RT&A, No 4 (65)
Volume 16, December 2021

A Reliability-Inventory Problem Under N-policy of replenishment of component

Achyutha Krishnamoorthy

•

Centre for Research in Mathematics
C.M.S. College, Kottayam - 686001, India
achyuthacusat@gmail.com krishnamoorthy@cmscollege.ac.in

Abstract

In this paper a new process is introduced. To some extent it has resemblance with Queueing-Inventory (Inventory with positive service time) (see Sigman and Simchi-Levy [2] and Melikov and Molchanov [1]. We consider a k - out - of - n: G system of identical components, each of which has exponentially distributed life time with parameter λ, independent of the others. When the number of working components goes down to N ($k \leq N \leq n$) due to failures, an order for $n - k + 1$ items is placed. Replenishment time is exponentially distributed with parameter β. On replenishment, all failed components are instantaneously replaced by the new arrivals, subject to a maximum of $n - k + 1$. This process is investigated and its long run system state distribution derived explicitly. An associated optimization problem is discussed. Throughout this paper the k - out - of - n system is assumed to be COLD.

Keywords: COLD system, System Reliability, k - out - of - n System, Replenishment policy, Serial and Parallel Systems

1. Introduction

The purpose of this paper is to introduce a notion similar to Queueing - Inventory (QI), introduced in 1992 by two groups of researchers: Sigman and Simchi-Levy [2] and Melikov and Molchanov [1], independently of each other. Until then service time associated with providing an inventoried item was assumed to be negligible. In reality, that assumption is rarely valid. A brief description of QI is as follows. In classical queue, if the server is ready to serve and customers are waiting then the service starts. The notion of the requirement of some materials is totally missing in it. However, to provide service some item(s) is often required. It was Kazimirsky [7] who came up with the idea of an item required to provide service. In the absence of such an item(s) service cannot be given. In classical inventory, it is assumed that the service time is negligible. That is to say, if the item is of demand is available, the server provides it to the customer in a negligible amount of time and the customer leaves the system. In case the item is not available, customers may wait until the inventory gets replenished. Thus absence of inventory alone results in customers joining a queue of demands. The waiting space may be of finite or infinite capacity. On replenishment, a certain number of waiting customers equal to min{number waiting, number of inventoried items replenished}, leave the system with the inventory - it is assumed that each customer asks for exactly one unit of the item. The assumption of negligible service time is often unrealistic. This is the one that prompted Sigman and Simchi-Levy as well as Melikov and Molchanov to introduce positive service time. This results in the formation of queue even when inventory is available. The reader may refer to the recent survey paper by Krishnamoorthy et al [4] for further details on the work done up to 2018 in QI.

We consider a service providing system, namely a k-out-of-n system. Such a system has n identical components/units. The system continuously operates. When the number of operational component hits $k-1$, the system fails. We assume that the life times of these n units are independent and identically distributed random variables with exponential distribution having parameter λ. Up on the number of working components going down to $N(k \leq N \leq n)$ due to failures, an order for $n-k+1$ items are placed. Lead time is exponentially distributed with parameter β. The life time of components and lead time are independent random variables. On replenishment, all failed components are replaced by the new arrivals, subject to a maximum of $n-k+1$. This process is analysed to derive its long run system state distribution. In this paper the case of COLD system alone is analysed and an associated optimization problem discussed. The system is referred to as COLD if the components that were operational at the time of system failure, do not deteriorate further until the system is again put back to operation by replacement/repair of failed components. We can consider different types of replenishment policies and also systems of that are WARM or HOT. In a warm system, components that remain operational at the time of system failure continue to deteriorate, but at a slower rate than when the system is up. We restrict the discussion to COLD system because the very purpose of this work is to announce the above indicated new direction of thoughts. For this reason we also assumed that all distributions involved are exponential.

Next we present a brief discussion in the investigation done on the reliability of k - out - of - n : G system. This system is extensively investigated. Its particular cases, serial and parallel systems are of special interest. A detailed discussion on these can be found in Sivazlian and Stanfel [3]. Krishnamoorthy and Ushakumari [5] extended a repairable k - out - of - n : G system to the case of retrial of failed components for repair. Krishnamoorthy et al [6] introduced the N-policy of repair in k - out - of - n : G system and investigated the optimal number N of failed components that should accumulate in order to start the repair of failed components in a cycle to maximize the reliability of the system. Here a cycle is defined as the time interval that starts at the epoch all the n components are in working condition until the moment all components that fail during this time period are repaired and the system is back with all components in operational state.

Barlow and Heidtmann [9] present a linear-time algorithm and its short computer program in BASIC for the computation of reliability of a k - out - of - n : G system. We now turn to a few more recent investigations on k - out - of - n : G system. Zhang et al. [10] analyse a k - out - of - n : G system with repairman's single vacation and shut off rule. The working times and repair times of components follow exponential distributions, and the duration of the repairman's vacation is governed by a phase type distribution. Both transient and long run system availability are obtained. Time-dependent behavior of the system performance measures under different initial system states, are obtained. Monte Carlo simulation and special cases of the system are investigated to check the correctness of the results obtained. Ji-EunByun et al [11] investigate the reliability growth of k-out-of- N systems using matrix-based system reliability method. To increase the reliability of a specific system, using redundant components is a common method which is called redundancy allocation problem (RAP). Some of the RAP studies have focused on k-out-of-n systems. However, all of these studies assumed predetermined active or standby strategies for each subsystem. Mahsa Aghaei et al [12] propose a k - out - of - n series - parallel system when the redundancy strategy can be chosen for each subsystem. Because the optimization of RAP belongs to the NP-hard class of problems, a modified version of genetic algorithm (GA) is developed. The exact method and the proposed GA are implemented on a well-known test problem and the results obtained demonstrate the efficiency of the approach of the authors compared to the previous studies.

In this paper we introduce the concept of replacement of failed components through a purchase of new items that have the same life time distribution as the failed components. The order for purchase is placed when the number of operational components in the system falls down to $N, k \leq N \leq n$. It takes an exponentially distributed amount of time, called the lead time, for the replenishment of items to take place. The order quantity is fixed at $n-k+1$. On physical realization of the order, failed components are replaced by the new arrivals. The time for replacement is assumed to be negligible. It may be noted that at most $n-k+1$ components need replacement at the time when replenishment takes place because operational components do not deteriorate when the system is down (COLD system). As a result none, one , ... up to a maximum of $N-k+1$ excess/spare components will be available as standby units. This means that the system is working now with all n components in operation and the remaining, if any, stay as spares. These are brought to operation, one at a time, as and when components fail. This process gets repeated.

The reader may wonder about the distinction from the classical queueing-inventory (QI) problem and may even ask the question: are they not the same if the number of customers in the QI is restricted to a finite number? *The answer is a firm NO. This is so because at a replenishment epoch the number of failed units of the k - out - of - n system can be smaller than* $n-k+1$, *the replenishment quantity. Thus there could be excess inventory to be stored, which are put into operation when failure of components of the system takes place. Those excess components alone have holding cost. However, in QI the inventory level may at most reach S at a replenishment epoch. Further all items held in the inventory have holding cost associated with them. Also notice that all components of the system that are in operation, deteriorate, though those on Şstandby (the excess remaining after failed components are replaced)Ť do not deteriorate (because the system is COLD)*. Thus there are valid reasons for analysing the reliability-inventory (RI) problem presented in the previous paragraph.

The remaining part of this paper is arranged as follows. In section 2, the mathematical model of the problem is presented. The long run system state distribution is explicitly computed. In section 3, we compute a few distributions of interest, associated with the model. Section 4 provides a cost function involving the decision variable N. Its analysis is then presented. This cost function is shown to be convex. Thus there is a global optimum value for N. Finally a concluding section tells about future plans for extensions and generalizations.

Notations and abbreviations:
In the sequel the following notations and abbreviations are employed:
i.i.d - independent and identically distributed.
rv(s) - random variable(s).
$CTMC$ - Continuous time Markov Chain.
IPV Ů initial probability vector.
$X(t)$ Ű Number of operational components in the system at time t.
$Y(t)$ Ű Number of spare/standby components available at time t.
Note that only when $X(t) = n$, the value of $Y(t)$ can be positive.

2. Mathematical Modeling and Analysis of the problem

The system under consideration is COLD: when the system fails in the absence of at least k operational components, the components that are still operational do not deteriorate until system again starts operation, with the failed components replaced by new ones. Though only one new component suffices to put the system back into operation, we follow the policy of replacing all failed components at the time when replenishment of the ordered items take place. The replenishment quantity is $n-k+1$. All of them may not be immediately required. Therefore the excess items are kept as spares/standby for future replacements as and when required. Life times of components are *i.i.d rv*s having exponential distribution with parameter λ. When number of operational components drops down to N, with $k \leq N \leq n$, an order is placed for $n-k+1$ new components. It takes an exponentially distributed time with parameter β for the materialization of this order. This is referred to as lead time in inventory management. During this time, none, one, ... , up to a maximum of $N-k+1$ components may fail. Up on replenishment, all failed components are replaced by new ones and the system continues to operate. It may be noted as stated earlier, that all of these $n-k+1$ units may not be required to bring back the number of operational components in the system to n. Therefore only that much of these new components that are required.

With $X(t)$ defined as the number of operational components at time t and $Y(t)$, that of spares, we see that $\{(X(t), Y(t)), t \geq 0)\}$ is a two-dimensional $CTMC$ with state space $\{(i,0)|i = k-1, 2, ..., n\} \cup \{(n,j)|j = 1, 2, ..., N-k+1\}$. This process is not skip - free to the right because, immediately after replenishment the number of operational components increases by at least $n-N$ (with none, one or more left as excess) and at most by $n-k+1$(without any unit left as standby). We employ the difference-differential equation technique to compute the long run system state distribution. The figure below provides the working of the system: 2 - out - of - 5 : G system.

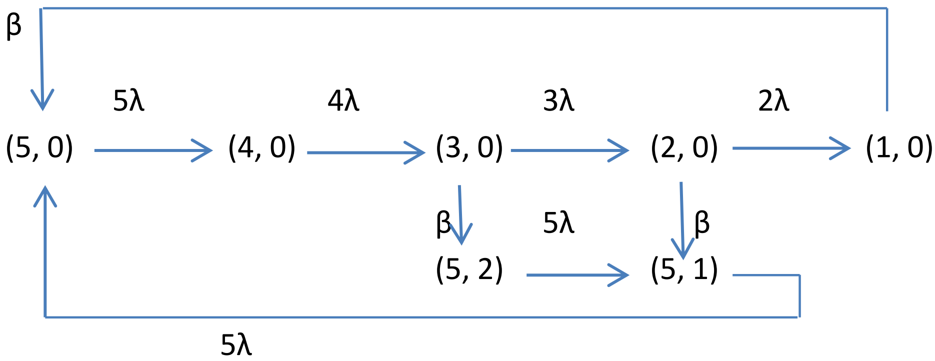


Figure 1: *Transition diagram of* 2 - *out* - *of* - 5 : G *system with* $N = 3$ *when the failure rate is* λ*.*

The transition rate matrix of the 2 - out - of - 5 : G system is as given below:

$$\begin{array}{c} \\ (1,0) \\ (2,0) \\ (3,0) \\ (4,0) \\ (5,0) \\ (5,1) \\ (5,2) \end{array} \begin{array}{c} \begin{array}{ccccccc} (1,0) & (2,0) & (3,0) & (4,0) & (5,0) & (5,1) & (5,2) \end{array} \\ \begin{pmatrix} -\beta & & & & \beta & & \\ 2\lambda & -(\beta+2\lambda) & & & & \beta & \\ & 3\lambda & -(\beta+3\lambda) & & & & \beta \\ & & 4\lambda & -4\lambda & & & \\ & & & 5\lambda & -5\lambda & & \\ & & & & 5\lambda & -5\lambda & \\ & & & & & 5\lambda & -5\lambda \end{pmatrix} \end{array}$$

In this we have, $n = 5$, $k = 2$ and $N = 3$. For that system we get the long run behavior of the system as

$$q_{4,0} = \tfrac{5}{4} q_{5,0};$$
$$q_{3,0} = \tfrac{4\lambda}{\beta+3\lambda} \tfrac{5}{4} q_{5,0};$$
$$q_{2,0} = \tfrac{3\lambda}{\beta+2\lambda} \tfrac{4\lambda}{\beta+3\lambda} \tfrac{5}{4} q_{5,0};$$
$$q_{1,0} = \tfrac{2\lambda}{\beta} \tfrac{3\lambda}{\beta+2\lambda} \tfrac{4\lambda}{\beta+3\lambda} \tfrac{5}{4} q_{5,0};$$
$$q_{5,2} = \tfrac{\beta}{5\lambda} q_{3,0} = \tfrac{\beta}{5\lambda} \tfrac{4\lambda}{\beta+3\lambda} \tfrac{5}{4} q_{5,0}$$
$$q_{5,1} = q_{5,0} - \tfrac{\beta}{5\lambda} q_{1,0} = q_{5,0} - \tfrac{\beta}{5\lambda} \tfrac{2\lambda}{\beta} \tfrac{3\lambda}{\beta+2\lambda} \tfrac{4\lambda}{\beta+3\lambda} \tfrac{5}{4} q_{5,0}.$$

Now we add these limiting probabilities. Since their sum is 1, we immediately get $q_{5,0}$.

A different failure rate case also will be discussed in the numerical section; this one considers inverse variation of rate of failure with the number of operational components: when the number of components in operation is j, the failure rate is λ/j. This leads to more compact expressions for the system state probabilities. Therefore, we can expect a much nicer expression for the optimal N value as well. The figure below provides the working of the system: 2 - out - of - 5 : G system.

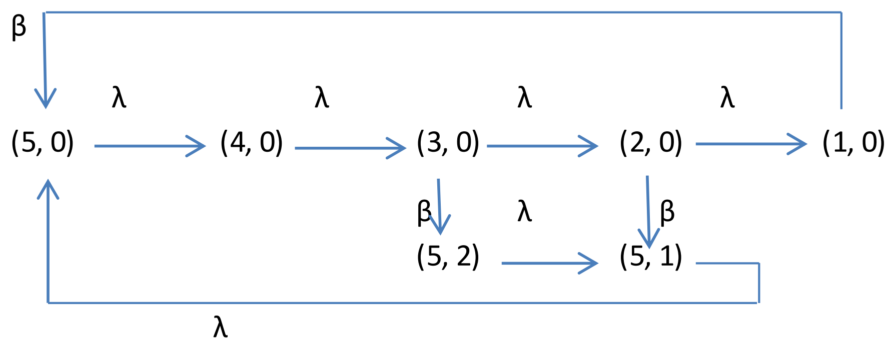


Figure 2: *Transition diagram of* 2 - *out* - *of* - 5 : G *system with* $N = 3$, *when the failure rate is* λ/j.

The transition rate matrix of the 2 - out - of - 5 : G system is as given below:

$$
\begin{array}{c} \\ (1,0) \\ (2,0) \\ (3,0) \\ (4,0) \\ (5,0) \\ (5,1) \\ (5,2) \end{array}
\begin{array}{c}
\begin{array}{ccccccc} (1,0) & (2,0) & (3,0) & (4,0) & (5,0) & (5,1) & (5,2) \end{array} \\
\begin{pmatrix}
-\beta & & & & \beta & & \\
\lambda & -(\beta+\lambda) & & & & \beta & \\
 & \lambda & -(\beta+\lambda) & & & & \beta \\
 & & \lambda & -\lambda & & & \\
 & & & \lambda & -\lambda & & \\
 & & & & \lambda & -\lambda & \\
 & & & & & \lambda & -\lambda
\end{pmatrix}
\end{array}
$$

Nevertheless, the long run system state probabilities are indicated below for the 2-out-of- 5 : G system, in the case where the failure rate is λ/j when j components are operating. We continue to use the same notation for the system state probability.

$$
\begin{aligned}
q_{5,1} &= (1 + \tfrac{\lambda}{\lambda+\beta})q_{5,2}; \\
q_{5,0} &= (1 + \tfrac{\lambda}{\lambda+\beta} + \tfrac{\lambda}{\beta}\tfrac{\lambda}{\lambda+\beta})q_{5,2}; \\
q_{4,0} &= (1 + \tfrac{\lambda}{\beta})q_{5,2}; \\
q_{3,0} &= (\tfrac{\lambda}{\beta})q_{5,2}; \\
q_{2,0} &= (\tfrac{\lambda}{\beta})(\tfrac{\lambda}{\lambda+\beta})q_{5,2}; \\
q_{1,0} &= (\tfrac{\lambda}{\beta})^2(\tfrac{\lambda}{\lambda+\beta})q_{5,2}.
\end{aligned}
$$

These, together with the normalizing condition, gives $q_{5,2}$.

Assume that the process initially starts in state $(n, 0)$. Up to the state $(N, 0)$, the process is a pure death process, with linear death rates (depending on the number of operational components). An order for replenishment for $n - k + 1$ units is placed on reaching $(N, 0)$. The replenishment may precede next failure or may be after the next failure and so on, could be even after the system fails. Therefore there is a chance of system reliability getting affected. Since the system is COLD, no more working component fails until they are again put into operation which can happen only after the replenishment. When replenishment takes place, all failed units are replaced instantaneously. Thus from $(N, 0)$ onwards the process is no more a pure death process nor it can be called a birth and death process because the replenishment is in bulk. Thus it is not skip-free to the right. Denote by $P_{ij}(t)$, the probability that the system is in state (i, j) at time t and $P'_{ij}(t)$ its derivative. Then the difference-differential equations satisfied by $P_{i,j}(t)$ are:

$$P'_{nj}(t) = -\lambda n P_{nj}(t) + \lambda n P_{nj+1}(t) + \beta P_{i0}(t); j = 0, 1, ..., i-k+1, i = k-1, ..., N;$$
$$P'_{i0}(t) = -(\lambda i + \beta) P_{i0}(t) + \lambda(i+1) P_{i+10}(t) \text{ for } i = k-1, ..., N;$$
$$P'_{i0}(t) = -\lambda i P_{i0}(t) + \lambda(i+1) P_{i+10}(t) \text{ for } i = N+1, ..., n-1.$$

These three systems of equations can be solved for computing the time dependent behavior of the system state probabilities (see Karlin and Taylor [13], Chapter 4). When transient effect fades, the system gets stabilized. Denote by q_{ij} the limit distribution, as $t \to \infty$, of $P_{ij}(t)$. The $CTMC$ under study is aperiodic and irreducible, though it may get absorbed into state $(k-1, 1)$, only to stay there for an exponentially distributed duration. Later on we will consider that state as an absorbing state for deriving the distribution of time during which the system provides failure free operation. Thus the above system of equations gives us:

$$n\lambda q_{n,j} = n\lambda q_{n,j+1} + \beta q_{i,0} \text{ for } j = 0, 1, ..., n-k+1-(n-i) : i = k-1, ..., N;$$
$$(\lambda i + \beta) q_{i,0} = \lambda(i+1) q_{i+1,0} \text{ for } i = k-1, ...N;$$
$$i\lambda q_{i,0} = (i+1)\lambda q_{i+1,0} \text{ for } i = N+1, ..., n-1.$$

These are recursively solved to arrive at the long run system state probability as given below.

Theorem 1. : With q_{ij} defined as the limit as $t \to \infty$ of $P_{ij}(t)$, we get:

$$q_{i,0} = \frac{i+1}{i} q_{i+1,0} \text{ for } i = N+1, ..., n-1;$$
$$q_{i,0} = \frac{\lambda(i+1)}{(\lambda i + \beta)} q_{i+1,0} \text{ for } i = k-1, ..., N \text{ and}$$
$$q_{n,j} = q_{nj+1} + \frac{\beta}{n\lambda} q_{i,0} \text{ for } j = 0, 1, ..., i-k+1 \text{ and } i = k-1, ..., N.$$

Proof. These show that we can express the system state probability in terms of $q_{n,0}$, for example. Then by total probability argument (the normalizing condition), we get $q_{n,0}$. Thus we have explicit analytical expressions for the system state probability. Next we use these to derive several system characteristics which, in turn, are used in analyzing a related optimization problem. ■

3. Performance Characteristics

- Mean number of operational components when the system is working (excluding spares, if any), $OCW = \sum_{j=1}^{j=N-k+1} n q_{n,j} + \sum_{i=k}^{n} i q_{i,0}$.
- Mean number of spare components, $SC = \sum_{j=1}^{N-k+1} j q_{n,j}$.
- Fraction of time system is down, $FTD = q_{k-1,0}$.
- Fraction of time the system is up, $FTU = 1 - q_{k-1,0}$.

FTU is the complement of FTD. Our objective is to make it as close to one as possible, subject to constraints of funds and at the same time the significance of the role of the machine. Thus N plays the most crucial role.

3.1. Related Distributions

In this section we derive a few distributions of interest that arise in the study of the system. We may assume, without loss of generality, that the system starts in state $(n, N-k)$. The distributions that are derived include the distribution of the time until first failure; distribution of the number of replenishments that take place before the system failure; distribution of the number of times the replenishment results in excess inventory and in particular, the distribution of the number of times the excess number of spares reached $N-k+1$ and those that resulted in no excess inventory.

Distribution of the time until first failure

We consider the Markov chain with state space $\{(i,0)|i = k, ..., n\} \cup \{(n, j)|j = 0, 1, ..., N-k+1\}$. Notice that we have dropped two states from the state space: The state $(k-1, 0)$ is excluded because we want the distribution of the time during which the process remains continuously in the transient states of the Markov chain. Because of that, in consequence to a replenishment, the excess inventory/spare parts level cannot be zero. The initial probability vector γ of the Markov chain has entries 1 at the place corresponding to $(n, 0)$ and 0 at the remaining positions. The reason for starting in state $(n, 0)$ is that a new cycle starts after the machine failed. Thus the state $(k-1, 0)$ is reached before replenishment of components. So after replacing all failed components by the new arrivals, the system is left with no spare unit. Our objective is to compute the distribution of the time T until the state $(k-1, 0)$ is reached for the first time. This is given in the following theorem.

Theorem 2. Starting in state one of the states in the set, the distribution of the time T until absorption takes place is phase type with representation (γ, U) of order $n + N - 2k + 1$. U is that part in the infinitesimal generator of the Markov chain corresponding to the set of states $\{(i,0)|i = k, ..., n\} \cup \{(n, j)|j = 0, 1, ..., N-k+1\}$ and γ is the IPV vector with 1 at the position corresponding to the state $(n, 0)$ and 0 at the remaining places.

NOTE We may relax the assumption that the initial state is $(n, 0)$ by associating probabilities for starting in any state. In that case there will be corresponding changes in the IPV γ. However, for computing the distribution of the time till next failure (i.e., distribution of the time duration between two successive failures of the system), the state $(n, 0)$ has to be the starting state. **Proof.** Write the difference - differential equations satisfied by the probabilities of the system occupying any state belonging to $\{(j,0)|j = k, ..., n\} \cup \{(n, j)|j = 0, 1, ..., N-k+1\}$. Now solve this matrix differential equation to get the tail distribution of T as $P(T > t) = \gamma e^{(Ut)}\mathbf{e}$, where $\mathbf{e}$ is a column vector of 1Šs having the same order as that of γ. Therefore $P(T < t) = 1 - \gamma e^{(Ut)}\mathbf{e}$. The expected time to failure is given by - $\gamma U^{-1}\mathbf{e}$.(see Neuts [8]). ■

Distribution of the number of times the replenishment results in excess inventory before absorption to $(k-1, 0)$

To compute this distribution we proceed as follows. We start at an epoch of replenishment that takes the state space to one of $(n, 1), ..., (n, N-k+1)$. These precisely correspond to those replenishments that take place while the system is in states $(k, 0), ..., (N, 0)$, respectively. The IPV will be defined accordingly. Further we assume that the immediately preceding replenishment took place only after reaching the state $(k-1, 0)$. The initial probability vector of the Markov chain associated with these states is $\Theta = (\theta_{k0}, ..., \theta_{n0}, ..., \theta_{nN-k})$ and at the remaining positions, including $(k-1, 0)$ and $(n, 0)$ the entries are all zeros. If we look at the time t (i.e., pre-event occurrence epoch), when the replenishment takes place during $[t, t+h)$ for h infinitesimally small, we notice that in the initial probability vector the only non-zero elements are $\theta_{k0}, ..., \theta_{N0}$. We introduce an additional component called *level*, as the first coordinate, into the state space of the process. We start at level 0 assuming that no replenishment order has so far materialized. It may happen that the process reaches $(0, k-1, 0)$ before the materialization of the replenishment order that was placed on reaching $(0, N, 0)$. In this case the required number turns out to be zero. We call this a *failure*. Suppose that replenishment against the order which was placed on reaching $(0, N, 0)$, materializes before dropping to $(0, k-1, 0)$. We label this as a *success*. Then the level goes up by 1 and the resulting state is an element of $\{(1, n, j) | j = 1, ..., N-k, N-k+1\}$. This is the first success. Thus the level, the first coordinate in the triplet, stands for the number of consecutive successes. The components start failing with passage of time and on reaching down to $(1, N, 0)$, the next replenishment order Is placed. The two possibilities thereafter are: i) replenishment only after the system breaks down (ie., state $(1, k-1, 0)$is reached) or ii) replenishment takes place before falling to state $(1, k-1, 0)$. In case the event mentioned as (ii) occurs, then we have the second success. The Śconsecutive success counting processŠ goes on like this. In this we notice that the time elapsed between consecutive replenishment epochs are *i.i.d.rv*s following the tail of the phase type distribution with representation $PH(\Theta, V)$ where V is the part of the infinitesimal generator corresponding to these transient states. It is important to note that, because $(k-1, 0)$ is absorbing state, we have not brought it into the above computational argument. For this reason the state $(., n, 0)$ also does not come into play.

Now back to the computation of the required probability distribution. Denote by Y, the random variable that represents the number of successes before the first failure where success and failure are in the context as described in the previous paragraph. Denote the tail of the $PH(\Theta, V)$ distribution described above by p and its complement by q. Then the distribution of Y is given by $P(Y = m) = p^m q$ for $m = 0, 1, ...$ which is the geometric distribution. We sum up these in the following theorem.

Theorem 3. The distribution of the number of times the replenishment results in excess inventory before absorption to $(k-1, 0)$ is given by the geometric distribution with parameter p where p is the tail of the $PH(\Theta, V)$ distribution which is the time until absorption into state $(k-1, 0)$ of the Markov chain describing the state space of the k - out - of - n : G system.

Corollary 1. From theorem 3.2, we conclude that the distribution of number of consecutive failures of the system between two successive failure free cycles is also geometrically distributed. Let Z denote this random variable. Then $P(Z = m) = q^m p$ for $m = 0, 1, 2, ...$

Corollary 2. From the state space description of the Markov chain of the k - out - of - n : G system, it is clear that the consecutive number of times the excess inventory is positive (i.e., it hits the set $\{1, 2, ..., N-k, N-k+1\}$ between two successive system failures, also has the geometric distribution: Denote this *rv* by D. Then $P(D = m) = p^m q$ for $m = 0, 1, 2, ...$

Remark 1. It can be easily proved that the distribution of the time between two successive visits to any state, say $(k-1, 0)$, is phase-type distributed with appropriate representation (see Theorem 3.1). A similar procedure can be adopted to compute the distribution of the time duration for successive visits to any state in the state space of the Markov chain.

4. An Optimization Problem

In this section we construct a cost function involving the decision variable N. The relevant costs are:
K - Fixed cost of placing an order for replenishment
C - Purchase cost/ unit item
h - Holding cost/excess units held/time
R - Penalty cost/time when system is down.
We consider the cost function: Average cost per unit time when the replenishment order for spare items is placed when number of operating components drops down to N,
$F(N) = [K + C(n-k+1)]/$(Expected time elapsed between two consecutive order placements)$+$
$h. \sum_{j=1}^{j=N-k+1} j q_{n,j} + R q_{k-1,0}$
First we compute the expected length of a cycle. Here a cycle time is the time duration, starting

from an epoch at which state $(N,0)$ is reached to the next epoch at which that state is revisited. . Denote the length of this cycle by W. We have to compute $E(W)$. First we compute the distribution of W. Figure 1 gives an idea about W in the special case discussed therein. In the general case also the state space was described earlier. We incorporate a major modification in the order in which the state space appears and also an Şadditional elementŤ to it for the computation of the distribution of $W : \{(N,0),(N-1,0),...,,(k,0),(k-1,0),(n,N-k+1),...,(n,N-k),....,(n,1),(n,0),...,(n-1,0),...,(N+1,0),*\}$. In this $*$ is an absorbing state and the remaining states are transient. This $*$ is the same as the state $(N,0)$; however the intention of using a distinct notation is to indicate that the state $(N,0)$ is revisited. Thus we can compute the distribution of the distribution of the time duration elapsed, starting from $(N,0)$ back to $(N,0)$ for the first time after the next replenishment at the same level or a lower level followed by deterioration of components. The infinitesimal generator of the corresponding $CTMC$ is given below.

$$\mathcal{G} = \begin{bmatrix} Q & Q^* \\ \mathbf{0} & 0 \end{bmatrix}$$

$$Q = \begin{bmatrix} Z_{11} & Z_{12} & \\ & Z_{22} & Z_{23} \\ & & Z_{33} \end{bmatrix}$$

and Q^* the column vector with entry $(N+1)\lambda$ in the last position.

$$Z_{11} = \begin{array}{c} \\ (N,0) \\ (N-1,0) \\ \vdots \\ (k,0) \end{array} \begin{array}{c} \begin{array}{ccccc} (N,0) & (N-1,0) & (N-2,0) & \dots & (k,0) \end{array} \\ \begin{pmatrix} -(\beta+N\lambda) & N\lambda & & & \\ & -(\beta+(N-1)\lambda) & (N-1)\lambda & & \\ & & \ddots & \ddots & \\ & & & & -(\beta+k\lambda) \end{pmatrix} \end{array}$$

$$Z_{12} = \begin{array}{c} \\ (N,0) \\ (N-1,0) \\ \vdots \\ (k,0) \end{array} \begin{array}{c} \begin{array}{ccccc} (k-1,0) & (n,N-k+1) & (n,N-k) & \dots & (n,1) \end{array} \\ \begin{pmatrix} & \beta & & & \\ & & \beta & & \\ & & & \ddots & \\ k\lambda & & & & \beta \end{pmatrix} \end{array}$$

$$Z_{22} = \begin{array}{c} \\ (k-1,0) \\ (n,N-k+1) \\ (n,N-k) \\ \vdots \\ (n,1) \end{array} \begin{array}{c} \begin{array}{ccccc} (k-1,0) & (n,N-k+1) & (n,N-k) & \dots & (n,1) \end{array} \\ \begin{pmatrix} -\beta & & & & \\ & -n\lambda & n\lambda & & \\ & & -n\lambda & & \\ & & & \ddots & \\ & & & & -n\lambda \end{pmatrix} \end{array}$$

$$Z_{23} = \begin{array}{c} \\ (k-1,0) \\ (n,N-k+1) \\ (n,N-k) \\ \vdots \\ (n,1) \end{array} \begin{array}{c} \begin{array}{cccc} (n,0) & \dots & (N+2,0) & (N+1,0) \end{array} \\ \begin{pmatrix} \beta & & & \\ 0 & & & \\ 0 & & & \\ \vdots & & & \\ n\lambda & & & \end{pmatrix} \end{array}$$

$$Z_{33} = \begin{array}{c} \\ (n,0) \\ (n-1,0) \\ \vdots \\ (N+2,0) \\ (N+1,0) \end{array} \begin{array}{c} \begin{array}{ccccc} (n,0) & (n-1,0) & \dots & (N+2,0) & (N+1,0) \end{array} \\ \begin{pmatrix} -n\lambda & n\lambda & & & \\ & -n\lambda & & & \\ & & \ddots & & \\ & & & -(N+2)\lambda & (N+2)\lambda \\ & & & & -(N+1)\lambda \end{pmatrix} \end{array}$$

It follows from it that the time until absorption to $*$ has the Coxian distribution with representation (δ, Q) where Q is that part of the infinitesimal generator sans the row and column corresponding to $*$. Its dimension is $n + N - 2(k-1)$ and δ is the initial probability vector with 1 at the first position and the remaining elements are 0sŠ. Its dimension is obvious from this description. Thus we have proved the following:

Theorem 4. The distribution of a cycle (starting from state (N,0), returning to it for the first time), has Coxian distribution with representation (δ, Q) of order $n + N - 2(k-1)$. Denoting by W the length of this cycle, we have $E(W) = \delta Q^{(-1)}\mathbf{e}$.

Now we go back to the cost function described above. We compute this for two cases:
(a) 2 - out - of - 5 : G system in which N can take values $2, 3, 4, 5$;
(b) 5 - out - of - 10 : G system in which N can take values $5, 6, 7, 8, 9$.
Fix the various costs as $K = \$10, C = \$1, h = \$3, R = \$20$.
We have computed the long run probability distribution of the system (a), as an illustration for the k - out - of - n : G system under N - policy for placing order for replenishment. First we take up that case. The expression for cost function is as follows: $F(N) = [10 + 1(5-2+1)]/$(Expected time elapsed between two consecutive order placements)$+3.\sum_{j=1}^{j=N-2+1} jq_{5,j} + 50q_{1,0}$. The results for various values of λ and β are summarized in the following table:

Table 1: *Effect of N on Cost Function for a* 2 - *out* - *of* - 5 : *G system.*

(λ, β)	$N = 2$	$N = 3$	$N = 4$	$N = 5$
$(1,1)$	28.0085	24.75	**24.444**	24.5
$(1,2)$	23.8617	**19.1453**	19.8726	21.0732
$(2,1)$	40.1261	**38.1405**	38.1920	38.3339

In the case when individual rate of failure is λ/j when the number of operating components is j, the system state probabilities are computed and given in section 2. (b) For this system the state space is$\{(i,0)|i = 4, ..., 10\} \cup \{(10, j)|j = 1, 2, ...N-4\}$.
The expression for cost function is as follows: $F(N) = [10 + 1(10-5+1)]/$(Expected time elapsed between two consecutive order placements)$+3.\sum_{j=1}^{j=N-2+1} jq_{10,j} + 50.q_{4,0}$. The results for various values of λ and β are summarized in the following table:

Table 2: *Effect of N on Cost Function for a* 5 - *out* - *of* - 10 : *G system.*

(λ, β)	$N = 5$	$N = 6$	$N = 7$	$N = 8$	$N = 9$
$(1,1)$	34.7209	34.0944	33.75	**33.6134**	33.64
$(1,2)$	28.9071	28.2580	**28.2174**	28.6125	29.3370
$(1.5,1)$	41.8150	41.6413	**41.6096**	41.6771	41.8181

A much more realistic but simple way of looking at the component deterioration would have been as follows: the rate of component deterioration is λ/j when j components in the system are operational. This leads to the system deterioration rate as $j.\lambda/j = \lambda$. This is the case describing the load balance on the system as stronger when a larger number of components are operational which is more realistic. In this case the expression for the system state probability gets much more simplified and looks more elegant.

The long run system state probabilities in this case are:

$$q_{(k-1,0)} = (\frac{\lambda}{\beta})(\frac{\lambda}{\lambda+\beta})^{(N-k)} q_{(N,0)};$$
$$q_{(N-i,0)} = (\frac{\lambda}{\lambda+\beta})^{(N-i)} q_{(N,0)} \text{ for } i = 1, 2, .., N-k;$$
$$q_{(N+1,0)} = q_{(N+2,0)} = ... = q_{(n-1,0)} = q_{(n,0)} = (\frac{\lambda+\beta}{\lambda}) q_{(N,0)}$$

The case when failure rate is inversely proportional to the number of operating components, the system state probabilities can be deduced from the above or directly computed. These are as given below:

$$q_{(N+1,0)} = q_{(N+2,0)} = ... = q_{(n-1,0)} = q_{(n,0)} = (1 + \frac{\beta}{\lambda}) q_{(N,0)};$$
$$q_{(n,N-k+1)} = (\frac{\beta}{\lambda}) q_{(N,0)};$$
$$q_{(n,N-k-j)} = (\frac{\beta}{\lambda})\{1 + (\frac{\lambda}{\lambda+\beta}) + ... + (\frac{\lambda}{\lambda+\beta})^{(j+1)}\} q_{(N,0)} \quad \text{for} \quad j = 0, 1, 2, .., N-k-1;$$
$$q_{(N-j,0)} = (\frac{\lambda}{\lambda+\beta})^j q_{(N,0)} \quad \text{for} \quad j = 0, 1, ..., N-k$$
$$q_{(k-1,0)} = \frac{\lambda}{\beta}(\frac{\lambda}{\lambda+\beta})^{(N-k)} q_{(N,0)}.$$

Table 3: *Effect of N on Cost Function for a* 2 *- out - of -* 5 : *G system, when failure rate is* λ/j *when the number of operating components in the system is j.*

(λ, β)	$N = 2$	$N = 3$	$N = 4$	$N = 5$
$(1, 1)$	11.4589	**7.4706**	7.7273	9.3385
$(1, 2)$	8.9415	**5.7808**	7.2535	9.7473
$(1.5, 1)$	16.67	11.5242	**10.8677**	11.5857

Table 1 shows that, for the 2−out-of−5 system, the optimal values of N for the various combinations of (λ, β) given by $(1, 1), (1, 2)$ and $(2, 1)$ are respectively, $4, 3, 3$ and the minimum costs are \$24.444, \$19.1453 and \$38.1405. In contrast to this, Table 3 shows pretty small values for the cost function. This shows the effect of reduced failure rate when the number of operating units is closer to the maximum value.

Table 4: *Effect of N on Cost Function for a* 5 *- out - of -* 10 : *G system, when failure rate is* λ/j *when the number of operating components in the system is j.*

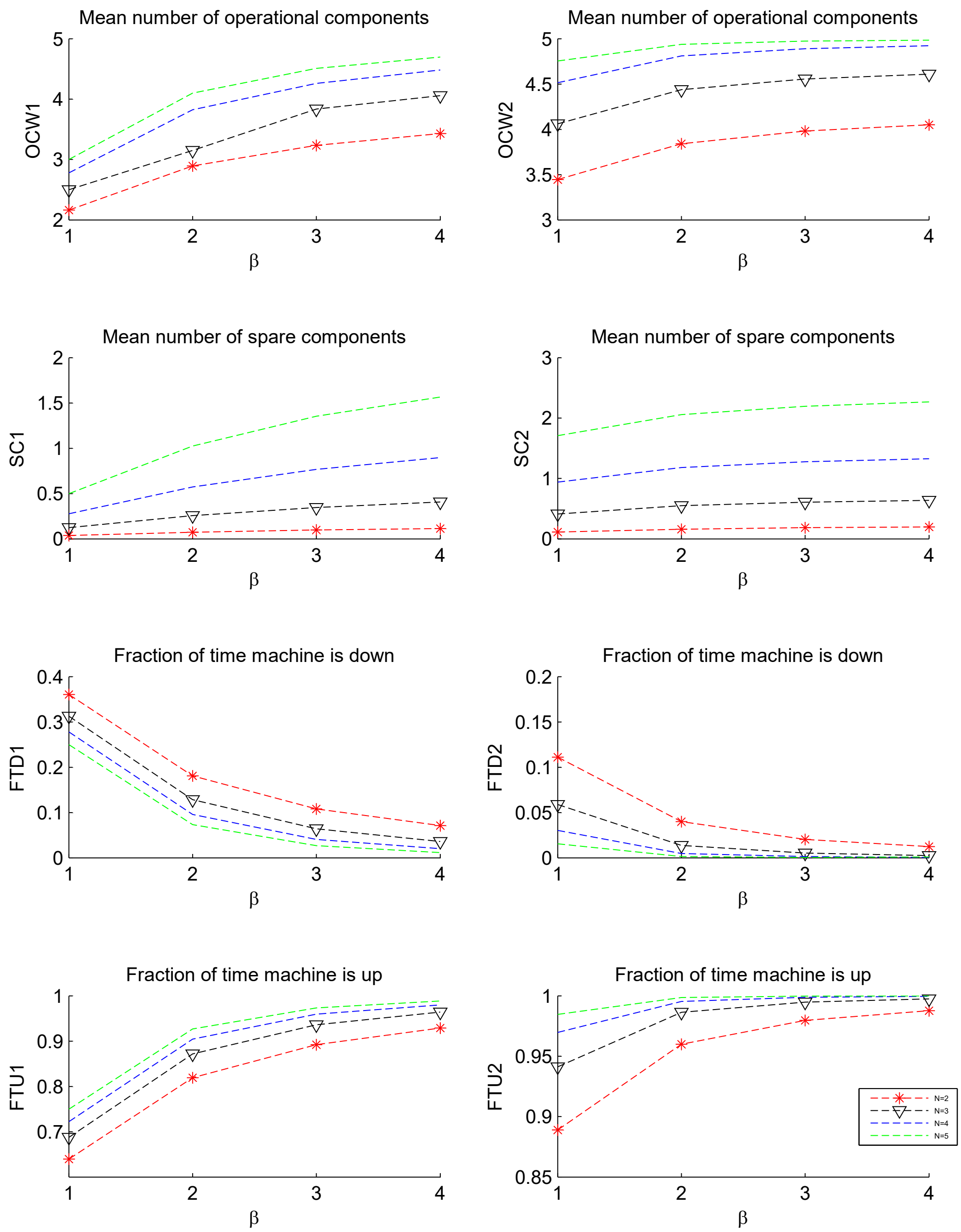
(λ, β)	$N = 5$	$N = 6$	$N = 7$	$N = 8$	$N = 9$
$(1, 1)$	6.5385	**5.4**	5.5306	6.5876	8.3679
$(1, 2)$	4.2703	**4.2018**	5.1754	6.8345	9.0552
$(1.5, 1)$	10.1739	8.4537	**7.9051**	8.3293	9.5734

Table 2 shows that, for the 5−out-of−10 system, the optimal values of N for the various combinations of (λ, β) given by $(1, 1), (1, 2), 1.5, 1)$ are respectively, \$33.6134, \$28.2174 and \$41.6096. In contrast to this, when failure rate is inversely proportional to the number of operating units, the cost gets considerably reduced and a shift in the optimal N value is also observed.

In this case we can easily see that the holding cost of excess (spare) components increases with the increase in value of N because $N - k$ increases under this condition; the order for replenishment is placed when N is closer to n and so there is higher chance of replenishment taking place before the system reaches the state $(k-1, 0)$, thereby ensuring smooth functioning of the system thereby reducing the risk involved due to system failure. Conversely, if we move down N towards k, the reliability of the system can get seriously affected because the order materialization may get delayed. Consequently the number of operating components could get reduced to $k - 1$, thus affecting system reliability. In other words the order for replenishment is placed when the number of operating components is closer to k. So the replenishment could get correspondingly delayed, endangering system reliability. Of course, one can argue that the replenishment time is exponentially distributed and so it lacks memory. In any case for the same parameter of the lead time exponential distribution, we will see the distinction through the examples. In the case of failure rate inversely proportional to number of operating components, we see that the cost function constructed is convex. In particular for parallel (1-out-of-n : G system) and serial (n-out-of-n : G system) systems we get the corresponding optimal N value from the general case.

The eight figures (titled as Figure 3) given below, provide a very clear picture of how the system performs. The first two among these indicate that, with faster replenishment rate the number of components in operation goes up in the two types of failure rates indicated. This trend is also seen to be true for the number of spares available (see the 3rd and 4th figures). Fraction of time the system is up, is considerably smaller when failure rate of the system is directly proportional to the number of operational components than when the system failure rate is inversely proportional to that number (the last two pair of figures). The third pair of figures tells us about the fraction of time the system is down in the two distinct scenarios.

Figure 3: *Effect of β and N on performance measures, when failure rate is λ and λ/j respectively.*



5. Concluding remarks

In this paper we considered a k - out - of - n : G system with N - policy for placing orders for replacement of failed components. The long run system state probability distribution is computed when failure rate is linear. The case of constant failure rate is shown to be a particular case of that. A number of distributions associated with the system are derived. In particular, the time duration between two successive failures of the system is shown to be of phase-type with appropriate representation. The distribution of consecutive number of failure free cycles (each replenishment taking place before the system drops to $(k-1,0)$, and thus system failure is averted) is shown to have geometric distribution. An optimization problem for determining the optimal value of the control variable N, is constructed and its optimal value is computed. Computational experience indicates that the function so constructed, is convex in N.

There are several extensions and generalizations of the problem investigated in this paper. For example, instead of exponential distribution any continuous distribution with non - negative part of real line as support which does not lack memory, could be introduced. However, this may result in the loss of $CTMC$ status for the system. The component life times also could be replaced by such distributions; however, this will lead to a very complex system. Yet another direction of investigation is the case of repair of failed components under N-policy. In this case, when the number of failed components reaches $n-N$, repair of failed units starts. Thus either a machinery/server for repair of failed components has to be hired. Questions such as immediate availability arises in this case just as the role played by the lead time in the model analysed. Also there arises the repair time. A comparison between the model analysed and the case of repair of failed components may lead to interesting results. There is a very important extension of the problem presented in this paper to what can be called Reliability - Queueing - Inventory problem. Another direction for future work is to have a permanent server for repair of failed components. He/she will also process items that can be used to replace failed components. The server does this while waiting for accumulation of $n-N$ failed components of the system. Work on these directions are in progress.

Acknowledgment: The author expresses his deep sense of gratitude to Ms. Anu Nuthan Joshua, Department of Mathematics, Union Christian College, Aluva, India, for carefully going through the manuscript and for correcting a few errors.

References

[1] Melikov, A.Z., Molchanov, A.A. (1992). Stock optimization in transportation/storage systems. *Cybernetics and Systems Analysis*, 28(3): 484 – 487.

[2] Sigman, K., Simchi-Levi, D. (1992) . Light traffic heuristic for an M/G/1 queue with limited inventory. *Annals of Operation Research*, 40: 371 – 380.

[3] Sivazlian B.D. Stanfel L.E. Analysis of Systems in Operations Research, Prentice-Hall international series in industrial and systems engineering, 1975 - 04 - 01.

[4] Krishnamoorthy, A., Dhanya Shajin, Viswanath C. Narayanan (2020). Inventory with Positive Service Time: a Survey, Queueing Theory 2, coordinated by Vladimir Anismov, Nikolaos Liminios. ISTE Editions.

[5] Krishnamoorthy, A.,ăUshakumari, P.V. (1999). Reliability of a k - out - of - n system with repair and retrial of failed units.ă*Topă7*, 293–304 doi:10.1007/BF02564728.

[6] Krishnamoorthy, A., ăUshakumari, P.V.,ăLakshmy, B. (2002). K - out - of - n - system with repair: The N - policy.*ăAsia -Pacific Journal of Operational Research*,ă19 (1): 47.

[7] Kazimirsky, A.V. (2006). Analysis of BMAP/G/1 queue with reservation of service. *Stochastic Analysis and Applications*, 24(4):703 – 718.

[8] Neuts, M.F.(1981). Matrix - Geometric Solutions in Stochastic Models: An Algorithmic Approach. The Johns Hopkins University Press, Baltimore.

[9] Barlow, R.E. and Heidtmann, K.D. (1984). Computing k - out - of - nSystem Reliability, *IEEE Transactions on Reliability*; 30(4):322 – 323.

[10] Zhang, Y., Wu, W., and Tang, Y.(2017). Analysis of an k -out-of-n : G system with repairmanŠs single vacation and shut off rule, *Operations Research Perspectives*,30(4): 29 – 38.

[11] Ji-EunByun, Hee-MinNoh, Junho Song (2017). Reliability growth analysis of k-out-of-N systems using matrix-based system reliability method, *Reliability Engineering and System Safety*, 165: 410 – 421.

[12] Mahsa Aghaei, Ali Zeinal Hamadani, Mostafa Abouei Ardakan (2017). Redundancy allocation problem for k-out-of-n systems with a choice of redundancy strategies, *Journal of Industrial Engineering International*, 13: 81 – 92

[13] Karlin S., Taylor H. E. (1975) *A first course in Stochastic Processes*, $2^n d$ ed., Academic Press, New York.