally divided by 4. It is divided by 4 since the average effect was simply an average of four contrasts, each contrast being the difference between a test result at both levels (-1 and 1).

Test	W	S	N	WS	WN	SN	WSN	Y
1	-	-	-	+	+	+	-	47.3
2	+	-	-	-	-	+	+	42.1
3	-	+	-	-	+	-	+	46.4
4	+	+	-	+	-	-	-	39.9
5	-	-	+	+	-	-	+	46.3
6	+	-	+	-	+	-	-	41.1
7	-	+	+	-	-	+	-	45.3
8	+	+	+	+	+	+	+	39.1
	(div) 4	(div) 4	(div) 4	(div) 4	(div) 4	(div) 4	(div) 4	(div) 8

Table 4.14 Calculation Matrix for 2^3 Design

Using Table 4.14, the mean is calculated from the last column:

$$\text{Mean} = \frac{47.3+42.1+46.4+39.9+46.3+41.1+45.3+39.1}{8} = 43.44$$

The rest of the effects are calculated from the respective columns using the sign and the divisor in that column. Note that the signs for the interaction effects are obtained by multiplying together the signs for the variables that form the interaction effect.

4.4.3 Development of an Empirical Model

The empirical model is given by the formula:

$$Y_i = \text{Mean} + \frac{W}{2}X_1 + \frac{S}{2}X_2 + \frac{N}{2}X_3 + \frac{WS}{2}X_1X_2 + \frac{WN}{2}X_1X_3 + \frac{SN}{2}X_2X_3 + \frac{WSN}{2}X_1X_2X_3$$

$$\begin{aligned}
\text{For } y_i &= y_1, & X_1 &= -1, & X_2 &= -1, & X_3 &= -1 \\
y_i &= y_2, & X_1 &= +1, & X_2 &= -1, & X_3 &= -1 \\
y_i &= y_3, & X_1 &= -1, & X_2 &= +1, & X_3 &= -1 \\
y_i &= y_4, & X_1 &= +1, & X_2 &= +1, & X_3 &= -1 \\
y_i &= y_5, & X_1 &= -1, & X_2 &= -1, & X_3 &= +1 \\
y_i &= y_6, & X_1 &= +1, & X_2 &= -1, & X_3 &= +1 \\
y_i &= y_7, & X_1 &= -1, & X_2 &= +1, & X_3 &= +1 \\
y_i &= y_8, & X_1 &= +1, & X_2 &= +1, & X_3 &= +1
\end{aligned}$$

To develop the empirical model, the estimation of standard errors for the effects must be calculated using replicated runs. The variation between the runs that are made at the same experimental conditions is a reflection of the total variability affecting runs made at different experimental conditions.

The pooled estimate of run variance is

$$s^2 = \frac{s_1^2 + s_2^2 + \ldots + s_8^2}{8}$$

The pooled estimate of run variance s^2 can be calculated from the last column in Table 4.7.

$$s^2_{pooled} = \frac{.23+.32+.27+.54+.23+.54+.46+.32}{8} = \frac{2.91}{8} = .36$$

$$s = \sqrt{0.36} = .60$$

Using this pooled estimate of run variance, we are able to estimate the variation associated with effects. Since each main effect and interaction is a statistic of the form $\bar{y}_+ - \bar{y}_-$, the total runs made in conducting a two-level factorial design was 80. As a result, the variance of each effect is given by

$$\text{Variance(effect)} = \text{Var}(\bar{y}_+ - \bar{y}_-)$$

$$s^2_{effect} = \frac{4}{N} s^2_{pooled} = .018$$

N = the total number of runs (80)

$s_{effect} = .13$

Using this information, we are able to identify those effects which are not significant from a statistical point of view, as seen in Table 4.15.

Effect	Effect ± standard error
Average	43.44 ± .065 (.13 / 2)
Main effects	
Workstation Speed, W	-5.77 ± .13
Server Speed, S	-1.53 ± .13
# of Workstation on LAN, N	-0.98 ± .13
Two-factors interactions	
WS	-0.57 ± .13
WN	-0.07 ± .13
SN	-0.03 ± .13
Three-factor interaction	
WSN	0.07 ± .13

Table 4.15 Effects and Standard Errors

From Table 4.15 we have an empirical model,

$$y = 43.44 + \frac{1}{2}(-5.77)W + \frac{1}{2}(-1.53)S + \frac{1}{2}(-0.98)N + \frac{1}{2}(-0.57)WS$$

The physical interpretations of this empirical model are

1. The workstation speed contributes the most to reduce the time taken to transfer files.
2. The interaction effect between workstation speed (W) and server speed (S) cannot be interpreted separately because of large WS interaction.
3. There was no interaction between workstation speed, server speed, and the number of workstations on the LAN.
4. The highest performance (shortest time to transfer files) will be achieved if using a faster workstation with a faster server while there are enough workstations on the LAN.
5. Higher performance is achieved when there are more workstations on the LAN. This is due to the less idle time for the file server.

From the experiment performed above, it is interesting to find out that server memory size has almost no effect on the performance of the LAN. It means that the extra memory planned to put into the current file server at the OACS is not necessary . Although the server speed contributed to the overall network performance improvement, it didn't improve as much as when the faster workstation is being used. With this experiment, we are able to verify that the network file server response faster when there are sufficient amount of workstations on the LAN, see section 4.4. However, in a real LAN environment setting (not everyone is doing file transfer), it did not made too much difference; the result should be more distinct than the experimental result according to the study conducted in section 4.4 (twice the more server throughput was achieved when the workstation increased from 2 to 8). As the mathematical model suggests, OACS should recommend and encourage each department to upgrade existing workstations to ones with faster clock speed to achieve higher performance LAN. Also, instead of adding extra memory to be used as cache memory, the way to increased the speed of the file server (processing power) should be considered, or possibly to upgrade to more efficient CPU such as Intel's Pentium CPU.

Chapter 5 New Network Technology

At the college of Behavioral and Social Sciences, the Geography department uses Lefrak's LAN to transfer large image files between workstations. As more users use the LAN to transfer large files and images, it is important to increase the current bandwidth of the network (10Mbps). Departments such as Geography will be an immediate beneficiary for such network improvements.

Higher speed networks will become indispensable as new applications and new uses of LANs become more common. Video, imaging, and fault-tolerance server all demand higher network speeds than are currently available. Most of the new high-speed technologies propose a network speed of 100 megabits per second (Mbps), which is considerably faster than today's 10Mbps Ethernet standard. There are already a number of applications that could benefit from such speedy networks; e.g. the transmission of images from one workstation to another or the mirroring of two file servers. Organizations planning multimedia and network teleconferencing projects will also need LAN connections in the 100Mbps range.

The leading architecture in the race toward high-speed LANs include four contenders that can work on unshielded twisted-pair wire (UTP): 100Base-VG; 100Base-X; Fiber Distributed Data Interface (FDDI) over copper; and asynchronous transfer mode (ATM). The 100Base-VG proposal is very different from the current Ethernet protocol. The 100Base-VG architecture uses a completely different media access control scheme from the standard Ethernet Demand Priority Access Method (DPAM). Under the proposed architecture, priorities can be assigned to specific message packets. The workstation signals the hub that it has a packet to send and requests either normal or high-priority service. If the network is idle, the hub allows nodes to transmit on a first-come, first-serve basis. Since the workstation has permission to send, it doesn't need to listen whether the medium (cable) is idle. 100Base-X is the latest enhancement of Ethernet, so it is commonly called fast Ethernet. A 100Base-X installation will look exactly like today's 10Base-T (twisted-pair cabling scheme discussed in the chapter 2). 100Base-X still relies on the same basic algorithm for regulating access to the network, carrier sense multiple access with collision detection (CSMA/CD). The 100Base-X network interface cards use faster signaling and space packets that are closer together than their 10Mbps ancestors, but the techniques are the same. Despite of its name, FDDI is not limited to systems that use fiber-optic cable. Under the latest definition of the American National Standard Institute (ANSI), the term FDDI can be used for schemes over fiber-optic, shielded twisted-pair, or UTP cable. FDDI is a networking scheme that delivers high reliability through redundancy and sophisticated data-handling protocols. Fiber-optic cable under FDDI enables signaling to travel as far as 2 kilometers between a hub and node, but the higher cost of fiber has limited its popularity. The FDDI protocols can run over copper cable if the distance is less than 100m and use a Level 5 UTP installation. ATM is a technology that has received a lot of attention. ATM is a cell-switching scheme that divides the stream of data into 48-byte cells. An additional 5-byte header provides routing information for a series of network cell switches. Because the cells are small, time-sensitive data such as voice and video can mix with other data without suffering much of a delay. If a few cells get out of sequence, a little buffering takes care of the tiny interruption.

All the technologies described above require a major upgrade of the existing network infrastructure to deliver more bandwidth. It means that existing network interface cards cannot be used and require new cables. However, switching hubs can be a much less disruptive and a more economical alternative. Switching hubs don't provide faster signaling, but they enable grater bandwidth and provide faster throughput. While the existing network interface card is used, each network interface card thinks it is the only one on the network. A faster processor in the hub routes packets across a backplane operating at hundreds of megabits per second. This is called a collapsed backbone architecture because the switching hubs act like a series of individual wiring hubs connected by a fast backbone link. In reality, most computers cannot take advantage of the full 10Mbps Ethernet channel, so these switching hubs offer product that allow you to share 10Mbps of bandwidth between anywhere from one to eight workstations. Generally, switching hubs can be installed in one of three different configurations: server front end, back end to group of hubs, high-speed wiring concentrator, and FDDI concentrator. As a server front-end, a switching hub is the only connection point for one or more servers. Each server gets the maximum bandwidth it can use while the workstations compete for more limited bandwidth. As a back end to a group of non-switching wiring hubs, a switching hub acts as a very fast yet economical backbone. As many as a dozen hubs can have 10Mbps bandwidth with no competition for the channel. As a high-speed wiring concentrator, a switching hub allows the LAN managers to give each workstation the necessary amount of bandwidth.

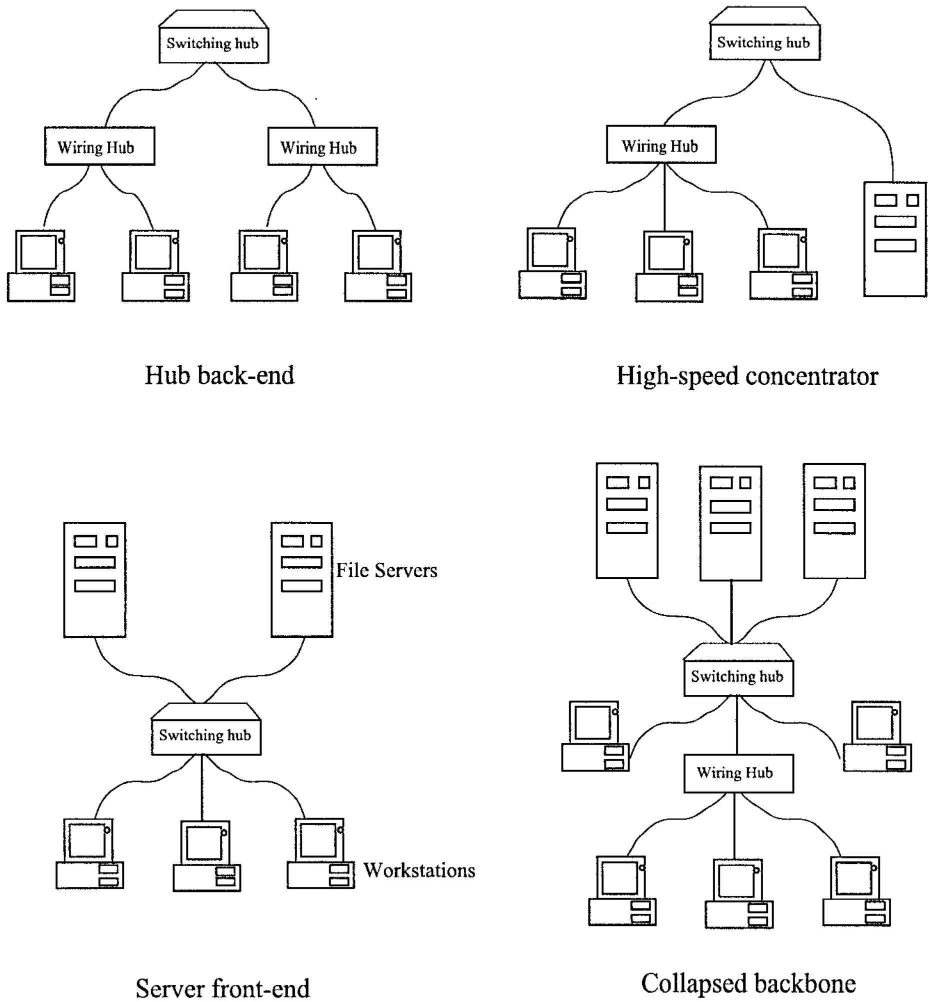


Figure 5.1 Switching Hubs

Chapter 6 Conclusion and Recommendation

The main objective of this thesis research is to apply the technique of performance analysis developed in the area of systems engineering to identify key systems parameters in local area network (LAN) systems for performance optimization. The research efforts are devoted to problem formulation of a LAN with a specific target to the current network system in the College of Behavioral Social Sciences. Under such an environment, workstation speed, server speed, memory size, and the number of computers are identified as four major system parameters in the LANs. The responding time for transmitting a 19 MB file between the server and a workstation is defined as the performance index to quantify the operational efficiency. A 2^4 full factorial design of experiments with five duplicates at each point is performed. Results from the eight runs are analyzed to establish an empirical model to characterize the main and interactional effects of the four system parameters on the communication efficiency. Significant findings are summarized as follows:

1) Server memory is not a key factor to improve the communication efficiency.
2) The other three systems parameters have significant effects with the workstation speed being most important.
3) The two-factor interaction between workstation speed and server speed indicates that the gain from increasing either workstation or server speed will be double-counted in improving the communication efficiency.

As a result of performance analysis, it is recommended that each department within the College of Behavioral and Social Sciences, an owner of a LAN, should upgrade existing workstations to ones with faster clock speed to achieve a higher performance.

Also, instead of adding extra memory to the file server, the clock speed of the file server should be increased, or the processor should be upgraded to more efficient CPU such as Intel's Pentium.

Recognizing the rapid advancement of computers, several recommendations are made for future improvement of LAN operational efficiency. Due to the increase of network traffic, it is important to look into ways to increase the bandwidth of the network. Devices such as switching hubs should be integrated into the LAN environment to enable greater bandwidth and provide faster throughput. Optimal hardware and software configurations alone will not ensure total LAN effectiveness. Consideration must be given to environmental constraints and to human needs, expectations, attitudes, and motivations. The Graphical User Interface (GUI) is a rapidly developing computing environment. GUI replaces text-based commands with graphical representations, so that users don't have to memorize commands. In the LAN environment, where there are many network printers, hard disks, and other devices, GUI can be a pleasant interface to network users. By replacing a command line with GUI, users do not have to memorize commands or optional parameters. Such an implementation makes the LAN easier to use and results in higher user satisfaction. Given that most organizations are faced with finite budgets, options available to a LAN manager will be limited. With this in mind, performance analysis is an important tool to prioritize the most cost effective way for improving LAN performance. These technical considerations must, of course, be balanced with human considerations such as the attractiveness of the GUI.

References

1. Kerzner, Harold, Project Management. New York: Van Nostrand Reinhold 1989.

2. Erol Gelenbe, Multiprocessor Performance. New York: John Wiley & Sons, 1989.

3. Ferrari, Domenico, Computer Systems Performance Evaluation. New Jersey: Prentice Hall 1978.

4. Pall, Gabriel A., Quality Process Management.. New Jersey: Prentice Hall 1987.

5. Booher, H. R. An Approach to System Integration. Van Nostrand Reinhold. 1990.

6. Carl Tropper, Local Computer Network Technologies, Academic Press. 1984.

7. Sinha, M. N., and W. O. Willborn, The Management of Quality Assurance. John Wiley & Son Inc. 1985.

8. Pyzdek, T., and R. W. Berger, Quality Engineering Handbook. ASQC Quality Press. 1992.

9. Blanchard, B. S., and W. J. Fabrycky, Systems Engineering and Analysis. Prentice Hall. 1990.

10. Dehnad, K., Quality Control, Robust Design, and Taguchi Method. Wadsworth & Brooks/Cole. 1989.

11. DeVor, R. E., Tsong-how Chang, and J. W. Sutherland, Statistical Quality Design and Control. Macmillan Publishing Company. 1992.

12. Sanders, M. S., and E. J. McCormick, Human Factors in Engineering and Design. McGraw-Hill, Inc. 1993.

13. Dimitri Bertsekas and Robert Gallager, Data Networks. New Jersey: Prentice-Hall, Inc. 1987.

14. Box, G. E. P, W. G. Hunter, and J. S. Hunter, Statistics for Experimenters. John Wiley & Sons, Inc. 1978.

15. Frank J. Derflet, Jr., Networking: Speed to Burn. PC Magazine, June 29 1993, pp. NE13-NE24.

16. Pramode K. Verma, Performance Estimation of Computer Communication

Networks. Computer Science Press. 1989.

17. J. G. Kalbfleisch, Probability and Statistical Inference. Springer-Verlag, 1985.

18. William J. Diamond, Practical Experiment Designs. Van Nostrand Reinhold. 1989.

19. D. Russell, The Principles of Computer Networking. Cambridge University Press, 1989.

20. Gianfranco Balbo and Giuseppe Serazi, Computer Performance Evaluation. Elsevier, 1991.

21. Stan Schatt, Understanding NetWare. Howard Wl Sams & Company, 1989.

22. Tom Brooks, The Local Area Network Reference Guide. Prentice-Hall, 1985.