Computing Limiting Average Availability of a Repairable System through Discretization

Debolina Chatterjee*, Jyotirmoy Sarkar

Department of Mathematical Sciences
Indiana University-Purdue University Indianapolis
402 N Blackford Street, Indianapolis, IN 46202-3216, USA

Abstract

Formulas for limiting average availability of a repairable system exist only for some special cases: (1) either the lifetime or the repair time is exponential; or (2) there is one spare unit and one repair facility. We consider a more general setting involving several spare units and several repair facilities; and we allow arbitrary life- and repair time distributions. Under periodic monitoring, which essentially discretizes the time variable, we compute the limiting average availability. The discretization approach closely approximates the existing results in the special cases; and increases the limiting average availability as we include additional spare unit or additional repair facility.

Keywords: Periodic monitoring, Perfect repair, Semi-Markov process, Transition probability, Sojourn time
2010 MSC: 90B25, 62N05, 60J27

1. Introduction

Reliability engineers have always been interested in different techniques to improve the functionality, quality and effectiveness of operating systems. Consequently, availability of a maintained system (that is, the probability that the system is fully functional) is a key quantity of interest. Many heavy industries such as power plants, metal casting, chemical production, space

*Corrosponding author
Email addresses: dchatter@iu.edu (Debolina Chatterjee), jsarkar@iupui.edu (Jyotirmoy Sarkar)

This is the author's manuscript of the article published in final edited form as:
Chatterjee, D., & Sarkar, J. (2020). Computing limiting average availability of a repairable system through discretization. Reliability Engineering & System Safety, 193, 106616. https://doi.org/10.1016/j.ress.2019.106616

administration etc. rely on expensive machineries for production and maintenance. Failure of these machineries is detrimental to the industry, resulting in both economic and logistic challenges. Therefore, the system should be actively maintained by setting up one or more repair facilities and also by keeping one or more back-up spare units to serve as replacement when any damaged/failed unit is sent for repair. Fire detection system, safety valves etc. especially use this kind of spare/repair management. The plan may sound straightforward; but there are many logistical issues to address. For instance, the system has to be monitored continuously to detect failure and switch the operation to the spare unit immediately. Also, one must determine the optimum number of repair facilities that should be established and the optimum number of spare units that should be kept on hand so that the overall system availability is not compromised, and at the same time the cost is within control.

We recall a well-studied model of a repairable system and some known results under that model. However, several restrictive assumptions in this otherwise attractive model severely limits its applicability. Here, we remove these restrictive assumptions by devising a discretization approach, which reduces the burden of monitoring the system continuously, reproduces the results in the known special cases, and extends to the most general setting.

1.1. Formulation of the Problem

Consider a continuously monitored one-unit repairable system supported by several identical spare units and several identical repair facilities. Initially, one unit is put on operation; and all spare units remain on cold standby (that is, spare units cannot fail). Upon failure of the operating unit, instantaneously a spare unit, if available, is put on operation (this is called instantaneous installation to operation) and the failed unit is sent to a repair facility (this is called instantaneous commencement of repair). Repair takes a random amount of time; and after repair the unit is restored back to a level equivalent to a new unit (this is called the perfect repair policy), which becomes a spare unit. We assume that lifetimes and repair times are stochastically independent. The system fails (and enters a down state) when the operating unit fails and there is no spare unit on standby to take over the operation. Thereafter, when at least one of the repairs is completed, the repaired unit is immediately put into operation; and the system is revived.

The most important measure of success of a repairable system is the long run probability that the system is functioning, or the limiting availability of the system. Oftentimes, under continuous life- and repair time distributions and continuous monitoring, the limiting availability exists; and then it equals the limiting average availability, or the limiting proportion of time the system is up; and is given by

$$A_{av} = \frac{MSUT}{MSUT + MSDT} \tag{1.1}$$

where MSUT denotes the mean system up time and MSDT denotes the mean system down time.

In the very special case of exponential lifetime and exponential repair time distributions with means μ and ν respectively, [1] (page 206), provided the limiting average availability for the case of one repair facility ($r = 1$) and either no or one spare unit ($s = 0$ or $s = 1$). More specifically,

$$A_{av}(r = 1, s = 0) = \frac{\mu}{\mu + \nu} = \frac{1/\nu}{1/\nu + 1/\mu} \tag{1.2}$$

since, in this case, in eq. (1.1) MSUT equals the mean time to failure and MSDT equals the mean time to repair; and

$$A_{av}(r = 1, s = 1) = \frac{\mu(\mu + \nu)}{\mu^2 + \mu\nu + \nu^2} = \frac{1/\nu}{1/\nu + 1/\mu - 1/(\mu + \nu)} \tag{1.3}$$

1.2. Availability in some other models

Allowing arbitrary distributions for the lifetime X and the repair time Y, [13] (page 283), derived the limiting average availability of a one-unit system supported by one repair facility and one spare unit as

$$A_{av}(r = 1, s = 1) = \frac{E[X]}{E[\max\{X, Y\}]} \tag{1.4}$$

Indeed, when eq. (1.4) is specialized to exponential life- and repair distributions, one can recover eq. (1.3).

In [9], for a maintained system under continuous monitoring and perfect repair policy, the instantaneous availability is determined using the Fourier transform technique. Here repair time is restricted to exponential, but lifetime is allowed to be either gamma or exponential. Further, using the same technique but incorporating several imperfect repairs before a replacement or a perfect repair, the availability is obtained for exponential lifetime and repair time distributions (with possibly different parameters) in [2].

Assuming periodic inspection, in [11], the system availability is determined when repair is perfect, lifetime is either gamma or exponential and repair time is constant. The work is extended in [3] by allowing an imperfect repair policy and a random repair time (specifically, exponential). Further in [12], a periodically inspected system supported by a spare unit and maintained with perfect repair or upgrade is considered; and both the instantaneous availability and the limiting average availability are determined for arbitrary lifetime, degenerate upgrade time and exponential repair time. The paper [4] adds to the results of [11] by assuming that the periodic inspections take place at fixed time points after repair or replacement in case of failure.

Allowing arbitrary continuous lifetime, but restricting to exponential repair times only, [10] derived the limiting average availability of a one-unit system under continuous monitoring when there are $s \geq 1$ spare units and $r \geq 1$ repair facilities, by studying the embedded Markov chain (tracked at selected observation times), which is said to be in *State* i where $(i = 0, 1, ..., s, s+1)$, if there are i failed units undergoing or awaiting repair by that observation time.

Apart from a one-unit system, availability has been studied also for a k-out-of-N system. For example, the authors of [14] study the interactions among several control variables such as preventive maintenance policy, spare part inventories and repair capacity while they affect the system availability. They present an exact as well as an approximate method to develop a trade-off among these control variables. These authors also advocate in [15] a block replacement policy in which all failed and degraded components are repaired by a single repair shop while spare units take over the operation. They provide two approximate methods to analyze the relation between system availability and control variables. In both papers they assume the component lifetimes and repair times are exponentially distributed.

For a k-out-of-$N : G$ system, [17] and [18] allow the repair time to have a general distribution, but assume the lifetime to be exponential. The former paper considered one repair man with a single vacation, while the latter considered a replaceable repair equipment which may fail during the repair period and then be replaced by a new one. Both papers used supplementary variable technique and Laplace transform to calculate the availability. The supplementary variable technique is implemented in [16] to derive state equations by defining the system state space and sojourn time in each state to calculate the availability of the system.

1.3. Overcoming the challenge

Let us highlight a serious drawback in the models mentioned above to set the stage for our current research. Although not realistic, researchers often assume exponential life- or repair time distribution to simplify mathematical derivations. They exploit the lack of memory property of the exponential distribution to ensure that the successive differences between life- or repair times are independent exponential variables (with different rates), and thereby they obtain closed form expressions for the limiting average availability.

Can we make the model more realistic by allowing arbitrary lifetime and arbitrary repair time distributions for any number of spare units and repair facilities? The challenge of obtaining the limiting average availability under this general setting is expressed in [10] as follows:

> "When repair time distribution is other than exponential, except for the case of $(r = 1, s = 1)$, one must keep track of the time on repair of all failed units at all times. Therefore, there is no hope of identifying an embedded discrete-time Markov chain, and the derivation of the limiting average availability will require a technique different from the one presented in this paper."

Some recent papers allow arbitrary life- and repair time distributions: In [5], the authors studied single-component repairable systems supporting different levels of workloads. They provide a numerical algorithm to evaluate the probability that the system will perform a specified amount of work within a specified mission time, and the associated conditional expected cost. The paper [6] models dynamic performance of multi-state series parallel systems with repairable elements that can function at different load levels and employs a universal generating function technique to assess system performance. Here the instantaneous availability is evaluated at different load levels. Further, in [7], the authors proposed a discrete-state continuous-time stochastic process to evaluate instantaneous availability for a common bus performance sharing (CBPS) system. The technique involves integration with respect to the joint distribution of $< T_j, X_j >$ (where T_j denotes the detection time of the failure of the j^{th} component and X_j denotes the operation time).

The current paper responds to the challenge posed in [10] by adopting a discretization approach: We inspect the system only at discrete time points; and we intervene only when during inspection we find a unit has failed or the failed system is ready for revival because at least one repair has been completed. In particular, if a repair has been completed, but the operating unit has not failed, we do not intervene at all! Thus, this approach essentially discretizes the time variable. Moreover, it relaxes the burden of monitoring the system continuously to monitoring it periodically (at inspection times only); hence, it is logistically preferable.

In Section 2, we revisit the case of $(r = 1, s = 1)$; model the stochastic process through discretization as a semi-Markov process; derive the limiting average availability; and exhibit its closeness to the analytic result (eq. (1.4)) of [13]. In Section 3, we extend the discretization method to the case of $(r = 2, s = 1)$; that is, we permit a second repair facility. Finally, Section 4 concludes the paper with a summary.

2. Discretization approach for $(r = 1, s = 1)$

We assume the following:

(1) Lifetimes of the units are independent and identically distributed (IID) continuous random variables with arbitrary cumulative distribution function (CDF) F on a positive support.

(2) Repair times are IID continuous random variables with arbitrary CDF G on a positive support.

(3) Lifetimes and repair times are stochastically independent.

(4) Repair is perfect; that is, a repaired unit is as good as new.

(5) The system is under periodic monitoring; that is, it is inspected at regular intervals.

(6) Interventions are made only at observation epochs when an operating unit is found to have failed or when the down system is ready for revival because at least one failed unit has been repaired.

(7) Whenever at inspection a unit is found to have failed, it is sent to the repair facility. Repair commences instantaneously if the facility is free. Otherwise, the failed unit awaits repair until the facility is free.

(8) Installation to operation happens immediately when a failed unit is sent to repair (at an inspection epoch) and there is a spare unit (as a result of an already completed repair), or when the failed system is ready for revival at an inspection epoch because one of the failed units has been repaired.

2.1. States of the system

Figure 1 depicts the states of the system (with explanations below), transition between them and the random variables determining such transitions.

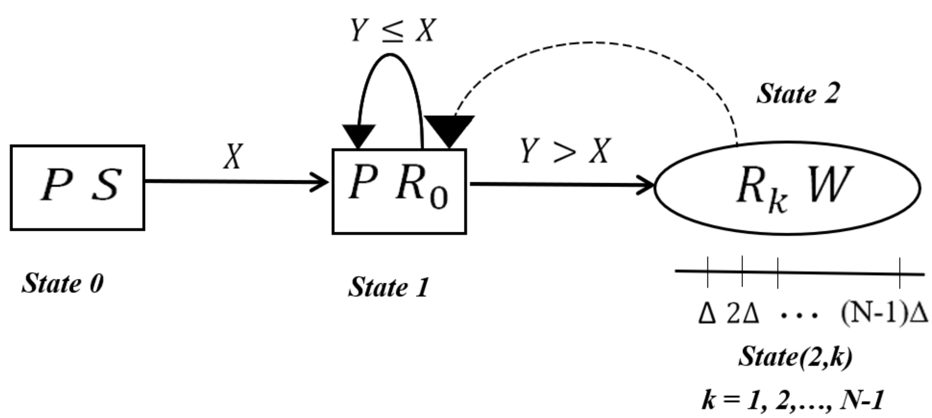

Figure 1: The state transition diagram for the $(r = 1, s = 1)$ case. A rectangle denotes an up state, and an oval denotes a down state. The status of each unit is denoted as follows: P for operation; S for stand-by; R for repair (with subscript indicating for how many inspection periods the repair has been going on); and W for waiting for repair.

We label the states of the system to indicate the number of failed units:

(0) *State 0* means there is no failed unit.

(1) *State 1* means there is one failed unit.

(2) *State 2* means there are two failed units. Additionally, we must use a second index to indicate how long the repair on the first failed unit has been going on when the system enters *State 2*, because that will determine how long the system will stay in *State 2*. This second index splits *State 2* into sub-states: We say the system is in *State* $(2, k)$ for

$k = 1, 2, \ldots, N-1$, if repair on the first failed unit has been going on for a duration $k\Delta$ when the other unit was detected to have failed. This is because we monitor the system only at epochs that are multiples of Δ from the start (or from system revival).

Note that by the time the system is detected to have failed, repair on the first failed unit has been going on for a positive duration. Hence, there is no *State* $(2, 0)$. Also, repair is surely completed in $N\Delta$ duration. Hence, there is no *State* $(2, N)$.

Let F and G denote the CDFs of the discretized lifetime and repair time X and Y respectively. Let p and q denote the corresponding probability mass functions (PMFs) calculated by taking successive differences $p_k = F(k\,\Delta) - F((k-1)\,\Delta)$ and $q_k = G(k\,\Delta) - G((k-1)\,\Delta)$ respectively, for $k = 1, 2, \ldots, N$. Let R denote the CDF of $\max\{X, Y\}$ calculated by taking product $R(k\,\Delta) = F(k\,\Delta)\,G(k\,\Delta)$, and let r denote the corresponding PMF of $\max\{X, Y\}$ obtained by successive differences $r_k = R(k\,\Delta) - R((k-1)\,\Delta)$ for $k = 1, 2, \ldots, N$.

We describe the transition probabilities between states of the system:

- At time $t = 0$, the system is in *State 0*, where one unit begins to operate and the other spare unit is on cold standby. The system goes from *State 0* to *State 1* when the operating unit is detected to have failed, repair starts on it and the spare unit is put on operation instantaneously. Hence,
$$P_{0\to 1} = 1 \tag{2.1}$$
The system never returns to *State 0*.

- From *State 1*, after an intervention, the system can go to two places: (i) If repair on the failed unit is completed before the operating unit is detected to have failed, then we do not record this transition at all. Instead, we wait until the operating unit is detected to have failed at epoch $k\Delta$. Then we interchange the roles of the two units; and say that the the system has re-entered *State 1*. This happens with probability
$$P_{1\to 1} = \sum_{k=1}^{N} p_k\; G(k\Delta) \tag{2.2}$$

(ii) On the other hand, if the operating unit is detected to have failed at epoch $k\Delta$, before the repair on the previously failed unit is completed, then the system goes to *State (2,k)* with probability

$$P_{1\to(2,k)} = p_k\{1 - G(k\Delta)\} \tag{2.3}$$

In this case, the freshly failed unit awaits repair to commence on it only after the repair on the previously failed unit is found to be completed at an inspection epoch. While the system is in *State 2* (that is, in any of the *States (2,k)*), no unit is operating; and the system is down.

- From *State (2,k)* the system surely goes to *State 1* when the ongoing repair on the first failed unit is found to be completed at an inspection time and the repair on the second failed unit begins. This happens with probability

$$P_{(2,k)\to 1} = 1 \tag{2.4}$$

In the proposed discretization approach, we split the repair time into N (to be determined momentarily) intervals each of length Δ; and observe the system at epochs $k\Delta$ for $k = 1, 2, \ldots, N$. For all practical purposes, we assume that repair is completed only at epochs $k\Delta$, since those are the observation epochs (and possible installation epochs).

We choose N large enough so that the probability that the larger of lifetime and repair time (hence, either lifetime or repair time) exceeds $N\Delta$ is very small (preferably under .001, say); that is, $\{1 - R(N\Delta)\} \approx .001$.

The continuous-time stochastic process, after discretization, can be described as a Semi-Markov Process: The probability distribution of the future state depends only on the current state (and not on the history of states visited so far); and the system stays in any state for a random duration whose distribution depends on the current state and the immediately next state.

Moreover, from the above discussion of transitions and associated probabilities, we note that the embedded discrete-time Markov chain is irreducible (that is, all states communicate with one another); and since the state space is finite, the chain is recurrent.

Using the theory of semi-Markov processes, see [8], we can find the limiting proportion of time the system spends in each state. First, we find the stationary probabilities $\{\pi_j, j \in S\}$ of the discrete-time Markov chain by solving the following state equations:

$$\pi_j = \sum_{i \in S} \pi_i P_{ij}, \text{ for all } j \in S; \text{ and } \sum_{j \in S} \pi_j = 1 \tag{2.5}$$

where P_{ij} denotes the transition probability from *State* $i \in S$ to *State* $j \in S$ and the transition probability matrix P, which is of dimension $(N+1) \times (N+1)$, is as follows:

$$P = \begin{array}{c} \begin{array}{ccccc} 0 & 1 & (2,1) & \dots & (2,N-1) \end{array} \\ \begin{pmatrix} P_{0,0} & P_{0,1} & P_{0,(2,1)} & \cdots & P_{0,(2,N-1)} \\ P_{1,0} & P_{1,1} & P_{1,(2,1)} & \cdots & P_{1,(2,N-1)} \\ P_{(2,1),0} & P_{(2,1),1} & P_{(2,1),(2,1)} & \cdots & P_{(2,1),(2,N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_{(2,N-1),0} & P_{(2,N-1),1} & P_{(2,N-1),(2,1)} & \cdots & P_{(2,N-1),(2,N-1)} \end{pmatrix} \end{array} \begin{array}{c} \\ 0 \\ 1 \\ (2,1) \\ \vdots \\ (2,N-1) \end{array}$$

$$= \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & * & * & \cdots & * \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & 0 & \cdots & 0 \end{bmatrix}$$

In the P-matrix above, the row and the column labels stand for the corresponding states. Note that although the transition matrix P is $(N+1) \times (N+1)$, it has non-zero entries (denoted by *) only in the second row corresponding to transition out of *State 1* and in the second column corresponding to transition into *State 1*. Therefore, it is straight-forward to solve eq. (2.5).

Second, we find the expected sojourn time in each state; that is, the expected time the system stays in that state before it moves to a new state. If a unit is found to have failed at inspection time $k\Delta$, it must have failed during the interval $((k-1)\Delta, k\Delta]$. For simplicity, we assume that it has failed at the midpoint of the interval; that is, it was operating for the initial $\Delta/2$ period in the interval and was in failed state during the last $\Delta/2$ period (but was undetected). Although this is a rather crude assumption, it serves our purpose as far as computation of limiting average availability is concerned.

The expected sojourn times μ_0 and μ_1 in *State 0* and *State 1* respectively, both equal $E(X) - \Delta/2 = \sum_{k=1}^{N} p_k \, k\Delta - \Delta/2$, since we do not record a repair until after the operating unit fails. We subtract $\Delta/2$ from the expected discretized lifetime to account for the fact that the system is actually down during the last $\Delta/2$ duration within each *State 0* and *State 1*.

The expected sojourn time $\mu_{(2,k)}$ in any *State* $(2, k)$ (a down state), is the expected additional repair time, given that the previously failed item has been undergoing repair for $k\Delta$ time. For $k = 1, 2, \ldots, N$, we have,

$$\mu_{(2,k)} = E[Y|Y > k] = \sum_{j=1}^{N-k} \frac{q_{j+k}\ j\Delta}{1 - G(k\Delta)} \tag{2.6}$$

There is no need to make a further adjustment of $\Delta/2$ in eq. (2.6) as the system is down the whole time while in *State (2,k)*.

Next, using Corollary to Proposition 4.8.1 of [8], the limiting probability that the stochastic process will be found in *State* j (where j runs over all N States $1, (2, 1), (2, 2), \ldots (2, N - 1)$) is independent of the initial state and is given by

$$\theta_j = \frac{\pi_j \mu_j}{\sum_{i=1}^{N} \pi_i \mu_i} \tag{2.7}$$

The denominator $\sum_{i=1}^{N} \pi_i \mu_i$ in (2.7) is called the **expected cycle time**; and it is the expected time between successive renewals (or entry into *State* 1). Having calculated all θ_j's, we define $\theta_2 = \theta_{(2,1)} + \cdots + \theta_{(2,N-1)} = 1 - \theta_1$, since *State 2* is the aggregate of *States* $(2, 1), (2, 2), \ldots, (2, N - 1)$.

Since the system is up in *States* 0 *and* 1, and down in *State* 2, but the system never returns to *States* 0, the limiting average availability of the system is given by

$$A_{av} = 1 - \theta_2 = \theta_1 \tag{2.8}$$

2.2. Computation and comparison

We want to compare the limiting average availability computed by eq. (2.8) under discretization approach to the value computed by eq. (1.4) under continuous monitoring. As a test case, let us assume a Weibull(shape=3, scale=1.12) lifetime distribution with mean lifetime 1, and a Weibull(shape=2, scale=2) repair time distribution with mean repair time 1.77.

Under discretization approach, since $F(12)G(12) < .001$, we decompose the time range $(0, 12]$ into $N = 120$ intervals of length $\Delta = 0.1$ each. We construct the CDFs of discretized life- and repair times, F and G, from the above mentioned Weibull distributions evaluated at $k\,\Delta$ for $k = 1, 2, \ldots, 120$. We construct the PMFs p, q, r as defined above by successive differences.

Using equations (2.1 - 2.4), we construct the transition probability matrix P, which in this case is of dimension 121×121. Recall from above that P has non-zero entries only in row 2 and column 2. Below we partially display the second row rounding each entry to 3 decimal places; all other entries of the second column are 1.

$$P = \begin{array}{c} \begin{array}{cccccccc} 0 & 1 & (2,1) & (2,2) & (2,3) & (2,4) & \ldots & (2,N-1) \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & .252 & .001 & .005 & .013 & .024 & \cdots & * \\ 0 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 \end{pmatrix} \end{array} \begin{array}{c} \\ 0 \\ 1 \\ (2,1) \\ \vdots \\ (2,N-2) \\ (2,N-1) \end{array}$$

Next, we calculate the stationary probabilities using (2.5): We find $\pi_0 = 0, \pi_1 = 0.572$; and for *State* $(2,k)$*'s* (for $k = 1, 2, \ldots, N-1$), the stationary probabilities, rounded to 4 decimal places, are: $\{\pi_{(2,1)}, \pi_{(2,2)}, \ldots, \pi_{(2,N-1)}\}$ $= \{.0004, .0028, .0075, .0140, .0218, .0300, .0375, .0433, .0464, .0465, .0435,$ $.0381, .0311, .0237, .0167, .0110, .0066, .0037, .0019, .0009, .0004, .0001,$ $.0001, 0, 0, 0, \ldots, 0\}$.

Lastly, the expected sojourn times in *State* 0 and *State* 1 are both obtained from $E(X) - \Delta/2 = \sum_{k=1}^{N} p_k\, k\Delta - \Delta/2$ as $\mu_0 = \mu_1 = 10.0014$. Likewise, for *State* $(2,k)$'s (for $k = 1, 2, \ldots, N-1$), we get the expected sojourn times (rounded to 4 decimal places) as $\{\mu_{(2,1)}, \mu_{(2,2)}, \ldots, \mu_{(2,N-2)}, \mu_{(2,N-1)}\}$ $= \{17.2677, 16.3901, 15.5837, \ldots, 1.4000, 1.000\}$.

Therefore, $\theta_0 = 0$ and θ_j's for $j = 1, (2,1), \ldots, (2, N-1)$ are calculated using eq. (2.7). In particular, $\theta_2 = 0.4665872$, and the expected cycle time $\sum_{i=1}^{N} \pi_i \mu_i = 10.72444$. Moreover, using eq. (2.8), the limiting availability to be $\theta_1 = 1 - \theta_2 = .5334128$.

Two comments follow: (1) The exact analytic result, given in (1.4), yields the limiting availability to be .5334131. Thus, our discretization approach closely approximates the analytic result previously derived by [13]. (2) For the case $(r = 1, s = 1)$, the limiting average availability is .53341, while for the case $(r = 1, s = 0)$, using eq. (1.1), the limiting average availability is only $1/2.77 = .361$. Thus, there is a significant increase (47.76%) in A_{av} with the introduction of a spare unit.

For $(r = 1, s = 1)$, having established the test case of Weibull life- and Weibull repair times, we carry out a more comprehensive study of various combinations of life- and repair time distributions, always ensuring mean lifetime=1 and mean repair time=1.77. We report in Table 1 the limiting average availability using both the analytic formula and the discretization approach. We extend the time range to $(0, 20]$ so that $F(20)G(20) < 0.001$, but we keep $\Delta = 0.1$, implying that there are 201 states.

Repair time / Lifetime	Exponential (1/1.77)	Gamma (2, 0.855)	Weibull (2, 2)
Weibull (3, 1.12)	.49341 *.49335*	.52055 *.52055*	.53341 *.53341*
Gamma (2, 0.5)	.48172 *.48167*	.50413 *.50413*	.51515 *.51515*
Inverse-Gauss (1 , 1)	.47221 *.47215*	.49058 *.49057*	.49867 *.49904*
Exponential (1)	**.46971** ***.46926***	.48787 *.48746*	.49677 *.49638*
Lognormal (-0.5 , 1)	.46263 *.46452*	.47865 *.48062*	.48946 *.48902*

Table 1: Availability under different life- and repair time distributions for the $(r = 1, s = 1)$ case. The top entry of each cell is the availability computed through discretization and the bottom entry using eq. (1.4).

Highlighted in the table is the special case when both life- and repair time distributions are exponential. The analytic result for this case is already given in [1](page 206), [13](page 283) and [10](Corollary 2.2). Here we demonstrate that the result of the discretization approach (.46971) closely approximates the analytic result (.46926). The slight discrepancy is due to crudely subtracting $\Delta/2$ from the expected sojourn times of the system up states; *State 0* and *State 1*.

To increase limiting average availability we have allowed a spare unit to take over the operation when the main unit has failed and is under repair. Of course, when there is only one repair facility (that is, $r = 1$), then when the system is down only the first failed unit is under repair while the other failed unit is awaiting repair. In order to improve the limiting average availability of the system, one strategy is to introduce one more repair facility to expedite the repair of the second failed unit. However, when there are multiple repair facilities, no analytic result exists in the literature to allow both life- and repair time distributions to be arbitrary. The close agreement between the values of eq. (1.4) and eq. (2.8) gives us confidence to proceed with the discretization approach in case $r > 1$.

3. The discretization approach for $(r = 2, s = 1)$

Having justified the discretization approach when $(r = 1, s = 1)$, we proceed to apply it to the case of a second repair facility, where no analytic result is available. Here, $(r = 2, s = 1)$; that is, there are one operating unit, one identical spare unit and two identical repair facilities.

3.1. The states of the system

Figure 2 shows the states of the system (with explanations below), transitions between them and the random variables determining the transitions.

Initially, the system is in *State 0*, where one unit begins to operate and the other unit is on cold standby. We write the state-space of the system in two different notation—using one or two indices—depending on the level of details required for the analysis:

$$S = \{0; 1; 2^+; 1^+\} = \{0; (1, 0); (2, 1), \ldots, (2, N-1); (1, 1), \ldots, (1, N-1)\}$$

where the first index i denotes how many units have been detected to have failed and are under repair, and the second index j tells us how long the repair

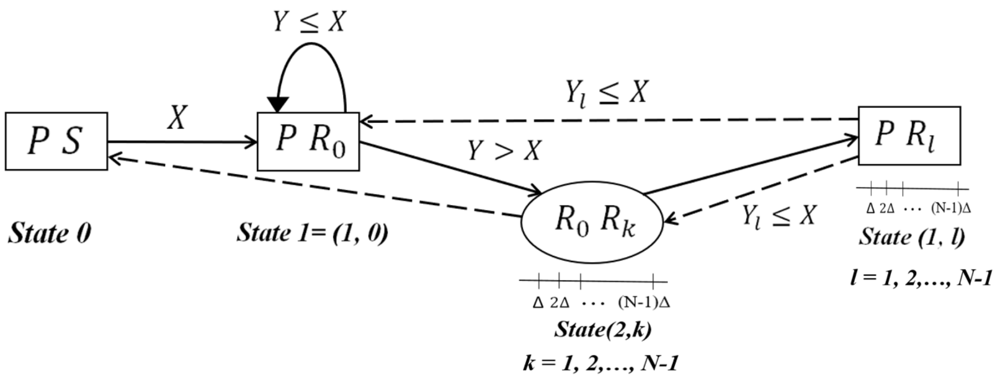

Figure 2: The state transition diagram for the $(r = 2, s = 1)$ case. The notation are the same as in Figure 1.

on the first failed unit has been going on when the repair on the second failed unit just starts.

Let us explain the state space notation in terms of several examples:

- *State* $1 = (1, 0)$ means that one unit has been detected to have failed; it has been placed on repair just now, so that its repair duration so far is 0; and the other unit has just been placed on operation.

- Note that there is no *State* $(2, 0)$ because by the time failure on the second unit is detected, the repair on the first failed unit has already started and it has been going on for a positive multiple of Δ. Also, there is no *State* $(2, N)$ because if repair has been going on for duration $N\Delta$, it must have been completed. Likewise, there is no *State* $(1, N)$.

- *State* $(2, 5)$ (provided, of course, $N > 5$) means that the system just entered *State 2* (that is, both units are known to have failed); repair on the first failed unit has been going on for 5Δ periods; and repair on the second failed unit has just started.

- *State* $(1, 7)$ (provided, of course, $N > 7$) means that repair on the only failed unit has been going on for 7Δ periods when the other unit is just put on operation (hence, there is only one failed unit).

Recall that we only record those inspection epochs when a failure is detected or when a down system is ready for revival because at least one unit has been repaired. In particular, we do not record epochs when a repair is completed, but the other unit is still operating.

Next, let us write down the recorded transitions between states and the associated transition probabilities. Recall that we monitor the system only at epochs $\Delta, 2\Delta, 3\Delta, \ldots$. As in the case of $(r = 1, s = 1)$, we assume that X is the discretized lifetime with CDF F and PMF p; and Y is the discretized repair time having CDF G and PMF q. Also, we choose N such that the larger of life- and repair times exceeds $N\Delta$ with probability at most .001.

- From *State 0*, the system surely goes to *State 1=(1,0)* after a random lifetime having PMF p. Therefore,

$$P_{0\to(1,0)} = 1 \tag{3.1}$$

- From *State 1=(1,0)*, if the operating unit is still functioning at epoch $k\,\Delta$, we do nothing. But if the operating unit is found to have failed at epoch $k\Delta$, then it must have failed in the interval $((k-1)\Delta, k\Delta]$, which happens with probability $p_k = F(k\,\Delta) - F((k-1)\,\Delta)$. There are two distinct cases to consider:

  (i) Repair is already completed by epoch $k\Delta$ (that is, repair is finished sometime during $(0, k\,\Delta]$), which happens with probability $\mathrm{P}\{Y \leq k\Delta\}$=$G(k\Delta)$. In this case, interchange the roles of the two units—the repaired unit takes over the operation and the failed unit is put on repair. Hence, the system re-enters *State 1=(1,0)*. Hence,

$$P_{(1,0)\to(1,0)} = \sum_{k=1}^{N} p_k\, G(k\,\Delta) \tag{3.2}$$

  (ii) Repair is not completed by epoch $k\Delta$, which happens with probability $\mathrm{P}\{Y \leq k\Delta\}$=$G(k\Delta)$. In this case, the system goes down, since both units have failed and there is no other spare unit to take over operation. More specifically, the system enters *State* $(2, k)$. Hence,

$$P_{(1,0)\to(2,k)} = p_k\,\{1 - G(k\,\Delta)\} \tag{3.3}$$

- When the system enters *State* $(2, k)$, we continue to observe the system at regular intervals of Δ, labeling those epochs as $(k+1)\Delta, (k+2)\Delta, \ldots$. Two distinct cases are possible:

(i) Both failed units are repaired during the same time interval, say, $((k+j-1)\Delta, (k+j)\Delta]$, where $j = 1, 2, \ldots, N-k$. To find the probability of this case happening, add over all j the product of two independent probabilities: Given that the repair of the first failed unit was not completed by time $k\Delta$, the conditional probability that it is completed during $((k+j-1)\Delta, (k+j)\Delta]$ is $\frac{q_{k+j}}{1-G(k\Delta)}$. The probability that the second failed unit on which repair started at epoch $k\Delta$ is repaired during the same time interval as the first failed unit is q_j. Finally, note that in this case, one of the repaired units (it does not matter which one, since the two units are identical) is put on operation and the other becomes a standby spare; that is, the system enters *State 0*. Therefore,

$$P_{(2,k)\to 0} = \sum_{j=1}^{N-k} q_j \, \frac{q_{k+j}}{1 - G(k\Delta)} \tag{3.4}$$

(ii) One of the repairs is completed, but not the other. In this case, the repaired unit is put on operation; and the repair on the other unit, which has been going on for $l\Delta$ time, continues on, causing the system to enter *State* $(1, l)$. The meaning of l is explained below in two sub-cases depending on which repair is completed—repair on the first failed unit, or repair on the second failed unit.

- (a) Suppose that the first failed unit, on which the repair has been going on for $k\Delta$ time, is repaired earlier; and it happens during interval $((k+l-1)\Delta, (k+l)\Delta]$. The conditional probability of this event is $\frac{q_{k+l}}{1-G(k\Delta)}$. The probability that the second failed unit, on which repair had started freshly at epoch $k\Delta$, will not be repaired within the additional $l\Delta$ duration is $\mathrm{P}\{Y > l\Delta\} = 1 - G(l\Delta)$.
- (b) Suppose that the second failed unit, on which repair started at epoch $k\Delta$, gets repaired earlier; and it happens during interval $((l-1)\Delta, l\Delta]$, which has probability q_{l-k}. Then the conditional probability that the first failed unit will not be repaired by epoch $l\Delta$, given that the repair was not completed by epoch $k\Delta$, is $\frac{1-G(l\Delta)}{1-G(k\Delta)}$.

  Combining the two sub-cases (a) and (b), we have

$$P_{(2,k)\to(1,l)} = [q_{k+l} + q_{l-k}] \left\{ \frac{1 - G(l\Delta)}{1 - G(k\Delta)} \right\} \tag{3.5}$$

where we interpret $q_t = 0$, unless $1 \le t \le N$.

- From *State (1,l)*, the system can go to one of two directions:

  (i) If repair is completed before the operating unit fails, we do not record that transition; instead, we wait until the operating unit fails, say during interval $((j-1)\,\Delta, j\Delta]$ (for $j = 1, 2, \ldots, N$), with probability p_j, and the system goes to *State (1,0)*. The conditional probability that repair is completed before this additional time $j\Delta$, given that the repair was not completed by time $l\Delta$, is $\frac{G((l+j)\Delta)-G(l\Delta)}{1-G(l\Delta)}$. Hence,

$$P_{(1,l)\to(1,0)} = \sum_{j=1}^{N} p_j \left\{ \frac{G((l+j)\Delta) - G(l\Delta)}{1 - G(l\Delta)} \right\} \tag{3.6}$$

  where we interpret $G(t\Delta) = 1$, whenever $t \ge N$.

  (ii) If the operating unit fails during interval $((k-l-1)\Delta, (k-l)\Delta]$, which happens with probability p_{k-l}, before repair of the failed unit is completed, then the system goes down and enters *State* $(2,k)$, where $k > l$. Given that the ongoing repair is not completed by time $l\,\Delta$, the conditional probability that the repair will not be completed in additional time $(k-l)\,\Delta$ (that is, by epoch $k\,\Delta$) is $\frac{1-G(k\Delta)}{1-G(l\Delta)}$. Hence,

$$P_{(1,l)\to(2,k)} = p_{k-l} \left\{ \frac{1 - G(k\Delta)}{1 - G(l\Delta)} \right\} \tag{3.7}$$

  for $k > l$.

Considering all the above state transition, the transition probability matrix P is of dimension $2N \times 2N$ and has the following structure:

$$P = \begin{array}{c} \begin{array}{cccccccc} 0 & 1 & (2,1) & \ldots & (2,N-1) & (1,1) & \ldots & (1,N-1) \end{array} \\ \left(\begin{array}{cccccccc} 0 & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & * & * & \cdots & * & 0 & \cdots & 0 \\ * & 0 & 0 & \cdots & 0 & * & \cdots & * \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ * & 0 & 0 & \cdots & 0 & * & \cdots & * \\ 0 & * & * & \cdots & * & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & * & * & \cdots & * & 0 & \cdots & 0 \end{array}\right) \end{array} \begin{array}{c} \\ 0 \\ 1 \\ (2,1) \\ \vdots \\ (2,N-1) \\ (1,1) \\ \vdots \\ (1,N-1) \end{array}$$

The row and column labels in above matrix represent the corresponding states. As in the case of $(r = 1, s = 1)$, here also the continuous-time stochastic process, after discretization, is a Semi-Markov Process. Hence, the analysis follows along similar lines.

First, we find the stationary probabilities $\{\pi_j, j \in S\}$ of the discrete-time Markov chain by solving the state equations that are similar in structure to eq. (2.5), but involve many more states.

Second, we find the expected sojourn time in each state. In fact, the expected sojourn times $\mu_0, \mu_{(1,0)}$ and $\mu_{(1,l)}$ in *States* $0, (1,0), (1,l)$, for $1 \leq l \leq N-1$, are all equal to $E(X) - \Delta/2 = \sum_{k=1}^{N} p_k\, k\Delta - \Delta/2$. [The subtraction of $\Delta/2$ accounts for the system being down during the last $\Delta/2$ duration within each state $0, (1,0), (1,l)$.] The expected sojourn time $\mu_{(2,k)}$ in *State* $(2,k)$ (a down state) is the expected value of the minimum of the two repair times Y_0 and Y_k having CDFs $G(j)$ and $\frac{G(k+j)-G(k)}{1-G(k)}$ for $0 \leq j \leq N$ (with $G(t) = 1$ for $t > N$) respectively. Using Problem 1.1 of [8], this expectation can be found as the sum of the survival function evaluated at non-negative integers. That is, for $k = 1, 2, \ldots, N$, we have

$$\begin{aligned} \mu_{(2,k)} = E[\min\{Y_0, Y_k\}] &= \sum_{j=0}^{N} P\{Y_0 \geq j, Y_k \geq j\} \\ &= \sum_{j=0}^{N-k} \frac{[1 - G(j\Delta)][1 - G((k+j)\Delta)]}{1 - G(k\Delta)}. \end{aligned} \tag{3.8}$$

Here, there is no need to make an additional adjustment of $\Delta/2$ as the system is down throughout the time it is in *State (2,k)*.

Next, using Corollary to Proposition 4.8.1 of [8], the limiting probability that the stochastic process will be found in *State* j is independent of the initial state and is given by expressions of the form (eq. (2.7)), but with many more states. Let us define *State* 1^+ as aggregate of *States* $(1,1), (1,2), \ldots, (1, N-1)$ and *State* 2 as aggregate of *States* $(2,1), (2,2), \ldots, (2, N-1)$.

Having calculated all θ_j's, we define $\theta_2 = \theta_{(2,1)} + \cdots + \theta_{(2,N-1)}$. Since the system is up in *States* $0, 1, 1^+$, and down in *State* 2, all states being recurrent, the limiting average availability of the system is given by

$$A_{av} = 1 - \theta_2. \tag{3.9}$$

3.2. Computations and comparison

We compute the limiting average availability for various life- and repair time distributions, always choosing mean lifetime 1 and mean repair time 1.77. We have truncated all distributions to have support $[0, 12]$, which we have partitioned into 120 equal sub-intervals; that is, we choose $\Delta = 0.1$. Consequently, there are 240 states in the state space S.

The transition probability matrix P is 240×240, whose entries, using equations (3.1 - 3.7) and rounded to 4 decimal places, are partially displayed:

$$
P = \begin{array}{c}
\begin{array}{cccccccc}
0 & 1 & (2,1) & \dots & (2,N-1) & (1,1) & \dots & (1,N-1)
\end{array} \\
\left(\begin{array}{cccccccc}
0 & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & .2517 & .0007 & \cdots & * & 0 & \cdots & 0 \\
.3213 & 0 & 0 & \cdots & 0 & .0075 & \cdots & * \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
.0025 & 0 & 0 & \cdots & 0 & .9975 & \cdots & 0 \\
0 & .2875 & * & \cdots & * & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & .9997 & 0 & \cdots & .0003 & 0 & \cdots & 0
\end{array}\right)
\end{array}
\begin{array}{c}
\\ 0 \\ 1 \\ (2,1) \\ \vdots \\ (2,N-1) \\ (1,1) \\ \vdots \\ (1,N-1)
\end{array}
$$

The stationary probabilities are obtained by using eq. (2.5). They are $\pi_0 = .010$, $\pi_{(1,0)} = .265$, and
$\{\pi_{(2,1)}, \pi_{(2,2)}, \pi_{(2,3)}, \dots, \pi_{(2,N-2)}, \pi_{(2,N-1)}\} = \{.0002, .0013, .0035, \dots, 0, 0\}$.

The expected sojourn times in *State* 0, *State* $(1, 0)$ and *State* $(1, l)$ for $l = 1, 2, \dots, N-1$ are all equal to 10.0016. And, using eq. (3.8)
$\{\mu_{(2,1)}, \mu_{(2,2)}, \mu_{(2,3)} \dots, \mu_{(2,N-2)}, \mu_{(2,N-1)}\} = \{12.549, 12.093, 11.665, .., 1.399, 1\}$.

Next, using eq. (2.7), we see that the limiting probabilities that the stochastic process will stay in a *State* j, for $j \in S$ are respectively $\theta_0 = .0106$, $\theta_{(1,0)} = .2794$, $\theta_{1^+} = .3764$ and $\theta_2 = .4666$. Also, the expected cycle time is 9.493. Finally, using eq. (3.9), the limiting average availability is obtained as .66650.

Furthermore, in Table 2, we display the limiting average availability calculated for the same set of life- and repair times as in the case $(r = 1, s = 1)$ and the percentage improvement when $(r = 2, s = 1)$.

Table 2 exhibits about 25-35% increase in limiting average availability when a second repair facility is included in the presence of one spare unit.

Life-time \ Repair-time	Exponential (1/1.77)	Gamma (2, 0.855)	Weibull (2 , 2)
Weibull (3, 1.12)	.65807 *33.37*	.66392 *27.54*	.66650 *24.94*
Gamma (2, 0.5)	.64764 *34.44*	.65057 *29.05*	.65171 *26.51*
Inverse-Gauss (1 , 1)	.63903 *35.33*	.63992 *30.44*	.63943 *28.23*
Exponential (1)	**.63676** ***35.56***	.63718 *30.61*	.63693 *28.21*
Lognormal (-0.5 , 1)	.63024 *36.23*	.63009 *31.64*	.62537 *29.88*

Table 2: We compare the limiting average availability between cases $(r = 1, s = 1)$ and $(r = 2, s = 1)$. The top entry in each cell is the computed availability for $(r = 2, s = 1)$; and the bottom entry is the percentage increase in availability compared to the $(r = 1, s = 1)$ case given in Table 1.

4. Conclusion

Recall from Section 2 that our discretization approach closely approximates the analytic result for the $(r = 1, s = 1)$ case. Also, from Section 3 we note that for the $(r = 2, s = 1)$ case under exponential life- and exponential repair times, the analytic result of [10], yields a limiting average availabiltiy of 0.63871, while our discretization approach using eq. (2.8) gives a limiting average availability of .63676. Hence, we claim that the discretization approach works reasonably well; and it can be used to compute the limiting average availability for *any* life- and repair time distributions. We also find that as we increase an additional spare unit from $(r = 1, s = 0)$ to $(r = 1, s = 1)$ or as we add an additional repair facility from $(r = 1, s = 1)$ to $(r = 2, s = 1)$ there is significant increase in the limiting average availability of the system.

Obviously, we anticipate a further increase in limiting average availability when the number of spare units and/or the number of repair facilities is increased. Of course, inclusion of an additional spare unit or an additional repair facility will invariably lead to an increase in the number of states and therefore inflate the computational burden. Nonetheless, the discretization approach will continue to yield the limiting average availability under any arbitrary continuous life- and repair time distributions for other systems as well. For example, in future we plan to extend the discretization method to study a k-out-of-$N : G$ system.

Thus, our main contribution in this paper is to provide a simple computational technique by utilizing the discretization approach that allows us to incorporate any arbitrary life- and repair time distributions as well as increase the number of repair facilities and/or the number of spare units.

5. Acknowledgement

We sincerely thank three reviewers for their valuable suggestions and advice. Not only their recommendations improved the readability and relevance of our work, but also helped us gather some resources for future research.

References

[1] Barlow, R.E., Proschan, F., 1975. Statistical theory of reliability and life testing: probability models. Technical Report. Florida State Univ Tallahassee.

[2] Biswas, A., Sarkar, J., 2000. Availability of a system maintained through several imperfect repairs before a replacement or a perfect repair. Statistics & Probability Letters 50, 105–114. URL: https://doi.org/10.1016/S0167-7152(00)00087-0.

[3] Biswas, A., Sarkar, J., Sarkar, S., 2003. Availability of a periodically inspected system, maintained under an imperfect-repair policy. IEEE Transactions on Reliability 52, 311–318. URL: https://ieeexplore.ieee.org/document/1248648.

[4] Cui, L., Xie, M., 2005. Availability of a periodically inspected system with random repair or replacement times. Journal of Statistical Planning and Inference 131, 89–100. URL: https://doi.org/10.1016/j.jspi.2003.12.008.

[5] Levitin, G., Xing, L., Dai, Y., 2015. Optimal loading of system with random repair time. European Journal of Operational Research 247, 137–143. URL: https://doi.org/10.1016/j.ejor.2015.05.033.

[6] Levitin, G., Xing, L., Dai, Y., 2017. Optimal loading of series parallel systems with arbitrary element time-to-failure and time-to-repair distributions. Reliability Engineering & System Safety 164, 34–44. URL: https://doi.org/10.1016/j.ress.2017.02.008.

[7] Levitin, G., Xing, L., Huang, H.Z., 2019. Dynamic availability and performance deficiency of common bus systems with imperfectly repairable components. Reliability Engineering & System Safety 189, 58–66. URL: https://doi.org/10.1016/j.ress.2019.04.007.

[8] Ross, S.M., Kelly, J.J., Sullivan, R.J., Perry, W.J., Mercer, D., Davis, R.M., Washburn, T.D., Sager, E.V., Boyce, J.B., Bristow, V.L., 1996. Stochastic processes. volume 2. Wiley New York.

[9] Sarkar, J., Chaudhuri, G., 1999. Availability of a system with gamma life and exponential repair time under a perfect repair policy. Statistics & Probability Letters 43, 189–196. URL: https://doi.org/10.1016/S0167-7152(98)00259-4.

[10] Sarkar, J., Li, F., 2006. Limiting average availability of a system supported by several spares and several repair facilities. Statistics & probability letters 76, 1965–1974. URL: https://doi.org/10.1016/j.spl.2006.04.046.

[11] Sarkar, J., Sarkar, S., 2000. Availability of a periodically inspected system under perfect repair. Journal of Statistical Planning and Inference 91, 77–90. URL: https://doi.org/10.1016/S0378-3758(00)00128-2.

[12] Sarkar, J., Sarkar, S., 2001. Availability of a periodically inspected system supported by a spare unit, under perfect repair or perfect upgrade. Statistics & probability letters 53, 207–217. URL: https://doi.org/10.1016/S0167-7152(01)00087-6.

[13] Sen, P.K., Bhattacharjee, M.C., 1984. Nonparametric Estimators of Availability Under Provisions of Spare and Repair. University of North Carolina at Chapel Hill. Institute of Statistics.

[14] de Smidt-Destombes, K.S., van der Heijden, M.C., van Harten, A., 2004. On the availability of a k-out-of-n system given limited spares

and repair capacity under a condition based maintenance strategy. Reliability engineering & System safety 83, 287–300. URL: https://doi.org/10.1016/j.ress.2003.10.004.

[15] de Smidt-Destombes, K.S., van der Heijden, M.C., van Harten, A., 2007. Availability of k-out-of-n systems under block replacement sharing limited spares and repair capacity. International Journal of Production Economics 107, 404–421. URL: https://doi.org/10.1016/j.ijpe.2006.08.013.

[16] Wang, N., Li, M., Xiao, B., Ma, L., 2018. Availability analysis of a general time distribution system with the consideration of maintenance and spares. Reliability Engineering & System Safety URL: https://doi.org/10.1016/j.ress.2018.06.025.

[17] Wu, W., Tang, Y., Yu, M., Jiang, Y., 2014. Reliability analysis of a k-out-of-n: G repairable system with single vacation. Applied Mathematical Modelling 38, 6075–6097. URL: https://doi.org/10.1016/j.apm.2014.05.020.

[18] Wu, W., Tang, Y., Yu, M., Jiang, Y., Liu, H., 2018. Reliability analysis of ak-out-of-n: G system with general repair times and replaceable repair equipment. Quality Technology & Quantitative Management 15, 274–300. URL: https://doi.org/10.1080/16843703.2016.1226712.