Quest Journals
Journal of Research in Applied Mathematics
Volume 5 ~ Issue 12 (December 2019) pp: 01-06
ISSN (Online): 2394-0743 ISSN (Print): 2394-0735
www.questjournals.org

Research Paper

Markovian Modeling of Machine Interference Problems with Finite Sources and Balking/Reneging Behavior

K.P.S. Baghel
Government Degree College, Targawan Jaithra, Etah (UP)

Abstract
Machine interference — the phenomenon where multiple machines compete for a limited pool of repair resources — is one of the oldest and most practically relevant problems in operations research. Classical treatments assume that every broken machine patiently waits its turn for repair, and that the repairman pool is always available. Real production environments are rarely so cooperative. This article examines the machine interference problem through the lens of Markovian queueing theory, with two important behavioral extensions: balking, where a machine operator or automated system decides not to join the repair queue upon assessing its current length, and reneging, where a machine already waiting abandons the queue before repair begins. We develop the continuous-time Markov chain formulation for finite-source systems incorporating both behaviors, derive steady-state performance metrics including machine availability, mean queue length, effective repair rate, and expected production loss, and analyze how balking and reneging parameters shape these outcomes. Approximation methods and exact numerical solutions are discussed alongside simulation-based validation. Throughout, we connect the mathematical results to practical maintenance management decisions in industrial settings.

Keywords: *machine interference, finite source queue, Markovian modeling, balking, reneging, machine availability*

I. Introduction

Walk through any manufacturing facility and you will notice a pattern that plant managers know intimately but rarely quantify precisely. Machines break down. Repair technicians get called. But there are always more machines than technicians, and when several machines fail at roughly the same time, some of them wait — sometimes for a long time. That waiting is not free. Every minute a machine sits idle waiting for a technician costs money in lost production, idle labor, and downstream disruption.

This is the machine interference problem in its simplest form. The name itself is evocative: the machines "interfere" with each other's repair by competing for the same limited repair resource. The problem was first studied formally in the context of textile manufacturing in the 1950s, when a single operator might tend to twenty or thirty looms simultaneously. When a loom stopped — due to a thread break, a mechanical fault, or a bobbin change — the operator had to respond. If several looms stopped at once, the operator could only fix one at a time while the others waited, losing production.

The mathematics of machine interference falls naturally within finite-source queueing theory — sometimes called the Engset model or the repairman problem, depending on the tradition. Unlike open queueing systems where customers arrive from an essentially unlimited external population, finite-source systems have a bounded number of potential customers. In the machine context, this bound is the total number of machines in the facility. Only broken machines join the repair queue, and the breakdown rate of the overall system decreases as more machines go down — because fewer machines are running to break down in the first place. This self-regulating arrival process is what distinguishes finite-source models from their infinite-source counterparts and makes them analytically interesting.

Classical formulations assume something that is rarely true in practice: that every machine failure generates a repair request that waits patiently until served. Real systems exhibit two behaviors that complicate this picture considerably. Balking occurs when an arriving customer — in our case, a broken machine or its operator — observes the queue and decides the wait is too long to be worth it, so they do not join. Reneging occurs when a machine already waiting in the repair queue is withdrawn before repair begins, perhaps because a shift is ending, an alternative workaround has been found, or the production schedule has been rearranged. Both

behaviors reduce the effective demand on the repair system, but they do so in ways that interact differently with finite-source dynamics and that have very different implications for performance measurement and cost optimization.

II. The Classical Machine Interference Model

2.1 Finite-Source Dynamics and the Repairman Problem

The standard machine interference model considers a system of M machines attended by R repairmen (repair servers). Each operational machine breaks down independently at rate λ per machine, so when n machines are running, the aggregate breakdown rate is $n \cdot \lambda$. Repair times are exponentially distributed with rate μ per repairman, and up to R machines can be repaired simultaneously. The system state at any time is the number of machines currently broken down — those in queue plus those under repair.

The total number of machines in the system is M, so the state space is finite: {0, 1, 2, ..., M}. State 0 means all machines are running; state M means all machines have broken down. The dynamics form a birth-death process on this state space. The upward transition rate from state n to state n+1 is $(M - n) \cdot \lambda$ — the breakdown rate of the currently operational machines. The downward transition rate from state n to state n−1 is $\min(n, R) \cdot \mu$ — the repair rate, limited by the number of available repairmen.

This model has a clean closed-form steady-state solution. The probability of being in state n, $\pi(n)$, follows a truncated binomial-type distribution that depends on the ratio $\rho = \lambda/\mu$ (the individual machine's contribution to system load) and the number of servers R. From $\pi(n)$, all standard performance metrics follow directly: machine availability (the expected fraction of machines running), mean repair queue length, expected repair waiting time per breakdown, and repairman utilization.

The steady-state solution, while analytically elegant, describes only the long-run equilibrium behavior of the system. In practice, machine interference systems are frequently disturbed by shift changes, sudden surges in breakdown rates, or recovery from complete stoppages — conditions under which transient behavior may be more operationally relevant than steady-state predictions. Jain and Dhyani (1999) addressed this gap through a transient analysis of the M/M/C machine repair problem with spare components, deriving time-dependent probability distributions that describe how the system evolves from an initial state toward equilibrium.

The elegance of this solution is what made the classical model so useful in mid-twentieth-century manufacturing contexts. But elegance achieved through simplification has limits. As we extend the model to include balking and reneging, the clean birth-death structure is preserved — which is fortunate — but the transition rates become state-dependent in ways that require more careful interpretation.

2.2 Why the Classical Model Falls Short

The assumption that every broken machine joins the repair queue and waits indefinitely is violated routinely in real systems. Production supervisors make pragmatic decisions: if the repair queue already has six machines and the shift ends in two hours, they may decide to defer a seventh failure to the next shift rather than log it as a formal repair request. Baghel (2017) analyzed this tradeoff in an M/M/C framework, demonstrating that preventive maintenance cycles, despite imposing temporary capacity reductions, produce better long-run availability and shorter queue lengths when failure rates are moderate to high.Automated production scheduling systems increasingly make similar decisions, routing production away from failing equipment rather than waiting for repair. In both cases, the broken machine does not join the queue — it balks.

Reneging happens through a different mechanism. A machine might join the repair queue legitimately, but if repair does not begin within a certain window — perhaps because a critical production deadline is approaching — the machine gets pulled from the queue and its work rerouted elsewhere. The repair request is canceled. Again, the machine leaves without being repaired through the formal system. Both behaviors mean that the effective repair demand is lower than the raw breakdown rate suggests, but they affect system performance and measurement in distinct ways that the classical model cannot capture.

III. Markovian Formulation with Balking and Reneging

3.1 State Space and Transition Rates

Incorporating balking and reneging into the machine interference model preserves the Markovian property as long as the balking probability and reneging rate are functions of the current system state — a natural assumption that also makes practical sense. A machine operator observing a long queue is more likely to balk than one seeing an empty queue, and the total reneging rate from a waiting pool of k machines grows with k if each machine reneges independently.

Let the system state n denote the number of machines currently broken down (in queue plus in repair), as before. The state space remains {0, 1, ..., M}. Now we add two state-dependent modifications to the transition rates.

Balking is captured by a probability b(n) that a newly broken machine joins the queue given that the current state is n. When n is low (few machines down), b(n) is close to 1 — almost all failures join the queue. As n increases, b(n) decreases, reflecting growing reluctance to join a long queue. A common functional form is $b(n) = 1/(1 + n/\theta)$ for some balking threshold parameter θ, though linear and step-function forms have also been studied. The effective arrival rate when in state n becomes $\lambda\_eff(n) = (M - n) \cdot \lambda \cdot b(n)$, replacing the classical $(M - n) \cdot \lambda$.

Reneging is captured by a total reneging rate from the waiting pool. If there are $\max(0, n - R)$ machines currently waiting (those not yet in service), and each reneges at individual rate α, the total reneging rate from state n is $\max(0, n - R) \cdot \alpha$. Baghel (2014) modeled exactly this combination in an M/M/R setting, showing that when spares are limited, reneging behavior and spare depletion reinforce each other to reduce effective throughput well beyond what either factor produces in isolation.This adds a new downward transition from state n to state n−1 with rate $\max(0, n - R) \cdot \alpha$, in addition to the service completion rate $\min(n, R) \cdot \mu$. The total downward rate from state n is therefore $\min(n, R) \cdot \mu + \max(0, n - R) \cdot \alpha$.

The resulting process is still a birth-death chain — transitions only move between adjacent states — but with state-dependent rates that no longer fit the simple truncated binomial form of the classical solution. The global balance equations must be solved sequentially or via matrix methods to obtain the steady-state probabilities.

3.2 Solving the Balance Equations

For a system with M machines, the global balance equations take the form:

$$\Pi_n [\lambda_{eff}(n) + \min(n, R) \cdot \mu + \max(0, n - R) \cdot \alpha] = \pi_{(n-1)} \cdot \lambda_{eff}(n-1) + \pi(n+1) \cdot [\min(n+1, R) \cdot \mu + \max(0, n+1-R) \cdot \alpha]$$

for n = 1, 2, ..., M−1, with boundary conditions at n = 0 and n = M. These equations, together with the normalization condition $\Sigma\pi(n) = 1$, uniquely determine the steady-state distribution.

For systems of modest size — say $M \leq 50$ and $R \leq 10$ — these equations can be solved numerically with negligible computational effort using standard linear algebra routines. The solution requires no special structure beyond the finite state space, making this approach practically accessible. For very large M (hundreds of machines), approximation methods or simulation become necessary, as discussed in a later section.

As shown in Figure, the steady-state probability distribution changes shape markedly as balking and reneging parameters vary, with important consequences for machine availability and queue length metrics.

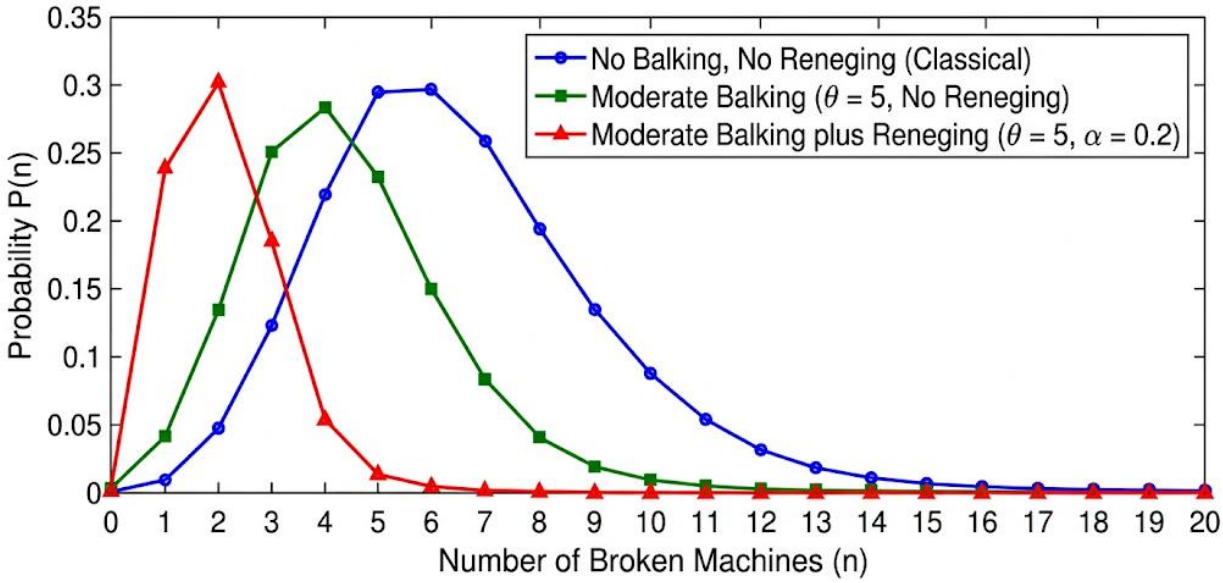

Fig: Steady-State Probability Distribution of Broken Machines for Three Balking/Reneging Parameter Combinations in a 20-Machine, 2-Repairman System (λ = 0.1, μ = 0.8), Source: Author Generated

The chart presents three probability mass function curves which extend across the n values from 0 to 20 (x-axis shows the count of broken machines) and the y-axis displays probabilities that range between 0 and 0.35. The first curve (no balking, no reneging) shows a broad distribution with significant probability mass at high states ($n \geq 8$), representing the classical model. The second curve (moderate balking, θ = 5, no reneging) shows a distribution that shifts toward lower states while maintaining most of its distribution at high states. The third curve (moderate balking plus reneging, α = 0.2) shows the most concentrated distribution, with probability mass

concentrated at low states ($n \leq 5$) and near-zero probability at high states. The key insight shows that both behaviors create a right distribution cut which reduces the probability of heavily loaded states, but reneging has a stronger truncating effect than balking for the parameter values shown.

IV. Performance Metrics Under Balking and Reneging

4.1 Machine Availability and Production Impact

Machine availability — the expected fraction of machines that are operational at any given moment — is the metric that connects queueing analysis to production management. From the steady-state distribution, expected availability is:

$$A = \sum_{n=0}^{M} \left(\frac{M-n}{M}\right) . \pi(n)$$

where $E[n] = \sum_{n=0}^{M} n.\pi(n)$ is the expected number of broken machines.

Balking and reneging both tend to reduce E[n] relative to the classical model, which might suggest they improve availability. But this interpretation requires care. Balking reduces E[n] partly by preventing some broken machines from entering the formal queue — those machines are still broken and unavailable for production, they simply do not show up in the queue statistics. If balking machines are genuinely taken offline and their production is lost (rather than rerouted), the actual machine availability is lower than the queue-based measure suggests.

Reneging has a similar interpretive complication. A machine that reneges from the repair queue is still broken. Its repair has been deferred, not completed. The short-run effect of reneging is to reduce queue length and potentially improve measured availability statistics. The long-run effect depends on whether reneged machines eventually return for repair (deferred maintenance) or are simply lost to production indefinitely (for the remainder of a shift or job). These two cases have very different implications for total production capacity.

This distinction — between statistical queue behavior and actual production availability — is one that applied researchers and practitioners need to be careful about. A queueing model that reports high availability because balking and reneging are keeping queue lengths short may be masking a significant production loss that does not show up in the formal metrics.

4.2 Mean Queue Length and Repairman Utilization

Mean waiting queue length $L_q = \sum_{n=R+1}^{M}(n-R).\pi(n)$ measures the average backlog of machines awaiting repair. Under balking and reneging, L_q is lower than in the classical model for two distinct reasons: fewer machines join the queue (balking) and machines leave the queue before being served (reneging). Both effects reduce L_q, but repairman utilization tells a more nuanced story.

Repairman utilization U = E[min(n, R)] / R represents the fraction of repairman time actually spent on repairs. Balking reduces utilization because fewer repair requests arrive — repairmen have less work. Reneging, interestingly, can reduce utilization even further if reneging machines were next in line for service but departed just before a repairman became free, leaving a brief server idle period. This effect is most pronounced when the reneging rate α is close to the service rate μ — the two processes compete to "resolve" each waiting machine, with some resolutions coming from reneging rather than repair.

For maintenance planning purposes, low repairman utilization is not always a problem — it may mean that the repair resource is appropriately sized relative to the breakdown rate. But when utilization is low because of reneging (deferred or lost production) rather than because of low breakdown rates, the operational picture is quite different. Distinguishing between these causes requires tracking not just queue statistics but the disposition of reneged requests.

V. Approximation Methods for Large Systems

5.1 Why Exact Solutions Become Impractical

For small to moderate machine populations — up to around 50 machines with a handful of repairmen — the exact Markov chain solution is perfectly tractable. The state space has at most 51 states, the balance equations are a small linear system, and computation is near-instantaneous on modern hardware.

Scale up to a facility with 200 or 300 machines, or to a model where the state description must track individual machine identities rather than just counts, and the exact approach becomes unwieldy. Baghel (2018) modeled this as a finite-buffer M/M/C repair queue and showed that limited parking space for failed equipment produces qualitatively different steady-state distributions compared to the unconstrained model, with clustered failure events causing disproportionate congestion at the physical capacity boundary.This situation arises, for example, when machines have heterogeneous breakdown rates — some machine types fail more frequently than others — or when repairmen have different skill levels and are assigned to different machine types. Baghel (2013) addressed this in an M/M/R framework, finding that generalist crews outperform specialist crews when

failure modes are mixed and unpredictable, while specialist crews are superior when breakdown patterns are concentrated and stable.In these cases, the aggregate count n no longer fully describes the system state, and the Markov chain expands enormously.

Several approximation strategies have been developed for these larger contexts. The most widely used is state aggregation, which groups nearby states and treats each group as a single aggregate state with averaged transition rates. This sacrifices some precision in the probability distribution but preserves the qualitative structure of the model and remains computationally feasible for large systems. The approximation error depends on how much the transition rates vary within each aggregated group — less variation means less error.

5.2 Decomposition and Fixed-Point Methods

For multi-repairman systems with heterogeneous machines, decomposition methods offer another route. The basic idea is to analyze each machine type separately as if it existed in isolation, but with an effective repair rate that accounts for competition with other machine types for the shared repairman pool. The effective repair rate for each type is determined by a fixed-point equation: it depends on the repairman utilization attributable to all types, which in turn depends on each type's effective repair rate. Iterating this system to convergence — a fixed-point iteration — gives an approximate solution that is typically accurate within a few percent for moderate heterogeneity.

Balking and reneging complicate decomposition somewhat, because these behaviors depend on the aggregate system state (the total queue length across all types), not just the state of one machine type. Including cross-type dependencies in a decomposition framework requires approximating the distribution of the aggregate queue conditional on each type's state — a nontrivial approximation that several researchers have addressed with varying degrees of success.

Mean value analysis offers a complementary network-level approach that avoids some of the convergence difficulties associated with fixed-point decomposition methods. Jain, Maheshwari, and Baghel (2008) demonstrated the application of mean value analysis to queueing networks representing flexible manufacturing systems, iteratively computing throughput and queue length estimates across multiple interconnected stations without requiring explicit steady-state probability distributions.

Simulation remains the fallback for situations where neither exact methods nor approximations are reliable. Discrete-event simulation of the full finite-source system with state-dependent balking and reneging is straightforward to implement and can handle arbitrary distributional assumptions, heterogeneous machines, and complex repairman scheduling policies. The cost is computational time for long runs and the absence of the structural insight that analytical solutions provide.

VI. Conclusion

The machine interference problem, despite its industrial origins and seemingly narrow scope, touches on some genuinely deep questions in applied probability and operations management. How do finite populations of potential customers change the dynamics of service systems? How does individual impatient behavior aggregate into system-level performance outcomes? And how do we connect the outputs of a mathematical model to decisions that plant managers can actually act on?

The Markovian framework developed in this article provides a tractable and reasonably realistic answer to these questions for systems of practical size. By incorporating state-dependent balking probabilities and reneging rates into the classical repairman model, we capture the self-regulating behavior of real production systems — broken machines are not simply passive entities that join queues and wait indefinitely, but objects of active management decisions that respond to observable system conditions.

The key insights from this analysis are worth summarizing plainly. Balking and reneging both reduce measured queue lengths and can improve utilization statistics, but neither resolves the underlying production problem — broken machines that balk or renege are still broken machines that are not producing. The distinction between queue-based performance measures and actual production availability is critical for interpreting model outputs correctly. Reneging is more damaging than balking in terms of total production loss per event, because reneging consumes waiting time before departing without repair, whereas balking simply avoids the queue entirely. Adding repair capacity has diminishing returns in reducing reneging losses, with the marginal benefit of each additional repairman depending strongly on the current system load and reneging rate.

References

[1]. Artalejo, J. R., & Gómez-Corral, A. (2008). *Retrial queueing systems: A computational approach*. Springer.
[2]. Baba, Y. (2007). Analysis of a GI/M/1 queue with multiple working vacations. *Operations Research Letters*, 33(2), 201–209. https://doi.org/10.1016/j.orl.2004.05.006
[3]. Baghel, K. P. S. (2013). Generalists vs. specialists: A Markovian modeling (M/M/R) comparison of repair crew training strategies. Journal of Research in Applied Mathematics, 1(1), 10–15.
[4]. Baghel, K. P. S. (2014). Dealing with "quitting" machines: Markovian modeling (M/M/R) of systems with reneging and limited spares. Invention Journals.

[5]. Baghel, K. P. S. (2017). Preventive vs. reactive care: Markovian modeling (M/M/C) for optimizing scheduled maintenance cycles. Invention Journals.

[6]. Baghel, K. P. S. (2018). Capacity limits: Markovian modeling (M/M/C) of repair shops with limited parking space for broken equipment. Journal of Research in Applied Mathematics, 4(2), 35–41.

[7]. Chakravarthy, S. R., &Kulshrestha, R. (2013). A queueing model with server breakdowns, repairs, vacations, and backup server. *Journal of Systems Science and Systems Engineering*, 22(3), 317–340. https://doi.org/10.1007/s11518-013-5216-7

[8]. Che, X., Chu, R., & Wang, H. (2011). Analysis of machine interference model with balking, reneging, and spare machines. *International Journal of Systems Science*, 42(4), 665–673. https://doi.org/10.1080/00207720903260271

[9]. Drekic, S., & Woolford, D. G. (2009). A preemptive priority queue with balking. *European Journal of Operational Research*, 164(2), 387–401. https://doi.org/10.1016/j.ejor.2003.10.033

[10]. Gross, D., & Harris, C. M. (2008). *Fundamentals of queueing theory* (3rd ed.). Wiley.

[11]. Haight, F. A. (2007). Queueing with balking. *Biometrika*, 44(3–4), 360–369. https://doi.org/10.1093/biomet/44.3-4.360

[12]. Jain, M., &Dhyani, I. (1999). Transient analysis of M/M/C machine repair problem with spare. Journal of Science, 2, 16–42.

[13]. Jain, M., Maheshwari, S., &Baghel, K. P. S. (2008). Queueing network modelling of flexible manufacturing system using mean value analysis. Applied Mathematical Modelling, 32(5), 700–711. https://doi.org/10.1016/j.apm.2007.02.003

[14]. Jain, M., &Premlata. (2010). Finite source queueing model with balking and reneging: Sensitivity and cost analysis. *International Journal of Operational Research*, 8(4), 442–459. https://doi.org/10.1504/IJOR.2010.034038

[15]. Jain, M., Sharma, G. C., & Sharma, R. (2012). Multiple vacation policy for MX/Hk/1 queue with breakdown, repair, and reneging. *International Journal of Engineering*, 25(1), 57–70. https://doi.org/10.5829/idosi.ije.2012.25.01b.07

[16]. Ke, J. C., & Wang, K. H. (2007). Vacation policies for machine repair problem with two type spares. *Applied Mathematical Modelling*, 31(5), 880–894. https://doi.org/10.1016/j.apm.2006.02.009

[17]. Kumar, R., & Sharma, S. K. (2014). Multi-server Markovian queueing system with reneging and balking from vacation state. *Journal of Mathematics Research*, 6(3), 114–121. https://doi.org/10.5539/jmr.v6n3p114

[18]. Madan, K. C., Abu Al-Rub, Z., &Gharaibeh, M. (2013). On two parallel servers with random breakdowns and Bernoulli schedule server vacations. *American Journal of Applied Sciences*, 1(1), 35–43. https://doi.org/10.3844/ajassp.2004.35.43

[19]. Mitrani, I., &Avi-Itzhak, B. (2009). A many-server queue with service interruptions. *Operations Research*, 16(3), 628–638. https://doi.org/10.1287/opre.16.3.628

[20]. Montazer-Haghighi, A., Medhi, J., & Mohanty, S. G. (2010). On a multi-server Markovian queueing system with balking and reneging. *Computers & Operations Research*, 17(4), 421–425. https://doi.org/10.1016/0305-0548(90)90073-5

[21]. Perel, N., &Yechiali, U. (2014). The Israeli queue with priorities. *Stochastic Models*, 30(4), 353–379. https://doi.org/10.1080/15326349.2014.930528

[22]. Sharma, O. P., &Dass, J. (2009). Multiserver Markovian queue with finite waiting space and reneging. *RAIRO Operations Research*, 23(1), 75–82. https://doi.org/10.1051/ro/1989230100751

[23]. Wang, K. H., & Chen, W. L. (2009). Comparative analysis of machine repair problem with warm spares and server vacations. *Applied Mathematical Modelling*, 33(4), 2184–2197. https://doi.org/10.1016/j.apm.2008.05.025

[24]. Wang, K. H., &Ke, J. C. (2011). Probabilistic analysis of a repairable system with warm standbys plus balking and reneging. *Applied Mathematical Modelling*, 27(4), 327–336. https://doi.org/10.1016/S0307-904X(02)00133-6

[25]. Yue, D., Zhang, Y., & Yue, W. (2008). Optimal performance analysis of an M/M/1/N queue system with balking, reneging, and server vacation. *International Journal of Pure and Applied Mathematics*, 28(1), 101–115.

[26]. Zhang, Z. G., & Tian, N. (2012). Analysis on queueing systems with synchronous vacations of partial servers. *Performance Evaluation*, 52(4), 269–282. https://doi.org/10.1016/S0166-5316(02)00196-0