

Article

The k -out-of- n : G System Viewed as a Multi-Server Queue

Achyutha Krishnamoorthy ^{1,*} , Anu Nuthan Joshua ^{2,*}  and Ambily P. Mathew ¹ ¹ Department of Mathematics, CMS College Kottayam, Kottayam 686001, India² Department of Mathematics, Union Christian College, Aluva 683102, India

* Correspondence: krishnamoorthy@cmscollege.ac.in (A.K.); anunuthanjoshua@uccollege.edu.in (A.N.J.)

† Current address: Department of Mathematics, Central University of Kerala, Kasargod 671316, India.

Abstract: This paper extends the k -out-of- n : G reliability system to a multi-server queue. We study a multi-server reliability-queuing model with the N -policy of repair. The queuing system considered here has n servers, each of which has identically and exponentially distributed service times with parameter μ . Servers are subject to breakdown at an exponential rate γ . The repair process follows the N -policy of repair. Although these servers work independently of each other, service can be provided only when k functional servers are available in the system. We study the model in the steady state, using the matrix analytic method. We evaluate some associated performance measures and provide graphical/numerical illustrations. We consider an optimization problem, and the results of the study are presented.

Keywords: reliability; k -out of n system; N -policy; HAP system; breakdown

MSC: 60K25; 60K10; 60K20; 90B22; 90B25



Citation: Krishnamoorthy, A.; Joshua, A.N.; Mathew, A.P. The k -out-of- n : G System Viewed as a Multi-Server Queue. *Mathematics* **2024**, *12*, 210. <https://doi.org/10.3390/math12020210>

Academic Editors: Gurami Tsitsiashvili, Alexander Bochkov and Alexander Zeifman

Received: 11 November 2023

Revised: 8 December 2023

Accepted: 2 January 2024

Published: 8 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

This paper extends the reliability system k -out-of- n : G to a multi-server queue (n parallel server, with server failure and N -policy of repair). As in the k -out-of- n : G system, at least k servers should be operational in order for the whole system to work (system can provide service). At the moment that the number of operational servers is reduced to $k - 1$, the service is disrupted. As a result, the system can start operation only after one working server is added to the system. This model can be applied directly to the high-altitude platform (HAP) system used for smooth telecommunication when all other facilities fail, due to natural calamities. The system collapses when the number of operational servers is reduced to $k - 1$, irrespective of the number of customers in the system. As with parallel (at least one operational component should be UP for the system to function)/serial (all components should be operational for the system to be functional) systems in the reliability context, we encounter a similar situation in the queuing model discussed in this paper. Thus, this paper extends the k -out-of- n reliability system to the queuing system and simultaneously to the classical multi-server queuing system. In the proposed queuing model, an element of dependency on the number of operational servers arises. Thus, it differs from the classical multi-server queuing models in that at least k ($k = 1$, for the classical multi-server system with server failures and their repair) servers should be operational to provide a service. This phenomenon can be interpreted as the system becoming overloaded when the number of operational servers reduces to $k - 1$.

The mathematical foundation of reliability was laid by the pioneering work of Barlow and Proschan (1965) [1]. The element of interest is the reliability of the machine in a given interval of time or failure-free operation up to time t . Naturally, one investigates the distribution of the number of failed components at any random time when the machine is in operation. The k -out-of- n system is classified as COLD, WARM or HOT, depending on whether the operational components are also subject to deterioration when the system is

down. In the COLD system, no operational components fail when the system is DOWN (not operational); for the WARM system, the operational components fail at a reduced rate when the system is DOWN compared to that when it is UP (operational); and in the case of a HOT system, the failure rate of components is the same, irrespective of whether the system is UP or DOWN. As such, for k -out-of- n systems, the notion of “providing service” was not in vogue until Krishnamoorthy, Sathian and Viswanath (2016) [2] introduced it. They considered separately two distinct cases of providing service. In one of them, the system serves (repairs) failed machines in an organization. The k -out-of- n system is considered as a single server subject to failure. Therefore, at most, one failed machine undergoes repair by the server at a time. The repair of failed components is according to some specific policy. In this case, the system state remains finite because it is assumed that there are only a limited number of machines in the organization. In the second model, the authors assumed that when the server (the k -out-of- n system) is idle, it provides a service to external customers. However, priority is given to the repair of internal customers (for example, failed machines in the organization). For these two cases, the authors investigated continuous-time server reliability.

The N -policy was introduced by Yadin and Naor in [3]. The N -policy of repair was introduced by Krishnamoorthy et al. (2002) [4] and is as follows: when the number of operational components comes down to level N ($k \leq N \leq n$), a repair facility starts repairing failed units, one at a time. This repair process continues until all failed components are brought back to an operational state. It is assumed that the components repaired are as good as new. This assumption is essential for mathematical tractability. There is another means of keeping the system reliability high—placing an order for $n - k + 1$ components when the number of operational components goes down to N . It takes a certain amount of time, deterministic or random, for the order to materialize. Sometimes, this may take place only after the system becomes non-operational. The higher the N value, the higher the probability of the system remaining UP until the order materialization. We can further improve the system reliability by combining the N -policy of repair and the order placement for new components. The order can be placed above the level at which the number of operating components drops to N or we can place the order at this level or even below this level. This policy can be further modified for the cancellation of a placed order if the number of operational components reaches the maximum n through the repair process before the materialization of the order placed. Thus, various extensions are possible. N -policy queuing systems with server breakdowns (both working and non-working breakdowns) are studied extensively in the queuing literature (for example, [5–7]).

Other than the N -policy of repair, for policies such as D (the accumulated work load reaching or crossing the level D (D is a continuous random variable)) and T , the time until the repair facility is activated after the system completely returns to a fully functional state (all components are in an operational state due to repair), are used. Additional solutions are different combinations of these policies. In the queuing literature, the N -policy of repair is the best- and the T -policy is the worst-performing. Similarly, comparisons among the combinations of policies can also be performed.

Highlights of This Work

- The present work is the first to consider the k -out-of- n system as a multi-server queue.
- Unlike the classical multi-server queue, in the present work, at least k servers should be operational to provide a service. The multi-server case can be deduced from the present work by assuming that $k = 1$.
- Two birth and death processes encountered in the analysis are (i) the customer’s arrival and service processes and (ii) the accumulation of servers that break while providing a service; until this number reaches $n - N$ (pure death process), the repair process starts from this point of time and, with this, we have a birth and death process. Note that we assume all random variables (inter-arrival times, service time, repair time of

failed servers) involved to be exponentially distributed. The combined process turns out to be a birth and death process.

- The system under consideration reduces to the classical $M/M/\infty$ queue if we assume that the servers (system components) do not fail and that $k = 1$. Then, we take the limit as $n \rightarrow \infty$.

The remaining part of this paper is as follows. In Section 2, the description of the problem is given. In Section 3, the mathematical modeling of the problem and its analysis are presented. The stability condition is derived in Section 4 and the steady state probabilities found. Section 5 outlines certain distributions and performance measures of the system's behavior. Numerical and graphical illustrations providing insights into the working of the system are included in Section 6. An optimization problem is considered in Section 7. Concluding remarks are provided in Section 8.

2. Model Description

The model considered here is one that extends the k -out-of- n : G reliability system to a multi-server queue, especially to a multi-server queue with n servers/units working in parallel.

We consider a multi-server queueing system where customers arrive according to a Poisson process with parameter λ . The service facility consists of n servers/units, each of which can provide service to individual customers. We are considering a k -out-of- n : G system. All the n parallel servers have identically and exponentially distributed service times with parameter μ . Although these servers work independently, service will take place only if at least k servers are operational or in working condition. These servers are susceptible to breakdown. Breakdown occurs at exponentially distributed time intervals with rate γ . Repair will take place according to the N -policy of repair as described in the previous section. When the number of operational components comes down to the level N ($k \leq N < n$), a repair facility starts repairing failed units one at a time. The repair process continues until all failed components are restored to an operational state. It is assumed that the time taken to repair each unit is exponentially distributed with parameter δ . The system considered here is COLD, i.e., when the system fails due to the lack of at least k operational units, the units that are operational do not deteriorate until the system restarts again, with the failed units replaced with new ones. If we are considering a k -out-of- n : F COLD system, system failure occurs when any k of the n units fail. Here, the model under consideration is a k -out-of- n : G COLD system.

We formulate the problem mathematically as follows.

3. Mathematical Formulation

Define, for $t \geq 0$,

$N_1(t)$: the number of customers in the system at time t ;

$N_2(t)$: the number of operational units in the system at time t ;

$R(t)$: the status of the server that repairs the failed components at time t .

$$R(t) = \begin{cases} 0 & \text{if the repair is OFF} \\ 1 & \text{if the repair is ON} \end{cases}$$

Then, $\{(N_1(t), N_2(t), R(t)) | t \geq 0\}$ is a regular irreducible CTMC on state space $\Omega = \{(n_1, n_2, 1) : n_1 \geq 0; k - 1 \leq n_2 \leq n - 1\} \cup \{(n_1, n_2, 0) : n_1 \geq 0; N + 1 \leq n_2 \leq n\}$.

Let $\pi = (\pi_{k-1}, \pi_k, \pi_{k+1} \dots \pi_{N-1}, \pi_N, \pi_{N+1}, \dots \pi_{n-1}, \pi_n)$ be the steady-state probability vector of the infinitesimal generator matrix A .

In other words,

$$\pi A = \mathbf{0}; \pi \mathbf{e} = 1. \tag{1}$$

Define

$$U_{n-1} = \frac{1}{\gamma} B_6$$

$$U_i = \begin{cases} (B_3 + B_7 U_{n-1})^{-1} (-B_4) & \text{for } i = n - 2 \\ (B_3 + B_5 U_{i+1})^{-1} (-B_4) & \text{for } N - 1 \leq i \leq n - 3 \\ (B_3 + B_5 U_{N+1})^{-1} (-B_1) & \text{for } i = N \end{cases}$$

For $k \leq i \leq N$,

$$\pi_i = \left(\frac{\delta}{\gamma}\right)^{i-(k-1)} \pi_{k-1}$$

For $N + 1 \leq i \leq n - 1$,

$$\pi_i = \pi_{i-1} U_{i-1}$$

$$\pi_n = \pi_{n-1} U_{n-1}$$

From the normalizing condition $\pi \mathbf{e} = 1$, we have

$$\pi_{k-1} \left[\sum_{i=0}^k \left(\frac{\delta}{\gamma}\right)^i + \left(\frac{\delta}{\gamma}\right)^k \sum_{l=N}^{n-1} \prod_{j=N}^l U_j \mathbf{e} \right] = 1. \tag{2}$$

The infinitesimal generator of this Markov chain indicates that it is a level-independent quasi-birth–death process. Thus, this queuing system is stable if and only if $\pi A_0 \mathbf{e} < \pi A_2 \mathbf{e}$; see Neuts [8]. The stability condition is given as follows:

Theorem 1. *The given system is stable if and only if*

$$\lambda < \sum_{i=0}^{N-k} (k+i) \mu \left(\frac{\delta}{\gamma}\right)^{i+1} \pi_{k-1} + \sum_{l=N+1}^n l \mu \prod_{j=N}^l U_j \mathbf{e} \left(\frac{\delta}{\gamma}\right)^k \pi_{k-1} \tag{3}$$

4.2. Steady-State Probability Vector

Assuming the stability of the system, we proceed to find the steady-state probability of the system states.

Let x be the steady-state probability vector of Q , i.e., x satisfies $xQ = \mathbf{0}$ and $x\mathbf{e} = 1$. We partition this vector as

$$x = (x_0, x_1, x_2 \dots),$$

where $x_i = x_{(i, n_2, r)}$ are of dimension $d = 2n - (N + k) + 1$.

$$x_{n+i} = x_n R^i, i \geq 1$$

where the matrix R is the minimal non-negative solution to the matrix quadratic equation

$$R^2 A_2 + R A_1 + A_0 = 0$$

and the vectors $x_0, x_1, \dots, x_n \dots$ are obtained by solving the equations

$$x_0 A_1^0 + x_1 A_2^1 = 0 \tag{4}$$

$$x_{i-1} A_0 + x_i A_1^i + x_{i+1} A_2^i = 0, \text{ for } i \leq i \leq n - 2 \tag{5}$$

$$x_{n-2} A_0 + x_{n-1} A_1^{n-1} + x_n A_2 = 0 \tag{6}$$

$$x_{n-1} A_0 + x_n (A_1 + R A_2) = 0 \tag{7}$$

$$C_2^{n-1} = [0 \quad 0 \quad (n-1)\mu]$$

$$C_1^{n-1} = [-(n-1)\mu]$$

The matrices $C_1^i; k-1 \leq i \leq N-1$ are square matrices of order $2n - (i+1+N)$ obtained by deleting $i-1$ rows and columns from $A_1^i : k-1 \leq i \leq N-1$. $C_1^i; N \leq i \leq n-1$ are square matrices of order $2n-2-i; N \leq i \leq n-1$ obtained by deleting $N-1+(i-N)$ rows and columns from $A_1^i; N \leq i \leq n-1$.

The matrix T_1^0 is a column matrix $[0, \dots, 0, T_1^{k-1}, \dots, T_1^{N-1}, T_1^N, \dots, T_1^{n-1}]$. $T_1^i; k-1 \leq i \leq N-1$ is a column matrix with the first entry $\lambda + \gamma$. $T_1^i; N \leq i \leq n-2$ is a column matrix with the first two entries $\lambda + \gamma$. $T_1^{n-1} = [\lambda + \gamma]$

Theorem 2. Let U be the random variable designating the server idle time. Then, $P(U > t) = \alpha \cdot e^{\mathcal{Q}_1 t} \mathbf{e}$. This distribution is also of the PH type with representation $PH(\alpha, \mathcal{Q}_1)$, where $\alpha = \frac{1}{d}(x_{(0,k-1,1)}, \dots, x_{(0,N,1)} \dots x_{(0,n,0)} \dots x_{(k-1,k,1)} \dots x_{(k-1,n,0)}, \dots x_{(n-2,n-1,0)}, \dots x_{(n-2,n,0)}, x_{(n-1,n,0)})$, and d is the normalizing constant.

5.2. Distribution of First Passage Time from an Inoperative State to n -Operational Server State

Under the assumption that a sufficiently large number of customers are present in the system, we compute the distribution of time taken for the system to pass from a state in which there are $k-1$ operational servers to a state in which there are n operational servers.

Consider the Markov chain $\{(N_2(t), 1) | t \geq 0\}$ on state space $\Omega_2 = \{(n_2, 1) : k-1 \leq n_2 \leq n-1\} \cup \{*\}$. Here, $\{*\}$ denotes the absorbing state indicating that all servers are working. The time to absorption of this CTMC to $\{*\}$ is the first passage time from state $(k-1, 1)$ to the state $\{*\}$. The infinitesimal generator matrix of this Markov chain, when the states are arranged in ascending order of n_2 , is of the form

$$Q_2 = \left[\begin{array}{cccccc|c} -\delta & \delta & & & & & 0 \\ \gamma & -(\delta + \gamma) & \delta & & & & 0 \\ & & \ddots & \ddots & \ddots & & \\ & & & \gamma & -(\delta + \gamma) & \delta & \\ & & & & \gamma & -(\delta + \gamma) & \delta \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right]$$

Theorem 3. Let V be the random variable denoting the first passage time from an inoperative state to an n -operational server state, under the assumption of a large number of customers in the system. Then, V has a PH distribution with representation $PH(\beta, Q_2)$, where $\beta = (1, 0, \dots, 0)$.

5.3. Distribution of First Passage Time from an n -Operational Server State to an Inoperative State

Here, we also compute the distribution of time taken for the system to pass from a state in which there are n operational servers to a state in which there are $k-1$ operational servers, under the assumption that a sufficiently large number of customers are present in the system.

Consider the Markov chain $\{(N_2(t), r) | t \geq 0\}$ on state space $\Omega_3 = \{(n_2, 1) : k \leq n_2 \leq n-1\} \cup \{(n_2, 0) : N+1 \leq n_2 \leq n\} \cup \{*\}$. Here, $\{*\}$ denotes the absorbing state indicating that only $k-1$ servers are working/operational. The time to absorption of this CTMC to $\{*\}$ is the first passage time from state $(n, 0)$ to the state $\{*\}$. When the states are arranged in decreasing order of n_2 , the infinitesimal generator matrix of this Markov chain is of the form

$$Q_3 = \left[\begin{array}{cc} S_1 & S_1^0 \\ \mathbf{0} & 0 \end{array} \right]$$

where

$$S_1 = \begin{bmatrix} S_{11} & S_{12} & \mathbf{0} \\ S_{21} & S_{22} & S_{23} \\ \mathbf{0} & S_{32} & S_{33} \end{bmatrix}$$

$$S_{11} = -\gamma$$

$$S_{12} = \gamma$$

$$S_{21} = e_1 \otimes [0, \delta]',$$

is e_1 is a column matrix of order $n - 1 - N$.

$$S_{22} = I_{(n-1-N)} \otimes \begin{bmatrix} -\gamma & \\ & -(\gamma + \delta) \end{bmatrix} + E^- \otimes \begin{bmatrix} 0 & 0 \\ 0 & \delta \end{bmatrix} + E^+ \otimes \begin{bmatrix} \gamma & 0 \\ 0 & \gamma \end{bmatrix}$$

where E^- and E^+ are square matrices (refer to page 19) of dimension $n - 1 - N$.

$$S_{23} = E_{(n-1)-N1} \otimes [\gamma \ \gamma]'$$

$E_{(n-1)-N1}$ is a matrix (refer to page 19) of order $(n - 1) - N \times (N - k)$.

$$S_{32} = E_{1(n-1)-N} \otimes [0 \ \delta]$$

$E_{1(n-1)-N}$ is a matrix of order $(N - k) \times (n - 1) - N$.

$$S_{33} = -(\lambda + \gamma)I_{N-k} + \delta E^- + \gamma E^+$$

where E^- and E^+ are square matrices (refer to page 19) of dimension $N - k$.

$$S_1^0 = \gamma e_{2n-(N+k)}$$

Theorem 4. Let W be the random variable denoting the first passage time from the n -operational server state to an inoperative state, under the assumption of a large number of customers in the system. Then, W has a PH distribution with representation $PH(\chi, Q_3)$, where $\chi = (1, 0, \dots, 0)$.

5.4. Distribution of Number of Times That the System Becomes Inoperative before Reaching the n Server State

Under the assumption of a large number of customers in the system, we compute the distribution of the number of times that the system hits the $k - 1$ server state before reaching the n server state. Let Δ denote the absorbing state indicating the number of working servers hitting n . Then, $(M(t), N_2(t))$, (where $M(t)$ is the number of times that the system hits the $k - 1$ server state before reaching the n server state) is the Markov chain on state space $\Omega_4 = \{\Delta\} \cup \{(m, n_2) : m \geq 0, k - 1 \leq n_2 \leq n - 1\}$. The infinitesimal generator is of the form

$$Q_4 = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots \\ \delta e_{n-k} & M_0 & M_1 & \mathbf{0} & \mathbf{0} & \dots \\ \delta e_{n-k} & \mathbf{0} & M_0 & M_1 & \mathbf{0} & \dots \\ \delta e_{n-k} & \mathbf{0} & \mathbf{0} & M_0 & M_1 & \dots \\ \vdots & & & \ddots & \ddots & \ddots \end{bmatrix}$$

where

$$M_0 = \begin{bmatrix} -\delta & \delta e'_{n-k} \\ \mathbf{0} & \gamma E^- + \delta E^+ - (\gamma + \delta)I_{n-k} \end{bmatrix}$$

$$M_1 = \gamma E_{21}$$

E_{21} is a matrix of order $n - k - 1$ with 1 in the $(2, 1)$ position and E^-, E^+ are matrices (refer to page 19) of order $n - k$.

Let y_m denote the probability that the system hits the $k - 1$ server state m times before reaching the n server state.

$$y_0 = -\delta\psi M_0^{-1}\mathbf{e}$$

$$y_m = (-1)^m \delta\psi (M_1 M_0^{-1})^m M_0^{-1}\mathbf{e}$$

where ψ is the initial probability vector $\psi = \frac{1}{d_1}(\pi_{k-1}, \dots, \pi_N, \dots, \pi_{(N+1,1)}, \dots, \pi_{(n-1,1)})$, and d_1 is the normalizing constant.

Theorem 5. *The expected number of times that the system hits the $k - 1$ server state before reaching the n server state is $\sum my_m$.*

5.5. Other Performance Measures

- Fraction of time for which the system is under repair:

$$F_{Repair} = \sum_{n_1=0}^{\infty} \sum_{n_2=k-1}^{n-1} x_{(n_1, n_2, 1)}$$

- Fraction of time for which the servers are idle:

$$F_{Idle} = \sum_{n_1=0}^{\infty} x_{(n_1, k-1, 1)} + \sum_{n_1=0, n_1 \leq n_2}^{\infty} \sum_{n_2=k-1}^n \sum_{r=0,1} x_{(n_1, n_2, r)}$$

- System reliability, i.e., the probability that at least k servers are operational:

$$S = \sum_{n_1=0}^{\infty} \sum_{n_2 \geq k}^n \sum_{r=0,1} x_{(n_1, n_2, r)} = 1 - \sum_{n_1=0}^{\infty} x_{(n_1, k-1, 1)}$$

- Average number of customers in the system:

$$E_{System} = \sum_{n_1=0}^{\infty} n_1 x_{n_1}$$

- Average number of failed servers in the system:

$$E_{FS} = \sum_{n_2=k-1}^{n-1} (n - n_2) \sum_{n_1=0}^{\infty} \sum_{r=0,1} x_{(n_1, n_2, r)}$$

- Average number of servers that are idle:

$$E_{IS} = \sum_{n_1=0}^{\infty} (k - 1) x_{(n_1, k-1, 1)} + \sum_{n_1=0, n_1 \leq n_2}^{\infty} \sum_{n_2=k-1}^n \sum_{r=0,1} (n_2 - n_1) x_{(n_1, n_2, r)}$$

6. Numerical Illustrations

In this section, we give some numerical examples that show the effect of the level N of the N -policy and the repair rate δ on certain performance measures. For this, we consider a 5-out-of-20 : G system.

6.1. Effect of Level N on the Performance of the System

In this numerical example, the following parameters are kept fixed with values as given below:

$$\lambda = 10; \mu = 4; \gamma = 2; \delta = 1$$

From Table 1 and Figures 1–4, the following conclusions can be made:

- The fraction of time for which the servers are under repair is the minimum for a value $N = 7$. To minimize F_{Repair} , we need to initiate the repair process when the number of non-operational servers is 13.

- The fraction of time for which the servers are idle can be minimized if we choose $N = 8$. It plays a crucial role in enhancing the cost-effectiveness of the model.
- The system reliability is maximum when $N = 7$. We can achieve a more reliable system if we take into consideration the results in Figure 3.
- The expected number of customers found waiting in the system is the minimum if we choose $N = 8$. The value of the expected number of customers is low as the servers are working in parallel.

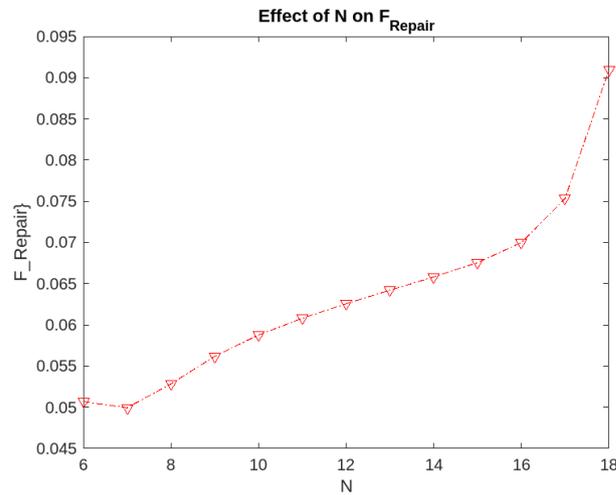


Figure 1. Effect of parameter N on the fraction of time for which the servers are under repair.

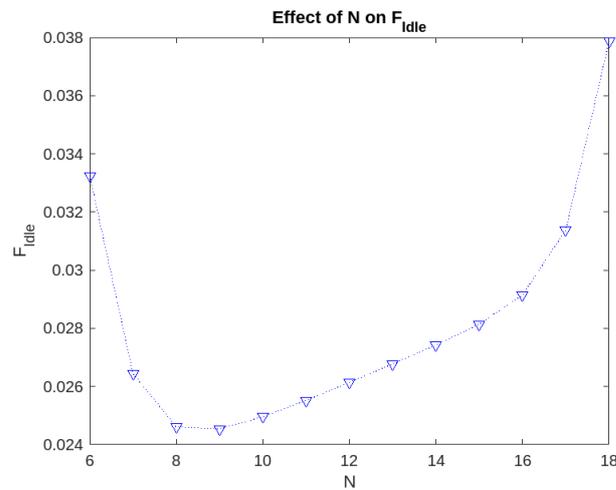


Figure 2. Effect of the level N on the fraction of time for which the servers are idle.

Table 1. Effect of N on F_{Repair} , F_{Idle} , S and E_{repair} .

N	F_{Repair}	F_{Idle}	S	E_{System}
6	0.0507	0.0332	0.9691	0.0651
7	0.0499	0.0265	0.9726	0.0530
8	0.0528	0.0246	0.9722	0.0505
9	0.0562	0.0245	0.9708	0.0510
10	0.0588	0.0250	0.9696	0.0523
11	0.0608	0.0255	0.9685	0.0537
12	0.0626	0.0261	0.9676	0.0551

Table 1. Cont.

N	F_{Repair}	F_{Idle}	S	E_{System}
13	0.0642	0.0268	0.9667	0.0564
14	0.0658	0.0274	0.9658	0.0578
15	0.0676	0.0281	0.9649	0.0594
16	0.0700	0.0291	0.9636	0.0615
17	0.0753	0.0314	0.9608	0.0662
18	0.0909	0.0379	0.9527	0.0799

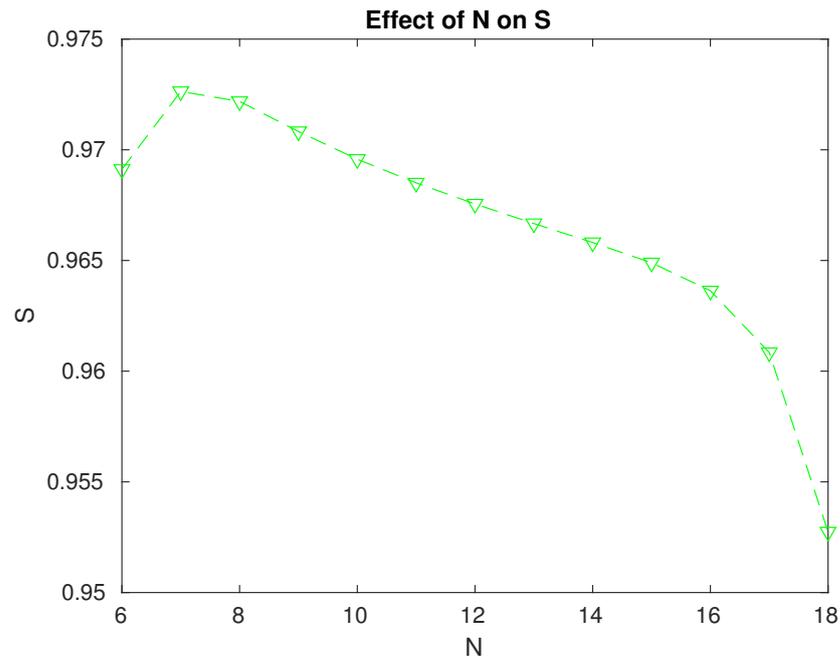


Figure 3. Effect of the level N on the reliability of the system.

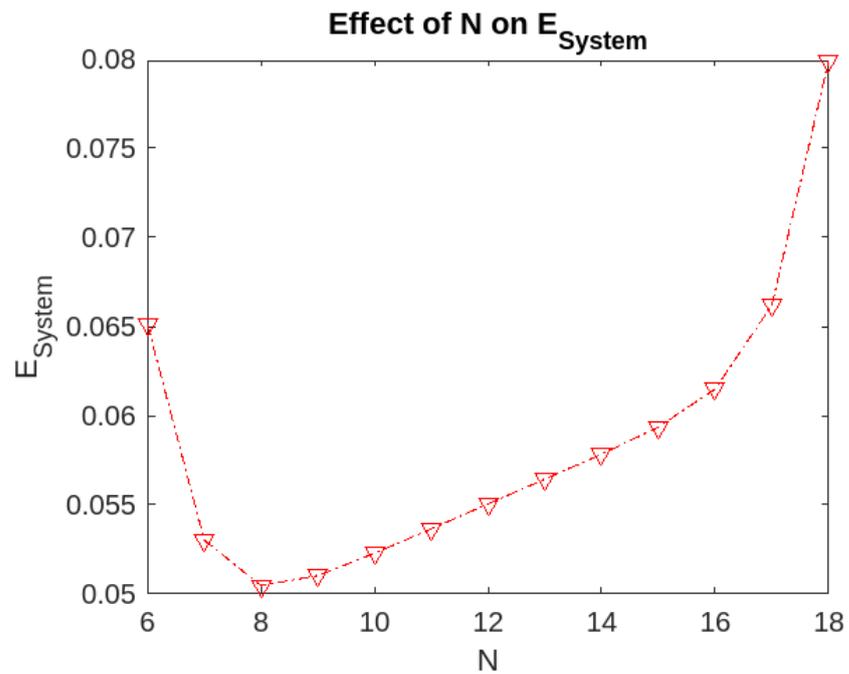


Figure 4. Effect of the level N on E_{System} .

The above example illustrates that the analysis can guide us in properly managing the repair policy with specific objectives.

6.2. Effect of the Repair Rate δ on the Performance of the System

In this section, we study the effect of the repair rate δ on the performance of the system. In this example, the following parameters are kept fixed with values as given below:

$$\lambda = 35; \mu = 26; \gamma = 3; N = 6$$

From Table 2 and Figures 5–8, the following conclusions can be made:

- The fraction of time for which the servers are under repair is the minimum for a value $\delta = 4$. We can decide to what extent we need to enhance the repair facilities or resources based on the data in Figure 5.
- The fraction of time for which the servers are idle is maximum for $\delta = 2$. The data in Figure 6 play a crucial role in specifically designing the facilities so that the idle time of the servers can be effectively utilized and the entire system can be managed accordingly.
- The system reliability increases with increasing values of δ as expected. The rate at which this increase occurs helps us to decide on the optimum value to be considered when compared to the resources or efforts involved in enhancing the repair facilities; see Figure 7.
- The expected number of customers found waiting in the system is maximum for $\delta = 2$.

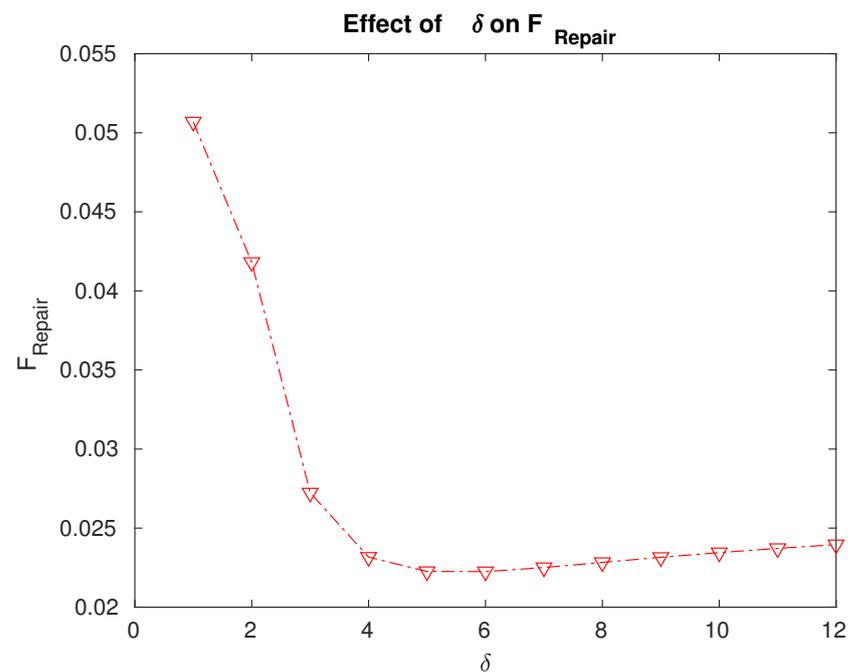


Figure 5. Effect of the repair rate δ on the fraction of time for which the servers are under repair.

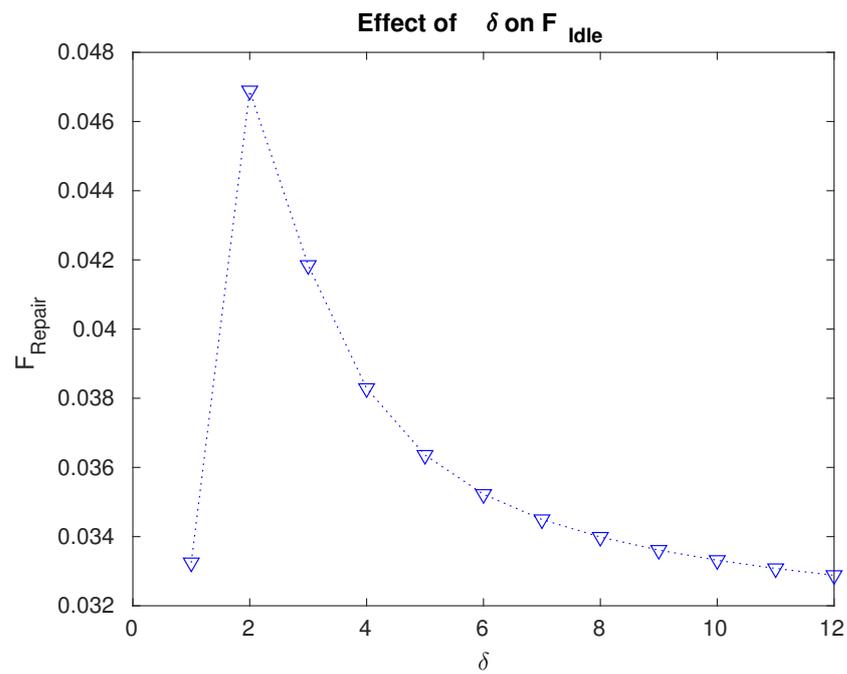


Figure 6. Effect of the repair rate δ on the fraction of time for which the servers are idle.

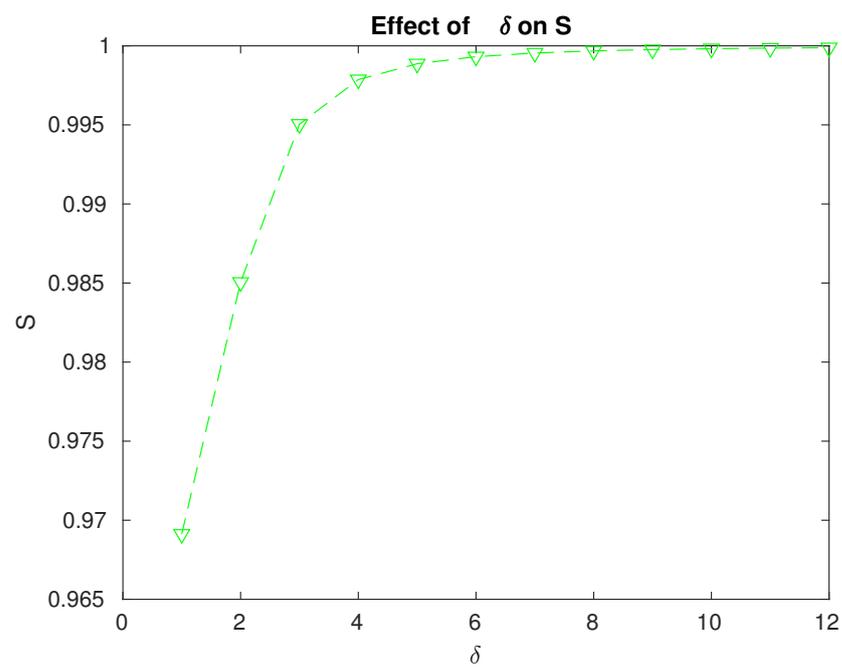


Figure 7. Effect of the repair rate δ on the reliability of the system.

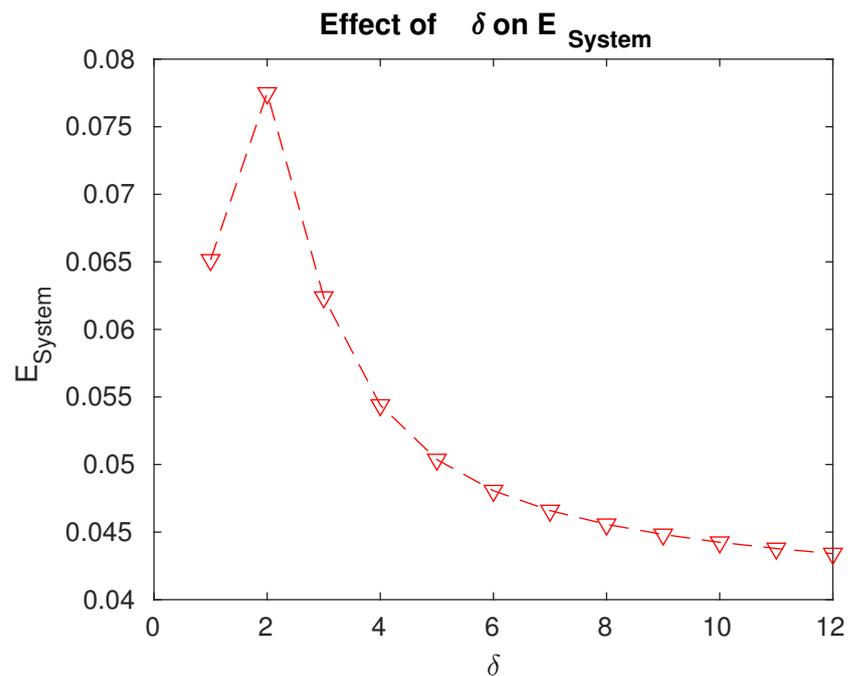


Figure 8. Effect of δ on E_{System} .

Table 2. Effect of repair rate δ on F_{Repair} , F_{Idle} , S and E_{repair} .

δ	F_{Repair}	F_{Idle}	S	E_{System}
1	0.0507	0.0332	0.9691	0.0651
2	0.0418	0.0469	0.9850	0.0775
3	0.0272	0.0418	0.9950	0.0624
4	0.0232	0.0383	0.9979	0.0544
5	0.0223	0.0364	0.9989	0.0504
6	0.0222	0.0352	0.9993	0.0481
7	0.0225	0.0345	0.9995	0.0466
8	0.0228	0.0340	0.9997	0.0456
9	0.0232	0.0336	0.9998	0.0448
10	0.0235	0.0333	0.9998	0.0442
11	0.0237	0.0331	0.9999	0.0438
12	0.0240	0.0329	0.9999	0.0434

7. Cost Analysis and Optimization Problem

Cost analysis plays an important role in decision making or in formulating policies related to the working of any system that we encounter in everyday life. In this section, we propose a cost function related to the system under consideration. We consider an optimization problem to find the value of the level N of the N -policy of the repair. With the help of numerical examples as well as graphical illustrations, we show that an optimal value N exists so that the total expected cost is the minimum. We also study the effect of the repair rate δ on the the total expected cost.

To determine the optimal level of N at which the repair facility can start working and to determine the effectiveness of enhancing the repair rate, we proceed as follows. For the cost analysis, we define the following costs.

- C_1 : Establishment cost or setup cost.
- C_2 : Holding cost per customer per unit time.
- C_3 : Unit time cost to run the repair mechanism.
- C_4 : Unit time cost incurred due to the idleness of the servers.
- C_5 : unit time revenue received from the busy servers.
- C_6 : Unit time revenue received when at least k servers are operational.

The expected total cost is

$$TC = C_1 + C_2 * E_{System} + C_3 * F_{Repair} + C_4 * F_{Idle} - C_5 * (1 - F_{Idle}) - C_6 * S$$

We fix the following values:

$$C_1 = 1000, C_2 = 75, C_3 = 100, C_4 = 200, C_5 = 500, C_6 = 160$$

The values of other parameters are the same as in Section 6.

- From Table 3 and Figure 9, we see that the optimal value of the level N is $N = 8$, for which the total expected cost is the minimum. Thus, it is optimal to wait until 12 servers are non-operational before starting the repair.
- From Table 3 and Figure 10, it can be seen that the expected total cost is the maximum when the repair rate is $\delta = 2$. This means that if we spend more to increase the rate at which the repair is performed from 1 to 2 or more, it will not be reflected in the total expected cost. Thus, a decision regarding the facilities to be arranged to ensure a specific repair rate can be made using this cost analysis. In this specific numerical example, it is enough to ensure a repair rate of at most $\delta = 1$.
- By analyzing the problem numerically, we can decide on the optimal values of the level N of the N -policy and also the optimal repair rate δ to be maintained.

Table 3. Effect of N and δ on expected total cost.

N	TC	δ	TC
6	528.2877	1	528.2877
7	523.1107	2	540.4254
8	521.8479	3	535.9006
9	521.9576	4	532.8554
10	522.4002	5	531.2721
11	522.9301	6	530.3771
12	523.4913	7	529.8193
13	524.0646	8	529.4412
14	524.6497	9	529.1666
15	525.2904	10	528.9562
16	526.1948	11	528.7882
17	528.1975	12	528.6496
18	534.0302		

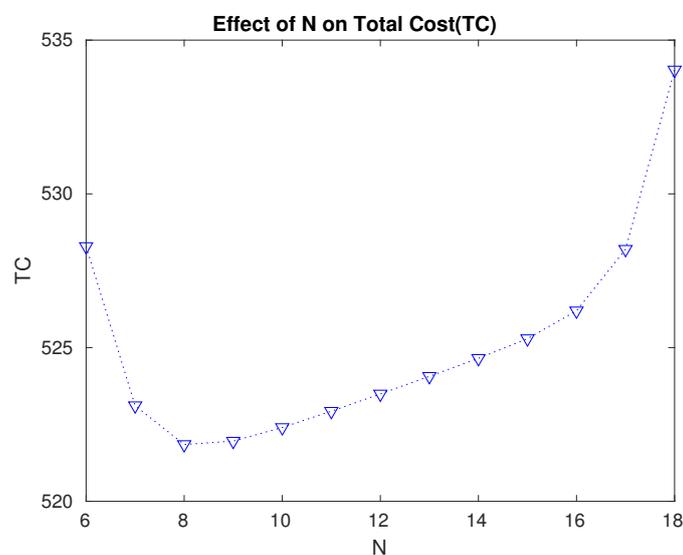


Figure 9. Effect of the changes in level N on the expected total cost.

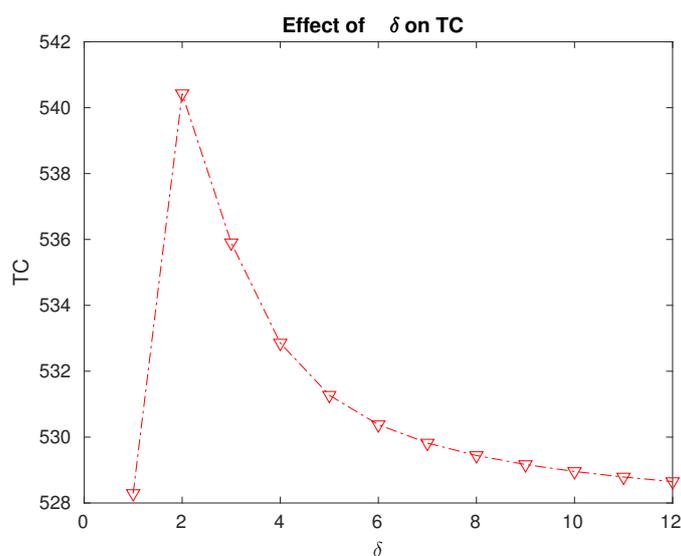


Figure 10. Effect of the repair rate δ on the expected total cost.

8. Conclusions

This paper studies a multi-server queuing system with the N -policy of repair, by viewing it as a k -out-of- n : G system. The steady-state distributions of various system states are computed. The distribution of server idle times is analyzed. Under the assumption of a sufficiently large number of customers in the system, the distribution of the first passage times from an inoperative state to an n server state and vice versa has been found. The assumption of a sufficiently large number of customers in the system is made to avoid complications that may arise due to future arrivals. Other system performance measures such as system reliability, the expected number of failed servers, etc., are computed.

The effect of an increase in the level N and the repair rate δ on various performance measures, when other parameters are kept fixed, is studied numerically and graphically. An increase in the repair rate significantly increases the system reliability and decreases the expected number of customers in the system. It reduces the idleness of servers and reduces the fraction of time for which the servers are under repair. On the other hand, an increase in N increases the fraction of time for which the servers are under repair. However, as N increases, the idleness of servers and the expected number in the system first decrease and then increase, while the system reliability first increases and then decreases. The optimal value of N depends on the parameters of the system. The examples illustrate how an optimal policy could be derived for a multi-server queuing system, keeping in mind certain specific objectives. A cost function has been constructed and the results of the cost analysis are presented.

There can be several extensions to the problem considered in this paper. An extension of the present work to one in which consecutive k -out-of- n systems provide a service to customers, either linearly or in a circular fashion, is proposed for future research. A similar study of the k -out-of- n : F system is underway. Moreover, the k -out-of- n system could be analyzed under both HOT and WARM conditions.

Author Contributions: Conceptualization, A.K.; Methodology, A.N.J.; Validation, A.P.M.; Formal analysis, A.N.J.; Data curation, A.P.M.; Writing—original draft, A.K., A.N.J. and A.P.M.; Writing—review & editing, A.K., A.N.J. and A.P.M.; Supervision, A.K. All authors have read and agreed to the published version of the manuscript.

Funding: Dr. Anu Nuthan Joshua and Dr. Ambily P. Mathew received support from DST-RSF research project number 22-49-02023 (RSF) and research project number 64800 (DST) for the preparation of this publication. <https://search.crossref.org/funding>.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

Notations and Abbreviations

The following abbreviations are used in this manuscript:

CTMC	Continuous-time Markov chain
QBD process	Quasi-birth–death process
\mathbf{e}	All one vector with appropriate dimension
I_n	Identity matrix of order $n \times n$
E^-	Square matrix of appropriate size with all zero entries except the entries $(E^-)_{(j,j-1)}$, which are equal to 1
E^+	Square matrix of appropriate size with all zero entries except the entries $(E^+)_{(j,j+1)}$, which are equal to 1
E_{ij}	Matrix of appropriate size with all zero entries except the entry (i, j) , which is equal to 1
e_i	Column vector of appropriate dimension with all zero entries except the entry i , which is equal to 1
A'	Transpose of matrix A
$\mathbf{0}$	Matrix whose entries are 0, of appropriate size
$A \otimes B$	Kronecker product; if $A = [a_{ij}]$ is a matrix of order $m \times n$ and if B is a matrix of order $p \times q$, then $A \otimes B = [a_{ij}B]$ will denote a matrix of order $mp \times nq$

References

1. Barlow, R.E.; Proschan, F. *Mathematical Theory of Reliability*; John Wiley and Sons: New York, NY, USA, 1965.
2. Krishnamoorthy, A.; Sathian, M.K.; Narayanan Viswanath, C. Reliability of a k -out-of- n System with Repair by a Single Server Extending Service to External Customers with Pre-emption. *Electron. J. Reliab. Theory Appl.* **2016**, *11*, 61–93.
3. Yadin, M.; Naor, P. Queueing system with a removable service station *Oper. Res. Q.* **1963**, *14*, 393–405. [[CrossRef](#)]
4. Krishnamoorthy, A.; Ushakumari, P.V.; Lakshmy, B. k -out-of- n -system with repair: The N -policy. *Asia-Pac. J. Oper. Res.* **2002**, *19*, 47–61.
5. Yen, T.-C.; Wang, K.-H.; Chen, J.-Y. Optimization Analysis of the N -Policy $M/G/1$ Queue with Working Breakdowns. *Symmetry* **2020**, *12*, 583. [[CrossRef](#)]
6. Vemuri, V.K.; Boppana, V.S.N.H.P.; Kotagiri, C.; Bethapudi, R.T. Optimal strategy analysis of an N -policy two-phase $M^X/M/1$ queueing system with server startup and breakdowns. *Opsearch* **2011**, *48*, 109–122. [[CrossRef](#)]
7. Singh, C.J.; Jain, M.; Kumar, B. Analysis of queue with two phases of service and m phases of repair for server breakdown under N -policy. *Int. J. Serv. Oper. Manag.* **2013**, *16*, 373–406 [[CrossRef](#)]
8. Neuts, M.F. *Matrix—Geometric Solutions in Stochastic Models: An Algorithmic Approach*; The Johns Hopkins University Press: Baltimore, MD, USA, 1981.
9. Gross, D.; Harris, C. *Fundamentals of Queueing Theory*, 3rd ed.; John Wiley: Chichester, UK, 1988.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.