ผลทางกายภาพของสภาวะแวดล้อมเครือข่ายต่อการประยุกต์เว็บ

นาย สรรสิริ เลิศกิจเจริญวงศ์

PHYSICAL EFFECTS OF THE NETWORK ENVIRONMENT ON WEB APPLICATIONS

Mr. Sansiri Lertkijcharoenwong

Thesis Title	PHYSICAL EFFECTS OF THE NETWORK ENVIRONMENT ON WEB APPLICATIONS
By	Mr. Sansiri Lertkijcharoenwong
Field of Study	Computational Science
Thesis Advisor	Assistant Professor Dr. Pattarasinee Bhattarakosol

Accepted by the Faculty of Science, Chulalongkorn University in Partial Fulfillment of the Requirements for the Master 's Degree

……………………………………….. Dean of Faculty of Science
(Associate Professor Dr. Wanchai Phothiphichitr)

THESIS COMMITTEE

………………………………………….. Chairman
(Associate Professor Dr. David Ruffolo)

………………………………….……. Thesis Advisor
(Assistant Professor Dr. Pattarasinee Bhattarakosol)

………………………………………….. Member
(Assistant Professor Dr. Pornpote Piumsomboon)

สรรสิริ เลิศกิจเจริญวงศ์ : ผลทางกายภาพของสภาวะแวดล้อมเครือข่ายต่อการประยุกต์เว็บ.
(PHYSICAL EFFECTS OF THE NETWORK ENVIRONMENT ON WEB APPLICATIONS)
อ. ที่ปรึกษา : ผศ. ดร. ภัทรสินี ภัทรโกศล, 42 หน้า. ISBN 974-03-0582-2.

       งานวิจัยนี้เป็นการทำการพิจารณาเวลาที่ใช้รอรับการประยุกต์เว็บ ซึ่งเกิดจากการขอจากเครื่อง
ลูกข่ายผ่านระบบเครือข่าย   ปัจจัยทางกายภาพในสภาวะแวดล้อมเครือข่ายที่จะพิจารณาเป็นตัวแปร
อิสระ คือ จำนวนของฮับ และสวิตช์ และ ระยะทางจากเครื่องลูกข่ายไปยังเครื่องแม่ข่ายโดยพิจารณา
ความยาวของสายส่งข้อมูล อีกทั้งเวลารอคอยนี้ยังได้รับผลกระทบจากขนาดของหน้าเว็บที่หลากหลาย
ด้วย ผู้วิจัยจะพิจารณาขนาดของหน้าเว็บเป็นตัวแปรอิสระอีกตัวแปรหนึ่ง การทดลองทั้งหมดได้กระทำ
ในเครือข่ายแบบอินทราเน็ตและตัวแปรอิสระทุกตัวได้ถูกกำหนดค่าไว้แล้ว   โดยที่ผู้วิจัยทำการวัดเวลา
รอคอยและบันทึกค่าหลังจากได้รับการประยุกต์เว็บอย่างสมบูรณ์ผ่านทางโปรแกรมท่องอินเทอร์เน็ต
ค่าของตัวแปรอิสระและค่าเวลาที่วัดได้เหล่านี้จะนำไปใช้เพื่อสร้างสมการถดถอยทำนายเวลาในการรอ
รับการประยุกต์เว็บที่เกิดขึ้น โดยสมการถดถอยที่ได้อยู่ในรูปของสมการถดถอยพหุคูณ ซึ่งแสดงถึงผล
ทางกายภาพของปัจจัยต่างๆในสภาวะเครือข่ายที่กระทบต่อเวลารอคอย ความถูกต้องของการทำนาย
ระยะเวลารอคอยจากกลุ่มข้อมูลที่ใช้ทดสอบมีความแม่นยำถึง 98% ซึ่งผู้วิจัยคาดว่าความสัมพันธ์ของ
ตัวแปรเหล่านั้นจะเป็นประโยชน์ในการส่งเสริมประสิทธิภาพของเครือข่ายให้ดีขึ้น หรือเป็นประโยชน์ใน
การสนับสนุนการขยายตัวของเครือข่าย อีกทั้งเป็นแนวทางในการศึกษาปัจจัยอื่นๆ ที่ส่งผลกระทบต่อ
เครือข่ายที่มีอยู่ในปัจจุบัน

ภาควิชา........คณิตศาสตร์...................... ลายมือชื่อนิสิต...
สาขาวิชา......วิทยาการคณนา.................. ลายมือชื่ออาจารย์ที่ปรึกษา.....................................
ปีการศึกษา....2544…………….............…

SANSIRI LERTKIJCHAROENWONG : PHYSICAL EFFECTS OF THE NETWORK ENVIRONMENT ON WEB APPLICATIONS. THESIS ADVISOR : ASSIST. PROF. DR. PATTARASINEE BHATTARAKOSOL, 42 pp. ISBN 974-03-0582-2

This research focuses on web applications retrieval time caused by a request from a client passed through the Intranet environment. In the Intranet environment, there are some physical factors that are considered as the independent variables such as the number of hubs, the number of switches, and the distance (total cable length) from a client to the server. Furthermore, this delay time also received major effects from various sizes of web pages. This factor is counted as another independent variable in this research. The experiment is performed in an Intranet for fixed values of the each of independent variables. A delay time is measured and recorded after loading web applications via an Internet browser. All observation values are applied for creating a model for predicting a proper delay time. The resulting model is in the form of a multiple regression model, which helps us to understand the effects from various physical factors over a network. The accuracy to predict a delay time using a testing data set is approximatly 98%. The regression model, which represents the relationship among mentioned variables, can be used to improve the network's performance or support the adjustment in a network's expansion. In addition, it can lead the way for studying effects from more physical factors in recent networks.

Department	Mathematics	Student's signature	
Field of study	Computational science	Advisor's signature	
Academic years	2001		

Acknowledgement

Contents

List of Tables

Page

List of Figures

Page

CHAPTER I

INTRODUCTION

1.1 Introduction

The issue in distributed resource locations problem had occurred in most of the organizations. If information are stored in various locations apart from users, this leads to the differences of the delay time when users retrieve data over the network. Currently the most popular information retrieval system is based on web application over the Internet services.

There is a research that considered the interaction of HTTP with several transport protocols [14]. This important work measured the delay time of HTML retrieved files on several transport protocols across a variety of networks. However, the HTML files in the experiment did not mention in details such as text, pictures, or videos that consisted. Nevertheless, this work helps us to take consideration on physical factors that affect the delay time of the HTML file retrieval.

This research considers the physical effects of the network environment on web applications, which consisted of text, pictures, and videos. The physical factors in this research are numbers of hubs and switches, distances between a client to the server, and sizes of web pages. We will measure the delay time relies on these physical factors in order to build the predictable delay time model.

1.2 Objective

The objective of this research is to analyze the relationship between physical factors as mentioned previously, and the delay time for retrieving a web page. Furthermore, all experiments are performed in the closed Intranet system. The result of this research is a general regression model for predicting the delay time under the certain factors.

1.3 Scope

- The web pages are not classified by the size of text or graphics that used to build it.

- Only the HTTP over TCP is used in our closed network system.

- There is only one client that is allowed to retrieve a web page at a time from a server.

- There is no congestion in our system.

- The network consists of hubs or switches, and no router.

- The delay time is measured in milliseconds.

- The delay time prediction model will be tested using different data sets.

1.4 Research Methodology

1. Study in networking areas, such as
   - Network topology
   - Network protocol
   - Cable installation,
2. Study Multiple Regression and Correlation,
3. Survey web sites and create web pages that are fixed sizes,
4. Write Java Script program used to measure the delay time,
5. Set up an Intranet system with tested plans,
6. Implement all web pages with Java scripts embedded,
7. Collect data from the experiment based on the tested plans,
8. Analyze the obtained delay time to build the regression model,
9. Collect data for testing model,
10. Test the result model and make conclusions.

1.5 Contribution

The obtained regression model represents the network physical effects to the web application retrieval delay time. These effects occur by physical network environment factors as mention earlier. This model guides the network designer to implement a suitable network topology and architecture to serve the real requirements of an organization. Additionally, the network designer can use this model to gain the improvement of their Intranet as users requested.

CHAPTER II

LITERATURE REVIEW

Network has become one of the most popular communicative factors in many kinds of organizations, firms and governments. Network is proposed to reduce cost of using resources between departments, moreover, communicate users. Additionally, it provides many services to users such as e-mail, e-commerce, e-news, etc.

In 1990, there were several studies about local area network (LAN) and its performance factors that affect the data receiving performance. A number of parameters can affect the delay of messages on the network, like the number of messages, the number of nodes on the network, the communications protocol, the buffer size, CPU power and available memory space [3][4]. Although these experiment considered many factors, but all factors are related to the techniques of transferring data over a network. Thus, there are many leftover factors that have not been considered but it can effect the delay time, such as the transferring distance and number of active devices over the network. Therefore, in our experiment, we will collect the delay time that caused different file sizes, various distances, various types and number of active devices. We expected that the result from our experiment will indicate the affect from these factors as same as the affect that is presented in [3][4].

In [5], there was more concerned with the performance of transport protocols which actually depends on protocol layering, processing overhead, operating system support and architectural constraints. Its experiment compared the performance derived from different transport protocols, i.e. TCP and UDP, SPP and IDP, over a 10 Mbps Ethernet LAN. In addition, it also described the method to measure the performance. However, [5] does not consider the performance of the HTTP that is a popular protocol for retrieving and linking data sources over a network.

In last five years of 20th century, the measurement of LAN had shown at a wide range of scales. More complicated researches were done in many ways due to the relationship between network resources (bandwidth and buffer capacity), network topology, distribution of file sizes and network model (or network relationship type: peer-to-peer and client/server). From [6], this research showed the relationship between file sizes, protocols, and the network traffic. It led us to consider on the recent distribution of file sizes which too large for the past.

The attempt to predict and prevent outages caused by overloading system in the Intranet was in [7]. This work addressed the retrieval time of files over FTP. The experiments showed the relationship between the delay time, number of users, and the file size in a controlled environment. In addition, the result model was applied to the WAN (Wide Area Network) environment as the basis model. The result from [7] was a data transfer model of FTP that was a linear regression model under non-congestion environment. According to the fact that FTP runs over the TCP as same as HTTP, it is possible that the delay time model for transferring a web page may be similar to the data transfer model of FTP protocol as presented in [7]. Furthermore, the prime factor should be the sizes of the transferred file as same as presented in [7].

Since it is very costly to setup a realistic huge LAN with several groups of nodes attached with different protocols for performing research. So a solution to observe the behaviour of a network is to monitor through the simulation programs. For example, a Network Simulator (NS) is an event driven simulation program, which combined one or more languages such as C/C++, TCL/TK, etc. Many features of NS make simulation results are the same as running the test over a real network. The simulation situation that can be generated or defined using the NS are, for instance, queue management, packet forwarding and scheduling, delay, links, web applications.

To use the NS, many parameters and conditions must be defined. Firstly, the network topology must be defined to the simulator program, included with the attached agents (TCP, UDP, etc) that are applied to communicate among nodes. Secondly, file sizes, packet sizes and transfer rate between sources and sinks should be assigned. The researcher must identify the time measurement interval before running the program. After setting all these parameters, the simulator program is ready to perform its tasks. During the testing period, all existing data are recorded in the logged file. The researcher can retrieve all stored data in the logged file to analyse using statistical methods.

Although the NS is a power tools for network simulation, there are some conditions that related to the architecture of computers over the network and the length of connection cannot be employed as parameters in the program. So, this program cannot examine experiments that concern these factors.

Prior works in networking field have proposed and improved solutions to each of problem. Furthermore, there are specific and useful tools that been created to make measurement and experiment go easier. Now, the represent of communicative world has gone by far and known as World Wide Web, the most popular Internet services. The World Wide Web allows user to view documents that contain text and graphics,

and to follow hypermedia links from one document to another. It 's been grown to become the largest source of traffic on the global Internet since 1994. The popularity of the web has resulted in a corresponding popularity for HTTP, the standard Hypertext Transfer Protocols and HTTP is layered over TCP.

In 1997, John Heidemann, et al. [8] presented analytic models for HTTP traffic over several transport protocols. They considered the interaction of HTTP with several transport protocols, including TCP, Transaction TCP, an UDP-based request-response protocol, and HTTP with persistent TCP connections. Each of these models used to evaluate network overhead carrying HTTP traffic across a variety of the transient effects of TCP slow-start. In addition, Comparing models to network packet traces measured with two protocols (HTTP, and persistent HTTP) over local and wide-area network validated these analytic models. Therefore, there are some interesting points on their work that became more considerate in this research.

In [9], it presented the relative differentiated services approach that grouped network traffic in a small number of service classes. These service classes are ordered based on their packet quality, delay time, or transfer rate. This work proposed the proportional differentiation model, which is the one of relative differentiated services model, aimed to provide the better Internet services for variety of usage.

In [10], it focused on characterizing information about World Wide Web resources and server responses that is relevant to web caching. The approach was to study a set of URLs (Universal Resource Location) at a variety of sites and gather statistics about the rate and nature of changes compared with the resource type. In addition, they gathered response header information reported by the servers with each retrieved resource. Results from their work indicate that there is potential to reuse more cached resources than is currently being realized due to inaccurate and nonexistent cache directives.

The high-performance web servers have been essential to meet the growing demands of the Internet and large-scale Intranets. In [11], they studied key factors affecting web server performance. This work presented empirical analysis illustrating how dynamic and static adaptation can enhance web server performance.

These last three works proposed the better Internet services and better services management to support the several of Internet usage. However, they focused on the non-physical effects on the wide scale of Internet. So we try to study on the physical effects to approach the same goal and apply the advantage to an Intranet.

If we look into the recent networks, they will be contained with a number of small to large Intranets (10 up to over thousand nodes).

Therefore in order to increase the performance of the available Intranet under the existing equipments, files, and applications, reorganizing of the network system might be necessary. This research is focussed on many physical factors that might affect to this performance so that the delay of the transmission time is minimized.

CHAPTER III

METHODOLOGY

3.1 Selected topology and UTP cable length

Generally Intranets install the star or tree topology, and all nodes in these Intranets are connected by a number of network equipment. If we trace the paths that use to establish connections between nodes (as clients and servers), Figure 3.1 shows connection path between client and server. In order to set the simulation system to be similar to the real connection system, this experiment will examine the effect from several length of UTP cables which are 10, 20, 50, 80, 90 and 100 meters long. The UTP cable used in research is DLink 's UTP Category 5 (10BASET 10Mbps 100 m between node and network hardware).
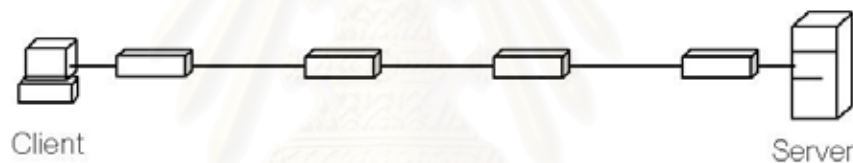
Figure 3.1 Connection line between client and server over Intranet

3.2 HTTP workloads distribution

From [8], it defined HTTP workloads in several potentials as follows:
- Small page (single 5 KB web page)
- Medium page (single 25 KB web page)
- Large page (single 100 KB web page)
- Small cluster (single 6651 B page with embedded 3883 B and 1866 B images)
- Medium cluster (single 3220 B page with three embedded images, sizes 57613 B, 2334 B, and 14910 B)
- Large cluster (single 100 KB page with 10 embedded 25 KB images)

Therefore, the web pages that are applied for this experiment for observing the HTTP workloads distribution are as listed.

1. 100 KB (103,290 bytes)
2. 500 KB (499,958 bytes)
3. 1 MB (997,858 bytes)
4. 5 MB (5,008,934 bytes)
5. 10 MB (10,005,681 bytes)

3.3 Making HTML files according to selected workloads

After the HTTP workloads were determined, we started to capture pictures, animations, and videos from the original sources over Internet. All these collections will be combined together to create the small web pages, medium web pages, and large web pages suitable for our experiment. In addition, we built the menu page that show links to those five files and set it as home page. This menu page made the measurement easier as shown in the next section.

3.4 Building a time-capturing program

Considering when a user wants to connect to a web site using browser to retrieve a web page, written by HTML. The user may use various browser applications that may or may not pay for the license such as Microsoft Internet Explorer, Netscape Navigator, or Opera. The connection starts when the user types the URL (Universal Resource Location) and presses enter. The HTTP protocol flows through the connection channel with the encapsulated requested address. Then, the communication with the destination web server starts. At the server, the requested will be interpreted to find the required web page and send back to the original request location via the same channel.

The time stamp in this research counts from the beginning of the request until the request has been served completely, the required web page presents to the user. This time stamp is called the delay time. From [3][4][6][7] have shown that the delay time depends on many factors such as the connection equipments – hub, switch, or router, the connection distance, and the transferred file size. In this research we will consider the delay time that is affected by the variation of equipment (hub and switch only), distances, and file sizes. Figure 3.2 shows the diagram of time measurement process. Figure 3.3 shows the time measurement process.
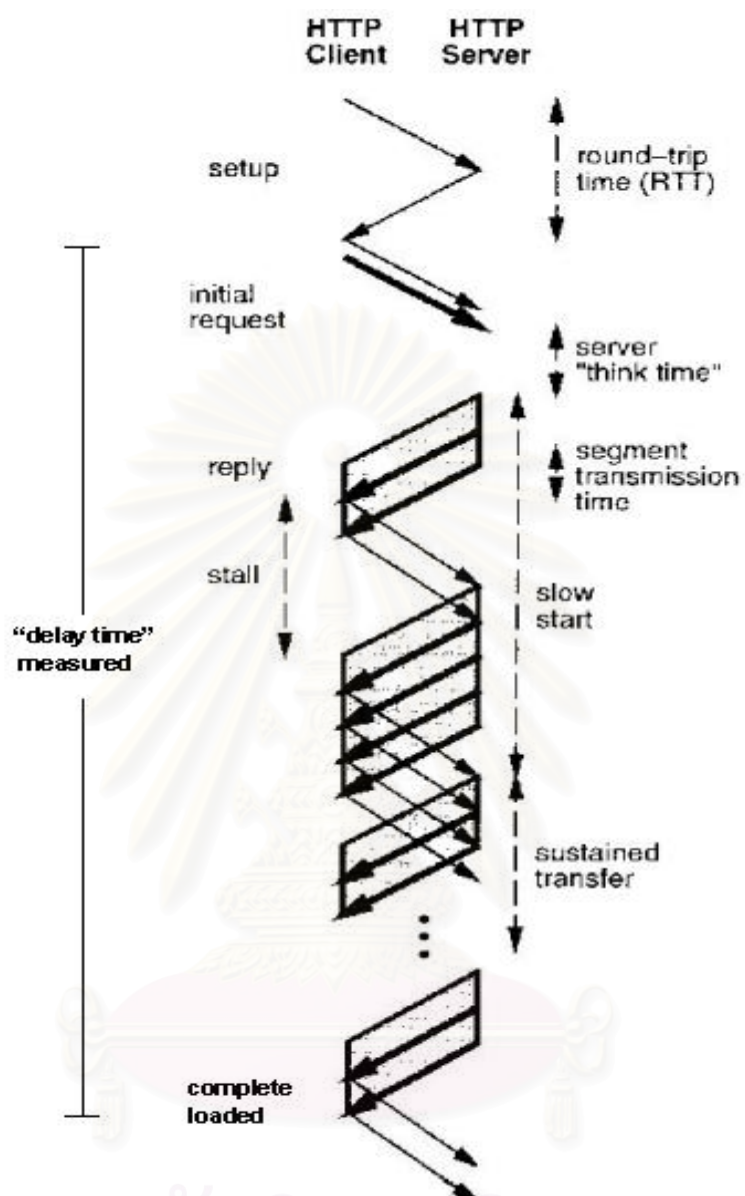
Figure 3.2 Time measurement process indicated by HTTP timing diagram. Bold arrows show data transfers, while thin arrows show acknowledge packets [8].
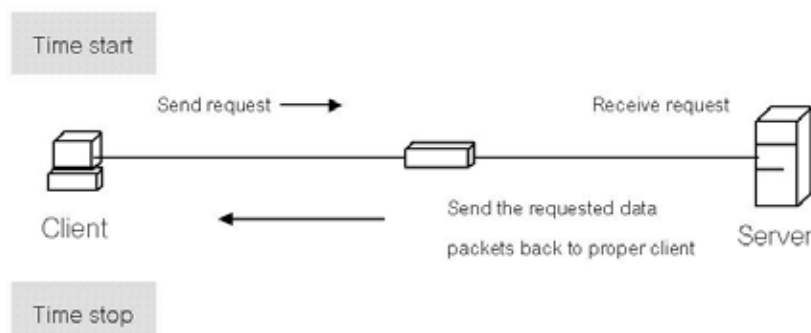
Figure 3.3 Time measurement process

To measure the delay time, Java Scripts were implemented. These scripts can be separated into 2 main parts. The first part is responsible for the user interface presentation for selecting the required file. Once a file name is called, the function "writecookie()" starts to create a temporary variable type "cookie," named as "start." This variable stores the calling time only. The time will be recorded in milliseconds. Figure 3.4 shows the first part of the time-capturing program, Start-time.

```
Program Start-time
<html>
<head>
        <title>Untitled Document</title>
        <meta http-equiv="Content-Type" content="text/html; charset="">
</head>
        <Script language="JavaScript">
        <!--
        function writecookie() {
                var time = new Date()
                document.cookie = "start="+escape(time.getTime())
                }   //-->
        </script>
        <body bgcolor="#FFFFFF">
        <p><font size="+3">Menu </font></p>
        <p><a   href="text100.html"   onClick="writecookie()">1.)Text   100%
        and Picture 0%</a></p>
        <p><a href="text75pic25.html" onClick="writecookie()">2.)Text 75%
        and Picture 25%</a></p>
        <p><a href="text50pic50.html" onClick="writecookie()">3.)Text 50%
        and Picture50%</a></p>
        <p><a href="text25pic75.html" onClick="writecookie()">4.)Text 25%
        and Picture75%</a></p>
        <p><a   href="pic100.html"   onClick="writecookie()">5.)Text  0%  and
        Picture100%</a></p>
        </body>
</html>
```

Figure 3.4 Program Start-time
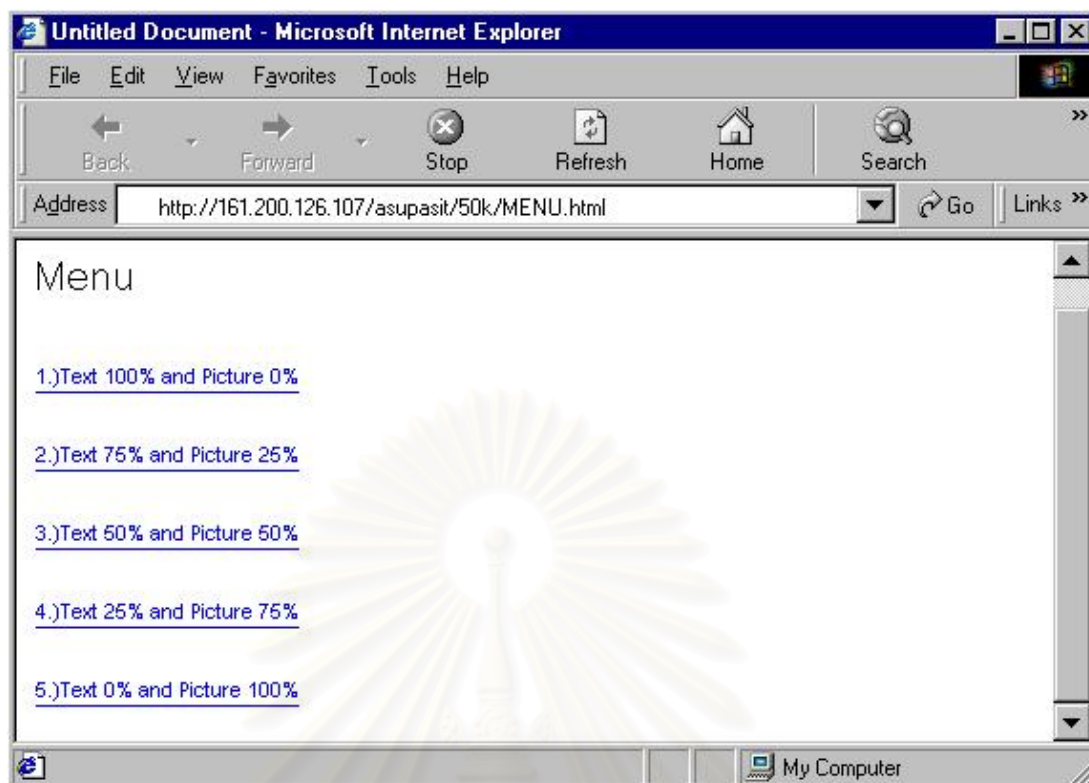
Figure 3.5 Menu web page

The second part is a set of scripts that embedded in any web pages that can be called. In this script, the function name "getcookie()" will be called to retrieve the time stamp value from the variable "start". Then the "writemsg()" function computes the delay time by subtracting the current time obtained from the system with the time recorded in the "start" variable, the measurement unit is in millisecond. The obtained delay time will be presented at the status bar of the presented window. Figure 3.6 presents the scripts of the second part, Stop-time.

In order to eliminate the covariate factor from the buffer usage of the system after browsing a web page. Temporary web pages that kept by Internet browser for quick viewing later must be deleted, and the computer must be started before loading a new page so that the loading buffer will be cleared off.

Program Stop-time

```html
<html>
<head>
        <title>text25%pic75%</title>
        <meta http-equiv="Content-Type" content="text/html; charset="">
<Script language="JavaScript">
<!--
        function getcookie(name) {
        var cookieName = name+"="
        if (document.cookie.length > 0){
                offset = document.cookie.indexOf(cookieName);
                If (offset != -1){
                        offset += cookieName.length;
                 end = document.cookie.indexOf(";",offset);
                        if (end == -1){
                 end = document.cookie.length;
                        }
                tempval = document.cookie.substring(offset, end);
                return unescape(tempval);
                }
        }
        }
function writemsg() {
        var time = new Date()
         var stp = time.getTime()
        var str = getcookie("start")
        window.status = stp-str + " msec";
}  //-->
</script>
</head>
<body onLoad="writemsg()">
<p><font size="+1">Performances of Distributed Systems Based on Ethernet
and Personal
        Computers</font> </p>
        {TEXT
</p>
<p align="left"><img src="40K.jpg" width="400" height="300"></p>
<p align="left"><img src="50K.jpg" width="600" height="342"></p>
<p align="left"><img src="60K.jpg" width="800" height="600"></p>
</body>
</html>
```

Figure 3.6 Program Stop-time

3.5 Setting up and controlling the network environment

In this topic, we will define the experimental environment in order to avoid the interrupting factors that can influence the measurement data to be unreliable and mistake. There are many criteria to be taken in the consideration as follows.

3.5.1 Client/Server specifications and details

The client is Pentium II, clock speed 400 MHz, RAM 64 MB with Windows 98 operating system and Microsoft Internet Explorer browser. In the other hands, the server will be Pentium III, clock speed 500 MHz, RAM128 MB implemented with the Windows NT 4.0 operating system and Microsoft Internet Information Server 3.0 as the web server of the system.

3.5.2 Hub and switch specifications

We use four hubs and one switch to perform the various connections, this will be described later. The specification of hubs and switch are as follows.

Hubs: There are 2 brands of hubs, Complex stackable hub, and SMC multi-segment stackable hubs. Each port in these hubs supplies 10 Mbps shared bandwidth to its attached device. Each hub also has a daisy-chain port that can be cascaded hubs quickly as the network grows.

Switch: A 3Com SuperStack II Switch 610 is used in the experiment. This switch features 24 switched 10 Mbps Ethernet ports and two Fast Ethernet ports to boost bandwidth for server and power users. It also features Store-and-forward, elastic port buffering, and intelligent flow management. The 3com switch can be configured from anywhere on network by using web-based management.

3.5.3 Cabling installation

After Client/Server, hub, and switch have been chosen, the next procedure is to install the cable. The DLink 's UTP Category 5 cable is selected to wire for our network.

As mentioned previously, the UTP cables are divided into several lengths: 10, 20, 40, 50, 60, 70, 80, 90, and maximum 100 meters. So, we can make various connection styles such as vary the distance of the connection length, or vary the number of hubs and switch installing over the network.

According to the type of cable, RJ-45 connectors are used as the links among active devices, hub or switch, and NIC. There are two methods of installing the connector: pin-to-pin, and crossover methods.

Pin-to-pin is the method that connects a node to an active device, or, between active devices with standard T568A [2] as show in Figure 3.7.
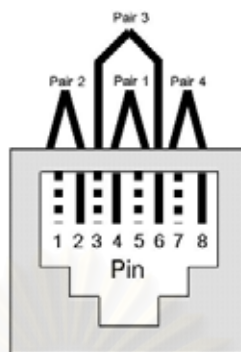
Figure 3.7 RJ-45 connector with standard T568A

The second method is the crossover method. This method is used to connect between two end-nodes (computer to computer) directly. Table 3.1 and Table 3.2 shows the pin-to-pin wiring and the crossover wiring of the pairs respectively.

Table 3.1 Pin-to-pin cable wiring

Pin Number	Wire Base Color	Wire Stripe Color
1	White	Orange
2	Orange	White
3	White	Green
4	Blue	White
5	White	Blue
6	Green	White
7	White	Brown
8	Brown	White

Table 3.2 Crossover cable wiring

Pin	Wire Base Color	Wire Stripe Color	Pin	Wire Base Color	Wire Stripe Color
1	White	Orange	1	White	Green
2	Orange	White	2	Green	White
3	White	Green	3	White	Orange
6	Green	White	6	Orange	White

3.5.4 Connection testing models

We separate total distance between two nodes, client/server, into 3 distances: 100, 200, and 300. Furthermore, there are variations of the number of active devices in each of distance. For instance, in the 100 and 200 meters, we put 1 to 4 hubs in the connection line, so there are 8 connection models that occur and have to be measured.

By the maximum limitation of cable length, and the number of active devices, the possibly connection models are showed in the Figure 3.8. Table 3.3 shows the experimental plan corespondent with connection models.
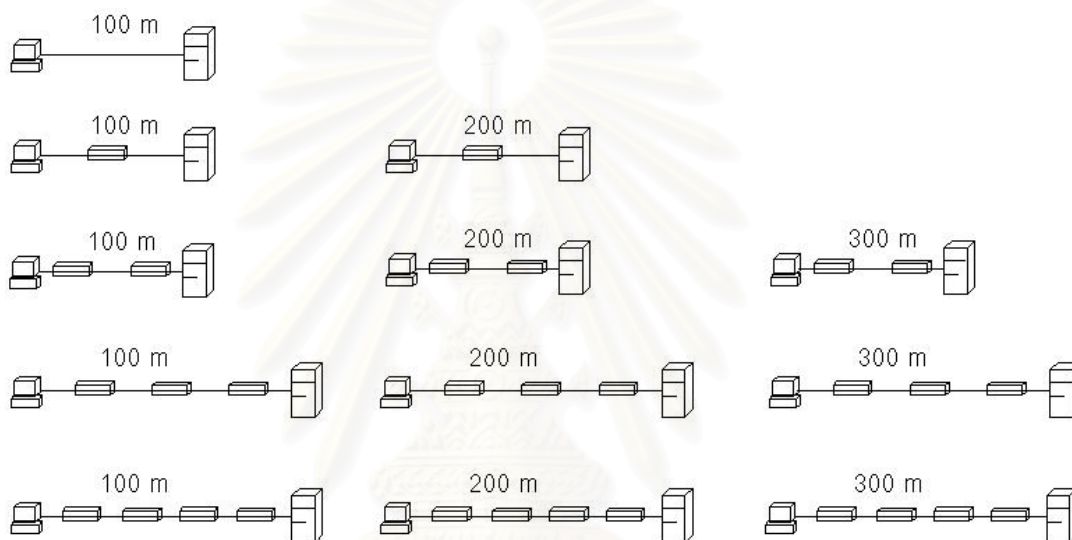
Figure 3.8 Connection models

Table 3.3 Experimental plan in connection details of cable length

Distance hubs	100 m	200 m	300 m
0	100		
1	50-50	100-100	
2	10-80-10	50-100-50	100-100-100
3	10-20-50-20	50-50-80-20	50-100-100-50
4	10-10-50-20-10	10-50-80-50-10	50-50-100-80-20

In addition, for each of models (except the crossover 100 meters without any active device) will be added a switch by replacing the hub

which connected directly to the server. Hence, there are 23 connection models need to set up in our experiment.

3.6 Collecting data

According to simulate the situation of Intranet of an organization as real and also prevent error in our experiment, all the web pages will be installed in the server located nearby clients. Additionally, there is no congestion occurs over the network. The experimental model that is applied in this research is repeated measurement model. The web page is retrieved from "http:\\161.200.126.160\sample\MENU.html". Each time the web is retrieved, the Stop-time runs and the delay time is presented. The total data measured in this experiment is based on 11 hubs only models plus 11 switch-replaced models and one cable only model. Each model is performed with 5 HTTP workloads and 10 times repeated measurement. So the total records in our experiment are $(11+11+1)\times 5\times 10$ which is equal to 1150 records. Table 3.4 shows the example of collected data.

Table 3.4 Example of collected data

FILESIZE(bytes)	SWITCH	HUB	DISTANCE(m)	TIME (ms)
997858	0	3	300	2530
997858	0	3	300	2740
997858	0	4	300	2860
997858	0	4	300	2700
997858	1	0	100	2360
997858	1	0	100	2420
997858	1	1	100	2410
997858	1	1	100	2420

3.7 Data Analysis

3.7.1 Incomplete Block Design [1]

The block in the experimental design provides control of the environment to reduce the experimental error. The experimental units are blocked into groups of similar units on the basis of a factor or factors that are expected or known to have some relationship with the response variable or the measurement that is hypothesized to respond differentially to the several treatments. However, there can be some experiments require a reduction in blocks or groups. Thus, this reductions cause the experimental plan be smaller than a complete replication of all treatments due to limitations of factors, or experimental environments. Complete block design can reduce the estimate of experimental error variances, but sometimes the limitation is important and reduction is insufficient. In our experiment, it requires the incomplete block experiment design.

According to the experimental plan in this research, it is impossible to have the completed block design. This is because the limitation of the cable length cannot exceed 100 meters without any active devices. Therefore, the experiment in this research is in the incomplete Block Design.

3.7.2 General Factorial Experimental Design [1]

In this research we defined many variable as factors that control the measurement of delay times. These factors are influencer of the delay time value since one parameter has been changed, for example the web size, the delay time will be increased or decreased according to the size of the web. To test the influencing from these factors, the analysis of variance (ANOVA) based on the Factorial Experimental Design is calculated under the significant level 0.05, or 95% confident interval. This analysis indicates factors that are the real influencers of the delay time, both for main effects and interaction among factors.

The treatment design used to address the research hypothesis can be placed in any compatible experiment design. In our work, all treatments are incomplete block design. Analysis of variance for the factorial experiments in incomplete block design presents the linearly relation between factors and dependent variable in the 95% significant. Moreover, it shows effects cause by any combination of two factors. For example, the linear model for a two-factor factorial, factor A with a levels and factor B with b levels which combination between them also affects the dependent variable in a randomized incomplete block design is,

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \phi_m + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + \ldots + (\gamma\phi)_{km} + e_{ijk}$$

$$i = 1,2\ldots a, \ j = 1,2,\ldots b, k = 1,2,\ldots c, \ m = 1,2,\ldots d$$

Where

μ is the general mean of population

α_i is the treatment effect of size of the web page (100 KB, 500 KB, 1 MB, 5 MB ,and 10 MB)

β_j is the treatment effect of number of hubs (0, 1,…,4 hubs)

γ_k is the treatment effect of number of switch (0, 1 switch)

ϕ_k is the treatment effect of distances (100, 200, and 300 m)

e_{ijk} is the experimental error with normal distribution.

When the combination of factors occurs, the term that composes of multiple of these factors will exist in the model.

3.7.3 Correlation [13]

The Bivariate Correlations procedure computes Pearson's correlation coefficient with their significance levels.

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

where

χ^2 call Pearson Chi – Square; O_{ij} are data frequency in cell(i,j); E_{ij} are data frequence experted in cell(i,j) if variable i and j are independent

$$E_{ij} = \frac{\left(r_i\right)\left(c_j\right)}{n}$$

i = 1,2,…,r ; j = 1,2,…c; r are frequency in row; c are frequency in column; n are number of data $= \sum r_i = \sum c_j$

Correlation measures how variables are related. Before calculating a correlation coefficient, the outliers data must have been removed. We use the scatter plot applies to all data and seek for the data that isolated. the Pearson's correlation coefficient is a measure of linear association. Two variables can be perfectly related, but if the relationship is not linear, Pearson's correlation coefficient is not an appropriate statistic for measuring their association.

For quantitative correlation coefficients with normally distributed variables, the Pearson correlation coefficient is suitable. Correlation coefficients range in value from -1 (a perfect negative relationship) and +1 (a perfect positive relationship). A value of 0 indicates no linear relationship. When interpreting results, any cause-and-effect conclusions due to a significant correlation are not used. Figure 3.9 shows the scatter plots of two variables, X and Y, related to Pearson correlation coefficient values.
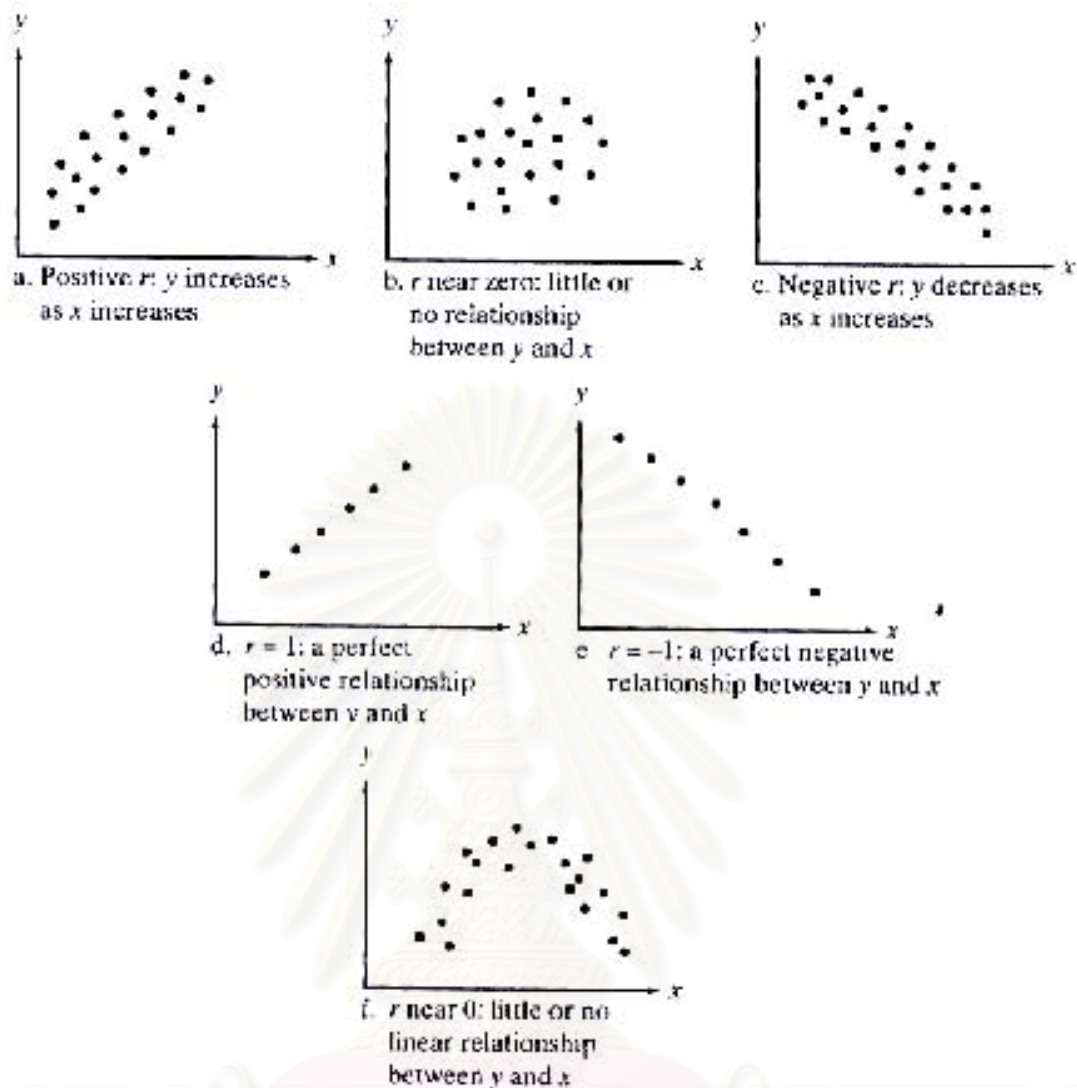
Figure 3.9 Scatter plots of two variables, X and Y, related to Pearson

correlation coefficient values [13]

3.7.4 Curve estimate [12]

The method choose data graphically to determine how the independent and dependent variables are related (linearly, quadratic, etc.). Residuals of a good model should be randomly distributed and normal. If a linear model is used, the following assumptions should be met. For each value of the independent variable, the distribution of the dependent variable must be normal. The variance of the distribution of the dependent variable should be constant for all values of the independent variable. The relationship between the dependent variable and the independent variable should be linear, and all observations should be independent.

If the scatter plot of an independent variable and dependent variable does not in the linear form, the data transformation of the independent variable needs to be computed. For example, if the scatter

plot of an independent and dependent variables is in the quadratic form, then we will compute the new variable that is the second degree of the independent variable and this variable will be applied in the regression model as if it is another independent variable. Then, the simple regression model with one independent variable will be the multiple regression model with two independent variables: the original independent variable, and the transformed variable as mentioned previously.

3.7.5 Multiple regression [13]

The multiple regression model is of the form

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \varepsilon$$

Where

y is the response variable

$\beta_0, \beta_1, ..., \beta_k$ are parameters with unknown values

$x_0, x_1, ..., x_k$ are information-contributing variables that are measured without error

ε is a random error component.

Since $\beta_0, \beta_1, ..., \beta_k$ and $x_0, x_1, ..., x_k$ are nonrandom, for each value of the independent variable, the distribution of the dependent variable is assumed to be normal. The variance of the distribution of the dependent variable should be constant for all values of the independent variable. The relationship between the dependent variable and each independent variable should be linear, and all observations should be independent.

Let's R-squared be the coefficient of determination. This value indicates correctness of the regression linear model derived from the raw data. For example, if R-squared = 0.987 means the correctness of prediction the value of dependent variable when all independents are defined will be 98.7%.

$$R^2 = \frac{SSR}{SST}$$

$$SST = SSR + SSE$$

Where

SSR is Regression Sum Square

SST is Total Sum Square or Total Sum Square

SSE is Residual Sum Square or Error Sum Square (SSE)

In the case that there are some independent variables related to each other appeared in the regression model, the adjusted R-Squared will be used to predict the correctness of the model. The adjusted R-Squared is obtained by adjusting the degrees of freedom of the model analysis.

$$Adjusted \ R^2 = 1 - \frac{\left[\dfrac{SSE}{DF_E}\right]}{\left[SST \times DF_T\right]} = 1 + \left[\dfrac{DF_T}{DF_E}\right] \times \left(R^2 - 1\right)$$

Where

DF_T is Degree of Freedom of Total Sum Square;
DF_E is Degree of Freedom of Error Sum Square

3.7.7 Nonlinear regression

Nonlinear regression is a method of finding a nonlinear model of the relationship between the dependent variable and a set of independent variables. However, the researcher must have an idea for the nonlinear form before transforming the original data to fit the model. Unlike traditional linear regression, which is restricted to estimating linear models, nonlinear regression can estimate models with arbitrary relationships between independent and dependent variables. This is accomplished using iterative estimation algorithms. Note that this procedure is not necessary for simple polynomial models of the form $Y = A + BX^2$. By defining $W = X^2$, we get a simple linear model, $Y = A + BW$, which can be estimated using traditional methods such as the Linear Regression procedure.

In term of nonlinear regression hypothesis, results are valid only if a function that accurately describes the relationship between dependent and independent variables has been specified. Additionally, the choice of good starting values is very important. Even if the correct functional form of the model has been specified, the poor starting values usage might causes model fail to converge or get a locally optimal solution rather than one that is globally optimal.

In term of data, the dependent and independent variables should be quantitative. Categorical variables such as religion, major, or region of residence need to be transform to binary (dummy) variables or other types of contrast variables.

Many models that appear nonlinear at first can be transformed to a linear model, which can be analyzed using the Linear Regression procedure. If the proper model is uncertain, the Curve Estimation procedure can help to identify useful functional relations in giving data if the proper model become uncertainty.

CHAPTER IV

RESULT

This chapter describes how the nonlinear regression model is analyzed with the interpreted data. The obtained model involved with General Factorial, Correlation, Curve Estimate, and Nonlinear Regression.

4.1 Model Assumptions

In order to obtain the reliable nonlinear regression analysis and more accurate prediction, our data have to be adjusted and considered to follow with hypothesis of nonlinear regression. Thus, the accurate model form has to be specified.

The independent variables in the regression analysis are web page size (HTTP workload), number of hub, switch, and distance between Client/Server.

4.2 Correlation Choosing Related Factors

The relationships among independent variables can be described by the correlation, Pearson correlation coefficient. These coefficients help us to choose whether independent variables should be in the regression model or not.

Table 4.1 shows that there is no coefficient between any pair of independent variables reach 0.5 so all independent variables should appear in the model. Furthermore, in term of delay time effect, correlation coefficients present that our regression model should consist of these independent variables as follow:

1. FILESIZE,

2. HUB,

3. SWITCH,

4. DISTANCE.

The variable "TIME" is our dependent variable of the delay time.

Table 4.1 Correlation coefficients

Correlations

		TIME	FILESIZE	HUB	SWITCH	DISTANCE
Pearson Correlation	TIME	1.000	.974	.083	-.023	.035
	FILESIZE	.974	1.000	.000	.000	.000
	HUB	.083	.000	1.000	-.316	.227
	SWITCH	-.023	.000	-.316	1.000	.047
	DISTANCE	.035	.000	.227	.047	1.000
Sig. (2-tailed)	TIME	.	.000	.046	.576	.401
	FILESIZE	.000	.	1.000	1.000	1.000
	HUB	.046	1.000	.	.000	.000
	SWITCH	.576	1.000	.000	.	.256
	DISTANCE	.401	1.000	.000	.256	.
N	TIME	575	575	575	575	575
	FILESIZE	575	575	575	575	575
	HUB	575	575	575	575	575
	SWITCH	575	575	575	575	575
	DISTANCE	575	575	575	575	575

The definition of each variable in Table 4.1 is described as follows.

− FILESIZE represents the size of the retrieved web page on our server. The measurement unit is Byte.

− HUB represents the numbers of hubs installed over our network. In this experiment the value of hub can be 0,1, 2, 3, or 4.

− SWITCH represents the numbers of switch installed over our network. According to the limitation of our experiment, the value of switch will be 0 or 1.

− DISTANCE represents the total cable length between the client and the server. The distance will be measured in meters.

− TIME represents the delay time measured in our experiment. The measurement unit of time is millisecond.

4.3 Relationships between *TIME* and variables

In this section, we describe relationships between dependent variable *TIME* and each of independent variables.

4.3.1 Relationship between *TIME* and *FILESIZE*
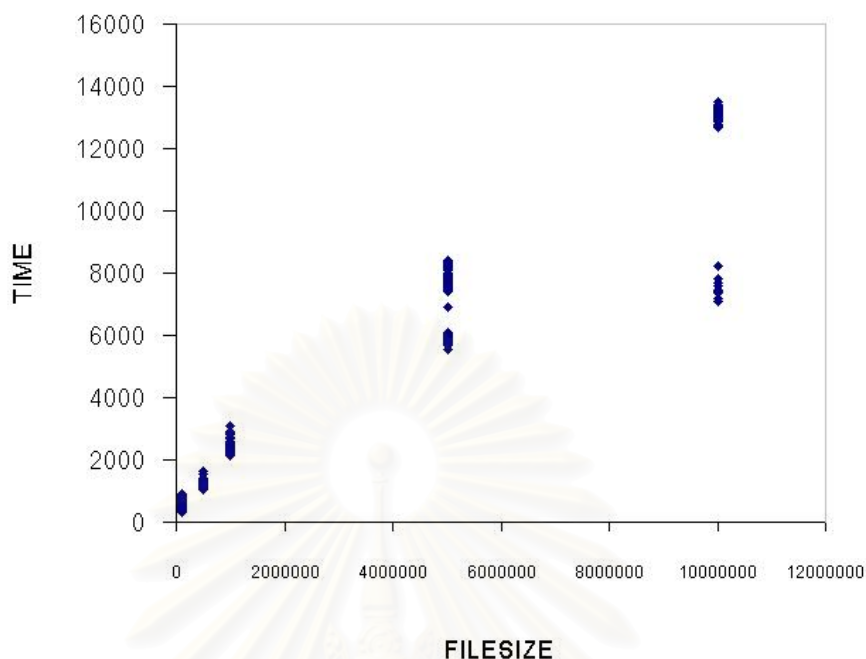
Figure 4.1 Scatter plot for *TIME* and *FILESIZE*

Figure 4.1 shows that the increasing of *FILESIZE* causes the increasing of *TIME*. However, we can see that there are two independent curves appeared in Figure 4.1. After consideration of the functions of switch, we can conclude that the splitting curve is caused by the existence of the switch over the network. This detail will be discussed in Section 4.4.

4.3.2 Relationship between *TIME* and *HUB*
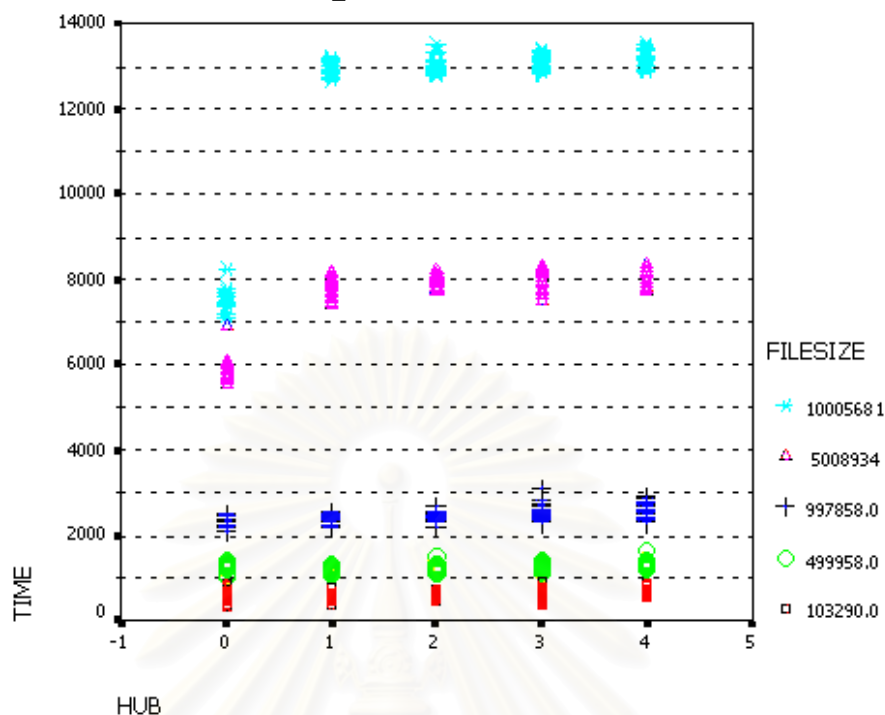
Figure 4.2 Scatter plot for *TIME* and *HUB*

Figure 4.2 shows that *HUB* trivially increases *TIME*. In addition, this increasing is separated into levels according to the values of *FILESIZE*.

4.3.3 Relationship between *TIME* and *SWITCH*
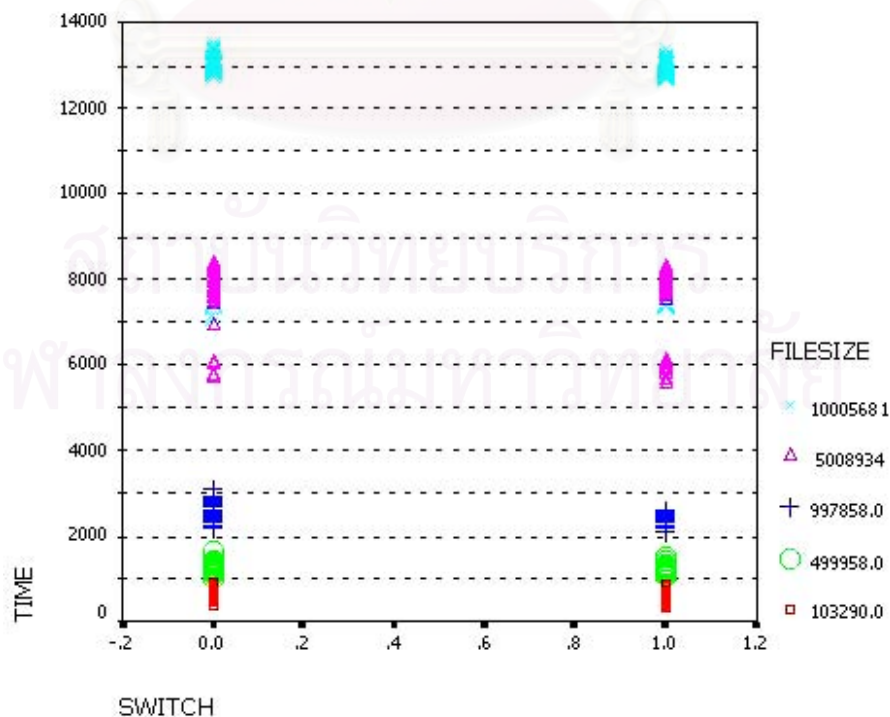
Figure 4.3 Scatter plot for *TIME* and *SWITCH*

Figure 4.3 shows that the increasing of *SWITCH* causes the reduction of *TIME*. This relationship is also separated by values of *FILESIZE*.

4.3.4 Relationship between *TIME* and *DISTANCE*
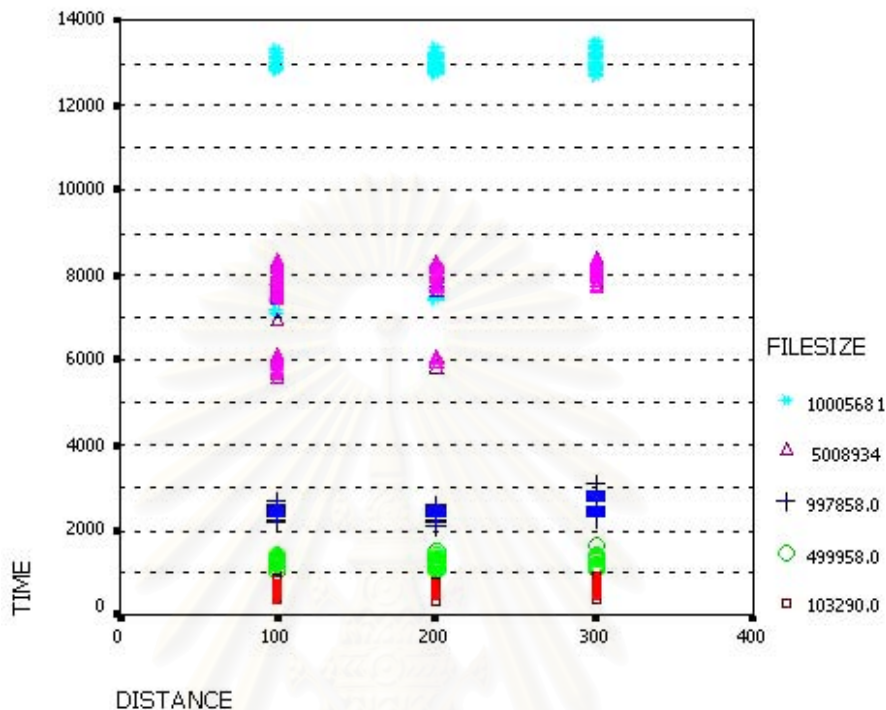
Figure 4.4 Scatter plot for *TIME* and *DISTANCE*

Figure 4.4 shows that the increasing of *DISTANCE* might cause *TIME* increase for each of *FILESIZE* levels.

4.4 Curve estimation and model specification

In this section, we consider the independent variable, FILESIZE, which is the factor that has the clearest effect on the delay time, as seen in scatter plots. Figure 4.5 shows the dominant relationship between *TIME* and *FILESIZE* with data points marked by values of *SWITCH*.
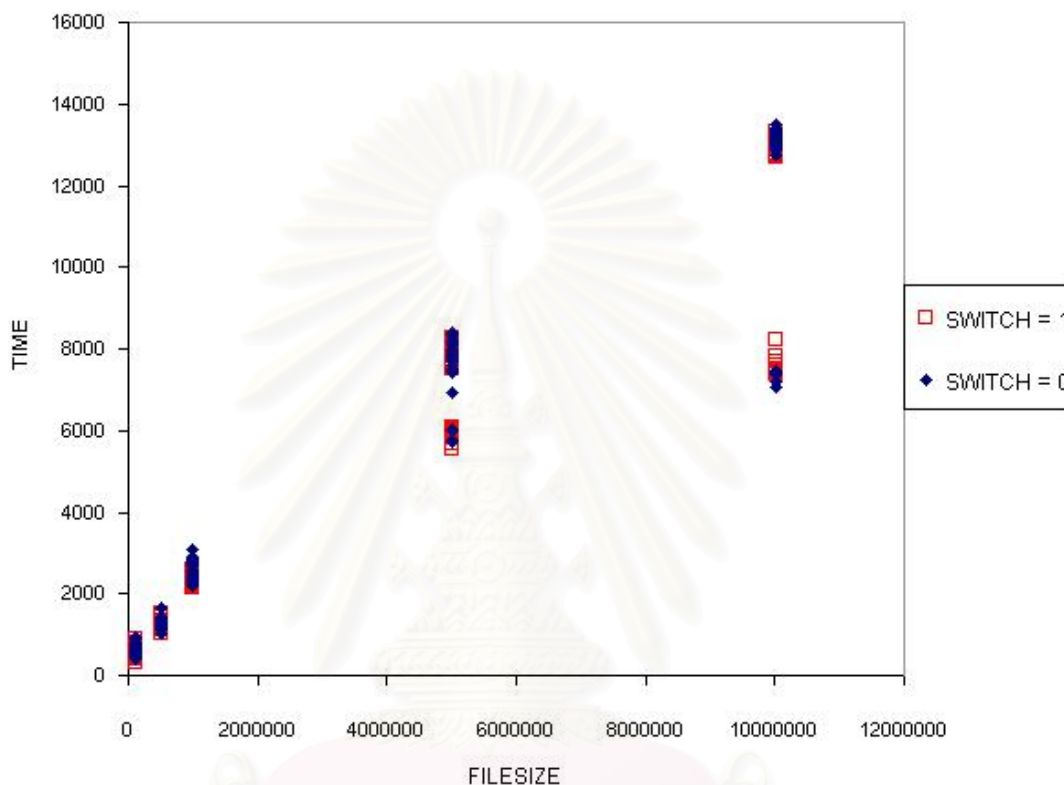
Figure 4.5 Scatter plot for TIME and FILESIZE marked by SWITCH

From Figure 4.5 we can see that the relationship between TIME and FILESIZE is not a linear. Therefore, the important point is to specify the accurate nonlinear regression model. Running Curve Estimation procedure of SPSS program produces curve estimation regression statistics and related plots for several different curve estimation regression models. A Power model is produced for our dependent variable as showed in Table 4.2.

Table 4.2 Curve estimate ANOVA

```
Independent:  FILESIZE
```

Dependent	Mth	Rsq	d.f.	F	Sigf	StdError
TIME	LIN	.965	574	15768.9	.000	1252.55451
TIME	LOG	.632	574	986.24	.000	4053.82116
TIME	QUA	.979	573	13348.4	.000	969.66067
TIME	CUB	.981	572	10046.6	.000	913.71080
TIME	POW	.999	574	549182	.000	.25895
TIME	EXP	.562	574	735.55	.000	5.30563

Result from Table 4.2 has done with six types of model such as Linear, Logarithmic, Quadratic, Cubic, Power, and Exponential. Comparisons among those models by consider R-squared and Standard Error hence us to choose the Power model which .999 R-squared and least of .25895 Standard Error. The results from curve estimation help us to decide whether model is suitable or not. So the power model is the most accurate model and a part of our result model has been constructed.

The Power model whose equation is $Y = b_0 \times X^{b_1}$

Where

Y is the dependent variable

X is the dependent variable

b_0 is the coefficient of X

b_1 is the power

Therefore, the general nonlinear regression model of TIME and FILESIZE is,

$$TIME = b_0 \times FILESIZE^{b_1}$$

The obtained results from Correlation, relationships between *TIME* and independent variables, and Curve estimate have been interpreted and combined. So the expected nonlinear regression model for predict delay time is in the form of

$$TIME = a_1(FILESIZE^{a_2}) + a_3\left(DISTANCE^{a_4}\right) + a_5(HUB)$$
$$+ a_6(SWITCH)(DISTANCE)$$

where
a_1, a_3, a_5, and a_6 are coefficients of each term.
a_2, and a_4 are powers of *FILESIZE* and *DISTANCE* respectively.

4.5 fitting the nonlinear regression model

The entire equation with all independent variables is running under nonlinear regression with the significant of 0.05. The starting values of a_1, a_2, a_3, a_4, a_5, and a_6 are assigned to 1. The nonlinear regression helps us to evaluate the coefficient and power of any factor in the equation. After a number of iteration, the coefficient and powers of factors have been converged as showed in Table 4.3, so the best result model is

$$TIME = 0.044(FILESIZE^{0.776}) - 824.077(DISTANCE^{-114.722})$$
$$+ 215.294(HUB) - 0.387(SWITCH)(DISTANCE)$$

Table 4.3 Value of a_i obtained from nonlinear regression

a_i	Estimate	Std. Error	Lower	Upper
a_1	0.044	0.011	0.022	0.067
a_2	0.776	0.016	0.745	0.807
a_3	-824.103	0	-824.103	-824.103
a_4	-114.722	0	-114.722	-114.722
a_5	215.295	24.554	167.066	263.523
a_6	-0.387	0.310	-0.996	0.221

The R-squared of this model is equal to 0.965, means that the accuracy to predict the delay time after knowing all independent variables using this model is 96.5%. So, we conclude that the nonlinear regression model obtained from our analytical process is appropriate for this set of data.

If we consider the term $-824.103(DISTANCE^{-114.722})$ with the values of DISTANCE, we see that this term is approach to zero.

From the result of our analysis shows that the nonlinear regression model to predict the delay time is,

$$TIME = 0.044(FILESIZE^{0.776}) + 215.294(HUB) - 0.387(SWITCH)(DISTANCE)$$

where

TIME is the delay time of loaded web application.

FILESIZE is the size of the retrieved web page on the server.

SWITCH is the numbers of switches installed over our network.

HUB is the numbers of hubs installed over our network.

DISTANCE is the total cable length between a client and the server.

4.6 Checking the nonlinear regression model

In order to test the nonlinear regression model obtained from the previous section, we use this model to analyse the testing data set, 575 observed values. To test the obtained model, we run this model with the testing data set in the nonlinear regression with the starting values 1 for a_1, a_2, a_3, a_4, a_5, and a_6 to obtain all constants and coefficients of variables. Table 4.4 shows the testing result that contains converged coefficients and powers of factors. So the testing model is

$$TIME = 0.045(FILESIZE^{0.776}) - 824.077(DISTANCE^{-114.718})$$
$$+ 211.371(HUB) - 0.409(SWITCH)(DISTANCE)$$

Table 4.4 Value of a_i obtained from testing nonlinear regression

a_i	Estimate	Std. Error	Lower	Upper
a_1	0.045	0.012	0.022	0.068
a_2	0.776	0.016	0.744	0.807
a_3	-824.077	0	-824.077	-824.077
a_4	-114.718	0	-114.718	-114.718
a_5	211.371	24.727	162.803	259.938
a_6	-0.409	0.312	-1.022	0.204

The R-squared of this new nonlinear regression model is equal to 0.965, means that the accuracy to predict the delay time after knowing all independent variables using this model is 96.5%. So, this nonlinear regression model obtained from our analytical process is appropriate.

Thus, we conclude that the general regression model to predict the TIME when FILESIZE, SWITCH, HUB, and DISTANCE are defined is

$$TIME = a_1(FILESIZE^{a_2}) + a_3(HUB) + a_4(SWITCH)(DISTANCE)$$

where

a_1, a_3, and a_4 are coefficients of each term.
a_2 is the power of *FILESIZE*.

CHAPTER V

CONCLUSION

5.1 Experiment Conclusion

The general regression model to predict the delay time is,
$$TIME = b_1\left(FILESIZE^{b_2}\right) + b_3 HUB$$
$$+ b_4(DISTANCE)(SWITCH)$$

Where
- $TIME$ is the delay time measured in our experiment.
- $FILESIZE$ is the size of the retrieved web page on our server.
- $DISTANCE$ is the total cable length between the client and the server.
- HUB is the numbers of hubs installed over the network.
- $SWITCH$ is the numbers of switches installed over the network.
- b_1, b_3, and b_4 are coefficients of each variable.
- b_2 is a power of a variable.

Where b_i 's value is presented in Table 5.1.

Table 5.1 Value of b_i obtained from nonlinear regression

b_i	Estimate	Std. Error	Lower	Upper
b_1	0.044	0.011	0.022	0.067
b_2	0.776	0.016	0.745	0.807
b_3	215.295	24.554	167.066	263.523
b_4	-0.387	0.310	-0.996	0.221

Considering the value of each variable b_i, we can see that the value of b_1, b_2 and b_5 are positive values while b_6 is a negative value. Therefore, we can say that the factors that increase the delay time are *FILESIZE*, and *HUB*, where the prime factor is the *FILESIZE*. Moreover, *SWITCH* can reduce the effect of *DISTANCE*.

5.2 Suggestions

This physical-effect experiment of Intranet was performed in a controllable environment so that there is no interference from users, or

Internet. However, there are many factors that are not existed in our experiment that may or may not be approved by the regression model of the delay time under this research.

For example, there is no congestion occurs in the Intranet system. However, if we want to have the real situation of an Intranet, the congestion of the traffic might be able to arise. Thus, the model that an environment under the congestion situation might be different from the obtained model in this research. Moreover, if the experiment had performed under the full scale of switches without hubs, the delay time model under this situation also might not similar to our presented regression model.

Consider all leftover factors such as number of active connection devices –switches and routers, congestion situation, number of users over the network, etc. The researcher might find that the regression model to predict the delay time is different from the obtained model in this experiment.

5.3 Contribution

This experiment has presented effects that are able to increase or decrease the delay time value while a web page is retrieved. Although there are many factors that are not taken into account, but at least ideas of using or implementing a network to support the web usages has shown in this research. In addition, the result model will help and optimize network's adjustment to gain more performance communications.

References

1. Kuehl, R. O. Design of Experiments: Statistical Principles of Research Design and Analysis. 2nd ed. California: Brooks/Cole, 2000.
2. Giles S. Workshop on data communication. Thai-Australia Science & Engineering Assistance Project. p. 42. June, 2000.
3. Kaste, V. A. T., Brodeen, A. E. M., and Broom B. D. Empirical evaluation of protocol performance over tactical networks. IEEE Transactions on Communication. 1992: pp. 533-535.
4. Kao, S. Variations on Ethernet traffic monitoring. IEEE TENCON'93 / Beijing. 1993: pp. 496-499.
5. Bimbo, A. D., and Vicario, E. Transport measurements over an Ethernet LAN. IEEE Transactions on Communication. Vol. 42. 1994: pp. 1486-1489.
6. Park, K., Kim, G., and Crovella M. On the relationship between file sizes, transport protocols, and self-similar network traffic. IEEE Transactions on Communication. 1996: pp. 171-180.
7. Bhattarakosol, P., and Vuthichaiworakul, J. A data transfer model for the file transfer protocol. The 2001 International Conference on Information Technology for the new Millennium. Bangkok 2000: pp. 213-221.
8. Heidemann, J., Obraczka, K., and Touch, J. Modeling the performance of HTTP over several transport protocols. IEEE/ACM Transactions on Networking. 1997: pp. 616-630.
9. Dovrolis, C., Stiliadis, D., and Ramanathan, P. Proportional differentiated services: Delay differentiation and packet scheduling. ACM SIGCOMM. 1999: pp. 109-120.
10. Wills, C. E., and Mikhailov, M. Towards a better understanding of web resources and server responses for improved caching [online]. 1999. Available from: http://www8.org/w8-papers/2a-webserver/towards/towards.html [2001, March 7]
11. Hu, J. C., Mungee, S., and Schmidt, D. C. Techniques for developing and Measuring high performance web servers over high speed networks. IEEE Transactions on Communication. 1998: pp. 1222-1231.
12. SPSS 7.5.1 for Windows [computer software]. U.S.A.: SPSS Inc., 1996. Available from: Department of Mathematics, Faculty of Science, Chulalongkorn University [2001, July 14]
13. McClave, J. T., Deitrich, F. H. II., and Sincich T. Statistics. 7th ed. (n.p.): Prentice Hall, 1997.

APPENDIX

Non-linear Regression

Derived model

Output Created		09 Oct 01 08:36:45
Comments		
Input	Data	D:\NoteBook1\papers\http\NetDiff\document\modify\Test1.sav
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	575
	Syntax	MODEL PROGRAM A1=1 A2=1 A3=1 A4=1 A5=1 A6=1 . COMPUTE PRED_ = A1*(Filesize**a2) + a3 * (distance**a4) + a5 * hub + a6*Distance*switch . NLR time /OUTFILE='C:\WINDOWS\TEMP\SPSSFNLR.TMP' /PRED PRED_ /CRITERIA SSCONVERGENCE 1E-8 PCON 1E-8 .
Resources	Elapsed Time	0:00:01.37

```
       All the derivatives will be calculated numerically.
 _
```

```
Iteration   Residual SS      A1          A2          A3          A4
                             A5          A6

   1       1.4506E+16   1.00000000  1.00000000  1.00000000  1.00000000
                        1.00000000  1.00000000
  1.1      1.6900E+26    .007852519  .999587988  -19.841390  4.32752764
                        291.462109   .021494571
  1.2      1.0536E+16    .936541539  .996019060  -824.10320  -144.72239
                        -66088.280   -790.35400
   2       1.0536E+16    .936541539  .996019060  -824.10320  -144.72239
                        -66088.280   -790.35400
  2.1      4.7580E+15    .789398944  .987363252  -824.10320  -144.72239
                        -204292.85   -2346.5614
   3       4.7580E+15    .789398944  .987363252  -824.10320  -144.72239
                        -204292.85   -2346.5614
  3.1      7.5762E+14    .449393932  .970374029  -824.10320  -144.72239
                        -191324.32   -1636.6493
   4       7.5762E+14    .449393932  .970374029  -824.10320  -144.72239
```

		-191324.32	-1636.6493		
4.1	3.1113E+11	.009580580	.969317408	-824.10320	-144.72239
		218.289492	-.36558834		
5	3.1113E+11	.009580580	.969317408	-824.10320	-144.72239
		218.289492	-.36558834		
5.1	29875535294	.009699944	.919196883	-824.10320	-144.72239
		218.147054	-.36662002		
6	29875535294	.009699944	.919196883	-824.10320	-144.72239
		218.147054	-.36662002		
6.1	465797945.4	.017008316	.838610382	-824.10320	-144.72239
		212.725382	-.40613531		
7	465797945.4	.017008316	.838610382	-824.10320	-144.72239
		212.725382	-.40613531		
7.1	2445618581	.035551456	.768048497	-824.10320	-144.72239
		210.475009	-.42253625		
7.2	431685093.5	.020597893	.822493411	-824.10320	-144.72239
		256.321479	-.08839039		
8	431685093.5	.020597893	.822493411	-824.10320	-144.72239
		256.321479	-.08839039		
8.1	654936042.4	.031705668	.790263827	-824.10320	-144.72239
		230.150660	-.27913345		
8.2	423340402.7	.022055192	.819270948	-824.10320	-144.72239
		253.853762	-.10637638		
9	423340402.7	.022055192	.819270948	-824.10320	-144.72239
		253.853762	-.10637638		
9.1	422304821.0	.024401212	.812825111	-824.10320	-144.72239
		248.356103	-.14644482		
10	422304821.0	.024401212	.812825111	-824.10320	-144.72239
		248.356103	-.14644482		
10.1	421052621.1	.027072613	.806379771	-824.10320	-144.72239
		242.761222	-.18722158		
11	421052621.1	.027072613	.806379771	-824.10320	-144.72239
		242.761222	-.18722158		

Iteration	Residual SS	A1	A2	A3	A4
		A5	A6		
11.1	419977106.2	.030044204	.799934946	-824.10320	-144.72239
		237.074360	-.22867095		
12	419977106.2	.030044204	.799934946	-824.10320	-144.72239
		237.074360	-.22867095		
12.1	419142090.8	.033342510	.793490744	-824.10320	-144.72239
		231.307829	-.27069958		
13	419142090.8	.033342510	.793490744	-824.10320	-144.72239
		231.307829	-.27069958		
13.1	418560286.3	.037002596	.787047306	-824.10320	-144.72239
		225.461634	-.31330825		
14	418560286.3	.037002596	.787047306	-824.10320	-144.72239
		225.461634	-.31330825		
14.1	418239626.8	.041063917	.780604798	-824.10320	-144.72239
		219.534631	-.35650699		
15	418239626.8	.041063917	.780604798	-824.10320	-144.72239
		219.534631	-.35650699		
15.1	417770266.5	.044691180	.775487951	-824.10320	-144.72239
		214.741378	-.39144137		
16	417770266.5	.044691180	.775487951	-824.10320	-144.72239
		214.741378	-.39144137		
16.1	417462495.4	.044316469	.776234513	-824.10320	-144.72239

```
                              215.376771   -.38681060
   17         417462495.4   .044316469   .776234513   -824.10320   -144.72239
                              215.376771   -.38681060
   17.1       417462448.5   .044392363   .776130337   -824.10320   -144.72239
                              215.281391   -.38750590
   18         417462448.5   .044392363   .776130337   -824.10320   -144.72239
                              215.281391   -.38750590
   18.1       417462447.9   .044382219   .776144564   -824.10320   -144.72239
                              215.294650   -.38740877
```

Run stopped after 39 model evaluations and 18 derivative evaluations.
Iterations have been stopped because the relative reduction between successive
residual sums of squares is at most SSCON = 1.000E-08
—

Nonlinear Regression Summary Statistics   Dependent Variable TIME

Source	DF	Sum of Squares	Mean Square
Regression	6	25222755352.1	4203792558.68
Residual	569	417462447.936	733677.41289
Uncorrected Total	575	25640217800.0	
(Corrected Total)	574	12058170188.9	

R squared = 1 - Residual SS / Corrected SS =   .96538

Parameter	Estimate	Asymptotic Std. Error	Asymptotic 95 % Confidence Interval Lower	Upper
A1	.044382219	.011398163	.021994610	.066769829
A2	.776144564	.015878832	.744956286	.807332843
A3	-824.1031974	.000000000	-824.1031974	-824.1031974
A4	-144.7223880	.000000000	-144.7223880	-144.7223880
A5	215.29465043	24.554326926	167.06646784	263.52283301
A6	-.387408772	.309906642	-.996109395	.221291851

Non-linear Regression

Tested model

Output Created	09 Oct 01 21:09:55	
Comments		
Input	Data	C:\My Documents\paper\modify\non_cut\model1_test.sav
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	575
	Syntax	MODEL PROGRAM A1=1 A2=1 A3=1 A4=1 A5=1 A6=1 . COMPUTE PRED_ = a1 * Filesize** a2 + a3 * (distance **a4) + a5* Hub + a6 *Switch*Distance. NLR time2 /OUTFILE='C:\WINDOWS\TEMP\SPSSFNLR.TMP' /PRED PRED_ /CRITERIA SSCONVERGENCE 1E-8 PCON 1E-8 .
Resources	Elapsed Time	0:00:01.76

```
            All the derivatives will be calculated numerically.

  _

   Iteration   Residual SS        A1           A2           A3           A4
                                  A5           A6

      1        1.4506E+16   1.00000000   1.00000000   1.00000000   1.00000000
                            1.00000000   1.00000000
     1.1       1.2890E+26    .007869284   .999586977  -19.684204    4.30511196
                            285.900783   -.01507096
     1.2       1.0536E+16    .936544052   .996019218  -824.07714   -144.71784
                            -66086.237   -790.33066
      2        1.0536E+16    .936544052   .996019218  -824.07714   -144.71784
                            -66086.237   -790.33066
     2.1       4.7583E+15    .789407555   .987363747  -824.07714   -144.71784
                            -204288.01   -2346.5162
      3        4.7583E+15    .789407555   .987363747  -824.07714   -144.71784
                            -204288.01   -2346.5162
     3.1       7.5771E+14    .449416564   .970375045  -824.07714   -144.71784
```

		−191343.45	−1636.8494		
4	7.5771E+14	.449416564	.970375045	−824.07714	−144.71784
		−191343.45	−1636.8494		
4.1	3.1470E+11	.009626208	.969312332	−824.07714	−144.71784
		214.206117	−.38822612		
5	3.1470E+11	.009626208	.969312332	−824.07714	−144.71784
		214.206117	−.38822612		
5.1	30267476965	.009746887	.919135422	−824.07714	−144.71784
		214.062667	−.38926057		
6	30267476965	.009746887	.919135422	−824.07714	−144.71784
		214.062667	−.38926057		
6.1	473181969.7	.017108625	.838343646	−824.07714	−144.71784
		208.649097	−.42871807		
7	473181969.7	.017108625	.838343646	−824.07714	−144.71784
		208.649097	−.42871807		
7.1	2484902121	.035865033	.767349328	−824.07714	−144.71784
		206.445292	−.44477910		
7.2	437391114.7	.020715175	.822185339	−824.07714	−144.71784
		252.723260	−.10748809		
8	437391114.7	.020715175	.822185339	−824.07714	−144.71784
		252.723260	−.10748809		
8.1	662791394.7	.031912016	.789873162	−824.07714	−144.71784
		226.457772	−.29892182		
8.2	429094077.7	.022181677	.818954620	−824.07714	−144.71784
		250.256053	−.12547135		
9	429094077.7	.022181677	.818954620	−824.07714	−144.71784
		250.256053	−.12547135		
9.1	428044300.4	.024547188	.812492269	−824.07714	−144.71784
		244.736574	−.16569977		
10	428044300.4	.024547188	.812492269	−824.07714	−144.71784
		244.736574	−.16569977		
10.1	426774403.5	.027241792	.806030423	−824.07714	−144.71784
		239.118667	−.20664387		
11	426774403.5	.027241792	.806030423	−824.07714	−144.71784
		239.118667	−.20664387		

—

Iteration	Residual SS	A1	A2	A3	A4
		A5	A6		
11.1	425682216.0	.030240038	.799569102	−824.07714	−144.71784
		233.408088	−.24826452		
12	425682216.0	.030240038	.799569102	−824.07714	−144.71784
		233.408088	−.24826452		
12.1	424832502.7	.033568822	.793108415	−824.07714	−144.71784
		227.617306	−.29046983		
13	424832502.7	.033568822	.793108415	−824.07714	−144.71784
		227.617306	−.29046983		
13.1	424238151.1	.037263712	.786648506	−824.07714	−144.71784
		221.746339	−.33326009		
14	424238151.1	.037263712	.786648506	−824.07714	−144.71784
		221.746339	−.33326009		
14.1	423907184.3	.041364737	.780189546	−824.07714	−144.71784
		215.793988	−.37664258		
15	423907184.3	.041364737	.780189546	−824.07714	−144.71784
		215.793988	−.37664258		
15.1	423476153.0	.045168498	.774850317	−824.07714	−144.71784
		210.787473	−.41313188		
16	423476153.0	.045168498	.774850317	−824.07714	−144.71784

```
                             210.787473  -.41313188
      16.1    423115663.4  .044773140  .775638306  -824.07714  -144.71784
                             211.458457  -.40824163
      17      423115663.4  .044773140  .775638306  -824.07714  -144.71784
                             211.458457  -.40824163
      17.1    423115609.6  .044855079  .775526985  -824.07714  -144.71784
                             211.356477  -.40898505
      18      423115609.6  .044855079  .775526985  -824.07714  -144.71784
                             211.356477  -.40898505
      18.1    423115608.9  .044844022  .775542339  -824.07714  -144.71784
                             211.370805  -.40888027
```

Run stopped after 39 model evaluations and 18 derivative evaluations.
Iterations have been stopped because the relative reduction between successive
residual sums of squares is at most SSCON = 1.000E-08
—

Nonlinear Regression Summary Statistics   Dependent Variable TIME2

Source	DF	Sum of Squares	Mean Square
Regression	6	25203510091.1	4200585015.18
Residual	569	423115608.902	743612.66942
Uncorrected Total	575	25626625700.0	
(Corrected Total)	574	12072266755.8	

R squared = 1 - Residual SS / Corrected SS =   .96495

| | | | Asymptotic 95 % |
| | | Asymptotic | Confidence Interval |
Parameter	Estimate	Std. Error	Lower	Upper
A1	.044844022	.011576888	.022105372	.067582673
A2	.775542339	.015961270	.744192139	.806892539
A3	-824.0771438	.000000000	-824.0771438	-824.0771438
A4	-144.7178377	.000000000	-144.7178377	-144.7178377
A5	211.37080527	24.727013567	162.80344161	259.93816892
A6	-.408880272	.312027344	-1.021746256	.203985712

VITA

Name: Mr. Sansiri Lertkijcharoenwong

Birth date: 01/09/77

Place of Birth: Bangkok, Thailand

Academic Background:
1994-1997   B.Sc. (Mathematics)   Chulalongkorn University, Bangkok
Thailand

Experiences:
2001   Teacher assistance for Introduction to Computer and
Programming Techniques
.