# A System Dynamics Based Methodology for Numerically Solving Transient Behaviour of Queuing Systems

Rahul Kr. Roy, Computer Centre
Indian Institute of Management Calcutta
Joka, D.H. Road, P.O. Box No. 16757, Calcutta 700 027, India

Dr. P. K. J. Mohapatra, Department of Industrial Engineering & Management
Indian Institute of Technology, Kharagpur
West Bengal 721302, India

Abstract : The paper proposes a methodology, of building *system dynamics* models for queuing systems. The methodology is applied to a variety of queuing systems and it is observed that, the models so developed are more transparent than conventional state-transition diagrams and incorporation of real life complexities are easier. In effect working out the transient and steady state behaviour of a wide variety of queuing systems becomes easy without going into much mathematical tedium.

## 1. Introduction

From a careful study of the literature on queuing theory one can make the following observations:

a. A majority of the queuing system models approximate the arrival and service process to Poisson process. As a result, it is normal to find that most of the exact solutions available in literature relate to queuing systems where the interarrival and the process times follow negative exponential distribution. These queuing systems are special cases of birth-death process.

b. Almost all the existing important results of queuing theory are obtained for the steady-state of queuing systems. Additionally, the parameters that describe the system (arrival and service rates) are time invariant or stationary. Few closed form expression exist for the transient behaviour of queuing systems and for systems where the arrival and service rates vary over time.

On the contrary, there are many situations in real life where the fluctuations of congestion with time that are of most interest to the analyst. As an example, we can consider the problem of determining personnel requirement in the railway booking counter. Normally, one would expect a variation in the number of people that would queue up for service, with the maximum demand occurring during 8 am and 11 am and during 4 pm and 8 pm on weekdays. It would be unwise to operate the same number of counters from early morning or late night hours as during peak passenger hours. Clearly, the situation calls for an analysis that recognizes explicitly the fact that demand for service in this case is time dependent and attempting a steady state analysis renders the recommendations made on the basis of steady-state-analysis, less useful.

Closed form solutions for transient behaviour are available for only a few queuing systems (Morse 1956). Several authors have tried approximate methods of analysis (Gaver 1966, Newell 1971, Neuts 1973, Moore 1975, Kotiah 1978). Numerical solution techniques have proved to be of considerable help in obtaining transient behaviour of queuing systems (Larson & Odoni 1983; Hengsbach and Odoni 1975; Kolesar, Rider, Crabill and Walker 1975). In all these cases, solutions have been found for the relevant differential equations. However, the technique of analysis followed by these authors requires construction of the differential equation which can be difficult in the case of non-markovian queues.

In analyzing the transient behaviour of queuing systems numerical solution techniques can often be of considerable help. Larson (1983) has described one methodology for analyzing M/M/m queuing system. Hengsbach (1975) and Kolesar (1975) has employed similar techniques to analyse transient behaviour of real world systems.

In this paper, we propose a generalized methodology, based on *system dynamics*, to model a wide variety of queuing systems. The methodology has been elucidated with the help of examples. It is observed that the models so built enables solution of transient behaviour of the respective queuing systems. Since every element of a system dynamics model has physical meaning, pictorial representation of queuing systems by system dynamics flow diagrams offer deeper understanding than a conventional state-transition diagram. This in turn enables incorporation of real life complexities without much of mathematical tedium.

## 2. System Dynamics Representation of Queuing Systems

In order to build the methodology, first we consider a queuing system that has features that are quite general in nature. An extensive class of well-known and often applied queuing systems can be simply viewed as special cases. We assume, that the system under discussions has the following characteristics:

(i) New users arrive at the system in a Poisson manner with a mean arrival rate of $\lambda_n$ expected arrivals per unit time, where n is the number of users, a user finds in the system (in queue plus in service) upon arrival.

(ii) There are m parallel, identical servers and for each of the servers service completion occurs in a Poisson manner with a mean service rate of $\mu_n$ per unit time, where n is the number of users in system.

(iii) The queuing system operates under a FCFS queue discipline.

(iv) The system has capacity to hold K ($K \geq m$) users. Any person who arrives when there are already K users in the system does not join the queue and leaves at once.

Given the above characteristics of the queuing system and assuming that there are N(t) users in the system at time t, the following four equations of conditional probability can be written:

$$P[N(t+\Delta t) = n+1 \mid N(t) = n] = \lambda_n \Delta t + o(\Delta t) \quad 1$$
$$P[N(t+\Delta t) = n-1 \mid N(t) = n] = \mu_n \Delta t + o(\Delta t) \quad 2$$
$$P[N(t+\Delta t) = n \mid N(t) = n] = 1 - \lambda_n \Delta t - \mu_n \Delta t - o(\Delta t) \quad 3$$
$$P[N(t+\Delta t) = k \mid N(t) = n] = o(\Delta t) \quad \text{for } |k-n| > 1 \quad 4$$

where $o(\Delta t)$ is a collection of terms that go to zero faster than k. $\Delta t$ as $\Delta t$ approaches zero. We can therefore ignore o($\Delta t$) if $\Delta t$ is sufficiently small.

Assuming the notation $P_n(t)$ to indicate P[N(t) = n], from equations 1 thou' 4 we can write

$$P_n(t+\Delta t) = P_{n+1}(t)\mu_{n+1}\Delta t + P_n(t)[1 - (\lambda_n + \mu_n)\Delta t] + P_{n-1}(t)\lambda_{n-1}\Delta t \quad 5$$

Since the number of people in the system at any point of time defines the state of the system at that point time then $P_n$(t) gives the probability that the system is in state n at time t. By our specification the system in question can accommodate any number of people from zero to K. Therefore there exists positive probability $P_n$(t) for every n = 0,1,2....K, and the following equation holds true.

$$\sum_{n=0}^{K} P_n(t) = 1, \quad \textit{for any } t \quad 6$$

$P_n$(t) shall henceforth be referred to as the *state probability*. Equation 5 after rearrangement can be written as the following

$$P_n(t+\Delta t) = P_n(t) + \Delta t\,[(P_{n+1}(t)\mu_{n+1} + P_{n-1}(t)\lambda_{n-1}) - P_n(t)(\lambda_n+\mu_n)] \quad 7$$

Clearly, equation 7 is the time dependent difference equation for $P_n$ (n = 0,1,2...K). $P_n$ can therefore be equated to a system dynamics level variable. The flow into $P_n$ during time interval t and t+$\Delta t$ is

| | | | |
|---|---|---|---|
| $P_1(t)\mu_1$ | for n | = | 0 |
| $P_{n+1}(t)\mu_{n+1} + P_{n-1}(t)\lambda_{n-1}$ | otherwise. | | |

The flow out of $P_n$ during time interval t and t+$\Delta t$ is

| | | | |
|---|---|---|---|
| $P_0(t)\lambda_0$ | for n | = | 0 |
| $P_n(t)(\lambda_n + \mu_n)$ | otherwise. | | |

The system dynamics model for the queuing system is now fully described.

## 3. Summary of the Methodology

The methodology to build system dynamics models for queuing systems, as proposed in section 2 can be summarized as follows:

1. The number of people present in the system defines the state which the system is in. For each state n (n = 0,1,2...K), that the system can attain, take one level variable and, name it as $P_n$. In case the number of states is unlimited, assume a very very large value for K.

2. For each level $P_n$ (n = 1,2...K-1) draw two rate variables flowing out of the level. One of these rates would go to the $P_{n+1}$ with the associated parameter as $\lambda_n$, the expected arrival rate for the queuing

system at state n. The other rate would go to $P_{n-1}$ with associated parameter for this flow as $\mu_n$, the expected service rate at state n. Both these rates will have information flows from the level $P_n$.

3. The level $P_0$, representing an empty system will have only one rate variable flowing out to the level $P_1$. The parameter associated with this is $\lambda_0$, the expected arrival rate at state zero and the rate will have information flow from the level $P_0$.

4. The level $P_K$, representing a saturated system will have a rate variable flowing out to the level $P_{K-1}$. The associated parameter for this rate variable is $\mu_K$, the expected service rate at state K, and the rate will have information flow from the level $P_K$.

5. The level variables can have any arbitrary mix of initial value provided equation 6 is satisfied for t=0, and in case a steady state for the system exists, all level variables will converge to the steady state value.

## 4. Examples

In this section we provide a few examples where the methodology described in section 3 has been used to build system dynamics models of well-known queuing systems and their variations. In each case the model has been simulated with a simulation interval (DT) equal to 0.01, expected arrival rate ($\lambda$) equal to 8, expected service rate ($\mu$) equal to 12. It was assumed that at the beginning of simulation, the system is empty. Consequently, the initial value for $P_0$ is set equal to one and for all other level $P_n$ (n > 0) the initial value is taken as zero. All along, the symbol $L_s$ has been used to denote the number of people in the system at time t and the symbol $L_q$ has been used to denote the number of people in the queue at time t. The system is assumed to have reached a steady state at a point when the value of $\left|1 - \frac{P_0(t+1)}{P_0(t)}\right|$ $(0 \leq t \leq 2000)$ is less than or equal to $10^{-8}$.

### 4.1 The M/M/1/K Queuing System

The M/M/1/K queuing system is a special case of the queuing system described in section 3. The value of system capacity K in this case is taken as 10.

$$\lambda_n = \lambda \qquad \text{for } n = 0,1,2....10$$
$$\mu_n = \mu \qquad \text{for } n = 0,1,2....10$$

The system dynamics flow diagram for the M/M/1/K queuing system is given in Figure 1.

The model was simulated for a total period of 20 time units(2000 simulation intervals). The dynamic behaviour of the system is shown in Figure 2 and 3.

Figure 2 shows the behaviour of $P_0(t)$, which, by our convention, denotes the probability of the system being idle at time t. According to the criteria set out in section 6 the steady state for the system is reached at time 10.0. The steady state value for $P_0$ is 0.33723. Figure 3 shows the dynamic behaviour of $L_s(t)$ *and* $L_q(t)$. Steady state values for $L_s(t)$ *and* $L_q(t)$ are 1,87 and 1.21 respectively. These also conforms to the theoretically obtained value.

### 4.2 The M/M/1 Queuing System

A M/M/1 queuing system is a special case of the M/M/1/K system where, by standard convention, the system has capacity to hold infinite number of users ($K = \infty$). In other words, the number of states the system can occupy is also infinite. As a result, by the methodology proposed in section 5, modelling such a system in system dynamics requires infinite number of states.

It should be noted here that level variables in system dynamics are quantities that have physical significance.

Therefore, modelling a system with $K = \infty$, needs to be done by means of a large number of level variables ($K \approx \infty$). However, the higher the K, the higher is the computational time. This calls for choosing a K that on one hand is sufficiently large and on the otherhand is good in terms of computation time.

Experiments ware conducted to see (i) the effect of K on computational time and (ii) the effect of $\lambda/\mu$ on the (simulation) time units to reach steady state.

In the first experiment, values of $\lambda$ was taken equal to 8, $\mu$ was taken equal to 12, K was varied from 10 to 250 in steps of 10. DT value was 0.01 and criteria for steady state was as spelt out in the beginning of section 6. The simulation was carried out on PC-XT with clock speed equal to 8 MHz.

For each value of K, from 10 to 250, the following procedure was repeated. In the beginning of simulation, $P_0(0)$ was set equal to 1, all other $P_i(0)$ (i = 1,2,...K) were set equal to zero. The clock time was saved. Simulation was carried out until $P_0$ reached steady state. At that point once again the clock time was noted down. The elapsed time between these two clock times gave the computation time for that value of K.

The result of the first experiment, in the form of variation of computation time with K, is shown in figure 4. The behaviour shows that initially, computation time increases sharply with increase in K (gradient $\approx$ 6.5 in 1). However, once K = 180, the increase was much slower (gradient $\approx$ 2 in 1).

In the second experiment, nine simulation runs of the model was carried out with values of $\lambda$ varying from 1.2 to 10.8 in steps of 1.2. In all the runs $\mu$ was taken equal to 12. In each run, K was varied from 10 to 200 in steps of 10. DT value was 0.01 and criteria for steady state was as spelt out in the beginning of section 6. The simulation was carried out on PC-XT with clock speed equal to 8 MHz. For each value of K, $P_0(0)$ was set initialized to 1 and all other $P_i(0)$ (i = 1,2,...K) were set equal to zero, and the simulation clock was initialized to zero. Simulation was carried out until $P_0$ reached steady state. The time by the simulation clock was saved.

Figure 5 shows the variation of simulation time with K and $\lambda$.

Based on the figures 4 & 5 the following observations can be made :

(a) Increase in $\lambda/\mu$ requires higher system capacity to reach steady state.

(b) For the same system capacity, increase in $\lambda/\mu$ requires higher simulation time units to reach steady state.

Based on these two experiments a it was thought that, K = 200 is a good magnitude to approximate infinite system capacity, because, it is observed that for all $0 < \lambda/\mu < 1$ the system reaches steady state within reasonable computation time.

Thus M/M/1 queuing system modelled with K = 200, reached a steady state $P_0$ value of 0.33333.

### 4.3 Variations of the M/M/1 queue

Given the model for the M/M/1 queuing system and the corresponding results of simulation that were shown to match with theoretical results, several variation on the initial model can be tried out.

#### 4.3.1 M/M/1 queue where users balk

In one such variation we can consider a system, where a prospective user of the queuing system decides, upon arrival at the system and observation of its state, not to wait for its use but to go elsewhere. Such situations are common in real life where users have options to choose among service facilities. Obtaining closed form solution for this kind of a system is difficult. However the methodology described in section 3 can be employed to numerically evolve the time dependent behaviour of such a system.

We consider the problem no. 4.2 given in Larson 1981 as and example.

The problem relates to one single server system with infinite system capacity, Poisson arrival process and negative exponential service times for which the rates of user arrival and service are

$$\lambda_n = \frac{\lambda}{n+1} \qquad n = 0,1,2...$$

$$\mu_n = \mu \qquad n = 1,2...$$

The system dynamics flow diagram for the problem is given in Figure 6. Interestingly, incorporation of this variation didn't call for significant modification of the model for the original M/M/1 system.

The model was simulated for a time period of 5 where the steady state was reached at time 1.6. The results obtained from the simulation is shown in Figure 7 and 8. We observe that the behaviour of $P_0$ (Fig. 7) is similar to that in M/M/1 queue where users do not balk. However since a portion of the customer are always lost, the steady state is reached faster with a higher steady state value for $P_0$ (0.51432), indicating a lower utilization of the service facilities. Figure 8 gives the time dependent behaviour of $L_s$ and $L_q$. Steady state value for $L_s$ and $L_q$ are 0.67 and 0.18 respectively.

### 4.3.2 M/M/1 queue with time dependent arrival rate.

In this variation of the normal M/M/1 queuing system, the rate at which customers arrive at the facility varies with time. Real life examples of such systems can be found at railway booking counters, cash counters of departmental stores where arrival of customers are not uniform during all hours of a normal day.

For our purpose we have considered a time dependent behaviour for the expected arrival rate as given in Figure 9. The system dynamics flow diagram for the queuing system is given in Figure 10.

The model was simulated for a time period of 25. The time dependent behaviour of $P_0$ is given in Figure 11. Since the arrival rate increases in steps, momentary steady states are observed in more than one place. The minimum value for $P_0$ (0.06395) is reached at time equal to 10, when $\lambda$ was equal to $\mu$. The behaviour of $L_s$ and $L_q$ is given in Figure 12.

### 4.4 $M/E_k/1$ Queuing System with Erlangian Service Time Distribution

In the case of $M/E_k/1$ queuing system, the arrival is Markovian, but the service time distribution is Erlangian of order k.

It is known that *when k sequential phases have independent identical exponential distribution, then the resulting density is known as k stage Erlang*. Thus the Erlang distribution with a mean of $1/\mu$ is equivalent to a series sum of k negative exponential distribution each with mean $1/(k\mu)$.

In sections 5.3.1 and 5.3.2, we have demonstrated the how a Markovian departure process i.e. an Erlangian process of order one, by means of a rate variable where the parameter is equal to $\mu$. Elsewhere in Roy and Mohapatra (90) it has been shown that, a k state transient Markov process is structurally equivalent to k order exponential delay (where k single order delay are connected in parallel). It may therefore be argued that an Erlangian departure process of order k can be represented by an exponential delay of order k.

The Erlangian departure process, of order k, between $P_i$ ($i = 1,2,...K$ ) to $P_{i-1}$ can thus be modelled, in system dynamics, by means of the level variable $P_i$, a rate which depends on $P_i$ ($i = 1,2,...K$ ) and has $k\mu$ as the associated parameter and an exponential delay of order k-1 and delay constant of $(k-1)/(k\mu)$. The resulting flow diagram for $M/E_k/1$ queuing system is given in figure 13.

It is easy to see why a delay of order k-1 is imposed on the rate variable. The Erlangian departure process of order k is equivalent to a series of k cascaded first order delay i.e. k cascaded level and rate pairs. The level variable $P_i$ and rate from $P_i$ ($i = 1,2,...K$ ) to $P_i$ ($i = 1,2,...K$ ) forms the first of these level rate pairs. The remaining k-1 level rate pair can be modelled as a delay of order k-1. The exponential delay of order k-1 will have a delay constant equal to $(k-1)/(k\mu)$.

### 4.4.1 Example

As an example of this kind of queuing system, we consider a system where the arrival follows a Poisson process with expected arrival rate $\lambda$ equal to 8. The service follows a Erlangian process of order 3. The expected service rate $\mu$ equal to 12.

Following the equivalence stated in section 5.4, the system is modelled as given in figure 14.

The model was simulated for a time period of 25. The resulting time dependent behaviour of $P_0$ is given in Figure 15. Observably, for some time initial period the system demonstrates an oscillatory behaviour around the steady state value. This can be attributed to the presence of third order delays (a level rate pair plus a second order delay) in the system. The time dependent behaviour of $L_s$ is given in Figure 16. The steady state, by criteria laid out in section 3, was reached after 6.25 time units. Steady state value for $P_0$ was 0.01439 and the same $L_s$ was 49.08

## 5. Conclusion

In conclusion we may make the following comments on the methodology proposed in this paper.

(i) The method is simple and has general applicability to build system dynamics models for queuing system with a variety of arrival and service processes. The same methodology has been employed to analyse queuing systems with $\lambda_n(t)$ or arrival rates that are time dependent.

(ii) In the system dynamics framework, incorporation of real life complexities can be done without going into much of mathematical tedium. Existence of non-linearity in the model does not impair analysis of the system. Analysis of complex queuing systems can be taken care without making simplifying assumptions.

(iii) Transient behaviour for the queuing systems can be worked out by simulating the system dynamics models over time. The steady state behaviour also can easily be arrived at.

(iv) System dynamics models are very transparent because the model variables and the parameters have real life meaning. Therefore, the models have great pedagogical value.

**7. Reference**


Coyle R. G.(1977), *Management System Dynamics*, John Wiley & Sons, 1977.

Forrester J. W.(1972), *Principles of Systems*, Wright Allen Press.

Gaver D. P. (1966), **'Observing Stochastic Processes and Approximate Transform Inversion'** in *Operations Research*, Vol 14, pp. 444-459

Hengsbach G. and Odoni A. R. (1975), **'Time Dependent Estimates of Delays and Delay Costs at Major Airports'**, Report R75-4, MIT Flight Transportation Laboratory, Cambridge, Massachusetts.

Kolesar P. J., Rider K. L., Crabill T. B. and Walker W. E.(1975), **'A Queuing-Linear Programming Approach to Scheduling Police Patrol Cars'** in *Operations Research*, Vol 23, pp. 1045-1062.

Kotiah T. C. T. (1978), **Approximate Transient Analysis of Some Queuing Systems'** in *Operations Research*, Vol 26, No. 2, pp.

Label J. D. (1978), **System Dynamics**, *Progress in Modelling And Simulation*, Addison Wesley Publishers,

Larson R. C. & Odoni A. R.(1981), *Urban Operations Research*, Prentice Hall, Englewood Cliffs, N.J. 07632

Moore S. C. (1975), **'Approximating the Behaviour of Nonstationary Single-Server Queues'** in *Operations Research*, Vol 23 , pp. 1011-1032

Mosekilde E. & Rasmussen S (1983), **'Random Processes in System Dynamics'** in *Dynamica*, Vol. 9 Part I, Summer, pp. 33-39

Neuts M. F. (1973), **'The Single Server Queue in Discrete Time Numerical Analysis I'**, *Naval Research Logistics Quarterly*, Vol 20, pp. 297-304

Newell G. F. (1971), *Applications of Queuing Theory*, Chapman and Hall Ltd. London

Roy R. & Mohapatra P. K. J.(1990), **'Markov Processes and System Dynamics - A Study of Equivalence'** in *Proceedings of the IVth National Conference on System Dynamics*, 1990 December, Tirupati. pp. 1-12

Trivedi K. S. (1982), *Probability and Statistics with Reliability, Queuing and Computer Science Application*, Prentice Hall International Inc. Englewood Cliff, N.J., U.S.A.

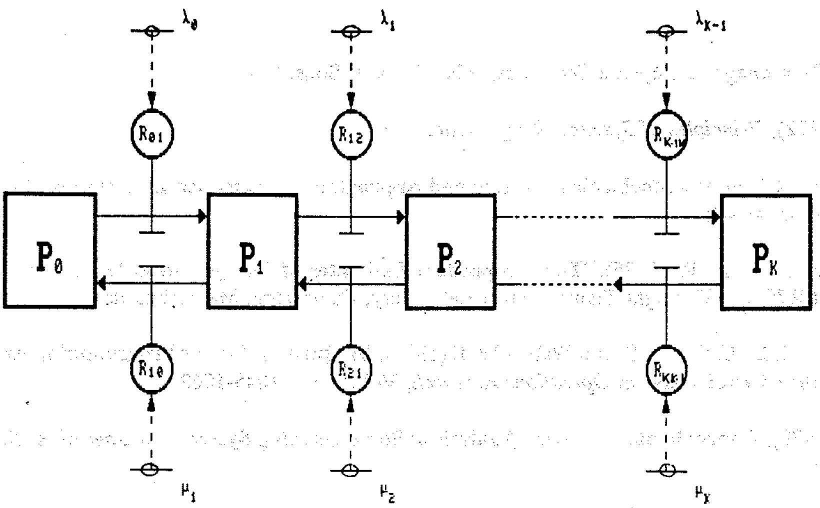


Figure 1 : Flow Diagram for the M/M/1/K Queuing System

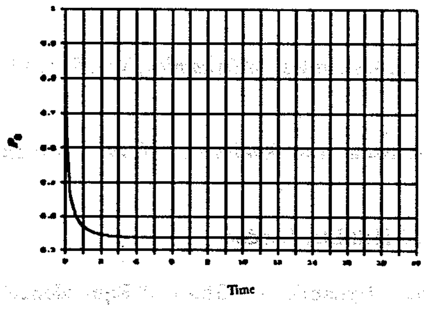


Figure 2 : Behaviour of $P_0$ for the M/M/1/K Queuing System.

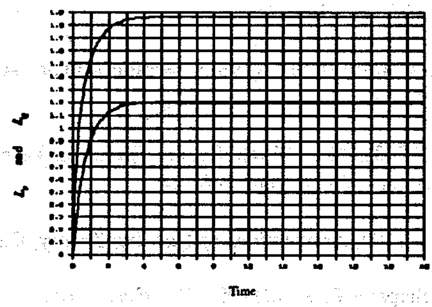


Figure 3 : Behaviour of $L_s$ and $L_q$ for the M/M/1/K Queuing System.

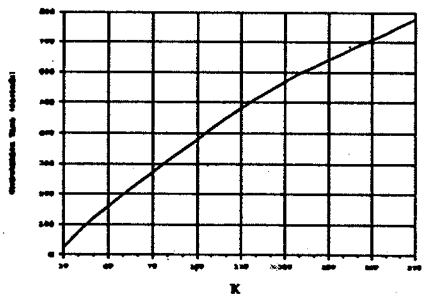


Figure 4 : Variation of computation time with K.

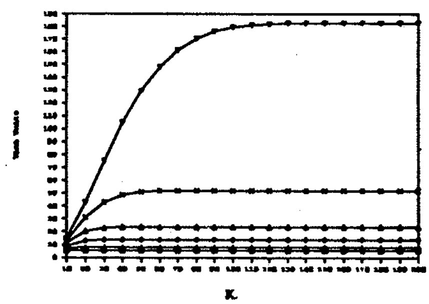


Figure 5 : Variation of simulation time with K and λ

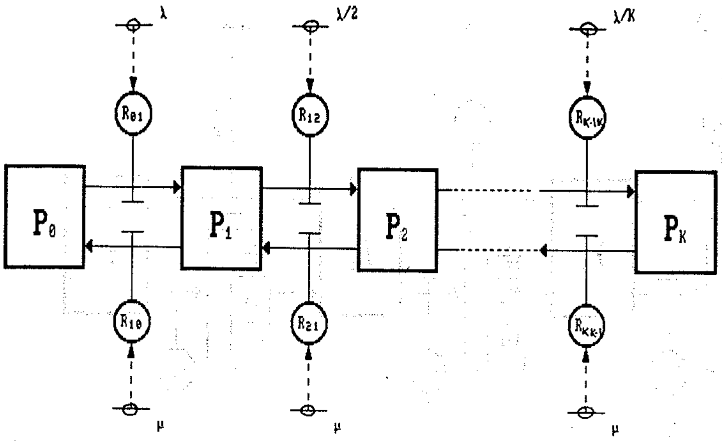


Figure 6 : Flow Diagram for the M/M/1 Queuing System that considers Balking.

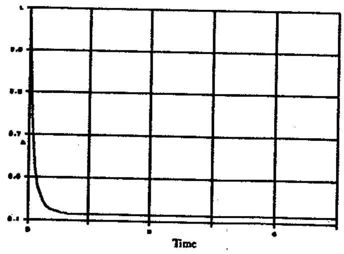


Figure 7 : Behaviour of $P_0$ for the M/M/1 Queuing System that considers Balking.

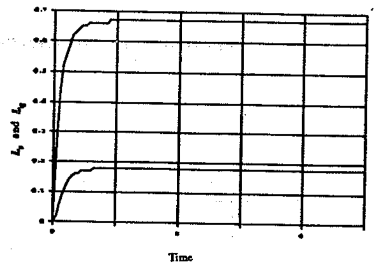


Figure 8 : Behaviour of $L_s$ and $L_q$ for the M/M/1 Queuing System that considers Balking.

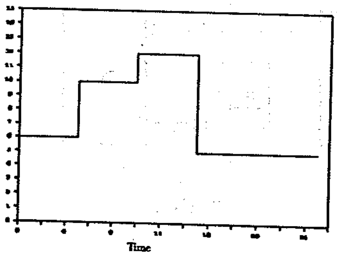


Figure 9 : Variation of λ over time.

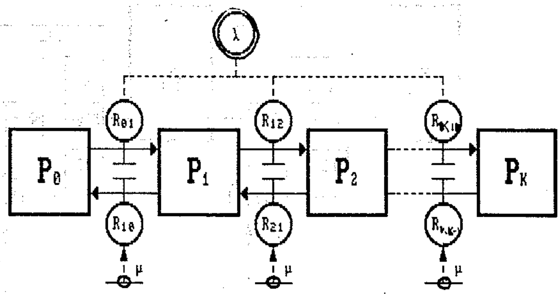


Figure 10 : Flow Diagram for the M/M/1 Queuing System where the arrival rate depends on time.

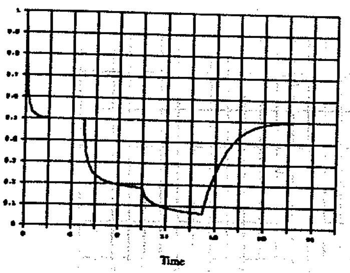


Figure 11 : Behaviour of $P_0'$ for the M/M/1 Queuing System where the arrival rate depends on time.

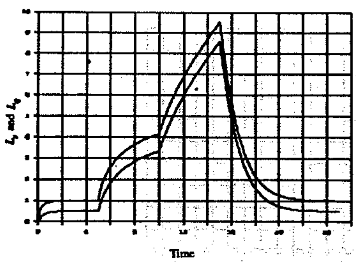


Figure 12 : Behaviour of $L_s$ and $L_q$ for the M/M/1 Queuing System where the arrival rate depends on time.

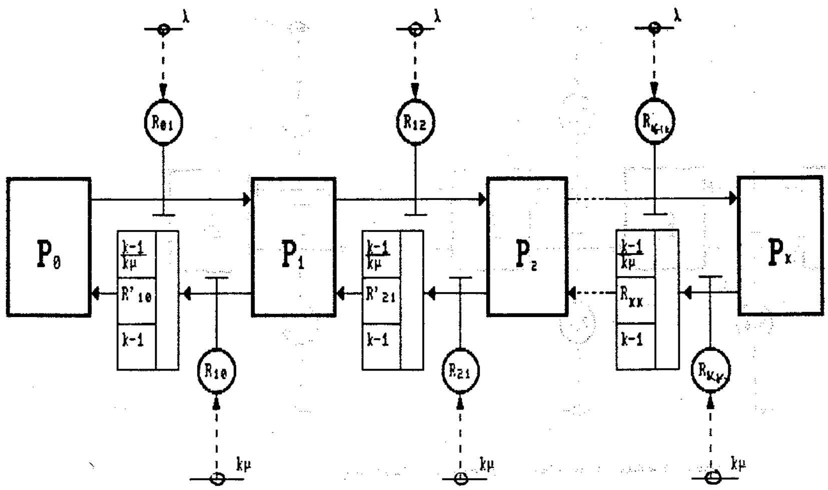


Figure 13 ; Flow Diagram for the $M/E_k/1$ Queuing System.

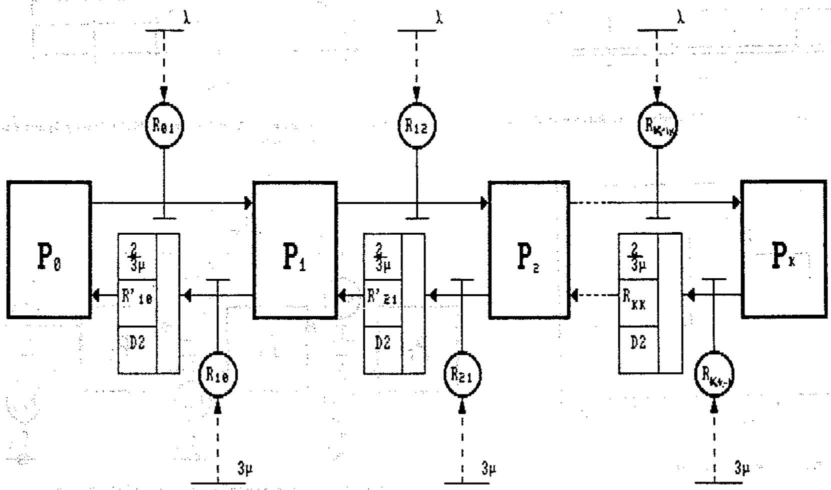


Figure 14 : Flow Diagram for the $M/E_3/1$ Queuing System.

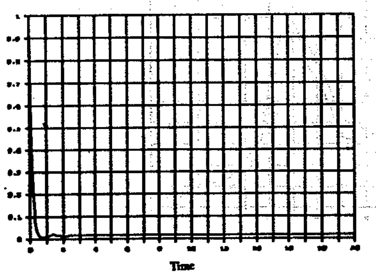


Figure 15 : Behaviour of $P_0$ for the $M/E_3/1$ Queuing System.

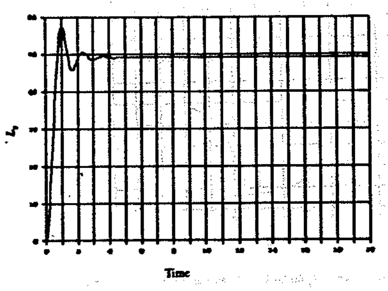


Figure 16 : Behaviour of $L_q$ for the $M/E_3/1$ Queuing System.