A Model for Network Performance Analysis in a Case of Transfer a Large Image Files

Veneta P. Aleksieva[1], Peter T. Antonov[2]

Abstract – **In this paper we propose the analytical model for quick determination and prognostic of network performance in case of transfer a large image files. In the creation of this model is used typical fragment of TCP/IP network, which may be expanded. The practical use of this model is based on simulations.**

Keywords- **network performance**

I. Introduction

The company's business today is dependent on computer systems and network technologies, which is accessible by Internet. Most of them dislocate your applications, and your earnings are depended on Web-sites, local networks and client-server applications.

II. Presentation

The aim of this elaboration is to develop a mathematical model, with which it have to prognosticate the performance of client-server network to detect and discard (if it is possible) bottleneck of the network and to achieve the desired admissible response time from server. The resulting analytical expressions may use as tools for prediction analysis when every network project is created, it doesn't matter the size or the application area. So, when the network architecture is created, it will be laid in advance the customers requirements and will be created the most proper network structure from view of future management and performance.

The network performance is base factor, which by bad parameters causes the productivity fall, the financial lose, the sale drop, the change for the worse company image. Because of these facts is important to find and discard the reasons for network delay and reach to acceptable quality of individual customer services. Many authors make researches in this direction [1],[2],[3],[4], it is published the patent software as tools for analysis of built working network[5].

In this paper is presented an analytical model, with which may be planned change of time for treatment of request for server in the network to improve the network performance when it has average or huge loads. Usually it is presented a quantity approach of performance measurement. In [1] is presented an experimental results about network delay in shared media, and with them it may predict approximately the network performance. In [2] and [4] it is presented graphics on experimental results for different modification of TCP-protocols in comparison and it is rated their influence on performance in the same network. In [3] is presented an analytical method for network performance analysis, which is defined from view of optimization of Web-server application. In this paper is presented a development of this method with reading of the following important aspects, which cause necessity of details network performance analysis:

1. a stage of a readiness for servicing of incoming requests to server, which come in an accidental moment of time ;
2. a difficult prognoses network load in the certain moment;
3. a quality of service on these requests
4. a permanently need of network expanding, if throughput is kept a constant
5. an insurance of acceptable quality of customer services, which is defined from:
    a. a subjective estimate of research-worker, which influence on the results of quality research;
    b. the used method for connection's diagnostic;
    c. the feedback is formed from the customers, and is presented distinctive lines of customers, which indicate determinate individual characteristics and allow ranging of research people based on group from comparable persons or an ideal.
6. insurance of high profitableness, determinate as ratio on expense to performance

Examining fragment of network structure is presented on figure 1, where: on PC2 are located image files in JPEG format with different size (from 4KB to 6345KB); to hub consequently are attached 1, 2, 3 and 4 clients, with same characteristics – pentiumIV, 2.4GHz, 256MB RAM DDR1, Windows XP Proffesional, ver. 2002, service pack 1, browser Internet Explorer 6.0. They simultaneously visualized the same image, which is downloaded through the network from PC2. The network delays are written with the program Ethereal. It is a typical network fragment, which may be extended, and the results are applicable to any other network.

In common case the delays of a web transaction are result from:

1. search in DNS
2. time to setup TCP
3. time to treatment from the server
4. time to delivering through the network
5. time to visualization, which is influenced from type of source code
6. the size of cache on the local browser
7. the type of information
8. the number of clients, which request the same file through the network

[1]Veneta P. Aleksieva, is with the Department of Computer Science and Engineering, Technical University of Varna, str."Studentska"1, 9010 Varna, Bulgaria,E-mail: ven7066@abv.bg

[2]Peter T.Antonov is with the Department of Computer Science and Engineering, Technical University of Varna, str."Studentska"1, 9010 Varna, Bulgaria,E-mail: peter.antonov@ieee.org

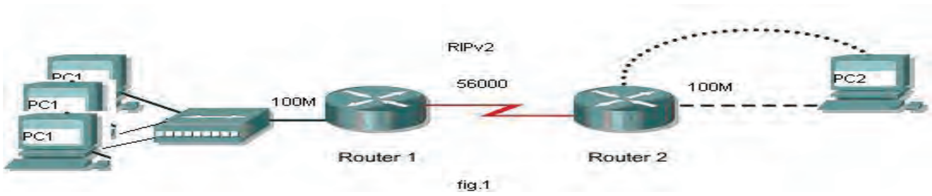

Fig.1. Model of typical fragment of network

9. the delays-caused collisions or faults of delivering, because of the busy communication medias

In this case the delays 1 and 3 may be ignored, because of static IP addresses and the constant treatment time of server during the experiment; 6 may be also ignored, because is the same in every of experiments and the images are visualized only once; 7 is the same, because for the aim of experiment are created alike files in JPEG with 4 stripes from basic colors, placed horizontal (it ignores the treatment way for images from the server).

In further explanation are used next symbols:

$T_{servicing}$ – time of service at message from network

Z – The size measured in bytes, including header and footer

T_{ij} – time from a point i to a point j /i and j may be host - router, router-router, router-host/

Lframes –the length of message, including all official information

$B_{x\text{-}y}$ - bandwidth of a link from x to y

H- the number of bytes official information - 20B from TCP header, 20B from IP header +38B from layer2 header and footer and a time between the frames

Dnic - delay time in a NIC

Dr – delay time in a router.

Moreover, the length of the sending message is increased, and it is calculated as:

$$\mathrm{Lframes} = Z + \left(\left[\frac{Z}{MTU}\right] + 1\right) * H \tag{1}$$

The expression $\left[\frac{Z}{MTU}\right]+1$ represent the number of frames, on which the message is divided, due to the length of MTU. If the message is less than MTU, then $\left[\frac{Z}{MTU}\right]=0$ and

$$\mathrm{Lframes} = Z + H \tag{2}$$

At that the time for transferring from one point to another during media with bandwidth = B is:

$$t = \frac{\mathrm{Lframes} * 8}{B} \tag{3}$$

When substitute Eq.(1) in Eq.(3) is calculated:

$$t = \frac{\left(Z + \left(\left[\frac{Z}{MTU}\right] + 1\right) * H\right) * 8}{B} \tag{4}$$

The delay time in the NIC for a frame is 0,556µs(Dnic), but in the router for a frame is 50µs(Dr). In that case the delay for the message in the host is:

$$\mathrm{Thost} = \mathrm{Dnic} * \left(\left[\frac{Z}{MTU}\right] + 1\right) \tag{5}$$

Then the delay for the message in the router is:

$$\mathrm{Trouter} = \mathrm{Dr} * \left(\left[\frac{Z}{MTU}\right] + 1\right) \tag{6}$$

The transition time of message during whole network from source to the destination in this example is:

$$T_{servicing} = T_{host1 - router} + T_{router1\text{-}router 2} + T_{router\text{-}host 2} + \\ +2*T_{router} + 2*T_{host} \tag{7}$$

When substitute Eq.(4), Eq. (5) and Eq.(6) in Eq.(7) is calculated:

$$\begin{aligned} T_{servicing} = {} & \frac{\left(Z + \left(\left[\frac{Z}{MTU}\right] + 1\right) * H\right) * 8}{B_{host1\text{-}router1}} + \\ & + \frac{\left(Z + \left(\left[\frac{Z}{MTU}\right] + 1\right) * H\right) * 8}{B_{router1\text{-}router2}} + \\ & + \frac{\left(Z + \left(\left[\frac{Z}{MTU}\right] + 1\right) * H\right) * 8}{B_{router2\text{-}host2}} + \\ & + 2 * \mathrm{Dnic} * \left(\left[\frac{Z}{MTU}\right] + 1\right) + \\ & + 2 * \mathrm{Dr} * \left(\left[\frac{Z}{MTU}\right] + 1\right) \end{aligned} \tag{8}$$

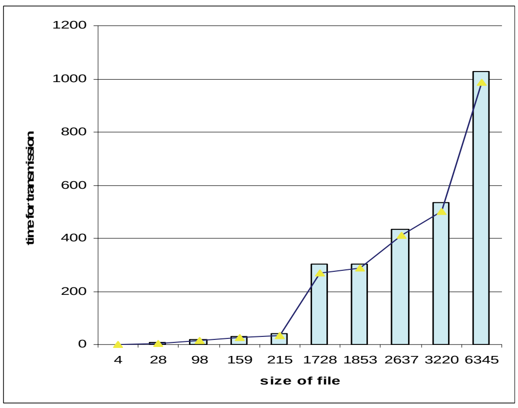

Fig. 2. The time of file transfer with different size of file

When it is measured the time of file transfer with different size of file, it is received a correspondence between a theoretical and an experimental times. It may be seen on Fig.2, as the data is presented in Table I.

When the receivers are more than one – for example N, and collisions don't have in the media, then from the source to router2 the number of frames are:

$\left(\left[\frac{Z}{MTU}\right]+1\right)*N$ and time of transfer to every receivers must be:

$$T_{servicingN} = T_{servicing} * N \qquad (9)$$

TABLE I
THE DIFFERENCE FOR 1 DESTINATION PC

size of file in KB	experimental time for 1PC	theoretical time for 1 PC	Difference	Difference as %
4	1,217471	0,803198	-0,414273	-34%
28	6,839668	5,549917	-1,289751	-19%
98	16,842333	19,207308	2,364975	14%
159	29,760572	31,184322	1,423750	5%
215	41,157712	42,139222	0,981510	2%
1728	301,896450	338,575240	36,678790	12%
1853	302,623950	363,041082	60,417132	20%
2637	435,454420	516,699550	81,245130	19%
3220	534,886030	630,921286	96,035256	18%
6345	1027,082400	1243,147080	216,064680	21%
			Average	6%

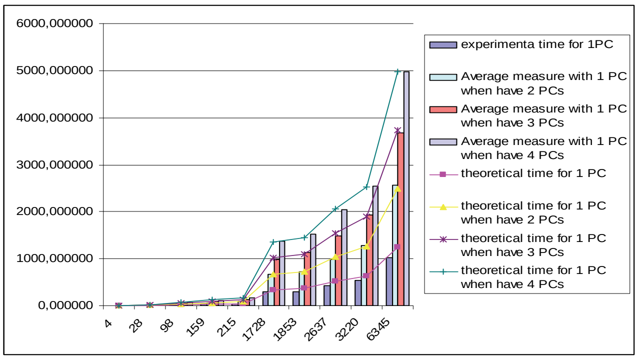

Fig.3. Theoretical and average measure times

Thus from view point of receiver the time to receive the file straight proportional of the increasing of the costumers number in the network (Fig.4). This can be watched at 2, 3, 4 receivers (Tables II and III and IV).

TABLE II
THE DIFFERENCE FOR 2 DESTINATION PCS

size of file in KB	theoretical time for 1 PC when have 2 PCs	Average measure with 1 PC when have 2 PCs	difference	Differenc e as %
4	1,606395	1,559643	-0,046752	-3%
28	11,099833	11,639395	0,539561	5%
98	38,414617	36,877514	-1,537103	-4%
159	62,368644	60,917755	-1,450890	-2%
215	84,278444	88,000552	3,722108	4%
1728	677,150480	669,790166	-7,360314	-1%
1853	726,082165	731,558122	5,475957	1%
2637	1033,399099	976,999486	-56,399614	-5%
3220	1261,842573	1276,562659	14,720086	1%
6345	2486,294159	2554,550637	68,256477	3%
			average	0%

TABLE III
THE DIFFERENCE FOR 3 DESTINATION PCS

size of file in KB	theoretical time for 1 PC when have 3 PCs	Average measure with 1 PC when have 3 PCs	difference	Difference as %
4	2,409593	2,348371	-0,061222	-3%
28	16,649750	16,514804	-0,134946	-1%
98	57,621925	60,761998	3,140073	5%
159	93,552966	87,740641	-5,812325	-6%
215	126,417667	132,995518	6,577851	5%
1728	1015,725719	976,966700	-38,759019	-4%
1853	1089,123247	1131,843270	42,720023	4%
2637	1550,098649	1489,349802	-60,748847	-4%
3220	1892,763859	1923,844201	31,080342	2%
6345	3729,441239	3682,679030	-46,762209	-1%
			average	0%

TABLE IV
THE DIFFERENCE FOR 4 DESTINATION PCS

size of file in KB	theoretical time for 1 PC when have 4 PCs	Average measure with 1 PC when have 4 PCs	difference	Difference as %
4	3,212790	3,146078	-0,066712	-2%
28	22,199666	21,260937	-0,938729	-4%
98	76,829234	78,010000	1,180766	2%
159	124,737288	119,194952	-5,542336	-4%
215	168,556889	174,846468	6,289579	4%
1728	1354,300959	1365,836432	11,535473	1%
1853	1452,164329	1522,303200	70,138871	5%
2637	2066,798198	2050,044590	-16,753608	-1%
3220	2523,685146	2537,800230	14,115084	1%
6345	4972,588319	4983,238101	10,649782	0%
			average	0%

The traffic in the network has a fluctuate character [2] and load moments of loading the time for answer from the view point of the receiver increase too much.

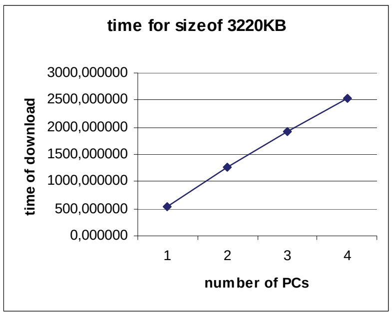

Fig. 4. Time of download for different number of PCs

The following of the development of the loading of the system is a key moment in planning of the performance, because the prediction of the future levels of loading the system in which is reached the moment of saturation define the most profitable ratio expenses – the performance. The planning of the development of the network, appropriate with the keeping of the parameters characterizing the performance leads to escaping financial losses and the dissatisfaction of the customers.

Moreover, for a random network, in which different sizes of request are downloaded from server, it is calculated the average size of file $Z_{average}$ by Eq. (10):

$$Z_{average} = \sum_{i=1}^{R} Z_i * r_i \qquad (10)$$

Where are used next symbols:

Zi –size of file,

R –number of every file, which may be downloaded

r –level of finding this document in percent [3].

Moreover, the time in the NIC, the network and routers by analogy with Eq.(8) is calculated by expression (11):

$$T_{servicing} = \frac{\left(Z_{average} + \left(\left[\frac{Z_{average}}{MTU}\right]+1\right)*H\right)*8}{B_{host1\text{-}router1}} + 2*\frac{\left(Z_{average} + \left(\left[\frac{Z_{average}}{MTU}\right]+1\right)*H\right)*8}{B_{router1\text{-}internet\ provider}} + Dnic*\left(\left[\frac{Z_{average}}{MTU}\right]+1\right) + Dr*\left(\left[\frac{Z_{average}}{MTU}\right]+1\right) \qquad (11)$$

When there are N clients simultaneously, it is calculated by Eq.(9).

The treatment time of server - Tserver is calculated from frequency with which the requests come. If the number of request for 1 second is M, it is calculated by Eq.(12):

$$T_{server} = \frac{1}{M} \qquad (12)$$

Furthermore, if is used the expressions Eq. (10), Eq.(11) and (12) easy may be check which network component is its bottleneck and if is possible, may be offer decision to overcome this problem. On TableV it is showed one example for $Z_{average=}$2MB, M=0,5, $B_{host\text{-}router1=}$100MB, MTU=1500B:

It is clearly to see, that the bottleneck in this system is the connection to ISP. If it is changed from 56к to E1, the bottleneck is shifted to the treatment of server. Thus the service time already completely correspond with notion of “acceptable” time from view point of customers. Here may be looked for optimal decisions about relation between an investment financial resources and a reached performance expectation.

TABLE V

$B_{router\text{-}internetprovider}$	56kB	2048MB
$T_{host\text{-}router}$	0,02702	0,02702
$T_{router\text{-}internet}$	394,97	0,00104
T_{router}	0,0683	0,0683
T_{server}	2	2
$T_{servicing}$	397,065	2,09636

EXAMPLE

III. CONCLUSION

In this paper, it is presented an analytical model and experimental results, from which it is reached to next common conclusions:

1. The analytical modeling predicts advantages of every variant before it is realized and this save time and money.
2. The network expansion and growth of complexity of contemporary and future systems approves the necessity of methods accepting of the performance valuation as a standard to system development and system realization. The analytical method of valuation in advance is quick and precise.
3. The development tracing of system load is a key moment when it is planning the system performance, because the prediction of future levels of system load, where the moment of saturation is reached, defines more profitable correlation of expense to performance. Now this evaluation in analytical model may be present as percent of customers increasing and as percent of increasing of size or number of contemporary download files.
4. The planning of network expansion is conformable with the parameters, which characterizing the performance, leads to overcome of financial loses and dissatisfied customers.

REFERENCES

[1] McCabe, J., Network Analysis, Architecture, and Design, Morgan Kaufmann publishers,2007

[2] C. Jin, D. Wei,” FAST TCP:From Theory to Experiments”, IEEE Network, vol. 19, pp.: 4- 11, Jan.-Feb. 2005

[3] Менаске, Д., В. Алмейда. Производительность Web-служб, анализ, оценка и планирование. М., изд. DiaSoft, 2003, 466 с.

[4] Li, Y.-T, D.Leith,, R.N.Shorten, “Experimental Evaluation of TCP Protocols for High-Speed Networks”, IEEE/ACM Transactions on networking, vol.15,pp.1109-1122,Oct.2007

[5] http://www.patentstorm.us/patents/6067412-fulltext.html, “Automatic bottleneck detection by means of workload reconstruction from performance measurements, US Patent Issued on May 23, 2000”