RT&A, No 4 (89)
Volume 20, December 2025

STOCHASTIC PETRI NET ANALYSIS OF RETRIAL REPAIRABLE SYSTEMS WITH TWO REPAIR MODES

[1*]Kheireddine Boudehane and [2] Samira Taleb

•

[1] University of Science and Technology Houari Boumediene USTHB,
Department of Probability and Statistics, Faculty of Mathematics,
RIIMA Laboratory, El-Alia, Bab Ezzouar, B.P. 32, 16111, Algiers, Algeria.
[2]University of Science and Technology Houari Boumediene USTHB,
Department of Probability and Statistics, Faculty of Mathematics,
MSTD Laboratory, El-Alia, Bab Ezzouar, B.P. 32, 16111, Algiers,
[1*]kboudehane@usthb.dz; [2]staleb@usthb.dz

Abstract

This paper presents a dependability analysis of an L-out-of-n system supported by a hierarchical standby structure with warm and cold units. The model accounts for realistic operating conditions, including a single unreliable repair server that may break down but can still provide limited service during failure periods. A retrial mechanism is also considered, where failed units that find the server busy join an orbit and retry for service later. The system is modeled using a Generalized Stochastic Petri Net (GSPN), which is then mapped to a Continuous-Time Markov Chain (CTMC) for steady-state performance evaluation. Furthermore, transient reliability analysis is conducted to derive the survival function of the system and the expectation of the first passage time to failure.

Keywords: K-out-of-n system, reliability, generalized stochastic Petri net, standby redundancy

1. Introduction

Reliability and availability are important indices for evaluating the performance and dependability of systems and devices. Reliability is a concept that concerns many areas of human activity: economic, scientific, technical, industrial and technological. It is related to notions of operational safety, quality, efficiency and performance. Its main objective is the analysis of the probability of a system failure and concerns in particular the mathematical domain. In some modelling approaches, the study of reliability is reduced to the study of the distribution of a positive random variable representing the lifetime of a system or component.On the other hand, Availability is the proportion of time that a system is expected to be operational and available for use. Interested readers on the subject are referred to Barlow and Proshan [1], David [2], Kececioglu [3], Lai and Xie [4], Birolini [5].

In general, real systems are composed of several components and have several fault modes, these systems are said to be complex. Functional modelling in the form of a block diagram of these systems makes it possible to determine their reliability. Each block represents a component or fault mode. Furthermore, the blocks are considered independent of each other. The most common coherent structures are serial, parallel and k-out-of-n systems (a system is said to be coherent if the normal functioning of all elements leads to the normal functioning of the structure and if the failure of all elements leads to the failure of the structure).

DOI: https://doi.org/10.24412/1932-2321-2025-489-665-676

A k-out-of-n:G system is a system of n components, that operates if and only if at least k ($1 \leq k \leq n$) of the n components operate. The system is referred to as a series system when $k = n$, and it is a simple parallel system when $k = 1$ and $n > 1$. k-out-of-n:G systems are successfully used in industrial and military systems, data processing systems and communication systems.

Except for series systems, systems usually have redundant structures. A redundant system is a system where one or more components may fail without the system ceasing to function. In standby redundancy, redundant components are put into use after the failure of the operating component. It makes the system much more reliable and efficient. In general, there are three kinds of standbys: warm, cold, and hot. A component is in "hot standby" if its failure rate in quiescent mode is the same as its failure rate in active mode, "cold standby" if its failure rate in quiescent mode is zero (i.e., the component cannot fail while in standby), and "warm standby" if it falls somewhere in the middle.

The study of k-out-of-n: G systems and standby redundancy kept the interest of several authors. Unuvar et al [6] found bounds for k-out-of-n and consecutive k-out-of-n reliabilities in the form of probabilistic constraints. Shekhar et al [7] dealt with the performance modeling and reliability analysis of a redundant machining system composed of several functional machines with switching failure and geometric reneging. Byun et al [8] extended the matrix-based system reliability (MSR) method to k-out-of-n systems. Eryilmaz [9] showed how to increase system lifetime of general coherent structure at system by adding cold standby redundancy to system and component levels. A retrial machine system with warm standby is studied in Chen [10] and Chen and wang [11]. In the first work the recovery policy is considered while in the second work the N policy is considered. An $M/G/1$ machine interference problem with imperfect switchover of standbys, in which an unreliable repairman maintains a group of machines is investigated by Ke et al [12]. Subasi et al [13] developed an algorithm to obtain robust and efficiently commutable bounds for the probability that exactly k-out-of-n events occur, where the underlying probability distribution is assumed to be uni-modal with known mode. Yang and Tsao [14] considered a repairable system consisting of m primary components, s spare components, and a repairman with multiple vacations. Sarkar and Tang [15] compared limiting availabilities of a k-out-of-n:G system and a k-component series system with $(n - k)$ spare components when there are different numbers of repair facilities. Yen et al [16] studied reliability and sensitivity analysis of a retrial machine repair problem with working breakdowns operating under the F-policy. Zhao et al [17] proposed new concepts of k-out-of-n:F balanced systems based on real engineering problems. Gao [18] investigated availability and reliability characteristics for a repairable fault-tolerant system with warm standbys and second optional repair service. Gao and Wang [19] performed a reliability and availability analysis of a repairable retrial system with mixed standbys and a single repair facility that undergoes preventive maintenance or experiences active breakdowns. Li et al [20] dealt with a repairable circular consecutive-k-out-of-n: retrial system where the repair time is assumed to follow a Weibull distribution. Yang and Wu [21] evaluated the availability and reliability of a standby repairable system incorporating imperfect switchovers and working breakdowns. Wang et al [22] proposed a novel mathematical model for reliability dealing with a (k_1, k_2)-out-of-n:G repairable system consisting of two different types of components and multiple vacations.Yuan [23] studied a k-out-of-n:G repairable system with n identical machines and R repairmen, and both the redundant dependency and the multiple vacation policy for repairmen were taken into account.

$GSPNs$ are bipartite mathematical graphs having the power to model complicated systems, to describe and analyze stochastic concurrent systems with synchronization properties. $GSPNs$ comprise Places (drawn as circles), timed transitions, and arcs. $GSPNs$ are an extension of Petri Nets (PNs), in which some transitions are exponentially distributed timed transitions(expressed as rectangles), and others are immediate (represented as thin bars). The state of a $GSPN$ is described by mean of markings containing the number of tokens in each place of the $GSPN$. Depending on whether the transition is timed or immediate, markings are classified into tangible and vanishing respectively. The tangible reachability graph is isomorphic to a continuous time Markov chain ($CTMC$). As a result, the states of the $CTMC$ correspond to the markings in the tangible reachability graph, and the state transition rates are the exponential firing rates of timed transitions in the $GSPN$. The stationary probability of the $CTMC$ allows us to derive some performance measures of the system. Interested readers on the $GSPNs$ can see Haas [24], Bause and Kritzinger [25], Reisig [26].

Numerous studies on $GSPNs$ are available in the literature. For example, Ikhlef et al [27] studied a Semi-Markovian single server retrial queues by means of Markov Regenerative Stochastic Petri Nets. Hakmi et al [28] used generalized stochastic Petri nets for the analysis of a retrial priority queue. YadollahzadehTabari and Mohammadizad [29] studied the efficiency of the S-MAC media access control protocol in wireless sensor networks by analyzing and modeling it using Generalized Stochastic Petri nets.

2. Model description and proposed $GSPN$

2.1. Model description

We consider an L-out-of-$(M+W+C)$ system involving M identical units operating in parallel, W units in warm standby (running in a partially ready state), C identical cold standby units (powered off and started only when needed), and a single unreliable repairer. The system stops functioning once the number of failed units reaches $k = M+W+C-L+1$. All these units are identical. When an operating unit experiences a failure, it is immediately replaced by an available warm unit. At the same time, an available cold standby unit becomes a warm unit. If no warm standby unit is available, a cold standby unit is activated to replace the failed operating unit. Once a standby unit (whether warm or cold) becomes active, it operates identically to the original operating unit. The transition from standby units (warm or cold) to operating units is considered perfect, with instantaneous switchover times. The lifetimes of operating and warm units are modeled as exponential distributions with failure rates λ and α, respectively, where $0 < \alpha < \lambda$.

The system does not include waiting space for the repairer handling the failed units. Therefore, if an operating or warm standby unit fails while the repairer is occupied with another failed unit, the newly failed unit is placed in an orbit for a random time following an exponential distribution with rate γ. This orbiting behavior represents a delayed repair process, waiting for the repairer to become available.

When a unit fails, it is dispatched to the repairer for inspection and restoration. This repairer, however, is also prone to breakdowns. In such cases, the repairer is restored by an external facility. During its downtime, it can provide partial repair services at a reduced rate μ_2, where $\mu_2 < \mu_1$ and μ_1 is the normal service rate. This scenario models a degraded service state during the repairer's breakdown.

To improve the repairer's utilization and ensure efficient operation, a recovery policy is implemented. This policy stipulates that the repairer can only undergo recovery when the number of failed units in the system meets or exceeds a threshold q ($q < L$). The repairer's lifetime and recovery time are modeled using exponential distributions with rates θ and σ, respectively. All failed units arriving during this period are placed in orbit, waiting until the repairer completes its recovery cycle. Once the recovery is completed, the repairer resumes normal operation at the full service rate μ_1. This enhancement reflects real-world practices where scheduled maintenance helps avoid costly downtimes caused by unexpected equipment failures.

Practical Example: $5G$ Edge System for Autonomous Vehicles

The proposed model can be applied to cloud service provider that operates M virtual machines (VMs) to serve customer requests. To improve reliability, the provider maintains W warm standby VMs running in a low-power state and C cold standby VMs powered off until needed. The system continues to operate as long as at least L VMs remain available, but it stops functioning once the number of failed VMs reaches $k = M+W+C-L+1$. Whenever an active VM fails, it is immediately replaced by a warm standby. At the same time, one of the cold standby VMs, if available, is promoted to warm status. If no warm standby is available, a cold standby is directly activated to replace the failed VM. Once promoted, standby VMs behave identically to the active ones.

Failed VMs are repaired by a dedicated repair server. If the repair server is already busy, failed VMs enter an orbit and retry for repair after a random time. The repair server itself is unreliable and may fail. When this occurs, it is repaired by an external facility. During its degraded state, it can still process repairs but at a reduced capacity. To optimize resource usage, the repair server is restored only when the number of failed VMs reaches a threshold q, with $q < L$.

This configuration reflects real-world data centers where backup servers and repair strategies are combined to ensure high availability despite random failures and repair delays.

3. *GSPN* Description

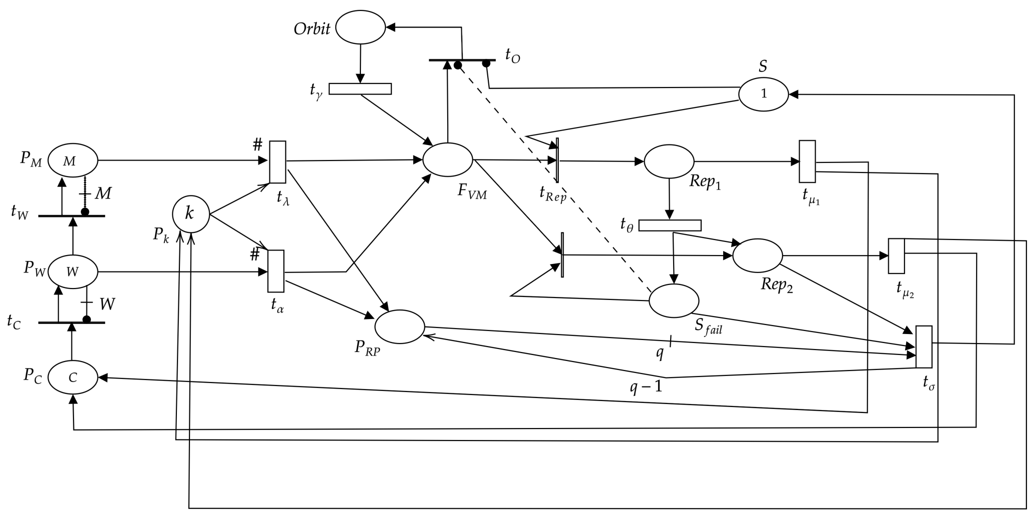


Figure 1: *The GSPN model of the system with warm and cold units, an unreliable repairer, and a recovery policy.*

In Fig. 1, we depict the GSPN model of the system with warm and cold *units*, an unreliable *repairer*, and a recovery policy. The places of the GSPN are defined as follows:

- Operating *units* are represented by place P_M with an initial marking of M.
- Warm *units* are depicted by place P_W with an initial marking of W.
- Cold *units* are represented by place P_C with an initial marking of C.
- Place F_U indicates whether a failed *unit* or a chosen *unit* from the orbit is ready for repair.
- The orbit is represented by the place *Orbit*.
- A *unit* under normal repair is represented by place Rep_1.
- A *unit* under partial repair (when the *repairer* is down) is represented by place Rep_2.
- The failed *repairer* is represented by place S_{fail}. This place signifies that the *repairer* is broken down due to a failure.
- The idle *repairer* is represented by place S. This place indicates that the *repairer* is operational and ready to address *unit* failures.
- The *repairer* recovery policy is represented by place P_{RP}.

Thus, the *GSPN* starts with an initial marking M_0:

$$M_0(\#P_M, \#P_W, \#P_C, \#Orbit, \#F_U, \#P_{RP}, \#S, \#Rep_1, \#Rep_2, \#S_{fail}) = M_0(M, W, C, 0, 0, 0, 1, 0, 0, 0)$$

where # denotes the marking in a place.

The failure of an operating *unit* is characterized by the firing of transition t_λ with rate $\#P_M\lambda$. The failure of a warm *unit* is represented by the firing of transition t_α with rate $\#P_W\alpha$. The arrival of a repeated *unit* is characterized by the firing of transition t_γ with rate γ.

The firing of transitions t_{μ_1} and t_{μ_2} represents the end of the normal repair period when the *repairer* is operational and the end of the partial repair period when the *repairer* is in a breakdown state, with rates μ_1 and μ_2, respectively. An active breakdown of the *repairer* is characterized by the firing of transition t_θ with rate θ. The end of the repair period for the failed *repairer* is characterized by the firing of transition t_σ with rate σ.

Transition t_{Rep_1} fires if there is a token in place F_U (representing the arrival of a *unit* from orbit or a newly failed *unit*) and the *repairer* is available (one token in place S). Transition t_{Rep_2} fires if there is a token in place F_U and the *repairer* is in a breakdown state (indicated by a token in place S_{fail}).

Transition t_O fires when the *repairer* is either under repair or busy, causing the failed *unit* to be placed into orbit. Transition t_W fires when P_M contains fewer tokens than M, replacing a failed primary *unit* with a warm *unit*. Transition t_C fires when P_W contains fewer tokens than W, replacing a failed warm *unit* with a cold *unit*.

Finally, transition t_σ fires when the number of failed *units* equals q, representing the end of the repair period for the failed *repairer*, with rate σ.

4. Steady-State Analysis

To simplify the notation of the *GSPN* marking, let us define:

- $I(t)$: The number of *units* being repaired at time t (in place Rep_1 for normal repair or Rep_2 for partial repair).
- $N(t)$: The number of *units* in the orbit (place *Orbit*) at time t.
- $J(t)$: The state of the *repairer* at time t (0 for available/working, 1 for failed).

Hence, the state of the system can be represented by the marking process $\{X(t), t \geq 0\} = \{(I(t), N(t), J(t)), t \geq 0\}$, which is a Continuous-Time Markov Chain (CTMC) with a finite state space Ω where: $\Omega = \{(i, n, j), 0 \leq i \leq 1, 0 \leq n \leq k-1, 0 \leq j \leq 1\}$.

The infinitesimal generator Q of the *CTMC* $\{X(t), t \geq 0\}$ is a $4k \times 4k$ ($k = M + W + C$) tridiagonal block matrix given as follows:

$$Q = \begin{pmatrix} D_1 & U_1 & 0 & \cdots & 0 \\ L_2 & D_2 & U_2 & \ddots & \vdots \\ 0 & L_3 & D_3 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & U_{k-1} \\ 0 & \cdots & 0 & L_k & D_k \end{pmatrix}_{4k \times 4k},$$

where the nonzero blocks are defined as follows:

$$L_i = \begin{pmatrix} 0 & \gamma & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad i = 2, \ldots, k,$$

$$U_i = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \lambda_{i+1} & 0 & 0 \\ 0 & 0 & \lambda_{i+1} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad i = 1, \ldots, k-1,$$

$$D_1 = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 \\ \mu_1 & -(\lambda_1 + \theta + \mu_1) & \theta & 0 \\ 0 & 0 & -(\lambda_1 + \mu_2) & \mu_2 \\ 0 & 0 & \lambda_0 & -\lambda_0 \end{pmatrix},$$

$$D_i = \begin{pmatrix} -(\lambda_{i-1} + \gamma) & \lambda_{i-1} & 0 & 0 \\ \mu_1 & -(\lambda_i + \theta + \mu_1) & \theta & 0 \\ 0 & 0 & -(\lambda_i + \mu_2) & \mu_2 \\ 0 & 0 & \lambda_{i-1} & -\lambda_{i-1} \end{pmatrix}, \quad i = 2, \ldots, q-1,$$

$$D_i = \begin{pmatrix} -(\lambda_{i-1} + \gamma) & \lambda_{i-1} & 0 & 0 \\ \mu_1 & -(\lambda_i + \theta + \mu_1) & \theta & 0 \\ 0 & \sigma & -(\lambda_i + \mu_2 + \sigma) & \mu_2 \\ \sigma & 0 & \lambda_{i-1} & -(\lambda_{i-1} + \sigma) \end{pmatrix}, \quad i = q, \ldots, k-1,$$

$$D_k = \begin{pmatrix} -(\lambda_{k-1} + \gamma) & \lambda_{k-1} & 0 & 0 \\ \mu_1 & -(\theta + \mu_1) & \theta & 0 \\ 0 & \sigma & -(\mu_2 + \sigma) & \mu_2 \\ \sigma & 0 & \lambda_{k-1} & -(\lambda_{k-1} + \sigma) \end{pmatrix}.$$

and

$$\lambda_n = \begin{cases} M\lambda + W\alpha, & 0 \le n \le C-1, \\ M\lambda + (W + C - n)\alpha, & C \le n \le W + C - 1, \\ (M + W + C - n)\lambda, & W + C \le n \le k-1. \end{cases}$$

Note that λ_n is the failure rate when there are n failed *units* in the system.

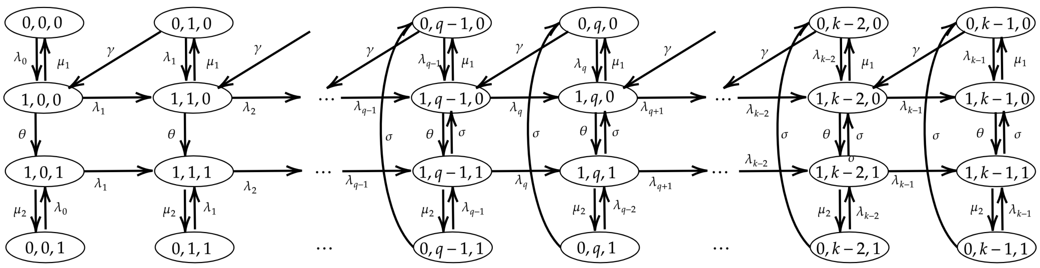


Figure 2: *CTMC associated to the GSPN model given in Fig. 1.*

4.1. Steady-State Distribution

Let $\pi = (\pi_0, \pi_1, \ldots, \pi_{k-1})$ be the stationary probability vector, where

$$\pi_j = (\pi_{0,j,0}, \pi_{1,j,0}, \pi_{1,j,1}, \pi_{0,j,1}), \quad j = 0, 1, \ldots, k-1.$$

Theorem 1. For the coherent repairable system, the steady-state probabilities π_n, $n = 0, 1, \ldots, k-1$, are given by

$$\pi_0 = \pi_{0,0,0}\, r_0,$$
$$\pi_n = \pi_0 \prod_{i=1}^{n} R_i, \quad n = 1, 2, \ldots, k-1, \tag{1}$$

where

$$\pi_{0,0,0} = \frac{1}{r_0\big(\mathbb{I}_4 + \sum_{n=1}^{k-1} \prod_{i=1}^{n} R_i\big) e_4},$$
$$r_0 = \left(1, \frac{\lambda_0}{\mu_1}, \frac{\theta\lambda_0}{\mu_1\lambda_1}, \frac{\mu_2\theta\lambda_0}{\mu_1\lambda_1}\right),$$

and

$$R_n = \begin{cases} -L_n\,(D_n + R_{n+1}U_n)^{-1}, & n = 1, 2, \ldots, k-2, \\ -L_{k-1}D_k^{-1}, & n = k-1, \end{cases}$$

with $\mathbb{I}_4$ denoting the 4×4 identity matrix and $e_4^T = (1, 1, 1, 1)$.

Proof. The stationary distribution vector π is obtained by solving the global balance equation

$$\pi Q = \mathbb{O}_{4k},$$

subject to the normalization condition

$$\sum_{n=0}^{k-1} \pi_n e_4 = 1,$$

where $\mathbb{O}_{4k}$ is a $1 \times 4k$ zero vector. This system yields the block equations:

$$\pi_0 D_0 + \pi_1 U_0 = \mathbb{O}_4, \tag{2}$$
$$\pi_{n-1}L_n + \pi_n D_n + \pi_{n+1}U_n = \mathbb{O}_4, \quad n = 1, 2, \ldots, k-2, \tag{3}$$
$$\pi_{k-2}L_{k-1} + \pi_{k-1}D_k = \mathbb{O}_4. \tag{4}$$

From (4), we obtain

$$\pi_{k-1} = \pi_{k-2}(-L_{k-1}D_k^{-1}) = \pi_{k-2}R_{k-1}.$$

By recursively substituting into (3), each π_n can be expressed in terms of π_{n-1} as

$$\pi_n = \pi_{n-1}R_n, \quad R_n = -L_n(D_n + R_{n+1}U_n)^{-1}.$$

Iterating this relationship yields the expression in (1). Finally, $\pi_{0,0,0}$ is obtained using the normalization condition and the above expressions. ∎

4.2. Performance Measure

From Theorem 1, we can derive the following performance measures:

- **Steady-state availability** $A(\infty)$: The stationary probability that the system is operational.

$$A(\infty) = 1 - (\pi_{1,k-1,0} + \pi_{1,k-1,1}) \tag{5}$$

- **Mean number of failed units in repair** $\overline{N}_U$: The average number of units in places Rep_1 or Rep_2.

$$\overline{N}_U = \sum_{n=0}^{k-1} (\pi_{1,n,0} + \pi_{1,n,1}) \tag{6}$$

- **Mean number of failed units in orbit** $\bar{O}_U$: The average number of units in the place *Orbit*.

$$\bar{O}_U = \sum_{n=0}^{k-1} n \cdot (\pi_{0,n,0} + \pi_{1,n,0} + \pi_{1,n,1} + \pi_{0,n,1}) \tag{7}$$

- **Mean number of failed units** $\bar{N}$:

$$\bar{N} = \overline{N}_U + \bar{O}_U \tag{8}$$

- **Mean rate of failed units** $\bar{\lambda}$:

$$\bar{\lambda} = \sum_{n=0}^{k-1} \lambda_n \cdot (\pi_{0,n,0} + \pi_{1,n,0} + \pi_{1,n,1} + \pi_{0,n,1}) \tag{9}$$

- **Mean response time** $\bar{T}$ (time from a unit's failure until its repair is complete), via Little™s formula:

$$\bar{T} = \frac{\bar{N}}{\bar{\lambda}} \tag{10}$$

- **Mean waiting time** $\bar{T}_w$ (time from a unit's failure until its repair begins), via Little™s formula:

$$\bar{T}_w = \frac{\bar{O}_U}{\bar{\lambda}} \tag{11}$$

5. Reliability Analysis

This section examines the survival function $R_{VM}(t)$ and the first-failure expectation (FFE_{VM}). The survival function $R_{VM}(t)$ is the probability that the system operates without failure up to time t. The FFE_{VM} is the expected average time until the first system failure. For this analysis, all system *units* are assumed to be operational at $t = 0$.

To determine the system reliability $R_{VM}(t)$, we modify the CTMC by making the failure states $(1, k-1, 0)$ and $(1, k-1, 1)$ absorbing states. This creates a new CTMC, $\{\tilde{X}(t), t \geq 0\}$, with an infinitesimal generator matrix $\tilde{Q}$:

$$\tilde{Q} = \begin{pmatrix} D_0 & U_0 & & & & & & & & \\ L_1 & D_1 & U_1 & & & & & & & \\ & L_2 & D_2 & U_2 & & & & & & \\ & & \ddots & \ddots & \ddots & & & & & \\ & & & L_{q-1} & D_{q-1} & U_{q-1} & & & & \\ & & & & L_q & D_q & U_q & & & \\ & & & & & \ddots & \ddots & \ddots & & \\ & & & & & & L_{k-2} & D_{k-2} & U_{k-2} \\ & & & & & & & L_{k-1} & D^*_{k-1} \end{pmatrix}$$

where the last block is given by: $$D^*_{k-1} = \begin{pmatrix} -(\lambda_{k-1}+\gamma) & \lambda_{k-1} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \sigma & 0 & \lambda_{k-1} & -(\lambda_{k-1}+\sigma) \end{pmatrix}.$$

Let us define:

- $\pi(t) = (\pi_0(t), \pi_1(t), \ldots, \pi_{k-1}(t))$ as the time-dependent state probability vector.
- $\pi^*(s) = \int_0^{+\infty} e^{-st}\pi(t)\,dt$ as the Laplace transform (LT) of $\pi(t)$.

The state probability vector $\pi(t)$ satisfies the differential equation $\pi'(t) = \pi(t)\tilde{Q}$. Assuming the system starts in a perfectly operational state, $\pi(0) = (1, 0, \ldots, 0)$, we can apply the Laplace transform to get the matrix equation:

$$\pi^*(s)(s\mathbb{I}_{4k} - \tilde{Q}) = \pi(0), \tag{12}$$

where $\mathbb{I}_{4k}$ is the identity matrix. This system can be solved numerically for $\pi^*(s)$. By applying the inverse Laplace transform to the solutions for the absorbing states, $\pi^*_{1,k-1,0}(s)$ and $\pi^*_{1,k-1,1}(s)$, we can find $\pi_{1,k-1,0}(t)$ and $\pi_{1,k-1,1}(t)$. Subsequently, we can calculate $R_{VM}(t)$ and FFE_{VM} as follows:

$$R_{VM}(t) = 1 - \pi_{1,k-1,0}(t) - \pi_{1,k-1,1}(t), \tag{13}$$

$$FFE_{VM} = \int_0^{+\infty} R_{VM}(t)dt. \tag{14}$$

6. Numerical Example

We analyzed our model by numerically evaluating its performance and reliability. A scenario was established with the following parameters: $\lambda = 0.4$, $\alpha = 0.05$, $\mu_1 = 2$, $\mu_2 = 1$, $\sigma = 2$, $\theta = 0.8$, $\gamma = 1$, $M = 3$, $W = 2$, $C = 1$, and $q = 2$. We then varied each parameter individually to observe its impact on key metrics. The table below provides a snapshot of how the primary unit failure rate (λ) affects the system's overall performance and reliability.

Impact of Unit Failure Rate (λ). Availability decreases as λ increases. Higher failure rates also increase mean response and waiting times, while reducing the First-Failure Expectation (FFE$_{VM}$).

Table 1: *Impact of Unit Failure Rate (λ) on Key Metrics*

Failure Rate λ	Availability $A(\infty)$	Resp. Time $\bar{T}$	Wait Time $\bar{T}_w$	FFE$_{VM}$
0.2	0.979	3.314	1.090	36.58
0.4	0.952	3.456	1.286	9.68
0.6	0.925	3.496	1.408	4.91

Impact of Normal Repair Rate (μ_1). Availability, response time, and FFE$_{VM}$ improve significantly as μ_1 increases. Faster repair shortens delays and boosts system durability.

Table 2: *Impact of Normal Repair Rate (μ_1) on Key Metrics*

Repair Rate μ_1	Availability $A(\infty)$	Resp. Time $\bar{T}$	Wait Time $\bar{T}_w$	FFE$_{VM}$
2.0	0.952	3.456	1.286	9.72
3.0	0.965	2.801	1.037	12.13
4.0	0.975	2.386	0.872	14.75

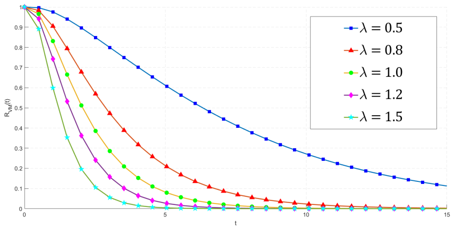
Impact of Repairer Breakdown Rate (θ). Higher breakdown rate θ reduces availability and FFE$_{VM}$, while slightly increasing response and waiting times.

As shown in Figs. 3a–3g, the survival function $R_{VM}(t)$ is sensitive to several parameters. An increase in the failure rate λ causes $R_{VM}(t)$ to decrease more rapidly, while the effect of α

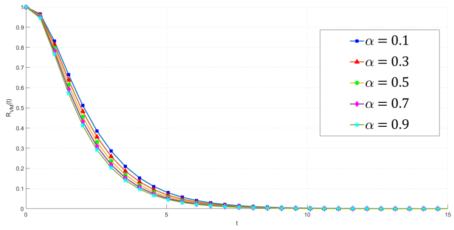
Table 3: *Impact of Repairer Breakdown Rate (θ) on Key Metrics*

Breakdown Rate θ	Availability $A(\infty)$	Resp. Time $\bar{T}$	Wait Time $\bar{T}_w$	FFE_{VM}
0.4	0.960	3.320	1.195	9.71
0.6	0.956	3.415	1.233	9.28
0.8	0.952	3.456	1.286	8.98

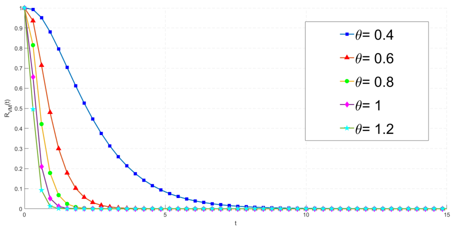
is relatively minor. A higher breakdown rate of the repair server θ accelerates the decline of reliability, highlighting its importance in sustaining system performance. In contrast, higher repair rates μ_1 and μ_2 improve $R_{VM}(t)$, with μ_1 having the stronger effect. Similarly, larger values of γ (retry rate) and σ help maintain higher reliability by reducing waiting and expediting recovery.



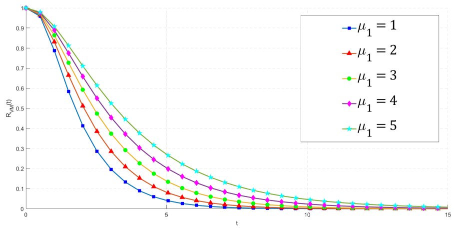
(a) $R_{VM}(t)$ *versus* λ



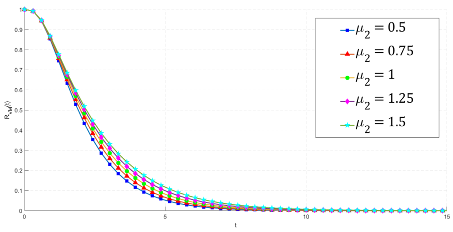
(b) $R_{VM}(t)$ *versus* α



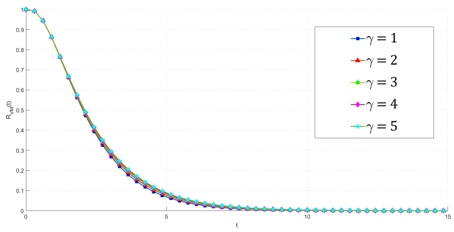
(c) $R_{VM}(t)$ *versus* θ



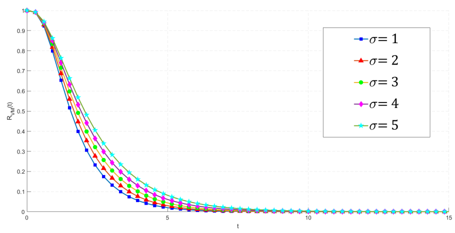
(d) $R_{VM}(t)$ *versus* μ_1



(e) $R_{VM}(t)$ *versus* μ_2



(f) $R_{VM}(t)$ *versus* γ



(g) $R_{VM}(t)$ *versus* σ

Figure 3: *Impact of various parameters on* $R_{VM}(t)$.

7. Conclusion

This study presents a reliability and availability modeling approach for a L-out-of-n system that rely on unit redundancy strategies. A hierarchical framework incorporating both warm and cold standby units was proposed to ensure high availability and service continuity. To evaluate the system's performance, a Generalized Stochastic Petri Net (GSPN) model was developed, which accounted for operational units, standby units, a single repairer, and retrial mechanisms. The results demonstrated that the proposed redundancy and recovery strategies effectively minimize downtime and ensure high system availability, which is crucial for mission-critical applications. The findings highlight that integrating warm and cold standby units with a well-designed repair and retrial process is a powerful approach for enhancing system dependability. Future work will explore the use of Markov Regenerative Stochastic Petri Nets (MRSPNs) to capture more realistic failure and repair behaviors where time distributions are not exponential. This extended framework promises greater accuracy in evaluating the reliability of complex operational systems.

References

[1] Barlow, R.E., & Proshan, F. Statistical Theory of Reliability and Life Testing. Holt, Rinehart & Winston Inc, New York (1975).
[2] David, J.S. Reliability, Maintainability and Risk. Butterworth-Heinemann (2001).
[3] Kececioglu, D. Reliability Engineering Handbook. DEStech Publications, Inc. (2002).
[4] Lai, C.D., & Xie, X. Stochastic Ageing and Dependence for Reliability. Springer (2006).
[5] Birolini, A. Reliability Engineering: Theory and Practice. Springer (2006).
[6] Unuvar, M., Ozguven, E.E., & Pr©kopa, A. Optimal capacity design under k-out-of-n and consecutive k-out-of-n type probabilistic constraints. Annals of Operations Research 226, 643"657 (2020).
[7] Shekhar, C., Jain, M., Raina, A.Z., & Mishra, R.P. Sensitivity analysis of repairable redundant system with switching failure and geometric reneging. Decision Science Letters 6, 337"350 (2017).
[8] Byun, J.E., Noh, H.-M., & Song, J. Reliability growth analysis of k-out-of-N systems using matrix-based system reliability method. Reliability Engineering and System Safety 165, 410"421 (2017).
[9] Eryilmaz, S. The effectiveness of adding cold standby redundancy to a coherent system at system and component levels. Reliability Engineering and System Safety 165, 331"335 (2017).
[10] Chen, W.L. System reliability analysis of retrial machine repair systems with warm standbys and a single server of working breakdown and recovery policy. System Engineering 21, 59"69 (2018).
[11] Chen, W.L., & Wang, K.H. Reliability Analysis of a Retrial Machine Repair Problem with Warm Standbys and a Single Server with N-policy. Reliability Engineering and System Safety 180, 476"486 (2018).
[12] Ke, J.C., Liu, T.H., & Yang, D.Y. Modeling of machine interference problem with unreliable repairman and standbys imperfect switchover. Reliability Engineering and System Safety 174, 12"18 (2018).
[13] Subasi, E., Subasi, M.M., Binmahfoudh, A., & Pr©kopa, A. Sharp bounds for the probability that exactly k-out-of-n events occur with unimodal distributions. Annals of Operations Research. https://doi.org/10.1007/s10479-018-2950-8 (2018).
[14] Yang, D.Y., & Tsao, C.L. Reliability and Availability Analysis of Standby Systems with Working Vacations and Retrial of Failed Components. Reliability Engineering and System Safety 182, 46"55 (2018).
[15] Sarkar, J., & Tang, Z. Comparing limiting availabilities of a k-out-of-n:G system and a k-component series system with $n - k$ spare components. Annals of Operations Research 312, 389"409 (2019).

[16] Yen, T.C., Wang, K.H., & Wu, C.H. Reliability-based measure of a retrial machine repair problem with working breakdowns under the F-policy. Computers & Industrial Engineering 150, 106885 (2020).

[17] Zhao, X., Wu, C., Wang, X., & Sun, J. Reliability analysis of k-out-of-n: F balanced systems with multiple functional sectors. Applied Mathematical Modelling 82, 108"124 (2020).

[18] Gao, S. Availability and reliability analysis of a retrial system with warm standbys and second optional repair service. Communications in Statistics, Theory and Methods 52(4), 1039"1057 (2021).

[19] Gao, S., & Wang, G. Reliability and availability analysis of a retrial system with mixed standbys and an unreliable repair facility. Reliability Engineering and System Safety 205, 107240 (2021).

[20] Li, M., Hu, L., Peng, R., & Bai, Z. Reliability modeling for repairable circular consecutive-K-out-of-n: F systems with retrial feature. Reliability Engineering and System Safety 216, 107957 (2021).

[21] Yang, D.Y., & Wu, C.H. Evaluation of the availability and reliability of a standby repairable system incorporating imperfect switch overs and working breakdowns. Reliability Engineering and System Safety 207, 107366 (2021).

[22] Wang, W., Hu, L., Zhang, T., & Wang, Y. Reliability modeling for a repairable (k_1, k_2)-out-of-n: G system with phase-type vacation time. Applied Mathematical Modelling 91, 311"321 (2021).

[23] Yuan, L. Reliability analysis for a k-out-of-n: G system with redundant dependency and repairmen having multiple vacations. Applied Mathematics and Computation 218, 11959"11969 (2021).

[24] Haas, P.J. Stochastic Petri Nets: Modelling, Stability, Simulation. Springer (2002).

[25] Bause, F., & Kritzinger, P.S. Stochastic Petri Nets: An Introduction to the Theory. Vieweg (2002).

[26] Reisig, W. Understanding Petri Nets: Modeling Techniques, Analysis Methods, Case Studies. Springer (2013).

[27] Ikhlef, L., Lekadir, O., & Assani, D. MRSPN analysis of Semi-Markovian finite source retrial queues. Annals of Operations Research 247, 141"167 (2016).

[28] Hakmi, S., Lekadir, O., & Assani, D. Analysis of priority queue with repeated attempts using generalized stochastic Petri nets. Miskolc Mathematical Notes 20, 925"939 (2019).

[29] YadollahzadehTabari, M., & Mohammadizad, P. Modeling and Performance Evaluation of Energy Consumption in S-MAC Protocol Using Generalized Stochastic Petri Nets. International Journal of Engineering 33, 1114"1121 (2020).