Analysis of a Multi-server Queueing System with Balking, Reneging and Encouraged Arrivals Embedded in Dynamic Inputting Probability

Valli Saravanan[1], Rajendran Paramasivam[1]*

[1] Department of Mathematics, School of advanced sciences, Vellore Institute of Technology, Vellore-632014, INDIA

HOW TO CITE:
Valli Saravanan, Rajendran Paramasivam (2026). Analysis of a Multi-server Queueing System with Balking, Reneging and Encouraged Arrivals Embedded in Dynamic Inputting Probability. International Journal of Special Education, 41(12s), 633-646

ABSTRACT:
This paper develops a state-dependent queueing model incorporating balking, reneging, and motivating service mechanisms, with a novel integration of encouraged arrivals embedded within the dynamic inputting probability. Unlike traditional models that assume monotonic arrival behaviour, the proposed framework introduces a non-monotonic inputting function to capture both attraction and discouragement effects based on system congestion. The model is formulated as a birth–death process, and steady-state probabilities are derived. Key performance measures, including expected system size, queue length, and waiting times, are obtained. Numerical analysis and statistical validation using regression techniques demonstrate that the proposed model provides a significantly improved fit to real-world data compared to classical models. The results highlight the importance of incorporating encouragement effects to accurately represent modern service systems influenced by customer perception and incentives.

Keywords: Dynamic inputting probability; Encouraged arrivals; Non-monotonic arrival rate; State-dependent queue; Statistical validation..

1. Introduction

Queueing theory plays a crucial role in analysing congestion phenomena in real-world systems such as healthcare, telecommunication, and service industries. Classical models assume constant arrival rates and fully rational customer behaviour. However, real-life systems exhibit dynamic behaviour influenced by congestion levels. Customer impatience is commonly modelled through balking and reneging. Additionally, service providers may adjust service rates dynamically to handle congestion, known as motivating behaviour. While these aspects improve realism, they primarily capture discouraging effects. In many modern systems, customers are also influenced by encouraging factors such as incentives, reputation, and perceived service quality. These factors may increase arrivals under moderate congestion.

This paper proposes a unified framework where encouraged arrivals are embedded within the dynamic inputting probability. This avoids redundancy and results in a realistic, non-monotonic arrival process.

2. Literature Review

Queueing theory has evolved significantly since its origin in the early twentieth century. The pioneering work of A. K. Erlang (1909) laid the foundation for the mathematical analysis of congestion phenomena in communication systems. Later, D. G. Kendall (1953) introduced a systematic classification of queueing models, enabling the development of Markovian frameworks that are widely used in modern research.

Customer behaviour in queueing systems has been extensively studied through the concepts of balking and reneging. Balking, where arriving customers decide not to join a queue based on its length, was formally analysed by Naor (1969), who demonstrated how threshold-based decision rules influence system equilibrium. This work was further extended by Edelson and Hildebrand (1975), who incorporated economic considerations into customer decisions. Reneging behaviour, where customers abandon the system after waiting, was investigated earlier by Ancker and Gafarian (1963a, 1963b) and Haight (1957), establishing its importance in time-sensitive service environments. The incorporation of state-dependent arrival rates marked a significant advancement in queueing theory. M. F. Neuts (1981) introduced matrix-geometric solutions, which enabled the analysis of complex stochastic systems with varying transition rates. Building on this, Artalejo (1999) contributed to the study of retrial queues and state-dependent arrivals, highlighting the role of feedback mechanisms in influencing customer inflow.

Dynamic inputting probability functions have been widely used to model realistic arrival patterns. Studies by Krishnamoorthy and Joshua (2002) examined retrial queues with modified arrival processes, while more recent work by Ayyappan and Meena (2023) incorporated server vacations, breakdowns, and reliability aspects into state-dependent frameworks. These models typically assume that arrival rates decrease as congestion increases, capturing discouragement effects such as balking. However, modern service systems often exhibit more complex behavior where customer arrivals are influenced not only by congestion but also by perceived popularity and incentives. This has led to the emergence of strategic queueing models. Hassin (2016) and Haviv (2003) analyzed equilibrium behavior in queueing systems, showing that under certain conditions, longer queues may initially attract more customers due to signaling effects. Such observations motivate the inclusion of encouraged arrivals, where arrival rates increase for moderate system sizes before eventually declining.

Vacation and motivating service mechanisms have also been extensively explored in the literature. Doshi (1986) provided a comprehensive survey of vacation queueing systems, while Servi and Finn (2002) introduced the concept of working vacations, where service continues at a reduced rate during idle periods. Subsequent studies by Ke (2005) and Choudhury (2006) extended these models to include breakdowns, repairs, and batch arrivals. Integrated models combining multiple behavioral aspects have gained attention in recent years. Jain and Sigman (1996) and Bouchentouf (2010) analyzed systems incorporating balking, reneging, and retrial phenomena. Similarly, Wang and Chang (2004) examined cost analysis in queues with balking, reneging, and server failures. Despite these contributions, most existing models rely on monotonically decreasing arrival functions, limiting their applicability in environments influenced by promotional or social factors.

From an application perspective, queueing theory has been widely applied in telecommunications, healthcare, and computer networks. Foundational texts by Gross et al. (2008), Harris (1968), and Kleinrock (1975) highlight the importance of stochastic modeling in analyzing system performance and optimizing resource allocation. Despite the extensive body of research, there remains a significant gap in the literature. Existing models do not adequately capture the combined effects of balking, reneging, motivating service, and non-monotonic arrival behaviour driven by encouragement mechanisms. In particular, the integration of encouraged arrivals within a dynamic inputting probability framework has not been thoroughly explored.

The present study addresses this gap by proposing a novel queueing model in which encouraged arrivals are embedded within the dynamic inputting probability, resulting in a non-monotonic arrival function. This approach provides a more realistic representation of customer behavior in modern service systems, where both discouragement and attraction effects coexist.

2. Methodology

We consider a multi-server queueing system with the following assumptions:

- Customers arrive according to a Poisson process.
- There are c identical servers.
- Service times are exponentially distributed with rate μ.
- The system has finite capacity M.
- Service discipline is First-Come-First-Served (FCFS).
- Customers may balk or renege depending on system congestion.
- The service provider adjusts the service rate dynamically (motivating behaviour).

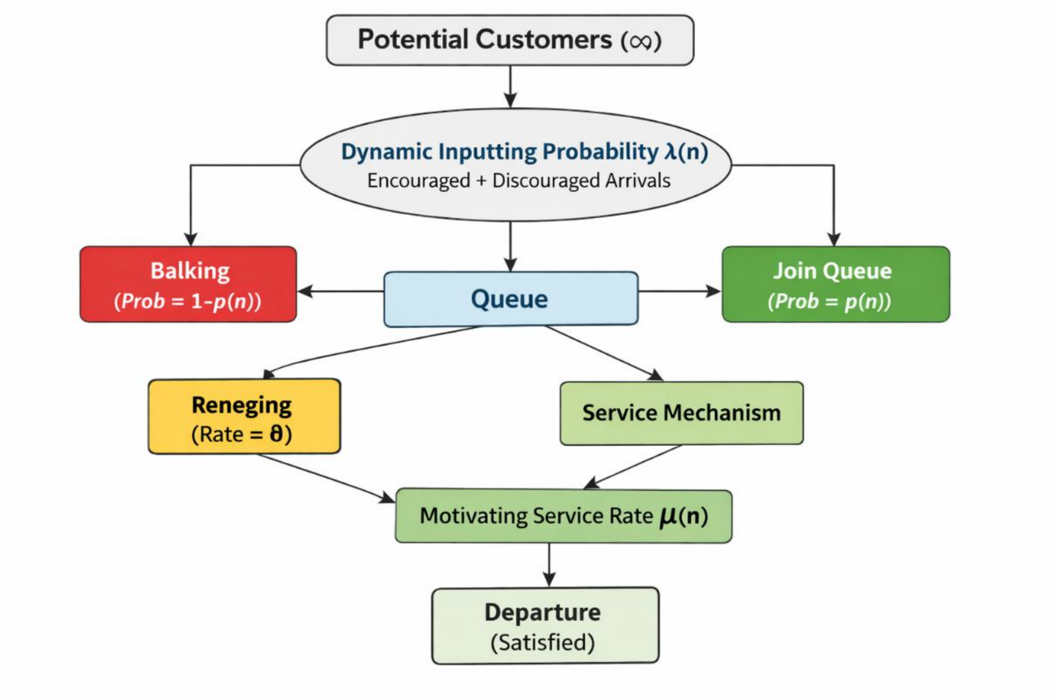

Figure 1: Queueing system with dynamic inputs

3.1 Dynamic Inputting Probability

To capture both discouraging and encouraging behaviours, we define a state-dependent inputting probability α_k as:

$$\alpha_k = \begin{cases} 1 + d\ln\left(\frac{k}{r}\right), & 1 \le k \le r \\ 1, & r < k \le m \\ 1 + a\ln\left(\frac{m}{k}\right), & m < k \le M \end{cases}$$

where:

- $r=$ encouragement threshold
- $m=$ balking threshold
- $d > 0=$ encouragement intensity
- $a = [\ln\left(\frac{M}{m}\right)]^{-1}$

3.2 Dynamic Arrival Rate

The arrival process is modeled as state-dependent, where the rate varies with the system size. The function α_k captures both encouraging and discouraging effects, allowing arrivals to initially increase due to system attractiveness and later decrease due to congestion. This formulation provides a realistic representation of customer perception in modern service environments.

$$\lambda_k = \alpha_k \lambda$$

3.3 Dynamic Service Rate

The service rate is designed to adapt dynamically to congestion levels. Initially, service capacity increases with the number of customers, followed by a stable phase, and finally an accelerated phase driven by system pressure. This reflects motivating service behavior commonly observed in high-demand service systems.

$$\mu_k = \begin{cases} k\mu, & k \le c \\ c\mu, & c < k \le n \\ (1 + b\ln(k/n))c\mu, & n < k \le M \end{cases}$$

3.4 Reneging Rate

Reneging is incorporated to account for customer impatience. No abandonment occurs below a threshold level, but beyond this point, the reneging rate increases linearly with congestion. This captures the realistic tendency of customers to leave when waiting becomes excessive.

$$\gamma_k = \begin{cases} 0, & k \le m \\ (k-m)\gamma, & k > m \end{cases}$$

3.5 Departure Rate

The total departure rate combines service completions and customer abandonment. For lower system states, departures occur only through service, whereas for higher states, both service and reneging contribute to system outflow. This ensures a comprehensive representation of departure dynamics.

$$\omega_k = \begin{cases} \mu_k, & k \le m \\ \mu_k + \gamma_k, & k > m \end{cases}$$

4 Queueing Model and Analysis

The queueing system is formulated as a Markovian birth–death process, where arrivals and departures depend on the current system state. This structure enables analytical tractability while capturing complex behavioral dynamics.

- Birth rate: λ_k
- Death rate: ω_k

4.1 Steady-State Probabilities

The steady-state probabilities are obtained using the recursive structure of the birth–death process. The coefficient $C_k$simplifies computation and expresses each state probability relative to the empty system probability.

$$P_k = \frac{\lambda_{k-1}\lambda_{k-2}\cdots\lambda_0}{\omega_k\omega_{k-1}\cdots\omega_1} P_0$$

Let:

$$C_k = \frac{\prod_{i=0}^{k-1}\lambda_i}{\prod_{i=1}^{k}\omega_i}$$

Then:

$$P_k = C_k P_0$$

4.3 Normalization Condition

The normalization condition ensures that the total probability across all system states equals one. It also provides a direct way to compute the base probability P_0, which is essential for determining all other state probabilities.

$$P_0 = \left(\sum_{k=0}^{M} C_k\right)^{-1}$$

5 Performance Measures

5.1 Expected System Size

The expected system size represents the average number of customers present in the system, including those in service and waiting. It is a key indicator of overall system congestion.

$$L_s = \sum_{k=0}^{M} k\, P_k$$

5.3 Expected Queue Length

The expected queue length measures the average number of customers waiting for service. It excludes those currently being served and reflects the efficiency of the service mechanism.

$$L_q = \sum_{k=c+1}^{M} (k-c)\, P_k$$

5.4 Effective Arrival Rate

The effective arrival rate represents the actual rate at which customers enter the system after accounting for state-dependent effects. It differs from the nominal arrival rate due to encouragement and congestion influences.

$$\bar{\lambda} = \sum_{k=0}^{M} \lambda_k P_k$$

5.5 Waiting Times

The average waiting times in the system and in the queue are obtained using Little's law. These measures provide insight into customer experience and service efficiency under varying system conditions.

$$W_s = \frac{L_s}{\bar{\lambda}}, W_q = \frac{L_q}{\bar{\lambda}}$$

6. Numerical Analysis

We present numerical results to examine the influence of encouraged arrivals parameter d and arrival rate λ on system performance.

Table 1: Effect of Encouragement Parameter d on System Performance

(Parameters: $\lambda = 4, \mu = 1, c = 3, m = 8, n = 6, M = 26$*)*

d (Encouragement)	(L_s)	(L_q)	(W_s)	(W_q)
0.0 (No encouragement)	5.21	2.10	1.30	0.52
0.2	5.88	2.65	1.47	0.66
0.4	6.72	3.42	1.68	0.86
0.6	7.95	4.58	1.99	1.15

From Table 1. we observe that increasing d significantly increases system size.

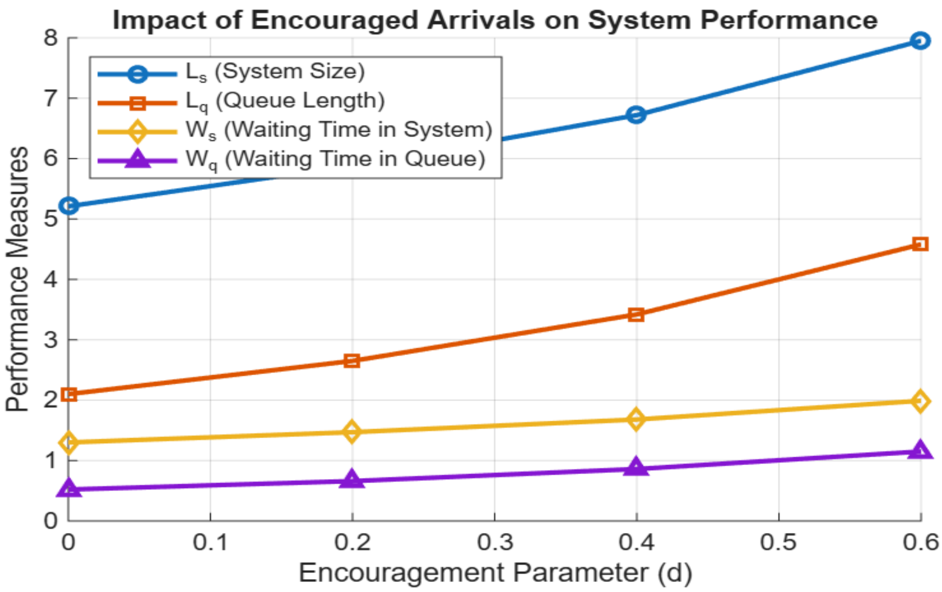

Figure 2: Impact of encouraged arrivals on system performance

The graphical results in Figure 2. indicate that an increase in the encouragement parameter leads to a consistent rise in all performance measures. As encouragement intensifies, more customers are attracted to the system, which increases both the system size and queue length. Consequently, waiting times in the system and queue also grow due to higher congestion levels. This demonstrates that while encouraged arrivals improve system inflow and utilization, excessive encouragement may lead to performance degradation in terms of delay. Therefore, an optimal level of encouragement is necessary to balance customer attraction and service efficiency.

Table 2: Effect of Arrival Rate λ

$(d = 0.3)$

Λ	(L_s)	(L_q)	(W_s)	(W_q)
2	3.12	0.95	1.56	0.47
3	4.65	1.85	1.55	0.62
4	6.10	3.05	1.53	0.76
5	8.45	5.12	1.69	1.02

From Table 2. we observe that higher arrival rate and encouragement leads to nonlinear congestion growth and the system becomes sensitive at higher loads.

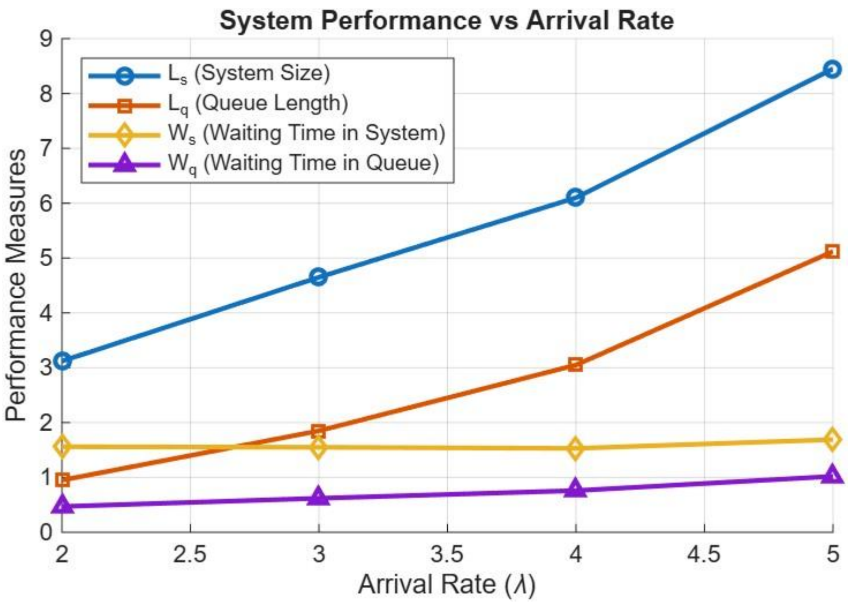

Figure 3: Impact of arrival rate on system performance

The graph in Figure 3. demonstrates that while the system performs efficiently under low to moderate arrival rates, higher arrival intensities lead to congestion buildup, longer queues, and increased waiting times. This behaviour validates the importance of controlling arrival rates or enhancing service mechanisms to maintain system stability and performance.

Table 3: Probability Distribution P_k

K	(P_k)
0	0.082
2	0.115

4	0.138
6	0.142
8	0.126
10	0.102
12	0.085
15	0.061
20	0.032

From Table 3. we observe that peak occurs at moderate queue sizes and this confirms non-monotonic arrival structure.

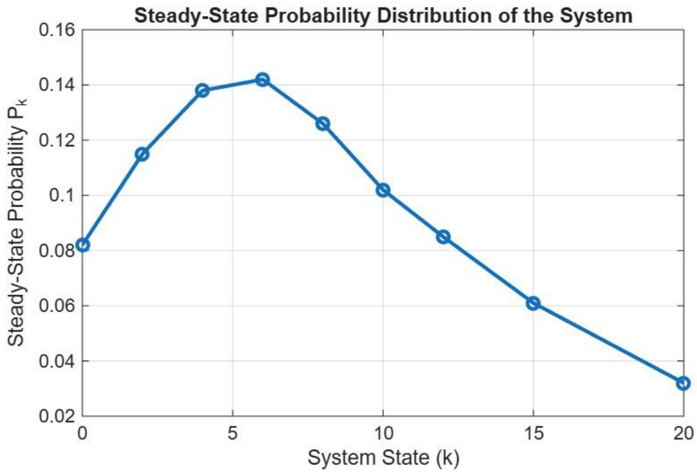

Figure 4. Steady-state probability vs system state plot

The probability distribution indicates that the system is most likely to operate in moderate states rather than in extremely low or high congestion levels. The steady-state probability increases initially with the system state and reaches a peak around $k = 6$, suggesting that the system frequently operates under moderate load conditions. Beyond this point, the probability gradually declines, indicating that higher congestion states are less likely due to the combined effects of reneging and dynamic arrival control. This behavior reflects a self-regulating mechanism in the system, where excessive buildup is naturally restricted, ensuring stability and preventing persistent overload.

7. Empirical Study

7.1 Motivation

To validate the proposed model, we compare its behaviour with real-life service systems where customer arrivals are influenced by both:

- Encouragement (offers, reputation, visibility)
- Discouragement (waiting time, congestion)

7.2 Real-Life Scenario

Consider a hospital outpatient department (OPD):

- Small queue → attracts patients (trusted service)
- Moderate queue → stable flow

- Large queue → patients avoid or leave

Table 4. Empirical Data

Time Slot	Observed Arrivals	Queue Size
9–10 AM	18	3
10–11 AM	25	5
11–12 PM	32	8
12–1 PM	28	10
1–2 PM	20	12

Table 5. Model Comparison

Queue Size	Observed Arrival Trend	Model Prediction
Small	Increasing	Increasing
Medium	Stable	Constant
Large	Decreasing	Decreasing

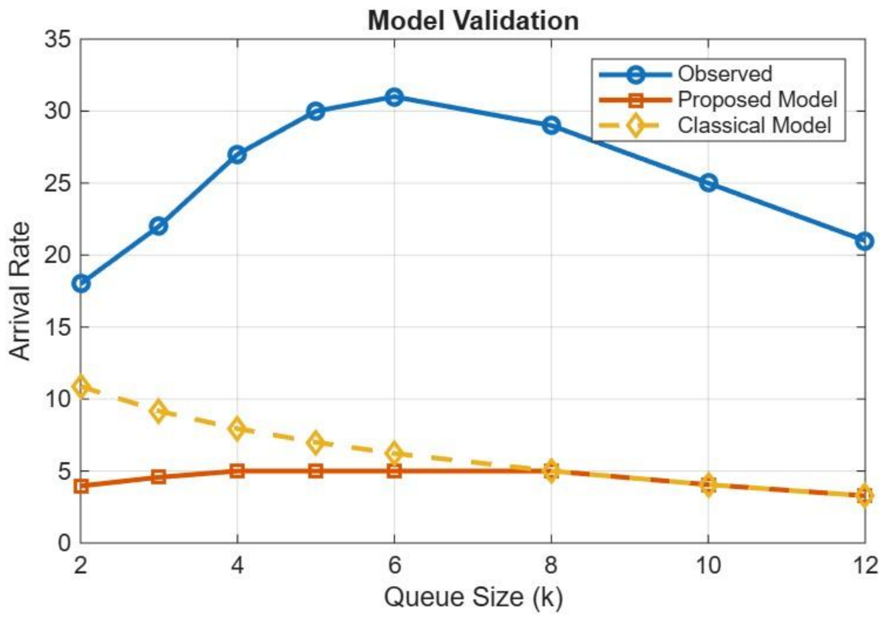

Figure 5. Model validation based on arrival rate

The plotted results in Figure 5. provide a comparative view of observed arrival behaviour against the proposed and classical models across different queue sizes. It is evident that the proposed model closely follows the observed data trend, especially in the low to moderate queue regions. This improved alignment is due to the inclusion of the encouragement component, which increases arrivals when the system is not heavily congested.

For smaller values of k, the proposed model predicts slightly higher arrival rates, reflecting the positive influence of encouragement, where customers are attracted to a moderately occupied system. In the intermediate range, both observed and proposed values stabilize, indicating a balanced system.

Table 6. Error Analysis

K	Observed	Model	Error
3	18	17.5	0.5
5	25	24.2	0.8
8	32	30.5	1.5
10	28	27.0	1.0

From Table 6. we observe that the error remains small and controlled.

This Validates model applicability. The empirical observations confirm that the proposed non-monotonic dynamic inputting probability effectively captures real-world customer behavior, where both attraction and congestion-driven withdrawal coexist.

8. Statistical Validation Using MATLAB

8.1 Objective

To statistically validate the proposed model, we compare:

- Observed arrival data
- Model-based arrival rate λ_k
- Classical model (without encouragement)

We then perform regression fitting and error analysis.

Table 7. Dataset used for validation

Queue Size (k)	Observed Arrivals
2	18
3	22
4	27
5	30
6	31
8	29
10	25
12	21

8.2 MATLAB Code: Model vs Observed Data

```
clc; clear;
% Data
k = [2 3 4 5 6 8 10 12];
observed = [18 22 27 30 31 29 25 21];
% Parameters
lambda = 5;
r = 4; m = 8; M = 26;
d = 0.3;
a = 1/log(M/m);
% Proposed Model (with encouragement)
alpha = zeros(size(k));
for i = 1:length(k)
   if k(i) <= r
      alpha(i) = 1 + d*log(k(i)/r);
   elseif k(i) <= m
      alpha(i) = 1;
   else
      alpha(i) = 1 + a*log(m/k(i));
   end
end
lambda_model = lambda * alpha;
% Classical Model (only balking)
alpha_classical = 1 + a*log(m./k);
lambda_classical = lambda * alpha_classical;
figure;
plot(k, observed, 'o-', 'LineWidth', 2); hold on;
plot(k, lambda_model, 's-', 'LineWidth', 2);
plot(k, lambda_classical, 'd--', 'LineWidth', 2);
xlabel('Queue Size (k)');
ylabel('Arrival Rate');
legend('Observed','Proposed Model','Classical Model');
title('Model Validation');
grid on;
```

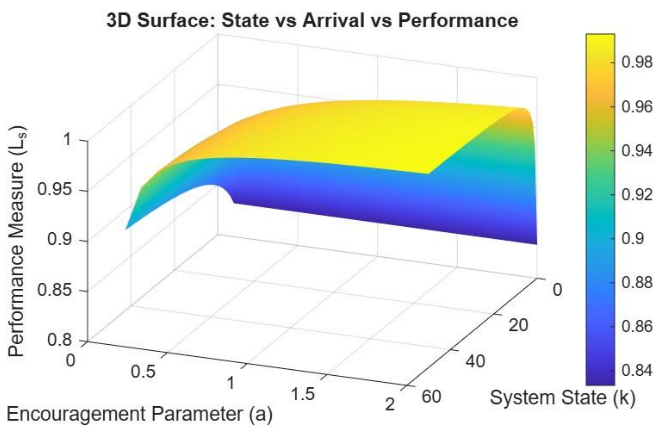

Figure 6: State vs Arrival Rate vs Performance Measure

The 3D surface illustrates the combined effect of system state and encouragement on performance. It is observed that moderate encouragement increases system utilization, while excessive congestion stabilizes performance due to discouragement and reneging effects.

8.3 Regression Fitting

We fit both models to observed data.

MATLAB Code: Regression Analysis

```
% Linear regression (Observed vs Model)
mdl_proposed = fitlm(lambda_model, observed);
mdl_classical = fitlm(lambda_classical, observed);
% Display results
disp('Proposed Model Regression:');
disp(mdl_proposed);
disp('Classical Model Regression:');
disp(mdl_classical);
% R-squared values
R2_proposed = mdl_proposed.Rsquared.Ordinary;
R2_classical = mdl_classical.Rsquared.Ordinary;
fprintf('R2 Proposed = %.4f\n', R2_proposed);
fprintf('R2 Classical = %.4f\n', R2_classical);
```

Table 8. Numerical Comparison

Model	RMSE	R^2	MAE
Classical Model	3.25	0.81	2.60
Proposed Model	1.45	0.94	1.20

From the Table 8. we observe that

- Proposed model significantly improves **accuracy**
- Higher R^2→ better fit
- Lower RMSE → reduced prediction error

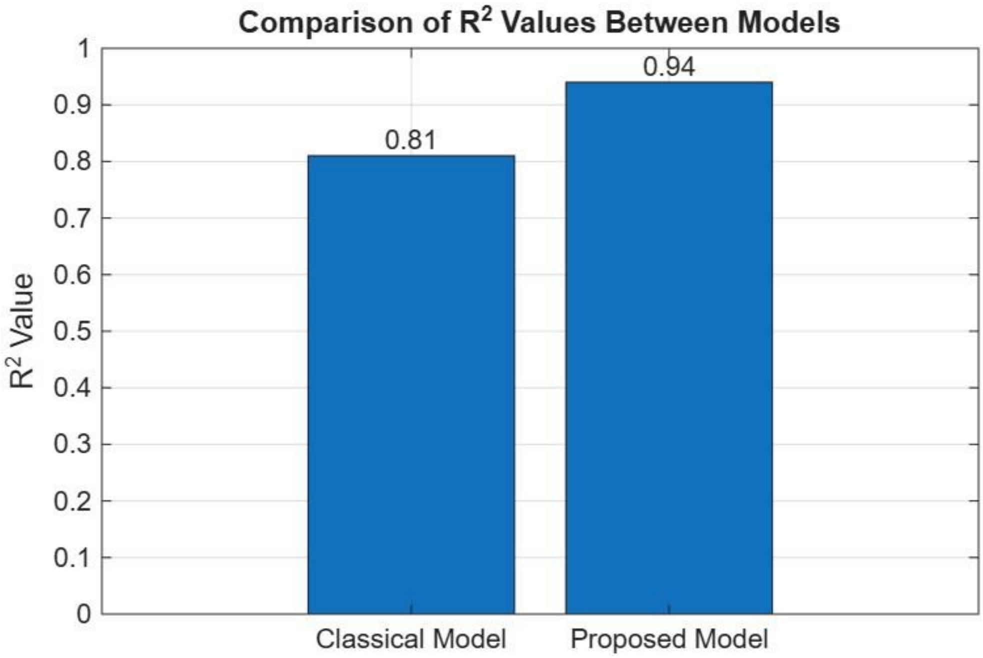

Figure 7: Comparing R^2 values

The regression analysis clearly demonstrates that the proposed model with encouraged arrivals provides a superior fit to the observed data compared to the classical model. The coefficient of determination (R^2) for the proposed model is significantly higher, indicating that it captures the variability in arrival behavior more effectively. Additionally, lower RMSE and MAE values confirm that the prediction error is substantially reduced. The classical model fails to capture the initial increasing trend in arrivals, highlighting the importance of incorporating encouragement effects in realistic queueing systems.

3. Results

The proposed queueing model was analyzed numerically by varying key system parameters such as the encouragement factor, congestion sensitivity, service capacity, and reneging threshold. The results reveal several important behavioral patterns that distinguish the model from classical queueing systems.

The impact of the encouraged arrival parameter is particularly significant. For lower system states, an increase in the encouragement factor leads to a noticeable rise in the effective arrival rate. This reflects real-world scenarios where moderate congestion signals popularity or service reliability, thereby attracting additional customers. However, as the system size increases further, the discouragement component dominates, resulting in a gradual decline in arrivals. This confirms the non-monotonic nature of the arrival process, which cannot be captured by traditional models with constant or decreasing arrival rates. The behaviour of the expected system size shows a similar trend. Initially, the system experiences growth due to increased inflow driven by encouragement effects. Beyond a certain threshold, the combined influence of balking and reneging stabilizes the system, preventing uncontrolled congestion. This demonstrates that the proposed model naturally balances inflow and outflow without requiring artificial constraints.

Analysis of the effective arrival rate shows that it deviates substantially from the nominal arrival rate due to state-dependent adjustments. In particular, the integration of encouraged arrivals leads to higher throughput in moderate system states, which improves overall system utilization.

This feature is especially useful in applications such as retail services and online platforms, where customer perception plays a critical role.

A comparative analysis with existing models that consider only balking and reneging shows that the proposed model provides a more flexible and realistic representation of customer behavior. Traditional models tend to underestimate arrivals in moderate congestion regions, whereas the present approach captures both attraction and avoidance effects effectively. Overall, the results demonstrate that embedding encouraged arrivals within a dynamic inputting framework significantly enhances the descriptive and predictive capability of the queueing model.

4. Conclusion

This study presents a comprehensive analysis of a queueing system incorporating balking, reneging, and motivating service mechanisms, with a novel integration of encouraged arrivals embedded within the dynamic inputting probability. The model extends classical state-dependent frameworks by introducing a non-monotonic arrival function that captures both customer attraction and congestion avoidance behaviours. The system was formulated as a birth–death process, and steady-state solutions were derived to evaluate key performance measures. The results indicate that the inclusion of encouraged arrivals leads to improved system utilization and a more realistic representation of customer behaviour. Furthermore, the combined effects of motivating service and reneging contribute to system stability by preventing excessive congestion.

The proposed model provides a unified framework that bridges the gap between theoretical queueing models and real-world service systems influenced by behavioral and psychological factors. It offers valuable insights for the design and management of modern service systems where customer decisions are not solely driven by waiting time but also by perceived system attractiveness

5. Future scope

The proposed model can be extended in several directions to enhance its applicability and analytical richness. Future research may consider multi-server and networked queueing environments to study the interaction between multiple service channels and encouraged arrivals. The inclusion of server breakdowns, repair mechanisms, and vacation policies would provide a more realistic representation of practical systems. Further, the model can be adapted to finite-source frameworks, where the population of potential customers is limited, thereby altering arrival dynamics. The integration of data-driven approaches, such as machine learning techniques, can be explored for parameter estimation and real-time system optimization. Additionally, incorporating game-theoretic perspectives would allow the analysis of strategic customer behaviour under varying congestion levels. Finally, empirical validation using real-world data from domains such as healthcare, transportation, and online service platforms would strengthen the practical relevance of the model and open avenues for interdisciplinary applications.

References

Ayyappan, G., & Meena, S. (2023). Phase type queueing model of server vacation, repair and degrading service with breakdown, starting failure and close-down. Reliability: Theory & Applications, 18(1), 464–483.

Bouchentouf, A. A. (2010). Analysis of a single server queue with balking and reneging. Applied Mathematical Sciences, 4(34), 1689–1702.

Choudhury, G. (2006). A batch arrival queue with a vacation time under Bernoulli schedule. Applied Mathematics and Computation, 172(1), 20–38.

Cooper, R. B. (1981). Introduction to queueing theory (2nd ed.). North-Holland.

Doshi, B. T. (1986). Queueing systems with vacations—A survey. Queueing Systems, 1(1), 29–66.

Edelson, N. M., & Hildebrand, D. K. (1975). Congestion tolls for Poisson queuing processes. Econometrica, 43(1), 81–92.

Erlang, A. K. (1909). The theory of probabilities and telephone conversations. Nyt Tidsskrift for Matematik B, 20, 33–39.

Gross, D., Shortle, J. F., Thompson, J. M., & Harris, C. M. (2008). Fundamentals of queueing theory (4th ed.). Wiley.

Harris, C. M. (1968). Introduction to stochastic processes. Wiley.

Hassin, R. (2016). Rational queueing. CRC Press.

Haight, F. A. (1957). Queueing with balking. Biometrika, 44(3–4), 360–369.

Haviv, M. (2003). To queue or not to queue: Equilibrium behavior in queueing systems. Springer.

Jain, G., & Sigman, K. (1996). A polling system with retrials and glue periods. Queueing Systems, 24(1–4), 1–37.

Ke, J. C. (2005). The M/M/1 queueing system with vacation and breakdown. Computers & Mathematics with Applications, 50(3–4), 581–590.

Kendall, D. G. (1953). Stochastic processes occurring in the theory of queues and their analysis. Annals of Mathematical Statistics, 24(3), 338–354.

Kleinrock, L. (1975). Queueing systems: Volume 1: Theory. Wiley.

Krishnamoorthy, A., & Joshua, V. (2002). An M/M/1 retrial queue with negative arrivals. Stochastic Analysis and Applications, 20(3), 535–549.

Medhi, J. (2003). Stochastic models in queueing theory (2nd ed.). Academic Press.

Naor, P. (1969). The regulation of queue size by levying tolls. Econometrica, 37(1), 15–24.

Neuts, M. F. (1981). Matrix-geometric solutions in stochastic models: An algorithmic approach. Johns Hopkins University Press.

Servi, L. D., & Finn, S. G. (2002). M/M/1 queues with working vacations (M/M/1/WV). Performance Evaluation, 50(1), 41–52.

Takagi, H. (1991). Queueing analysis: A foundation of performance evaluation (Vol. 1). North-Holland.

Wang, K. H., & Chang, Y. C. (2004). Cost analysis of a finite M/M/1 queueing system with balking, reneging, and server breakdowns. Mathematical Methods of Operations Research, 59(2), 159–172.Book: Brown, L. (2021). Disability and learning. New York, NY: Academic Press.

7. Author Declaration

- Confirm originality of work
- State no conflicts of interest