Design and Implementation of G/G/1 Queuing Model Algorithm for its Applicability in Internet Gateway Server

Laxmi Singh[1] and Riktesh Srivastava[2]
[1]Department of Solid State Physics and Electronics, Dr. RML Avadh University, India
[2]Project Head-Internet Technologies, Academecia Software Solution, India

Abstract: *How to competently apportion system resource to process the client request by gateway servers is a tricky predicament. In this paper, we propose an enhanced proposal for autonomous recital of gateway servers under highly dynamic traffic loads. We have developed G/G/1 queuing model algorithm and premeditated its intricacy, so that there is lossless information repossession at each node of gateway server. This facilitates to reduce response time variance in existence 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, so that there is lossless information retrieval of data at every node of internet gateway servers. The paper portrays the assessment of buffer size using G/G/1 queuing model to estimate the final value of the memory size and then examine its implementation at the gateway servers. The obtained output is based on the 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, G/G/1, internet gateway server, queuing process.*

Received February 20, 2007; accepted June 8, 2007

1. Introduction

Internet traffic, as epitomized by web browsing behaviour, is very bursty, or equivalently, the ratio of the peak to average data rate is quite high. As e-commerce continues to grow swiftly [9], the original architecture providing the service of e-commerce is becoming more and more important. According to [11], slow performance of internet server will cost as much as $4.35 billion annually in lost revenues. The architecture of internet gateway server(s) is in general referred to as "web services". The term "web services" describes specific business functionality through internet connections, for providing a way for another entity to use the services it provides [7]. Web services are the building blocks for the future generation of applications and solutions on the internet.

The services provided through internet gateway 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 gateway server(s) [2, 5, 10].

However, web traffic is highly dynamic and volatile [1, 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 internet gateway server(s). These situations make the working of memory of internet gateway server(s) - 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 [3, 4]. In fact, queuing models have been successfully applied to areas such as capacity planning and performance analysis [10].

In internet, computers of a city are coupled to a city gateway server. The city gateway server can be associated with gateway servers of other town. The connectivity can be through using current telecommunication system, i.e., using optical fibers, microwave links, tele-line links, satellite links and the like.

In these associations, a computer can transmit its data to any other computer of the internet network. This formulates the internet connectivity of a country having one or two country gateway servers. The country gateway server accepts data from servers of erstwhile towns. The town server receives data from assorted clients of the town. The data flow from clients to town server is random. The data coming in the country server is further random. The data leaving the server after staying for some times in the memory is also random. This makes the picture for data flow in server's memory, random arrival and random departure process, with temporarily staying in the buffer memory. The problem becomes thus, a typical queuing theory model. The layout of the system is shown in Figure 1.

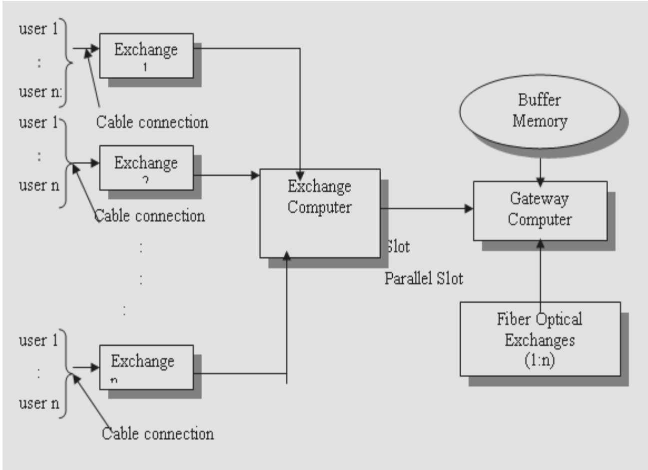


Figure 1. Layout of internet gateway server(s).

If the Figure is further consolidated, it will be reduced to a single arrival process and a single departure process along with processing system. In this system, memory provides platform for data to stay in it. As arrival and departure processes are random, the estimation of memory size becomes a major issue. In this paper, we have formulated the algorithm to define the buffer size and waiting time of the request, which is based upon the following principles:

- No data should be overflowed/ freezed out.
- Memory should not be abnormally large.
- As there are millions of gateway servers placed either at towns or at country, the algorithm must be such designed to ensure stable functioning of the entire system.

These three problems are of prime importance of electronically designing internet server(s). The gateway mode resembles the working of queuing theory, it is therefore planned to carry out the study by employing queuing theory.

2. Queuing Theory for Stable Working of Gateway Server(s)

Let the random arrival of data at the gateway server is λ. The random arrival is summation of data at any instance from the links:

- Satellite.
- Tele-line.
- Optical Fiber.
- Radio-net.
- Microwave links.
- Local users.

Similarly, the departure of data is denoted by μ, then, it should be sum of departure data of:

- Satellite.
- Tele-line.
- Optical Fiber.
- Radio-net.
- Microwave links.
- Local users.

Pictorially, it is shown in Figure 2.

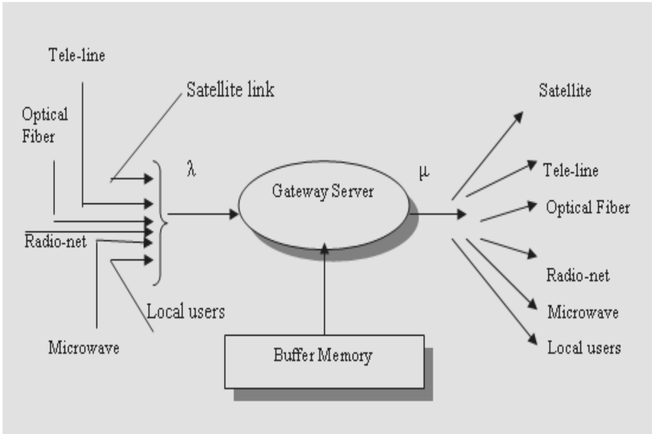


Figure 2. Arrival and departure process at internet gateway server.

λ is summation of arrival data per unit time, i.e.,

$$\lambda=[a_1+a_2+a_3+a_4+a_5+a_6]/T \quad (1)$$

where, a_1,a_2,a_3,a_4,a_5,a_6 are the total data arrived at the gateway server through links, satellite, tele-line, optical fiber, radio-net, microwave links, local users, respectively. T is the total time in which the data has arrived at server. Similarly, the departure rate is defined as:

$$\mu=[d_1+d_2+d_3+d_4+d_5+d_6]/T \quad (2)$$

where, d_1,d_2,d_3,d_4,d_5,d_6 are total data departed from the gateway server through links, satellite, tele-line, optical fiber, radio-net, microwave links, local users, respectively. T is the total time in which the data has departed from server. Now, there can be 3 possibilities:

- If arrival rate is greater than the departure rate, i.e., $\lambda>\mu$, this case is called *transient*. If arrival rate is greater than the departure rate at any gateway server, then, there will be overflow of data at the buffer memory. This makes the system unstable. Hence, it is concluded that the arrival rate at no circumstances at the server be greater than the departure rate. It is a very important constraint and must be met for stable working of internet at each node/ server.
- if arrival rate is equal to departure rate, i.e., $\lambda=\mu$, this is a very typical case, called NULL case. Such phenomenon randomly occurs. This case is typical

for academic studies only. Practically, this neither occurs nor is desirable.

- if arrival rate is less than the departure rate, i.e., $\lambda<\mu$, this case is called as *ergodic state*. If this condition is maintained, then there will be finite queue length of data, which will be essential to be accumulated. The finite queue length thus guarantee's finite memory size. The study should be consequently made for the assessment of optimal memory size, which endow with all the randomly arriving data for temporarily accommodation.

In the current study, when a gigantic quantity of data is received at the gateway server, and enormous random departure occurs, predicament becomes thorny to be solved systematically This quandary gets further dense when the data arrives at the internet gateway server from diverse channels at different rates. Had it been arriving from a single channel and departing through another single channel, the queuing model would have been proximated to M/M/1, where the first *M* describes the arrival process to be Markovian. Markovian arrival process is a poisson arrival where inter-arrival distribution is negative exponential. The second *M*, describes the departure process with processing unit 1 (one in number). In case of multi-arrival channels, this postulation does not fit in. The most effectual distribution can be measured for the process of arrival or departure to be a general distribution. On such cases, the queue model becomes either, G/M/1 or M/G/1. However, if arrival or departures both are considered multi-channel, which is the current functioning of internet gateway server, the suitable replica becomes G/G/1. These cases occur simultaneously in one or other gateway server(s). It is worth to repeat that M/M/1 queue model have been analytically studied for the estimation of average queue length in variance and average waiting time and its variance. However, in case of general distribution, it is very complicated to compute the queuing parameters mathematically. We, therefore hereby carried the study of G/G/1 and develop the algorithm to calculate the average queue length.

3. Formation of Arrival Instances

As it is planned to clutch out stimulation study for G/G/1 model, we need two sets of arrival instances, each having same arrival rates but different distribution, namely:

- Negative-Exponential Distribution.
- 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. Following equation is used for transformation of equiprobable data to negative-exponential sequence:

$$y = 1/k(\log(1-x)) \tag{3}$$

where, y=Negative-Exponential data, x=Equiprob data, k=constant.

Let these arrival instances are $a_1, a_2, \ldots\ldots a_N$. Then, the arrival instances which is nothing but markov chain is general using equation

$$A_N=a_{N-1}+a_n \tag{4}$$

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 five distributions having pre-assigned rate of arrivals. These distributions are:

- Bernaulli Distribution
- Geometric Distribution
- Equiprobable Distribution
- Gaussian Distribution
- Negative-Exponential Distribution

Bernaulli distribution: the formula used for the transformation of equiprobable to bernaulli distribution as:

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

Equiprobable distribution = x(i)
Bernaulli dstribution = y(i)

then,

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

where,
$b=1/\lambda$ in arrival process, $b=1/\mu$ in departure process, and a=1-b.
Transformation of equiprobable to *N-E*

$$y = 1/k(\log(1-x)) \tag{6}$$

where,
y=Negative-Exponential *data,* x=Equiprobable random between 0 and 1, *and* k= appropriate constant.

Geometric distribution

$$f(y(i))=a/(1-by(i)) \text{ for } i\in 1 \text{ to } N \tag{7}$$

Equiprobable distribution = x(i)
Bernaulli distribution = y(i)

then,

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

where,
$\delta=b/a$, $b=1-a$, and $b=\lambda/\mu$ for arrival/ departure.

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 $x_1, x_2, \ldots\ldots\ldots\ldots x_{12}$. Then, gaussian distribution

$$y = 1/12\ [\sum_{i=1}^{12} x(i)] \quad (8)$$

Equiprobable distribution: equiprobable numbers are spawned between 0 and 1 for the same size using command available with the system.

4. Formation of Departure Instances

Arrival process as shown in Figure 3-(a), 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:

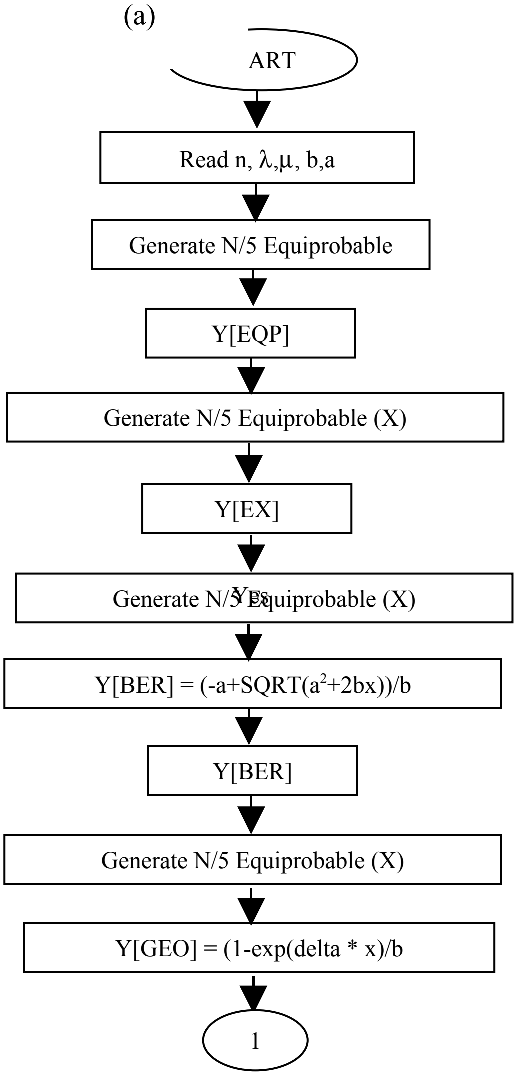
$$D_n = D_{n-1} + d_n \quad (9)$$

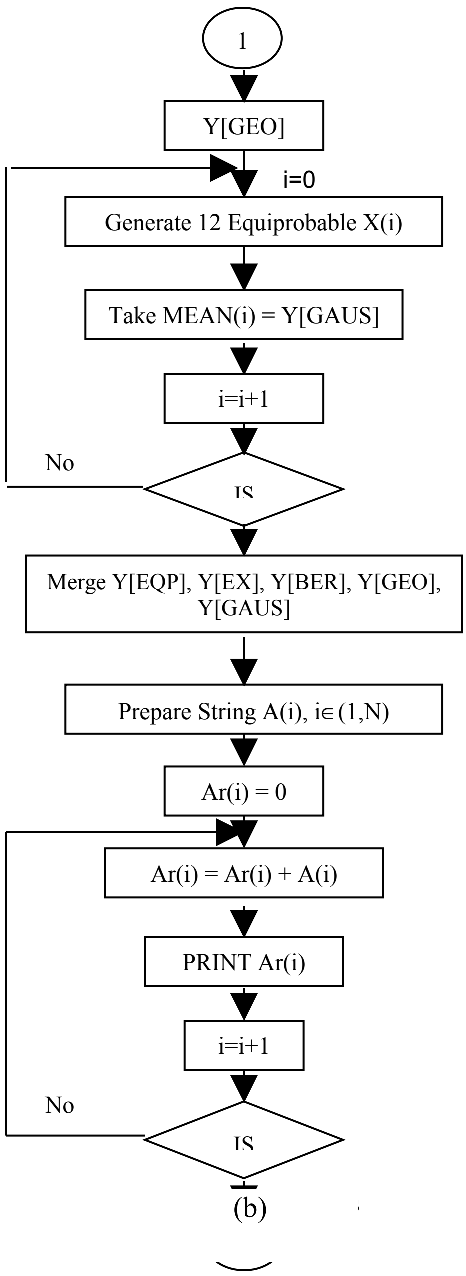
where n=0 to N and $d=n^{th}$ departure duration. To guarantee that departure doesn't take situate before arrival, the departure chain is shifted by 1 second.

5. Generation of General Distribution of Arrival

Based on the five arrival channels, we have formulated the flowchart for the general distribution of arrival. It is elucidated as shown in Figure 3.

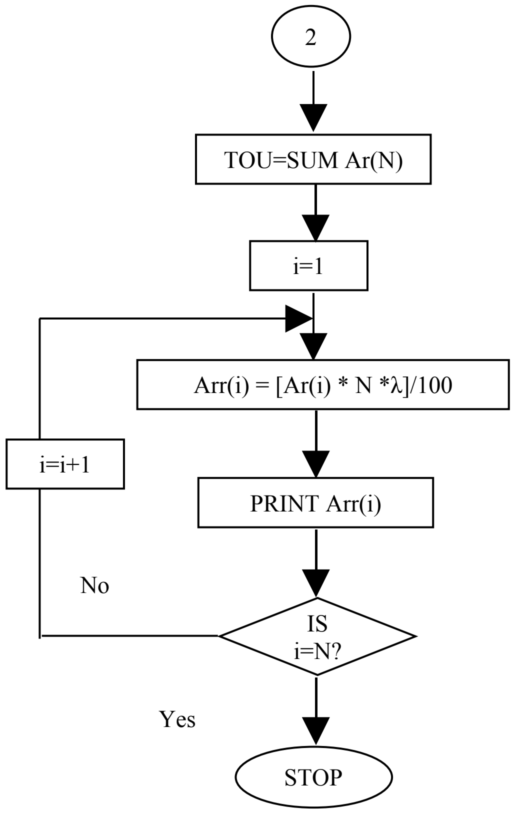
(a)





(b)

(b)

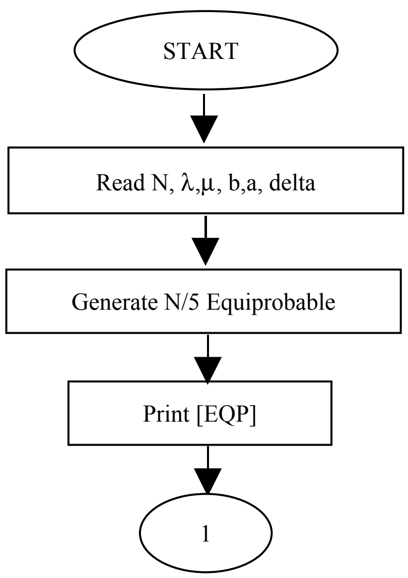


(c)

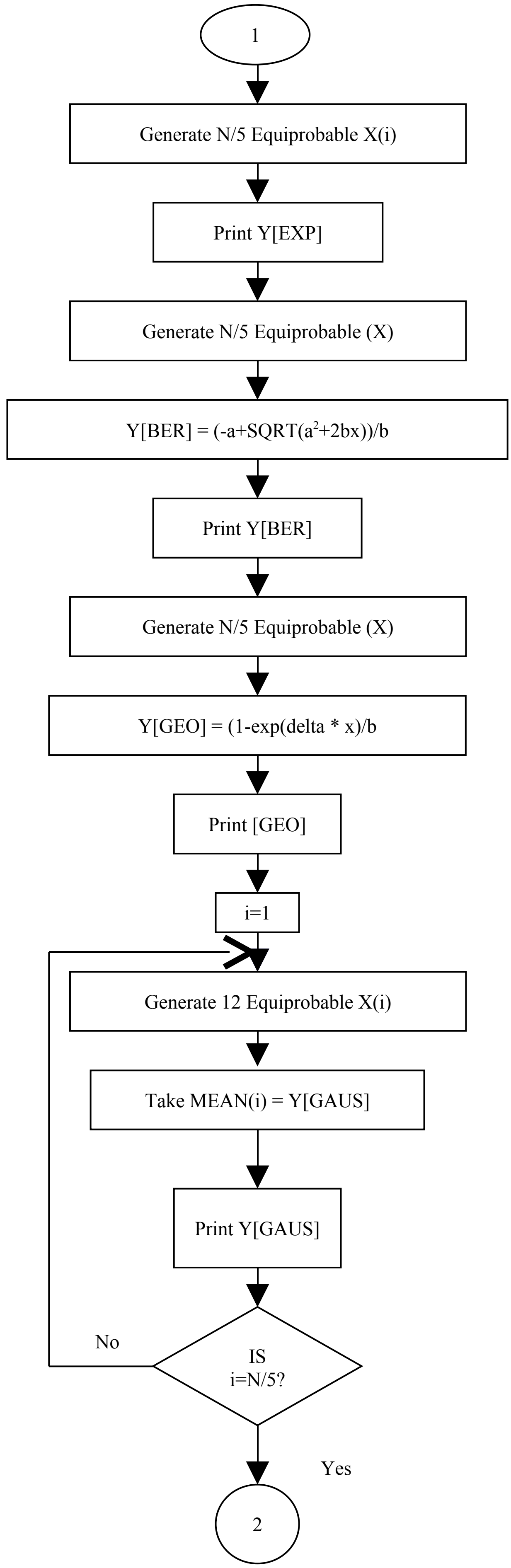
Figure 3. Flowchart for generation of general distribution of arrival

6. Departure Process

The flowchart of departure process for G/G/1 queuing model is illustrated as shown in Figure 4.



(a)



(b)

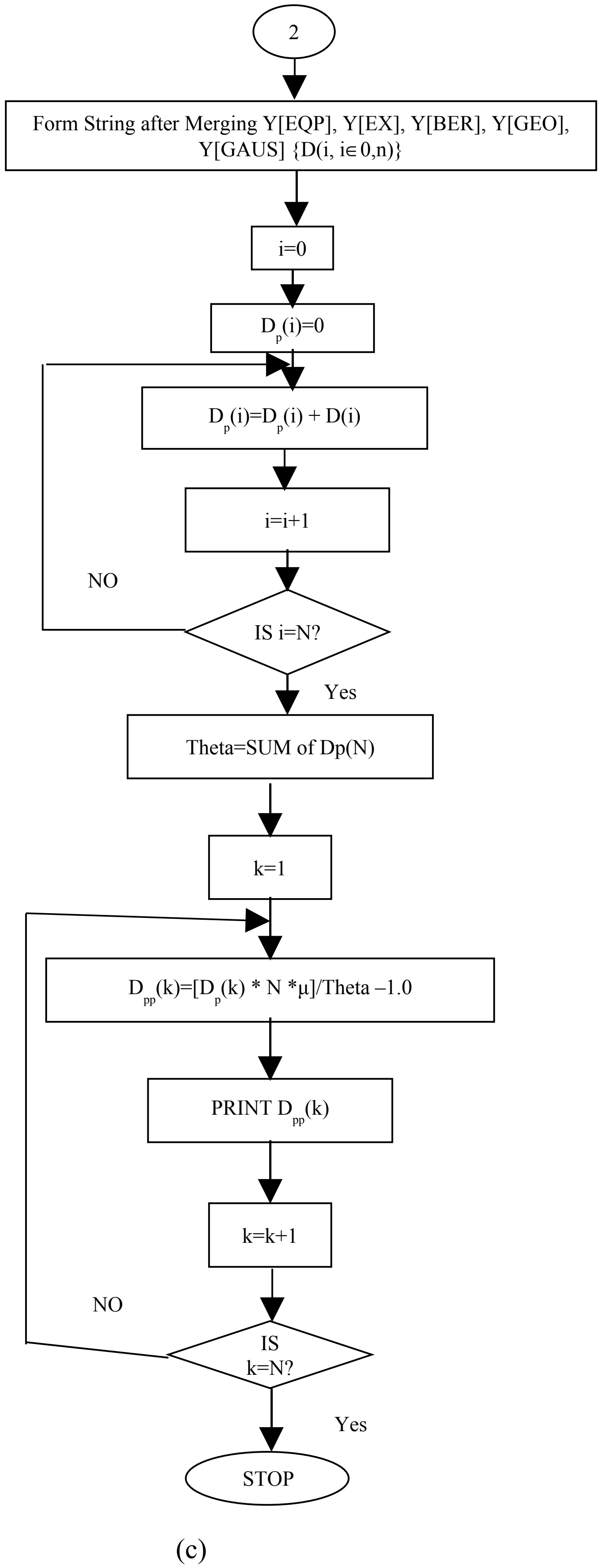


Figure 4. The flowchart of departure process for G/G/1 queuing model

7. Computation of Average Queue Length in Queue Model

Based on the algorithm depicted in segment 5 and 6, the pragmatic algorithm for computing the queue length, using G/G/1 model is given in Figure 5:

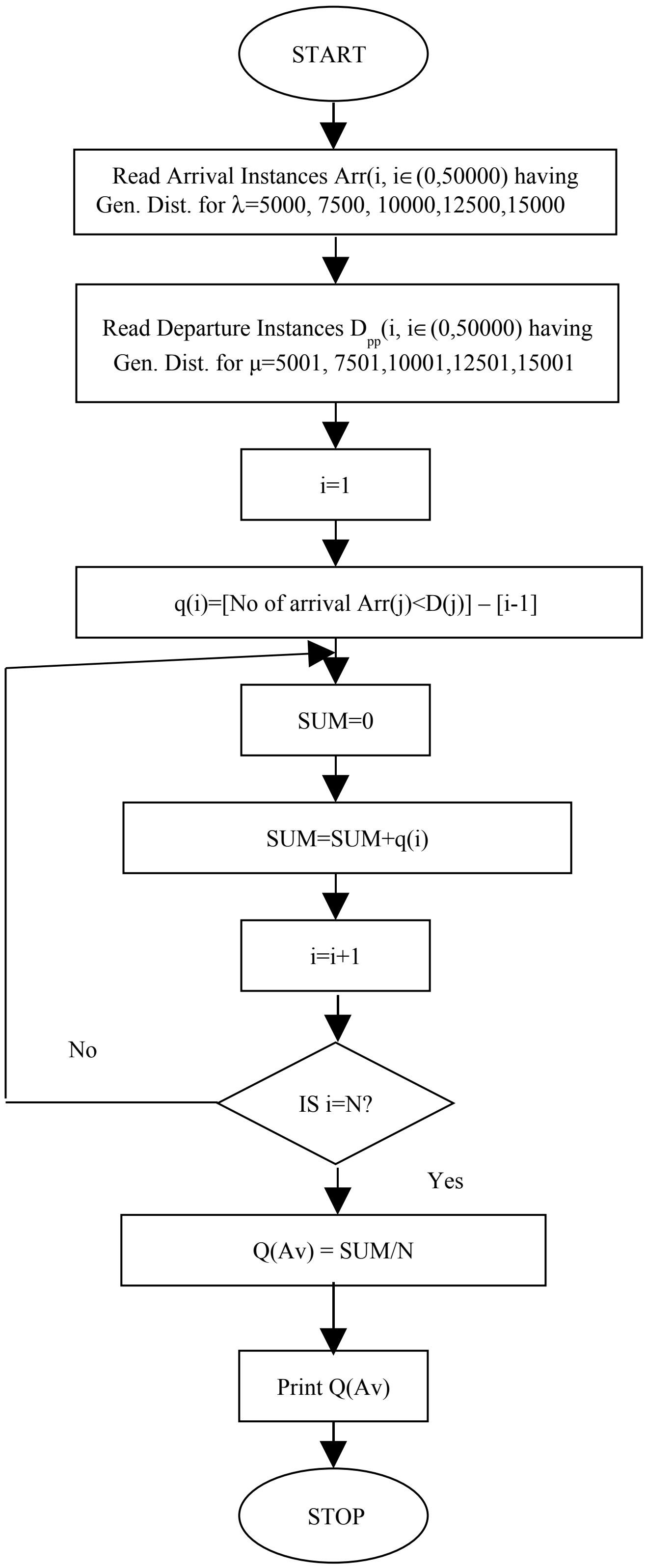


Figure 5. The flowchart for computation of average queue length in queue model G/G/1

8. Calculated Results

In the segment, we have Figured queue length and waiting time for M/M/1 and G/G/1 respectively.

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, $q_1, q_2, \ldots\ldots, q_n$. Then, average queue length

$$\overline{Q(L)} = y_n \sum_{i=1}^{n} Q(i) \qquad (10)$$

$$\overline{W(t)} = 1/\mu[Q(i)] \qquad (11)$$

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.

Using the algorithm, we have computed the two queuing models (M/M/1, G/G/1) in single fondle. 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 1. Estimation of queue length.

Rate of Arrival	Q(L) for M/M/1	Q(L) for G/G/1
5000	5000	5538
7500	7500	8400
10000	10000	11190
12500	12500	14126
15000	15000	17695

9. Estimation of Buffer Size Using Non-Linear Regression Technique

We have calculated queue length for 5 rates 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:

$$y_\phi = a_0 + a_1\lambda + a_2\lambda^2 \qquad (12)$$

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:

$$\left.\begin{aligned} na_0 + a_1\Sigma\lambda_i + a_2\Sigma\lambda_i^2 &= \Sigma y_i \\ a_0\Sigma\lambda_i + a_1\Sigma\lambda_i^2 + a_2\Sigma\lambda_i^3 &= \Sigma\lambda_i\, y_i \\ a_0\Sigma\lambda_i^2 + a_1\Sigma\lambda_i^3 + a_2\Sigma\lambda_i^4 &= \Sigma\lambda_i^2\, y_i \end{aligned}\right\} \qquad (13)$$

There are three variables and three equations. These are solved for values a_0,a_1 and a_2 and are given in Table 2.

Table 2. Computation of a_0, a_1, and a_2.

G/G/1 model
a_0=1483.5
a_1=0.873377
a_2=0.000006

Then, the queue length for G/G/1 is given as under:

$$y = 1483.5 + 0.873377\lambda + 0.000006\lambda^2 \qquad (14)$$

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 3.

Table 3. Computation of queue length using regression technique.

Rate of Arrivals	Q(L) for M/M/1	Q(L) for G/G/1
10^{10}	10^{10}	1.1076×10^{10}
10^{11}	10^{11}	1.1076×10^{11}
10^{12}	10^{12}	1.1076×10^{12}

10. Conclusion

We proposed that for internet gateway server(s), the buffer estimation is contemporary need in designing 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 gateway 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 internet gateway server.

The study confirms that the system must be such that at each internet gateway we 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] Barford P. and Crovella M., "Generating Representative Web Workloads for Network and Server Performance Evaluation," *Measurement and Modeling of Computer Systems*, pp. 151-160, 1998.

[2] Chen X. and Mohapatra P., "Performance Evaluation of Service Differentiating Internet Servers," *IEEE Transactions on Computers*, vol. 51, no. 11, pp. 1368-1375, 2002.

[3] Kleinrock L., "Queueing Systems," *Applications,* John Wiley, vol. 2, 1976.

[4] Lazowska E., Zahorjan J., Graham G., and Sevcik K., *(Eds.)*, *Quantitative System Performance: Computer System Analysis Using Queueing Network Models*, Prentice-Hall, Inc., 1984.

[5] Low S. and Srikant R., "A Mathematical Framework for Designing a Low-Loss, Low-Delay Internet," *IEEE Transactions on Communications*, 2002.

[6] Menasce D. and Almeida V., *Capacity Planning for Web Services: Metrics, Models, and Methods*, Prentice Hall, 2001.

[7] Mogul J., "Operating Systems Support for Busy Internet Servers," *in the 5th Workshop on Hot Topics in Operating Systems (HotOS-V)*, Orcas Island, WA, 1995.

[8] Singh L. and Srivastava R., "Estimation of Buffer Size of Internet Gateway Server via G/M/1 Queuing Model," *International Journal of Applied Science, Engineering and Technology*, vol. 4, no. 1, pp. 474-482, 2007.

[9] UDDI, "*Universal Description, Discovery, and Integration of Business for the Web*," 2004.

[10] Ying L., Dullerud G., and Srikant R., "Global Stability of Internet Congestion Controllers with Heterogeneous Delays," *IEEE Transactions on Networking*, vol. 14, no. 3, pp. 579-590, 2006.

[11] Zari M., Saiedain H., and Naeem M., "Understanding and Reducing Web delays," *IEEE Journal for Electronics and Computer Science*, vol. 34, no. 12, pp. 30-37, 2001.

Laxmi Singh is presently professor and head of Department of Solid State Physics and Electronics, Dr. RML. Avadh University, Faizabad, India. He is also the director of Institute of Engineering and Technology, University Campus, Dr. RML Avadh University, Faizabad.

Riktesh Srivastava received his MS degree in electronics in 1998. Since 2004, he is working towards his PhD at the Department of Solid State Physics and Electronics of the Dr. RML Avadh University. His research is focused on internet server buffer estimation using queuing theory.