Kasu Journal of Computer Science Vol. 1 No. 2 [June, 2024], pp. 167-182
https://doi.org//10.47514/kjcs/2024.1.2.003

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

An Optimized Model for Service Rates' Control in M/M/1 Queue Networks with Server Vacations

Imran Ademola Adeleke[1], Adegbuyi David Gbadebo*[1], Hanat Yetunde Raji-Lawal[2]
[1]Department of Computer Science, Lagos State University of Education, Oto / Ijanikin, Lagos, Nigeria
[2]Department of Computer Science, Lagos State University, Ojo, Lagos, Nigeria
adelekeia@lasued.edu.ng[1], gbadeboad@lasued.edu.ng*[1], halaw313@yahoo.com[2]

Abstract
Evidence from literature had revealed that several studies on servers' failures and customers' behaviour in M/M/1 queue networks had been done. However, none of these studies had sufficiently established a relationship between optimality in servers' usage and control of queue network resources. Consequently, this study proposes a fuzzy-based optimal control system of the threshold type such that all arrivals finding K customers in queue and service are rejected. The objective is to choose the service rate dynamically in order to minimize average cost. OMNeT++ was used as a simulation framework while dataset was generated randomly. The system was simulated with the system starting from an initial state k= 0, s = 0 and c = 0, employing the fuzzy control at each decision epoch. The evolution of k and s for the first 1500 time units indicates that an idle server is not turned on immediately after a customer arrives. This shows the hysteretic property of the system indicating the hystereticity of the optimal policy. Consequently, it is concluded that the proposed approach is efficient and promising.

Keywords: Severs' vacations, Threshold policy, Fuzzy-based controller, Servers' cost

1. Introduction

Queues are part of everyday's life. This is because we wait in banks, hotels, supermarkets, airports, hospitals, etc. There are other types of queues which are invisible such as queues of voice calls or data packets in communication channels. Generally, queues are often undesirable as they come with cost of time and other resources. However, queues exist because service resources are not sufficient to satisfy demand. This often results from the fact that servers may be unavailable because of space or cost limitations, or the system is unable to provide the level of service necessary to prevent waiting.

A queuing system is a stochastic dynamic system which involves customers' arrival, waiting in queue until they are served by one or more servers and consequent departures. The evolution of a queuing system can be described by observing the states of its servers, including the type of service, busy or idle state as well as the number of customers in each queue as itevolve over time. A queuing system with variable rate is one where the mean service rate may be chosen from a set of finite service rates of: $\{\mu_k \mid k \in K\}$, where K is a fixed set of service types. In this system, the service time of a customer is a random variable which is independent of the arrival process and the previous service times.

The problem of determining when to turn servers 'on' or 'off' is multi-dimensional. In order to do this successfully, it is essential to determine the savings from turning servers off and comparing this with the cost of serving customers which are awaiting service. Badian-Pessot, Lewis and Down (2019) opined that

Kasu Journal of Computer Science Vol. 1 No. 2 [June, 2024], pp. 167-182
https://doi.org//10.47514/kjcs/2024.1.2.003

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

it is equally important to observe that while servers are on, it is possible to adopt a dynamic control of the service rate and consequently choosing to increase efficiency by using a fast service rate.

A stationary policy is an *N*-policyif it calls for turning the server on when *N* jobs are in the system and may turn off the server only when the system is empty. *N*-policies are optimal for *M/M/*1 queues with startup costs and a single service rate (Heyman, 2018). For *M/M/*1 queues with dynamic service rate control, Lippman (2015) showed the optimal service rate structure is monotone as far as the number of customers awaiting service turns is concerned.

Majority of studies on the problem of optimal control by the number of active servers suggest that the strategy of control is determined by the class of threshold strategies. The threshold strategy is generally defined by the set of integer numbers. In this case, a new server is activated when the number of customers in the system or in the queue exceeds one of thresholds while some of the active servers are switched off at the service completion moment in which the number of customers in the system or in the queue is below the corresponding threshold level.

The threshold strategies are relevant as they react by activating additional server when the system becomes congested and by switching off one of the active servers when the number of customers reduces. This enables the system to attain some trade-off between the requirement to provide good quality of service for customers as well as the obvious desire of the system manager to reduce expenses associated with the maintenance of redundant servers.

The common problem of threshold strategy is the effect of necessity for frequent switching servers on and off. Constant switching is often not desirable if it requires some time such as setup and removal times, hiring as well as releasing times, during which the server cannot provide service, or if the system manager has to pay for every switching. In order to avert this problem, hysteresis strategies are often employed as such methods are quite popular in optimal dynamic control by the arrival and (or) service rates in different queuing models.

Although many researches had been done on ensuring optimality on the use of servers in queue networks, not enough studies had been in minimizing servers' overall cost covering such areas as holding and switching costs. This is necessary to minimize costs of servers' usage in the organization. In order words, to the best of our knowledge, not much work had been done in modeling a fuzzy-based system for the control of service rates with the aim of increasing systems' efficiency and reducing wastages of servers' time. This in turn will provide optimal usage of available servers while also minimizing customers' wait time. In this study, a fuzzy-based control system was modeled to dynamically manage service rates in an *M/M/1* queue network. In other words, we modeled the proposed system from a managerial perspective as it is useful in developing more sophisticated controls for implementation in *M/M/1* queue systems.

2. Related Works

Research works on queuing systems involving removable servers in networks include Yadin and Naor (2013) which analyzed *M/G/*1 queues under *N*-policies and application. The optimality of *N*-policies for *M/G/*1 queues including the costs of system warm-up using average and total discounted reward criteria,

Kasu Journal of Computer Science Vol. 1 No. 2 [June, 2024], pp. 167-182
https://doi.org//10.47514/kjcs/2024.1.2.003
Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

respectively was investigated by Bell (2011). Related to this, a study of *M/M/*1 queues under *N*-policies with both start-up costs and exponentially distributed warm-up times was done by Baker (2003) with no consideration for service rate control in the system. Borthakur, Medhi and Gohain (2017) extended this study by including parameters for general warm-up times, providing steady-state probabilities, average wait time and number of customers in queue.

Feinberg and Kella (2022) performed an analysis of *M/G/*1 queue where the service time is known upon arrival. Similar variations of these models include queues with server breakdowns and warm-up times as investigated by Wang, Wang and Pearn (2017), queues with server vacations (Ke, 2023) and queues with batch arrivals (Federgruen and So, 2021), among many others. The first to model a system with varying service rates using MDPs wasCrabill, (1974). Recently, Lippman (2015) carried out an examination of the optimal policy for *M/M/*1 queues with dynamic service rate control proving the monotonicity of the optimal service rate.

A single server system with dynamic service rate control andMarkov modulated arrivals was modeled by Kumar, Lewis, and Topaloglu (2023) in which the service rates are chosen from a closed subset of [0,∞]. An investigation of the value of dynamic service rate control in combination with admission control in *M/M/1* queues was done by Dimitrakopoulos and Burnetas (2017). The use of sleep-state control under various policies in *M/G/*1 queues was studied by Gebrehiwot, Aaltoand Lassila (2020). Related to this, Gandhi, Gupta, Harchol-Balter and Kozuch (2020) examined an *M/G/*1 queue with warm-up times and many sleep states for a specific cost function: the product of the mean power consumption and mean response time in steady state. Other related studies considered multi-server systems such as in Gandhi, Harchol-Balterand Adan (2020) as well as Maccio and Down (2022).

In a study of a single server vacation queue by Vijayalakshmi and Kalidass(2018) as well as Yang and Chen (2018), a relationship between geometric abandonments and Bernoulli's feedbacks was established. Similarly, Deepa and Kalidass(2018) investigated the steady-state behaviour of Markovian queue with finite capacity, working breakdowns and server vacations. An *M/M/1* queuing model with second optional service, in which the server is subject to working breakdowns was studied by Dong-Yuh and Yi-Hsuan (2018). Related to this, Vijayalakshmi, Kalidass and Pavithra (2018) considered an *M/M/1/N* queuing system with a working breakdown and two-phase services.

In real life queuing systems, it is often considered that a server stops service entirely during the breakdown period. However, it is often possible that the server provides service at a slower rate rather than entirely stopping the service during the breakdown period. This scenario of server breakdowns is called working breakdowns. This idea was first introduced by Kalidass and Kasturi (1995), consequently studing the steady-state analysis of the *M/M/1* queue with working breakdowns.

Liou (2023) adopted the matrix-geometric method to examine a finite capacity Markovian queue with an unreliable server subject to working breakdowns and impatient customers. An analysis of thesteady-state analysis of *M/M/1/N* queue with working breakdowns and Bernoulli feedbacks was studied byKalidass and Pavithra (2016).

Fuzzy set theory is a paradigm shift which helps to resolve classical and non-classical problems in a more convenient way than crisp systems by softening set boundaries. Modern management of queues has taken

Kasu Journal of Computer Science Vol. 1 No. 2 [June, 2024], pp. 167-182
https://doi.org//10.47514/kjcs/2024.1.2.003

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

advantage of fuzzy set theory by using it as a way to manage imprecise and uncertain situations to model and process flexible queue controls. Fuzzy control of queuing systems is an application of fuzzy logic theory to the control of queues. A fuzzy logic controller provides a decision mechanism that dynamically determines the parameters, patterns and policies of a queuing system in some specified optimal sense. In essence, a fuzzy queuing control is a combination of artificial intelligence, operations research and optimal control policies.

3. Methodology

This section is discussed under the following sub-sections: analysis of the steady state for *M/M/*1 queues, problem definition, fuzzy Markov chain, design of the fuzzy-based controller and simulation of the Markov queuing system.

3.1. Analysis of the steady state of *M/M/*1 queues

If *N(t)* is the number of customers in the system at time *t*, Vijayalakshmi, Kalidass and Deepa (2021) opined that the dual-variate process $\{(C(t),\ N(t)),\ t \geq 0\}$ can be defined as a continuous-time Markov chain. Consequently,

$$\text{Let: } P_i,\ n(t) = \Pr ob\{C(t) = i,\ \mathrm{N(t)} = \mathrm{n}\},\ \ i = 1, 2, 3, 4 \text{ while } n \geq 0$$

The steady-state probability equations, according to Vijayalakshmi and Kalidass (2018) are as follows:

$$\mu p P_{1,1} + \mu_1 P_{2,1} = \lambda P_{1,0};$$

$$\lambda P_{1,n-1} + \gamma P_{3,n} + \mu p P_{1,n+1} + \mu_1 P_{2,n+1} = (\lambda + \beta + \mu) P_{1,n}, 1 \leq n \leq N-1;\ \lambda P_{1,n-1} + \gamma P_{3,n} = (\beta + \mu) P_{1,n},\ n = N;$$

$$\mu q P_{1,1} + \xi P_{4,1} = (\mu_1 + \alpha + \lambda) P_{2,1}\ \lambda P_{2,n-1} + \mu q P_{1,n} + \xi P_{4,n} = (\mu_1 + \alpha + \lambda) P_{2,n} + 2 \leq n \leq N;$$

$$\lambda P_{2,n-1} + \mu q P_{1,n} + \xi P_{4,n} = (\mu_1 + \alpha) P_{2,n}, n = N;$$

$$\mu_b p P_{3,1} + \mu_{b1} P_{4,1} = \lambda P_{3,0};$$

$$\mu_1 p P_{3,n+1} + \lambda P_{3,n-1} + \beta P_{1,n} + \mu_{b1} P_{4,n+1} = (\gamma + \lambda + \mu_1) P_{3,n},\ 1 \leq n \leq N-1;$$

$$\lambda P_{3,n-1} + \beta P_{1,n} = (\gamma + \mu_1) P_{3,n},\ n = N;$$

$$\alpha P_{2,1} + \mu_1 q P_{3,1} = (\xi + \lambda + \mu_{b1}) P_{4,1};$$

$$\alpha P_{2,N} + \mu_1 q P_{3,N} + \lambda P_{4,N-1} = (\xi + \lambda + \mu_{b1}) P_{4,n},\ 2 \leq n \leq N;$$

$$\alpha P_{2,n} + \mu_1 q P_{3,n} + \lambda P_{4,n-1} = (\xi + \mu_{b1}) P_{4,n}, n = N.$$

From the normalizing condition,

$$P_{1,0} + P_{1,e} + P_{2,0} + P_{2,e} + P_{3,0} + P_{3,e} + P_{4,0} + P_{4,e} = 1$$

Kasu Journal of Computer Science Vol. 1 No. 2 [June, 2024], pp. 167-182
https://doi.org//10.47514/kjcs/2024.1.2.003

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

where ***e*** is the unit column vector of dimension *N*, resulting in:

$$P_{1,0} = 1/\psi$$

where

$$\Psi = 1 - A_{12}A_{22}^{-1}e + \Theta_7 A_{62}A_{22}^{-1}e + \Theta_8 e + \Theta_7 e + \Theta_5\Theta_6^{-1}e$$

In this case, a continuous-time Markov chain X_t, $t \in (-\infty, \infty)$ according to Zhang, Phillis and Kouikoglou (2005) is a stochastic process that moves in a countable set of states $\{S_1, S_2, ...\}$ and satisfies the Markovproperty:

$$P(X_{tn}, | X_{tn-1}, X_{tn-2}, ..., X_{t()}) = P(X_{tn}, | X_{tn-1})$$

for all choices of $-\alpha < t_0 < t_1 < ... < t_n < \alpha$

3.2. Problem definition

A queuing system in which the server may be turned off is a system with vacations. This is illustrated in figure 1.

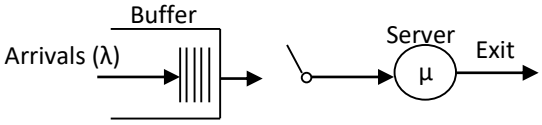


Figure 1: A single server system with vacations

Arrivals into the system identified in figure 1 are according to a Poisson process with parameter λ while the buffer has infinite capacity. The system has one exponential server with service rate μ, where $\mu > \lambda$. This is a queuing system in which it is possible to adjust the service rate to zero during certain intervals of time depending on varying system characteristics. Consequently, the queuing system has three distinctive costs:

a. The power-on cost *R*: This is the cost incurred whenever the server is turned on and it is expressed as: R_k, k = 0, 1, per unit time. The corresponding values of $k = 1$ and $k = 0$ when it is turned on and off respectively;
b. The service cost *r*: This is the cost per time unit when the server is on. This is r_k, *where* k = 0, 1, per unit time when the server is turned on and off with corresponding values of $k = 1$ and $k = 0$ respectively while $r_0 \leq r_1$;
a. The holding cost *h*: This is the holding cost per time unit per customer in the system, including the one in service (if any).

The goal is to find an optimal control policy, which tells when the server is turned off resulting in the minimization of the average cost rate of the system over a period of time. This is a semi-Markov decision process as a result of the presence of the controller. Consequently, the times between successive service completions are no longer exponentially distributed. This stochastic process is synonymous with a Markov process at the instants of state transitions.

Kasu Journal of Computer Science Vol. 1 No. 2 [June, 2024], pp. 167-182
https://doi.org//10.47514/kjcs/2024.1.2.003

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

According to Heyman (2018), it is optimal to keep the server always on when:

$$\left[\frac{2r(1-\rho)}{h}+1\right]^2 - \frac{8\lambda R(1-\rho)}{h} < 0 \quad \text{.....(1)}$$

where $\rho = \lambda / \mu$ is the traffic intensity. It is necessary to switch on the server when there are *N*customers waiting to be served and to power it off when there is no customer to be served, i.e. while the server is idle. In this case, the least cost is given as:

$$n^* = \sqrt{\frac{2\lambda R(1-\rho)}{h}} \quad \text{.....(2)}$$

In (2) above, *N = 0* when (1) is true. In this case, it is cost effective to keep the server always on when the power offcost *R*is sufficiently larger than the service cost rate *r*.

3.3. Birth and death processes in the proposed method

In the proposed model, service and inter-arrival times of customers are independent, exponential random variables with mean values *1/μ* and *1/λ* respectively. Consequently, its optimal control is of the threshold type. This implies that all arrivals finding *K* customers ahead both in queue and in service are rejected for some specified threshold value *K*. The state transitions are shown in figure 2.

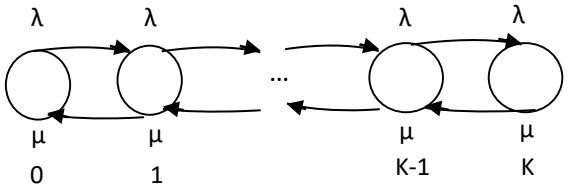


Figure 2: State transitions of the queue model

The rates of transition of each state in figure 2 can be expressed as:

$$\Lambda_{n,n+1} = \begin{cases} \lambda & \text{if } n < K, \\ 0 & \text{if } n = K. \end{cases}$$

$$\Lambda_{n,n-1} = \begin{cases} \mu & \text{if } n > 0, \\ 0 & \text{if } n = 0. \end{cases}$$

$$\Lambda_n = \begin{cases} \lambda & \text{if } n = 0, \\ \lambda + \mu & \text{if } 0 < n < K, \\ \mu & \text{if } n = K. \end{cases}$$

Kasu Journal of Computer Science Vol. 1 No. 2 [June, 2024], pp. 167-182
https://doi.org//10.47514/kjcs/2024.1.2.003
Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

3.4. Fuzzy Markov chain

It is possible to define a fuzzy Markov chain for an *M/M/1* queue system. Let $X = (\tilde{S})$ represent the fuzzy stochastic process which is fuzzy Markovian by an evaluation of the system after the completion of service of a customer and the commencement of service of the next customer in queue. If $\tilde{P}_i$ is the probability of i arrival during service time $\tilde{S}$, then $\tilde{P}_i$ is a fuzzy function such that there exists a possibility distribution induced by $\mu_{\tilde{S}}$ on each of its points. Since arrival is a Poisson process with λ, defined the fuzzy probability function can be expressed as:

$$\mu_{\tilde{P}_i}(x) = \sup_{t \in R^+} \left\{ \mu_{\tilde{S}}(t) \mid x = \frac{\exp(-\lambda t)(\lambda t)^i}{i!} \right\}$$

A transition from state i to j requires the arrival of j customers during the service interval. Moving from state i to j where $j \geq i$, $-1 > 0$ requires that $(j-i+1)$ customers arrive during a service interval. Consequently, moving from i to j where $j < i-1$ is impossible as long as service occurs only at a time.

Let us denote the transition probability matrix of the fuzzy Markov chain by $\tilde{P} = [P_{ij}]$. Then it is possible to express all $j \geq i-1$, $i > 1$ as follows:

$$\tilde{P} = [P_{ij}] = \begin{bmatrix} \tilde{P}_0, & \tilde{P}_1, & \tilde{P}_2, & \tilde{P}_3, & \ldots \\ \tilde{P}_0, & \tilde{P}_1, & \tilde{P}_2, & \tilde{P}_3, & \ldots \\ 0\ , & \tilde{P}_0, & \tilde{P}_1, & \tilde{P}_2, & \ldots \\ 0\ , & 0, & \tilde{P}_0, & \tilde{P}_1, & \ldots \\ . & . & . & . & \ldots \\ . & . & . & . & \ldots \\ . & . & . & . & \ldots \end{bmatrix}$$

where P_{ij} is defined as:

$$\forall j \geq i-1, \quad i \geq 1;$$

$$\mu_{\tilde{P}_{ij}}(x_{ij}) = \sup_{t \in R^+} \left\{ \mu_{\tilde{s}}(t) \mid x_{ij} = \frac{\exp(-\lambda t)(\lambda t)^{j-i+1}}{(j-i+1)!} \right\}$$

3.5. Simulation of Random Variables Applicable in Markov Queuing Models

The stochastic nature of queuing systems exists such that customers arrive at a queue network and complete service at random times. Consequently, in order to simulate a queuing system, it is necessary to generate sequences of random variables, which is a representation of the possible realization of the customer arrival

Kasu Journal of Computer Science Vol. 1 No. 2 [June, 2024], pp. 167-182
https://doi.org//10.47514/kjcs/2024.1.2.003

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

and processing times. These sequences are generated by means of "random-variable generators". Generally, random values are sample variables of the universal nature, *U* which is distributed uniformly on the interval [0, 1]. The linear congruential generators are the most common random number generators applicable in queuing problem applications. These generators usually begin with an initial integer value zero (Z0) and consequently yield successive random numbers of the nature: u_0, u_1, ... u_∞ by using the following:

$$Z_{n+1} = (aZ_n + b) \; mod \; c,$$

$$\mu_{n+1} = Z_{n+1} / c$$

In this case, "$(aZ_n + b)$ *mod* c " is the remainder of the division of $(aZ_n + b)$ by c. The variables *a*, *b* and *Z* are non-negative integers which are all lesser than c. Since the value of Z_n determines the next random value while *c* is integer, the above scheme generates at most c which differs from random numbers ranging from 0 to (c - 1) before it reverts to value Z from which it began. A good random number generator need to have a long period and a uniform coverage of the interval [0, 1].

Let *X* be a real random variable with known distribution function:

$$F_x(x) = P(X \leq x)$$

This defines a new random variable with the following distributions:

$$F_U(\mu) = P(U \leq \mu) = P[F_x(X) \leq \mu] = P[X \leq Fx^{-1}(\mu)] = \mu$$

If it is assumed that the inverse function, Fx^{-1} exists, then F_x and *U* range over [0, 1]. Consequently, the existence of Fx^{-1} is guaranteed if F_x is strictly increasing. If the distribution function is not increasing but it contains flat segments and discontinuities, as in the case of discrete random variables, then the inverse distribution function is defined as:

$$Fx^{-1}(\mu) = \infty\{t : F(t) \geq \mu\}$$

This is valid for every $\mu \in [0, 1]$. In this case, *x* is the smallest *t* such that $F_x(t) \geq \mu$.

3.6. Design of the fuzzy-based controller

The design of the fuzzy-based controller is premised on the assumption that the server is kept active as long as there are customers awaiting service turns and powered off once there are no customers awaiting service turns. The state of the system can be represented as: *(x, y)*, where $x = xt$ is the state of the server at time *t*, with value 0 while the server is powered off and 1 while the server is in the ready state. Similarly the number of customers in the system awaiting service turns including the one in service (if any) at time *t is*

$y = y(t)$. The value of *y* changes whenever a new customer arrives or a customer exits the system as a result of service completion. The value of *x* changes anytime the server is powered on or off. Consequently, the average cost rate of the system over a given period of time can be expressed as:

Kasu Journal of Computer Science Vol. 1 No. 2 [June, 2024], pp. 167-182
https://doi.org//10.47514/kjcs/2024.1.2.003

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

$$g = \lim_{T\to\alpha}\left[\sum_{t=1}^{T}\frac{hs(t)}{T} + \sum_{t=1}^{T}\frac{rk(t)}{T} + \sum_{t=1}^{T}\frac{Rx1_{\{k(t-1)=0,\ k(t)=1\}}}{T}\right]$$

where $1\{k(t-1)=0,\ k(t)=1\}$ is a function with value *1* if condition *x* is true and *0*, if untrue. This implies that $Rx1_{\{k(t-1)=0,\ k(t)=1\}}$ is the cost of switching on the server at time *t*. Similarly, the cost, *c* is expressed as:

$$c = h\sum_{i=1}^{n} s_i$$

where *i* is the i^{th} consecutive time unit within the current server state, *n* is the sum of time the server was idle in the current state *k* starting from the last time it was switched to that state while s_i is the number of customers present in the system at the i^{th} time.

There are three cases when it might be optimal to power on the server. These include:

a. Provided there is no powering on costs, it is optimal to turn on the server when there is an arrival. This is because the existence of a powering-on cost might result in a delay to turn on the server even while there are customers present in the system awaiting service turns. This is to ensure a reduction or avoidance in this cost. In this case, the optimal switching-on time will depend on the relationships between holding and powering-on costs. This relationship results in the fact that when the accumulation of holding cost *c* during the server's idle time is reasonable enough to compete with the switching cost, it is optimal to turn on the server.
b. The traffic intensity ρ is defined using the following function:

$$\rho = \lambda / \mu$$

In this case, ρ is the average use of the server over an infinite time 1- ρ. This quantity is the fraction of the time the server can be idle. If $\rho \geq 1$, then the number of customers in the system increases infinitely. In this case, if the server is switched off during a fraction of time*T*, this will save a constant cost rate *r* while the holding cost rate will grow at an average rate of $h\lambda$. It is optimal to keep the server active during this time. However, $\rho=0$ *and* $\lambda=0$, it is optimal to serve all customers in the system after which the server is switched off permanently. In order to avert this from occurring, it is assumed that $\rho > 0$. If ρ is close to zero, then it is necessary to keep the server idle until the accumulation of holding cost ***c***during the server's idle time. In this case, it is optimal to turn on the server. This can be succinctly stated that the higher the ρ, the easier it is to make the decision to switch on the server.
c. If $h = 0$, keeping the server idle always off achieves the minimum cost $g = 0$. This is so despite the fact the number of customers in the system increases as time *T* increases. In this case, it is assumed that $h > 0$. Hence, it is obvious that the higher the *h*, the easier it is to make the decision to switch on the server.

The inputs of the rule base include: *c*, *h* and ρ with each input represented by four linguistic values: ZE, SP, PO and EP representing 'Zero', 'Slightly Positive', 'Positive' and 'Extremely Positive' respectively. The integer weight for each linguistic input value was assigned as 0, 1, 2 and 3 for ZE, SP, PO and EP respectively. The output of each rule combination, denoted by *d*.. It is the decision regarding when to switch the switch the server 'on' or 'off'. This is represented by the linguistic values ON and OFF. The relationship among *c*, *h*, ρ and *d* are indicated in table 1.

Kasu Journal of Computer Science Vol. 1 No. 2 [June, 2024], pp. 167-182
https://doi.org//10.47514/kjcs/2024.1.2.003

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

Table 1: Relationship among *c*, *h*, *ρ* and *d*

Rule sequence	*c*	*h*	*ρ*	*d.*
1	ZE	ZE	ZE	OFF
2	SP	ZE	ZE	OFF
3	PO	ZE	ZE	OFF
4	EP	ZE	ZE	ON
5	ZE	SP	ZE	OFF
6	SP	SP	ZE	OFF
7	PO	SP	ZE	ON
8	EP	SP	ZE	ON
9	ZE	PO	ZE	OFF
10	SP	PO	ZE	ON
11	PO	PO	ZE	ON
12	EP	PO	ZE	ON
13	ZE	EP	ZE	ON
14	SP	EP	ZE	ON
15	PO	EP	ZE	ON
16	EP	EP	ZE	ON
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
49	ZE	ZE	EP	ON
50	SP	ZE	EP	ON
51	PO	ZE	EP	ON
52	EP	ZE	EP	ON
53	ZE	SP	EP	ON
54	SP	SP	EP	ON
55	PO	SP	EP	ON
56	EP	SP	EP	ON
57	ZE	PO	EP	ON
58	SP	PO	EP	ON
59	PO	PO	EP	ON
60	EP	PO	EP	ON
61	ZE	EP	EP	ON
62	SP	EP	EP	ON
63	PO	EP	EP	ON
64	EP	EP	EP	ON

The decision (*d.)* is OFF if the sum is less than or equal to 2; otherwise it is ON. The mathematical weight of rule 1, for instance is 0 + 0 + 0 = 0. The sum 0 < 2 and therefore the *d.* is OFF. For rule 13, the mathematical weight is ZE (0) + EP (3) + ZE (0) = 3. The sum i.e. 3 >2, therefore the *d.* is ON. Similarly, for rule 59, the mathematical weight is PO (2) + PO (2) + EP (3) = 7. The sum i.e. 7 > 2, therefore the *d.* is ON.

The membership function of the input variables is either triangular or trapezoidal. The functional forms of the membership functions of inputs these values depend on the corresponding input. The interval $[0, +\infty)$

Kasu Journal of Computer Science Vol. 1 No. 2 [June, 2024], pp. 167-182
https://doi.org//10.47514/kjcs/2024.1.2.003

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

depicts the physical domain of the accumulated cost *c*. In other words, when $c \geq R$, then *c* belongs to the fuzzy set HP with membership grade 1.0. In essence, HP for *c* has a trapezoidal membership function. The membership functions of ZE, SP and PO for *c* are triangular with their peak values distributed between the values 0 and *R*. In order to obtain the normalized values of *c*, the observed values were multiplied by the scaling factor 6/*R*. The membership functions of the normalized input value *c* are depicted in figure 3.

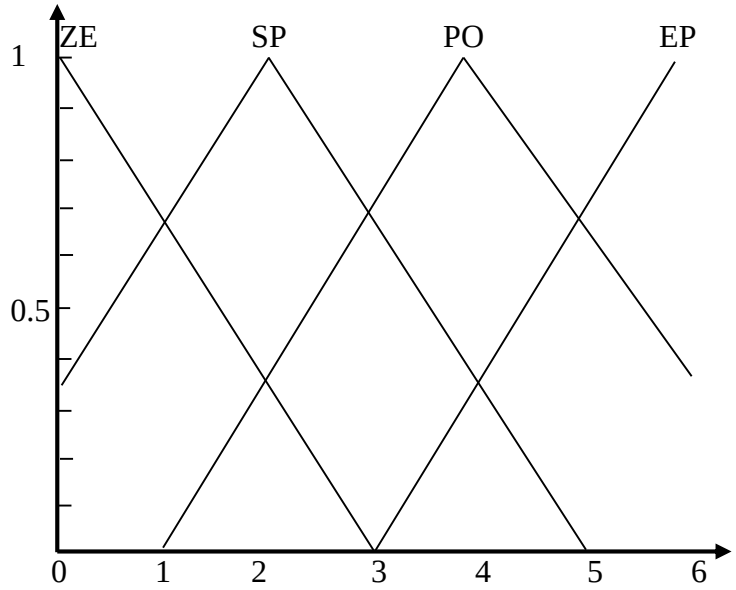
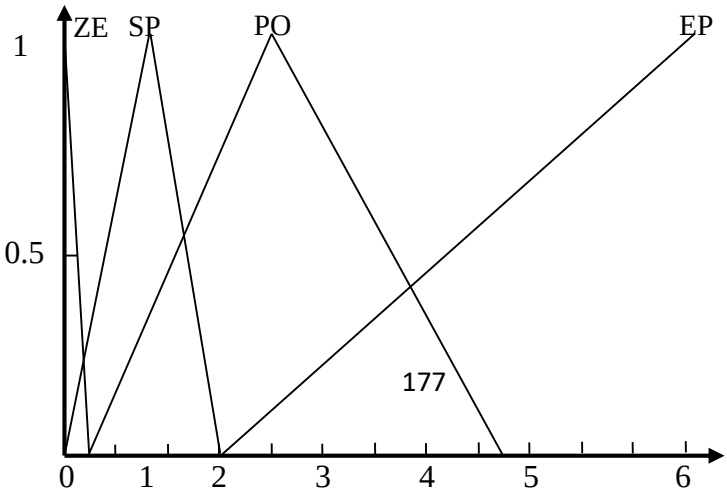


Figure 3: Membership functions for normalized input *c*.

Since $h \geq R$, then *h is of the* set EP with membership grade 1.0. Hence when $h \rightarrow 0$, then *h* is of the set ZE with membership grade 1.0. Similarly, $\rho \in (0, 1)$ as EP for ρ with membership grade 1.0 is at 1 and ZE is at 0. Consequently, the scaling factor of *h* and ρ is 6/*R* and 6 respectively.

If all customers have been served and no more customer is available in the system, the server is switched off at this time. In this case the server is off during *n* time units,

$i = 1, 2, \ldots, n$. The arrival of a new customer in the system corresponds to a time unit and $s_i = s_{i+1} + 1$. The expected number *n* of arrivals in any interval is therefore proportional to ρ. Figure 4 shows the membership functions for normalized input ρ.



Kasu Journal of Computer Science Vol. 1 No. 2 [June, 2024], pp. 167-182
https://doi.org//10.47514/kjcs/2024.1.2.003
Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

Figure 4: Membership functions for normalized input ρ

The membership functions for *h* are the same for ρ, except that the normalized values of *h* are unbounded. Figure 5 shows the membership functions for normalized input *h*.

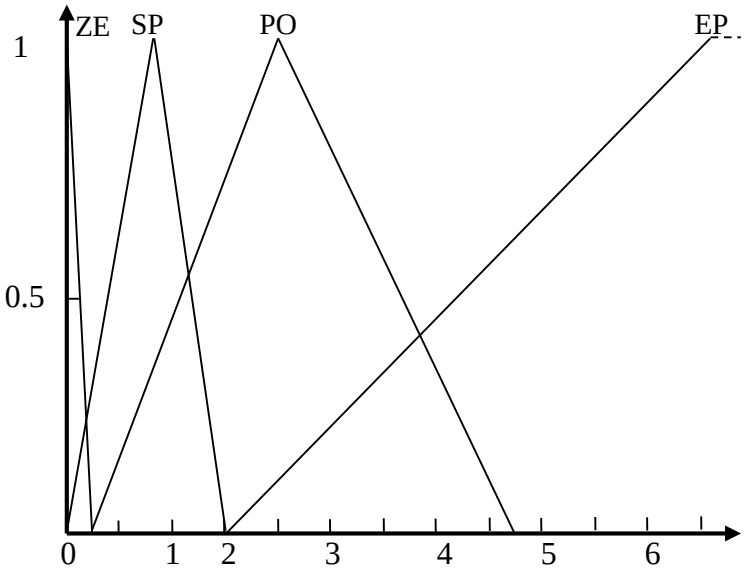


Figure 5: Membership functions for normalized input *h*

3.6. Simulation of the Markov queuing system

Simulation is a modeling technique which involves the use of a computer to generate a sample path and test attributes of a stochastic system. Consequently, it is a symbolic expression of the underlying laws that relate the current state to past states. Thus, if given the state at time zero, a simulation model generates the next states and the corresponding transition epochs in sequence.

4. Results and discussion

The sojourn time in state S_i is of the exponential distribution with λ_i. In addition, the probability of the system moving from state S_i to S_j is λ_{ij} / λ_i, which is independent of the time spent in S_i .. The queuing system has a single state variable that ranges over the set $\{S_1, S_2, ...\}$. If attime zero, the system is in state S_i, then it implies that the system will remain in thisstate for a random interval of time with an exponential

Kasu Journal of Computer Science Vol. 1 No. 2 [June, 2024], pp. 167-182
https://doi.org//10.47514/kjcs/2024.1.2.003
Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

distribution of mean $1/\lambda_i$. From this state, it proceeds to state S_j with probability λ_{ij}/λ for some random time.

The system was simulated beginning with an initial state $k = 0, s = 0$ and $c = 0,$ while employing the fuzzy control at each decision epoch. The first 100 time units were considered the warm-up period for the system. The evolution of *k*and *s* for the first 1,500 time units is depicted in figure 6.

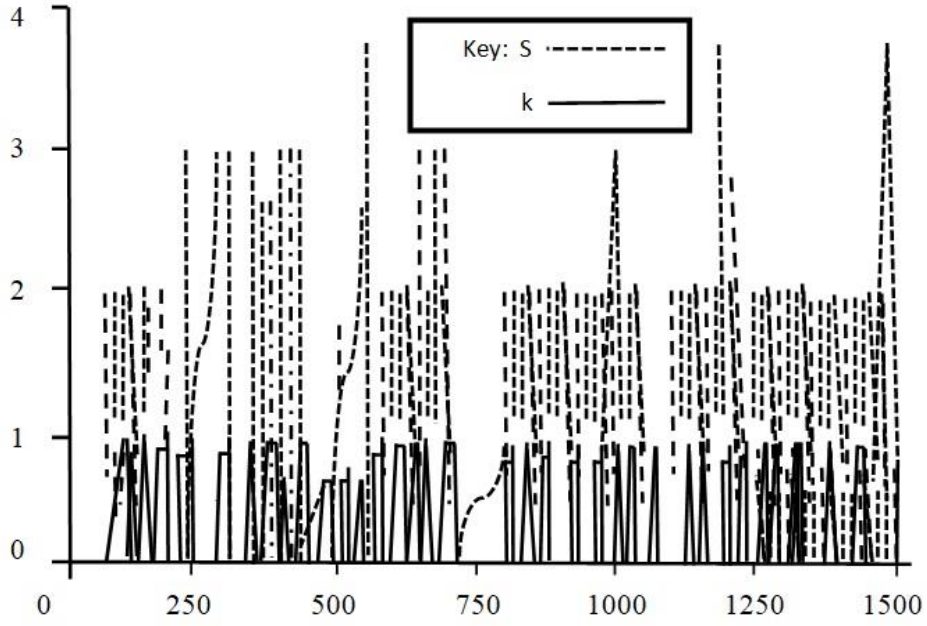


Figure 6: Evolution of *k* and *s*.

An evaluation of figure 6 indicates that it follows an idle server is not immediately switched on to begin service upon the arrival of a customer but rather after multiple arrivals. This is an indication of the hysteretic property of the systemwhich establishes a proportional relationship between the numbers of customers and the number of times the server is turned on. In other words, it shows the interplay between the number of times the server is turned on as a function of *s* during a period of 15,000 time units. We see that the server is most often turned on when there are multiple servers in the system.

If the server capacity is $r \geq 1$ and the server takes *r* customers per service, then If by the end of service less than r customers are waiting service turns, an idle period begins and the server consequently goes on vacation. The server resumes service only when the queue size becomes *r*. In this system, service is resumed when the queue crosses *N* from below and turns off when it crosses *r* from above. This problem is encountered in transportation systems. Queuing systems where the server capacity becomes random following a vacation period are encountered in distributed operating systems and it is a common example of a well-known processor allocation problem.

Kasu Journal of Computer Science Vol. 1 No. 2 [June, 2024], pp. 167-182
https://doi.org//10.47514/kjcs/2024.1.2.003

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

The input process is an orderly stationary Poisson process of $\{r_n; n \in N\}$ with intensity $\lambda > 0$. While r_n is the arrival rate of the N^{th} customer, *N(t)* is the associated counting process. Similarly, if σ_n denotes the service time of Nth batch of customers, it is assumed that $\{\sigma_n; n \in N\}$ and $\{r_n; n \in N\}$ are independent. Let $T_n (T_0 = 0)$, represent the time of the N^{th} service completion, If the queue size is greater than r, then $T_{n+1} - T_n = \sigma_n$ but if the queue size is less than r, $T_{n+1} - T_n = (\text{time for the queue to reach N}) + \sigma_n$. Similarly, if V_n is the number of customers that arrive the queue during the N^{th} service σ_n and *Q(t)* represent the queue size at any time *t*. If at time T_n, at least r customers are awaiting service turns, the server takes a batch of exactly r customers. The service time σ_n in this case is distributed according to a probability distribution function β having a finite first moment *b*. If less than *r* customers are awaiting service turns at time T_n, then an idle period begins. In figure 5, the idle period is more of the values less than 1. Idle period starts every time the queue drops to *r*. Once $N \geq r$, customers are in the queue, service begins.

It is obvious from the graph that hysteresis control essentially complicates analysis as a result of the large numbers of boundary states in which dynamics of the system essentially varies. In contrast to the threshold strategy, the hysteresis strategy adopts possible non-uniqueness of the server whenever it is in its active state under a fixed number of customers in the system. Consequently, construction of a Markov chain that describes the dynamics of the system and its generator is a cumbersome task.

5. Conclusion

The use of a fuzzy-based approach in ensuring optimality in the usage of network resources including servers' time had proven effective as evident in the results. Consequently, it had been shown that a fuzzified process of optimal control and usage of servers is efficient in the management of queue network. In essence, the proposed method had proven to be efficient in the control of admission and servers' resources in an *M/M/1* queue system with server vacations. Consequently, it is suitable in the control of admission in traffic and communications control systems since the control of admission is of its threshold type.

References

Badian-Pessot, P., Lewis, M. E. and Down, D. G. (2019). Optimal Control Policies for an *M/M/*1Queue with a

Removable Server and Dynamic Service Rates. *Probability in the Engineering and Informational Sciences*. 11(4). pp.1-21. DOI:10.1017/s0269964819000299.

Bell, C. E. (2011). Characterization and Computation of Optimal Policies for Operating an M/G/1 Queuing System with Removable Server. *Operations Research*. 19(1). pp. 208-218.

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

Borthakur, A., Medhi, J. andGohain. A. (2019). Poisson Input Queuing System with Startup Time and Under Control-Operating Policy. *Computers and Operations Research*. 14(1). pp. 33-40.

Crabill, T. B. (1974). Optimal Control of a Maintenance System with Variable Service Rates. *Operations Research*. 22(4). pp.736-745.

Deepa, B. and Kalidass, K. (2018). Markovian Queue with Finite Capacity, Working Breakdowns and Server Vacations. *International Journal of Pure and Applied Mathematics*. 10(18). pp. 859-873.

Dimitrakopoulos, Y. andBurnetas, A. (2017). The Value of Service Rate Flexibility in an M/M/1 Queue with Admission Control. *IISE Transactions*. 49(6). pp. 603-621.

Dong-Yuh, Y. and Yi-Hsuan, C. (2018). Computation and Optimization of a Working Breakdown Queue with Second Optional Service. *Journal of Industrial and Production Engineering.* 35(3) pp. 181-188.

Federgruen, A. and So, K. C. (2021). Optimality of Threshold Policies in Single-Server Queuing Systems with Server Vacations. *Advances in Applied Probability*. 23(2). pp. 388-405.

Feinberg, E. A. and Kella. O. (2002). Optimality of D-Policies for an M/G/1 Queue with a Removable Server. *Queuing Systems*. 42(4). pp. 355-376.

Gandhi, A., Gupta, V., Harchol-Balter, M. and Kozuch, M. A. (2020). Optimality Analysis of Energy-Performance Trade-Off for Server Farm Management. *Performance Evaluation*. 67(11). pp.1155-1171.

Gandhi, A., Harchol-Balter, M. and Adan, I. (2020). Server Farms with Setup Costs. *Performance Evaluation*. 67(11). pp.1123-1138.

Gebrehiwot, M. E., Aalto, S. A. and Lassila, P. (2020). Optimal Sleep-State Control of Energy-Aware *M/G/1* Queues. *Proceedings of the 8th International Conference on Performance Evaluation Methodologies and Tools*. pp. 82-89. 2014.

George, J. M. and Harrison, J. M. (2021). Dynamic Control of a Queue with Adjustable Service Rate. *Operations Research*. 49(5). pp. 720-731.

Heyman, D. P. (2018). Optimal Operating Policies for M/G/1 Queuing Systems. *Operations Research*. 16(12) pp. 362-382.

Kalidass, K. and Kasturi, R. (1995). A Queue with Working Breakdowns. *Computers in Industrial Engineering.* 18(17) pp. 779-783.

Kalidass, K. and Pavithra. K. (2016). An *M/M/1/N* Queue with Working Breakdowns and Bernoulli Feedbacks. *International Journal of Applied Mathematics*. 6(2). pp. 1212-1228.

Kim, B. K. and Lee, D H. (2021). The *M/G/1* Queue with Disasters and Working Breakdowns. *Applied Mathematical Modeling.* 38(15).1788-1798.

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

Ke, J. C. (20123). The Optimal Control of an M/G/1 Queuing System with Server Start-Up and Two Vacation Types. *Applied Mathematical Modeling*. 27(6). pp. 437-450.

Kumar, R., Lewis, M. E. and Topaloglu, H. (2023) Dynamic Service Rate Control for a Single-Server Queue with Markov-Modulated Arrivals. *Naval Research Logistics (NRL).* 60(8). pp. 661-677.

Liou, C. D. (2023). Markovian Queue Optimization Analysis with an Unreliable Server Subject to Working Breakdowns and Impatient Customers. *International Journal of Systems Science*. pp. 1121-1130.

Lippman, S. A. (2015). Applying a New Device in the Optimization of Exponential Queuing Systems. *Operations Research.* 23(4) pp. 687-710.

Maccio, V. J. and Down, D. G. (2022). On Optimal Control for Energy-Aware Queuing Systems". *In Tele-traffic Congress (ITC 27)*. 27th International, pp. 98-106. IEEE.

Vijayalakshmi, V. and Kalidass, K. (2018). The Single Server Vacation Queue with Geometric Abandonments and Bernoulli's Feedbacks. *International Journal of Engineering and Technology*. 7 (2) pp. 172-179.

Vijayalakshmi, V., Kalidass, K. and Deepa, B. (2021). Cost Analysis of *M/M/1/N* Queue with Working Breakdowns and a Two-Phase Services. *Journal of Physics*: Conference Series 1850. IOP Publishing. DOI:10.1088/1742-6596/1850/1/012026

Wang, K. H., Wang, T. Y. and Pearn, W. L (2017). Optimal Control of the N-Policy M/G/1 Queuing System with Server Breakdowns and General Startup Times. *Applied Mathematical Modeling*. 31(10). pp. 2199-2212.

Yadin, M. and Naor, P. (2013). Queuing Systems with a Removable Service Station. *Journal of the Operational Research Society*. 14(4). pp 393-405.

Yang. D. and Chen, Y. H. (2018). Optimize the *M/M/1* Queue with Second Optional Service and Working Breakdowns. *Journal of Industrial and Production Engineering*. 35(9). pp.181-188.

Zhang, R., Phillis, Y. A. and Kouikoglou, V. S. (2005). *Fuzzy Control of Queuing Systems*. Springer. United States of America. pp 17-21.