

Memory Estimation of Internet Server Using Queuing Theory: Comparative Study between M/G/1, G/M/1 & G/G/1 Queuing Model

L. K. Singh, and Riktesh Srivastava

Abstract—How to effectively allocate system resource to process the Client request by Gateway servers is a challenging problem. In this paper, we propose an improved scheme for autonomous performance of Gateway servers under highly dynamic traffic loads. We devise a methodology to calculate Queue Length and Waiting Time utilizing Gateway Server information to reduce response time variance in presence of bursty traffic. The most widespread contemplation is performance, because Gateway Servers must offer cost-effective and high-availability services in the elongated period, thus they have to be scaled to meet the expected load. Performance measurements can be the base for performance modeling and prediction. With the help of performance models, the performance metrics (like buffer estimation, waiting time) can be determined at the development process. This paper describes the possible queue models those can be applied in the estimation of queue length to estimate the final value of the memory size. Both simulation and experimental studies using synthesized workloads and analysis of real-world Gateway Servers demonstrate the effectiveness of the proposed system.

Keywords—M/M/1, M/G/1, G/M/1, G/G/1, Gateway Servers, Buffer Estimation, Waiting Time, Queuing Process.

I. INTRODUCTION

AS E-Commerce continues to grow swiftly [1], the original Architecture providing the service of E-Commerce is becoming more and more important. According to [2], slow performance of Internet Server will cost as much as \$4.35 billion annually in lost revenues.

The architecture of Gateway Server is in general referred to as “Web services”. The term “Web Services” describes specific business functionality through Internet connections, for the purpose of providing a way for another entity to use the services it provides [3]. Web services are the building blocks for the future generation of applications and solutions on the Internet.

The services provided through Internet Server(s) are colossal. It transfers information in form of text, images and voice. The whiz of information technology is because of

consistent and steady functioning of Internet Server(s)[4], [5], [6].

However, Web traffic is highly dynamic and volatile [3],[7]. The data arrives and departs from different nodes randomly. Thus, we can envisage that, “number of channels” for arrival and “number of channels” for departing must be identical. The incoming data can be stochastically treated as a “process” and so will be the case of departing from the memory of Web Servers. These situations make the working of Memory of Web Servers - a typical case of - “Queuing Process”[8]. During the last 40 years, research has shown that queuing models serve as a fundamental tool to model computing systems [9],[10]. In fact, queuing models have been successfully applied to areas such as capacity planning [11] and performance analysis [11] etc.

The layout of Gateway Server is depicted in Figure 1, in which data arriving and data departing at a node has been revealed:

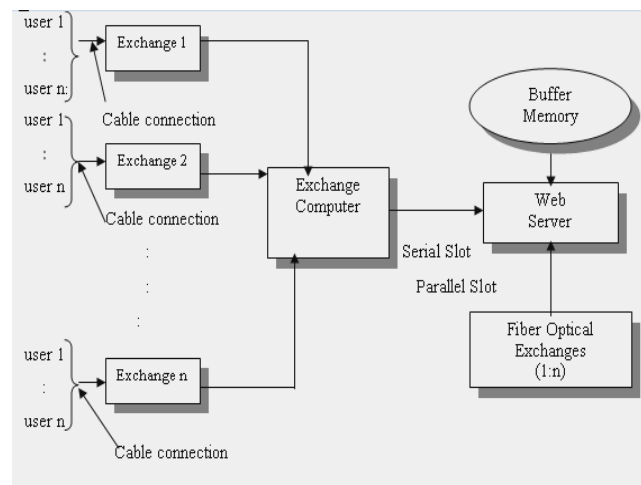


Fig. 1 Layout of Gateway Computer-Data Arrival/ Data departure

It has been implicated that memory passes through diverse situations of Queue Models, i.e., M/M/1, M/G/1, G/M/1 and G/G/1. In all these four models, M/M/1 model is most disciplined and can be analyzed analytically to estimate the queuing parameters, i.e., “Queue Length” and “Waiting time” [11], [12]. These are derived as given in equations (1) and (2):

Dr. L. K. Singh is Professor and Head, Department of Solid State Physics and Electronics, Dr. R.M.L. Avadh University, Faizabad, India.

Riktesh Srivastava is Research Scholar, Department of Solid State Physics and Electronics, Dr. R.M.L. Avadh University, Faizabad, India and presently working as Project Head-Internet Technologies, Academecia Software Solutions, India (e-mail: rikteshsri1@rediffmail.com/riktesh@academecia.com).

$$\bar{Q}(L) = \frac{\lambda / \mu}{(1 - \lambda / \mu)} \quad (1)$$

$$\bar{W}(t) = 1/\mu \left\{ \frac{\lambda / \mu}{(1 - \lambda / \mu)} \right\} \quad (2)$$

The other models (M/G/1, G/M/1, G/G/1) emerges to be most undisciplined and provides situations for calculating queuing parameter's analytically incredibly tough. One can relatively envision that this will be case, which requires bigger value of memory size, thereby, making task a challenging application. However, to get all cases verified, we propose to carry out the study for all the four cases employing stimulation techniques on workstation. Stimulation technique is a viable method when problem arises to computer analytically. Stimulation techniques for queuing models is possible if arrival instances and departing instances are generated having pre-defined rate of arrivals and rate of departures for a given distribution. After gazing the models, it is mandatory to delineate two types of strings, for the given average rate of arrivals and departures,

A String of departure or arrival for pre-assigned average values having negative exponential distribution.

The other set of strings having pre-assigned average rate of arrivals/average rate of departure having general distribution. General distribution is a case when there is no distribution for which the string belongs more than, say, 50.

The computer for estimation of Queue Length and Waiting time can employ the two strings. For stable maneuver of the system, it is to be ensured that processing speed of workstation be larger than the arrival rates. This condition when convene, is called "*functioning of the system in Ergodic State*".

The reminder of this paper is planned as follows. Section 2 gives the background of Computer Simulation study for designing IGS. Section 3 provides the formal description of formal of arrival instances at IGS. Section 4 provides the description of formal of departure instances at IGS and its implementation. Based on the description of Section 4 and 5, Section 5 reviews the Mathematical study of waiting time and Queue Length w.r.t. 4 queuing models. Section 6 and 7 describes the computational results of the proposal. Finally, Section 8 summarizes our conclusions and offers possible directions for future work.

II. COMPUTER SIMULATION

As has been conferred, to figure two long strings of size more than 50000 for pre-assigned average values and either having the general distribution or having negative exponential distribution. Then, at each departing time, the Queue Length present in the system helps to compute – "Memory Size" and "Waiting Time".

III. FORMATION OF ARRIVAL INSTANCES

As, it is planned to clutch out stimulation studies for model(s) M/M/1, M/G/1, G/M/1, and G/G/1, We need two sets of Arrival Instances, each having same arrival rates but

different distribution, namely

A. *Negative-Exponential Distribution*

B. *General Distribution*

General N-E distribution

The arrival process for N number of arrival instances is accumulated by generating N number of equiprobable distributions. Subsequently, these are transformed to Negative-Exponential distribution for pre-assigned rate of arrivals. Equation (3) is used for transformation of equiprobable data to Negative-Exponential sequence:

$$y = 1/k (\log(1-x)) \quad (3)$$

where,

y=Negative-Exponential data

x=Equiprobable random between 0 and 1

k= appropriate constant

Let these arrival instances are a_1, a_2, \dots, a_N . Then, the arrival instances which is nothing but Markov chain is general using equation (4):

$$A_N = a_{N-1} + a_n \quad (4)$$

where, A_N for $n=0$ to N are time instances of the arrival of data. The second string for the arrival is having general distribution. The general distribution has been achieved in the study by merging 5 distributions having pre-assigned rate of arrivals. These distributions are:

1. Bernaulli Distribution
2. Geometric Distribution
3. Equiprobable Distribution
4. Gaussian Distribution
5. Negative-Exponential Distribution

The equiprobable method is provided with the system are used for generation of any distribution employing equal area criteria method using following formula:

A. *Bernaulli Distribution*

The equation for the transformation of equiprobable to Bernaulli distribution is given in equation (5):

$$f(y(i)) = a + by(i) \text{ for } i \in 1 \text{ to } N \quad (5)$$

Equiprobable Distribution = $x(i)$

Bernaulli Distribution = $y(i)$

then,

$$y(i) = -a \pm \sqrt{a^2 + 2bx(i)/b} \quad (6)$$

where,

$b = 1/\lambda$ in arrival process

$b = 1/\mu$ in departure process

$a = 1 - b$

B. Transformation of Equiprobable to N-E

$$y = \frac{1}{k} (\log(1-x)) \quad (7)$$

where,

y=Negative-Exponential data

x=Equiprobable random between 0 and 1

k= appropriate constant

C. Geometric Distribution

$$f(y(i)) = \frac{a}{(1-by(i))} \text{ for } i \in 1 \text{ to } N \quad (8)$$

Equiprobable Distribution= $x(i)$ for $i = 1 \in 1 \text{ to } N$

Geometric Distribution = $y(i)$ for $i = 1 \in 1 \text{ to } N$

then,

$$y(i) = \frac{1}{b(1-e^{-\delta x(i)})} \quad (9)$$

where,

$$\delta = b/a$$

$$b = 1-a$$

$$b = \lambda/\mu \text{ for arrival/departure.}$$

D. Gaussian Distribution

There is varied procedure of generating the Gaussian Distribution. The accepted practice is by intriguing mean of 12 Equiprobable numbers and arranges them in a String.

Mathematical expression is specified beneath:

Let Equiprobable number are x_1, x_2, \dots, x_{12} .

Then, Gaussian distribution

$$y = \frac{1}{12} \left[\sum_{i=1}^{12} x(i) \right] \quad (10)$$

E. Equiprobable Distribution

Equiprobable numbers are spawned between 0 and 1 for the identical range using command offered with the system.

IV. FORMATION OF DEPARTURE INSTANCES

As epitomized in arrival process, Departure process will also have two elongated strings for departure instances are to be stimulated for:

Negative-Exponential Distribution

Average rate of departure, $\mu = \lambda + 1$

The identical approach is employed, as been depicted in the arrival process. The other string is made of, for $\mu = \lambda + 1$, having general distributions with combination of 5 distributions as conferred on arrival process. What is done actually, 5 sets of N/5 numbers are generated having distributions – Bernaulli, Gaussian, Negative-Exponential, Geometric, Equiprobable and are merged randomly.

From these two strings, two departure chains are computed using formulae:

$$D_n = D_{n-1} + d_n \quad (11)$$

where, $n = 0 \text{ to } N$ and $d = n^{\text{th}}$ departure duration. To guarantee that departure doesn't take situate before arrival, the departure chain is shifted by 1 second.

V. QUEUE LENGTH AND WAITING TIME

In the segment, we have computed "Queue Length" and "Waiting Time" for 4 models, namely, M/M/1, M/G/1, G/M/1 and G/G/1, respectively.

A. M/M/1 Model

Two strings are chosen, both having Negative-Exponential Distribution. One should be of arrival process and other for the departure process. At each departing instance, Queue Length, is computed. Let these Queue Length be, q_1, q_2, \dots, q_n . Then, Average Queue Length

$$\overline{Q(L)} = y_n \sum_{i=1}^n Q(i) \quad (12)$$

$$W(t) = \frac{1}{\mu} (Q(i)) \quad (13)$$

B. M/G/1 Model

In this case, arrival process is selected having distribution Negative-Exponential. The departure process is selected having general distribution at the each instance of the departure, Queue Length is counted and using equation(s) 10 and 11 and the average queue length is computed.

C. G/M/1 Model

In this case, the arrival string is chosen having general distribution and the departure distribution is picked up for negative exponential distribution. The Queue Length and Waiting time is computed as described in the segment affirmed above.

D. G/G/1 Model

This is a emblematic case and offers largest Queue length and largest Waiting time. In this case, a string from arrival process is chosen having general distribution. Similarly, a string from departure process is also chosen having general distribution. The value of $Q(L)$ and $W(t)$ are computed as the technique portrayed previously.

VI. COMPUTATIONAL RESULTS

The program has been developed using VC++ to compute all the four queuing models in single fondle. The program is appexed for models M/M/1, M/G/1, G/M/1 and G/G/1. The result so acquired for rate of arrivals 5000, 7500, 10000, 12500, 15000. The rate of departure has been assumed to be 5001, 7501, 10001, 12501, 15001. The computed results are revealed in the subsequent table:

TABLE I
ESTIMATION OF QUEUE LENGTH

Rate of arrival (λ)	$Q(L)$ FOR M/M/1	$Q(L)$ FOR M/G/1	$Q(L)$ FOR G/M/1	$Q(L)$ FOR G/G/1
5000	5000	5102	5098	5538
7500	7500	7653	7647	8400
10000	10000	10204	10194	11190
12500	12500	12755	12740	14126
15000	15000	15306	15283	17695

VII. ESTIMATION OF MEMORY SIZE USING NON-LINEAR REGRESSION TECHNIQUE

We have calculated Queue Length for 5 rate of arrivals. These 5 points are on the curve of Queue Lengths. Queue Length at very high rate will be also on the curve of Queue Length.

Let the equation of the curve is represented as:

$$y_{\phi} = a_0 + a_1\lambda + a_2\lambda^2 \quad (14)$$

Now, using Non-linear regression techniques, we have to calculate the value of a_0, a_1 , and a_2 , for the minimal error. These leads to develop formation equations as given below:

$$\left. \begin{aligned} na_0 + a_1 \sum \lambda_i + a_2 \sum \lambda_i^2 &= \sum y_i \\ a_0 \sum \lambda_i + a_1 \sum \lambda_i^2 + a_2 \sum \lambda_i^3 &= \sum \lambda_i y_i \\ a_0 \sum \lambda_i^2 + a_1 \sum \lambda_i^3 + a_2 \sum \lambda_i^4 &= \sum \lambda_i^2 y_i \end{aligned} \right\} \quad (15)$$

There are three unknown variables, i.e., a_0, a_1 and a_2 and three equations for given values of λ .

These values are computed and are given as:

$$\left. \begin{aligned} a_0 &= 1484.5 \\ a_1 &= 0.873201 \\ a_2 &= 0.0000051 \end{aligned} \right\} \quad (16)$$

where, λ is the rate of arrival and $Q(L)$ is the Queue Length.

The results for the Queue Length and hence estimated Memory Size are as given Table II:

TABLE II
COMPUTATION OF a_0, a_1 AND a_2

M/G/1 MODEL	G/M/1 MODEL	G/G/1 MODEL
$a_0=1484.0$	$a_0=1484.5$	$a_0=1483.5$
$a_1=0.873200$	$a_1=0.873201$	$a_1=0.873377$
$a_2=0.000005$	$a_2=0.0000051$	$a_2=0.000006$

Then, the Queue Length for M/G/1, G/M/1 and G/G/1 model is given as under:

$$y = 1484.0 + 0.873200\lambda + 0.000005\lambda^2 \quad (17)$$

$$y = 1484.5 + 0.873201\lambda + 0.0000051\lambda^2 \quad (18)$$

$$y = 1483.5 + 0.873377\lambda + 0.000006\lambda^2 \quad (19)$$

where, λ is the rate of arrival and y is the Queue Length.

The results for the Queue Length so obtained are given in Table III:

TABLE III
COMPUTATION OF QUEUE LENGTH USING REGRESSION TECHNIQUE

Rate of arrival (λ)	$Q(L)$ FOR M/M/1	$Q(L)$ FOR M/G/1	$Q(L)$ FOR G/M/1	$Q(L)$ FOR G/G/1
10^{10}	10^{10}	1.0204 $\times 10^{10}$	1.0196 $\times 10^{10}$	1.1076 x 10^{10}
10^{11}	10^{11}	1.0204 $\times 10^{11}$	1.0196 $\times 10^{11}$	1.1076 x 10^{11}
10^{12}	10^{12}	1.0204 $\times 10^{12}$	1.0196 $\times 10^{12}$	1.1076 x 10^{12}

VIII. CONCLUSION AND FUTURE WORK

We proposed that for Internet Gateway Servers, the buffer estimation is contemporary need in crafting the design of future Internet system. Comparing various queuing algorithms, the aim is to define and design pseudo-code and algorithm to calculate the accurate Buffer Size of Internet Server, under a wide range of workload conditions including bursty traffic. This is achieved by utilizing the server internal queue length measurements. Extensive simulation study shows that the new scheme can provide smooth performance control and better track in Web server systems.

The study confirms that the system must be such that at each Internet Gateway Server (IGS) should have $\lambda < \mu$. This will guarantee the stability of the implementation. The buffer estimated by queuing theory will ensure that there is no overflow of data at any stage and size will be optimal.

With respect of contemporary progress, further research also includes the implementation of Computation Grid, using the same technology.

REFERENCES

- [1] UDDI (2004), "Universal description, discovery, and integration of business for the web,".

- [2] Zari ,Mazen, et. al(2001), "Understanding and Reducing Web delays", pp. 30-37, IEEE Journal for Electronics and Computer Science, Vol. 34, No.12.
- [3] Mogul ,J. C. (1995), "Operating systems support for busy internet servers," in Fifth Workshop on
- [4] Low , Steven H., Srikant , R. (2002), "A Mathematical Framework for Designing a Low-Loss, Low-Delay Internet", IEEE Transactions on Communications.
- [5] Chen ,Xiangping, Mohapatra , Prasant (2002), "Performance Evaluation of Service Differentiating Internet Servers", pp. 1368-1375, Vol. 51, No. 11.
- [6] Ying , Lei et. al. (2003), "Global Stability of Internet Congestion Controllers with Heterogeneous Delays", IEEE Transactions on Communications.
- [7] Barford , P., Crovella ,M. (1998), "Generating representative web workloads for network and server performance evaluation," in Measurement and Modeling of Computer Systems, pp. 151–160.
- [8] Dr. L.K. Singh, Riktesh Srivastava, "Estimation of Buffer Size of Internet Gateway Server via G/M/1 Queuing Model" International Journal of Applied Science, Engineering and Technology, Volume 4 Number 1, pp. 474-482, January 2007.
- [9] Hot Topics in Operating Systems(HotOS-V), Orcas Island, WA.
- [10] Kleinrock , L. (1976), Queueing Systems, Vol. 2, Applications. John Wiley Publications, NY.
- [11] Menasce , D., Almeida , V. (2001), Capacity Planning for Web Services: Metrics, Models, and Methods. Prentice Hall PTR.
- [12] Lazowska , E. D. et. al(1984), Quantitative system performance: computer system analysis using queueing network models. Prentice-Hall, Inc.
- [13] Anderson , Darrell et. al.(1999), "A Case for Buffer Servers", p. 82, IEEE Seventh Workshop on Hot Topics in Operating Systems.
- [14] Roberts , Jim W. (2001), "Traffic Theory and the Internet", IEEE Transactions on Communications.

Dr. L. K. Singh is presently Professor and Head, Department of Solid State Physics and Electronics, Dr. R.M.L. Avadh University, Faizabad, India. Additionally, he is also the Director(officiating) Institute of Engineering and Technology, University Campus, Dr. R.M.L. Avadh University, Faizabad. He also published numbers of in the field of Spectroscopy and Fiber Optics.

Riktesh Srivastava presently Project Head, Web Technologies, Academecia, is also research scholar under Dr. L.K. Singh. His areas of specialization include programming, Web technologies. He has written two books on – Operating System and C++. He holds MSc(Electronics) Degree.