Innovating Across Boundaries: Bridging Science, Technology, and Humanities
September 2024
ISBN Number: 978-93-95305-66-2

Exploring Single-Server Retrial Queueing Models: A Comprehensive Review of Batch Arrivals and Vacation Dynamics

Vinita Yadav
Ph.d scholar Mathematics department
Baba Mastnath University Asthal bohar , Rohtak

Abstract

This review presents a comprehensive analysis of single-server retrial queueing models, focusing on the critical aspects of batch arrivals and vacation dynamics. Retrial queues are vital in service systems where customers may need to reattempt service after an initial failure, making them relevant in various practical applications. The study emphasizes the significance of understanding how batch arrivals affect server behavior and system performance, particularly regarding congestion and waiting times. Additionally, the review explores the implications of server vacations, which further complicate the dynamics of retrial queues by introducing periods of unavailability. The main contributions of this review include a detailed examination of queue length and waiting time distributions, the development of mathematical modeling techniques, and the integration of insights from existing literature. By synthesizing key findings and methodologies, this review aims to provide a foundational understanding of the complex interactions in retrial queueing systems, offering valuable implications for future research and practical implementations.

Keywords: Single-server retrial queues, batch arrivals, vacation dynamics, queue length.

1. Introduction

Retrial queueing systems are a specialized class of queueing models designed to capture the behavior of customers who cannot be served immediately upon arrival and must wait before attempting to access the server again. In these systems, customers that find the server busy do not simply leave; instead, they enter a retrial orbit, where they wait for a specific period before making another attempt to receive service. This behavior is particularly relevant in various real-world contexts, such as telecommunications, where calls may be blocked or busy, and customers must retry later, or in customer service environments like call centers, where agents may be occupied. The dynamics of retrial queueing systems can significantly differ from those of traditional single-server queues due to the additional complexities introduced by retrial behavior. Understanding the dynamics of batch arrivals within retrial queueing systems is crucial. Batch arrivals occur when multiple customers arrive at the system simultaneously, which can lead to sudden increases in demand for service.

ISBN Number: 978-93-95305-66-2

This situation presents unique challenges for system performance, as larger groups of customers can exacerbate congestion, leading to longer waiting times and increased queue lengths. Modeling batch arrivals helps researchers and practitioners assess how varying batch sizes affect the overall performance of the queue, allowing for better predictions and management of service capabilities. Moreover, the behavior of customers within the retrial orbit can also be influenced by batch arrivals, necessitating the need for accurate mathematical modeling to reflect these interactions. In addition to batch arrivals, the phenomenon of server vacations adds another layer of complexity to retrial queueing systems. Server vacations refer to periods when the server is temporarily unavailable, whether due to scheduled breaks, maintenance, or unforeseen circumstances. During these vacation times, customers who are already in the queue cannot receive service, leading to potential increases in waiting times and overall system congestion.

The interaction between server vacations and retrial behavior must be considered to accurately capture the system's performance metrics. Understanding how vacation dynamics affect customer behavior and system performance is essential for effective operational management and resource allocation. Retrial queueing systems are vital for accurately modeling many real-world service scenarios. The interplay between batch arrivals and server vacations can significantly influence the system's efficiency and customer satisfaction. Therefore, developing robust models that incorporate these factors is essential for optimizing service delivery in various applications, including telecommunications, healthcare, and customer service industries. This background lays the foundation for a comprehensive exploration of performance measures within retrial queueing systems, focusing on the effects of batch arrivals and vacation dynamics.

2. Literature Review

The exploration of retrial queueing systems has garnered significant attention in recent years, particularly with respect to their complex dynamics involving customer behaviors, service interruptions, and system performance metrics. This literature review synthesizes key findings from recent studies focusing on various aspects of retrial queues, including non-Markovian dynamics, performance comparisons using different modeling techniques, and the effects of server vacations. Kalaiselvi and Saravanarajan (2024) delve into non-Markovian dynamics by analyzing a single-server retrial queue that incorporates recurrent clients and balking behavior, specifically under extended Bernoulli vacations.

ISBN Number: 978-93-95305-66-2

Their stochastic analysis highlights the intricacies of customer behavior in response to server availability and illustrates how vacation policies impact overall system efficiency. This work emphasizes the need for comprehensive models that account for the non-exponential inter-arrival and service time distributions, which are often present in real-world applications. Liu, Xu, and Liu (2024) contribute to the understanding of retrial queues by examining parallel customer arrivals and standby server configurations. Their methodology analyzes various strategies that can enhance service efficiency and reduce customer waiting times. The findings suggest that incorporating parallel service mechanisms in retrial systems can significantly improve performance metrics, thereby offering valuable insights for the design of more resilient service infrastructures. Aarthi (2024) compares the performance of fuzzy queuing models against intuitionistic fuzzy queuing models within the context of single-server retrial queues. This comparative analysis demonstrates the strengths of fuzzy logic in handling uncertainties related to customer behavior and service disruptions. The study illustrates that fuzzy models can provide a more nuanced understanding of queue dynamics, particularly in environments where traditional probabilistic approaches may fall short.

The comprehensive literature review by Mathavavisakan and Indhira (2024) on retrial queueing systems with Bernoulli vacations reveals a growing body of research focused on vacation policies. This review outlines various modeling approaches and their implications for system performance, highlighting how vacation strategies can influence queue length and waiting time distributions. By synthesizing existing research, the authors provide a valuable framework for understanding the interplay between server availability and customer behavior in retrial systems. Saravanan et al. (2023) extend this discussion by analyzing a multi-server retrial queueing system characterized by unreliable servers, discouragement factors, and vacation policies. Their performance analysis illustrates the challenges posed by server unreliability and how vacation dynamics can exacerbate congestion in the system. This work emphasizes the importance of considering multiple factors that affect service delivery in complex environments. Dimitriou (2023) introduces a novel perspective by studying a single-server retrial queue with event-dependent arrival rates. This approach allows for the modeling of varying customer arrival patterns, reflecting real-world scenarios where demand fluctuates based on specific events. The findings underscore the significance of adaptability in queueing models to capture dynamic customer behaviors effectively. Wang et al. (2023) investigate the strategic interactions of customers in a single-server retrial queue characterized by noncooperative and cooperative joining strategies.

ISBN Number: 978-93-95305-66-2

Their research highlights how different policies, including N-policy and multiple server vacations, can affect customer decisions and overall system performance. The results suggest that cooperative strategies may lead to improved outcomes for both customers and service providers. Sanga and Jain (2022) propose a fuzzy model for a single-server double orbit retrial queue, emphasizing the role of fuzzy logic in capturing customer behavior across multiple orbits. Their model provides a comprehensive framework for analyzing service dynamics in more complex scenarios, where customers may enter different service paths based on their experiences. The literature on retrial queueing systems reveals a diverse range of approaches and methodologies that enhance our understanding of customer behavior, service dynamics, and performance optimization. The interplay between batch arrivals, server vacations, and various customer strategies continues to be a focal point for research, offering significant implications for the design and management of service systems across various industries. Future research should further explore the integration of advanced modeling techniques to address the complexities of real-world service environments.

3. Model Formulation and Assumptions

3.1 General Structure of Retrial Queues

In retrial queueing systems, the general structure revolves around a scenario where customers arrive to receive service from a single server. These systems are unique in how they handle situations when the server is busy. Unlike traditional queueing models where customers wait in line, in a retrial queue, customers who arrive to find the server occupied do not form a typical waiting queue. Instead, they enter what is called a "retrial orbit," a virtual holding area where they wait for an opportunity to retry for service after some time. This feature of retrying differentiates retrial queues from more conventional models. The server in this type of system alternates between different states: idle, busy, or on vacation. When the server is idle, it is immediately available to serve any arriving customer. If it is busy, customers are directed to the retrial orbit. The retrial process is an essential part of the system's dynamics, where customers make independent attempts to access the server after waiting for a random amount of time. In some models, the server also takes scheduled or random vacations, during which it is unavailable for service. This adds another layer of complexity, as customers not only compete with each other to get service but must also account for times when the server is on break.

In terms of arrivals, retrial queueing systems can also model batch arrivals, where multiple customers arrive simultaneously. The batch size might follow a specific probability distribution, such as Poisson or geometric. If the server is free when a batch arrives, customers are served one by one. However, if the server is busy, the entire batch (or sometimes individual customers from the batch) enters the retrial

ISBN Number: 978-93-95305-66-2

orbit, further increasing the number of retries needed before service can be completed. The retrial process is governed by a probability distribution, typically exponential, that determines when customers retry for service. This general structure of retrial queues, involving customers being directed to the retrial orbit when the server is busy and retrying for service, forms the basis for analyzing complex queueing models. The retrial and vacation dynamics introduce unpredictability and impact key performance metrics, such as waiting times and queue lengths, making the study of these systems highly relevant in fields like telecommunications and manufacturing where these behaviors are common.

- **Arrival Process (Batch Arrivals)**

The batch arrival process be modeled as a Poisson process with rate \(\lambda\). The size of the arriving batch is a random variable \(B\) with probability distribution:

$$\boldsymbol{P(B = k) = p_k \; k = 1, 2, \ldots}$$

The average batch size can be given as:

$$\boldsymbol{E[B] = \sum_{k=1}^{\infty} k \, . \, p_k}$$

This describes that on average(E[B] customers arrive at a time.

- **Service Process**

Let the service times be independent and identically distributed (i.i.d.) with an exponential distribution. If the service rate is denoted by \(\mu\), the service time \(S\) follows:

$$\boldsymbol{P(S \leq t) = 1 - e^{-\mu t} \; for \; t \geq 0}$$

Here, μ is the service rate, which represents the speed at which the server completes service for one customer.

- **Retrial Process**

Customers that find the server busy are sent to the retrial orbit. The retrial times are assumed to be exponentially distributed with rate \(\theta\), meaning that each customer in the orbit attempts to access the server after a random time:

$$\boldsymbol{P(Retrial \; time \leq t) = 1 - e^{-\theta t} for \; t \geq 0}$$

Here, θ is the retrial rate, determining how frequently a customer retries for service.

ISBN Number: 978-93-95305-66-2

Vacation Process

The server may take vacations, which follow an exponential distribution. Let V denote the length of the vacation with rate *v*:

$$\boldsymbol{P(V \leq t) = 1 - e^{-vt} for\ t \geq 0}$$

Here, *v* is the vacation rate, indicating how long the server remains unavailable for service during a vacation.

- **System State**

Let X(t) be the state of the system at time t, where:

X(t) = 0 if the server is idle,

X(t) = 1 if the server is busy serving a customer,

X(t) = 2, 3,… for the number of customers in the retrial orbit.

The probability that there are n customers in the system at time t is denoted by P_n t.

3.2 Assumptions

In modeling a single-server retrial queue with batch arrivals and vacation dynamics, certain assumptions are made to simplify and structure the system's behavior. These assumptions relate to batch size distribution, retrial rate, vacation dynamics, and service time distributions, which are essential in formulating the queue's performance.

- **Batch Size Distribution:** It is assumed that customers arrive in batches, where the size of each batch is a random variable, denoted as B. The batch size follows a specified probability distribution $P(B = k) = p_k$, where k represents the number of customers in the batch, and p_k is the probability that a batch contains k customers. For simplicity, the batch size could follow common distributions such as Poisson, geometric, or binomial, depending on the system being modeled. The expected batch size is denoted by E[B], which helps in determining the arrival load on the system.
- **Retrial Rate:** Customers who find the server busy upon arrival do not queue in the conventional sense but enter a virtual waiting area (the retrial orbit). It is assumed that customers in this orbit make independent attempts to retry for service after some random time. The time between retries is exponentially distributed with parameter θ, where θ represents the

ISBN Number: 978-93-95305-66-2

retrial rate, governing how frequently a customer retries to access the server. This exponential assumption simplifies the analysis of retrial behavior, as the memoryless property of the exponential distribution is useful in deriving system dynamics.

- **Vacation Dynamics:** The server is allowed to take vacations, which are periods during which it is unavailable for serving customers. It is assumed that vacation times follow an exponential distribution with parameter *v*, where *v* is the vacation rate. This means that the duration of the server's vacation is random but follows the exponential distribution. During vacation periods, arriving customers are either sent to the retrial orbit or must wait until the server returns. The server's vacations could be modeled as single or multiple vacations, and the length of each vacation is assumed to be independent of the others.
- **Service Time Distributions:** The service time for each customer is assumed to be an independent and identically distributed random variable following an exponential distribution with rate μ. This means that the time a customer spends being served by the server is memoryless and random, and its probability distribution is:

$$\boldsymbol{P(S \leq t) = 1 - e^{-\mu t} \; for \; t \geq 0}$$

Here, μ is the service rate, determining the speed at which the server processes each customer. Exponential service times are a standard assumption in queueing theory as they allow for tractable mathematical analysis.

4. Analysis of Performance Measures

In retrial queueing systems, understanding the performance of the system under various conditions is critical. The key performance measures include queue length, waiting times, server utilization, and the probability that the server is idle or busy. This section focuses on analyzing these measures to assess system behavior, particularly under the influence of retrial dynamics, batch arrivals, and server vacations.

4.1 Queue Length Distribution

Queue length distribution is one of the fundamental performance measures in any queueing system, as it reflects the number of customers in the system at any given time, including both those waiting in the retrial orbit and those being served. In retrial queueing systems with batch arrivals and vacation dynamics, the queue length distribution is more complex than in traditional queueing models due to the unique behaviors associated with retries and the server's unavailability during vacations.

ISBN Number: 978-93-95305-66-2

4.1.1 Analytical Methods for Calculating Queue Length in Retrial Systems

To calculate the queue length distribution in retrial systems, analytical methods such as Markov chains, probability generating functions (PGF), or balance equations are employed. These methods take into account the stochastic nature of customer arrivals, retrial attempts, service times, and server vacations. The goal is to determine the steady-state probability distribution P_n, which represents the probability that there are n customers in the system at a given time. In the context of a retrial queue with batch arrivals, the arrival process is modeled as a Poisson process with batch size distribution. For each batch, customers either enter service (if the server is idle) or are sent to the retrial orbit (if the server is busy). Customers in the retrial orbit retry for service after an exponentially distributed time. The queue length distribution can be found by solving the system's Kolmogorov forward equations (also known as balance equations), which describe the transitions between different states of the system, such as moving from a state where n customers are present to a state with n+1 or n-1 customers.

For instance, the balance equations take into account:

By solving these balance equations, typically using generating functions or numerical methods, the steady-state probabilities P_n for each possible queue length n are obtained. Once the queue length distribution is known, other performance metrics like average queue length, waiting times, and system utilization can be derived. The general form of the balance equations is given by:

$$\boldsymbol{\lambda P_0 = \mu P_{n+1} + \theta P_v \; for \; n \geq 1}$$

$$\boldsymbol{(\lambda + \mu) = \mu P_{n+1} + \theta P_r \; for \; n \geq 1}$$

Where:

P_v is the probability that the server is on vacation.

P_r is the probability that a customer retries from the orbit.

P_n represents the probability of n customers in the system.

Probability Generating Function (PGF)

To simplify the analysis, we use the probability generating function (PGF) for the queue length distribution:

$$\boldsymbol{G(z) = \sum_{n=0}^{\infty} P_n z^n}$$

ISBN Number: 978-93-95305-66-2

The generating function allows us to manipulate the queue length probabilities more easily by converting the infinite sum of balance equations into a functional equation.

From the balance equations, the PGF is derived as:

$$G(z) = \frac{(1-\rho)(1-z)}{1-z-\lambda(1-z^k)/\mu}$$

Where:

ρ is the server utilization,

z is a complex variable.

4.2 Waiting Time Distribution

Waiting time distribution is a critical performance measure in retrial queueing systems as it reflects how long a customer has to wait before being served. In systems with server vacations and retrial dynamics, waiting time analysis becomes more complex due to the intricate interactions between customer arrivals, retries, and the server's availability.

Effects of Server Vacations and Retrials on Waiting Time

The waiting time of a customer in a retrial queue is composed of two key components:

- **Time spent in the retrial orbit:** This is the time a customer spends retrying to access the server after their initial attempt.
- **Time spent waiting for service:** Once a customer is successfully admitted to the queue, they still have to wait for the server to become available.

When a customer arrives at the system and finds the server busy, they are sent to the retrial orbit. The time a customer spends in the orbit is determined by the retrial rate θ, which follows an exponential distribution. This means the longer a customer waits in the retrial orbit, the more likely they are to retry for service. The mean retrial time is inversely proportional to θ, with the average time in the orbit given by:

$$E[W_{orbit}] = 1/\theta$$

ISBN Number: 978-93-95305-66-2

For systems with batch arrivals, multiple customers may enter the retrial orbit simultaneously, increasing the number of customers retrying for service. This creates congestion in the orbit, leading to longer average waiting times for service.

Waiting Time due to Server Vacations have a direct impact on waiting times. When the server goes on vacation, it temporarily becomes unavailable, causing the queue to build up. Customers arriving during the vacation period are either added to the queue (if allowed) or sent to the retrial orbit. The time they wait during the server's absence adds to their overall waiting time. The vacation time V typically follows an exponential distribution with rate *v*. The mean vacation time E[V] is:

$$\boldsymbol{E[V] = 1/v}$$

During this period, no customers are served, so the waiting time for customers arriving during a vacation is prolonged by the entire duration of the vacation. Additionally, customers who have entered the retrial orbit will continue retrying, but even if they succeed in gaining access to the server, they must wait until the server returns from vacation. This means that retrials during vacations are wasted attempts, adding to the effective waiting time.

5. Conclusion

This review of single-server retrial queueing models with batch arrivals and vacation dynamics has provided valuable insights into the complex interplay between retrials, customer arrivals, and server availability. Retrial queueing systems are highly relevant in settings where customers or tasks that do not receive immediate service must retry. The retrial process helps manage these cases efficiently, ensuring the server is utilized optimally, and preventing system overload. The retrial rate, denoted by θ, significantly influences the system's congestion, affecting both queue length and waiting time distributions. The role of batch arrivals in retrial systems is particularly impactful.

When customers arrive in batches, the server faces intense bursts of traffic, causing longer waiting times and larger queue lengths. The distribution of batch sizes influences these outcomes, making it important to model batch arrival patterns accurately. Larger batch sizes can overwhelm the server, leading to more customers being sent to the retrial orbit, further increasing congestion and waiting times. Server vacations introduce an additional layer of complexity to these models. During vacation periods, the server becomes unavailable, which increases both the waiting time and the queue length as customers must either wait longer or retry from the orbit. Different vacation policies, such as multiple vacations or exhaustive vacations, have varying impacts on system performance. For instance,

ISBN Number: 978-93-95305-66-2

multiple vacations prolong the unavailability of the server, while exhaustive vacation policies ensure that all customers in the system are served before the server takes a break, leading to shorter waiting times overall. The queue length distribution reveals how batch arrivals and vacation periods interact with retrial behavior. The mathematical models, including probability generating functions (PGFs) and balance equations, capture the system's dynamics and provide insights into how various factors contribute to system congestion. The analysis demonstrates that the combination of batch arrivals, retrial attempts, and server vacations significantly impacts the queue length, making these models suitable for studying real-world scenarios. Moreover, the waiting time distribution is deeply influenced by retrials and server vacations. Customers who cannot access the server immediately may experience prolonged waiting times, particularly when the server is on vacation or when multiple customers are retrying simultaneously. The use of Laplace-Stieltjes transforms (LSTs) and other mathematical techniques allows for precise estimation of waiting times, providing a detailed understanding of how system parameters affect customer experience.

References

1. Kalaiselvi, J., & Saravanarajan, M. C. (2024). Exploring non-Markovian dynamics: Stochastic analysis of a single-server retrial queue with recurrent clients and balking behavior under extended Bernoulli vacations. *Heliyon*, *10*(13).
2. Liu, X., Xu, X., & Liu, M. (2024). Strategy Analysis of Retrial Queue with Parallel Customer and Standby Server. *Methodology and Computing in Applied Probability*, *26*(3), 1-22.
3. Aarthi, S. (2024). COMPARISON OF SINGLE SERVER RETRIAL QUEUING PERFORMANCE USING FUZZY QUEUING MODEL AND INTUITIONISTIC FUZZY QUEUING MODEL WITH INFINITE CAPACITY. *Reliability: Theory & Applications*, *19*(2 (78)), 218-231.
4. MICHEAL MATHAVAVISAKAN, N. G. S., & INDHIRA, K. (2024). A LITERATURE REVIEW ON RETRIAL QUEUEING SYSTEM WITH BERNOULLI VACATION. *Yugoslav Journal of Operations Research*, *34*(1).
5. Saravanan, V., Poongothai, V., & Godhandaraman, P. (2023). Performance analysis of a multi server retrial queueing system with unreliable server, discouragement and vacation model. *Mathematics and Computers in Simulation*, *214*, 204-226.
6. Dimitriou, I. (2023). A single server retrial queue with event-dependent arrival rates. *Annals of Operations Research*, *331*(2), 1053-1088.
7. Mathavavisakan, N. G. M., & Indhira, K. (2023). A literature review on retrial queueing system with Bernoulli vacation. *Yugoslav Journal of Operations Research*, *34*(1), 109-134.
8. Wang, Z., Liu, L., Zhao, Y. Q., Li, L., & Xu, W. (2023). Joining strategies of noncooperative and cooperative in a single server retrial queue with N-policy and multiple server vacations. *Communications in Statistics-Theory and Methods*, *52*(4), 1076-1100.

ISBN Number: 978-93-95305-66-2

9. Sanga, S. S., & Jain, M. (2022). Fuzzy modeling of single server double orbit retrial queue. *Journal of Ambient Intelligence and Humanized Computing*, *13*(9), 4223-4234.
10. Sun, K., & Wang, J. (2021). Equilibrium joining strategies in the single-server constant retrial queues with Bernoulli vacations. *RAIRO-Operations Research*, *55*, S481-S502.
11. Jain, M., & Rani, S. (2021). Markovian model of unreliable server retrial queue with discouragement. *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences*, *91*, 217-224.
12. Jain, M., & Sanga, S. S. (2021). Unreliable single server double orbit retrial queue with balking. *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences*, *91*, 257-268.
13. Ayyappan, G., & Thilagavathy, K. (2021). Analysis of MAP (1), MAP (2)/PH/1 non-preemptive priority queueing model under classical retrial policy with breakdown, repair, discouragement, single vacation, standby server, negative arrival and impatient customers. *International Journal of Applied and Computational Mathematics*, *7*(5), 184.
14. Chakravarthy, S. R. (2020). A retrial queueing model with thresholds and phase type retrial times. *Journal of applied mathematics & informatics*, (3-4), 351.
15. Sun, K., Wang, J., & Zhang, G. (2019). Strategic joining in a single-server retrial queue with batch service. In *Queueing Theory and Network Applications: 14th International Conference, QTNA 2019, Ghent, Belgium, August 27–29, 2019, Proceedings 14* (pp. 183-198). Springer International Publishing.
16. Cui, S., Su, X., & Veeraraghavan, S. (2019). A model of rational retrials in queues. *Operations Research*, *67*(6), 1699-1718.
17. Ahuja, A., Jain, A., & Jain, M. (2019). Finite population multi-server retrial queueing system with an optional service and balking. *International Journal of Computers and Applications*, *41*(1), 54-61.
18. Bruneel, H., & Wittevrongel, S. (2017). Analysis of a discrete-time single-server queue with an occasional extra server. *Performance Evaluation*, *116*, 119-142.
19. Upadhyaya, S. (2016). Queueing systems with vacation: an overview. *International journal of mathematics in operational research*, *9*(2), 167-213.
20. Nobel, R. (2016). Retrial queueing models in discrete time: a short survey of some late arrival models. *Annals of Operations Research*, *247*, 37-63.
21. Sakurai, H., & Phung-Duc, T. (2016). Scaling limits for single server retrial queues with two-way communication. *Annals of Operations Research*, *247*, 229-256.
22. Wang, J., & Zhang, F. (2016). Monopoly pricing in a retrial queue with delayed vacations for local area network applications. *IMA Journal of Management Mathematics*, *27*(2), 315-334.
23. Wang, F., Wang, J., & Zhang, F. (2015). Strategic behavior in the single-server constant retrial queue with individual removal. *Quality Technology & Quantitative Management*, *12*(3), 325-342.
24. Artalejo, J. R., & Lopez-Herrero, M. J. (2012). The single server retrial queue with finite population: a BSDE approach. *Operational Research*, *12*, 109-131.
25. Krishnamoorthy, A., Gopakumar, B., & Narayanan, V. C. (2012). A retrial queue with server interruptions, resumption and restart of service. *Operational Research*, *12*, 133-149.