

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
NAAS Rating (2026): 4.49
Maths 2026; 11(2): 108-112
© 2026 Stats & Maths
<https://www.mathsjournal.com>
Received: 19-11-2025
Accepted: 20-01-2026

Dr. Helena Costa
Department of Operations
Research, Faculty of Industrial
Engineering and Management,
Atlantic School of Engineering,
Porto, Portugal

Dr. Miguel Ferreira
Department of Operations
Research, Faculty of Industrial
Engineering and Management,
Atlantic School of Engineering,
Porto, Portugal

Dr. Inês Almeida
Department of Operations
Research, Faculty of Industrial
Engineering and Management,
Atlantic School of Engineering,
Porto, Portugal

Dr. Tiago Ribeiro
Department of Operations
Research, Faculty of Industrial
Engineering and Management,
Atlantic School of Engineering,
Porto, Portugal

Corresponding Author:
Dr. Helena Costa
Department of Operations
Research, Faculty of Industrial
Engineering and Management,
Atlantic School of Engineering,
Porto, Portugal

Balking, reneging, server vacations, and optional repair in M/M/1-type queues: A literature review

Helena Costa, Miguel Ferreira, Inês Almeida and Tiago Ribeiro

DOI: <https://www.doi.org/10.22271/math.2026.v11.i2b.2614>

Abstract

Real queueing systems rarely match the clean assumptions of the classical M/M/1 model: customers decline to join a long queue, grow impatient and leave before being served, and the server itself may be unavailable, whether on a scheduled vacation or undergoing repair after a breakdown. This review traces the literature on four such departures from the classical model balking, reneging, server vacations, and repair from Haight's foundational 1950s papers through the vacation-queue survey of Doshi and the joint balking-reneging models of Ancker and Gafarian, to more recent studies that combine two or three of these features in a single Markovian queue. Particular attention is given to how each combination is analysed, generally via steady-state balance equations, a probability generating function, or a matrix-geometric method, and to how many of the four features a single study manages to combine at once. Across ten studies spanning nearly seven decades, the largest number of features combined in any one study is three; none of the reviewed literature analyses a single M/M/1-type queue with server vacations, balking, reneging, and optional repair together, which is the gap this review sets out to document.

Keywords: M/M/1 queue, balking, reneging, server vacation, optional repair, steady-state analysis, queueing theory, customer impatience, server breakdown, performance measures

Introduction

The classical M/M/1 queue assumes Poisson arrivals, exponential service, and a server that is always available whenever a customer is present. Real queueing situations rarely cooperate with all three assumptions at once. Customers who see a long queue may simply not join it, a behaviour Haight termed balking in a 1957 paper that derived the steady-state distribution of a queue in which each arriving customer joins with a probability depending on the number already present ^[1]. Two years later, Haight turned to the complementary behaviour, reneging, in which a customer who has already joined the queue loses patience and departs before being served ^[2].

Ancker and Gafarian brought the two behaviours together in a pair of 1963 papers, deriving steady-state results for a queue in which customers may both balk on arrival and renege after joining, and establishing much of the algebraic machinery birth-death balance equations with state-dependent arrival and service rates that later authors would reuse when adding further features ^[3, 4].

A separate line of work addressed server unavailability. Doshi's 1986 survey organised the by-then substantial literature on server vacations, periods during which the server is unavailable either because it has finished all waiting work or because it is attending to some secondary task, and showed how a wide range of specific vacation models could be understood through a common decomposition result ^[5]. Server breakdown and repair form a related but distinct line: rather than a scheduled or voluntary absence, the server becomes unavailable unpredictably and must be restored, sometimes through a single repair phase and, in related models elsewhere in the literature, through a first essential and second optional repair phase.

This review traces how these four features balking, reneging, vacations, and repair have been combined in the literature on single-server Markovian queues, with particular attention to which combinations have been analysed together and which have not.

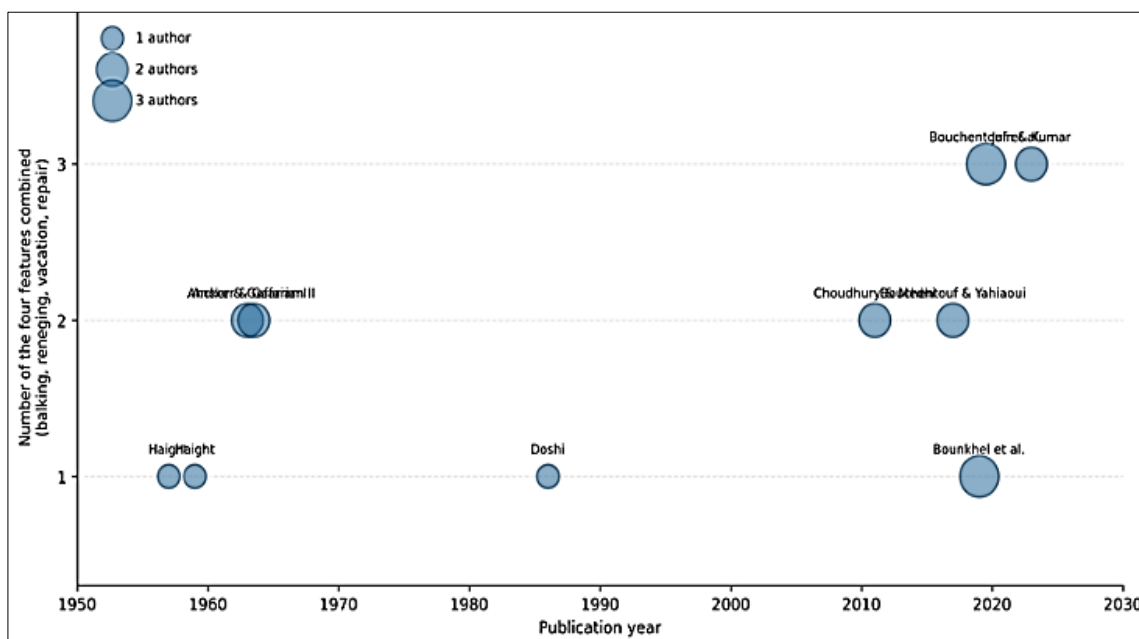


Fig 1: Number of the four features combined against publication year; bubble size shows the number of authors ($n = 10$)

Single-feature studies (Haight's two papers and Doshi's survey) sit at the bottom of the plot across a thirty-year span, while every study combining three features falls after 2018. No study reaches a height of four, which is the empirical starting point for the gap discussed later in this review. Bubble size shows no clear relationship with either publication year or feature count, suggesting author-team size in this literature is driven by factors other than how many features a given study combines.

Foundations: Methods for analysing extended M/M/1 queues

Two methodological approaches dominate this literature. The first, used by Haight and by Ancker and Gafarian, treats the queue as a birth-death process with state-dependent rates: the balking probability lowers the effective arrival rate as the queue grows, and the reneging rate adds an extra state-dependent term to the effective service rate, after which the steady-state distribution follows from the usual birth-death balance equations ^[1, 4]. This approach is analytically direct but becomes unwieldy once vacations or repair are added, since the state space must be expanded to track the server's status alongside the queue length.

The second approach, associated with Doshi's survey and the vacation-queue literature more broadly, treats the server's status busy, idle, on vacation, or in repair as an additional dimension of the state space and derives a probability generating function or matrix-geometric solution over the joint state ^[5]. Doshi's central decomposition result, that the queue length or waiting time in a vacation system can be written as the sum of two independent pieces, one matching the corresponding no-vacation system and one capturing the residual vacation time, gives later authors a template for extending the basic M/M/1 vacation model without re-deriving the whole steady-state solution from scratch.

Later studies that combine vacations with balking and reneging generally adopt the second approach, augmenting the state space with a vacation indicator and applying the birth-death or matrix-geometric machinery to the combined process. Bouchentouf and Yahiaoui's 2017 model, for instance, combines multiple working vacations with reneging and a feedback mechanism in a single continuous-time Markov chain, solved via the standard balance equations extended to the vacation states ^[7].

Table 1: Chronological summary of reviewed studies

Study (Year)	Ref.	Feature(s) Modelled	Analysis Method	Key Contribution
Haight (1957)	[1]	Balking	Steady-state distribution	Introduced balking to queueing theory
Haight (1959)	[2]	Reneging	Steady-state distribution	Introduced reneging to queueing theory
Ancker & Gafarian I (1963)	[3]	Balking + Reneging	Steady-state, birth-death	Joint balking-reneging model
Ancker & Gafarian II (1963)	[4]	Balking + Reneging	Steady-state, extended measures	Extended joint model
Doshi (1986)	[5]	Vacation	Survey / decomposition	Foundational vacation-queue survey
Choudhury & Medhi (2011)	[6]	Balking + Reneging	Finite-buffer steady-state	Balking/reneging under finite buffer
Bouchentouf & Yahiaoui (2017)	[7]	Vacation + Reneging (+ feedback)	Steady-state, cost model	Vacation combined with reneging and feedback
Bounkhel <i>et al.</i> (2019)	[8]	Repair (breakdown)	Steady-state, linear operator	Flexible service with breakdown/repair
Bouchentouf <i>et al.</i> (2019)	[9]	Vacation + Balking + Reneging	Steady-state, cost/profit	First three-feature combination reviewed here
Jain & Kumar (2023)	[10]	Vacation + Balking + Repair	Supplementary variable, probability generating function	Closest precedent to a four-feature model

Combining balking, reneging, and vacations

Combining balking with reneging was the first extension to be worked out in detail. Ancker and Gafarian's two 1963 papers

established the steady-state queue-length distribution and several performance measures for a single-server queue in which both a state-dependent balking probability and a state-

dependent reneging rate operate simultaneously [3, 4]. Choudhury and Medhi later revisited the same combination under a finite buffer, showing that the finite capacity changes several of the steady-state expressions but leaves the qualitative balking-reneging interaction intact [6]. Combining vacations with reneging came later and required tracking the server's vacation status as part of the Markov chain state. Bouchentouf and Yahiaoui's model adds reneging, feedback, and a retention mechanism, in which a reneging customer may be persuaded to stay, on top of a multiple-working-vacation queue, solving the resulting chain for steady-state probabilities and a cost function [7]. Bouchentouf, Cherfaoui, and Boualem went further, combining vacations with both balking and reneging together in a feedback queue and deriving a full economic analysis alongside the steady-state solution, making theirs the first three-feature combination in the literature reviewed here [9].

Repair and the search for a four-feature model

Repair has been treated somewhat separately from the balking-reneging-vacation cluster. Bounkhel, Tadj, and Hedjar analysed a flexible queue in which the server can switch between single and batch service and may break down during either mode, deriving the steady-state distribution via a linear-operator approach rather than the birth-death or matrix-geometric methods used elsewhere in this review [8]. Jain and Kumar's 2023 model is the closest existing precedent to a four-feature combination, incorporating an unreliable server subject to breakdown and repair, working vacations with interruption, and balking, within a retrial-queue rather than a plain M/M/1 framework, and solved via the supplementary-variable technique [10]. Their model does not include reneging, which is the one feature it is missing relative to the full combination named in this review's title.

Table 2: Key concepts referenced in this review

Concept	Ref.	Description
Balking	[1]	An arriving customer declines to join the queue at all, typically with a probability that depends on the number of customers already present
Reneging	[2]	A customer who has already joined the queue departs before completing service, typically after an exponentially distributed patience time
Server vacation	[5]	The server becomes unavailable for a random period after completing service or emptying the queue, resuming service upon return
Server breakdown and repair	[8, 10]	The server fails during service according to a Poisson process and is restored after a repair period; in related models elsewhere in the literature this repair period is split into a first essential and second optional phase

Table 2 defines, in words, the four features around which this review is organised; every study discussed here models one or

more of these four concepts on top of a base Markovian queue.

Table 3: Reference distribution by journal

Journal	Number of References
Biometrika	1
Metrika	1
Operations Research	2
Queueing Systems	1
RAIRO-Operations Research	1
Arabian Journal of Mathematics	1
Entropy	1
OPSEARCH	1
Proceedings of the National Academy of Sciences, India, Section A	1

Operations Research accounts for two of the ten references, both by Ancker and Gafarian; every other journal in the list contributes exactly one reference, spreading this literature

across probability, operations-research, and applied-mathematics venues rather than a single specialist queueing journal.

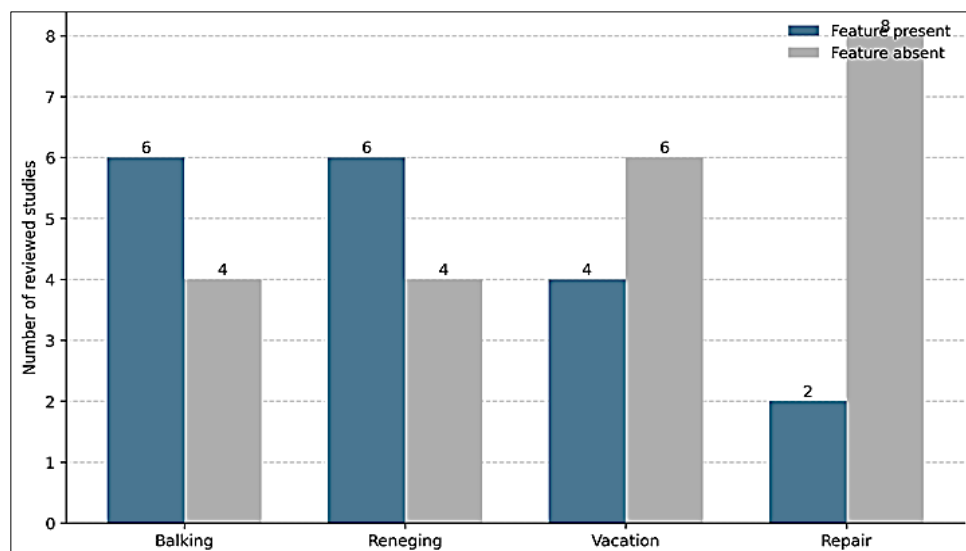


Fig 2: Reviewed studies by presence of each of the four features (n = 10)

Balking and renegeing are each present in six of the ten reviewed studies, and vacation in four, but repair appears in only two. Repair is therefore the least-modelled of the four

features considered here, which the empirical basis for the gap is identified in the Synthesis below.

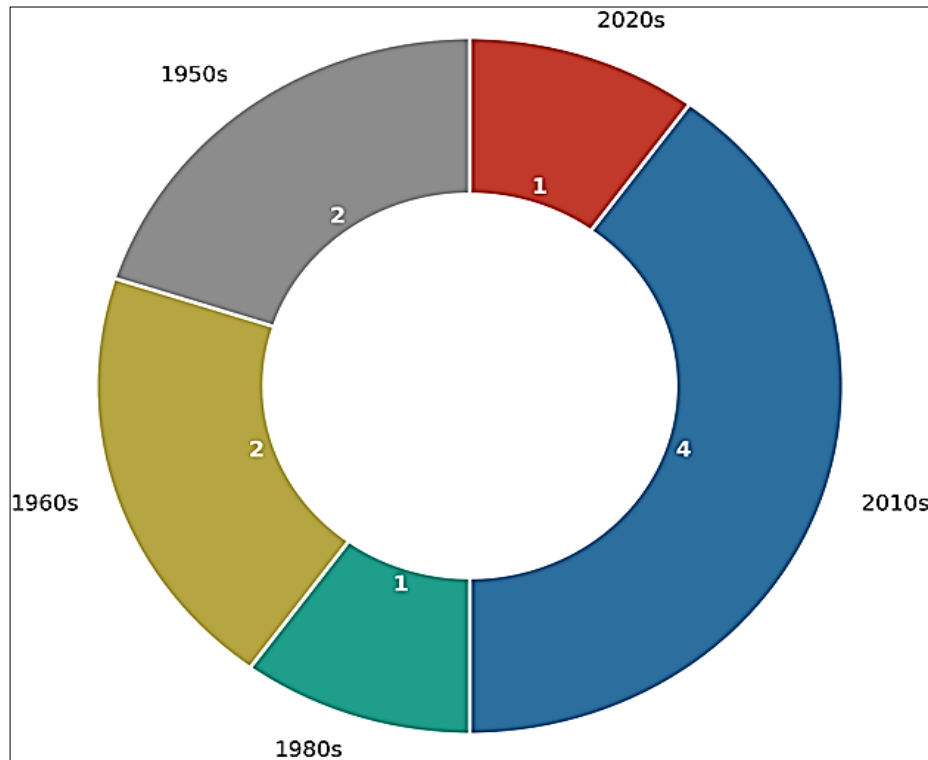


Fig 3: Reviewed references by decade of publication ($n = 10$)

Four of the ten references, the largest single share, come from the 2010s, when vacation-renegeing and vacation-balking-renegeing combinations were first worked out. The 1950s and 1960s together account for four more, reflecting the field's

origins in Haight's and Ancker and Gafarian's single- and dual-feature models, with a twenty-year gap before Doshi's 1986 survey and a further twenty-five years before the 2010s cluster begins.

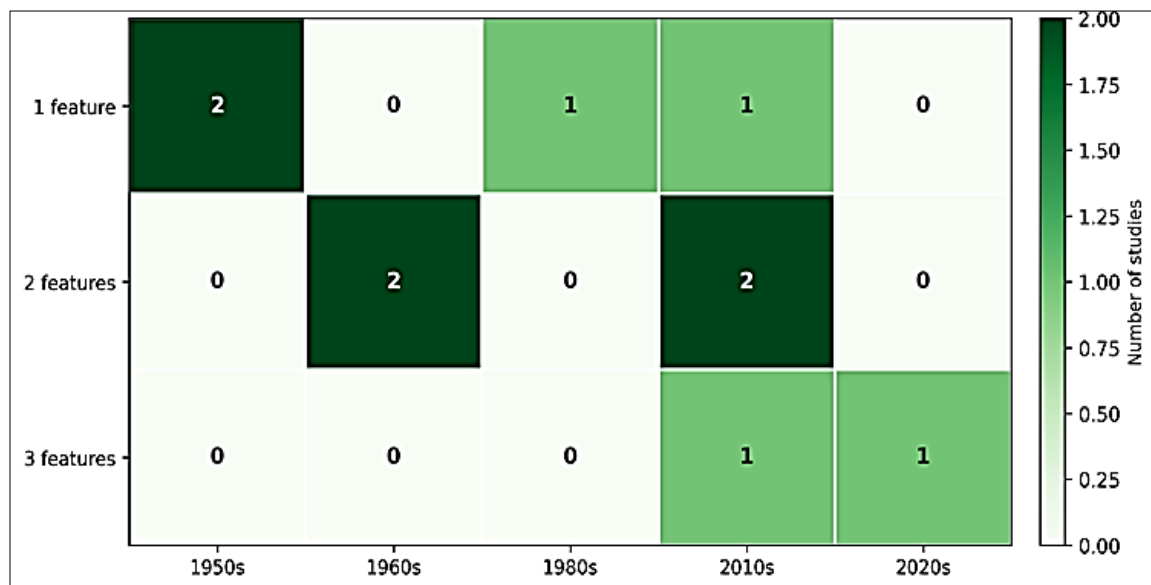


Fig 4: Number of features combined by decade of publication ($n = 10$)

Figure 4 cross-tabulates feature count against decade. Every three-feature study falls in either 2019 or 2023, the two most recent years covered here, while every 1950s and 1980s study combines exactly one feature. The 1960s and 2010s columns are the only ones containing two-feature studies, three decades apart, which underlines how gradually this literature has added features one combination at a time.

Synthesis and research gaps

Read together, these ten studies show steady, incremental combination rather than a single unified framework. Balking and renegeing were combined first, in the 1960s, using birth-death methods that remain the standard toolkit for any two-feature combination even today [3, 4, 6]. Vacations were treated separately for two decades before being combined with renegeing, and then with both balking and renegeing, using the

state-space-augmentation approach that Doshi's survey had organised^[5, 7, 9]. Repair has followed the most separate path of all, appearing alongside vacations and balking in Jain and Kumar's retrial-queue model but never alongside reneging, and never within a plain, non-retrial M/M/1 framework^[8, 10].

The count in Figure 2 makes the resulting gap explicit: repair is the least-modelled of the four features considered here, present in only two of the ten reviewed studies, and neither of those two studies also includes reneging. No reviewed study combines all four features vacations, balking, reneging, and optional repair in a single M/M/1 queue. Jain and Kumar's model is the closest precedent, missing only reneging, and it is set in a retrial rather than a direct-access queue, which changes several of the balance equations relative to the simpler M/M/1 structure implied by this review's scope^[10].

A second observation is methodological rather than substantive. The birth-death and matrix-geometric approaches used throughout this literature are, in principle, additive: each new feature typically adds one dimension or one term to an already-worked-out set of balance equations rather than requiring a fundamentally new solution method. This suggests a four-feature M/M/1 model combining vacations, balking, reneging, and optional repair is technically within reach using tools already established in the papers reviewed here, even though no study has yet assembled all four in one place.

Conclusion

This review has traced how four departures from the classical M/M/1 assumptions balking, reneging, server vacations, and repair have been introduced and progressively combined over nearly seven decades, from Haight's original single-feature models through Ancker and Gafarian's balking-reneging combination to the three-feature vacation models of the late 2010s and Jain and Kumar's near-complete 2023 model.

The clearest gap identified here is the absence of a single-server Markovian queue that combines all four features, server vacations, balking, reneging, and optional repair, together, analysed via the same birth-death or matrix-geometric methods already well established for smaller combinations. Given that each feature has already been combined with at least two of the other three in some prior study, assembling all four together looks like an incremental rather than a fundamentally new undertaking, and would complete a natural progression that this literature has been building toward since Bouchentouf, Cherfaoui, and Boualem's three-feature model.

Declarations

Conflict of interest: The author(s) declare no conflict of interest.

Funding: This review received no specific funding.

References

1. Haight FA. Queueing with balking. *Biometrika*. 1957;44(3-4):360-369.
2. Haight FA. Queueing with reneging. *Metrika*. 1959;2(1):186-197.
3. Ancker CJ, Gafarian AV. Some queueing problems with balking and reneging. I. *Oper Res*. 1963;11(1):88-100.
4. Ancker CJ, Gafarian AV. Some queueing problems with balking and reneging. II. *Oper Res*. 1963;11(6):928-937.
5. Doshi BT. Queueing systems with vacations: A survey. *Queueing Syst*. 1986;1:29-66.

6. Choudhury A, Medhi P. Some aspects of balking and reneging in finite buffer queues. *RAIRO Oper Res*. 2011;45(3):223-240.
7. Bouchentouf AA, Yahiaoui L. On feedback queueing system with reneging and retention of reneged customers, multiple working vacations and Bernoulli schedule vacation interruption. *Arab J Math*. 2017;6(1):1-11.
8. Bounkhel M, Tadj L, Hedjar R. Steady-state analysis of a flexible Markovian queue with server breakdowns. *Entropy*. 2019;21(3):259.
9. Bouchentouf AA, Cherfaoui M, Boualem M. Performance and economic analysis of a single server feedback queueing model with vacation and impatient customers. *Opsearch*. 2019;56:300-323.
10. Jain M, Kumar A. Unreliable server M[X]/G/1 retrial feedback queue with balking, working vacation and vacation interruption. *Proc Natl Acad Sci India Sect A Phys Sci*. 2023;93:57-73.