Information Sciences Letters
An International Journal

http://dx.doi.org/10.18576/isl/120205

Availability Modeling and Analysis Based on Repair Policy for Selector Distribution Unit (SDU) Platform

M. Malkawi[1,], O. Murad[2] and M. Kadhum[3]*

[1]Department of Software Engineering, College of ICT, Jordan University of Science and Technology, Jordan
[2]Association of Arab Universities, Amman, Jordan
[3]Department of Biomedical Engineering, Ashur University College, Bagdad, Iraq

Received: 2 Jun. 2022, Revised: 12 Jul. 2022, Accepted: 18 Sep. 2022
Published online: 1 Feb. 2023

Abstract: This paper presents the impact of repair policies on the availability of a selector distribution unit (SDU) in a telecommunication system. Three repair policies are analyzed, One-Shot Repair Policy, Sequential Repair Policy, and Simultaneous Repair Policy. The availability of the SDU unit is measured using each of the investigated repair policies. Failure rates and average availability are used as primary performance indexes. The results show that the impact of repair policy is minimal when the number of faults is relatively low. When the number of faults to be repaired is relatively large, the on-site simultaneous repair policy outperforms the other policies. The results were obtained using fault MEADEP reliability and availability analysis tool. The data used for measuring the performance of the repair policies are obtained from environment and practical numbers. The investigated SDU has availability budget of 0.99999933, which is required to achieve a 5 NINES availability for the entire telecom network.

Keywords: Availability, SDU, Reliability, Repair policies.

1. Introduction

1.1 Availability modeling

Availability modeling is a process that translates the availability design features of a particular system into a mathematical or logical model in order to evaluate the ability of the system to provide its services as and when required [1, 2, 3]. The availability of a system is a product of two major sub-models. One sub-model defines how the system breaks or gets in an outage state. The other sub-model defines how to deal with an outage, particularly how to repair a system in order to restore its full functionality [3, 4, 5]. [4][5]. A general view of the activities involved in availability modeling and the availability design features that the model should translate are depicted in Figure 1 [3, 5, 8].

In this paper, we provide a general view of availability models for a selection distribution unit (SDU) [6, 7] , which is one of core units in telecommunication infrastructure system [5, 8]. However, we will emphasize the impact of repair policies on the overall availability of SDU system. We will use the modeling tool (MEADEP) and a simulation method, e.g. UltraSan will be considered for further analysis. Both tools are designed for Markov chain process analysis and/or simulation [9, 10].

There are two conventional approaches to reliability and availability prediction: 1) modeling of a system in the design phase, or 2) assessment of the system at a later phase, typically by test or relying on the field data from already installed systems. The first approach relies on probabilistic models that use component level failure rates [10] published in handbooks or supplied by the manufacturers. This approach provides an early indication of system reliability, but the model as well as the underlying data later need to be validated by actual measurements. The second approach typically uses test data and reliability growth models [9]. It involves fewer assumptions than the first, but it can be very costly. The higher the reliability specified for a system, the longer the required test or the larger number of systems have to be tested. A further difficulty arises in the translation of reliability data obtained by test into those applicable to the operational environment.

The greatest benefit of availability modeling comes when its results are used at the early design stage. In that case the deficiencies in system architecture may be identified and corrected at relatively low cost and with a few side effects.

*Corresponding author e-mail: mimalkawi@just.edu.jo

Fig. 1. Availability Modeling Work Products

1.2 Definitions

In this paper, we will adopt and use the following definitions and terminology for modeling purposes [3] [3][5].

- Availability: Probability that a system is available to perform its function, measured in number of NINES.
- Fault: A defect, imperfection or flaw that occurs in hardware or software
- Error: The occurrence of an incorrect value in some unit of information within a system. Error is a manifestation of a fault.
- Failure: A deviation in the expected performance of a system (Outages and Service Disruptions)
- Detection: Recognizing that a fault/error has occurred Containment/Isolation: Isolating a fault and preventing its effects from propagating throughout a system.
- Recovery: Restoring the system to a stable (operational) state.
- Availability / 5NINES Requirements: The requirements for a system, e.g., SDU platform to meet allocated availability requirements derived from the goal of achieving the 5NINES system availability. The actual numbers will be shown in the following sections. The specific end-to- end system availability models should focus on successes/failures of end-to-end transaction events.

For example, in a 5-NINES system, the general availability requirement for end-to-end systems specifies that System Availability must be at least 99.999%.

- Equipped Capacity: The total capacity installed in the system.
- Provisioned Capacity: Equipped capacity minus the spare capacity for redundancy (i.e., cards that provide redundancy is not counted in the provisioned capacity).
- Total SDU Outage: Total loss of the provisioned capacity for more than N seconds (e.g., 30 seconds), due to causes affecting the availability of the system.
- Partial SDU Outage: Loss of 66% or more of the provisioned capacity for more than 30 seconds, due to causes affecting the availability of the system.
- Downtime: Time for Total Outages + (Time for Partial Outages X percentage of provisioned capacity lost.).
- Service disruption: Any outage that is less than 30 seconds (provisioned capacity loss).

Note that the total and partial outage Measurements are necessary for an accurate view of system availability. However, the automatic generation of these outage measurements (which involve the Network Manager) is subject to several constraints.

In particular, the measurements must be coordinated with other network equipment outages to avoid double counting of outages.

Also, the outages must not be included in the measurements until the network equipment is cut over to commercial service. Furthermore, the system must be aware of provisioned capacity to properly calculate partial outages

1.3 Literature review

In the context of reliability for telecommunication systems, numerous areas of focus in the setting of system reliability.

Some works address the requirement for reliability at the edge both for source and destinations. For example in[11] at various network layers, explain how can be decreasing delay that can impact latency. Such as at physical layer, explain the effect of frame size network latency. Another work that development an algorithm that improving retransmission delay when faulty packets occur. On the other hand, in [12][13] different algorithmic approaches were studded , such as multi connectivity, machine learning, and algorithm based on neural network at the receiver and transmitter to enhancing reliability due to channel failures.

Another area of focus is in the study of the reliability of wireless links in general.

Another area of focus is in the study of the reliability of wireless links in general. [14][15], where introduce a framework for modeling, predicting and analyzing the theoretical reliability and availability of a wireless link. A new framework based on the reliability block diagram of the wireless link is established, and its reliability data (e.g. the Bit Error Rate (BER) or Packet Error Rate (PER)) are predicted by the part stress analysis method. The framework is motivated to achieve the functionality of successfully transmitting a packet of size n within a given deadline while accounting for two factors: channel fading and retransmissions [16].

In [14] Dealt with the renewed process as a repair index () and wireless transmission case failure (), and demonstrated reliability modelling and analyses of wireless transmission as a repairable system by Using the standard Markov chains.

The Network Element for Mobile Switching Center Server (MSS) and Telecommunication Application Server (TAS) deployed in a cloud environment is reviewed in terms of availability and dependability. The calculation was done through the use of a reliability block diagram and a simulation method. The N+M was caused as a result of all this work; a matters N of components has been backed up by another number M of additional components. Furthermore, the factor

M =N will be necessary to accomplish six nine's availability [17].

In [18] the authors Deployment of the tool chine for fault-tolerant communications system to overcome the limitations of existing approaches such as Reliability block diagrams (RBD) where each component can be between two states (active or failed). A tool has been created to perform numerical Markov.

Reward (MRMs) analysis. The dependability in a cellular situation in [8] was calculated using data from manual failure logs. The authors also present estimates for the mean time between failures and the mean time to repair, demonstrating that both follow a Weibull or two-stage hyper-exponential distribution. Furthermore, network failures are thought to be reliant on changes and work performed in the network, rather than being wholly random. These assumptions are backed up by the results of the tests. While in [14] the authors use a semi-Markov transition approach to build renewal models for normal and degraded network connection states based on long-term dependability data such as link availability. The amount of time network links spend in normal states is characterized as link reliability, and the network has two transition actions i.e., failure and repair exist in the network. The authors find that lower reliability states are highly absorbing states where the link is sufficiently degraded i.e., failure (λ) rate larger than repair (μ). To provide reliability analysis. In [19] researchers applied multiple machine learning techniques like as neural networks, NBC, and SVM, as well as a Continuous Time Markov Chain (CTMC) analytical model. It has forecast fault occurrences by collecting historical data and estimating the maximum likelihood of the next fault site using Bayesian Network prediction models to predict the next fault using real network data. The results suggest that Deep Neural Networks with Auto encoders were the most successful of the strategies studied.

2. Methodologies

2.1. Modeling and Analysis Tool

One of the first activities required for the system availability modeling is evaluation and selection of the modeling method, and consequently the tools to be used for modeling. The tools and methods depend on the complexity of the model and the level of details that are required for the model. Three levels of modeling complexity are recognized in this work. The first level aims at obtaining a quick view of the system availability parameters, without incorporating repair and recovery techniques. A simple spreadsheet method is sufficient for this purpose. [20][21].

The second level of modeling incorporates detailed knowledge about the various components of the system, the recovery rate, the recovery and repair modes, and the various sensitivity parameters [22]. This level requires the use of a more advanced modeling tool. The White smoke modeling group has decided to use a tool known as Measurement of Availability and Dependability **(MEADEP**) [9]. The most powerful feature of MEADEP is that it allows the decomposition of a complex system into simple hierarchical design of the availability components. It combines reliability block diagrams and Markov processes. For more information on MEADEP, A third level of modeling may become required if the system decomposition is no longer possible under MEADEP [23][24]. Also, for transient state analysis and simulation, another level of modeling is required. For this purpose, UltraSan (a stochastic activity networks based tool) is being studied and considered by the modeling team() [10].

2.2. Reliability and Redundancy Models

The High Availability of the SDU [6] subsystem is achieved through the use of highly reliable components and implementation of physical and logical redundancy. The reliability of individual components is reflected in the relatively large mean time to failure (MTTF). [25] For software, there are well defined methods for building robust and reliable software components. The authors of this document will address the subject of software reliability and availability modeling in a separate document [26]. In order to achieve high availability, several forms of redundancy may be used:[27]

1. 2N redundancy: Every component is backed up by another, identical component. If N components are sufficient for capacity, another N components are provided for increased availability. Two variations of 2N redundancy are used:

- **Active/Active:** Both components are used simultaneously. When one fails, the second component picks up the capacity lost either due to the failure, or the system experiences a 50% outage.
- **Active/Standby:** The redundant component dues not carry any load while the primary component is running and active. If the primary component fails, the standby component provides service in its place. Success depends upon the availability of the standby at the time of transition, i.e., upon the standby experiencing no faults while waiting, and on the successful fail over to the stand by component. The fail over is mainly software dependent parameter, except in some cases where the hardware can provide fail over mechanism. The fail over is characterized by the success rate, and the duration of the fail over operation. Depending on the length of the duration, the system may be subject to outage (complete or partial) or disruption of service.

2) N+1 Redundancy: N components are used to provide service to the system load, while one additional component is in a standby (backup) mode. When an active component fails, the standby is brought into service. Several variations exist for the N+x mode of redundancy:

- N-Pool: N active elements form a pool of resources that provides service at full capacity. As N decreases, a loss of capacity may occur. A total outage is reached when the number of active elements goes below a predetermined threshold. This case is known as N-out-of-M, where M = N+x.
- N distinct resources: Each element provides a service that cannot be substituted by other elements. All N elements are required for operation. Redundancy is achieved with (x) standby elements.

The SDU modeling in this paper includes all of the above redundancy schemes.

2.3 Repair Policy

The repair policy defines how a hardware or software component is repaired once a failure has occurred. Repair is defined as the process, which allows a component to return to a state that is at least partially functional. That description captures the semantics of automatic recovery, like through a software reboot or application restart, and the semantics of manual recovery, like through a technician site visit.[28]. Different recovery paths can exist within a single repair policy. For example, it may be possible to recover from a software error automatically, but it may not be possible to recover from hardware errors without a manual service call [29]. We considered three different policies: one-shot, sequential, and simultaneous. The repairs in reliability engineering is defined by the μ [repairs/hour].

2.3.1. One-Shot Repair Policy

This repair policy describes the case when all components are repaired in one site visit. The assumption is that the repair engineer will travel to the site once the first failure occurs. While at the site, the repair engineer will fix all faulty equipment, including any components that fail while he is making repairs. The mean time to repair (four hours) assumes an average case of three hours of travel time, and one hour for actual repair. This time is assumed to be enough for all repairs. [30]

2.3.2. Sequential Repair Policy

This policy describes the sequential repair of each failed element. As elements are repaired one by one they are returned to service. The mean time to repair increases with the number of failed components, but an average value of four hours is assumed [31].

2.3.3. Simultaneous Repair Policy

In this policy, all failed components are repaired at the same time, each by a different engineer. It was considered for comparison purposes only, since it is obviously not a realistic scenario [31].

3. Results

We evaluate the system availability under the three repair policies with the following set of assumptions:

1. actual repair time to fix a problem is equal to one hour
2. travel time of the on-site maintenance personnel is negligible
3. travel time of the off-site maintenance personnel is equal to three hours
4. One shot repair policy requires one travel to fix N possible faults
5. Simultaneous repair policy implies that N repairmen repair N faults
6. Sequential repair policy assumes an individual visit of a repairman to fix each fault.

Since we are designing a high availability system, we will use the simplified definition of system availability to estimate the impact of different repair policies [32]

$$A \approx 1 - \frac{MTTR}{MTTF} = 1 - \frac{failure\_rate}{repair\_rate} \tag{1}$$

The simplified definition of system availability was obtained from the classical system availability definition

$$A = \frac{MTTF}{MTTF + MTTR} \tag{2}$$

By applying to it the expansion in series for the case:

$$\frac{MTTR}{MTTF} \prec\prec 1.$$

One shot repair policy (all N faults repaired during one travel) takes the following time to fix all N faults

$$Tr(N) := Tt + Trt \cdot N \tag{3}$$

Where:

Tt – time to travel, and Trt – time to repair a single fault.

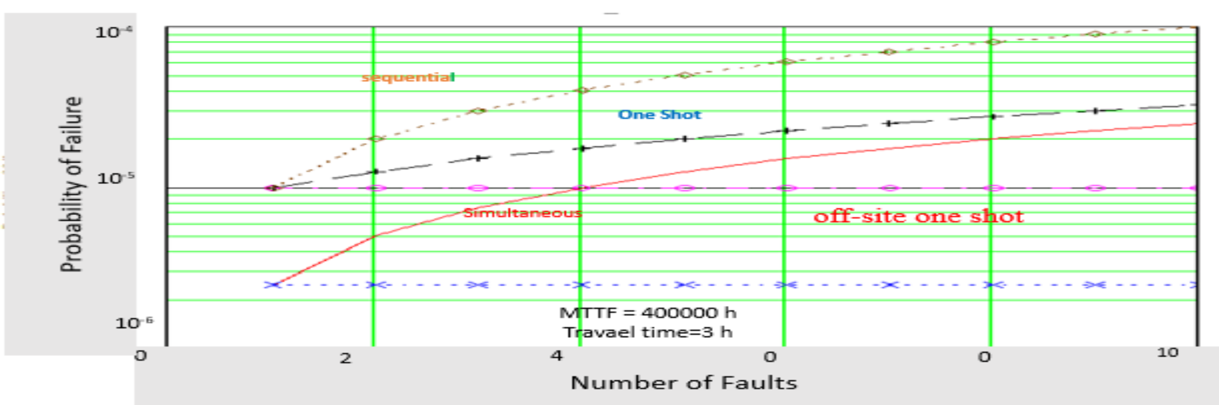

Fig. 2. probability of failure vs. number of faults of faults fixed according to different recovery procedures.

For the most of faults we will assume Trt = 1 (hour), average travel time for the off-site personnel Tt = 3 hours and negligible time Tt = 0 for the on-site personnel. It is more efficient to compare repair policies looking at their impact on the probability of system failure

Probability_of system_failure = 1 – System_availability (4)

Figure 2 shows the set of graphs corresponding to different repair policies for the cases of having on- and off-site maintenance personnel. The typical number of faults is just one. Just to see how critical the failure rate to the number of faults is, we extended a range of the number of faults to ten. As shown by Figure 2, having the on-site maintenance personnel will add about 0.000007 or more to system availability for the given system MTTF. Figure 2 shows the impact of different repair strategies for the fixed set of repair times of one and four hours (see Eq (3) for N = 1). Figure 3 helps to see how critical system availability numbers are to different repair times or to its inverse value - repair rates. The plots of Figure 2 represent the case of fixing a single fault.

4. Discussion

In general, when the ratio of the repair rate to the failure rate is relatively large the difference between all policies is negligible and the system availability is on the order of a few nines. Naturally, that statement is valid only for the most typical single fault case. When, the ratio of the repair rate to the failure rate is relatively low we can see that both the one shot repair and the simultaneous repair provide very close availability numbers. The simultaneous repair is slightly better than the one-shot repair. However, due to the high cost of deploying more than one engineer to a site, we recommend to use the one-shot repair policy [33].

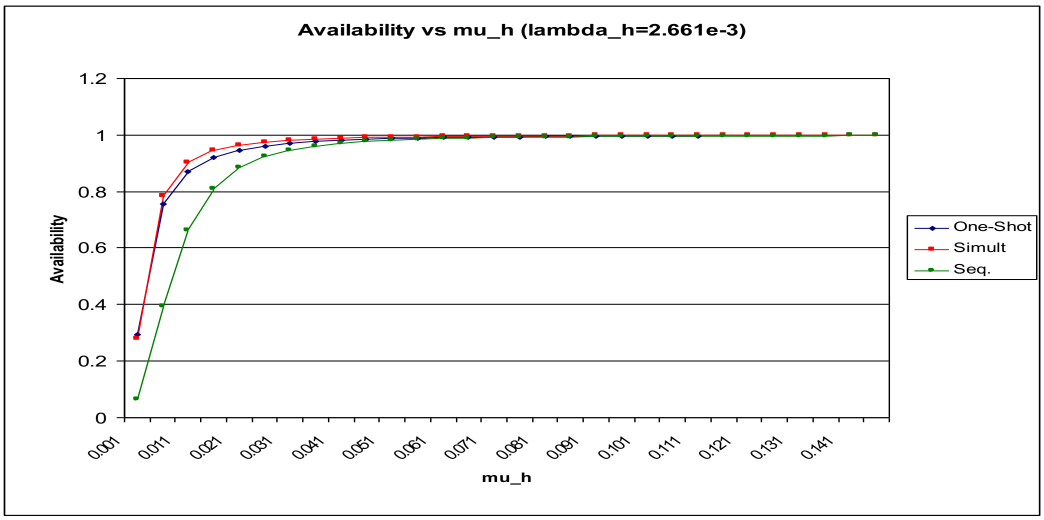

Fig. 3. System availability vs. repair rate μ for the system having failure rate λ

The results of the availability analysis, described in this report, take into account only possible hardware (HW) failures. To meet the assigned SDU availability requirements of 0.99999933, derived from the goal of attaining five nines system availability, it was assumed an equal contribution of faults between HW, SW and procedural errors. It means that HW availability has to be not less than 0.99999978 to meet the SDU availability requirements. Analysis of the current proposed SDU architecture and its components revealed that the current architecture is capable to meet the availability goal only for the high coverage value of fault detection for the SDU configuration having at least two shelves. The fault coverage value has to be not less than 95% while the system recovery time has to be within one hour.

5. Recommendations

A single shelf SDU is not capable to meet availability budget allocation numbers. In general, if N shelves are meeting performance requirements, to meet high availability requirement an SDU system has to be configured as N+1 shelves. The most efficient repair policy, when the MTTF is relatively low is the simultaneous repair policy. We recommend using the policy for systems to have been reliable enough with relatively low MTTF value. For less reliable systems, with larger MTTF values, it is more cost effective to us sequential repair, which is less costly repair policy.

Acknowledgment

The authors would like to extend special thanks to the high availability team at Motorola, where the research was conducted.

Conflict of interest

The authors declare that there is no conflict regarding the publication of this paper.

References

[1] R. Kaur, A. L. Sangal, and K. Kumar, “Modeling and simulation of adaptive neuro-fuzzy based intelligent system for predictive stabilization in structured overlay networks,” *Eng. Sci. Technol. an Int. J.*, vol. 20, no. 1, pp. 310–320, (2017).

[2] L. Ortigoza-Guerrero, “Performance Analysis of Fractional Guard Channel Policies in Mobile Cellular Networks,” *IEEE Trans. Wirel. Commun.*, vol. 5, no. 2, pp. 301–305, doi: 10.1109/TWC.2006.1611053 (2006).

[3] M. I. Malkawi, “The art of software systems development: Reliability, Availability, Maintainability, Performance (RAMP),” *Human-centric Comput. Inf. Sci.*, vol. 3, no. 1, pp. 1–17, Dec., doi: 10.1186/2192-1962-3-22/TABLES/2 (2013).

[4] E. Marcus and H. Stern, “Blueprints for high availability,” p. 587, (2003).

[5] G. Ignatius, B. Moore, M. Malkawi, and L. Votta, “Availability Work Products–A Strategic Approach,” (2001).

[6] M. H. Mohammad, “Cellular diagnostic systems using hidden Markov models.” Virginia Tech, 2006.

[7] A. Mishra, “Architecture and performance of a network-processor-based frame selection and distribution unit (SDU) for CDMA-2000,” in *Technologies, Protocols, and Services for Next-Generation Internet*, vol. 4527, pp. 37–41 (2001).

[8] S. M. Matz, L. G. Votta, and M. Malkawi, “Analysis of failure and recovery rates in a wireless telecommunications system,” *Proc. 2002 Int. Conf. Dependable Syst. Networks*, pp. 687–693, doi: 10.1109/DSN.2002.1029014 (2002,).

[9] D. Tang, M. Hecht, J. Miller, and J. Handal, “MEADEP: A dependability evaluation tool for engineers,” *IEEE Trans. Reliab.*, vol. 47, no. 4, pp. 443–450, (1998).

[10] [W. H. Sanders, W. D. Obal II, M. A. Qureshi, and F. K. Widjanarko, “The UltraSAN modeling environment,” *Perform. Eval.*, vol. 24, no. 1–2, pp. 89–115, (1995).

[11] Z. Ma, M. Xiao, Y. Xiao, Z. Pang, H. V. Poor, and B. Vucetic, “High-reliability and low-latency wireless communication for internet of things: challenges, fundamentals, and enabling technologies,” *IEEE Internet Things J.*, vol. 6, no. 5, pp. 7946–7970, (2019).

[12] M. S. Elbamby *et al.*, “Wireless edge computing with latency and reliability guarantees,” *Proc. IEEE*, vol. 107, no. 8, pp. 1717–1737, (2019).

[13] K. He, Z. Wang, D. Li, F. Zhu, and L. Fan, “Ultra-reliable MU-MIMO detector based on deep learning for 5G/B5G-enabled IoT,” *Phys. Commun.*, vol. 43, p. 101181, (2020).

[14] R. Sattiraju, P. Chakraborty, and H. D. Schotten, “Reliability analysis of a wireless transmission as a repairable system,” in *2014 IEEE Globecom Workshops (GC Wkshps)*, pp. 1397–1401(2014).

[15] R. Sattiraju and H. D. Schotten, “Reliability modeling, analysis and prediction of wireless mobile communications,” in *2014 IEEE 79th Vehicular Technology Conference (VTC Spring)*, pp. 1–6 (2014).

[16] O. Hiari, R. Mesleh, and N. Aljanini, “A Reliability Analysis Framework for Space Modulation Techniques,” *IEEE Trans. Commun.*, vol. 69, no. 7, pp. 4795–4806, (2021).

[17] A. Hilt, G. Járó, and I. Bakos, “Availability prediction of telecommunication application servers deployed on cloud,” 2016.

[18] K. Lampka, M. Siegle, and M. Walter, “An easy-to-use, efficient tool-chain to analyze the availability of telecommunication equipment,” in *International Workshop on Parallel and Distributed Methods in Verification*, , pp. 35–50 (2006).

[19] Y. Kumar, H. Farooq, and A. Imran, “Fault prediction and reliability analysis in a real cellular network,” in *2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*, pp. 1090–1095 (201).

[20] S. Bahri and H. Ben Bacha, “A study of asymptotic availability modeling for a Failure and a repair rates following a Gamma distribution,” (2009).

[21] Q. Qiu and L. Cui, “Reliability evaluation based on a dependent two-stage failure process with competing failures,” *Appl. Math. Model.*, vol. 64, pp. 699–712, (2018).

[22] L. Junliang, C. Yueliang, Z. Yong, Z. Zhuzhu, and F. Weijie, “Availability modelling for periodically inspected

systems under mixed maintenance policies," *J. Syst. Eng. Electron.*, vol. 32, no. 3, pp. 722–730, (2021).

[23] R. B. Jadhav, S. D. Joshi, U. G. Thorat, and A. S. Joshi, "Software Defect Prediction Utilizing Deterministic and Probabilistic Approach for Optimizing Performance through Defect Association Learning," *Int. J.*, vol. 8, no. 6, (2020).

[24] M. Cinque, D. Cotroneo, and A. Pecchia, "Event logs for the analysis of software failures: A rule-based approach," *IEEE Trans. Softw. Eng.*, vol. 39, no. 6, pp. 806–821, (2012).

[25] T. Nakagawa, "Studies on Reliability and Maintenance," *Reliab. Model. With Appl. Essays Honor Profr. Toshio Nakagawa His 70th Birthd.*, p. 349, (2013).

[26] T. Nakagawa, "Two-unit redundant models," *Stoch. Model. Reliab. Maint.*, pp. 165–191, (2002).

[27] T. Nakagawa, "Redundant Systems," in *Stochastic Processes*, Springer, pp. 199–217(2011).

[28] Q. Qiu, L. Cui, and D. Kong, "Availability and maintenance modeling for a two-component system with dependent failures over a finite time horizon," *Proc. Inst. Mech. Eng. Part O J. Risk Reliab.*, vol. 233, no. 2, pp. 200–210, (2019).

[29] T. Nakagawa and S. Mizutani, "A summary of maintenance policies for a finite interval," *Reliab. Eng. Syst. Saf.*, vol. 94, no. 1, pp. 89–96, (2009).

[30] W. Yun, Y. Han, and H. Kim, "Simulation-based inspection policies for a one-shot system in storage over a finite time span," *Commun. Stat. Comput.*, vol. 43, no. 8, pp. 1979–2003, (2014).

[31] T. Nakagawa, S. Mizutani, and M. Chen, "A summary of periodic and random inspection policies," *Reliab. Eng. Syst. Saf.*, vol. 95, no. 8, pp. 906–911, (2010).

[32] J. Barabady, "Improvement of system availability using reliability and maintainability analysis." Luleå tekniska universitet, (2005).

[33] J. Barabady and U. Kumar, "Availability allocation through importance measures," *Int. J. Qual. Reliab. Manag.*, (2007).