Malaysian Journal of Computer Science, Vol. 9 No. 1, June 1996, pp. 6-11

A LINEAR ALGEBRAIC APPROACH IN ANALYZING THE M/GE/1 AND GE/M/1 QUEUING SYSTEMS AT EQUILIBRIUM

Mohammad Saleh bin Mustafa
Computer Science Department
43400 Universiti Pertanian Malaysia
Serdang, Selangor
Malaysia
email: mohammad@cs.upm.my

Abu Talib bin Othman
Computer Centre
43400 Universiti Pertanian Malaysia
Serdang, Selangor
Malaysia
email: abu@pp.ccrisc.upm.my

ABSTRACT

Uses the algebraic approach in the queuing theory to derive the M/G/1 equilibrium solution for the number of jobs in the system when the probability distribution function representing the general distribution is the generalized exponential (GE-type). Similarly the GE/M/1 system is solved. Furthermore, it has been shown that as expected the solutions are equivalent to the maximum entropy solutions of the M/G/1 and G/M/1 systems respectively at equilibrium.

Keywords: ***Queuing systems, Performance evaluation, Computer networks, Linear algebraic approach.***

1.0 INTRODUCTION

Transforming the queuing system problems from one of the integral equations to one of the algebraic equations over a finite dimensional vector space [1] is called the algebraic approach in queuing theory. A system of the algebraic approach can be described by representing each non-exponential service time (or interarrival time) distribution by a collection of exponential servers, with the constraint that the collection can only be accessed by one customer at a time [2].

The algebraic approach is one of the most important tools for modeling and analyzing the queuing systems. Many problems in the queuing theory that are traditionally solved by unrelated mathematical technique can now be solved in a consistent integrated fashion. On the other hand, many systems performance measures which are normally ignored because of their computational and formulation difficulties can be dealt with easily in the linear algebraic approach of the queuing theory. And any problem which can be cast into matrix-vector format can easily be adapted to make use of the high-speed parallel and vector processor available today [3].

Moreover, with the algebraic approach queuing systems can be given formal descriptions which are more easily understandable and more easily modifiable than the ones obtained by many other models.

The M/G/1 is a FCFS queuing system with Poisson arrivals and general service time distribution while the G/M/1 is of G-type arrival pattern and single exponential server. The equilibrium solutions for the number of jobs in the systems vary with the probability distribution function chosen to represent the general distribution, and the system performance is affected by the distribution form of interarrival and service times. The GE distribution represents a 'natural' information theoretic choice for the G-type and of course, the GE pdf is an 'external' case of a family of distributions with a mixture of exponential models, e.g. hyperexponential-2 (H_2). The interest in the GE distribution, as a universal model, to approximate G-type distributions (when only the first two moments are known) is motivated further by some of its robust and versatile properties, which make it particularly useful for the analysis of multiserver queues and general queuing networks [4].

There is a very large number of two-stage models representing the G-type distribution with the same mean value and coefficient of variation. That means, if the mean and coefficient of variation of the two-stage system are given to be fixed constants and equal to the mean and coefficient of variation of a given G-type distribution, then solving the equations of the two-stage system will result in having a solution with a free parameter which means infinitely many solutions since this parameter can take infinitely many values.

The procedure we follow is to write the algebraic solution of the $M/H_2/1$ queuing system, then we change the parameters of the hyperexponential function [5], by new parameter k and the two parameters: the mean and coefficient of variation. After that we follow the convergence of the two-stage function to the GE distribution when the parameter k approaches infinity [4]. Similarly, the work is done for the $H_2/M/1$ queuing system taking the advantage from [2] in representing the system.

Finally we see that when the G-type distribution is presented by the GE model, a linear algebraic approach in analyzing the M/G/1 and G/M/1 queuing systems becomes exact.

2.0 THE M/GE/1 QUEUE USING THE LINEAR ALGEBRAIC APPROACH

First, we represent the equilibrium solution for the queue length in the M/H_2/1 queuing system in an algebraic form, based on [3]. The system is described in fig. 1. There are two phases in S1 with completion rates μ_1 and μ_2. A customer upon entering goes to phase 1 with probability α, or to phase 2 with probability 1-α, and then leaves S1 when finished.

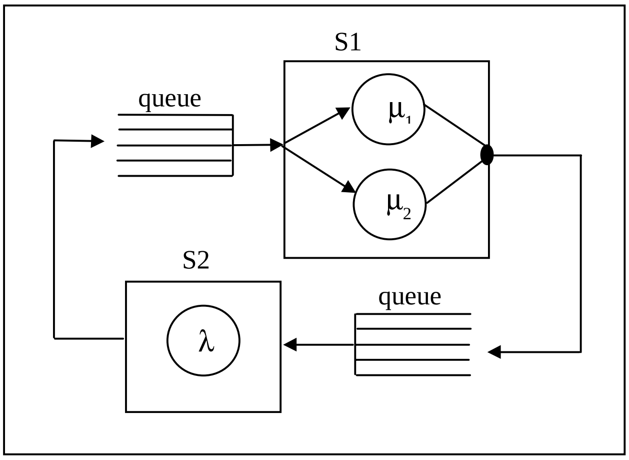

Fig. 1**:** M/H_2/1 queuing system

The following notations will be used:-
p is the ***entrance vector***, whose ith component is the probability that a customer, upon entering S1, will go to phase i.
q' is the ***exit vector*** whose ith component is the probability that a customer, upon completing service at phase i, will leave S1.
P is the ***sub-stochastic matrix***, whose ijth component is the probability that a customer who has just finished service at i will go to j.
M is the ***completion rate matrix***, whose diagonal elements are the completion rates of the individual phases [μi = Mii].
ε' is the vector, whose all elements are ones.
I is the identity matrix.
O is the zero matrix.
r(n) is the steady state probability that there are n customers at S1.
λ is the mean arrival rate.
μ is the mean service rate.
ρ is the utilization factor.
C_s^2 is the squared coefficient of variation for the service time distribution.
C_a^2 is the squared coefficient of variation for the inter-arrival time distribution.

The algebraic solution for the M/H_2/1 queuing system can be written as the solution for M/G/1 system [3].

$$r(n) = (1-\rho)\Psi\left[\mathbf{U}^n\right] \tag{1.1}$$

$$\rho = \frac{\lambda}{\mu} = \lambda\Psi[\mathbf{V}] \tag{1.2}$$

$$\bar{x} = \Psi[\mathbf{V}] = \mathbf{pV}\varepsilon' \tag{1.3}$$

$$\mathbf{V} = \mathbf{B}^{-1} \tag{1.4}$$

$$\mathbf{B} = \mathbf{M}(\mathbf{I}-\mathbf{P}) \tag{1.5}$$

$$\mathbf{U} = \mathbf{A}^{-1} \tag{1.6}$$

$$\mathbf{A} = \mathbf{I} + \left(\frac{1}{\lambda}\right)\mathbf{B} - \mathbf{Q} \tag{1.7}$$

$$\mathbf{Q} = \varepsilon'\mathbf{p} \tag{1.8}$$

Now, we follow the definition of the M/H_2/1 queuing system to construct its algebraic solution:

Since the probability for a customer to enter phase 1 is equal to α, then the probability to enter phase 2 is equal to 1-α since the summation of the entrance probabilities should equal to 1. So, the entrance probability vector is:

$$\mathbf{p} = [\alpha \quad 1-\alpha] \tag{2.1}$$

And the completion rate matrix has the completion rates of the two phases in its main diagonal with the rest of its elements zeros.

$$\mathbf{M} = \begin{bmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{bmatrix} \tag{2.2}$$

A customer entering S1 has the right to be served by one and only one of the two servers, and after finishing service at the chosen server it can not be served by the other one, i.e., service is finished for that customer and it has to exit from S1. In matrices this means that the travel probability matrix is the zero matrix and the exit vector is the unit vector, i.e.,

$$\mathbf{P} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \tag{2.3}$$

$$\mathbf{q}' = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{2.4}$$

Substituting (2.2) and (2.3) in (1.5) we get the service rate matrix **B**,

$$\mathbf{B} = \mathbf{M}\,(\mathbf{I} - \mathbf{P}) = \mathbf{M}\,(\mathbf{I} - \mathbf{O}) = \mathbf{M} \tag{2.5}$$

From (1.4) the service time matrix is:

$$\mathbf{V} = \mathbf{B}^{-1} = \mathbf{M}^{-1} = \begin{bmatrix} \frac{1}{\mu_1} & 0 \\ 0 & \frac{1}{\mu_2} \end{bmatrix} \tag{2.6}$$

The mean service time is given by substituting (2.1) and (2.6) in (1.3),

$$\bar{x} = \Psi[\mathbf{V}] = \mathbf{pV}\varepsilon' \begin{bmatrix} \frac{\alpha}{\mu_1} & \frac{1-\alpha}{\mu_2} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{\alpha}{\mu_1} + \frac{1-\alpha}{\mu_2} \tag{2.7}$$

The transition rate matrix is given by substituting (2.1) in (1.8),

$$\mathbf{Q} = \varepsilon'\mathbf{p} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} \alpha & 1-\alpha \end{bmatrix} = \begin{bmatrix} \alpha & 1-\alpha \\ \alpha & 1-\alpha \end{bmatrix} \tag{2.8}$$

Substituting (2.5) and (2.8) in (1.7) we get the matrix,

$$\mathbf{A} = \mathbf{I} + \left(\frac{1}{\lambda}\right)\mathbf{B} - \mathbf{Q}$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \left(\frac{1}{\lambda}\right) \begin{bmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{bmatrix} - \begin{bmatrix} \alpha & 1-\alpha \\ \alpha & 1-\alpha \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\mu_1}{\lambda} + 1 - \alpha & \alpha - 1 \\ -\alpha & \frac{\mu_2}{\lambda} + \alpha \end{bmatrix} \tag{2.9}$$

Taking the inverse of **A** from (2.9), we get the important matrix,

$$\mathbf{U} = \mathbf{A}^{-1} = \frac{1}{\frac{\mu_1\mu_2}{\lambda^2} + \frac{\alpha\mu_1}{\lambda} + \frac{(1-\alpha)\mu_2}{\lambda}} \times$$

$$\begin{bmatrix} \frac{\mu_2}{\lambda} + \alpha & 1-\alpha \\ \alpha & \frac{\mu_1}{\lambda} + 1 - \alpha \end{bmatrix} \tag{2.10}$$

Finally, the queue length probabilities are given by,

$$r(n) = (1-\rho)\Psi\left[\mathbf{U}^n\right] = (1-\rho)\mathbf{pU}^n\varepsilon'$$

$$= (1-\rho)\begin{bmatrix} \alpha & 1-\alpha \end{bmatrix} \left(\frac{1}{\frac{\mu_1\mu_2}{\lambda^2} + \frac{\alpha\mu_1}{\lambda} + \frac{(1-\alpha)\mu_2}{\lambda}} \right)^n \times$$

$$\left(\begin{bmatrix} \frac{\mu_2}{\lambda} + \alpha & 1-\alpha \\ \alpha & \frac{\mu_1}{\lambda} + 1 - \alpha \end{bmatrix} \right)^n \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{2.11}$$

For the two-stage M/G/1 queuing system we have the following:

$$\bar{x} = \frac{1}{\mu} = \Psi[\mathbf{V}] = \mathbf{pV}\varepsilon' = \begin{bmatrix} \frac{\alpha}{\mu_1} & \frac{1-\alpha}{\mu_2} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$= \frac{\alpha}{\mu_1} + \frac{1-\alpha}{\mu_2} \tag{3.1}$$

and,

$$C_s^2 = \frac{2\left(\frac{\alpha}{\mu_1^2} + \frac{1-\alpha}{\mu_1^2}\right)}{\left(\frac{\alpha}{\mu_1} + \frac{1-\alpha}{\mu_2}\right)^2} - 1 \tag{3.2}$$

We can have infinitely many two-stage systems with the same mean and coefficient of variation by changing one of its parameters. We can write [5],

$$\mu_1 = k\alpha\mu \tag{4.1}$$

$$\mu_2 = \frac{k(1-\alpha)\mu}{(k-1)} \tag{4.2}$$

Constraint (3.1) is still true.
We substitute the new representation of μ_1 from (4.1) and μ_2 from (4.2), in (3.2), and solve for α to get,

$$\alpha(k) = \frac{C_s^2 - 1}{2(C_s^2+1)} + \frac{2}{k(C_s^2+1)} + \frac{\left((C_s^2-1)^2 + 8(C_s^2-1)/k + (1-C_s^2)/k^2\right)^{\frac{1}{2}}}{2(C_s^2+1)} \quad (4.3)$$

In the M/H$_2$/1 queuing system as the 'tuning' parameter k approaches infinity H$_2 \rightarrow$ GE.

By substituting (4.1), (4.2) and (4.3) in (2.11) we get,

$$r(n) = (1-\rho)[\alpha(k) \quad 1-\alpha(k)] \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix}^n \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (5.1)$$

Taking the limit of (5.1) as $k \longrightarrow \infty$,

$$\lim_{k\to\infty}[r(n)] = (1-\rho)\lim_{k\to\infty}([\alpha(k) \quad 1-\alpha(k)]) \times \left(\lim_{k\to\infty}\begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix}\right)^n \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$= (1-\rho)\begin{bmatrix} \frac{C_s^2-1}{C_s^2+1} & \frac{2}{C_s^2+1} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & \frac{(C_s^2+1)\rho}{(C_s^2-1)\rho+2} \end{bmatrix}^n \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (5.2)$$

Substituting $y_s = \frac{C_s^2-1}{2} \Rightarrow C_s^2 - 1 \rightarrow 2y_s$ and $C_s^2 + 1 \rightarrow 2(y_s+1)$ in (5.2) we get,

$$r(n) = (1-\rho)\begin{bmatrix} \frac{y_s}{y_s+1} & \frac{1}{y_s+1} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & \frac{(y_s+1)\rho}{\rho y_s+1} \end{bmatrix}^n \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$= (1-\rho)\frac{1}{y_s+1}\left(\frac{(y_s+1)\rho}{\rho y_s+1}\right)^n \quad (5.3)$$

This solution is equivalent to the maximum entropy solution of the M/G/1 queuing system given in [5].

3.0 THE GE/M/1 QUEUE USING THE LINEAR ALGEBRAIC APPROACH

What we are going to do here is the same as we did in the previous section i.e., from the solution of H$_2$/M/1 we get the solution of GE/M/1 by taking the limit when the parameter k goes to infinity, and compare the result with the maximum entropy solution of the G/M/1 queuing system.

The G/M/1 queue is a FCFS queue with a general arrival process and exponential service time distribution.

Similarly, the matrix **U** representing the G/M/1 queuing system can be obtained by changing the ρ to its reciprocal in the matrix **U** representing the M/G/1 queuing system [2].

So,

$$\lim_{k\to\infty} \mathbf{U} = \begin{bmatrix} 0 & 0 \\ 0 & \dfrac{\frac{C_a^2+1}{\rho}}{\frac{C_a^2-1}{\rho}+2} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & \frac{C_a^2+1}{C_a^2-1+2\rho} \end{bmatrix} \quad (6.1)$$

The only eigenvalue of **U** is equal to

$$\frac{C_a^2+1}{C_a^2-1+2\rho} \quad (6.2)$$

According to [2] the queue length solution for the M/G/1 system is given in the following formula:

$$r(n) = \rho w^{-n}(w-1) \quad (6.3)$$

where, w is an eigenvalue of **U** whose magnitude is strictly larger than the magnitude of all other eigenvalues of **U**

Hence, substituting (6.2) in (6.3) we get,

$$r(n) = \rho\left(\frac{C_a^2+1}{C_a^2-1+2\rho} - 1\right)\left(\frac{C_a^2+1}{C_a^2-1+2\rho}\right)^{-n}$$

$$= \frac{2\rho(1-\rho)}{C_a^2-1+2\rho}\left(\frac{C_a^2-1+2\rho}{C_a^2+1}\right)^n \quad (6.4)$$

Letting $y_a = \frac{C_a^2 - 1}{2}$ then we have,

$$C_a^2 - 1 \rightarrow 2y_a \tag{7.1}$$
$$C_a^2 + 1 \rightarrow 2(y_a + 1) \tag{7.2}$$

Substituting (7.1) and (7.2) in (6.4) we get,

$$r(n) = (1 - \rho)\frac{\rho}{y_a + \rho}\left(\frac{y_a + \rho}{y_a + 1}\right)^n \tag{8}$$

which is also equivalent to the maximum entropy solution of the G/M/1 queuing system.

4.0 NUMERICAL RESULTS

The probabilities of finding n customers in the queue of the coaxial server "Cs" has been calculated for the two systems i.e., (M/GE/1 and GE/M/1) when the coefficient of variation is 1, 1.25, 1.50, 1.75, 2.00, 2.25, 2.50, 2.75, 3.00, 3.50, varying the utilization factor ρ from 0.1 to 0.9 (by an increase of 0.1) for each value of the coefficient of variation. In each case the probabilities were calculated for n "the number of customers in the queue of the coaxial server" started from 0 i.e., the probability of finding no customer in the queue until n = 30.

The generated data is studied by comparing the probabilities of the two systems and the effects caused on both by changing the parameters C_s^2 and ρ.

Firstly, it is known that the queue length probabilities are exactly equal and are equal to $(1-\rho)\rho^n$ when the coefficient of variation is equal to one. As an example, this is shown in fig. 2 when $\rho = 0.6$.

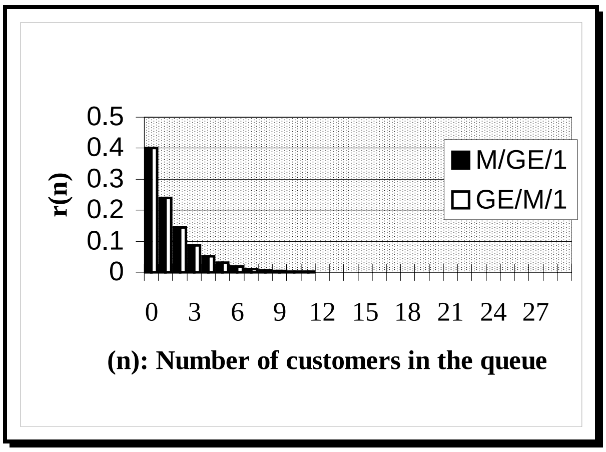

Fig. 2: The queue length probabilities for the M/GE/1 and GE/M/1 queuing systems with utilization factor = 0.6 and coefficient of variation = 1

For a fixed value of $C_s^2 > 1$ the difference between the probabilities of the two systems is clearly noticed especially for the first few n's i.e., (n = 1, 2, 3) an example shows this in fig. 3.

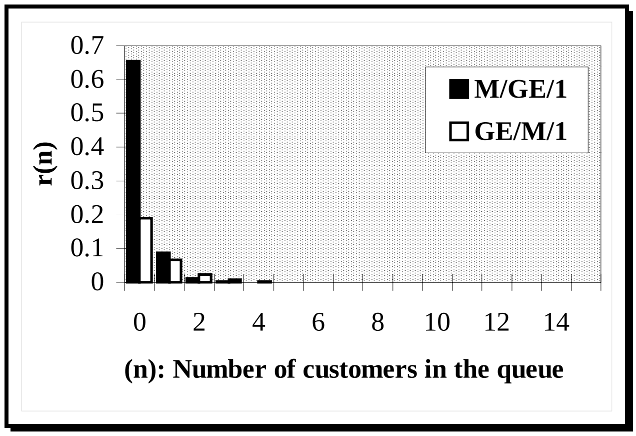

Fig. 3: The queue length probabilities for the M/GE/1 and GE/M/1 queuing systems with $\rho = 0.1$ and $C_s^2 = 1.75$

On the other hand, when the utilization factor ρ gets larger the difference between the probabilities of the two systems is getting smaller and smaller until it becomes valueless (around 0.001) for large values of ρ (like 0.7, 0.8, 0.9). This is shown in fig. 4.

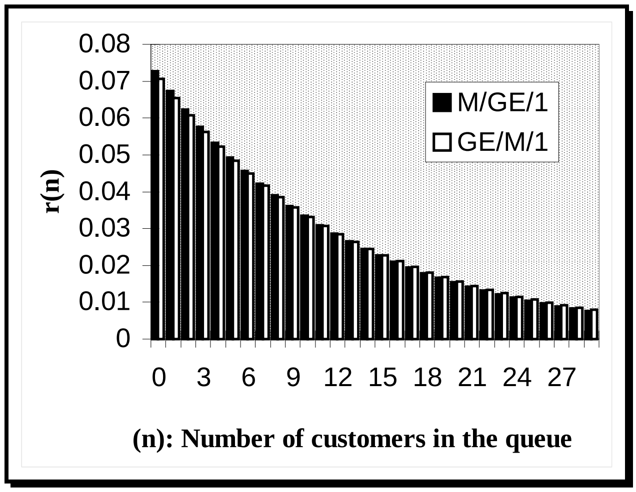

Fig. 4: The queue length probabilities for the GE/M/1 and M/GE/1 queuing systems with $\rho = 0.9$ and $C_s^2 = 1.75$

For a fixed ρ the difference between the two probabilities gets smaller but slowly, as the coefficient of variation increases.

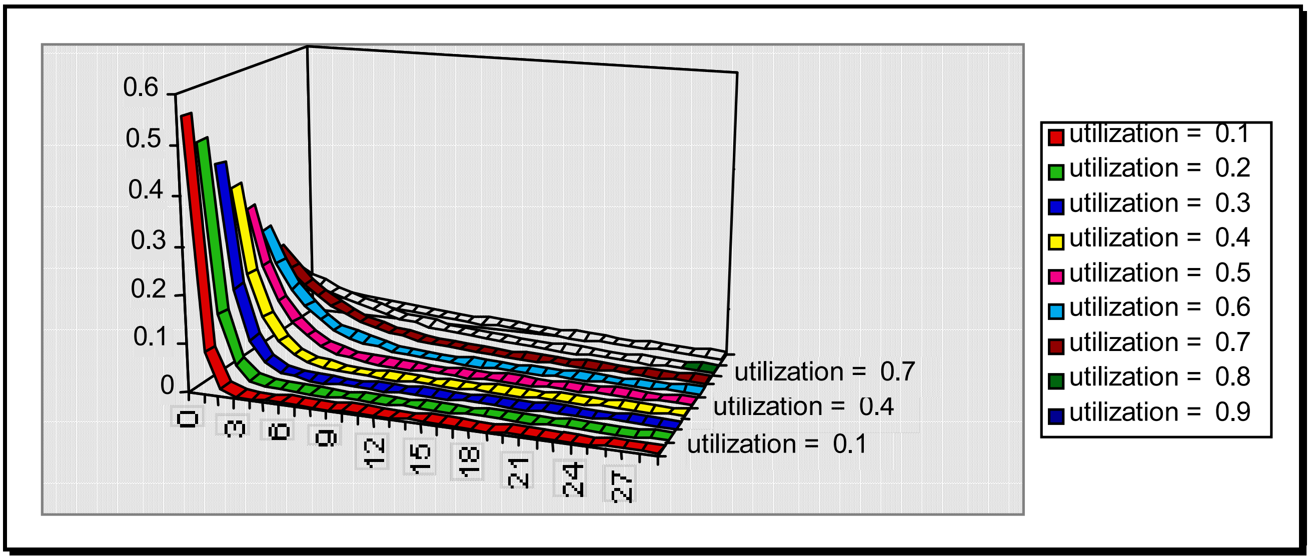

Fig. 5: An M/GE/1 queuing system with coefficient of variation = 2.25

Finally, for the M/GE/1 or the GE/M/1 as the utilization factor increases the sharpness in the beginning of the curve representing the probabilities, decreases to get closer to the other probabilities, fig. 5.

5.0 CONCLUSION

The queue length probability formulas for both the M/GE/1 queuing system and the GE/M/1 queuing system have been successfully derived through the algebraic approach of the queuing systems. The formulas are the same as those of the maximum entropy solutions of M/G/1 and G/M/1 queuing systems respectively.

The systems have been studied and compared with each other under various conditions and some of these cases have been represented in suitable figures.

As a future work, the queue length formulas for both the M/GE/C and GE/M/C queuing systems will be derived by following the same procedure, i.e., representing the algebraic solutions of $M/H_2/C$ and $H_2/M/C$ systems.

REFERENCES

[1] Lester Lipsky, "Explicit Solutions of M/G/C/ /N-type Queuing Loops with Generalizations", *Operations Research*, Vol. 33, No. 4, July-August, 1985, pp. 911-927.

[2] J. L. Carrol and others, "Solutions of M/G/1/ /N-type Loops with Extensions to M/G/1 and GI/M/1 Queues", *Operations Research*, Vol. 30, No. 3, May-June, 1982, pp. 490-514.

[3] Lester Lipsky, *Queuing Theory A Linear Algebraic Approach.* Singapore: Macmillan Publishing Company, 1992.

[4] Demetres D. Kouvatsos, "A Maximum Entropy Analysis of the G/G/1 Queue at Equilibrium", *J. Opl. Res. Soc.* Vol. 39, No. 2, 1988, pp. 183-200.

[5] M. A. El-Affendi and D. D. Kouvatsos, "A Maximum Entropy Analysis of the M/G/1 and G/M/1 Queuing Systems at Equilibrium", *Acta Informatica* Vol. 19, 1983, pp. 339-355.

BIOGRAPHY

Mohammad Saleh bin Mustafa obtained his Bachelor of major in Mathematics and minor in Computer Science at Yarmouk University, Jordan in 1991. He is currently a student in the Master Programme at the Department of Computer Science, University Pertanian Malaysia. His research interest lies in the Linear Algebraic Approach in the Queuing Theory.

Abu Talib bin Othman is an Associate Professor. He obtained his Doctor of Philosophy from University of Bradford, England in 1988. Currently, he is the Head of the Computer Centre, University Pertanian Malaysia. His research interest includes CASE, network performance and simulation, Queuing systems and business reengineering.