ปัจจัยที่มีความสัมพันธ์ต่อเวลาการรอคอยบนสภาวะแวดล้อมของเอชทีทีพี

นาย ศุภสิทธิ์ อภิวัฒนาพงศ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาการคณนา ภาควิชาคณิตศาสตร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2544
ISBN 974-03-0581-4

FACTORS RELATED TO THE DELAY TIME OVER THE HTTP ENVIRONMENT

MR. SUPASIT APIWATTANAPONG

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Computational Science
Department of Mathematics
Faculty of Science
Chulalongkorn University
Academic Year 2001
ISBN 974-03-0581-4

Thesis Title	FACTORS RELATED TO THE DELAY TIME OVER THE HTTP ENVIRONMENT
By	Mr. Supasit Apiwattanapong
Field of Study	Computational Science
Thesis Advisor	Assistant Professor 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)

ศุภสิทธิ์ อภิวัฒนาพงศ์ : ปัจจัยที่มีความสัมพันธ์ต่อเวลาการรอคอยบนสภาวะแวดล้อมของเอชทีทีพี. (FACTORS RELATED TO THE DELAY TIME OVER THE HTTP ENVIRONMENT) อ.ที่ปรึกษา : ผู้ช่วยศาสตราจารย์ ดร. ภัทรสินี ภัทรโกศล, จำนวนหน้า37. ISBN 974-03-0581-4.

งานวิจัยนี้ได้ทำการพิจารณาเวลาการรอคอยบนสภาวะแวดล้อมของเอชทีทีพี ซึ่งเกิดจากเครื่องลูกข่ายหนึ่งตัวเรียกเว็บเพจจากเครื่องแม่ข่ายโดยผ่านฮับหนึ่งตัว ปัจจัยทางกายภาพที่จะพิจารณาเป็นตัวแปรอิสระของเครื่องลูกข่ายที่มีผลกระทบกับเวลาการรอคอย คือ ความเร็วของหน่วยประมวลผล และขนาดของหน่วยความจำ อีกทั้งเวลาการรอคอยนี้ยังได้รับผลกระทบจากขนาดของเว็บเพจ และชนิดของเว็บเพจ ดังนั้นผู้วิจัยจะพิจารณาขนาดของเว็บเพจที่หลากหลาย และแบ่งชนิดของเว็บเพจออกตามอัตราส่วนของตัวอักษรและรูปภาพเป็นตัวแปรอิสระ นอกจากนี้ผู้วิจัยจะพิจารณาชนิดของบราวร์เซอร์ต่างๆ อีกตัวแปรหนึ่งด้วย การทดลองทั้งหมดนี้ได้ทำในระบบเครือข่ายแบบอินทราเน็ตและตัวแปรอิสระทุกตัวได้ถูกกำหนดค่าไว้แล้ว ผู้วิจัยได้ทำการวัดเวลาการรอคอยและบันทึกค่าหลังจากได้รับเว็บเพจสมบูรณ์ผ่านบราวเซอร์ ค่าของตัวแปรอิสระและค่าเวลาการรอคอยที่วัดได้เหล่านี้จะนำไปใช้เพื่อสร้างโมเดลทำนายเวลาการรอคอย โดยโมเดลที่ได้อยู่ในรูปของสมการถดถอยเชิงพหุคูณ ซึ่งแสดงปัจจัยต่างๆ ในสภาวะเครือข่ายที่กระทบกับเวลาการรอคอย หลังจากนั้นได้ทำการทดสอบโมเดล พบว่า ความถูกต้องของการทำนายเวลาการรอคอยกับกลุ่มข้อมูลที่ใช้ทดสอบมีความแม่นยำถึง 85% ซึ่งผู้วิจัยคาดว่าความสัมพันธ์ของตัวแปรเหล่านี้จะเป็นประโยชน์ในการออกแบบระบบอินทราเน็ตที่มีประสิทธิภาพได้ดีขึ้น

ภาควิชา.....คณิตศาสตร์.....ลายมือชื่อนิสิต.....

สาขาวิชา.....วิทยาการคณนา.....ลายมือชื่ออาจารย์ที่ปรึกษา.....

ปีการศึกษา.....2544.....

##4272411423
KEY WORD: delay time/ Intranet/multiple regression/ HTTP
MR. SUPASIT APIWATTANAPONG : FACTORS RELATED TO THE DELAY TIME OVER THE HTTP ENVIRONMENT. THESIS ADVISOR : ASSIST. PROF. PATTARASINEE BHATTARAKOSOL ,Ph.D. 37 pp. ISBN 974-03-0581-4

This research considered the delay time over the HTTP environment by a client requested web page from a server by pass through a hub. The physical factor of client affect the delay time that are considered as the independent variables such as types of processors and sizes of memory. Furthermore, the delay time is also effected from sizes of web page and types of web page. Considering various sizes of web pages and classifying web page by the ratio between the texts and graphics used for each page, these are counted as independent variables in this experiment. Additionally we also determine the differences between various browsers as an independent variable of our research. The experiment is done in intranet for each of fixed independent variables. A delay time is measured and recorded after completed reloading a web page. All observation values are applied for creating a model for predicting a proper delay time. The model is in the form of multiple regression that various factors related to the delay time. Using the obtained model, the correctness of a delay time prediction is 85% accuracy. Thus, the researcher believes that this result will help improving the efficiency of any Intranets designing.

Department.....Mathemaics.....Student's signature.....S. Apiwattanapong.....
Filed of study.....Computational science..Advisor's signature.....C. Bhattarakosol.....
Academic years...2001.....

ACKNOWLEDGEMENT

I would like to express my sincere gratitude and deep appreciation to Assist. Prof. Pattarasinee Bhattarakosol, my principal supervisor for her guidance, invaluable advice, supervision and encouragement through this research which enable me to complete this thesis successfully. I also would like to thank Mr. Sansiri Lertkijcharoenwong and Mr. Werawat Cheunpanyanun for supporting me during my experiment.

Sincere thanks also go to Assoc. Prof. Dr. David Ruffolo and Assist. Prof. Dr. Pornpote Piumsomboon for their valuable efforts to read, edit, and comment this thesis.

I owe unending gratitude to my family whose support and encouragement helped me to finish my thesis properly. I wish to thank Ms. Kanoktip Koatsamran and my friends for encouragement. I wish to thank Mr. Satit Wongwaiwaeh and the staff of the Department of Mathematics, Chulalongkorn University for their co-operation and generous supports that lead me for my graduation.

Mr. Supasit Apiwattanapong

Contents

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Introduction

The world wide web has rapidly become one of the most popular Internet services. The popularity of the web has resulted in a corresponding popularity for HTTP, the standard Hyper-Text Transport Protocol. HTTP is layered over TCP. Therefore, the efficiencies of clients, workstations, or servers, causes the high demands of network usages. Thus, the traffic over the network is increased. Consequently, the data transfer rate is reduced, or we can say that the delay time is increased.

From [6] research, the sizes of servers and client affect the delay time in TCP/IP network and [5] has developed a simple analytic model for HTTP performance over different networks and transport protocols From these experiments, our research will consider in the performance of a client and a server relates to the speed of a processor, size of memory, size of the web page, and the ratio of text and graphics on the web page. The expected result is the model that can predict the delay time caused by retrieving a web page.

1.2 Objectives

1. Analyze and classify types of web pages (Text and graphics).
2. Analyze the pattern of the relation between delay time and type of web pages accessed.
3. Analysis the pattern of relationships among types of servers and types of web pages that could minimize the delay time.

1.3 Scope of Thesis

- Applications and implementation methods classify the web pages.
- 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 occurs in our system.
- The network consists of one hub.
- The delay time is measured in the millisecond unit.

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

1.4 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 multiple regression model indicates the relations among the delay time and considered factors over the HTTP environment. This indication leads to the improvement of performance of an Intranet system. Additionally, the model can be used for obtaining the web page design condition that supports the small delay time with satisfaction of users. Moreover, the suitable client and server specifications can be drawn out to obtain the fast retrieval Intranet system.

CHAPTER II

LITERATURE REVIEW

This chapter describes several researches that related to the performance measurement in the closed Local Area Network (LAN) under several factors. These factors normally affect to the performance of any network application, especially LAN.

In 1990s, There were several studies about Ethernet LAN and its performance factors that affect the sending and receiving of data. Factors in [1] that were taken in the consideration are message length, message arrival rate, and transmission mode. The results can be applied to the design of communications architectures and protocols.

[2] analyses Ethernet traffic under the changes of buffer size, CPU power, and available memory space. The results show that a higher speed of CPU is required to reduce the transferring period.

In [3] there was more concerned with the performance evaluation of transport protocol that actually depends on protocol layering, processing overhead, operating system support, and architectural constraints. Its experiment compared the performance derived from different transport protocol, i.e. TCP and UDP, SPP and IDP, over an Ethernet LAN. In addition, it also described the method to measure the performance of network. However [3] does not consider the performance of HTTP that is a popular protocol for retrieving and linking data sources over a network. From [4] this research showed that the relationship between files sizes, network topology, protocol, and the network traffic. It led us to consider on the recent distribution of file sizes which too large for the past.

In [5] it has developed a simple analytic model for HTTP performance over different networks and transport protocols. This model is used to evaluate network overhead carrying HTTP traffic across a variety of network characteristics. An analysis of the transient effects of TCP slow-start had been presented. The accuracy of the model is within 5% of measured performance for wide-area networks. The model had shown that the connection setup costs are relatively unimportance for current modem, ISDN, and LAN users.

In [6] it had performed the experiment on the Ethernet-based personal computer systems in real time applications such as teleconferencing, computer telephony, multimedia broadcasting, real-time control of distributed systems, and distributed simulations. The results

show that there are some differences between the delay time of the frame transfer for different frame sizes. Additionally, the capabilities of each operating system, Microsoft Windows NT 4.0, Linux 2.1, Microsoft Windows 95, and MsDos, including various types of network cards, can be counted as factors that affect to the frame transfer system.

In [7], It has presented a new load-measurement-based ER (explicit-rate) calculation algorithm (Load Based Explicit Rate Control, or LBEPC) for an explicit-rate mode ABR(Available Bit Rate) service in Asynchronous Transfer Mode(ATM) Networks. This algorithm performs very effectively with TCP/IP traffic over ATM. The simulation to evaluate this algorithm is considered on World Wide Web traffic and a client-server model. The client-server model that is used in [14] consists of a web client, a web server, and a Finite Source Function (FSF) to represent the characteristics of users. The results had shown that the algorithm of ABR is more efficient than the UBR because the loss packet in ABR is small.

In [8], It provided the evidence that the delay time of CGI (Common Gateway Interface) requests follows a Pareto distribution with infinite mean and infinite variance. Moreover it had proved that the traffic to a web server as seen by the server itself and as seen by a proxy server was similar.

From [3] [5] [6] these works proposed the better Internet services and the better services management to support the several of Internet usage. However, they focused on the non-physical effects to the delay time but lacked other [3][5]. In [6] focused on the physical effects to the delay time but lacked other. So we try to study on the physical and non-physical effects to approach the same goal and apply the advantage to an Intranet.

If we consider network, they will be contained with a server, a client and a hub. Therefore in order to increase the performance of the available Intranet under configure of client, files, and applications. This research is focused on many factors that might affect to this performance so that the delay of the transmission time is minimized.

CHAPTER III

METHODOLOGY

This chapter will explain details of the data source and collection techniques, and the analytical procedures.

3.1 System Environment

The simulation system consists of a web server, a client, and a connection hub as shown in Figure 3.1. The specification of the web server is Pentium Pro 233 MHz, RAM size is 64 MB, network interface card is 10/100 MBPS, and operating system is Microsoft Windows NT 4.0. This experiment will vary configurations of the client so that the differences of the delay time will be shown. The standard system for the client is that the speed for its network interface card is 10/100 MBPS, and the operating system is Microsoft Windows 98. The other specifications of the client are as follows.

- CPU type: Pentium II at 350MHz, Pentium II at 400MHz, Pentium III at 733MHz,
- RAM size: 32 MB and 64 MB.

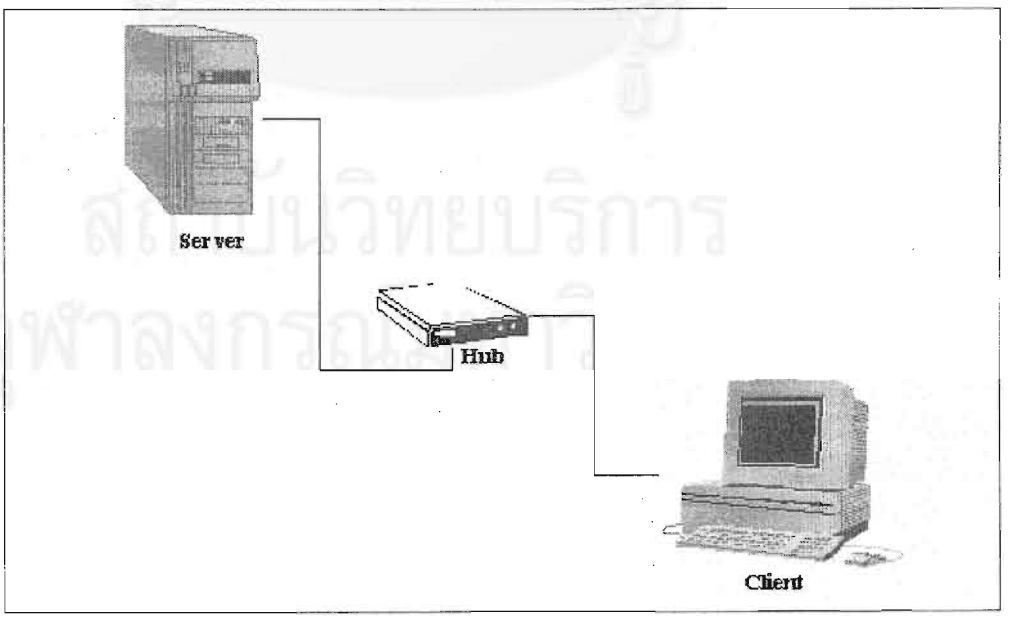

Figure 3.1: The measurement configuration

There are 3 sizes of the web pages: 50 KB, 100 KB, and 200 KB. Table 3.1: shows the file sizes that are used in the experiment, measurement unit is Byte. Additionally each type will be separated to be 4 subtypes of the web pages based on the ratio of text and graphics on the web page, as shown in Table 3.1.

Table 3.1: The real file sizes that are used in our experiment

	Type of web page				
Web page size	**Text 100%**	**Text75% and Picture 25%**	**Text50% and Picture 50%**	**Text25% and Picture 75%**	**Picture 100%**
50K	51,940	52,051	52,190	52,197	53,625
100K	102,916	102,162	103,290	105,217	102,177
200K	204,803	204,591	205,142	205,047	207,167

(unit: Bytes)

In order to determine the ability of famous web browsers, this research chooses the two most common license-paid browsers –Microsoft Internet Explorer 5.0 and Netscape Navigator 4.7, and a free popular browser –Opera 5.0.

3.2 Measurement Parameters

In this research, the environment of an Intranet is simulated in the computer laboratory. This is because we would like to have a close consideration of factors that affect to the delay time over the HTTP environment. These factors are size of the web page, types of processors, types of Browsers, size of memory, and ratio between the text and graphics in the page.

3.3 File sizes and type of web page

Currently, World Wide Web has rapidly become one of the most popular Internet services. Many web sites have emerged such as computer equipment and applications, entertainment (music, games, movies, etc.), social service information, business organization information, and others. This research will classify these web sites into 5 categories: education, news, entertainment, business, and search engine.

From [5], sizes of web pages are separated into 50KB, 100KB, and 200KB. However, there are some other factors that should be taken in the consideration of the delay time of browsing a web page. In order to confirm the web pages existing in the real world, 470 web sites had been randomly selected and saved into a local hard disk. These web sites are

classified to be 5 classes of the web applications: web for education (series-1), web for news (series-2), web for business (series-3), web for entertainment (series-4), and web for search engine (series-5). The distribution of each class is presented in Figure 3.2.

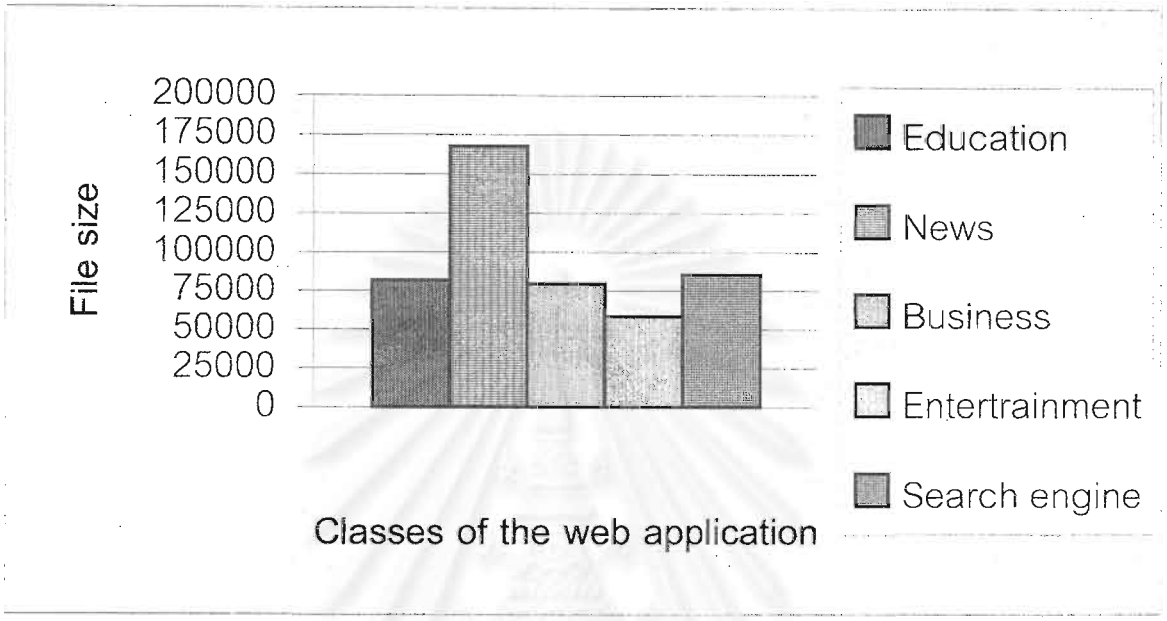

Figure 3.2: The web size distributions

According to the graph, it shows that there is some different among each type of web. Therefore the Analysis of Variance (ANOVA) under the Completed Randomize Design (CRD) is performed. However, the homogeneity of variance testing to confirm the variances among groups of webs are not significant different must be computed using the significant level equal to 0.05. The result of this test is that the variance of each group is significant difference ($p < 0.05 = \alpha$, α is the significant level). Thus, the ANOVA testing based parametric testing of CRD cannot be calculated.

As the result of significant different of variance among groups of web sites, the non-parametric test for determining the significant of means among web sites is performed using Kruskal-Wallis method. The result of the analysis can be concluded as follows.

The mean of sizes of web pages for education is significant different from the means of the sizes of web pages for news, entertainment, and search engine ($p < 0.05 = \alpha$). The mean of sizes of web pages for news is significant different from the means of sizes of web pages for business, entertainment, and search engine ($p < 0.05 = \alpha$).

According to the observation of the researcher, there are some differences between quantity of text and graphics presented on the web

According to the observation of the researcher, there are some differences between quantity of text and graphics presented on the web page. This ratio of the text and graphics may effect to the browser's loaded time. Therefore, the types of the web page will be grouped based on the ratio of the text and graphic on the web page. These groups are full text web page, text 75% and graphic 25%, text 50% and graphic 50%, text 50% and graphic 75%, and pure graphic web page.

3.4 Time-capturing program [9]

There are many kinds of browsers available for users on Internet. Some browsers are free such as Internet Explorer or Opera, but browsers have the license fees such as Netscape Navigator. These browsers normally have the same standard of retrieval method that is browsing the web page or connecting to a web site using a URL name. Once the URL is typed, the connection process using the HTTP starts. At the destination, web server, the requested will be interpreted to find the certain page that the user would like to retrieve. Then, that page will be sent as encapsulated bytes of information over the network back to the requested location, client. After the browser at the client receives the encapsulated information, these bytes of information will be de-encapsulated and presented to the user. In this research, the delay time interval is measured from the time starts the request until the full page is presented. Figure 3.3 shows the diagram of time measurement process. And Figure 3.4 shows the time measurement process.

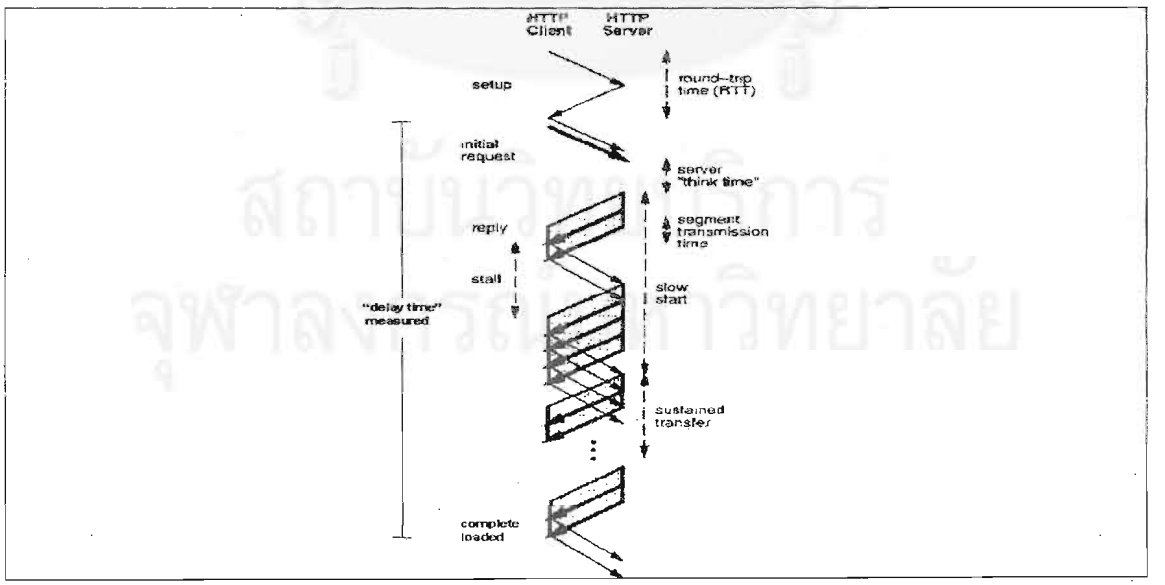

Figure 3.3: Time measurement process indicated by HTTP timing diagram [5]

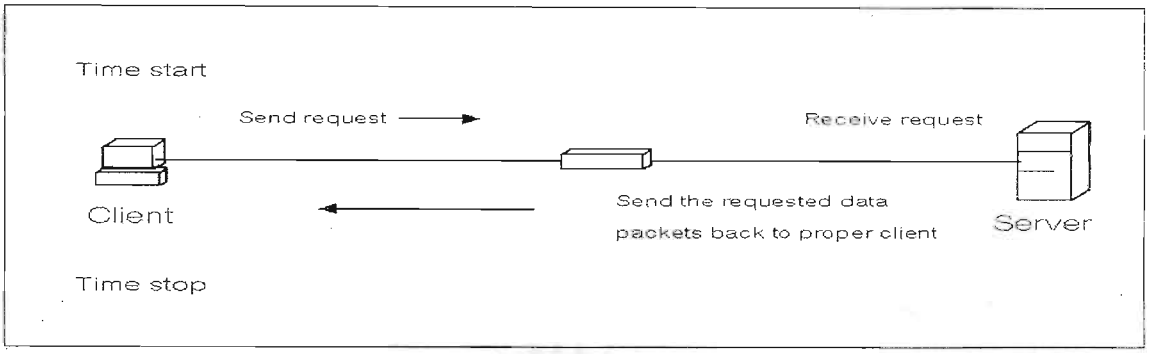

Figure 3.4: Time measurement process

In the time-captured program, implemented using the Java Scripts, will consist of two sub-programs. The first sub-program, Start-time is responsible for locating the web page. The screen presents the menu that lists various kinds of web pages as we separated in the previous section. This screen is called "HTML-files menu", Figure 3.5. To retrieve a web in the list, the user just presses mouse on the file name that is the link to the web server.

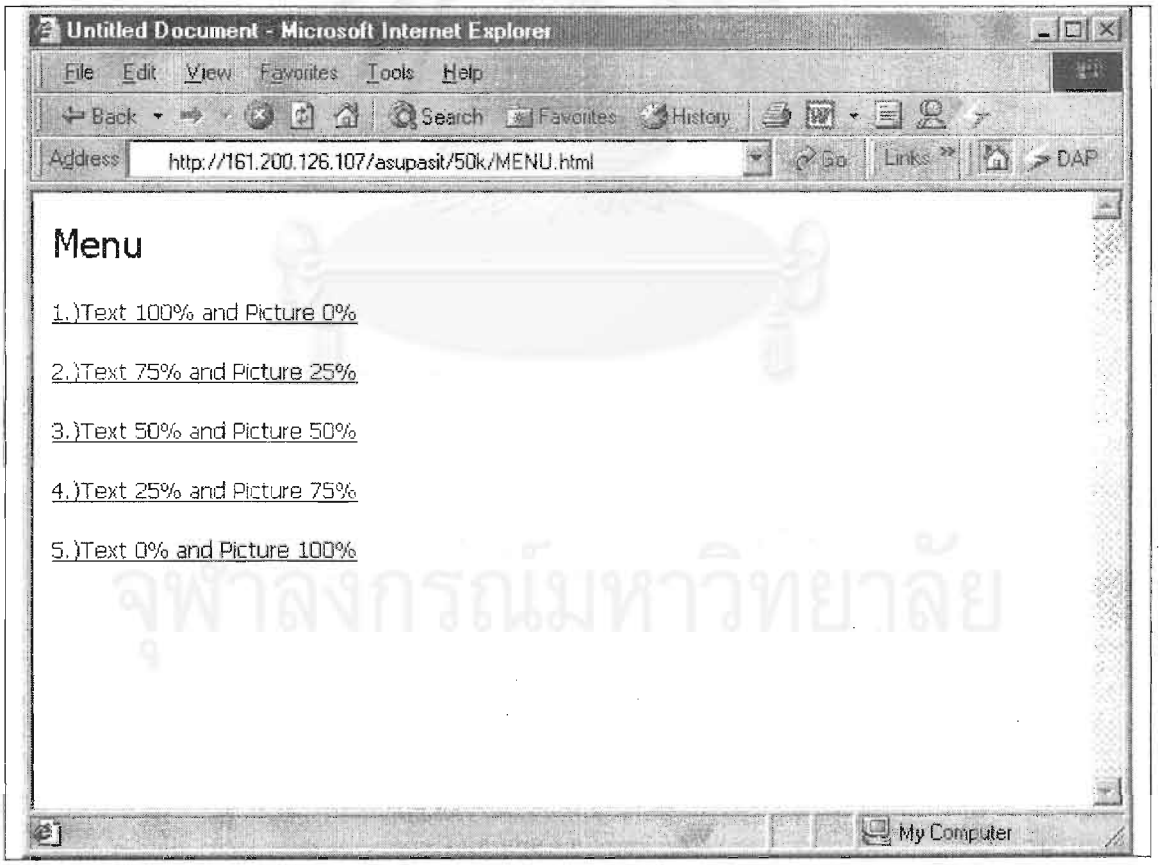

Figure 3.5: Menu web page

The second sub-program, Stop-time, is responsible for creating the finishing retrieving time for each web page. Figure 3.6 shows that the delay time of the web page is 330 millisecond.

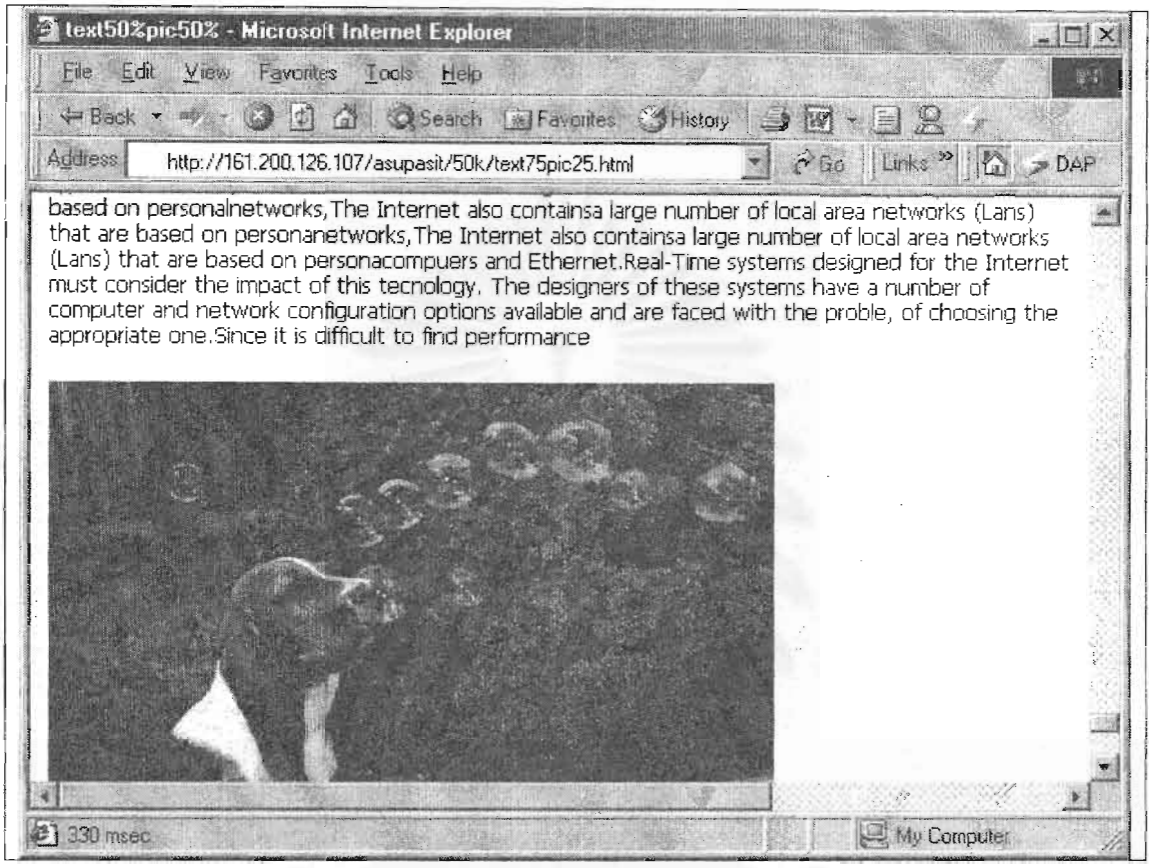

Figure 3.6: Show the delay time of the web page

3.5 Data Collection Method

In the topic, we present method that is applied for this experiment to collect data. The Intranet was simulated in a closed computer laboratory in order to protect the external interference to the delay time. In this research we install the web server at URL name http://161.200.126.107/asupasit/50k/MENU.html which stores various page sizes of web as required in our experiment. The screen of this URL is presented in Figure 3.5.

The experiment has performed as a repeated experiment, with 2700 data values of delay times are measured. For collecting the right time values, each time the retrieval process was completed, the history of visited web page and all files in the temporary Internet files folder such as cookies of IE must be deleted. This is because the history of each browser recorded URL names of web sites and the temporary Internet files collected web pages of web sites. So without deleting these files, the

second round of retrieval the web page will reduce the delay time. Therefore, in order to obtain the right delay time without effects from the historical records, these files must be deleted.

Not only deleted the historical files, the client system must be restarted to clear the cache. Thus, the effects from the buffer usage will be eliminated. Each of time measurement in this experiment spends time about 8 minutes.

3.6 Data Analysis

3.6.1 Factorial Experimental Design [10]

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 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 complete block design. Analysis of variance for the factorial experiments in Complete Block Design presents the linearly relation between factors and dependent variable in the 95% significant. Moreover, it shows effects cause by any combination of 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 Complete Block Design with r blocks is,

$$y_{ijkmno} = \mu + \tau_i + \beta_j + \gamma_k + \phi_m + \sigma_n + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\gamma\phi)_{im} + (\gamma\sigma)_{in} + (\beta\gamma)_{ik} + (\beta\phi)_{jm} + (\beta\sigma)_{jn} + (\gamma\phi)_{km} + (\gamma\sigma)_{kn} + (\phi\sigma)_{mn} + (\tau\beta\gamma)_{ijk} + (\tau\beta\phi)_{ijm} + (\tau\beta\sigma)_{ijn} + (\tau\gamma\phi)_{ikm} + (\tau\gamma\sigma)_{ikn} + (\tau\phi\sigma)_{imn} + (\beta\gamma\phi)_{jkm} + (\beta\gamma\sigma)_{jkn} + (\beta\phi\sigma)_{jmn} + (\gamma\phi\sigma)_{kmn} + (\tau\beta\gamma\phi)_{ijkm} + (\tau\beta\gamma\sigma)_{ijkn} + (\tau\beta\phi\sigma)_{ijmn} + (\tau\gamma\phi\sigma)_{ikmn} + (\beta\gamma\phi\sigma)_{jkmn} + (\tau\beta\gamma\phi\sigma)_{ijkmn} + \varepsilon_{ijkm}$$

Where

y_{ijkmno} the observation value under size of the web page level i, types of processors level j, types of Browsers level k, size of memory level m, ratio between the text and graphics in the page level n, and the observation value p when

i=1,2,...a, j=1,2,...b, k=1,2,...c, m=1,2,...d, n=1,2,...e, o=1,2,...f.

μ is the general mean of population.

τ_i is the treatment effect of size of the web page level i by $\sum_{i=1}^{a} \tau_i = 0$.

β_j is the treatment effect of types of processors level j by $\sum_{j=1}^{a} \beta_j = 0$.

γ_k is the treatment effect of types of browsers level k by $\sum_{k=1}^{a} \gamma_k = 0$.

ϕ_m is the treatment effect of size of memory level m by $\sum_{m=1}^{a} \phi_m = 0$.

σ_n is the treatment effect of ratio between the text and graphics in the page level n by $\sum_{n=1}^{a} \sigma_n = 0$.

$(\tau\beta)_{ij}$ is the treatment effect of interaction of size of the web page level i and Types of processors level j by $\sum_{i=1}^{a}\sum_{j=1}^{b}(\tau\beta)_{ij} = 0$.

$(\tau\gamma)_{ik}$ is the treatment effect of interaction of size of the web page level i and types of browsers level k by $\sum_{i=1}^{a}\sum_{k=1}^{c}(\tau\gamma)_{ik} = 0$.

$(\tau\phi)_{im}$ is the treatment effect of interaction of size of the web page level i and size of memory level m by $\sum_{i=1}^{a}\sum_{m=1}^{d}(\tau\phi)_{im} = 0$.

$(\tau\sigma)_{in}$ is the treatment effect of interaction of size of the web page level i and ratio between the text and graphics in the page level n by $\sum_{i=1}^{a}\sum_{n=1}^{e}(\tau\sigma)_{in} = 0$.

$(\beta\gamma)_{jk}$ is the treatment effect of interaction of types of processors level j and types of browsers level k by $\sum_{j=1}^{b}\sum_{k=1}^{c}(\beta\gamma)_{jk} = 0$.

$(\beta\phi)_{jm}$ is the treatment effect of interaction of types of processors level j and size of memory level m by $\sum_{j=1}^{b}\sum_{m=1}^{d}(\beta\phi)_{jm} = 0$.

$(\beta\sigma)_{jn}$ is the treatment effect of interaction of types of processors level j and ratio between the text and graphics in the page level n by $\sum_{j=1}^{b}\sum_{n=1}^{e}(\beta\sigma)_{jn}=0$.

$(\gamma\phi)_{km}$ is the treatment effect of interaction of types of browsers level k and size of memory level m by $\sum_{k=1}^{c}\sum_{m=1}^{d}(\gamma\phi)_{km}=0$.

$(\gamma\sigma)_{kn}$ is the treatment effect of interaction of types of browsers level k and ratio between the text and graphics in the page level n by $\sum_{k=1}^{c}\sum_{n=1}^{e}(\gamma\sigma)_{kn}=0$.

$(\phi\sigma)_{mn}$ is the treatment effect of interaction of size of memory level m and ratio between the text and graphics in the page level n by $\sum_{m=1}^{d}\sum_{n=1}^{e}(\phi\sigma)_{mn}=0$.

$(\tau\beta\gamma)_{ijk}$ is the treatment effect of interaction of size of the web page level i, types of processors level j, and types of browsers level k by $\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}(\tau\beta\gamma)_{ijk}$.

$(\tau\beta\phi)_{ijm}$ is the treatment effect of interaction of size of the web page level i, types of processors level j, and size of memory level m by $\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{m=1}^{d}(\tau\beta\phi)_{ijm}=0$.

$(\tau\beta\sigma)_{ijn}$ is the treatment effect of interaction of size of the web page level i,types of processors level j, and ratio between the text and graphics in the page level n by $\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{n=1}^{e}(\tau\beta\sigma)_{jn}=0$.

$(\tau\gamma\phi)_{ikm}$ is the treatment effect of interaction of size of the web page level i, types of browsers level k, and size of memory level m by $\sum_{i=1}^{a}\sum_{k=1}^{c}\sum_{m=1}^{d}(\tau\gamma\phi)_{ikm}=0$.

$(\tau\gamma\sigma)_{ikn}$ is the treatment effect of interaction of size of the web page level i, types of browsers level k, and ratio between the text and graphics in the page level n by $\sum_{i=1}^{a}\sum_{k=1}^{c}\sum_{n=1}^{e}(\tau\gamma\sigma)_{ikn}=0$.

$(\tau\phi\sigma)_{imn}$ is the treatment effect of interaction of size of the web page level i, size of memory level m, and ratio between the text and graphics in the page level n by $\sum_{i=1}^{b}\sum_{m=1}^{d}\sum_{n=1}^{e}(\tau\phi\sigma)_{imn}=0$.

$(\beta\gamma\phi)_{jkm}$ is the treatment effect of interaction of types of processors level j, types of browsers level k, and size of memory level m by $\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{m=1}^{d}(\beta\gamma\phi)_{jkm}=0$.

$(\beta\gamma\sigma)_{jkn}$ is the treatment effect of interaction of types of processors level j, types of browsers level k, and ratio between the text and graphics in the page level n by $\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{n=1}^{e}(\beta\gamma\sigma)_{jkn}=0$.

$(\beta\phi\sigma)_{jmn}$ is the treatment effect of interaction of types of processors level j, size of memory level m, and ratio between the text and graphics in the page level n by $\sum_{j=1}^{b}\sum_{m=1}^{d}\sum_{n=1}^{e}(\beta\phi\sigma)_{jmn}=0$.

$(\gamma\phi\sigma)_{kmn}$ is the treatment effect of interaction of types of browsers level k, size of memory level m and ratio between the text and graphics in the page level n by $\sum_{k=1}^{c}\sum_{m=1}^{d}\sum_{n=1}^{e}(\gamma\phi\sigma)_{kmn}=0$.

$(\tau\beta\gamma\phi)_{ijkm}$ is the treatment effect of interaction of size of the web page level i, types of processors level j, types of browsers level k, and size of memory level m by $\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{m=1}^{d}(\tau\beta\gamma\phi)_{ijkm}=0$.

$(\tau\beta\gamma\sigma)_{ijkn}$ is the treatment effect of interaction of size of the web page level i, types of processors level j, types of browsers level k, and ratio between the text and graphics in the page level n by $\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{n=1}^{e}(\tau\beta\gamma\sigma)_{ijkn}=0$.

$(\tau\beta\phi\sigma)_{ijmn}$ is the treatment effect of interaction of size of the web page level i, types of processors level j, size of memory level m, and ratio between the text and graphics in the page level n by $\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{m=1}^{d}\sum_{n=1}^{e}(\tau\beta\phi\sigma)_{ijmn}=0$.

$(\tau\gamma\phi\sigma)_{ikmn}$ is the treatment effect of interaction of size of the web page level i, types of browsers level k, size of memory level m and ratio between the text and graphics in the page level n by $\sum_{i=1}^{a}\sum_{k=1}^{c}\sum_{m=1}^{d}\sum_{n=1}^{e}(\tau\gamma\phi\sigma)_{ikmn}=0$.

$(\beta\gamma\phi\sigma)_{jkmn}$ is the treatment effect of interaction of types of processors level j, types of browsers level k, size of

memory level m and ratio between the text and graphics in the page level n by $\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{m=1}^{d}\sum_{n=1}^{e}(\beta\gamma\phi\sigma)_{jkmn}=0$.

$(\tau\beta\gamma\phi\sigma)_{ijkmn}$ is the treatment effect of interaction of size of the web page level i, types of processors level j, types of browsers level k, size of memory level m and ratio between the text and graphics in the page level n by $\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{m=1}^{d}\sum_{n=1}^{e}(\tau\beta\gamma\phi\sigma)_{ijkmn}=0$.

ε_{ijkmno} is the experimental error with normal distribution.

The block effect ρ_k represents the average deviation of the units in block k from general mean.

3.6.2 Multiple Regression and Correlation [10]

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

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

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,

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

$i=1,2\ldots r,\ j=1,2,\ldots c$

r is the frequency in rows; c are frequency in columns; n are number of data $=\sum r_i=\sum c_j$.

Correlation coefficient measures how variables are related. Before calculating a correlation coefficient, outliners must have been removed. Pearson's correlation coefficient is a measure of linear association only. Two variables can be perfectly related (Pearson-r → 1), but if the relationship is not linear, Pearson's correlation coefficient is not an appropriate statistic for measuring their association. Therefore, the scatter plot of two variables should be plotted before calculating the Pearson correlation coefficient to ensure the relation model between two variables.

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. Figure 3.9 is the scatter plots that show relation between two variables, (x, y), based on Pearson-r value.

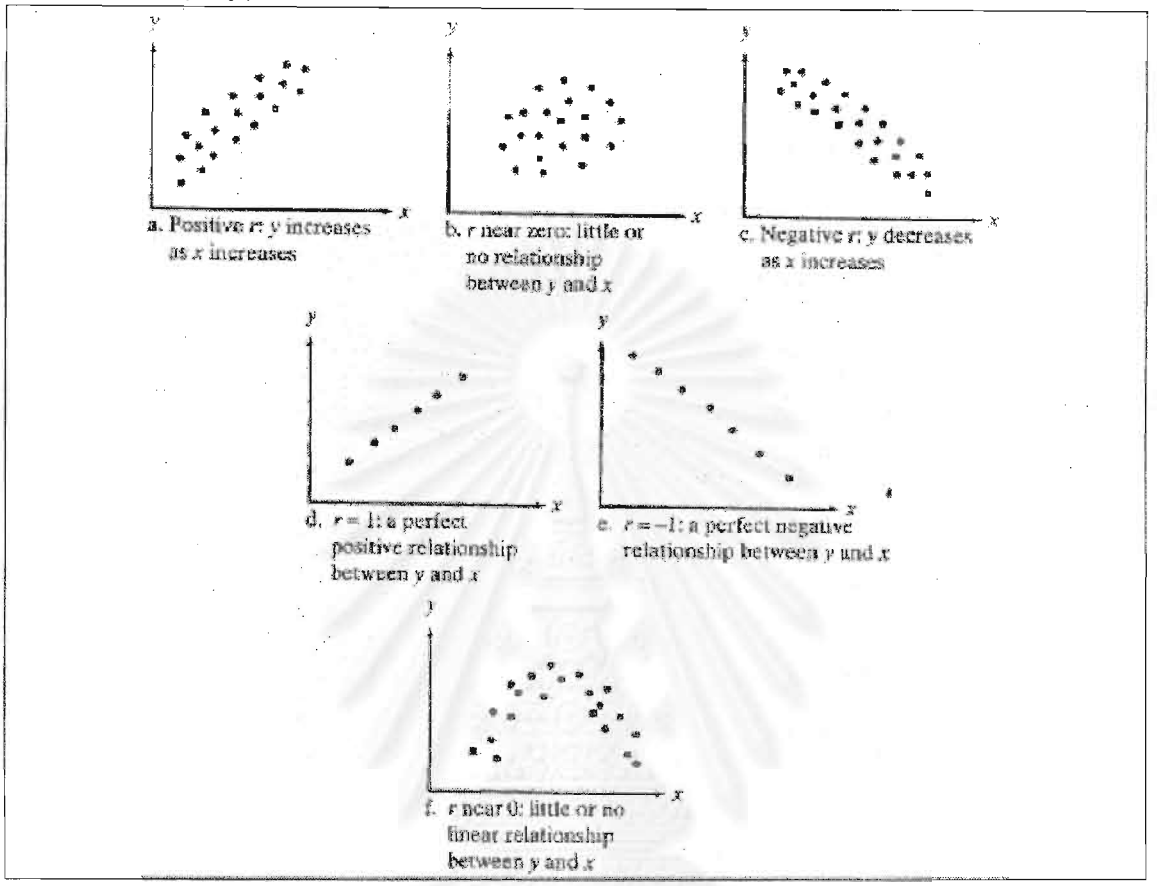

Figure 3.7: Value of r and their implications

3.6.3 Curve estimate [11]

The method choose data graphically to determine how the independent and dependent variables are related (linearly, quadratic, etc.). The 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.

Choose one or more curve estimation regression models. To determine which model to use, plot data. If variables appear to be related linearly, use a simple linear regression model. When variables are not linearly related, the transforming data is the useful step. The following

models are available in the Curve Estimation procedure: linear, logarithmic, inverse, quadratic, cubic, power, compound, S-curve, logistic, growth, and exponential. If the data set does not exactly represent which model is suitable. More curve estimation with several models will help us to make a decision in which model is the most approachable.

3.6.4 Multiple regression [10],[11]

The multiple regression model is of the form

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 (x_1 \times x_2) + \beta_7 (x_1 \times x_3)$$
$$+ \beta_8 (x_1 \times x_4) + \beta_9 (x_1 \times x_5) + \beta_{10} (x_2 \times x_3) + \beta_{11} (x_2 \times x_4) + \beta_{12} (x_2 \times x_5)$$
$$+ \beta_{13} (x_3 \times x_4) + \beta_{14} (x_3 \times x_5) + \beta_{15} (x_4 \times x_5) + \beta_{16} (x_1 \times x_2 \times x_3) + \beta_{17} (x_1 \times x_2 \times x_4)$$
$$+ \beta_{18} (x_1 \times x_2 \times x_5) + \beta_{19} (x_1 \times x_3 \times x_4) + \beta_{20} (x_1 \times x_3 \times x_5) + \beta_{21} (x_1 \times x_4 \times x_5)$$
$$+ \ldots + + \beta_k (x_1 \times x_2 \times x_5) + \beta_{k+1} (x_1 \times x_2 \times x_3 \times x_4) + \beta_{k+2} (x_1 \times x_2 \times x_3 \times x_5) + \ldots +$$
$$+ \beta_l (x_2 \times x_3 \times x_4 \times x_5) + \beta_{l+1} (x_1 \times x_2 \times x_3 \times x_4 \times x_5)$$

Where

y is the response variable,

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

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

ε is a random error component.

Since $\beta_0, \beta_1, \ldots, \beta_k$ and $x_0, x_1, \ldots, x_k$ are nonrandom

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 each independent variable should be linear, and all observations should be independent.

Let's R-square be the coefficient of determination. This value indicates correctness of the regression linear model derived from the raw data. For example, if R-square = 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-Square will be used to predict the correctness of the model. The adjusted R-Square obtained from adjusting the degrees of freedom of the model analysis.

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

Where

DF_T is Degree of Freedom of Total Sum Square,

DF_E is Degree of Freedom of Error Sum Square.

3.6.5 Linear Regression Variable Selection Methods[11]

The two methods of linear regression, stepwise method and enter method, are used in our regression analysis. The major method in this research, stepwise linear regression, is used to evaluate the most effective model whereas enter linear regression is used to verify the result model with the new data set.

Stepwise linear regression method is a method that chooses independent variables in the model from consideration of the partial correlation that derived within the processes. This method will eliminate the interference of the selected variables before choosing the new independent variable into the regression model.

3.6.6 Nonlinear regression[11]

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 by transforming the original data set to be new data set. These transformed data can be analyzed using the Linear Regression procedure.

CHAPTER IV

RESULTS

This chapter described results of our experiment, an obtained general model to forecast the delay time under all expected independent variables, and testing result using the derived model with testing data set.

4.1 ANOVA of Factorial Experimental Design

The hypothesis to test the main effects from all factors in our experiment is to perform the ANOVA with significant level equal 0.05. Our factors are size of the web page, types of Browsers, size of memory, types of processors, and ratio between the text and graphics in the page. Therefore, our hypothesises are listed below.

The hypothesises of the analysis are

H_{01}: Mean values of the delay times obtained from different size of the web page are not significant different

H_{02}: Mean values of the delay times obtained from different types of Browsers are not significant different

H_{03}: Mean values of the delay times obtained from different size of memory are not significant different

H_{04}: Mean values of the delay times obtained from different types of processors are not significant different

H_{05}: Mean values of the delay times obtained from different ratio between the text and graphics in the page are not significant different

H_{06}: Mean values of the delay times obtained from different size of the web page and types of Browsers are not significant different

H_{07}: Mean values of the delay times obtained from different size of the web page and size of memory are not significant different

H_{08}: Mean values of the delay times obtained from different size of the web page and types of processors are not significant different

H_{09}: Mean values of the delay times obtained from different size of the web page and ratio between the text and graphics in the page are not significant different

H_{010}: Mean values of the delay times obtained from different types of Browsers and size of memory are not significant different

H_{011}: Mean values of the delay times obtained from different types of Browsers and types of processors are not significant different

H_{012}: Mean values of the delay times obtained from different types of Browsers and ratio between the text and graphics in the page are not significant different

H_{013}: Mean values of the delay times obtained from different size of memory and types of processors are not significant different

H_{014}: Mean values of the delay times obtained from different size of memory and ratio between the text and graphics in the page are not significant different

H_{015}: Mean values of the delay times obtained from different types of processors and web types are not significant different

H_{016}: Mean values of the delay times obtained from different size of the web page, types of Browsers, and size of memory are not significant different

H_{017}: Mean values of the delay times obtained from different size of the web page, types of Browsers, and types of processors are not significant different

H_{018}: Mean values of the delay times obtained from different size of the web page, types of Browsers, and ratio between the text and graphics in the page are not significant different

H_{019}: Mean values of the delay times obtained from different size of the web page, size of memory, and types of processors are not significant different

H_{020}: Mean values of the delay times obtained from different size of the web page, size of memory, and ratio between the text and graphics in the page are not significant different

H_{021}: Mean values of the delay times obtained from different size of the web page, types of processors, and ratio between the text and graphics in the page are not significant different

H_{022}: Mean values of the delay times obtained from different types of Browsers, size of memory, and types of processors are not significant different

H_{023}: Mean values of the delay times obtained from different types of Browsers, size of memory, and ratio between the text and graphics in the page are not significant different

H_{024}: Mean values of the delay times obtained from different types of Browsers, types of processors, and ratio between the text and graphics in the page are not significant different

H_{025}: Mean values of the delay times obtained from different size of memory, types of processors, and ratio between the text and graphics in the page are not significant different

H_{026}: Mean values of the delay times obtained from different size of the web page, types of Browsers, size of memory, and types of processors are not significant different

H_{027}: Mean values of the delay times obtained from different size of the web page, types of Browsers, and size of memory and ratio between the text and graphics in the page are not significant different

H_{028}: Mean values of the delay times obtained from different size of the web page, types of Browsers, and types of processors and ratio between the text and graphics in the page are not significant different

H_{029}: Mean values of the delay times obtained from different size of the web page, size of memory, and types of processors and ratio between the text and graphics in the page are not significant different

H_{030}: Mean values of the delay times obtained from different types of Browsers, size of memory, and types of processors and ratio between the text and graphics in the page are not significant different

H_{031}: Mean values of the delay times obtained from different size of the web page, types of processors, types of Browsers, size of memory, and ratio between the text and graphics in the page are not significant different

From the ANOVA testing, the conclusions to determine main effects and interaction effects among factors are summarized as follows.

There is at least one mean value of the delay timed obtained from a web size is significant different from other mean values from other size of the web page, with p=0.004 < 0.05.

There is at least one mean value of the delay times obtained a size of the web page and a ratio between the text and graphics in the page is significant different from other mean values obtained from other size of the web page and ratio between the text and graphics in the page, with p=0.007 < 0.05. This means that there is an interaction effect between size of the web page and ratio between the text and graphics in the page.

There is at least one mean value of the delay times obtained from a size of the web page, a types of Browser, and a types of processors that is significant different from other mean values obtained from other size of the web page, types of Browsers, and types of processors, with p=0.007 < 0.05. This means that there is an interaction effect causes by size of the web page, types of Browsers, and types of processors.

There is at least one mean value of the delay times obtained from a size of the web page, a size of memory, and a types of processor is significant different from other mean values obtained from other size of the web page, size of memory, and types of processors, with p=0.001 < 0.05. This means that there is an interaction effect causes by size of the web page, size of memory, and types of processors.

There is at least one mean value of the delay times obtained from a web size, size of memory, and ratio between the text and graphics in the page is significant different from other mean values obtained from other size of the web page, size of memory, and ratio between the text and graphics in the page, with p=0.038 < 0.05. This means that there is an interaction effect causes by size of the web page, size of memory, and ratio between the text and graphics in the page.

There is at least one mean value of the delay times obtained from a types of Browser, a size of memory, and a types of processor is significant different from other mean values obtained from types of Browsers, size of memory, and types of processors, with p=0.00 < 0.05. This means that there is an interaction effect causes by types of Browsers, size of memory, and types of processors.

4.2 Correlation Choosing Related Factors

The correlation is used to measure the relationship between two variables, independent variable and dependent variable. The Measurement of correlation coefficient in this analysis is Pearson coefficient of correlation.

Let we define the variables used in our analysis program.

Size of the web page is *FS* (Unit: Bytes; 50 KB, 100 KB, 200 KB.)

Types of processors is *CP* (Unit: MHz.; 350MHz, 400MHz, 733 MHz.)

Types of browsers will be interpreted from two dummy variables: *NS* and *OP* as presented below.

- If the experiment used Netscape Navigator, the value of *NS* = 1; otherwise the value of *NS* = 0.
- If the experiment used Opera, the value of *OP* = 1; otherwise *OP* = 0.
- If the experiment used Internet Explorer, the values of both *NS* and *OP* are equal 0.

Size of memory is *RM* (Unit: MB.; 32 MB, 64 MB.)

Ratio between the text and graphics in the page is *TX* (Unit: Percentage of Text in a file; 100, 75, 50, 25, 0.).

Table 4.1: No relationship among independent variables

		BR	CP	FS	RM	TX
BR	Pearson Correlation	1.000	.000	.000	.000	.000
	Sig. (2-tailed)	.	1.000	1.000	1.000	1.000
	N	1350	1350	1350	1350	1350
CP	Pearson Correlation	.000	1.000	.000	.000	.000
	Sig. (2-tailed)	1.000	.	1.000	1.000	1.000
	N	1350	1350	1350	1350	1350
FS	Pearson Correlation	.000	.000	1.000	.000	.000
	Sig. (2-tailed)	1.000	1.000	.	1.000	1.000
	N	1350	1350	1350	1350	1350
RM	Pearson Correlation	.000	.000	.000	1.000	.000
	Sig. (2-tailed)	1.000	1.000	1.000	.	1.000
	N	1350	1350	1350	1350	1350
TX	Pearson Correlation	.000	.000	.000	.000	1.000
	Sig. (2-tailed)	1.000	1.000	1.000	1.000	.
	N	1350	1350	1350	1350	1350

Table 4.1 shows that the delay time is affected by all independent variables according to the value of coefficient of correlations. Therefore we can define that the dependent variable of our multiple regression linear model is the delay time, where the independent variables of our multiple regression linear model are

1. Size of the web page
2. Types of processors
3. Types of Browsers
4. Size of memory
5. Ratio between the text and graphics in the web page

Table 4.2: Relationship among independent variables

		TTIME	BR	CP	FS	RM	TX
TTIME	Pearson Correlation	1.000	-.104**	-.281**	.597**	-.252**	-.171**
	Sig. (2-tailed)	.	.000	.000	.000	.000	.000
	N	1350	1350	1350	1350	1350	1350
BR	Pearson Correlation	-.104**	1.000	.000	.000	.000	.000
	Sig. (2-tailed)	.000	.	1.000	1.000	1.000	1.000
	N	1350	1350	1350	1350	1350	1350
CP	Pearson Correlation	-.281**	.000	1.000	.000	.000	.000
	Sig. (2-tailed)	.000	1.000	.	1.000	1.000	1.000
	N	1350	1350	1350	1350	1350	1350
FS	Pearson Correlation	.597**	.000	.000	1.000	.000	.000
	Sig. (2-tailed)	.000	1.000	1.000	.	1.000	1.000
	N	1350	1350	1350	1350	1350	1350
RM	Pearson Correlation	-.252**	.000	.000	.000	1.000	.000
	Sig. (2-tailed)	.000	1.000	1.000	1.000	.	1.000
	N	1350	1350	1350	1350	1350	1350
TX	Pearson Correlation	-.171**	.000	.000	.000	.000	1.000
	Sig. (2-tailed)	.000	1.000	1.000	1.000	1.000	.
	N	1350	1350	1350	1350	1350	1350

**. Correlation is significant at the 0.01 level (2-tailed).

From Table 4.2 we can interpret the difference in the delay time is much smaller if speed of processors increase. By comparing size of files, we conclude that the difference in the delay is much smaller if the large file sizes are used. The delay time of size of memory decrease if large size of memory and the delay time of ratio between the text and graphics in the page decrease if ration of text less than ration of graphics.

4.3 Choosing the independent variable

The multiple regression model is of the from

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \mathrm{K} + \beta_k x_k$

Where

y is the delay time

$\beta_0, \beta_1, \mathrm{K}, \beta_k$ $x_0, x_1, \mathrm{K}, x_k$ are the unstandardized coefficients

$x_0, x_1, \mathrm{K}, x_k$ are the set of independent variables

To consider the degree of each variable in the model, we consider the scatter plot and applied the results from curve estimate methods to predict and describe the degree of variables, as shown in Figure 4.1-Figure 4.5.

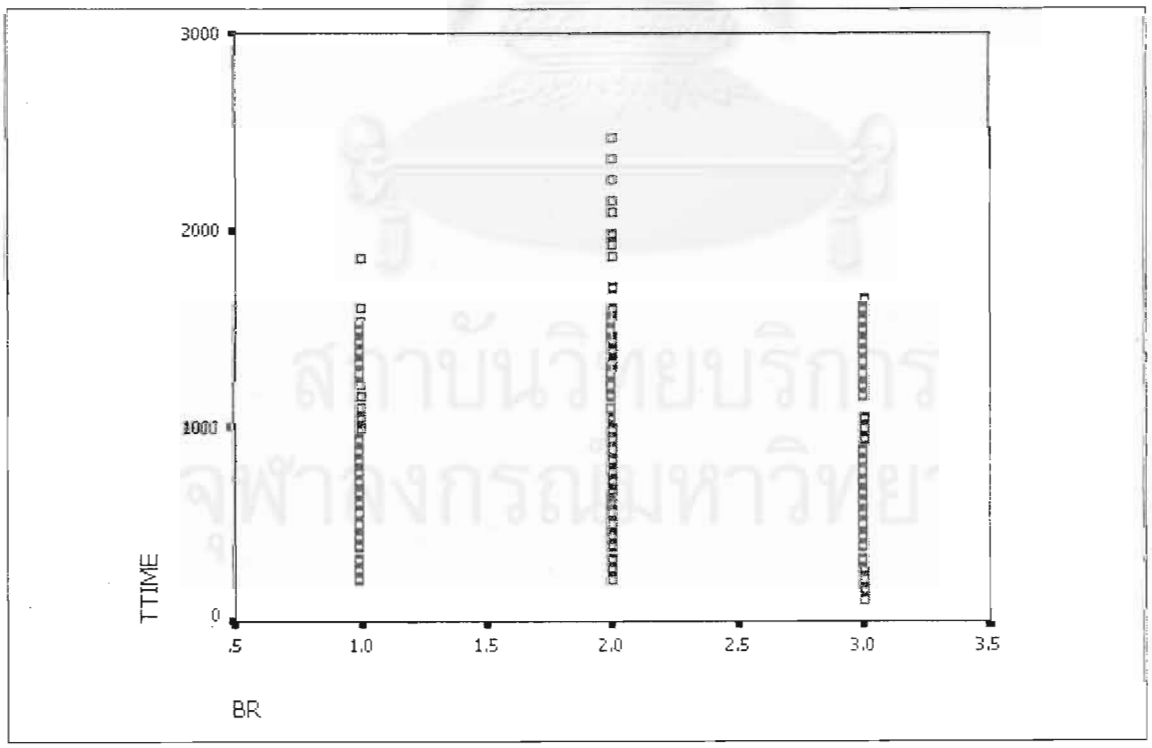

Figure 4.1: The scatter plot of type of browser

Figure 4.1 show that type of browser assigned code number 1 is Internet Explorer 5.0, code number 2 is Netscape Navigator 4.7, and code

number 3 is Opera 5.0. By comparing, we conclude there is no significant different in the delay time of Internet Explorer and Opera. On the other hands, there is a significant different of the delay time of Netscape Navigator comparing with other browsers. The results show that the delay time obtained from Netscape Navigator is larger than other browsers.

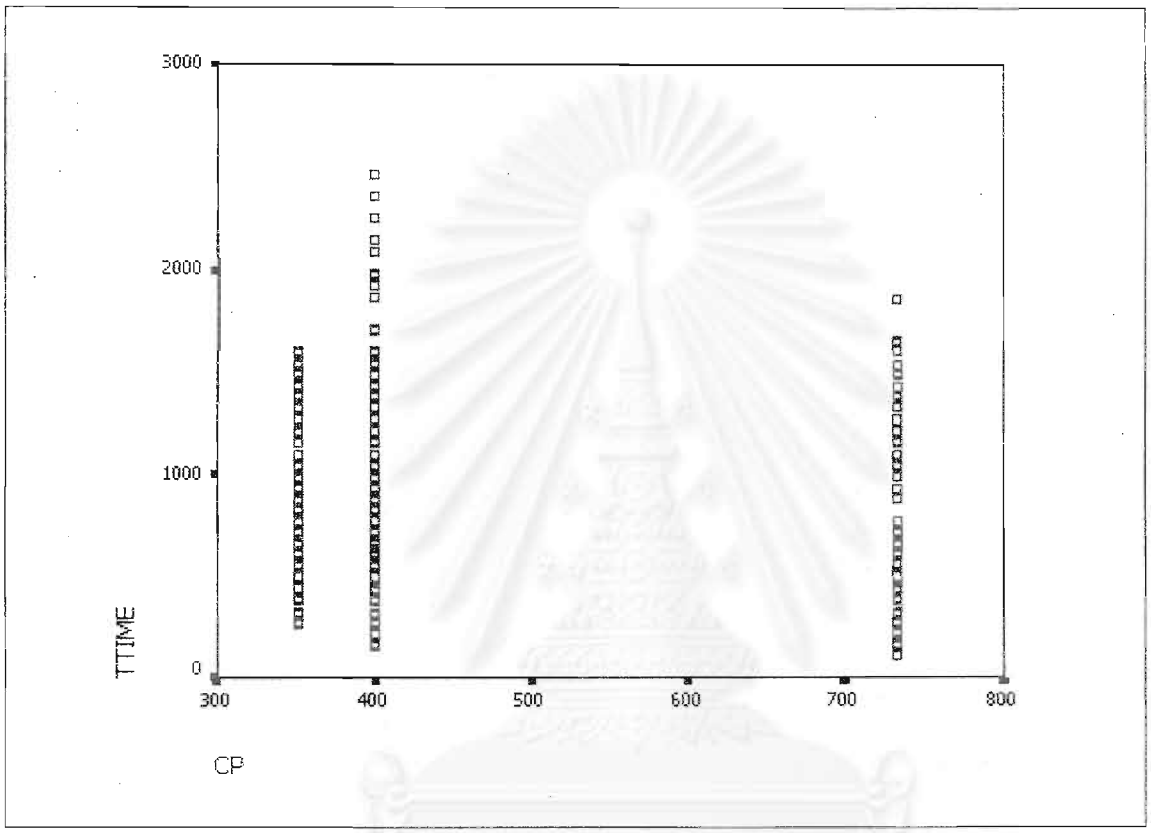

Figure 4.2: The scatter plot of type of processor

Figure 4.2 compares three Processors, using Pentium II at 350 MHz, Pentium II at 400 MHz, and Pentium III at 733 MHz. In this plot, it does not clearly indicate the affect of high performance CPU to the delay time.

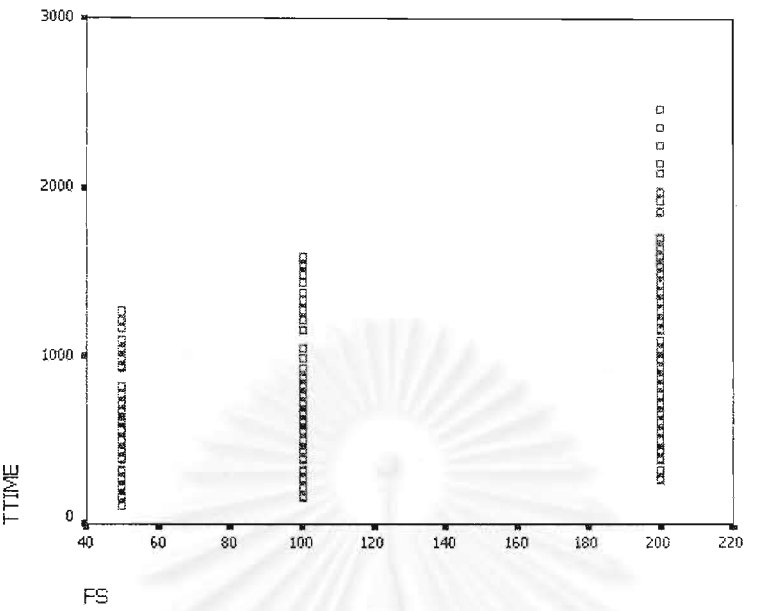

Figure 4.3: The scatter plot of size of web page

The comparison of three different sizes of web pages is presented in Figure 4.3. It is clearly seen that the delay time increases if the sizes of the web increased.

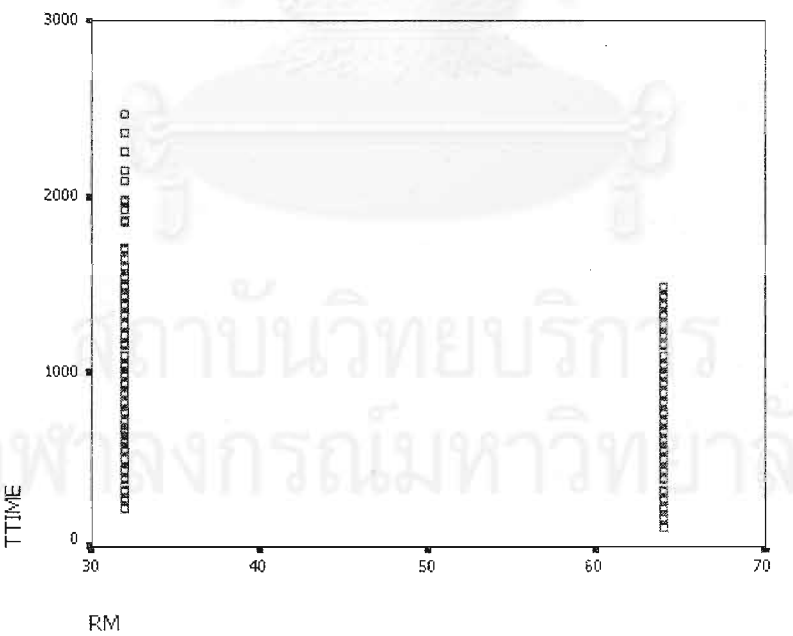

Figure 4.4: The scatter plot of size of memory

Figure 4.4 shows that using a large memory will reduce the delay time.

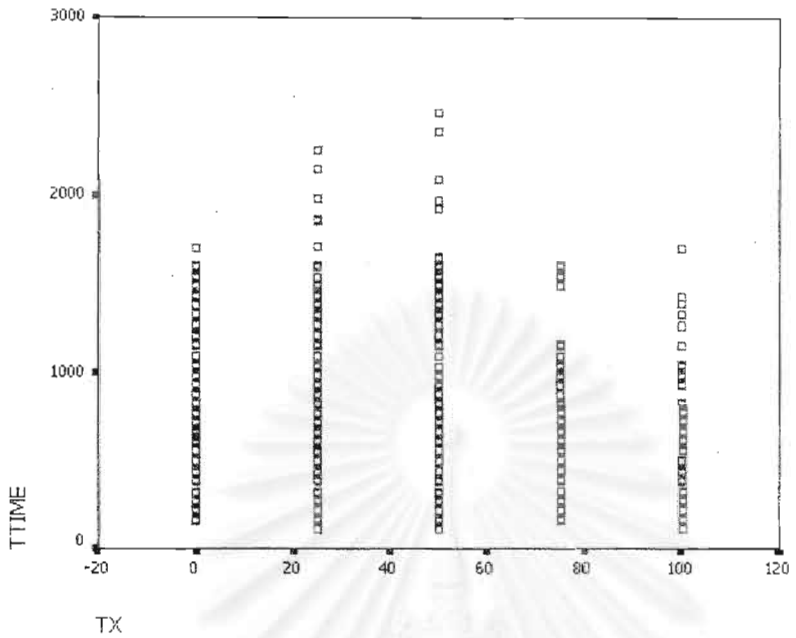

Figure 4.5: The scatter plot of ratio of text and graphic in web page

The comparisons of five different proportions of text and graphics in a web page are given in Figure 4.5. We conclude that the delay times of pure text files or pure graphic files are small comparing with other proportions of text and graphics. Additionally, 50% text and graphic files, as presented in Figure 4.5 cause the slowest delay time.

According to the scatter plots, the expected regression model of the delay time is as follows.

$$TIME = f(NS, OP, CP, FS, RM, TX)$$

$$TIME = b_1 NS + b_2 OP + b_3 CP + b_4 FS + b_5 RM + b_6 CP^2 + b_7 TX^3 + b_8 (OP \quad FS) + b_9 (NS \quad TX^3) + b_{10} (CP \quad FS) + b_{11} (CP^2 \quad FS) + b_{12} (RM \quad FS) + b_{13} (RM \quad TX^3) + +b_{14} (FS \quad TX^3) + b_{15} (NS \quad FS) + b_{16} (NS \quad CP^2)$$

Where

$TIME$ is the delay time.

b_1 is the coefficient of Netscape.

b_2 is the coefficient of Opera.

b_3 is the coefficient of type of processor.

b_4 is the coefficient of size of web page.

b_5 is the coefficient of size of memory.

b_6 is the coefficient of square of type of processor.

b_7 is the coefficient of power three of ratio of text and graphic in web page.

b_8 is the coefficient of Opera and size of web page.

b_9 is the coefficient of Netscape and power three of ratio of text and graphic in web page.

b_{10} is the coefficient of type of processor and size of web page.

b_{11} is the coefficient of square of type of processor and size of web page

b_{12} is the coefficient of size of memory and size of web page.

b_{13} is the coefficient of size of memory and power three of ratio of text and graphic in web page.

b_{14} is the coefficient of size of web page and power three of ratio of text and graphic in web page.

b_{15} is the coefficient of Netscape and size of web page.

b_{16} is the coefficient of Netscape and square of type of processor.

Using the 1350 data and the stepwise procedure under the significant is 0.05, the constant and coefficient of variables are shown in Table 4.3

Table 4.3: the constant, coefficient of variables, standard error, and asymptotic confidence interval

No.	Model	Estimate	Standard Error	Lower	Upper
1	NS	190.69121	32.61899	126.70126	254.68115
2	OP	-202.14521	26.90588	-254.92767	-149.36276
3	CP	1.66656	0.18012	1.31319	2.01992
4	FS	11.36646	1.48444	8.45436	14.27857
5	RM	-2.91642	0.73512	-4.35855	-1.47429
6	CP^2	-0.00162	0.00019	-0.00200	-0.00124
7	TX^3	-0.00011	0.00005	-0.00021	-0.00001
8	OP×FS	0.93538	0.20341	0.53633	1.33442
9	NS×TX^3	0.00018	0.00002	0.00012	0.00023
10	CP×FS	-0.02149	0.00609	-0.03345	-0.00953
11	CP^2×FS	0.00001	5.5740E-06	6.6636E-06	0.00002
12	RM×FS	-0.03578	0.00517	-0.04594	-0.25627
13	RM×TX^3	4.2857E-06	8.6154E-07	2.5956E-06	5.9588E-06
14	FS×TX^3	-3.0283E-06	2.2111E-07	-3.4621E-06	-2.5945E-06
15	NS×FS	0.83047	0.20341	0.43100	1.22909
16	NS×CP^2	-0.00063	0.00005	-0.00074	-0.00051

This multiple linear regression model has R^2= 0.728, this means the prediction of the delay time using this model after knowing all defined independent variables has 72.8% accuracy. Therefore, we conclude that the multiple regression is a reliable model.

4.4 Testing the multiple regression

Applying the multiple regression model to the testing data set with 1350 data. The multiple regression model is

$$TIME = b_1 NS + b_2 OP + b_3 CP + b_4 FS + b_5 RM + b_6 CP^2 + b_7 TX^3 + b_8 (OP \times FS) + b_9 (NS \times TX^3) + b_{10} (CP \times FS) + b_{11} (CP^2 \times FS) + b_{12} (RM \times FS) + b_{13} (RM \times TX^3) + +b_{14} (FS \times TX^3) + b_{15} (NS \times FS) + b_{16} (NS \times CP^2)$$

Where

$TIME$ is the delay time.

b_1 is the coefficient of Netscape.

b_2 is the coefficient of Opera.

b_3 is the coefficient of type of processor.

b_4 is the coefficient of size of web page.

b_5 is the coefficient of size of memory.

b_6 is the coefficient of square of type of processor.

b_7 is the coefficient of power three of ratio of text and graphic in web page.

b_8 is the coefficient of Opera and size of web page.

b_9 is the coefficient of Netscape and power three of ratio of text and graphic in web page.

b_{10} is the coefficient of type of processor and size of web page.

b_{11} is the coefficient of square of type of processor and size of web page

b_{12} is the coefficient of size of memory and size of web page.

b_{13} is the coefficient of size of memory and power three of ratio of text and graphic in web page.

b_{14} is the coefficient of size of web page and power three of ratio of text and graphic in web page.

b_{15} is the coefficient of Netscape and size of web page.

b_{16} is the coefficient of Netscape and square of type of processor.

Table 4.4: The constant, coefficient of each variable of testing, standard error, and asymptotic confidence interval

No.	Model	Estimate	Standard Error	Lower	Upper
1	NS	173.33164	33.94878	106.73281	239.93046
2	OP	-212.46812	28.00284	-267.40253	-157.53371
3	CP	1.70352	0.18747	1.33575	2.07129
4	FS	9.11876	1.54496	6.08793	12.14959
5	RM	-2.70958	0.76509	-4.21051	-1.20865
6	CP^2	-0.00162	0.00020	-0.00202	-0.00123
7	TX^3	-0.00013	0.00005	-0.00023	-0.00002
8	OP×FS	0.99264	0.21170	0.57733	1.40795
9	NS×TX^3	0.00018	0.00003	0.00012	0.00024
10	CP×FS	-0.01221	0.00634	-0.02467	0.00023
11	CP^2×FS	8.9702E-06	5.8012E-06	-2.4103E-06	0.00002
12	RM×FS	-0.03820	0.00539	-0.04878	-0.02763
13	RM×TX^3	4.1592E-06	8.9667E-07	2.4002E-06	5.9183E-06
14	FS×TX^3	-2.8584E-06	2.3012E-07	-3.3098E-06	-2.4069E-06
15	NS×FS	1.05073	0.21170	0.63542	1.46605
16	NS×CP^2	-0.00063	0.00006	-0.00075	-0.000511

Using the coefficient values in Table 4.4, we obtained $R^2 = 0.716$ means that the model is acceptable.

CHAPTER V

CONCLUSION AND DISCUSSION

5.1 Conclusion

From the experiment, the multiple regression model of the delay time over HTTP environment is presented as follow

$$TIME = f(NS, OP, CP, FS, RM, TX)$$

$$TIME = b_1 NS + b_2 OP + b_3 CP + b_4 FS + b_5 RM + b_6 CP^2 + b_7 TX^3 + b_8 (OP \times FS) + b_9 (NS \times TX^3) + b_{10} (CP \times FS) + b_{11} (CP^2 \times FS) + b_{12} (RM \times FS) + b_{13} (RM \times TX^3) + +b_{14} (FS \times TX^3) + b_{15} (NS \times FS) + b_{16} (NS \times CP^2)$$

Where

$TIME$ is the delay time.

b_1 is the coefficient of Netscape.

b_2 is the coefficient of Opera.

b_3 is the coefficient of type of processor.

b_4 is the coefficient of size of web page.

b_5 is the coefficient of size of memory.

b_6 is the coefficient of square of type of processor.

b_7 is the coefficient of power three of ratio of text and graphic in web page.

b_8 is the coefficient of Opera and size of web page.

b_9 is the coefficient of Netscape and power three of ratio of text and graphic in web page.

b_{10} is the coefficient of type of processor and size of web page.

b_{11} is the coefficient of square of type of processor and size of web page

b_{12} is the coefficient of size of memory and size of web page.

b_{13} is the coefficient of size of memory and power three of ratio of text and graphic in web page.

b_{14} is the coefficient of size of web page and power three of ratio of text and graphic in web page.

b_{15} is the coefficient of Netscape and size of web page.

b_{16} is the coefficient of Netscape and square of type of processor.

From the experimental and testing model presented in the previous Chapter, we can see that the coefficient of variables can be considered as shown in Table 5.1.

Table 5.1: Coefficients of variables

Positive values	Negative values
NS	OP
CP	RM
FS	CP^2
Op×FS	TX^3
$NS \times TX^3$	CP×FS
$CP^2 \times FS$	RM×FS
$RM \times TX^3$	$FS \times TX^3$
NS×FS	$NS \times CP^2$

5.2 Discussion

According to the multiple regression model, we can conclude that if the user used Netscape to browse a web page, then the delay time will be higher than using other browsers. Additionally, if the large web page is retrieved, the delay time will be increased.

If a large web page is called from a client using Opera, the delay time will be larger than calling a small web page. However the experiment has shown that increasing the size of memory will not reduce the delay time as requested.

5.3 Suggestions

According to the experiment of this thesis under the close system, the multiple regression model for predicting the delay time over the HTTP environment fairly limited. Therefore to study the real environment, all possible factors that effect to the delay time must be taken in account, such as network topologies, communication protocols, types of graphics (.bmp, .jpg, etc.), and operating system usage in the environment.

5.4 Advantage Of Research

The result of a study of factors related to the delay time over the HTTP environment can be used as a guidance of Intranet system design phase for an organization. Thus, the investment cost and maintenance cost can be limited while the organization will gain the high benefits for this design.

References

1. Virgina A. T. Kaste, Ann E. M. Brodeen, and Bartbara D. Broome, "Empirical Evaluation of Protocol Performance over Tactical Networks": U S Army Ballistic Research Laboratory, Aberdeen Proving Ground, Maryland, IEEE 1992
2. Shang-Juh Kao, "Variations on Ethernet Traffic Monitoring": Department of Allpied Mathematics, National Chung-Heing University, Taichung, Taiwan, China,IEEE 1993
3. Alberto Del Bimbo, member, IEEE, and Enrico Vicario, "Transport Measurements Over an Ethernet LAN": IEEE 1994
4. Kihong Park, Giate Kim, and Mark Crovella, "On the relationship between files sizes, transport protocols, and self-similar network traffic": IEEE 1996
5. John Heidemann, Katia Obraczka, and Joe Touch, "Modeling the Performance of HTTP Over Several Transport Protocols": IEEE/ACM transaction on networking, vol. 5 no. 5, OCTOBER 1997.
6. Andro Milanovic, Sinisa Srbljic, and Josko Radej, "Performance of Distributed System Based on Ethernet and Personal Computers": University of Zagreb, IEEE 1999
7. Toshio Soumiya, Koji Nakamichi, and Arnold Bragg, "Performance Evaluation of TCP over ATM Using World Wide Web Traffic": Fujitsu Laboratories Ltd.,
8. Tai Minh Le, Maurice Castro, "A study of the Effects of Dynamic Content on Web Server Performance": Department of Computer Science, RMIT University, 25 October,1999.
9. S. Lertkijcharoenwong, "Physical Effects of the Network Environment on Web Applications": M. Sc. Thesis, Department of Mathematics, Faculty of Science, Chulalongkorn University
10. Jame T. McClave, Frank H. Deitrich II , and Terry Sincich, "Statistics": Prentice Hall, 7th Edition,1997
11. "SPSS 8.0 for Windows" [computer software]. U.S.A.: SPSS Inc., 1996. Available from: Department of Mathematics, Faculty of Science, Chulalongkorn University [2001, July]

APPENDIX A

RESULT OF THESIS FROM SPSS SOFTWARE

Non-linear Regression

All the derivatives will be calculated numerically.

Iteration	Residual SS	A1	A2	A3	A4
		A5	A6	A7	A8
		A9	A10	A11	A12
		A13	A14	A15	A16
1	1.8290E+19	1.00000000	1.00000000	1.00000000	1.00000000
		1.00000000	1.00000000	1.00000000	1.00000000
		1.00000000	1.00000000	1.00000000	1.00000000
		1.00000000	1.00000000	1.00000000	1.00000000
1.1	50557583.85	118.526278	-126.54345	1.66481154	11.5229311
		-2.9221380	-.00165944	-.00015958	.184380146
		.000321152	-.02142111	.000017518	-.03576706
		.000004269	-.00000304	1.05561338	-.00052644
2	50557583.85	118.526278	-126.54345	1.66481154	11.5229311
		-2.9221380	-.00165944	-.00015958	.184380146
		.000321152	-.02142111	.000017518	-.03576706
		.000004269	-.00000304	1.05561338	-.00052644
2.1	48336523.82	190.691200	-202.14522	1.66656078	11.3664707
		-2.9164258	-.00162636	-.00011274	.935380689
		.000180380	-.02149496	.000017598	-.03578870
		.000004286	-.00000303	.830047561	-.00063042
3	48336523.82	190.691200	-202.14522	1.66656078	11.3664707
		-2.9164258	-.00162636	-.00011274	.935380689
		.000180380	-.02149496	.000017598	-.03578870
		.000004286	-.00000303	.830047561	-.00063042
3.1	48336523.82	190.691210	-202.14522	1.66656062	11.3664696
		-2.9164255	-.00162636	-.00011274	.935380804
		.000180379	-.02149495	.000017598	-.03578870
		.000004286	-.00000303	.830047484	-.00063042

Run stopped after 6 model evaluations and 3 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	16	667565776.183	41722861.0114
Residual	1334	48336523.8170	36234.27572
Uncorrected Total	1350	715902300.000	
(Corrected Total)	1349	178056779.926	

R squared = 1 - Residual SS / Corrected SS = .72853

Parameter	Estimate	Asymptotic Std. Error	Asymptotic 95 % Confidence Interval Lower	Upper
A1	190.69121010	32.618910767	126.70126134	254.68115887
A2	-202.1452155	26.905884537	-254.9276700	-149.3627610
A3	1.666560619	.180126789	1.313197992	2.019923247
A4	11.366469605	1.484446569	8.454365624	14.278573587
A5	-2.916425486	.735126736	-4.358555865	-1.474295107
A6	-.001626357	.000195189	-.002009268	-.001243445
A7	-.000112744	.000051564	-.000213899	-.000011589
A8	.935380804	.203413062	.536336473	1.334425135
A9	.000180379	.000029255	.000122988	.000237771
A10	-.021494954	.006097064	-.033455832	-.009534076
A11	.000017598	5.57401E-06	6.66360E-06	.000028533
A12	-.035788699	.005179603	-.045949754	-.025627645
A13	4.28574E-06	8.61547E-07	2.59561E-06	5.97588E-06
A14	-3.02833E-06	2.21113E-07	-3.46210E-06	-2.59456E-06
A15	.830047484	.203413046	.431003185	1.229091784
A16	-.000630424	.000058665	-.000745509	-.000515340

Vita

Name: Mr. Supasit Apiwattanapong

Birth date: 14/09/76

Place of Birth: Bangkok, Thailand

Academic Background:

1995-1998 B.Sc. (Mathematics) King Mongkut's Institute of technology North Bangkok, Bangkok Thailand

Experiences:

1997-1998 Officer at the Organizational Management, Faculty of Applied Science, King Mongkut's Institute of technology North Bangkok