Performance analysis of hybrid (M/M/1 and M/M/m) client server model using Queuing theory

Saptarshi Gupta [1], Dr Rajan John [2], Tannia Rubavathy [3]
[1]*Department of Electronics and Communication Engineering*
[2 3]*Department of Computer Science & Engineering*
[1 2 3] *All Nations University, Koforidua, Ghana, West Africa*
[1] *Email- ece.saptarshi@gmail.com*

Abstract- Internet use packet switching and it is called delay system. When any request comes from client side, server may serve that request immediately or it goes into queue for some time. A client is the computer, which requests the resources (mail, audio, video etc), equipped with a user interface (usually a web browser) for presentation purposes. DNS (Domain name server) will map the web address to its corresponding Internet protocol address. All communication takes place using transfer of packets. Packets arrive according to a Poisson process with rate λ. Router will route the request to that particular Internet Protocol (IP) of the application server. The application server task is to provide the requested resources (mail, audio, video, authentication), but by calling on another server (Data server), which provides the application server with the data it requires. This paper deals with single server and multiple server queues. This paper intends to find out the Performance (average queue length, average response time, average waiting time) analysis of hybrid (M/M/1, M/M/m) client server model using queuing theory.

Keywords – Queuing theory, Client server model, Single server queue, Multiple server queues.

I. INTRODUCTION

In 3-tier architecture, there is an intermediary level, meaning the architecture is generally split up between:

a) A client, i.e. the computer, which requests the resources, equipped with a user interface (usually a web browser) for presentation purposes.
b) The application server (also called middleware), whose task it is to provide the requested resources, but by calling on another server.
c) The data server, which provides the application server with the data it requires.

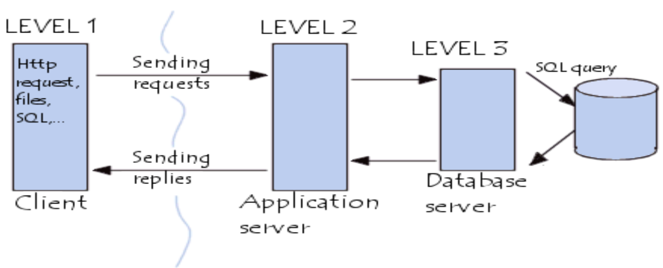

Figure 1.1 Three-tier architecture of client server model [1]

In 3-tier architecture, each server (tier 2 and 3) performs a specialized task (a service). A server can therefore use services from other servers in order to provide own service. As a result, 3-tier architecture is potentially an n-tiered architecture [1].

However, Web traffic is highly dynamic and volatile. 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” [3]. It has been implicated that memory passes through diverse situations of Queue Models, i.e., M/M/1, M/M/m. M/M/1 model is most disciplined and can be analyzed analytically to estimate the queuing parameters [3]. The standard queuing notation, A/B/C, ‘A’ represents the arrival distribution, ‘B’ the service distribution, and ‘C’ the number of servers. For M/M/1, model number of server is 1 and for M/M/m model ‘m’ denotes multiple servers. ‘M’ means “memory less”, which in this context implies Poisson distribution for arrival rates and exponential distribution for service times [2].

II. HYBRID QUEUING ARCHITECTURE FOR CLIENT SERVER MODEL

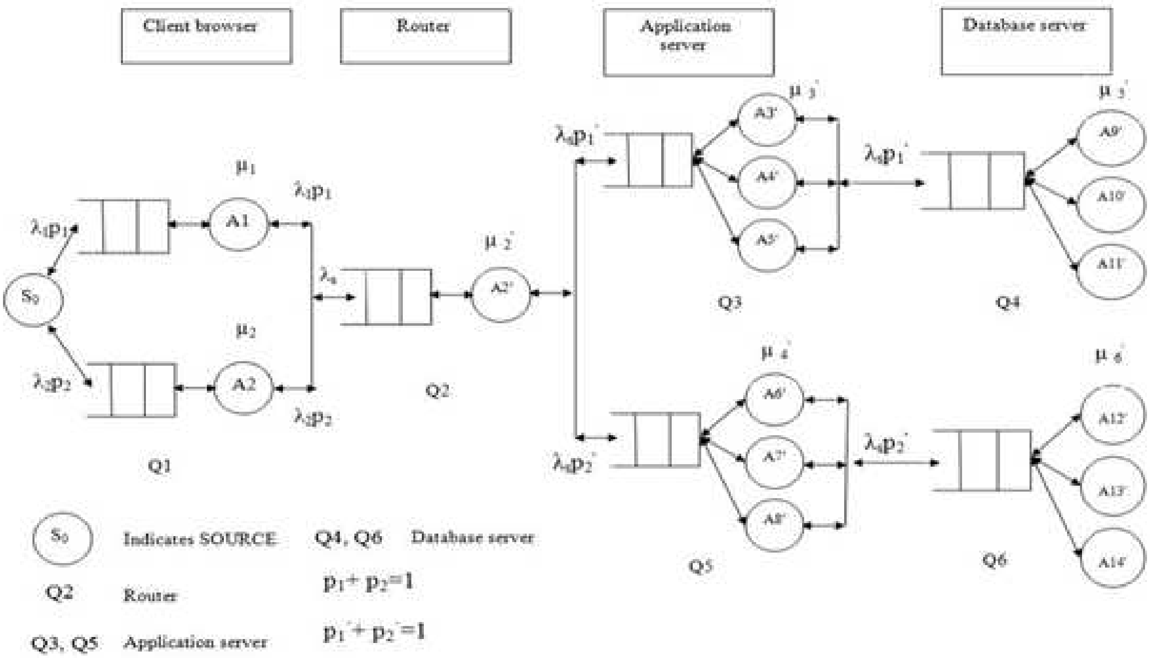

Figure 2.1 Hybrid queuing (M/M/1, M/M/m) model architecture (client server model)

There are 2-inputs in the queuing network considered in Figure 2.1 with the arrivals at the 2-inputs being $\lambda_1 p_1$ and $\lambda_2 p_2$. The service rates are μ_1 and μ_2. The arrival rate at source (S_0) is λ. The probabilities of arrivals at source input S_0 are p_1, p_2 respectively.

Let λ_1 be the arrival rate at first queue, λ_2 be the arrival at second queue at the client browser. Let the service rate of the servers A1 and A2 be μ_1 and μ_2 respectively. Client browser will map the web address to its corresponding Internet protocol address.

After getting serviced by server A1 and A2, the job arrives at queue Q2 (Router). Let λ_s be the arrival rate of the jobs at Q2 (Router), so the arrival rate at Q2 is λ_s. The service rate of the server A2′ is μ_2'. Here, Q2 (router) checks the

International Journal of Electronics and Computer Science Engineering
WWW.IJECSE.ORG

IP address (Internet protocol address) of the particular application server which is mapped by the DNS (Domain name server) and the router forwards that request to that particular application server.

Upon getting serviced by server A2′, the job arrives at queue Q3 (Application server) or Q5 (Application server). This transaction totally depends on the IP address of Application server which has been mapped by the DNS (Domain name server).

Now, the jobs arrive at the queues Q3, Q5 with probabilities p_1' and p_2', where $p_1' + p_2'$=1. So, the arrival rate at Q3 is $\lambda_s\, p_1'$, and that at Q5 is $\lambda_s\, p_2'$.

If the job arrives at Q3 with probability p_1' with the arrival rate $\lambda_s\, p_1'$ then the arrival rate of Q4 will be $\lambda_s\, p_1'$ because Q3 and Q4 are in serial connection. The service rate of Q3 Application servers A3′, A4′ and A5′ is μ_3' and that of Q4 Database servers A9′, A10′, A11′ is μ_5'.The service rate of Q3 (Application server A3′, A4′ and A5′) are equal because they are M/M/m queues, Likewise the service rates of Q4 (Database server A9′, A10′ and A11′) are equal because they are M/M/m queues as well.

Finally, jobs after service completion at servers Q2 and Q3 arrive at the sink Q4 (database), the database server, which provides the application server Q3 with the data it requires. Here, Application server acts like a web server which serves pages for viewing in a Web browser.

If the job arrives at Q5 with probability p_2' with the arrival rate $\lambda_s\, p_2'$ then the arrival rate of Q6 will be $\lambda_s\, p_2'$ because Q5 and Q6 are in serial connection. The service rate of the Q5 Application servers A6′, A7′ and A8′ is μ_4' and the service rate of Q6 Database servers A12′, A13′, A14′ is μ_6'.The service rate of Q5 (Application server A6′, A7′ and A8′) are equal because they are M/M/m queue, likewise the service rate of Q6 (Database server A12′, A13′ and A14′) are equal because they are M/M/m queues as well. Here m is the number of servers.

Finally, jobs after service completion at servers Q2 and Q5 arrive at sink Q6 (database), the database server, which provides the application server Q5 with the data it requires. Here, Application server acts like a web server which serves pages for viewing in a Web browser.

2.1. Performance measures for hybrid queuing client server model

The performance of the single and multiple servers is measured by the average queue lengths, average waiting time, average response time and the average no of jobs in the system [4]. The queues in the model are assumed to be M/M/1 and M/M/m (Hybrid). Here client browsers (Q1) and router (Q2) are M/M/1 type. Application servers (Q3 and Q5) and database servers (Q4 and Q6) are M/M/m type.

Performance measures such as (a) average queue length, (b) average response time, (c) average waiting time are derived in this section.

2.1.1. Average queue lengths

The average queue lengths in A1 and A2 (client browser) are

$$E[N_1^{(A1)}] = \frac{\rho_1^{2(A1)}}{1-\rho_1^{(A1)}} = \frac{(\lambda_1 \mathrm{p}_1)^2}{\mu_1(\mu_1 - \lambda_1 \mathrm{p}_1)},\quad E[N_2^{(A2)}] = \frac{\rho_2^{2(A2)}}{1-\rho_2^{(A2)}} = \frac{(\lambda_2 \mathrm{p}_2)^2}{\mu_2(\mu_2 - \lambda_2 \mathrm{p}_2)}. \tag{2.5}$$

Similarly, the average queue length in A2′ (router) is $E[N_2'^{(A2')}] = \dfrac{\rho_2'^{2(A2')}}{1-\rho_2'^{2(A2')}} = \dfrac{\lambda_s^{\,2}}{\mu_2'(\mu_2' - \lambda_s)}$. (1)

IJECSE
International Journal of Electronics and Computer Science Engineering
WWW.IJECSE.ORG

The average queue length in application servers and database servers (Q3, Q5), (Q4, Q6) is $E[N_Q] = \dfrac{\rho P_Q}{1-\rho}$, where, probability that an arriving customer has to wait in queue (m customers or more in the system) is $P_Q = \dfrac{p_0 (m\rho)^m}{m!(1-\rho)}$, and in the M/M/m case, the Probability of n customers in the system is $p_0 = \left[\sum_{k=0}^{m-1} \dfrac{(m\rho)^k}{k!} + \dfrac{(m\rho)^m}{m!(1-\rho)} \right]^{-1}$.

According to Figure2.1 the application server Q3 and Q5 consist of 3 sub servers each so here m=3 for each Q3 and Q5.

After simplification the average queue length for application server Q3 is $E[N_{Q3}] = \dfrac{\lambda_S p_1' P_{Q3}}{3\mu_3' - \lambda_S p_1'}$. (2)

Where, probability that an arriving customer has to wait in queue is

$P_{Q3} = \dfrac{P_{03}(\lambda_S p_1')^3}{2(3\mu_3' - \lambda_S p_1')\mu_3'^2}$, and in the M/M/m case, the Probability of n customers in the system is

$$P_{03} = \frac{2\mu_3'^2(\mu_3' - \lambda_S p_1') + 2\lambda_S p_1' \mu_3'(\mu_3' - \lambda_S p_1') + (\lambda_S p_1')^2(\mu_3' - \lambda_S p_1') + (\lambda_S p_1')^3}{2\mu_3'^2(\mu_3' - \lambda_S p_1')}.$$

After simplification the average queue length for application server Q5 is $E[N_{Q5}] = \dfrac{\lambda_S p_2' P_{Q5}}{3\mu_4' - \lambda_S p_2'}$. (3)

Where, $P_{Q5} = \dfrac{P_{05}(\lambda_s p_2')^3}{2(3\mu_4' - \lambda_S p_2')\mu_4'^2}$ and $P_{05} = \dfrac{2\mu_4'^2(\mu_4' - \lambda_S p_2') + 2\lambda_S p_2' \mu_4'(\mu_4' - \lambda_S p_2') + (\lambda_S p_2')^2(\mu_4' - \lambda_S p_2') + (\lambda_S p_2')^3}{2\mu_4'^2(\mu_4' - \lambda_S p_2')}$.

According to Figure 2.1 database server Q4 and Q6 consist of 3 sub servers each so here m=3 for each Q4 and Q6.

After simplification the average queue length for database server Q4 is $E[N_{Q4}] = \dfrac{\lambda_S p_1' P_{Q4}}{3\mu_5' - \lambda_S p_1'}$. (4)

Where, $P_{Q4} = \dfrac{P_{04}(\lambda_s p_1'^3)}{2(3\mu_5' - \lambda_S p_1')\mu_5'^2}$ and $P_{04} = \dfrac{2\mu_5'^2(\mu_5' - \lambda_S p_1') + 2\lambda_S p_1' \mu_5'(\mu_5' - \lambda_S p_1') + (\lambda_S p_1')^2(\mu_5' - \lambda_S p_1') + (\lambda_S p_1')^3}{2\mu_5'^2(\mu_5' - \lambda_S p_1')}$.

After simplification the average queue length for database server Q6 is $E[N_{Q6}] = \dfrac{\lambda_S p_2' P_{Q5}}{3\mu_6' - \lambda_S p_2'}$. (5)

Where, $P_{Q6} = \dfrac{P_{06}(\lambda_s p_2')^3}{2(3\mu_6' - \lambda_S p_2')\mu_6'^2}$ and $P_{06} = \dfrac{2\mu_6'^2(\mu_6' - \lambda_S p_2') + 2\lambda_S p_2' \mu_6'(\mu_6' - \lambda_S p_2') + (\lambda_S p_2')^2(\mu_6' - \lambda_S p_2') + (\lambda_S p_2')^3}{2\mu_6'^2(\mu_6' - \lambda_S p_2')}$.

2.1.2. Average response times

The average response time in A1 and A2 are

$$E[R_1^{(A1)}] = E[W_1^{(A1)}] + \frac{1}{\mu_1} = \frac{\lambda_1 p_1}{\mu_1(\mu_1 - \lambda_1 p_1)} + \frac{1}{\mu_1}, \quad E[R_2^{(A2)}] = E[W_2^{(A2)}] + \frac{1}{\mu_2} = \frac{\lambda_2 p_2}{\mu_2(\mu_2 - \lambda_2 p_2)} + \frac{1}{\mu_2}.$$

Similarly, the average response time in A2′ (router) is $E[R_2'^{(A2')}] = E[W_2'^{(A2')}] + \frac{1}{\mu_2'} = \frac{\lambda_s}{\mu_2'(\mu_2'-\lambda_s)} + \frac{1}{\mu_2'}$. (6)

The average response time in application servers and database servers (Q3, Q5), (Q4, Q6) be E[R] $= \frac{\rho P_Q}{\lambda(1-\rho)} + \frac{1}{\mu}$,

where, probability that an arriving customer has to wait in queue (m customers or more in the system) is $P_Q = \frac{p_0 (m\rho)^m}{m!(1-\rho)}$

, and in M/M/m case Probability of *n* customers in the system is $p_0 = \left[\sum_{k=0}^{m-1} \frac{(m\rho)^k}{k!} + \frac{(m\rho)^m}{m!(1-\rho)}\right]^{-1}$.

According to Figure2.1 application server Q3 and Q5 consist of 3 sub servers each so here m=3 for each Q3 and Q5. After simplification the average response time for application server Q3 is

$$E[R_{Q3}] = \frac{P_{Q3}}{3\mu_3' - \lambda_S p_1'} + \frac{1}{\mu_3'}. \quad (7)$$

Where, $P_{Q3} = \frac{P_{03}(\lambda_s p_1')^3}{2(3\mu_3' - \lambda_s p_1')\mu_3'^2}$ and $P_{03} = \frac{2\mu_3'^2(\mu_3'-\lambda_S p_1') + 2\lambda_S p_1'\mu_3'(\mu_3'-\lambda_S p_1') + (\lambda_S p_1')^2(\mu_3'-\lambda_S p_1') + (\lambda_S p_1')^3}{2\mu_3'^2(\mu_3'-\lambda_S p_1')}$.

After simplification the average response time for application server Q5 is $E[R_{Q5}] = \frac{P_{Q5}}{3\mu_4' - \lambda_S p_2'} + \frac{1}{\mu_4'}$. (8)

Where, $P_{Q5} = \frac{P_{05}(\lambda_s p_2')^3}{2(3\mu_4' - \lambda_S p_2')\mu_4'^2}$ and $P_{05} = \frac{2\mu_4'^2(\mu_4'-\lambda_S p_2') + 2\lambda_S p_2'\mu_4'(\mu_4'-\lambda_S p_2') + (\lambda_S p_2')^2(\mu_4'-\lambda_S p_2') + (\lambda_S p_2')^3}{2\mu_4'^2(\mu_4'-\lambda_S p_2')}$.

According to Figure 2.1 the database server Q4 and Q6 consist of 3 sub servers each so here m=3 for each Q4 and Q6.

After simplification the response time for database server Q4 is $E[R_{Q4}] = \frac{P_{Q4}}{3\mu_5' - \lambda_S p_1'} + \frac{1}{\mu_5'}$. (9)

Where, $P_{Q4} = \frac{P_{04}(\lambda_s p_1')^3}{2(3\mu_5' - \lambda_s p_1')\mu_5'^2}$ and $P_{04} = \frac{2\mu_5'^2(\mu_5'-\lambda_S p_1') + 2\lambda_S p_1'\mu_5'(\mu_5'-\lambda_S p_1') + (\lambda_S p_1')^2(\mu_5'-\lambda_S p_1') + (\lambda_S p_1')^3}{2\mu_5'^2(\mu_5'-\lambda_S p_1')}$.

After simplification the average response time for database server Q6 is $E[R_{Q6}] = \frac{P_{Q5}}{3\mu_6' - \lambda_S p_2'} + \frac{1}{\mu_6'}$. (10)

Where, $P_{Q6} = \frac{P_{06}(\lambda_s p_2')^3}{2(3\mu_6' - \lambda_S p_2')\mu_6'^2}$ and $P_{06} = \frac{2\mu_6'^2(\mu_6'-\lambda_S p_2') + 2\lambda_S p_2'\mu_6'(\mu_6'-\lambda_S p_2') + (\lambda_S p_2')^2(\mu_6'-\lambda_S p_2') + (\lambda_S p_2')^3}{2\mu_6'^2(\mu_6'-\lambda_S p_2')}$.

2.1.3. Average waiting time:

The average waiting time in A1 and A2 (client browser) are

$$E[W_1^{(A1)}] = \frac{E[N_1^{(A1)}]}{\lambda_1 p_1} = \frac{\lambda_1 p_1}{\mu_1(\mu_1 - \lambda_1 p_1)}, \quad E[W_2^{(A2)}] = \frac{E[N_2^{(A2)}]}{\lambda_2 p_2} = \frac{\lambda_2 p_2}{\mu_2(\mu_2 - \lambda_2 p_2)}.$$

Similarly, the average waiting time in A2′ (router) is $E[W_2'^{(A2')}] = \dfrac{E[N_2'^{(A2')}]}{\lambda_s} = \dfrac{\lambda_s}{\mu_2'(\mu_2' - \lambda_s)}$. (11)

The average waiting time in application servers and database servers (Q3, Q5), (Q4, Q6) be E[W] = $\dfrac{\rho P_Q}{\lambda(1-\rho)}$, where, probability that an arriving customer has to wait in queue (m customers or more in the system) is $P_Q = \dfrac{p_0 (m\rho)^m}{m!(1-\rho)}$ and in M/M/m case Probability of n customers in the system is $p_0 = \left[\sum_{k=0}^{m-1} \dfrac{(m\rho)^k}{k!} + \dfrac{(m\rho)^m}{m!(1-\rho)}\right]^{-1}$.

According to Figure 2.1 the application server Q3 and Q5 consist of 3 sub servers each so here m=3 for each Q3 and Q5. After simplification the average waiting time for application server Q3 is

$$E[W_{Q3}] = \frac{P_{Q3}}{3\mu_3' - \lambda_S p_1'}. \quad (12)$$

Where, $P_{Q3} = \dfrac{P_{03}(\lambda_s p_1')^3}{2(3\mu_3' - \lambda_S p_1')\mu_3'^2}$ and $P_{03} = \dfrac{2\mu_3'^2(\mu_3' - \lambda_S p_1') + 2\lambda_S p_1' \mu_3'(\mu_3' - \lambda_S p_1') + (\lambda_S p_1')^2(\mu_3' - \lambda_S p_1') + (\lambda_S p_1')^3}{2\mu_3'^2(\mu_3' - \lambda_S p_1')}$.

After simplification the average waiting time for application server Q5 is $E[W_{Q5}] = \dfrac{P_{Q5}}{3\mu_4' - \lambda_S p_2'}$. (13)

Where, $P_{Q5} = \dfrac{P_{05}(\lambda_s p_2')^3}{2(3\mu_4' - \lambda_S p_2')\mu_4'^2}$ and $P_{05} = \dfrac{2\mu_4'^2(\mu_4' - \lambda_S p_2') + 2\lambda_S p_2' \mu_4'(\mu_4' - \lambda_S p_2') + (\lambda_S p_2')^2(\mu_4' - \lambda_S p_2') + (\lambda_S p_2')^3}{2\mu_4'^2(\mu_4' - \lambda_S p_2')}$.

According to Figure 2.1 the database server Q4 and Q6 consist of 3 sub servers each so here m=3 for each Q4 and Q6.

After simplification the average waiting time for database server Q4 is $E[W_{Q4}] = \dfrac{P_{Q4}}{3\mu_5' - \lambda_S p_1'}$. (14)

Where, $P_{Q4} = \dfrac{P_{04}(\lambda_s p_1')^3}{2(3\mu_5' - \lambda_S p_1')\mu_5'^2}$ and $P_{04} = \dfrac{2\mu_5'^2(\mu_5' - \lambda_S p_1') + 2\lambda_S p_1' \mu_5'(\mu_5' - \lambda_S p_1') + (\lambda_S p_1')^2(\mu_5' - \lambda_S p_1') + (\lambda_S p_1')^3}{2\mu_5'^2(\mu_5' - \lambda_S p_1')}$.

After simplification the average waiting time for database server Q6 is $E[W_{Q6}] = \dfrac{P_{Q6}}{3\mu_6' - \lambda_S p_2'}$. (15)

Where, $P_{Q6} = \dfrac{P_{06}(\lambda_s p_2')^3}{2(3\mu_6' - \lambda_S p_2')\mu_6'^2}$ and $P_{06} = \dfrac{2\mu_6'^2(\mu_6' - \lambda_S p_2') + 2\lambda_S p_2' \mu_6'(\mu_6' - \lambda_S p_2') + (\lambda_S p_2')^2(\mu_6' - \lambda_S p_2') + (\lambda_S p_2')^3}{2\mu_6'^2(\mu_6' - \lambda_S p_2')}$.

2.1.4. Total number of jobs, response times and waiting times for different paths:

For path (Q1 (A1), Q2, Q3, Q4), the total number of jobs from 1st browser to data base server is

$$E[N_x] = E[N_1^{(A1)}] + E[N_2'^{(A2')}] + E[N_{Q3}] + E[N_{Q4}]$$

$$= \frac{(\lambda_1 p_1)^2}{\mu_1(\mu_1 - \lambda_1 p_1)} + \frac{\lambda_s^2}{\mu_2'(\mu_2' - \lambda_s)} + \frac{\lambda_S p_1' P_{Q3}}{3\mu_3' - \lambda_S p_1'} + \frac{\lambda_S p_1' P_{Q4}}{3\mu_5' - \lambda_S p_1'}. \quad (16)$$

For path (Q1 (A1), Q2, Q3, Q4), the total response time from 1st browser to data base server is

E [Rx] = $E[R_1^{(A1)}] + E[R_2'^{(A2')}]$ + E[R_{Q3}] + E[R_{Q4}]

$$= \frac{\lambda_1 p_1}{\mu_1(\mu_1 - \lambda_1 p_1)} + \frac{1}{\mu_1} + \frac{\lambda_s}{\mu_2'(\mu_2' - \lambda_s)} + \frac{1}{\mu_2'} + \frac{P_{Q3}}{3\mu_3' - \lambda_S p_1'} + \frac{1}{\mu_3'} \frac{P_{Q4}}{3\mu_5' - \lambda_S p_1'} + \frac{1}{\mu_5'} . \quad (17)$$

For path (Q1 (A1), Q2, Q3, Q4), the total waiting time from 1st browser to data base server is

E [Wx] = $E[W_1^{(A1)}] + E[W_2'^{(A2')}]$ + E[W_{Q3}] + E[W_{Q4}]

$$= \frac{\lambda_1 p_1}{\mu_1(\mu_1 - \lambda_1 p_1)} + \frac{\lambda_s}{\mu_2'(\mu_2' - \lambda_s)} + \frac{P_{Q3}}{3\mu_3' - \lambda_S p_1'} + \frac{P_{Q4}}{3\mu_5' - \lambda_S p_1'} . \quad (18)$$

For path (Q1 (A2), Q2, Q5, Q6), the total number of jobs from 2nd browser to data base server is

E [Ny] = $E[N_2^{(A2)}] + E[N_2'^{(A2')}]$ + E[N_{Q5}] + E[N_{Q6}]

$$= \frac{(\lambda_2 p_2)^2}{\mu_2(\mu_2 - \lambda_2 p_2)} + \frac{\lambda_s^{\ 2}}{\mu_2'(\mu_2' - \lambda_s)} + \frac{\lambda_S p_2' P_{Q5}}{3\mu_4' - \lambda_S p_2'} + \frac{\lambda_S p_2' P_{Q5}}{3\mu_6' - \lambda_S p_2'} . \quad (19)$$

For path (Q1 (A2), Q2, Q5, Q6), the total response time from 2nd browser to data base server is

E [Rx] = $E[R_2^{(A2)}] + E[R_2'^{(A2')}]$ + E[R_{Q5}] + E[R_{Q6}]

$$= \frac{\lambda_2 p_2}{\mu_2(\mu_2 - \lambda_2 p_2)} + \frac{1}{\mu_2} + \frac{\lambda_s}{\mu_2'(\mu_2' - \lambda_s)} + \frac{1}{\mu_2'} + \frac{P_{Q5}}{3\mu_4' - \lambda_S p_2'} + \frac{1}{\mu_4'} + \frac{P_{Q5}}{3\mu_6' - \lambda_S p_2'} + \frac{1}{\mu_6'} . \quad (20)$$

For path (Q1 (A2), Q2, Q5, Q6), the total waiting time from 2nd browser to data base server is

E [Wx] = $E[W_2^{(A2)}] + E[W_2'^{(A2')}]$ + E[W_{Q5}] + E[W_{Q6}]

$$= \frac{\lambda_2 p_2}{\mu_2(\mu_2 - \lambda_2 p_2)} + \frac{\lambda_s}{\mu_2'(\mu_2' - \lambda_s)} + \frac{P_{Q5}}{3\mu_4' - \lambda_S p_2'} + \frac{P_{Q6}}{3\mu_6' - \lambda_S p_2'} . \quad (21)$$

2.2. Numerical results for Hybrid queuing client server model

2.2.1. Queue length vs. Arrival rate for link Q1 (A1) to Q4 and link Q1 (A2) to Q6

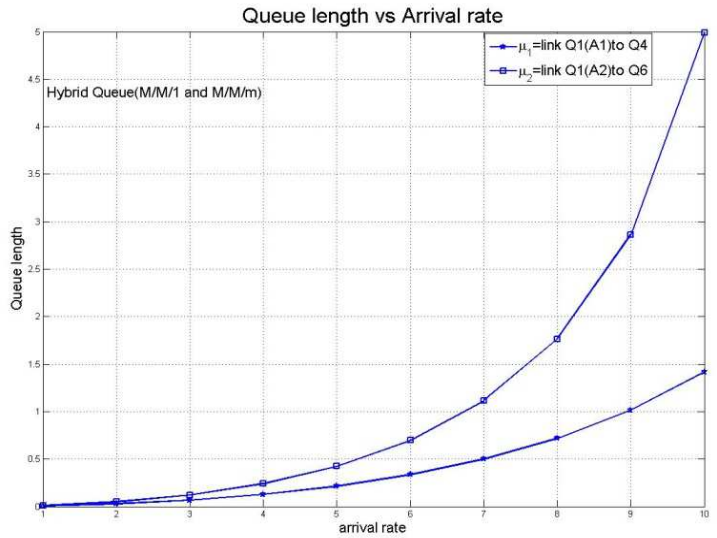

Figure 2.2 Queue length vs. Arrival rate for link Q1 (A1) to Q4 and link Q1 (A2) to Q6.

Let λ be the total number of arrivals in the 2-input queuing network. In the example considered in this section, the arrival rate, λ=1, 2 ,…, 10. The other specifications include

- Probability of arrivals at queues Q1 (A1) and Q1 (A2) are 0.4 and 0.6.
- The service rate specifications of different servers in the network are μ_1= 7.8521, μ_2=7.2356, μ_2' = 16.7454, μ_3' = 13.7303, μ_5' = 12.1244, μ_4' =8.5400, μ_6' = 10.3267.
- Probability of arrivals at queues Q3 (A3', A4' and A5') and Q5 (A6', A7' and A8') are 0.4 and 0.6

For each value of λ, the average queue lengths in all nodes of the 2-input queuing client server network is computed. The average queue lengths in paths Q1 (A1) to Q4 and Q1 (A2) to Q6 are computed from (1) to (5). The total queue lengths in paths Q1 (A1) to Q4 and Q1 (A2) to Q6 are computed in (16) and (19) respectively. The total queue lengths of link Q1 (A1) to Q4 and link Q1 (A2) to Q6 vs. different arrivals rate are plotted in Figure 2.2. From Figure 2.2 it is found that as the arrival rate increases, the queue length of the link Q1 (A2) to Q6 increases more than the queue length of the link Q1 (A1) to Q4. Because the service rates for server Q3 (μ_3') and Q4 (μ_5') are more than the service rates of Q5 (μ_4') and Q6 (μ_6').

2.2.2. Response time vs. Arrival rate for link Q1 (A1) to Q4 and link Q1 (A2) to Q6

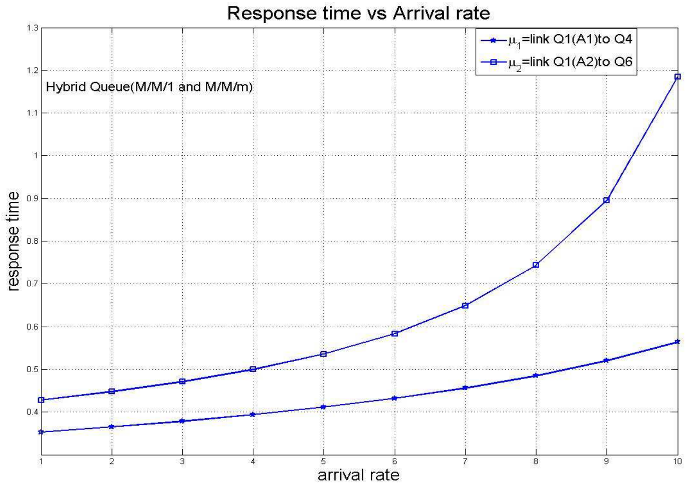

Figure 2.3 Response time vs. Arrival rate for link Q1 (A1) to Q4 and link Q1 (A2) to Q6

For each value of λ, the utilizations, average response times in all nodes of the 2-input queuing client server network is computed. The average response times in paths Q1 (A1) to Q4 and Q1 (A2) to Q6 are computed from (6) to (10). The total response times in paths Q1 (A1) to Q4 and Q1 (A2) to Q6 are computed in (17) and (20). The total response times of link Q1 (A1) to Q4 and link Q1 (A2) to Q6 vs. different arrivals rate are plotted in Figure 2.3. From Figure 2.3 it is found that as the arrival rate increases, the response time of the link Q1 (A2) to Q6 increases more than the response time of the link Q1 (A1) to Q4. Because the service rates for server Q3 (μ_3') and Q4 (μ_5') are more than the service rates of Q5 (μ_4') and Q6 (μ_6'), **(2.2.1).**

2.2.3. Waiting time vs. Arrival rate for link Q1 (A1) to Q4 and link Q1 (A2) to Q6

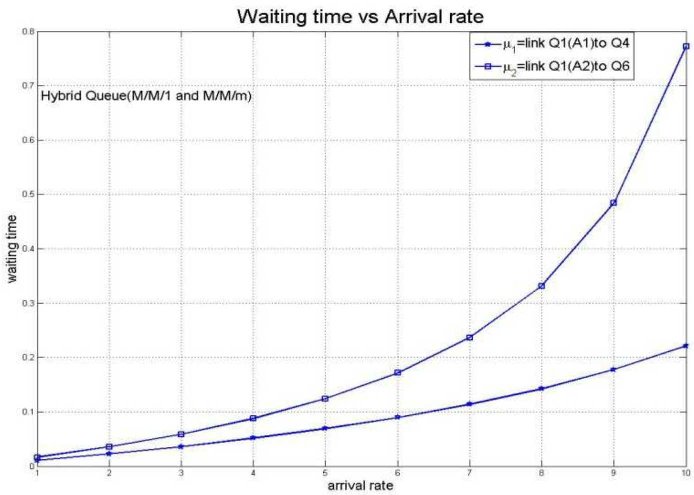

Figure 2.4 Waiting time vs. Arrival rate for link Q1 (A1) to Q4 and link Q1 (A2) to Q6

For each value of λ, the utilizations, average waiting times in all nodes of the 2-input queuing client server network are computed. The average waiting times in paths Q1 (A1) to Q4 and Q1 (A2) to Q6 are computed from (11) to (15). The total waiting times in paths Q1 (A1) to Q4 and Q1 (A2) to Q6 are computed in (18) and (21). The total waiting times of link Q1 (A1) to Q4 and link Q1 (A2) to Q6 vs. different arrivals rate are plotted in Figure 2.4. From Figure 2.4 it is found that as the arrival rate increases, the waiting time of the link Q1 (A2) to Q6 increases more than the waiting time of the link Q1 (A1) to Q4. Because the service rates for server Q3 (μ_3') and Q4 (μ_5') are more than the service rates of Q5 (μ_4') and Q6 (μ_6'), **(2.2.1).**

III.CONCLUSION

In this paper, performance measures for client server model such as average queue lengths, average response times and average waiting times are derived for M/M/1 and M/M/m (Hybrid) type model and the results shows that response time is less. Because of this less response time probability of packet loss reduce and system become faster. The service rate of the equivalent server is also derived and computed numerically. Queue lengths vs. Arrival rate, Response time vs. Arrival rate, waiting time vs. Arrival rate are plotted using MATLAB 7.5.0 Software. Decision for routing is made at the last node in each stage of the network as to which path to choose for obtaining the least response time.

IV. REFERENCE

[1] http://en.kioskea.net/contents/cs/cs3tier.php3#q=Security+in+client%2Fserver+model&cur=2&url=%2F
[2] Dimitri Bertsekas, Robert Gallager, *Data Networks*, Second Edition, Prentice Hall of India Private Limited, 1997.
[3] Singh, and Srivastava, "Memory Estimation of Internet Server Using Queuing Theory: Comparative Study between M/G/1, G/M/1 & G/G/1 Queuing Model", *World Academy of Science, Engineering and Technology*, 33 2007 pp. 153-157.
[4] Vidhyacharan Bhaskar & Patrick Lallement, "Activity routing in a distributed supply chain: Performance evaluation with two inputs". *Journal of Network and Computer Applications*, *Science Direct*, 31 (2008) 402–428.