Original Research Article.

Performance assessment of queuing networks with intermittently accessible servers, failed servers, corrective maintenance, feedback, and customer abandonment

Abstract

The aim of this work is to evaluate the performance of M/M/K (K>2) multiserver queuing networks with intermittently accessible servers, failed servers, corrective maintenance, feedback and client dropouts. We first studied the two-server heterogeneous M/M/2 model, in which Server 1 is considered to be subject to outages with corrective maintenance or repair, and Server 2 is intermittently accessible, i.e. accessibility is possible but sometimes interrupted. Using the Geometric Matrix Method (GMM), we obtained the equilibrium equations of the system, and solving them by substitution enabled us to obtain the steady-state probabilities. Using these probabilities, we obtained measures of system performance. In a second step, we extended the M/M/2 model to include a variable number of servers (M/M/K (K>2) model). As this model is more complex to analyze numerically, we used the algorithmic method. Firstly, we used the PSO algorithm to minimize operational costs by dynamically adjusting the arrival rate λ, the service rate μ and the number of servers K. Secondly, we used the PSO algorithm to minimize the average waiting time and the abandonment rate in order to maximize customer satisfaction. This will benefit both the operator and the customer.

Keywords: *Queue networks, M/M/K model, PSO algorithm, intermittent servers, servers down, corrective maintenance, aborting, operational costs*

MSC: 60K20, 60K25, 60K30

1 Introduction

Queuing systems with heterogeneous servers, customer feedback and abandonment have several applications, including manufacturing systems, computer systems, telecommunications systems, etc.

Several recent studies have dealt with queue models with heterogeneous servers, feedback and customer abandonment.

In this paper, we first consider a system of two heterogeneous servers (M/M/2 model), one of which is intermittently accessible, i.e. temporarily unavailable, while the other is subject to failures requiring corrective maintenance or repair. In a second step, we will extend this model to include a variable number of servers (M/M/K (K>2) model).

Our work will be structured in two main parts:

- In the first part, for the M/M/2 model, we will first use the Geometric Matrix Method (GMM) to obtain the equilibrium equations, then we will solve these equations by substitution to obtain the equilibrium probabilities, and finally we will deduce the system performance measures,
- In the second part, for the M/M/K (K>2) model, we will integrate optimization algorithms to minimize operational costs or maximize customer satisfaction, provide a detailed analysis of performance metrics and explore strategies to optimize performance, add a component to model customer impatience and numerically evaluate the influence of parameters on system performance.

.

1.1 Survey of literature

A great deal of research has focused on the study of the 2-server heterogeneous M/M/2 queue model, where one server operates normally without interruption and the other is temporarily unavailable. This is referred to as an intermittently accessible server.

Agassi Melikov and Sevinj Aliyeva [1] studied a system with heterogeneous servers, a Markov modulated Poisson flow and instantaneous feedback. They used approximate algorithms to obtain steady-state probabilities of models with finite and infinite queues and presented results from numerical experiments.

R. Sethi, M. Jain, R.K. Meena and D. Garg [2] studied the impatient behavior of customers in the case of an unreliable Markovian M/M/1 server queuing system operating under policy N. The failed server is repaired according to the threshold policy, i.e. repair begins when the system reaches capacity. They used a recursive technique to obtain the steady-state distribution of the queue length. Various performance indices have been established and numerical experiments are carried out. The numerical results obtained are compared with the soft computing technique based on an adaptive neuro-fuzzy inference system. A cost function is formed and a quasi-Newton technique is applied to reduce the total cost associated with multiple activities in the queuing system.

Toky Basilide Ravaliminoarimalalason and Falimanana Randimbindrainibe [3] have studied discrete resource sharing in a queuing system. They have built an analytical model of the distribution of occupied resources that can help with resource sizing. Infinite and finite quantities of discrete server resources are highlighted and validated with special cases of individual resource requirements following the Poisson and binomial distributions.

C. Shekhar, S. Varshney and A. Kumar Matrix-geometric solution [4] treated a multi-server finite-capacity queuing system with Bernoulli's modified scheduled vacation policy and realistic non-subscriber customer retention policy. They used the matrix analytical approach to obtain steady-state probabilities with which various system performance measures were also developed with practical justification. Finally, the expected cost minimization problem was formulated and treated with the meta-heuristic approach.

S. Thakur, A. Jain and M. Jain [5] have studied the Markovian finite-capacity queuing model M/M/1/N with vacations and impatient customer behavior. Using input and output transition rates, they constructed state probability equations for different system states. To calculate the length of the stationary queue, they performed a matrix-geometric analysis. Sensitivity analysis was carried out to validate system performance measures. To examine the scope of the adaptive neuro-fuzzy inference system (ANFIS), computational results were presented using matrix-geometric and ANFIS approaches.

M. Seenivasan, and K. S. Subasri [6] analyzed a Markov queueing system with two heterogeneous servers, one with intermittent availability and the other with unreliable clients and performance problems. Server 1 is always available and Server 2 is intermittently available and also prone to failures. Incoming clients form a single queue, and arrival follows the Poisson process with parameter λ. The heterogeneous servers follow exponential service rates μ_1 for server 1 and μ_2 for server 2 ($\mu_1 \neq \mu_2$). They used the geometric matrix technique to obtain the steady-state probabilities. The numerical study was analyzed in various performance measures for different parameter values.

M.Seenivasan, M.Indumathi and V. J. Chakravarthy[7] analyzed the performance of two heterogeneous server queue models with one server that can be accessed intermittently. By introducing the bivariate process, he obtained steady-state probabilities using the matrix geometric method. Some numerical results were obtained.

Toky Basilide Ravaliminoarimalalason and F. Randimbindrainibe [8]used game theory to optimize resource sharing in queuing networks according to customer aspirations. She used analytical methods and algorithms for static and dynamic models applicable to distributed systems such as cloud computing.

M. Seenivasan, R. Senthilkumar and K.S. Subasri [9] studied a heterogeneous M/M/2 queuing system, which features two types of server failures with disaster recovery. Server 1 is always available and Server 2 is intermittently available and may be subject to failures. They used the matrix-geometric technique (MGT) to describe the model and derive probability vectors. At the end, they gave some numerical models and found performance measures using the probability vectors at different values of contingent parameters of the queuing model.

M. Seenivasan, M. Kameswari and M. Indumathi [10] considered two heterogeneous server queuing models with different types of service rates. They obtained the static probability line vectors using the geometric matrix method. In addition, they performed numerical analysis as Pemp, PRes in 2nd Int, PRes in 2nd work, P2nd Int, P2nd working as well as MNS assuming specific values for the parameter.

M. Seenivasan and R. Abinaya[11] studied a queueing model with single working vacations and catastrophes in which he considered a system with a single holiday and different service rates. He used the geometric matrix method (GMM) to obtain steady-state probability vectors and some performance measure have been determined.

Divya Kothandaraman and Indhira Kandaiyan [12] investigated the concept of a Markovian queueing model with heterogeneous, intermittently available servers and feedback as part of a hybrid vacation policy. They addressed both the asymmetric transition representation and the hybrid vacation policy. They presented necessary and sufficient conditions for system stability. In addition, they derived the steady-state probability distribution of the queueing model using the geometric matrix method. In addition, they constructed some formulas to determine the model's performance indicators. Finally, they studied the influence of system parameters using a few numerical examples.

S. Wang, Z. Ma, X. Niu and Y. Liu [13] analyzed the performance of a vacation-based queuing system with failover and spare servers in the MP2P network. They built a multiple-vacation-based M/M/c+d queuing model that synchronizes with fail-repairable and spare servers, and variable service rates, then used this model to simulate the performance of an MP2P network in various wireless environments. They used the quasi-birth-and-death (QBD) process, the geometric matrix solution method and the Gauss-Seidel iteration method to obtain expressions for the system's performance indicators. Numerical experiments enabled them to study the effects of system parameters on MP2P system performance indicators, such as online energy consumption. Furthermore, using Nash equilibrium and the optimal social strategy, the value of the maximum social benefit under social optimization is obtained, providing a decision framework for the MP2P system to improve network performance and reduce online energy consumption.

N. Singh, M. Jain and S. Dhibar [14] studied the M/M/∞ queuing system with a server failure and two different types of customers called lossy customers and delayed customers. The stochastic decomposition property of the system size was established. Sensitivity analysis was performed using numerical examples. The Adaptive Neuro Fuzzy Inference System (ANFIS) technique is implemented to develop the ANFIS controller to evaluate the simulation results by including the realistic feature of fuzzy descriptors that can be noticed in several queuing scenarios.

Darvinder Kumar and Gulab Singh Bura, [15] have considered a finite-capacity Markovian queuing system with two identical servers under two environmental conditions. Changing environmental conditions also affect the state of the queuing system. In addition, the system is also subject to random disasters, which destroy all current clients in the system. A repair process is then initiated and, once the repair is successful, the system is ready to operate. In this case, the repair time is called the recovery time. They modeled this queuing system to the solution of the transient state using the probability generating function technique.

Anderson Ribeiro Duarte [16] presented an approach for optimal server allocation in several topologies (series, merge and split). The methodology employs a multi-objective heuristic strategy using the simulated annealing algorithm. Server utilization performance was maximized simultaneously with the minimization of total expected client time in the queuing network. The results of the computer experiments demonstrated the effectiveness of the proposal.

K. Divya [17] studied a model of queues with 2 heterogeneous servers with feedback, one of which is accessible intermittently and the other is assumed to always be accessible without disturbances. He used the matrix geometry method to determine the probability vectors that allowed him to evaluate metrics such as server state probabilities and the average number of customers in the system.

1.1.1 Model construction

Consider a system of queues of 2 heterogeneous servers with feedback and customer abandonment. Customers arrive in the system through a Poisson process of λ rate. The service levels for servers 1 and 2 are μ_1 and μ_2 respectively and follow an exponential distribution. Server 1 is prone to γ_1 rate failures and benefits from ρ_1 rate corrective maintenance. Server 2 is either idle or intermittently accessible. Customers enter the system based on the status of the servers, i.e. whether they are active or inactive or intermittently accessible. Inactive Server 2 can either become active to provide service at a rate η_0, or enter a period of intermittent accessibility to provide a service at rate η_1. After receiving service, a dissatisfied customer may decide to request additional service with a $\overline{\theta}$ rate (feedback) or permanently abandon the system with a $\theta = 1 - \overline{\theta}$ rate.

The state of the system is described by: $S(t) = (N(t), S_1(t), S_2(t))$ with:

- $N(t)$, the number of customers in the system.
- $S_1(t)$ describes the state of server 1 and:

$$S_1(t) = \begin{cases} 0 \text{ inactive i.e. no customer in service} \\ 1 \text{ active, i.e. serving a customer} \\ 2 \text{ i.e. down} \\ 3 \text{ i.e. under corrective maintenance} \end{cases}$$

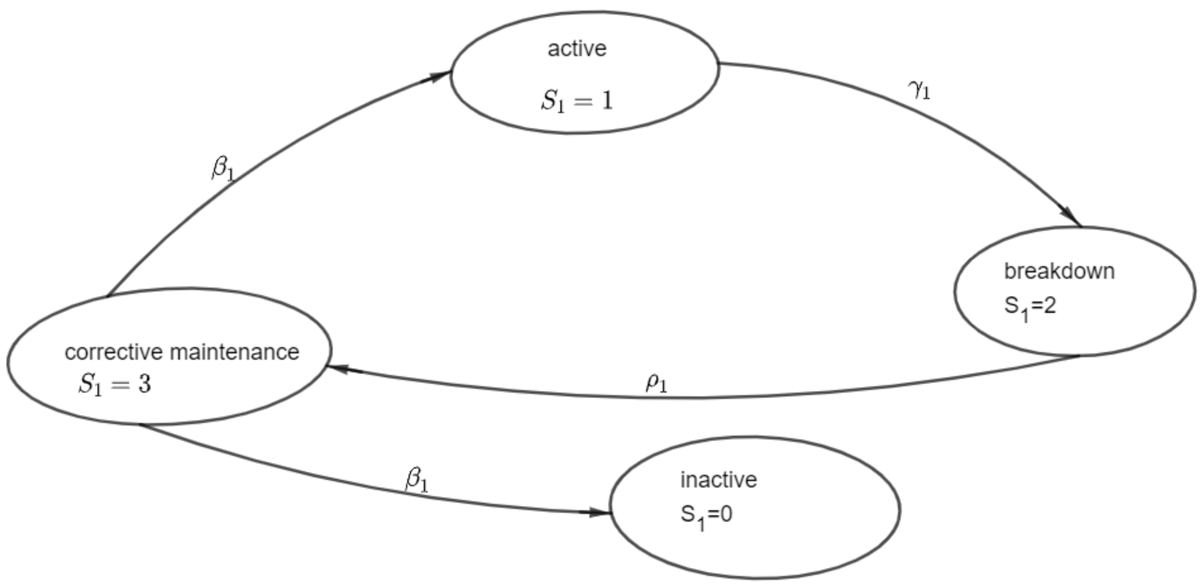
- $S_2(t)$ describes the state of server 2 and:

$$S_2(t) = \begin{cases} 0 \text{ inaccessible i.e. cannot serve a customer} \\ 1 \text{ accessible i.e. can serve a customer} \end{cases}$$

Hence the state space is:

$\Omega = \{(N, S_1, S_2) : N \geq 0, S_1 \in \{0, 1, 2, 3\}, S_2 \in \{0, 1\}\}$

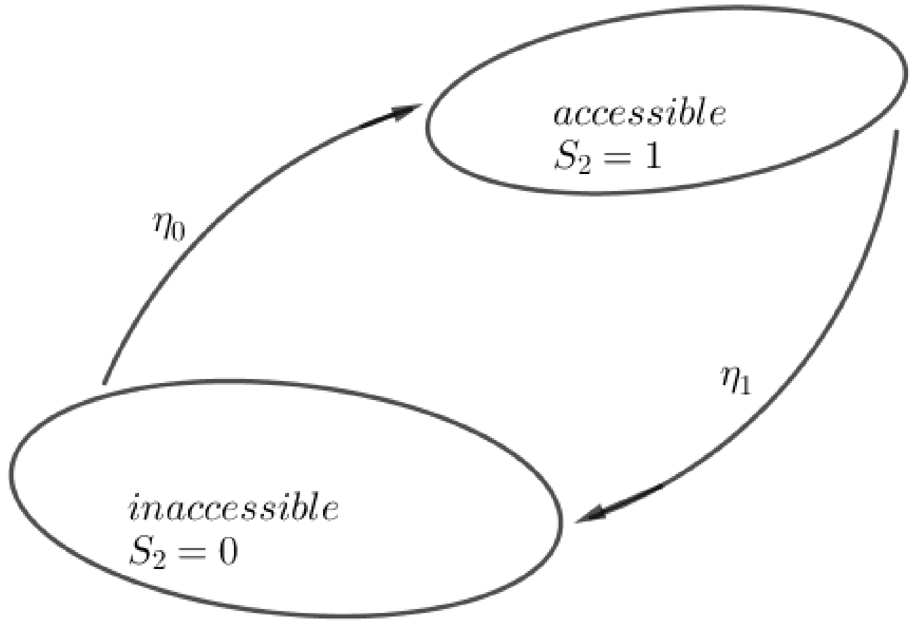
GRAPH 1 **Internal transition graph for server 1 (subject to failures with corrective maintenance)**



The graph shows that server 1 goes from the active state ($S_1 = 1$) to the failed state with a failure rate of γ_1 then from the failure state to the corrective maintenance state ($S_1 = 3$) with a rate of ρ_1 and then from the corrective maintenance state ($S_1 = 3$) to the idle state ($S_1 = 0$) or active ($S_1 = 1$) with a repair rate of β_1. When $S_1 = 1$, it can serve a customer with a rate

μ_1 reducing the number of customers in the system by 1.

GRAPH 2 **Internal transition graph for server 2 (intermittent)**



The graph shows that server 2 goes from the accessible state ($S_2 = 1$) to the inaccessible state ($S_2 = 0$) with a rate of η_1 and then from the inaccessible state ($S_2 = 0$) to the accessible state ($S_2 = 1$) with a rate of η_0. When Server 2 is accessible, it can serve a customer with a μ_2 rate, but inaccessible, it cannot serve customers.

The infinitesimal generator G is defined by:

$$G = \begin{pmatrix} A_0 & B_0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ B_2 & B_1 & B_0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & B_2 & B_1 & B_0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & B_2 & B_1 & B_0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & B_2 & B_1 & B_0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots \end{pmatrix}$$

with:

• A_0, the matrix of internal transitions of servers. It takes into account failures, repairs, corrective maintenance, and alternations between accessible and inaccessible states for Server 2.

• B_0, the matrix of customer arrivals that occur at a λ rate.

• B_1, the matrix of customer departures after service. These departures depend on the active servers $S_1 = 1$ or $S_2 = 1$.

• and B_2, the matrix of customer abandonments that occur at a θ. Applying the matrix geometric method, we get:

$$A_0 = \begin{pmatrix} -(\beta_1+\eta_0) & \eta_0 & \beta_1 & \beta_1+\eta_0 \\ \eta_1 & -(\beta_1+\eta_1) & \beta_1+\eta_1 & \beta_1 \\ \beta_1 & \beta_1+\eta_0 & -(\gamma_1+\eta_0) & \eta_0 \\ \beta_1+\eta_1 & \beta_1 & \eta_1 & -(\gamma_1+\eta_1) \end{pmatrix}$$

$$B_0 = \begin{pmatrix} 0 & \lambda & 0 & 0 \\ 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & \lambda \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$B_1 = \begin{pmatrix} -(\mu_1+\mu_2+\theta) & \eta_1 & \gamma_1 & 0 \\ \eta_0 & -(\mu_1+\theta) & 0 & 0 \\ \rho_1 & 0 & -(\mu_2+\theta) & 0 \\ \rho_1 & \beta_1 & 0 & -(\mu_2+\theta) \end{pmatrix}$$

$$B_2 = \begin{pmatrix} -\theta & 0 & 0 & 0 \\ 0 & -\theta & 0 & 0 \\ 0 & 0 & -\theta & 0 \\ 0 & 0 & 0 & -\theta \end{pmatrix}$$

Let be B the matrix of the system's global transitions.

We have:

$$\begin{aligned} B &= B_0+B_1+B_2 \\ &= \begin{pmatrix} 0 & \lambda & 0 & 0 \\ 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & \lambda \\ 0 & 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} -(\mu_1+\mu_2+\theta) & \eta_1 & \gamma_1 & 0 \\ \eta_0 & -(\mu_1+\theta) & 0 & 0 \\ \rho_1 & 0 & -(\mu_2+\theta) & 0 \\ \rho_1 & \beta_1 & 0 & -(\mu_2+\theta) \end{pmatrix} + \begin{pmatrix} -\theta & 0 & 0 & 0 \\ 0 & -\theta & 0 & 0 \\ 0 & 0 & -\theta & 0 \\ 0 & 0 & 0 & -\theta \end{pmatrix} \end{aligned}$$

Hence:

$$B = \begin{pmatrix} -(\mu_1+\mu_2+2\theta) & \lambda+\eta_1 & \gamma_1 & 0 \\ \eta_0 & -(\mu_1+2\theta) & \lambda & 0 \\ \rho_1 & 0 & -(\mu_2+2\theta) & \lambda \\ \rho_1 & \beta_1 & 0 & -(\mu_2+2\theta) \end{pmatrix}$$

Theorem 1.1. *Let $P=(p_0,p_1,p_2,p_3)$ be the probability vector. The probability vector is ob-*

tained by solving the equations $PB = 0$ and $\sum_{i=0}^{3} p_i = 1$. We have the following system:

$$-\left(\mu_1 + \mu_2 + 2\theta\right) p_0 + \eta_0 p_1 + \rho_1 p_2 + \rho_1 p_3 = 0 \tag{1}$$

$$\left(\lambda + \eta_1\right) p_0 - \left(\mu_1 + 2\theta\right) p_1 + \beta_1 p_3 = 0 \tag{2}$$

$$\gamma_1 p_0 + \lambda p_1 - \left(\mu_2 + 2\theta\right) p_2 = 0 \tag{3}$$

$$\lambda p_2 - \left(\mu_2 + 2\theta\right) p_3 = 0 \tag{4}$$

Proof 1.1. *Solving the system by substitution, we get:*

$$p_0 = \frac{\left(\mu_1 + 2\theta\right)\left[\left(\mu_1 + \mu_2 + 2\theta\right)\left(\mu_2 + 2\theta\right)^2 - \gamma_1 \rho_1 \left(\mu_2 + 2\theta + \lambda\right)\right] - \beta_1 \lambda \left[\lambda\left(\mu_1 + \mu_2 + 2\theta\right) + \eta_0 \gamma_1\right]}{C + D} \tag{5}$$

with:

$$C = \left(\mu_1 + 2\theta + \lambda + \eta_1\right)\left[\left(\mu_1 + \mu_2 + 2\theta\right)\left(\mu_2 + 2\theta\right)^2 - \rho_1 \gamma_1 \left(\mu_2 + 2\theta + \lambda\right)\right]$$

and:

$$D = \left[\lambda\left(\mu_1 + \mu_2 + 2\theta\right) + \eta_0 \gamma_1\right]\left[-\beta_1 \lambda + \left(\lambda + \eta_1\right)\left(\mu_2 + 2\theta\right) + \lambda\left(\lambda + \eta_1\right)\right]$$

$$p_1 = \frac{\left(\lambda + \eta_1\right)\left[\left(\mu_1 + \mu_2 + 2\theta\right)\left(\mu_2 + 2\theta\right)^2 - \rho_1 \gamma_1 \left(\mu_2 + 2\theta + \lambda\right)\right]}{C + D} \tag{6}$$

$$p_2 = \frac{\left(\lambda + \eta_1\right)\left(\mu_2 + 2\theta\right)\left[\lambda\left(\mu_1 + \mu_2 + 2\theta\right) + \eta_0 \gamma_1\right]}{C + D} \tag{7}$$

$$p_3 = \frac{\lambda\left(\lambda + \eta_1\right)\left[\lambda\left(\mu_1 + \mu_2 + 2\theta\right) + \eta_0 \gamma_1\right]}{C + D} \tag{8}$$

Using these probabilities, the following performance measures of the system are deduced:

- Let L_q, the average number of customers in the queue, and ρ, the occupancy rate of the system.

  We have:

  $$L_q = \frac{p_2 \rho}{2\left(1 - \rho\right)^2} \tag{9}$$

  Since:

  $$\rho = \frac{\lambda}{2\mu} \tag{10}$$

Then:

$$L_q = \frac{p_2 \lambda}{4\mu \left(1 - \dfrac{\lambda}{2\mu}\right)^2} \tag{11}$$

Replacing p_2 with its expression, we have:

$$L_q = \frac{\lambda \left(\lambda + \eta_1\right) \left(\mu_2 + 2\theta\right) \left[\lambda \left(\mu_1 + \mu_2 + 2\theta\right) + \eta_0 \gamma_1\right]}{4\mu \left(1 - \dfrac{\lambda}{2\mu}\right)^2 (C + D)} \tag{12}$$

- Let be L, the average number of customers in the system.

  We have:

$$L = L_q + \frac{\lambda}{\mu} \tag{13}$$

with $\dfrac{\lambda}{\mu}$, the average number of customers in service. Replacing L_q with its expression, we have:

$$L = \frac{\lambda \left(\lambda + \eta_1\right) \left(\mu_2 + 2\theta\right) \left[\lambda \left(\mu_1 + \mu_2 + 2\theta\right) + \eta_0 \gamma_1\right]}{4\mu \left(1 - \dfrac{\lambda}{2\mu}\right)^2 (C + D)} + \frac{\lambda}{\mu} \tag{14}$$

- Let W_q, the average wait time in the queue.

  We have:

$$W_q = \frac{L_q}{\lambda} \tag{15}$$

Replacing L_q with its expression, we have:

$$W_q = \frac{\left(\lambda + \eta_1\right) \left(\mu_2 + 2\theta\right) \left[\lambda \left(\mu_1 + \mu_2 + 2\theta\right) + \eta_0 \gamma_1\right]}{4\mu \left(1 - \dfrac{\lambda}{2\mu}\right)^2 (C + D)} \tag{16}$$

- Let be R, the customer abandonment rate and $T_{threshold}$, the threshold time set by customers.

  We have:

$$R = \begin{cases} 1 \text{ if } W_q > T_{threshold} \\ \dfrac{W_q}{T_{threshold}} \quad \text{otherwise} \end{cases} \tag{17}$$

Replacing W_q with its expression, we have:

$$R = \begin{cases} 1 \text{ if } W_q > T_{threshold} \\ \dfrac{\left(\lambda + \eta_1\right) \left(\mu_2 + 2\theta\right) \left[\lambda \left(\mu_1 + \mu_2 + 2\theta\right) + \eta_0 \gamma_1\right]}{4\mu \left(1 - \dfrac{\lambda}{2\mu}\right)^2 (C + D)\, T_{threshold}} \quad \text{otherwise} \end{cases} \tag{18}$$

2 M/M/K Network Model (K>2)

2.1 Model Construction

Consider a multi-server network of queues of type M/M/K (K>2). Customers arrive in the network through a Poisson process of global rate λ and service times are exponentially distributed by global rate μ and there are K servers (K>2) divided into two categories: some are prone to outages with corrective maintenance and other servers that can be accessed intermittently. Let θ be the overall rate of customer abandonment when they are dissatisfied with the quality of service. A subset of the servers ($K_{outages}$) can fail with a probability of p_{outage} and the repair time follows an exponential distribution of rates μ_r. Another subset of servers ($K_{intermittent}$) can become temporarily unavailable with a probability of p_i. The servers are restored to efficiency after repair.

2.2 Modeling Objectives

As the model is more complex to analyze digitally, this requires the use of algorithms in order to:

- minimize operational costs, i.e. reduce costs related to servers (operating or under repair), reduce costs related to breakdowns, reduce costs related to corrective maintenance, reduce costs related to intermittencies, reduce costs related to customer losses i.e. abandonments,
- maximize customer satisfaction, i.e. minimize the abandonment rate and the average waiting time.

We will therefore first use the Particle Swarm Optimization (PSO) algorithm to dynamically adjust the λ arrival rate, the μ service level, the number of servers K and minimize the total operational cost C. In a second step, we will use the PSO to minimize the abandonment rate and the average waiting time in order to maximize customer satisfaction, which will therefore be beneficial for both the operator and the consumers (or customers).

2.3 Minimizing Operational Costs

Using the PSO algorithm and Using Matlab software, we obtain the optimal values for the arrival rate λ, the service level μ, the number of servers K, the total operational cost C and the performance parameters of the M/M/K networks.

2.3.1 Results obtained

TABLE 1The results obtained are recorded in the following table:

Arrival rate λ	10.00
Service Rates μ	10.00
Number of Servers K	2
Minimum total cost$C$	273.54
Occupancy rate ρ	0.50
Average number of customers in queue L_q	0.00
Average Wait Time W_Q	0.00 Hour
Average number of customers in the system$L$	1.00
Average total time W	0.10 hour

This table gives the best configurations of the arrival rate, service level, and number of servers to minimize operational costs. The minimum total operational cost is 273.54 and the optimal values of the arrival rate, service level and number of servers to achieve this minimum cost are 10, 10 and 2 respectively. The average number of customers in the queue as well as the average waiting time are zero.

2.3.2 Numerical simulations

By numerical simulation, the PSO algorithm gives the evolution of the minimum total operational cost and the performance parameters of the M/M/K network as a function of the arrival rate, the service rate and the number of servers. The following curves are obtained:

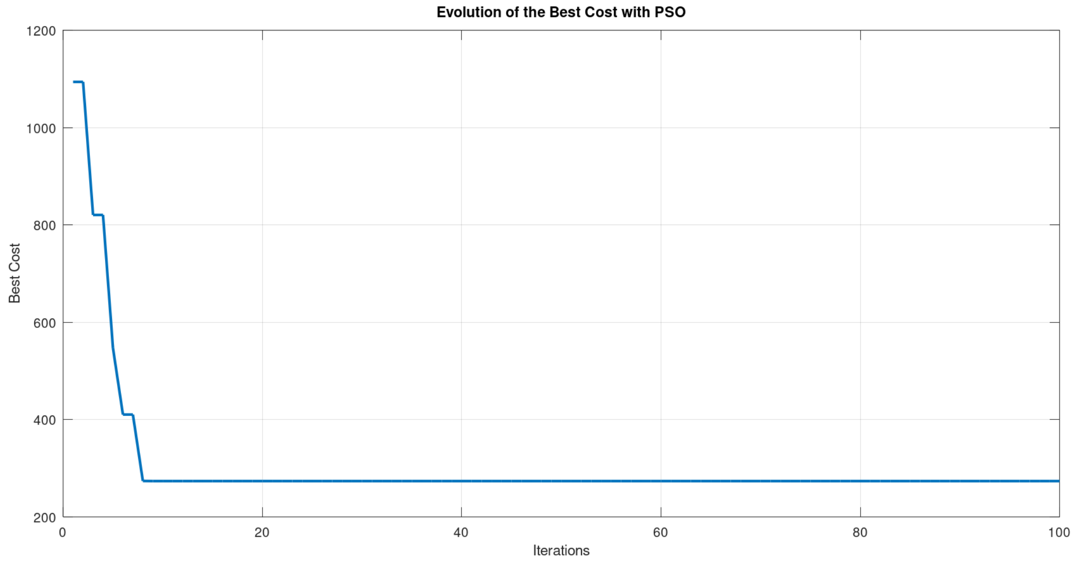


Figure 1:Evolution of the Best cost with PSO

The figure shows that from the first to the tenth iteraton, the minimum total operational cost decreases and from the tenth to the hundredth iteration, the cost remains constant. On the one hand, this indicates better operational efficiency. Increasing the number of servers allows for better load management, reducing outages and downtime. This reflects value for money that improves the overall profitability of infrastructure. On the other hand, the cost remaining fixed despite the increase in the arrival rate, the service level and the number of servers, can be explained by inefficiency in the management of resources.

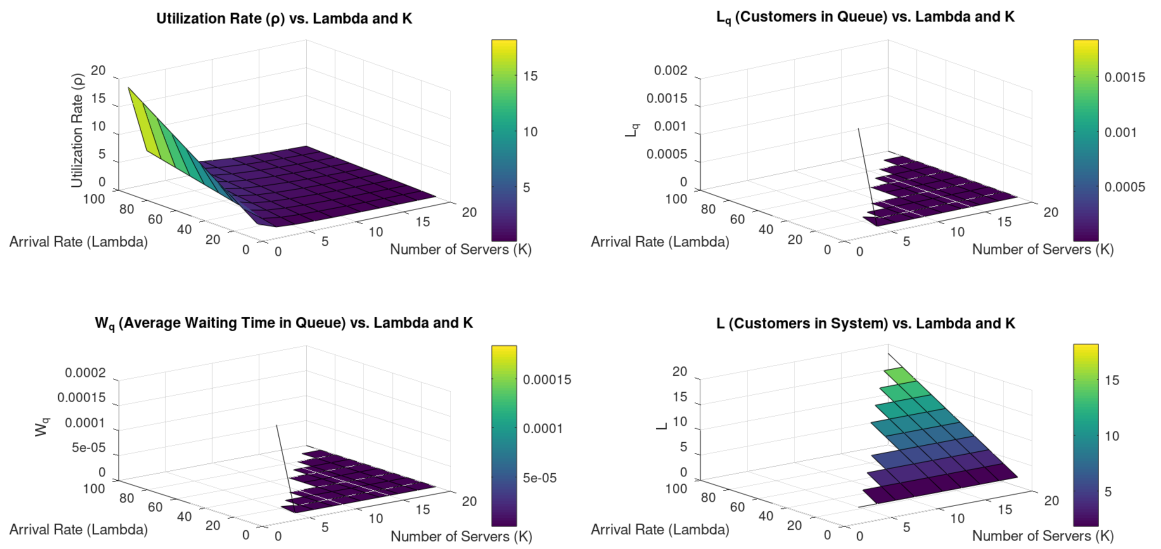


Figure 2: Numerical simulation for high performing system

In these graphs, the blue and green areas indicate a high-performing system with low values for the average number of customers in the queue L_q, the average wait time W_q, the average number of customers in the system L, and moderate occupancy. Orange areas indicate increasing load and high pressure on the system. These areas are therefore points of attention. Black areas indicate critical conditions where performance is unstable or infinite ($\rho \geq 1$).

2.4 Maximizing Customer Satisfaction

To maximize customer satisfaction, we use the PSO algorithm to reduce wait time and abandonment rate by using a weighted cost function to combine them.

2.4.1 Results obtained

Using the R software, we obtain the optimal values for the arrival rate λ, the service level μ, the number of servers K, the minimum total cost C, the minimum average wait time (W_q) and the minimum abandonment rate (R). These results are recorded in the following table:

Arrival rate λ	10
Service Rates μ	10
Number of Servers K	20
Minimum total cost C	0
Minimum average wait time W_q	0
Minimum Abandonment Rate R	0

Table 2. Best configurations of different parameters

This table gives the best configurations of arrival rate, service rate, number of servers, and minimum total cost to minimize abandonment rate and average wait time. The optimal values for the arrival rate, service level, number of servers and minimum total cost are respectively 10; 10; 20 and 0. The minimum average wait time and minimum dropout rate are 0 and 0, respectively.

2.4.2 Numerical simulations

By numerical simulation, the PSO algorithm gives the evolution of the abandonment rate and the average waiting time as a function of the minimum total cost. The following curves are obtained:

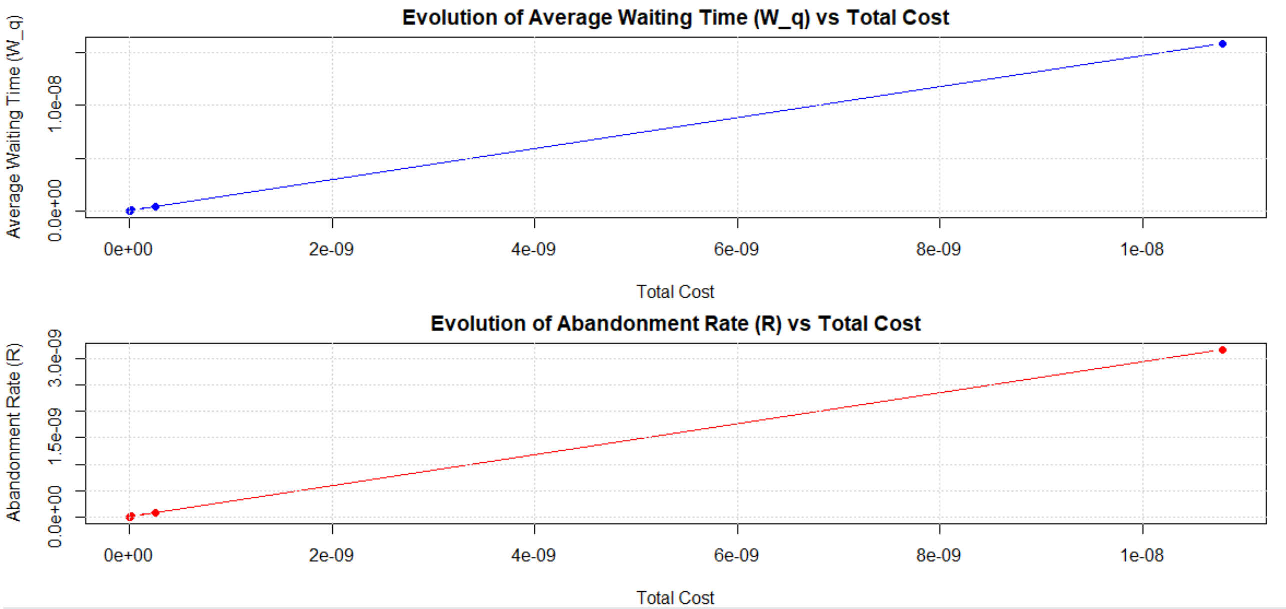


Figure 3: Evolution of waiting time and abandonment

According to the figure, the increase in total cost leads to an increase in wait time and abandonment rate. Increased wait time leads to an increased risk of customer dissatisfaction as they spend more time in the queue. The simultaneous increase in total cost, average wait time, and abandonment rate is due to a mismatch between demand and resources. This requires the right sizing to maintain acceptable service levels while keeping costs under control.

2.4.3 Some proposals

- Increase the number of servers to reduce overhead, which directly decreases the average wait time,

- Ensure that the waiting time remains below the impatience threshold by adjusting the service level or the number of servers,
- Manage demand by implementing strategies to smooth the load,
- Optimize weights, i.e. increase the weight for wait time and weight for abandonment rate so that the OSP prioritizes customer satisfaction.

3 Conclusion and Perspectives

In this paper, we studied the performance of M/M/K-type queuing networks with intermittently accessible servers, failed servers with corrective maintenance and customer dropouts. Firstly, we studied the two-server heterogeneous case, using the Geometric Matrix Method to obtain steady-state probabilities in order to deduce system performance metrics. Secondly, we have considered the case where the number of servers is strictly greater than two. As the number of servers increases, it becomes difficult to study the model using numerical methods. We therefore used an algorithmic method to evaluate the system's performance. We began by using the Particle Swarm Optimization (PSO) algorithm to obtain the best configurations of arrival rate λ, service rate μ and number of servers K, in order to minimize the total operational cost C. This cost is linked to servers, breakdowns, corrective maintenance, intermittency and customer abandonment. In a second step, we used the PSO to explore different regions of space, to avoid getting stuck in a local minimum which is not necessarily the best solution, and to increase the chances of finding a global minimum which is therefore the best solution. In other words, the PSO will make it possible to manage local minimums. This will reduce waiting times and abandonment rates, maximizing customer satisfaction. Finally, we made recommendations to help further maximize customer satisfaction.
In perspectives, we want to make an extension to the M/G/K model, i.e., by considering general service distributions (G) instead of the exponential distribution (M). We also intend to explore other potential extensions, such as priority queues or different maintenance strategies.

References

[1] **Agassi Melikov, Sevinj Aliyeva and Janos Sztrik** *Analysis of Instantaneous Feedback Queue with Heterogeneous Servers* In: Mathematics 2020, 8, 2186
doi: 10.3390/math8122186

[2] **R. Sethi, M. Jain, R.K. Meena and D. Garg** *Cost optimization and ANFIS computing of an unreliable M/M/1 queueing system with customers impatience under N-policy* Int. J. Appl. Comput. Math. 6 (2020) 1–14.

[3] **T.B. Ravaliminoarimalalason, F. Randimbindrainibe** *Distribution of Occupied Resources on A Discrete Resources Sharing in A Queueing System* In: International Journal of Engineering Research and Technology (IJERT), ISSN 2278-0181, pp 638-643, Vol. 10 Issue 1, 2021

[4] **C. Shekhar, S. Varshney and A. Kumar** *Matrix-geometric solution of multi-server queueing systems with Bernoulli scheduled modified vacation and retention of reneged customers: a meta-heuristic approach* In: Qual. Technol. Quant. Manage. 18 (2021) 39–66.

[5] **S. Thakur, A. Jain and M. Jain** *ANFIS and cost optimization for markovian queue with operational vacation* In: Int. J. Math. Eng. Manage. Sci. 6 (2021) 894.

[6] **M. Seenivasan, and K. S. Subasri** *Performance analysis of multi-server markovian queue with intermittently available and unreliable server* AIP Conference Proceedings 2364, 020036 (2021); https://doi.org/10.1063/5.0062891

[7] **M.Seenivasan, M.Indumathi and V. J. Chakravarthy** *Performance Analysis Of Two Heterogeneous Server Queueing Model with Intermittently Obtainable Server Using Matrix Geometric Method* In: Journal of Physics: Conference Series 1724 (2021) 012001 doi: 10.1088/1742-6596/1724/1/012001

[8] **T.B. Ravaliminoarimalalason, F. Randimbindrainibe,** *Occupied resources in a system with resource sharing between service-based customers* In: International Journal of Engineering Research and Technology (IJERT), ISSN 2278-0181, pp 158-164, Vol. 10 Issue 10, 2022.

[9] **M. Seenivasan, R. Senthilkumar and K.S. Subasri** *M/M/2 heterogeneous queueing system having unreliable server with catastrophes and restoration* In: Mater. Today: Proc. 51 (2022) 2332–2338.

[10] **M. Seenivasan, M. Kameswari and M. Indumathi** *Markovian Queueing Model with Intermittently Accessible Server and Restoration Using Matrix Geometric Method* IEEE (2022) 1–7.

[11] **M. Seenivasana[a,*], R. Abinaya[b]** *Markovian queueing model with single working vacation and catastrophic* Materials Today:Proceedings51(2022)2348–2354

[12] **Kothandaraman, D.; Kandaiyan, I** *Analysis of a Heterogeneous Queuing Model with Intermittently Obtainable Servers under a Hybrid Vacation Schedule* Symmetry 2023, 15, 1304. https:// doi.org/10.3390/sym15071304

[13] **S. Wang, Z. Ma, X. Niu and Y. Liu** *Performance analysis of a Queueing system based on vacation with fault repairable and spare servers in the MP2P network* In: Wireless Networks 29 (2023) 2321–2336.

[14] **N. Singh, M. Jain and S. Dhibar** *ANFIS computing for M/M/∞ queue with two types of service interruption and balking* In: Proc. Nat. Acad. Sci. India Sect. A: Phys. Sci. 94 (2024) 63–73

[15] **Darvinder Kumar[1] and Gulab Singh Bura[2]** *Transient Solution of a Two Homogeneous Servers Markovian Queueing System with Environmental, Catastrophic and Restoration Effects* International Journal of Mathematics and Statistics Studies, 12 (1), 45-53, 2024 doi: https://doi.org/10.37745/ijmss.13/vol12n14553

[16] **Anderson Ribeiro Duarte** *The server allocation problem for Markovian queueing networks* 256-271https://doi.org/10.1504/IJSOM.2024.138931

[17] **K. Divya and K. Indhira** *Performance analysis and ANFIS computing of a markovian queuing model with intermittently accessible server under a hybrid vacation policy* In: RAIRO Operations Research 58 (2024) 1257–1279
https://doi.org/10.1051/ro/2024044