Al-Nahrain Journal of Science
ANJS, Vol. 28(3), September, 2025, pp. 140 – 146

A New M/M/1 Queueing System with Bridging Gap Minimization

Farah L. Joey[1,2,*], Wah June Leong[2], C. Y. Chen[2], M. L. Othman[3]

[1] Department of Mathematics and Computer Applications, College of Sciences, Al-Nahrain University, Jadriya, Baghdad, Iraq.
[2] Department of Mathematics and Statistics, Faculty of Science, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia.
[3] Department of Electrical and Electronic Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

Article's Information

Received: 28.04.2025
Accepted: 24.06.2025
Published: 15.09.2025

Keywords:
Queueing theory,
Kendall notation,
Exponential distribution,
Bridging function,
Moment generating function,
PSO algorithm.

Abstract

In queueing systems, the Kendall notation a/b/x/q/y/z is a standard format used to describe six categories: the arrival distribution of clients, the service distribution of servers, servers' number, the queue capacity, the system capacity, and the queue discipline. In this paper, a new category, referred to as the bridging function with best alpha (Br_∝*) between the arrival and service distributions, is introduced into the Kendall notation. It is placed between the arrival_distribution and the service_distribution, replacing the queue capacity (b), so that the notation becomes (a/ Br_∝*/b/x/y/z). Based on this new notation, the capacity rate of the queueing line can be predicted and represents the gap between the client arrival rate and the client service rate. It is shown that the proposed bridging function with best alpha follows the same distribution of arrival and service, both of which are assumed to be exponentially distributed. The Particle Swarm Optimization (PSO) algorithm is employed to find the optimal rate of best alpha, using the bridging function for the arrival and service distributions as the objective function. Simulation results of M/ Br_∝*/M/1 suggest that the bridging function with best alpha (Br_∝*) can reduce the delay time in the queueing system and converge to the optimal solution when applied to a standard M/M/1 system simulated by MATLAB code.

http://doi.org/10.22401/ANJS.28.3.18
*Corresponding author: farahlateefjoey@gmail.com

1. Introduction

Many real-life models involve random processes, which are defined as stochastic processes. These random processes are essentially a collection of random variables that can be classified into two categories: continuous and discrete [1, 2, 3]. In general, a random process relies on a probability distribution to obtain all probabilities considering random variables based on a cumulative distribution function (CDF). In a discrete process, the probability related to discrete random variables is called the Probability Mass Function (PMF) [9, 17]:

$$F_X(x) = \sum_i px(x_i) \quad \dots (1)$$

On the other hand, a Probability Density Function (PDF) with continuous random variables is defined as:

$$f_X(x) = \frac{dF_X(x)}{dx} \quad \dots (2)$$

where

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\ du \quad \dots (3)$$

The moment generating function (MGF) of random variables associated with a probability distribution, is then characterized as the expectation or moment of the random variable X. So that the m^{th}moment generating function ($mMGF$) about the origin of the random variable X is referred to as:

$$\mu'_m = E(X^m) = \begin{cases} \sum_i x^m p_X(x) & \text{for discrete } R.V., \\ \int_{-\infty}^{\infty} x^m f_X(x)dx & \text{for countinuous } R.V. \end{cases} \quad \dots(4)$$

The expectation of random variables of the first moment represents the mean as follows:

$$\mu = \mu'_1 = E(X) \quad \dots(5)$$

And the expectation of the second moment is defined as the variance:

$$\sigma^2 = E(X^2) - \big(E(X)\big)^2 = E(X^2) - (\mu)^2 \quad \dots(6)$$

where $\mu'_2 = E(X^2)$.

We can then define the mth moment generating function alternatively as below.

$$M_X(t) = E(e^{tX}) = \begin{cases} \sum_i e^{tx} p_X(x) & \text{for discrete } R.V, \\ \int_{-\infty}^{\infty} e^{tx} f_X(x)dx & \text{for countinuous } R.V, \end{cases}$$

where

$$\mu'_m = \frac{d^m\, M_X(t)}{dt^m}\Big|_{t=0} \quad \dots(7)$$

Many facilities need to systematize their queues; therefore, they can be modeled using queueing systems. Banik et al. (2022) studied the effectiveness of the relationship during successive inter-batch arrival periods in the queueing models GIX/D/c and BMAP/D/c. They employed steady-state equations and the roots of a characteristic equation to derive the probability distribution of the queueing model. As a result, numerical calculations of waiting time and queue length were obtained, and the extracted roots performed well for all types of inter-arrival distributions. Furthermore, the waiting time was found to be shorter compared to the matrix-analytic method, and the calculation time using the roots increased rapidly with the number of servers [11]. In addition, Khalid et al. (2022) investigated the terminal departure queueing system at Cairo International Airport. The standard waiting time was compared to the benchmarks provided by IATA's Level of Service (LoS) concept. They utilized the standard Kendall notation to model the departure queue of passengers. It was found that during peak (rush) hours, 50% more servers were required, along with an expansion of the departure area [12]. According to Anni et al. (2021), a queueing system was implemented for managing ration shop operations using an Android application. Customers received two alerts of SMS messages, one notifying them of product availability and another instructing them to proceed to the counter for service or to reschedule. As a result, the application notified the appropriate counter, stored customer information, and facilitated online payments, thereby reducing the length of the waiting queue [13]. To construct a queueing model, the essential components needed to define are arrival rate (λ_a), service rate (μ_s), and number of servers [1, 2, 4]. Furthermore, there is a major representation of the queueing system, which is called Kendall notation, depending on the six categories characterized as (a/b/x/q/y/z) that is illustrated below in Table 1:

Table 1. Kendall's description of the queueing system

Categories	Sym	Classifications of Distributions
a: Arrival category b: Service category	Ma D E$_i$ M$_i$ Pha G	Markovian distribution (Poisson and exponential). Deterministic. Erlang $\forall\, i$. Mixture of Exponential. Phase. General.
x: Servers number	X	Positive Integer Indicates the Number of Servers.
q: queueing capacity or the maximum queueing length.	Q	Positive Integer Indicates the Queue Length.
y: System capacity	Y	Positive Integer Indicates the System Capacity.
z: Queue discipline	FCFS LCFS ROS P GD	First Come First Serve. Last Come First Serve. Random Order of Service. Priority. General Discipline.

The Kendall notation described in Table 1, has a category (L) that is the queueing capacity or the maximum queueing length, which is often fixed as a constant by default. In this paper, we propose replacing the L category with a function, referred to as the bridging function (Br), positioned between the server and arrival category. This function estimates the queueing line rate τ as a probability distribution. Accordingly, the objective of this study is to regulate the queue length based on the arrival_rate λ_a and the service_rate μ_s, both of which follow the same probability distribution. To determine the optimal queueing capacity rate, we employ an alternative approach using the Particle Swarm Optimization (PSO) algorithm. According to the model of M/M/1 queueing system, the utilization parameter $\rho = \frac{\lambda_a}{\mu_s}$ is used in the queueing system to estimate the busyness of servicing customers. If $\mu_s > \lambda_a$, then ρ will be less than 1 and it implies that the system is stable. On the contrary, if $\lambda_a > \mu_s$, then the system will be unstable. Moreover, if $\rho = 1$ is obtained, it leads to a system being busy most of the time [5].

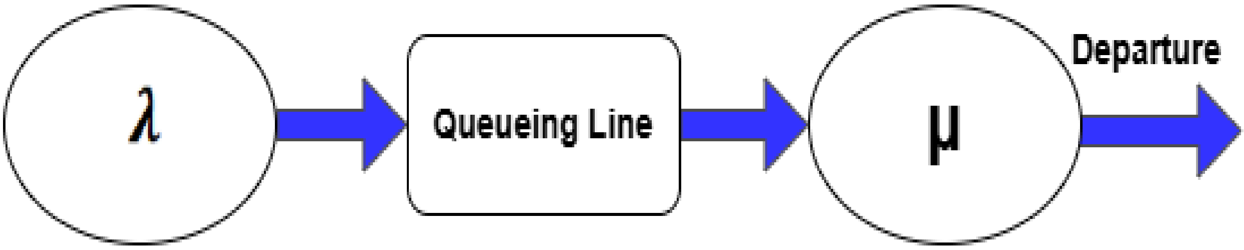


Figure 1. M/M/1 queueing model.

By assuming $\rho < 1$, the probability of servicing can be determined by the probability of the number of clients, $P_0 = 1 - \rho$ and particularly the probability of having n clients in the system is denoted by $P_n = \rho^n P_0 = \rho^n(1-\rho)$. Hence, the clients' number in the system, namely Ls can be estimated as:

$$Ls = \sum_{n=0}^{\infty} nP_n = \frac{\rho}{1-\rho} = \frac{\lambda_a}{\mu_s - \lambda_a} \quad \dots (8)$$

Subsequently, the clients' number in queue Lq:

$$Lq = Ls - \rho = \frac{\rho}{1-\rho} - \rho = \frac{{\lambda_a}^2}{\mu_s(\mu_s - \lambda_a)} \quad \dots (9)$$

Furthermore, by using Little's law $W = \frac{L}{\lambda_a}$ as a performance measurement, the average time spent in the queue is given by:

$$Wq = \frac{Lq}{\lambda_a} = \frac{\lambda_a}{\mu_s(\mu_s - \lambda_a)} \quad \dots (10)$$

And the average time spent in the system, Ws is obtained as

$$Ws = \frac{Ls}{\lambda_a} = \frac{1}{\mu_s - \lambda_a} \quad \dots (11)$$

Or

$$Ws = Wq + \frac{1}{\mu_s} = \frac{\lambda_a}{\mu_s(\mu_s - \lambda_a)} + \frac{1}{\mu_s} = \frac{1}{\mu_s - \lambda_a} \quad \dots (12)$$

The structure of this paper is as follows: The first section gives the basic definitions and reviews related to the queueing theory. In the next section, the methodology for the proposed new notation is presented. Results and their discussion are also given in Section 3, and finally, the conclusion is stated in the last section.

2. Methodology

Traditionally, Kendall notation, represented as a/b/x/q/y/z, is employed to describe a queueing model. In this study, we propose an extended notation by introducing a new category referring to the bridging function (Br_∝*), which connects the arrival (a) and service (b) distributions. The modified notation is expressed as (a/Br_∝*/b/x/y/z). The bridging function (Br_∝*) is characterized by its convexity, a key concept that plays a significant role in areas such as optimal control theory, game theory, linear programming, and convex optimization. The following section presents the formal definitions of a convex set and a convex function.

Definition 1 [15]: A set $S \subseteq R^n$ is said to be a convex set if $\forall\, a, b \in S$, we have:

$$\propto a + (1{-}\propto)b \in S, where \ \propto \in [0,1] \quad \dots(13)$$

Graphically, this means that any two points belong to S, the line segment between them belongs to S as well.

Therefore, to motivate the bridging function, F_{Br} for (a/Br_∝*/b/x/y/z) using the convexity, we can define F_{Br} as follows:

$$F_{\mathrm{Br}} = \propto b + (1{-}\propto)a, \qquad \propto \in [0,1] \quad \dots (14)$$

where F_{Br} is a convex combination of arrival distribution a and service distribution b. This combination is derived by taking into consideration the two main distributions of the queueing system and may give better insight into the gap between arrival rate λ_a and service rate μ_s. The aim is to obtain the minimum of F_{Br} with respect to λ_a and μ_s. Moreover, if the Markovian distributions of the continuous random variables x and y are assumed to follow an exponential distribution. Then, their probability density function (PDF) is defined as:

$$f_X(x) = \lambda_a e^{-\lambda_a x}, \quad x \geq 0 \quad \dots(15)$$

where the corresponding mean is $\frac{1}{\lambda_a}$ and the variance, $\sigma^2 = \frac{1}{{\lambda_a}^2}$. Therefore, F_{Br} in eq. (14) can be given as

$$F_{\mathrm{Br}} = \propto \mu_s e^{-\mu_s t} + (1 - \propto)\lambda_a e^{-\lambda_a t} \; \dots(16)$$

where the mean of F_{Br} between the server and arrival is:

$$\tau = E(t) = \int_0^\infty t\, f(t) dt$$
$$= \int_0^\infty t\left[\propto \mu_s e^{-\mu_s t} + (1 - \propto)\lambda_a e^{-\lambda_a t}\right] dt$$
$$= \frac{1}{\lambda_a} + \frac{\propto}{\mu_s} - \frac{\propto}{\lambda_a} \quad \dots(17)$$

Thus, at $\propto = 0$, the mean of F_{Br} is given by $\frac{1}{\lambda_a}$ which is also the mean of the arrival distribution, and at $\propto = 1$, the mean of F_{Br} is equal to $\frac{1}{\mu_s}$, which is exactly the mean of the server distribution.

2.1. Particle Swarm Optimization (PSO) Algorithm

Particle Swarm Optimization (PSO) is a widely used algorithm for solving complex optimization problems. Its concept comes from the collective movement of bird groups searching for optimal food sources. In this analogy, each bird in the flock is represented as a particle, and the entire group forms a swarm [7, 8]. The PSO process is considered stochastic, as it relies on random behavior observed in nature [10].

During the optimization process, each particle adjusts its position in search of the optimal solution based on its current position, velocity, and objective value. This mechanism is illustrated in Figure 2, where the position of each particle is updated based on the velocity update which incorporates either local and global best positions (or objective values), along with the particle's inertia using the equations below [16]:

$$Vel(i)(t+1) = W.Vi(t) + k1.r1.(P.best - xi(t)) + k2.r2.(G.best - xi(t)) \; \dots(18)$$
$$xi(t+1) = xi(t) + Vel(i)(t+1) \; \dots(19)$$

where

$Vel(i)$:velocity of ith particle, xi:ith particle position, W:inertia weight of velocity, $k1, k2$: the coefficients of acceleration, $r1, r2$: random numbers between (0,1), $P.best$: local best position, $G.best$: global best position. The algorithm continues to update positions and velocities until a stopping criterion is met. Ultimately, the global best objective is achieved by identifying the particle with the best objective value throughout iterations [6, 10, 18].

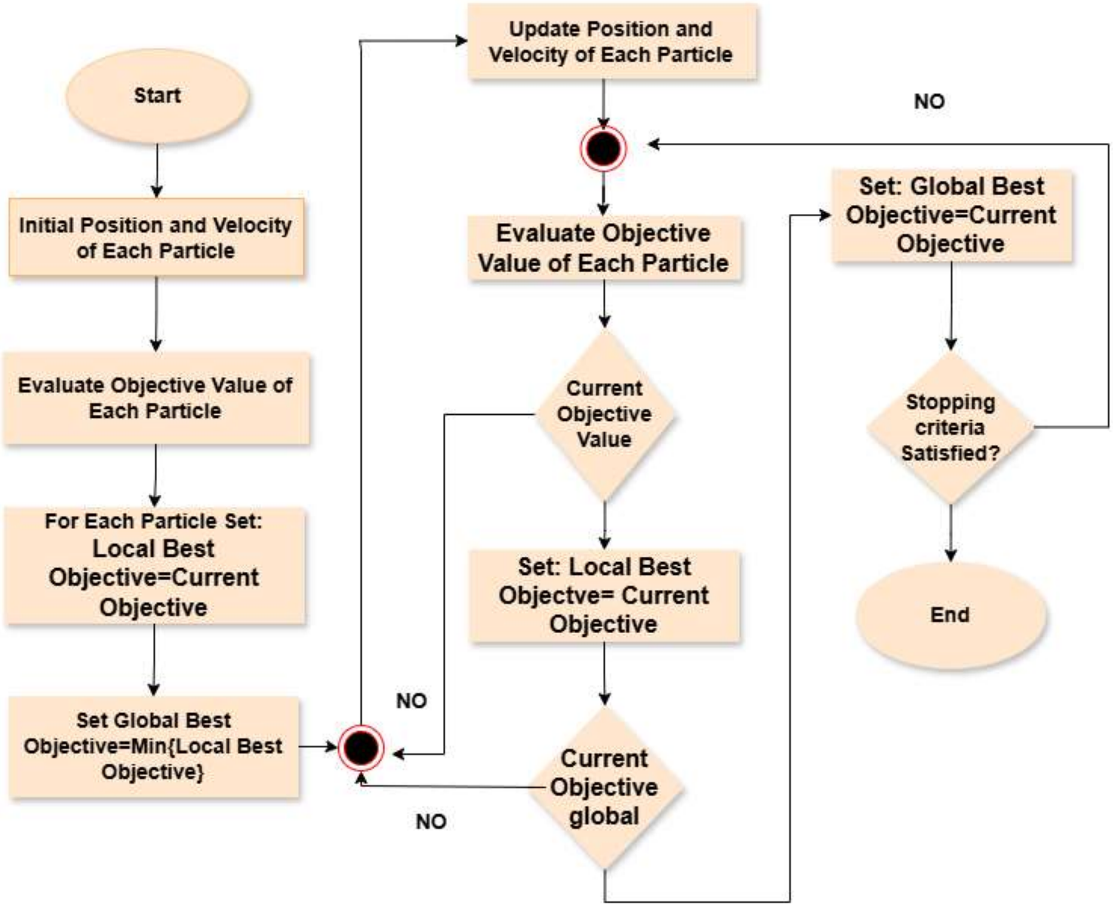


Figure 2. Flow chart of PSO algorithm.

In this study, the PSO algorithm is used to compute an optimal queue length by letting the mean of the bridging function with best alpha, namely $\frac{1}{\lambda_a}+\frac{\propto}{\mu_s}-\frac{\propto}{\lambda_a}$ as the objective function where the upper bound $ub = 100$, and lower bound $lb = 1$ are selected with population size $N = 1000$. We perform 10 trials and collect the results as follows:

$$Ave(\tau_{PSO}) = \frac{\sum_{i=1}^{m} \tau_{PSO_i}}{m} \quad \dots (20)$$

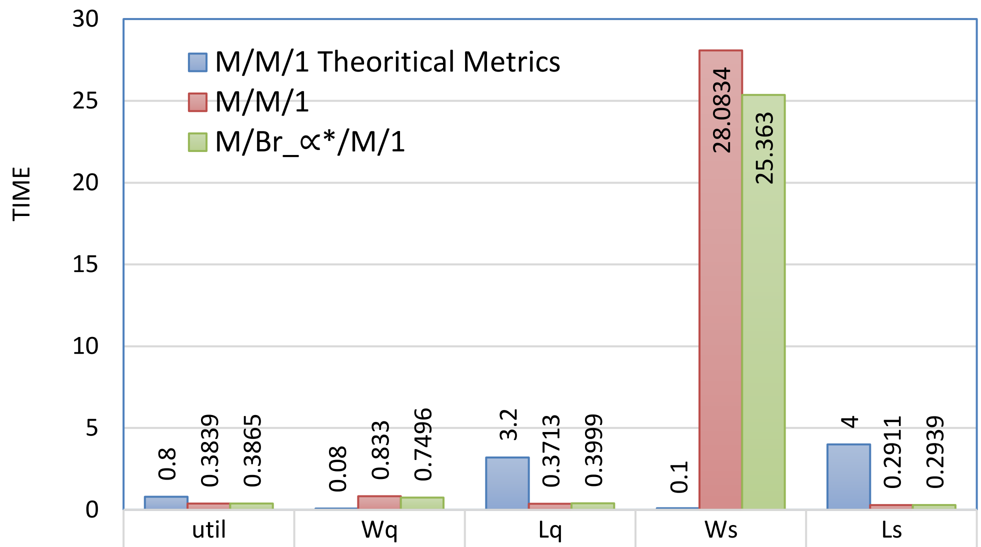
where m is the trials number.

3. Results and Discussion

Numerical experiments are conducted by assuming randomly the arrival rate $\boldsymbol{\lambda_a = 40}$, and service rate $\boldsymbol{\mu_s = 50}$. The results are given in Table 2 below.

	Table 2. Comparisons of modified queueing systems.		
	M/M/1Theoretical Metrics	M/M/1	M/ Br_∝*/M/1
$Util$	0.8	0.3839	0.3865
Wq	0.08	0.8330	0.7496
Lq	3.2	0.3713	0.3999
Ws	0.1	28.0834	25.363
Ls	4	0.2911	0.2939

Table 2 demonstrates the comparison results of the simulation performance measurements of standard M/M/1 and the new approach of M/Br_∝*/M/1 with the theoretical performance measurements of M/M/1 and shown in Figure 3:



	util	Wq	Lq	Ws	Ls
M/M/1 Theoritical Metrics	0.8	0.08	3.2	0.1	4
M/M/1	0.3839	0.833	0.3713	28.0834	0.2911
M/Br_∝*/M/1	0.3865	0.7496	0.3999	25.363	0.2939

PERFORMANCE MEASUREMENTS

Figure 3. Comparison of the standard M/M/1 simulation and the new approach of M/Br_∝*/M/1.

Table 2 and Figure 3 illustrate the notable differences in system performance of queueing simulations; it suggests that M/Br_∝*/M/1 exhibits greater efficiency than M/M/1. A reduction in measurements was observed in M/Br_∝*/M/1 compared to M/M/1, leading to a decrease in service delay time. Specifically, the new approach M/Br_∝*/M/1 leads to the decline of the waiting time in queue Wq and waiting time in system Ws. Additionally, the results of M/Br_∝*/M/1 converge to the theoretical measurements of M/M/1 which are represented as the optimal solution. Consequently, a slight increase in system utilization ρ was noted, resulting in a higher number of customers entering the system being confirmed by computational analysis in Lq and Ls, which demonstrated an increased capacity for receiving and servicing customers through the system.

4. Conclusions

As a conclusion, the higher performance of M/Br_∝*/M/1 simulation highlights the best compared to the M/M/1 simulation. It also clarifies that some factors may contribute to this best performance, which are: queueing simulation employs a more efficient queuing mechanism, reduces waiting time, and enhances throughput. So, M/Br_∝*/M/1 implies that it achieves short service time, balance in arrival rates, and has a proper operational algorithm. Moreover, the whole results of M/Br_∝*/M/1 simulation converge to the optimal results of the theoretical metric of standard M/M/1, and this could be implemented as an optimal queueing system in real-world applications.

Funding: No funds have been received for this research.

Acknowledgments: The authors would like to thank Universiti Putra Malaysia and Al-Nahrain University for the support in writing this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] Yang, X.S.; "Optimization Techniques and Applications with Examples". 1st ed.; Wiley Blackwell: Hoboken, NJ, USA, 2018.
[2] Kobayashi, H.; Turin, W.; Mark, B.L.; "Probability, Random Processes, and Statistical Analysis". 1st ed.; Cambridge University Press: New York, NY, USA, 2012.
[3] Abo-Alsabeh R., Daham H. A.; Salhi A.; "An evolutionary approach for solving the minimum volume ellipsoid estimator problem". In: Innovations in bio-inspired computing and

applications; IBICA 2020, Abraham, A.; Sasaki, H.; Rios, R.; et al. (eds); Advances in Intelligent Systems and Computing, Springer: Cham, Switzerland, pp. 23–31, 2021
[4] Narayan, B.U.; "An Introduction to Queueing Theory: Modeling and Analysis in Applications". 1st ed.; Birkhäuser: Boston, MA, USA, 2015.
[5] Smith, J.M.; "Introduction to Queueing Networks: Theory Practice". 1st ed.; Springer: Berlin, Germany, 2019.
[6] Salam, N.A.; Farah, F.L.; Abdalmaaen, H.F.; "Improve Indoor Position Localization Based on FFNN with PSO Optimization Algorithm". Int. J. Adv. Sci. Technol, 29 (4): 7246–7258, 2020.
[7] Stewart, W. J.; "Probability, Markov Chains, Queues, and Simulation: The Mathematical Basis of Performance Modeling". 1st ed.; World Publishing Corporation: Beijing, China, 2009.
Youming, Z.; Xingchen, H.; "Application of Video Image Processing in Sports Action Recognition Based on Particle Swarm Optimization Algorithm". Prev. Med, 173: 107592, 2023. https://doi.org/10.1016/j.ypmed.2023.107592
[8] Su, B.; Lin, W; Wang, J.; Rui, C.; "Sewage Treatment System for Improving Energy Efficiency Based on Particle Swarm Optimization Algorithm". Energy Rep.,8: 8701–8708, 2022.
[9] Solano-Rojas, B.J.; Villalón-Fonseca, R.; Batres, R.; "Micro Evolutionary Particle Swarm Optimization (MEPSO): A New Modified Metaheuristic". Syst. Soft Comput., 5: 200057, 2023.
https://doi:10.1016/j.sasc.2023.200057
[11] Banik, A.D.; Chaudhry, M.L.; Wittevrongel, S.; Bruneel, H.; "A Simple and Efficient Computing Procedure of the Stationary System-Length Distributions for GIX/D/c and D/c/1 Queues". Comput. Oper. Res., 138: 105564, 2022. https://doi.org/10.1016/j.cor.2021.105564
[12] Khalid, A.; Awad, K.; Mohamed, A.; Wagdy, A.; "Airport Terminal Building Capacity Evaluation Using Queuing System". Alexandria Eng. J., 61 (12): 10109–10118, 2022. https://doi.org/10.1016/j.aej.2022.03.055
[13] Sheebha, A. D.J.; V P, N.; Abdulla, N.; Nihara, N.; Amiyan, K. S.S.; "An Android App: Virtual Queuing System for Public Distribution System". In: Proceedings of the 2021 IEEE International Conference on Mobile Networks and Wireless Communications (ICMNWC), Tumkur, India, December 2021, Editors: Rajashree V. Biradar, Vijayakumar P., Sanjeev Kunte. IEEE: Piscataway, NJ, USA, 2021.
[14] Singh, S.K.; Acharya, S.K.; Cruz, F.R.B.; Da Costa Quinino, R.; "Bayesian Sample Size Determination in a Single-Server Deterministic Queueing System". Math. Comput. Simul., 187: 17–29, 2021. https://doi.org/10.1016/j.matcom.2021.02.010
[15] Boyd, S.; Vandenberghe, L.; "Convex Optimization". 1st ed.; Cambridge University Press: New York, NY, USA, 2013.
[16] Tiza, V.; Arun N.K.; Gopakumar P.; "Congestion management in power grids using multi-agent systems and particle swarm optimization". Franklin Open, 12: 100303,2025. https://doi.org/10.1016/j.fraope.2025.100303
[17] Shaymaa R. Th.; "Comparison Between the Simulated Prediction Methods of the Markov and Mixed Models". Al-Nahrain J.Sci., 27(4): 12-20, 2024. https://doi.org/10.22401/fw29aw06
[18] Ban, A. M.; Suhaib, A.; "Pattern Recognition Using Particle Swarm Optimization with Proposed a New Conjugate Gradient Parameter in Unconstrained Optimization". Al-Nahrain J. Sci., , 19(3): 138-147, 2016. https://doi.org/: 10.22401/JNUS.19.3.19.